

Swiss PID Hub : Création d'un Hub pour la gestion des identifiants pérennes en Suisse

Mémoire de recherche réalisé par :

Matthieu CEVEY

Julien Antoine RAEMY

Sous la direction de :

René SCHNEIDER, professeur HES

Carouge, le 14 janvier 2020

**Information documentaire
Master HES-SO en Sciences de l'Information
Haute école de gestion de Genève (HEG-GE)**

Remerciements

Nous tenons ici à vivement remercier toutes les personnes qui nous ont aidées à réaliser ce projet de recherche effectué dans le cadre du Master of Science HES-SO en Sciences de l'Information à la Haute école de gestion de Genève.

Tout particulièrement, nous sommes très reconnaissants d'avoir bénéficié des nombreux conseils avisés et de la disponibilité du Professeur René Schneider qui nous a encadré tout au long de ce projet.

Nous sommes également redevables envers John A. Kunze, Identifier Systems Architect à la California Digital Library et créateur du schème ARK, ainsi qu'à Bertrand Caron, expert métadonnées de préservation à la Bibliothèque nationale de France, d'avoir pris le temps de répondre à nos sollicitations et de nous avoir aiguillé vers les bonnes ressources.

Merci aussi aux personnes qui ont répondu à nos messages ou qui ont participé à nos entretiens. Dans l'ordre chronologique :

- Christine Pirinoli, Isabelle Lucas, Christophe Batteur (HES-SO)
- Gabi Schneider (swissuniversities)
- Gwënola Dos Santos, Virginie Keller (Infothèque HEG-GE)
- Jon Stroop (Princeton University Library)
- Andreas Nef (Docuteam)
- Patrick Ruch (HEG-GE et Swiss Institute of Bioinformatics)
- Didier Grange, Xavier Ciana (Archives de la Ville de Genève)
- Alexis Rivier, Sylvain Fejoz (Bibliothèque de Genève)

Enfin, nous remercions nos collègues Claire Wuillemin, Anouk Santos, Luc Mottin et Benoît Epron qui ont émis des critiques constructives dans le cadre de l'élaboration et de la présentation de notre poster scientifique.

Résumé

Le projet de recherche « Swiss PID Hub : création d'un Hub pour la gestion des identifiants pérennes en Suisse » s'inscrit dans la continuité du projet de recherche de Espasandin, Jaquet et Lefort, étudiantes du Master en Sciences de l'Information (*Master of Science HES-SO in Information Science – Master IS*) entre 2016 et 2018, qui a permis d'établir un panorama des identifiants pérennes utilisés de nos jours, et au projet Identités de Confiance pour les Données de l'Art et du Design (ICOPAD), réalisé entre juin 2017 et décembre 2018 à la Haute école de gestion de Genève (HEG-GE).

Actuellement, la situation en matière d'attribution et de coordination des identifiants pérennes (*persistent identifiers – PID*) en Suisse n'est pas satisfaisante. Ce travail a donc pour but d'étudier les possibilités de création d'un Hub national offrant une solution « identifiants pérennes pour tous et tout ». L'enjeu principal réside dans la proposition d'un modèle conceptuel de cette infrastructure facilitant l'attribution et la gestion de PID, et en particulier d'Archival Resource Keys (ARK), en analysant les moyens techniques, juridiques, organisationnels et financiers nécessaires à sa création et à son fonctionnement.

Ce projet fait également office d'étude préparatoire au projet Infrastructure Nationale d'un Complément d'Identifiants Pérennes, Interopérables et Traçables (INCIPIT) soumis en août 2019 au fonds « P-5 Information scientifique » de swissuniversities lors de l'appel à projets pour l'année 2020 (Call 192).

Ce travail s'applique à suivre une méthodologie de recherche cohérente et adaptée. En nous basant sur une revue de la littérature, nous avons pu analyser l'existant et extraire des études de cas qui nous ont permises de modéliser un prototype du Hub et de ses services. Des entretiens avec de potentiels partenaires et clients nous ont amenés à proposer une série de recommandations destinées à faciliter la mise en place du Hub.

Les services principaux de celui-ci consisteront en la possibilité de créer, réserver, modifier et supprimer un ARK tout en gardant une traçabilité grâce aux déclarations de permanence des objets. Techniquement, nous estimons que l'utilisation d'ARK est la solution la plus pertinente et efficace dans le contexte helvétique, celle-ci répondant au mieux aux besoins de la communauté scientifique, sans compter qu'elle s'inscrit parfaitement dans la mouvance de l'Open Science.

Ce travail établit les conditions nécessaires à l'implémentation de ce Hub national, en matière d'outils et de critères indispensables permettant d'assurer la pérennité du système, à savoir l'unicité des identifiants, la responsabilité des agents assignant des ARK et la possibilité de modifier les métadonnées.

La création d'une telle infrastructure est une opportunité qui permettra à la Suisse de continuer à jouer un rôle de premier plan en matière d'innovation scientifique et technologique, tout en participant pleinement aux mouvements d'ouverture de la science.

Mots-clefs : AITO, API, ARK, cool URI, DOI, ePIC, ERC, FAIR, HEG-GE, HES-SO, ICOPAD, identifiants pérennes, IIF, INCIPIT, Linked Open Data, métadonnées, N2T, Open Science, PID, SIB, swissuniversities, URN, URL, Web sémantique

Table des matières

Remerciements.....	i
Résumé	ii
Liste des tableaux	vi
Liste des figures.....	vii
Liste des abréviations et des acronymes	viii
1. Introduction.....	1
2. Contexte	2
2.1 Cadre général	2
2.2 Nature du mandat.....	2
2.3 Origines du projet	2
2.4 Objectifs généraux et spécifiques	3
2.4.1 Objectif 1 – Conceptualisation	3
2.4.2 Objectif 2 – Prototypage.....	3
2.4.3 Objectif 3 – Recommandations	3
2.5 Questions de recherche	4
3. Revue de la littérature	5
3.1 Périmètre.....	5
3.2 Vue d'ensemble des PID.....	8
3.3 Archival Resource Key (ARK)	8
3.3.1 Principes et gouvernance.....	9
3.3.2 Structure du schème ARK.....	10
3.3.3 Métadonnées associées et déclaration de permanence	11
3.3.4 Avantages	13
3.3.5 ARK en Suisse	14
3.4 Implémentation d'un système d'attribution d'identifiants ARK.....	15
3.4.1 Processus d'implémentation générique de PID	15
3.4.2 Exemples d'implémentation d'un système ARK.....	15
3.4.3 Implémentation d'un Hub en Suisse	16
4. Méthodologie	19
4.1 Démarche globale	19
4.1.1 Cahier des charges (04/2019).....	19
4.1.2 Data Management Plan (05/2019).....	20
4.1.3 Poster scientifique (12/2019)	20
4.1.4 Billets de blog axé données (06/2019) et axé projet (12/2019).....	20
4.2 Recensement des informations	20
4.2.1 Résultats du projet ICOPAD	21
4.2.2 Entretiens	21
4.2.3 Revue de la littérature	21
4.2.4 Échanges avec l'initiative AITO.....	22

4.3	Analyse des données et modélisation	22
4.3.1	Études de cas	22
4.3.2	Prototypage.....	22
4.3.3	Recommandations pour le projet INCIPIT	23
5.	Études de cas.....	24
5.1	Sélection	24
5.1.1	Études de cas du projet ICOPAD.....	24
5.1.2	Études de cas de la FAQ ARK.....	24
5.1.3	Études de cas d'ArODES.....	24
5.1.4	Études de cas du SIB.....	25
5.1.5	Études de cas sélectionnées	25
5.2	Analyse.....	28
6.	Modélisation.....	31
6.1	Comparaison de services similaires	31
6.1.1	EZID (CDL)	31
6.1.1.1	Processus d'attribution et fonctionnalités	32
6.1.1.2	Domaines d'utilisation.....	33
6.1.1.3	Tarification	33
6.1.2	DOI-Desk (Bibliothèque de l'EPFZ)	34
6.1.2.1	Processus d'attribution et fonctionnalités	34
6.1.2.2	Domaines d'utilisation.....	34
6.1.2.3	Tarification	34
6.1.3	Obtention d'URN et d'identifiants ePIC en Suisse	35
6.1.3.1	Service URN de la Bibliothèque nationale suisse	35
6.1.3.2	ePIC par le biais du Centre suisse de calcul scientifique.....	35
6.2	Vue d'ensemble des critères de prototypage.....	35
6.2.1	Considérations techniques.....	36
6.2.1.1	Générateur d'identifiants ARK et résolveur local.....	36
6.2.1.2	Résolveur global	36
6.2.1.3	Schémas de métadonnées et déclarations de permanence	37
6.2.1.4	Infrastructure.....	38
6.2.2	Considérations juridiques.....	39
6.2.2.1	Forme juridique du Hub	39
6.2.2.2	Contrat de prestations de service.....	40
6.2.2.3	Hébergement des données	41
6.2.3	Considérations organisationnelles	41
6.2.3.1	Prestations offertes.....	41
6.2.3.2	Ressources humaines	41
6.2.4	Considérations financières.....	42
6.2.4.1	Estimation des coûts	42
6.2.4.2	Propositions de tarification	43
6.2.4.3	Projections sur cinq ans	44
6.3	Processus d'attribution	45
6.4	Blueprint de service	47
7.	Recommandations.....	48
7.1	Récits utilisateurs	48

7.2	Fonctionnalités requises	50
7.3	Viabilité financière et organisationnelle.....	51
7.4	Facteurs clefs de succès.....	51
7.5	Critères d'innovation	52
8.	Conclusion	53
	Bibliographie	55
Annexe 1 :	<i>Ten persistent myths about persistent identifiers</i>	63
Annexe 2 :	Poster scientifique	64
Annexe 3 :	Analyse détaillée des études de cas	65
Annexe 4 :	Maquette de l'interface utilisateur	67
Annexe 5 :	Vue d'ensemble des processus d'attribution	68
Annexe 6 :	Blueprint de service (format paysage)	69

Liste des tableaux

Tableau 1 : Description détaillée de la structure ARK	10
Tableau 2 : Objet ARK, métadonnées associées et déclaration de permanence	12
Tableau 3 : Avantages financiers, organisationnels et techniques du schème ARK.....	13
Tableau 4 : Organisations établies en Suisse inscrites au registre des NAAN	14
Tableau 5 : ARK - exemples d'implémentation.....	15
Tableau 6 : Études de cas pour le Swiss PID Hub	25
Tableau 7 : Études de cas par typologie d'objet.....	29
Tableau 8 : Opérations du service EZID.....	32
Tableau 9 : Scénario optimiste d'estimation des coûts.....	43
Tableau 10 : Scénario pessimiste d'estimation des coûts	43
Tableau 11 : Proposition de tarification 1	43
Tableau 12 : Proposition de tarification 2	44
Tableau 13 : Récits utilisateurs et conséquences sur le déploiement du Hub	48
Tableau 14 : Analyse détaillée des études de cas	65

Liste des figures

Figure 1 : Comparaison entre PID et cool URI	7
Figure 2 : Object Curation Domains	8
Figure 3 : Premières étapes de création d'un service ARK	9
Figure 4 : Anatomie d'un identifiant ARK	10
Figure 5 : Description et déclaration de permanence (ERC)	12
Figure 6 : Politique de permanence ARK de la BnF	13
Figure 7 : Recording to Archiving Architecture (identité de confiance)	17
Figure 8 : Grille de comparaison des études de cas (métadonnées et permanence)	30
Figure 9 : Architecture ARK à la CDL	31
Figure 10 : Processus d'enregistrement des DOI via la Bibliothèque de l'EPFZ	34
Figure 11 : Structure du résolveur global N2T	37
Figure 12 : Interface de recherche d'EZID	39
Figure 13 : projection du chiffre d'affaires	44
Figure 14 : Projection de rentabilité	45
Figure 15 : Logigramme de création d'ARK	45
Figure 16 : Logigramme de réservation d'ARK	46
Figure 17 : Logigramme de modification ou de suppression d'ARK	46
Figure 18 : Blueprint de service	47
Figure 19 : Poster scientifique du Swiss PID Hub	64
Figure 20 : Maquette basse fidélité de la page web « Création d'ARK »	67
Figure 21 : Logigramme d'attribution d'identifiants ARK (grand format)	68
Figure 22 : Blueprint de service (grand format)	69

Liste des abréviations et des acronymes

AITO	ARKs-in-the-Open
ANVL	A Name-Value Language
API	Application programming interface
ArODES	Archive institutionnelle des domaines de la HES-SO
ARK	Archival Resource Key
ASCII	American Standard Code for Information Interchange
BN	Bibliothèque nationale suisse
BnF	Bibliothèque nationale de France
BnL	Bibliothèque nationale du Luxembourg
CA	Chiffre d'affaires
CC	Creative Commons
CDL	California Digital Library
CDNL	Conference of Directors of National Libraries
CENL	Conference of European National Librarians
CIDOC-CRM	CIDOC Conceptual Reference Model
CSCS	Centre suisse de calcul scientifique
CSV	Comma-separated value
DaSCH	Data and Service Center for the Humanities
DC	Dublin Core
DHS	Dictionnaire historique de la Suisse
DK	Dublin Kernel
DNS	Domain Name System
DOI	Digital Object Identifier
EAD	Encoded Archival Description
EPFZ	École polytechnique fédérale de Zürich
ePIC	Persistent Identifiers for eResearch
EPT	Équivalent plein temps
ERC	Electronic Resource Citation
FAIR	Findable, Accessible, Interoperable, Reusable

FOAF	Friend of a Friend
GND	Gemeinsame Normdatei
GRID	Global Research Identifier Database
HEG-GE	Haute école de gestion de Genève
HES	Haute école spécialisée
HES-SO	Haute école spécialisée de Suisse occidentale
HTTP(S)	The Hypertext Transfer Protocol (Secure)
ICA	International Council on Archives
ICOPAD	Identités de confiance pour les données de l'art et du design
ID	Information documentaire
IETF	Internet Engineering Task Force
IIIF	International Image Interoperability Framework
INCIPIT	Infrastructure nationale d'un complément d'identifiants pérennes, interopérables et traçables
IPFS	InterPlanetary File System
ISNI	International Standard Name Identifier
JSON-LD	JavaScript Object Notation for Linked Data
LOCKSS	Lots of Copies Keep Stuff Safe
LOD	Linked Open Data
LODZ	Linked Open Data Zürich
MARC	Machine-Readable Cataloging
Master IS	Master of Science HES-SO in Information Science
NAAN	Name Assigning Authority Number
NBN	National Bibliography Number
NMA	Name Mapping Authority
NOID	Nice Opaque Identifiers
N2T	Name-to-Thing
OAI-PMH	The Open Archives Initiative Protocol for Metadata Harvesting
ÖNB	Österreichische Nationalbibliothek
ORCID	Open Researcher and Contributor ID

OWL	Web Ontology Language
PFPDT	Préposé fédéral à la protection des données et à la transparence
PID	Persistent identifier
P5	swissuniversities – P5 Information scientifique
RAID	Redundant Array of Independent Disks
RDA	Resource Description and Access
RDF	Resource Description Framework
RFC	Request for Comments
RIC	Records in Context
ROR	Research Organization Registry
SIB	Swiss Institute of Bioinformatics
SIK-ISEA	Schweizerische Institut für Kunstwissenschaft – Institut suisse pour l'étude de l'art
SPARQL	SPARQL Protocol and RDF Query Language
Turtle	Terse RDF Triple Language
URL	Uniform Resource Locator
URN	Uniform Resource Name
UUID	Universally Unique Identifier
UX	User experience
VIAF	Virtual International Authority File
W3C	World Wide Web Consortium
XML	eXtensible Markup Language
ZB	Zentralbibliothek Zürich
ZHdK	Zürcher Hochschule der Künste

1. Introduction

Aujourd'hui, à l'heure où les citoyens cherchent à reprendre le contrôle de leurs données personnelles, à apprendre à maîtriser des flux d'information toujours plus conséquents et nombreux, à l'heure où les gouvernements promettent la transparence et où le concept de *smart cities* fait son chemin, en ce jour où jamais la science n'avait autant produit de données et résultats, il nous est apparu incroyable qu'il existe si peu de recommandations officielles quant à la pérennisation de toute cette information. Information qui dans la plupart des cas est produite grâce aux subventions de la Confédération. Information qui n'est donc pas toujours accessible, pas toujours interoperable, pas toujours réutilisable.

Mais le mouvement Open Science prend de l'ampleur, s'étend, et le travail que nous avons réalisé ici, que, nous espérons, vous prendrez plaisir à lire, participe pleinement de ce progrès en proposant de rationaliser les efforts et les investissements des institutions culturelles et des organismes de recherche tout en offrant un but et une vision commune à l'ensemble de la communauté scientifique suisse. Travailler ensemble, conjuguer les compétences et œuvrer pour le bien commun, autant de buts louables que nous nous proposons d'atteindre en créant un service et des outils servant concrètement les intérêts de la recherche.

Le Web, fabuleux outil dont nos ancêtres n'auraient même pas osé rêver, est un moyen idéal de partager des informations, et pourtant, celles qui y sont publiées se perdent facilement et se modifient perpétuellement. L'archivage du Web avec des initiatives communautaires, telle qu'Internet Archive, est une des réponses possibles *a posteriori* au problème. Celle que nous présentons dans ces lignes propose de s'y attaquer à sa racine, en identifiant dès leur création, voire au préalable, les données produites, quel que soit le contexte.

La réponse à cette problématique, initiée par l'apparition des catalogues des premières bibliothèques et sans cesse développée par l'humanité depuis, se trouve, selon nous, dans une gestion intelligente et adaptée des identifiants pérennes. C'est pourquoi nous nous sommes intéressés à la création et à la mise en place d'un Hub d'attribution et de gestion d'Archival Resource Keys, un service d'envergure nationale, qui servirait de façon adéquate les intérêts de la communauté scientifique.

Pour poser les fondations de ce Hub, nous avons analysé le contexte dans lequel il prendrait place, avons parcouru la littérature existante sur le sujet, et défini une méthodologie adaptée. Sur ces bases, nous avons sélectionné des études de cas, qui nous ont inspirées pour effectuer une modélisation du service, en prenant en compte les aspects techniques, juridiques, organisationnels et financiers. Cela nous a également permis d'établir des processus de fonctionnement et d'émettre des recommandations ciblées pour que ce projet puisse, un jour, bénéficier d'une réalisation effective.

2. Contexte

2.1 Cadre général

La gestion des données est une problématique actuelle, tant au niveau d'une institution qu'au niveau national, voire mondial. Pour faciliter la diffusion et l'utilisation de connaissances, et pour être en accord avec les principes FAIR (*Findable, Accessible, Interoperable, Reusable*) (Wilkinson et al., 2016), une solution a été de créer des identifiants pérennes (*persistent identifiers* - PID) permettant d'améliorer la gestion des données, sur le plan de leur référencement, de leur accessibilité et donc de leur visibilité.

Or, il existe une quantité importante de ces identifiants, dont les principaux ont été recensés dans le travail de recherche « Panorama et modélisation d'identifiants pérennes pour la création d'identités de confiance » (Espasandin et al., 2018) sous la direction de M. René Schneider, Professeur HES, ce qui, au final, n'aide que partiellement à la bonne gestion des données de la recherche. Notre travail consistera donc en l'évaluation des différentes possibilités et contraintes liées à la création d'un Hub national permettant l'attribution logique d'identifiants pérennes complémentaires aux Digital Object Identifiers (DOI).

2.2 Nature du mandat

La nature de ce projet est partagée entre des aspects techniques, juridiques, organisationnels et financiers. L'enjeu est de proposer au moins un modèle conceptuel de Hub facilitant l'attribution et la gestion de PID, notamment d'Archival Resource Keys (ARK), sur le territoire suisse.

2.3 Origines du projet

Proposé par le Professeur René Schneider dans le cadre de ses travaux en Linked Open Data (LOD), ce mandat fait suite au projet de recherche de Espasandin, Jaquet et Lefort (2018), étudiantes du Master en Sciences de l'Information (*Master of Science HES-SO in Information Science – Master IS*) entre 2016 et 2018, qui a permis d'établir un panorama des identifiants pérennes utilisés de nos jours, et au projet Identités de Confiance pour les Données de l'Art et du Design (ICOPAD¹), réalisé entre juin 2017 et décembre 2018 à la Haute école de gestion de Genève (HEG-GE). La situation en Suisse concernant l'attribution et la coordination de ces identifiants n'est actuellement pas satisfaisante (Schneider, Raemy, 2019a ; Raemy, Schneider, 2019c ; Brocard, 2019), nous allons donc étudier dans ce travail les possibilités de création d'un Hub offrant une solution « identifiants pérennes pour tous et tout » au niveau national.

Swiss PID Hub fait également office d'étude préparatoire (cf. 3.4.3) au projet Infrastructure Nationale d'un Complément d'Identifiants Pérennes, Interopérables et Traçables (INCIPIT) soumis en août 2019 au fonds « P-5 Information scientifique » de swissuniversities (P5) lors de l'appel à projets pour l'année 2020 (*Call 192*), visant à financer des projets qui se basent sur les mesures de la Stratégie de mise en œuvre 2017 – 2020 (swissuniversities, 2019a ; Dieudé, 2019).

¹ Projet ICOPAD : https://campus.hesge.ch/id_bilingue/projekte/icopad/index_fr.asp (consulté le 10 janvier 2020)

Les contraintes liées à la création d'un tel Hub sont nombreuses, et devront être analysées avec soin dans un premier temps afin d'en tirer les conclusions nécessaires à sa réalisation, dans un second temps, une proposition de modélisation.

Ce projet, de par l'implication d'un grand nombre d'acteurs, d'institutions et de concepts techniques, est ambitieux mais offrirait, s'il peut être réalisé, de grands avantages, notamment dans le développement du LOD et la création d'un modèle de référence.

2.4 Objectifs généraux et spécifiques

Pour ce projet de recherche, nous avons défini trois objectifs généraux (en gras) déclinés en objectifs spécifiques (listes à points).

2.4.1 Objectif 1 – Conceptualisation

Créer un cadre conceptuel autour d'une solution helvétique, complémentaire aux DOI, permettant d'attribuer des ARK à tous les types de données

- Rédiger un état de l'art sur l'implémentation de systèmes, d'initiatives ou de projets attribuant des PID ;
- Établir une grille d'entretien pour réaliser des entretiens semi-dirigés ;
- Contacter des personnes clés travaillant dans les organisations suisses ayant les capacités de devenir des Hub ;
- Analyser les entretiens et ajuster la direction du projet de recherche si nécessaire.

2.4.2 Objectif 2 – Prototypage

Modéliser un prototype du Hub

- Rechercher et choisir des études de cas hétérogènes provenant des données de potentiels partenaires ;
- Modéliser les études de cas en les déclinant en différentes entités ;
- Établir un workflow des processus d'attribution ;
- Établir une liste de critères indispensables pour la création d'outils de génération et de résolution d'identifiants ;
- Créer un Blueprint de service du Hub.

2.4.3 Objectif 3 – Recommandations

Établir des recommandations pour la création du Hub

- Obtenir des commentaires par des professionnels sur nos études de cas ;
- Établir des recommandations techniques (logiciels, résolveur de liens, métadonnées, *mapping* selon le type d'entité et les métadonnées disponibles) ;
- Faire une estimation globale des coûts de création et de fonctionnement ;
- Établir une liste de critères indispensables nécessaires à la création du Hub.

2.5 Questions de recherche

Ce projet vise à répondre aux trois questions de recherche suivantes :

- Quelles sont les conditions (techniques, juridiques, organisationnelles et financières) pour créer une infrastructure permettant d'attribuer des identifiants pérennes complémentaire aux DOI pour tous ?
- Comment définir le moment optimal pour l'attribution d'un PID selon les domaines du *Data Curation Continuum Model*² définis par Andrew Treloar (*private, collaboration, publication*) ?
- Comment assurer une couche LOD pointant vers d'autres autorités ?

² Cf. Figure 2

3. Revue de la littérature

3.1 Périmètre

Depuis que la diffusion des connaissances peut se faire au moyen de l'écrit, il a été nécessaire d'identifier les sources de ce savoir, généralement des livres, rouleaux et parchemins à l'époque, puis des documents électroniques dès la fin du XX^e siècle. D'abord simples étiquettes accrochées aux rouleaux de parchemin, puis colophons en fin de volume, jusqu'à la publication de catalogues recensant des notices bibliographiques, parfois propres à un exemplaire particulier, et enfin l'idée d'identifiants pérennes qui s'impose au XXI^e siècle. De tout temps il a donc fallu, pour l'avancée de la recherche, pour la compréhension de notre monde et pour notre divertissement (ne l'oublions pas), pouvoir identifier un document. Ce besoin d'identification suppose donc des méthodes de classement et, plus important encore, des règles et standards permettant d'harmoniser la description des objets identifiés et ainsi permettre une diffusion facilitée très large, au bénéfice de l'ensemble de la société.

L'avènement de l'informatique, et d'Internet, a considérablement modifié les façons de faire, en offrant entre autres des outils et des méthodes inédites jusque-là. Mieux encore, les technologies permettent aujourd'hui la mise à disposition des données (issues de la recherche, d'expériences) utilisées pour la création du document. Assez logiquement, des besoins d'identification des concepts, des auteurs et des terminologies se sont fait sentir avec l'évolution de la pratique des sciences, et donc des solutions ont été trouvées. Celle sur laquelle nous nous penchons dans ce travail est celle des PID.

Cette solution, ou plutôt ce faisceau de solutions est fortement lié au LOD : outre le besoin d'adjointre des métadonnées descriptives à un objet, il est nécessaire aujourd'hui de pouvoir faire le lien entre ces données et en favoriser ainsi la recherche et le partage. Le LOD ne peut fonctionner qu'en présupposant l'existence d'identifiants fiables et pérennes, rendant ainsi nécessaire l'adoption d'une solution, ou du moins en assurant l'interopérabilité entre plusieurs solutions d'identification des objets.

Si vous avez déjà atterri sur une page « *404 Not Found* » sur Internet, et c'est plus que probable, vous savez à quel point ne pas trouver une page web censée être disponible à une adresse précise est mauvais pour l'image de l'institution ou de l'organisme qui l'héberge. L'on parle carrément de « désastre culturel » dans ce cas (Kunze, Caron, 2019a), particulièrement pour les bibliothèques, centres d'archives, musées et autres institutions liées au patrimoine et à la mémoire d'une société. Il en va dès lors de leur réputation en plus de l'enjeu primordial de l'accès à l'information.

