

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données



Projet de recherche réalisé par :

Marion FAVRE

Héloïse HAHLING

Sous la direction de :

Pierre-Yves BURGI, Directeur du projet DLCM

Genève, le 17 janvier 2020

Master en Sciences de l'information

Haute École de Gestion de Genève (HEG-GE)

Remerciements

Nous tenons à remercier tout particulièrement les personnes qui nous ont apporté une aide précieuse et sans qui le projet n'aurait pas pu être mené à bien :

- Messieurs Pierre-Yves Burgi et Hugues Cazeaux, qui nous ont accompagnées tout au long de ce travail et qui se sont montrés disponibles. En tant que mandants, ils nous ont permis de nous plonger dans les problématiques liées à la gestion des données de recherche, domaine qui nous n'était pas familier, mais pour lequel nous manifestons un intérêt tout particulier.
- Monsieur René Schneider, qui nous a conseillées et orientées vers des outils qui ont facilité la réalisation de notre travail.
- Madame Caroline Wilhem, qui nous a aidées dans l'organisation des entretiens avec les chercheurs de l'Université de Genève et qui nous a donné des compléments d'information essentiels dans la réalisation de notre travail.
- Madame Eliane Blumer, qui nous a mises en contact avec des chercheurs de l'EPFL, dans le but de mener nos entretiens.
- Monsieur Andreas Fürholz, qui nous a prêté une aide indispensable dans l'organisation et la réalisation des tests utilisateurs avec les chercheurs de la ZHAW.
- Toutes les chercheuses et tous les chercheurs de l'Université de Genève, de l'Université de Lausanne, de l'EPFL, ainsi que de la ZHAW, qui se sont montrés intéressés par notre problématique et qui ont bien voulu nous consacrer du temps pour la réalisation d'entretiens et de tests utilisateurs, sans quoi notre travail n'aurait pas pu aboutir.
- Les équipes travaillant pour les logiciels Axure et Loop11, qui nous ont accordé des licences gratuites pour réaliser ce projet.

Nous remercions aussi chaleureusement les proches qui nous ont soutenues, aidées et encouragées pendant cette période intense :

- Romain Bodinier et Thomas Dériaz, pour leur soutien et leurs innombrables conseils.
- Sarah Schupbach, pour ses conseils et ses relectures.
- Julien Raemy, pour ses conseils en termes de tests utilisateurs.
- Estelle Hug, Valérie Jacquod et Louise-Anne Thévoz, pour leur écoute et leur soutien.

Résumé

Le DLCM est un consortium suisse dont le but est d'aider les chercheurs dans la gestion de leurs données de recherche. Il est en voie de compléter sa gamme de services par une plateforme de préservation à long terme des données de recherche – Yareta pour Genève et Olos pour la Suisse – et nous a mandatées pour enquêter sur les besoins des utilisateurs d'une telle infrastructure, afin d'en gérer les aspects ergonomiques et esthétiques. Ce projet s'inscrit dans la perspective de l'ouverture des données de recherche demandée par le Fonds national suisse de la recherche scientifique (FNS) dès 2020.

Le projet a commencé par une phase de recherche, lors de laquelle nous avons étudié la littérature, des plateformes comparables et rencontré des chercheurs incarnant les utilisateurs potentiels du nouveau service du DLCM. Le principe de cette stratégie de recherche était de toujours placer l'utilisateur final au cœur de la réflexion pour ensuite concevoir un prototype dont le design serait centré sur l'utilisateur. Au cours de cette étape, nous avons notamment pris connaissance des fonctionnalités principales inhérentes à ce genre d'infrastructure, ainsi que des besoins, attentes et contraintes des chercheurs en termes de gestion et de partage de données de recherche.

Ensuite, le projet s'est poursuivi par une phase plus créative : nous avons utilisé les informations de la phase de recherche et les exigences formulées par le DLCM pour créer une première maquette de Yareta, sachant qu'Olos serait basé sur le même modèle.

Finalement, la création de maquettes et la réalisation de tests utilisateurs se sont succédé de manière itérative, de façon à obtenir des retours sur le prototype et à l'améliorer pour qu'il soit au plus proche des besoins et attentes des futurs utilisateurs.

Les tests utilisateurs sur les maquettes nous ont permis d'identifier des éléments satisfaisants à maintenir, des problèmes d'utilisabilité à résoudre, ainsi que des questions à transmettre au DLCM. Les problèmes d'utilisabilité résidaient essentiellement dans des formulations à modifier ou des fonctionnalités à ajouter. Les questions, quant à elles, portaient davantage sur le fonctionnement de l'infrastructure et les droits des usagers. Nous avons également constaté que les usagers ne lisaient les contenus que rapidement et qu'ils ne souhaitaient pas avoir à chercher des informations peu visibles au premier abord, ce qui nous a poussé à hiérarchiser les contenus à afficher. Lors du dernier test utilisateur, l'utilisabilité de la maquette a été jugée bonne, voire excellente, par l'ensemble des participants.

En guise de conclusion, ce travail nous a obligées à confronter nos idées et à les tester, ce qui nous a permis de fournir au DLCM une maquette finale dont l'utilisabilité satisfait les futurs utilisateurs potentiels. Celle-ci est accompagnée de recommandations formulées sur la base des remarques et interrogations des participants aux tests qui ne portaient pas directement sur la maquette en elle-même, mais sur des éléments plus proches de la stratégie du DLCM. Les remarques clés portant sur l'interface ont également été incluses dans ce document.

Table des matières

Remerciements	i
Résumé	ii
Liste des tableaux	vi
Liste des figures	vii
1. Introduction	1
2. Méthodologie	3
2.1 Phase de recherche	3
2.1.1 Travail déjà accompli par le DLCM	3
2.1.2 Etude de plateformes comparables	4
2.1.3 Revue de la littérature	4
2.1.4 Entretiens	4
2.1.5 Persona	5
2.2 Maquette et tests utilisateurs	5
2.2.1 Prototypage de Yareta	5
2.2.2 Tests utilisateurs et itération	6
2.3 Recommandations	9
3. Phase de recherche	11
3.1 Recommandations du DLCM et première version de Yareta	11
3.2 Etude de plateformes comparables	14
3.3 Revue de la littérature	14
3.3.1 Les données et leur archivage	15
3.3.2 Les chercheurs et leurs données de recherche	15
3.3.3 Les dépôts de données existants	18
3.3.4 Le design web centré sur l'utilisateur	21
3.3.5 Conclusion intermédiaire	23
3.4 Entretiens	24
3.5 Persona	25
4. Maquette et tests utilisateurs	26
4.1 Prototypage de Yareta	26
4.1.1 Analogie avec la méthodologie de Jesse James Garrett	26
4.1.2 Conception de la première maquette	26
4.2 Tests utilisateurs et itération	36

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

4.2.1	Retours	36
4.2.2	Maquette finale	38
4.2.3	Comparaison de l'utilisabilité des première et dernière maquettes.....	46
5.	Recommandations.....	48
6.	Difficultés rencontrées et améliorations possibles	50
7.	Conclusion.....	51
	Bibliographie	52
Annexe 1 :	Grille d'observation des plateformes existantes	58
Annexe 2 :	Formulaire de consentement pour la participation aux entretiens en français.....	63
Annexe 3 :	Guide d'entretien en français	64
Annexe 4 :	Entretien 1	67
Annexe 5 :	Entretien 2	77
Annexe 6 :	Entretien 3	85
Annexe 7 :	Entretien 4	91
Annexe 8 :	Entretien 5	100
Annexe 9 :	Persona.....	106
Annexe 10 :	Récapitulatif des exigences du DLCM nous concernant pour la plateforme	108
Annexe 11 :	Première maquette – Page de détail d'un jeu de données	109
Annexe 12 :	Première maquette - Première page du formulaire de dépôt ...	111
Annexe 13 :	Première maquette – Formulaire de création d'une unité organisationnelle.....	113
Annexe 14 :	Première maquette – Boîte de dialogue permettant de rejoindre une unité organisationnelle	114
Annexe 15 :	Première maquette – Page de détail d'une unité organisationnelle	115
Annexe 16 :	Formulaire de consentement pour la participation aux tests utilisateurs en anglais	116
Annexe 17 :	Guide pour la participation aux tests utilisateurs en anglais ..	117
Annexe 18 :	Guide de discussion des tests utilisateurs en anglais	120

Annexe 19 :	Tests utilisateurs – Consignes affichées sur le logiciel Loop11 (troisième test).....	122
Annexe 20 :	Compte rendu du premier test utilisateur	125
Annexe 21 :	Compte rendu du deuxième test utilisateur	138
Annexe 22 :	Compte rendu du troisième test utilisateur	154
Annexe 23 :	Liste des modifications apportées aux maquettes.....	164
Annexe 24 :	Troisième maquette – Page de gestion des jeux de données..	167
Annexe 25 :	Dernière maquette – Page de détail d'un jeu de données	168
Annexe 26 :	Dernière maquette – Première page du formulaire de dépôt ...	170
Annexe 27 :	Synthèse des recommandations générales	172
Annexe 28 :	Recommandations relatives aux métadonnées	193
Annexe 29 :	Questionnaire <i>System Usability Scale</i>	202

Liste des tableaux

Tableau 1: Eléments des recommandations du DLCM relatifs à notre projet.....	11
Tableau 2: Synthèse des retours obtenus lors des différentes séries de test	37

Liste des figures

Figure 1: Rubriques de la grille de lecture	4
Figure 2: Page d'accueil de Yareta	12
Figure 3: Page "Dépôts" de Yareta	13
Figure 4: Durée de conservation des données de recherche estimée par des chercheurs de l'Université de Lausanne (N=310)	16
Figure 5: Niveaux de la conception centrée sur l'utilisateur	22
Figure 6: Pyramide des besoins de Maslow adaptée aux interfaces web	23
Figure 7: Prototype papier	27
Figure 8: Prototype <i>Google Drawings</i>	27
Figure 9: Document de travail présentant la navigation entre les pages de la première maquette	27
Figure 10: Message d'avertissement affiché lorsque des fonctionnalités ne sont pas implémentées	28
Figure 11: Page d'accueil de la première maquette lorsque l'utilisateur est connecté	29
Figure 12: Bandeau de la première maquette lorsque l'utilisateur n'est pas connecté	30
Figure 13: Page d'inscription de la première maquette	31
Figure 14: Page de connexion de la première maquette	31
Figure 15: Page de résultats de recherche de jeux de données de la première maquette	31
Figure 16: Recherche avancée de la première maquette	32
Figure 17: Première page du formulaire de dépôt de jeux de données de la première maquette	33
Figure 18: Sélection des fichiers avant leur versement dans la première maquette	34
Figure 19: Boutons de la fin de la première page du formulaire de la première maquette	34
Figure 20: Page de gestion de ses propres jeux de données sur la première maquette	35
Figure 21: Focus sur les boutons d'action spécifiques à certains statuts des jeux de données sur la première maquette	36
Figure 22: Page de gestion des unités organisationnelles sur la première maquette	36
Figure 23: Page d'accueil de la dernière maquette	39
Figure 24: Page de recherche de jeux de données de la dernière maquette	40

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Figure 25: Page de détail d'un jeu de données dans la dernière maquette.....	41
Figure 26: Première page du formulaire de dépôt de jeux de données de la dernière maquette	42
Figure 27: Sélection des fichiers avant leur versement dans la dernière maquette	42
Figure 28: Boutons de la fin de la première page du formulaire de la dernière maquette.....	43
Figure 29: Page de gestion des jeux de données de la dernière maquette.....	44
Figure 30: Page de gestion des unités organisationnelles sur la dernière maquette	45
Figure 31: Page de détail d'une unité organisationnelle.....	46
Figure 32: Évaluation de l'utilisabilité de la première maquette (n=6)	47
Figure 33: Évaluation de l'utilisabilité de la troisième maquette (n=7)	47
Figure 34: Recommandations générales en chiffres	48

1. Introduction

Avec le développement de l'*Open Science*¹ et des principes FAIR² (*Findable, Accessible, Interoperable and Reusable*), le Fonds national suisse de la recherche scientifique (FNS) s'est positionné et souhaite que l'intégralité des résultats des recherches qu'il finance soit publié en *Open Access*³ d'ici à 2020. Selon lui, les résultats de la recherche financée par les fonds publics devraient être accessibles gratuitement, afin de pouvoir être réutilisés (Fonds national suisse 2019a) ; cela permettrait de satisfaire les critères de trouvabilité, d'accessibilité et de réutilisabilité mentionnés dans les principes FAIR.

Par ailleurs, l'Université de Genève a mis en place un Programme « Infrastructures et services numériques pour la recherche » pour soutenir l'évolution de la recherche en développant l'infrastructure numérique et en mettant en place de nouveaux services adaptés aux besoins des chercheurs (Université de Genève 2019a). En 2017, pour soutenir cette initiative, le Grand Conseil a adopté le projet de loi 12'146 qui accorde aux hautes écoles du canton un crédit d'investissement de plus de 15 millions CHF pour mettre en place les services inhérents au programme de l'UniGE (Grand Conseil de la République et Canton de Genève 2017).

Ces questions concernent directement le DLCM (*Data Life Cycle Management*), qui est un consortium suisse né dans le cadre de la numérisation croissante de la recherche et l'importance croissante accordée à la gestion du cycle de vie des données⁴ (DLCM 2019a). En effet, son but est d'aider les chercheurs à gérer leurs données de recherche et à satisfaire les exigences des agences de financement de la recherche, comme le FNS (Tan 2018). Au début du projet, le DLCM proposait notamment des formations en termes de gestion des données de recherche, un service de conseil personnalisé, des outils susceptibles d'aider à la mise en place de plan de gestion des données (*data management plan* ou DMP), ainsi que des outils accompagnant la mise en place de politiques de gestion des données (DLCM 2019a).

Dans ce contexte, le DLCM souhaite compléter sa gamme de services avec une infrastructure de préservation à long terme des données de recherche, qui s'apparente à une archive définitive, et non à un *repository*, qui est un dépôt de données intermédiaire, généralement utilisé dans une perspective d'*open access* (Burgi 2018, Schneider 2018c et HEG 2019). Au

¹ L'*Open science* est un « mouvement visant à rendre la recherche scientifique et les données qu'elle produit accessibles à tous et dans tous les niveaux de la société » (Wikipédia 2009).

² Dans le cadre de l'*Open science*, les principes du *FAIR data* évoquent « les manières de construire, stocker, présenter ou publier des données de manière à permettre que la donnée soit « trouvable, accessible, interoperable et réutilisable » » (Wikipédia 2017a).

³ L'*Open access* est la « mise à disposition en ligne de contenus numériques, qui peuvent eux-mêmes être soit libres, soit sous un des régimes de propriété intellectuelle » (Wikipédia 2006).

⁴ Le cycle de vie des données comprend leur création, leur traitement, leur analyse, leur préservation, leur mise à disposition, ainsi que leur réutilisation (Burgi 2018).

vu du double contexte suisse et genevois, le DLCM développera deux plateformes : Olos pour la Suisse et Yareta pour Genève.

Pour ce faire, le DLCM bénéficie d'une équipe de professionnels de l'Université de Genève capable d'assumer la dimension technique du projet⁵. En parallèle, nous avons été mandatées par Pierre-Yves Burgi, le Directeur du projet DLCM, pour enquêter sur les besoins des utilisateurs d'une telle plateforme et d'en gérer les aspects ergonomiques, comme esthétiques. Nous nous concentrerons sur Yareta, car la plateforme a été développée pendant que nous réalisons le projet ; de plus Olos devrait bénéficier de la même conception que Yareta⁶.

Le mandat est à la fois de natures théorique et pratique. En effet, d'une part, l'aspect théorique réside dans l'étude de la littérature et de l'existant en termes de plateforme web de dépôt de données de recherche. D'autre part, un aspect pratique domine dans la mesure où nous avons dû mener une étude de terrain en rencontrant des chercheurs afin de cibler leurs besoins. De plus, ce projet s'inscrit dans une démarche concrète de réalisation d'une plateforme de dépôt de données en collaboration directe avec Pierre-Yves Burgi, Directeur du projet DLCM, et Hugues Cazeaux, Responsable du pôle e-Research de l'Université de Genève⁷.

Ce rapport comporte deux grandes parties : nous commencerons par présenter notre méthodologie, en abordant les éléments dans un ordre chronologique ; puis nous nous concentrerons sur les résultats. Ces derniers sont, quant à eux, divisés en trois chapitres : le premier rassemble les résultats obtenus lors de la phase de recherche ; le second porte sur la réalisation de la première maquette du portail et son amélioration grâce aux tests utilisateurs ; et le dernier traite des recommandations d'implémentation de la plateformes émises à la fin du projet.

⁵ Entretien avec M. Pierre-Yves Burgi, Directeur du projet DLCM, et M. Hugues Cazeaux, Responsable du pôle e-Research de l'Université de Genève, Genève, le 8 mars 2019.

⁶ réf. 5

⁷ réf. 5

2. Méthodologie

Le mandat porte sur l'ergonomie ; aussi, nous avons adopté une méthode de travail destinée à mettre l'utilisateur au centre de notre recherche et à nous permettre de proposer un design d'interface capable de lui offrir la meilleure expérience de service possible. En effet, Sylvie Daumal souligne que la qualité des services fait la différence auprès du public et qu'elle poussera le public à privilégier un service plutôt qu'un autre (2018) ; or, la qualité d'un service est perçue à travers l'expérience des utilisateurs. Erika Hall appuie ces propos : « L'utilisabilité est le standard minimal absolu pour toute chose conçue pour être utilisée par des êtres humains. Si votre design ne permet pas aux utilisateurs ciblés d'utiliser le produit comme ils le souhaitent, ce design est un échec du point de vue du design orienté utilisateur » (2015).

Le projet a débuté par une phase de recherche, lors de laquelle nous avons utilisé des données existantes, à savoir les recommandations du DLCM, le schéma de métadonnées de DataCite, la première version de Yareta, les plateformes existantes et les informations trouvées dans la littérature. Nous avons aussi utilisé des données récoltées par nos soins auprès de chercheurs suisses, de manière à placer l'utilisateur final au centre de notre démarche et à pouvoir proposer un produit qui lui corresponde au mieux. La phase de recherche constitue une étape essentielle lors de l'utilisation de design centré sur l'utilisateur (*user-centered design*) (Hall 2015).

Ensuite, le projet s'est poursuivi par une phase plus créative : nous avons utilisé les informations de la phase de recherche pour créer une première maquette de Yareta. Puis, nous avons réalisé plusieurs séries de tests utilisateurs, de manière à adapter la maquette en fonction des retours et du comportement des utilisateurs.

Finalement, l'ensemble du travail accompli nous a permis de formuler des recommandations hiérarchisées pour que le DLCM puisse prendre en compte nos conseils selon une certaine priorité.

2.1 Phase de recherche

2.1.1 Travail déjà accompli par le DLCM

D'abord, nous avons pris connaissance du travail déjà accompli par le DLCM, ce qui comprend les recommandations formulées par rapport à la plateforme à concevoir (Cazeaux 2018) et l'état initial de la plateforme Yareta déjà développée (Université de Genève 2019b). Comme les métadonnées utilisées sur la plateforme pour les jeux de données doivent être conformes au schéma de métadonnées de DataCite^{9,10}, nous avons comparé les métadonnées mentionnées dans les recommandations du DLCM et celles du schéma de DataCite pour en

⁸ « *User-centered design* (UCD) est un principe de conception qui donne la priorité aux attentes des utilisateurs à chaque étape de développement d'un produit ou d'un service » (Garrett 2011).

⁹ DataCite est une organisation mondiale à but non lucratif qui fournit des identifiants permanents (DOI) pour les données de recherche et autres résultats de recherche (DataCite 2019a).

¹⁰ réf. 5

mesurer les similitudes dans le but de les implémenter dans notre maquette le moment venu (DataCite 2019b).

2.1.2 Etude de plateformes comparables

Ensuite, nous nous sommes intéressées aux plateformes de dépôt de données de recherche déjà existantes. Afin de les observer de façon cohérente et régulière, nous avons travaillé avec une grille d'observation (cf. Annexe 1) ; celle-ci suggérait d'observer notamment les institutions responsables, les modalités d'utilisation, les processus, les fonctionnalités ou encore l'ergonomie. Les plateformes Zenodo, Figshare, Dryad et Dataverse ont été examinées sur les conseils de MM. Pierre-Yves Burgi, Directeur du projet DLCM, et René Schneider, professeur HEG et Responsable du Master en Sciences de l'information.

2.1.3 Revue de la littérature

En ce qui concerne la revue de la littérature, nous nous sommes concentrées sur quatre axes de recherche définis avec Pierre-Yves Burgi :

1. Les dépôts de données
2. L'ergonomie de sites web
3. Les besoins des utilisateurs
4. Les retours d'expérience

Les bases de données spécialisées, la littérature professionnelle et le réseau de bibliothèques RERO ont été sollicités en priorité pour l'identification des sources intéressantes.

Afin de nous aider dans cette procédure, nous avons d'abord synthétisé les sources et les informations identifiées dans une grille de lecture, dont les rubriques étaient les suivantes :

Figure 1: Rubriques de la grille de lecture

Référence bibliographique	Date	Type (presse, académique, professionnel, norme, loi...)	Thématique	Sous-thèmes	Résumé	Intérêt pour notre recherche	Remarques
---------------------------	------	---	------------	-------------	--------	------------------------------	-----------

2.1.4 Entretiens

Après avoir étudié les informations existantes, nous avons choisi d'interroger le public cible de la plateforme, à savoir les chercheurs des hautes écoles et universités suisses. Le but était de poser des questions plus précises aux utilisateurs finaux, de manière à les mettre au centre de notre démarche et à proposer un produit qui leur corresponde au mieux. Aussi, nous avons rédigé un formulaire de consentement et un guide d'entretien en français et en anglais ; ces documents ont été validés par notre mandant et les versions françaises se trouvent en annexe de ce dossier (cf. Annexes 2 et 3).

Le guide d'entretien était structuré en cinq parties, de manière à aborder l'archivage des données de recherche, les données produites, le cadre imposé par les institutions, ainsi que les besoins et attentes des chercheurs vis-à-vis d'une plateforme de dépôt comme Yareta. Les formulations utilisées visaient à mener des entretiens semi-dirigés : d'une part, des questions fermées permettaient d'obtenir des réponses claires sur une problématique et, d'autre part, des questions ouvertes invitaient le répondant à raconter son histoire et ses expériences (Martin et Hanington 2013 et Daumal 2018).

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

En parallèle de la préparation du matériel, nous avons constitué un panel de chercheurs intéressés à aborder la gestion de leurs données avec nous. Pour ce faire, nous avons bénéficié de l'aide précieuse de Mme. Caroline Wilhem, conseillère à la recherche à l'Université de Genève, qui a bien voulu transmettre un message de présentation de notre projet dans une liste de diffusion rassemblant des chercheurs intéressés par la gestion des données de recherche. Mme. Eliane Blumer, bibliothécaire à l'EPFL, et M. René Schneider nous ont également transmis les coordonnées de personnes susceptibles de nous aider dans la constitution de notre panel. Dans une moindre mesure, nous avons également disposé des affichettes de présentation du projet sur des lieux de passage en espérant être contactées de manière spontanée par des chercheurs. Ainsi, nous avons constitué un panel de six chercheurs affiliés à l'Université de Genève, à l'Université de Lausanne et à l'EPFL. Les entretiens ont tous eu lieu en juillet 2019, dans les lieux de travail de nos répondants et ont été enregistrés pour faciliter leur transcription¹¹.

2.1.5 Persona

Lorsque l'on construit un produit ou un service, il convient d'en définir le public cible ; sans cela, le produit ou service risque fortement de ne pas convenir aux utilisateurs (Martin et Hanington 2013). Un persona est « un personnage imaginaire représentant un groupe ou segment cible dans le cadre du développement d'un nouveau produit ou service » (Bathelot 2018). Le persona rassemble les caractéristiques d'un groupe, un récit utilisateur et des détails fictifs, de manière à visualiser les besoins des utilisateurs et à faciliter la communication au sein de l'équipe de travail ; généralement, il se présente sous la forme de la fiche signalétique du profil d'un véritable individu (Hall 2015 et Daumal 2018).

Pour réaliser ce travail, il faut recueillir suffisamment d'informations pour décrire plusieurs utilisateurs, puis rechercher les comportements et des caractéristiques récurrents (Martin et Hanington 2013) ; c'est ce que nous avons pu faire, une fois la phase de recherche terminée.

2.2 Maquette et tests utilisateurs

2.2.1 Prototypage de Yareta

Une fois la phase de recherche terminée, le projet s'est poursuivi par une phase plus créative, lors de laquelle nous avons utilisé les informations de la phase de recherche pour créer une première maquette de la plateforme Yareta. Il nous semble important de souligner que, comme notre mandat consistait à accorder une attention particulière aux utilisateurs finaux, nous nous sommes concentrées sur les chercheurs et l'interface utilisateur de base ; aussi, nous n'avons pas travaillé sur l'interface des administrateurs de la plateforme. A cette étape, nous nous sommes focalisées sur les bonnes pratiques en termes de design centré sur l'utilisateur, les fonctionnalités principales implémentées dans les dépôts existants et les besoins, attentes et contraintes des chercheurs en termes de gestion et de partage de données de recherche. Le but était de concevoir un site qui corresponde aux besoins et aux attentes des usagers tout en étant attrayant et aisé à utiliser.

¹¹ Les enregistrements ont été supprimés pour garantir l'anonymat de nos répondants.

Nous avons commencé par dessiner plusieurs versions de la structure générale du site (emplacement des menus, des contenus, du logo, etc.) sur papier, en nous inspirant de la structure de la première version de Yareta déjà développée par l'équipe du DLCM (cf. Figure 2: Page d'accueil de Yareta) ; nous pensions qu'il serait plus utile au DLCM que nous utilisions leur travail comme base à améliorer. Rapidement, nous avons privilégié le logiciel gratuit *Google Drawings*, qui permet de créer des schémas vectoriels de manière collaborative ; le but était de modifier facilement les différentes pages, que ce soit en termes de disposition ou de contenu, et de visualiser simplement l'ensemble des parcours utilisateurs possibles (succession des pages en fonction des parcours décrits dans le personas) (Daumal 2018).

Les bonnes pratiques de design web suggèrent de réaliser des tris de cartes¹², de manière à soumettre l'architecture de l'information aux usagers finaux, mais nous ne l'avons pas fait : tout d'abord, Yareta n'est pas un site d'information, mais une plateforme de préservation à long terme de données de recherche qui n'a pas vocation d'organiser de l'information ; ensuite, elle disposait déjà d'une structure réalisée par le DLCM sur laquelle nous nous sommes basées ; finalement, la constitution du panel et l'organisation des entretiens ont été difficiles à réaliser et il nous a semblé peu judicieux d'organiser des rencontres supplémentaires au vu du temps imparti.

Après avoir mis au point une structure satisfaisante, nous avons utilisé le logiciel de prototypage payant Axure, pour lequel nous avons obtenu une licence RP9 Team gratuite pendant la durée du projet. Cela nous a permis d'implémenter certaines fonctionnalités et, ainsi, de rendre la maquette dynamique. L'objectif était que la maquette interactive puisse être testée sur le support final (Daumal 2018).

2.2.2 Tests utilisateurs et itération

Un test utilisateur est indispensable dans la mesure où les concepteurs du site en question ne sont pas représentatifs de la plupart des utilisateurs du web et que des participants externes au projet sont capables d'amener le recul qui manque souvent aux concepteurs (Krug 2012). Il vise à évaluer l'utilisabilité d'un site en s'intéressant à l'aisance des participants dans la réalisation de tâches, à leur compréhension des intitulés ou encore à la pertinence de l'information ; il s'agit donc d'identifier des problèmes dans le but de les corriger (Hall 2015 et Daumal 2018). Cette méthode a aussi été décrite sous le nom de *usability testing* (Martin et Hanington 2013).

2.2.2.1 Conception du test

Nous avons opté pour une itération de tests en présentiel, pour favoriser les échanges qualitatifs. Aucun test à distance n'a donc été prévu. Chaque test organisé en deux temps doit,

¹² « Le tri de cartes (*card sorting*) est une méthode très utilisée, notamment pour son apport dans l'architecture de l'information. Elle révèle la façon dont les utilisateurs regroupent naturellement les choses, ce qui permet de prendre des décisions sur les systèmes de navigation, les métadonnées, les relations (hiérarchiques, sémantiques...) entre les différents contenus » (Daumal 2018).

tout d'abord, permettre aux participants d'effectuer des tâches sur la maquette, puis d'échanger sur leur perception de la maquette et de l'exercice réalisé.

Une fois que la maquette a été suffisamment avancée pour tester les fonctionnalités principales de la plateforme, à savoir la recherche et la soumission de jeux de données, le premier test utilisateur a pu être paramétré sur le logiciel Loop11, pour lequel nous avons bénéficié d'une licence gratuite le temps de réaliser notre projet. Nous avons choisi cet outil, car il est conçu pour mener des tests à distance : ainsi, une fois lancé, le test ne nécessite plus l'intervention d'un modérateur, les consignes peuvent être paramétrées, l'écran des participants peut être enregistré et il génère des statistiques automatiques. De plus, il est possible d'effectuer le test avec la maquette Axure. Avant chaque série de tests, nous avons vérifié que tout était fonctionnel, de façon à identifier les problèmes d'affichage ou les éventuelles ambiguïtés.

Lors du premier test, trois tâches devaient être accomplies :

1. L'inscription sur la plateforme et la demande d'adhésion à une unité organisationnelle, dans le but d'évaluer la compréhension du fonctionnement des entités spécifiques à la plateforme ;
2. La recherche d'un jeu de données, car il s'agit d'une fonctionnalité clé de la plateforme ;
3. Le dépôt d'un jeu de données, car il s'agit également d'une fonctionnalité essentielle de la plateforme.

Les deuxième et troisième tâches correspondent aux deux expériences utilisateurs décrites dans le persona. Pour les tests suivants, une quatrième tâche a été ajoutée : la validation d'un jeu de données. Elle n'a pu être implémentée dans le premier test, car la maquette n'était pas suffisamment aboutie pour le permettre.

Les consignes des tests et les questions posées par l'intermédiaire de Loop11 sont disponibles en annexe (cf. Annexe 19). Après chacune des tâches accomplies par les participants, il leur serait demandé d'évaluer la difficulté de celles-ci et de commenter leur expérience.

Ensuite, en prévision des tests, nous avons rédigé un formulaire de consentement, afin que les participants nous autorisent à récolter des données et à les utiliser, un guide de test à distribuer aux participants et une liste de questions destinées à nous aider à mener une discussion semi-dirigée en deuxième partie de test. Tous ces documents sont disponibles en annexe (cf. Annexes 16, 17 et 18). Nous avons également créé une présentation PowerPoint dans le but d'exposer brièvement notre projet, la maquette et le déroulement du test aux participants ; la présentation incluait aussi les éventuels problèmes inhérents à la maquette et le rappel du fait que le test porterait sur la maquette et non leurs compétences (Daumal 2018). La finalité de cette introduction était aussi de familiariser les participants au visuel du logiciel, afin qu'ils ne se perdent pas dans la recherche des consignes et des boutons. Le support PowerPoint a été adapté au fil des séances en fonction des questions et problèmes soulevés par les participants.

2.2.2.2 Déroulement des tests

Nous avons eu la chance de collaborer avec des chercheurs de la Zürcher Hochschule für Angewandte Wissenschaften (ZHAW) à Zurich, pour procéder aux tests sur la maquette et recueillir des critiques et des remarques, dans l'idée d'améliorer le prototype en termes d'ergonomie et de design centré sur l'utilisateur. Nous avons également sollicité les membres du panel constitué pour les entretiens qui avaient manifesté leur intérêt pour les tests utilisateurs. Chaque test a mobilisé entre six et sept participants ; en effet, Krug estime que davantage de problèmes d'utilisabilité sont identifiés en testant une interface plusieurs fois avec peu de participants plutôt qu'en la testant une seule fois avec un grand nombre de participants (2012). De plus, il est conseillé de sélectionner quatre à huit participants par persona et nous n'en avons défini qu'un seul (Hall 2015). L'ensemble des vingt et un participants ont réalisé le test depuis leur propre ordinateur et, dans la mesure du possible – soit lors des tests réalisés à Genève –, les participants ont pu faire le test dans leur contexte de travail, c'est-à-dire dans leur bureau, ces conditions favorisant la détente des participants (Daumal 2018). Notons encore que les répondants n'ont participé chacun qu'à une session de test.

Chaque session a duré entre deux heures et deux heures et demi, discussion comprise, et s'est déroulée selon le schéma suivant :

1. Accueil des participants, présentation du projet et du test grâce au support PowerPoint et distribution des documents évoqués plus haut ;
2. Réalisation des tests individuels sur la maquette par chaque participant, alors que nous les observons silencieusement en prenant des notes ;
3. Pause et encas ;
4. Discussion¹³ ;
5. Remerciement et clôture de la session.

2.2.2.3 Analyse des résultats et adaptation de la maquette

Une fois les tests réalisés, nous avons créé une grille d'analyse qui, pour chaque tâche, proposait de lister les remarques des participants et d'associer en regard de chacune d'elles les modifications à apporter à la maquette avant le test suivant. Les résultats de tests utilisateurs se présentent sous la forme d'une liste de modifications à prévoir et leurs justifications (Krug 2012 et Hall 2015). La grille a été remplie selon quatre sources :

- Les observations faites pendant que les participants réalisaient les tests sur la maquette ;
- La discussion menée en groupe en deuxième partie de session ;
- Les enregistrements vidéo des écrans des participants permis par le logiciel Loop11 ;
- Les réponses et remarques saisies directement par les participants sur le logiciel Loop11 lors des sessions de tests.

¹³ Le guide de discussion n'a finalement pas été utilisé, car il s'est avéré plus intéressant d'aborder les tâches une à une, sous la forme d'un échange moins formel.

Une fois le compte rendu des tests rédigé, il était enfin possible d'adapter la maquette en fonction des différents types de retours obtenus. Les problèmes dont la solution n'avaient pas été discutée avec les participants ont été résolus en imaginant plusieurs prototypes papier, et en choisissant ensuite la solution à implémenter (*parallel prototyping* décrit dans Martin et Hanington 2013).

2.2.2.4 Itération

L'itération fait référence à l'action de répéter les tests, ce qui permet d'améliorer la maquette dans une phase cyclique. Le but est d'affiner le prototype et de « mettre à l'épreuve du feu ce qui se construit peu à peu » ; cela permet également de tester différentes variantes d'une solution (Daumal 2018). Ainsi, trois séries de tests ont été menées à Zurich et à Genève, entre le 12 septembre et le 8 octobre 2019, et quatre maquettes ont été conçues, la dernière n'ayant pas été testée auprès de futurs utilisateurs.

2.3 Recommandations

Une fois la dernière maquette finalisée et le matériel nécessaire réuni, nous avons pu rédiger des recommandations pour le DLCM relatives à l'implémentation de Yareta et, plus spécifiquement, à son ergonomie. Les recommandations ont été rédigées en deux parties : d'une part, les recommandations relatives aux métadonnées à associer aux jeux de données lors du processus de dépôt ; d'autre part, les recommandations générales. Les recommandations sont fondées en priorité sur les entretiens et les tests utilisateurs.

D'abord, les recommandations relatives aux métadonnées ont été organisées sous forme de comparaison entre les recommandations du DLCM en termes de métadonnée et le schéma de métadonnées de DataCite (cf. Travail déjà accompli par le DLCM) Un code couleur a été retenu pour traduire visuellement le statut obligatoire, recommandé et optionnel des métadonnées.

Les recommandations générales ont aussi été organisées sous forme de tableau. Dans un premier temps, nous avons fusionné les informations issues des entretiens et des tests utilisateurs dans un tableau pour les compter et rassembler les remarques semblables. Ensuite, les recommandations ont été organisées selon les thématiques suivantes :

- Menus
- Page d'accueil
- Inscription et gestion du profil
- Unités organisationnelles
- Recherche et téléchargement
- Dépôt de jeux de données
- Gestion des jeux de données

Finalement, à l'intérieur des thématiques, les recommandations ont été classées selon trois niveaux d'importance, de manière à hiérarchiser la lecture du DLCM (Hall 2015) :

- Faible : problèmes à faible impact qui affectent peu d'utilisateurs.
- Moyenne : problèmes modérés de faible fréquence ou problèmes mineurs de fréquence modérée.
- Haute : problèmes à impact élevé qui entravent souvent l'utilisateur et dont la non résolution représente un risque pour le succès du produit.

3. Phase de recherche

3.1 Recommandations du DLCM et première version de Yareta

Les recommandations du DLCM portaient sur la plateforme de dépôt de données de recherche nationale « Olos ». Cependant, vu qu'Olos n'était pas encore en développement, que Yareta aurait la même esthétique et que nous travaillons dans le canton de Genève, nous avons pris connaissance de ces éléments dans l'intention de les inclure dans notre maquette de Yareta. Nous nous sommes concentrées sur les éléments portant sur l'ergonomie, les fonctionnalités et l'esthétisme.

Tableau 1: Eléments des recommandations du DLCM relatifs à notre projet

Esthétisme	Fonctionnalité	Ergonomie
<ul style="list-style-type: none">Charte graphique	<ul style="list-style-type: none">Gestion des jeux de donnéesDépôt de jeux de donnéesRecherche de jeux de donnéesGestion des unités organisationnelles¹⁴.Rôles utilisateursMétadonnées à associer aux jeux de données déposés	<ul style="list-style-type: none">Eléments d'architecture du site

(Cazeaux 2018)

La comparaison des métadonnées figurant dans les recommandations du DLCM et dans le schéma de métadonnées de DataCite est disponible en annexe (cf. Annexe 28). Ce document a évolué en cours de projet, de manière à pouvoir formuler des recommandations spécifiques aux métadonnées au DLCM. Ce comparatif nous a permis d'identifier trois types d'informations :

- Les différences de formulation dans les intitulés des métadonnées ;
- Les différences de niveau de détail de certaines métadonnées, car le schéma de DataCite permet une description de haute précision, ce qui n'est pas toujours nécessaire ;

¹⁴ « Une unité organisationnelle est un concept logique pour exprimer une organisation. En général, elle peut représenter un laboratoire de recherche, un projet de recherche ou un espace de travail collaboratif partagé entre chercheurs » (traduit de Cazeaux 2018).

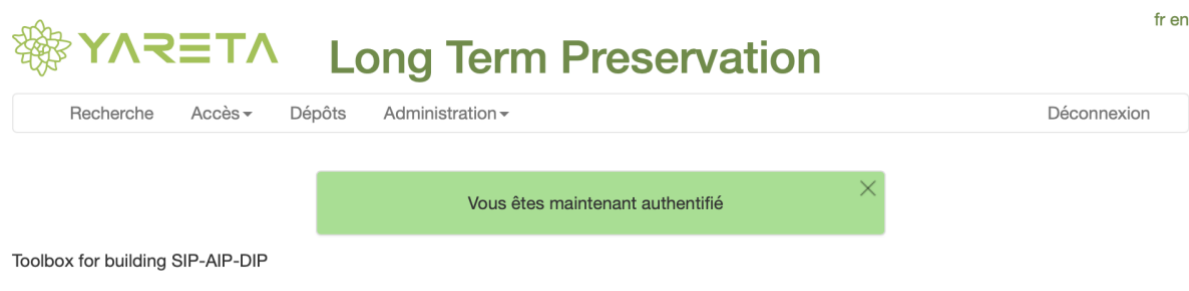
- Des lacunes de descriptions, dans la mesure où certains aspects de la recherche n'étaient pas décrits, alors qu'ils sont importants pour la compréhension des données ou de leur contexte de collecte.

Les métadonnées ont été implémentées dans nos maquettes, de façon à tester le processus de dépôt de données et à demander l'avis des futurs usagers lors des tests utilisateurs (cf. Retours). La hiérarchie des métadonnées a été maintenue dans le but de ne pas décourager l'utilisateur lors du processus de dépôt. Ainsi, les métadonnées recommandées ont également été adaptées en fonction des besoins spécifiques des futurs utilisateurs de la plateforme.

La première version de Yareta nous a permis d'observer les fonctionnalités et les éléments de contenus prévus par le DLCM sur sa plateforme de dépôt de données de recherche, ainsi que leur organisation. Au cours de cette étape, nous avons remarqué plusieurs éléments sur lesquels il nous faudrait réfléchir, même si nous étions conscientes que la version que nous examinions n'était pas la version finale du portail.

D'abord, de manière générale, la page d'accueil ne donnait pas suffisamment d'informations aux usagers sur la plateforme et son fonctionnement, ce qui impliquait le risque de perdre des usagers. De plus, certaines terminologies étaient ambiguës et susceptibles d'induire de la confusion pour les utilisateurs. Par exemple, dans le bandeau de navigation, l'utilisateur ignore ce à quoi le menu « Accès » donne justement accès ; l'utilisateur peut également se demander si le menu « Dépôts » permet de déposer un jeu de données ou s'il permet de visualiser l'ensemble de ses dépôts.

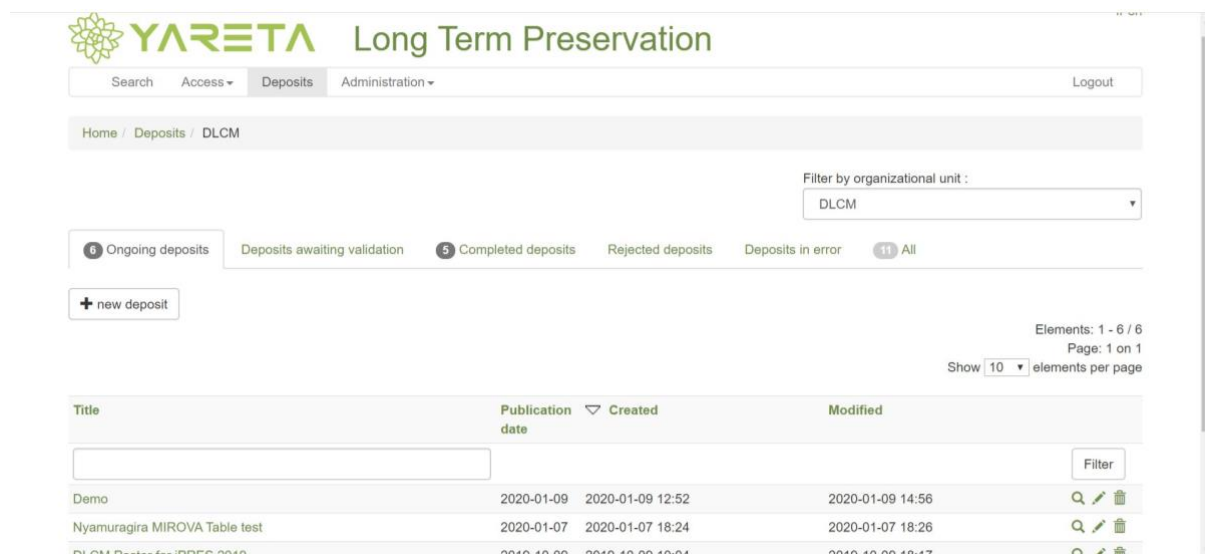
Figure 2: Page d'accueil de Yareta



(Université de Genève 2019b)

En réalité, il se trouvait que le menu « Dépôts » donnait accès aux deux hypothèses avancées ci-dessus :

Figure 3: Page « Dépôts » de Yareta



(Université de Genève 2019b)

En effet, d'une part, l'utilisateur a accès aux jeux de données déposés par ses soins, mais aussi par les unités organisationnelles¹⁵ auxquelles il est affilié ; ce rassemblement peut prêter à confusion. D'autre part, l'utilisateur a accès au formulaire de dépôt de jeux de données qui se présente en trois étapes ; l'ajout des métadonnées, le dépôt des fichiers et la validation du dépôt. Les métadonnées nous semblent claires et détaillées, et les étapes de validations (cf. Figure 3: Page « Dépôts » de Yareta) logiques et faciles à intégrer. Si le processus est complet et bien construit, il nous semble un peu long et limité en termes d'explications.

Ensuite, la page dédiée à la recherche de jeux de données déjà déposés sur la plateforme était basique et ne permettait ni recherche avancée, ni filtre des jeux de données retournés lors de la recherche.

¹⁵ Au sein d'une unité organisationnelle, des rôles sont attribués de manière à ce que les membres d'une unité organisationnelle n'aient pas tous les mêmes droits (DLCM 2019b).

3.2 Etude de plateformes comparables

Si les domaines couverts, les volumes déposés autorisés, les publics visés et les prix varient, nous avons remarqué que tous les formats de données sont autorisés, que la plupart des plateformes sont en *open source*¹⁶ et qu'elles ont tendance à recommander l'utilisation des licences *Creative Commons*¹⁷.

De manière générale, cette étape nous a permis de réaliser que certaines dispositions sont plus engageantes que d'autres et que certaines fonctionnalités sont susceptibles de faciliter l'utilisation de la plateforme à ses usagers. En effet, nous avons accordé une attention particulière à l'observation des pages d'accueil, des bandeaux de navigation ou encore des fonctions de recherche, de manière à nous inspirer des éléments facilitateurs et à éviter les complications.

Par ailleurs, l'expérimentation de différents processus semblables nous a aussi permis de reconnaître les éléments facilitateurs des démarches correspondantes. Par exemple, nous avons déposé des jeux de données sur toutes les plateformes observées, de manière à tester les différents processus de dépôt en place. Nous avons ainsi imité Zenodo en proposant de permettre aux usagers de glisser leurs données directement de leur gestionnaire de fichiers à la plateforme (fonctionnalité dite de *drag & drop*).

Ce travail a été décisif pour la mise en place de la première maquette de Yareta. La grille d'observation est disponible en annexe ; cependant, il s'agit d'un document de travail qui ne contient que les informations utiles à notre réflexion (cf. Annexe 1). Elle fait mention de critères qui ne sont pas liés à l'ergonomie – comme les prix, par exemple –, mais ceux-ci sont présents à titre indicatif, car nous avons préféré nous concentrer sur l'ergonomie, comme le prévoyait notre mandat.

3.3 Revue de la littérature

Comme présenté en introduction, le FNS souhaite que l'intégralité des résultats des recherches qu'il finance soit publié en *Open Access* d'ici à 2020. Ainsi, les résultats de la recherche financée par les fonds publics seront accessibles gratuitement, et pourront être réutilisés, dans l'idée de satisfaire les critères de trouvabilité, d'accessibilité et de réutilisabilité mentionnés dans les principes FAIR (Fonds national suisse 2019a). Dans cette même logique,

¹⁶ « Un logiciel *Open Source* est un programme informatique dont le code source est distribué sous une licence permettant à quiconque de lire, modifier ou redistribuer ce logiciel. Il se distingue cependant du logiciel libre, au sens où l'*Open Source* est davantage tourné vers un objectif de développement collaboratif que vers la défense des libertés individuelles » (Dabi-Schwebel 2018).

¹⁷ « Creative Commons est une organisation à but non lucratif, dont le but est de faciliter la diffusion et le partage des œuvres, tout en accompagnant les nouvelles pratiques de création à l'ère du numérique. Le but [...] était donc de développer des droits de propriété intellectuelle plus souples. Les licences Creative Commons ont été créées en partant du principe que la propriété intellectuelle était fondamentalement différente de la propriété physique, et du constat selon lequel les lois actuelles sur le copyright étaient un frein à la diffusion de la culture » (Wikipédia 2004).

le FNS impose aux chercheurs qui souhaitent obtenir des subventions de prévoir un *Data Management Plan*¹⁸ (DMP) et leur suggère de déposer les données qu'ils produisent dans des bases de données numériques en *Open Access* (Fonds national suisse 2019b).

Ces questions concernent directement le DLCM qui, dans le cadre du développement de son nouveau service, est amené à poursuivre ses interrogations sur les données et leurs enjeux, et à s'intéresser plus particulièrement aux besoins des utilisateurs, aux dépôts existants, ainsi qu'aux aspects ergonomiques que la plateforme devra prendre en considération.

3.3.1 Les données et leur archivage

L'Organisation de coopération et de développement économiques (OCDE) définit les données de recherche comme étant « des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche » (2007). Et, comme le souligne C. Borgman, ces données comprennent également les formats et les bases de données qui doivent être lus avec des ordinateurs ou des logiciels spécifiques (2012).

La définition de l'OCDE précise que les données sous-tendent la science, dont la finalité est d'appuyer un raisonnement ; pour que celui-ci puisse être validé, il doit être accessible et partagé (Milhit 2012). Il est acquis que les articles scientifiques doivent être publiés, mais un changement s'installe progressivement : le partage des documents complémentaires à la recherche commence à être demandé par les organismes de financement et les éditeurs, si bien que, depuis 2008, le nombre de citations de jeux de données tend à augmenter (Borgman 2012 et Kraker, Lex & Gorraiz 2015).

Les enjeux liés au partage des données de recherche prennent une place croissante dans les discussions, car, avec l'apparition du *Big Data* et le développement des technologies qui permettent de collecter ou de générer d'impressionnants volumes de données, le partage se justifie notamment pour des raisons d'efficacité, de puissance et de rigueur (Thelwall & Kousha 2015).

3.3.2 Les chercheurs et leurs données de recherche

Les chercheurs sont les premiers concernés par l'archivage et le partage des données de recherche, car c'est à eux de préparer leurs données et d'entreprendre les démarches inhérentes à ces initiatives. De plus, ils incarnent le public cible de la future plateforme de dépôt de données de recherche à long terme du DLCM. Il convient donc de s'intéresser à leur perception avant concevoir la plateforme. Cette partie se fonde essentiellement sur deux enquêtes qualitatives cohérentes menées auprès des chercheurs : l'une réalisée à l'Université de Lausanne par Carmen Jambé (2015), l'autre réalisée à l'Université de Rennes 2 par

¹⁸ Un data management plan (ou plan de gestion des données) est un « document évolutif qui aide le(s) chercheur(s) ou le chargé de projet de la recherche à définir un plan pour gérer les données utilisées et générées dans le cadre de son activité ou de son projet de recherche » (Wikipédia 2017b).

Alexandre Serres (2017). Des constats d'autres scientifiques viendront compléter et appuyer les deux chercheurs susmentionnés.

3.3.2.1 Les données avant l'archivage

Dans les études de C. Jambé et A. Serres, les chercheurs interrogés mentionnent essentiellement trois dispositifs de stockage de leurs données : l'ordinateur de l'université, un disque dur externe et leur ordinateur personnel (Jambé 2015 et Serres 2017). A Lausanne, le dépôt de données n'arrivait qu'en douzième position et la moitié des répondants n'associaient aucune métadonnée à leurs données. Nous verrons plus tard que ce manque de documentation initial freine l'archivage des données de recherche. En ce qui concerne les volumes ainsi stockés, seuls 9% s'élèvent à plus d'un téraoctet, alors que la plupart (29%) se situent entre 1 et 50 gigaoctets ; relevons que 24% des chercheurs ne savent pas estimer le volume de leurs données (Jambé 2015). Ces valeurs sont cohérentes avec celles collectées à Rennes, même si Alexandre Serres rappelle qu'il est difficile de prévoir les véritables besoins des chercheurs en termes d'espace de stockage, car leurs pratiques sont trop différentes les unes des autres (2017).

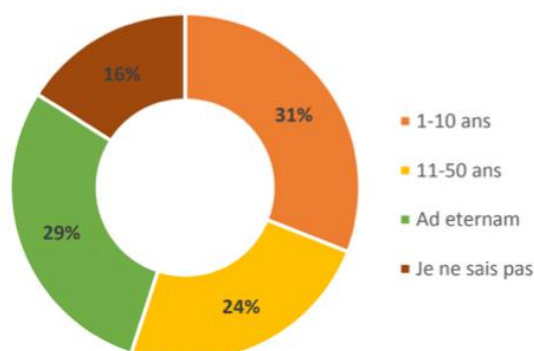
3.3.2.2 L'archivage des données

En Europe, le programme européen pour la recherche et le développement couvrant la période entre 2014 et 2020 s'appelle Horizon 2020 et vise à développer les initiatives *Open access* et *Open science* (Commission européenne 2019). Pourtant, à Rennes, même si les chercheurs sont invités à archiver leurs données de recherche de manière pérenne, l'archivage ne semblait pas encore être un sujet de préoccupation en 2017, puisque 93% des interrogés disent ne pas avoir de politique en la matière au niveau de leur laboratoire, ou l'ignorent (Serres 2017).

Dans les faits, 24% des chercheurs lausannois avaient déjà ou avaient l'intention de déposer leurs données sur une plateforme à cet effet et il en va de même à Rennes (Jambé 2015 et Serres 2017). Notons cependant que près de la moitié ignorerait encore cette possibilité (Jambé 2015).

Les chercheurs ont des avis disparates quant à la durée idéale de l'archivage de leurs données.

Figure 4: Durée de conservation des données de recherche estimée par des chercheurs de l'Université de Lausanne (N=310)



(Jambé 2015, p. 340)

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

3.3.2.3 Le partage et la réutilisation des données

3.3.2.3.1 Facteurs incitatifs

En premier lieu, il convient d'observer les éléments susceptibles de contraindre les chercheurs à archiver leurs données de recherche. En Suisse, la loi fédérale sur l'archivage stipule que « [t]ous les documents de la Confédération qui ont une valeur juridique, politique, économique, historique, sociale ou culturelle » doivent être conservés et archivés (LAr ; RS 152.1), ce qui inclut les données produites dans le cadre de recherches subventionnées par le FNS (Milhit 2012). Par ailleurs, la pression des éditeurs et des organes de financement peut aussi pousser les chercheurs à archiver leurs données.

Ensuite, le premier avantage offert par l'archivage est qu'il permet la préservation à long terme des données (Jambé 2015). La sécurité des données associée à l'archivage est également un atout (Jambé 2015). En effet, souvent, l'archivage des données de recherche se fait sur une plateforme de dépôt et ce genre de dispositif garantit la sécurité des données de trois façons : d'abord, ces plateformes suivent les meilleures pratiques en termes de sécurité des données ; puis, elles conçoivent des politiques d'archivage et de partage qui guident les chercheurs ; enfin, elles prévoient des politiques d'accès et de réutilisation des données à l'attention des utilisateurs. Ces éléments sont susceptibles de rassurer les dépositaires soucieux de l'avenir de leur travail, même en cas de partage et de réutilisation (Dubois, Strait & Walsh 2018). En termes de sécurité, l'archivage permet également de gérer les droits d'accès aux données (Jambé 2015).

Par ailleurs, les DMP demandés par certains organes de financement et la capacité des chercheurs à présenter leurs données et leur gestion influencent les montants accordés (Borgman 2012). L'archivage peut donc augmenter les subventions aux projets de recherche.

Un dernier aspect peut entrer en ligne de compte. Il semblerait qu'une certaine pression sociale qui considère l'archivage des données de recherche comme étant une bonne pratique peut motiver les chercheurs à entreprendre cette démarche (Kim & Adler 2015).

3.3.2.3.2 Facteurs dissuasifs

D'une part, les chercheurs voient surtout que le temps et le financement alloués aux tâches inhérentes à l'archivage – à savoir l'organisation, le tri et la documentation des données –, qui est souvent considéré comme une opération « secondaire », ne sont pas dédiés aux opérations qui alimentent leur carrière, comme la collecte de données, le traitement, l'analyse, les conférences, les publications, etc. (Borgman 2012 et Serres 2017). De plus, ce travail supplémentaire n'est pas valorisé et le chercheur n'est ni récompensé, ni reconnu pour cela (Borgman 2012). Notons que, lorsque l'archivage des données sera devenu une pratique courante, il sera plus facile pour les chercheurs de prendre ces tâches en considération lors de l'élaboration de leur planning et de leur budget (Dubois, Strait & Walsh 2018). En attendant, pour compenser ces freins, les chercheurs auraient besoin que le processus de dépôt soit simple et rapide (ex : mise en place de règles de nommage des fichiers), que l'interface de dépôt soit *user-friendly* et que la démarche de gestion des données soit soutenue et valorisée au sein des institutions dont sont issus les scientifiques (Serres 2017).

D'autre part, la conservation des données de recherche sur le long terme est coûteuse. Ainsi, toutes les données générées par la recherche scientifique ne peuvent être retenues pour l'archivage à long terme et leur sélection soulève des questions auxquelles les chercheurs ne peuvent pas toujours répondre seuls (Borgman 2012).

3.3.2.4 Besoins à satisfaire lors de l'archivage de données de recherche

Finalement, les enquêtes de Carmen Jambé (2015) et d'Alexandre Serres (2017) identifient sept besoins des utilisateurs qui doivent être satisfaits lors de l'archivage de données de recherche sur un dépôt de données de recherche : la sécurité des données doit être garantie, les données doivent être accessibles, partagées et diffusées, le volume de stockage doit être suffisant, l'interface doit être *user-friendly*, et la démarche de dépôt doit être soutenue par les institutions et valorisée. Par ailleurs, il semblerait qu'une liste de métadonnées à fournir lors de l'archivage et qu'une règle de nommage des fichiers ne soient pas des priorités pour les chercheurs (Serres 2017). Le besoin des chercheurs de bénéficier d'une aide à la gestion des données de recherche pour leur partage et leur archivage transparaît clairement à travers ces deux études, mais pas seulement. En effet, d'autres études sur les chercheurs ont été identifiées, comme celles d'Ethan P. White (2013) ou de Carol Tenopir (2011), mais les arguments se recoupaient et aucun ne donnait autant de chiffre que Carmen Jambé (2015) et Alexandre Serres (2017). De plus, les deux études choisies donnaient des chiffres cohérents et se corroboraient.

3.3.3 Les dépôts de données existants

La plupart des articles identifiés abordaient des aspects très techniques de la construction des dépôts de données – comme les profils d'application pour métadonnées décrit par Edward Krause (2015) –, restaient très superficiels ou n'étaient plus d'actualité. Toutefois, au vu de l'orientation de notre mandat, nous nous sommes concentrées sur les sources axées sur les fonctionnalités.

