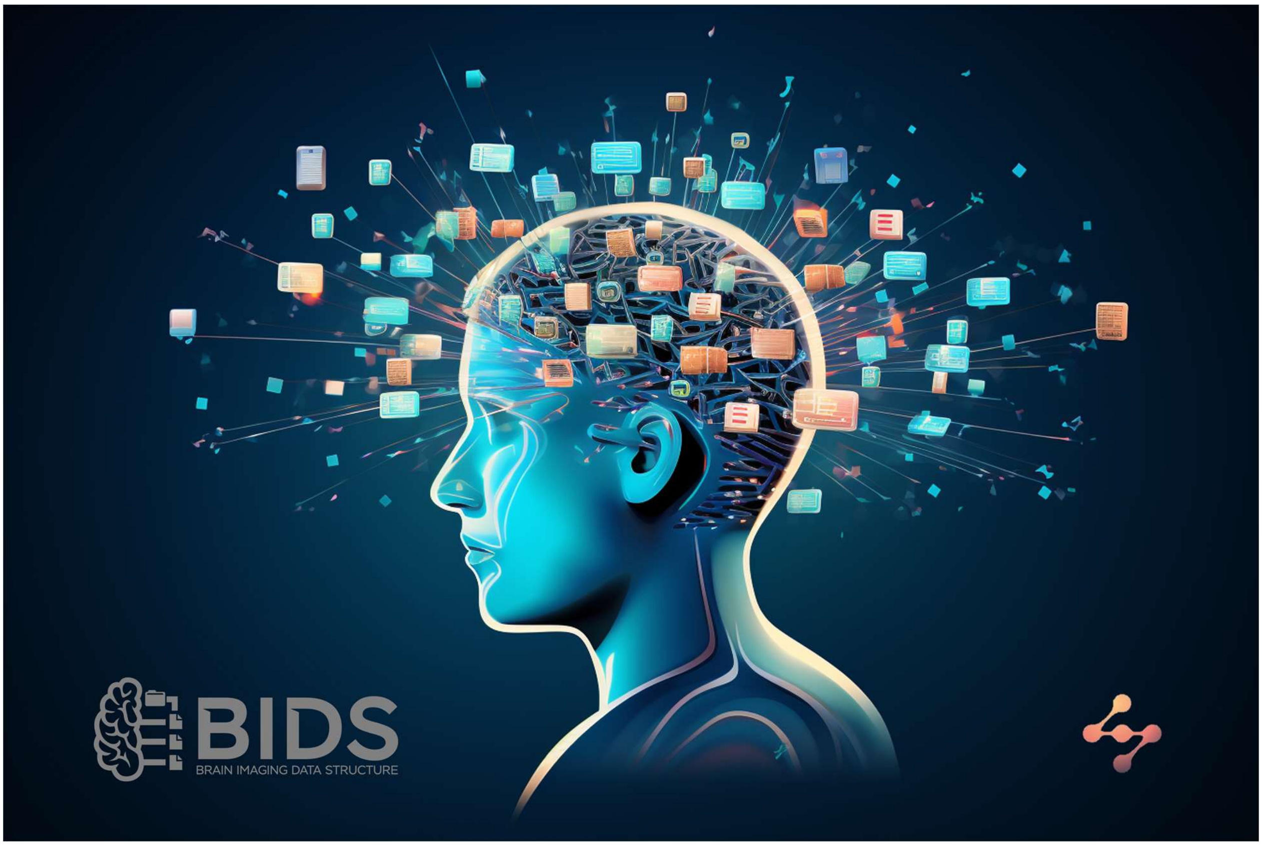
Hes·so VALAIS WALLIS

Haute Ecole de Gestion
Hochschule für Wirtschaft

Filière Informatique de gestion
Studiengang Wirtschaftsinformatik

Bachelor thesis 2023

Metadata management and search for Brain Imaging datasets.



Student : Théo Choffat

Teacher : Jean-Paul Calbimonte

Submitted on : 20.11.2023

www.hevs.ch

Sources of illustration on title page

Background image: generated by AI with MidJourney.