En format Dublin Core, un identifiant est « *une chaîne de caractères identifiant une ressource de façon unique* » (Espasandin et al., 2018), et un identifiant pérenne se caractérise par sa constitution en trois parties (préfixe, autorité nommante et identifiant) offrant la possibilité d'assurer l'identification de la ressource dans le temps. Espérer le maintien d'un identifiant pour un temps infini est complètement illusoire, aussi il est plutôt question de garantir une maintenance à moyen ou long terme grâce à la désignation de l'institution responsable de celle-ci dans l'identifiant lui-même (autorité nommante).

Dès lors, on comprend que c'est le service chargé du PID qui permettra d'assurer la pérennité de l'accès à une ressource, et donc que c'est le choix d'un système d'identifiants pérennes qui s'avère délicat, même si l'interopérabilité est possible. Les notions de trouvabilité,

d'accessibilité, d'interopérabilité et de réutilisabilité, ou autrement dit, les principes FAIR (Wilkinson et al., 2016) à la base de l'Open Science, et toutes les actions et infrastructures sous-tendant l'implémentation de solutions propres à soutenir ces principes, incitent vivement le monde de la recherche à adopter une solution simple, peu coûteuse et fonctionnelle pour identifier une ressource.

L'amélioration de la conception, de la mise à disposition, de la documentation et du référencement des identifiants est à même de résoudre la plupart des problèmes liés au cycle des données, et permettrait donc une science plus efficace et efficiente. Cependant, une bonne pratique ne sera pas adoptée qu'en raison des avantages qu'elle offre à la collectivité, mais également en raison des facilités de mise en œuvre et de gestion, y compris financière, par ses responsables. Les défis ne sont pas les mêmes pour les émetteurs d'identifiants que pour ceux qui les référencent, mais la seule solution pour développer une solution « pour tous et tout », est d'inciter tous les acteurs de cet écosystème à travailler ensemble avec le but commun d'harmoniser les identificateurs (Van de Sompel et al., 2014 ; McMurry et al., 2017).

Lorsque l'on sait que pour une série d'articles publiés entre 1997 et 2012, 75% des références citées ont disparu ou été modifiées depuis (Jones et al., 2016), la question de l'intégrité à long terme des publications scientifiques devient particulièrement préoccupante. Et s'il est tout à fait concevable d'imaginer qu'en ce qui concerne les ressources web non scientifiques, ce taux est bien plus important, l'on peut se poser la question de savoir si pour certaines sources, il est plus ou moins préjudiciable de rencontrer des problèmes d'identification. Dans tous les cas, c'est la collectivité qui en pâtit.

Emmanuelle Bermès (2006) soulève quelques problèmes liés au choix des identifiants, notamment celui de leur finalité (*le but est-il la citabilité ou l'unicité et la pérennité ?*) et de leur significativité. En effet, dans l'optique d'identifiants fonctionnant à long terme, un identifiant signifiant risque de perdre son sens avec le temps (changement de nom de l'institution, modification de la nature de la ressource, problèmes de la langue et de son évolution, etc.).

Dès lors, la solution d'identifiants générés par des machines, construits par un algorithme normalisé sur la base d'informations techniques et temporelles pourraient pallier ce problème, mais en fait apparaître un autre : la nécessité de conserver un lien entre la ressource décrite et son identifiant, puisque celui-ci ne comporte en son sein aucune information sur cette ressource. La dualité de cette problématique reflète exactement l'éternelle position des bibliothèques : doivent-elles privilégier la communication (consultation/indexation) ou la conservation (préservation/archivage) ? L'un des défis du choix d'identifiants sera donc de parvenir à réconcilier ces deux aspects pour offrir un service adéquat à la collectivité (Bermès, 2006).

Pour une institution, assurer la persistance des identifiants s'assimile à offrir un service selon Kunze et al. (2017), ce qui implique donc de prévoir quelles sont les ressources nécessaires pour ce faire, aussi bien au niveau matériel et technologique qu'au niveau financier et humain. L'idée d'un fournisseur unique qui gère un résolveur pour la réorientation d'identifiants afin de soutenir l'accès à long terme à des objets de toutes sortes est soulevée, mais il est indispensable dès lors de disposer d'une terminologie normalisée commune afin que les fournisseurs puissent rendre, d'une certaine façon, leurs politiques et leurs pratiques interopérables.

Historiquement, les bibliothèques et autres dépôts de savoir ont été les premiers à développer un système d'identificateur persistant, permettant l'identification unique - et parfois même l'accès - à une ressource. Plus récemment, le succès du Web a amené les chercheurs et experts au concept de *Cool URI* comme moyen universel d'identification et d'accès aux ressources électroniques (Berners-Lee, 1998 ; Sauermann, Cyganiak, 2008 ; Cevey, Raemy, 2020). Développées en parallèle, ces deux méthodes ont chacune leurs avantages et leurs inconvénients, et leur coexistence n'a pas toujours été simple (Bazzanella et al., 2013).

De nombreuses initiatives se concentrent encore sur l'approche *Cool URI*, idée promue par le World Wide Web Consortium (W3C) et principalement utilisée dans le domaine LOD. Des projets antérieurs gérés par la HEG-GE, tels que linked.swissbib.ch et Linked Open Data Zurich (LODZ), ont montré que les PID sont essentiels pour avoir une infrastructure solide (Prongué et al., 2017 ; Raemy, Schneider, 2019a). Seule leur utilisation permet le développement de services stables tout en réglant des problèmes tels que l'autorité, le niveau de confiance, le caractère unique, la durabilité et le rapprochement des métadonnées, pour n'en citer que quelques-uns (cf. Figure 1).

Figure 1 : Comparaison entre PID et cool URI

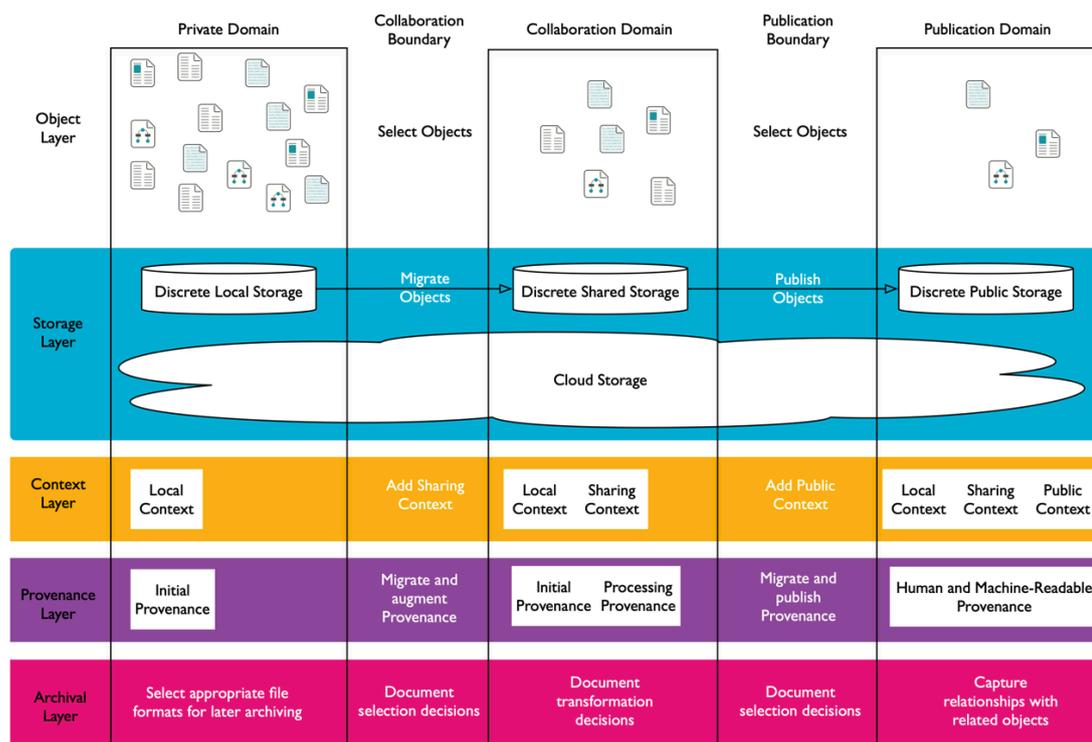
Feature	PIDs	Cool URIs
Resolver	YES	NO
Authority	YES	NO
Naming authorities	YES	NO
Level of trust	HIGH	LOW
Policies	YES	NO
Persistence	YES	NO
Actionability of IDs	Partially	YES
Uniqueness	YES	NO
Content change	NO	YES
Content negotiation	NO	YES
Cross linkage	NO	YES
Effort for implementation	HIGH	LOW
Costs for users	Potentially HIGH	LOW
Sustainability issues	MANY	FEW
Identified entities	Mainly digital objects	Everything
Bridge metadata	NO	YES

(Schneider, Raemy, 2019b inspirés de ; Bazzanella et al., 2013)

Généralement, un PID est attribué à un objet lors de sa publication ou de son archivage, ce qui correspond à la troisième phase (Treloar, Klump, 2019) des *Object Curation Domains* (cf. Figure 2 ci-après). Si l'enrichissement de métadonnées est toujours plus important au fil d'une recherche, l'attribution d'identifiants (pérennes ou dynamiques) est néanmoins essentiel durant tout le processus de la gestion des données (McMurry et al., 2017 ; Raemy, 2018 ; Raemy, Schneider, 2019a).

Pour éviter la multiplication d'identifiants pour un seul objet et ses composantes, réserver un PID avant leur publication et gérer leur versionnage sont désormais des fonctionnalités que certains dépôts proposent, comme par exemple Zenodo (DoRANum, 2017 ; Zenodo, 2019).

Figure 2 : Object Curation Domains



Version 2.2, @atreloar, 3 Mar 2019
 Research Objects image CC-BY <http://researchobject.org/>

(Treloar, Klump, 2019)

3.2 Vue d'ensemble des PID

Un recensement des PID et de leurs potentialités a été effectué dans le cadre du projet de recherche mené par (Espasandin et al., 2018), et un magnifique panorama très pratique fournit une très bonne vue d'ensemble du domaine³. Notre projet s'inscrivant directement dans sa lignée, nous nous abstenons de reproduire ici les résultats obtenus par cette première étude et vous incitons vivement à consulter l'excellent travail réalisé dans le cadre de l'obtention du Master IS.

Pour rappel, les PID sont généralement découpées en trois parties (Prongué, Raemy, 2017) :

1. Le protocole d'accès (*https* par exemple) et le nom de domaine
2. L'autorité ayant attribué l'identifiant
3. Le numéro de l'identifiant

3.3 Archival Resource Key (ARK)

Cette section fournit les principes et la gouvernance du système ARK, la structure d'un identifiant ARK, les avantages d'un tel schéma, ainsi que son application en Suisse. Si nous avons choisi de nous focaliser sur ce système, c'est parce que de nombreux travaux sur le sujet (Espasandin et al., 2018) ont, en conclusion, préconisé cette solution, ce d'autant plus qu'elle est parfaitement complémentaire au système DOI déjà bien établi dans notre pays.

³ Le panorama des PID peut être consulté et zoom facilement à cette adresse URL : <http://free.iiifhosting.com/iiif/996d2d88ad5b7901c0c770fdb6215a4eeecb360b81b45d19ed045444245910b/> (consulté le 10 janvier 2020)

3.3.1 Principes et gouvernance

ARK est un type, ou schème, d'identifiants décentralisé qui a été créé en 2001 par la California Digital Library (CDL). Cette dernière, avec l'aide d'institutions partenaires, maintient le schème ARK et fournit gratuitement à toute institution intéressée à attribuer des identifiants ARK un numéro unique d'autorité nommante (*Name Assigning Authority Number - NAAN*) (Kunze, Bermès, 2019 ; CDL, 2019a).

Dès qu'une organisation obtient un NAAN, elle est inscrite au registre⁴ et peut réaliser sa propre politique d'attribution, c'est-à-dire établir des règles de nommage et des déclarations de permanence. Elle décide également quelles seront les autorités d'adressage (*Name Mapping Authority - MMA*). Autrement dit, chaque institution attribuant des identifiants ARK prend la responsabilité « *dans la permanence de l'accès à ses ressources* » (BnF, 2019b). Un MMA équivaut généralement à un sous-domaine d'une institution. Cette dernière a la possibilité de déployer autant de MMA, reflétant généralement une structure organisationnelle particulière. A la Bibliothèque nationale de France (BnF), le schème ARK est notamment utilisé pour les documents numériques (gallica.bnf.fr), pour les notices bibliographiques (catalogue.bnf.fr) ainsi que sur data.bnf.fr, leur plateforme LOD (Caron, 2018 ; Caron et al., 2018).

Figure 3 : Premières étapes de création d'un service ARK



Le champ d'application des identifiants ARK est très varié car il peut être assigné à n'importe quel type de données (Ferguson et al., 2018). Selon le registre des numéros d'autorités nommantes (NAAN), plus de 500 organisations utilisent le schème ARK pour identifier leurs ressources. Parmi celles-ci se trouvent des documents numérisés ou nés-numériques, des vocabulaires, des personnages historiques, des données de la recherche, des journaux, etc. Un ARK peut même être attribué à des données qui n'existent pas encore et être donc temporairement lié vers une page web expliquant par exemple le processus de création ou les buts du projet. A ce jour, il a été estimé qu'au moins 3.2 milliards d'ARK ont été attribués (Kunze, Caron, 2019a).

Pour créer ses identifiants ARK, il existe plusieurs solutions (BnF, 2019a) :

- Développer ses propres outils d'attribution et de résolution ;
- Réutiliser des logiciels Open Source ;
- Opter pour une solution intégrant nativement l'attribution d'identifiants ARK ;
- Souscrire à une offre telle que EZID (*easy-eye-dee*) de la CDL (cf. 3.4.2).

Depuis 2018, la CDL et DuraSpace⁵ (un fournisseur de dépôt Open Source) collaborent au projet *ARKs in the Open* (AITO) qui vise à bâtir une communauté internationale ouverte en collaboration avec des bibliothèques, centres d'archives, musées, universités, éditeurs et

⁴ Name Assigning Authority Number (NAAN) Registry : https://n2t.net/e/pub/naan_registry.txt (consulté le 10 janvier 2020). Ce registre est également tenu par la BnF et la *US National Library of Medicine* qui hébergent toutes deux des miroirs.

⁵ DuraSpace : <https://duraspace.org/> (consulté le 10 janvier 2020)

instituts de recherche (Tripp, Kunze, 2019) qui puisse assurer une gouvernance partagée autour du schème ARK, notamment en assurant la maintenance du registre des NAAN, en promouvant le développement et le partage de logiciels Open Source gérant l’attribution et la résolution d’identifiants ou encore en soumettant la spécification ARK à l’*Internet Engineering Task Force* (IETF⁶) afin de la standardiser en tant que *Request for Comments* (RFC) officielle (Kunze, Caron, 2019a).

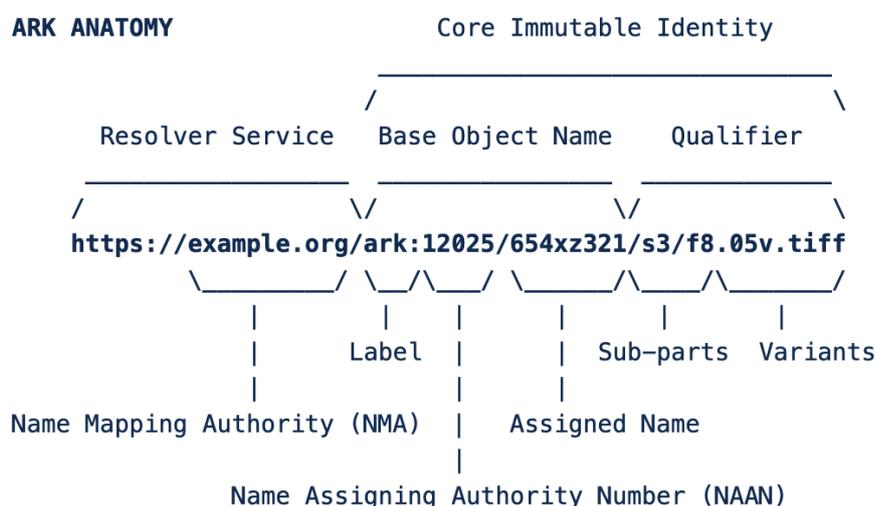
3.3.2 Structure du schème ARK

La structure complète du schème ARK est la suivante (CDL, 2019a) :

[http(s)://NMA/]ark:/NAAN/ARK Name/[Qualifier]

Sur la Figure 4, ci-dessous, se trouve un exemple d’un identifiant ARK découpée en trois parties distinctes. Quant au Tableau 1, il fournit une description plus approfondie de la structure.

Figure 4 : Anatomie d'un identifiant ARK



(Kunze, Caron, 2019a)

Tableau 1 : Description détaillée de la structure ARK

<p>Service de résolution (<i>Resolver Service</i>)</p>	<p>Un identifiant ARK est constitué par une séquence de caractères contenant le label « ark: ». Lorsqu’il est intégré dans une adresse URL, il est précédé du protocole d’accès (http:// ou https://) et de l’autorité d’adressage (NMA). Le tout formant le service de nommage chargé de rendre actionnable les identifiants ARK à la ressource identifiée et aux métadonnées qui lui sont liées. Cette autorité d’adressage est modifiable et remplaçable car ni le serveur web ni les protocoles actuels ne sont supposés durer plus longtemps que les objets identifiés.</p>
<p>Nom de base de l’objet (<i>Base Object Name</i>)</p>	<p>L’identifiant immuable et globalement unique suit le label « ark: ». Cela comprend un numéro d’autorité nommante (NAAN) suivi du nom attribué à l’objet (<i>assigned name</i>). Le nom de base peut parfois être subdivisé en trois sous-composantes : un préfixe (<i>shoulder</i>) permettant de faciliter la délégation de noms selon différents critères (typologie de l’objet, persistance, etc.), un numéro alphanumérique ainsi qu’un caractère de contrôle pour éviter toutes erreurs de transcription.</p>

⁶ Internet Engineering Task Force : <https://ietf.org/> (consulté le 10 janvier 2020)

Qualificatifs (<i>Qualifiers</i>)	Les identifiants ARK peuvent également assurer une granularité très fine grâce à l'utilisation de qualificatifs (<i>qualifier</i>) permettant par exemple d'identifier une ressource contenue dans une autre (<i>sub-parts</i>), comme un folio d'un manuscrit, un chapitre d'un document ou encore un sous-fonds d'archives, ou une variante distincte (<i>variants</i>) d'une ressource telle que le format de fichiers ou le versionnage.
---	--

(Kunze, Bermès, 2019 ; CDL, 2019a ; 2019b ; BnF, 2019b ; Kunze, Caron, 2019a ; 2019b)

L'identité immuable de base (*Core Immutable Identity*) d'un ARK, autrement dit l'identifiant ARK principal, comprend le nom de base de l'objet et parfois un ou des qualificatifs (Kunze, Caron, 2019b). Lorsqu'un identifiant ARK est sous la forme d'une URL (comme sur la Figure 4), on parle d'un « ARK actionnable ». Cette URL peut être présentée sous différentes formes, soit celle hébergée sur le site de l'institution ou du fournisseur, autrement dit la page de destination, soit celle qui pourra la rediriger vers la page de destination à l'aide d'un résolveur local ou global tel que Name-To-Thing (N2T) (CDL, 2019a ; Wimalaratne, Fenner, 2018).

Il faut également relever que la spécification ARK recommande l'opacité dans la structure des identifiants afin de survivre à un changement de libellé de l'organisation et aux évolutions de la ressource nommée (BnF, 2019a).

3.3.3 Métadonnées associées et déclaration de permanence

Vu que les ARK ont été créés pour identifier n'importe quel type d'objets, ils sont ordinairement accompagnés⁷ d'un jeu de métadonnées composé d'éléments très génériques.

Depuis 2001, le schéma de métadonnées régulièrement utilisé est le Dublin Kernel (DK) qui est un sous-ensemble inspiré des éléments du Dublin Core (DC). Il a été conçu pour faciliter la gestion ordonnée des collections en permettant la création de descriptions brèves mais extrêmement uniformes d'objets. DK a été structuré pour répondre notamment aux questions suivantes (Kunze, Turner, 2010) :

- **Who?** *An entity responsible for expressing the object, such as creating it or making it available. Examples of "who" include a person, an organization, or a service.*
- **What?** *A human-oriented name given to the resource, or what this "telling" of the resource was called.*
- **When?** *A point or period of time associated with an event in the lifecycle of the resource, often when it was expressed, created or made available.*
- **Where?** *An access-oriented name given to the resource, or where this resource was expressed. is to identify the resource by means of a string or number conforming to a formal identification system.*
- **How?** *Reserved for a coded value indicating how the object was expressed.*
- **Huh?** *Reserved to indicate the character set encoding and language of the metadata record.*

On parle également de Electronic Resource Citation (ERC) lorsqu'un noyau est composé des quatre premières questions (Kunze, Turner, 2010). La conception de ce « noyau fondamental » reconnaît que la principale contribution du DC est le consensus international et

⁷ Lorsqu'on parle de « métadonnées associées », il s'agit de celles de l'identifiant et non des métadonnées descriptives ou techniques de l'objet.

interdisciplinaire qui a permis d'identifier une série d'éléments, sans se soucier de leur labellisation (Kunze, Bermès, 2019).

Selon l'implémentation de la spécification ARK, une description sommaire de création de l'identifiant en ERC peut être obtenu par la fonctionnalité d'inflexion qui consiste à ajouter un point d'interrogation « ? » à la fin d'un ARK actionnable. A cela s'ajoute la déclaration de permanence expliquant les conditions de permanence de l'objet. Ces conditions peuvent être obtenues par la fonctionnalité de double inflexion, autrement dit en ajoutant deux points d'interrogation « ?? » à la fin d'un ARK actionnable. Dans cette déclaration de permanence de l'objet se trouve normalement un lien vers la politique de permanence de l'institution.

Pour illustrer ces propos, le Tableau 2 synthétise le cas d'un objet (*A Study of Rhythm in Bach's Orgelbüchlein*) détenu et numérisé par la University of North Texas.

Tableau 2 : Objet ARK, métadonnées associées et déclaration de permanence

ARK actionnable	Description
https://n2t.net/ark:/67531/metadc107835/	Objet possédant un identifiant ARK
https://n2t.net/ark:/67531/metadc107835/?	Description sommaire de l'objet en ERC
https://n2t.net/ark:/67531/metadc107835/??	Description et déclaration de permanence en ERC (cf. Figure 5)
https://digital.library.unt.edu/ark:/67531/	Politique de permanence ARK de l'institution

La description et la déclaration de permanence sont sérialisées avec la syntaxe d'enregistrement A Name-Value Language (ANVL), construite de la même manière que les en-têtes d'email ou ceux du protocole HTTP (Kunze et al., 2006).

Figure 5 : Description et déclaration de permanence (ERC)

```

erc:
  who: Austin, Larry
  what: A Study of Rhythm in Bach's Orgelbüchlein
  when: 1952
  where: https://digital.library.unt.edu/ark:/67531/metadc107835/
erc-support:
  who: University of North Texas Libraries
  what: Permanent: Stable Content:
  when: 20081203
  where: https://digital.library.unt.edu/ark:/67531/

```

Description de l'objet

Déclaration de permanence

La spécification ARK recommande aux institutions qui assignent des identifiants de rédiger une politique de permanence comprenant des informations au sujet des trois points suivants : la réaffectation ou non des identifiants, l'opacité de ceux-ci ainsi que l'intégration d'un caractère de contrôle au sein du nom ARK. Sur la Figure 6 ci-après se trouve la politique de permanence des identifiants ARK de la BnF⁸.

⁸ Cette politique peut se trouver en ajoutant « .policy » (la BnF traite cette information comme un qualificatif de service) à la fin de chaque ARK se trouvant sur leur bibliothèque numérique Gallica : <https://gallica.bnf.fr/ark:/12148/bpt6k107371t.policy> (consulté le 10 janvier 2020)

Figure 6 : Politique de permanence ARK de la BnF

La BnF attribue des identifiants dans le domaine ARK 12148 (Bibliothèque nationale de France) selon les principes suivants :

Aucun identifiant ARK ne sera attribué à nouveau ; c'est-à-dire qu'une fois qu'un lien entre un identifiant ARK et un objet aura été publié, ce lien devra être considéré comme unique, et ce pour une durée indéfinie.
 Les identifiants ARK attribués par la BnF ne contiennent, dans la mesure du possible, aucune information sémantique facilement reconnaissable ; ceci contribue à faciliter leur utilisation indépendamment d'un contexte de temps ou de lieu.
 Les identifiants ARK attribués par la BnF contiennent un caractère de contrôle qui les garantit contre des erreurs de caractère isolées et des erreurs de transposition.

La BnF garantit la pérennité des identifiants ARK attribués, mais l'étendue précise des garanties est en cours de définition. En effet, la gestion d'un objet numérique pourra exiger des modifications appropriées de cet objet pour en assurer la persistance. Par ailleurs, le niveau déclaré des garanties pourra évoluer au fur et à mesure que les normes et méthodes de pérennisation seront mieux maîtrisées, et que la BnF mettra en application des procédures pour maintenir ces garanties et les associer aux objets qu'elle gère.
 Certaines des collections numériques de la BnF dupliquent des éléments de collections d'autres institutions.

(Caron, 2018)

Il est important de signaler que la spécification ARK n'exige aucun schéma particulier de métadonnées et que chaque institution peut déployer le standard de son choix, utiliser le noyau ERC pour ensuite l'étendre si nécessaire, ou encore décider ne pas associer de métadonnées aux identifiants (Kunze, Caron, 2019a ; Kunze, Bermès, 2019).

3.3.4 Avantages

Outre la grande flexibilité et le niveau de granularité élevé qu'offrent les identifiants ARK, voici quelques exemples des avantages financiers, organisationnels et techniques du déploiement d'un système ARK.

Tableau 3 : Avantages financiers, organisationnels et techniques du schème ARK

Types d'avantages	Description des avantages du schème ARK
Avantages financiers	<ul style="list-style-type: none"> • Obtenir un numéro d'autorité nommante gratuitement. • Pouvoir bénéficier de logiciels Open Source et gratuits existants. • Maintenir les charges d'attribution à un coût peu élevé.
Avantages organisationnels	<ul style="list-style-type: none"> • Être capable de gérer le problème de la perte ou de la division du nom de domaine sans perdre le contrôle des identifiants. • Pouvoir changer de prestataire de service ou d'infrastructure sans avoir à coordonner les transferts de bases de données. • Utiliser une infrastructure conforme aux valeurs d'une organisation donnée.
Avantages techniques	<ul style="list-style-type: none"> • Lier n'importe quel schéma de métadonnées aux objets identifiés. • Créer un identifiant sans métadonnées. • Créer un identifiant avant même que l'objet n'existe. • Créer un identifiant temporaire. • Créer un identifiant permettant d'activer autant de liens que possible grâce à l'utilisation du <i>suffix passthrough</i>. • Lier l'objet identifié directement vers son contenu plutôt que sur une page de destination. • Assigner un PID supplémentaire, comme un DOI ou un Handle, à n'importe quel identifiant ARK. • Accéder aux métadonnées et aux déclarations de permanence de l'objet par les fonctionnalités d'inflexion.