Claire Austin a étudié plus de trente dépôts de données de recherche canadiens et internationaux et a ainsi remarqué que la plupart des plateformes montraient une forte affiliation à un gouvernement ou à une université, ce qui s'explique notamment par la nature des organes de financement (2016). Elle constate aussi que les dépôts détiennent des données multidisciplinaires, même si la majorité d'entre eux sont orientés vers un domaine particulier.

Comme les dépôts de données apparaissent dans un contexte d'*Open science* et dans une optique de partage, la plupart des données sont en *Open access* (Peters et al. 2017).

3.3.3.1 Les fonctionnalités des dépôts

Même si les affiliations et les domaines des dépôts de données varient, les fonctionnalités prévues sont souvent semblables.

3.3.3.1.1 La préservation à long terme

La préservation des jeux de données sur le long terme est une fonctionnalité clé des dépôts, avec la recherche et l'extraction de données (Kim 2018). En 2016 déjà, 56% des dépôts étudiés par Claire Austin annonçaient offrir cette fonctionnalité avec une politique et des

pratiques formalisées et 44% avaient mis en place une redondance des fichiers pour une meilleure sécurité de préservation (2016).

3.3.3.1.2 *La gestion des métadonnées*

Les métadonnées jouent un rôle primordial dans les dépôts puisqu'elles permettent de décrire les données pour faciliter leur compréhension et peuvent être utilisées pour mettre en place des fonctionnalités telles que la recherche de jeux de données, l'interopérabilité des données, etc. (Crosas 2011). Pourtant, le contrôle de leur qualité est une opération compliquée et la plupart des dépôts n'abordent pas cet aspect (Austin et al. 2016). On considère généralement que les métadonnées sont de bonne qualité si elles donnent les informations bibliographiques essentielles de découverte, d'utilisation, de provenance, d'actualité, d'authenticité et d'administration, tout en respectant des normes et en permettant l'interopérabilité (Park 2009). Malgré cela, on constate que les métadonnées sont disparates : les schémas utilisés sont soit propres au dépôt, soit issus d'un standard, comme le Dublin Core ou le DataCite, par exemple (Austin et al. 2016).

Il est opportun de fournir des aides et des explications quant à la saisie des métadonnées, car elles sont souvent ajoutées par les déposants eux-mêmes qui ne savent souvent pas comment s'y prendre (Greenberg et al. 2009 et Austin et al. 2016). Si plus des deux tiers des plateformes étudiées par C. Austin fournissent une aide à la saisie des métadonnées, leur qualité n'est généralement pas vérifiée (2016). En effet, Dimitris Rousidis a identifié les problèmes récurrents et suggère par conséquent que les champs libres soient décrits de façon à éviter les erreurs de saisie, que les noms de personnes soient normalisés et contrôlés, et que les formats de dates soient imposés (2014). C'est l'une des raisons pour lesquelles il est conseillé d'utiliser des systèmes de saisie automatique pour certaines métadonnées, comme les dates, de façon à éviter les erreurs de format ; d'ailleurs, 72% des plateformes étudiées par C. Austin utilisent un système d'horodatage des fichiers (2016).

Les jeux de données sont surtout valorisés lorsqu'ils sont utilisés et cités, et la citation est facilitée lorsque le jeu de données est associé à un identifiant unique pérenne (Crosas 2011). C'était en général le cas pour les jeux de données archivés sur les dépôts étudiés par C. Austin : près de la moitié des plateformes avaient recours au DOI (Austin et al. 2016 et Thelwall & Kousha 2015). En outre, 63% des dépôts étudiés par C. Austin génèrent automatiquement des citations pour favoriser la réutilisation de jeux de données (2016).

Figshare propose une utilisation supplémentaire des identifiants pérennes : il utilise les identifiants ORCID (*Open Researcher and Contributor ID*) des chercheurs pour permettre des liens entre les différents travaux et faciliter l'identification des auteurs des jeux de données (Reed 2017). Cette fonctionnalité existe aussi sur d'autres dépôts.

Dans le cadre de l'*Open science* et des principes FAIR, l'interopérabilité est souhaitable, puisqu'elle facilite les échanges et le partage des données (Crosas 2011). Elle se réalise notamment par l'utilisation de standards de métadonnées. D'ailleurs, en utilisant le standard de métadonnée Dublin Core, le dépôt Dryad facilite la collecte des métadonnées, la recherche de jeux de données, ainsi que l'échange de métadonnées avec d'autres formats (Greenberg et al. 2009).

3.3.3.1.3 Les licences et le contrôle des accès

Trois quarts des plateformes étudiées par C. Austin ont recours à des licences ouvertes du type Creative Commons pour les jeux de données et 84% des plateformes permettent de restreindre l'accès aux données ; 80% des dépôts imposent tout de même aux utilisateurs d'être identifiés pour y accéder (2016). En outre, 66% des dépôts interrogés disposent d'une politique sur le partage des données, les conditions d'utilisation et la propriété (Austin et al. 2016).

Dataverse prévoit trois niveaux d'accès aux données (Crosas 2011) :

1. Les jeux de données publics avec des conditions d'utilisation permettent de consulter librement les informations descriptives, mais l'utilisateur doit accepter des conditions d'utilisation propres à l'étude pour accéder aux fichiers.
2. Les jeux de données avec dossier(s) restreint(s) permettent toujours de consulter librement les informations descriptives, mais certains fichiers ne sont accessibles (consultation et téléchargement) qu'aux utilisateurs autorisés.
3. Les jeux de données restreints ne permettent pas une consultation complète des descriptions de l'étude.

Sur le dépôt Zenodo, 89% des jeux de données offrent un accès ouvert et seul 2,5% font mention d'un embargo ; ces proportions sont en accord avec la volonté de Zenodo de promouvoir le partage ouvert des résultats de recherche financés par la Commission européenne (Sicilia, Garcia-Barriocanal & Sanchez-Alonso 2017).

3.3.3.1.4 Les statistiques

La moitié des dépôts interrogés par C. Austin donnent accès aux statistiques d'utilisation des données de recherche, comme le nombre de téléchargements (2016). C'est notamment le cas de Figshare for Institution, une version de Figshare qui peut être achetée par une institution pour ne rassembler que les données de l'institution en question et qui propose des statistiques personnalisées (Reed 2017).

3.3.3.2 La mise en place d'un dépôt

Peu de documents sur la mise en place de dépôts de données de recherche ont été identifiés. Aux USA, l'Université de Carnegie Mellon s'est dotée d'une plateforme construite sur Figshare ; il s'agit d'un dépôt institutionnel prenant également en charge les données de recherche (Scherer, Zilinski & Valen 2018). Le projet est né suite à une demande d'archivage de données auprès de la bibliothèque de l'Université qui s'est rendu compte du manque d'infrastructures adaptées. Le retour d'expérience publié par David Scherer, Lisa Zilinski et Dan Valen (2018) aborde surtout la question des rôles à attribuer aux administrateurs de la plateforme et de leur intégration dans les différents processus de travail, à savoir celui de dépôt et celui de validation du dépôt. Les exigences formulées envers la plateforme sont destinées à éviter des lacunes fonctionnelles : certaines métadonnées doivent être soumises au moyen d'un fichier « README.txt » associé aux jeux de données et une règle de nommage des fichiers doit être prévue. Finalement, le document liste aussi les critères de facilité de dépôt : la soumission doit être rapide et facile, le délai d'exécution doit être court et un DOI doit être attribué aux données.

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

3.3.3.3 L'utilisation des données déposées

3.3.3.3.1 Zenodo

Pour faciliter la réutilisation des données et leur donner davantage de visibilité, les données déposées sur Zenodo sont indexées par d'autres bases de données, comme Web of Science (Peters et al. 2017).

Sur Zenodo, 54% des ressources sont des textes, 34% des logiciels et 12% des jeux de données ; la présence des logiciels sur le podium des typologies s'explique par le contexte d'*Open science* qui demande que les recherches puissent être reproduites et vérifiées, ce qui n'est parfois pas possible sans le logiciel adapté (Sicilia, Garcia-Barriocanal & Sanchez-Alonso 2017). Toutefois, les logiciels et les jeux de données contenus dans Zenodo n'ont que peu, voire pas été cités, mais ceux qui le sont l'ont été à plusieurs reprises (Peters et al. 2017).

3.3.3.3.2 Dryad

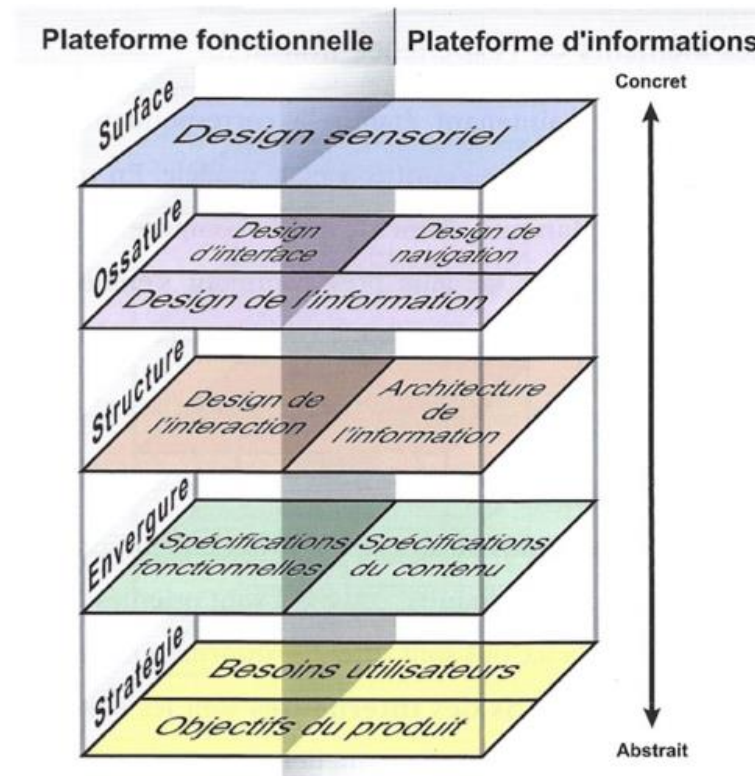
Le nombre de dépôts et de citations de données archivées sur Dryad varient en fonction des domaines d'étude, mais les citations restent largement inférieures aux données déposées (He & Nahar 2016). Par contre, He et Nahar constatent que plus de 70% des données déposées sur la plateforme sont citées dans un délai de deux ans après leur publication. En termes de domaine d'étude, l'agriculture, la biologie et la médecine obtiennent le plus de citations et, en termes de typologie de données, les données expérimentales – notamment relative à la méthodologie – rencontrent le plus de succès (2016).

3.3.4 Le design web centré sur l'utilisateur

La conception de la maquette, pour le DLCM, formalisera l'aboutissement de la recherche sur les chercheurs qui utiliseront la plateforme et l'observation des plateformes existantes. Afin d'améliorer son utilisabilité, nous avons mené des recherches sur le design de sites Internet centré sur l'utilisateur, de façon à ce que son expérience sur la plateforme soit optimale. Les besoins des utilisateurs abordés plus haut sont au cœur de la démarche.

Jesse James Garrett définit l'expérience utilisateur comme « l'expérience que confère le produit [ou service] à ceux qui l'utilisent dans la vraie vie [...] Elle ne concerne pas le fonctionnement interne d'un produit ou d'un service. Elle concerne le fonctionnement externe, celui auquel une personne est confrontée » (2011). Pour qu'un site Internet soit centré sur l'utilisateur, Garrett suggère une approche en cinq étapes successives qui placent toutes l'utilisateur au cœur de la réflexion (2011) :

Figure 5: Niveaux de la conception centrée sur l'utilisateur



(Garrett 2011)

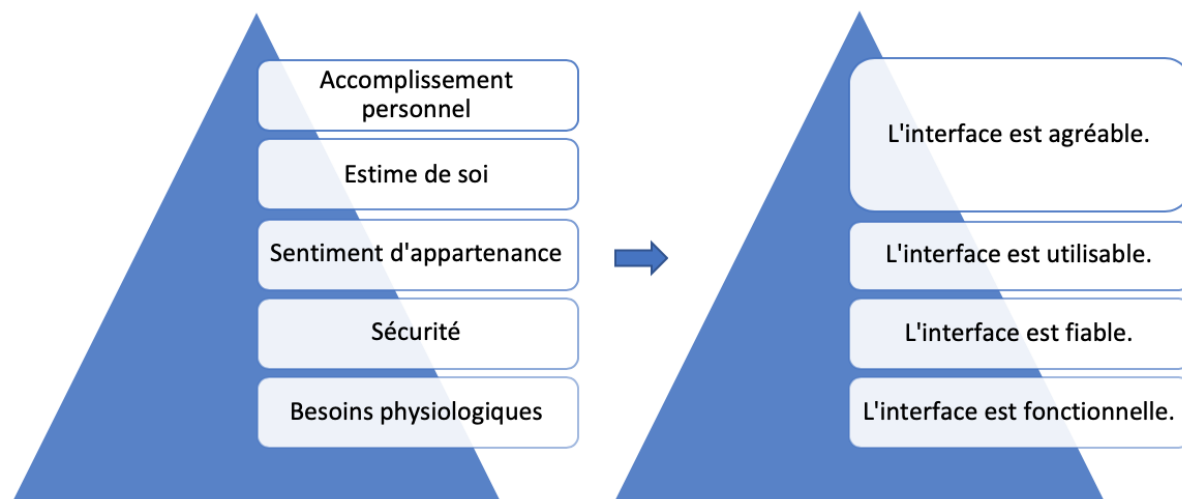
1. La stratégie pose les bases et permet de concilier les objectifs du site web avec les besoins des utilisateurs. Il s'agit d'apprendre à connaître les utilisateurs finaux pour pouvoir construire des personas ; cette étape suppose une phase de recherche, bien décrite par Sylvie Daumal, nécessaire à la conceptualisation de la stratégie (2018).
2. L'envergure fixe les fonctionnalités et les éléments de contenu que le site doit présenter pour atteindre la stratégie. Cette étape définit, en quelque sorte, le cahier des charges du site web.
3. La structure définit l'organisation conceptuelle des éléments pensés dans l'envergure. Le design d'interaction décrit le comportement de l'utilisateur sur la plateforme et la manière dont la plateforme lui répond, alors que l'architecture décrit l'organisation des informations.
4. L'ossature est la première étape de concrétisation. Il s'agit de prévoir la navigation du site, ainsi que la disposition des fonctions et du contenu sur les pages. Le design d'interface fait référence à la disposition des fonctionnalités (ex : boutons) ; le design de navigation s'apparente à la présentation des espaces alloués à l'information ; et le design d'information rassemble les deux éléments précédents pour présenter l'information de manière efficace.
5. La surface rassemble toutes les couches précédentes et leur donne un certain aspect visuel ; c'est à elle que sera confronté l'utilisateur lors de son expérience utilisateur. Il s'agit du design sensoriel, car il fait appel au toucher (emplacement des différents éléments), à l'ouïe et, surtout, à la vue.

Aaron Walter va plus loin en parlant de « design émotionnel », un courant de design « qui utilise la psychologie et l'artisanat pour donner l'impression qu'il y a une personne, et non une

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

machine, à l'autre bout de la ligne ». Il s'agit de faciliter l'interaction homme-machine et de considérer que les émotions positives suscitent chez l'utilisateur une confiance et un attrait pour le site (2011). Pour ce faire, il utilise la pyramide des besoins de Maslow pour identifier les besoins des utilisateurs et faire en sorte que l'interface web y réponde.

Figure 6: Pyramide des besoins de Maslow adaptée aux interfaces web



(Walter 2011)

En plus des besoins utilisateurs liés aux fonctionnalités offertes par l'interface, il faut prendre en compte le comportement de l'utilisateur et son utilisation du web. Trois règles clés se détachent des écrits en la matière. D'abord, l'internaute ne doit pas avoir besoin de chercher : tout doit être clair et intuitif, de façon à ce qu'il comprenne l'organisation et le fonctionnement du site Internet (Krug 2012) ; le visuel de la navigation joue un rôle clé dans ce domaine. Ensuite, le visuel doit permettre de hiérarchiser les contenus et de mettre en avant le principal (Krug 2012). Finalement, en ce qui concerne les contenus, ils doivent être précis, positifs et objectifs (Garrett 2011 et Krug 2012). Par exemple, la page d'accueil doit être simple, sans contenu superflu et attirer l'attention de l'utilisateur sur l'essentiel (Walter 2011 et Krug 2012).

La conception d'un site Internet évolue d'abord à travers la réalisation de prototypes soumis à de nombreux tests utilisateurs : il s'agit de vérifier la conformité de l'interface par rapport aux besoins des utilisateurs et de faire évoluer le prototype vers une version aboutie qui pourra être implémentée sur l'interface web (Daumal 2018).

3.3.5 Conclusion intermédiaire

Cette étude de la littérature a permis de réaliser que les chercheurs ne sont pas familiers du processus d'archivage de leurs données de recherche sur des dépôts à cet effet et que cette démarche est surtout perçue comme une contrainte gourmande en ressources et imposée par les organes de financement. Il convient donc de rendre les procédures aussi faciles et intuitives que possible, tout en leur fournissant l'assistance nécessaire.

L'enjeu principal des dépôts réside dans la définition et la gestion des métadonnées, car elles soutiennent non seulement la mise en place de fonctionnalités parallèles, mais aussi la description des jeux de données. Elles doivent être complètes mais ciblées, ouvertes mais

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

cadrées, pour que les chercheurs qui participent à leur saisie ne compromettent pas leur qualité.

Le design centré sur l'utilisateur se présente comme une boîte à outil obligeant le concepteur à garder l'internaute au cœur de sa réflexion. Il s'agit de ne faire appel qu'à l'intuition et aux sens de ce dernier : il ne doit rien chercher, tout doit être à sa portée pour rendre son expérience aussi fluide que possible.

Finalement, la gestion des données de recherche est un domaine interdisciplinaire qui rassemble des professionnels d'horizons différents (juristes, informaticiens, chercheurs, concepteurs d'interface web et spécialistes de l'information) qui doivent se comprendre, malgré les différences de langage et de culture professionnelle, afin de construire un service adapté aux besoins de ses utilisateurs.

3.4 Entretiens

Les entretiens nous ont permis de comprendre le contexte de travail des chercheurs suisses romands, leurs contraintes, leurs besoins, ainsi que leurs attentes en termes de gestion et de partage de données de recherche.

En premier lieu, la démarche d'archivage des données de recherche est une demande du FNS que les chercheurs perçoivent comme une contrainte demandeuse en ressources, puisque les tâches inhérentes à l'archivage ne sont pas familières aux chercheurs. Ainsi, elles leur prennent du temps – et de l'argent –, sans qu'elles ne soient reconnues ou valorisées dans leur carrière ; aussi, la démarche d'archivage doit être simple, claire et rapide. En plus, les chercheurs redoutent de publier leurs données rapidement, de peur de ne pas pouvoir les exploiter eux-mêmes, et que d'autres récoltent le fruit de leur travail de collecte de données. Pour prévenir ce problème, ils aimeraient que leurs données puissent être archivées de manière privée, ou sous embargo.

Ensuite, nos répondants se sont montrés curieux quant aux volumes et formats autorisés par Yareta. Ils nous ont aussi demandé de prévoir des fonctionnalités spécifiques, comme :

- Liens avec d'autres publications
- Téléchargements de lots de fichiers
- Visualisation des dépôts avant soumission
- Sauvegarde automatique des dépôts en cours
- Moteur de recherche
- Affichage des résultats de recherche sous forme de liste
- Mise à jour des jeux de données déjà publiés
- Valorisation des dépôts

Finalement, de manière générale, nous avons senti nos répondants préoccupés par le temps que leur prendra la démarche d'archivage de leurs jeux de données et les questions de propriété et de partage des données. Il sera donc important de prendre leurs remarques en considération et de mettre en place des aides destinées à faciliter l'utilisation de la plateforme.

Les transcriptions anonymisées des entretiens sont disponibles en annexe de ce document (cf. Annexes 4, 5, 6, 7 et 8).

3.5 Persona

Une fois la phase de recherche terminée, nous avons réalisé un seul persona focalisé sur les chercheurs. En effet, ces derniers incarnent les utilisateurs prioritaires de Yareta ; de plus, notre récolte d'information n'aurait pas permis de réaliser d'autres personas.

Notre persona synthétise les éléments identifiés au cours des entretiens, souvent confirmés dans la revue de la littérature. Il comprend également deux récits utilisateurs portant sur deux tâches réalisables grâce à Yareta ; pour chaque étape des récits – qui s'apparentent à des expériences utilisateurs¹⁹ –, les émotions du persona sont traduites par des émoticônes. Ceux qui ne sourient pas signalent les étapes ou les fonctionnalités susceptibles de nuire à l'expérience utilisateur : il faudra donc y prêter une attention particulière. Le persona se trouve en annexe (cf. Annexe 9).

¹⁹ « L'expérience utilisateur ne concerne pas le fonctionnement interne d'un produit ou d'un service. Elle concerne le fonctionnement externe, celui auquel une personne est confrontée » (Garrett 2011).

4. Maquette et tests utilisateurs

4.1 Prototypage de Yareta

4.1.1 Analogie avec la méthodologie de Jesse James Garrett

Dans la revue de la littérature, l'approche de Jesse James Garrett a été abordée et un parallèle avec notre projet semble essentiel ; en effet, les étapes dont parle Garrett ont été suivies (cf. Le design web centré sur l'utilisateur) :

1. Les objectifs de la plateforme avaient été définis en amont par le DLCM (Cazeaux 2018) et les besoins utilisateurs ont été étudiés en première partie de projet à travers la revue de littérature et les entretiens ;
2. Les spécificités fonctionnelles de base avaient également été prévues par le DLCM dans les recommandations à l'égard de la plateforme, ainsi que dans la première version de Yareta (Cazeaux 2018 et Université de Genève 2019b) (cf. Annexe 10). En revanche, peu d'éléments de contenu étaient rédigés, puisque Yareta est une plateforme de préservation à long terme de données de recherche et que son contenu principal comprendra les données versées par ses utilisateurs ; en effet, les contenus devant être rédigés par le DLCM et l'Université de Genève portent sur la présentation du portail et de son fonctionnement. Nous verrons plus tard que des recommandations en termes de contenu ont été formulées, même si le mandat porte sur l'ergonomie de la plateforme et non sur son contenu (cf. Recommandations).
3. L'organisation des éléments et l'architecture de l'information avaient aussi été prévues par le DLCM dans la version initiale de Yareta (Université de Genève 2019b). Cependant, comme nous le verrons, certains aspects ont été revus en fonction de la revue de littérature pour offrir plus de clarté aux futurs usagers.
4. Lorsque nous avons débuté le prototypage, nous avons commencé par ce que Garrett appelle l'« ossature », puisque les étapes précédentes avaient déjà été éclaircies par le DLCM.
5. Nous avons également abordé la surface, même si nous nous sommes concentrées davantage sur l'ergonomie. En ce qui concerne l'esthétisme, la charte graphique de Yareta avait également été définie en amont de notre projet (DLCM 2019b).

4.1.2 Conception de la première maquette

Comme évoqué dans la méthodologie, nous avons débuté le processus de création de la maquette en travaillant sur papier, puis sur le logiciel *Google Drawings*, car le papier ne nous permettait pas d'ajouter et de modifier des détails aussi facilement que nous le souhaitions :

Figure 7: Prototype papier

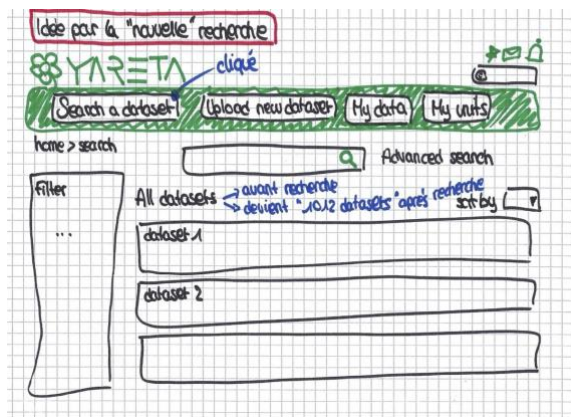
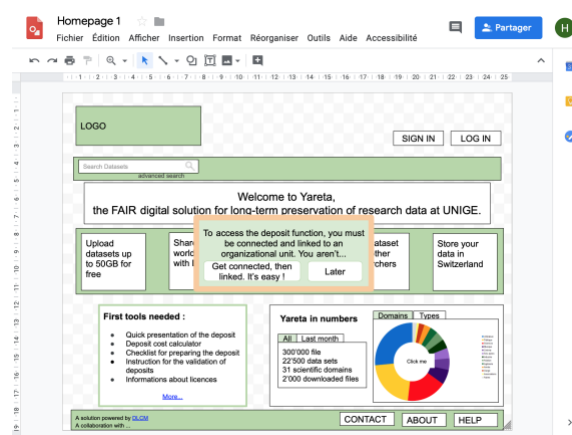
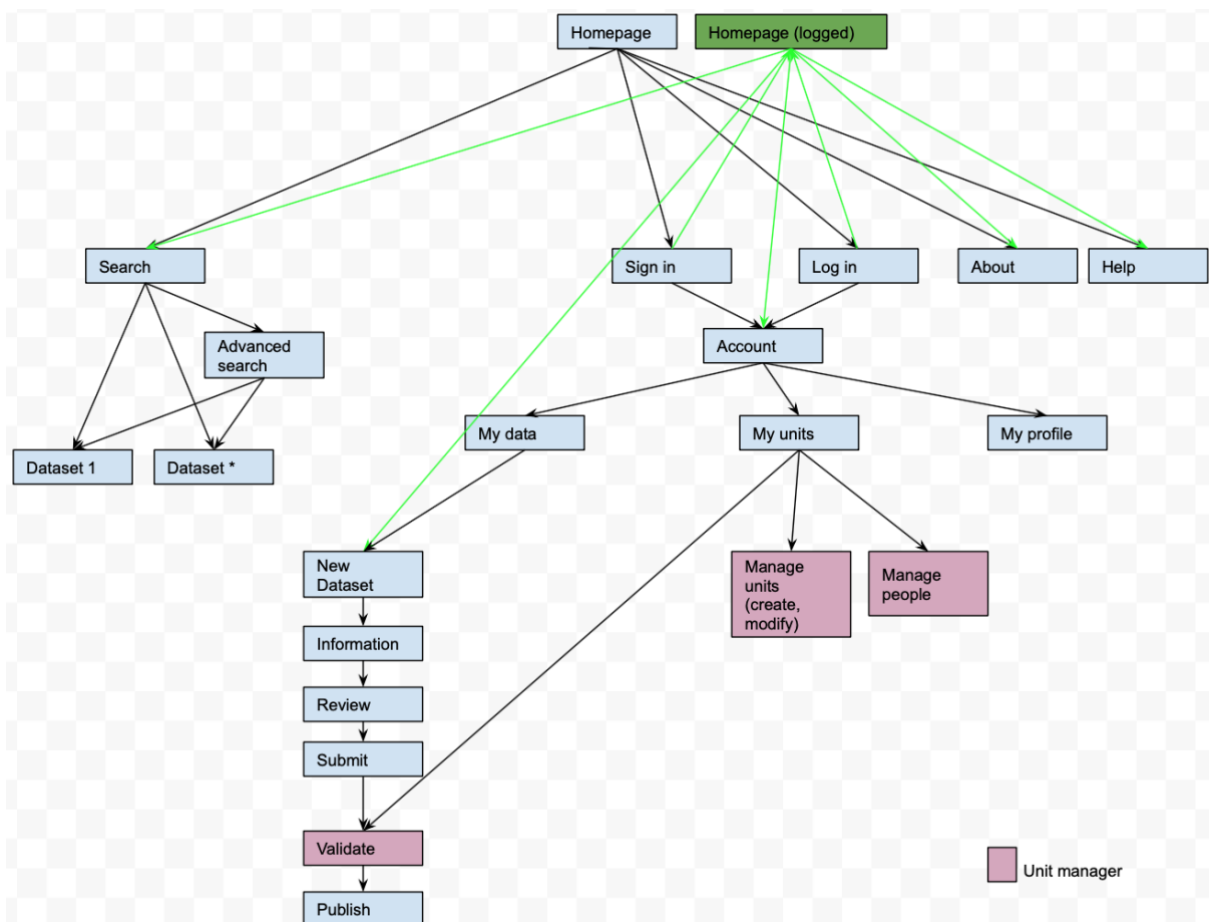


Figure 8: Prototype *Google Drawings*



En parallèle du prototypage, nous avons réfléchi à la navigation entre les pages, dont le schéma montre que les accès ne sont pas les mêmes si l'utilisateur est connecté ou non à la plateforme (cf. Figure 9: Document de travail présentant la navigation entre les pages de la première maquette). Les cases pourpres représentent les pages gérées par les administrateurs des unités organisationnelles ; ces éléments n'ont pas été approfondis lors de ce projet.

Figure 9: Document de travail présentant la navigation entre les pages de la première maquette



Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Le schéma montre également que les pages sont liées à des fonctionnalités clés de la plateforme ; en effet, l'utilisateur doit pouvoir rechercher des jeux de données, se connecter, déposer des jeux de données, ainsi que gérer ses jeux de données, son profil et ses unités organisationnelles.

Ensuite, nous avons utilisé le logiciel Axure RP9 Team, pour travailler le design d'interface, le design d'information et le contenu, ainsi que pour simuler les interactions (Daumal 2018). Dans la suite de ce chapitre, nous aborderons successivement les pages principales de la première maquette. L'intégralité de la première maquette peut être consultée via le lien suivant : <https://xmfcgv.axshare.com/>.

De manière générale, nous avons utilisé un modèle de page identique sur l'ensemble de la plateforme, pour favoriser le confort de l'utilisateur (Schneider 2018b). Nous avons également fait en sorte que l'utilisateur de la plateforme n'ait jamais l'impression de chercher une information, sans pour autant surcharger les pages, car les utilisateurs ne lisent pas l'entier des contenus (Krug 2012) (cf. Annexe 20). Par exemple, les textes doivent être rédigés de manière claire et concise et les boutons doivent comprendre des mots univoques et non ambigus (Krug 2012). Par ailleurs, nous avons prévu des aides sous plusieurs formes, en fonction des besoins : elles peuvent être rassemblées dans un endroit (ex : « Starter kit » de la page d'accueil), présentées sous forme d'exemples dans des champs de saisie (ex : formulaire de dépôt de données de recherche) ou présentées sous forme de points d'interrogation développés au passage de la souris (ex : formulaire de dépôt de données de recherche) (Schneider 2018b).

Certaines fonctionnalités ou actions n'ont pas été implémentées dans la maquette. Nous les avons signalées aux participants au moyen d'un message qui s'affiche lorsqu'ils cliquent sur une fonctionnalité non disponible.

Figure 10: Message d'avertissement affiché lorsque des fonctionnalités ne sont pas implémentées



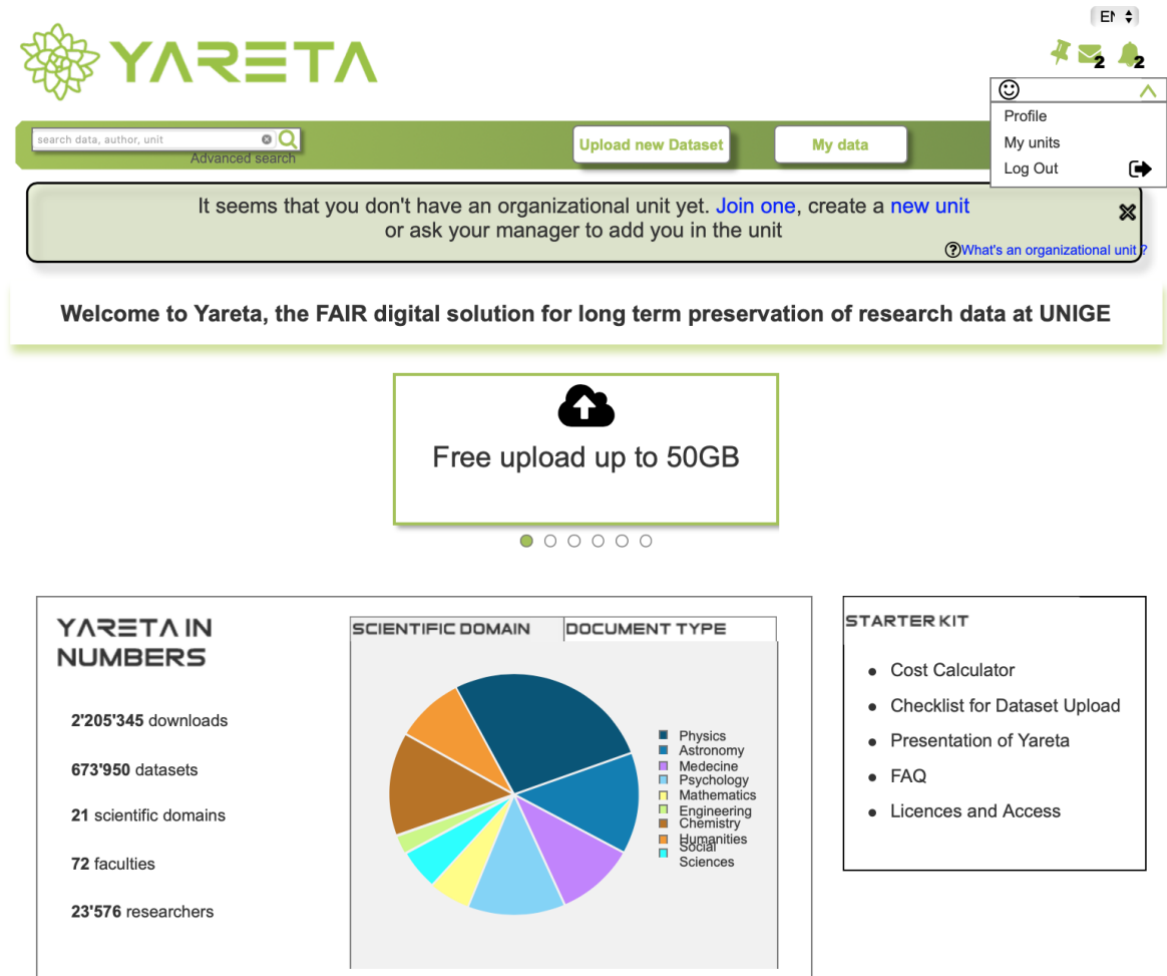
Feature not available in this mockup

4.1.2.1 Page d'accueil

La page d'accueil est la première page consultée par les utilisateurs ; elle doit donc les inciter à utiliser la plateforme, tout en :

- Présentant l'identité, la fonction et le contenu de la plateforme ;
- Stimulant la curiosité de l'utilisateur en promouvant le contenu et la fonction ;
- Permettant à l'utilisateur de se connecter ;
- Orientant l'utilisateur vers ce qu'il cherche (Krug 2012).

Figure 11: Page d'accueil de la première maquette lorsque l'utilisateur est connecté



Le logo se trouve en haut, à gauche, par convention (Schneider 2018a). Il permet à l'utilisateur de revenir à la page d'accueil depuis toutes les pages de la plateforme. Le menu de gestion du profil et des unités organisationnelles se trouve en haut, à droite, comme sur la plupart des plateformes qui permettent une connexion (ex : Google, Facebook, Amazon)²⁰. Les trois icônes se trouvant au-dessus du profil (la punaise, l'enveloppe et la cloche) représentent respectivement les fonctionnalités permettant d'enregistrer des jeux de données, de recevoir des messages à l'interne de la plateforme et de recevoir des notifications – si un jeu de données a été validé et publié, par exemple.

Le bandeau vert propose la navigation principale qui permet d'accéder aux fonctionnalités clés de la plateforme, à savoir la recherche de jeux de données, le dépôt de jeux de données et la gestion de ses propres jeux de données. Le *pop-up* vert clair, qui se trouve juste au-dessous du bandeau de navigation, doit orienter l'utilisateur suite à sa première connexion ; en effet,

²⁰ Cet affichage n'a pas été maintenu, car les participants aux tests utilisateurs ne trouvaient pas les informations contenues dans ce menu déroulant. Finalement, seul l'accès au profil a été maintenu à cet endroit de la page (cf. Annexe 20).

pour pouvoir déposer un jeu de données, il est indispensable d'être affilié à une unité organisationnelle.

La partie supérieure de la page d'accueil contient un mot d'introduction destiné à présenter la fonction de la plateforme et un carrousel destiné à présenter les avantages de la plateforme par rapport à ses concurrentes ; le but est d'informer l'utilisateur sur l'identité et la fonction du site sur lequel il se trouve (Krug 2012).

Ensuite, l'encart de gauche présente des statistiques sur les jeux de données déjà déposés sur la plateforme ; cet élément d'information a été demandé par des participants lors des entretiens (cf. Annexe 4). Les deux onglets affichant des graphes sont pensés comme des éléments cliquables qui redirigerait l'utilisateur sur la page de recherche en affichant des jeux de données correspondant aux critères des graphes. Par exemple, un utilisateur qui sélectionnerait le quartier bleu foncé se verrait redirigé sur une page de résultat de recherche rassemblant les jeux de données rattachés au domaine de la physique.

Finalement, le « Starter kit » de droite rassemble les éléments d'information importants pour l'utilisateur lors de l'utilisation de la plateforme. Les éléments prévus à ce stade du projet sont indicatifs et non exhaustifs, et les contenus correspondants n'ont pas été rédigés dans le cadre de ce travail.

Figure 12: Bandeau de la première maquette lorsque l'utilisateur n'est pas connecté



Lorsque l'utilisateur n'est pas connecté, il n'a accès qu'à la page d'accueil et à la fonctionnalité de recherche de jeux de données, comme le montre la barre de recherche située sur la gauche²¹. Les fonctionnalités d'inscription et de connexion se trouvent en haut de la page, à droite, par convention (Schneider 2018b).

4.1.2.2 Pages d'inscription et de connexion

Les pages d'inscription et de connexion sont prévues, mais les éléments réellement demandés aux usagers dépendent du DLCM et de l'Université de Genève. Aussi, les contenus que nous avons imaginés sont indicatifs.

²¹ Cet affichage n'a pas été maintenu, car il ne permet pas à l'utilisateur non connecté de comprendre les fonctionnalités principales du site (Krug 2012) (cf. annexe 20).

Figure 13: Page d'inscription de la première maquette

Figure 14: Page de connexion de la première maquette

La deuxième figure illustre aussi le bandeau inférieur de la page (*footer*). Il contient les liens vers les pages « A propos » et « Aide » dont les contenus n'ont pas été rédigés dans le cadre de ce projet. Leur emplacement est le fruit d'une convention (Schneider 2018a).

4.1.2.3 Page de recherche

La page de recherche est accessible aussi bien par les utilisateurs connectés que par les utilisateurs non connectés. La possibilité ou non de sauvegarder des jeux de données dans les « favoris » (fonctionnalité de la punaise, en haut de l'écran, à droite), et de consulter les pages des auteurs de jeux de données et des unités organisationnelles sera réservée aux utilisateurs connectés. De manière générale, la page est organisée comme les pages de recherche auxquelles les utilisateurs du web sont habitués ; ce schéma peut être observé sur RERO Explore, par exemple (RERO 2019).

Figure 15: Page de résultats de recherche de jeux de données de la première maquette

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Pour la première fois, un fil d'Ariane a été mis en place pour faciliter l'orientation de l'utilisateur sur la plateforme (Krug 2012 et Schneider 2018b).

Dans la partie supérieure de la page, nous avons déplacé la barre de recherche, initialement située dans le bandeau de navigation au-dessous de celui-ci, centrée, de manière à occuper la page qui lui est finalement dédiée. Le bouton « Advanced search » permet de naviguer entre deux modes de recherche : la recherche simple propose une recherche par mots clés et des options de filtre, tandis que la recherche avancée propose de combiner des mots clés et d'autres critères de recherche – non exhaustifs sur la figure ci-dessous –, les options de filtres restant à disposition. A terme, il pourrait être intéressant d'envisager d'intégrer l'autocomplétion²² et la suggestion en temps réel²³ à la barre de recherche pour faciliter les saisies des utilisateurs (Schneider 2018b).

Figure 16: Recherche avancée de la première maquette

Sur la gauche de la page sont disposées les options de filtre évoquées plus haut. La liste de filtres n'est pas exhaustive.

Le centre de la page est dédié aux résultats de recherche, divisés en grandes catégories (jeux de données, auteurs et unités organisationnelles) grâce à des onglets. Nous suggérons que les termes utilisés pour la recherche soient surlignés dans les résultats de recherche, pour une meilleure visualisation. Chaque résultat serait décrit avec des métadonnées de base. Sur la droite des encarts dédiés à chaque résultat se trouve une icône²⁴ décrivant l'accès aux données et la punaise permettant d'ajouter le jeu de données à ses « favoris ».

4.1.2.4 Page de détail d'un jeu de données

Depuis les résultats de recherche, les utilisateurs doivent pouvoir cliquer sur un jeu de données et accéder à sa page de détail. Cette page doit comprendre les métadonnées du jeu de données, un aperçu des fichiers et une fonctionnalité de téléchargement. Dans la première maquette, cette page est une ébauche disponible en annexe (cf. Annexe 11).

²² « [L']autocomplétion fournit des suggestions de valeurs en temps réel destinées à un champ de saisie » (Scott et Neil 2010).

²³ « [La] suggestion en temps réel fournit des suggestions de termes en temps réel pour mettre au point une recherche » (Scott et Neil 2010).

²⁴ Cet affichage n'a pas été maintenu, car les participants aux tests utilisateurs ne comprenaient pas la signification des icônes (cf. annexe 20).

4.1.2.5 Formulaire de dépôt d'un jeu de données

Le formulaire de dépôt de jeux de données est essentiel, dans la mesure où il permet à la plateforme de réaliser son objectif prioritaire, à savoir celui d'enrichir son dépôt. Nous avons opté pour un processus en deux grandes étapes : sur la première page, l'utilisateur verse ses fichiers, puis les décrits avec des métadonnées ; sur la deuxième page, l'utilisateur vérifie les métadonnées saisies et soumet son jeu de données pour une validation ultérieure par le manager de son unité organisationnelle.

Figure 17: Première page du formulaire de dépôt de jeux de données de la première maquette

The screenshot shows the 'New Dataset' form in the YARETA platform. The form is titled 'New Dataset' and is in the 'Review' stage. It features a green header with the YARETA logo, a search bar, and buttons for 'Upload new Dataset' and 'My data'. The main content area is divided into 'File(s)' and 'Metadata' sections. The 'File(s)' section has a large area for file upload with a 'Drag & Drop' instruction and a 'Choose file(s)' button. The 'Metadata' section has three rows for 'MANDATORY', 'RECOMMENDED', and 'OPTIONAL' fields. At the bottom, there are 'Cancel', 'Save', and 'Next' buttons.

D'abord, en haut de la première page du formulaire, au centre, on remarque une ébauche de barre de progression destinée à guider l'utilisateur dans le processus de dépôt (Schneider 2018b). Ensuite, un espace est prévu pour permettre à l'usager de déposer ses fichiers de deux façons : soit il peut glisser et déposer ses fichiers depuis son gestionnaire de fichiers vers la page (*drag and drop*), soit il peut ouvrir son gestionnaire de fichiers grâce à un bouton. Une fois les fichiers versés, l'usager peut les envoyer ou non sur la plateforme grâce à un système de cases à cocher.

Figure 18: Sélection des fichiers avant leur versement dans la première maquette

Files and metadata > Review

New Dataset

File(s)

	Title	Size	Format	Creation date
<input type="checkbox"/>	Datas.csv	200 Mo	Tabl	27/07/19
<input type="checkbox"/>	Readme.txt	13 Mo	Text	27/07/19
<input type="checkbox"/>	Datas_2.doc	14 Mo	Text	27/07/19
<input type="checkbox"/>	Transcripts	15 Go	Folder	27/07/19
<input type="checkbox"/>	Interview_1.flacc	5 Go	Video	27/07/19
<input type="checkbox"/>	Interview_2.flacc	4 Go	Video	27/07/19
<input type="checkbox"/>	Interview_3.flacc	6 Go	Video	27/07/19

Ensuite, l'utilisateur est invité à décrire les fichiers préalablement sélectionnés grâce à des métadonnées. Nous avons opté pour cette procédure, de manière à s'assurer que les métadonnées décrivent bien les fichiers déposés et qu'aucune confusion ne soit possible. Les métadonnées sont divisées en trois niveaux hiérarchiques : les métadonnées obligatoires, recommandées et optionnelles. L'affichage sous forme d'accordéon permet de ne pas submerger l'utilisateur de métadonnées et d'aborder le formulaire progressivement (Scott et Neil 2010 et Schneider 2018b). Des aides sont prévues pour les métadonnées dans les champs, sous forme d'exemples, ou dans la marge de droite, sous forme de points d'interrogation se développant au passage de la souris. A ce stade, les métadonnées implémentées sont celles présentées dans les recommandations du DLCM (Cazeaux 2018). Certaines métadonnées, comme l'embargo, nous ont semblé déjà performantes dans le formulaire de la première version de Yareta ; aussi, nous les avons reprises telles quelles, sans en modifier l'apparence. L'intégralité du formulaire est disponible en annexe (cf. Annexe 12).

Figure 19: Boutons de la fin de la première page du formulaire de la première maquette

A la fin de la page, les boutons permettent à l'utilisateur d'annuler sa démarche, d'enregistrer le formulaire sans poursuivre le processus ou de poursuivre en accédant à la page suivante²⁵. En effet, il est important que l'utilisateur ne soit pas contraint dans sa démarche s'il souhaite quitter le processus (Schneider 2018b). Une fonction de sauvegarde automatique du formulaire est souhaitée et fait l'objet d'une recommandation (cf. Annexe 27).

La deuxième page du formulaire récapitule les métadonnées saisies dans la première page. Après vérification, l'utilisateur peut revenir à la page précédente pour apporter des modifications à son dépôt, enregistrer le formulaire pour le modifier ultérieurement ou

²⁵ Cet affichage n'a pas été maintenu, par manque de clarté pour les participants aux tests utilisateurs (cf. annexe 20).

soumettre son jeu de données pour une validation par le manager de son unité organisationnelle.

4.1.2.6 Pages de gestion des jeux de données

Cette page a été inspirée de la première version de Yareta, car la succession entre les différents statuts accordés aux jeux de données permet aux usagers de mieux gérer leurs dépôts (cf. Figure 3: Page « Dépôts » de Yareta). Cependant, nous avons ajouté des icônes à ces statuts pour offrir une aide visuelle aux utilisateurs qui consultent l'ensemble de leurs dépôts. Les boutons d'actions situés dans la colonne de droite disposent d'une aide affichée sur le passage de la souris (*tooltip*).

Par ailleurs, nous avons ajouté une option permettant de passer de la page de gestion de ses propres jeux de données à la page de gestion des jeux de données des unités organisationnelles auxquelles l'utilisateur est affilié. Cette option est incarnée sous la forme du bouton (*toggle*) situé sous le bandeau de navigation, à droite²⁶. Comme la page rassemblant les jeux de données des unités n'est pas développée dans cette maquette, nous nous concentrerons sur la page rassemblant ceux du chercheur.

Figure 20: Page de gestion de ses propres jeux de données sur la première maquette

YARETA

search data, author, unit Advanced search

Upload new Dataset My data

Home / My Data

My dataset ☐ Linnaeus Team (default) ☐

All 6 ☒ Completed 2 ☐ In progress 1 ☐ Waiting for validation 1 ☐ Rejected 1 ☐ Incorrect 1

Status	Title	Unit	Size	Upload date	Validation date	Action
✓	Cells preservation in a qualitative way for introducing viral bacteries.	Linnaeus Team	3 GB	27/08/2018	29/08/2018	
⌚	Decision making time based on academic resources	Linnaeus Team	570 MB	12/08/2019	-	
⌚	Cells preservation in a qualitative way for introducing viral bacteries.	Linnaeus Team	3 GB	27/08/2018	-	
✗	Cells preservation in a qualitative way for introducing viral bacteries.	Linnaeus Team	3 GB	27/08/2018	-	
⚠	Cells preservation in a qualitative way for introducing viral bacteries.	Linnaeus Team	3 GB	27/08/2018	-	
✓	Cells preservation in a qualitative way for introducing viral bacteries.	Linnaeus Team	3 GB	27/08/2018	29/08/2018	

Suivant les onglets consultés, lorsque ceux-ci rassemblent des jeux de données qui nécessitent une action de la part de leur auteur, des boutons d'actions sont ajoutés :

²⁶ Cet affichage n'a pas été maintenu, car les participants aux tests utilisateurs ne le voyaient pas et ne comprenaient pas son fonctionnement (cf. annexe 20).

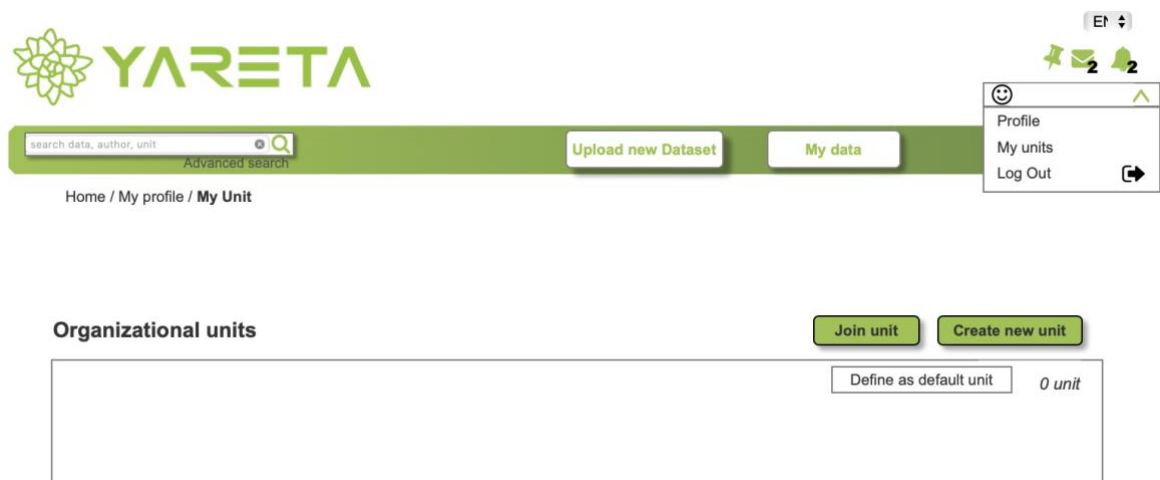
Figure 21: Focus sur les boutons d'action spécifiques à certains statuts des jeux de données sur la première maquette

All 6	✓ Completed 2	🕒 In progress 1	⌚ Waiting for validation 1	🚫 Rejected 1	⚠️ Incorrect 1	
<div>Accept dataset(s)</div> <div>Refuse dataset(s)</div>						
	Title	Unit	Size	Upload date	Validation date	Action
<input type="checkbox"/>	Cells preservation in a qualitative way for introducing viral bacteries.	DLCM Team	3 GB	27/08/2018	29/08/2018	

4.1.2.7 Page de gestion des unités organisationnelles

Une page est prévue pour la gestion des unités organisationnelles, puisque les utilisateurs sont obligés d'être affiliés à ce genre d'entité pour pouvoir déposer des jeux de données. L'utilisateur accède à cette page par le menu déroulant situé en haut de la page, à droite. A ce stade, il s'agit encore d'une ébauche.

Figure 22: Page de gestion des unités organisationnelles sur la première maquette



L'utilisateur a trois options : d'abord, il peut demander à créer une unité organisationnelle (cf. Annexe 13) ; ensuite, il peut demander à rejoindre une unité organisationnelle existante (cf. Annexe 14) ; finalement, s'il est déjà affilié à une unité organisationnelle, il peut accéder à la page de l'unité en question (cf. Annexe 15). Cette dernière page est également accessible depuis la page de recherche ; une icône permet aux utilisateurs d'envoyer directement une demande d'adhésion au manager de l'unité.

En ce qui concerne l'affichage, il est prévu que les unités auxquelles sont rattachés les utilisateurs soient rassemblées sur cette page, sous forme de vignettes.

4.2 Tests utilisateurs et itération

4.2.1 Retours

Lors de ces tests, nous avons identifié des éléments satisfaisants à maintenir et des problèmes d'utilisabilité à résoudre – qui concerne spécifiquement la maquette –, mais aussi des questions à transmettre au DLCM. Les problèmes d'utilisabilité résidaient essentiellement dans des formulations à modifier ou des fonctionnalités à ajouter. Les questions seront abordées dans la partie dédiée aux recommandations (cf. Recommandations).

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Cette étape nous a permis d'améliorer la maquette de test en test, qu'il s'agisse de l'architecture de l'information, des terminologies utilisées ou de l'adaptation du fonctionnement de la plateforme à celui de ses utilisateurs. L'ensemble des informations récoltées lors des trois séries de tests utilisateurs a été consigné dans des comptes rendus détaillés disponibles en annexe (cf. Annexes 20, 21 et 22)²⁷.

Tableau 2: Synthèse des retours obtenus lors des différentes séries de test

Première série		
Avantages	Inconvénients	Autres remarques
Messages d'aide Simplicité du formulaire	Terminologies ambiguës Icônes ambiguës Dissimulation de certains menus Barre de recherche ambiguë Amélioration possible dans l'organisation des métadonnées Fonctionnalité de téléchargement de lots de données manquantes Difficulté d'accès aux jeux de données des unités organisationnelles Barre de progression manquante dans le processus de dépôt Impossibilité de modification des métadonnées après la publication Aides pour la complétion des métadonnées insuffisantes Fin du formulaire de dépôt peu claire	Non lecture des contenus par les participants Informations de la page d'accueil mal hiérarchisées Visibilité limitée du bandeau de navigation Concept d'unité organisationnelle ambigu Fil d'Ariane cliquable à prévoir Aide à l'orthographe à prévoir Traduction des métadonnées à prévoir Suppression des jeux de données publiés à prévoir
Deuxième série		
Simplicité de la procédure d'inscription	Disposition des informations dans les résultats de recherche à revoir	Concept d'unité organisationnelle ambigu

²⁷ Les retours obtenus n'ont pas tous fait l'objet de recommandation.

<p>Mise en évidence de la liste des unités organisationnelles que les utilisateurs peuvent demander à rejoindre</p> <p>Notifications</p> <p>Simplicité du processus de dépôt</p> <p>Barre de progression dans le processus de dépôt</p> <p>Fonctionnalité de <i>drag and drop</i></p> <p>Hiérarchie des métadonnées demandées dans le formulaire de dépôt</p>	<p>Disposition des informations dans la page de détail des jeux de données à revoir</p> <p>Confusion entre la gestion de ses propres jeux de données et ceux de ses unités organisationnelles</p> <p>Certaines pages à renommer</p> <p>Incompréhension de certaines métadonnées</p>	<p>Questions sur les modalités de dépôt (prix, formats, etc.)</p> <p>Complétion collaborative des métadonnées du formulaire de dépôt à prévoir</p> <p>Autorisation du manager à corriger les métadonnées avant le dépôt à prévoir</p>
Troisième série		
<p>Simplicité de la procédure d'inscription</p> <p>Possibilité de demander à rejoindre une unité organisationnelle</p> <p>Affichage du nombre de résultat de recherche</p> <p>Simplicité du formulaire de dépôt</p>	<p>Terminologie utilisée dans le bandeau de navigation à revoir</p> <p>Confusion entre la gestion de ses propres jeux de données et ceux de ses unités organisationnelles</p> <p>Fonctionnalité de citation des jeux de données à prévoir</p> <p>Mise en évidence l'auteur des jeux de données et son affiliation à prévoir dans les résultats de recherche</p> <p>Aides supplémentaires à prévoir</p>	<p>Concept d'unité organisationnelle ambigu</p> <p>Questions sur la propriété des données et les possibilités de modification des jeux de données après publication</p>

4.2.2 Maquette finale

Le processus itératif a permis au projet d'aboutir à une quatrième et dernière maquette soumise au DLCM. Les deuxième et troisième maquettes qui incarnent différents niveaux de maturité du prototype ne sont pas présentées dans ce rapport, mais peuvent être consultées aux liens suivants :

- Deuxième maquette : <https://zlrgwh.axshare.com>
- Troisième maquette : <https://6m98fn.axshare.com>

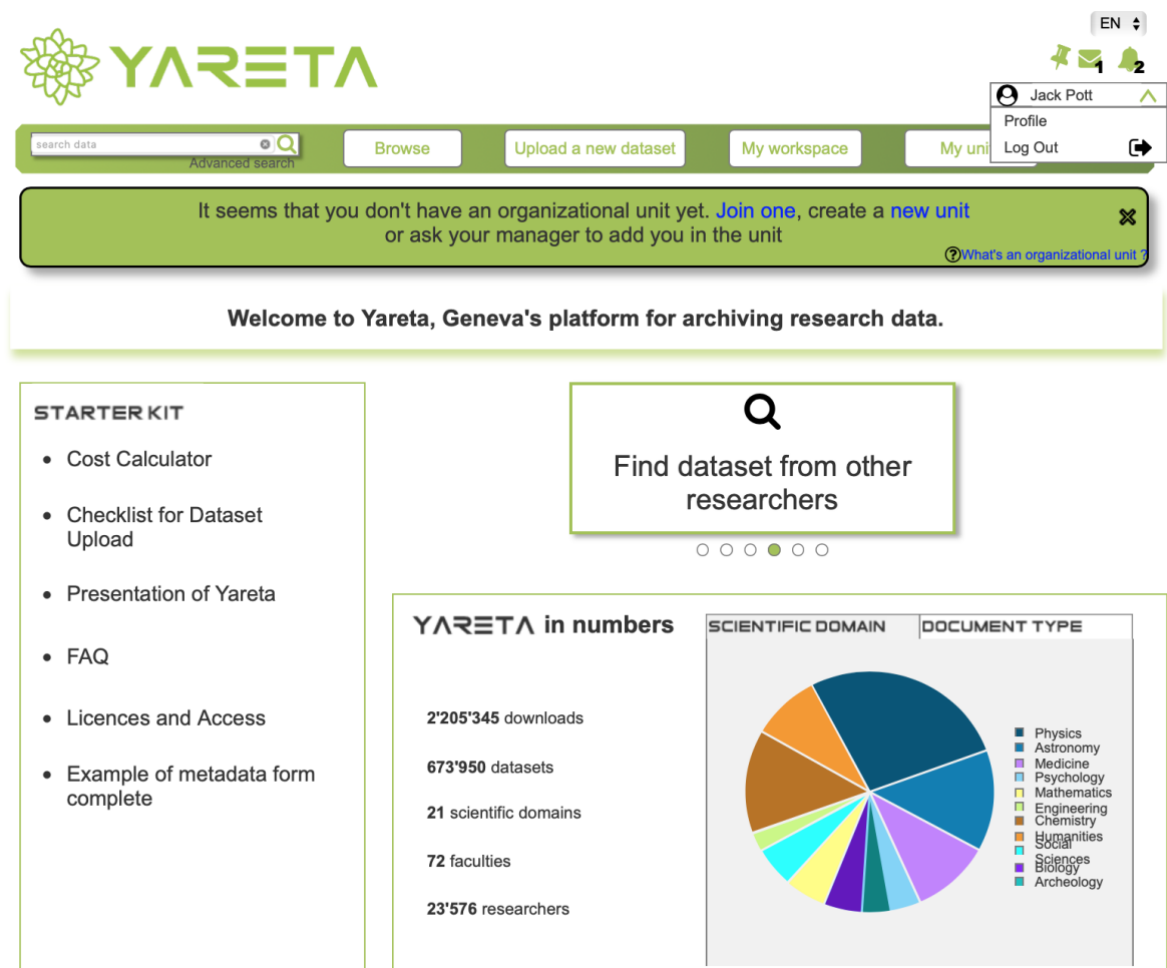
Quant à la quatrième maquette, elle va être présentée ci-après, de manière à montrer les évolutions implémentées par rapport à la première maquette (cf. Maquette finale) ; elle peut être consultée dans son intégralité au lien suivant : <https://v4aj02.axshare.com>. La liste des

modifications apportées entre les différentes maquettes est disponible en annexe (cf. Annexe 23).