BIDS logo: https://upload.wikimedia.org/wikipedia/commons/thumb/d/de/BIDS_Logo.png/800px-BIDS_Logo.png

The SENSE logo: https://media.licdn.com/dms/image/C4E0BAQGQ6IevuCoolw/companylogo_200_200/0/1642694087699/thesense_research_center_logo?e=1707955200&v=beta&t=ikpMYGjxEBNQregwGJ__PwX_cQG1apNBcfCQvB5lBHk

Abstract

This thesis, in collaboration with the SENSE (Swiss Innovation and Research Center), aims to enhance EEG data management at CHUV by implementing an effective data organization technology. Addressing the challenges posed by current unstructured systems in neuroimaging, the research progresses through three key phases: a theoretical exploration of EEG, an evaluation of data organization standards (focusing on DICOM and BIDS), and the selection of suitable technologies for practical application.

Key findings include the selection of the BIDS standard for its compatibility with CHUV's requirements and the adoption of EEGLAB for converting raw EEG data into a structured format. Additionally, a custom Python tool was developed to facilitate efficient data querying within the reorganized database.

This work establishes a more structured and efficient approach to managing EEG data, enhancing research capabilities at CHUV. The developments made from this thesis is a foundation for future advancements in medical data organization for the CHUV.

Key words: Electroencephalography (EEG), Data Management, Standardization (Focusing on DICOM and BIDS), Technology Implementation

Foreword

This thesis marks the result of my training at HES-SO Valais, leading to a Bachelor's degree in Business Informatics. It was conducted during the final semester of a three-year full-time course. The primary objective of this is to showcase the skills and knowledge that I have acquired as a future graduate throughout my academic journey.

The subject of this work was developed in collaboration with Jean-Paul Calbimonte, my supervising professor, and enriched by researchers Chrysa Retsa and Micah Murray at the Centre Hospitalier Universitaire Vaudois (CHUV). The mission was to delve into the complexities of medical data organization, with a specific focus on electroencephalography (EEG) data within CHUV. A significant portion of the research involved a detailed exploration and analysis of medical data standards such as DICOM (Digital Imaging and Communications in Medicine) and BIDS (Brain Imaging Data Structure), analyzing their suitability and application for optimizing EEG data management at CHUV.

This report adheres to the APA 6 (American Psychological Association) formatting standard for scientific reports, taking into consideration the aspects of writing style, layout, figures and illustrations, and references. This standard was chosen to ensure clarity, consistency, and academic rigor in presenting the research findings.

For ease of reading and to maintain consistency in the narrative, the masculine form will be used in the texts. This choice is purely for linguistic simplicity and does not intend to overlook the importance of gender inclusivity in academic and professional environments.

Acknowledgements

I would like to express my deepest thanks to the people who guided and supported me during the completion of this bachelor's thesis.

Firstly, I would like to thank Jean-Paul Calbimonte, my teacher and mentor during this adventure. His encouragement, feedback and advice were very important in helping me to complete this work. His experience and expertise not only guided me through this work but also inspired me to improve its quality and rigour.

I am very grateful to Chrysa Retsa and Micah Murray, researchers at the CHUV, for their indispensable contribution to this project. Their willingness to share their knowledge of neuroscience was a great help in understanding and approaching this subject. I would also like to thank them for their efforts in sharing and accessing the CHUV's data, which was crucial to this project.

I would also like to extend my thanks to my peers Daniel Coimbra and Benjamin Morel. Their support extended beyond friendship; their advice and suggestions greatly enhanced the quality of this work. Their re-reading of the project gave me a critical and constructive eye which was the source of my motivation and improvement through the writing of this thesis.

Finally, I would like to thank my family, friends, and all who have supported me during these years of study. This thesis is not only a reflection of my efforts but also the support and belief of everyone who has been a part of this journey.