	<ul style="list-style-type: none"> • Migrer aisément les identifiants ARK vers une autre infrastructure lors d'une modification de gouvernance. • Pouvoir combiner ARK avec d'autres spécifications techniques telles que les interfaces de programmation applicative (API) conçues par la communauté <i>International Image Interoperability Framework</i> (IIIF – se prononce « <i>triple eye eff</i> ») pour faciliter la dissémination, la citabilité et le partage d'images et d'objets audiovisuels.
--	--

(CDL, 2007 ; Snyderman et al., 2015 ; BnF, 2019b ; Schneider, Raemy, 2019b ; Raemy, Schneider, 2019b ; Kunze, Caron, 2019a ; 2019b)

3.3.5 ARK en Suisse

En consultant le registre des NAAN, nous constatons que dix organisations suisses (ou installées en Suisse) y sont inscrites⁹ :

Tableau 4 : Organisations établies en Suisse inscrites au registre des NAAN

Organisation(s)	Inscription
World Intellectual Property Organization (WIPO)	12.07.2002
Data and Service Center for the Humanities (DaSCH)	16.10.2017
Office des archives de l'État de Neuchâtel, Bibliothèque de la Ville de la Chaux-de-Fonds, Bibliothèque publique et universitaire de Neuchâtel	26.11.2018
Docuteam	05.12.2018
Musée d'histoire des sciences de la Ville de Genève, Archives de la Ville de Genève, Bibliothèque de Genève, Musée Voltaire	16.09.2019

(CDL, 2019d)

A priori, toutes ces institutions attribuent des identifiants ARK à l'interne et n'offrent pas de service de prestations externes. Pour l'instant, seul le DaSCH (cf. 3.4.2) a déjà assigné des identifiants ARK de manière publique.

La plupart des autres organisations qui se sont inscrites au registre attribueront des ARK dans un futur proche, comme par exemple Docuteam qui attend d'intégrer une telle fonctionnalité dans le contexte d'une migration vers une nouvelle version du système qu'ils utilisent¹⁰ ou encore les quatre institutions genevoises qui se sont inscrites en septembre 2019 qui vont le faire dans le cadre d'un projet commun de diffusion des fonds d'archives et de documents numérisés.

⁹ Une carte Google intitulée « Global ARK Registrants » est disponible à l'adresse suivante : <https://www.google.com/maps/d/viewer?mid=1ALGeRERECL36f2pg7pqrthUYNmuU43UM&usp=sharing> (consulté le 10 janvier 2020)

¹⁰ Correspondance électronique du 14 novembre 2019 avec un collaborateur de Docuteam

3.4 Implémentation d'un système d'attribution d'identifiants ARK

3.4.1 Processus d'implémentation générique de PID

Le coût d'un processus d'implémentation peut être décomposé en fonction des différents outils (ou chaînes) qui le composent. Ainsi, il est possible d'estimer quel investissement est nécessaire pour chaque outil : gestion de contenu, hébergement, surveillance et redirection. Sans oublier qu'il est également possible de gérer soi-même, en tant qu'institution, l'un ou l'autre de ces outils et donc de réduire les coûts de fonctionnement. Deux des avantages des ARK, comme des URL d'ailleurs, est que l'institution qui en achète ne sera pas facturée séparément pour les identifiants. De plus, ils ne fonctionnent pas en silo, ce qui permet une interopérabilité avec les autres identifiants (Kunze, Caron, 2019a).

Évidemment, avant d'entamer la rédaction d'une politique d'attribution d'identifiants et une politique de permanence, plusieurs questions essentielles se posent, notamment en ce qui concerne les buts et le choix des technologies à adopter, de manière à cerner immédiatement le périmètre de travail et les moyens à mettre en œuvre (Nicholas et al., 2009 ; Simons, 2012). Ces questions, nous ne manquerons pas de nous les poser et reviendrons dessus dans notre travail.

Les contraintes à la création d'un Hub pouvant attribuer des ARK sont évidemment nombreuses : qu'elles soient techniques, légales, technologiques, ou financières, il est indispensable de les identifier très en amont afin de prévenir les problèmes, d'éviter les mauvaises surprises et de gérer au mieux les finances allouées.

3.4.2 Exemples d'implémentation d'un système ARK

ARK étant un système décentralisé, ses applications sont très variées. Dans le tableau ci-dessous se trouvent trois exemples d'implémentation du système ARK ainsi que la solution EZID proposée par la CDL.

Tableau 5 : ARK - exemples d'implémentation

Institution Prestataire	Plateforme(s)	Spécificités
Université de Bâle - DaSCH	https://www.salsah.org/ http://dasch.swiss	Les ARK sont utilisés pour identifier des objets numérisés au sein d'une plateforme LOD utilisant l'ontologie Knora développée par le laboratoire des humanités numériques de l'Université de Bâle (Raemy et al., 2017 ; DHLab Basel, 2019). Chaque objet reçoit un ARK et les qualificatifs sont déclinés de la sorte : un identifiant alphanumérique pour identifier une page ainsi qu'un horodatage (<i>timestamp</i>) pour permettre aux chercheurs de citer un objet dans un contexte donné. Exemple : http://ark.dasch.swiss/ark:/72163/0803-31fb126ca48-e.201992
BnF - Gallica	https://gallica.bnf.fr/	Gallica a utilisé les qualificatifs ARK pour identifier les numéros de pages ou folios, les différents formats ainsi que les vignettes. Exemples : https://gallica.bnf.fr/ark:/12148/bpt6k107371t/f24.image https://gallica.bnf.fr/ark:/12148/bpt6k107371t/f24.thumbnail https://gallica.bnf.fr/ark:/12148/bpt6k107371t/f24.image.jpeg

Institution Prestataire	Plateforme(s)	Spécificités
Durham University Library	https://library.dur.ac.uk	Durham University Library utilise les identifiants ARK pour chaque objet numérique compatible avec IIIF (manifeste) et un qualificatif pour chaque image (canevas) contenue dans cet objet (Durham University Library, 2017). Exemple : https://n2t.durham.ac.uk/ark:/32150/t1mcz30ps64z/t1t4b29b613b.html (en enlevant le « .html » à la fin de l'URL, nous obtenons des informations sérialisées en JSON-LD).
05.02.2020 11:03:00C DL - EZID	https://ezid.cdlib.org/	EZID est un service de prestation <i>out-of-the-box</i> payant permettant de créer des ARK (ainsi que des DOI) à partir du profil de métadonnées ERC. Il est également possible de créer des ARK à partir d'un autre schéma de métadonnées comme le DC qualifié (CDL, 2019c). La tarification varie de \$500 (département) à \$1'500 (institution) par année. Ce tarif comprend la création et la maintenance d'un million d'identifiants.

3.4.3 Implémentation d'un Hub en Suisse

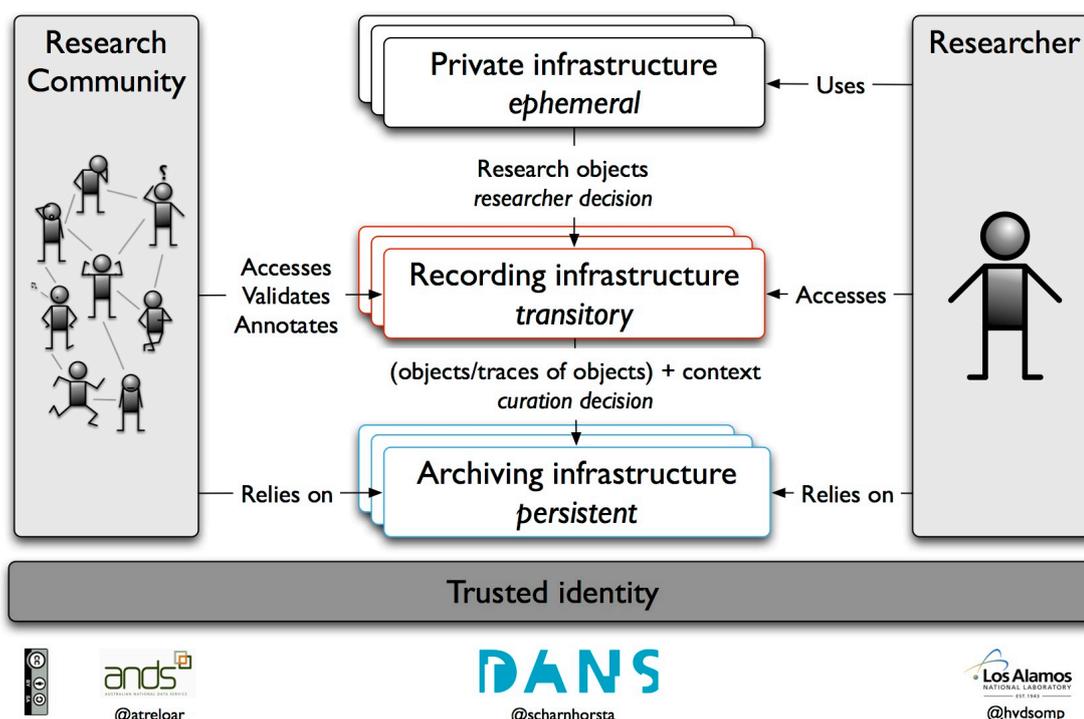
Le Projet INCIPIT (192-001) initialisera l'implémentation d'un Hub pouvant attribuer des identifiants ARK en Suisse (swissuniversities, 2019b), en complémentarité des DOI que les institutions helvétiques peuvent obtenir depuis 2009 par le biais du DOI-Desk de la Bibliothèque de l'École Polytechnique Fédérale de Zürich (EPFZ) (cf. 6.1.2).

Le but de ce Hub est d'offrir aux institutions suisses des ARK leur permettant autant un haut niveau de granularité dans la description de leurs jeux de données que d'encourager leurs chercheurs à établir une « identité de confiance » (cf. Figure 7 ci-après) tout au long du processus de recherche (Schneider, Raemy, 2019a).

L'identification des objets doit tenir compte des nouvelles frontières de la recherche qui sont encore mouvantes et imprécises (Lavoie et al., 2014) et qui incluent maintenant à la fois les composantes matérielles (données, publications de tous types de formats, etc.), les différentes parties prenantes (chercheurs, institutions, bibliothécaires, etc.), mais aussi toutes les étapes du processus de recherche, en partant du stockage éphémère jusqu'à l'assurance d'une conservation à long terme (étapes, résultats, révisions, archives, etc.).

Pour garantir cette « confiance », les PID ont un rôle primordial à jouer (Espasandin et al., 2018, p. 8-9) et ce dès le début du partage des données des chercheurs à leurs pairs, autrement dit bien souvent avant la phase de publication ou d'archivage.

Figure 7 : Recording to Archiving Architecture (identité de confiance)



(Treloar et al., 2015)

La mise en place de ce Hub serait un bénéfice pour toute la communauté patrimoniale et scientifique de Suisse et permettrait de créer une véritable unité dans un pays possédant quatre langues nationales et vingt-six cantons, et dont le fédéralisme incite à une certaine décentralisation des prises de décision.

Par ailleurs, ce projet s'inscrit parfaitement dans la stratégie pour l'Open Science menée par P5 incitant les hautes écoles à créer les conditions adéquates pour réutiliser, propager et reproduire des données et méthodes de recherche par voie électronique. Cette approche est bien entendue basée sur les principes FAIR.

Dès lors, une uniformisation des pratiques, en plus des arguments financiers (rationalisation des coûts grâce à la centralisation), organisationnels et techniques habituels, permettrait de confirmer et développer encore plus la position de tête de la Suisse dans le domaine de la recherche et de l'enseignement, favorisant ainsi sa renommée de pays parmi les plus innovants au monde, et, d'une certaine manière, « montrant la voie » à titre d'exemple. D'ailleurs, la devise suisse se prête particulièrement à ce projet : « Un (PID) pour tous, et tous pour un (PID) » !

Ainsi les objectifs de ce Hub de gestion des PID s'inscrivent dans une logique à la fois politique et d'amélioration des conditions de la recherche. Ce Hub a pour vocation de devenir le centre d'expertise sur le sujet, une vitrine pour la recherche mettant en avant la réutilisation et la préservation des connaissances, au sein des institutions suisses évidemment, mais à l'international également. Financièrement, une telle infrastructure permettra de réduire les coûts de gestion des PID, tout en proposant des solutions « sur mesure » à une grande variété d'institutions. De plus, cela permettra aussi d'économiser du temps et des ressources.

L'uniformisation de la technologie employée pour identifier des ressources offrira également des avantages non négligeables, comme par exemple la possibilité, grâce à un niveau de granularité élevé de l'identification des ressources, d'extraire bien plus facilement des résultats pertinents en vue d'effectuer des statistiques précises. En exploitant bien l'anatomie de l'identifiant ARK, il sera possible de retrouver des ressources au moyen d'informations n'étant pas forcément renseignées dans les métadonnées associées. Cette granularité favorisera également l'assurance de la provenance, voire de la migration, de la ressource si la gouvernance d'une institution cliente du Hub venait à être modifiée (changement de nom de l'organisme, transfert d'un projet académique d'une entité à une autre, etc.).

Du point de vue client, un changement de prestataire « [...] *sans avoir à coordonner des transferts de bases de données avec une autorité centrale* » (Kunze, Caron, 2019b) est dès lors facilité grâce à la structure ARK et au résolveur global N2T qui arrivent à gérer la scission d'espace de noms.

De plus, nous préconisons vivement la participation à l'initiative mondiale AITO qui, au travers de plusieurs groupes d'intérêt, développe et améliore les possibilités des ARK. Suivre attentivement ce développement nous semble indispensable, car il est certain que la spécification et le mode de gouvernance du registre des NAAN seront amenés à évoluer dans le courant des prochaines années, et qu'une participation active à cette initiative ne pourra qu'être bénéfique pour le déploiement et la reconnaissance du Hub au fil du temps.

Les avantages de ce Hub sont nombreux et variés, mais cette centralisation comporte aussi quelques risques, à commencer par le fait que la gestion des PID n'est pas si simple, demande de véritables compétences, et surtout un engagement sur le long terme. Une harmonisation des pratiques ne va pas sans idées reçues, et les principales ont été relevées par John Kunze (2018)¹¹.

¹¹ Cf. Annexe 1 : *Ten persistent myths about persistent identifiers*

4. Méthodologie

Ce projet de recherche, et donc la méthodologie inhérente à celui-ci, s'inscrit dans la continuité du projet ICOPAD. Dans ce chapitre sont présentées la démarche méthodologique globale dans son contexte académique, la récolte d'information ainsi que l'analyse des données qui ont pu être synthétisées sous forme d'un Blueprint de service, de critères essentiels et de recommandations pour l'établissement d'une infrastructure ARK en Suisse.

4.1 Démarche globale

Dans cette section sont présentés les différents livrables qui étaient à rendre en parallèle au projet de recherche qui s'est déroulé entre les deuxième et troisième semestres académiques du Master IS à la HEG-GE, autrement dit de février 2019 à janvier 2020.

4.1.1 Cahier des charges (04/2019)

Première étape indispensable, nous avons commencé par créer le cahier des charges définissant la nature du mandat, ses contraintes, ses objectifs (généraux et spécifiques), la méthodologie de travail ainsi qu'un calendrier.

Si l'on se reporte à la version de notre cahier des charges validée en avril 2019 et que nous la comparons au travail que vous tenez entre vos mains, vous remarquerez certaines différences. En effet, pour diverses raisons, comme l'adaptation à une réalité changeante et l'évolution rapide des recommandations émanant d'organismes officiels (tels que swissuniversities), nous avons choisi d'ajuster nos objectifs afin de réaliser un travail assurément utile et réaliste pour la communauté. Dès lors, voici une liste des modifications que nous y avons apporté.

Objectif 1 : Conceptualisation

Si l'objectif général lui-même reste inchangé (« Créer un cadre conceptuel autour d'une solution helvétique, complémentaire aux DOI, permettant d'attribuer des ARK à tous les types de données »), nous avons adapté les objectifs spécifiques en basculant les questions de modélisation et de choix d'études de cas à analyser dans le second objectif général, justement intitulé « Prototypage ».

Objectif 2 : Prototypage

Comme dit ci-dessus, nous avons rassemblé sous cet objectif tout ce que touchait à la modélisation du Hub, bien que nous ayons laissé de côté l'objectif secondaire de création d'une maquette du front-end et d'une RESTful API, objectif qui nous a paru un peu prématuré dans le cadre de ce projet. Nous nous contenterons, à ce sujet, d'émettre des recommandations en matière de critères indispensables.

Les conseils financiers et techniques initialement prévus ici ont été basculés assez logiquement dans l'objectif suivant, renommé « Recommandations ».

Objectif 3 : Recommandations

Initialement intitulé « Négociation et déploiement », ce dernier objectif est celui qui a connu le plus de changements. Nous espérons réussir dans le temps imparti à déployer une solution avec des partenaires clefs, et ainsi pouvoir tester la solution et l'améliorer au besoin. Malheureusement, l'implémentation d'une solution ne pouvait se faire qu'avec l'appui de

certaines institutions, dont les politiques en la matière ne sont pas encore bien définies. Étant largement tributaires de décisions prises à un niveau politique bien plus élevé, nous avons donc adapté cet objectif pour qu'il soit réalisable. Dès lors, il sera passablement tourné vers des recommandations financières et techniques, appuyées par des retours de professionnels sur nos études de cas. Notre travail se terminera par l'établissement d'une liste de critères indispensables et généraux nécessaires à la création d'un Hub national.

4.1.2 Data Management Plan (05/2019)

Afin de planifier au mieux le cycle de vie de nos données, et évidemment de documenter précisément leurs types, les méthodes de collecte, de gestion et d'analyse de celles-ci, un *data management plan* (DMP) a été produit. Le but de ce DMP est bien sûr de faciliter leur recherche, leur accès, leur réutilisation et leur citation, et ceci afin de se conformer aux principes FAIR.

4.1.3 Poster scientifique (12/2019)

La réalisation d'un poster scientifique (cf. Figure 19 en Annexe 2) sur le sujet de notre travail de recherche était également prévue dans notre cursus, et nous avons donc mis en forme, et de façon plutôt esthétique, une sélection des informations trouvées ou produites dans le cadre de nos recherches. Une présentation de ce poster face à des publics professionnels (ou non) nous a permis d'affiner nos arguments en prenant en compte les retours, critiques et débats qui suivaient l'explication de notre projet. Cette session fut une excellente occasion de promouvoir notre travail, tout en élargissant notre cercle de connaissances professionnelles.

Rapidement, nous avons choisi de symboliser le futur Hub helvétique par une montagne dont le sommet est recouvert de neige. Un téléphérique en fait l'ascension, gravissant les différentes étapes du processus de création d'identifiants ARK, jusqu'à arriver au sommet où l'on atteint le niveau de service le plus élevé ainsi que l'assurance que la référence à la ressource est pérenne et précise au possible, à l'image de la persistance et de la beauté limpide des neiges éternelles.

Tout autour, nous retrouvons diverses informations, telles que les objectifs du projet ainsi que du Hub lui-même, l'anatomie d'un identifiant ARK et les métadonnées ERC l'accompagnant, le processus de création et les outils nécessaires à sa réalisation. Nous tenions également à disposition des intéressés des copies du Blueprint de service que nous avons réalisé, celui-ci reflétant plus largement le travail de réflexion que nous effectuons pour ce projet (cf. 6.4).

4.1.4 Billets de blog axé données (06/2019) et axé projet (12/2019)

Toujours dans le cadre de notre cursus, nous avons écrit et publié deux billets de blog sur le sujet de notre travail. Le premier était axé sur les données de notre projet (Cevey, Raemy, 2019b), le second présentait quelques outils, services et fonctionnalités du Hub, ainsi que sa modélisation sous forme de Blueprint de service (Cevey, Raemy, 2020). Ces deux billets ont été publiés sur le blog du Master IS, « Recherche d'IdéeS »¹².

4.2 Recensement des informations

La recherche d'information a été décisive au sein de ce projet de recherche, car cela a permis à la fois de valider ou de consolider les résultats obtenus par le projet ICOPAD et de faire

¹² Recherche d'IdéeS : <https://campus.hesge.ch/blog-master-is/> (consulté le 10 janvier 2020)

évoluer les caractéristiques requises que devrait comporter un Hub national attribuant des identifiants ARK.

4.2.1 Résultats du projet ICOPAD

Tous les résultats importants du projet ICOPAD sont disponibles en ligne et ont été en partie réalisés par un des auteurs de ce mémoire de recherche. En fonction de ces résultats, cinq scénarios, une matrice de décision ainsi qu'un diagramme de séquence ont été imaginés. Ces trois livrables peuvent être considérés comme faisant partie d'un modèle informationnel qui pourrait être mis en œuvre au sein d'un Hub national. C'est sur la base de ce modèle, que le processus d'attribution d'identifiants ARK et certaines considérations techniques, telles que les fonctionnalités principales du service, ont été conçus.

Huit études de cas issues des trois institutions partenaires du projet ICOPAD ont également été retenues au sein de ce mémoire de recherche.

4.2.2 Entretiens

Effectuer des entretiens, par téléphone, logiciel vidéo ou en présentiel, nous a semblé l'un des meilleurs moyens d'obtenir des informations sur les processus mis en place ailleurs dans le monde et la réalité du terrain. Ces entretiens, semi-dirigés et sur la base d'une grille de questions, nous ont permis en premier lieu d'identifier les personnes clefs de notre domaine de recherche. Dès lors, nous avons adapté chaque grille d'entretien en fonction du type d'institution à laquelle appartenait la personne ressource, en ciblant nos questions en fonction de l'existant en matière de gestion des PID et/ou de leurs besoins.

En plus de nous permettre d'établir un état des lieux fiable, ces entretiens ont permis de récolter des avis et opinions de professionnels sur les différentes manières de traiter et gérer des PID, autant d'un point de vue technique qu'organisationnel, mais également d'identifier de nouveaux besoins et d'affiner nos recommandations.

4.2.3 Revue de la littérature

La revue de la littérature effectuée (cf. 3) s'est faite en premier lieu sur la base des recherches menées dans le cadre d'un précédent projet de recherche de Master (Espasandin et al., 2018) et du projet ICOPAD (Prongué, Raemy, 2017 ; Raemy, 2018 ; Raemy, Schneider, 2019a ; Schneider, Raemy, 2019a). Pour compléter ces informations, nous nous sommes tournés vers d'autres sources, telles que les présentations réalisées lors de la conférence PIDapalooza¹³, « *the open festival for persistent identifiers* » (Meadows, 2017), organisée annuellement depuis 2016 par CrossRef, DataCite, Open Researcher and Contributor ID (ORCID) et la CDL. Les différents guides, livres blancs et articles stratégiques et techniques sur l'implémentation de PID ont également été des sources d'information très utiles pour nous rendre compte de l'existant en la matière.

Bien entendu, les nombreux retours d'expériences d'institutions, de services ou de projets attribuant des PID se sont révélés fort intéressants, comme ceux réalisés lors du Sommet ARK du 21 mars 2018 à la BnF (La Tribune des Archivistes, 2018) et ceux issus des diverses activités du projet AITO initié par DuraSpace et la CDL.

¹³ PIDapalooza : <https://pidapalooza.org/> (consulté le 10 janvier 2020)

4.2.4 Échanges avec l'initiative AITO

ARKs In The Open (AITO) est une initiative ayant généré une certaine quantité de littérature de qualité sur le sujet des ARK, à commencer par une FAQ (traduite en français). Nous y avons trouvé de nombreuses réponses à nos interrogations lors de la phase de rédaction de la revue de la littérature. A noter qu'il est possible d'y apporter des modifications dans une version commentée du document, ce que nous avons fait à propos de quelques éléments, dont l'avantage offert par la structure ARK en matière d'accueil de spécifications telles que les API développées par IIRF.

Sur le site web du projet AITO, nous avons également eu la possibilité de participer à une étude ayant pour objectif d'améliorer l'évaluation des intérêts de la communauté ARK et d'orienter la constitution et les priorités de l'organisation (Kunze, 2019c)¹⁴.

Le projet AITO se compose de trois groupes de travail (*Outreach Working Group*, *Sustainability Working Group*, *Technical Working Group*) et d'un groupe consultatif (*Advisory Group*). Il y aurait une possibilité de poursuivre les discussions via un ou plusieurs de ces groupes étant donné que le projet INCIPIT sera probablement financé par le fonds P5, ce qui permettrait aux responsables du Hub de se tenir au courant et d'adapter les fonctionnalités proposées si, par exemple, la spécification ARK est modifiée ou si de nouveaux outils sont développés.

4.3 Analyse des données et modélisation

A la suite et en parallèle au recensement des informations, une phase d'analyse et de modélisation a été effectuée.

4.3.1 Études de cas

Pour faciliter la modélisation technique, notamment le *mapping* des métadonnées, les liens vers des notices d'autorité ainsi que les déclarations de permanence, nous avons sélectionné une vingtaine d'études de cas (cf. 5) représentant une variété de types d'objets qui ont déjà ou qui pourraient avoir besoin d'identifiants ARK.

4.3.2 Prototypage

Sur la base des informations issues de notre revue de la littérature et des études de cas analysées, nous avons pu commencer à prototyper un modèle de Hub répondant aux besoins et contraintes que nous avons identifiés. Les considérations et propositions concernant la charge humaine de travail se sont faites généralement sur la base des entretiens effectués, lors desquels nous avons pu rassembler quantité d'informations issues d'expériences vécues et de nombreux conseils de professionnels du domaine.

En croisant toutes ces données avec les moyens techniques nécessaires à la mise en place d'un Hub destiné à la gestion des PID, et donc en retenant la solution ARK, il a été possible de créer un Blueprint de service (cf. 6.4) permettant de visualiser graphiquement les relations entre les différents composants et processus du service. C'était également un moyen d'éviter des oublis, de ne pas passer à côté d'évidences et de pouvoir retracer, pas à pas, le parcours utilisateur d'une institution désirant utiliser les services du Hub. Dès lors, nous avons pu affiner nos réponses aux questions des ressources, financières et humaines.