4.2.2.1 Page d'accueil

La page d'accueil permet de constater les modifications générales apportées à la maquette. D'abord, le bandeau de navigation a été homogénéisé : que l'utilisateur soit connecté ou non, il aura accès au même menu ; toutefois, les pages de soumission et de gestion des jeux de données, ainsi que la page de gestion des unités organisationnelles ne peuvent toujours être consultées que si l'usager est connecté. De plus, le bandeau donne dorénavant aussi accès à la page de gestion des unités organisationnelles auparavant accessibles depuis le menu situé en haut, à droite. Les intitulés des menus ont été modifiés pour plus de clarté pour les utilisateurs et un menu « Browse » a été ajouté pour donner accès directement à l'ensemble des jeux de données déposés sur la plateforme ; cet ajout est le fruit d'une demande des participants aux tests utilisateurs (cf. Annexe 22).

Figure 23: Page d'accueil de la dernière maquette



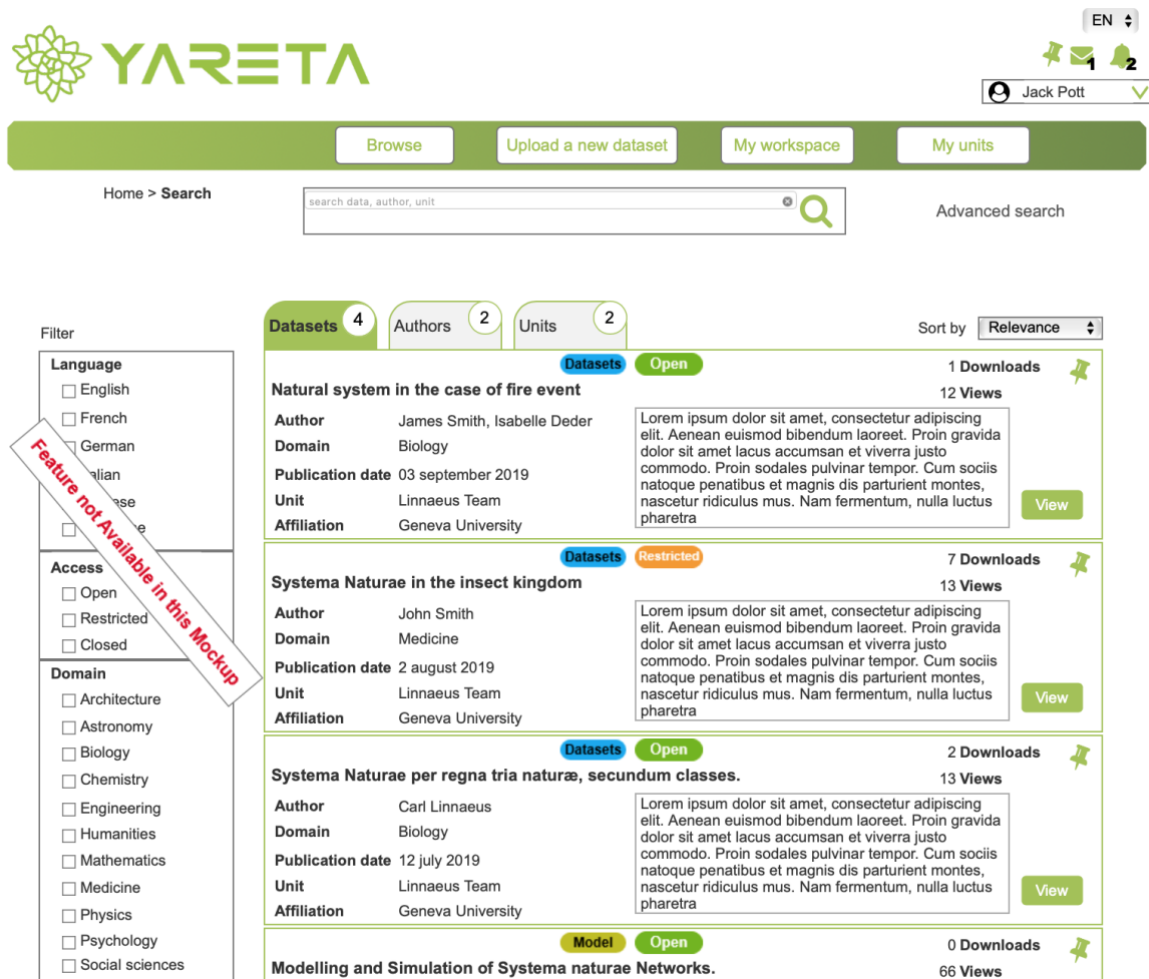
En ce qui concerne spécifiquement la page d'accueil, seules deux modifications sont observables : d'une part, le message d'accueil a été raccourci et modifié pour plus de clarté pour les utilisateurs ; d'autre part, les aides du « Starter kit » ont été déplacées à droite pour prendre plus de place et avoir la priorité de lecture sur les statistiques (Schneider 2018a).

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

4.2.2.2 Page de recherche

La fonctionnalité de recherche est désormais divisée en trois parties : la recherche simple et la recherche avancée qui ont déjà été présentées, ainsi que la fonctionnalité de parcours de l'ensemble des jeux de données (« Browse »). Pour l'option « Browse », le total des jeux de données disponibles est affiché en dessus de l'option de tri « Sort by », à droite.

Figure 24: Page de recherche de jeux de données de la dernière maquette



En ce qui concerne l'affichage des résultats de recherche, les métadonnées affichées ont été enrichies et associées à des libellés pour plus de clarté et une meilleure description, et les icônes décrivant l'accès aux jeux de données ont été remplacées par des étiquettes textuelles colorées. De plus, la qualification du jeu de données figure désormais sous forme d'étiquette colorée à gauche de l'indication d'accès. Ensuite, les statistiques de téléchargement et de consultation des jeux de données ont été ajoutées sur la partie supérieure droite des encarts dédiés à chaque jeu de données. Finalement, un bouton « View » donnant accès au détail des jeux de données a été ajouté, car les participants aux tests ne comprenaient pas qu'ils pouvaient atteindre ces informations en cliquant sur le titre du jeu de données.

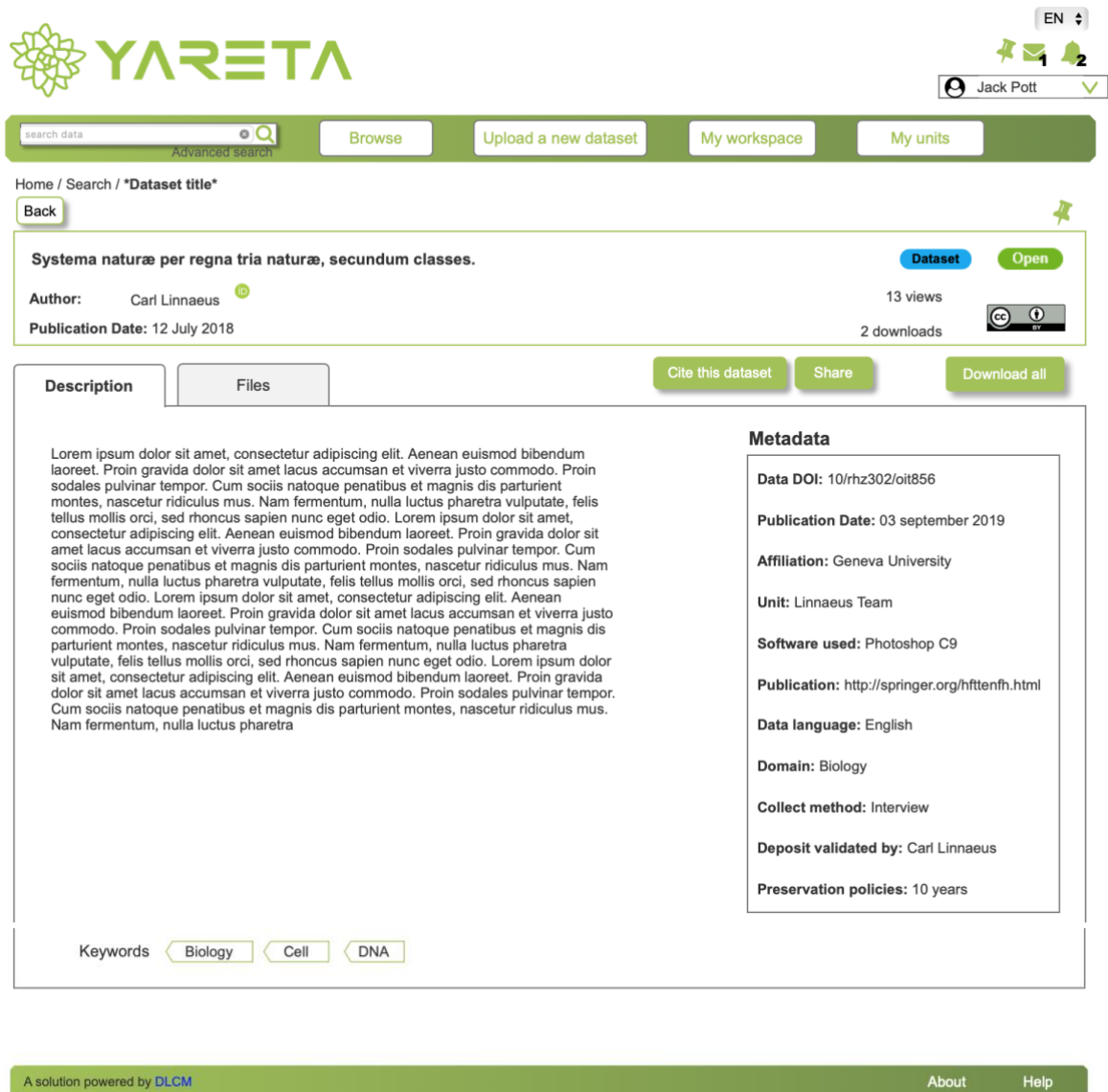
4.2.2.3 Page de détail d'un jeu de données

Cette page a été enrichie pour que l'ensemble des métadonnées demandées dans le formulaire de dépôt y figurent. Des fonctionnalités de citation, de partage et de téléchargement

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

de lots de fichiers ont été ajoutées à la demande des participants aux tests (cf. Asnnexe 21 et 22).

Figure 25: Page de détail d'un jeu de données dans la dernière maquette



4.2.2.4 Formulaire de dépôt des jeux de données

De manière générale, la structure du formulaire en plusieurs étapes décrite précédemment a été maintenue (cf. Conception de la première maquette).

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Figure 26: Première page du formulaire de dépôt de jeux de données de la dernière maquette

Home / New dataset

Before uploading a new dataset, consult out [starter kit](#) to be sure you don't forget anything

New Dataset

Files > Metadata > Preview > Submit

1. File(s)

*** Mandatory step**

Feature not available in this mockup

Drag & Drop or Choose file(s)

Upload conditions (size max, number of files max, general informations for the upload)

2. Metadata

	MANDATORY
	RECOMMENDED
	OPTIONAL

Request a DOI ? Cancel Save & quit Next

Dans la partie supérieure de la première page, on remarque d'abord qu'un *pop-up* destiné à orienter l'utilisateur vers les aides susceptibles d'être utiles avant de débuter le processus. Ensuite, la barre de progression a été détaillée, de manière à mieux annoncer la structure du processus.

Figure 27: Sélection des fichiers avant leur versement dans la dernière maquette

1. File(s)

	Title	Size	Format	Creation date
<input type="checkbox"/>	Data.csv	200 Mo	Tabl	27/07/19
<input type="checkbox"/>	Readme.txt	13 Mo	Text	27/07/19
<input type="checkbox"/>	Data_2.doc	14 Mo	Text	27/07/19
<input checked="" type="checkbox"/>	Transcripts	15 Go	Folder	27/07/19
<input type="checkbox"/>	Interview_1.flacc	5 Go	Video	27/07/19
<input type="checkbox"/>	Interview_2.flacc	4 Go	Video	27/07/19
<input type="checkbox"/>	Interview_3.flacc	6 Go	Video	27/07/19

Delete file(s) Back Upload files ?

En ce qui concerne le dépôt des fichiers, les deux options de sélection sont conservées. Cependant, une fois les fichiers sélectionnés, les boutons ont été développés pour éviter toute confusion auprès des utilisateurs quant aux actions possibles et l'option de retour en arrière a été ajoutée pour permettre plus de flexibilité dans la démarche (cf. Annexe 20).

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Quant à la partie dédiée à la saisie des métadonnées, si sa structure a été maintenue, certaines métadonnées ont toutefois été ajoutées pour correspondre au mieux au résultat de la comparaison entre les métadonnées recommandées par le DLCM et celles du schéma de métadonnées DataCite (cf. Annexe 28). De plus, certaines ont changé de position hiérarchique. Par exemple, la métadonnée « Primary data type » est passée du statut obligatoire au statut optionnel. Suite aux discussions menées avec les participants aux différentes sessions de tests, les intitulés de certaines métadonnées ont évolué pour plus de clarté et certaines d'entre elles peuvent être dupliquées pour une meilleure description. De manière générale, les aides ont été développées pour rassurer les utilisateurs, même si leur contenu reste basique, car ils ne constituent pas le cœur de notre travail.

Figure 28: Boutons de la fin de la première page du formulaire de la dernière maquette



Finalement, en bas de page, le bouton « Save » a été complété pour éviter toute confusion auprès des utilisateurs et le bouton « Request a DOI » a été ajouté, car certains participants aux tests pensaient que le DOI demandé dans le formulaire était celui de la publication dans le cadre de laquelle avaient été générées les données (cf. Annexe 20).

L'intégralité du formulaire est disponible en annexe (cf. Annexe 26).

La deuxième page fait office de résumé du dépôt : la liste des fichiers versés a été ajoutée pour que l'aperçu du dépôt soit complet.

4.2.2.5 Page de gestion des jeux de données

La page la plus complexe à concrétiser a été la page rassemblant les données de recherche des utilisateurs et de leur unité organisationnelle, que nous avons appelée successivement « My data », « My dataset », puis « My workspace » ; en effet, « My data » aurait aussi pu faire référence au profil de l'utilisateur et « My dataset » était trop restrictif, puisque nous voulions que les jeux de données des unités organisationnelles soient également accessibles depuis cette page.

Figure 29: Page de gestion des jeux de données de la dernière maquette

Unit's datasets

☐ Show only my dataset

Unit : All

Status	Title	Author	Unit	Size	Upload date	Validation date	Action
⚠	Nerve cell study software	Jack Pott	Linnaeus Team	50 Go	20 september 2019		👁️🔍
🕒	Systema totoria investigation	Marius Favrot	Linnaeus Team	3 Go	12 july 2019		👁️📄✎
🕒	My favorite flower	Albert Kunz	Linnaeus Team	3 mo	2 august 2019		👁️✅❌
🕒	Study on the four-leaf clover	Jack Pott	Linnaeus Team	1 Tb	14 september 2019		👁️📄✎
❌	Deposit test	Isabelle Deder, Marc Patek	Linnaeus Team	2 Go	2 june 2019		👁️🔍
🕒	Natural system in the case of fire event	James Smith, Isabelle Deder, Pierre Matek	Linnaeus Team	6 Go	20 june 2019		👁️✅❌
✅	Systema Naturae	John Smith	Linnaeus Team	10 Go	1 august 2019	2 august 2019	👁️
✅	The transformation of protein no. 4 into amino acid	John Smith	Linnaeus Team	50 mo	18 june 2019	20 june 2019	👁️

La page de gestion des jeux de données n'a pas seulement évolué : elle a été complètement transformée. En effet, dans la première maquette, nous avons complètement séparé les jeux de données de l'utilisateur de ceux de l'ensemble des unités organisationnelles au moyen d'un *toggle* (cf. Pages de gestion des jeux de données). Dans la deuxième maquette, nous avons fait en sorte que le *toggle* fonctionne, afin de pouvoir, pour la première fois, ajouter la tâche de validation d'un jeu de données au test utilisateur ; c'est donc à partir de cette maquette que cette fonctionnalité a pu être testée. Dans la troisième maquette, puisque le *toggle* n'était pas compris par les participants au test, nous l'avons remplacé par un bouton analogue à un onglet (cf. Annexe 24). Comme la fonctionnalité de transition entre les jeux de données de l'utilisateur et ceux de ses unités organisationnelles n'était toujours pas convaincante pour nos répondants, nous proposons, dans la dernière maquette, d'afficher l'ensemble des jeux de données des unités organisationnelles sur une page en proposant une option de filtre intitulée « Show only my data » (cf. annexe 22).

Nous proposons également de prévoir des *pop-up* dans la partie supérieure de la page, afin d'avertir les utilisateurs des tâches ponctuelles qui leur incombent.

4.2.2.6 Pages liées aux unités organisationnelles

4.2.2.6.1 Page de gestion des unités organisationnelles

Figure 30: Page de gestion des unités organisationnelles sur la dernière maquette



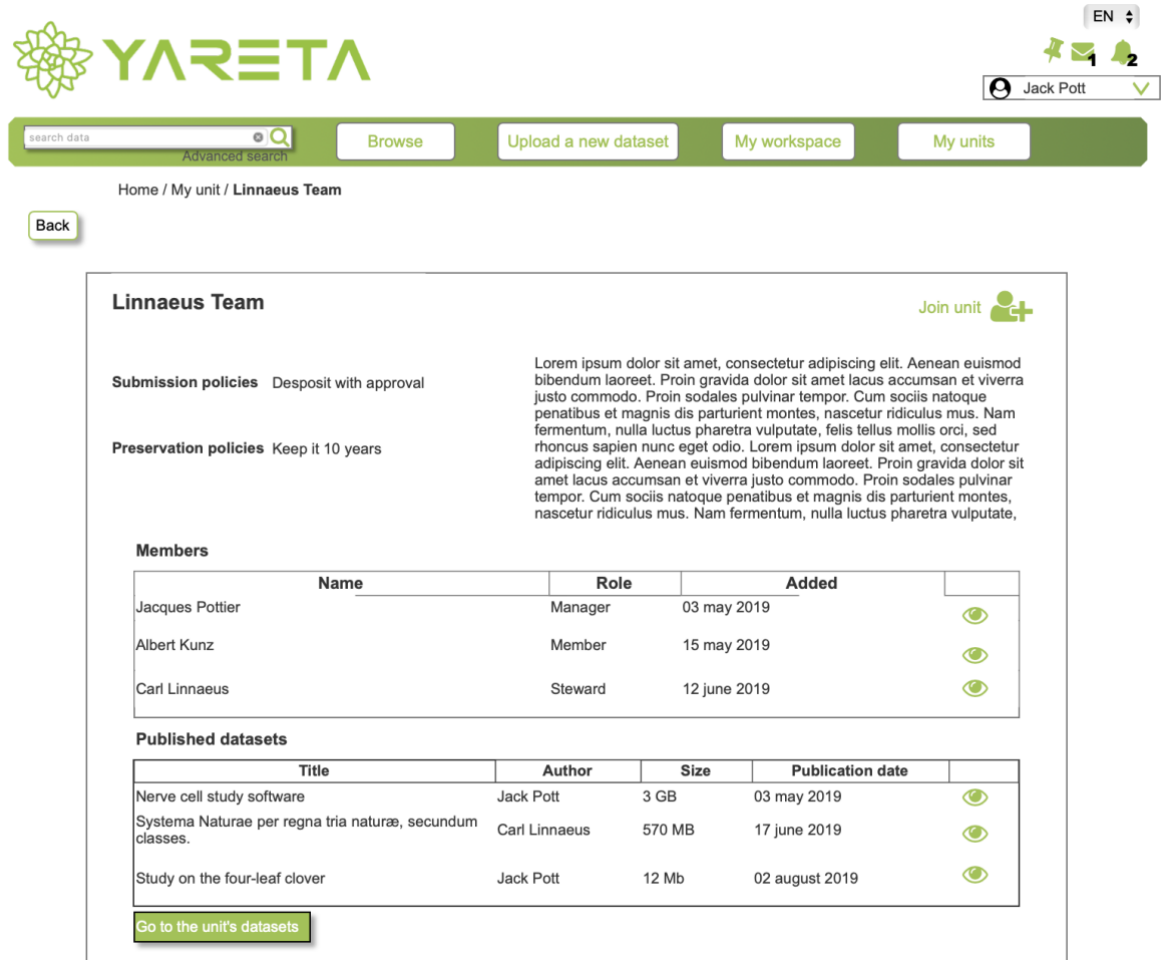
Au premier abord, cette page n'a pas beaucoup évolué. Les boutons de droite « Find a unit » et « Create new unit » ont été renommés et l'aperçu de l'unité a été ajouté. Nous suggérons aussi de prévoir différents affichages des unités organisationnelles – sous forme de vignettes ou sous forme de listes – en fonction des préférences de l'utilisateur (Schneider 2018b).

Sur la figure ci-dessus, on remarque la mention « Waiting for approval » sur la vignette de l'unité : cela signifie que l'utilisateur a demandé à rejoindre l'unité et que le manager de l'unité sollicitée doit valider la candidature de l'utilisateur.

4.2.2.6.2 Page de détail d'une unité organisationnelle

Cette page est accessible en cliquant sur l'unité organisationnelle visible sur la Figure 30: Page de gestion des unités organisationnelles sur la dernière maquette ou en cherchant l'unité depuis la page de recherche.

Figure 31: Page de détail d'une unité organisationnelle



YARETA

EN

Jack Pott

search data Advanced search

Browse Upload a new dataset My workspace My units

Home / My unit / Linnaeus Team

Back

Linnaeus Team Join unit

Submission policies Desposit with approval

Preservation policies Keep it 10 years

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus accumsan et viverra justo commodo. Proin sodales pulvinar tempor. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Nam fermentum, nulla luctus pharetra vulputate, felis tellus mollis orci, sed rhoncus sapien nunc eget odio. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus accumsan et viverra justo commodo. Proin sodales pulvinar tempor. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Nam fermentum, nulla luctus pharetra vulputate,

Members

Name	Role	Added	
Jacques Pottier	Manager	03 may 2019	
Albert Kunz	Member	15 may 2019	
Carl Linnaeus	Steward	12 june 2019	

Published datasets

Title	Author	Size	Publication date	
Nerve cell study software	Jack Pott	3 GB	03 may 2019	
Systema Naturae per regna tria naturæ, secundum classes.	Carl Linnaeus	570 MB	17 june 2019	
Study on the four-leaf clover	Jack Pott	12 Mb	02 august 2019	

Go to the unit's datasets

Comme cette page est aussi accessible à des utilisateurs qui n'y seraient pas affiliés, un bouton de demande d'adhésion est prévu dans la partie supérieure droite de l'encart ; l'icône est associée à du texte pour plus de clarté pour les usagers. Les métadonnées principales liées à l'unité figurent sur la page.

Les utilisateurs affiliés à l'unité disposent d'un bouton leur permettant d'accéder aux jeux de données de l'unité en cours de validation, grâce au bouton « Go to the unit's datasets » situé dans la partie inférieure gauche de la page.

4.2.3 Comparaison de l'utilisabilité des première et dernière maquettes

Lors des tests utilisateurs, la dernière question posée aux participants portait sur l'évaluation de l'utilisabilité de la maquette. Sur cette base, nous avons réalisé les graphes suivants portant sur la première et la troisième maquette qui s'est avérée être la dernière maquette testée :

Figure 32: Évaluation de l'utilisabilité de la première maquette (n=6)

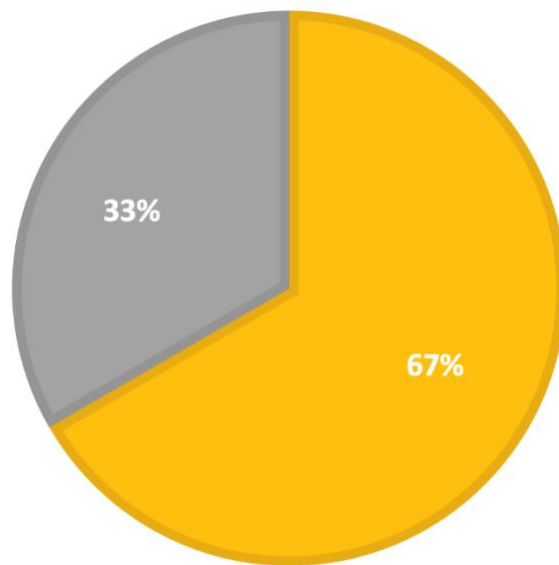
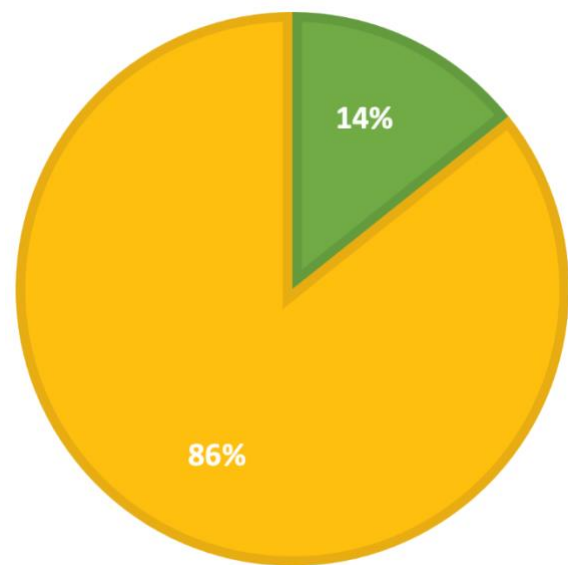


Figure 33: Évaluation de l'utilisabilité de la troisième maquette (n=7)



■ Excellent ■ Good ■ Fair

Ces graphes montrent que les retours des utilisateurs potentiels et l'itération de tests permettent indéniablement d'améliorer le prototype et de se rapprocher d'un produit qui corresponde aux attentes des usagers. Bien que notre maquette finale n'ait pu être testée, nous nous sommes basées sur les retours obtenus lors de la troisième série de tests pour la réaliser ; aussi, son utilisabilité ne peut être jugée moins satisfaisante que celle de la troisième maquette.

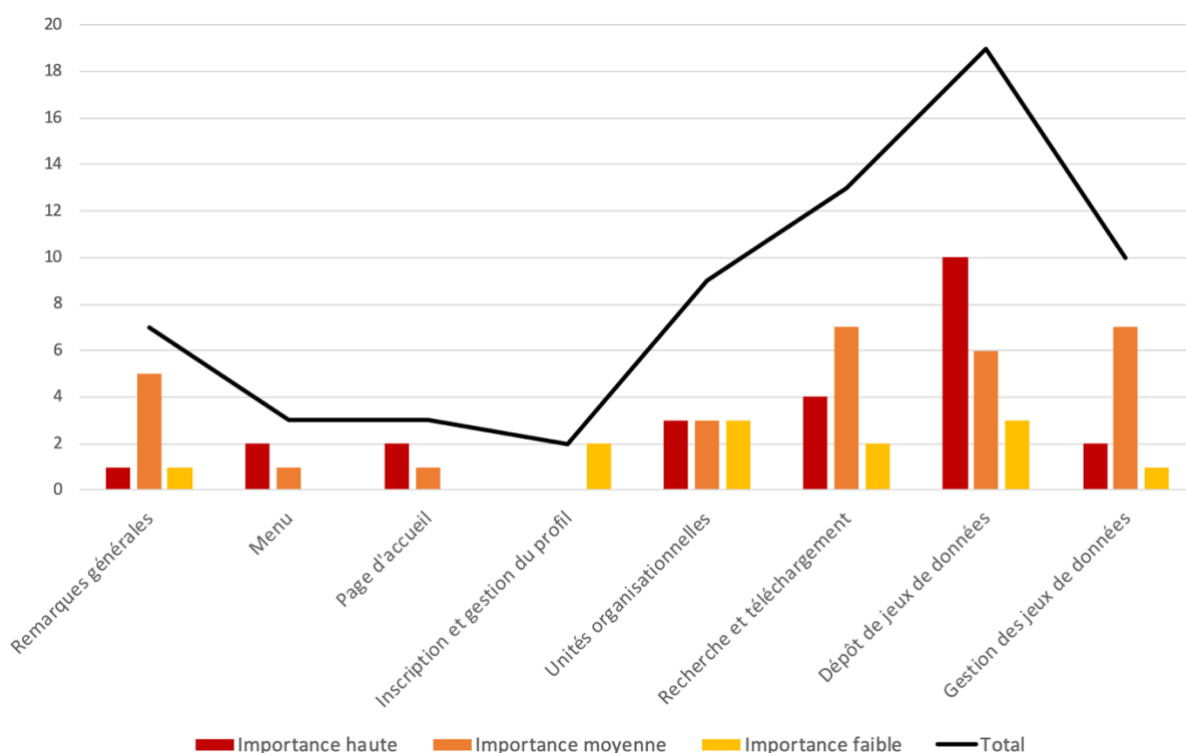
5. Recommandations

Comme présenté dans la méthodologie, les recommandations ont été divisées en deux parties distinctes : les recommandations relatives aux métadonnées à associer aux jeux de données d'une part, et les recommandations générales d'autre part (cf. Recommandations). Nous avons procédé de la sorte pour permettre une meilleure lisibilité.

En ce qui concerne les recommandations liées aux métadonnées, elles sont issues de la comparaison entre les exigences du DLCM et le schéma de métadonnées de DataCite, ainsi que des échanges avec nos répondants. Il s'agit essentiellement de renommage des intitulés et de niveau de détail à prévoir pour la description des jeux de données. Le document complet, disponible en annexe, comprend une quarantaine de rubriques (cf. Annexe 28). Nous aimerions souligner l'importance des métadonnées, car elles soutiennent la description des jeux de données, leur recherche, et, ainsi, leur réutilisation.

Quant aux soixante-six recommandations générales, elles peuvent être synthétisées sous la forme du graphe ci-dessous :

Figure 34: Recommandations générales en chiffres



En termes de nombre, on remarque que le dépôt de jeux de données a donné lieu au plus grand nombre de recommandations en général, mais aussi au plus grand nombre de recommandations importantes. Viennent ensuite la recherche et le téléchargement, puis la gestion des jeux de données. En effet, nous avons vu plus haut que la page de gestion des jeux de données a été celle sur laquelle nous avons le plus travaillé (cf. Page de gestion des jeux de données). On remarque aussi que le podium décrit ci-avant s'apparente aux trois fonctionnalités clés de la plateforme.

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

De manière générale, les recommandations portent sur l'attention à accorder aux différents niveaux de navigation prévus sur la plateforme, aux aides et aux informations essentielles à transmettre aux usagers. Nous avons également suggéré l'implémentation de fonctionnalités secondaires susceptibles de faciliter et d'améliorer l'expérience des utilisateurs sur Yareta.

Finalement, nous souhaitons souligner le fait que les chercheurs collaborent beaucoup et que cette collaboration doit être prise en compte dans le développement de cette plateforme. Par exemple, le concept d'unité organisationnelle doit être assoupli pour permettre aux chercheurs d'unités différentes de publier des jeux de données communs et la complétion du formulaire de dépôt devrait pouvoir être collaborative.

Pour plus de détails, l'ensemble des recommandations générales est aussi disponible en annexe (cf. Annexe 27).

6. Difficultés rencontrées et améliorations possibles

Même si ce projet nous a beaucoup intéressées et que son déroulement s'est bien passé, nous avons rencontré quelques difficultés que nous aimerions relever. D'abord, en termes de calendrier, le début du projet – entre mars et juin – a été ponctué par des rendus intermédiaires, comme la rédaction d'un plan de gestion des données et d'un billet de blog, qui nous a demandé du temps que nous n'avons pas pu consacrer au cœur du projet. Aussi, la constitution du panel de chercheurs que nous avons besoin de solliciter pour les entretiens et les tests utilisateurs n'a pu commencer que mi-avril. Comme l'été approchait et que les entretiens auraient lieu sur cette période, il nous a été difficile de trouver des chercheurs intéressés par notre recherche et disposés à nous accorder du temps. Heureusement, grâce à M. Andreas Fürholz, l'organisation des tests utilisateurs n'a finalement posé aucun problème.

Ensuite, lorsque nous avons débuté le prototypage de Yareta, nous avons décidé de nous inspirer de la première version de la plateforme, en pensant que cela faciliterait le travail des développeurs qui implémenteraient nos suggestions. Cette prévision ne s'est absolument pas réalisée, puisque nous ignorions qu'une version complètement différente était en cours de développement. Heureusement, cela n'empêchera pas le DLCM de prendre en compte notre travail, même si le design de la plateforme n'est pas le même.

En ce qui concerne les tests utilisateurs, il est généralement conseillé qu'ils soient mis en place par d'autres personnes que celles qui ont créé la maquette à tester, car les concepteurs de la maquette risqueraient de ne pas rester observateurs des tests et objectifs dans l'analyse des résultats (Hall 2015 et Daumal 2018). Malheureusement, cela n'a pas été possible dans notre cadre de travail, même si nous avons fait de notre mieux pour ne pas induire de biais dans l'analyse.

Le questionnaire accompagnant le test prévoyait des questions sur l'évaluation de l'utilisabilité des maquettes testées. Cependant, il existe une méthode avérée pour évaluer l'utilisabilité d'un site Internet : le questionnaire *System Usability Scale*. Il s'agit de dix questions qui visent à obtenir le point de vue de l'utilisateur sur le service testé (cf. Annexe 29) (Raphaël 2011). Malheureusement, nous n'avons pensé à cette méthode qu'après le premier test ; nous n'avons donc pas souhaité l'utiliser pour les tests suivants, de manière à rester cohérentes dans notre processus.

Par ailleurs, lors des rencontres avec des chercheurs, que ce soit dans le cadre des entretiens ou dans celui des tests d'utilisabilité, notre position d'étudiantes mandatées par le DLCM n'a pas toujours été facile à gérer, car nos répondants auraient parfois préféré interagir directement avec des membres du DLCM, de façon à obtenir des réponses plus claires à leurs questions ou à se voir garantir la prise en compte de leurs interventions.

7. Conclusion

Ce travail théorique et pratique nous a amenées à nous intéresser aux utilisateurs potentiels de Yareta, de manière à prendre en compte leurs besoins et leur réalité dans la conception du prototype de la plateforme en cours de développement. Nous avons ainsi compris que, s'ils ne sont pas familiers du concept de préservation à long terme, ils sont tout de même favorables à l'archivage et au partage de leurs données de recherche, pour autant que leur travail soit protégé et reconnu, et que la prise en main de l'outil leur soit aisée. Nous avons également remarqué que, même si une plateforme de dépôt de données de recherche s'articule autour de trois fonctionnalités de base apparemment simples – à savoir la recherche, le dépôt et la gestion des jeux de données –, les pages qui y correspondent nécessitent un travail plus important que les autres, car elles comportent les plus grands enjeux en termes d'expérience utilisateur.

Par ailleurs, l'itération de tests utilisateurs nous a permis de confronter nos idées, de valider certains aspects de la maquette et d'en interroger d'autres susceptibles d'induire des problèmes d'utilisabilité. La plupart du temps, les solutions étaient trouvées en travaillant les terminologies ou la navigation. Nous souhaitons insister sur le fait que, sans les rencontres répétées avec les utilisateurs potentiels de Yareta, nous aurions pu copier des plateformes existantes et respecter les bonnes pratiques, mais qu'en aucun cas nous n'aurions pu appliquer les méthodes de design centré sur l'utilisateur et construire une maquette de plateforme qui soit au plus proche des besoins et attentes du public cible. Les graphes portant sur l'évaluation de l'utilisabilité des maquettes par nos répondants (différents pour chaque test) prouvent d'ailleurs que la phase itérative est à l'origine de l'amélioration du sentiment des usagers par rapport au prototype. En effet, les utilisateurs de ce type de plateformes n'ont pas toujours la même logique ou le même fonctionnement que leurs concepteurs ; aussi, leurs retours sont d'un intérêt précieux.

Enfin, nous avons mentionné plus haut que nous nous étions concentrées sur la collecte de données qualitatives. Toutefois, si nous en avons à nouveau l'occasion, nous aurions aussi prévu une collecte de données quantitatives en réalisant des tests utilisateurs à distance, susceptibles de mobiliser un nombre significatif de participants. De plus, nous aurions utilisé le questionnaire du *System Usability Scale*. Ces deux modifications méthodologiques nous auraient permis d'accorder plus de crédit à l'évaluation de l'utilisabilité de la maquette par les testeurs. Au terme de ce mandat, nous suggérons au DLCM de continuer à recourir aux tests utilisateurs lors du développement de Yareta, puis d'Olos.

Bibliographie

AUSTIN, Claire C., et al., 2016. Research Data Repositories: Review of Current Features, Gap Analysis, and Recommendations for Minimum Requirements. *IASSIST Quarterly* [en ligne]. [Consulté le 31 juillet 2019]. Publié le 8 avril 2016. Disponible à l'adresse : <https://iassistdata.org/conferences/2015/presentation/7042>.

BATHELOT, Bertrand 2018. Persona en marketing. *Définitions Marketing* [en ligne]. [Consulté le 8 janvier 2020]. Disponible à l'adresse : <https://www.definitions-marketing.com/definition/persona-en-marketing/>.

BORGMAN, Christine L., 2012. The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology* [en ligne]. Publié le 5 avril 2012. Vol. 63, No 6, pp. 1059–1078. [Consulté le 19 août 2019]. Disponible à l'adresse : <https://doi.org/10.1002/asi.22634>.

BURGI, Pierre-Yves, 2018. *(Help) Manage research data: it's (still) worth it!* [document PowerPoint]. Mis à jour en 2018.
Support de cours : Gestion des données de recherche, Haute école de gestion de Genève, 2018.

CAZEAUX, Hugues, 2018. *Solution requirements* [document PDF]. Publié le 29 octobre 2018, 46 p.
Document interne au DLCM.

CARVALHO AMORIM, Ricardo, ROCHA, João, AGUIAR CASTRO, João, RIBEIRO, Cristina, 2016. A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Universal Access in the Information Society* [en ligne]. Publié en juin 2016. [Consulté le 2 janvier 2020]. Disponible à l'adresse : https://www.researchgate.net/publication/303918099_A_comparison_of_research_data_management_platforms_architecture_flexible_metadata_and_interoperability.

COMMISSION EUROPEENNE, 2019. What is Horizon 2020? *European Commission* [en ligne]. Mis à jour en 2019. [Consulté le 25 août 2019]. Disponible à l'adresse : <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>.

CROSAS, Mercè, 2011. The Dataverse Network®: An Open-Source Application for Sharing, Discovering and Preserving Data. *D-Lib Magazine* [en ligne]. Janvier-février 2011. Vol 17, No 1/2. [Consulté le 05 août 2019]. Disponible à l'adresse : <https://dx.doi.org/10.1045/january2011-crosas>.

DABI-SCHWEBEL, Gabriel, 2018. Open-source (logiciel). *1 min 30: enjoy marketing* [en ligne]. Publié en 2018. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.1min30.com/dictionnaire-du-web/open-source-logiciel>.

DAUMAL, Sylvie, 2018. *Design d'expérience utilisateur : Principes et méthodes UX*. 3^e édition. Paris : Eyrolles. ISBN 978-2-212-67456-9.

DATA CITE, 2019a. Missions. DataCite.org [en ligne]. [Consulté le 2 janvier 2020]. Disponible à l'adresse : <https://datacite.org/mission.html>.

DATA CITE, 2019b. DataCite Metadata Schema: Documentation for the Publication and Citation of Research Data. *Schema.DataCite.org* [en ligne]. [Consulté le 2 janvier 2020].

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Disponible à l'adresse : https://schema.datacite.org/meta/kernel-4.3/doc/DataCite-MetadataKernel_v4.3.pdf.

DATAVERSE PROJECT, 2017. A Comparative Review of Various Data Repositories. *dataverse.org* [en ligne]. Publié le 25 juillet 2017. [Consulté le 2 janvier 2020]. Disponible à l'adresse : <https://dataverse.org/blog/comparative-review-various-data-repositories>.

DATAVERSE PROJECT, 2019. The Dataverse Project. *dataverse.org* [en ligne]. Mis à jour en 2019. [Consulté le 2 janvier 2020]. Disponible à l'adresse : <https://dataverse.org>.

DLCM, 2019a. *DLCM.ch* [en ligne]. [Consulté le 20 août 2019]. Disponible à l'adresse : <https://www.dlcm.ch/>.

DLCM, 2019b. DLCM Integration Guide. *sandbox.dlcm.ch* [en ligne]. Mis à jour le 17 décembre 2019. [Consulté le 9 janvier 2020]. Disponible à l'adresse : <https://sandbox.dlcm.ch/administration/docs/DLCM-IntegrationGuide.html#organizational-unit> (accès par abonnement).

DRYAD, 2019. *Dryad* [en ligne]. Mis à jour en 2020. [Consulté le 2 janvier 2020]. Disponible à l'adresse : <https://datadryad.org/stash>.

DUBOIS, James M., STRAIT, Michelle, WALSH Heidi, 2018. Is It Time to Share Qualitative Research Data? *Qualitative Psychology* [en ligne]. Publié en 2018. Vol. 5, No 3, pp. 380-393. [Consulté le 18 août 2019]. Disponible à l'adresse : <https://psycnet.apa.org/record/2017-12030-001>.

FIGSHARE, 2020. Figshare.com [en ligne]. Mis à jour en 2020. [Consulté le 2 janvier 2020]. Disponible à l'adresse : <https://figshare.com>.

FONDS NATIONAL SUISSE, 2015. Règlement des subsides. *SNF.ch* [en ligne]. Publié le 27 février 2015. [Consulté le 20 août 2019]. Disponible à l'adresse : http://www.snf.ch/fr/encouragement/documents-telechargements/Pages/reglement-reglement-des-subsides.aspx#br_a_47.

FONDS NATIONAL SUISSE, 2019a. Open Access to Publications. *SNF.ch* [en ligne]. [Consulté le 20 août 2019]. Disponible à l'adresse : <http://www.snf.ch/fr/leFNS/points-de-vue-politique-de-recherche/open-access/Pages/default.aspx>.

FONDS NATIONAL SUISSE, 2019b. Open Research Data. *SNF.ch* [en ligne]. [Consulté le 20 août 2019]. Disponible à l'adresse : http://www.snf.ch/fr/leFNS/points-de-vue-politique-de-recherche/open_research_data/Pages/default.aspx.

GARRETT, Jesse James, 2011. Les éléments de l'expérience utilisateur : placer l'utilisateur au cœur de la conception des produits web et mobiles. 2e édition. Paris : Pearson. ISBN 978-2-7440-2485-6.

GRAND CONSEIL DE LA REPUBLIQUE ET CANTON DE GENEVE, 2017. Projet de loi ouvrant un crédit au titre de subvention cantonale d'investissement de 15 546 000 F pour financer l'évolution des infrastructures et services associés de calcul à haute performance et de stockage long terme afin d'optimiser et faciliter l'utilisation des données issues de la recherche dans les hautes écoles universitaires genevoises (PL 12146) [document PDF]. *Grand Conseil* [en ligne]. Publié le 21 juin 2017. [Consulté le 3 décembre 2019]. Disponible à l'adresse : <http://ge.ch/grandconseil/data/texte/PL12146.pdf>.

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

GREENBERG, Jane, et al., 2009. A metadata best practice for a scientific data repository. *Journal of Library Metadata* [en ligne]. December 2009, Vol 9, Issue 3-4, pp. 194-212. [Consulté le 14 août 2019]. Disponible à l'adresse : <https://doi.org/10.1080/19386380903405090>.

G2, 2020. figshare pricing. *G2.com* [en ligne]. Mis à jour en 2020. [Consulté le 2 janvier 2020]. Disponible à l'adresse : <https://www.g2.com/products/figshare/pricing>.

HALL, Erika, 2015. *La phase de recherche en Web Design*. Paris : Eyrolles. A book apart, n°9. ISBN 978-2-212-14146-7.

HARVARD DATAVERSE, 2020. *Harvard Dataverse* [en ligne]. Mis à jour en 2020. [Consulté le 2 janvier 2020]. Disponible à l'adresse : <https://dataverse.harvard.edu>.

HE, Lin, NAHAR, Vinita, 2016. Reuse of scientific data in academic publications: An investigation of Dryad Digital Repository. *Aslib Journal of Information Management* [en ligne]. Publié en 2016. Vol. 68, No. 4, pp. 478-494. [Consulté le 16 août 2019]. Disponible à l'adresse : <https://www.emerald.com/insight/content/doi/10.1108/AJIM-01-2016-0008/full/html>.

HEG, 2019. *M4 – Propositions de projets de recherche* [document PDF]. Mis à jour en février 2019.

Support de cours : Projet de recherche, Haute école de gestion de Genève, 2019.

JAMBE, Carmen, 2015. *La gestion des données de recherche à l'Université de Lausanne : enjeux transdisciplinaires* [document PDF]. Genève : Haute école de gestion de Genève. Travail de bachelor. [Consulté le 13 août 2019]. Disponible à l'adresse : <http://doc.rero.ch/record/258023>.

KIM, Suntae, 2018. Functional Requirements for Research Data Repositories. *International Journal of Knowledge Content Development & Technology* [en ligne]. Mars 2018. Vol 8, No 1, pp. 25-36. [Consulté le 02 juillet 2019]. Disponible à l'adresse : <http://ijkcddt.net/xml/13281/13281.pdf>.

KIM, Youngseek et ADLER, Melissa, 2015. Social scientist's data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *International Journal of Information Management* [en ligne]. Mai 2015, Vol 35, pp. 408-418. [Consulté le 25 juillet 2019]. Disponible à l'adresse : <https://dx.doi.org/10.1016/j.ijinfomgt.2015.04.007> (accès par abonnement).

KRAKER, Peter, LEX, Elisabeth, GORRAIZ, Juan et al., 2015. Research Data Explored II: the Anatomy and Reception of figshare. *arxiv.org* [en ligne]. Publié le 26 juin 2015. [Consulté le 18 août 2019]. Disponible à l'adresse : <https://arxiv.org/pdf/1503.01298.pdf>.

KRAUSE, Edward M., CLARY, Erin, OGLETREE, Adrian, GREENBERG, Jane, 2015. Evolution of an Application Profile: Advancing Metadata Best Practices through the Dryad Data Repository. *DCMI International Conference on Dublin Core and Metadata Applications, Sao Paulo, 2015* [en ligne]. [Consulté le 18 août 2019]. Disponible à l'adresse : <https://dcpapers.dublincore.org/pubs/article/view/3763>.

KRUG, Steve, 2012. *Don't make me think.fr : optimisez la navigation de vos sites*. Paris : Pearson. ISBN 978-2-7440-2552-5.

Loi fédérale du 26 juin 1998 sur l'archivage (LAr ; RS 152.1). *Le portail du Gouvernement suisse* [en ligne]. Publié le 26 juin 1998. Mis à jour le 1er mai 2013. [Consulté le 20 août 2019].

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Disponible à l'adresse : <https://www.admin.ch/opc/fr/classified-compilation/19994756/index.html>.

MARTIN, Bella, HANINGTON, Bruce, 2013. *100 méthodes de design*. Paris : Eyrolles, 2013. ISBN 9782212135626.

MILHIT, Igor, 2012. *Enjeux de l'archivage à long terme des données primaires de la recherche scientifique : implications pour le GIPDIR* [document PDF]. Genève : Haute école de gestion de Genève. Travail de bachelor. [Consulté le 11 août 2019]. Disponible à l'adresse : <http://doc.rero.ch/record/30352>.

ORGANISATION DE COOPERATION ET DE DEVELOPPEMENT ECONOMIQUES, 2007. *Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics* [en ligne]. Paris : Les Éditions de l'OCDE, 2017, 29 pages. [Consulté le 20 août 2019]. Disponible à l'adresse : <http://www.oecd.org/fr/sti/inno/38500823.pdf>.

PARK, Junh-Ran, 2009. Metadata Quality in Digital Repositories: A Survey of the Current State of the Art. *Cataloging & Classification Quarterly* [en ligne]. Vol. 47, No 3-4, pp. 213-228. [Consulté le 13 août 2019]. Disponible à l'adresse : <https://doi.org/10.1080/01639370902737240> (accès par abonnement).

PETERS, Isabella, et al, 2017. Zenodo in the Spotlight of Traditional and New Metrics. *Frontiers in Research Metrics and Analytics* [en ligne]. Décembre 2017, Vol. 2, No 13. [Consulté le 26 août 2019]. Disponible à l'adresse : <https://www.frontiersin.org/articles/10.3389/frma.2017.00013/full>.

RAPHAEL, 2011. SUS pour System Usability Scale. *Le bloc-notes, UX & Design d'expérience utilisateur* [en ligne]. Publié le 18 novembre 2011. [Consulté le 12 janvier 2020]. Disponible à l'adresse : <https://blocnotes.iergo.fr/concevoir/les-outils/sus-pour-system-usability-scale/>.

REED, Robin B., 2017. Figshare for Institutions. *Journal of the Medical Library Association* [en ligne]. Vol. 104, No 4. ISSN 1558-9439. [Consulté le 27 août 2019]. Disponible à l'adresse : <https://doi.org/10.5195/jmla.2016.167>.

RERO, 2019. *Réro explore*. [en ligne]. [Consulté le 10 janvier 2020]. Disponible à l'adresse : https://explore.rero.ch/fr_CH/rero.

ROUSIDIS, Dimitris et al., 2014. Data Quality Issues and Content Analysis for Research Data Repositories: The Case of Dryad. In: *Let's Put Data to Use: Digital Scholarship for the Next Generation*. Proceedings of the 18th International Conference on Electronic Publishing [en ligne]. pp. 49-58. [Consulté le 30 juillet 2019]. Disponible à l'adresse : <https://doi.org/10.3233/978-1-61499-409-1-49>.

SAYOGO, Djoko Sigit et PARDO, Theresa A., 2013. Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly* [en ligne]. Janvier 2013, Vol 30, supp 1, pp. s19-s31. [Consulté le 29 juillet 2019]. Disponible à l'adresse : <https://doi.org/10.1016/j.giq.2012.06.011> (accès par abonnement).

SCHERER, David, ZILINSKI, Lisa, VALEN, Dan, 2018. Balancing Multiple Roles of Repositories: Developing a Comprehensive Institutional Repository at Carnegie Mellon University [document PowerPoint]. In: 2018 Open Repositories Conference, Montana State University, 5 juin 2018. Mis à jour le 5 juin 2018. [Consulté le 18 août 2019]. Disponible à l'adresse :

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

https://kilthub.cmu.edu/articles/Balancing_Multiple_Roles_of_Repositories_Developing_a_Comprehensive_Institutional_Repository_at_Carnegie_Mellon_University/6448013.

SCHNEIDER, René, 2018a. *User-centered design : orientation - navigation* [document PowerPoint]. Mis à jour en mars 2018.

Support de cours : User-centered design, Haute école de gestion de Genève, 2018.

SCHNEIDER, René, 2018b. *User-centered design : Design patterns* [document PowerPoint]. Mis à jour en mars 2018.

Support de cours : User-centered design, Haute école de gestion de Genève, 2018.

SCHNEIDER, René, 2018c. *Du plancton et des baleines : dépôts de données* [document PowerPoint]. Mis à jour en novembre 2018.

Support de cours : Gestion des données de recherche, Haute école de gestion de Genève, 2018.

SCOTT, Bill, NEIL, Theresa, 2010. *Interface web interactives : 75 modèles de design pour concevoir des applications web riches*. Paris : Pearson, 2010. ISBN 978-2-7440-2398-9.

SERRES, Alexandre, 2017. *Données de la recherche en SHS : Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2* [document PDF]. Rennes : Université Rennes 2, 2017. [Consulté le 13 août 2019]. Disponible à l'adresse : <https://digitallibrary.hypotheses.org/1042>.

SICILIA, Miguel-Angel, GARCIA-BARRIOCANAL, Elena, SANCHEZ-ALONSO, Salvador, 2017. Community curation in open dataset repositories: insights from Zenodo. *Procedia Computer Science* [en ligne]. Vol 106, pp. 54-60. [Consulté le 13 août 2019]. Disponible à l'adresse : <https://doi.org/10.1016/j.procs.2017.03.009>.

SPRINGER NATURE, 2020. Recommended Data Repositories. *Nature.com* [en ligne]. Mis à jour en 2020. [Consulté le 2 janvier 2020]. Disponible à l'adresse : <https://www.nature.com/sdata/policies/repositories>.

TAN, Vicheth Jérémy, 2018. Rapport de stage : Conception du prototype d'un portail de gestion des données de recherche dans le cadre du projet national DLCM [document PDF]. 2ème version. Mis à jour en 2018. Document interne à l'Université de Genève.

TENOPIR, Carol et al., 2011. Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* [en ligne]. Juin 2011, Vol 6, No 6. [Consulté le 14 août 2019]. Disponible à l'adresse : <https://doi.org/10.1371/journal.pone.0021101>.

THELWALL, Mike, KOUSHA, Kayvan, 2015. Figshare: a universal repository for academic resource sharing? *Online Information Review* [en ligne]. Publié le 18 décembre 2015. Vol. 40 No. 3, pp. 333-346. [Consulté le 17 août 2019]. Disponible à l'adresse : https://www.researchgate.net/publication/287640807_Figshare_A_universal_repository_for_academic_resource_sharing.

UNIVERSITE DE GENEVE, 2019a. e-Research : Le programme. *Université de Genève* [en ligne]. [Consulté le 3 décembre 2019]. Disponible à l'adresse : <https://www.unige.ch/eresearch/fr/propos/le-programme/>.

UNIVERSITE DE GENEVE, 2019b. *Yareta: The Research Data Repository of Geneva's Higher Education Institutions* [en ligne]. [Consulté le 2 janvier 2020]. Disponible à l'adresse : <https://yareta.unige.ch>.

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

WALTER, Aaron, 2011. *Design émotionnel*. Paris : Eyrolles. A book apart, n°5. ISBN 978-0-937557-3.

WHITE, Ethan P., 2013. Nine simple ways to make it easier to (re)use your data. *Ideas in Ecology et Evolution* [en ligne]. Août 2013. Vol 6, No 2, pp. 1-10. [Consulté le 05 août 2019]. Disponible à l'adresse : <https://dx.doi.org/10.4033/iee.2013.6b.6.f>.

WIKIPEDIA, 2004. Licence Creative Commons. *Wikipédia* [en ligne]. Publié le 3 mars 2004. Mis à jour le 30 octobre 2019. [Consulté le 9 janvier 2020]. Disponible à l'adresse : https://fr.wikipedia.org/wiki/Licence_Creative_Commons.

WIKIPEDIA, 2006. Libre accès (édition scientifique). *Wikipédia* [en ligne]. Publié le 27 juillet 2006. Mis à jour le 2 juillet 2019. [Consulté le 24 août 2019]. Disponible à l'adresse : [https://fr.wikipedia.org/wiki/Libre_accès_\(édition_scientifique\)](https://fr.wikipedia.org/wiki/Libre_accès_(édition_scientifique)).

WIKIPEDIA, 2009. Science ouverte. *Wikipédia* [en ligne]. Publié le 28 novembre 2009. Mis à jour le 10 juin 2019. [Consulté le 24 août 2019]. Disponible à l'adresse : https://fr.wikipedia.org/wiki/Science_ouverte.






WIKIPEDIA, 2011. Dryad (repository). *Wikipedia* [en ligne]. Publié le 14 avril 2011. Mis à jour le 8 novembre 2019. [Consulté le 2 janvier 2020]. Disponible à l'adresse : [https://en.wikipedia.org/wiki/Dryad_\(repository\)](https://en.wikipedia.org/wiki/Dryad_(repository)).

WIKIPEDIA, 2017a. Fair data. *Wikipédia* [en ligne]. Publié le 15 mai 2017. Mis à jour le 25 juillet 2019. [Consulté le 24 août 2019]. Disponible à l'adresse : https://fr.wikipedia.org/wiki/Fair_data.

WIKIPEDIA, 2017b. Plan de gestion des données. *Wikipédia* [en ligne]. Publié le 26 janvier 2017. Mis à jour le 11 décembre 2017. [Consulté le 24 août 2019]. Disponible à l'adresse : https://fr.wikipedia.org/wiki/Plan_de_gestion_des_données.