TABLE OF CONTENTS

List of Tables

List of Figures

Abbreviations List

ADHD: Attention Deficit Hyperactivity Disorder

AE: Application Entity

AI: Artificial Intelligence

APA: American Psychological Association

BDF: Bitmap Distribution Format

BIDS: Brain Imaging Data Structure

CHUV: Centre Hospitalier Universitaire Vaudois

CNS: Central Nervous System

CT: Computed Tomography

CSV: Comma-Separated Values

DA: Date

DB: Database

DICOM: Digital Imaging and Communications in Medicine

EDF: European Data Format

EEG: Electroencephalography

ERP: Enterprise Resource Planning

EXE: Executable

fMRI: Functional Magnetic Resonance Imaging

GUI: Graphical User Interface

HES-SO: Haute École Spécialisée de Suisse occidentale

Hz: Hertz

iEEG: Intracranial Electroencephalography

INCF: International Neuroinformatics Coordination Facility

IT: Information Technology

JPG: Joint Photographic Experts Group

JSON: JavaScript Object Notation

LORIS: Longitudinal Online Research and Imaging System

MRI: Magnetic Resonance Imaging

MWL: Modality Worklist

PACS: Picture Archiving and Communication System

PNG: Portable Network Graphics

PO: Product Owner

Q&A: Questions and Answers

SENSE: Swiss Innovation and Research Center

TSV: Tab-Separated Values

TXT: Text

UI: Unique Identifier

USB: Universal Serial Bus

VL: Value Length

VPN: Virtual Private Network

VR: Value Representation

1. INTRODUCTION

1.1. Subject

In the rapidly evolving landscape of medical research, effective data organization is crucial, especially in the field of neuroimaging and electrophysiology. This thesis delves into the critical issue of data organization within Brain Image Systems, with a focus on electroencephalograms (EEGs) at the Centre Hospitalier Universitaire Vaudois (CHUV). The challenge lies in the current state of data management: a decentralized and unsystematic approach where individual researchers manage data based on personal preferences, leading to a fragmented and unmanageable database.

This inefficient approach lowers research efficiency, particularly as the volume of data grows, making it increasingly difficult to conduct meaningful research. Recognizing the imperative need for a structured, cohesive data management strategy, this thesis seeks to explore and establish a more efficient and standardized system for EEG data organization.

The investigative journey of this thesis unfolds in three distinct yet interconnected phases. The initial phase is theoretical in nature, discussing the state-of-the-art is crucial to our understanding and approach. This includes an in-depth exploration of electroencephalography (EEG) as a technique, the definition and role of standards in medical data organization, and an introduction to actual standards like DICOM (Digital Imaging and Communications in Medicine) and BIDS (Brain Imaging Data Structure). This foundational knowledge is the basic tools needed for the next analytical steps.

In the second phase, the thesis will discuss about a comparative analysis of the DICOM and BIDS standards. This involves evaluating each standard's applicability and effectiveness specifically in the context of EEG data management at CHUV. The goal is to meticulously assess and rate these standards to identify the one that aligns best with the needs and objectives of CHUV's EEG database system.

The third phase shifts focus to the technological aspects. It involves a comprehensive comparison and selection of various technologies that are crucial for the practical implementation of the chosen standard. This phase is pivotal as it addresses the challenges of converting and restructuring the existing EEG database. A significant emphasis is placed on identifying or developing a conversion technology that facilitates the transformation of the current system into a more streamlined, accessible, and efficient one. This technological evolution is critical to realizing the goal of optimized EEG data management at CHUV.

The final segment of the thesis will provide a comprehensive conclusion, encapsulating the journey of transforming EEG data management at CHUV. It will offer insights into the result of the implementation, reflect on the learning experience, and discuss the potential future developments in this field. The thesis aims to address the central question: "How can we effectively organize and search within an EEG database at CHUV?" By answering this, the thesis will not only contribute to enhancing research efficiency at CHUV but also offer a blueprint for similar challenges in medical data management elsewhere.

1.2. Objectives

The primary objective of this thesis is to analyze and select the most effective data management technology for organizing medical imagery and electroencephalography (EEG) data. This objective is in collaboration with the SENSE (Swiss Innovation and Research Center), a collaboration between the Centre Hospitalier Universitaire Vaudois (CHUV) and HES-SO Valais/Wallis. The goal is to identify and demonstrate the potential of a data management that can enhance the searching and utilization of EEG data within CHUV.

Key objectives include:

- **Comprehensive Analysis**: Conduct a detailed examination of current data management practices in neuroimaging and electrophysiology, focusing on the challenges posed by the existing unstructured systems.
- **Standard Evaluation**: Evaluate and compare existing standards in medical data organization, such as DICOM (Digital Imaging and Communications in Medicine) and BIDS (Brain Imaging Data Structure), before choosing the most suitable standard for EEG data management at CHUV.
- **Technology Selection**: Identify and assess various technologies for the practical implementation of the chosen standard. Discussing about the technologies that facilitate the conversion and restructuring of the current EEG database into a format that is more organized, accessible, and efficient. Furthermore, choosing the right tool to perform queries inside this organized database.
- **Practical Application**: Prepare a proof of concept (or a full implementation, according to the time left) to demonstrate the feasibility and effectiveness of the chosen data management strategy.