¹⁴ Les résultats de ce sondage sont disponibles à la page web suivante : <https://wiki.lyrasis.org/display/ARKs/2019-11-14+Outreach+WG+Agenda+and+Notes> (consulté le 10 janvier 2020)

Le prototypage du Hub nous aura permis de nous rendre compte des scénarios d'utilisation typiques du service proposé, tout en estimant le volume des données à traiter et les besoins d'interactions avec le « client ». Bien entendu, notre modèle, bien que basé sur des réalités et des estimations aussi précises que possible, reste un modèle théorique propre à être affiné par de nouvelles itérations et adaptations aux différentes évolutions, techniques, financières et humaines du domaine auquel il appartient.

4.3.3 Recommandations pour le projet INCIPIT

A la suite de l'étape de modélisation, des recommandations ont été listées (cf. 7) à l'adresse des parties prenantes du projet INCIPIT qui démarrera l'initialisation du Hub. Ces recommandations ont été effectuées premièrement sous la forme de récits utilisateurs, qui sont des phrases courtes permettant de cerner au mieux les besoins des clients du Hub, ainsi que sous la forme de listes à points détaillant les fonctionnalités techniques requises de l'infrastructure.

Les parties suivantes concernent la viabilité du service sur le long terme, les facteurs clefs de succès nécessaires aux personnes en charge de ce service ainsi que les critères indispensables pour rester innovant.

5. Études de cas

Ce chapitre sur les études de cas qui ont déjà ou qui auraient besoin d'un identifiant ARK a été divisée en deux sections : sélection (cf. 5.1) et analyse (cf. 5.2).

5.1 Sélection

Les études de cas qui ont été sélectionnées proviennent du projet ICOPAD, de la FAQ ARK du projet AITO ainsi que d'ArODES, l'archive institutionnelle des domaines de la HES-SO. Nous nous sommes également penchés sur le cas du Swiss Institute of Bioinformatics (SIB), sans pour autant retenir une étude de cas précise, mais en analysant les différents types d'objets ayant recours aux ARK.

Ces choix ont été faits sur la base de la revue de la littérature, qui nous a permis d'identifier rapidement quelques cas intéressants, sur les conseils du Prof. Schneider, spécialiste du milieu, et sur les informations récoltées lors des entretiens menés avec des professionnels et d'éventuelles institutions intéressées par utiliser les futurs services du Hub. Les études de cas menée dans le cadre du projet ICOPAD nous ont également été très utiles.

5.1.1 Études de cas du projet ICOPAD

Ces études de cas sont issues des interactions avec les trois partenaires du projet : la Zentralbibliothek de Zürich (ZB), l'Institut suisse pour l'étude de l'art (SIK-ISEA) et la Zürcher Hochschule der Künste (ZHdK). Parmi toutes les études de cas répertoriées dans ce projet, nous avons effectué une sélection précise répondant spécifiquement au contexte de nos recherches et retenu huit scénarios (cf. Tableau 6) couvrant tous les cas de figures imaginés.

5.1.2 Études de cas de la FAQ ARK

En étudiant attentivement la foire aux questions réalisée dans le cadre de l'initiative AITO pour démontrer le potentiel des identifiants ARK (Kunze, Caron 2019), nous avons réalisé que nombre d'interrogations répertoriées ici fournissaient d'excellentes études de cas, mais nous avons réduit notre sélection uniquement à celles qui présentaient un identifiant ARK, soit directement dans l'URL de la page de destination, soit sous sa forme N2T trouvée dans les métadonnées.

5.1.3 Études de cas d'ArODES

L'archive institutionnelle des domaines de la HES-SO, qui souhaite entre autres garantir un accès stable et un archivage pérenne aux travaux de recherche, fut un vivier d'études de cas fort intéressantes, notamment en raison de la grande variété des types de données issues de leurs sept collections¹⁵ disponibles sur la plateforme institutionnelle. ArODES est également un potentiel partenaire de choix pour la création du Hub.

Néanmoins, durant la réalisation de ce projet de recherche, nous avons appris que la HES-SO avait acheté cinquante DOI courant 2019, DOI qui ont tous été attribués depuis, mais dont les ressources qu'ils identifient ne sont pas encore visibles sur ArODES. S'il est prévu d'approfondir la réflexion en la matière en 2020, toute pratique sera évidemment établie sur la base de discussions avec les bibliothécaires des hautes écoles spécialisées (HES-SO, 2019).

¹⁵ Documents de conférence, articles scientifiques, articles professionnels, chapitres de livres, rapports de recherche, livres, masters et doctorats.

Actuellement, la HES-SO ne possède pas encore son propre préfixe DOI. Au sein d'ArODES, la génération d'identifiants est effectuée à la chaîne (*numerus currens*) et la granularité se construit sur la base du permalien lorsque le fichier plein texte est disponible. Comme par exemple :

- <https://arodes.hes-so.ch/record/2393>
- https://arodes.hes-so.ch/record/2393/files/santos_2018_guide_gestion_affiches.pdf

5.1.4 Études de cas du SIB

Nous n'avons pas retenu d'études de cas particulières au sein du SIB, mais étant donné qu'il sera un partenaire de choix du projet INCIPIT, nous avons tout de même analysé certaines de ses spécificités, notamment en matière de la variété d'objets pouvant nécessiter l'emploi d'identifiants ARK. A ce sujet, nous estimons que les niveaux de granularité ayant besoin d'être identifiés pour le SIB sont : la publication, la micro-publication, le jeu de données, l'instance au sein d'un jeu de données, ainsi qu'une donnée au sein d'un jeu de données.

Le SIB ne possède pas de politique uniforme, celle-ci étant actuellement en est train d'être définie, en fonction de différents projets. Les DOI sont déjà utilisés dans certains cas, mais cela demeure marginal, le SIB recourant massivement aux URL. Malgré tout, le SIB sera un excellent cas d'étude pour générer des ARK car leurs objets sont très variés et requièrent un niveau de granularité particulièrement fin.

5.1.5 Études de cas sélectionnées

Ci-après, vous trouverez un tableau récapitulatif des études de cas sélectionnées, spécifiant leur provenance, leur intitulé, une courte description ainsi qu'une URL d'accès.

Tableau 6 : Études de cas pour le Swiss PID Hub

Cas	Provenance	Intitulé de l'objet	Courte description et URL (plateforme)
01	ICOPAD - ZB	<i>Bibliothecae novae Tigurinorum publico privatae album / Donatorenbuch</i>	Manuscrit numérisé https://www.e-manuscripta.ch/zuz/doi/10.7891/e-manuscripta-45784 (e-manuscripta)
02	ICOPAD - ZB	<i>Album von Zürich-See</i>	Illustration numérisée https://www.e-rara.ch/zuz/doi/10.3931/e-rara-53198 (e-rara)
03	ICOPAD - ZHdK	<i>Blue My Mind</i>	Vidéo d'une étudiante de la ZHdK réalisée dans le cadre d'un Travail de Master of Arts in Film https://www.emuseum.ch/objects/199242/blue-my-mind (eMuseum)

Cas	Provenance	Intitulé de l'objet	Courte description et URL (plateforme)
04	ICOPAD - ZHdK	<i>Museum für Gestaltung - Type 1926</i>	Description d'un objet muséal https://www.emuseum.ch/objects/149981/type-1926 (eMuseum)
05	ICOPAD - ZHdK	<i>Menschenhaar</i>	Entrée de dictionnaire de cheveux humains http://materialarchiv.ch/app-tablet/#detail/1701/menschenhaar (Materialarchiv)
06	ICOPAD - ZHdK	<i>Sound Colour Space</i>	Projet de recherche artistique https://www.zhdk.ch/forschungsprojekt/426348
07	ICOPAD - ZHdK	<i>Archiv Performativ</i>	Performance artistique http://archivperformativ.zhdk.ch/
08	ICOPAD - SIK-ISEA	<i>Dieter Roth</i>	Entrée de dictionnaire d'un artiste http://www.sikart.ch/KuenstlerInnen.aspx?id=4000336 (SIKART)
09	FAQ ARK - BnF	<i>Jorge Luis Borges / François Taillandier</i>	Livre numérisé https://gallica.bnf.fr/ark:/12148/bpt6k3333931q (Gallica)
10	FAQ ARK - Industry Documents Library	<i>International Climate Change Partnership Meeting March 20, 2003</i>	Procès-verbal numérisé https://n2t.net/ark:/88122/nyfl0228
11	FAQ ARK - RMap Hub	<i>RMap DiSCO</i>	Graphe Linked Data interactif https://rmap-hub.org/discos/ark%3A%2F87281%2Ft2vh6csn
12	FAQ ARK - Internet Archive	<i>nNASOS1.8 / Nintendo</i>	Jeu vidéo https://n2t.net/ark:/13960/t04z22f41

Cas	Provenance	Intitulé de l'objet	Courte description et URL (plateforme)
13	FAQ ARK - Smithsonian	<i>Bombus unidentified</i>	Inventaire de spécimen https://n2t.net/ark:/65665/3a01b5a62-5dbd-4f5e-8c01-166c38d59ad5 (Entomology Collection)
14	FAQ ARK - Social Networks and Archival Context	<i>Curie, Marie, 1867-1934</i>	Entrée de dictionnaire d'une chercheuse https://snaccooperative.org/ark:/99166/w6qv3kx7
15	FAQ ARK - Periodo	<i>Bronze Age Malta (ca. 2,500-700 BC)</i>	Terme de vocabulaire https://n2t.net/ark:/99152/p03wskdjgsh
16	ArODES	<i>Data-driven approach for measuring the severity of the signs of depression using reddit posts: women and men in the orchestra</i>	Document de conférence https://arodes.hes-so.ch/record/4111
17	ArODES	<i>Le concept de qualité en archivistique contemporaine : quelques pistes...</i>	Article scientifique https://arodes.hes-so.ch/record/397
18	ArODES	<i>Résoudre des problèmes</i>	Article professionnel https://arodes.hes-so.ch/record/4064
19	ArODES	<i>Classer numérique</i>	Chapitre de livre https://arodes.hes-so.ch/record/703
20	ArODES	<i>Suggested measures for deploying IIF in Swiss cultural heritage institutions</i>	Rapport de recherche https://arodes.hes-so.ch/record/3607

Cas	Provenance	Intitulé de l'objet	Courte description et URL (plateforme)
21	ArODES	<i>Un intranet institutionnel 2.0 : transfert de l'intranet des Hôpitaux universitaires de Genève d'une technologie Java vers une technologie PHP</i>	Livre https://arodes.hes-so.ch/record/410
22	ArODES	<i>Étude sur la définition et la mesure des qualités des archives définitives issues d'une évaluation</i>	Doctorat https://arodes.hes-so.ch/record/310

5.2 Analyse

Ci-dessous, nous vous proposons un retour sur les analyses des études de cas sur lesquelles nous nous sommes penchés. Pour une vue d'ensemble plus pratique, nous avons réalisé un tableau classant ces études de cas par typologie d'objet, puis une grille de comparaison ordonnant les ressources selon leurs métadonnées et leur permanence. Une analyse a été effectuée pour chacune de ces présentations qui exposent de façons différentes les études de cas sélectionnées.

Il est difficile de classer les objets selon leur typologie, mais pour représenter la diversité de nos études de cas, nous nous sommes essayés à cet exercice comme le font par exemple l'élément *resourceType* du vocabulaire DC (DC.Type) (Cox et al., 1998) ou encore la propriété *resourceTypeGeneral* du schéma DataCite, utilisée comme l'un de six champs obligatoires de métadonnées pour les DOI (de Smaele et al., 2018).

A noter que pour les objets nés numériques ou numérisés, il ne s'agit pas du format d'enregistrement mais bien du genre de l'élément qui est essentiellement lié à la signification ou au contenu de la ressource. Nous avons ainsi déterminé six catégories génériques :

- **Texte** : acte de colloque, article, livre, loi, prépublication, thèse, etc.
- **Image** : photographie, œuvre picturale, dessin, scan, etc.
- **Audiovisuel** : film, son, jeu vidéo, etc.
- **3D** : expression, image, modèle mathématique, etc.
- **Organique** : molécule, spécimen, personne, etc.
- **Intangible / Immatériel** : processus, projet, performance artistique, etc.

En se basant sur cette typologie, nous avons classé dans le Tableau 7 ci-après les études de cas sélectionnées. Dès lors, nous remarquons que les études de cas du projet ICOPAD et celles de la FAQ ARK de l'initiative AITO sont très variées de par leur nature. Quant à celles

issues d'ArODES, assez logiquement vu les buts poursuivis par l'archive institutionnelle, elles se concentrent autour de données de type texte.

Tableau 7 : Études de cas par typologie d'objet

	Texte	Image	Audiovisuel	3D	Organique	Intangible / Immatériel
ICOPAD	01		03	04	05, 08	06, 07
		02				
FAQ ARK	09, 10	11	12		13, 14	15
ArODES	16, 17, 18, 19, 20, 21, 22					

La grille de comparaison (Figure 8, ci-après) donne une vision globale des études de cas selon deux axes (cf. Tableau 14 en Annexe pour une vue objet par objet). Chaque axe de la grille est découpé en trois quantiles, ce qui donne un maximum de neuf combinaisons différentes.

- L'axe des abscisses partage les études de cas selon les métadonnées utilisées pour identifier l'objet¹⁶ :
 - **Personnalisées** : un profil « maison » de métadonnées a été utilisé ;
 - **Mixte** : un mixte entre un profil « maison » et des éléments standardisés comme le DC ;
 - **Standardisées** : les métadonnées utilisées proviennent uniquement d'un standard bien établi, comme par exemple MARC.
- L'axe des ordonnées classe les études de cas en fonction de leur permanence, autrement dit de la disponibilité, supposée ou déclarée, des ressources. Les trois catégories qui ont été retenues ou inspirées de la nomenclature de (Kunze et al., 2017) sont :
 - **Indéterminée** : aucune information sur la disponibilité de l'objet ou aucun engagement particulier de l'institution fournissant son accès ;
 - **Supposée permanente** : lorsqu'il y a une mention de permaliens ou lorsqu'un PID a été utilisé pour identifier l'objet sans pour autant fournir une mention sur sa permanence ;
 - **Permanente** : lorsqu'il y a une déclaration de permanence explicite de l'objet.

Les informations complémentaires que l'on peut trouver au sein de cette grille de comparaison sont les types d'identifiants utilisés ou mentionnés (ARK, DOI, Handle, etc.), si l'objet contient des liens externes vers des sites web, catalogues (WorldCat¹⁷) ou notices d'autorité (telles que le GND allemand¹⁸ ou VIAF¹⁹) et finalement si l'objet est disponible dans une visualisation ou sérialisation en LOD (JSON-LD, RDF, Turtle) ou encore par le biais des API de IIIF.

Nous pouvons ainsi remarquer que le Hub devrait être en capacité d'accueillir, en identifiants ARK, un large éventail d'objets sur le court, moyen et long terme en créant des

¹⁶ A noter qu'il s'agit bien des métadonnées disponibles sur la page de destination de l'objet et non les métadonnées associées de l'identifiant comme celles du noyau ERC qui peuvent être parfois obtenues si on infléchit un ARK actionnable.

¹⁷ WorldCat : <https://www.worldcat.org/> (consulté le 10 janvier 2020)

¹⁸ Gemeinsame Normdatei (GND) : <https://d-nb.info/> (consulté le 10 janvier 2020)

¹⁹ Virtual International Authority File (VIAF) : <https://viaf.org/> (consulté le 10 janvier 2020)

inférences vers des bases de données externes, en LOD ou non. Cela présuppose donc un service dynamique et flexible qui puisse identifier aisément tous types de données et d'associer n'importe quel schéma de métadonnées selon les besoins des institutions clientes.

Figure 8 : Grille de comparaison des études de cas (métadonnées et permanence)

<p>Etude(s) de cas</p> <p>Identifiant(s) : ARK, DOI, Handle, ISBN, URL</p> <p>Liens externes : GND, VIAF, Wikipédia, etc.</p> <p>LOD : JSON-LD, IIIF, RDF, Turtle</p>	<p>01</p> <p>DOI</p> <p>Wikipédia</p> <p>IIIF</p>	<p>02</p> <p>DOI</p> <p>GND, Wikipédia</p> <p>IIIF</p>	<p>09</p> <p>ARK</p> <p>IIIF</p>	<p>Permanente</p>	
<p>03, 04</p> <p>URL</p> <p>12</p> <p>ARK (via N2T)</p> <p>14</p> <p>ARK</p> <p>LOC, VIAF, Wikipédia, WorldCat</p> <p>JSON-LD</p>	<p>10, 13</p> <p>ARK (via N2T)</p>	<p>15</p> <p>ARK (via N2T)</p> <p>JSON-LD, Turtle</p>	<p>11</p> <p>ARK</p> <p>Graphe RDF</p>		<p>Supposée permanente</p>
<p>05, 06, 07</p> <p>URL</p> <p>08</p> <p>URL</p> <p>DHS, european-art.net, GND</p>			<p>16, 17, 18, 21</p> <p>URL</p> <p>19</p> <p>URL + ISBN</p> <p>20</p> <p>URL + DOI externe</p> <p>22</p> <p>URL + Handle externe</p>		<p>Indeterminée</p>
<p>Métadonnées personnalisées</p>	<p>Métadonnées mixtes</p>	<p>Métadonnées standardisées</p>			

6. Modélisation

Le chapitre de modélisation se consacre à la conceptualisation du Hub en explorant les services similaires d'attribution de PID (cf. 6.1), en investiguant les critères essentiels sur lesquels pourront se baser les parties prenantes du projet INCIPIT à la création d'un service ARK en Suisse (cf. 6.2), ainsi qu'en imaginant un processus d'attribution des identifiants sous forme de logigrammes (cf. 6.3) et un parcours utilisateur sous la forme d'un Blueprint de service qui retrace le chemin d'une organisation voulant s'inscrire au Hub pour assigner des identifiants ARK à ses ressources (cf. 6.4).

6.1 Comparaison de services similaires

Afin de proposer une modélisation du Hub réaliste, basée sur des services existants, nous avons mené une comparaison entre le service EZID de l'Université de Californie proposant, entre autres des identifiants ARK, et le DOI-Desk de la Bibliothèque de l'EPFZ qui, comme son nom l'indique, fournit des DOI aux institutions suisses.

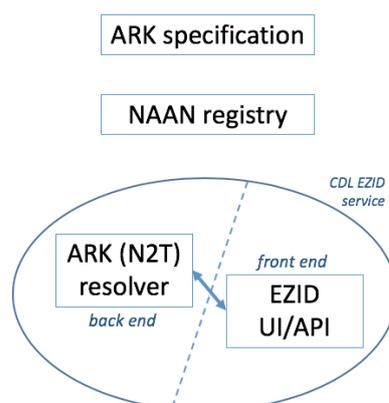
Nous mentionnons également à la fin de cette section deux autres organisations importantes en Suisse qui attribuent des PID. Il s'agit de la Bibliothèque nationale suisse (BN) qui assigne des Uniform Resource Name (URN) et du Centre suisse de calcul scientifique (CSCS) qui attribue pour sa part des identifiants ePIC (*Persistent Identifiers for eResearch*).

6.1.1 EZID (CDL)

EZID est un service de gestion de PID à la CDL permettant de créer et de gérer des DOI, des ARK et, dans un futur proche, des identifiants d'autres types. La CDL héberge d'ailleurs le résolveur global N2T (cf. 6.2.1.2) qui redirige tous les PID gérés par EZID (ainsi que d'autres, comme LSID, LSRN, Handle, et certaines classes d'URN).

EZID a pour principale vocation de servir d'agence de maintenance pour le schème ARK (cf. Figure 9 ci-dessous qui représente l'architecture du système ARK à la CDL).

Figure 9 : Architecture ARK à la CDL



(Kunze, 2019a)

6.1.1.1 Processus d'attribution et fonctionnalités

EZID est un moyen simple d'obtenir, de décrire et de gérer les PID objet, entre autres numériques. Pour y accéder, il existe une interface utilisateur, qui propose quelques fonctions de gestion de compte, mais également une API donnant accès aux cinq opérations du service. Le Tableau 8 ci-dessous décrit succinctement les fonctionnalités d'EZID (CDL, 2014).

Tableau 8 : Opérations du service EZID

Opération	Description
1) <i>Get identifiant metadata</i>	Les métadonnées peuvent être récupérées pour n'importe quel identifiant, et une authentification n'est pas requise.
2) <i>Create identifiant</i>	<p>Créer un identifiant signifie établir un enregistrement de l'identifiant dans EZID en envoyant une requête URL. Ici, l'authentification est nécessaire et l'utilisateur doit avoir la permission de créer des PID dans l'espace de noms (<i>shoulder</i>) nommé par le préfixe de l'identificateur. Ces espaces de noms sont visibles par les utilisateurs dans l'interface (onglet « Créer un identifiant »).</p> <p>En ce qui concerne les métadonnées, un corps de requête optionnel permet de définir les métadonnées de départ de l'identifiant, sans restrictions sur celles qui peuvent être soumises, même s'il existe une convention pour en nommer les différents éléments. EZID propose également un support intégré pour certains ensembles de métadonnées. A noter que certains schémas d'identifiants possèdent des règles de normalisation différentes, par exemple les DOI en matière de majuscules.</p>
3) <i>Mint identifiant</i>	Générer un identifiant est assez similaire à sa création, mais le client, plutôt que de fournir un identifiant complet, spécifie seulement un espace de noms formant le préfixe de l'identifiant, et le service EZID s'occupe de générer une chaîne aléatoire et opaque pour le suffixe. L'authentification reste nécessaire pour créer un identifiant ainsi que pour en générer un. EZID incorpore automatiquement le nouveau PID dans certains types de métadonnées et remplace toutes les occurrences de la chaîne dans l'URL cible par le nouvel identifiant créé.
4) <i>Update identifiant</i>	Les métadonnées d'un identifiant peuvent être mises à jour par l'envoi d'une requête URL. Ici, l'authentification du client est bien entendu requise car seuls le propriétaire et certains utilisateurs définis ont le droit de le modifier. Dans ce cas, les éléments des métadonnées sont exploités individuellement, et les valeurs peuvent être supprimées, remplacées ou ajoutées. Toutes les métadonnées ne peuvent par contre pas être mises à jour.
5) <i>Delete identifiant</i>	Seuls le propriétaire et certains utilisateurs définis ont le droit de supprimer un identifiant. A noter que les identifiants réservés peuvent également être détruits.

EZID propose aussi un « service démo »²⁰ permettant au client de tester le service en créant des identifiants dont le NAAN est 99999. Les identifiants créés par ce moyen ont une durée de vie de deux semaines environ.

²⁰ Service démo de EZID : <https://ezid.cdlib.org/demo/simple> et <https://ezid.cdlib.org/demo/advanced> (consultés le 10 janvier 2020)

6.1.1.2 Domaines d'utilisation

Il est évident que la plupart des utilisateurs de ce service sont des départements, facultés, centres de recherche et bibliothèques universitaires anglo-saxonnes, en particulier celles présentes sur le campus de l'Université de Californie. L'exemple de Princeton est intéressant, car avant de faire appel à EZID, cette université générait ses propres identifiants ARK.

« At Princeton, we mint NOIDs and bind ARKs via California Digital Library's EZID service. We ran our own instance of the ARK/NOID software for several years, but, considering the relatively low cost, moving to EZID and eliminating our local maintenance burden was an easy business decision. »

Jon Stroop, Director of Library Information Technology and Digital Services (Princeton University Library)²¹

Plus généralement, les ARK sont surtout utilisés par les institutions culturelles telles que les musées, les archives et les bibliothèques²².

6.1.1.3 Tarification

Le prix pour l'utilisation du service EZID varie en fonction de plusieurs critères, notamment celui du type d'organisation cliente (CDL, 2020). Dans nos recommandations financières, nous proposons d'adopter un genre de tarification similaire. Les prix affichés ci-dessous comprennent les cotisations annuelles (pour la gestion et la maintenance de base) ainsi que les frais de création et de maintenance pour un million d'identifiants par année (Kunze, 2019b).

- Toutes les institutions liées à l'Université de Californie bénéficient de la **gratuité**.
- Tarifs spéciaux appliqués aux écoles et organisations non-lucratives :
 - Niveau Bachelor : **\$300**
 - Niveau Master : **\$600**
 - Institut de recherche / organisation sans but lucratif : **\$1'500**
 - Groupe/département/équipe de recherche : **\$500**
- Tarifs appliqués aux organisations lucratives :
 - Petites (< 100 employés) : **\$1'500**
 - Moyennes (100 à 500 employés) : **Prix sur demande**
 - Grandes (> 500 employés) : **Prix sur demande**

Les consortiums liés à l'éducation et comportant trois ou plus institutions ou organisations en leur sein bénéficient par ailleurs d'une réduction de 20% des coûts.

²¹ Échange de messages directs entre l'un des auteurs de ce mémoire et Jon Stroop sur l'instance Slack de IIF (<https://iif.slack.com/>) le 22 novembre 2019.

²² On parle généralement du secteur GLAM en anglais – (*Galleries, Libraries, Archives, Museums*).

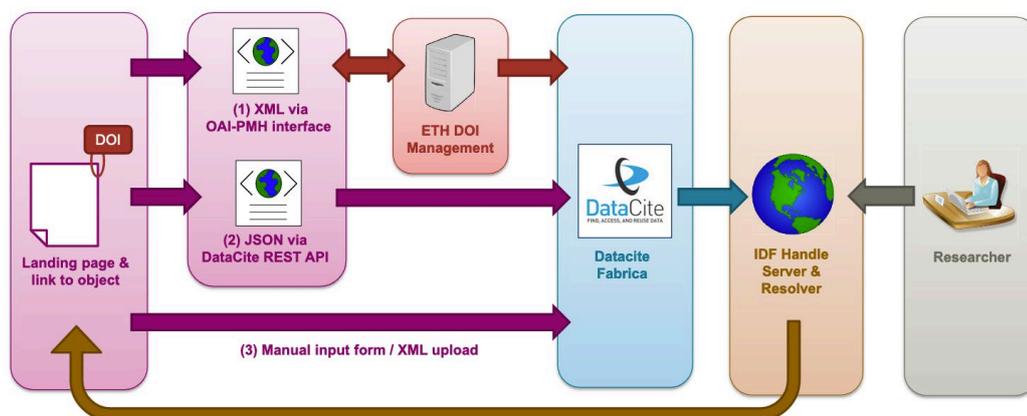
6.1.2 DOI-Desk (Bibliothèque de l'EPFZ)

Le DOI-Desk de la Bibliothèque de l'EPFZ, qui est membre de DataCite, attribue des DOI en Suisse depuis 2009 (Hirschmann, 2019). Le service est offert aux institutions académiques et de recherche désirant obtenir des identifiants pour leurs objets numériques.