ZENODO, 2020. *Zenodo.org* [en ligne]. Mis à jour en 2020. [Consulté le 02 janvier 2020]. Disponible à l'adresse : <https://zenodo.org>.

Annexe 1 : Grille d'observation des plateformes existantes

Nom		 figshare	 DRYAD	 The Dataverse Project	 YARETA
Institution	Cern	MacMillan Group	National Evolutionary Synthesis Center	Harvard	DLCM-UNIGE
Domaine(s) couvert(s)	Tous	Tous	Essentiellement pour le domaine scientifique et médical	Tous (dans le monde entier)	Tous
Type(s) de données autorisé(s)	Tous (données de recherche et article)	Données de recherche inédites, même si résultat négatif (aucune données sensibles tolérées).	Données de recherche en tous genres (aucune données personnelles tolérées).		Tous
Volume autorisé	max 50 GB par dépôt	250MB-1GB par fichier et 1-20GB en tout par personne selon la formule. Max 1 TB	20 GB max pour tous les fichiers (sinon, frais supplémentaire)	Max 2GB par fichier	50 GB gratuits - pas de limites de taille
Formats autorisés	Tous	Tous, tant que la visualisation dans le navigateur est possible.	Tous, mais formats faciles à extraire encouragés.	Tous	Tous
Public cible	Priorité à l'Europe, mais pas de restriction	Institutions, éditeurs, chercheurs		Chercheurs, revues, institutions et développeurs du monde entier	Hautes écoles du canton de Genève
Prix	Free for individuals	0-15\$ par mois	120\$US pour la publication des données (20GB) + 50\$ par 10GB supplémentaire		100 CHF/TB/an pour deux répliques
Principe	Open source, construit sur la base d'Invenio (biblio numérique)	Cloud service	Open source	Open source	Open source
Licences autorisées	CC0 CC BY	Les licences Creative Commons sont encouragées, mais d'autres options sont possibles (FAQ).	CC0 ou licence compatible		CC BY, CC0, etc. (définies par le responsable de l'unité orga.)

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Fonctionnalités	L'interface web simple est complétée par une API riche qui permet aux outils et services tiers d'utiliser Zenodo comme backend dans leur workflow.	Il y a une API et un "desktop uploader".	Les données reçoivent un DOI. Il y a un DryadLab (modules éducatifs gratuits pour que les étudiants puissent s'engager dans la recherche en utilisant des données réelles). possibilité d'ajouter de nouvelles versions de fichiers de données afin d'effectuer des mises à jour ou des corrections. Possibilité de rendre les données disponibles pour une révision par les pairs si le manuscrit a été soumis à une revue partenaire participante.	Mise à disposition de bonnes pratiques, présence d'un blog.	Recherche. Dépôt. Gestion des unités organisationnelles, <i>dashboard</i>
Bandeau de navigation	Barre de recherche, Upload, Communities, My account	Browse, Barre de recherche, Activity, My Account	About, For Researchers, For organizations, Contact us, My Account	Search (barre de navigation à afficher), About, User Guide, Support, My Account	Recherche, Accès, Dépôts, Administration, Déconnexion
Page d'accueil	Recent uploads, les datas sets mis en avant. Pas d'options de tri	Bandeau contenant "my data, projects, collections, activity". Le chercheur/communauté mis en avant. Possibilité de Drag/Drop directement	Data sets ajoutés récemment, barre de recherche, "submit data now", tweets, stats, news,... Beaucoup d'informations, difficile de trouver rapidement ce que l'on veut.	Add Dataset/Dataverse, barre de recherche, tri par sujet, data sets ajoutés récemment, activité (stats).	En deux parties : la première est une page d'accueil générale donnant des informations générales et redirigeant vers d'autres sources d'informations ; la seconde est le portail, mais elle est en cours de développement.
Design (visibilité, lisibilité, rôle)	Bleu, Peu d'informations	Blanc, assez sobre.	Vert, Compact, peu lisible.	Blanc, sobre, pas de charte	Le vert ne permet pas d'établir un lien avec l'UNIGE

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

des images, convivialité)	en dehors des data sets, sobre.			graphique reconnaissable.	ou le DLCM, mais est en ligne avec le nom Yareta
Ergonomie (navigation, accueil, liens, sommaire, structure, recherche interne)				Il semble y avoir plusieurs plateformes ; distinction peu claire.	Fil d'Ariane sommaire, les termes du bandeau de navigation ne sont pas clairs.
Ajout d'un Data set: marche à suivre	Upload > new upload > Files (Drag/Drop or Choose Files), communities, upload type, basic information, license, funding, Identifiers, > Publish	My data > Drag/Drop or Browse for files > remplir les informations > publish > save changes	Submit data now > describe publication > Upload Data files > Review > Check out & submit	Add a dataset > citation metadata, files (Drag/Drop or select files) > Save dataset	Dépôts > nouveau dépôt > metadata > fichier (données primaires et secondaires) > Valider
Ajout d'un Data set: difficulté	Facilement compréhensible. Possibilité de sauvegarder pour terminer plus tard. Différents type de documents possible. Métadonnées comprises. Propose une classification des informations (required, recommended, optional). Pas de fil d'ariane	Facile d'ajouter un data set. Pas de catégorisation des informations, tout à la suite. Rond vert pour les informations indispensables. Différents type de documents possible. Métadonnées comprises, plus maigre que Zenodo. La case à cocher pour publier n'est pas très visible. Pas de fil d'ariane	Beaucoup de textes pour expliquer les étapes. Pas forcément accueillant. * pour obligatoire. Pas de Drag/Drop. Métadonnées comprises. Possibilité de joindre un ReadMe file. Fil d'ariane. Pas de confirmation avant la validation du data set. Facilité de se tromper...	Facile de trouver l'ajout de dataset. Tout sur une page pour les ajouts. Métadonnées comprises. * pour obligatoire. Lisible et clair	L'ajout d'un dataset n'est pas accessible directement. Ajout des fichiers dans une 2ème partie pas très clair. Formulaire classique et lisible.
Recherche d'un Data set	Barre de recherche page d'accueil, facette disponible, tri disponible, pas de recherche avancée. Liste	Barre de recherche page d'accueil, tri disponible, peu de facettes (type et licence), enregistrement de la recherche, icône, aide pour recherche plus avancée (dans barre de recherche simple).	Barre de recherche page d'accueil, facette disponible, tri disponible, recherche avancée (basique), liste, visuellement pas attractif.	Barre de recherche page d'accueil, facettes disponibles, tri disponible, recherche avancée, liste avec icône pour différencier type documents.	Pour le moment, la recherche ne se fait que par mot clé, sans possibilité de tri ou de recherche avancée. Les résultats sont décrits par leur unité organisationnelle, le titre du jeu de données, l'année de publication, ainsi que le niveau d'accès.

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

					Ensuite, une fois sur le jeu de données, plus de métadonnées sont disponibles, certaines en format XML.
Contenu (info institutionnelle, offre, logique de l'organisation, pertinence des liens, langues)	Informations institutionnelles, tout en anglais, FAQ clair et complet.	Information institutionnelle. Page d'accueil immense avec tout et n'importe quoi. Menu en bas de page. Super FAQ ! Existence d'un blog sur ???	Informations institutionnelles . Super FAQ !	Informations institutionnelles , info sur le fonctionnement . Pas de FAQ.	Français et anglais. Pas d'information sur le dépôt, hormis sur la page d'accueil, avant la connexion (bref lien avec UNIGE et DLCM). Les informations à donner sont encore à définir.
Accessibilité (référencement, popularité, compatibilité, vitesse de chargement, langues)	Dépôt référencé par les moteurs de recherche courant (Google, DuckDuckGo, Ecosia, Swisscows). Plateforme uniquement en anglais.	Dépôt référencé par les moteurs de recherche courant (Google, DuckDuckGo, Ecosia, Swisscows). Plateforme uniquement en anglais.	Dépôt référencé par les moteurs de recherche courant (Google, DuckDuckGo, Ecosia, Swisscows). Plateforme uniquement en anglais.	Dépôt référencé par les moteurs de recherche courant (Google, DuckDuckGo, Ecosia, Swisscows). Plateforme uniquement en anglais.	Le dépôt peut être trouvé par les moteurs de recherche, mais à travers des pages de l'Université de Genève. La vitesse de chargement des langues est rapide.
Création de compte	Email, MDP, pseudo. Validation d'email. Rapide	Nom, prénom, email, MDP. Validation email. Rapide	Email, puis validation. Nom, prénom, langue (EN uniquement), MDP, Lent	Inscription via OrCID ou Github possible. Pseudo, MDP, nom, prénom, email, affiliation, position. Validation email. Rapide	Le compte Yareta est lié à un compte AAI et SWITCHeduID. L'inscription à la plateforme est possible par l'intermédiaire de l'équipe de Yareta. Quant à l'inscription à une unité organisationnelle, elle se fait par le manager de l'unité en question.
Autres remarques	La recherche n'a pas besoin d'être publiée pour que les données puissent être déposées (protection). Toutes les métadonnées	Les données sont stockées aussi longtemps que le dépôt existe. Utilisation de DOI.	Les données reçoivent un DOI. Par défaut, les données sont sous embargo jusqu'à la publication de l'article. Les descriptions de	L'idée centrale derrière Dataverse est d'automatiser une grande partie du travail de l'archiviste professionnel, de fournir des services au	L'interface utilisateur est encore en cours de développement (basé sur technologie Angular pour assurer la convivialité).

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

	sont accessibles sous licence CC0, et tout le contenu ouvert est accessible via des API ouvertes. Chaque copie de fichier possède deux répliques situées sur des serveurs différents.		données doivent être en anglais.	créateur de données et de distribuer des crédits à ce dernier.	D'autre part, ce développement va tenir compte de cette présente étude pour améliorer l'ergonomie de l'outil.
--	---	--	----------------------------------	--	---

Annexe 2 : Formulaire de consentement pour la participation aux entretiens en français

Formulaire de consentement libre et éclairé

Soumis dans le cadre du projet de recherche :

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs à déposer leurs données

Dans le cadre du Master en Sciences de l'information de la Haute école de Gestion de Genève, le consortium DLCM a mandaté deux étudiantes, pour participer au développement du portail de dépôt des données de recherche du DLCM et, plus particulièrement, se concentrer sur les aspects ergonomiques et esthétiques de l'interface. Pour ce faire, il est nécessaire de récolter des données sur les besoins des chercheurs pour le stockage de leurs données de recherche. C'est pourquoi des entretiens individuels sont réalisés.

Le projet est réalisé par Marion Favre et Héroïse Hahling.

Je, soussigné, donne mon accord pour participer à l'étude susmentionnée, et accepte que les informations recueillies soient utilisées dans le cadre de ce projet de recherche, puis archivées dans un dépôt facilitant le partage des données de recherche. Je validerai la retranscription de l'entretien avant la publication du projet de recherche.

J'ai été informé-e que toutes les informations me concernant seront traitées de manière anonyme.

- ☐ J'accepte que l'entretien soit enregistré et archivé sous une forme anonymisée.
- ☐ J'accepte que l'entretien soit enregistré, mais je souhaite qu'il soit détruit une fois le verbatim validé.

Lieu et date :

Signature du participant :

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Annexe 3 : Guide d'entretien en français

Etude des besoins utilisateurs

Guide d'entretien

Introduction

Le consortium DLCM, dont le but est d'aider les chercheurs dans la gestion de leurs données, souhaite compléter sa gamme de services avec une infrastructure de préservation à long terme des données de recherche. Bien que le DLCM collabore avec une équipe de professionnels de l'Université de Genève capable d'assumer la dimension technique du projet, nous avons été mandatées en tant qu'étudiantes du Master en Sciences de l'Information dispensé à la Haute école de gestion de Genève pour enquêter sur les besoins des utilisateurs d'une telle plateforme et d'en gérer les aspects ergonomiques, comme esthétiques.

Questions préliminaires

Intitulé du projet :	Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs à déposer leurs données		
NOM, prénom		Date	
Sexe		Âge	
Institution/Faculté		Statut	

Entretien

N°	Question
Généralités/Introduction	
1.1	Êtes-vous familier du processus d'archivage de données de la recherche ? Pourquoi ?

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

1.2	Distinguez-vous l'archivage de vos articles de celui de vos données de recherche ?
1.3	Différenciez-vous l'archivage à court, moyen et long terme ?
Données	
2.1	Quels types de données produisez-vous ?
2.2	En général, quelles tailles font vos données (fourchette) ?
2.3	Selon vous, combien de temps un dépôt de données à long terme devrait-il conserver les données ?
Stockage/Archivage des données de recherche	
3.1	Archivez-vous vos données de recherche ? (Pourquoi ?)
3.1.1	Stockez-vous vos données à court/moyen terme (format, outil, dépôt, temps) ?
3.1.2	Stockez-vous vos données à long terme (format, outil, dépôt, temps) ? (Pourquoi ?)
3.1.3	Avez-vous déjà stocké des sets de données sur un dépôt ? Si oui, pourriez-vous nous raconter comment vous avez procédé ?
3.1.4	Quelles seraient les éventuelles remarques que vous pourriez faire à propos du dépôt que vous aviez alors utilisé (nom/avantages/inconvénients) ?
3.1.5	Quelles sont les difficultés que vous avez rencontré durant l'archivage de vos données dans un dépôt ?
3.2	Lors d'un travail de recherche, à quel moment vous interrogez-vous pour la première fois sur l'archivage des données de la recherche ?
Cadre	

4.1	Réalisez-vous un DMP (<i>Data Management Plan</i> , ou plan de gestion des données) dans le cadre de vos projets ?
4.2	Votre institution vous impose-t-elle une politique, une procédure, un dépôt, ou quoi que ce soit par rapport au dépôt de vos données ?
Besoins et attentes	
5.1	Qu'attendez-vous d'un dépôt de données de la recherche ?
5.1.1	Quelles sont les fonctionnalités indispensables pour vous ? Que doit vous permettre de faire l'interface de dépôt ?
5.1.2	Pour vous, quelle importance (rôle) ont : <ul style="list-style-type: none"> • Le fil d'Ariane ? • Le design (sobriété ? couleur ?) ? • Les aides contextuelles ?
5.1.3	Selon vous, comment doivent être les fonctionnalités de : <ul style="list-style-type: none"> • dépôt ? • recherche ? • gestion du compte ? Comment les imaginez-vous ?
5.2	Si besoin, montrer Figshare et Zenodo et demander l'impression quant à l'esthétisme et aux trois fonctions clés (dépôt, recherche, gestion du compte).

Conclusion

Avez-vous des commentaires à rajouter ou des questions au terme de cet entretien ?

Remerciements

Nous arrivons au terme de cet entretien et vous remercions chaleureusement pour votre participation. Nous vous transmettrons le verbatim de cet entretien prochainement, afin que vous puissiez le valider avant son utilisation dans la recherche et son archivage.

Annexe 4 : Entretien 1

L'entretien dure 67 minutes. Il est enregistré, afin de faciliter sa transcription.

Questions préliminaires

Intitulé du projet:	Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs à déposer leurs données		
		Date	05.07.2019
Sexe	Homme (1)	Âge	50
Institution/Faculté	Université de Genève Section de physique		
Sexe	Homme (2)	Âge	38
Institution/Faculté	Université de Genève Section de physique		

Entretien

N°	Question
Généralités/Introduction	
1.1	Êtes-vous familier du processus d'archivage des données de la recherche ? Pourquoi ?
	<i>(1) La démarche d'archivage des données de recherche ne vient pas des chercheurs qui ont plutôt tendance à le faire à reculons. Il faut donc que l'outil soit attractif et le processus simple et rapide. En ce qui me concerne, oui, je pense savoir ce dont il s'agit. J'ai rédigé le data management plan pour la prochaine requête adressée au Fonds national suisse par mon chef de groupe de recherche (groupe de 10-12 personnes). Je n'ai pas en tête tous les détails de ce que le Fonds national attend, mais il me semble que ce n'est pas non plus très clair pour lui.</i>
1.2	Distinguez-vous l'archivage de vos articles de celui de vos données de recherche ?
	<i>(1) Oui, ce sont deux choses très différentes. Nous faisons déjà l'archivage de nos articles depuis longtemps. En revanche, l'archivage des données de recherche dans l'optique de les rendre accessibles au public est quelque chose de nouveau.</i>
1.3	Différenciez-vous l'archivage à court, moyen et long terme ?
	<i>(1) Je ne peux parler que personnellement. Pour le moment, l'archivage de mes données de recherche personnelles est laissé à la responsabilité des chercheurs.</i>

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

	<p><i>Dans le groupe de recherche dont je fais partie, il n'y a pas de système d'archivage centralisé. Nous y avons réfléchi dans le cadre de la rédaction du DMP dont je vous parlais, nous sommes en train de mettre en place une solution dans ce sens. Le problème qui se pose est que nous sommes confrontés à des données qui sont remarquablement volumineuses (de l'ordre du terabyte) et nous ne pouvons pas envisager de stocker ces volumes pour de longues durées. Ça ne fait pas beaucoup de sens dans la mesure où ces données peuvent être régénérées par du calcul dans un intervalle de temps raisonnable. Dans ces conditions, il faut archiver le programme qui permet de générer les données.</i></p> <p><i>Donc, pour ce qui est des données de taille raisonnable qui sont utilisées pour rédiger les publications scientifiques, l'archivage est laissé à la responsabilité des chercheurs. Dans mon cas, c'est archivé sur des backups démultipliés ; c'est a priori stable, mais ce n'est pas dans un système institutionnalisé. On considère qu'un chercheur doit être capable de fournir les éléments qui permettent de reconstituer ses données ou de refaire ses calculs au moins à l'échelle de dix ans. Donc, pour le moment, le chercheur conserve lui-même ses données de recherche pendant dix ans en les fournissant sur demande uniquement. Maintenant, sous la pression du Fonds national, on met en place un système sur le serveur de l'Université de Genève, sur lequel il nous sera demandé de déposer nos données. Un lien DOI sera mis dans la publication pour que les données soient accessibles. Cependant, le choix sur les données archivées et les données non archivées est discuté au sein de notre groupe de recherche ; c'est une décision qui doit être prise au cas par cas en fonction des données et des possibilités.</i></p> <p><i>Au sein de notre groupe, nous avons établi une liste de ce que l'on veut stocker pour chaque publication et je suis chargé de tester cette idée et de courir après les chercheurs pour qu'ils remplissent leur part de travail en termes d'archivage des données de recherche. Nous sommes encore en pleine exploration, afin d'affiner notre système. Notre idée de base est de faire un stockage local au sein du groupe de recherche ; une partie des données stockées restera privée et une partie sera préparée pour la publication dans une solution institutionnelle (Yareta). Par contre je n'ai aucune idée de la durée de conservation des données sur cet outil.</i></p>
Données	
2.1	Quels types de données produisez-vous ?
	<p><i>(1) Je fais partie d'un groupe de physique théorique, même si notre département compte essentiellement des groupes de physique expérimentale pour lesquels mes réponses ne sont pas représentatives. Les groupes expérimentaux font des mesures qui constituent leurs données de recherche. En tant que groupe théorique, nous écrivons un code qui permet de générer des données numériques binaire (format que l'on lit avec un code qui doit être livré avec les données) ou des données en chiffre qui sont stockées dans un fichier en format texte autodocumenté (mention du type de données, des unités, etc.). En physique, il n'y a aucun consensus sur les formats de données ; le système basique est ASCII. Certains chercheurs utilisent des logiciels propriétaires (ex : Mathematica, MATHLAB, etc.) qui ont parfois des formats propriétaires. Dans ce cas, les données sont ce qu'elles sont et on ne peut pas en changer le format.</i></p>
2.2	En général, quelles tailles font vos données (fourchette) ?

(1) Tout à l'heure, je vous disais qu'on générerait des données dont le volume était de l'ordre du terabyte, mais ce n'est pas ce qui sera finalement archivé. Tous les projets, tous les processus de recherche ne mènent pas à des volumes pareils ! Il y a de petits projets, pour lesquels on a un code qui génère un nombre, même si le code peut tourner pendant trois semaines. La donnée ainsi obtenue n'est pas difficile à stocker. D'autres types de calcul permettent de générer une quantité de fichiers qui représentent des grandeurs physiques ; ces fichiers, qui peuvent être très grands, sont utilisés comme inputs pour faire du processing (analyse), dont le résultat sera publié dans un article. Dans ce cas, les données sont trop volumineuses pour être archivées ; ce ne serait pas réaliste. Nous les conserverons nous-mêmes le temps de les analyser, ou un peu plus longtemps si on en a la place, mais elles ne seront pas gardées dix ans. Par contre, nous avons pris la décision d'archiver toutes les données qui apparaissent dans une publication. Dans notre groupe de recherche ou dans le département de physique, il n'y a pas de consensus pour archiver tout le processus de génération des données simplement parce que ce n'est pas réaliste. Quand j'écris un code, je le compile, j'utilise des bibliothèques qui ne m'appartiennent pas ; dois-je les archiver ? Les bibliothèques sont-elles compilées avec un compilateur que je devrais également archiver ? Le compilateur tourne sur un système d'exploitation ; dois-je archiver le système d'exploitation ? On se rend compte que l'on est incapable d'archiver l'intégralité de la chaîne qui mène à une donnée. On essaie donc d'archiver de manière raisonnable.

De mon point de vue, il y a un problème de principe à archiver de manière publique un code ; car le code naît d'un savoir-faire. Ainsi, le DMP que j'ai rédigé précise que nos codes seront archivés de manière privée au moment de la publication. Par exemple, un chercheur en astronomie dira qu'il publie toutes ses images, mais il ne publiera pas les réglages du télescope qui font que celui-ci donne des meilleures images qu'un autre. Ce savoir-faire fait partie de la propriété intellectuelle du chercheur et il n'est pas supposé être divulgué. De la même manière, si l'on passe un an à étudier les techniques présentées dans un bouquin, le savoir-faire acquis sera stocké dans notre cerveau et personne ne pourra nous l'enlever. Par contre, si l'on étudie un algorithme public, il faut beaucoup de travail pour l'intégrer dans un programme qui fonctionne (analogie avec le contenu du livre) ; forcer le chercheur à divulguer son savoir-faire le dépossède de son travail. Nous avons des doctorants qui investissent beaucoup d'énergie pour réussir à comprendre un algorithme et le traduire dans un code ; ensuite, ils font un calcul qu'ils publient. Si on leur demande de publier le code avec la publication, un autre professeur qui n'a pas le temps d'écrire un code, mais qui est plus avancé pour identifier les problèmes importants, il pourra prendre le code, améliorer le calcul et scier le travail du doctorant. Je trouve qu'il y a quelque chose d'injuste dans cette exigence et il me semble que cet avis est partagé par les personnes qui développent des codes à l'Université de Genève. Après, quand la méthode est avancée et connue de tout le monde, on peut envisager publier les codes en Open Source et partager l'outil. Mais comme nos codes sont souvent à l'avant-garde, la publication de nos codes n'est pas une option qui nous plaît beaucoup.

(2) Je trouve que, pour les données, c'est un peu pareil. Hormis les questions éthiques de données personnelles – dans la médecine notamment –, les données sont souvent complexes et pas traitables en une seule fois. Ainsi, les chercheurs ont tendance à reprendre les données en plusieurs fois pour s'intéresser à différents aspects et publier plusieurs articles sur la base d'un même jeu de données. C'est une raison pour laquelle les chercheurs ne sont pas très ouverts à publier – en tout cas rapidement – leurs données ; il faut leur laisser le temps de traiter ce qu'ils souhaitent traiter.

	<p><i>J'ai moi-même déjà eu l'occasion de réutiliser des données qui avaient été publiées sous forme d'article ; récupérer les points d'une courbe n'était pas facile et c'est vrai qu'avoir accès aux fichiers originaux aurait été agréable.</i></p> <p><i>(1) Voilà pourquoi, dans le DMP que j'ai rédigé, je propose à mon groupe de faire un service minimum en archivant les données qui sont dans les articles en format numérique. Le chercheur peut, s'il le souhaite, donner des informations complémentaires. Par exemple, il arrive que, dans un article, il soit dit que telle ou telle méthode est utilisée, mais qu'elle ne fonctionne pas pour x ou y raison ; les archives qui prouvent ce genre d'information pourrait également être publiées. Il est difficile d'être systématique sur ces questions et il faut faire confiance aux chercheurs.</i></p> <p><i>(2) C'est probablement aussi une question de temps. Il est demandé énormément de choses au chercheur qui doit publier, aller en conférence, enseigner, être dans des commissions, faire sa recherche, développer de nouveaux outils, suivre des doctorants. C'est bien plus qu'un travail à cent pourcent ! Donc, si on leur demande, en plus, de faire l'archivage de leurs données – donc, donner les données, leur associer des métadonnées, expliquer les conditions de récolte, etc. –, on leur ajoute un travail conséquent. Le travail d'archivage ne peut être réalisé que par le chercheur, puisque c'est lui qui a fait les mesures ; il est difficile de sous-traiter ces tâches. Après, comme il s'agit d'une demande du Fonds national, les gens vont être obligé de le faire, mais ils ne vont pas faire de zèle, juste le minimum pour obtenir des fonds. Ils ne vont pas passer deux mois à préparer leurs données pour l'archivage ; ce n'est pas possible. Le FNS ne va pas leur donner un budget supplémentaire pour l'archivage, mais si celui-ci peut être budgétisé.</i></p> <p><i>Pour les données expérimentales, tout ne peut pas être archivé, car certaines machines peuvent générer des données de l'ordre du terabyte. Par contre, la démarche sera similaire à celle décrite par (1), de manière à publier au moins les données publiées, ce qui réduit la quantité de données à quelques mégabytes.</i></p> <p><i>Ici, la plupart des chercheurs utilisent des outils scientifiques ou des logiciels open source pour créer leurs graphiques qui découlent souvent de données brutes (fichier texte avec des colonnes qui définissent les données). Si on leur demande d'archiver cela, ils archiveront sans doute les données brutes pas trop complexes.</i></p> <p><i>(1) Ces éléments relèvent de la responsabilité des groupes de recherche. Il y a deux points de vue : soit les données sont numériques, soit ce sont des images.</i></p>
2.3	<p>Selon vous, combien de temps un dépôt de données à long terme devrait-il conserver les données ?</p>
	<p><i>(1) Il faut être réaliste. Aller chercher des données qui ont cinquante ans pour les traiter est anecdotique. La recherche avance un peu trop vite pour que des archivages si longs soient généralisés. Après, bien sûr, il arrive que l'on cherche un papier de 1956, parce que les chercheurs de l'époque avaient mesuré certaines choses, que personne ne l'a refait depuis et que les données nous intéressent. Comme le disait (2), on regarde les courbes et on essaie de récupérer les données point par point, mais si cela nécessite une infrastructure ou un budget déraisonnable, c'est exagéré. Pour moi, pendant dix ans, il faut être capable de répondre à quelqu'un qui aurait des doutes sur nos données. [...]</i></p>

	<p>(2) Quand je suis allé voir la présentation de Yareta – à l’occasion de son inauguration –, il a été dit que la durée de conservation de base serait de dix ans. Il a aussi été dit que, si l’infrastructure était payée pour vingt ans, les sommes investies permettraient apparemment une conservation indéfinie, car, avec l’évolution des technologies, on pourra stocker beaucoup plus de données dans les vingt prochaines années. A priori, dix ans me semblent aussi raisonnables.</p> <p>(1) Dix ans, c’est la durée de deux thèses. (2) a fini sa thèse il y a dix ans, et on lui demande encore des informations sur ses données, peut-être pas au niveau du public, mais au niveau du groupe de recherche. A l’intérieur du groupe, nous n’avons aucune raison de ne pas garder les données que l’on archive ad vitam aeternam ; on dispose des technologies nécessaires pour les dupliquer et, si on doit changer de système d’archivage, on va certainement aussi les garder une fois le nouveau système en place. Je ne vois pas de nécessité de purger les données après tant d’années.</p>
Stockage/Archivage des données de recherche	
3.1	Archivez-vous vos données de recherche ? (Pourquoi ?)
	<p>(1) Oui. Une fois, j’ai dû le faire sur un serveur pour un dépôt de données. Mais, sinon, je fais toujours mon archivage personnel.</p> <p>(2) Oui, j’ai dû le faire une fois pour un article publié dans un journal, PLOS One, qui me demandait de le faire ; c’était une exigence de sa part. Le journal en question fournissait la solution : je leur donnais les données et il les archivait. [...]. Maintenant, je ne fais plus de recherche, mais, avant, quand ce n’était pas une exigence du Fonds national, je ne le faisais pas. Enfin, comme (2), j’archivais mes recherches de mon côté, car il arrive que des gens me demandent des données ou des informations sur des recherches antérieures ou que l’on doive réutiliser son propre matériel.</p>
3.1.1	Stockez-vous vos données à court/moyen terme (format, outil, dépôt, temps) ?
	Éléments déjà abordés. Les données sont stockées de manière privée sur le système personnel des chercheurs ou sur le système du groupe de recherche.
3.1.2	Stockez-vous vos données à long terme (format, outil, dépôt, temps) ? (Pourquoi ?)
	Éléments déjà abordés.
3.1.3	Avez-vous déjà stocké des sets de données sur un dépôt ? Si oui, pourriez-vous nous raconter comment vous avez procédé ?
	(1) Oui, nous l’avons évoqué tout à l’heure. Lorsque j’avais dû le faire, le journal dans lequel j’avais publié n’avait rien fait et ne proposait aucune solution ; il demandait juste à ce que l’archivage soit fait. Le serveur sur lequel j’avais archivé avait une interface relativement simple, mais je ne me souviens pas des détails. J’avais repris mes fichiers, je les avais renommés pour que les noms correspondent à ce qui figurait dans l’article. C’était relativement simple ; d’ailleurs, j’avais envoyé mes

	<p>données au premier auteur de l'article (fruit d'une collaboration européenne) pour qu'il les dépose, donc je n'avais même pas eu à déposer mes données moi-même.</p> <p>(2) Oui, nous l'avons également évoqué. Je ne sais pas si vous êtes familières des plateformes de publication d'articles scientifique ; si ce n'est pas le cas, il serait peut-être intéressant d'y jeter un œil. Elles sont toutes un peu différentes, mais il y a trois ou quatre grandes éditions qui ont leur propre système. Il faut rentrer toutes les informations personnelles et les autres éléments caractérisant l'article et il en va de même pour la partie concernant les données, que ce soit les figures ou les fichiers. Ainsi, l'interface que j'avais utilisée pour déposer mes données était assez basique, dans le sens où les informations demandées n'étaient pas standardisées. J'aurais pu déposer un fichier avec des nombres, sans aucune documentation, et ce serait certainement passé. C'est ennuyeux, car, s'il le demande, il ne sert à rien de donner des nombres si on ne sait pas ce dont il s'agit. En l'occurrence, chaque fichier devait être téléchargé séparément, mais il est parfois possible d'envoyer des fichiers zippés (compressés).</p>
3.1.4	<p>Quelles seraient les éventuelles remarques que vous pourriez faire à propos du dépôt que vous aviez alors utilisé (nom/avantages/inconvénients) ?</p> <p>(2) Si je compare avec ce qu'on m'a présenté de Yareta, j'ai trouvé très intéressant que Yareta utilise un système d'analyse de fichiers qui lui permet de savoir quels types de fichiers et de contenus sont déposés. Cela permet à l'utilisateur qui recherche des données de faire un tri plus précis. Ce sont des informations qui me semblent cruciales, car, finalement, la personne qui va envoyer ses données peut essayer de tricher ou se tromper.</p> <p>(1) Oui, par exemple, si le chercheur utilise un Mac, il y a parfois des fichiers que l'on ne voit pas et que l'on peut déposer par erreur au moment de la compression des fichiers. Dans ce processus de soumission, il devrait y avoir une étape où l'utilisateur voit de manière transparente ce qui va être stocké (type de fichier, taille, etc.), de manière à pouvoir corriger les éventuelles erreurs.</p>
3.1.5	<p>Quelles sont les difficultés que vous avez rencontré durant l'archivage de vos données dans un dépôt ?</p> <p>(2) Non. Dans les faits, ce n'est pas compliqué, mais ça prend un peu de temps.</p> <p>(1) Quand on nous a montré la première interface pour les archives ouvertes de l'Université de Genève pour que les chercheurs puissent y archiver leurs articles scientifiques – la démarche est obligatoire, mais il n'y a aucun contrôle –, l'interface était en français, alors que la plupart des chercheurs de l'Université ne sont pas francophones ; ils ne pouvaient donc pas utiliser cet outil. Donc, au moment où l'on réfléchit à une interface, il est absolument impératif de commencer par la faire en anglais. Les traductions faites après coup sont souvent incomplètes. Les systèmes bilingues doivent faire attention à ce que les traductions soient justes, qu'aucun malentendu ne soit possible et que toutes les pages soient traduites !</p>
3.2	<p>Lors d'un travail de recherche, à quel moment vous interrogez-vous pour la première fois sur l'archivage des données de la recherche ?</p> <p>(1) Quand on travaille sur une recherche, on ne s'occupe pas trop de ce genre de questions. Je pense que c'est au moment où l'on commence à remplir les formalités</p>

	<p><i>pour publier que cette problématique commence à apparaître. Ça vient assez tard... Bien sûr, il y a probablement des gens qui ont cette question en tête dès le début et qui vont organiser leurs données de manière à ce que la procédure d'archivage soit plus facile à la fin.</i></p> <p><i>Lorsque j'ai conçu le DMP pour mon groupe de travail, j'ai proposé des idées, une structure, que l'on est en train de tester, mais la procédure qui consiste à nourrir cette base de données vient après la publication. Je ne vais pas intervenir sur la façon dont les chercheurs gèrent leur travail, mais je leur présente ce qu'ils doivent me fournir à la fin (quoi, dans quel ordre, comment, etc.) ; soit ils s'orientent vers cette façon de fonctionner dès le début, soit ils se débrouillent pour que j'obtienne ce que j'ai demandé à la fin. Je pense qu'il va falloir toquer à la porte des chercheurs pour leur rappeler d'archiver leurs données et comment le faire – dans mon groupe, ce sera moi. Ça risque d'être pénible et je risque de devoir beaucoup insister. Donc, pour la plupart des chercheurs, la question va se poser tard ; ils auront leurs données, mais ils n'auront peut-être pas réfléchi au nommage. Chacun a sa méthode de travail avec ses avantages et ses inconvénients : on peut avoir un répertoire avec un nommage de fichier univoque ou pleins de sous-répertoire avec un nommage partiel. A mon avis, il devra y avoir un processus de transformation au moment de la publication en vue de produire l'information que l'on veut archiver.</i></p> <p><i>(2) Effectivement, entre la recherche et la publication – qui est le but de la recherche –, il y a énormément d'étapes. On récolte des données, on les traite, on en extrait des informations importantes, etc., et c'est à la fin, quand on a un bel article qui nous plaît, que l'on peut dire ce qui est pertinent à archiver. Avant, il y a tellement de données et de processus que, si les chercheurs devaient tout archiver, ils ne s'en sortiraient pas.</i></p> <p><i>(1) Oui ! En termes de quantité, c'est comme si l'on mettait le contenu des corbeilles à papier dans des classeurs, au cas où les déchets contiendraient une idée de génie ! Dans la recherche aussi, il y a aussi pleins de choses qui disparaissent.</i></p> <p><i>(2) Du coup, l'archivage, tel qu'il est demandé, va vraiment se faire à la fin, après la publication, même s'il faut que ce soit prêt avant la publication puisque les données sont souvent demandées lors de la publication de l'article. En même temps, ce n'est que lorsque la publication sera bien formatée et que toutes les parties prenantes la valideront que l'on va se poser la question de l'archivage.</i></p> <p><i>(1) Il faudra un travail assez fastidieux de la part des chercheurs pour qu'ils le fassent assez tôt et que le DOI des données puisse être donné lors de la publication de l'article. Si les journaux ne l'exigent pas, je sens que ça va être compliqué de demander que ce travail soit fait à temps pour que le DOI figure dans la publication ; le DOI risque d'être sur le site web de l'éditeur ou ailleurs et non sur la publication.</i></p>
Cadre	
4.1	Réalisez-vous un DMP (<i>Data Management Plan</i> , ou plan de gestion des données) dans le cadre de vos projets ?
	<i>(1) Oui, j'en ai écrit un, mais c'est encore un projet. La solution est en train d'être implémentée.</i>

	<p><i>(2) J'ai eu l'occasion de participer à des demandes de fonds adressées au FNS, mais je n'ai jamais rédigé un tel document lorsque je faisais de la recherche. En tant qu'informaticien, on va peut-être venir me demander des documents qui peuvent être intégrés à des demandes de fonds. Après, vu qu'une solution standardisée est en train de voir le jour [Yareta], les gens vont certainement commencer à partager ce genre de documents, puisqu'ils sont liés à une solution destinée à résoudre les mêmes problèmes pour tout le monde. Ce n'est pas un document qui va être rédigé par chaque groupe de recherche individuellement ; après, peut-être que certains éléments seront propres à chaque groupe, je ne sais pas.</i></p> <p><i>[...]</i></p>
4.2	<p>Votre institution vous impose-t-elle une politique, une procédure, un dépôt, ou quoi que ce soit par rapport au dépôt de vos données ?</p>
	<p><i>(1) Il y a certainement des recommandations en la matière, mais je ne sais pas. L'Université fonctionne si elle reçoit des fonds du FNS et, pour ça, elle doit satisfaire ses exigences en matière d'archivage des données de recherche, notamment. Après, comme l'Université vient d'investir des sommes conséquentes [dans Yareta], j'espère que c'est pour que ses chercheurs l'utilisent.</i></p> <p><i>(2) En ce qui concerne les archives ouvertes pour les articles, il me semble que la demande vient du rectorat. Peut-être qu'il compte imposer le même genre de règle pour les données ; je ne sais pas.</i></p> <p><i>(1) Comme les chercheurs se lancent dans l'archivage à reculons, ils ne lisent pas toujours les communications trop abondantes du rectorat. On sait ce qu'on doit faire, mais pas toujours d'où vient la contrainte.</i></p>
Besoins et attentes	
5.1	<p>Qu'attendez-vous d'un dépôt de données de la recherche ? Quelles sont les fonctionnalités indispensables pour vous ? Que doit vous permettre de faire l'interface de dépôt ?</p> <p>(Ex: fil d'Ariane pour suivre l'avancée de la procédure, design sobre pour ne pas interférer dans la démarche, beaucoup de couleur pour rendre l'activité plus vivante, des aides contextuelles, etc.)</p>
	<p><i>(1) Nous avons déjà parlé des questions de langues et de typologie de données, mais il faut surtout que ce soit simple. arXiv.org est un exemple de ce que les chercheurs aiment.</i></p> <p><i>(2) Il faut que l'interface soit claire. Par exemple, le site Internet des CFF est extrêmement clair, limpide, sans cinquante millions de boutons : on entre la destination, l'heure et on obtient le résultat. C'est transparent et utilisable par un enfant. A l'inverse, le site de la SNCF est le pire exemple que l'on puisse donner en termes de complexité et de fonctionnement [rire]. Donc mieux vaut s'inspirer des CFF que de la SNCF !</i></p> <p><i>(1) On n'a pas besoin de trop de graphiques, de trop d'éléments design. arXiv est un système d'archivage d'articles scientifiques qui doit avoir une trentaine d'années. Les chercheurs y mettent leurs articles avant publication. L'interface est restée simple, rapide, efficace et fonctionnelle. Pour moi, c'est un bon exemple d'interface qui a su rester simple, même en modernisant ce qui devrait l'être. Il faut être « to the point » : quand on veut soumettre quelque chose, il faut un formulaire dans lequel on se retrouve facilement. Pour moi, il est important qu'au moment où l'on arrive sur</i></p>

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

	<p><i>l'interface, on soit invité à faire ce que l'on est venu faire. Si l'on tape « Yareta » sur Google, on arrive sur la page d'accueil du site facilement ; par contre, le premier truc que l'on voit sur l'interface est l'historique, quoi, comment, pourquoi, etc. Si je veux soumettre un truc, je dois chercher un bon moment. Je n'ai pas expérimenté longtemps l'interface, je voulais juste savoir où mettre mes données et cela m'a pris un moment. Il est donc important que l'on arrive rapidement sur un écran qui nous invite et qui nous accompagne à faire ce pourquoi l'on est sur le site, si possible avec un processus linéaire et pas une arborescence dans laquelle on ne sait pas où on en est. La suite des informations demandées doit être logique ! Il est sûrement nécessaire de fournir des données sur la personne qui dépose, sur la publication liée aux données, etc., avant de déposer les données en soi. Je ne sais pas exactement comment le site sera organisé. Il est souvent utile que le formulaire de soumission comprenne des exemples pour que les utilisateurs sachent ce que l'on attend d'eux immédiatement.</i></p> <p><i>(2) Oui. C'est par exemple très utile pour les formats et les séparateurs : les éléments de dates sont-ils séparés par des points ou des virgules, la casse est-elle prise en compte, etc. ?</i></p> <p><i>(1) Il y a probablement un moment de prise de contact où l'on entre des informations de base destinées à préparer l'upload des données.</i></p> <p><i>(2) Une barre de progression dans le processus avec des titres explicites serait agréable pour se situer. Si l'utilisateur oublie de fournir certaines données dans le processus, elle devrait pouvoir reprendre le processus plus tard sans perdre ce qui aura déjà été complété ou pouvoir revenir en arrière. Une liste des éléments obligatoires pour le dépôt permettrait aux utilisateurs de bien se préparer au dépôt et de gagner du temps. Si l'utilisateur doit recommencer de zéro en cas d'interruption, il sera vite agacé.</i></p>
5.2	<p>Sur la page d'accueil, préférez-vous que l'on vous propose directement de déposer vos données ou de consulter les derniers dépôts effectués avec une option de recherche ?</p>
	<p><i>(1) Cela va beaucoup dépendre de l'expérience des utilisateurs finaux, qui ne sera pas grande la plupart du temps. S'il sait déjà comment fonctionne l'interface, il sera facile de déposer les données. Mais s'il arrive sur l'interface pour la première fois, il ne saura pas comment faire, il s'interrogera sur les modalités et des conditions de dépôt. Donc, si le système est capable d'extraire des données bibliographiques contenues dans un fichier compressé, ce serait super et le déposant n'aurait qu'à vérifier les informations trouvées par le système. Comme il ne sera probablement jamais possible de faire cela, il faut un processus linéaire et logique : par exemple, dans les archives ouvertes, on entre les auteurs d'abord, les données bibliographiques ensuite et on upload l'article à la fin.</i></p> <p><i>Ce qui m'a frappé sur les autres interfaces du genre que j'ai pu voir, c'est qu'il n'est jamais possible de prendre connaissance de l'activité qui se déroule sur la plateforme. On ne sait pas combien il y a d'articles, les sujets sur lesquels ils portent, etc.</i></p> <p><i>La fonctionnalité de recherche dans la base de données doit être distincte de la soumission ! Donc, si je veux soumettre, je ne veux pas arriver sur une fonctionnalité de recherche qui interférerait dans mon processus de dépôt. Et si je veux chercher des données, je veux une page bien faite, simple et efficace, qui me permette d'avoir une idée de ce qui se trouve sur le dépôt. Il faut que les deux activités soient distinctes sur la page d'accueil et qu'il n'y ait pas de confusion possible.</i></p>

	<p>[...]</p> <p>(2) Sur la page d'accueil, il faudrait deux gros blocs qui permettent d'arriver sur le processus de soumission ou sur le processus de recherche.</p> <p>(1) La page d'accueil peut même proposer d'autres services. Par exemple, en tant que chercheur, on peut être intéressé à connaître les statistiques de dépôts : qu'y a-t-il comme données dans ce dépôt ? Pour moi, la recherche de données est quelque chose de très spécifique qu'il n'est pas facile de bien faire. La partie soumission, comme je l'ai dit, doit être simple, linéaire et sans « chichi ». A la fin de la soumission, je trouverais important qu'il y ait une page de synthèse qui permette à l'utilisateur d'avoir une vue d'ensemble de son dépôt et de le vérifier. Finalement, il y a probablement un temps à partir duquel les données sont verrouillées et durant lequel il n'est plus possible de les modifier et ce temps devrait être suffisant pour que l'utilisateur puisse effectuer des corrections ou des modifications, avant que le dépôt soit définitif ; une période tampon d'une semaine ne semble pas problématique dans le cas d'un dépôt à long terme. Sinon, il faudrait pouvoir versionner les données, car il ne devrait pas être possible de modifier les données après coup. Ainsi, un chercheur qui aurait triché ne pourrait pas modifier ses données par la suite.</p> <p>Après, il faut aussi que les utilisateurs sachent quels sont leurs droits par rapport aux données déposées. Restent-ils propriétaires ? Ai-je la possibilité de retirer des données que j'aurais déposé au préalable ? Je ne connais pas la position de Yareta sur ces points. Quelles sont les modalités d'accès, non pas en lecture, mais en écriture, à ces données ?</p> <p>Avoir une période tampon et une possibilité de versionnage me semble être une bonne solution.</p>
--	---

Conclusion

Avez-vous des commentaires à rajouter ou des questions au terme de cet entretien ?

	<p>(1) Vous parliez de faire des tests sur des maquettes ; je pense que le gros du travail se fera à ce moment. Pour moi, il serait important de ne pas trop attendre avant d'avoir une version bêta qui fonctionne pour pouvoir vraiment prendre en compte le feedback des gens sur l'interface. Lorsque la question s'était posée pour les archives ouvertes de l'Université de Genève, j'avais vu que nos commentaires d'utilisateurs étaient bien accueillis et pris en compte. Pour vous, nous aurons des données de recherche prêtes à être soumises pour les tests utilisateurs.</p> <p>(2) Je vois un peu ça avec les sites Internet. Si on commence à demander l'avis de tout le monde, c'est bien d'avoir des avis avant de vraiment développer l'interface, car c'est un véritable cauchemar de recommencer le travail après chaque avis.</p> <p>(1) Oui : certains feedback ont moins d'importance que d'autres, surtout lorsqu'il s'agit de choix de couleur. Il faut faire quelque chose d'attractif pour les chercheurs.</p> <p>(2) Mettez l'accent sur la simplicité et l'ergonomie ; c'est l'essentiel. Évitez le superflu.</p>
--	--

Annexe 5 : Entretien 2

L'entretien dure 69 minutes. Il est enregistré, afin de faciliter sa transcription.

Questions préliminaires

Intitulé du projet :	Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs à déposer leurs données		
		Date	11.07.2019
Sexe	Homme	Âge	33
Institution/Faculté	Université de Genève Faculté des lettres Département des sciences de l'Antiquité		

Entretien

N°	Question
Généralités/Introduction	
1.1	Êtes-vous familier du processus d'archivage des données de la recherche ? Pourquoi ?
	<i>Oui, relativement, en tout cas pour l'archivage à titre personnel, et non public. J'archive mes propres données selon un système de classement ou de bases de données pour moi-même. Justement, je participe parce que je suis intéressé à apprendre comment mes données puissent être mises à disposition.</i>
1.2	Distinguez-vous l'archivage de vos articles de celui de vos données de recherche ?
	<i>J'ai déjà publié des articles scientifiques et je distingue un peu les deux éléments. Dans mon cas, ils se rejoignent pour ma thèse – que j'ai fini il y a un an –, parce qu'il serait intéressant que les données – que ce soit des fiches, des photos ou autres – soient aussi accessibles d'une manière ou d'une autre, même si elles sont séparées de la publication. Je suis en train de préparer la publication – à la fois Open Access et imprimée – de ma thèse et j'aimerais aussi donner accès à la base de données, aux photos, aux cartes, etc. Dans le livre, il y aura suffisamment d'illustrations ou autres, mais, si les gens veulent avoir accès à l'ensemble de ce que j'ai fait, il faudra qu'il y ait accès par ailleurs.</i>
1.3	Différenciez-vous l'archivage à court, moyen et long terme ?
	<i>Oui, je pense. Pour le moment, tout est conservé dans une optique de long terme, pour autant que l'informatique tienne. Je ne range pas mes fichiers juste pour une</i>

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

	<i>semaine. Enfin, je garde toutes les données, mais je ne garde pas toutes les versions de mes chapitres, par exemple ; je dois avoir deux ou trois versions préliminaires, mais, globalement, j'efface au fur et à mesure. Je pars du principe que la dernière version est plus aboutie. Par contre, pour ce qui est des données de terrains, des photos, etc., j'ai toujours tout gardé.</i>
Données	
2.1	Quels types de données produisez-vous ?
	<p><i>Je produis pleins de données, qui vont des textes aux photographies. Si l'on parle de données de base, il y a des points GPS qui sont insérés dans des cartes qui, elles-mêmes, sont aussi des données. Je travaille avec des systèmes d'informations géoréférencés (SIG, GIS en anglais) ; le logiciel avec lequel je travaille s'appelle ArcGIS et il me permet de faire des cartes combinées à une base de données en quelque sorte. Sur chaque point ou forme inséré sur le plan, des informations peuvent être ajoutées. Par exemple, pour un site archéologique, il y aura des dimensions, les nombres de vestiges, etc. GoogleMaps est un SIG. L'avantage est que cet outil sert à la fois de base de données de travail et de résultat final, puisque, à la fin, on a une carte. Le type de données principal que j'aimerais archiver et mettre à disposition est des cartes. Elles ne pourront pas toutes être illustrées, ou en tout cas pas dans toutes leurs versions, dans une publication imprimée, voire Open Access. Sinon, j'ai quelques plans (≠ cartes), qui ne sont parfois pas produit par moi, mais juste modifiés, quelques itinéraires de parcours que j'ai fait à pied sur le territoire que j'ai étudié ; cela peut aussi être intéressant à stocker ou à illustrer. J'ai aussi des tableaux Excel et d'autres documents bureautiques, sans que ce soit toujours des données intéressantes à sauvegarder sur une plateforme comme celle dont il est question dans votre étude ; ce peut être des choses à conserver temporairement et qui ne serait pas compréhensible ou utilisable par d'autres personnes. C'est aussi un challenge d'avoir des choses compréhensibles par d'autres personnes du domaine ou non. J'oublie peut-être quelque chose, mais voilà les éléments clés : photos, données GPS, cartes.</i></p> <p><i>Les cartes doivent être modifiées grâce au logiciel ArcGis, mais elles peuvent être extraites en format JPEG, PDF ou autre, et insérées dans des outils du type de GoogleEarth. Il y a des formats divers et des formats lisibles plus facilement – pour faire des illustrations, par exemple. Elles peuvent donc être stockées sous un format qui permet de les lire facilement. Après, – mais je ne suis pas sûr de le faire – beaucoup de gens mettent à disposition leurs cartes techniques sous des formats propres aux logiciels avec lesquelles elles sont créées pour les utiliser. Il y a eu, par exemple, une couverture des terres agricoles de toute l'Europe, avec les types de terre, et ça a été mis à disposition sur Internet, parce que l'Europe l'a voulu ; ce sont des données utilisables, pour autant que l'on utilise le logiciel qui a permis de les créer.</i></p>
2.2	En général, quelles tailles font vos données (fourchette) ?
	<i>Pour ma thèse, je ne sais plus exactement, mais je dirais que j'ai quelques gigabits, ou quelques dizaines de gigabits. Si l'on prend tout ce que j'ai produit ou modifié pendant ma thèse, c'est plus, mais je pense que je vais archiver des données de l'ordre que quelques dizaines de gigabits. Les logiciels qui permettent de faire les</i>

	<i>cartes génèrent de lourds fichiers, mais, une fois extraits dans d'autres formats, ils ne pèsent presque plus rien – comme un fichier Word ou Excel.</i>
2.3	Selon vous, combien de temps un dépôt de données à long terme devrait-il conserver les données ?
	<p><i>Je ne sais pas. [...] Le plus longtemps possible ; il n'y a pas de raison que les données disparaissent. L'intérêt, si l'on propose une solution au sein de l'Université, réside aussi dans le fait que l'on a moins peur que le système disparaisse. Au sein du Département des sciences de l'Antiquité, nous avons réfléchi à créer une interface spécifique ; je faisais partie des quelques motivés, mais nous avons eu de la peine à trouver des gens soit "geek", soit intéressés pour poursuivre la réflexion. Il y a toujours ceux qui disent que c'est trop compliqué ou trop long, et ceux qui pensent que l'ensemble de leurs prises de notes sur trente ans sont candidates à l'archivage à long terme.</i></p> <p><i>On entend souvent parler de dix ans. Dans le cas de ma thèse, les données nécessaires à la compréhension seront publiées, donc les gens susceptibles de vouloir en savoir plus se manifesteront dans ce laps de temps et ils ne représenteront pas des centaines de personnes.</i></p> <p><i>Quelque part, on peut dire que les données archéologiques devraient être conservées sur du plus long terme que des données en sciences dures, parce que les données – de base, en tout cas – restent vraies. Par exemple, les dessins réalisés par quelqu'un qui travaillerait sur de la céramique antique – je ne parle pas des cartes, car elles peuvent toujours être enrichies – seront encore justes dans cent ans ! Une fois que quelqu'un a documenté correctement de la céramique (dessiner, photographier, analyser la composition éventuellement), des monnaies ou autres, les données ne bougent pas ; ce ne sont que les datations qui peuvent parfois changer, sachant que l'on n'invente pas souvent de nouvelles techniques. L'objet ne se retrouve pas sur un système informatique, contrairement aux données de base qui en découlent et qui peuvent être utiles très longtemps. Il y a moins de révolutions, de nouveaux systèmes de calcul ou autre que dans d'autres domaines. Disons que dix ans restent un laps de temps relativement court pour conserver des données archéologiques ; on travaille encore avec beaucoup de choses qui ont été documentées il y a entre cinquante et deux-cents ans. Ceux qui travaillent sur des statues travaillent parfois sur des textes anciens de plusieurs siècles, mais ça ne veut pas dire que les analyses n'ont pas changé. Par exemple, certains scientifiques ont pu faire des dessins de frises du Parthénon avant son explosion (1687) et ce sont les seules représentations du monument datant de l'époque des Ottomans. Cela ne va pas arriver avec mes données de thèse, mais, sur le principe, les données archéologiques peuvent encore être utiles après plusieurs siècles.</i></p> <p><i>Pour mes données de thèse, dix ans devraient suffire, mais pour l'archéologie en général, des gens peuvent vous répondre deux-mille ans, ce qui ne serait pas complètement faux, même si c'est un peu prétentieux. Typiquement, les relevés d'objets restent intéressants, mais je n'ai pas ce genre de chose dans ma thèse. On trouve souvent des données qui ont plusieurs dizaines d'années et sur lesquelles personne n'a plus jamais travaillé.</i></p> <p><i>Pour ma thèse, j'ai travaillé sur une ville en Calabre pendant la période romaine ; d'autres personnes vont sans doute aussi travailler dessus, mais pas beaucoup. En archéologie, on a bien assez de sujets qui nous occupent pour ne pas tous travailler sur la même chose.</i></p>
Stockage/Archivage des données de recherche	

3.1	Archivez-vous déjà archivé vos données de recherche sur un outil public ? (Pourquoi ?)
	<p><i>Globalement, non. Ma thèse est maintenant sur une archive ouverte, mais j'ai choisi de restreindre l'accès pour qu'il ne soit possible que pour les gens de l'Université de Genève. J'ai vu qu'on pouvait bloquer l'accès complètement, mais je n'en voyais pas l'intérêt. Sinon, mes données sont sauvegardées dans plusieurs endroits, mais pas de manière publique. Pour les autres recherches que j'ai faites, c'est pareil dans l'ensemble, notamment parce qu'il n'y a pas encore de plateforme idéale ; il n'y a pas vraiment de dépôt prévu pour les données archéologiques et je ne sais pas s'il existe ce genre de choses pour les archéologues en science, parce qu'ils ont plus d'analyses au carbone 14 ou des choses comme ça. C'est l'un des défauts de l'archéologie : même si l'entente est bonne, chacun a son projet et sa façon de faire. Par exemple, la manière de documenter les choses n'est pas toujours la même et cela complique le stockage ou la mise en série.</i></p> <p><i>Donc non, mes données n'ont jamais été archivées de manière publique.</i></p> <p><i>Pour ma thèse, les données sont surtout chez moi, et un peu chez mon professeur et vers d'autres chercheurs. Par exemple, dans mon bureau, nous sommes trois à travailler sur la même ville à différentes périodes, donc ça fait sens de partager nos données. Avant, j'ai fait beaucoup de recherches en Grèce et, dans ce cadre, l'archivage partiellement public des données était plus développé. Il y avait une archive commune et un serveur, et chaque chercheur devait donner tout ce qu'il avait produit, mais ce n'était pas une règle écrite. Donc, dans ce cas, les données étaient plus partagées. Ici, le serveur contient des dossiers par thèmes et les chercheurs ont accès aux dossiers thématiques sur lesquels ils travaillent. Donc ce n'est pas totalement public.</i></p> <p><i>De manière générale, je partage volontiers mes données. Mais là, par exemple, je ne vais pas partager mes données de thèse avec tous mes sites, parce que ça va bientôt être publié, mais après, pourquoi pas !</i></p>
3.1.1	Quels formats informatiques utilisés vous en général pour vos données ?
	<p><i>J'ai différents formats d'image : tiff, jpeg et autres classiques. Il y a les formats de ArcGIS, dont je parlais, qui sont propriétaires et parfois générés sans que l'on en soit conscient. Quand j'ouvre le dossier dans lequel il y a mes SIG, il y a pleins de formats que je ne connais pas (CPG, DBF, etc.), mais que je n'ai pas besoin d'ouvrir. En réalité, le seul fichier que l'on ouvre et qui va chercher les informations qui se trouve dans les formats inconnus – pour autant que tout soit dans le même dossier –, c'est le fichier ArcGIS de base, dont je ne connais par contre pas l'extension. Le tout pèse en général deux mégaoctets. [...]</i></p> <p><i>Il m'arrive aussi souvent d'aller chercher des données d'autres chercheurs – pas forcément dans des domaines proches – sur Internet, même si ce n'est pas toujours facile à obtenir, et, souvent, c'est pareil : on télécharge un lot de fichiers dont tous les formats ne sont pas connus, on en ouvre un qui utilise les autres. Par exemple, quand je cherche des données sur l'hydrographie de l'Italie, les données ne sont pas collectées par des archéologues. [...] Les données de recherche de chercheurs qui ont une septantaine d'années et qui ne les ont jamais publiées risquent d'être bientôt perdues, car, souvent, les chercheurs n'autorisent pas leurs collaborateurs à se les approprier ou à les gérer à leur place. Par exemple, une collègue est en conflit avec un professeur de Genève, qui estime avoir encore des droits scientifiques sur du</i></p>