The goal is to provide a structured, standardized approach to EEG data organization at CHUV, addressing the question: "How can we effectively organize and search within an EEG database at CHUV?".

1.3. Scope and limits

This thesis, to be completed between September 18, 2023, and November 18, 2023, operates within a tight timeframe, influencing its scope and depth. Given this constraint, the focus is on conducting a comprehensive analysis and providing informed recommendations to the SENSE research and innovation center concerning the future implementation of an optimal metadata strategy for EEG data management.

While the ideal outcome would include a complete implementation of the chosen strategy, the limited duration may necessitate pivoting towards developing a proof of concept. This proof of concept is intended to demonstrate the feasibility of the strategy, rather a fully operational solution. It is designed for future development and implementation.

Moreover, an essential aspect of this thesis is the creation of detailed documentation, which will include user guide. This documentation is aimed at ensuring that the work conducted can be effectively continued.

2. Background & State of the Art

2.1. Context

In the field of data management and technology, a standard refers to a set of established norms or requirements designed to ensure consistency, compatibility, and reliability within a specific field (ISMS, 2023). When implementing a standard, it is important to know what we are looking for, the different options available, and the characteristics of each standard. While other standards do exist in the field of medical data organization, in our case, we selected two standards that could be potential candidates for implementation in our project with the CHUV (ISMS, 2023). On the one hand, BIDS is the new emergent technology, with a very active community and promising new advances in the field. On the other, DICOM is the old standard that is already well established and has already proved its worth. The aim of this section will therefore be to differentiate between the two standards, compare them, and finally decide which one to choose so that we can move on to implementation.

But first, we need to have a better understanding of electroencephalography (EEG) in general. This is the subject of the very next chapter, that delves into the basic information that one might need to comprehend the explanation and choice of the DICOM and BIDS standards.

2.2. What is an electroencephalography?

When a human being thinks, his brain sets up a myriad of very complex bio-ingenious events resulting in the simple act of thinking. This works, in a very simplified way, by connecting neurons together. A brain has multiple billions of neurons, and the connection is made through chemical and electrical signals. The EEG provides a window for us to be able to visualize and interpret what is happening in real time (Kirschstein, T. & Köhling, R., 2009).

An EEG analysis of the brain simply measures the electrical currents passing through our brain while stimulating the patient to certain tasks to measure, analyze and finally diagnose the activity of a brain. This diagnosis allows the doctor to be able to see the presence of certain diseases such as Alzheimer’s, schizophrenia, or ADHD (difficulty of attention and hyperactivity) (Lenartowicz, A. & Loo, S. K., 2014). To do this, the doctor places between 20 and 100+ electrodes on the patient’s brain. To optimize the connection between the electrodes and the brain, it is possible to apply a conductive gel that will ensure better connectivity by ignoring the hair. Then, the doctor asks the patient to do several activities to put the brain to the test and finally to be able to detect any problems that might arise (Angus-Leppan, 2007). In the most used exercises, we can find:

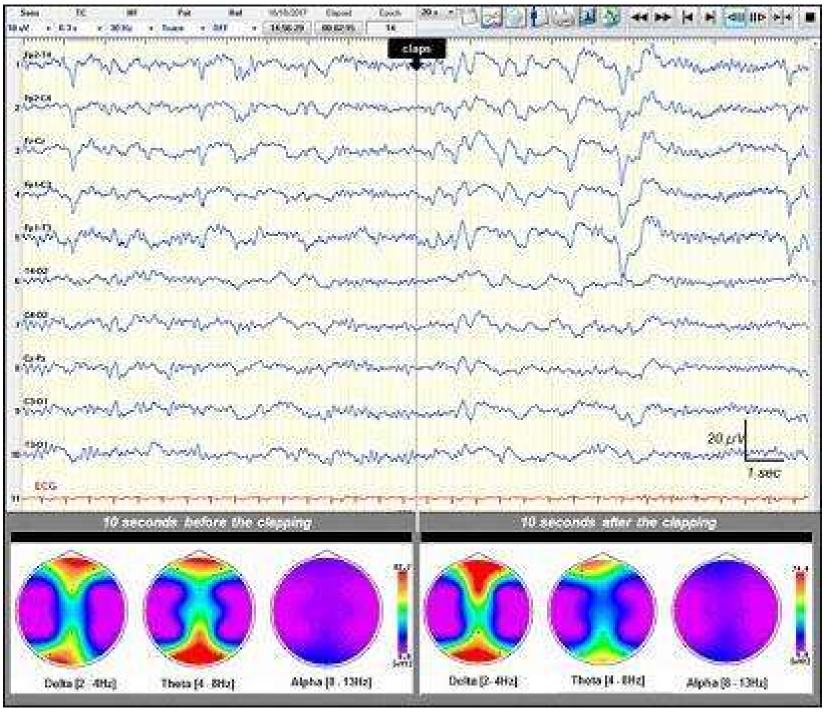
Sleep: The most used method to analyze the brain. Indeed, when we are awake, our brain receives all kinds of commands, many of which should not be analyzed. For example, all information related to muscle contraction, even involuntarily, will "pollute" the electrocardiogram. It is therefore recommended to do the test on a sleeping patient.