6.1.2.1 Processus d'attribution et fonctionnalités

Dès qu'une organisation helvétique est inscrite au DOI-Desk, elle a la possibilité de générer automatiquement des DOI via l'API de DataCite, ou semi automatiquement via l'interface de moissonnage OAI-PMH (ETHZ Library, 2019). Le service propose également de passer par l'outil *DataCite Fabrica* pour obtenir manuellement des identifiants. La Figure 10 suivante montre une vision d'ensemble des différents processus d'enregistrement des DOI.

Figure 10 : Processus d'enregistrement des DOI via la Bibliothèque de l'EPFZ



(Hirschmann, 2019)

Pour enregistrer un objet ayant un DOI, cinq métadonnées, sans compter le numéro d'identifiant, sont obligatoires : *Creator*, *Title*, *Publisher*, *Publication Year* and *Resource Type General*. Ces champs peuvent être remplis à la main ou par le biais du versement d'un fichier XML conforme aux exigences de DataCite.

6.1.2.2 Domaines d'utilisation

Le service peut être utilisé par les hautes écoles (HES, EPF et universités), les centres de recherche ainsi que par les organisations à but non lucratif en Suisse. Les organisations à but lucratif ont cependant aussi la possibilité de s'inscrire au DOI-Desk (ETHZ Library, 2018).

Si les publications comme les articles scientifiques (articles et données de la recherche) représentent la très grande majorité des types de données ayant un DOI, n'importe quel objet numérique peut obtenir un DOI via le DOI-Desk.

6.1.2.3 Tarification

Pour les établissements académiques et les organisations à but non lucratif, la tarification est la suivante :

- Frais d'initialisation (paiement unique à la conclusion du contrat) : **CHF 200**
- Tarif de base (payable annuellement pour un maximum de 50 DOI) : **CHF 200**
- Tarif « DOI top-up » (payable annuellement en plus du tarif de base pour enregistrer un nombre illimité de DOI) : **CHF 800**

Autrement dit, la première année, l'utilisation du DOI-Desk coûte soit CHF 400 (jusqu'à 50 DOI), soit CHF 1'200 (nombre illimité de DOI). Dès la seconde année, les frais d'initialisation tombent et le tarif varie donc de CHF 200 (jusqu'à 50 DOI) à CHF 1'000 (nombre illimité de DOI). A l'image du fonctionnement de l'EZID avec les départements et facultés de l'Université de Californie, il faut relever que le service DOI-Desk est gratuit pour les entités appartenant à l'EPFZ.

Dès 2020, il y aura au sein de l'association DataCite un nouveau modèle d'adhésion. Cela impliquera que le modèle de tarification sera également modifié pour les institutions clientes du DOI-Desk (Hirschmann, 2019).

6.1.3 Obtention d'URN et d'identifiants ePIC en Suisse

Cette rubrique présente deux autres services attribuant des PID en Suisse dans le domaine du patrimoine culturel ainsi que dans le domaine scientifique.

6.1.3.1 Service URN de la Bibliothèque nationale suisse

L'URN est un identificateur univoque pour des ressources sur Internet. C'est la BN qui attribue ces URN en Suisse sur la base du National Bibliography Number (NBN), domaine créé suite à une initiative commune des bibliothèques nationales, initiative issue des activités de la Conference of Directors of National Libraries (CDNL) et de la Conference of European National Librarians (CENL). La BN collabore avec la Deutsche Nationalbibliothek (DNB) et la Österreichische Nationalbibliothek (ÖNB) pour la gestion des espaces noms d'URN externes, pour la résolution, ainsi que pour l'enregistrement des données relatives à l'URN (BN, 2016).

La BN dispose, comme l'avait déjà relevé le travail d'Espasandin et al. (2018, p. 25), d'une politique d'attribution dans laquelle l'espace nom est `urn:nbn:ch`, et où l'attribution et l'enregistrement sont réalisés de façon centralisée par la bibliothèque nationale (ou les archives de longues durées). Par ailleurs, tous les objets numériques traités par la BN reçoivent un URN, même s'ils possèdent déjà un identifiant d'un autre type. En plus d'être gratuit, l'URN garantit assurément une bien meilleure stabilité que les URL, tout en permettant un accès durable à la ressource en raison de la disponibilité technique du service.

6.1.3.2 ePIC par le biais du Centre suisse de calcul scientifique

Depuis septembre 2018, le CSCS de Lugano²³ est membre du consortium d'ePIC, qui vise à fournir des services d'identifiant pour la communauté européenne de la recherche. Ce PID, tout comme les DOI, est basé sur le système Handle.

La cotisation annuelle pour obtenir un maximum de 500'000 identifiants ePIC avec son propre préfixe (commençant soit par `21.17xxx`, soit par `21.T17xxx`) de la part du CSCS est de CHF 600. La gestion et la création d'identifiants ePIC peut se faire par le biais de deux interfaces (CSCS, 2019).

6.2 Vue d'ensemble des critères de prototypage

Afin de proposer une solution, ou du moins des pistes de réflexion quant à la création d'un Hub d'attribution de PID au niveau national, nous avons axé nos recommandations sur la base de quatre types de critères : techniques, juridiques, organisationnels et financiers. Au travers

²³ The Swiss PID service for scientific data : <https://pid.cscs.ch/home/> (consulté le 10 janvier 2020)

ceux-ci, nous avons pu identifier les outils existants ainsi que les différents scénarios à prendre en compte, produire une estimation des coûts et de la charge humaine de travail. Tout ceci dans le but de permettre une réalisation effective optimale.

6.2.1 Considérations techniques

Ces considérations techniques portent sur différents outils utilisés en matière de création d'ARK, de résolveur, de déclaration de permanence, et sur l'infrastructure nécessaire.

6.2.1.1 Générateur d'identifiants ARK et résolveur local

Les institutions attribuant des ARK sont libres de choisir comment elles souhaitent générer et résoudre de manière locale leurs identifiants. La création des chaînes de caractères qui composent le nom ARK doivent tout de même remplir les conditions suivantes : utiliser les chiffres et lettres respectant la norme informatique ASCII (*American Standard Code for Information Interchange*), la sensibilité de la casse (*xy*, *Xy* et *XY* sont par exemple distincts) et exclure les caractères suivants²⁴ :

= ~ * + @ _ \$. /

La BnF a, par exemple, déployé sa propre « application d'aiguillage des ARK »²⁵ qui redirige les identifiants ARK vers les NMA appropriés.

Si le Hub voulait s'appuyer sur une solution existante, le logiciel Nice Opaque Identifier (NOID), écrit dans le langage de programmation Perl et développé par la CDL²⁶, pourrait être utilisé. NOID est généralement configuré comme un résolveur fonctionnant en arrière-plan d'un serveur web (CDL, 2019e). Il fait ainsi office de « convertisseur » à la volée d'adresses URL publiées de manière pérenne. Dans ce type d'arrangement, NOID maintient une table de correspondance qui est consultée chaque fois qu'un navigateur web demande un ARK actionnable (ou autre schème de PID) au serveur. NOID permet également d'associer des métadonnées aux PID. Il faut également noter que le répertoire que NOID utilise pour générer des noms ARK contient seulement des chiffres et des consonnes, excepté la lettre « L » en minuscule.

La génération d'ARK est également dépendante des besoins de granularité ou non dans l'identification des ressources de la part des institutions clientes. Le Hub pourrait tout à fait proposer des paquets comportant des fonctionnalités de base ou avancées et par conséquent une tarification différente en fonction des besoins de chaque institution (cf. 6.2.4.2).

6.2.1.2 Résolveur global

N2T est un service global de résolution partagé qui se base sur le registre des NAAN pour résoudre les identifiants ARK et qui utilise le logiciel Open Source Eggnog²⁷. Il est également maintenu par la CDL.

La mise sur pied d'une telle infrastructure permet en quelque sorte de se prémunir de la disparition d'une organisation et/ou de ses serveurs web car N2T est par exemple le seul résolveur de PID qui « *ne nécessite pas de modifications du navigateurs web et qui offre une*

²⁴ Les deux derniers caractères sont des caractères réservés par la spécification ARK pour divulguer des relations de granularité ou de service.

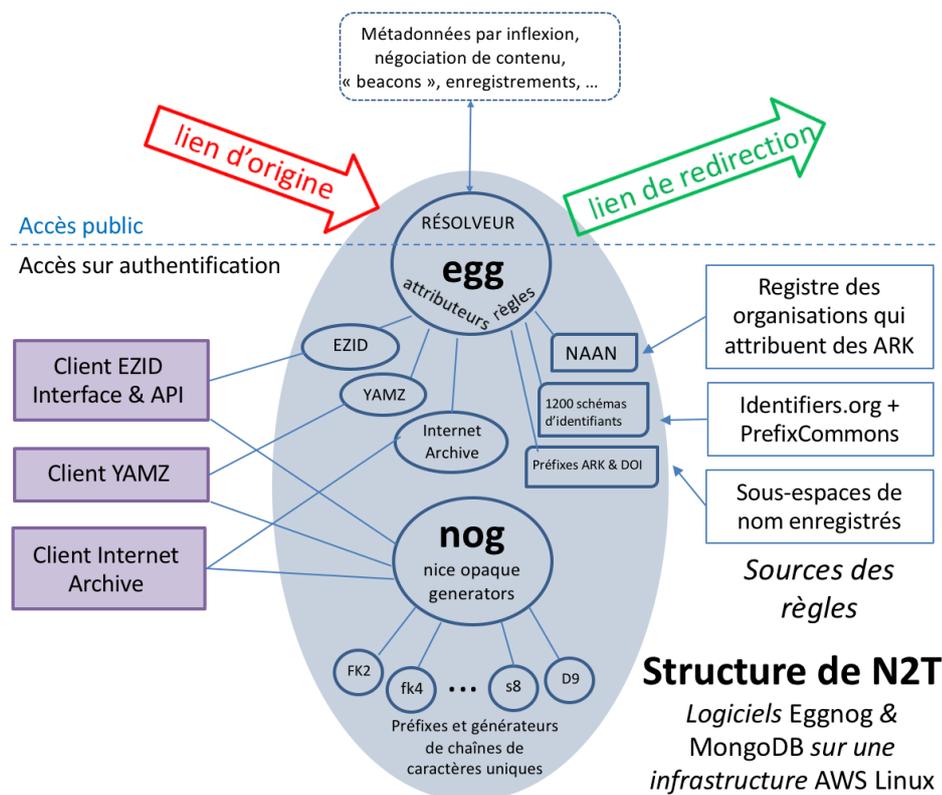
²⁵ Résolveur ARK de la BnF : <http://ark.bnf.fr/> (consulté le 10 janvier 2020)

²⁶ Nice Opaque Identifier : <https://n2t.net/e/noid.html> (consulté le 10 janvier 2020)

²⁷ Eggnog : <https://bitbucket.org/cdl/n2t-eggnog/src/default/> (consulté le 10 janvier 2020)

solution à l'inévitable problème de la scission d'espace de noms », autrement dit lorsqu'une organisation se scinde en plusieurs entités et que de nouveaux serveurs web doivent être déployés (CDL, 2007 ; Kunze, Caron, 2019a). La demande d'un NAAN supplémentaire et le transfert d'une partie ou de la totalité des NMA et de ses préfixes sont ainsi facilités grâce à l'anatomie flexible des ARK et de la structure du résolveur globale (cf. Figure 11 ci-après).

Figure 11 : Structure du résolveur global N2T



(Kunze, Caron, 2019b)

La maintenance de N2T est assurée par Internet Archive et la CDL, par le biais d'EZID et du métadictionnaire YAMZ²⁸. Un partenariat du Hub avec N2T serait une opportunité intéressante car cela permettrait à l'infrastructure de bénéficier d'un accès direct à la table de redirection N2T et cela éviterait au Hub de devoir jouer les intermédiaires si une de ses institutions clientes venaient à se scinder en deux ou à disparaître.

6.2.1.3 Schémas de métadonnées et déclarations de permanence

Une vision agnostique au sujet des métadonnées associées serait le meilleur moyen de répondre aux besoins des clients et des utilisateurs finaux. Chaque institution cliente du Hub devrait pouvoir choisir un ou plusieurs standards de métadonnées ou encore avoir la possibilité de soumettre des fichiers CSV, XML ou JSON où les informations seraient aisées à extraire et analyser.

Par défaut, le noyau ERC devrait être proposé pour associer des métadonnées aux identifiants ARK et il serait même envisageable de ne pas remplir les champs *who*, *what* et *when*²⁹.

²⁸ YAMZ : <https://yamz.herokuapp.com/> (consulté le 10 janvier 2020)

²⁹ Le champ *where* doit être rempli car il constitue un objectif lors de la création d'ARK.

Les schémas de métadonnées qui pourraient être proposés ou soumis pour contenter un large ensemble d'organisations culturelles sont les suivants :

- Le noyau ERC ;
- Les éléments de base et le set qualifié du DC ;
- MARC ;
- Resource Description and Access (RDA) ;
- Encoded Archival Description (EAD) ;
- Records in Context (RiC)³⁰ ;
- CIDOC-Conceptual Reference Model (CRM)³¹.

Le Hub pourrait également déployer une solution en LOD en se basant sur les métadonnées des identifiants ou les informations obtenues via des liens externes. Cela pourrait notamment être effectué par le biais d'une sérialisation des hiérarchies et des variantes des identifiants (qualificatifs de granularité et de service) grâce aux vocabulaires et ontologies du Web sémantique (schema.org, FOAF, OWL, etc.) ainsi qu'aux liens vers des plateformes telles qu'ORCID, International Standard Name Identifier (ISNI), VIAF, Global Research Identifier Database (GRID), Research Organization Registry (ROR), Rameau, GND, ou encore Wikidata.

En revanche, ce qui ne devrait pas être paramétrable par les clients du Hub sont les métadonnées techniques issues du contexte de création, réservation, modification ou suppression d'ARK pour des raisons de traçabilité.

Quant aux déclarations de permanence renseignant sur la disponibilité des objets, le Hub pourrait se baser sur la nomenclature proposée par Kunze et al. (2017)³² et exposer ces informations aux utilisateurs finaux par le biais des inflexions.

Les objets qui auront une durée limitée et qui ne seront plus disponibles devront soit rediriger vers une autre ressource ou vers une page web expliquant les raisons de suppression tout en affichant les métadonnées descriptives, soit vers une page web générique du Hub donnant des explications sur la non-disponibilité de la ressource si aucune initiative n'est prise par l'institution cliente du service.

Néanmoins, cette option est uniquement possible si un ARK sous sa forme N2T est actionné. Pour vérifier la disponibilité des ARK, cela présuppose autant une communication entre les institutions clientes et le Hub qu'un suivi automatisé des identifiants attribués.

6.2.1.4 Infrastructure

Le Hub devra se munir d'une infrastructure comprenant des serveurs abritant une table de redirection (Mozilla Developer Network, 2019) et hébergeant notamment les données suivantes : métadonnées associées des identifiants, données des institutions clientes, site web et API.

³⁰ International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 0.1 : <https://www.ica.org/standards/RiC/ontology.html> (consulté le 10 janvier 2020)

³¹ CIDOC-CRM : <http://www.cidoc-crm.org/> (consulté le 10 janvier 2020)

³² *Finite, indefinite, lifetime, subinfinite*

Concernant le site web, celui-ci devra être autant une plateforme d'information et de promotion auprès des institutions clientes et des prospects qu'une interface utilisateur offrant les fonctionnalités de création et de gestion d'ARK. L'API aura pour fonction d'être un accès privilégié pour créer de manière semi-automatique un grand nombre d'identifiants. Ces deux accès doivent également proposer différentes manières de créer des noms ARK prenant en compte les besoins du client dans l'automatisation ou non du processus et ceci sans oublier la question de l'opacité.

Une plus-value d'EZID consiste à proposer une interface de recherche³³ (cf. Figure 12). Le Hub pourrait tout à fait s'en inspirer et exposer à cet endroit tous les ARK comportant un statut public.

Figure 12 : Interface de recherche d'EZID

UNIVERSITY OF CALIFORNIA **EZID** Identifiers made easy

SEARCH LEARN CONTACT LOGIN

Search across EZID identifiers

Full text search using words about or describing the identifier. **SEARCH**

EZID search uses AND logic which means ALL terms entered into a search box must be present in order to show up in the results. [Read more](#)

^ **ADVANCED SEARCH** ?

Identifier/Identifier Prefix: e.g. doi:10.17605/OSF.IO/QXUPF

Object Title (What): e.g. Welfare: current vs. never

Object Creator (Who): e.g. Schofield, Timothy

Object Publisher: e.g. Open Science Framework

Object Publication Year: From: e.g. 2015 To: e.g. 2016

Object Type: Select a type of object

ID Type: ARK **SEARCH**

Une telle infrastructure devra bien évidemment fonctionner 24h/24 et détenir assez de sauvegardes de sécurité. Autrement dit, pour que l'infrastructure soit stable, elle doit avoir de la redondance, en utilisant par exemple des serveurs de type *Redundant Array of Independent Disks 10* (RAID-10) tout en les dispersant géographiquement (Danilak, 2012). En Suisse, cela peut être effectué au moyen d'une solution cloud telle que SWITCHengines³⁴, un service de stockage et de puissance de calcul chez SWITCH qui a déjà proposé soit la gratuité, soit des tarifs privilégiés aux projets de recherche financés par le fonds P5.

6.2.2 Considérations juridiques

Cette section porte sur les différentes formes juridiques que le service pourrait prendre lors de sa phase de maintenance, sur les différents modèles de contrats que le Hub pourrait proposer à ses clients ainsi que sur la question de l'hébergement des données.

6.2.2.1 Forme juridique du Hub

La première étape, l'initialisation du Hub, s'effectuera au sein de la filière Information documentaire (ID) de la HEG-GE via le projet INCIPIT en collaboration avec le SIB. Par la

³³ EZID Search : <https://ezid.cdlib.org/search> (consulté le 10 janvier 2020)

³⁴ SWITCHengines : <https://www.switch.ch/engines/> (consulté le 10 janvier 2020)

suite, plusieurs scénarios sont envisageables selon un plan de viabilité. Ces scénarios, explorés lors du projet INCIPIT, proposeront des solutions organisationnelles pour une mise en place effective du Hub.

De notre côté, en prévision du démarrage du projet INCIPIT, nous avons déjà réfléchi aux solutions organisationnelles les plus évidentes, et nous estimons que trois options principales, pouvant être déclinées de différentes manières, sont envisageables pour une gestion efficace du service :

- Prise en charge par la HEG-GE (via la filière ID, l'Infothèque, voire un service dédié) ;
- Création d'un service dédié de la HES-SO, via ArODES ;
- Regroupement en un consortium d'institutions composé d'un nombre important de parties prenantes liées au Hub.

Dans les deux premiers cas, il faut prendre en compte le fait que la HES-SO Genève est un établissement autonome de droit public « *doté de la personnalité morale placé sous la surveillance du Conseil d'État qui l'exerce par l'intermédiaire du département de l'instruction publique, de la formation et de la jeunesse* » (République et Canton de Genève, 2013, art. 1). Dès lors, et selon la mission fixée par la loi fédérale, la HES-SO au travers de ses filières peut proposer des prestations de services aux entreprises, administrations et institutions, ce qui a l'avantage de renforcer sa proximité avec son environnement économique, industriel, social et culturel (HEG-GE, 2012).

Dans le cas de la création d'un consortium composé de plusieurs institutions, il faudra veiller à respecter bien évidemment les dispositions légales prévues, à savoir l'inscription au registre du commerce et le fait que le but poursuivi doit être social et non lucratif, même si pour atteindre ce but idéal, l'association « *exerce une industrie en la forme commerciale* » (Code civil suisse, 2020, art. 60). Par ailleurs, les cotisations perçues de la part des institutions membres pourraient être calculées de manière à couvrir leurs besoins en matière d'identifiants, ce qui pourrait inciter des organisations à rejoindre le consortium du Hub.

Il faudra également réfléchir à une solution adéquate pour assurer la durabilité technique du service. Nous avons dès lors imaginé faire appel à un organisme externe pouvant assurer la maintenance 24h/24 des serveurs stockant les données, un organisme comme RERO ou SWITCH, possédant un savoir-faire certain dans le déploiement d'une infrastructure qui soit disponible à toute heure. La puissance de calcul que ces organismes pourraient offrir est également un argument nous faisant envisager cette solution.

6.2.2.2 Contrat de prestations de service

Assez classiquement, les contrats de prestations de service devront comporter toutes les informations habituelles des parties prenantes : dispositions générales, objet et champ d'application, obligations du fournisseur, conditions de travail et dispositions diverses.

Des modèles de contrats pour des prestations de service (mandat, expertise, consulting) existent déjà au sein de la HES-SO, et les adapter au Hub ne posera aucun problème. Il faudra simplement veiller à bien stipuler quel type de prestations pour quel type de clients (cf. Tableau 11 et Tableau 12).

6.2.2.3 Hébergement des données

Pour une protection optimale des données personnelles des clients du Hub, celles-ci devront être hébergées en Suisse (PFPDT, 2019). Cela peut être réalisé si l'on choisit d'utiliser les serveurs de la HEG-GE, de la HES-SO Genève ou encore si un service de cloud comme SWITCHengines est utilisé.

Les métadonnées des identifiants créés et hébergées sur le Hub auront une licence Creative Commons 0 (CC0), consistant « *au renoncement maximal des droits d'auteur* » (Creative Commons, 2009), pour pouvoir être valorisées et extraites facilement sur une plateforme de recherche, via une API ou encore par le biais d'une interface de requête SPARQL (Harris, Seaborne, 2013). Les institutions clientes auront tout de même la possibilité de notifier si l'identifiant d'un objet a un statut public ou non, mais ne devraient pas pouvoir modifier le type de licence des métadonnées qu'elles enregistrent et associent aux identifiants ARK pour des raisons de compatibilité.

Quant aux données des clients (adresse, nom de contact, date d'inscription, etc.), elles ne devront bien évidemment pas être publiées sans leur consentement.

6.2.3 Considérations organisationnelles

L'organisation du Hub doit prendre en compte deux aspects principaux essentiels : les prestations offertes et les ressources humaines.

6.2.3.1 Prestations offertes

La prestation principale qui sera proposée par le Hub sera bien entendu la création et la maintenance d'identifiants ARK sur le territoire suisse afin de garantir un accès pérenne, ou au moins une traçabilité sûre aux connaissances.

La génération d'identifiants ARK pourra être réalisée par le biais de fonctionnalités de base lors de la création ou réservation d'identifiants si les objets n'ont pas de hiérarchies particulières. L'utilisation des qualificatifs de granularité et de service ainsi qu'un soutien étendu aux institutions clientes dans la mise en place de politiques de permanence feront partie des fonctionnalités avancées du Hub.

De par son statut « officiel » validé par le fonds de recherche P5, le Hub aura pour vocation d'être un centre d'expertise sur les PID, notamment en collaborant avec l'initiative AITO, et devra donc se positionner en tant que soutien à la rédaction de politiques de permanence pour les institutions suisses, particulièrement pour les organisations du patrimoine culturel et académiques.

6.2.3.2 Ressources humaines

Il va sans dire que certaines compétences spécifiques en gestion de PID sont absolument nécessaires pour mener à bien ce projet et faire fonctionner le Hub, mais la formation à celles-ci ne demande pas excessivement de temps, et les profils demandés existent déjà. L'aspect technique étant essentiellement informatique, notamment lorsqu'il faudra « patcher » les serveurs en cas de dysfonctionnements, il pourrait être judicieux d'imaginer engager deux personnes complémentaires se répartissant les compétences. Si des connaissances et compétences étendues en informatique sont assez courantes sur le marché du travail, celles en matière de gestion de PID se trouvent préférentiellement parmi les diplômés du Master IS de la HEG-GE ou les personnes porteuses d'un titre jugé équivalent.

La charge de travail en équivalent plein temps (EPT) a été estimée sur la base des scénarios imaginés (cf. Tableau 9 et Tableau 10), des entretiens avec des professionnels du domaine et de l'analyse de services similaires. Pour pallier toute éventualité, nous proposons deux estimations permettant ainsi d'obtenir une fourchette réaliste dont les coûts sont détaillés dans la section 6.2.4.1. Nous estimons donc qu'un 0.2 EPT serait la charge minimum pour proposer un Hub fonctionnel et représenterait environ 360 heures de travail par an, charge qui pourrait aller jusqu'à 0.5 EPT et correspondre à 900 heures annuelles de travail.

6.2.4 Considérations financières

Estimer les coûts d'initialisation et de fonctionnement du Hub ne fut pas une tâche facile. Même lors d'entretiens avec des responsables de services similaires, il est très clairement apparu qu'une estimation financière dépendait de facteurs très variés, parfois difficilement identifiables. Bertrand Caron, expert en métadonnées de préservation à la BnF, nous a confirmé qu'il était difficile pour lui d'extraire les coûts relatifs au fonctionnement de son service étant donné que plusieurs solutions techniques avaient été développées à l'interne et qu'il était complexe de séparer les coûts de la gestion de la ressource de ceux de la gestion de l'identifiant lui-même.

Par ailleurs, en matière de finances, se posent les questions pratiques de comment vendre un service d'archivage numérique, et surtout de la quantification de la préservation numérique : quelle est concrètement la valeur monétaire d'un tel service ? Cette question rejoint plus globalement celle de la valeur des institutions culturelles, comme les bibliothèques.

Pour trouver des clients, le Hub devra compter sur un soutien politique de la part de la HES-SO et des institutions de financement de la recherche, sous formes de directives, de bonnes pratiques ou d'incitation à l'utilisation du service comme condition d'obtention de subventions.

Dans tous les cas, il est certain que les coûts de fonctionnement évolueront avec le temps et le développement d'outils technologiques plus appropriés.

6.2.4.1 Estimation des coûts

Très concrètement, nous estimons l'initialisation du Hub à environ 3'000 heures de travail, ce qui correspond à 1.65 EPT, charge qui sera intégrée au projet INCIPIT qui s'initiera sur la base de ce travail.

Au sujet de la maintenance, nous avons imaginé, comme expliqué dans le chapitre précédent, deux possibilités pour calculer les coûts de main-d'œuvre : la première propose au minimum une charge de travail équivalant à 0.2 EPT (environ 360 heures), la seconde à 0.5 EPT (environ 900 heures). Les estimations de coûts, calculées sur la base d'une rémunération de CHF 60 par heure, se trouvent dans les tableaux ci-après (cf. Tableau 9 et Tableau 10).