	<p><i>matériel qu'il a étudié il y a cinquante ans, qu'elle étudie actuellement et qu'elle veut publier ; il essaie donc de la bloquer. L'esprit Open Access n'est pas encore très implanté dans notre domaine, ou cela dépend des générations. Pourtant, en archéologie, il s'agit de données de base, pas comme un code informatique qui demande beaucoup de travail.</i></p> <p><i>Pour revenir aux autres formats que je peux avoir, certaines cartes sont en KML. Il y a aussi des Word, des PDF, du FileMaker pour les bases de données. J'ai un certain nombre de dessins, avec Photoshop et autres, mais rien de très original. J'ai aussi des fichiers qui comprennent des données GPS et qui sont générés par le logiciel de Garmin, BaseCamp. Je n'ai aucun son ou vidéo, par contre. Je crois que c'est tout.</i></p>
3.1.2	Stockez-vous vos données à long terme (format, outil, dépôt, temps) ? (Pourquoi ?)
	<i>Éléments déjà abordés.</i>
3.1.3	Avez-vous déjà stocké des sets de données sur un dépôt ? Si oui, pourriez-vous nous raconter comment vous avez procédé ?
	<i>Éléments déjà abordés.</i>
3.1.4	Quelles seraient les éventuelles remarques que vous pourriez faire à propos du dépôt que vous aviez alors utilisé (nom/avantages/inconvénients) ?
	<i>Éléments déjà abordés.</i>
3.1.5	Quelles sont les difficultés que vous avez rencontré durant l'archivage de vos données ?
	<p><i>Je n'ai jamais perdu de données. Je suis un peu un maniaque des backups. Tout récemment, une collègue s'est fait voler son sac et, avec ça, près de six mois de données sur le terrain et mille francs suisses cash, parce qu'elle organisait un colloque. Heureusement, je n'ai jamais eu ce genre de mauvaises surprises. Si je suis dans une phase où je produis des données assez intensément, je fais des sauvegardes tout le temps. J'ai une partie de ce que j'écris sur Dropbox, mais je n'utilise pas les serveurs de l'Université. On nous a proposé SwitchDrive, mais les gens qui l'ont utilisé autour de moi ont eu des bugs qui m'ont un peu refroidi. En plus, pour commencer à l'utiliser, il faut demander des accès et on nous prend par la main de manière un peu fatigante ; Dropbox est beaucoup plus facile à utiliser. Avec l'Université, nous avons aussi droit à un Terabits de stockage sur OneDrive, ce que j'utilise un peu. Aucune de mes données n'est à un seul endroit.</i></p>
3.2	Lors d'un travail de recherche, à quel moment vous interrogez-vous pour la première fois sur l'archivage des données de la recherche ?
	<p><i>Pour ma thèse, cette problématique s'est posée relativement tôt par mon propre intérêt, parce que je n'avais aucune obligation. Par contre, maintenant, si quelqu'un fait une demande équivalente à celle que j'avais faite pour ma thèse au FNS – c'était un doc.ch –, il faut présenter un data management plan qui prend du temps à être rédigé et qui semble être un peu en décalage avec notre branche, parce que nos volumes de données ne sont pas énormes. Je vous parlais de dix gigabits, mais peut-être que d'autres vous diront dix ou cent terabits ; je ne sais pas. L'autre problème,</i></p>

	<p>avec le DMP, est qu'il faut tout quantifier à l'avance et que l'on n'en est pas souvent capable.</p> <p>Le système doc.ch, qui est pour chercheur individuel et non pour des projets, permet une sacrée liberté. Ça peut avoir des défauts si on n'est pas rigoureux, mais on peut faire ce qu'on veut avec nos données une fois que le projet est lancé. Je n'ai donc jamais discuté avec mon professeur pour savoir ce qu'on allait faire de nos données. Initialement, nous voulions lancer un projet de groupe, qui n'a pas passé, mais dans lequel la gestion des données aurait peut-être pris plus de place, puisque le projet aurait rassemblé des chercheurs d'horizons différents (historiens, archéologues, biologistes, etc.).</p> <p>Maintenant, je suis en train d'essayer de lancer un projet de post-doctorat, dont je vais surtout gérer les données moi-même en espérant que des systèmes vont se développer pour l'archivage. Mais je ne mène pas une immense réflexion sur des conditions précises à fixer à l'avance. D'ailleurs, pour les « early post doc » – le type de financement que je suis en train de demander –, aucun DMP n'est demandé ; c'est peut-être au vu du laps de temps relativement court sur lequel se déroulera le projet (dix-huit mois), ou parce que les jeunes chercheurs qui débutent un doctorat doivent prouver qu'ils maîtrisent aussi ces aspects-là. Je ne sais pas.</p>
Cadre	
4.1	Réalisez-vous un DMP (<i>Data Management Plan</i> , ou plan de gestion des données) dans le cadre de vos projets ?
	<i>Non, on ne m'en a jamais demandé et j'archive au fur et à mesure. Ce n'est pas inutile d'y réfléchir pour éviter des problèmes plus tard, mais je ne l'ai jamais fait.</i>
4.2	Votre institution vous impose-t-elle une politique, une procédure, un dépôt, ou quoi que ce soit par rapport au dépôt de vos données ?
	<p><i>Non, l'Université de Genève ne m'a rien demandé de spécifique, hormis que les résultats [publication] soient déposés dans une archive ouverte. Les discussions avec le professeur qui dirigeait ma thèse portaient sur l'importance de pouvoir rendre disponible et archiver, parce que le FNS demande à ce que tout soit en Open Access ; mais ces discussions portaient plus sur la publication des articles et des résultats que des données. Le FNS commence à demander l'archivage des données de recherche également, mais je ne faisais pas encore partie de ceux à qui l'on demande d'avoir les données en Open Access. Apparemment, cette exigence concerne surtout les scientifiques des sciences dures ; nos données restent relativement ridicules en termes de volume par rapport à des données de calculs de physiciens ou autres. L'idée est de tout avoir online, mais ça ne m'a pas été demandé fermement.</i></p> <p><i>Les sciences de l'Antiquité sont rattachées aux Lettres, donc ce ne sont pas les gens les plus informatisés du monde, mais nous devons quand même nous y mettre de par l'importance des technologies qui nous aident à faire des modélisations, des plans, des cartes, etc. Après, c'est vrai qu'un latiniste sera moins concerné par le problème, même s'il utilise probablement Word et Excel...</i></p>
Besoins et attentes	

5.1	<p>Qu'attendez-vous d'un dépôt de données de la recherche ? Quelles sont les fonctionnalités indispensables pour vous ? Que doit vous permettre de faire l'interface de dépôt ?</p> <p>(Ex: fil d'Ariane pour suivre l'avancée de la procédure, design sobre pour ne pas interférer dans la démarche, beaucoup de couleur pour rendre l'activité plus vivante, des aides contextuelles, etc.)</p>
	<p><i>[...] Idéalement, il me plairait beaucoup que, quand on publie un livre ou un article en Open Access, il y ait directement les liens nécessaires entre les publications et les données. C'est-à-dire que, au lieu d'avoir un bête PDF qui reproduit le livre, l'image serait petite, comme sur un site web, on l'ouvrirait et on pourrait faire défiler trente-deux images associées. Je pense qu'on y arrivera à un moment donné ou à un autre. [...]</i></p> <p><i>Sinon, dans le cadre d'un dépôt, il faudrait que l'on puisse stocker des données de manière durable, comme on en parlait avant. Il faudrait aussi qu'il puisse devenir mon propre dépôt aussi : je trouverais pratique que l'endroit où l'on dépose nos données de manière rangée pour qu'elles soient visibles et accessibles au public ne soit pas qu'une vitrine. Sinon, les images et les cartes devraient être rangées dans des dossiers pour être vues et téléchargées par d'autres chercheurs ; ça n'a pas besoin d'être plus sexy qu'une arborescence sur un ordinateur, mais il serait bien de pouvoir visualiser les données finales, si c'est adéquat du moins, avant de les télécharger. Ce qui serait bien, c'est que les données soient facilement trouvables et compréhensibles (présence de légendes ou autres), pour que les gens qui les consultent et qui ne connaissent pas le domaine puissent identifier les données et s'y retrouver. Typiquement, mes cartes peuvent être comprises par quelqu'un qui ne serait pas archéologue. L'interface devrait être explicative, et pas seulement un dépôt froid, ce qui est parfois le problème des serveurs privés qui contiennent une arborescence dans laquelle il faut cliquer douze fois pour trouver quelque chose qui pourrait nous être utile.</i></p> <p><i>Un moteur de recherche serait pratique, même si ce n'est pas facile à mettre en place. Il serait génial que les fautes d'orthographe soient corrigées, que plusieurs orthographes soient possibles pour un terme et que la casse ne soit pas prise en compte. Peut-être que les dépositaires pourraient indexer leurs jeux de données. Et, peut-être qu'il serait intéressant d'avoir une page par thème étudié, pour éviter d'avoir un immense moteur de recherche qui doit chercher partout et pour rassembler des recherches semblables et leurs données, et pour présenter les études. Rédiger cinq lignes pour présenter ma recherche et les données que je déposerais me prendrait trente secondes ; ce n'est pas un immense travail supplémentaire. En plus, il pourrait éventuellement y avoir deux arborescences : celle-ci par thème de recherche, et une par type de données, de manière à pouvoir parcourir toutes les cartes existantes dans le dépôt.</i></p> <p><i>Si je m'imagine en tant qu'utilisateur, il serait bien de pouvoir trouver des cartes réalisées par d'autres personnes, dans d'autres domaines, de manière à voir comment ils travaillent et comment ils construisent leurs cartes, même si ce ne sont pas des données brutes que je réutiliserais.</i></p> <p><i>[...]</i></p> <p><i>De manière générale, tous les chercheurs ne sont pas à l'aise avec l'informatique ; ils ont donc besoin que l'interface soit claire et facile d'accès. Des aides contextuelles peuvent faciliter l'usage aussi. Sans que ça ne ressemble à un site web, il est important de pouvoir voir les données en miniature (aperçu), afin de se les représenter. Un bouton permettant de télécharger un jeu de données dans son ensemble simplifierait aussi la tâche des chercheurs qui cherchent des datasets sur</i></p>

	<p><i>le dépôt ; si un téléchargement prend deux heures, les gens ne vont pas passer par ce biais pour obtenir ou verser les données.</i></p> <p><i>Le lien entre la publication (article) et les données doit être mis en évidence !</i></p>
5.2	<p>Avez-vous déjà déposé votre thèse sur le dépôt institutionnel de l'Université de Genève ? Si oui, comment ça s'était passé ?</p>
	<p><i>Sans problème, ce n'était pas trop compliqué. Par contre, le portail n'est absolument pas sexy ! On retrouve les choses en tapant le nom de famille du chercheur par exemple. Il y a aussi un problème au niveau du nom : on appelle le dépôt une archive ouverte, mais les accès peuvent être restreints partiellement ou complètement par les dépositaires. Le problème, c'est qu'on ne peut pas ouvrir l'accès complètement, avant que la publication ne soit officiellement publiée ou en Open Access. Pour votre plateforme, il faudrait imaginer la possibilité de mettre des embargos en durée sur les jeux de données. Pour l'archive ouverte, on nous demande de mettre notre manuscrit, donc ce qui nous permet de soutenir la thèse et d'avoir un imprimatur, mais l'imprimatur n'est pas lié à la publication finale, mais au droit de déposer auprès de l'Université. [...]</i></p>

Conclusion

Avez-vous des commentaires à rajouter ou des questions au terme de cet entretien ?

Lors de vos tests utilisateurs, il pourrait être intéressant de faire tester par des gens qui n'ont jamais été confrontés au dépôt de données, à l'archivage de données ou à la recherche de jeux de données, car ils ne connaîtront pas le système.

Annexe 6 : Entretien 3

L'entretien dure 32 minutes. Il est enregistré, afin de faciliter sa transcription.

Questions préliminaires

Intitulé du projet:	Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs à déposer leurs données		
		Date	11.07.2019
Sexe	Femme	Âge	48 ans
Institution/Faculté	Université de Genève et HUG Faculté de médecine		

Entretien

N°	Question
Généralités/Introduction	
1.1	Êtes-vous familière du processus d'archivage des données de la recherche ? Pourquoi ?
	<i>Au sein de mon département, on le fait, mais pas sous la forme d'une plateforme qui serait accessible à l'extérieur ; c'est-à-dire qu'on a le serveur académique NAS, sur lequel tout le monde a un espace de stockage et sur lequel on peut archiver nos données. C'est un serveur sécurisé qui ne met par contre pas les données à disposition d'autres chercheurs : c'est de l'archivage sans partage. Enfin, le partage n'est possible qu'au sein du groupe de recherche.</i>
1.2	Distinguez-vous l'archivage de vos articles de celui de vos données de recherche ?
	<i>Oui, parce qu'on ne s'occupe pas vraiment nous-même de l'archivage de nos articles à partir du moment où ils sont publiés – donc accessible dans le domaine public –, car on considère que c'est l'éditeur qui va en gérer l'archivage. Maintenant, la plupart des journaux dans lesquels on publie sont online, donc, au moment de la soumission, l'article et les données qui lui sont attachées se retrouvent sur un serveur qui appartient à l'éditeur du journal. A partir de ce moment, le matériel électronique est entre les mains des éditeurs qui gère leur processus.</i>
1.3	Différenciez-vous l'archivage à court, moyen et long terme ?
	<i>Non. Pour le moment, nous fonctionnons de manière très primaire ; c'est-à-dire que l'on pose tout sur NAS et on arrête de réfléchir. Donc il y a des données que l'on va réutiliser, et d'autres pas. Dans mon groupe, on n'est pas très bon avec les métadonnées : en général, celui qui a déposé des données sur le serveur sait ce que</i>

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

	<p><i>c'est, mais celui qui veut les reprendre trois ans plus tard n'y comprendra rien et sera incapable de les réutiliser, parce qu'elles auront mal été archivées. Il n'y a pas beaucoup de rigueur. On prend maintenant un peu conscience, parce qu'on en entend de plus en plus parler ; on a des exigences au niveau des organismes de financement qui nous demande de dire comment on compte archiver les données. Donc je pense que ce sont des choses qui vont entrer dans les mœurs, mais, en ce qui me concerne, c'est encore vraiment le début. Je pense qu'on serait content ; enfin, on râlerait au début, mais ça nous rendrait vraiment service à moyen terme d'avoir une structure qui nous permette de déposer les choses de façon un peu cadrée, avec des exigences. On ne sait pas de quelles données contextuelles ou autres on a besoin. Au moment où l'on dépose nos données, il faut nous obliger à trier (identifier ce qui est réutilisable ou non) et à décrire correctement ce que l'on archive. C'est surtout à ce niveau que je vois l'utilité d'une plateforme comme celle sur laquelle vous travaillez. Il faut un mi-chemin entre l'espace de stockage où l'on pose tout au fur et à mesure du travail et la publication qui est vraiment la forme très aboutie du travail : il y a une zone grise entre les deux avec des données qui pourraient profiter à d'autres ou à nous-mêmes, mais que l'on ne gère pas très bien. On a besoin d'un guide qui nous aide à trier et qui nous oblige à décider de ce que l'on va archiver à long ou moyen terme et de ce que l'on ne va pas garder, ni réutiliser.</i></p>
Données	
2.1	Quels types de données produisez-vous ?
	<p><i>Il y a différents types de données. On a beaucoup d'appareillages scientifiques qui mesurent différentes choses : ce peut être toutes sortes de choses qui sont souvent développées précisément pour une utilité expérimentale. Les fichiers créés ne sont lisibles que par les logiciels (logiciels) qui vont avec la machine en question. Les raw data sont les premières choses que l'on obtient. Après, il y a toujours un moment où l'on traite les données ; là, on aura des feuilles de calculs de type Excel. On fait aussi des analyses statistiques avec le logiciel GraphPad et d'autres logiciels plus génériques. De la même façon, on a pas mal d'images qui sont numérisées dans un format spécifique, mais que l'on va pouvoir extraire en TIFF ou JPEG. Sinon, j'ai aussi du traitement de texte.</i></p> <p><i>On collabore aussi avec des plateformes – la plateforme de génomique, par exemple – qui utilise des logiciels spécifiquement dédiés à une technique expérimentale ; dans ce cas, les volumes de données générés sont énormes par rapport à ce que notre labo produit. Ces plateformes ont une façon différente de la nôtre de stocker les données : ce n'est pas forcément dans le flux de nos données. On y accède par un portail sur Internet.</i></p>
2.2	En général, quelles tailles font vos données (fourchette) ?
	<p><i>C'est dur à dire. J'ai fait ce genre de calcul, lorsque j'ai soumis une demande au Fonds National et que j'ai rempli un DMP, mais je n'ai plus les chiffres en tête. Ce n'est pas énorme par rapport à d'autres collègues. Je dirais maximum cinquante gigabits.</i></p>
2.3	Selon vous, combien de temps un dépôt de données à long terme devrait-il conserver les données ?

	<i>C'est difficile à dire, mais je ne suis pas sûre que ce soit que les chercheurs qui vont fixer cette limite, parce que je pense qu'il y aura des exigences des organismes de financement. Le chiffre que l'on entend souvent, c'est dix ans. Personnellement, il ne m'arrive pas souvent de reprendre des données que j'ai récolté il y a dix ans, mais ça arrive quand même de temps en temps.</i>
2.3.1	Vous arrive-t-il de chercher des données d'autres chercheurs qui ont plus de dix ans ?
	<i>Oui, ça m'arrive fréquemment. Mais, à ce moment-là, ce ne sont que les données que l'on trouve dans la littérature qui sont disponibles, donc des études publiées. C'est vrai qu'il me serait certainement utile d'avoir également accès aux données attachées à ces études... Donc c'est aussi vrai que, si mes données étaient disponibles pour d'autres plus de dix ans, ce serait bien. Après, je ne sais pas comment on va gérer le côté informatique lié à leur utilisation ; j'ai moi-même des choses sur mon ordinateur qui ont plus de dix ans, mais je ne peux plus ouvrir la moitié, car je n'ai plus les logiciels qui permettent de les lire. A l'époque, je ne me suis pas souciée de leur archivage, de la conversion des formats, etc. En plus, convertir des fichiers dans des formats ouverts est un boulot immense que l'on ne fait pas ; on imagine peut-être le faire avec les fichiers précieux, mais il est difficile de prévoir ce qui deviendra précieux avec le temps. C'est compliqué.</i>
Stockage/Archivage des données de recherche	
3.1	Archivez-vous vos données de recherche ? (Pourquoi ?)
	<i>Comme on l'a déjà dit, j'arrive sur un serveur à l'interne du CMU. Et, comme je l'ai aussi dit avant, on ne fait pas ça bien, parce qu'on pose les fichiers tel quel sur le serveur. En fait, les ordinateurs que l'on a ici sont des stations de travail reliées directement au serveur. On n'a pas un système sur lequel on travaille et un système sur lequel on stocke joliment : on travaille directement sur l'archive. C'est pour ça que c'est un peu le Bronx ; on ne fait pas attention à la façon dont on archive. Il existe des journaux qui demandent à ce que les données soient stockées sur des plateformes en particulier, mais je n'ai jamais eu à le faire. Il existe des serveurs dédiés qui doivent un peu ressembler à la plateforme sur laquelle vous travaillez, mais il semble qu'il s'agisse de données déjà mises en forme et complémentaires à l'étude publiée, et pas forcément de données brutes. Je ne suis pas sûre.</i>
3.1.1	Stockez-vous vos données à court/moyen terme (format, outil, dépôt, temps) ?
	<i>Éléments déjà abordés.</i>
3.1.2	Stockez-vous vos données à long terme (format, outil, dépôt, temps) ? (Pourquoi ?)
	<i>Éléments déjà abordés.</i>
3.1.3	Nous nous avez dit que vous aviez déjà déposé des articles auprès de revues au moyen de formulaires sur Internet (analogie). Pourriez-vous nous raconter comment vous aviez procédé ?

	<p>Tous les journaux dans lesquels on publie ici, en science de la vie, disposent d'une interface pour la soumission des articles. Chaque journal a ses propres exigences en termes de format, de genre de fichiers, de taille, de libellé, de contenu, etc. On doit donc se conformer aux exigences du journal que l'on cible et déposer les données en fonction.</p> <p>On génère un compte utilisateur au moment où on commence à utiliser la plateforme en question. Évidemment, si l'on veut utiliser plusieurs fois le même journal, on réutilise le même profil. Il y a toute une partie administrative où on s'identifie – affiliation, etc. – et, après, il y a un volet pour la soumission des données de recherche à proprement parler.</p>
3.1.4	Quelles seraient les éventuelles remarques que vous pourriez faire à propos du dépôt que vous aviez alors utilisé (nom/avantages/inconvénients) ?
	<i>Pas de commentaire spécial.</i>
3.1.5	Quelles sont les difficultés que vous avez rencontré durant l'archivage de vos données dans un dépôt ?
	<i>Pas de commentaire spécial.</i>
3.2	Lors d'un travail de recherche, à quel moment vous interrogez-vous pour la première fois sur l'archivage des données de la recherche ?
	<i>L'archivage se fait naturellement, mais, du coup, il n'est pas réfléchi.</i>
Cadre	
4.1	Tout à l'heure, vous avez mentionné un DMP (<i>Data Management Plan</i> , ou plan de gestion des données). Est-ce vous qui l'avez rédigé ?
	<p><i>Oui, c'est moi qui l'ai écrit. C'était horrible ! Je ne m'étais jamais posée les questions mentionnées dans les « guideline » et ne savais vraiment pas comment y répondre ! Donc, même avec les guidelines, je ne comprenais pas bien ce que l'on attendait de moi. C'est vraiment une autre façon de penser ; ce n'est pas quelque chose que l'on avait l'habitude de faire.</i></p> <p><i>D'après ce que j'ai compris, le FNS ne va pas le valider au premier tour, parce qu'il considère que c'est un « work in progress ». Les chercheurs doivent soumettre un DMP qui soit à peu près plausible ; je ne sais pas ce que le FNS a comme critère, mais il ne vérifie pas si tout est bien fait, parce qu'il considère que l'on est supposé le mettre à jour régulièrement au cours de la recherche, en fonction de son évolution.</i></p>
4.2	Votre institution vous impose-t-elle une politique, une procédure, un dépôt, ou quoi que ce soit par rapport au dépôt de vos données ?
	<i>Non, c'est le FNS qui pose le cadre. D'ailleurs, ce n'est peut-être pas si bien. C'est pour ça que je trouvais que le fait d'avoir une solution suisse un peu centralisée et avec des normes était bien. Pour le moment, chacun y va de son système et chacun réinvente l'eau chaude, sans que ce soit uniforme, communiqué ou guidé. J'ai l'impression qu'il serait plus efficace s'il y avait – je ne sais pas – un groupe de</i>

	<p><i>réflexion qui s'y connaisse, qui y réfléchisse et qui définisse un système et une manière de faire qui fonctionne.</i></p> <p><i>J'ai entendu parler des démarches du DLCM en cherchant un peu comment faire et je suis tombée dessus par hasard sur Internet, mais je trouve qu'il ne fait pas beaucoup de communication. Sinon, je suis allée à un atelier de la bibliothèque [de l'Université] pour que l'on m'aide à remplir mon DMP et on m'a plus parlé du DLCM.</i></p>
Besoins et attentes	
5.1	<p>Qu'attendez-vous d'un dépôt de données de la recherche ? Quelles sont les fonctionnalités indispensables pour vous ? Que doit vous permettre de faire l'interface de dépôt ?</p> <p>(Ex: fil d'Ariane pour suivre l'avancée de la procédure, design sobre pour ne pas interférer dans la démarche, beaucoup de couleur pour rendre l'activité plus vivante, des aides contextuelles, etc.)</p>
	<p><i>Je vous ai déjà parlé de guidelines et d'aide pour le tri des données. Sinon, je trouverais plus facile si tout était un peu uniformisé, de façon à ce que chacun puisse poser ses données de recherche de la même façon et qu'il soit plus facile de les retrouver. L'utilité que je verrais à cette plateforme serait d'avoir un intermédiaire entre l'endroit où l'on pose nos données pas très bien rangées pour notre propre utilisation et la publication aboutie dans laquelle les données sont sélectionnées, parce qu'on ne peut publier que des histoires complètes, finies, qui ont suffisamment éveillé d'intérêt chez les gens qui les ont évaluées. Et il y a pleins de données – des données négatives, par exemple – qu'on génère à longueur d'année, qu'on sait qu'on ne peut pas publier, mais qui ne seront pas utilisables par d'autres si elles ne sont nulle part ; et les autres devront refaire le même travail à double. C'est un gaspillage de ressources et d'énergie. Ce serait donc très pratique de pouvoir poser ce genre de ressources quelque part de manière accessible en disant bien « qu'elles ne forment pas une histoire complète, qu'il n'y aura pas de joli papier avec une introduction et une conclusion, mais on a fait ça, ça a donné ça et, si ça vous intéresse, venez piocher » ! Mon attente, ce serait d'avoir une plateforme où l'on puisse mettre à disposition des résultats qui n'auraient pas été exploités dans une publication, mais qui existent. Je ne sais même pas s'il existe déjà quelque chose qui permette de retrouver ces données. C'est le partage qui m'intéresse le plus, parce que, finalement, on a déjà une solution d'archivage pur et dur.</i></p> <p><i>Dans tout ce qui est relatif à la génomique, au Big Data, il y a des plateformes avec des données qui sont accessibles, où on peut aller chercher des choses qui nous intéressent. Les plateformes que j'utilise sont thématiques ; c'est-à-dire qu'elles ne proposent que des données sur le séquençage du génome, par exemple. Donc, quand on a des données du même type, on peut les archiver au même endroit, mais on a aussi des données d'autres types pour lesquels il n'y a pas de plateforme du genre. Je sais qu'il existe des plateformes génériques, mais je ne les trouve pas conviviales : la navigation y est difficile, la recherche aussi. Et, j'avais l'impression que, pour déposer des données, il fallait qu'elles soient relativement élaborées et mises en contexte ; je n'ai pas eu l'impression que les gens y déposaient des fichiers contenant des raw data. Ce que j'y ai cherché et trouvé ne m'a pas plus apporté que les publications auxquelles ils étaient liés. Finalement, les données qui servent à la publication et qui se retrouvent sur les dépôts donnent la même information que la publication, en plus détaillé.</i></p>

5.2	Démonstration de Figshare et Zenodo.
	<p><i>Préférence pour le fonctionnement de Zenodo.</i></p> <p><i>Ce que je trouverai le plus utile pour aller chercher les données des autres – ou même les miennes –, ce serait qu'il y ait un outil de recherche qui soit facile d'accès et qui permette de chercher des gens, des domaines ou des mots clés. Cela permettrait de mieux cibler la recherche sur la cellule ou la molécule qui nous intéresse, par exemple. Enfin, c'est comme ça que je cherche des informations. Les résultats peuvent être affichés sous forme de liste ; on y est habitué avec PubMed. Un pictogramme ou un dessin ne m'aiderait pas forcément. Après, ce qu'il y a de sympa dans PubMed, c'est que, depuis la liste de résultats, on peut facilement avoir un résumé, sans pour autant afficher la page complète de l'article ; parfois, on n'a pas envie d'aller sur le lien de l'éditeur.</i></p> <p><i>[...]</i></p>

Conclusion

Avez-vous des commentaires à rajouter ou des questions au terme de cet entretien ?

Je pense qu'il faut qu'on puisse corriger ou versionner les données si elles montrent des erreurs.

Annexe 7 : Entretien 4

The interview lasts 56 minutes. It is recorded, in order to facilitate its transcription.

Preliminary questions

Project title	Design of a portal for archiving research data and proposal of interfaces to encourage researchers to submit their data		
		Date	18.07.2019
Sex	Homme	Age	34
Institution/Faculty	EPFL Faculté des Sciences et Techniques de l'ingénieur (STI)		

Entretien

N°	Question
General/Introduction	
1.1	Are you familiar with the research data archiving process? Why?
	<i>Yes. I was explained by the DLCM when we met, so I understand the difference between "storage" and "archival".</i>
1.2	Do you differentiate between archiving your articles and archiving your research data?
	<i>Yes. We do storage and not archival in the sense that the DLCM means. We have our own repository for research data and it's called Materials Cloud Archive. There we archive data – at the latest, when the corresponding scientific paper is published – because, since January 2019, there is a requirement from the Swiss National Science Foundation [FNS]. So the papers will go on a journal in open access normally, or, if it's not open access, we use the green road and publish on arXiv. And the data is published and sent in our repository.</i>
1.3	Do you differentiate between short, medium and long-term archiving?
	<i>We keep on the computing clusters the data used only during the research; anyway, they're not really used afterwards, unless you need it to restart calculations. So, we can call this short-term preservation maybe. Those that are kept are in the laboratory directly, either on the supercomputer where we are doing the simulations, on the group machines or on the workstation where we do backups. And then we have the long-term archiving – meaning ten years in our case – where basically our repository also guarantees that the data is preserved in the long term. Data is stored at CSCS [swiss national supercomputing center] who stores it for at least ten years.</i>

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

We encourage researchers in Marvel, a NCCR [National Centre of Competences in Research] – NCCRs are Swiss projects founded by the Swiss National Science Foundation-to think about which data to store. It means also to encourage people to think of the cost of storing data versus recomputing it. The cost we have for the data, in the sense of developing and maintaining the archive, is ~140 CHF per terabyte per year. The cost of computing depends of the cost of CPU and on the salaries of the people who are running the simulations. We are working on platforms like AIDA [software] which are used to run simulations with a few clicks. You encode in a Python form the whole workflow that you run and, when run it is then stored in the form of a graph, with full provenance automatically tracked. The human time should be minimal because you have perfect track of what has been run when you ran on supercomputers. It reduces the human cost. Still you have the CPU cost because you run on supercomputers for days or months. We tell researchers the cost of data if you store it, and they should think of how much it costs to recompute, and if you really need to put everything involving checkpoint files, or just the final files and the input files together with the AiiDA provenance.

We have these two distinctions, short-term is not published in general. It also includes our checkpoint files, that the code needs when you need to restart. Typically, when you run a code on a supercomputer you can run, e.g., for 24 hours maximum, and then you must restart anyway. Then the code should write something about the simulation on the disk, also an overview in case you need to use these data one month after the experiment.

Instead the final files will be published in the long-term which means ten years or more on our hard drive.

The repository that we use has been developed mainly within NCCR MARVEL. The web address is nccr-marvel.ch. It is a collaborative repository involving 40 professors and over 150 postdocs and PhDs across Switzerland from Geneva to Lausanne, Zurich, Bern, Fribourg etc... In EPFL there are several groups involved, including my group but there are also other groups in physics or chemistry. For MARVEL and other funding projects, like the European MaX Centre of Excellence, people should store in our Materials Cloud repository, and we are then opening the repository to the world. We have however a moderation phase, so when someone submits data, one of us checks the content. We don't do a scientific peer-review but we assume it will be done for the publication so we do a moderation phase where we check the topic. For the world, we decided to limit to computational science and experimental sciences is in general not accepted, except if the people are from MARVEL and MaX. In this case, we can make an exception. If anybody in the world contributes, the limit is 5GB per project for any project in the domain of computational material science. It is 50 GB if they use AiiDA which is a software tracking the provenance, because we think there is additional value in having this information. This will also be limited to computational science only.

These archives are shared and open to the world. There is no password or restriction. You get the DOI and the files are downloadable. This Material Cloud Archive is part of the Materials Cloud portal. In case you use AiiDA, you have links to what we call "discover" and "explore" sections which are interactive views on the web where you can browse the graph of data and calculations that AiiDA generated. You can also see the inputs and the outputs. The "Explore" section is like GitHub for git. Git is the tool for storing the history of sourcecode and GitHub is a service used to visualize the git

	<i>repositories. For us, AiiDA stores the “history” of the data generation process, and the section called Explore is the place where we visualize the AiiDA graph. Moreover, in the discover section we show curated data, linked to the Explore section. [...]</i>
Data	
2.1	What types of data do you produce?
	<p><i>As a computational scientist, I run simulations and I have no experimental data. There are different categories of data produced, the first is the code development. The data produced is the code. Every code or around 95% of the codes we develop are open-access on Github. We don't typically archive the code even if it is possible with a link between GitHub and Zenodo. We plan in the future to release versions to archive them on Zenodo but we don't do it yet.</i></p> <p><i>The second category is the research data. This is for example the input files we give to the simulations. These files are the inputs to codes to simulate materials properties. It is giving as an input the atomic positions of the atoms, for example. There are then different kind of files used; the input files which includes the position of the atoms and the parameters for the code, the output files which are text files, excel-like tabular files or binary files. We store these in the long-term usually.</i></p> <p><i>There are some intermediate data such as the wave function that the codes compute and need for the computations. They are big (10-100 GB). These are stored on the supercomputer and left there. They are archived as a “.tar” file and moved in another folder after the simulation. This is done because the simulations run on a scratch folder which is wiped out every 30 days. So, if you want to keep the data, you must move it in a different folder. This data is possible to be regenerated as we use also AiiDA as said earlier. Just giving the inputs and gather the output wouldn't be necessarily sufficient to know what we did so we give also these graphs in an AiiDA format. The inputs nodes could include the crystal structure, the code used for calculations, and a set of parameters and potentials. The nodes returned as output could be another structure modified by the code, with some parameters as well.</i></p>
2.2	In general, what size do your data make (range)?
	<p><i>The files stored in the long-term are from less than 1GB to 10-15 GB. The biggest files can go to 1 or 2 Terabytes. It depends if we analyze one material, with only a few atoms that represents only a few kB. If we do a high-throughput computation with ten thousand materials, the simulation will have thousands of data. Also, we do some molecular dynamics calculations where we take a crystal structure and you let it evolve over time and for every pico-second you have an snapshot. This is an information of the position of the atoms. The longer you run and the bigger the output becomes. This is how we get Terabytes files. Even for GB files of input data you can produce hundreds of GB on the supercomputer.</i></p> <p><i>If you ask other communities these are quite small files. For example, climate community they have difficulties to share their files and store it on tapes.</i></p>
2.3	How long do you think a long-term data repository should keep the data?
	<i>I said that it was for ten years with our system. It is hard to say if it is enough. The problem is the cost and the aging of the data. There are some data which are never used and big so it is a waste of money. Also with the evolution of technologies, it is</i>

	<p><i>hard to justify paying for much longer. I think that less than 10 years is too short. 10 years is reasonable, up to 20 years. I feel that we need to understand the value of these data. There are multiple ways to do it. You can take the data that has not been accessed for 5 years and move it to a less expensive storage and it would be inaccessible from the web. What we normally use is accessible from the web. The next step is that if it has not been accessed for two years, you take it and put it on a tape. If you request it, you must wait for instance for a day to get it.</i></p> <p><i>Secondly, if it hasn't been used for 5 or 10 years it is maybe ok to delete it. If the data is valuable, it is probably going to be archived and taken over to move it. I think that 10 to 20 years is a reasonable range. We need to think what to do in 10 years otherwise, if we just delete it, we keep losing value constantly. There should be in the future at least a way to know which data will be deleted in, say, the next two years. If you really care you can move it somewhere else in a different repository and keep a page saying that it has been moved there.</i></p>
Storage/Archive research data	
3.1	Do you archive your research data? (Why?)
	<p><i>I have two ways to see it. One is the developer of the archive and the other is the researcher that produces the data. As a researcher, I use the archive and develop my own archives. I think in the context of DLCM, it is more storage than archive in the sense that there is no ISO standard we follow to make sure that it is replicated in multiple places.</i></p> <p><i>We do archive our data; the repositories are managed by MARVEL. It is not an EPFL repository. This is for the long-term so ten years in our case.</i></p>
3.1.1	Do you store your data in the short to medium term (format, tool, repository, time)?
	<p><i>Either it is stored on the supercomputer or on group machines. It can also be work stations. We have policies for periodic backups. For MARVEL, we have written a data management plan. We explain very carefully what we do so the policy promises that the backups exist for at least 90 days and that there is at least a weekly backup. This is the minimum policy that was accepted by everybody in the institution. There is no way to force to do it but we send a questionnaire once a year and everybody must say If they are doing it or not. It is depending on every researcher and on the size of their data. Sometimes, I delete the intermediate files after one week and if I keep them, it means that I still need them for analysis.</i></p> <p><i>If I know that I don't need the intermediate data for the paper, I delete it. Or I just keep the input and the output to make it reproducible. If I know it is a final simulation, I keep it. There are some cases when I run the input, I get the output with a lot of intermediate files but I will never use this simulation so I delete it directly. If I really think that the calculation is wrong, I must restart it from scratch with different parameters so there is no point in storing these files. In case I need to do the first step, which takes one hour, and another step which takes two days, I prefer to run all the shorter or cheap calculation first and then all the "expensive" calculations later. I need to run, store temporarily and when I am ready I take these files back and do the second step of the research.</i></p>
3.1.2	Do you store your data in the long term (format, tool, repository, time)? (Why?)

	<i>[Question not asked]</i>
3.1.3	Have you ever stored datasets on a repository? If so, could you tell us how you did it?
	<p><i>I did store on Zenodo because it was required at a conference. Since we have our archive I usually archive on our repository when it is related to computational material science. For people that submit experimental data to our repository then we tell them in the moderation phase to simply send all the files to Zenodo. Many of the data repository entries are below 5 GB or 50 GB limit but there are few cases when we needed to store much more. Zenodo told us that there was no legal way for us to pay them to store more data on their repository. We had to develop something on our own. We are now working to implement ourselves the management of our archive. We are now migrating to INVENIO (software used behind Zenodo). In the beginning of winter, we should have migrated everything to INVENIO so that we can scale easily the repositories to archive more data. We have a relatively small repository with a hundred deposit per month but still the human work involved is quite important.</i></p> <p><i>The process for uploading is the following. There is a web form to fill in on our implemented system. You fill in the metadata and send it. Then you can see the preview. When you are ready, you send an email to our archive email address. One of us will reply in the moderation phase. This is for technical reasons and in the future we will allow them to create an account and submit from a web page. You write all the metadata; first name and surname, affiliation, the title, description (basically the abstract), keywords (at least 3), license (the default one is "CC-BY"). We accept any license that is in the SPDX licenses list [Software Package Data Exchange]. Then you must load the files. For each file, there is a short text description (one line). There are some guidelines that are published on the website. We suggest inputting a TXT file with longer explanations. It is very inconvenient to download 1GB and discover it is not the files that you want so this TXT files helps avoiding it. We always ask to put an email address, to use standard formats, if possible to use text formats like CSV rather than excel and other formats. If there is a PDF, it is used for describing the data. It is however not usable for machine-reading the data. Then you submit, you can check the preview and if everything is alright, you can send an email. One of us will reply, guaranteed within a week. We might ask for some changes, like the filenames, remove some files, add a TXT file. When the author replies, we accept it and the DOI is released. You can ask to create a new version of a previous entry when you do a new entry.</i></p> <p><i>For the preparation phase, I will tell you what I do personally but everyone does as he wants. I take my files and delete the ones that were wrong. I make sure that I have the inputs and outputs properly organized and make sure it is sorted out and put in folders. If they are parsed and make sense to keep, I create a folder for these files. For example, when I do a certain number of simulations, the structure changes and only the final simulation is relevant, then I create a folder with coordinates only and not the full output of the code. I try to choose a format that is easy to use and to visualize. After this, I prepare also some TAR files that are compressed, depending on how I think it is best to organize. There are four categories for the files. I write a REA ME that explains everything; how to use the data and ideally how to rerun the simulation with the input and the output files. This is done in all cases.</i></p>

	<p><i>Then, if I use AiiDA, I include the AiiDA database and if possible I also make it available on the web in the Explore and the Discover sections. It is then easy to find it and download the entire archive and understand it.</i></p> <p><i>Also, for the first time ever in the field probably, we are doing something for a paper that we will publish in a month. We are releasing a virtual machine where we reproduced the whole set of codes that we used for the paper. We release the scripts to automatically recreate this machine (using Ansible) and the virtual machine itself. Anybody can take it and start it. The codes are already installed with the right version used for the paper. They can then rerun the simulations. We put an example input, the script runs it and you have an explanation telling you to start the machine, go into the home folder and run the script, wait, get the output and then check if the output is the same than the one that we published. We are going beyond what people does in the field and we really want to guarantee the reproducibility. This is a complex process because you must spend full days checking that the codes that interact with each other are the exact versions that were used for the publication, which version of AiiDA and plug-ins are installed. We check also the workflows because there are six simulations in a row that run in a specific order.</i></p>
3.1.4	<p>Do you have any comments about the deposit you used at the time (name/advantages/disadvantages)?</p>
	<p><i>Many things could be improved. I will tell you my comments but I think they are valid for many other repositories.</i></p> <p><i>First thing, if you work with many authors, it is very annoying to introduce the authors and affiliations. Finding ways to facilitate this would be good. It could be in a text string format with commas and the script puts it in the right form or you could write emails and the data would be taken from a source and automatically fill the form. A format where you write in a text file the name, surname and the affiliation would be practical. For example, if you have 30 authors, you would fill 30 authors instead of 90 fields.</i></p> <p><i>What I want also to improve in the next version is to make it easier to update the reference of the data to the original scientific publication. This step involves a human person to approve the change.</i></p> <p><i>We already have a preview but I tell you anyway because I think it is very important to see what is going on.</i></p> <p><i>One thing that we could improve even more is to make it very clear in the form if there are requirements that we will check during the moderation phase. For example, there are technical requirements, like they should write at least 3 keywords. More practical requirements like the necessity to write the text file of description and to compress the files, if you write it, usually people do it.</i></p> <p><i>People are all right if you give them a task to improve their data, they are not neglecting it. And usually they reply quite fast to the archive to thank because we checked their data. The first time that you submit, you don't know how to do it, so if you are clear it helps a lot. But if it is all written on a long documentation page, they won't read it so it is better to put one point close to the form field corresponding to it.</i></p>
3.1.5	<p>What difficulties did you encounter during the archiving of your data in a repository?</p>
	<p><i>There is the problem with many authors but it is more a problem of wasted time. If you put it online, there is often the question of which license we should use. The typical thing people do is to put the default choice.</i></p> <p><i>I think that there is a difficulty in how to prepare the data such that it can be reused. We should use standard formats, it is done in the end usually. What happens is that</i></p>

	<p>when the research is done, you must publish and you want to curate the data. The minimum to publish is done because you want to go ahead. Even if you have a good will, one will spend 3 days to make it perfect and then stop. Sometimes you can make mistakes. For example, if you write an input file for a code, you run it and produce the output file. Then you try to understand the output and change the input but you realize that the changes are wrong. Then, you have an input file that is not corresponding to the output. This happens very often, not because people want to publish wrong data. This is in the process of the research, you try to simulate and then check the output. Doing the archiving after analyzing is tricky and many errors result from this. What we do with AiiDA is to automatically archive while you run the simulation so you can't change the output. There is no mistake and you can't forget any parameters. It requires some time to learn how to use AiiDA and add the right plug-ins. Now, most of the entries don't use AiiDA, we don't want to force people to use it. Still, we hope that people will see the advantages and how good it is to have a database that follows standard formats. Still for most researchers, the greatest difficulty is to make the simulations and their outcome reproducible and following standards. Often, they don't follow the standards. There is no format specific for my data so we just publish them sometimes saying which version of the code was used.</p>
3.2	<p>When do you first ask yourself about archiving research data during a research project?</p>
	<p>For a scientific project, I try to take care about this question very early. Before I start the project, I know I want to publish so I start running with AiiDA. In the end, there is still a curation phase but my AiiDA database is nearly already curated. Of course, there are possibilities to improve. In this case, for the reproducible work that we are doing with the virtual machines, we didn't know in the beginning that we were doing this. Now, after 2 years, we have all the technology to do it and in the end we reproduced the working environment of our actual machine with the exact version. Luckily, we used GIT so we know exactly which version we have and we could easily (in a few days of simulations) reproduce all data of the paper. The more we can try to think about it in advance, the better it is, but practically, it is always something that you will do in the end.</p>
Framework	
4.1	<p>Do you produce a DMP (Data Management Plan) as part of your projects?</p>
	<p>I took part of the redaction of this plan. For MaX, it happened in the first phase, 3 years ago, within the updates and the new DMP for the new phase of the project. For MARVEL, the swiss national science foundation asked for a research data strategy which is an extended version of a DMP with additional questions. These are huge projects involving many groups. I have been writing it.</p>
4.2	<p>Does your institution impose a policy, procedure, repository, or anything related to your data management?</p>
	<p>We usually should follow the procedures of the swiss national foundation and the EU. All projects are funded by these institutions mainly. First, we used the support of the EPFL library, particularly in the beginning. Now, they have a template and they provide it to us. We prepared new templated where we prefilled some contents in case people use the Materials Cloud Archive and/or AiiDA.</p>

	<p><i>If you use Material Cloud, you can already say that the data is kept for ten years and that it is a non-profit repository. If you use AiiDA, you can tell them that research data is automatically saved and in which format it saved. If you use one of them or both, you have to write down the technical requirements needed in the DMP.</i></p> <p><i>I don't know if the EPFL has some requirements. I know that there are a lot of efforts to have open science and I am aware of the technical efforts for this archiving process but I am not sure about their requirements. I always put my data on Material Cloud because it is already covering more than the minimal requirements by Material or EPFL.</i></p>
Needs and expectations	
5.1	Have you ever used data from other researchers to conduct your studies? If so, how did you get them ?
	<p><i>Yes, I used quite a lot of external databases. Some of them are open-source without fees and some are commercial. There are some other materials repositories online which contain materials data. We also take source from there as starting point of some research. In one case, the output of our research was used by other researchers for a completely different topic. Ours was a purely material science project. Another group of computational sciences took the data and, using a machine learning prediction, they could calculate the time needed to compute the simulation given the data of input. This is quite important and useful for us.</i></p>
5.2	What do you expect for a long-term repository? What are the essential features?
	<p><i>I expect support for a minimum ten years aligning to the requirements of the funding agency. [...] One thing that we want to have (and don't have yet) is an Embargo but with a specific way to do it. The procedure is to put the data online, the moderators can check it. You can give a link to a reviewer of paper to check the data. You are asked also to put a specific date after which the data becomes public. The condition is that the maximum time is two years until it becomes public but you can reduce it. The goal is that if you have a data and you want to write a second paper, you can keep it private. Two years is a reasonable time because if you don't publish anything on this data for two years then it should be good to release it. On the other hand, it gives you enough time to publish. After two years, if you don't ask to curate the data before it is online, the human cost to curate it then is too high (students might have left the group etc.). This kind of feature is important and can be guaranteed by the system.</i></p> <p><i>The cost is a tricky part. I would expect that it would be free for public funded research. However, I don't know if it is the right thing to do for big datasets. I think that funding agencies should have a specific way of supporting publishing data that start having costs. The cost should be fair ideally. One way could be that archive is a commercial part where people pay and it offers free support for public funded research. If researchers must pay too much, they won't do it. On the other hand, there can be some mechanism to request the storage if it is more than 50Gb, you should do a proposal to be allowed to deposit more data, or put some effort to reduce the size.</i></p> <p><i>Then, one technical feature we were discussing with DLCM, is that we have an agreement with the CSCS (Swiss supercomputer center). They have a lot of storage and we pay them for storage. The archive we have, we pay a storage both for the current storage and we pay upfront for the next 10 years. This required some legal agreements and we do some projects together so they can get some money directly from the funding agencies to this aim. With DLCM, if you must pay someone else; for us, there should be a way to pay directly to CSCS storage and a small part should be</i></p>

	<i>given to DLCM because it guarantees the archiving. I think both ways are important. One that can pay DLCM and take care of everything and another one that can pay the storage place and DLCM would use it with a minimal fee for additional services. This is mainly for big projects, not so much for only one person.</i>
5.3	Do you have some comments about the research part or the platform that we will create?
	<p><i>I expect explanations of the additional guarantees that you have when archiving respectively storing. A clear way that tells you where you can store it, such that you know where they are. Also, to know the costs, the different prices should be given, for which guarantees and what timeframe. For my research, I don't need it but I think it would be interesting to have a choice between online and offline storage. So, it would mean that you could either access immediately to your data or ask for them and have it after a day or more.</i></p> <p><i>Then a way is to have EMBARGOS, with a clear explanation of what EMBARGOS is and why it happens and that nobody can change it after the approval. You can do a new version but you cannot remove it.</i></p> <p><i>One thing that would be interesting to know for a user is to know what copyright is and which consequences it has. A description of the process of withdrawal and how long it takes would be good to have. Also, what happens if I withdraw my data: the DOI is there but the data has been withdrawn and all the consequences. I think these are useful information because it is quite complicated. There should be a clear way to deal with withdrawal procedure and to know who decides, what time does it take to decide, can it be withdrawn only temporarily, etc.</i></p> <p><i>Then the guarantees should be clear, either if I pay or not for the archival, what are the guarantees for the long-term depository and who is paying. I don't know if it is useful to have legal guarantees at some point for institutions like EPFL. [...]. For a single researcher, I think that it is not necessary.</i></p>

Conclusion

Do you have any comments or questions at the end of this interview?

One comment that comes to my mind, if you should store or upload big files (>1Gb) it would be good to have a robust way uploading it. If there are some problems of connexion it should hold.

Also, one important things about the forms, make sure that there are ways to save it as soon as possible to avoid losing some time restarting everything from scratch if there is any problem !

Annexe 8 : Entretien 5

L'entretien dure 37 minutes. Il est enregistré, afin de faciliter sa transcription.

Questions préliminaires

Intitulé du projet :	Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs à déposer leurs données		
		Date	18.07.2019
Sexe	Femme	Âge	40 ans
Institution/Faculté	Université de Lausanne Faculté de Géosciences et environnement		

Entretien

N°	Question
Généralités/Introduction	
1.0	Quel est votre rôle auprès des chercheurs ? A quelle(s) faculté(s) sont-ils liés ?
	<p><i>J'ai un rôle de conseil et d'orientation. C'est-à-dire que les chercheurs peuvent venir me voir pour presque tous les aspects de leurs recherches – par exemple le financement de projets, la publication, l'open data, l'éthique, la carrière, etc. Soit je leur réponds, soit je me renseigne avant de leur répondre, soit je les redirige vers les services adéquats. Il m'est aussi possible de les conseiller pour ce qui est de la gestion de leurs données de recherche. Ces questions sont assez récentes, donc je vais vous répondre du mieux que je peux.</i></p> <p><i>Beaucoup de gens savent qu'au prochain délai FNS ils devront faire un DMP, alors qu'ils n'en ont jamais fait ; donc je pense que j'aurais des demandes avant cette deadline. Des gens viennent me voir à propos du DMP et à propos des dépôts d'archives.</i></p>
1.1	Êtes-vous familière du processus d'archivage des données de la recherche ? Pourquoi ?
	<p><i>Oui ! Je ne l'ai jamais fait pour moi, en tant que chercheuse, mais je connais assez bien le principe et les politiques actuelles relatives aux données de recherche. Je sais que la plupart des bailleurs et des éditeurs demandent une certaine gestion des données.</i></p> <p><i>Au niveau des chercheurs, c'est assez variable : certains – même parmi les expérimentés – ne sont pas du tout au courant de ce genre de choses ou découvrent ça peu à peu. Dans notre faculté, il y a des domaines différents : il y a des gens qui sont plus dans les sciences humaines et social, et des gens qui sont plus dans les sciences naturelles (géologie). En sciences naturelles, il y a déjà beaucoup plus de</i></p>

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

	<p><i>pression de la part des journaux qui demandent systématiquement les données de recherche depuis plusieurs années. En sciences humaine et sociale, il n'y a pas cette pression-là ; les pratiques et les sensibilités à ces questions sont donc un peu différentes. C'est une faculté assez hétérogène ! Certains n'ont encore jamais déposé de données sur des dépôts.</i></p> <p><i>A mon arrivée, j'avais mené une petite enquête pour savoir où les chercheurs déposaient leurs données ; d'après leurs réponses, ils n'avaient pas tous compris ce que signifie l'archivage des données et, sauf erreur, aucun ne m'a donné le nom d'un dépôt FAIR et accepté par le FNS. Je sais que certains archivent correctement, mais c'est très nouveau pour une grande majorité d'entre eux.</i></p> <p><i>[...]</i></p> <p><i>En fait, je réalise qu'il m'est déjà arrivé de déposer des données de recherche pour mon propre compte. Au moment où j'ai voulu publier un article chez un éditeur qui m'a demandé de déposer mes données de recherche sur un dépôt. Je ne sais plus si c'était sur Dryad ou Figshare. J'imagine que le processus de dépôt a changé, car il me semble que l'on ne m'avait pas demandé de métadonnées ou de choses comme ça, alors que, maintenant, il me semble que ces informations sont un peu plus réglementées. J'ai l'impression d'avoir vraiment fait le service minimum ; je dois avoir fait ça très mal [rires]. Je devais le faire, mais j'avais vraiment l'impression que personne n'allait utiliser ces données et probablement que beaucoup de chercheurs pensent comme moi pour leurs données. On ne voit pas bien comment des données de recherche peuvent être utiles à quelqu'un d'autre, car elles sont liées à une question précise, les analyses ont été faites et, finalement, s'il y a besoin de données, il y aura sans doute besoin d'explications ou d'un contact. Il m'est déjà arrivé que des chercheurs me demandent mes données pour une publication et que je les leur donne et que je leur explique ce que j'ai fait, ce que ça signifie, etc. : il y a eu un contact assez poussé. Je trouve qu'il n'est pas évident de réutiliser des données mises en ligne par d'autres. J'étais en biologie et j'étudiais le comportement animal, donc c'était des choses assez spécifiques ; lorsque j'ai mis mes données en ligne, je pense que je ne les ai pas bien décrites. Le système était probablement moins développé que maintenant et je ne connaissais pas les standards de métadonnées ; je pense que les dépôts et leurs standards sont beaucoup mieux expliqués maintenant.</i></p> <p><i>La démarche était très simple et j'ai facilement obtenu un DOI que j'ai pu mettre dans mon article. Je n'ai pas eu de difficulté particulière, mais je pense qu'il faut clarifier la description des données, car la manière dont on stocke nos propres données n'a rien à voir avec la façon dont on archive des données sur un dépôt pour qu'elles soient réutilisées. Pour que les données soient réutilisées, il faut que leur description soit très bien faite et, vu que les chercheurs ne sont pas dans cette culture d'avoir envie que les données soient réutilisées – au contraire, on veut se les garder –, je ne sais pas s'ils le font bien. Finalement, si on fait le strict minimum, est-ce que les données pourront être réutilisées ? Je n'en suis pas sûre.</i></p>
1.2	Distinguez-vous l'archivage des articles de celui des données de recherche ?
	<p><i>Oui, ce sont deux choses différentes, même si c'est encore assez compliqué. Maintenant, il y a de plus en plus de liens pour que les données puissent être liées aux articles et citées. Toutes les pratiques vont changer. Les données vont devenir un peu comme des articles ; vont-elles entrer dans un CV académique, saura-t-on</i></p>

	<i>combien de fois elles auront été citées ? J'imagine que ça va se rapprocher de la publication d'un article finalement.</i>
1.3	Les chercheurs de votre faculté différencient-ils l'archivage à court, moyen et long terme ?
	<i>Non. Au vu des réponses que j'avais obtenue à mon sondage, non. Après, c'est vrai que les termes sont spécifiques : la première fois que l'on m'a dit que stocker et archiver n'étaient pas la même chose, je n'ai pas compris. Et, effectivement, ça n'a rien à voir. Dans mon questionnaire, peut-être que je n'avais pas bien expliqué les termes comme « stockage à court terme » et « archivage à long terme » que les gens n'ont peut-être pas compris, parce que c'est peut-être encore un jargon qu'ils n'ont pas l'habitude d'utiliser. Pour le moment, les données ne sont pas encore citées, ni valorisées ; l'archivage est donc perçu comme une contrainte sans avantage. Actuellement, lorsqu'ils déposent leurs données, ils le font le plus tard possible, ou avec un embargo, pour pouvoir se les réserver et les exploiter le plus longtemps possible. Ils voient leurs données comme quelque chose de personnel qu'ils aimeraient garder pour eux, pour leurs propres recherches. Donc ils y voient surtout des contraintes.</i>
Données	
2.1	Quels types de données produisent les chercheurs avec lesquels vous travaillez ?
	<i>Ils produisent des données très variables. Il y a des domaines de recherche très différents. D'une part, les géologues, les scientifiques des sciences naturelles, les géographes, etc. produisent des données de géologie, de géographie, etc. D'autre part, les gens des sciences sociales ont plutôt des entretiens, des questionnaires. Du coup, la structure des données peut être très différente : en géologie, il y a beaucoup d'images et d'analyses chimiques – ce peut être des données très volumineuses qui peuvent poser problème pour l'archivage. C'est pareil avec les entretiens qui peuvent se présenter sous forme d'enregistrement sonores ou vidéos – cela peut poser des problèmes éthiques et des problèmes de volumes de données.</i>
2.2	En général, quelles tailles font leurs données (fourchette) ?
	<i>Il m'est difficile de vous répondre, car c'est vraiment très variable. Ça peut aller de quelques gigabits – s'il s'agit de simples tableaux – à plusieurs terabits. Dans le questionnaire que j'avais fait, j'avais appris qu'on compte 32 gigabits par image. Ces volumes rendent l'archivage difficile.</i>
2.3	Selon vous, combien de temps un dépôt de données à long terme devrait-il conserver les données ?
	<i>Personnellement, je me pose une question financière : on est censé payer par an, mais que se passe-t-il si le chercheur part à la retraite, si le financement du projet se termine, etc. ? Est-ce que quelqu'un prend le relais ? Est-ce que les données sont détruites ? Est-ce que cela dépend de si les données sont utilisées ou non ? Est-ce qu'on fait comme dans les cimetières, avec une sorte de roulement – quelle mauvaise image ! – ? [rires]</i>

	<p><i>On parle souvent de dix ans, mais c'est le minimum, parce que c'est ce qui est exigé par certains éditeurs, puisque l'on est censé pouvoir donner les données attachées à un article jusqu'à dix ans après sa publication. Après, on peut dire qu'on ne les a plus et ce serait dommage qu'elles soient détruites. Finalement, que fait-on des données une fois les dix ans passés ? Et qui finance, surtout ? Les projets s'accumulant, je ne pense pas que les chercheurs seront prêts à continuer de financer les archivages de données. Parce que, suivant les projets, ils doivent participer au financement. Si le FNS finance l'archivage à hauteur de 10 ou 15 mille Francs suisses, cela comprend le financement du dépôt, mais aussi de la préparation des données à l'archivage, donc je ne suis pas sûre que la somme suffise pour l'archivage pendant dix ans...</i></p> <p><i>[...]</i></p> <p><i>Ce qu'on fait des données après dix ans doit peut-être dépendre de l'utilisation des données ; par exemple, des données qui sont beaucoup utilisées ne devraient pas être détruites, alors que l'on peut se poser la question pour des données qui n'auraient pas été ressorties. C'est un choix à faire.</i></p>
Stockage/Archivage des données de recherche	
3.1	Aidez-vous les chercheurs à archiver leurs données de recherche ? (Pourquoi ?)
	<i>Eléments déjà abordés.</i>
3.1.1	Savez-vous s'ils stockent leurs données à court/moyen terme (format, outil, dépôt, temps) ? Si oui, savez-vous comment ils procèdent ?
	<i>Eléments déjà abordés.</i>
3.1.2	Stockent-ils leurs données à long terme (format, outil, dépôt, temps) ? (Pourquoi ?)
	<p><i>Certains chercheurs de la faculté le font sur Zenodo, par exemple. A l'Unil, FORS est souvent mis en avant ; c'est un dépôt en sciences sociales. Ce sont les deux outils qui sont souvent mis en avant. Je ne sais pas s'il existe d'autres dépôts plus spécifiques ; je n'en ai pas vu auprès des chercheurs que je côtoie et personne n'en a fait mention dans les questionnaires que j'ai fait passer.</i></p> <p><i>Je ne suis pas sûre que la majorité des chercheurs stockent leurs données ; à mon avis, ce n'est pas le cas.</i></p>
3.1.3	Avez-vous déjà stocké ou aidé à stocker des sets de données sur un dépôt ? Si oui, pourriez-vous nous raconter comment vous avez procédé ?
	<i>Non, je ne les ai qu'orienté vers des dépôts en disant « ce dépôt existe et correspond à vos besoins », mais ce sont eux qui ont réalisé la démarche technique.</i>
3.1.4	Quelles seraient les éventuelles remarques que vous pourriez faire à propos du dépôt que vous aviez alors utilisé (nom/avantages/inconvénients) ?