Hyperventilation: A person who is in hyperventilation will be more likely to cause a risk event and, therefore, it will be easier for the doctor to detect it. This is because too much oxygen is toxic to the CNS (Central nervous system).

Photic simulation: Technique used to detect the risk of epilepsy in the patient. It is simply to agitate the sense of vision by showing colors and flashes to the patient to be able to provoke and measure an epileptic event.

At the output, the machine takes out the voltage of each electrode (calculated in Hz) over time. All these outputs are analyzed in the form of a “brainwave” allowing the doctor or researcher to visualize the variation in brain activity. What interests the medical profession is the presence of spikes that can indicate the presence of an anomaly. As an example, here is the result of an EEG showing a patient with impaired consciousness reacting to auditory stimulation (claps). We can see the difference between the part before and after the clap. The researchers will then analyze the different spikes to draw a conclusion (Azabou, É., et al., 2018).

Figure 1 : Example of an electroencephalogram



Source: https://ccforum.biomedcentral.com/articles/10.1186/s13054-018-2104-z/figures/1

Then, the researchers will usually create a Word document or an Excel spreadsheet to add useful information regarding the experiment, the patient or other type of information. Here is what a (simple) EEG Word document typically looks like:

Figure 2 : Word document of a researcher

participant_id	gender	age	Ethnicity	Air_conditioning
sub-001	F	44	Indian	on
sub-002	F	32	Indian	on
sub-003	F	28	Non_indian	off
sub-004	M	35	Indian	off

Description

Data collection took place at the Meditation Research Institute (MRI) in Rishikesh, India under the supervision of Arnaud Delorme, PhD. The project was approved by the local MRI Indian ethical committee and the ethical committee of the University of California San Diego (IRB project # 090731).

Participants sat either on a blanket on the floor or on a chair for both experimental periods depending on their personal preference. They were asked to keep their eyes closed and all lighting in the room was turned off during data collection. An intercom allowed communication between the experimental and the recording room.

Participants performed three identical sessions of 13 minutes each. 750 stimuli were presented with 70% of them being standard (500 Hz pure tone lasting 60 milliseconds), 15% being oddball (1000 Hz pure tone lasting 60 ms) and 15% being distractors (1000 Hz white noise lasting 60 ms). All sounds took 5 milliseconds to ramp up and 5 milliseconds to ramp down. Sounds were presented at a rate of 1 per second with a random gaussian jitter of standard deviation 25 ms. Participants were instructed to respond to oddball by pressing a key on a keypad that was resting on their lap.

Instructions

Close your eyes. Press the button when you hear the high pitch sound and ignore other sounds.

Task Description

Participants were instructed to close their eyes and respond to oddball sounds. There were 3 types of sounds interspaced by about 1 second each, standard (500 Hz pure sounds lasting 60 ms), oddball (1000 Hz pure sounds lasting 60 ms), or distractor sounds (white noise lasting 60 ms). Participants were instructed to press a keypad on their lap when hearing an oddball and to ignore other types of stimuli. See experimental code for additional details.

Conditions

condition 1	response
condition 2	standard
condition 3	ignore
condition 4	oddball
condition 8	noise

Source: https://www.youtube.com/watch?v=LIiRj8-BVk4&t=278s

There is multiple useful information like the age, the gender, the ethnicity, and some information that may be useful for this experimentation (like here if the air conditioning was on during the test).