En ce qui concerne les coûts fixes, nous proposons deux scénarios, l'un optimiste, l'autre plutôt pessimiste. Le premier fait état d'un coût de CHF 15'000 annuel pour l'infrastructure de base auquel il faut ajouter CHF 5'000 pour un partenariat avec N2T, le second prévoit un coût d'infrastructure de CHF 20'000 et d'un partenariat de CHF 10'000 avec N2T. Cette fourchette de coûts permet de prendre en compte la difficulté d'estimation déjà relevée et de laisser une certaine marge de main-d'œuvre quant à la réalisation effective.

La fourchette d'estimations des coûts pour maintenir une infrastructure d'attribution d'ARK en Suisse se situe entre CHF 41'600 et CHF 84'000 par an.

Dans tous les cas, ces coûts fixes sont indispensables à la mise en place d'un Hub d'attribution d'identifiants ARK, et les éventuels coûts dus à de la main-d'œuvre supplémentaire pourront être dilués sur l'ensemble des budgets de l'institution accueillant le service.

Tableau 9 : Scénario optimiste d'estimation des coûts

Coûts annuels	Possibilité 1 (0.2 EPT)	Possibilité 2 (0.5 EPT)
Coûts de la main-d'œuvre	CHF 21'600	CHF 54'000
Coûts fixes	CHF 20'000	CHF 20'000
Coûts totaux	CHF 41'600	CHF 74'000

Tableau 10 : Scénario pessimiste d'estimation des coûts

Coûts annuels	Possibilité 1 (0.2 EPT)	Possibilité 2 (0.5 EPT)
Coûts de la main-d'œuvre	CHF 21'600	CHF 54'000
Coûts fixes	CHF 30'000	CHF 30'000
Coûts totaux	CHF 51'600	CHF 84'000

6.2.4.2 Propositions de tarification

Pour proposer un système financier viable, nous nous sommes penchés sur des propositions de tarifications inspirées par le fonctionnement d'autres services similaires (EZID, DOI-Desk, etc.). Il nous semblait évident d'offrir des tarifs adaptés en fonction du type d'institution cliente et des fonctionnalités souhaitées. Dans une logique de soutien à la recherche et de promotion de l'Open Science, les partenaires du Hub et les organisations à but non-lucratif bénéficieront de tarifs préférentiels.

La cotisation annuelle comprendrait la création et la maintenance d'un nombre illimité d'ARK (cf. Tableau 11 et Tableau 12). La deuxième proposition de tarification, celle de rendre le service gratuit à tous les partenaires du Hub, ne pourrait être viable que si une participation financière de la HES-SO ou d'un consortium est confirmée.

Tableau 11 : Proposition de tarification 1

Types de client	Fonctionnalités de base	Fonctionnalités avancées
Partenaires du Hub (dont la HES-SO)	CHF 150	CHF 300
Organisations à but non-lucratif	CHF 250	CHF 500
Organisations à but lucratif	CHF 1'000	CHF 1'500

Tableau 12 : Proposition de tarification 2

Types de client	Fonctionnalités de base	Fonctionnalités avancées
Partenaires du Hub (dont la HES-SO)	Gratuit	
Organisations à but non-lucratif	CHF 250	CHF 500
Organisations à but lucratif	CHF 1'000	CHF 1'500

6.2.4.3 Projections sur cinq ans

Ci-après (Figure 13), nous avons réalisé une projection du chiffre d'affaires (CA) estimé du Hub sur cinq ans, en prenant en compte un nombre potentiel de clients pour chaque type de contrat (fonctionnalités de base et avancées) et en fonction du type d'organisation. Une augmentation annuelle et graduelle de 200% du nombre de clients a été estimée. Cette estimation a été faite sur la base du Tableau 11, mais la dernière ligne (« Totaux proposition 2 ») donne les chiffres totaux en fonction du Tableau 12.

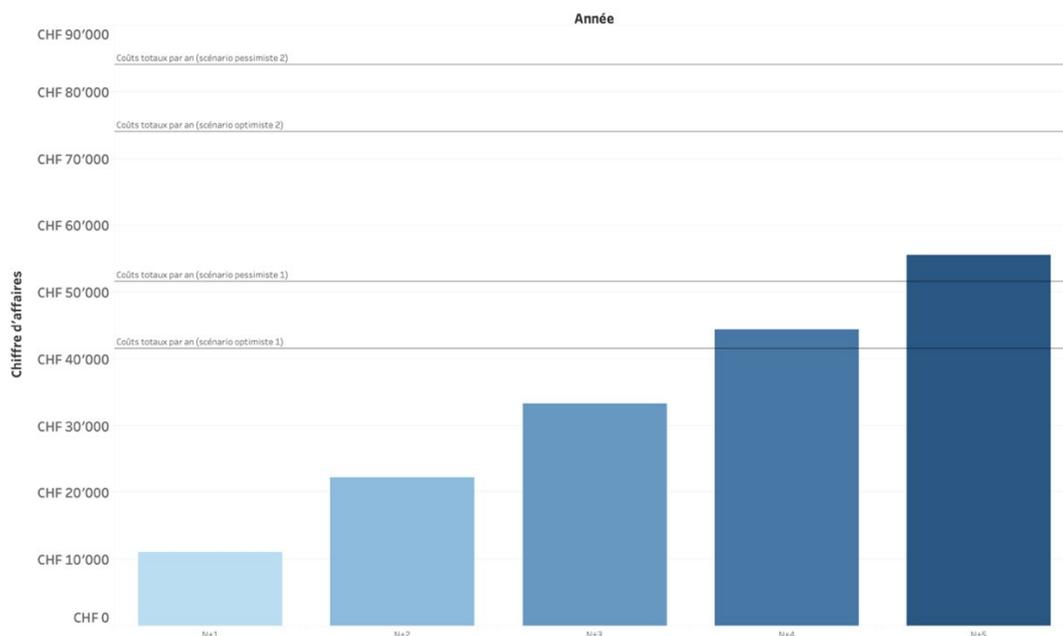
Figure 13 : projection du chiffre d'affaires

	N+1		N+2		N+3		N+4		N+5	
	Clients	CA	Clients	CA	Clients	CA	Clients	CA	Clients	CA
Niveau 1 (fondions de base)	5	CHF 750	10	CHF 1'500	15	CHF 2'250	20	CHF 3'000	25	CHF 3'750
Niveau 1 (fondions avancées)	2	CHF 600	4	CHF 1'200	6	CHF 1'800	8	CHF 2'400	10	CHF 3'000
Niveau 2 (fondions de base)	15	CHF 3'750	30	CHF 7'500	45	CHF 11'250	60	CHF 15'000	75	CHF 18'750
Niveau 2 (fondions avancées)	5	CHF 2'500	10	CHF 5'000	15	CHF 7'500	20	CHF 10'000	25	CHF 12'500
Niveau 3 (fondions de base)	2	CHF 2'000	4	CHF 4'000	6	CHF 6'000	8	CHF 8'000	10	CHF 10'000
Niveau 3 (fondions avancées)	1	CHF 1'500	2	CHF 3'000	3	CHF 4'500	4	CHF 6'000	5	CHF 7'500
Totaux proposition 1 (tous les niveaux)	30	CHF 11'100	60	CHF 22'200	90	CHF 33'300	120	CHF 44'400	150	CHF 55'500
Totaux proposition 2 (niveaux 2 à 3)	23	CHF 9'750	46	CHF 19'500	69	CHF 29'250	92	CHF 39'000	115	CHF 48'750

Le graphique ci-après (cf. Figure 14) ne prend en compte que la première proposition, autrement dit celle de faire payer toutes les institutions clientes, y compris celles de la HES-SO. En abscisse, nous retrouvons le code couleur des années de la Figure 13.

Le seuil de rentabilité du scénario optimiste 1 (CHF 41'600) serait dès lors atteint la quatrième année et celui du scénario pessimiste 1 (CHF 51'600) la cinquième année.

Figure 14 : Projection de rentabilité

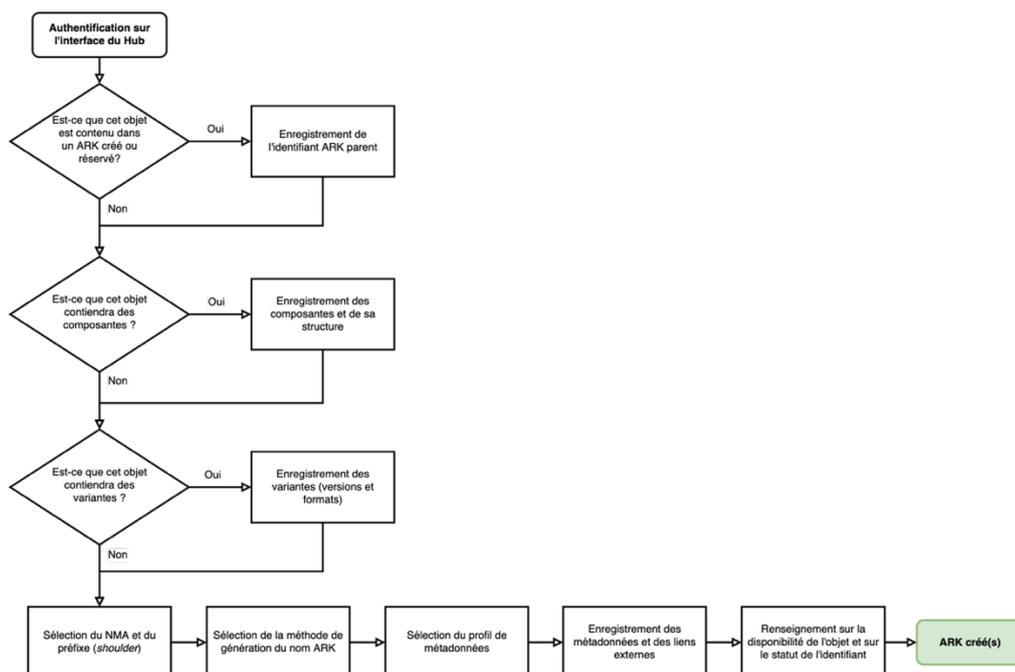


6.3 Processus d'attribution

Pour mieux visualiser les différents processus liés aux services proposés par le Hub, nous avons créé des logigrammes d'attribution d'identifiants. Le premier (cf. Figure 15) montre le processus de création d'identifiants ARK, le second (cf. Figure 16) celui de réservation d'identifiants, et le troisième (cf. Figure 17) celui de modification ou de suppression d'identifiants.

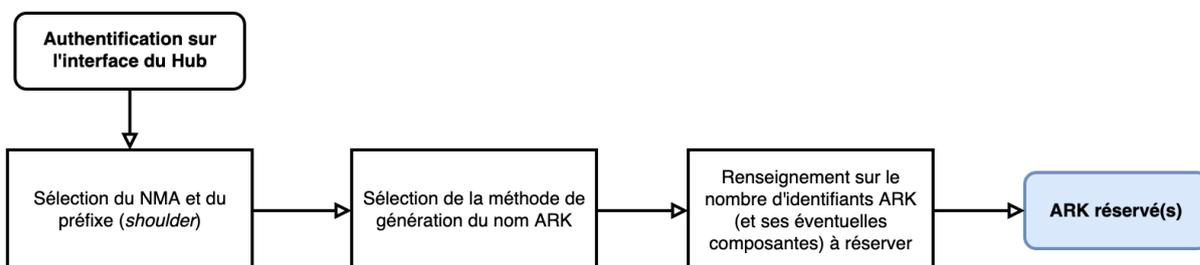
L'étape la plus complexe dans la création d'un ARK est celle relative aux fonctionnalités avancées du service, autrement dit la gestion des qualificatifs de granularité et de service. Toutes les étapes de la Figure 15 sont également illustrées dans une maquette basse fidélité d'une page web du Hub permettant la création d'identifiants (cf. Figure 20 en Annexe 4).

Figure 15 : Logigramme de création d'ARK



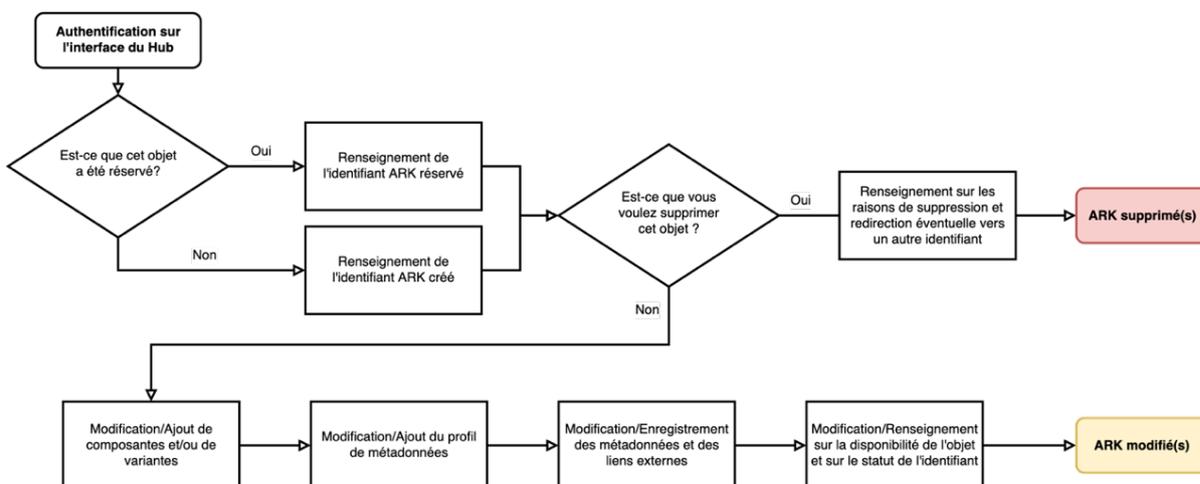
La réservation d'identifiant(s) ARK est l'étape d'attribution la plus simple, surtout si une méthode automatique de génération de chaînes de caractères a été sélectionnée pour générer des noms ARK. Cette fonctionnalité est essentielle pour les clients qui désirent anticiper la mise en ligne de ressources mais également pour celles qui voudraient générer des URL qui correspondent aux ARK réservés.

Figure 16 : Logigramme de réservation d'ARK



Le logigramme suivant répond en quelque sorte à notre deuxième question de recherche sur le moment optimal d'attribution d'un identifiant, car il sera possible de renseigner les raisons du changement et de rediriger éventuellement un ARK vers un nouvel objet. Ces fonctionnalités font du Hub un service dynamique tout en gardant à l'esprit une initiative de traçabilité des ressources dans le temps par le biais notamment des déclarations de permanence.

Figure 17 : Logigramme de modification ou de suppression d'ARK



A noter que les étapes « Sélection du NMA et du préfixe » et « Sélection de la méthode de génération du nom ARK » ne peuvent pas être modifiées *a posteriori* pour des raisons techniques.

Une autre fonctionnalité lors du processus d'attribution pour une institution cliente est sa gestion des autorités d'adressage (NMA) et des préfixes (*shoulder*) sous-jacents. Par exemple, l'ajout par une institution cliente d'un nouveau préfixe pour un ou plusieurs sous-domaines pourrait soit être effectué lors de la création, réservation ou modification d'ARK, soit de manière séparée (comme suggéré par la maquette en Annexe 4).

Une vue d'ensemble de ces logigrammes combinés est disponible en Annexe 5 (cf. Figure 21).

6.4 Blueprint de service

Comme déjà expliqué à la section 4.3.2, le prototypage de notre infrastructure nous a amené à la réalisation d'un Blueprint de service définissant les interactions lorsqu'une institution désire s'inscrire au Hub pour attribuer des identifiants ARK à ses ressources.

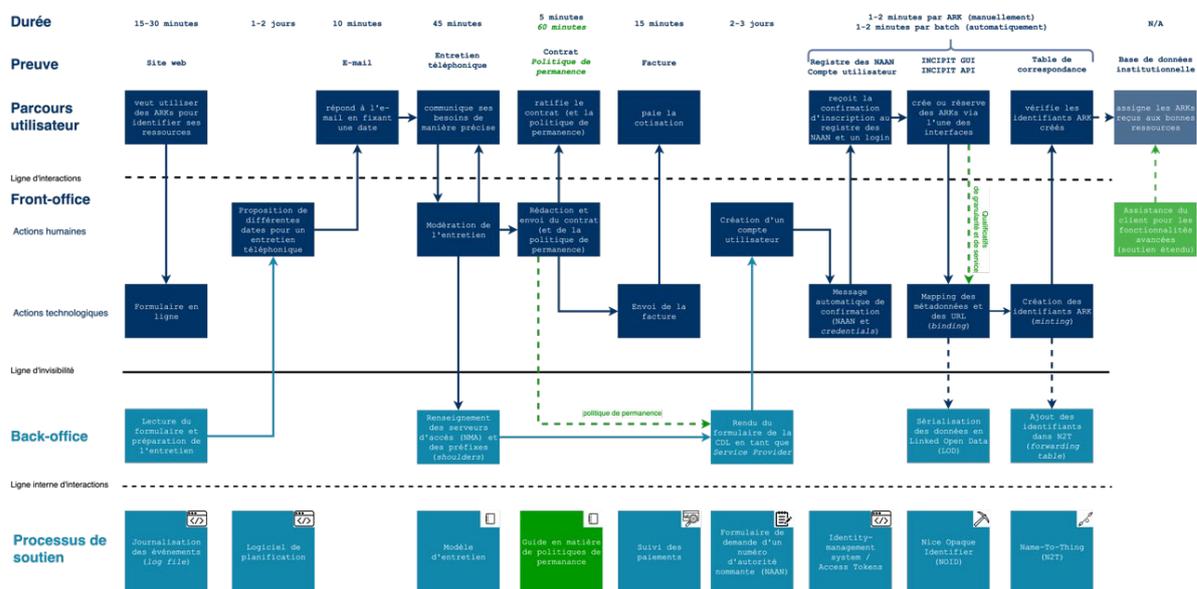
Les six étapes principales suivantes sont exposées dans le Blueprint de service :

1. Une demande d'inscription de la part d'un prospect NAAN via un formulaire de contact ;
2. Une clarification des besoins dans le but de proposer un contrat au futur client ;
3. Une ratification du contrat de la part du client et éventuellement une proposition de politique de permanence ;
4. Une inscription au registre des NAAN et une création d'un compte utilisateur pour se connecter aux interfaces du Hub ;
5. Une génération d'identifiants ARK, comportant potentiellement des qualificatifs de granularité et de service, ayant des métadonnées associées ;
6. Une sérialisation des métadonnées en LOD.

Après l'étape de sérialisation, l'institution cliente devra encore assigner les ARK créés aux bonnes ressources au sein des pages web de destination ainsi que dans sa propre base de données.

Dans la Figure 18 ci-dessous, tous les éléments en bleu représentent les fonctionnalités de base du service et ceux en vert les fonctionnalités avancées.

Figure 18 : Blueprint de service



Une version grand format plus lisible vous est proposée en Annexe 6 (cf. Figure 22).

7. Recommandations

Sur la base des objectifs listés ci-dessous, nous avons pu effectuer des recommandations propres à destination du projet INCIPIT :

- Obtenir des commentaires par des professionnels sur nos études de cas, exposées ici sous forme de récits utilisateurs ;
- Établir des recommandations techniques en matière de logiciels, résolveur de liens, métadonnées et de *mapping* (selon le type d'entité et les métadonnées disponibles), ce qui permet d'établir les fonctionnalités requises ;
- Faire une estimation globale des coûts de création et de fonctionnement afin d'assurer la viabilité financière et organisationnelle du Hub ;
- Établir une liste de critères indispensables nécessaires à la création du Hub correspondant aux facteurs clefs de succès.

Ce chapitre de recommandations est découpé en cinq sections, dont les quatre premières sont issues des points ci-dessus, et la dernière section présente les aspects d'innovation essentiels à tout nouveau service.

Ces recommandations représentent l'avis des auteurs de ce rapport, mais il est évident que les choix que nous proposons devront être analysés plus finement par la suite, dans le cadre du projet INCIPIT.

7.1 Récits utilisateurs

Nous avons décidé de rassembler les commentaires des professionnels et ceux issus de la revue de la littérature sous forme de récits utilisateurs employés fréquemment dans le contexte de projets dans le domaine UX (*User experience*). Ces récits sont énoncés de la manière suivante (Schneider, 2019) :

« *En tant que [rôle], je voudrais <objectif/souhait>, (pour <bénéfice>).* »

Dans le Tableau 13 ci-dessous, chaque récit utilisateur (sur la gauche) a une ou plusieurs conséquences sur l'implémentation de l'infrastructure ARK.

Tableau 13 : Récits utilisateurs et conséquences sur le déploiement du Hub

Récits utilisateurs <i>En tant qu'institution cliente du Hub, je voudrais...</i>	Conséquence(s)
... attribuer des ARK à tous les types d'objets pour avoir une traçabilité des ressources publiées dans un but de préservation numérique.	<ul style="list-style-type: none">• Fonctionnalité « Création d'ARK ».
... créer des ARK avec le profil de métadonnées de mon choix pour répondre aux attentes de mon milieu professionnel et de mes publics cibles.	<ul style="list-style-type: none">• Proposer les schémas de métadonnées les plus usités et laisser les clients proposer d'autres formats d'où l'information peut être facilement extraite.
... créer des ARK rapidement sans (trop de) métadonnées.	<ul style="list-style-type: none">• Ne pas avoir de métadonnées obligatoires (sauf pour le champ où l'URL est renseignée) ;• Avoir une API dédiée.

Récits utilisateurs <i>En tant qu'institution cliente du Hub, je voudrais...</i>	Conséquence(s)
... réserver des ARK avant leur publication pour pouvoir intégrer les permaliens aisément au sein de ma plateforme.	<ul style="list-style-type: none"> Fonctionnalité « Réserve d'ARK ».
... que les métadonnées des identifiants ARK puissent être lues par des humains et des machines.	<ul style="list-style-type: none"> Déployer un mécanisme de négociation de contenu.
... renseigner sur la disponibilité de chaque objet ayant un ARK.	<ul style="list-style-type: none"> Déployer des déclarations de permanence individuelles.
... avoir la possibilité de modifier des ARK.	<ul style="list-style-type: none"> Fonctionnalité « Modification d'ARK ».
... avoir la possibilité de supprimer des ARK dont je n'ai plus besoin.	<ul style="list-style-type: none"> Fonctionnalité « Réserve d'ARK ».
... rediriger des ARK ayant une date de péremption vers un nouvel objet ou vers une page affichant les métadonnées et la raison de suppression.	<ul style="list-style-type: none"> Pouvoir mettre à jour la table de redirection N2T pour des objets individuels.
... représenter la hiérarchie d'un corpus d'objets grâce aux ARK.	<ul style="list-style-type: none"> Implémenter les qualificatifs de granularité et de service.
... réutiliser facilement les métadonnées des ARK créés dans des applications en LOD.	<ul style="list-style-type: none"> Créer des assertions RDF ; Proposer un SPARQL <i>endpoint</i>.
... inscrire de nouveaux sous-domaines dans lesquels seront publiés des objets.	<ul style="list-style-type: none"> Fonctionnalité « Gestion des NMA et des préfixes ».
... rediriger des ARK qui appartiennent désormais à une autre entité.	<ul style="list-style-type: none"> Fonctionnalité « Gestion des NMA et des préfixes ».

Pour répondre aux besoins des institutions clientes et de leurs utilisateurs finaux, le Hub devra avoir une vue agnostique et être flexible sur les questions de métadonnées, de permanence des objets et de sérialisation en LOD d'une partie ou de la totalité des métadonnées associées aux ARK, mais devra cependant toujours renseigner sur les raisons de modification ou de suppression. Cette notion de traçabilité est importante à tous les niveaux :

- Pour le Hub qui sera jugé sur les procédures de redirection ;
- Pour les institutions clientes qui n'apparaîtront pas sérieuses dans un contexte de préservation numérique si des PID renvoient vers des pages HTTP 404 ;
- Pour les utilisateurs finaux dans leur recherche et vérification d'information.

Toutes ces fonctionnalités et leurs spécificités ne devront pas toutes être implémentées lors du projet INCIPIT, mais pourront être déployées soit de manière incrémentale lorsqu'il y a croisement des besoins, soit de manière fortuite pour tester un besoin sous-jacent. Un bon exemple est celui de IIF qui intègre certaines spécifications selon les études de cas soumises par la communauté au sein de chacune de leurs API, mais si cela n'a pas été implémenté par au moins deux organisations différentes, la spécification est directement supprimée lorsqu'une prochaine version de l'API aura été définie (Appleby et al., 2018a ; 2018b).

7.2 Fonctionnalités requises

Pour fonctionner techniquement, le Hub requiert :

- Un logiciel tel que NOID pour générer et résoudre des ARK localement en comptant sur une modification de ce dernier ou sur le développement d'un outil similaire pour que les institutions clientes puissent aisément créer des patterns pour le déploiement des qualificatifs de granularité ;
- Une génération flexible de noms ARK qui peut soit être automatisée et opaque, soit basée sur un système existant comme l'Universally Unique Identifier (UUID) ou encore qui soit faite manuellement tout en gardant un minimum de sécurité au moyen d'un caractère de contrôle qui peut par exemple se calculer avec une opération *modulo* ;
- Une participation politique et financière au consortium N2T pour pouvoir facilement modifier la table de redirection objet par objet ;
- Une interface utilisateur pour promouvoir et documenter les fonctionnalités du Hub ainsi que pour créer, réserver, modifier et supprimer manuellement ou semi automatiquement des identifiants ;
- Une API pour gérer rapidement et de manière automatisée ce qui est possible via l'interface utilisateur ;
- Un système qui puisse proposer voire ingérer une variété de schémas de métadonnées ;
- Un système qui puisse sérialiser les métadonnées pour des applications LOD ;
- Des serveurs ayant un niveau de redondance garantissant l'accès aux données 24h/24.

Un exemple d'infrastructure ARK au niveau national qui est en train de se réaliser est celui effectué par la Bibliothèque nationale du Luxembourg (BnL) qui a par exemple décidé de « [...] publier une documentation exhaustive sur la structure du service de PID, les modèles de données ainsi que les politiques d'affectation et de résolution » (Maurer-Popistașu, 2018). Il serait donc très intéressant de nouer des contacts avec cette institution, soit directement soit par le biais du projet AITO auquel participe la BnL.