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

	<i>Question non pertinente.</i>
3.1.5	Quelles sont les difficultés que vous avez rencontré ou que les chercheurs ont rencontrées et dont ils vous ont parlé durant l'archivage de vos/leurs données dans un dépôt ?
	<i>Question non pertinente.</i>
3.2	Lors d'un travail de recherche, à quel moment les chercheurs s'interrogent-ils pour la première fois sur l'archivage des données de la recherche ?
	<p><i>Soit au moment où ils font une demande auprès d'un bailleur de fonds qui demande un DMP – ce peut être la Commission européenne ou le Fonds national, par exemple –, soit au moment où ils publient un article auprès d'une revue qui demande les données de recherche attachées à la publication. Donc la démarche se fait soit tout au début, soit tout à la fin.</i></p> <p><i>Pas mal de journaux proposent – de manière un peu malhonnête je trouve – des dépôts très chers et qui invitent les chercheurs à y déposer leurs données. Souvent, ils ne sont pas au courant que des solutions gratuites existent. Dans les faits, les journaux obligent de plus en plus un dépôt de données et proposent parfois – mais n'obligent pas – leur propre solution. Dans ce cas, l'article ne peut pas être publié tant que le DOI attaché aux données n'est pas fourni.</i></p>
Cadre	
4.1	Réalisez-vous un DMP (<i>Data Management Plan</i> , ou plan de gestion des données) dans le cadre de vos projets ? Ou avez-vous déjà aidé/discuté avec un chercheur qui devait en réaliser un ? Savez-vous comment il/elle avait vécu l'expérience ?
	<i>J'ai arrêté la recherche il y a deux ou trois ans, donc j'ai déjà déposé des données parce que des journaux l'exigeaient, mais je n'ai jamais réalisé de DMP. Quant aux chercheurs, je ne les ai jamais pris par la main pour en réaliser un, mais je leur ai conseillé des outils et montré des exemples. Justement, le DLCM a réalisé un très bon modèle que je montre souvent.</i>
4.2	Votre institution vous impose-t-elle une politique, une procédure, un dépôt, ou quoi que ce soit par rapport au dépôt des données ?
	<i>Non, pas du tout. Il existe un service de stockage des données de recherche pendant le projet qui vient d'ouvrir, mais il n'y a aucune solution d'archivage. Je crois qu'il y a des projets pour que cela change. FORS est basé à l'Unil et porte, pour le moment, sur les sciences sociales, mais peut-être que l'Unil veut le généraliser ou le développer ; je ne sais pas. Il y a des recommandations en termes d'archivage des données de recherche, mais aucune obligation. Les obligations viennent des organes de financement.</i>
Besoins et attentes	

5.1	<p>Savez-vous ce qu'attendent les chercheurs d'un dépôt de données de la recherche ? Quelles sont les fonctionnalités indispensables ? Que doit leur permettre de faire l'interface de dépôt ?</p> <p>(Ex: fil d'Ariane pour suivre l'avancée de la procédure, design sobre pour ne pas interférer dans la démarche, beaucoup de couleur pour rendre l'activité plus vivante, des aides contextuelles, etc.)</p>
	<p><i>Pour moi, ce qui est super important pour donner une motivation à déposer des données, c'est qu'il y ait des mesures d'impact qui puissent être valorisées par les chercheurs. Il faut que les chercheurs puissent valoriser le fait qu'ils aient déposé des données et que leurs données aient été réutilisées ; il faut que ça leur serve dans leur carrière académique. Il faut aussi pouvoir lier les données aux articles publiés. Je pense que ce serait vraiment un moteur pour que les choses se développent et aillent dans le bon sens, parce que, effectivement, s'il ne s'agit que de déposer, d'oublier et d'avoir son petit numéro, ce n'est pas très intéressant.</i></p> <p><i>[...]</i></p> <p><i>Pour moi, il reste le problème du financement : comment fait-on lorsque l'on veut stocker plusieurs terabits de données ? Que peut-on stocker ? Quelles sont les règles ?</i></p> <p><i>Après, les chercheurs qui travaillent avec des entretiens font face à des questions d'anonymisation qui peuvent prendre un temps très long qui pose également problème. Finalement, si le fait de mettre les données à disposition prend un mois de travail, cela va être difficile à négocier !</i></p>
5.2	<p>Savez-vous si les chercheurs ont recours à des données de recherche d'autres chercheurs pour mener leurs études ? Si oui, comment procèdent-ils ? A quoi font-ils attention lorsqu'ils mènent une recherche sur une interface, par exemple ?</p>
	<p><i>Je n'ai jamais vu de chercheur qui fasse ça. Je leur ai posé la question et n'ai jamais eu de retour dans ce sens. Beaucoup utilisent des données qu'ils n'ont pas produit eux-mêmes, mais ce ne sont pas des données issues de dépôt ; ce sont des données sur la biodiversité ou autres, issues d'associations, etc. La réutilisation de données issues de dépôt pourrait devenir une pratique.</i></p>

Conclusion

Avez-vous des commentaires à rajouter ou des questions au terme de cet entretien ?

Non.

Annexe 9 : Persona



JACQUES POTTIER

CARACTERISTIQUES

37 ans

Citadin de classe moyenne

Carrière académique

A l'aise avec les technologies de base

PROFESSION

Professeur en biologie à l'Université de Genève :

Ce poste l'amène à diriger les recherches de son laboratoire et à enseigner.

Valeurs :

Partage et collaboration.

Sauvegarde du patrimoine, pour autant que ça ne demande pas un investissement trop important.

CONTEXTE D'UTILISATION DE YARETA

Objectifs 

Obtenir un financement du FNS en archivant ses données de recherche sur un portail à cet effet (obligation du FNS).

Trouver des données d'autres chercheurs susceptibles d'enrichir son étude.

Enjeux 

Peu de ressources à disposition (temps et argent)

Gestion de plusieurs projets de recherche en parallèle.

Motivations 

Obtenir le financement du FNS.

Assurer la pérennité de ses données de recherche.

Curiosité pour les opportunités offertes par la plateforme.

Freins 

Manque de ressources à disposition.

Aucune notion d'archivage.

Peu d'expérience avec ce genre de plateforme.

USER STORIES

En tant que chercheur, j'aimerais...

- ✓ déposer mes données de recherche simplement, pour obtenir le financement du FNS ;
- ✓ lier les jeux de données que je dépose aux articles dans lesquels ils sont exploités ;
- ✓ consulter et utiliser des données d'autres domaines ;
- ✓ publier des données négatives ou associer des embargos à mes jeux de données.

En tant que déposant, j'aimerais...

- ✓ rester propriétaire de mes données, afin de pouvoir les retirer de la plateforme si je le souhaite ;
- ✓ être au courant de la procédure de traitement et de curation de mes données par la plateforme.

USER JOURNEY

Déposer un jeu de données

1. Afin d'obtenir un financement du FNS pour mon dernier projet de recherche, j'ai dû réaliser un DMP. L'article est prêt à être publié et je dois préparer l'archivage de mes données de recherche. Je n'ai pas beaucoup de temps et ne souhaite pas donner accès à mes données tout de suite. 😞
2. Comme je travaille à l'Université de Genève, je me renseigne sur la solution institutionnelle : Yareta. 😞
3. Une fois sur la page d'accueil, je suis obligé de me connecter pour accéder à la fonction de dépôt. 😞
4. En regardant mieux la page d'accueil, la box « Starter kit » me donne les informations utiles pour envisager un dépôt. J'en prends connaissance et y réfléchit. 😊
5. Après réflexion, Yareta me semble adéquat. Je m'y inscris et, en tant que responsable de laboratoire, crée une unité organisationnelle. Je ne suis pas très à l'aise avec cette notion. 😞
6. En attendant la validation de cette étape, je prépare mes données pour le dépôt grâce à la *checklist* du « Starter kit ». Heureusement, je suis organisé et cette préparation est relativement rapide. 😊
7. La création de l'unité est validée rapidement et je peux entreprendre la démarche de dépôt. 😊
8. Je trouve facilement le formulaire de soumission du jeu de données. Les champs de métadonnées sont expliqués et semblent complets. Je reçois des aides et me sens accompagné dans un processus simple. J'apprécie la possibilité de pouvoir compléter les métadonnées en plusieurs fois et mettre un embargo. Il m'est aussi très pratique d'obtenir un DOI que je peux lier à ma publication. 😊
9. Une fois mes données soumises, elles sont rapidement validées et publiées. J'en suis averti par un canal que j'ai choisi. 😊

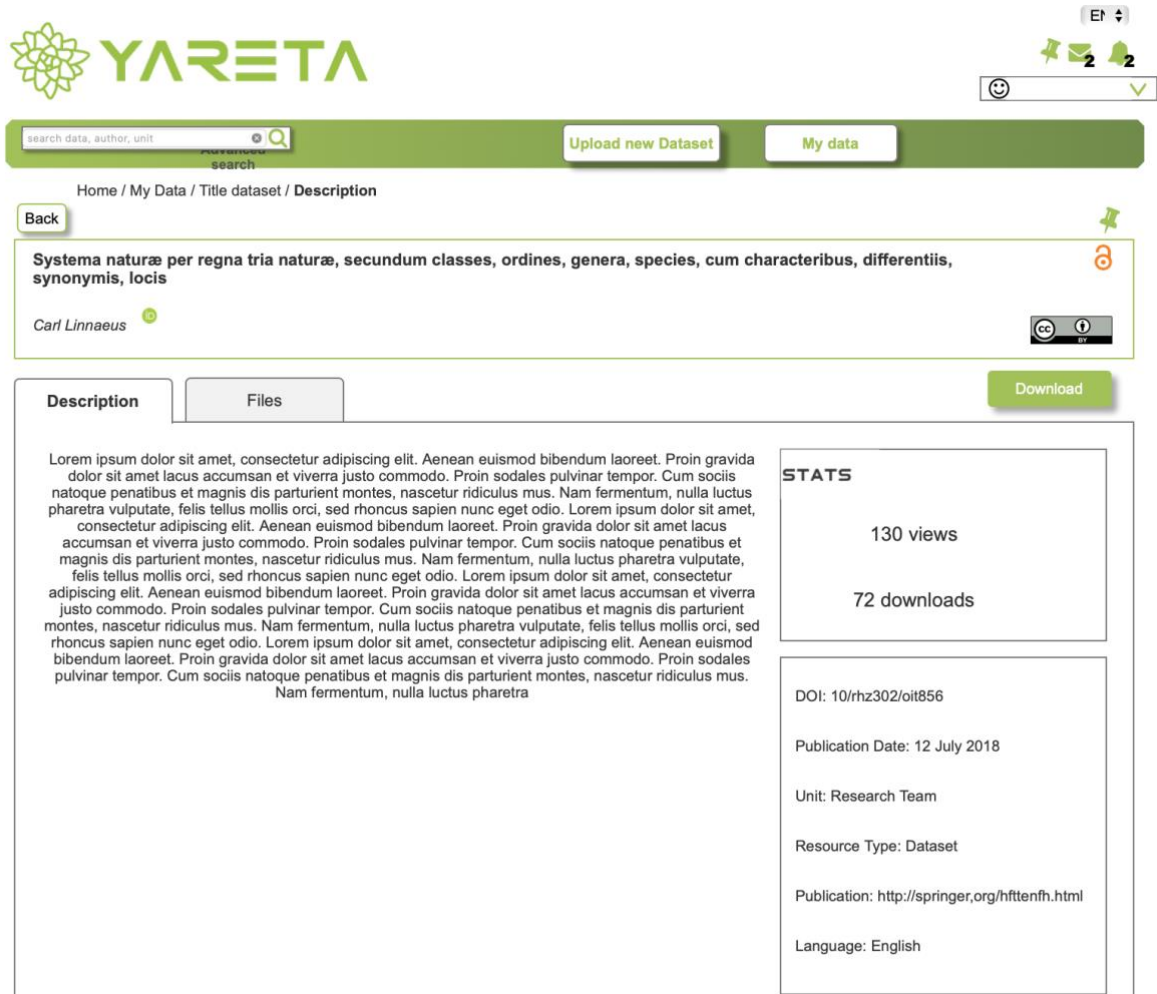
Chercher un jeu de données

1. J'ai récolté des données pour un de mes projets de recherche, mais il me manque le temps et les ressources pour collecter des données supplémentaires que d'autres ont certainement déjà observées. Je suis un peu stressé à l'idée de ne pas trouver des données qui m'aideraient. 😞
2. Je me rends sur Yareta, notre plateforme institutionnelle, pour y trouver des données susceptibles de m'aider. Je n'ai jamais utilisé de dépôt de données et ne sais pas à quoi m'attendre. 😞
3. Le menu met en évidence la fonction de recherche qui me permet d'entrer des mots clés. Trouver la fonctionnalité qui m'intéresse rapidement me rassure. 😊
4. La page de résultat me propose de filtrer les résultats selon des critères qui correspondent à mes besoins. Le processus est rapide et simple ; c'est agréable. 😊
5. Les informations affichées me permettent de choisir les jeux de données susceptibles de m'aider. 😊
6. Je navigue facilement entre les résultats de recherche et les détails des jeux de données. Je vois facilement les aperçus des fichiers, ainsi que leurs métadonnées, ce qui m'aide à en choisir deux que je télécharge en deux clics, afin d'en prendre connaissance et, peut-être, de les utiliser. 😊
7. Le premier jeu ne correspondait pas à mes attentes, mais le second me semble exploitable. Je suis satisfait de ma visite sur le dépôt, car la manœuvre a été rapide et concluante, même si je vais réitérer le processus sur d'autres plateformes afin de compléter ma recherche. 😊

Annexe 10 : Récapitulatif des exigences du DLCM nous concernant pour la plateforme

- Respecter OAIS
- Home page -> Data deposit, data research, organizational unit management.
- Breadcrumb
- Main button colored
- Home page: dataset research, top with data deposit and organizational unit management
- Data deposit: List of projects created, in progress, submitted.
 - Create a new deposit & set the deposit information
 - Add data file to an existing deposit
 - Modify an existing deposit
 - Validate a deposit
 - Approve a deposit
- Data set research: Google-like using keywords, faceted search
- Organizational unit management: list of units he has access to, create or modify. Manager (of the people in the unit)
- Metadata: Mandatory:
 - Title
 - Creator(s)
 - Institution(s) (for each creator)
 - AccessType
 - Licence (if access type is public)
 - Description
 - RessourceType
 - Steward (attributed automatical using submitter's institution)
 - Approver
 - Software used (multiple values)
- Optional:
 - Submission project
 - DOI (can be prerequested)
 - External references DOI
 - External references URL
 - External references Title /Creator(s) /Publisher /PublicationYear
 - Language
 - Contributor(s),
 - Relation(s) and Institution(s)
 - ORCID (Creator, Contributor)
 - Format
 - Related Identifier and relation type
 - Subject and optional vocabulary
 - Additional metadata and Schema
 - Geolocation related info
 - Funding information: grant ID, funder
 - DMP file

Annexe 11 : Première maquette – Page de détail d'un jeu de données



Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis

Carl Linnaeus



Description

Files

[Download](#)


Sort by **Name**

Title	Size	Format	
Title title title	3 GB	JPEG	
Decision making time based on academic resources	570 MB	PDF	
Title of the file	12 Mb	CSV	

PREVIEW

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Annexe 12 : Première maquette - Première page du formulaire de dépôt




ET
2
2

Home / My data / New dataset

Files and metadata > Review

New Dataset

File(s)



Feature not available in this mockup

Drag & Drop

or

Upload conditions

Metadata

MANDATORY

Organizational unit *

Linnaeus Team

Title *

Author(s) *

Brown, James

+ Add me

+ New author

Description *

Resource type *

Dataset

Primary data type *

Observational

Publication date *

01

January

2019

Access *

Open

Embargo *

No

Licence *

CC by

DOI *

https://doi.org/10.1038/s41586-01-1534-3

Request a DOI

Submission policies *

Deposit with approval

Preservation policies *

Keep it for 10 years

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

RECOMMENDED

Language
English

Collection begin
0
January
2019

Collection end
0
January
2019

Keywords
Biology, Linnaeus, Nature

Contributor(s)
Brown, James

+ New contributor

Subject

Collect method

Software used

Domain
Architecture

OPTIONAL

Collect context

External DOI
https://doi.org/10.1038/s41586-019-1534-3

External URL

Additional metadata and Schema

Geolocalisation related info

Funding information


DMP file
Choose...

ReadMe file
Choose...




Cancel
Save
Next



Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données


Annexe 13 : Première maquette – Formulaire de création d'une unité organisationnelle



EN

  2  2

search data, author, unit 

Advanced search

Upload new Dataset



My data

Home / My Unit

New Unit

Unit Name *

Members *


Manager   + Add me

+ add member


Description

Funding agencies

Submission policies

Deposit with approval 

Preservation policies

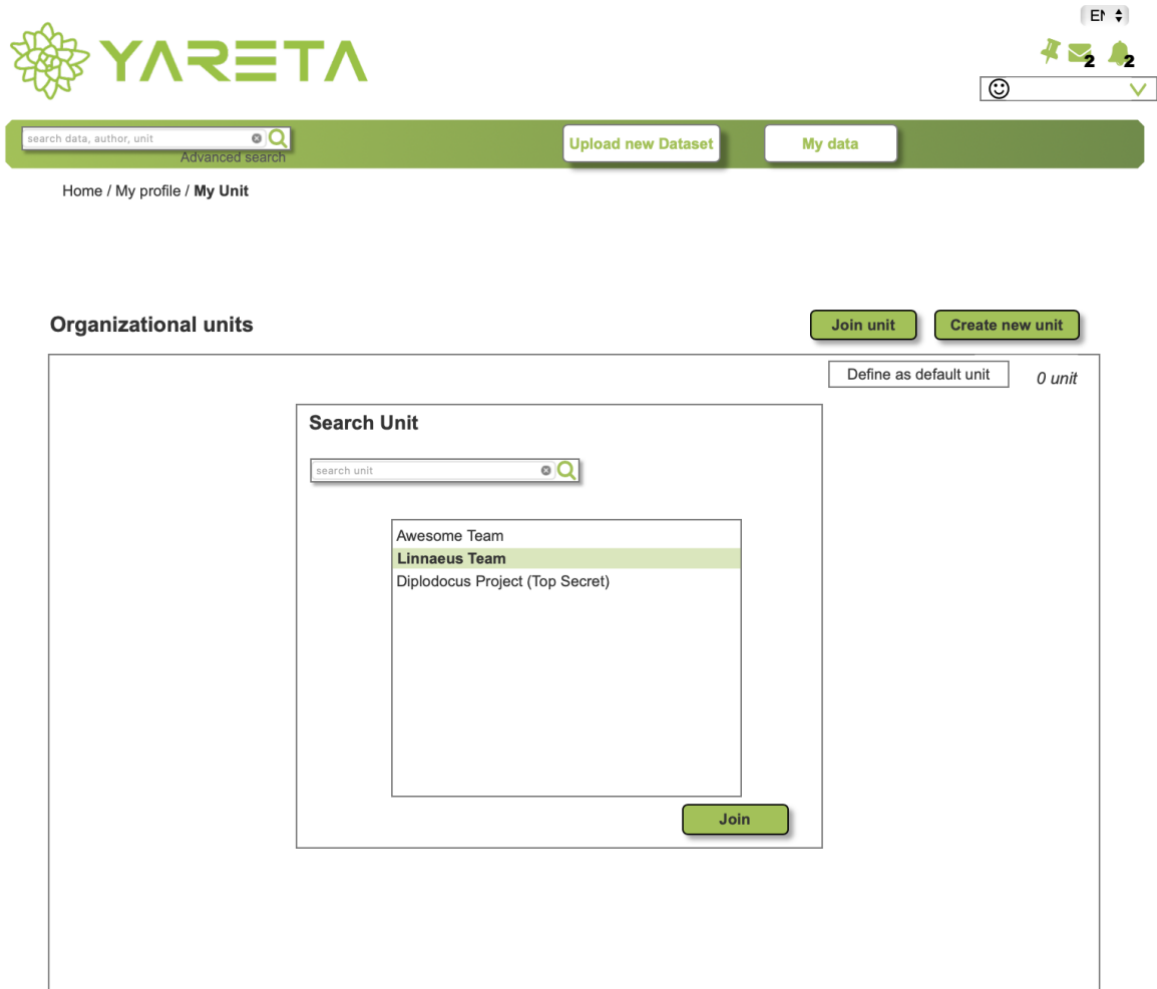
Keep it for 10 years 

Cancel

Submit


Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Annexe 14 : Première maquette – Boîte de dialogue permettant de rejoindre une unité organisationnelle



Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Annexe 15 : Première maquette – Page de détail d'une unité organisationnelle



ET
2
2

[Home](#) / [My Data](#) / [Title dataset](#) / [Description](#)

Linnaeus Team

Submission policies

Desposit with approval

Preservation policies

Keep it 10 years

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus accumsan et viverra justo commodo. Proin sodales pulvinar tempor. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Nam fermentum, nulla luctus pharetra vulputate, felis tellus mollis orci, sed rhoncus sapien nunc eget odio. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean euismod bibendum laoreet. Proin gravida dolor sit amet lacus accumsan et viverra justo commodo. Proin sodales pulvinar tempor. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Nam fermentum, nulla luctus pharetra vulputate,

Members

Sort by

Name	Rank	Added
Marion Favre	Manager	03 may 2019
Heloise Hahling	Member	15 may 2019
Tom Sawyer	Steward	12 june 2019

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Annexe 16 : Formulaire de consentement pour la participation aux tests utilisateurs en anglais

Consent form

Submitted as part of the research project:

Design of a portal for archiving research data and proposal of interfaces to encourage researchers to submit their data

As part of the Master's degree in Information Sciences at the Geneva's HEG, the DLCM consortium asked two students to participate in the development of the DLCM research data repository portal and, more specifically, to focus on the ergonomic and aesthetic aspects of the interface.

A mock-up of the platform has been designed and will be tested with potential future users, in order to help DLCM design a tool tailored to the needs of researchers.

The project is conducted by Marion Favre et Héloïse Hahling.

I agree to participate in the above-mentioned study, and agree that the information collected may be used in this research project. All the recording will be deleted after treatment for this project.

I have been informed that all information concerning me will be treated anonymously.

- I agree that my screen will be recorded.

Place and date:

Participant's signature:

Annexe 17 : Guide pour la participation aux tests utilisateurs en anglais

Research project	<i>Design of a portal for archiving research data and proposal of interfaces to encourage researchers to submit their data</i>		
N° of the test run	<input type="checkbox"/> ₁ <input type="checkbox"/> ₂ <input type="checkbox"/> ₃		
Firstname, LASTNAME		Date	
Gender		Age	
Institution/Faculty		Fonction	

Introduction

Welcome to this user test conducted for the DLCM !

This test will be conducted in two steps:

1. Four tasks will have to be completed on the mockup from your personal computer (individual tasks). We are not giving you any other information than what is on this guide and on the mockup in order to identify what is not clear for you. Your screen will be recorded for later analysis.
2. A discussion will be held to talk about the features of the mockup. This will guide the improvement of the mockup.

Remember that we test the mockup, not you ! We are not evaluating your skills or proficiency. If something is not clear or not working, it's not your fault, but the mockup's.

Part one: Tasks

At the end of each task, take a moment to write your feeling about the mockup for the later discussion (strength, weakness, advantages, disadvantages, etc.). There are no unnecessary or futile comments.

You will know you have completed the task when you'll see the message "**Congratulations, you have completed Task X**".

Do the tasks in order, and one at a time. Take your time, this is not a race :)

Disclaimer: Some options provided on the model are not available, such as filter options. A message will inform you of them.

Link to the user-test : <https://tinyurl.com/ut3dlcm>

Please, use the following information to connect to the platform :

Username	Jack Pott
E-mail address	jack.pott@team.ch
Password	jackpot

Background

Welcome Jack Pott. You are a researcher in Biology, part of the "Linnaeus Team". Your team leader asked you to connect to Yareta and join their unit, so you can add your dataset on the repository.

Task 1 : Get connected and join the organizational unit "Linnaeus Team".
Comments
Task 2 : Search and download a dataset about "Systema naturae".
Comments

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Task 3 : Upload a dataset and submit it.
Comments
Task 4 : Validate a dataset in order to publish it.
Comments

Annexe 18 : Guide de discussion des tests utilisateurs en anglais

Part two : Discussion

Questions

1	What is your general feeling about the mockup ?
2	Can you give us disadvantages of the mockup ?
3	And some advantages ?
4	Is there something not clear in the task you did ?
5	What task gave you the hardest time ? why ?

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

6	Which part(s) need improvement ?
7	What do you think about the page "My data"? How do you think it could be improved? Would you have any other logic to suggest?
8	Is there anything else you'd like to tell us about?

Annexe 19 : Tests utilisateurs – Consignes affichées sur le logiciel Loop11 (troisième test)

Introduction

Welcome to this user test conducted for the DLCM.

Each task has a simple objective. You will know that you have completed the task when you see the message "Congratulations, you have completed Task X".

At the end of each task, take a moment to write your feeling about the mockup for the later discussion (strength, weakness, advantages, disadvantages, etc.). Remember that we test the mockup, not you : there are no unnecessary or futile comments.

Do the tasks in the given order, and one at a time. Take your time, this is not a race :)

Disclaimer: Some options provided on the model are not available. A message will inform you of them.

Première tâche : Log/Sign in and join the organizational unit "Linnaeus Team"

Consigne	<p>Welcome to the first task Jack Pott !</p> <p>You are a researcher in Biology. You are part of the "Linnaeus Team". Your team leader asked you to connect to Yareta and join his unit, so you can add your dataset to the repository.</p> <p>Get connected, find the unit and join it.</p> <p>To connect, use this:</p> <p>Name: Jack Pott</p> <p>Email: jack.pott@team.ch</p> <p>Password: jackpot</p>
Question 1	<p>How would you rate the difficulty of this task ?</p> <ul style="list-style-type: none">• Hard• Moderate• Easy
Question 2	<p>Do you have any comments about this task (strength, weakness, advantages, disadvantages, improvements...) ?</p>

Deuxième tâche : Search and download a dataset

Consigne	<p>Now that you are related to a unit, you need to find a dataset of Carl Linnaeus with a title containing "Systema naturae" in it.</p> <p>You need to download it to complete this task.</p>
Question 1	<p>How would you rate the difficulty of this task ?</p> <ul style="list-style-type: none">• Hard• Moderate• Easy
Question 2	<p>Do you have any comments about this task (strength, weakness, advantages, disadvantages, improvements...) ?</p>

Troisième tâche : Upload a dataset and submit it

Consigne	<p>Your team leader want you to publish your last project's research data.</p> <p>Upload it on Yareta.</p> <p>Info: The folder you want to upload is called "Data" ; you can invent its metadata.</p>
Question 1	<p>How would you rate the difficulty of this task ?</p> <ul style="list-style-type: none">• Hard• Moderate• Easy
Question 2	<p>Do you have any comments about this task (strength, weakness, advantages, disadvantages, improvements...) ?</p>

Quatrième tâche : Validate a dataset

Consigne	<p>Time has passed and you are now the manager of the "Linnaeus Team" organizational unit. You have been informed that your collaborator, Albert Kunz, has submitted a dataset. He is waiting for his latest dataset to be validated and published.</p>
----------	---

	Find the dataset, consult it, then validate it.
Question 1	How would you rate the difficulty of this task ? <ul style="list-style-type: none"> • Hard • Moderate • Easy
Question 2	Do you have any comments about this task (strength, weakness, advantages, disadvantages, improvements...) ?

Questions finales

Question 1	How would you rate the usability of this mockup ? <ul style="list-style-type: none"> • Poor • Fair • Good • Excellent
Question 2	Do you have any other comments ?

Annexe 20 : Compte rendu du premier test utilisateur

Modalités du test

La première maquette a été créée sur le logiciel *Axure* et est accessible par le lien suivant : <https://xmfcgv.axshare.com>.

Le test a été réalisé au moyen du logiciel *Loop11*, qui permet de donner accès au test via un lien URL et d'enregistrer l'écran des participants.

Le test s'est déroulé sur deux jours :

Zurich, jeudi 12 septembre 2019	Genève, jeudi 19 septembre 2019
4 participants qui ont signé un formulaire de consentement.	2 participants qui ont signé un formulaire de consentement.

Les participants ont dû réaliser trois tâches :

1. Se connecter à la plateforme et rejoindre une unité organisationnelle donnée ;
2. Chercher et télécharger un jeu de données indiqué.
3. Soumettre (*upload*) un jeu de données à la plateforme.

Après chacune d'entre elles, il était demandé aux participants de juger de la difficulté de l'exercice. La réalisation des tâches a duré entre 12 et 35 minutes et s'est suivie par une discussion d'une heure et demie. Ce compte-rendu est structuré en tâche, les informations récoltées lors du test et lors de la discussion étant rassemblées par tâche.

Le rapport PDF de *Loop11* est disponible en annexe de ce document. Il contient certains commentaires des participants, ainsi que les informations sur les tâches. Toutefois, il ne faut pas tenir compte des statistiques de réussite, celles-ci étant parfois faussées lorsque les participants n'étaient pas sur la page finale de la tâche au moment où il la signalait comme atteinte.

Compte-rendu

Tâche 1 : se connecter à la plateforme et rejoindre une unité organisationnelle donnée

Difficulté de la tâche : quatre participants ont trouvé la tâche facile, alors qu'un ne l'a pas terminée, et deux participants ont jugé sa difficulté moyenne.

Temps nécessaire : 2'50" – 6'35"

Éléments soulevés par les participants	Répercussions sur la maquette
Les participants hésitent entre « Log in » et « Sign in ». Ils suggèrent de remplacer « Sign in » par « Register ».	○ Remplacer « Sign in » par « Register ».
Les participants se demandent si le choix de l'identifiant des personnes ne devrait pas se porter sur Switch, plutôt que sur ORCID. Un participant affilié à l'Université de Genève nous demande si son compte Yareta sera lié à son compte ISIS.	<i>Ces questions doivent être discutées par le DLCM et la réponse doit figurer dans les aides de la plateforme.</i> <i>En ce qui concerne ISIS, nous suggérons que ce ne soit pas le cas, car l'utilisateur doit pouvoir changer d'institution tout en gardant son compte Yareta.</i>
Yareta est un service de l'Université de Genève : le compte utilisateur est-il privé ou institutionnel ? Il nous est suggéré de faire en sorte que la plateforme soit capable de récupérer les informations de l'Université pour compléter les informations du profil utilisateur, comme les informations du compte AAI par exemple.	<i>Cette fonctionnalité figurera dans nos recommandations d'implémentation.</i>
La page d'inscription oblige les utilisateurs à accepter les conditions d'utilisation qui n'existent pas encore.	<i>Ce problème ne sera pas résolu dans la maquette, mais il sera recommandé au DLCM de prévoir la page en question.</i>
La touche « Tab » du clavier ne permet pas de naviguer d'un champ à l'autre.	<i>Ce problème ne sera pas résolu sur la maquette, mais il sera recommandé de prévoir cette fonctionnalité dans le portail final.</i>

<p>Les participants apprécient le message qui constate que l'utilisateur n'est pas lié à une unité et qui le dirige vers la page adéquate.</p> <p>En revanche, certains participants aimeraient avoir davantage d'informations sur ce que sont les unités organisationnelles, ce à quoi elles servent, pourquoi il est nécessaire de se lier à l'une d'entre elles, comment elles évoluent dans le temps, comment les créer, qui peut les créer et comment l'inscription à une unité est validée.</p>	<p><i>Le concept d'unité organisationnelle n'est pas clair pour les participants et il faut impérativement le clarifier et prévoir de l'expliquer aux usagers.</i></p>
<p>Les participants qui passent par « My unit » sont déstabilisés de se retrouver sur une page « vide » et suggèrent que la liste des unités figure sur cette page. En effet, certains participants ne trouvent pas le bouton « Join » immédiatement.</p>	<ul style="list-style-type: none"> ○ Lors de la première connexion, remplacer la page vide par la page de recherche d'unité.
<p>Sur la page « My unit », le bouton « Join unit » ne semble pas clair. Un participant n'a même pas cliqué et a renoncé à la tâche.</p>	<ul style="list-style-type: none"> ○ Renommer le bouton « Join unit » par « Find a unit ».
<p>Pour les participants qui cherchent l'unité avec la barre de recherche, l'icône pour rejoindre l'unité ne semble pas claire : deux participants ont renoncé à cette voie pour réaliser la tâche par la page « My unit ».</p>	<ul style="list-style-type: none"> ○ Ajouter le texte « Join unit » à côté de l'icône.
<p>Une fois que les participants demandent à rejoindre une unité, rien ne s'affiche sur leur page « My unit » et cela les rend confus.</p>	<ul style="list-style-type: none"> ○ Faire en sorte que l'unité sollicitée s'affiche sur la page « My unit » avec le statut « Waiting for validation ».
<p>Un participant n'a pas trouvé la page « My unit » et n'a pas réalisé la tâche.</p>	<ul style="list-style-type: none"> ○ Afficher « My unit » dans le <i>header</i>, comme « My data ».
<p>Le fil d'Ariane n'est pas cliquable, alors qu'il serait utile qu'il puisse être utilisé comme moyen de navigation.</p>	<p><i>Le fil d'Ariane ne sera pas cliquable sur la maquette, mais il sera recommandé aux développeurs de prévoir cette fonctionnalité.</i></p>
<p>Les unités organisationnelles restent un concept difficile à saisir et confus pour les utilisateurs. Il est important de le clarifier. Ils</p>	<p><i>Ces discussions appartiennent au DLCM, mais les conclusions doivent être annoncées aux utilisateurs par la FAQ, par exemple.</i></p>

se demandent notamment quels seront leurs possibilités d'accès aux jeux de données qu'ils ont déposés en lien avec une unité s'ils la quittent plus tard, ou quels seront leurs possibilités de dépôt de données en lien avec une unité qu'ils auraient quitté.

Certains participants s'opposent à l'implémentation de ce système et ont le sentiment que les unités organisationnelles ont pour but de faciliter le travail des administrateurs de la plateforme et non celui des chercheurs. Ils trouvent que ce concept ne fait pas sens, puisque la composition des équipes de projet dépend du projet et que les laboratoires peuvent collaborer. Ainsi, la plateforme risque de voir les unités organisationnelles aussi nombreuses que les jeux de données, ce qui mettrait en péril leur utilisation. De plus, ils s'inquiètent de la sécurité de la composition des unités : qui pourra s'y inscrire ? La seule utilité perçue est la possibilité de compléter les métadonnées des jeux de données à déposer à plusieurs et cette option n'est pas prévue.

Ensuite, la question de la validation des jeux de données pose également problème. Un manager, s'il est chef de projet, n'aura probablement pas le temps de valider les données de son équipe (il n'en aura donc pas envie) et n'aura probablement pas une connaissance suffisante des données pour valider leur dépôt ; l'option de désigner un membre de l'unité comme validateur ne leur semble pas meilleure. En effet, le déposant n'a pas besoin que ses jeux de données soient approuvés par un tiers, alors qu'il est le plus à même de juger de leur qualité. Par ailleurs, ils se demandent ce que le validateur devrait vérifier.

Cette question de validation des données ajoute une couche administrative lourde qui risque d'orienter les utilisateurs vers d'autres

<p>dépôts. De plus, les articles sont déjà revus par des tiers ; cela peut être perçu comme une tâche effectuée à double.</p> <p>Apparemment, les participants préféreraient interagir directement avec la plateforme, sans passer par une unité organisationnelle.</p>	
Éléments remarqués par nous-mêmes	Répercussions sur la maquette
La plupart du temps, les participants entrent leurs véritables informations (et non le nom donné dans la mise en contexte) ; la vidéo ne peut donc pas être anonymisée.	<ul style="list-style-type: none"> ○ Fournir aux participants les éléments nécessaires à leur inscription (nom, prénom, mail, mot de passe).

Tâche 2 : chercher et télécharger un jeu de données

Difficulté de la tâche : la moitié des participants a trouvé la tâche facile et l'autre moitié a jugé la difficulté comme étant moyenne.

Temps nécessaire : 1'15" – 6'00"

Éléments soulevés par les participants	Répercussions sur la maquette
Deux participants n'ont pas trouvé la barre de recherche instantanément et se sont d'abord rendus sur d'autres pages.	<ul style="list-style-type: none"> ○ Nous pourrions mieux mettre en évidence la recherche. Une solution est prévue plus bas.
Un participant se demande quel est l'opérateur booléen par défaut dans la recherche simple.	<i>Nous suggérons que l'opérateur booléen par défaut soit le « AND ».</i>
Un participant chercher avec le mot clé « linnaeus », ce qui fait que les mots surlignés dans les résultats de recherche ne sont pas les bons.	<ul style="list-style-type: none"> ○ Enlever le surlignement dans les résultats de recherche, du moins pour les tests utilisateurs.
Les participants demandent qu'une aide pour l'orthographe soit prévue.	<i>L'aide relative à l'orthographe sera mentionnée dans les recommandations d'implémentation, car cette fonctionnalité ne peut pas être mise en place sur la maquette.</i>

Certains participants trouvent qu'il n'est pas clair de devoir cliquer sur le titre du jeu de données pour pouvoir consulter le détail.	<ul style="list-style-type: none"> ○ Faire en sorte que le titre se souligne au passage de la souris.
Les icônes relatives à l'accès aux jeux de données ne sont pas claires pour les participants. Ils suggèrent de mettre en place un code couleur.	<ul style="list-style-type: none"> ○ Remplacer les icônes par un texte coloré.
Le nombre de vues et de téléchargements des jeux de données devrait être affiché sur la page de résultat de recherche.	<ul style="list-style-type: none"> ○ Ajouter cette option.
<p>Sur la page du détail du jeu de données, le titre « Stats » n'est pas utile.</p> <p>Le DOI devrait être mis en avant, et non l'unité organisationnelle.</p> <p>En général, il faut prioriser les métadonnées, notamment en vue du <i>Responsive design</i>, et l'ensemble des métadonnées doivent figurer sur la page de présentation des jeux de données.</p>	<ul style="list-style-type: none"> ○ Revoir la disposition des éléments d'information sur la page « Data description ». ○ Toutes les métadonnées doivent figurer sur cette page : ajouter les métadonnées manquantes.
Le bouton « Download » devrait être modifié en « Download all ».	<ul style="list-style-type: none"> ○ Remplacer « Download » par « Download all ». ○ Les icônes de téléchargement individuel devraient également être vert, pour plus de cohérence.
De manière générale, les participants semblent vouloir une liste des jeux de données déposés sur la plateforme ; la fonction de recherche ne leur suffit pas. Cependant, d'autres participants jugent ce souhait comme étant irréaliste, car ce serait trop lourd pour la plateforme. Ils suggèrent que la liste soit remplacée par les derniers jeux de données déposés.	<ul style="list-style-type: none"> ○ Proposer une liste complète des jeux de données sur la plateforme en proposant des options de filtre. ○ Remplacer le champ de recherche (à gauche) par un bouton « Search a dataset » qui mènerait à la liste. La fonction de recherche serait en haut de cette nouvelle page, comme elle l'est actuellement une fois que l'on se trouve sur la page des résultats. <p><i>Les développeurs de la plateforme seront plus à même de juger de la faisabilité de cette fonctionnalité et de choisir l'option à privilégier.</i></p>

Les participants aimeraient pouvoir accéder à la liste de l'ensemble des jeux de données d'une unité organisationnelle.	<ul style="list-style-type: none"> ○ Cela est déjà possible dans la maquette actuelle, mais la page sera améliorée.
Éléments remarquables par nous-mêmes	Répercussions sur la maquette
Le tri des jeux de données retournés par la recherche ne fonctionne pas.	<i>C'est exact, mais cet élément a été mentionné dans le guide de test, de façon à ce que les participants ne cherchent pas à faire fonctionner ces options. La maquette les fait figurés, mais elle n'a pas vocation à fonctionner comme la plateforme finale.</i>
La fonctionnalité de la punaise ne fonctionne pas et n'est pas claire pour les participants.	<ul style="list-style-type: none"> ○ Si l'on décide de ne pas la faire fonctionner pour les tests utilisateurs, ajouter le message « Feature not available », ainsi que « Save » au passage de la souris. ○ Créer la page des favoris, même si cette option ne sera pas fonctionnelle sur la maquette.

Tâche 3 : soumettre un jeu de données à la plateforme

Difficulté de la tâche : la moitié des participants a trouvé la tâche facile et l'autre moitié a jugé la difficulté comme étant moyenne.

Temps nécessaire : 1'28" – 9'48"

Éléments soulevés par les participants	Répercussions sur la maquette
<p>Au moment où le participant choisit le fichier à télécharger, les deux boutons « Delete » et « Add » prêtent à confusion : plusieurs participants se sont demandés s'il s'agissait déjà de publier le jeu de données.</p> <p>D'autres demande à ce qu'un <i>pop-up</i> soit associé à « Delete » pour que les fichiers ne soient pas supprimés par erreur.</p>	<ul style="list-style-type: none"> ○ Remplacer le bouton « Add » par un bouton « Back » qui réinitialiserait le champ. ○ Ajouter une option « Upload » pour finaliser l'ajout de fichiers. ○ Remplacer « Delete » par « Delete file(s) ». <p><i>Il sera recommandé que les fichiers puissent être organisés dans la fenêtre où ils figurent.</i></p>
Un participant apprécierait d'avoir une barre de progression dans le processus de soumission.	<ul style="list-style-type: none"> ○ Améliorer la barre de progression et la rendre cliquable.

Les participants demandent à ce qu'un champ soit ajouté pour que l'auteur puisse traduire le titre du jeu de donnée en plusieurs langues.	<ul style="list-style-type: none"> ○ Ajouter ce champ dans les métadonnées optionnelles.
Les participants aimeraient pouvoir associer les jeux de données déposés à plusieurs unités organisationnelles.	<ul style="list-style-type: none"> ○ Ajouter un bouton « + » à côté de la liste déroulante des unités organisationnelles qui permettrait de dupliquer la liste déroulante pour ajouter plusieurs unités, comme il est déjà possible de le faire pour les auteurs.
Il est suggéré de séparer le champs « Author » en « First name » et « Last name ».	<ul style="list-style-type: none"> ○ Adapter le champ en question.
<p>La métadonnée « Ressource type » n'est pas claire. Un participant nous demande pourquoi elle est collectée, comment juger de sa fiabilité. Selon lui, il est compliqué de ne choisir qu'un seul élément de la liste.</p> <p>Le participant rencontre les mêmes questions par rapport à la métadonnée « Primary data type » : comment être suffisamment précis, alors que la liste n'est pas expliquée ?</p>	<p><i>Toutes les métadonnées devront se voir liées à une aide adéquate ; cependant, la mise en place d'aides détaillées ne sera pas possible dans la maquette. Il faut aussi qu'il soit annoncé que le modèle de métadonnées implémenté est basé sur DataCite.</i></p>
Lorsque des dates sont demandées, il nous est suggéré de proposer un ajout automatique de la date du jour.	<ul style="list-style-type: none"> ○ Faire en sorte que la date de publication suggérées par la plateforme soit la date du jour.
Il manque une aide pour les licences.	<ul style="list-style-type: none"> ○ Une aide associée à la métadonnée « Licence » devrait afficher une présentation des licences autorisées.
<p>Le DOI est ajouté automatiquement à chaque jeu de données ; il n'est donc pas nécessaire dans le formulaire.</p> <p>De plus, certains participants se demandent s'il s'agit ici du DOI du jeu de données, ou de celui de la publication.</p>	<ul style="list-style-type: none"> ○ Laisser uniquement l'option « Request un DOI » en bas du formulaire, avec une aide. ○ Renommer le champ « Data's DOI ».

Le champs « Language » n'est pas clair et les usagers aimeraient pouvoir associer plusieurs langues à leurs jeux de données.	<ul style="list-style-type: none"> ○ Remplacer le nom de cette métadonnée par « Data language ». ○ Prévoir une option pour dupliquer le champ.
Le champs « Preservation policies » ne devrait pas contenir de termes en latin, car tout le monde ne le comprend pas.	<ul style="list-style-type: none"> ○ Remplacer « Ad Eternam », par « Forever ».
Les participants ne comprennent pas la différence entre « Title », « Subject » et « Domain ».	<ul style="list-style-type: none"> ○ Supprimer « Subject », puisqu'il s'apparente à « Domain ».
Le champ « Submission policies » prête à confusion pour les utilisateurs qui suggèrent qu'il soit rempli par le manager au moment de la validation.	<i>Ce point dépendra des discussions du DLCM à propos des unités organisationnelles. Aussi, nous décidons de maintenir ce champ dans le formulaire de soumission d'un dépôt et d'aborder ce point dans nos recommandations.</i>
<p>Le champ « Domain » n'est pas un champ obligatoire, alors qu'il est utilisé pour trier les données : ce n'est pas cohérent.</p> <p>Par ailleurs, les participants nous rendent attentives au fait que déterminer la liste des domaines sera une tâche compliquée, car il faut être général, tout en couvrant l'ensemble des domaines scientifiques. Ils suggèrent de pouvoir dupliquer ce champ pour qu'un jeu de données puisse être interdisciplinaire.</p>	<ul style="list-style-type: none"> ○ Déplacer ce champ de métadonnées dans les métadonnées obligatoires. ○ Ajouter une option « + » pour que le champ puisse être dupliqué.
Les participants ne comprennent pas la différence entre les auteurs et les contributeurs, et le champ prévu à cet effet n'est pas complet.	<ul style="list-style-type: none"> ○ Associer le champ existant à une liste déroulante des rôles des contributeurs envisagés par DataCite et prévoir une aide. ○ Renommer le champ « Contributor » « Other contributor ».
De manière générale, les contenus attendus dans les champs ne sont pas clairs.	<ul style="list-style-type: none"> ○ Tous les champs ouverts vides doivent donner des exemples ou clarifier le type de contenu attendu.
Les participants nous interrogent sur les métadonnées choisies, même s'ils les trouvent le plus souvent adéquates.	<ul style="list-style-type: none"> ○ Revoir les métadonnées en fonction des recommandations du DLCM et du schéma de métadonnées DataCite.

<p>Les participants se demandent comment saisir des métadonnées en plusieurs langues.</p> <p>Ils aimeraient que les mots clés soient traduits.</p>	<p><i>La possibilité de traduire les métadonnées devra être explorée par les développeurs du DLCM. Nous ne pouvons que suggérer cette possibilité.</i></p> <p><i>Envisager la traduction des mots clés comme recommandations lors de la rédaction du rapport.</i></p>
<p>Le bouton « Save » du formulaire prête à confusion. Les participants qui sauvegardent le formulaire de dépôt dès la première page n'atteignent pas la page « Review » et sont déstabilisés de rester sur la même page, comme si rien ne se produisait.</p> <p>Les participants souhaitent que ce bouton « Save » soit expliqué.</p>	<ul style="list-style-type: none"> ○ Supprimer le bouton « Save » sur la page « Review ». ○ Ajouter une information sur la localisation du jeu de données créé sous forme de <i>pop-up</i> au moment de la soumission. <p><i>Il sera recommandé de mettre en place une sauvegarde automatique sur le formulaire de la plateforme finale, et de créer un menu qui reste affiché en tout temps sur la page.</i></p>
<p>Lorsque les participants utilisent le bouton « Next » sans avoir complété les champs obligatoires, ils savent qu'il manque des informations, mais ils ne savent pas lesquelles.</p>	<p><i>Il sera recommandé que le bouton « Next » renvoie l'utilisateur aux informations manquantes obligatoires qu'il lui reste à compléter.</i></p>
<p>Certains participants souhaitent que les intitulés des métadonnées (obligatoires, recommandées, optionnelles) figurent sur la page « Review ».</p>	<ul style="list-style-type: none"> ○ Prévoir cette distinction.
<p>Sur la page « My data », les participants ne comprennent pas ce à quoi sert « Rejected », ni sa différence avec « Incorrect ».</p>	<ul style="list-style-type: none"> ○ Sur l'onglet en question, il faudrait afficher le message « Rejected datasets must be resubmitted ».
<p>Les participants demandent à ce qu'un moyen soit prévu pour que le déposant des données puisse supprimer ses jeux de données s'il le souhaite. Cependant, dans ce cas, le manager ne devrait pas pouvoir prendre la décision seul.</p>	<ul style="list-style-type: none"> ○ Mettre en place l'option de suppression dans « My data ». ○ Ajouter un champ dans le formulaire de dépôt pour que le déposant puisse choisir le membre de l'unité qui validera son jeu de donnée. <p><i>Le DLCM devrait s'intéresser à la suppression de jeux de données et la procédure précise à prévoir.</i></p>

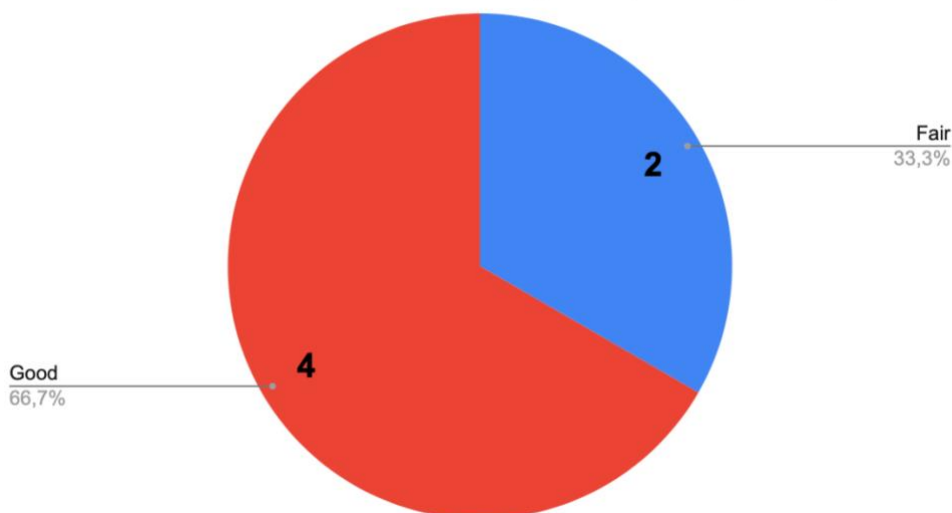
De plus, les participants aimeraient pouvoir choisir la personne de l'unité qui validera ou non leurs jeux de données.	
De manière générale, les participants trouvent le formulaire « assez facile à manipuler » et apprécient l'option de retour en arrière pour corriger les données erronées ou incomplètes.	
Éléments remarquables par nous-mêmes	Répercussions sur la maquette
Les métadonnées « Access » et « Embargo » doivent être améliorées. Le fait de déclencher les boutons radio dans un second temps attirera l'attention des participants dessus, car certains participants déclarent ne pas les avoir vus.	<ul style="list-style-type: none"> ○ Le choix « Restricted » doit afficher les boutons radio : le premier bouton radio déclenche une liste déroulante avec les unités possibles, alors que le second bouton radio déclenche un champ libre avec un texte indicatif, comme « Describe the conditions of access to the dataset. E.g.: if you want to access this dataset, contact Mr. Smith at the address max.smith@dlcm.ch ». Les boutons radio doivent pouvoir être décochés si le choix d'accès est modifié. ○ Le choix « Closed » ne doit rien afficher. ○ Lorsque l'utilisateur choisit l'embargo, il doit pouvoir entrer la date de fin d'embargo.
Il pourrait être intéressant de mettre en valeur les mots clés.	<i>Il sera recommandé de faire en sorte que les mots clés soient validés par la touche « Enter » ou le séparateur « ; » pour que les mots clés soient mis en valeur dans des bulles.</i>
Puisque la complétion des métadonnées semble compliquée pour les utilisateurs, suggérer d'ajouter au « Starter kit » de la page d'accueil un modèle de dépôt.	<ul style="list-style-type: none"> ○ Ajouter sur le « Starter kit » de la page d'accueil une ligne « Example of filing ». <i>Cet élément figurera dans les recommandations d'implémentation de la maquette.</i>
Un participant a pu passer à la page « Review » sans avoir chargé de jeu de données. Cela ne devrait pas être possible,	<ul style="list-style-type: none"> ○ Ajouter une information à propos de l'obligation de déposer des fichiers,

car la zone « Drag and Drop » est obligatoire.	et rendre obligatoire pour passer à la page « Review ».
La page « Review » ne donne aucun aperçu des fichiers chargés et cette page devrait être nommée « Preview ».	<ul style="list-style-type: none"> ○ Ajouter cette option. ○ Remplacer « Review » par « Preview ».
Lors des tests, nous partons du principe que les participants sont nouveaux utilisateurs de la plateforme ; aussi, la page « My data – My dataset » présente des incohérences.	<ul style="list-style-type: none"> ○ Il faut supprimer les jeux de données « Completed ». ○ « In progress » ne devrait contenir que le fichier <i>uploadé</i>, pour autant que le participant ait sauvegardé, et non soumis, son jeu de données depuis le formulaire de dépôt. ○ « Waiting for validation » doit se voir supprimer les boutons « Accept » et « Refuse ». ○ « Rejected » et « Incorrect » ne doivent pas contenir de jeux de données.
Lors des tests, nous partons du principe que les participants n'occupent pas le rôle de « manager » pour leur unité organisationnelle ; aussi, la page « My data – Linnaeus Team dataset » présente des incohérences.	<ul style="list-style-type: none"> ○ « In progress » ne doit pas avoir de bouton « Submit », ni d'icône d'action.
Le prochain test doit prévoir une tâche liée à la page « My data ». Pour cette tâche, le participant doit jouer le rôle de manager d'une unité organisationnelle. Ainsi, il doit déjà avoir des données déposées dans « My data – My datasets », être lié à plusieurs unités qui ont des jeux de données en attente de validation.	<ul style="list-style-type: none"> ○ Préparer cette tâche.
Sur la page « My data », « Rejected » et « Incorrect », le bouton d'information ne semble pas suffisamment clair.	<ul style="list-style-type: none"> ○ Mettre le contenu du bouton dans une colonne à cet effet.

Remarques générales

D'abord, tous les participants ont été satisfaits de l'utilisabilité de la maquette :

Evaluation de l'utilisabilité de la maquette par les participants



Ensuite, la plupart des participants ne lisent pas le contenu des pages. Aussi, il faudrait que les informations indispensables à l'utilisation des fonctionnalités principales de la plateforme soient mises en évidence.

Par ailleurs, deux participants considèrent que la page d'accueil donne trop d'importance aux informations superflues : les statistiques et les fonctionnalités essentielles devraient être mises en avant. Aussi, il nous est suggéré de donner plus d'importance au *header*, d'afficher « My data » et « My units » en permanence et de rediriger vers la connexion si l'utilisateur n'est pas encore connecté.

Sur cette même page, le carrousel qui présente les avantages du portail annonce que le dépôt est gratuit pour les jeux de données jusqu'à 50 GB. Deux participants nous informent que, en ce qui les concerne, 50 GB s'apparente plus à un fichier qu'à un jeu de donnée. Ils demandent aussi à ce que le carrousel fasse mention de la vitesse de téléchargement des fichiers (*upload* et *download*).