Next, we have a description of the data, to understand where, how and why they did this. We can also find in these types of document information like the instructions, the explanation of the task and other types of parameters.

The main problem with this structure is that it is hard for a researcher that was not part of the experiment to find a specific information as this word document vary a lot between the person that made it. Also, if there is a big database containing multiple experiments (datasets), it is really time consuming to find and filter the information.

If an EEG is not invasive, an iEEG (intracranial Electroencephalography) is invasive. The main difference is that the electrodes are placed surgically directly on top of the tissues of the brain. The results of an iEEG are way more precise than a normal EEG. It is mainly used on patients that are experiencing drug-resistant epilepsy to localize the seizure in the brain to surgically remove it afterwards (Lachaux, J., Rudrauf, D., & Kahane, P., 2003). It is also way more expensive and may cause complications. The output of the non-invasive and the invasive method is a file with the ".bdf" or ".edf" extension containing the brainwave data. The researcher also adds notes that will describe the patients state (left or right-handed, air conditioning on/off, age, gender, etc.) and the detailed description of the experiment as well as other useful information.

The main issue with this is that we have a lot of useful information, but they are currently not structured. The researchers sometimes write the description inside a Word document, sometimes inside an Excel spreadsheet or even directly into a Notepad. This is not a problem if a researcher works alone but starts to be an issue when he needs to share this information. Another researcher, that is, for example, not used to work with Excel or simply does not know where the information is stored, can lose a lot of time searching for the right document. Especially if the database is containing hundreds of experiments. That is the reason why we need to find a data structure capable of organizing datasets and for researchers to be able to search easily useful information.

2.3. Brain Imaging Data Structure (BIDS)

2.3.1. History and Evolution of BIDS

“The Brain Imaging Data Structure (BIDS) is an emerging standard for the organization of neuroimaging data.”. (BIDS, 2023) The project was instantiated as OBIDS (Open Brain Imaging Data Structure) in January 2015 at Stanford University during a conference called INCF (International Neuroinformatics Coordination Facility). OBIDS was developed by the community for the community in response of the lack of organization in the field of brain imaging. At this time, the project was spearheaded by Chris Gorgolewski. (BIDS, 2023) In early 2019, the project was carried on by other researchers from Stanford (Stephan Appelhoff, Franklin Feingold and a Stanford Laboratory called Poldrack Lab. Soon after, in October 2019, “the BIDS community voted to ratify a new governance structure and to elect five members to oversee the development and adoption of the standard.” (BIDS, 2023) This group is still up to date today and is called the BIDS Steering Group. Because it is a community effort, multiple tools and platforms have emerged and a lot of them are currently still in development. The community is very active and is really trying to facilitate the procedure to motivate more people to adopt this standard. However, because it is a relatively recent project, there is still a lack of tools and procedures. Of course, we will talk more about this later.

2.3.2. BIDS: Purpose and Need

The BIDS project was originally created in response to a lack of organization and consensus in the medical imagery field. At the time, in 2015, DICOM was the most used standard. The problem is that DICOM is quite complex to set up and has few issues. The main one is that DICOM has too many optional fields to enter during the creation of an image object. (Springer, 2023) This results in many data inconsistencies and blank fields that are sometimes filled with errors only to be validated. Another problem is that the different search groups all have their unique ways of organizing data, making it hard to share, compare, consolidate, etc. All these problems of not having a consistent and solid data organization are developing other collateral issues like the lack of transparency and reproducibility as other doctors might skip crucial information. Having a poor data organization also makes it impossible to scrape and, therefore, impossible to analyze.