Une des tendances sur le Web depuis quelques années, tendance à laquelle le Hub devra veiller, est de le décentraliser dans un modèle pair-à-pair en utilisant des protocoles comme l'*InterPlanetary File System* (IPFS) au lieu d'HTTP (Alam et al., 2016 ; Graber, 2019). Un système ARK pourrait tout à fait fonctionner avec IPFS étant donné que la spécification anticipe un changement de protocole. Les logiciels et interfaces devront par contre s'adapter si une telle transition s'effectuait de manière radicale. Si cette option d'adresser des PID via des protocoles Internet décentralisés a déjà été mentionnée comme une solution envisageable (Golodoniuc et al., 2017 ; Car et al., 2017) ou viable (Sicilia et al., 2019), elle comporte toutefois un risque comme le mentionne David Rosenthal, spécialiste de la préservation numérique qui a initié le programme *Lots of Copies Keep Stuff Safe* (LOCKSS) à Stanford :

« Hash-based names, as used for example in IPFS, are thought to be persistent because they are computed from the content itself and are extremely likely to be unique. [...] on the timescales needed for digital preservation it is inevitable that the chosen hash algorithm will become vulnerable to collisions (as we saw recently with SHA-1) and thus the names will no longer be unique. »

(Rosenthal, 2017)

Afin de s'assurer de la pérennité d'un système d'identification dans un contexte technologique en constante évolution, Car et al. (2017) proposent trois critères indispensables :

- **L'unicité** dans la génération des identifiants pour éviter les heurts ;
- **La responsabilité** dans la gestion des identifiants qui doivent être créés par des « agents » déterminés (institutions ou individus) ;
- **La modification des métadonnées** associées aux identifiants doit être garantie pour permettre aux propriétaires d'affiner l'adressage d'un objet qui inévitablement va changer d'emplacement au fil du temps.

7.3 Viabilité financière et organisationnelle

Pour assurer la viabilité financière et organisationnelle du Hub, nous recommandons fortement de le créer au sein de la HES-SO et d'appliquer la seconde proposition de tarification (cf. Tableau 12) qui, pour rappel, offre la gratuité aux institutions partenaires. Ce « manque à gagner » pourrait être comblé par la perception de cotisations adaptées en fonction du type d'institution.

Créé au sein de la HES-SO, le Hub bénéficierait d'un soutien politique fort, mais il faut également qu'il obtienne le soutien des différents organismes de recherche suisses pour asseoir sa position. Ce serait là l'occasion d'harmoniser au niveau national les recommandations en matière de PID en vue de promouvoir l'Open Science et ses valeurs.

Nous imaginons que d'engager un assistant au sein de la HES-SO pourrait convenir dans un premier temps, mais il faudra alors s'assurer qu'il reste en poste suffisamment longtemps. Idéalement, il faudrait établir un contrat à cinq ans, ou à durée indéterminée, avec une personne d'expérience, ou réfléchir à externaliser certaines compétences auprès d'un organisme externe, comme RERO ou SWITCH. Cela permettrait d'assurer le fonctionnement du Hub, du moins d'un point de vue technique et organisationnel.

Dans la situation peu probable où l'infrastructure ne pourrait se maintenir grâce aux moyens de la HEG-GE, de la HES-SO ou d'un consortium, les parties prenantes du Hub devront actionner une des stratégies de succession élaborée lors du projet INCIPIT afin de mettre à disposition de la communauté la documentation et les outils développés.

7.4 Facteurs clefs de succès

Avant d'estimer tout facteur de succès du Hub, nous devons nous assurer de transmettre le projet à INCIPIT de la meilleure façon possible. Comme ce projet sera mené au sein de la HEG-GE, en collaboration avec l'un des auteurs de ce travail, cela ne devrait pas poser trop de difficultés. Nous prévoyons toutefois de rester aussi disponibles que possible pour répondre aux éventuelles interrogations, que cela soit sur notre méthodologie, nos résultats ou nos recommandations.

Pour permettre au Hub de rencontrer le succès escompté, il faudra être particulièrement attentif à plusieurs facteurs, non seulement lors de son initialisation, mais aussi à long terme, à commencer par la nécessité de bénéficier de compétences solides pour garantir la durabilité technique de l'infrastructure. Cet élément est absolument indispensable puisqu'il est à la base du service proposé par l'infrastructure ARK. Aucune économie de moyens ne pourra être faite à ce niveau : cette durabilité est, au final, la raison d'être et la carte de visite du Hub. C'est pourquoi il faudra impérativement engager du personnel compétent, formé et expérimenté si possible, faisant figure d'expertise en matière de PID.

Un fonctionnement optimal du Hub sera dès lors sa meilleure publicité, mais il faudra bien sûr prévoir un plan marketing pour faire connaître le service. Aller à la rencontre des institutions potentiellement intéressées dans une démarche proactive pour se faire connaître dans un premier temps est une bonne idée, et il serait judicieux de s'appuyer sur la réputation des partenaires du Hub et de ses soutiens politiques pour convaincre les éventuels prospects. Idéalement, il faudrait convaincre les organismes de soutien à la recherche de mettre en avant ce service sur leurs plateformes, voire de rendre obligatoire son utilisation si les chercheurs souhaitent bénéficier de leurs subsides.

Un autre facteur de succès qui nous semble important est de garantir la participation active du Hub aux discussions du projet AITO afin de suivre l'évolution de l'écosystème ARK, ce qui assurerait un accès aux dernières évolutions techniques, permettrait de bénéficier des derniers outils créés, et surtout de s'imposer comme un acteur sérieux et incontournable de la communauté ARK. Un bon positionnement du Hub dans son domaine professionnel d'application lui garantira d'office un certain succès.

7.5 Critères d'innovation

Nous avons identifié, dans ce travail, trois critères de réussite de l'innovation que représente ce projet. En premier lieu, la question de sa désirabilité, puis sa viabilité, et enfin sa faisabilité.

En matière de désirabilité humaine, le projet ICOPAD a déjà démontré le besoin, en Suisse, de bénéficier d'une infrastructure ARK. Il se trouve que les partenaires du projet INCIPIT ont également manifesté un fort intérêt pour la création du Hub.

En ce qui concerne la viabilité économique, même si nous avons effectué une projection du chiffre d'affaires sur cinq ans, il sera nécessaire d'en faire une plus approfondie sur cinq, voire dix ans. Le fait que INCIPIT sera le troisième projet à se pencher sur cette réalisation garantit une certaine durabilité, si ce n'est un intérêt renouvelé pour le sujet. Il faudra néanmoins, pour pallier toute éventualité, trouver des partenaires ou organisations ayant la capacité d'accueillir la structure du Hub si celui-ci ne pouvait, pour une raison ou une autre, plus être géré au sein de l'organisme où il aura été créé. L'estimation que nous avons effectuée à la section 6.2.4.3 permet de se faire une idée des impératifs financiers du Hub pour prétendre à sa viabilité.

Et pour finir, la faisabilité technique : rediriger des URL ou générer des noms ARK n'est pas si difficile, et si offrir une possibilité de granularité adéquate l'est un peu plus, le grand défi réside dans la capacité à gérer une infrastructure disponible à toute heure. C'est pourquoi il faudra impérativement s'assurer du niveau de compétence des personnes engagées, ainsi que de bénéficier de cadres juridiques et organisationnels parfaitement adaptés, prenant en compte le maximum d'éventualités.

8. Conclusion

Ce projet a pour but d'apporter une petite pierre à un édifice potentiellement gigantesque, de par sa portée et ses implications pour la communauté scientifique suisse. Une petite pierre, mais une pierre qui, avec d'autres, établit les fondations d'un grand projet d'envergure nationale. Initié avec ICOPAD, il sera prolongé avec INCIPIT et, espère-t-on, finalisé par la création à proprement dite d'un Hub national d'attribution et de gestion d'identifiants pérennes.

Pensé comme au service de la science, cet outil presque philosophique s'inscrit évidemment pleinement dans les objectifs énoncés par swissuniversities et autres organismes actifs dans le monde de la recherche, de même qu'il participe à « ouvrir » la science, à rendre disponibles les résultats des chercheurs et les données qu'ils produisent au plus grand nombre de citoyens. Les principes FAIR, jamais très loin lorsque l'on parle d'Open Science, sous-tendent bien sûr le développement de ce Hub, puisqu'au final, le but de celui-ci est de rendre plus facilement accessible la connaissance produite par l'humanité, et ceci à tous et à tous les niveaux de la société.

C'est pourquoi il s'agit là d'une véritable opportunité. Grâce au projet INCIPIT qui prendra la relève, il offrira la possibilité aux institutions suisses actives dans la recherche et la science en général de travailler efficacement, en complémentarité, et surtout en accord avec les principes liés au partage et à l'ouverture des données. Ce projet national permettra peut-être d'ouvrir la voie en vue d'une rationalisation des moyens destinés à la recherche, tout en rendant aux citoyens les fruits de celle-ci, qui leur appartiennent. Dans tous les cas, il permettra une meilleure cohésion entre les institutions, voire un meilleur rapport social entre les institutions du patrimoine culturel, le secteur académique et les utilisateurs finaux des données.

Il est certain que la mise en place du Hub permettra d'assurer une offre adéquate en matière de gestion de PID au niveau national, ce qui, comme l'a démontré le projet ICOPAD, n'est actuellement pas le cas. Au vu de l'existant, ce sont les bibliothèques, archives et musées qui bénéficieront en premier lieu de cette infrastructure, car ce sont ces institutions qui, souvent, ont des besoins très spécifiques pour hiérarchiser des objets et y associer divers schémas de métadonnées. Bien entendu, le monde académique y trouvera un énorme bénéfice, particulièrement en raison du dynamisme des identifiants ARK pour créer des objets, parfois limités dans leur disponibilité, mais conservant une traçabilité, un versionnage, voire des informations sur leur persistance ou leur indisponibilité.

Imaginer un tel outil d'envergure nationale peut tomber sous le sens, mais sa réalisation n'en est pas pour autant évidente. Avec ce travail, nous balisons le terrain pour nos successeurs, notamment en ce qui concerne certains choix d'outils ou de formats, mais l'essentiel du travail reste à faire. Ces choix que nous avons faits, nous les avons réfléchis, pesés, et ce n'est pas par hasard que nous recommandons fortement l'utilisation du système ARK : il nous semble être, actuellement, le meilleur choix possible en matière de PID objet. Sa flexibilité, sa compatibilité, les possibilités qu'il offre sont immenses et correspondent exactement aux besoins que nous avons identifiés auprès de potentiels partenaires, voire investisseurs, du Hub. Bien sûr, il reste un gros travail à effectuer avant que celui-ci ne voie le jour et soit opérationnel, mais le travail de réflexion précédant cette étape est sur le bon chemin, entre les mains de professionnels, et nous sommes aujourd'hui certains, convaincus que ce travail n'aura pas été vain.

Si, sur le plan technique, la réalisation de Hub ne pose que peu de défis hormis celui de la garantie de persistance, sur le plan philosophique, sa portée est énorme : en créant une infrastructure nationale, la Suisse se positionnerait en exemple pour le monde de la recherche, telle une influenceuse, porteuse du flambeau de l'Open Science éclairant l'ignorance, terrassant les obscures et redoutées pages HTTP « *404 Not Found* » et mettant fin à la perte ou l'incompatibilité des données. Le bénéfice ne sera dès lors pas que financier, mais aussi social : la communauté scientifique d'abord, mais également la population, tous pourront rendre compte d'une amélioration technique et organisationnelle dans l'accès au savoir.

Ce Hub, c'est une opportunité incroyable de faire rayonner l'expertise scientifique suisse, et donc sa communauté, dans le monde entier, c'est l'opportunité de rassembler des forces aujourd'hui éparses et de toutes les aiguiller dans une même direction, vers un avenir meilleur, plus rationnel et pragmatique.

Bibliographie

ALAM, Sawood, KELLY, Mat et NELSON, Michael L., 2016. InterPlanetary Wayback: The Permanent Web Archive. In : *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries - JCDL '16* [en ligne]. Newark, New Jersey, USA : ACM Press. 2016. p. 273-274. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <http://dl.acm.org/citation.cfm?doid=2910896.2925467>.

APPLEBY, Michael, CRANE, Tom, SANDERSON, Robert, STROOP, Jon et WARNER, Simeon, 2018a. IIIF Design Patterns. In : *International Image Interoperability Framework* [en ligne]. 22 mars 2018. [Consulté le 10 janvier 2020]. Disponible à l'adresse : https://iiif.io/api/annex/notes/design_patterns/.

APPLEBY, Michael, CRANE, Tom, SANDERSON, Robert, STROOP, Jon et WARNER, Simeon, 2018b. IIIF Editorial Process. In : *International Image Interoperability Framework* [en ligne]. 12 novembre 2018. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://iiif.io/community/policy/editorial/>.

BAZZANELLA, Barbara, BORTOLI, Stefano et BOUQUET, Paolo, 2013. Can persistent identifiers be cool? In : *International journal of digital curation*. 14 juin 2013. Vol. 8, n° 1, p. 14-28. DOI [10.2218/ijdc.v8i1.246](https://doi.org/10.2218/ijdc.v8i1.246).

BERMÈS, Emmanuelle, 2006. *Des identifiants pérennes pour les ressources numériques : l'expérience de la BnF* [en ligne]. Paris, France. Bibliothèque nationale de France. [Consulté le 10 janvier 2020]. Disponible à l'adresse : https://web.archive.org/web/20181006042857/http://www.bnf.fr/documents/ark_presentation_bermes_2006.pdf.

BERNERS-LEE, Tim, 1998. Cool URIs don't change. In : *W3C* [en ligne]. 1998. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.w3.org/Provider/Style/URI>.

BN, 2016. Uniform Resource Name URN. In : [en ligne]. 27 octobre 2016. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.nb.admin.ch/snl/fr/home/fachinformationen/e-helvetica/urn.html>.

BNF, 2019a. Identifiants internationaux. In : *Bibliothèque nationale de France* [en ligne]. 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.bnf.fr/fr/centre-d-aide/identifiants-internationaux>.

BNF, 2019b. L'identifiant ARK (Archival Resource Key). In : *Bibliothèque nationale de France* [en ligne]. 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.bnf.fr/fr/lidentifiant-ark-archival-resource-key>.

BROCARD, Martine, 2019. L'autorité scientifique face à l'open access. In : *Revue Hémisphères* [en ligne]. 10 juillet 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://revuehemispheres.com/l-autorite-scientifique-face-a-l-open-access/>.

CAR, Nicholas, GOLODONIUC, Pavel et KLUMP, Jens, 2017. The Challenge of Ensuring Persistency of Identifier Systems in the World of Ever-Changing Technology. In : *Data Science Journal*. 4 avril 2017. Vol. 16, n° 0, p. 13. DOI [10.5334/dsj-2017-013](https://doi.org/10.5334/dsj-2017-013).

CARON, Bertrand, 2018. Préconisations pour l'implémentation d'ARK par les sous-autorités nommantes et autorités d'adressage BnF. In : *Bibliothèque nationale de France* [en ligne]. 12 mars 2018. [Consulté le 10 janvier 2020]. Disponible à l'adresse : https://multimedia-ext.bnf.fr/pdf/ark_preconisations_bnf.pdf.

CARON, Bertrand, HOUSSAYE, Jordan de La, LEDOUX, Thomas, PEYRARD, Sébastien, REECHT, Stéphane et TRAMONI, Jean-Philippe, 2018. ARK is in the Air: ARKs Trending in the French-speaking Area and the BnF's Role in the ARK Story. In : [en ligne]. 21 septembre 2018. [Consulté le 10 janvier 2020]. DOI 10.17605/OSF.IO/R6ZJY. Disponible à l'adresse : <https://osf.io/r6zjy/>.

CDL, 2007. Name-to-Thing (N2T) Resolver. In : *N2T* [en ligne]. 2 novembre 2007. [Consulté le 10 janvier 2020]. Disponible à l'adresse : https://n2t.net/e/n2t_vision.html.

CDL, 2014. The EZID API, Version 2. In : *EZID* [en ligne]. 2014. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://ezid.cdlib.org/doc/apidoc.html>.

CDL, 2019a. Archival Resource Key (ARK) Identifiers. In : *N2T* [en ligne]. 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : https://n2t.net/e/ark_ids.html.

CDL, 2019b. Identifier Basics. In : *EZID* [en ligne]. 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : https://ezid.cdlib.org/learn/id_basics.

CDL, 2019c. Identifier Concepts and Practices at the California Digital Library. In : *EZID* [en ligne]. 2019. [Consulté le 13 décembre 2019]. Disponible à l'adresse : https://ezid.cdlib.org/learn/id_concepts.

CDL, 2019d. Name Assigning Authority Number (NAAN) Registry. In : *N2T* [en ligne]. 22 novembre 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : https://n2t.net/e/pub/naan_registry.txt.

CDL, 2019e. Nice Opaque Identifier (minter and name resolver). In : *N2T* [en ligne]. 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://n2t.net/e/noid.html>.

CDL, 2020. Learn About EZID. In : *EZID* [en ligne]. janvier 2020. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://ezid.cdlib.org/learn/>.

CEVEY, Matthieu et RAEMY, Julien A., 2019a. Swiss PID Hub : création d'un Hub pour la gestion des identifiants pérennes en Suisse. In : *Session des posters scientifiques* [en ligne]. Poster scientifique. Haute école de gestion de Genève. 12 décembre 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://zenodo.org/record/3551008>.

CEVEY, Matthieu et RAEMY, Julien A., 2019b. Un PID pour (enfin ?) les gouverner tous. In : *Recherche d'IdéeS* [en ligne]. 2 septembre 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://campus.hesge.ch/blog-master-is/un-pid-pour-les-gouverner-tous/>.

CEVEY, Matthieu et RAEMY, Julien A., 2020. Ce Hub a plus d'une corde à son arc ! In : *Recherche d'IdéeS* [en ligne]. 28 janvier 2020. [Consulté le 15 janvier 2020]. Disponible à l'adresse : <https://campus.hesge.ch/blog-master-is/ce-hub-a-plus-dune-corde-a-son-arc/>.

CODE CIVIL SUISSE, 2020. RS 210 Code civil suisse du 10 décembre 1907. In : *Confédération suisse* [en ligne]. 1 janvier 2020. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.admin.ch/opc/fr/classified-compilation/19070042/index.html#a60>.

COX, Simon, GUENTHER, Rebecca et RUSCH-FEJA, Diann, 1998. Type Element Working Draft. In : *DCMI* [en ligne]. 23 octobre 1998. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.dublincore.org/specifications/dublin-core/type-element/>.

CREATIVE COMMONS, 2009. CC0 1.0 Universal. In : [en ligne]. 2009. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://creativecommons.org/publicdomain/zero/1.0/>.

CSCS, 2019. Manage PID through an API. In : *Persistent Identifiers (PID) at CSCS* [en ligne]. 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://pid.cscs.ch/access/manage-pid-through-an-api/>.

DANILAK, Radoslav, 2012. System, method, and computer program product for providing data redundancy in a plurality of storage devices [en ligne]. US8090980B2. 3 janvier 2012. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://patents.google.com/patent/US8090980B2/en>. US11/942,623

DE SMAELE, Madeleine, HART, Amy Hatfield, ASHTON, Jan, MARTÍNEZ, Isabel Bernal, DIETIKER, Stefanie, ELLIOTT, Jannean, FENNER, Martin, HUGO, Wim, JAKOBSSON, Stefan, ROY, Sophie, RÜCKNAGEL, Jessika, STARR, Joan, YAHIA, Mohamed et ZOLLY, Lisa, 2018. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.2. In : *DataCite* [en ligne]. 2018. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://doi.org/10.5438/bmjt-bx77>.

DHLAB BASEL, 2019. The Knora Base Ontology. In : *Knora Documentation* [en ligne]. 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://docs.knora.org/paradox/02-knora-ontologies/knora-base.html>.

DIEUDÉ, Aude, 2019. Call 192: Appel projets pour l'année 2020. In : *swissuniversities* [en ligne]. 1 mai 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : http://web.archive.org/web/20191126143928/https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Organisation/SUK-P/SUK_P-2/Call192_20190501__FR.pdf.

DORANUM, 2017. Versionnage des DOI dans Zenodo. In : *Données de la Recherche : Apprentissage Numérique* [en ligne]. 21 juin 2017. [Consulté le 29 novembre 2019]. Disponible à l'adresse : <https://doranum.fr/2017/06/21/versionnage-des-doi-dans-zenodo/>.

DURHAM UNIVERSITY LIBRARY, 2017. ARK URLs. In : *Durham Priory Library Recreated* [en ligne]. 26 mars 2017. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.durhampriory.ac.uk/about-the-project/technology/ark-urls/>.

ESPASANDIN, Kate, JAQUET, Aurélie, LEFORT, Lise et SCHNEIDER, René (dir), 2018. *Panorama et modélisation d'identifiants pérennes pour la création d'identités de confiance* [en ligne]. Mémoire de recherche. Carouge, Suisse : Haute école de gestion de Genève. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://doc.rero.ch/record/309479>.

ETHZ LIBRARY, 2018. Organisational Requirements. In : *DOI-Desk* [en ligne]. 5 juin 2018. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://documentation.library.ethz.ch/display/DOID/Organisational+Requirements>.

ETHZ LIBRARY, 2019. Registering DOIs via Interface. In : *DOI-Desk* [en ligne]. 12 août 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://documentation.library.ethz.ch/display/DOID/Registering+DOIs+via+Interface>.

FERGUSON, Christine, MCENTRYE, Jo, BUNAKOV, Vasily, LAMBERT, Simon, SANDT, Stephanie van der, KOTARSKI, Rachael, STEWART, Sarah, MACEWAN, Andrew, FENNER, Martin, CRUSE, Patricia, HORIK, René van, DOHNA, Tina, KOOP-JACOBSEN, Ketil, SCHINDLER, Uwe et MCCAFFERTY, Siobhan, 2018. D3.1 : *D3.1 Survey of Current PID Services Landscape* [en ligne]. S.I. European Union. [Consulté le 10 janvier 2020]. FREYA : Connected Open Identifiers for Discovery, Access and Use of Research Resources. Disponible à l'adresse : <https://zenodo.org/record/1324296>.

GOLODONIUC, Pavel, CAR, Nicholas J. et KLUMP, Jens, 2017. Distributed Persistent Identifiers System Design. In : *Data Science Journal*. 28 juin 2017. Vol. 16, n° 0, p. 34. DOI [10.5334/dsj-2017-034](https://doi.org/10.5334/dsj-2017-034).

GRABER, Jay, 2019. Comparing IPFS and Dat. In : *Medium* [en ligne]. 26 décembre 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://medium.com/@graber.jay/comparing-ipfs-and-dat-8f3891d3a603>.

HARRIS, Steve et SEABORNE, Andy, 2013. SPARQL 1.1 Query Language. W3C Recommendation. In : *W3C* [en ligne]. 21 mars 2013. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.w3.org/TR/sparql11-query/>.

HEG-GE, 2012. Prestations de service. In : *Haute école de gestion de Genève* [en ligne]. 24 octobre 2012. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.hesge.ch/heg/prestations-service>.

HES-SO, 2019. Comment publier en Open Access ? In : *Open HES-SO* [en ligne]. 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://openscience.hes-so.ch/fr/comment-publier-open-access-14504.html>.

HIRSCHMANN, Barbara, 2019. 10 Years DOI Desk at ETH Zurich. In : *Persistent Identifiers in Research – Celebrating 10 Years of DOI Desk at ETH Zurich* [en ligne]. Eidgenössische Technische Hochschule Zürich : ETH Zurich. 13 septembre 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.research-collection.ethz.ch/handle/20.500.11850/365054>.

JONES, Shawn M., SOMPEL, Herbert Van de, SHANKAR, Harihar, KLEIN, Martin, TOBIN, Richard et GROVER, Claire, 2016. Scholarly Context Adrift: Three out of Four URI References Lead to Changed Content. In : *PLOS ONE*. 2 décembre 2016. Vol. 11, n° 12, p. e0167475. DOI [10.1371/journal.pone.0167475](https://doi.org/10.1371/journal.pone.0167475).

KUNZE, John, 2018. Ten persistent myths about persistent identifiers. In : *UC Office of the President: California Digital Library* [en ligne]. 24 août 2018. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://n2t.net/ark:/13030/c7gb1xh09>.

KUNZE, John, 2019a. ARKs @ CDL Architecture. In : *ARKs in the Open Project* [en ligne]. 15 mars 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://wiki.lyrasis.org/display/ARKs/ARKs+%40+CDL+Architecture>.

KUNZE, John, 2019b. ARKs @ CDL cost estimates. In : *ARKs in the Open Project* [en ligne]. 15 mars 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://wiki.lyrasis.org/display/ARKs/ARKs+%40+CDL+cost+estimates>.

KUNZE, John, 2019c. ARKs Interest Survey, 10 minutes. In : *ARKs in the Open Project* [en ligne]. 16 septembre 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://wiki.lyrasis.org/display/ARKs/ARKs+Interest+Survey%2C+10+minutes>.

KUNZE, John et BERMÈS, Emmanuelle, 2019. The ARK Identifier Scheme. In : *Internet Engineering Task Force* [en ligne]. 22 juin 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://tools.ietf.org/html/draft-kunze-ark-22>.

KUNZE, John, CALVERT, Scout, DEBARRY, Jeremy D., HANLON, Matthew, JANÉE, Greg et SWEAT, Sandra, 2017. Persistence Statements: Describing Digital Stickiness. In : *Data Science Journal*. 14 août 2017. Vol. 16, n° 0, p. 39. DOI [10.5334/dsj-2017-039](https://doi.org/10.5334/dsj-2017-039).

KUNZE, John et CARON, Bertrand, 2019a. ARK Identifiers FAQ. In : *ARKs in the Open Project* [en ligne]. 9 décembre 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://wiki.lyrasis.org/display/ARKs/ARK+Identifiers+FAQ>.

KUNZE, John et CARON, Bertrand, 2019b. FAQ sur les identifiants ARK. In : *ARKs in the Open Project* [en ligne]. 9 décembre 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://wiki.lyrasis.org/pages/viewpage.action?pageId=178880619>.

KUNZE, John, KAHLE, Brewster, MASANES, Julien et MOHR, Gordon, 2006. A Name-Value Language (ANVL). In : *Internet Engineering Task Force* [en ligne]. 28 février 2006. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://tools.ietf.org/search/draft-kunze-anvl-02>.

KUNZE, John et TURNER, Adrian, 2010. DCMI Kernel Metadata Community: Kernel Metadata and Electronic Resource Citations (ERCs). In : *Dublin Core Metadata Initiative* [en ligne]. 13 mai 2010. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.dublincore.org/groups/kernel/spec/>.