Finalement, certains participants n'étaient pas familiers de la pastille verte de *Loop11* qui permet de consulter les consignes, valider une tâche ou renoncer à une tâche. Lors des prochains tests, nous devons présenter cette fenêtre de manière plus claire aux répondants.

Annexe 21 : Compte rendu du deuxième test utilisateur

Modalités du test

La deuxième maquette a été créée sur le logiciel *Axure* et est accessible par le lien suivant : <https://zlrghw.axshare.com>.

Le test a été réalisé au moyen du logiciel *Loop11*, qui permet de donner accès au test via un lien URL et d'enregistrer l'écran des participants ; voici le lien URL menant au deuxième test : https://www.loop11.com/ui/?l11_uid=61366.

Le test s'est déroulé sur deux jours :

Zurich, vendredi 27 septembre 2019	Genève, mardi 1 ^{er} octobre 2019
5 participants qui ont signé un formulaire de consentement.	2 participants ont signé un formulaire de consentement commun, car ils ont réalisé le test sur le même écran.

Les participants ont dû réaliser quatre tâches :

4. Se connecter à la plateforme et rejoindre une unité organisationnelle donnée ;
5. Chercher et télécharger un jeu de données indiqué.
6. Soumettre (*upload*) un jeu de données à la plateforme.
7. Valider la publication d'un jeu de données sur la plateforme.

Après chacune d'entre elles, il était demandé aux participants de juger de la difficulté de l'exercice. La réalisation des tâches a duré entre 16'09" et 39'37" minutes et s'est suivie par une discussion d'une heure et demie la première fois, et trente minutes la deuxième. Ce compte-rendu est structuré en tâche, les informations récoltées lors du test et lors de la discussion étant rassemblées par tâche.


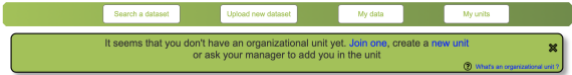
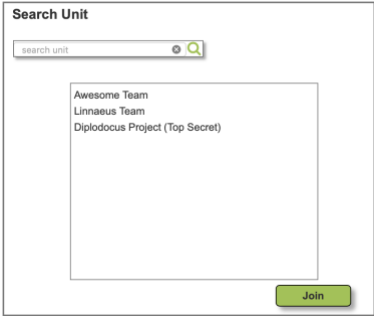
Le rapport PDF de *Loop11* est disponible en annexe de ce document. Il contient certains commentaires des participants, ainsi que les informations sur les tâches. Toutefois, il ne faut pas tenir compte des statistiques de réussite, celles-ci étant parfois faussées lorsque les participants n'étaient pas sur la page finale de la tâche au moment où il la signalait comme atteinte ; de plus, les statistiques comptent six tests, alors qu'il y a eu sept participants (deux ayant dû faire le test sur le même ordinateur).

Compte-rendu

Tâche 1 : se connecter à la plateforme et rejoindre une unité organisationnelle donnée

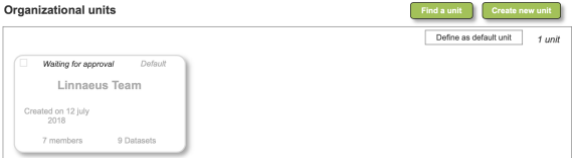


Difficulté de la tâche : Six personnes ont jugé la tâche facile, et une l'a évalué comme moyenne. Notons que la personne qui a jugé la difficulté de la tâche comme moyenne a d'abord essayé de passer par la recherche ; une fois que la personne s'est connectée, la tâche a été accomplie rapidement et sans problème.

Temps nécessaire : 1'58" - 3'52"

Éléments soulevés par les participants	Répercussions sur la maquette
<p>Sur le formulaire d'inscription à la plateforme, certains participants ont été troublés de devoir accepter des « Terms & Conditions » qu'ils ne pouvait pas consulter.</p>  <p>Hormis ce détail, les participants ont trouvé la procédure d'enregistrement habituelle et simple, et le message d'orientation vers la liste des unités organisationnelles agréables.</p> 	<ul style="list-style-type: none"> ○ Souligner « Terms & Conditions », comme s'il s'agissait d'un bouton cliquable. ○ Ajouter la mention « Feature not available » au passage de la souris. <p><i>Le contenu de la future page « Terms & Conditions » devra être construit par le DLCM.</i></p>
<p>Sur la page « M units », les participants ont apprécié la mise en évidence de la liste des unités existantes. Cela a facilité la réalisation de la tâche 1.</p> 	
<p>Sur la page « My units », les participants s'interrogent sur le bien-fondé de l'option « Create new unit ». Le DLCM ne risque-t-il</p>	<p><i>Ces discussions, déjà évoquées dans le compte-rendu du premier test, appartiennent aux DLCM. Les conclusions auront</i></p>

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données


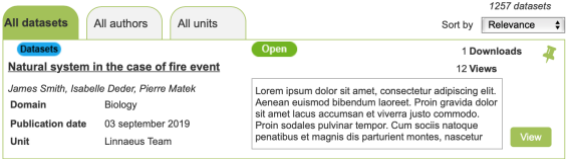
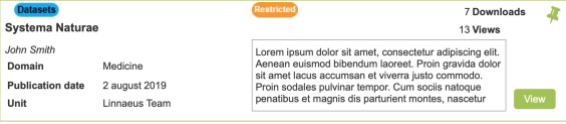
<p>pas d'être submergé de demande de création d'unités organisationnelles ? Il risque également des enregistrements multiples par personne.</p> <div data-bbox="336 360 715 405"> Find a unit Create new unit </div>	<p><i>forcément des répercussions sur la plateforme. Cependant, nous choisissons de maintenir la possibilité de créer une nouvelle unité organisationnelle pour le moment, de manière à simplifier la communication entre les utilisateurs et le DLCM à ce sujet.</i></p>
<p>Deux participants estiment que les usagers ne devraient pas pouvoir s'inscrire d'eux-mêmes dans une unité : ils devraient être ajoutés par le manager. Ainsi, certains participants pensent que l'utilisateur devrait être invité dans une unité pour pouvoir, ensuite, créer un compte. Ce commentaire est à associer avec une certaine réticence au partage de données sans condition avec tout utilisateur (principes FAIR).</p> <p>Ils craignent aussi que le nombre d'inscrits ne soit multiplié par des utilisateurs qui auraient plusieurs comptes ou qui n'auraient « rien à faire sur la plateforme ».</p>	<p><i>Ces discussions appartiennent aux DLCM ; leurs conclusions auront forcément des répercussions sur la plateforme. Comme il s'agit de décision stratégique, nous ne nous positionnons pas sur la question et maintenons la fonctionnalité en l'état sur la maquette.</i></p> <p><i>En ce qui concerne les demandes spontanées pour rejoindre des unités, nous avons proposé une alternative à ce qu'avait déjà imaginé le DLCM, à savoir une association sur invitation. Si notre option trouble les usagers, il faut effectivement privilégier l'option initiale du DLCM.</i></p> <p><i>Quant à la réticence éventuelle des usagers à partager leurs données de manière ouverte, il est toujours possible de publier les données selon un accès restreint ou clos.</i></p> <p><i>Nous suggérons que les inscriptions puissent se faire sans que l'utilisateur ne soit nécessairement associé à une unité, puisque les principes FAIR implique un accès le plus large possible aux données et que la connexion offre plus de fonctionnalités de recherche.</i></p>
<p>De manière générale, les participants au deuxième test, tout comme les participants aux entretiens et au premier test, s'interrogent sur le concept d'unité organisationnelle, sa signification, son fonctionnement et ses implications. Ils s'inquiètent de ce qu'il adviendra en cas de départ à la retraite, de décès, etc.</p>	<p><i>Ces discussions, déjà évoquées dans le compte-rendu du premier test, appartiennent aux DLCM. Les conclusions auront forcément des répercussions sur la plateforme. Une meilleure communication à ce sujet semble fondamentale.</i></p>

<p>Sur la page « My units », les participants se demandent si l’affichage des unités en bloc est pertinent. Si l’usager est associé à beaucoup d’unités, ils préféreraient que l’affichage soit sous forme de liste, comme dans la page de recherche.</p> 	<ul style="list-style-type: none"> ○ Prévoir une option pour que l’usager puisse choisir entre l’affichage en blocs ou en liste, par exemple : 
<p>Deux participants sont passés par la recherche pour rejoindre l’unité organisationnelle et n’ont rencontré aucune difficulté.</p>	
Éléments remarquables par nous-mêmes	Répercussions sur la maquette
<p>Les participants essaient d’utiliser la touche « Tab » pour naviguer entre les champs de connexion, alors que cela ne fonctionne pas sur la maquette.</p>	<p><i>Cette fonctionnalité pourrait être paramétrée sur la maquette, mais nous avons choisi d’y renoncer au vu du temps supplémentaire nécessaire au paramétrage de chaque champ, alors que ce maque n’incarne pas un problème d’utilisabilité pour la plateforme. En revanche, nous recommanderons aux développeurs du DLCM d’implémenter cette option.</i></p>
<p>Plusieurs participant ont cherché les unités organisationnelles par la page de recherche. Le bouton « Search a dataset » peut prêter à confusion dans la mesure où des auteurs et des unités organisationnelles peuvent aussi être cherchés et trouvés par ce biais.</p> 	<ul style="list-style-type: none"> ○ Renommer le bouton « Search a dataset » du <i>header</i> en « Search ».

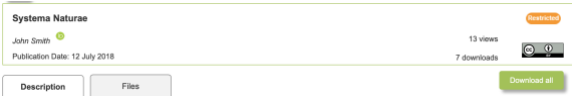

Tâche 2 : chercher et télécharger un jeu de données

Difficulté de la tâche : Cinq personnes ont jugé la tâche facile et deux l'ont évalué comme moyenne.

Temps nécessaire : 2'21" - 13'34"

Éléments soulevés par les participants	Répercussions sur la maquette
<p>Dès la deuxième tâche, le nom du participant ne s'affiche plus dans la barre du profil. Aussi, les utilisateurs ne comprennent pas l'utilité de la barre ; ils la prennent parfois pour une barre de recherche.</p> 	<p><i>En effet, il a fallu fractionner la maquette pour mieux simuler les tâches du test utilisateur ; aussi, les informations de l'étape de connexion (tâche 1) ne sont pas reprises dans la tâche 2 pour que le mail de connexion s'affiche dans la barre du profil. Nous résoudrons ce problème pour le prochain test.</i></p>
<p>Sur la page des résultats de recherche, il nous est suggéré de revoir la disposition des éléments d'information : l'auteur et l'affiliation doivent figurer, contrairement à l'unité organisationnelle qui ne leur semble pas utile.</p> 	<ul style="list-style-type: none"> ○ Revoir la disposition et hiérarchiser les éléments d'information présentant les jeux de données. ○ Remplacer l'affichage de l'unité organisationnelle par l'affichage de l'affiliation. ○ Afficher l'auteur en gras, et non en italique.
<p>Il était demandé aux participants de chercher un jeu de données « à propos » de "Systema Naturae" de l'auteur Carl Linnaeus. Dans les résultats de recherche, il se trouve que ces termes se retrouvent dans plusieurs jeux de données. De plus, un jeu porte ce titre, sans que son auteur ne soit Carl Linnaeus, et son accès est restreint. Les participants n'ont pas apprécié être induit en erreur.</p>  <p>En plus, ils nous font remarquer qu'il n'est pas cohérent d'avoir une option de</p>	<ul style="list-style-type: none"> ○ Reformuler la consigne de façon plus précise. ○ Griser les boutons de téléchargement du jeu de données dont le titre est « Systema Naturae », puisque son accès est restreint.

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données


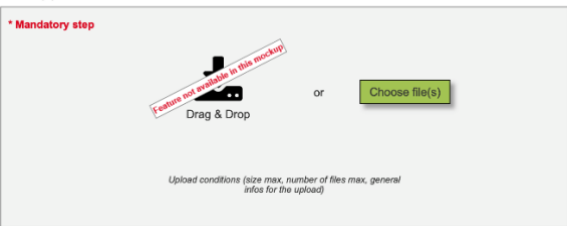
<p>téléchargement sur la page de détail d'un jeu de données dont l'accès est restreint.</p> 	
<p>Certains participants demandent à ce que les termes utilisés dans la recherche soient surlignés dans les résultats de recherche, de façon à accélérer le parcours des résultats.</p>	<p><i>Nous avons prévu cette option dans le premier test, mais elle a été enlevée, car le surlignage n'était pas adaptable en fonction des termes de recherche entrés. Nous ne comptons pas la rétablir, mais il sera recommandé au DLCM de prévoir cette fonctionnalité.</i></p>
<p>Sur la page des résultats de recherche, la fonctionnalité de la punaise n'est pas claire.</p> 	<ul style="list-style-type: none"> ○ Renommer l'aide « Add to favorites ». ○ Créer la page « Favorites » si nous avons le temps avant le prochain test.
<p>L'affiliation de l'auteur doit figurer sur la page de détail des jeux de données.</p>	<ul style="list-style-type: none"> ○ Ajouter cet élément d'information.
<p>Deux participants nous demandent de prévoir un thésaurus pour l'indexation des jeux de données, ainsi que des suggestions de jeux de données en fonction des préférences des utilisateurs.</p>	<p><i>Il n'est pas réaliste de prévoir cet outil sur la maquette au vu du temps à disposition pour le projet. Cependant, cette demande sera transmise au DLCM sous forme de recommandation.</i></p>

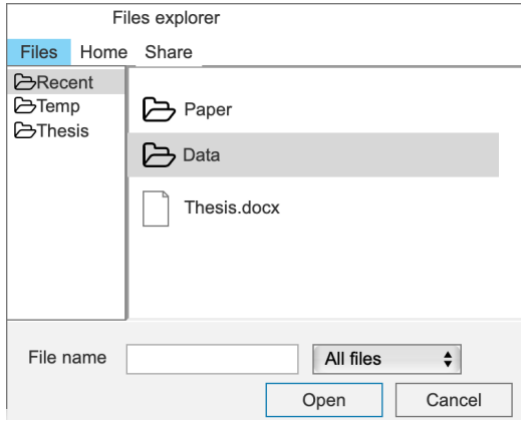


Tâche 3 : soumettre un jeu de données à la plateforme


Difficulté de la tâche : Cinq personnes ont jugé la tâche facile, alors que deux l'ont évalué comme moyenne.




Temps nécessaire : 2'52" – 19'50"

Éléments soulevés par les participants	Répercussions sur la maquette
<p>Les participants ont trouvé le processus de soumission très bien ! Ils ont apprécié que les champs présentent des exemples des contenus attendus, mais suggèrent que les</p>	<ul style="list-style-type: none"> ○ Inverser les champs « Last name » et « First name ».

<p>champs de nom de famille et de prénom soient inversés.</p> <p>Author(s) * <input type="text" value="Brown"/> <input type="text" value="James"/> + Add me + New author</p> <p>Other contributor(s) <input type="text" value="Brown"/> <input type="text" value="James"/> <input type="text" value="ContactPerson"/> + New contributor</p>	
<p>Un participant considère notre formulaire de soumission comme étant plus facile à utiliser que celui de la plateforme <i>sandbox</i> du DLCM. Il trouve les aides utiles, même si certaines sont encore à travailler (ex : « Submission policies »).</p>	<p><i>Nous ne travaillerons pas le contenu exact des aides, car cette tâche sort de notre mission qui porte sur l'amélioration de l'utilisabilité de la plateforme.</i></p>
<p>Les participants ont apprécié la barre de progression du formulaire de dépôt de jeux de données. Certains souhaiteraient néanmoins une barre plus développée.</p> 	<ul style="list-style-type: none"> ○ Séparer la barre en plusieurs parties : <i>files, metadata, preview, submit.</i>
<p>La fonctionnalité de <i>Drag and Drop</i> est très appréciée.</p> 	
<p>Dans le formulaire de dépôt d'un jeu de données, le champ « File name » de l'organisateur de fichier prête à confusion. Les participants sont confus, car lorsque l'on clique sur le dossier « Data » le titre devrait s'afficher dans le champ « File name » ; comme ce n'est pas le cas, les participants ont l'impression de devoir renommer leur dossier à télécharger.</p>	<p><i>En réalité, il s'agit d'un champ qui affiche le nom du fichier sélectionné avec la souris.</i></p> <ul style="list-style-type: none"> ○ Simuler le titre du dossier dans le champ.

	
<p>Les participants ont apprécié la hiérarchisation des métadonnées et leur présentation sous forme de menu déroulant. Cependant, ils nous signalent que les flèches indiquant cette possibilité ne sont pas dans le bon sens.</p> 	<ul style="list-style-type: none"> ○ Changer le sens des flèches des catégories de métadonnées : la flèche pour dérouler les métadonnées doit pointer vers le bas, alors que celle pour les enrouler doit pointer vers le haut.
<p>La métadonnée « Primary data type » n'est pas forcément adéquate. Les participants suggèrent de pouvoir choisir plusieurs types par jeu de données.</p>	<ul style="list-style-type: none"> ○ Déplacer cette métadonnée dans les métadonnées optionnelles. ○ Prévoir la répétition du champ, à l'aide d'un bouton « + ».
<p>Un participant nous demande pourquoi la personne de contact serait ajoutée dans la métadonnée « Other contributor(s) », alors que l'auteur pourrait être une personne de contact.</p>	<p><i>Les rôles proposés dans « Other contributor(s) » sont issus du schéma de métadonnées de DataCite. Le champ en question permet d'orienter les usagers vers la bonne personne si l'auteur ne souhaite pas être la personne à contacter en cas de besoin.</i></p>
<p>Sur le formulaire de soumission d'un jeu de données, les participants ne comprennent pas qu'il est possible de sauvegarder le travail en cours, afin de le terminer plus tard.</p> 	<ul style="list-style-type: none"> ○ Renommer le bouton « Save » de la première page du formulaire en « Save and quit » et l'associer à un <i>pop-up</i> qui indiquerait où retrouver les données sauvegardées, à savoir sur la page « My data – In progress ». ○ Mettre le bouton « Cancel » de la première page du formulaire sur la gauche.


<p>Deux participants veulent qu'il soit possible d'ajouter des métadonnées au niveau du fichier. Un champ ouvert leur suffirait, mais cette fonctionnalité est essentielle pour permettre la réutilisation des fichiers et donner du sens.</p> <p>Les deux participants en question sont archéologues et se demandent comment rassembler plusieurs photos d'un objet à différent stade de sa vie (ex : objet sur la fouille archéologique, avant nettoyage, après nettoyage et dans un musée).</p>	<p><i>Avec la deuxième partie de la remarque du participant, nous avons l'impression qu'il attend que la plateforme prépare automatiquement ses données, alors que ce n'est pas possible. La fonctionnalité à laquelle il s'intéresse ne sera possible qu'à travers l'organisation des fichiers en dossiers au moment du versement.</i></p>
<p>Un participant a été perturbé, par les exemples de complétion des champs libres du formulaire : il les a pris pour des informations pré-remplies.</p> 	<p><i>Il s'agit d'exemples destinés à aider les participants. Nous ne pouvons pas être plus claires à ce sujet, d'autant plus qu'il s'agit d'une pratique courante dans la réalisation de formulaire en ligne.</i></p>
<p>Deux participants aimeraient que les informations requises dans le formulaire soient connues à l'avance, en cas de délégation de la tâche de dépôt à un collaborateur différent de celui qui a enregistré les données.</p>	<ul style="list-style-type: none"> ○ Lors de l'ouverture du formulaire de dépôt, prévoir un <i>pop-up</i> qui inviterait les usagers à consulter les outils du « Starter kit » de la page d'accueil.
<p>Certains participants sont inquiétés par les formats et les tailles de fichiers que la plateforme peut supporter. Ils souhaitent que ces informations figurent sur la plateforme.</p>	<p><i>Nous ne sommes pas à même de répondre à cette question, mais nous suggérons que le DLCM indique cette information sur le « Starter kit » de la page d'accueil.</i></p>
<p>De manière générale, les participants aimeraient que les métadonnées puissent être complétées par plusieurs personnes, car personne n'a toutes les informations nécessaires, et c'est là que réside, selon eux, l'intérêt des unités.</p>	<ul style="list-style-type: none"> ○ Prévoir la gestion des droits des membres des unités sur la page des unités existantes. ○ Prévoir la possibilité d'indiquer les personnes autorisées à compléter les métadonnées dans le formulaire de dépôt de jeux de données.
<p>Éléments remarquables par nous-mêmes</p>	<p>Répercussions sur la maquette</p>


Les participants ont de la peine à lire les aides du formulaire de dépôt.	<ul style="list-style-type: none"> ○ Agrandir la police de l'affichage des aides du formulaire.
<p>Une erreur persiste dans le formulaire du dépôt, au niveau du champ « Access ». Si l'utilisateur clique sur « Restricted », les boutons radio s'affichent et fonctionnent, mais s'il clique ensuite sur « Open » ou « Closed », il n'est plus possible de réafficher les boutons radio associés à « Restricted ».</p> 	<p><i>Malheureusement, ce problème ne peut pas être résolu pour le moment. De plus, il s'agit d'une erreur de fonctionnement de la maquette et non d'un problème d'utilisabilité au niveau de la plateforme.</i></p>
<p>Sur le formulaire de dépôt, si l'utilisateur souhaite associer un embargo à son jeu de données, aucun champ ne s'affiche pour lui permettre de fixer la date de fin d'embargo.</p> 	<p><i>Le champ en question ne sera pas ajouté, actuellement, le DLCM a prévu cette fonction et l'a mise en place d'une façon très agréable pour les usagers. Nous lui recommanderons de maintenir cette fonctionnalité.</i></p>
<p>Lorsque les participants essaient de finaliser le formulaire de dépôt sans avoir rempli les champs obligatoires, il n'est pas possible d'arriver sur la page « Preview », mais rien d'orienter les usagers vers les champs incomplets.</p>	<p><i>Il n'est pas possible de mettre en place cette fonctionnalité sur la maquette, mais il sera recommandé au DLCM de l'implémenter sur la plateforme finale.</i></p>
<p>La métadonnée « Access » n'est pas reportée sur la page « Preview ».</p> 	<ul style="list-style-type: none"> ○ Chercher la cause de l'erreur et y remédier si possible.
<p>Certains participants ont été troublés par les boutons de certains messages d'avertissement : ils nous ont fait remarquer que la formulation de boutons n'était parfois pas explicite.</p>	<ul style="list-style-type: none"> ○ Vérifier les boutons, de façon à les clarifier.

Tâche 4 : valider la publication d'un jeu de données sur la plateforme

Difficulté de la tâche : Cinq personnes ont jugé la tâche moyenne, alors que deux l'ont évalué comme difficile. Notons que deux des personnes qui ont jugé la tâche moyenne n'ont pas réussi l'exercice et qu'une autre personne a dû abandonner, en grande partie à cause d'une question de temps.

Temps nécessaire : 4'17" - 9'03"

Éléments soulevés par les participants	Répercussions sur la maquette
Un participant trouve la distinction entre « My data » et « My units » confuse.	<i>Nous ne prévoyons pas de rassembler ces pages, car « My units » joue un rôle essentiellement administratif, alors que « My data » joue un rôle opérationnel pour l'utilisateur dans la gestion de ces jeux de données. En revanche, une solution est envisagée plus bas pour clarifier le header.</i>
La plupart des participants essaient d'accéder à la page de validation de jeux de données en passant par la page « My units », alors que rien n'est prévu dans ce sens.	<ul style="list-style-type: none"> ○ Prévoir un lien pour que l'utilisateur puisse accéder aux jeux de données de ses unités organisationnelles depuis la page « My units ».
<p>Certains participants ont essayé de réaliser la tâche en passant par les notifications, alors que nous n'avions pas envisagé cette option.</p>  <p>D'ailleurs, à propos des trois icônes du haut (punaise, message et notification), les participants s'interrogent sur leur affichage.</p>	<ul style="list-style-type: none"> ○ Faire en sorte que les notifications se déroulent devant le <i>header</i> ; idem pour les messages. ○ Lorsque l'utilisateur clique sur la notification en question, il doit être dirigé sur la page des jeux de données de l'unité organisationnelle en question. <p><i>Rien ne sera entrepris pour modifier l'affichage des icônes, car il est habituel de les placer en haute à droite de la page web.</i></p>
Certains participants passent par la recherche pour trouver l'auteur, puis accéder à ses jeux de données, mais cela ne fonctionne pas.	<p><i>Le jeu de données n'étant pas encore publié, il ne sera pas affiché dans les résultats de la recherche. Rien ne sera donc entrepris pour cela.</i></p> <ul style="list-style-type: none"> ○ Créer la page de l'auteur impliqué dans la tâche sur la page de résultat de recherche, tout en indiquant que

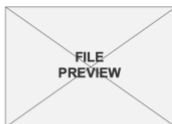
	les jeux de données affichés sont ceux qui ont déjà été publiés.																																
<p>Les participants ne trouvent pas logique que les jeux de données de leur unité organisationnelle se trouve sur la page « My data », c'est-pourquoi ils n'ont pas tout de suite trouvé la bonne page.</p> <p>Ils suggèrent de renommer la page « My data » en « Dataset ».</p> <div><div>My data</div></div>	<ul style="list-style-type: none">○ Remplacer le nom de la page « My data » en « My workspace ». <p><i>Nous ne voulons pas utiliser le terme « Dataset », car nous estimons que cela pourrait induire une confusion entre cette page et la page de recherche.</i></p>																																
<p>Les participants qui passent par la page « My data » ne voient pas le <i>toggle</i> ou ne savent pas comment l'utiliser. Certains jugent que ce bouton est associé à trop de fonctionnalités.</p> <div><div>My dataset</div><div></div><div>Linnaeus Team (default) ▾</div></div>	<ul style="list-style-type: none">○ Supprimer le <i>toggle</i> et dissocier l'option de navigation entre « My data » et « My unit's data », de l'option de navigation entre les jeux de données des différentes unités organisationnelles. <p><i>Quelle que soit les options choisies, elles doivent être visibles et claires.</i></p>																																
<p>Lorsque le manager voit le détail du jeu de données à valider, un participant s'étonne de voir un « Publication date », alors que le jeu de données n'est justement pas encore publié.</p> <div><div>Preview</div><div><table><thead><tr><th>Title</th><th>Size</th><th>Format</th><th>Creation date</th></tr></thead><tbody><tr><td>Data.csv</td><td>200 Mo</td><td>Tabi</td><td>27/09/19</td></tr><tr><td>Readme.txt</td><td>13 Mo</td><td>Text</td><td>27/09/19</td></tr><tr><td>Data_2.doc</td><td>14 Mo</td><td>Text</td><td>27/09/19</td></tr><tr><td>Transcripts</td><td>15 Go</td><td>Folder</td><td>27/09/19</td></tr><tr><td>Interview_1.flacc</td><td>5 Go</td><td>Video</td><td>27/09/19</td></tr><tr><td>Interview_2.flacc</td><td>4 Go</td><td>Video</td><td>27/09/19</td></tr><tr><td>Interview_3.flacc</td><td>6 Go</td><td>Video</td><td>27/09/19</td></tr></tbody></table><div><div>FILE</div><div>PREVIEW</div></div></div><div><div>Mandatory</div><div><div>Organizational unit *</div><div>Manager *</div><div>Title *</div><div>Author(s) *</div><div>Description *</div><div>Resource type *</div><div>Primary data type *</div><div>Publication date *</div></div><div><div>Linnaeus Team</div><div>Jack Pott</div><div>My favorite flower</div><div>Albert Kunz</div><div>I present to you my best flower from my garden</div><div>Dataset</div><div>Observation</div><div>27 September 2019</div></div></div></div>	Title	Size	Format	Creation date	Data.csv	200 Mo	Tabi	27/09/19	Readme.txt	13 Mo	Text	27/09/19	Data_2.doc	14 Mo	Text	27/09/19	Transcripts	15 Go	Folder	27/09/19	Interview_1.flacc	5 Go	Video	27/09/19	Interview_2.flacc	4 Go	Video	27/09/19	Interview_3.flacc	6 Go	Video	27/09/19	<ul style="list-style-type: none">○ Supprimer le champ « Publication date » puisque, de toute façon, il s'agira d'un champ automatique, tout comme le DOI dont il était question dans le compte-rendu du premier test utilisateur.
Title	Size	Format	Creation date																														
Data.csv	200 Mo	Tabi	27/09/19																														
Readme.txt	13 Mo	Text	27/09/19																														
Data_2.doc	14 Mo	Text	27/09/19																														
Transcripts	15 Go	Folder	27/09/19																														
Interview_1.flacc	5 Go	Video	27/09/19																														
Interview_2.flacc	4 Go	Video	27/09/19																														
Interview_3.flacc	6 Go	Video	27/09/19																														
<p>En ce qui concerne l'affichage du détail des jeux de données en cours de complétion, les participants ne savent pas s'ils préféreraient disposer du même affichage que celui du</p>	<ul style="list-style-type: none">○ Maintenir l'affichage actuel.																																

détail des jeux de données déjà publiés ou de l’affichage actuel.

Affichage d’un jeu de données en attente de validation :

Preview

Title	Size	Format	Creation date
Data.csv	200 Mo	Tabl	27/09/19
Readme.txt	13 Mo	Text	27/09/19
Data_2.doc	14 Mo	Text	27/09/19
Transcripts	15 Go	Folder	27/09/19
Interview_1.flacc	5 Go	Video	27/09/19
Interview_2.flacc	4 Go	Video	27/09/19
Interview_3.flacc	6 Go	Video	27/09/19



Mandatory

Organizational unit *	Linnaeus Team
Manager *	Jack Pott
Title *	My favorite flower
Author(s) *	Albert Kunz
Description *	I present to you my best flower from my garden
Resource type *	Dataset
Primary data type *	Observation
Publication date *	27 September 2019

Affichage d’un jeu de données publié :

Natural system in the case of fire event

James Smith Isabelle Dedier Pierre Matek

Publication Date: 12 July 2018

12 views 1 downloads

Open Download all

Description Files

Download all

Metadata

Data DOI: 10.1101/2020.08.05.335556

Publication Date: 03 september 2019

Unit: Linnaeus Team

Software used: Photoshop CS

Publication: <http://springer.org/httmh.html>

Data language: English

Domain: Biology

Collect method: Interview

Deposit validated by: Interview

Preservation policies: 20 years

Keywords:

Les participants aimeraient que le manager qui vérifie les jeux de données puissent modifier lui-même les métadonnées.

- Prévoir l’édition des métadonnées depuis l’affichage du détail des jeux de données en attente de validation.

Certains participants se sont interrogés sur la gestion des droits au sein des unités organisationnelles.

Ils veulent savoir comment est prévue cette fonctionnalité clé. Ils considèrent que, selon le statut, les utilisateurs n’ont pas besoin d’avoir la même interface.

Nous prenons bonne note du commentaire des participants, mais rappelons qu’un utilisateur peut être associé à plusieurs unités, mais qu’il n’occupe pas nécessairement le même rôle dans les différentes unités auxquelles il est lié. Ainsi, il faut mieux que l’affichage des fonctionnalités soit adapté au niveau des unités et de la gestion des données, plutôt qu’au niveau de la plateforme.

	<p><i>En l'état actuel, la maquette prévoit cette gestion lors de la création d'une nouvelle unité organisationnelle :</i></p> <p>New Unit</p> <p>Unit Name * <input type="text"/></p> <p>Members * <input type="text"/> Manager ? + Add me</p> <p>+ add member</p> <ul style="list-style-type: none"> ○ Prévoir la gestion des droits sur la page des unités déjà existantes également. ○ Prévoir le bouton d'édition des réglages des unités organisationnelles existantes.
Pour toutes les notifications, un participant suggère que les usagers concernés reçoivent la notification par mail avec un lien qui ne l'obligerait pas à se connecter à la plateforme et qui l'orienterait sur la bonne page.	<i>Cela n'est pas réalisable sur la plateforme, mais il sera suggéré au DLCM d'étudier cette possibilité.</i>
Deux participants ont jugé l'utilisabilité du site comme étant mauvaise parce qu'ils n'étaient vraiment pas satisfaits de devoir chercher des informations, essentiellement par rapport à la tâche 4.	<i>De manière générale, réfléchir à ce qui pourrait être simplifié.</i>

Remarques générales

D'abord, la majorité des participants ont été satisfaits de l'utilisabilité de la maquette :



Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Notons cependant que les deux participants ayant jugé l'utilisabilité de la maquette mauvaise l'on fait, car ils n'étaient pas satisfaits de devoir chercher des informations, essentiellement par rapport à la quatrième tâche.

Ensuite, la plupart des participants demandent à ce que les métadonnées puissent être rééditées si nécessaire après la publication des jeux de données, comme le permettent d'autres solutions d'archivage de données de recherche. En effet, dans une perspective institutionnelle, il n'est pas admissible pour les participants que des métadonnées erronées soient publiées sans possibilité de les corriger. Les participants comprennent le rôle des unités organisationnelles à travers cette fonctionnalité : en cas de départ d'un auteur de jeux de données d'une unité organisationnelle, celle-ci deviendrait responsable de ses jeux de données et pourrait les éditer au besoin.

Par ailleurs, permettre de « versionner » et de supprimer les jeux de données si nécessaire semblent aussi être des fonctionnalités importantes aux yeux des participants au test. Ceux-ci souhaiteraient que le DLCM envisage cette possibilité.

Certains participants ont été perturbés dans la réalisation des différentes tâches du test, car le fil d'Ariane n'était pas correct sur toutes les pages. En effet, certaines pages sont accessibles de plusieurs façons, mais il serait trop compliqué de le mettre en place sur Axure ; la maquette n'a pas la vocation de fonctionner comme une véritable plateforme. Autrement dit, nous ne résoudrons pas ce problème sur la maquette, mais nous recommanderons au DLCM d'implémenter un fil d'Ariane dynamique et cliquable sur la plateforme finale.

Par ailleurs, la question des langues est apparue dans le débat, tout comme dans le premier test. Pour les participants issus de la Faculté des Lettres, l'anglais ne suffit pas : il faut prévoir la traduction de l'interface, mais aussi des métadonnées. Ils signalent que, en l'état, la Faculté des Lettres sera réticente à utiliser la plateforme, car l'anglais n'est pas pratiqué par tous et que certains mots n'ont pas leur équivalent dans d'autres langues. Nous recommanderons donc au DLCM de prévoir une interface en anglais, français, allemand et italien, et de s'interroger sur les possibilités en termes de traduction des métadonnées dans les mêmes quatre langues.

Finalement, les participants se sont exprimés sur la page d'accueil :

- Le message de bienvenue est jugé trop long et les termes « FAIR » et « long term preservation » semblent poser problème. Il nous est suggéré de remplacer le message actuel par « Welcome to Yareta, the Geneva's platform for archiving research data », ce que nous ferons sur la prochaine maquette.
- Certains ont avoué ne pas connaître la signification de « FAIR » ; ils le comprennent plutôt dans le sens de « équitable ». Il suggère que l'acronyme soit expliqué ou que le logo officiel des principes FAIR figure.
- Le carrousel est trop grand pour son importance. Il pourrait être plus petit, voire supprimer, car les avantages de la plateforme figureront certainement dans la page « About » ou dans la présentation de la plateforme proposée dans le « Starter kit » de la page d'accueil.
- Le « Starter kit » devrait expliquer les différents rôles qui peuvent être associés à un utilisateur.

- La police utilisée pour le bloc « Yareta in numbers » compromet la lisibilité. « Yareta » pourrait conserver sa police de titre de plateforme, mais « in numbers » devrait avoir la même police que les autres textes de la maquette.
- Dans une optique de *responsive design*, il nous est demandé d'indiquer les systèmes d'exploitation, les navigateurs, ainsi que les versions de ces outils depuis lesquels il est conseillé d'accéder à la plateforme. En effet, même parmi les jeunes, certains usagers ne sont plus à l'aise depuis un ordinateur. Il sera donc recommandé au DLCM de se pencher sur les enjeux du *responsive design* et de donner les informations demandées aux utilisateurs depuis le « Starter kit ».
- De manière général, la page d'accueil leur semble trop riche en information, le principe qui devrait primer n'est pas « What you see is what you get », mais « What you see is what you need ». Il faut donc revoir la hiérarchie des informations sur la page d'accueil et être certaines que les utilisateurs n'auront rien besoin de chercher.

Annexe 22 : Compte rendu du troisième test utilisateur

Modalités du test

La troisième maquette a été créée sur le logiciel *Axure* et est accessible par le lien suivant : <https://6m98fn.axshare.com>. Le test, quant à lui, a été réalisé au moyen du logiciel *Loop11*, qui permet de donner accès au test via un lien URL et d'enregistrer l'écran des participants.

Le test s'est déroulé le mardi 8 octobre à Zurich et a rassemblé 7 participants.

Les participants ont dû réaliser quatre tâches :

1. Se connecter à la plateforme et rejoindre une unité organisationnelle donnée ;
2. Chercher et télécharger un jeu de données indiqué.
3. Soumettre (*upload*) un jeu de données à la plateforme.
4. Valider la publication d'un jeu de données sur la plateforme.

Après chacune d'entre elles, il était demandé aux participants de juger de la difficulté de l'exercice. La réalisation des tâches a duré entre 11'22" et 29'22" minutes et s'est suivie par une discussion d'une heure. Ce compte-rendu est structuré en tâche, les informations récoltées lors du test et lors de la discussion étant rassemblées par tâche.

Le rapport PDF de *Loop11* est disponible en annexe de ce document. Il contient certains commentaires des participants, ainsi que les informations sur les tâches. Toutefois, il ne faut pas tenir compte des statistiques de réussite, celles-ci étant parfois faussées lorsque les participants n'étaient pas sur la page finale de la tâche au moment où il la signalait comme atteinte.

Compte-rendu

Tâche 1 : se connecter à la plateforme et rejoindre une unité organisationnelle donnée

Difficulté de la tâche : Tous les participants ont jugé la tâche comme étant facile.

Temps nécessaire : 1'41" - 6'59"

Éléments soulevés par les participants	Répercussions sur la maquette
Tous les participants ont réalisé la tâche sans difficulté. Ils sont satisfaits de pouvoir demander eux-mêmes à rejoindre une unité.	
Tout comme dans les entretiens et les tests utilisateurs précédents, la question des unités organisationnelles et de leur fonctionnement resurgit. Les participants s'interrogent notamment sur la propriété des données.	<i>Ces discussions, déjà évoquées dans les compte-rendus précédents, appartiennent aux DLCM. Les conclusions auront forcément des répercussions sur la plateforme. Cependant, nous choisissons de maintenir le concept, puisque c'est une volonté du DLCM.</i>
Éléments remarqués par nous-mêmes	Répercussions sur la maquette
Lorsque les participants passent par la page « My unit » pour atteindre la page de l'unité Linnaeus Team, puis cliquer sur le lien vers les données de l'unité, ils arrivent sur la page « My datasets » et non sur la page « My unit's datasets ».	<i>Ce problème est inhérent à la maquette et n'arrivera pas sur la plateforme finale du DLCM.</i>

Tâche 2 : chercher et télécharger un jeu de données

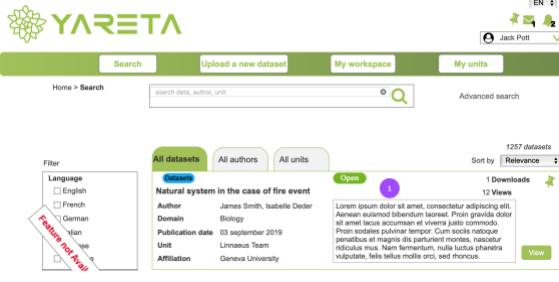
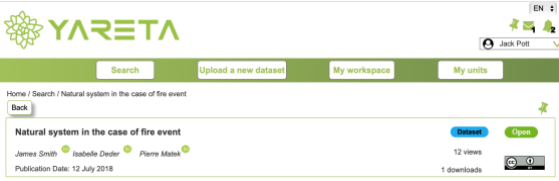
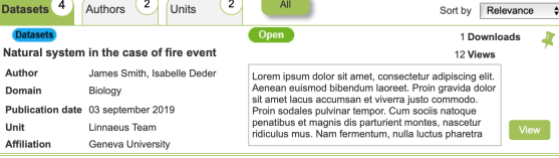
Difficulté de la tâche : Trois participants ont jugé la tâche comme étant facile et quatre l'ont jugé moyenne.

Notons que quatre personnes qui ont jugé la tâche comme étant de difficulté moyenne n'ont pas réussi, car elle est passée par la page des jeux de données de l'unité organisationnelle et, comme nous n'avions pas envisagé ce chemin, le lien du jeu de données de la consigne ne renvoie pas vers la bonne page de détail de jeu de données.

Temps nécessaire : 1'48" - 4'18"

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Éléments soulevés par les participants	Répercussions sur la maquette
<p>Le bouton « Search » ne semble pas clair. Certains participants s'interrogent sur le bouton « Search » : faut-il le renommer (<i>Find</i> ou <i>Browse</i>), réintégrer la barre de recherche ?</p> <p>Sur les sept participants, seuls deux sont partisans du <i>browse</i>.</p>	<p><i>La discussion au sujet du bouton « Search » est très enrichissante. A chaque proposition des participants, nous leur répondons en fonction des tests précédents. Suite à nos explications, ils sont finalement convaincus du bien-fondé de la solution présentée dans la troisième maquette, même s'ils regrettent de devoir cliquer sur le bouton « Search » pour accéder à la barre de recherche.</i></p> <ul style="list-style-type: none"> ○ Au vu de cette discussion, nous comptons remettre la barre de recherche présente dans les maquettes précédentes, car cette tâche n'a jamais posé problème auparavant. ○ Transformer le bouton « Search » en « Browse », pour donner accès à la liste de l'ensemble des jeux de données disponibles sur le dépôt.
<p>Six participants sur sept sont passés par la page « My unit » ou « My workspace » pour voir les jeux de données de leur unité (Linnaeus Team), puis télécharger ces jeux-ci. Ils associent l'unité « Linnaeus Team » à l'auteur « Carl Linnaeus ».</p>	<p><i>En l'occurrence, cela fonctionne ; le problème réside dans le fait que la ligne du tableau correspondant au jeu de données de la consigne ne pointe pas vers le bon jeu de données, car nous n'avions pas envisager ce cheminement.</i></p> <ul style="list-style-type: none"> ○ Corriger le lien qui dirige le jeu de données de la consigne (sur la page « Unit's datasets ») vers la page de détail du jeu de données pour qu'il pointe vers la bonne page.
<p>Un participant n'est pas persuadé que la page de recherche avec l'ensemble des jeux de donnée soit réaliste. Il pense qu'une page de ce genre serait trop lourde à charger.</p>	<p><i>Ce commentaire sera rapporté au DLCM, dont les développeurs seront plus à même de juger de la faisabilité de cette fonctionnalité.</i></p>

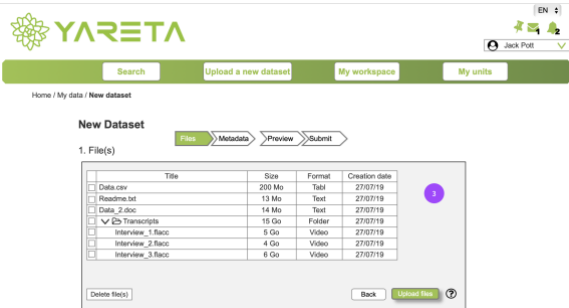
	
<p>Les participants apprécient que le nombre de jeux de données soit affichés sur la page de recherche, mais ils demandent à ce qu'il soit mieux mis en évidence pour la liste de l'ensemble des jeux de données présents sur la plateforme.</p>	<ul style="list-style-type: none"> ○ Mettre en évidence le nombre total de jeux de données présents sur la plateforme sur la page de recherche.
<p>Sur la page de détail des jeux de données, il est demandé de mettre en évidence l'auteur.</p> 	<ul style="list-style-type: none"> ○ Afficher les libellés des métadonnées pour clarifier l'affichage du détail des jeux de données.
<p>Un participant demande à ce qu'une fonctionnalité de citation soit envisagée.</p>	<ul style="list-style-type: none"> ○ Prévoir cette fonctionnalité.
<p>Les participants qui se sont intéressés à la page de détail des auteurs signalent que l'affiliation des auteurs devrait être plus détaillée.</p>	<ul style="list-style-type: none"> ○ Détailler l'affiliation. ○ Ajouter un bouton d'édition de la page des auteurs sur la page « Profile ».
<p>Éléments remarqués par nous-mêmes</p>	<p>Répercussions sur la maquette</p>
<p>Sur la page de résultats de recherche, pour accéder à la page de détail d'un jeu de données, un participant insiste pour cliquer sur « Open » (métadonnée relative à l'accès aux données), plutôt que sur « View ».</p> 	<ul style="list-style-type: none"> ○ Revoir l'affichage des métadonnées sur la page des résultats de recherche.

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Tâche 3 : soumettre un jeu de données à la plateforme

Difficulté de la tâche : Deux participants ont jugé la tâche comme étant facile et cinq l'ont jugé moyenne.

Temps nécessaire : 3'27'' - 11'50''

Éléments soulevés par les participants	Répercussions sur la maquette
Un participant exprime sa satisfaction par rapport au formulaire : il le trouve claire et facile à prendre en main. Les aides sont appréciées.	
Un participant nous demande de prévoir une fonctionnalité pour pouvoir enregistrer des <i>templates</i> de métadonnées pour faciliter la complétion du formulaire de dépôt.	<i>Il n'est pas réaliste de mettre cette fonctionnalité en place sur la maquette, mais ce commentaire sera transmis au DLCM.</i>
Plusieurs participants nous interrogent sur les possibilités d'organisation des fichiers à travers l'organisateur de fichiers : l'organisateur de fichiers permet-il de réorganiser l'arborescence des fichiers chargés ? Ils signalent que cette fonctionnalité leur semble importante.	<i>Il n'est pas possible de permettre l'organisation de fichier sur la maquette, mais nous comptons suggérer cette fonctionnalité au DLCM, car elle semble être importante pour les chercheurs que nous avons rencontrés.</i>
	
Il nous est signalé que, lorsque les participants cliquent sur la barre de progression, les métadonnées saisies se réinitialisent (s'effacent).	<i>Il n'est pas envisageable de remédier à ce problème sur la maquette, mais nous suggérerons aux développeurs du DLCM de penser à la sauvegarde des données saisies dans le formulaire s'ils implémentent une barre de progression cliquable.</i>
Certains participants rappellent le manque de clarté des unités organisationnelles. Ils nous demandent de remplacer l'aide	<ul style="list-style-type: none"> ○ Remplacer l'aide associée à l'unité organisationnelle du formulaire par

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

associée à l'unité organisationnelle du formulaire par l'aide associée au *pop-up* de la page d'accueil et de repenser leur affichage dans le formulaire de dépôt de jeux de données.

l'aide associée au *pop-up* de la page d'accueil.

Un participant s'interroge sur la métadonnée d'embargo : peut-on en modifier la date après la publication du jeu de données ? Cette possibilité lui tient à cœur.


Comme cette demande apparaît pour la première fois, nous ne prévoyons pas cette fonctionnalité sur la maquette finale. Toutefois, il sera recommandé au DLCM de s'intéresser à cette demande pour l'implémenter, ou non.

Les licences restent un élément flou pour les participants. Les aides, en l'état actuel, ne leur suffisent pas.

Visiblement, l'aide que nous proposons dans le formulaire ne suffit pas aux utilisateurs. Il sera donc recommandé au DLCM de prêter une attention toute particulière sur la formulation du texte de l'aide. Une communication à propos des licences pourrait être opportune pour les chercheurs.

Plusieurs participants aimeraient que la gestion des droits d'accès puisse se faire par fichier.

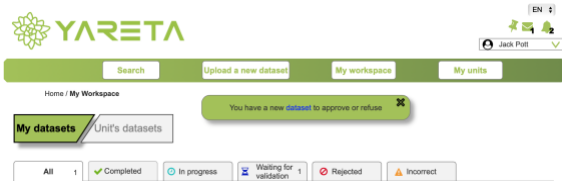
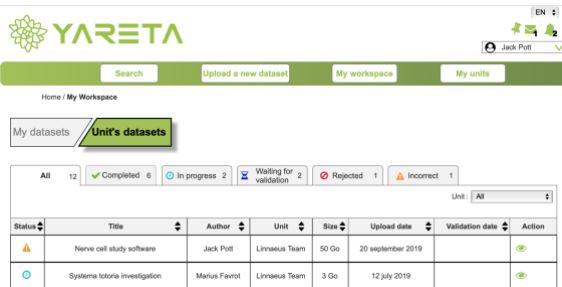
Nous ne prévoyons pas cette fonctionnalité dans la maquette, mais nous recommanderons au DLCM de s'interroger sur la faisabilité de cette demande.

Il nous est demandé de mettre en évidence les publications associées aux jeux de données en évidence, car la métadonnée en question semble avoir été perdue entre la première et la troisième maquette.	<ul style="list-style-type: none"> ○ Revoir les métadonnées liées aux références externes.
Dans le formulaire de dépôt, un participant ne comprend pas quels champs compléter lorsqu'il clique sur « Next » et que des champs requis ne sont pas complétés.	<i>Il ne sera pas possible de remédier à ce problème sur la maquette, mais il sera recommandé au DLCM de mettre en évidence les champs à remplir lorsque les utilisateurs sont empêchés de passer à la page « Preview ».</i>
Éléments remarqués par nous-mêmes	Répercussions sur la maquette
<p>Dans le formulaire de dépôt de jeu de données, un participant remplit le champs « Title » avec son propre titre : « Dr ». Il ne semble donc pas clair que le titre demandé porte sur le jeu de données.</p>  <p>The screenshot shows a '2. Metadata' form with a 'MANDATORY' label and an upward arrow. It contains several fields: 'Organizational unit' with a dropdown menu showing 'Linnaeus Team' and a '+ Add unit' button; 'Manager (for validation)' with a dropdown menu showing 'John Smith (Linnaeus Team)' and a '?' icon; 'Title' with an empty text input field; and 'Author(s)' with two input fields containing 'James' and 'Brown', and buttons for '+ Add me' and '+ New author' with a '?' icon.</p>	<ul style="list-style-type: none"> ○ Renommer la métadonnée « Title » « Dataset's title ».

Tâche 4 : valider la publication d'un jeu de données sur la plateforme

Difficulté de la tâche : Deux participants ont jugé la tâche comme étant facile, quatre l'ont jugé moyenne et un la trouve difficile.

Temps nécessaire : 1'35" - 8'12"

Éléments soulevés par les participants	Répercussions sur la maquette
<p>Plusieurs participants n'ont pas vu – ou du moins pas tout de suite – le <i>pop-up</i> de la page « My datasets ».</p> 	<ul style="list-style-type: none"> ○ Mettre en évidence le <i>pop-up</i>.
<p>Sur la page « My workspace », la navigation entre « My datasets » et « Unit's datasets » n'est pas claire.</p>	<p><i>Une solution est proposée plus bas.</i></p>
<p>Les participants peinent encore à trouver la page « My unit's dataset ».</p>  <p>Il nous est suggéré de poser le problème dans l'autre sens et d'avoir une page « Unit's datasets », qui comprendrait un bouton « Show only my datasets ».</p> <p>Le filtre par unité organisationnelle devrait être mis en évidence, d'après certains participants.</p>	<ul style="list-style-type: none"> ○ Sur la page « My workspace », n'afficher que les « Unit's datasets » et prévoir une case à cocher « Show only my datasets ». ○ Sortir le filtre des unités organisationnelles des onglets.
<p>Sur l'onglet « All » de la page « Unit's datasets », les participants jugent qu'il manque les boutons d'action, car ils ne voient pas toujours les différents onglets.</p>	<ul style="list-style-type: none"> ○ Ajouter les boutons d'action dans la colonne de droite du tableau.

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Certains participant sont passés par la page « My units » pour atteindre la page de leur unité, puis consulter les jeux de données des unités. Ils demandent à ce que le lien entre l'unité et ses jeux de données soit mis en évidence.

- Afficher le « Go to the unit's datasets » comme un bouton.

Éléments remarquables par nous-mêmes

Si les participants arrivent jusqu'à la page de détail du jeu de données à valider, ils ne regardent pas la page jusqu'en bas et ne voient pas les boutons de rejet et de validation.

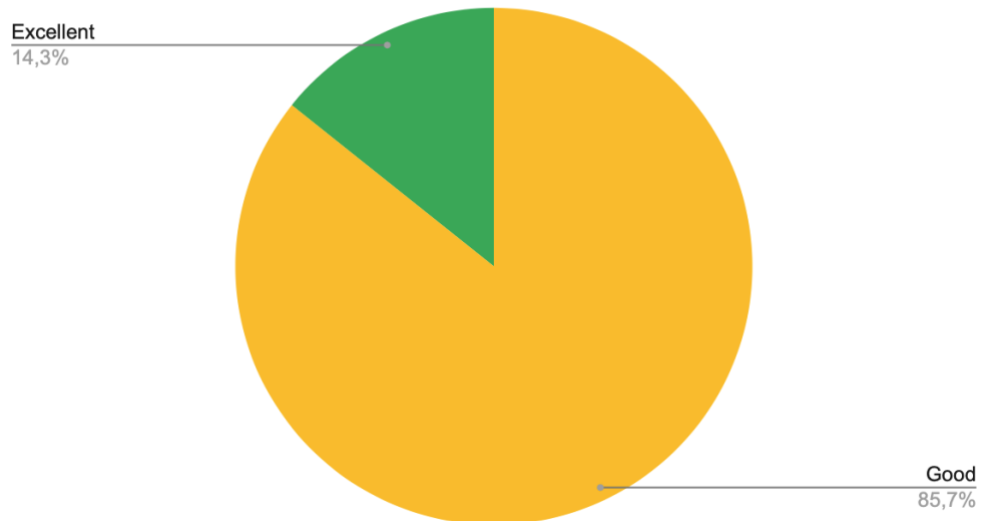
Répercussions sur la maquette

Il est trop compliqué de mettre en place cette fonctionnalité sur la maquette. Toutefois, il sera suggéré au DLCM de prévoir des boutons de soumission, validation et enregistrement qui suivent le déroulement de la page du formulaire de soumission et de validation des jeux de données (sticky en anglais).

Remarques générales

D'abord, l'ensemble des participants ont été satisfait de de l'utilisabilité de la maquette :

Evaluation de l'utilisabilité de la maquette par les participants



Finalement, deux remarques concernent spécifiquement la page d'accueil. D'une part, certains participants aimeraient pouvoir faire défiler plus vite les informations qui se trouve dans le carrousel. Cette fonctionnalité ne fonctionne pas sur la maquette, mais elle est généralement inhérente au carrousel : il sera cependant demandé au DLCM d'être attentif à cette demande s'il choisit d'implémenter un carrousel. D'autre part, un participant nous signale que les onglets des statistiques ne fonctionnent pas ; nous résoudrons donc ce problème pour la maquette finale.

Annexe 23 : Liste des modifications apportées aux maquettes

V4.0 (last maj: 28.10.2019)

Général:

- Renommer bouton “search” par “browse”
- Ajout d’une barre de recherche dans le bandeau de navigation
- Création de la page de profil
- Mise à jour du fil d’Ariane sur les différentes pages de la maquette

Homepage:

- Réparer les statistiques

Dataset 1:

- Ajout d’une option de partage et de citations
- Modifier affichage auteur

New Dataset:

- Modifier “title” par “dataset’s title”
- Modifier le contenu de l’aide du champ “organizational unit”
- Ajouter la possibilité d’ajouter le lien de la publication liée au jeu de donnée
- Ajout d’une option d’accès durant l’embargo

Search:

- Suppression du bouton “all”
- Déplacement des icônes de description des résultats

My unit:

- Modifier le lien pour accéder aux datasets de l’unité, par un bouton

My workspace:

- Modifier (agrandir) le pop-up
- Déplacer le filtre des unités
- Suppression de l’onglet “my dataset”
- Ajout d’une case à cocher “show only my dataset” et paramétrage de la maquette pour filtrer les résultats en fonction.
- Ajout d’icônes pour les datasets

Author:

- Modifier “institution” par “affiliation” dans la page de l’auteur

V3.0 (last maj: 07.10.2019)

Général:

- Augmenter taille langue du site
- Modifier ordre d’affichage des notifications et des messages
- Remplacer “ma data” par “my workspace” dans le bandeau de navigation
- Remplacer “search a dataset” par “search” dans le bandeau de navigation
- Modification de la zone de profil, avec nom et prénom plutôt qu’email, et icône différent

Homepage:

- Changement du descriptif de la page d’accueil
- Changement de l’affichage de la page d’accueil

Design d’un portail pour l’archivage des données de la recherche et proposition d’interfaces pour inciter les chercheurs/structures à déposer leurs données

Dataset 1:

- Afficher bouton “download” en grisé pour les datasets restreints
- Modification affichage de la page de dataset

New Dataset:

- Augmenter taille des aides
- Changement sens des flèches des bandeaux de métadonnées
- Rajouter “data” dans le champ “folder name” du gestionnaire de fichier
- Inverser prénom et nom dans le formulaire
- Modifier la barre de progression du formulaire d’ajout de dataset
- Suppression du champ “publication date” dans le formulaire
- Déplacement de “primary data type” en métadonnées optionnelles
- Ajout d’une option de date pour l’embargo
- Réparer l’affichage d’”Access”
- Suppression de collect context dans le formulaire
- Modification des options de sauvegarde pour le formulaire
- Ajout d’un pop-up pour le formulaire

Search:

- Changement affichage auteur
- Ajout d’un champ “affiliation” dans les résultats de la recherche
- Pin: changement de tooltip: “save as favorite”

My unit:

- Ajout d’une option de changement d’affichage des unités
- Ajout d’une option d’édition pour les managers d’unités et ajout de membres
- Modifications générales
- Ajout d’un lien pour les datasets depuis les unités

My Data/Workspace:

- Rediriger la notification “dataset to approve” sur la bonne page
- Suppression du toggle
- Changement de mise en page
- Modification de l’affichage entre les datasets de l’auteur et de l’unité
- Ajout d’un pop-up

Register:

- Souligner “terms & conditions”

V2.0 (last maj: 25.09.2019)

Général:

- Mettre “My units” dans le bandeau de navigation, et le supprimer du profil
- Modifier “Sign in” en “Register”
- Suppression de la barre de recherche pour un bouton “search a dataset” et réarrangement des boutons.
- Rajouter les boutons dans le bandeau de navigation sans être logger
- Corrections de fautes d’orthographe
- Modification visuel de la page de datasets et en rajouter pour correspondre aux résultats de recherche
- Modification visuel de la page d’unité
- Modification visuel de la page “My data”

Homepage:

- Mettre un lien pour la notion de “fair”
- Rajout dans le starter kit, d’une aide pour remplir le formulaire de dépôt de dataset
- Rajouter les mêmes domaines sur les stats de la page d’accueil que sur le formulaire

Dataset 1:

- Remplacer “download” par “download all”
- Aligner le texte à gauche

New Dataset:

- Remplacer “Ad Eternam” par “Forever”
- Ajouter une option pour choisir les unités qui peuvent accéder à un dataset “restricted”
- Suppression du champ “DOI” et déplacer “request a DOI” dans le menu “save/next”
- Ajout de fichier: ajout d’un bouton “upload” et “back”
- Preview du nouveau dataset: suppression du bouton “save”
- Suppression de “Add files” dans le formulaire de nouveau dataset
- Modification de l’affichage de la progression pour le formulaire
- Ajout d’une étape de téléchargement des fichiers du formulaire
- Ajout d’une mention de champ obligatoire pour l’ajout de fichier dans le formulaire
- Ajout d’une option pour ajouter des unités
- Déplacement du champ “domain” de Recommended à Mandatory, et option d’en mettre plusieurs.
- Modifier “language” par “data language”
- Modification du champ “Contributor” par “Other contributor(s)” et ajout d’options à sélectionner dans le formulaire
- Rajout de langues et option d’en mettre plus qu’un dans le formulaire
- Rajout d’un champ pour mettre des titres dans différentes langues dans le formulaire (optionnel)
- Rajout d’un champ pour choisir le manager qui validera le dataset
- Suppression de “Subject”
- Mise en page de la partie “access” du formulaire
- Séparer les métadonnées de “preview” par type (obligatoire, recommandé, optionnel)


Search:

- Ajout d’un tooltip “save” pour les “pin” de la recherche
- Remplacer les icônes “open, closed, restricted”, par du texte
- Créer une page avec tous les datasets, auteurs et unités.
- Modification de l’affichage des datasets, des auteurs et des unités.
- Rajouter des filtres.
- Rajout de “join unit” dans les résultats.
- Supprimer le surlignement de mots dans les résultats.