LA TRIBUNE DES ARCHIVISTES, 2018. Choisir des URL persistantes pour la mise en ligne de sa base de données : ARK pas à pas... In : *La Tribune des Archivistes* [en ligne]. 21 octobre 2018. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <http://latribunedesarchives.blogspot.com/2018/10/choisir-des-url-persistantes-pour-la.html>.

LAVOIE, Brian, CHILDRESS, Eric R., ERWAY, Ricky, FANIEL, Ixchel M., MALPAS, Constance, SCHAFFNER, Jennifer et VAN DER WERF, Titia, 2014. The Evolving Scholarly Record. In : [en ligne]. 5 juin 2014. [Consulté le 10 janvier 2010]. Disponible à l'adresse : <https://www.oclc.org/research/publications/2014/oclcresearch-evolving-scholarly-record-2014-overview.html>.

MAURER-POPISTAȘU, Roxana, 2018. What, who, where, how? Persistent Identifiers at the National Library of Luxembourg - Digital Preservation Coalition. In : *Digital Preservation Coalition* [en ligne]. 21 novembre 2018. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://dpconline.org/blog/idpd/what-who-where-how>.

MCMURRY, Julie A., JUTY, Nick, BLOMBERG, Niklas, BURDETT, Tony, CONLIN, Tom, CONTE, Nathalie, COURTOT, Mélanie, DECK, John, DUMONTIER, Michel, FELLOWS, Donal K., GONZALEZ-BELTRAN, Alejandra, GORMANNS, Philipp, GRETHE, Jeffrey, HASTINGS, Janna, HÉRICHÉ, Jean-Karim, HERMJAKOB, Henning, ISON, Jon C., JIMENEZ, Rafael C., JUPP, Simon, KUNZE, John, LAIBE, Camille, NOVÈRE, Nicolas Le, MALONE, James, MARTIN, Maria Jesus, MCENTYRE, Johanna R., MORRIS, Chris, MUILU, Juha, MÜLLER, Wolfgang, ROCCA-SERRA, Philippe, SANSONE, Susanna-Assunta, SARIYAR, Murat, SNOEP, Jacky L., SOILAND-REYES, Stian, STANFORD, Natalie J., SWAINSTON, Neil, WASHINGTON, Nicole, WILLIAMS, Alan R., WIMALARATNE, Sarala M., WINFREE, Lilly M., WOLSTENCROFT, Katherine, GOBLE, Carole, MUNGALL, Christopher J., HAENDEL, Melissa A. et PARKINSON, Helen, 2017. Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. In : *PLOS Biology*. 29 juin 2017. Vol. 15, n° 6, p. e2001414. DOI 10.1371/journal.pbio.2001414.

MEADOWS, Alice, 2017. PIDapalooza – the open festival for persistent identifiers. In : *Insights*. 8 novembre 2017. Vol. 30, n° 3, p. 161-164. DOI [10.1629/uksg.393](https://doi.org/10.1629/uksg.393).

MOZILLA DEVELOPER NETWORK, 2019. Redirections in HTTP. In : *MDN Web Docs* [en ligne]. 19 novembre 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://developer.mozilla.org/en-US/docs/Web/HTTP/Redirections>.

NICHOLAS, Nick, WARD, Nigel et BLINCO, Kerry, 2009. A policy checklist for enabling persistence of identifiers. In : *D-Lib magazine* [en ligne]. janvier 2009. Vol. 15, n° 1/2. [Consulté le 10 janvier 2020]. DOI 10.1045/january2009-nicholas. Disponible à l'adresse : <http://www.dlib.org/dlib/january09/nicholas/01nicholas.html>.

PF PDT, 2019. Transmission à l'étranger. In : *Confédération suisse* [en ligne]. 2019. [Consulté le 9 janvier 2020]. Disponible à l'adresse : <https://www.edoeb.admin.ch/edoeb/fr/home/datenschutz/handel-und-wirtschaft/uebermittlung-ins-ausland.html>.

PRONGUÉ, Nicolas et RAEMY, Julien A., 2017. *Revue de la littérature : identifiants pérennes (PID), Linked Data, Données de la recherche* [en ligne]. Carouge, Suisse. Haute école de gestion de Genève. [Consulté le 10 janvier 2020]. Disponible à l'adresse : https://campus.hesge.ch/id_bilingue/projekte/icopad/doc/Prongue_Raemy_Revue_Litterature_2017.pdf.

PRONGUÉ, Nicolas, RICCI, Fabio, SCHNEIDER, René et SCHURTE, René, 2017. Art and design as linked data: the LODZ project (Linked Open Data Zurich). In : *Libellarium: journal for the research of writing, books, and cultural heritage institutions* [en ligne]. 2 mars 2017. Vol. 9, n° 2. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <http://www.libellarium.org/index.php/libellarium/article/view/256>.

RAEMY, Julien A., 2018. *Identifiants pérennes (PID) : Processus d'obtention, mapping et approches d'attribution, modélisation, glossaire* [en ligne]. Carouge, Suisse. Haute école de gestion de Genève. [Consulté le 10 janvier 2020]. Disponible à l'adresse : https://campus.hesge.ch/id_bilingue/projekte/icopad/doc/Raemy_PID_Processus_Approches_Modelisation_2018.pdf.

RAEMY, Julien A., FORNARO, Peter et ROSENTHALER, Lukas, 2017. Implementing a Video Framework based on IIIF: A Customized Approach from Long-Term Preservation Video Formats to Conversion on Demand. In : *Archiving Conference*. 15 mai 2017. Vol. 2017, n° 1, p. 68-73. DOI [10.2352/issn.2168-3204.2017.1.0.68](https://doi.org/10.2352/issn.2168-3204.2017.1.0.68).

RAEMY, Julien A. et SCHNEIDER, René, 2019a. Assigning persistent identifiers to art and design entities. In : *Fourth Swiss Congress for Art History* [en ligne]. Mendrisio. 7 juin 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://zenodo.org/record/3241453>.

RAEMY, Julien A. et SCHNEIDER, René, 2019b. *Suggested measures for deploying IIIF in Swiss cultural heritage institutions (White paper)* [en ligne]. White paper. Genève. Haute école de gestion de Genève. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://zenodo.org/record/2640416>.

RAEMY, Julien A. and SCHNEIDER, René Martin, 2019c. Towards Trusted Identities for Swiss Researchers and their Data. *International Journal of Digital Curation*. 11 septembre 2019. Vol. 14, no. 1, p. 303–314. DOI [10.2218/ijdc.v14i1.596](https://doi.org/10.2218/ijdc.v14i1.596).

RÉPUBLIQUE ET CANTON DE GENÈVE, 2013. Loi sur la Haute école spécialisée de Suisse occidentale – Genève (LHES-SO-GE). C 1 26. In : *République et Canton de Genève* [en ligne]. 29 août 2013. [Consulté le 10 janvier 2020]. Disponible à l'adresse : https://www.ge.ch/legislation/rsg/f/s/rsg_c1_26.html.

ROSENTHAL, David, 2017. Bad Identifiers. In : *DSHR's Blog* [en ligne]. 19 décembre 2017. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://blog.dshr.org/2017/12/bad-identifiers.html>.

SAUERMANN, Leo et CYGANIAK, Richard, 2008. Cool URIs for the Semantic Web: W3C Interest Group Note. In : *W3C* [en ligne]. 3 décembre 2008. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.w3.org/TR/cooluris/>.

SCHNEIDER, René, 2019. Introduction au UCD. In : *Cours de User-centred design*. Support de cours Bachelor en Information documentaire. Haute école de gestion de Genève. 20 février 2019.

SCHNEIDER, René et RAEMY, Julien A., 2019a. Résultats du projet ICOPAD. In : *ID Bilingue* [en ligne]. février 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : https://campus.hesge.ch/id_bilingue/projekte/icopad/results_fr.html.

SCHNEIDER, René et RAEMY, Julien A., 2019b. Towards Trusted Identities for Swiss Researchers and their Data. In : *14th International Digital Curation Conference (IDCC)* [en ligne]. Melbourne, Australia. 6 février 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://doi.org/10.5281/zenodo.2415995>.

SICILIA, Miguel-Angel, GARCÍA-BARRIOCANAL, Elena, SÁNCHEZ-ALONSO, Salvador et CUADRADO, Juan-Jose, 2019. Decentralized Persistent Identifiers: a basic model for immutable handlers. In : *Procedia Computer Science*. 1 janvier 2019. Vol. 146, p. 123-130. DOI [10.1016/j.procs.2019.01.087](https://doi.org/10.1016/j.procs.2019.01.087).

SIMONS, Natasha, 2012. Implementing DOIs for research data. In : *D-Lib magazine* [en ligne]. mai 2012. Vol. 18, n° 5/6. [Consulté le 10 janvier 2020]. DOI doi:10.1045/may2012-simons. Disponible à l'adresse : <http://www.dlib.org/dlib/may12/simons/05simons.html>.

SNYDMAN, Stuart, SANDERSON, Robert et CRAMER, Tom, 2015. The International Image Interoperability Framework (IIIF): A community & technology approach for web-based images. In : *Archiving Conference* [en ligne]. Los Angeles, CA : IS&T. mai 2015. p. 16-21. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://purl.stanford.edu/df650pk4327>.

SWISSUNIVERSITIES, 2019a. P-5 Information scientifique. In : *swissuniversities* [en ligne]. 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://www.swissuniversities.ch/fr/themes/digitalisation/p-5-information-scientifique>.

SWISSUNIVERSITIES, 2019b. Project applications of the eleventh call for proposals (August 2019). In : *swissuniversities* [en ligne]. 7 novembre 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : http://web.archive.org/web/20191108083241/https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Organisation/SUK-P/SUK_P-2/ProjectApplications_August2019-1.pdf.

TRELOAR, Andrew et KLUMP, Jens, 2019. Updating the Data Curation Continuum. In : *International Journal of Digital Curation*. 11 septembre 2019. Vol. 14, p. 87-101. DOI [10.2218/ijdc.v14i1.643](https://doi.org/10.2218/ijdc.v14i1.643).

TRELOAR, Andrew, SCHARNHORST, Andrea et VAN DE SOMPEL, Herbert, 2015. Recording to Archiving Architecture. In : *andrew.treloar.net* [en ligne]. 2015. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://andrew.treloar.net/research/diagrams/recording-to-archiving-architecture.jpg>.

TRIPP, Erin et KUNZE, John, 2019. Project Vision. In : *ARKs in the Open Project* [en ligne]. 1 février 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://wiki.lyrasis.org/display/ARKs/Project+Vision>.

VAN DE SOMPEL, Herbert, SANDERSON, Robert, SHANKAR, Harihar et KLEIN, Martin, 2014. Persistent Identifiers for Scholarly Assets and the Web: The Need for an Unambiguous Mapping. In : *International Journal of Digital Curation* [en ligne]. 23 juillet 2014. Vol. 9, n° 1. [Consulté le 10 janvier 2020]. DOI [10.2218/ijdc.v9i1.320](https://doi.org/10.2218/ijdc.v9i1.320). Disponible à l'adresse : <http://www.ijdc.net/index.php/ijdc/article/view/320>.

WILKINSON, Mark D., DUMONTIER, Michel, AALBERSBERG, IJsbrand Jan, APPLETON, Gabrielle, AXTON, Myles, BAAK, Arie, BLOMBERG, Niklas, BOITEN, Jan-Willem, DA SILVA SANTOS, Luiz Bonino, BOURNE, Philip E., BOUWMAN, Jildau, BROOKES, Anthony J., CLARK, Tim, CROSAS, Mercè, DILLO, Ingrid, DUMON, Olivier, EDMUNDS, Scott, EVELO, Chris T., FINKERS, Richard, GONZALEZ-BELTRAN, Alejandra, GRAY, Alasdair J.G., GROTH, Paul, GOBLE, Carole, GRETHE, Jeffrey S., HERINGA, Jaap, 'T HOEN, Peter A.C, HOOFT, Rob, KUHN, Tobias, KOK, Ruben, KOK, Joost, LUSHER, Scott J., MARTONE, Maryann E., MONS, Albert, PACKER, Abel L., PERSSON, Bengt, ROCCA-SERRA, Philippe, ROOS, Marco, VAN SCHAİK, Rene, SANSONE, Susanna-Assunta, SCHULTES, Erik, SENGSTAG, Thierry, SLATER, Ted, STRAWN, George, SWERTZ, Morris A., THOMPSON, Mark, VAN DER LEI, Johan, VAN MULLIGEN, Erik, VELTEROP, Jan, WAAGMEESTER, Andra, WITTENBURG, Peter, WOLSTENCROFT, Katherine, ZHAO, Jun et MONS, Barend, 2016. The FAIR Guiding Principles for scientific data management and stewardship. In : *Scientific Data*. 15 mars 2016. Vol. 3, p. 160018. DOI 10.1038/sdata.2016.18.

WIMALARATNE, Sara et FENNER, Martin, 2018. D2.1 : *D2.1 PID Resolution Services Best Practices* [en ligne]. S.I. European Union. [Consulté le 10 janvier 2020]. FREYA : Connected Open Identifiers for Discovery, Access and Use of Research Resources. Disponible à l'adresse : <https://zenodo.org/record/1324300>.

ZENODO, 2019. Frequently Asked Questions. In : *Zenodo* [en ligne]. 2019. [Consulté le 10 janvier 2020]. Disponible à l'adresse : <https://help.zenodo.org/>.

Annexe 1 : *Ten persistent myths about persistent identifiers*

Ci-dessous se trouvent la liste complète des « dix mythes persistants sur les identifiants pérennes » réalisée par John Kunze (2018). Celle-ci a été tout d'abord publiée sur Twitter³⁵ et est également disponible dans une traduction française réalisée par Bertrand Caron³⁶, collaborateur travaillant au Département des métadonnées à la BnF.

- *Myth 1: PIDs guarantee access. Get real. All PID services run on evolving software/hardware that no vendor warranties. How could anyone, let alone poor non-profits, change that?*
- *Myth 2: PIDs rarely break. Nonsense. Millions of PIDs are broken. Updating redirection tables is real work for you and your successors. (You do have a succession plan, right?)*
- *Myth 3: PIDs must not be URLs. What a crock. PIDs not carried inside clickable URLs are irrelevant or only of academic interest.*
- *Myth 4: PIDs aren't vanity URLs. Nope. Tons of PIDs contain organizational acronyms, making them nearly as fragile as vanity URLs.*
- *Myth 5: PIDs, unlike URLs, are not locations. Wrong. No URL or clickable PID is a location. No expert can look at it and tell where the content is assembled, via what redirect chain, DNS routing tables, proxies, etc.*
- *Myth 6: PIDs aren't needed, so just use "Cool URLs". Sorry. How do you know a URL is cool? The average URL breaks in 44 days, and most URLs shouldn't persist. Assigning a PID tells people there's hope.*
- *Myth 7: PID resolver technology is hard. Fiddlesticks. Global resolvers have always been simple: table lookup plus HTTP redirection plus \$20 a year to rent a hostname. Every URL shortener is an identifier resolver.*
- *Myth 8: PIDs require vendor lock in. Poppycock. No database system discriminates among identifier types unless directed to discriminate.*
- *Myth 9: PIDs must be centralized. False. Any PID with a globally unique core after the "carrier" URL hostname is persistable. In fact if it cannot be served by other hosts, it cannot persist. No hostname or protocol lasts forever.*
- *Myth 10: PIDs should be free. No. While you can choose a PID that avoids fees and vendor lock in, every PID is a service commitment involving at least some sweat equity.*

³⁵ <https://twitter.com/jakkbl/status/1033059109346586624> (consulté le 10 janvier 2020)

³⁶ <https://framalistes.org/sympa/arc/arks-forum-fr/2018-09/msg00000.html> (consulté le 10 janvier 2020)

Annexe 2 : Poster scientifique

La Figure 19 ci-dessous a été réalisée dans le cadre de la session des posters scientifiques du jeudi 12 décembre 2019 à la HEG-GE³⁷.

Figure 19 : Poster scientifique du Swiss PID Hub



h e g
Haute école de gestion
Genève

Swiss PID Hub

Création d'un Hub pour la gestion des identifiants pérennes en Suisse



Hes-SO
Haute école spécialisée
de Suisse occidentale

Objectifs du projet

- Créer** un cadre conceptuel complémentaire aux DOI permettant d'attribuer des *Archival Resource Keys* (ARK) à tous les types de données
- Modéliser** un prototype de Hub
- Établir** des recommandations pour la création d'un Hub

Objectifs du Hub

- Être** un centre d'expertise sur les identifiants pérennes (PID)
- Assurer** l'accès aux connaissances
- Réduire** les coûts de gestion
- Collaborer** au sein de l'initiative mondiale *ARKs-in-the-Open* (AITO)

Anatomie de l'identifiant ARK

Service de résolution (Resolver Service) | Nom de base de l'objet (Base Object Name)

<https://example.org/ark:/12345/bor73502194t/f42.jpg>

Protocole d'accès	Serveur d'accès de l'autorité d'adressage (NMA)	Schéma	Numéro d'autorité nommante (NAAN)	Nom ARK (préfixe, numéro d'identifiant, caractère de contrôle)	Qualificatifs
-------------------	---	--------	-----------------------------------	--	---------------

Métadonnées ERC

Electronic Resource Citation (ERC)
<https://n2t.net/ark:/12345/bor73502194t/f42.jpg>
 who: Cevey, Matthieu ; Raemy, Julien A.
 what: Swiss PID Hub
 when: 2019.12.12
 where: <https://example.org/ark:/12345/bor73502194t/f42.jpg>

Obtention des NAAN via le registre ARK tenu par la California Digital Library 

Processus de création d'ARK

- Fonctionnalités de base** : étapes 1 / 2 / 3
Fonctionnalités avancées : étape 4
- Initialisation** : 1.65 EPT (3000 h.)
Maintenance : 0.2-0.5 EPT (360-900 h.)
- Contrat adapté selon la forme juridique de l'organisation et ses buts



Outils du Hub

- Interface utilisateur (UI) et interface de programmation applicative (API) 
- Nice Opaque Identifier (NOID) Minter and Name resolver 
- Name-To-Thing (N2T) Global ARK resolver 



- 4** Qualificatifs de granularité et de service
- 3** Nom ARK selon le type d'entité
- 2** Métadonnées ERC et création d'inférences en Linked Open Data (LOD)
- 1** Numéro d'autorité nommante (NAAN) et serveur d'accès de l'autorité d'adressage (NMA)

MATTHIEU CEVEY ET JULIEN A. RAEMY
 matthieu.cevey@hesge.ch | <https://orcid.org/0000-0003-3602-4584>
 julien.raemy@hesge.ch | <https://orcid.org/0000-0002-4711-5759>

SESSION DES POSTERS SCIENTIFIQUES - 12 DÉCEMBRE 2019
 Projet de recherche sous la direction du Prof. Dr. René Schneider
 Master en Sciences de l'Information (2018-2020)



DOI: 10.5281/zenodo.3551008

(Cevey, Raemy, 2019a)

³⁷ NB : A défaut d'avoir un identifiant ARK, ce qui devrait être bientôt possible en Suisse via le Hub, notre poster scientifique a un DOI (<https://doi.org/10.5281/zenodo.3551008>).

Annexe 3 : Analyse détaillée des études de cas

Ci-dessous, vous trouverez une analyse détaillée des études de cas mises en avant dans la Figure 8.

Tableau 14 : Analyse détaillée des études de cas

Cas	Identifiant	Métadonnées	Liens externes	LOD	Disponibilité de l'objet
01	DOI	Mixtes (éléments du DC)	Wikipédia	Manifeste IIF (JSON-LD)	Permanente
02	DOI	Mixtes (éléments du DC)	GND, Wikipédia	Manifeste IIF (JSON-LD)	Permanente
03	URL	Personnalisées	-	-	Supposée permanente
04	URL	Personnalisées	-	-	Supposée permanente
05	URL	Personnalisées	-	-	Indéterminée
06	URL	Personnalisées	-	-	Indéterminée
07	URL	Personnalisées	-	-	Indéterminée
08	URL	Personnalisées	GND, Dictionnaire historique de la Suisse (DHS), european-art.net	-	Indéterminée
09	ARK	Standardisées (DC)	-	Manifeste IIF (JSON-LD)	Permanente
10	ARK (via N2T)	Mixtes (éléments du DC)	-	-	Supposée permanente
11	ARK	Standardisées (DC, Fabio, FOAF)	-	Graphe RDF	Supposée permanente
12	ARK (via N2T)	Personnalisées	-	-	Supposée permanente
13	ARK (via N2T)	Mixtes (éléments du DC)	-	-	Supposée permanente
14	ARK	Personnalisées	Wikipédia, VIAF, LOC, WorldCat	JSON-LD	Supposée permanente
15	ARK (via N2T)	Mixtes (éléments du DC)	-	JSON-LD Turtle	Supposée permanente
16	URL	Standardisées (MARC, MARCXML, DC)	-	-	Indéterminée

Cas	Identifiant	Métadonnées	Liens externes	LOD	Disponibilité de l'objet
17	URL	Standardisées (MARC, MARCXML, DC)	-	-	Indéterminée
18	URL	Standardisées (MARC, MARCXML, DC)	-	-	Indéterminée
19	URL + ISBN	Standardisées (MARC, MARCXML, DC)	-	-	Indéterminée
20	URL + DOI externe	Standardisées (MARC, MARCXML, DC)	-	-	Indéterminée
21	URL	Standardisées (MARC, MARCXML, DC)	-	-	Indéterminée
22	URL + Handle externe	Standardisées (MARC, MARCXML, DC)	-	-	Indéterminée

Annexe 4 : Maquette de l'interface utilisateur

Figure 20 : Maquette basse fidélité de la page web « Création d'ARK »

Mon navigateur préféré

← → ↻ <https://swisspidhub.ch/my-account>

Création d'ARK	Réservation d'ARK	Modification/Suppression d'ARK	Gestion des NMA et des préfixes	Paramètres du compte
----------------	-------------------	--------------------------------	---------------------------------	----------------------

Qualificatifs de granularité et de service

Est-ce que cet objet est contenu dans un ARK déjà créé ou réservé ? Oui Non

Est-ce que cet objet contiendra des composantes ? Oui Non

Est-ce que cet objet aura des variantes ? Oui Non

Sous-domaine et nom ARK

Autorité d'adressage (NMA)

Préfixe (*shoulder*)

Méthode de génération du nom ARK

Métadonnées et liens externes

Where

Who

What

When

Déclaration de permanence et statut de l'identifiant

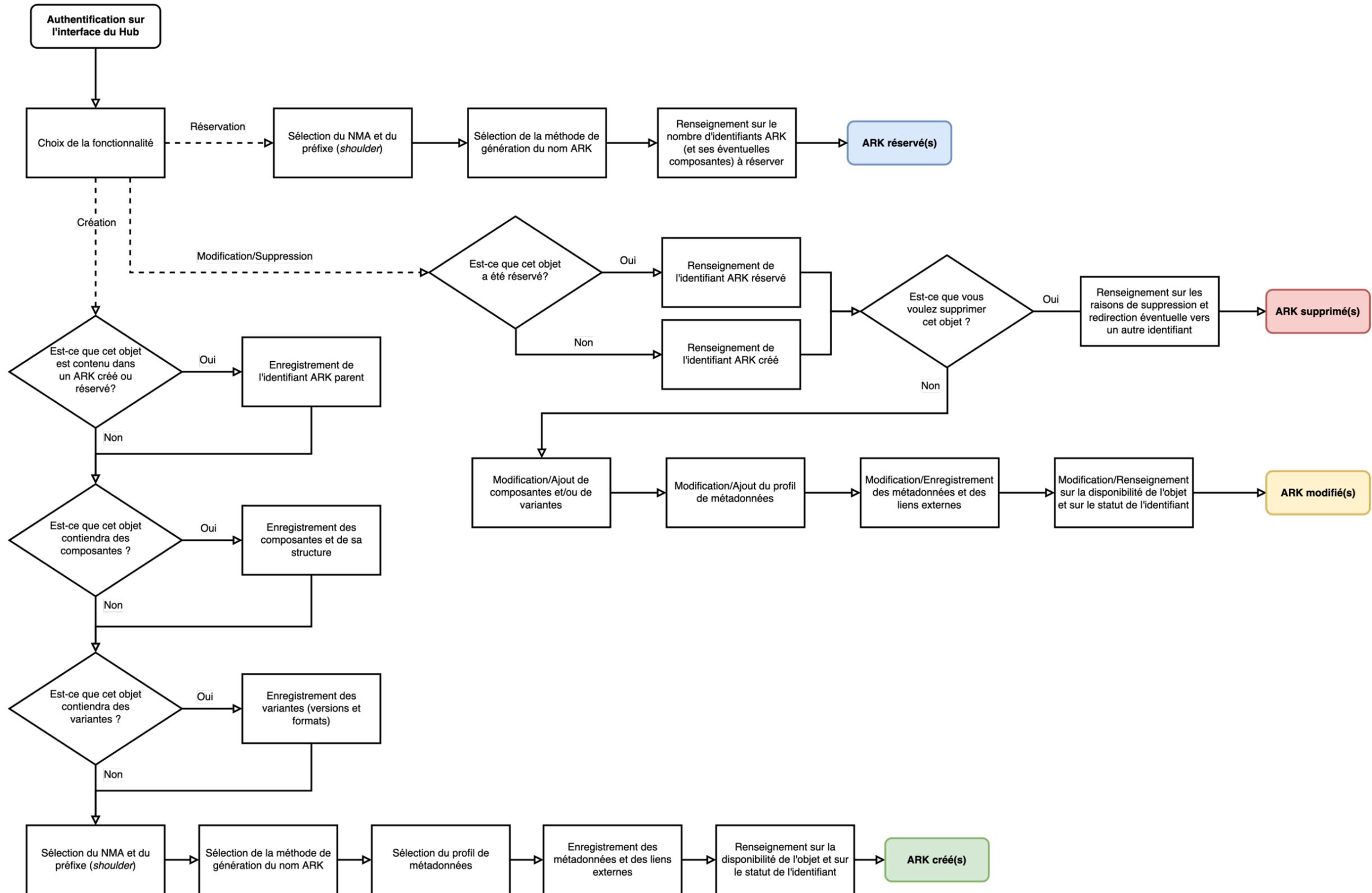
Disponibilité de l'objet

Statut

Autorisation d'indexer et de moissonner l'identifiant Oui Non

Annexe 5 : Vue d'ensemble des processus d'attribution

Figure 21 : Logigramme d'attribution d'identifiants ARK (grand format)



Annexe 6 : Blueprint de service (format paysage)

Figure 22 : Blueprint de service (grand format)