My unit:

- Si l’usager n’a pas d’unit, la page affiche directement la fonction recherche.
- Ajout de “waiting for validation” pour les demandes d’unités
- Remplacer “Join unit” par “Find a unit”

Annexe 24 : Troisième maquette – Page de gestion des jeux de données



EN
1
2
Jack Pott

Search
Upload a new dataset
My workspace
My units

Home / My Workspace

You have a new [dataset](#) to approve or refuse

My datasets
Unit's datasets

All 12
Completed 6
In progress 2
Waiting for validation 2
Rejected 1
Incorrect 1

Unit : All

Status	Title	Author	Unit	Size	Upload date	Validation date	Action
	Nerve cell study software	Jack Pott	Linnaeus Team	50 Go	20 september 2019		
	Systema totoria investigation	Marius Favrot	Linnaeus Team	3 Go	12 july 2019		
	My favorite flower	Albert Kunz	Linnaeus Team	3 mo	2 august 2019		
	Study on the four-leaf clover	Jack Pott	Linnaeus Team	1 Tb	14 september 2019		
	Deposit test	Isabelle Deder, Marc Patek	Linnaeus Team	2 Go	2 june 2019		
	Natural system in the case of fire event	James Smith, Isabelle Deder, Pierre Matek	Linnaeus Team	6 Go	20 june 2019		
	Systema Naturae	John Smith	Linnaeus Team	10 Go	1 august 2019	2 august 2019	
	The transformation of protein no. 4 into amino acid	John Smith	Linnaeus Team	50 mo	18 june 2019	20 june 2019	

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données



Jack Pott

search data  [Advanced search](#)

My units

Back

Open

13 views

2 downloads

Files

Share

Metadata

Preservation policies: 10 years

168

Systema naturæ per regna tria naturæ, secundum classes.

Dataset

Open

Author: Carl Linnaeus

13 views

Publication Date: 12 July 2018

2 downloads



Description

Files

Cite this dataset

Share


Download all

Sort by Name

Title	Size	Format	
Title title title	3 GB	JPEG	↓
Decision making time based on academic resources	570 MB	PDF	↓
Title of the file	12 Mb	CSV	↓

PREVIEW

Annexe 26 : Dernière maquette – Première page du formulaire de dépôt



EN
1
2
Jack Pott

Home / New dataset

Before uploading a new dataset, consult out [starter kit](#) to be sure you don't forget anything

New Dataset

Files
Metadata
Preview
Submit

1. File(s)

*** Mandatory step**

Feature not available in this mockup

or

Choose file(s)

Drag & Drop

Upload conditions (size max, number of files max, general informations for the upload)

2. Metadata

MANDATORY

Organizational unit *

Linnaeus Team

+ Add unit ?

Manager (for validation) *

John Smith (Linnaeus Team)

?

Dataset's title *

Author(s) *

James

Brown

+ Add me

+ New author

?

Description *

Present the dataset, or copy the abstract of the related article

Resource type *

Dataset

?

Domain *

Architecture

+ Add Domain

?

Access *

Open

?

Embargo *

No

?

Licence *

CC BY

?

Submission policies *

Deposit with approval

?

Preservation policies *

Keep it for 10 years

?

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

RECOMMENDED

Publication's link

Data language

English

+ Add language

Data collection begin

0

January

2019

Data collection end

0

January

2019

Keywords

Other contributor(s)

James

Brown

ContactPerson

+ New contributor

Collect method

Software used
+ Add Software

OPTIONAL

Additional title

Title in other languages

English

+ Add title

External reference DOI

External reference URL

Primary data type

Observational

+ Add type

Additional metadata and Schema

Geolocalisation related info

Funding information

DMP file

ReadMe file

Request a DOI

Cancel

Save & quit

Next

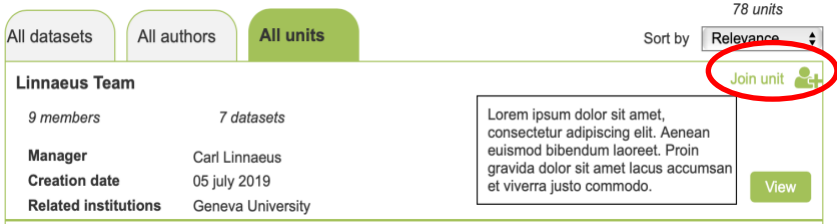
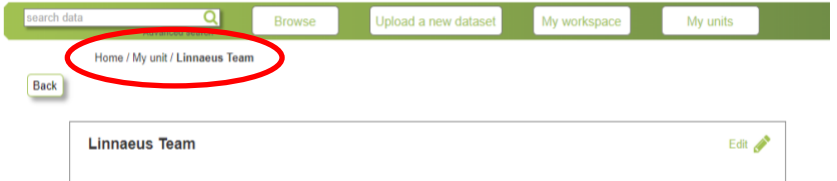
Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données



Annexe 27 : Synthèse des recommandations générales

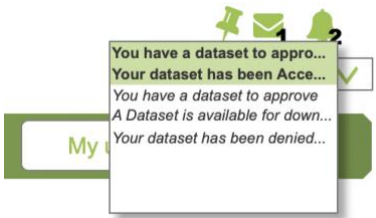
Ce document regroupe les recommandations essentielles formulées suite aux tests utilisateurs. La majorité d'entre elles sont implémentées dans la maquette, sauf exceptions dues aux limites du logiciel *Axure*. En annexe de ce document, vous trouverez les recommandations concernant les métadonnées demandées lors de la soumission d'un jeu de données. Veuillez consulter la maquette pour prendre connaissance des recommandations d'ergonomie qui ne sont pas toutes listées ci-après. Lien vers la maquette : <https://v4aj02.axshare.com>

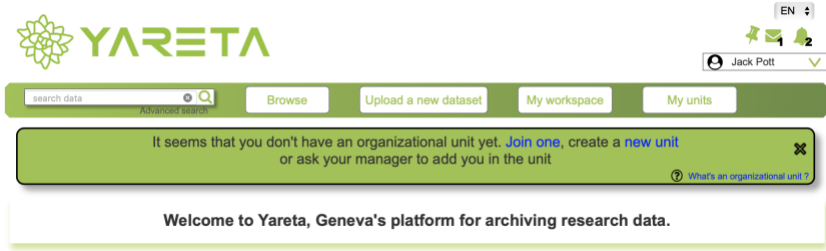
Importance	Fonctionnalité concernée	Page(s) concernée(s)	Description
Remarques générales			
Haute	Liens entre les entités et leurs données	Search Browse Data linnaeus (open) MyWorkspace Unit Linnaeus Author	Prévoir autant de liens que possible entre les pages dédiées aux personnes, aux unités organisationnelles et aux données, pour faciliter la recherche et la consultation de jeux de données apparentés.
Moyenne	Texte	Toutes les pages	Rédiger des textes de présentation le plus clair, mais aussi le plus court possible (ex : message d'accueil, aides, etc.).
Moyenne	Utilisation d'icônes	Toutes les pages	Ajouter du texte à côté des icônes pour clarifier leur utilité auprès des utilisateurs. Il devrait être visible en permanence, et pas uniquement au passage du curseur.

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

			<p><i>Exemple issu de la maquette :</i></p> 
Moyenne	Fil d'Ariane	Toutes les pages	<p>Mettre en place un fil d'Ariane cliquable, pour connaître sa localisation sur le site à tout moment et faciliter la navigation entre les pages.</p> <p><i>Exemple issu de la maquette :</i></p> 
Moyenne	Traduction	Toutes les pages	Prévoir la traduction de la plateforme en anglais, allemand, français et italien.
Moyenne	Navigation entre les champs	Register New dataset	Faciliter la navigation dans les champs des formulaires grâce à la touche Tab.
Faible	Accès à la plateforme		Permettre de se connecter sur la plateforme sans forcément rejoindre une unité.

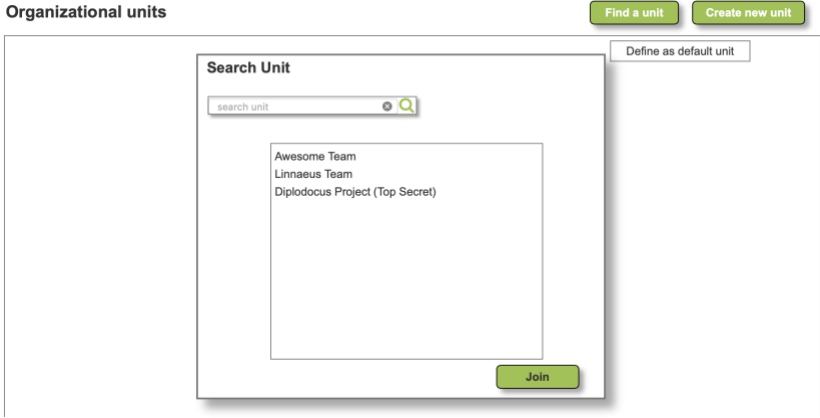
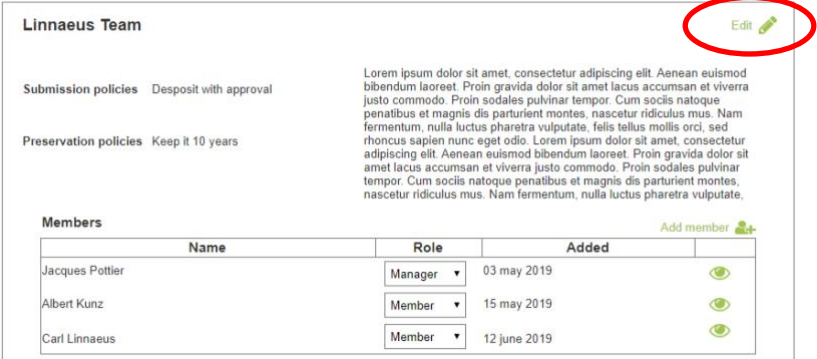
Menu			
Haute	Recherche	Toutes les pages	<p>Prévoir une barre de recherche dans le bandeau de navigation pour rechercher des jeux de données, des auteurs ou des unités organisationnelles.</p> <p><i>Exemple issu de la maquette :</i></p> 
Haute	Menu de navigation	Toutes les pages	<p>Les fonctionnalités de base – à savoir le dépôt de données, la recherche, la gestion des jeux de données, ainsi que la gestion des unités organisationnelles – doivent figurer dans le bandeau de navigation pour plus de visibilité.</p> <p><i>Exemple issu de la maquette :</i></p> 
Moyenne	Notifications	Toutes les pages	<p>Mettre en place un système de notifications. L'utilisateur doit pouvoir choisir dans son profil le(s) type(s) de notifications qu'il souhaite recevoir (sur la plateforme, par email avec lien de direction sur la page concernée).</p> <p>L'usager doit également pouvoir recevoir des messages relatifs à ses jeux de données ou à son activité sur la plateforme (ex : si quelqu'un souhaite avoir accès à un jeu de données qu'il aurait maintenu privé).</p> <p><i>Exemple issu de la maquette :</i></p>

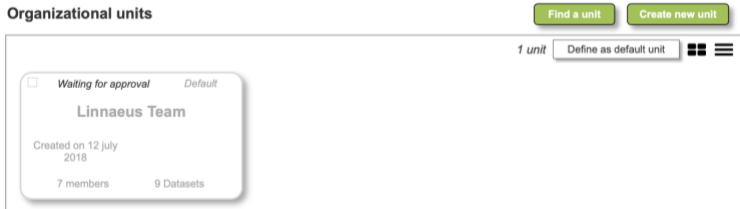
			
Page d'accueil			
Haute	Documentation	Homepage	Sur la page d'accueil, afficher les informations principales relatives à la plateforme (ex : le prix, le temps de conservation, la taille max, etc.).
Haute	Documentation	Homepage	<p>Créer une documentation sur les points importants de la plateforme et de son fonctionnement, et la mettre sur la page d'accueil, dans le « Starter kit » par exemple.</p> <p><i>Exemple issu de la maquette :</i></p> <div data-bbox="1758 667 1971 973"> <p>STARTER KIT</p> <ul style="list-style-type: none"> • Cost Calculator • Checklist for Dataset Upload • Presentation of Yareta • FAQ • Licences and Access • Example of metadata form complete </div>
Moyenne	Documentation	Homepage	Afficher les statistiques relatives aux jeux de données, ainsi que les avantages essentiels sur la page d'accueil de la plateforme.
Inscription et gestion du profil			
Faible	Inscription	Register	Nommer la fonctionnalité d'inscription « Register », plutôt que « Sign in », qui peut prêter à confusion (confusion entre « Sign in » et « Log in »).

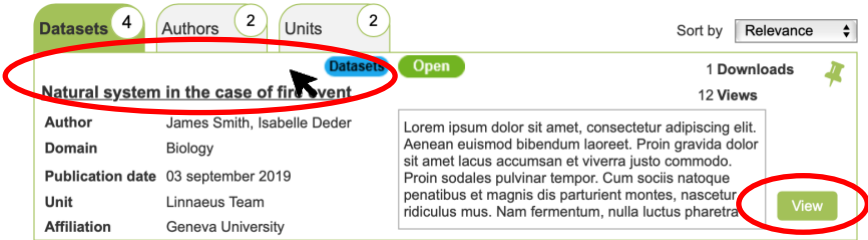
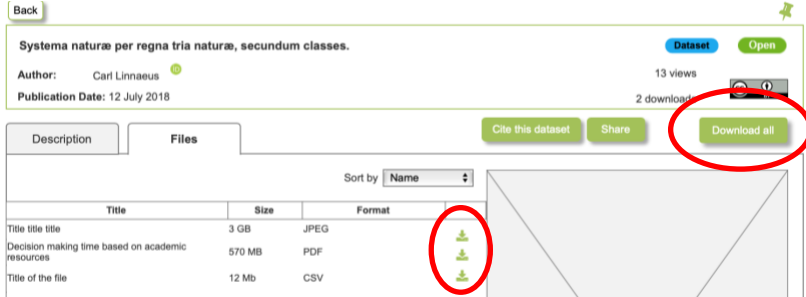
Faible	Récupération des informations liées au profil des usagers	Profile	Prévoir de récupérer automatiquement les informations détenues par l'institution à laquelle est affilié l'utilisateur pour compléter les informations du profil utilisateur (ex : informations liées au compte AAI).
Unités organisationnelles			
Haute	Unités organisationnelles	Homepage	Mettre à disposition une documentation claire et complète sur le fonctionnement des unités organisationnelles, la validation des données/jeux de données, la gestion des droits au sein des unités et les droits des usagers s'ils quittent une unité. Cette aide pourrait se trouver dans le « Starter kit » de la page d'accueil, par exemple.
Haute	Unités organisationnelles	Homepage	<p>Lors de la première connexion, guider les utilisateurs vers une page pour rejoindre une unité organisationnelle.</p> <p><i>Exemple issu de la maquette :</i></p> 
Haute	Unités organisationnelles	Data linnaeus (open)	Sur la page des unités organisationnelles, afficher la liste des jeux de données déjà publiés sur la plateforme par l'unité correspondante.

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

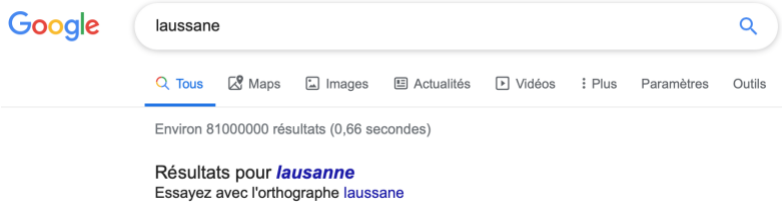
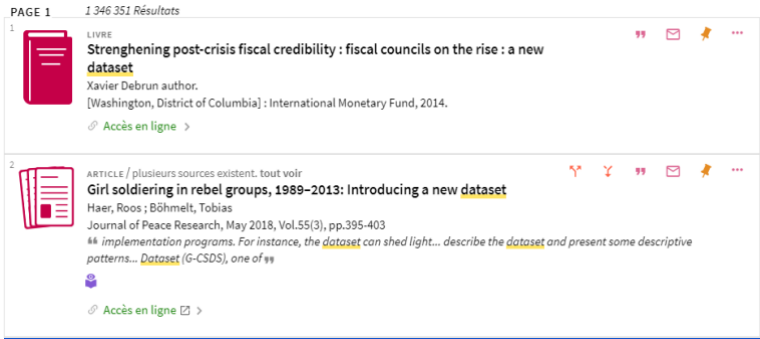
			<div><div>Exemple issu de la maquette :</div><div><div><div>Linnaeus Team</div><div><div>Join unit</div><div></div></div></div><div><div>Submission policies</div><div>Desposit with approval</div></div><div><div>Preservation policies</div><div>Keep it 10 years</div></div><div><div>Members</div><table><thead><tr><th>Name</th><th>Role</th><th>Added</th><th></th></tr></thead><tbody><tr><td>Jacques Pottier</td><td>Manager</td><td>03 may 2019</td><td></td></tr><tr><td>Albert Kunz</td><td>Member</td><td>15 may 2019</td><td></td></tr><tr><td>Carl Linnaeus</td><td>Steward</td><td>12 june 2019</td><td></td></tr></tbody></table></div><div><div>Published datasets</div><table><thead><tr><th>Title</th><th>Author</th><th>Size</th><th>Publication date</th><th></th></tr></thead><tbody><tr><td>Nerve cell study software</td><td>Jack Pott</td><td>3 GB</td><td>03 may 2019</td><td></td></tr><tr><td>Systema Naturae per regna tria naturae, secundum classes.</td><td>Carl Linnaeus</td><td>570 MB</td><td>17 june 2019</td><td></td></tr><tr><td>Study on the four-leaf clover</td><td>Jack Pott</td><td>12 Mb</td><td>02 august 2019</td><td></td></tr></tbody></table></div></div></div>	Name	Role	Added		Jacques Pottier	Manager	03 may 2019		Albert Kunz	Member	15 may 2019		Carl Linnaeus	Steward	12 june 2019		Title	Author	Size	Publication date		Nerve cell study software	Jack Pott	3 GB	03 may 2019		Systema Naturae per regna tria naturae, secundum classes.	Carl Linnaeus	570 MB	17 june 2019		Study on the four-leaf clover	Jack Pott	12 Mb	02 august 2019	
Name	Role	Added																																					
Jacques Pottier	Manager	03 may 2019																																					
Albert Kunz	Member	15 may 2019																																					
Carl Linnaeus	Steward	12 june 2019																																					
Title	Author	Size	Publication date																																				
Nerve cell study software	Jack Pott	3 GB	03 may 2019																																				
Systema Naturae per regna tria naturae, secundum classes.	Carl Linnaeus	570 MB	17 june 2019																																				
Study on the four-leaf clover	Jack Pott	12 Mb	02 august 2019																																				
Moyenne	Unités organisationnelles	MyUnit	Offrir la possibilité aux usagers de demander à rejoindre une unité organisationnelle directement sur la plateforme.																																				
Moyenne	Unités organisationnelles	MyUnit	Ne pas afficher de page vide sur la page des unités organisationnelles lorsque les usagers n'en ont pas encore : afficher soit une page de recherche pour en rejoindre une, soit un <i>pop-up</i> pour envoyer une demande d'ajout au PI.																																				

			<p><i>Exemple issu de la maquette :</i></p> 
Moyenne	Unités organisationnelles	Unit Linnaeus	<p>Prévoir une option d'édition des unités organisationnelles, ainsi que des rôles des membres.</p> <p><i>Exemple issu de la maquette :</i></p> 


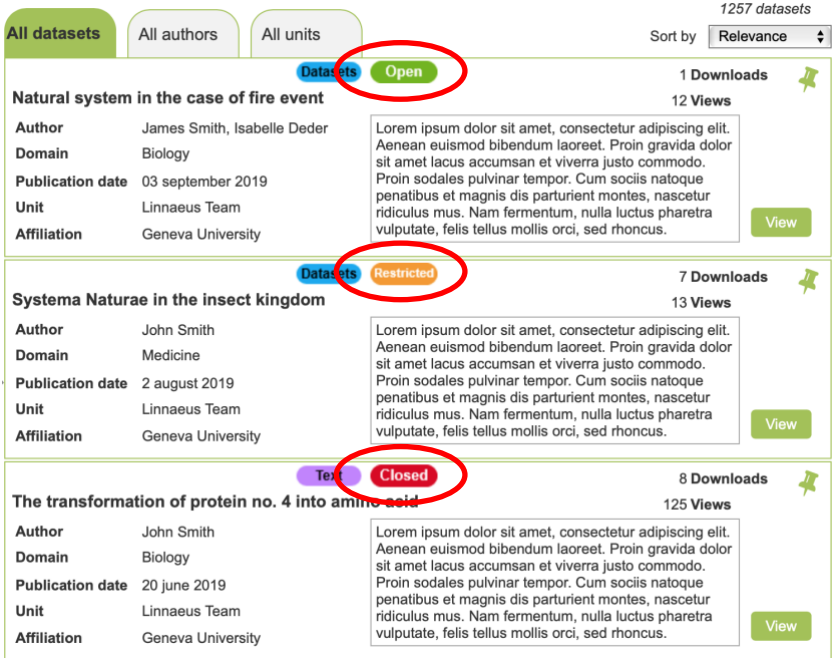
Faible	Unités organisationnelles	MyUnit	Il est plus clair que le bouton qui permet de visualiser les unités organisationnelles avant de les rejoindre soit nommé « Find a unit » et non « Join a unit ».
Faible	Unités organisationnelles	MyUnit	<p>Prévoir d'afficher les unités organisationnelles sollicitées en grisé sur la page à cet effet entre le moment de la demande d'adhésion et la validation de la demande.</p> <p><i>Exemple issu de la maquette :</i></p> 
Faible	Affichage des unités organisationnelles	MyUnit	Permettre aux usagers de choisir un affichage en liste ou en bloc de leurs unités organisationnelles.
Recherche et téléchargement			
Haute	Recherche des jeux de données	Browse	Prévoir une page (ex : Browse) qui regroupe l'ensemble des jeux de données, avec des facettes pour les filtrer, pour autant que la page puisse être chargée en un temps raisonnable.
Haute	Affichage des jeux de données dans la fonctionnalité de recherche	Search Browse	Mettre en place un système visible et clair pour l'utilisateur pour afficher le détail d'un « item » dans les résultats de la recherche. Le bouton « View » et le soulignement du titre sont les moyens choisis dans la maquette.

			<p>Exemple issu de la maquette :</p> 
Haute	Affichage des jeux de données	Data (open) linnaeus	Afficher toutes les métadonnées dans le détail des jeux de données.
Haute	Téléchargement de jeux de données	Data (open) linnaeus	<p>Prévoir un bouton « Download all » en plus de la fonctionnalité de téléchargement de fichiers individuels pour permettre le téléchargement du jeu de données complet.</p> <p>Exemple issu de la maquette :</p> 
Moyenne	Aide à l'orthographe	Search Browse	Prévoir une aide pour l'orthographe dans la recherche.


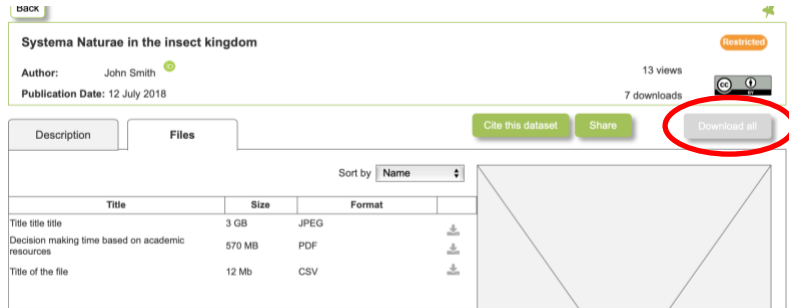
Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

			<p>Exemple issu du moteur de recherche Google :</p> 
Moyenne	Soulignement	Search	<p>Dans la recherche, souligner les mots de la recherche dans les résultats afin de faciliter l'analyse aux utilisateurs.</p> <p>Exemple issu de Renouvaud :</p> 
Moyenne	Favoris	Search	<p>Prévoir une punaise (ou un autre moyen) pour sauvegarder des jeux de données.</p> <p>Nommer le <i>tooltip</i> « Add to favorites » pour éviter l'incompréhension des utilisateurs.</p>

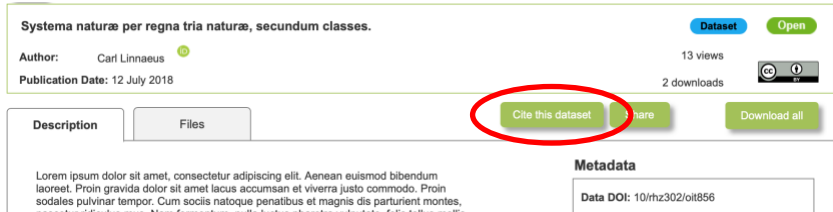

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

			<p>Exemple issu de Renouvaud :</p> 
Moyenne	Affichage des jeux de données dans la fonctionnalité de recherche	Search Browse	<p>Utiliser un code couleur pour définir les différents accès aux jeux de données. Les icônes ne sont pas suffisamment claires pour les utilisateurs.</p> <p>Exemple issu de la maquette :</p> 

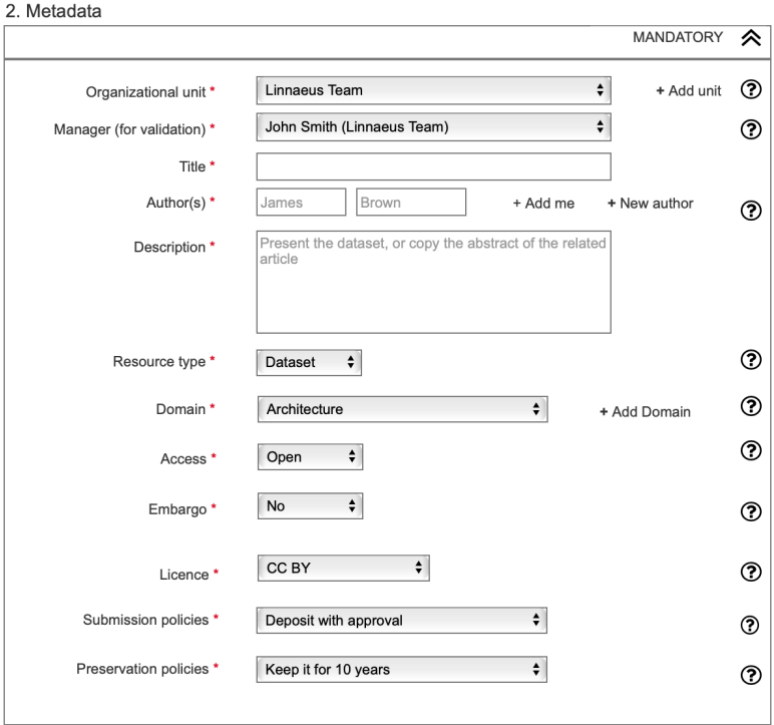
Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Moyenne	Affichage des jeux de données dans la fonctionnalité de recherche	Search Browse	<p>Afficher le nombre de vues et de téléchargements dans la liste des résultats des jeux de données.</p> <p><i>Exemple issu de la maquette :</i></p> 
Moyenne	Jeu de données	Systema Naturae (restricted)	<p>Si un jeu de données est restreint ou fermé, le bouton de téléchargement doit être grisé ou invisible, et une information doit clairement être indiquée sur la page pour expliquer le non-téléchargement.</p> <p><i>Exemple issu de la maquette :</i></p> 
Moyenne	Citation des jeux de données	Data linnaeus (open)	<p>Prévoir une fonctionnalité de citation automatique des jeux de données suivant la norme choisie par l'utilisateur.</p>

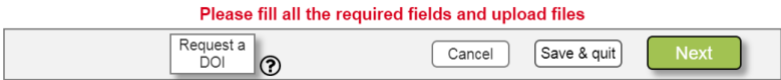
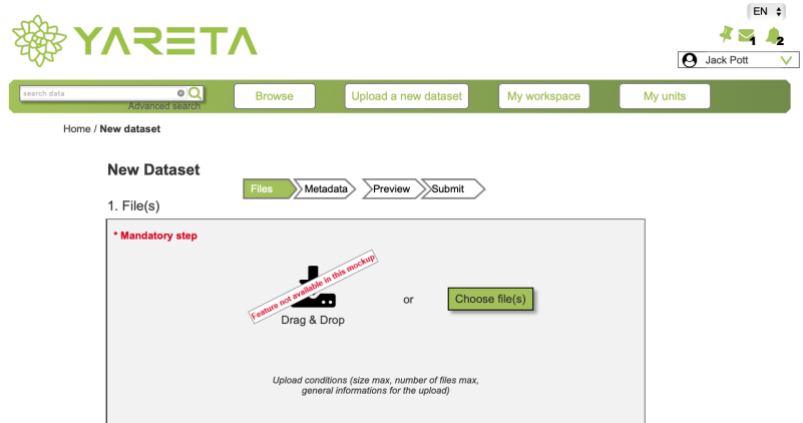
Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

			<p><i>Exemple issu de la maquette :</i></p> 
Faible	Affichage des jeux de données dans la fonctionnalité de recherche	Browse	<p>Mettre en valeur le nombre total de jeux de données sur la page les répertoriant tous.</p> <p><i>Exemple issu de la maquette :</i></p> 
Faible	Suggestion		Envisager de faire des suggestions de jeux de données personnalisées en fonction des préférences des utilisateurs.
Dépôt de jeux de données			
Haute	Formulaire de dépôt	New dataset	Ajouter une fonction de sauvegarde automatique du formulaire, pour que les utilisateurs ne perdent pas leurs données en cas de déconnexion.

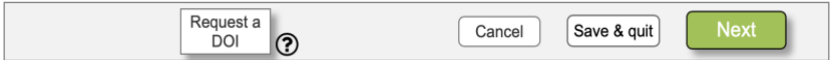
Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Haute	Formulaire de dépôt	New dataset	Mettre en place une aide pour le remplissage du formulaire (sur la page d'accueil et sur la page du formulaire). Le « Starter kit » pourrait par exemple proposer un exemple de formulaire complété.
Haute	Formulaire de dépôt	New dataset	<p>De manière générale, associer chaque métadonnée à une aide – que ce soit sous la forme de champ pré rempli avec des exemples (<i>hint</i>) ou de texte explicatif. Ce dispositif est notamment important pour les licences, car les usagers ne sont pas familiers avec cet élément.</p> <p><i>Exemple issu de la maquette :</i></p>  <p>The screenshot shows a '2. Metadata' form with a 'MANDATORY' label and an upward arrow icon. The form contains the following fields:</p> <ul style="list-style-type: none"> Organizational unit *: Dropdown menu with 'Linnaeus Team' selected, and a '+ Add unit' button with a help icon. Manager (for validation) *: Dropdown menu with 'John Smith (Linnaeus Team)' selected, and a help icon. Title *: Empty text input field. Author(s) *: Text input field with 'James' and 'Brown' entered, and '+ Add me' and '+ New author' buttons with a help icon. Description *: Text area with placeholder text 'Present the dataset, or copy the abstract of the related article'. Resource type *: Dropdown menu with 'Dataset' selected, and a help icon. Domain *: Dropdown menu with 'Architecture' selected, and a '+ Add Domain' button with a help icon. Access *: Dropdown menu with 'Open' selected, and a help icon. Embargo *: Dropdown menu with 'No' selected, and a help icon. Licence *: Dropdown menu with 'CC BY' selected, and a help icon. Submission policies *: Dropdown menu with 'Deposit with approval' selected, and a help icon. Preservation policies *: Dropdown menu with 'Keep it for 10 years' selected, and a help icon.

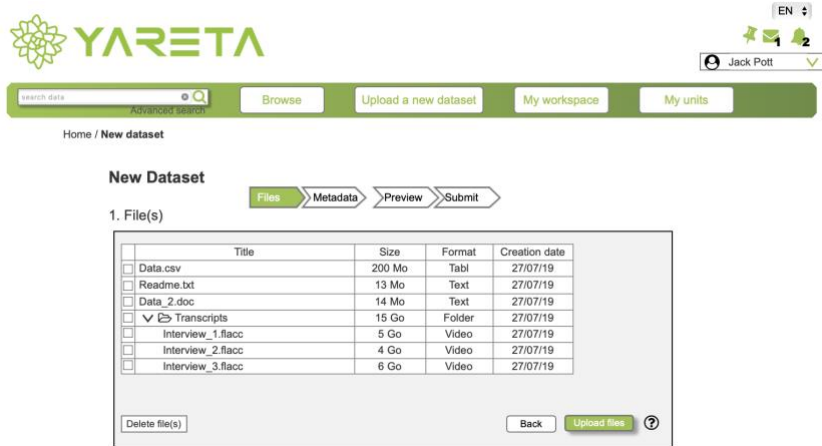
Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Haute	Formulaire de dépôt	New dataset	<p>Obliger le dépôt et l'<i>upload</i> de données, ainsi que la complétion des métadonnées obligatoires pour passer à la page suivante du formulaire.</p> <p>Afficher un <i>pop-up</i> pour avertir l'utilisateur qu'il doit remplir certains champs avant de pouvoir soumettre le jeu de données.</p> <p>Mettre les champs manquants en évidence (par exemple en rouge) pour que l'utilisateur sache quels champs remplir, afin de pouvoir passer à la page suivante.</p> <p><i>Exemple issu de la maquette :</i></p> 
Haute	Formulaire de dépôt	New dataset	<p>Offrir la possibilité de « Drag & drop » les fichiers ou dossiers pour l'ajout d'un nouveau jeu de données.</p> <p><i>Exemple issu de la maquette :</i></p> 

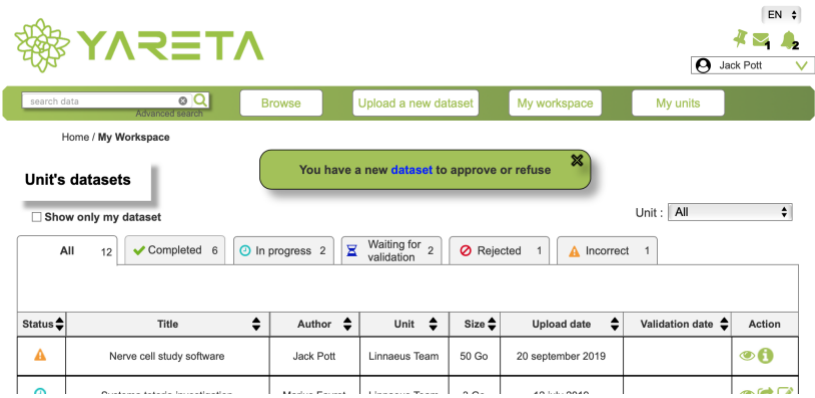
Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Haute	Formulaire de dépôt	New dataset	Demander aux usagers de charger les fichiers avant de les décrire, de façon à ce qu'ils sachent ce sur quoi porte les métadonnées.
Haute	Formulaire de dépôt	New dataset	Permettre d'affilier un jeu de données à plusieurs unités organisationnelles s'il est commun à différents groupes de recherche.
Haute	Formulaire de dépôt	New dataset	<p>Avoir un bandeau avec les boutons « Save & quit », « Next », « Cancel » et « Request a DOI » de manière « sticky ».</p> <p>Prévoir des <i>pop-up</i> avec les différentes options pour avertir l'utilisateur quand il quitte le formulaire ou lorsqu'il soumet son jeu de données.</p> <p>L'utilisateur doit toujours avoir les options dont il a besoin sous les yeux.</p> <p><i>Exemple issu de la maquette :</i></p> 
Haute	Formulaire de dépôt	MyWorkspace	Donner des informations claires sur les stades des jeux de données non publiés, et les étapes à suivre en cas de jeu de données incorrect ou rejeté.
Haute	Formulaire de dépôt	Preview dataset new	Prévoir une page de « Preview » pour l'ajout de jeu de données pour que l'utilisateur ait une vue d'ensemble de son dépôt avant de le soumettre.
Moyenne	Aide à la complétion des champs	New dataset	Prévoir le remplissage automatique (ou la proposition de remplissage) du formulaire de dépôt de jeux de données en fonction des informations déjà connues par la plateforme. D'une part, le formulaire devrait être capable de récupérer des informations inhérentes au dépôt en fonction des fichiers choisis et de la personne qui remplit le formulaire (ex : nom de l'auteur, unité organisationnelle, etc.) ;

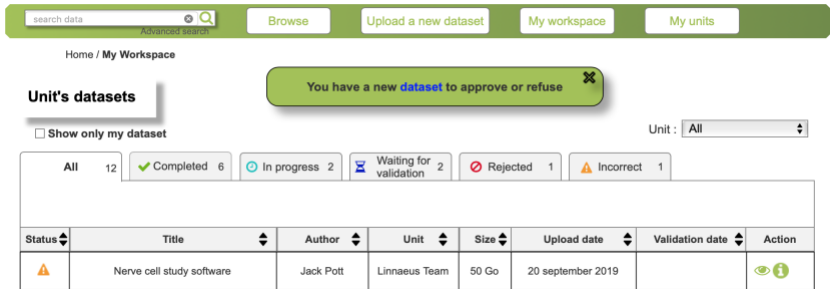
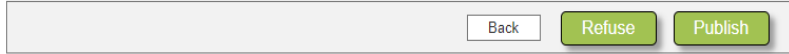
			<p>d'autre part, il devrait être possible de suggérer des contenus en fonction des premières lettres entrées dans certains champs libres (ex : mots clés et logiciels).</p> <p><i>Exemple issu du moteur de recherche Google :</i></p> 
Moyenne	Formulaire de dépôt	New dataset	<p>Prévoir l'enregistrement des métadonnées en cours de complétion et permettre aux usagers de revenir en arrière dans le formulaire d'ajout de jeux de données pour en modifier les informations avant la soumission du dépôt.</p>
Moyenne	Formulaire de dépôt	New dataset	<p>Créer une barre de progression cliquable pour le formulaire d'ajout d'un nouveau jeu de données.</p> <p><i>Exemple issu de la maquette :</i></p> 

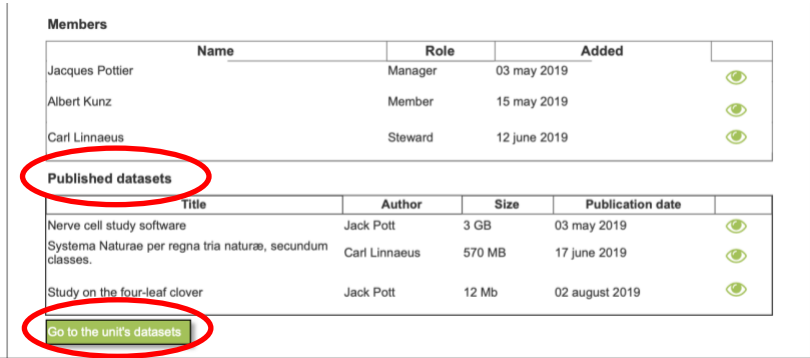
Moyenne	Formulaire de dépôt	New dataset	Ajouter le DOI automatiquement lors de la publication du jeu de données et donner la possibilité de réserver un DOI si l'utilisateur en a besoin avant la validation de la publication.																																
Moyenne	Formulaire de dépôt	Preview dataset new	Permettre de prévisualiser les fichiers du jeu de données dans la page de vérification avant soumission.																																
Moyenne	Formulaire de dépôt	New dataset	Prévoir la traduction des métadonnées dans les langues de la plateforme.																																
Faible	Formulaire de dépôt	New dataset	Prévoir la possibilité d'enregistrer des modèles de formulaire de dépôt, pour permettre aux chercheurs de gagner du temps lors du remplissage.																																
Faible	Formulaire de dépôt	New dataset	<p>Offrir la possibilité aux usagers d'organiser leurs fichiers dans le gestionnaire de fichiers du formulaire de dépôt.</p> <p><i>Exemple issu de la maquette :</i></p>  <table border="1"> <thead> <tr> <th>Title</th> <th>Size</th> <th>Format</th> <th>Creation date</th> </tr> </thead> <tbody> <tr> <td><input type="checkbox"/> Data.csv</td> <td>200 Mo</td> <td>Tabl</td> <td>27/07/19</td> </tr> <tr> <td><input type="checkbox"/> Readme.txt</td> <td>13 Mo</td> <td>Text</td> <td>27/07/19</td> </tr> <tr> <td><input type="checkbox"/> Data_2.doc</td> <td>14 Mo</td> <td>Text</td> <td>27/07/19</td> </tr> <tr> <td><input checked="" type="checkbox"/> Transcripts</td> <td>15 Go</td> <td>Folder</td> <td>27/07/19</td> </tr> <tr> <td> <input type="checkbox"/> Interview_1.flacc</td> <td>5 Go</td> <td>Video</td> <td>27/07/19</td> </tr> <tr> <td> <input type="checkbox"/> Interview_2.flacc</td> <td>4 Go</td> <td>Video</td> <td>27/07/19</td> </tr> <tr> <td> <input type="checkbox"/> Interview_3.flacc</td> <td>6 Go</td> <td>Video</td> <td>27/07/19</td> </tr> </tbody> </table> <p>At the bottom of the file list, there are buttons for 'Delete file(s)', 'Back', and 'Upload files' with a help icon.</p>	Title	Size	Format	Creation date	<input type="checkbox"/> Data.csv	200 Mo	Tabl	27/07/19	<input type="checkbox"/> Readme.txt	13 Mo	Text	27/07/19	<input type="checkbox"/> Data_2.doc	14 Mo	Text	27/07/19	<input checked="" type="checkbox"/> Transcripts	15 Go	Folder	27/07/19	<input type="checkbox"/> Interview_1.flacc	5 Go	Video	27/07/19	<input type="checkbox"/> Interview_2.flacc	4 Go	Video	27/07/19	<input type="checkbox"/> Interview_3.flacc	6 Go	Video	27/07/19
Title	Size	Format	Creation date																																
<input type="checkbox"/> Data.csv	200 Mo	Tabl	27/07/19																																
<input type="checkbox"/> Readme.txt	13 Mo	Text	27/07/19																																
<input type="checkbox"/> Data_2.doc	14 Mo	Text	27/07/19																																
<input checked="" type="checkbox"/> Transcripts	15 Go	Folder	27/07/19																																
<input type="checkbox"/> Interview_1.flacc	5 Go	Video	27/07/19																																
<input type="checkbox"/> Interview_2.flacc	4 Go	Video	27/07/19																																
<input type="checkbox"/> Interview_3.flacc	6 Go	Video	27/07/19																																

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Faible	Formulaire de dépôt	Preview dataset	new	Nommer la page de prévisualisation du formulaire « Preview ».
Gestion des jeux de données				
Haute	Gestion des jeux de données	MyWorkspace		<p>Créer un espace commun avec les jeux de données de l'unité et du chercheur et mettre en place une case à cocher pour ne sélectionner que ses propres jeux de données (ex : page « My workspace »).</p> <p>Ne pas utiliser de <i>toggle</i>, car les utilisateurs ne le voient pas ou ne savent pas l'utiliser.</p> <p><i>Exemple issu de la maquette :</i></p> 
Haute	Gestion des jeux de données	MyWorkspace		<p>Prévoir la gestion des jeux de données par plusieurs personnes, pour permettre une complétion collaborative des métadonnées au sein des unités organisationnelles.</p>

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

Moyenne	Intitulé du menu de gestion des jeux de données	MyWorkspace	Nommer la page qui regroupe les jeux de données de l'utilisateur, ainsi que ceux de son(ses) unité(s), « My workspace ».
Moyenne	Gestion des jeux de données	MyWorkspace	<p>Créer un <i>pop-up</i> pour la page « My workspace » pour mettre en évidence les actions que les usagers doivent réaliser sur certains jeux de données des unités (valider, corriger, etc.).</p> <p><i>Exemple issu de la maquette :</i></p> 
Moyenne	Validation des jeux de données	Preview dataset	<p>Sur la page de détail des jeux de données à valider, prévoir un menu « sticky », tout comme pour le formulaire, afin qu'il soit toujours visible sur la page.</p> <p><i>Exemple issu de la maquette :</i></p> 
Moyenne	Gestion des jeux de données	MyWorkspace	Permettre de supprimer un jeu de données.

Moyenne	Gestion des jeux de données	MyWorkspace	Permettre de versionner les jeux de données.
Moyenne	Gestion des jeux de données	MyWorkspace	Prévoir l'édition des métadonnées par la personne qui valide le jeu de données.
Moyenne	Gestion des jeux de données	Unit Linnaeus	<p>Prévoir un moyen d'accéder aux jeux de données (publiés ou non) des unités dont l'utilisateur est membre, depuis la page de l'unité en question.</p> <p><i>Exemple issu de la maquette :</i></p> 
Faible	Gestion des jeux de données	MyWorkspace	Prévoir des boutons d'actions pour les différents stades de maturité des jeux de données (ex : vue, validation, suppression, etc.).

Annexe 28 : Recommandations relatives aux métadonnées

Les recommandations portant sur les métadonnées sont rassemblées dans ce document qui permet également de comparer le schéma de métadonnée de DataCite avec les métadonnées demandées par le DLCM. Les recommandations concernent essentiellement la page de formulaire de dépôt (« New dataset », sur la maquette). Leur disposition et les options qui leur sont liées sont visibles sur la maquette à l'adresse : https://v4aj02.axshare.com/#g=1&id=vifv72&p=new_dataset. Nous vous rendons attentif au fait que certaines métadonnées ne sont pas implémentées sur la maquette.

De manière générale, afin de simplifier la complétion du formulaire de dépôt de données de recherche, nous recommandons de hiérarchiser les métadonnées selon trois niveaux : obligatoires, recommandées et optionnelles. Il est plus agréable pour les usagers qu'elles soient affichées au besoin, grâce à un menu déroulant.

Légende :

Métadonnées obligatoires

Métadonnées recommandées

Métadonnées optionnelles

Métadonnées demandées par le DLCM	Métadonnées du modèle DataCite	Recommandations	Commentaires
Creator(s)	Creator	Author: <ul style="list-style-type: none"> • name • surname • ORCID 	Ce champ doit être séparé en deux et pouvoir être dupliqué en cas de collaboration. Les ORCID devraient être récupérés automatiquement dans les profils des utilisateurs pour ne pas être demandés dans le formulaire de dépôt, mais tout de même être affichés dans le détail des jeux de données publiés.

Institution(s)		Organisational unit	<p>Ce champ doit pouvoir être dupliqué en cas de collaboration interinstitutionnelle.</p> <p>En plus de cela, il faudrait permettre aux auteurs de désigner la personne chargée de valider leurs jeux de données, car tous les membres d'un groupe de recherche ne connaissent pas tous les projets de leur groupe de la même façon.</p>
Title	Title	Dataset's title	Cet intitulé offre plus de clarté pour les usagers.
Description	Description	Description	Il est essentiel de prévoir un champ de description du jeu de données.
	Subject	Domain	La métadonnée « Subject » de DataCite doit être renommée « Domain » pour plus de clarté. Comme il est difficile de définir les domaines pertinents à prendre en compte, nous suggérons de rester assez général, mais de pouvoir dupliquer ce genre, afin d'affiner la précision. Nous proposons aussi qu'il soit obligatoire pour pouvoir proposer l'option de filtre correspondante dans la fonctionnalité de recherche.
		Collect method	Ces métadonnées complètent les métadonnées relatives à la description et au domaine. Elles sont importantes, car les éventuels réutilisateurs de jeux de données ont besoin de connaître le contexte de collecte.
		Data collection begin	
		Data collection end	
		Resource Type:	Ces deux métadonnées se complètent et permettent de définir les

ResourceType	ResourceType	<ul style="list-style-type: none"> • Audiovisual • Collection • DataPaper • Dataset • Event • Image • Interactive Resource • Model • PhysicalObject • Service • Software • Sound • Text • Workflow • Other 	<p>données rassemblées dans un jeu. Les aides associées doivent être claires</p>
		<p>Primary data type:</p> <ul style="list-style-type: none"> • Observational • Experimental • Simulation • Derived • Reference • Digitalized 	
Access Type		Access	<p>Ce champ doit être lié à d'autres options de métadonnées.</p> <p><i>Exemples issus de la maquette :</i></p> <ul style="list-style-type: none"> • Accès ouvert : <p>Access * <input type="button" value="Open"/></p>

Design d'un portail pour l'archivage des données de la recherche et proposition d'interfaces pour inciter les chercheurs/structures à déposer leurs données

			<ul style="list-style-type: none"> Accès restreint : <div> <div>Access * <input type="text" value="Restricted"/></div> <div>Embargo * <input type="text" value="Yes"/></div> <div> <input checked="" type="radio"/> Only my unit(s) <input type="radio"/> Customised: <div> <input type="checkbox"/> Linnaeus Team <input type="checkbox"/> Awesome Team <input type="checkbox"/> Checkbox </div> </div> </div> <div> <div>Access * <input type="text" value="Restricted"/></div> <div>Embargo * <input type="text" value="Yes"/></div> </div> <div> <input type="radio"/> Only my unit(s) <input checked="" type="radio"/> Customised: <div> Example of condition: Please contact linnaeus@team.ch to obtain this dataset </div> </div> Accès fermé : <div>Access * <input type="text" value="Closed"/></div>
		Embargo	Ce champ doit faire apparaître un champ supplémentaire, si l'utilisateur sélectionne « Yes », afin de délimiter la date de fin d'embargo et l'accès à la fin de ce dernier. Le DLCM a déjà très bien mis en place cette métadonnée sur son portail.
Licence	Rights	Licence	Cette métadonnée doit être associée à une aide, car les usagers éprouvent des difficultés avec les licences.
		Submission policies	Ce champ est lié aux unités organisationnelles, mais, dans le formulaire, il permet aux déposants d'adapter la métadonnée.
		Preservation policies	Ce champ est lié aux unités organisationnelles, mais, dans le formulaire, il permet aux déposants d'adapter la métadonnée.

Steward		Steward	Cette métadonnée est associée aux unités organisationnelles et aux rôles attribués à ses membres. Elle doit donc être paramétrée depuis l'unité.
Approver		Approver	Cette métadonnée est associée aux unités organisationnelles et aux rôles attribués à ses membres. Elle doit donc être paramétrée depuis l'unité.
DOI	Identifier	Data's DOI	Cette métadonnée doit être obligatoire, mais automatisée. Il serait avantageux que les usagers qui auraient besoin d'un DOI avant la publication du jeu de données puissent en « réserver » un et le connaître aussitôt.
Format	Size	Size	Ces deux métadonnées doivent être recueillies automatiquement lors du dépôt ; c'est pourquoi elles ne figurent pas dans le formulaire, mais uniquement dans le détail des jeux de données. La version doit être obligatoire, pour autant que le DLCM implémente cette fonctionnalité.
	Format	Format	
	Version	(Version)	
	Date	Dataset's publication date	Ce champ est obligatoire, mais il doit être automatisé en fonction de la date de validation du jeu de donnée.
Subject and optional vocabulary		Keywords	Ce champ est renommé pour plus de simplicité. Il est nécessaire à la fonctionnalité de recherche.
Software used		Software used	Cette information est importante pour les réutilisateurs.

External references Title		Publication's link	Les informations relatives à la publication de l'article doivent permettre d'en retrouver facilement la référence et de remettre les données dans leur contexte. Comme nous suggérons que ces métadonnées soient récupérées à partir du DOI de l'article, seul le lien vers la publication figure sur la maquette.
		Publication's title	
External references Creator(s)		Publication's author(s)	
External references Publisher	Publisher	Publication's publisher	
External references PublicationYear	PublicationYear	Publication's year	
Language	Language	Data Language	La métadonnée doit être renommée pour plus de clarté. Il faut aussi que ce champ puisse être dupliqué en cas de besoin.
Contributor(s)	Contributor	Other contributor(s) (name, surname, ORCID, rôle): <ul style="list-style-type: none"> ContactPerson DataCollector DataCurator DataManager Distributor Editor Hosting Institution 	Cette métadonnée enrichit le champ « Author(s) ». Il doit être renommé « Other contributor(s) » pour éviter une confusion avec « Author(s) ». L'ORCID devrait être récupéré automatiquement depuis les profils des contributeurs. La liste permet de qualifier la contribution. Il faudra porter une attention particulière à la rédaction de l'aide associée à ce champ.

		<ul style="list-style-type: none"> • Producer • ProjectLeader • ProjectManager • ProjectMember • Registration Agency • Registration Authority • RelatedPerson • Researcher • ResearchGroup • RightsHolder • Sponsor • Supervisor • WorkPackage Leader • Other 	
		Additional title	Il faut permettre aux usagers de traduire le titre de leur dépôt pour en faciliter la recherche.
External references DOI		External references DOI	
External references URL		External references URL	

Additional metadata and Schema		Additional metadata and Schema	Cette métadonnée doit permettre à l'utilisateur de lier un fichier.
Geolocation related information	GeoLocation	Geolocation related information	Cette métadonnée est décrite en détail par le document de DataCite ; toutefois, nous recommandons de ne pas demander trop de détail aux usagers, de peur de les décourager lors de la complétion du formulaire de dépôt.
DMP file		DMP file	Il faut permettre aux usagers d'attacher ce genre de document à leur dépôt, puisque c'est un document demandé par le FNS.
Relation(s) and Institution(s)			Cette métadonnée ne figure pas dans nos recommandations, car les informations sont comprises dans la métadonnée « Organisational unit », ainsi que sur le profil des personnes et qu'il n'est pas utile de demander aux déposants de remplir des informations déjà détenue par la plateforme.
ORCID (Creator, Contributor)			Cette métadonnée ne figure pas dans nos recommandations, car elle est comprise dans les profils des usagers et qu'il n'est pas utile de demander aux déposants de remplir des informations déjà détenues par la plateforme.
Related Identifier and relation type	RelatedIdentifier AlternateIdentifier	Other identifier (champ libre) Identifier type (liste)	DataCite détaille clairement les trois métadonnées retenues – avec des listes contrôlées notamment – et recommande de détailler encore davantage l'identifiant, mais nous craignons que tant de détails ne décourage les utilisateurs de la plateforme.

		Relation with the identifier (liste)	Cette métadonnée n'a pas été testée auprès des utilisateurs.
Funding information: grant ID, funder	FundingReference	Funding information: <ul style="list-style-type: none"> Name of the funding provide Type of funding (grant, ...) 	DataCite propose de détailler ces informations d'une façon précise. Afin de ne pas alourdir le formulaire, ce champ n'a pas été détaillé, ni testé. Nous suggérons toutefois que le nom et le type de financement figurent dans le formulaire.

Annexe 29 : Questionnaire *System Usability Scale*

	Pas du tout d'accord				Tout à fait d'accord
1. Je pense que je vais utiliser ce service fréquemment	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	1	2	3	4	5
2. Je trouve ce service inutilement complexe	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	1	2	3	4	5
3. Je pense que ce service est facile à utiliser	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	1	2	3	4	5
4. Je pense que j'aurai besoin de l'aide d'un technicien pour être capable d'utiliser ce service	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	1	2	3	4	5
5. J'ai trouvé que les différentes fonctions de ce service ont été bien intégrées	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	1	2	3	4	5
6. Je pense qu'il y a trop d'incohérence dans ce service.	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	1	2	3	4	5
7. J'imagine que la plupart des gens serait capable d'apprendre à utiliser ce services très rapidement.	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	1	2	3	4	5
8. J'ai trouvé ce service très lourd à utiliser.	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	1	2	3	4	5
9. Je me sentais très en confiance en utilisant ce service.	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	1	2	3	4	5
10. J'ai besoin d'apprendre beaucoup de choses avant de pouvoir utiliser ce service.	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	1	2	3	4	5