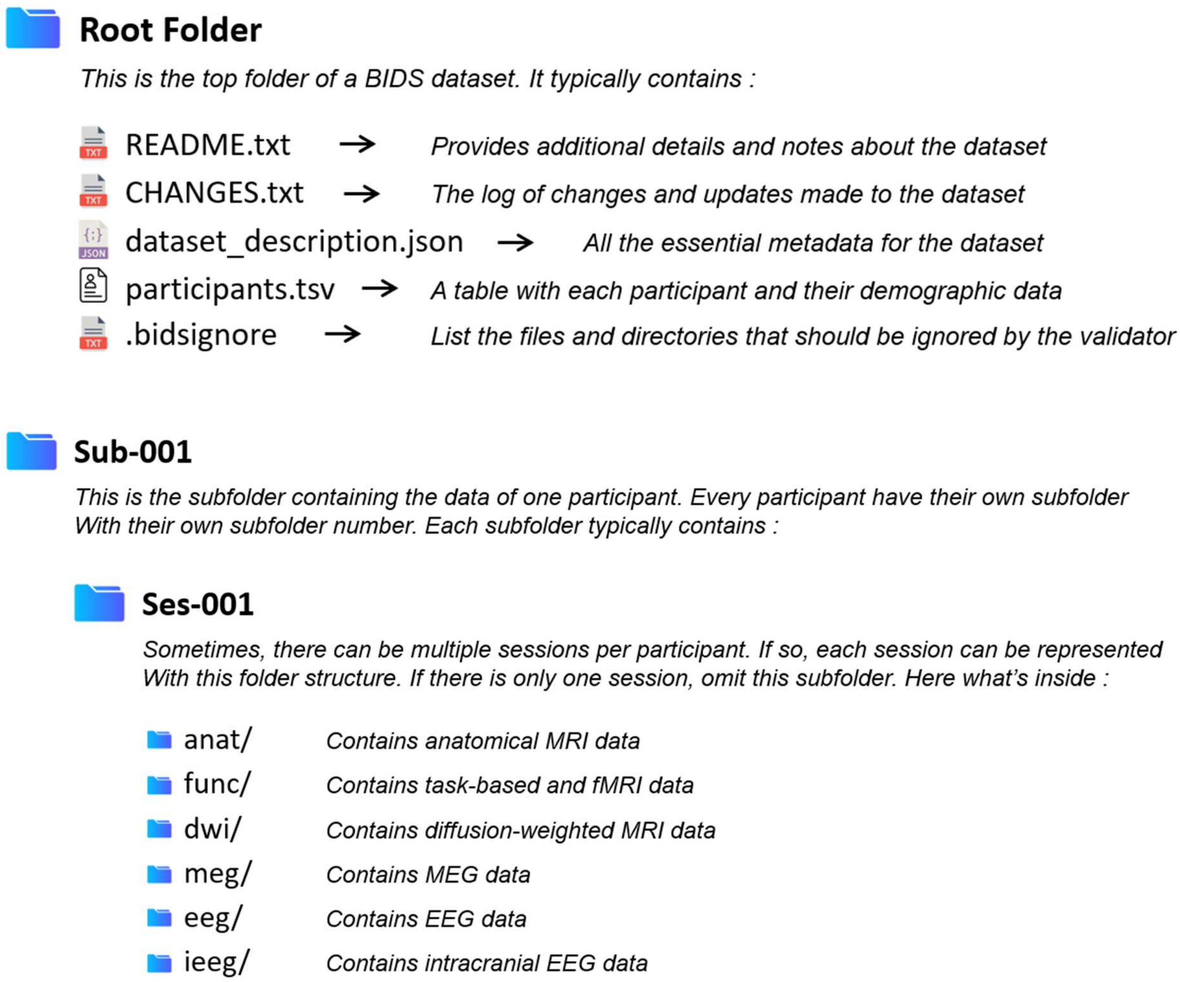
This is where BIDS comes for rescue with its clear and comprehensive data structure and organization. This standard also provides a clear way to implement the metadata (the information that goes with the image like the patient's name, age, sex, etc..). This clear representation facilitates data sharing, makes easier to understand the data and the methodologies used as well as the pre-processing methods increasing the data transparency. Regarding the issue that they currently have at the CHUV, this implementation could also solve the searching problem that they currently have. With this standard, the hospital may proceed to the analysis of their dataset, promoting the reuse of the data to respond to new research questions or meta-analysis. (Gorgolewski, K. J., et al., 2016)

Because the BIDS is open source, many tools have been developed to be able to implement this strategy easily. In the case of this thesis and the CHUV, the implementation part needs to be done as quick as possible while keeping a perfect data integrity. We do believe that the BIDS standard can be a good response to this project of transforming a non-organized dataset into a validated data structure capable of performing searching and analysis actions.

2.3.3. BIDS Structure and Components

The BIDS standard has been made famous thanks to the good structure and organization that it has. This chapter will explain more in detail how it is done and will also explain some definitions. A dataset is separated and organized with the help of a simple folder structure. This structure is hierarchical so that the user can first have an overall comprehension of the dataset and then dig into more fine details. The hierarchy can be represented like so:

Figure 3 : BIDS structure example



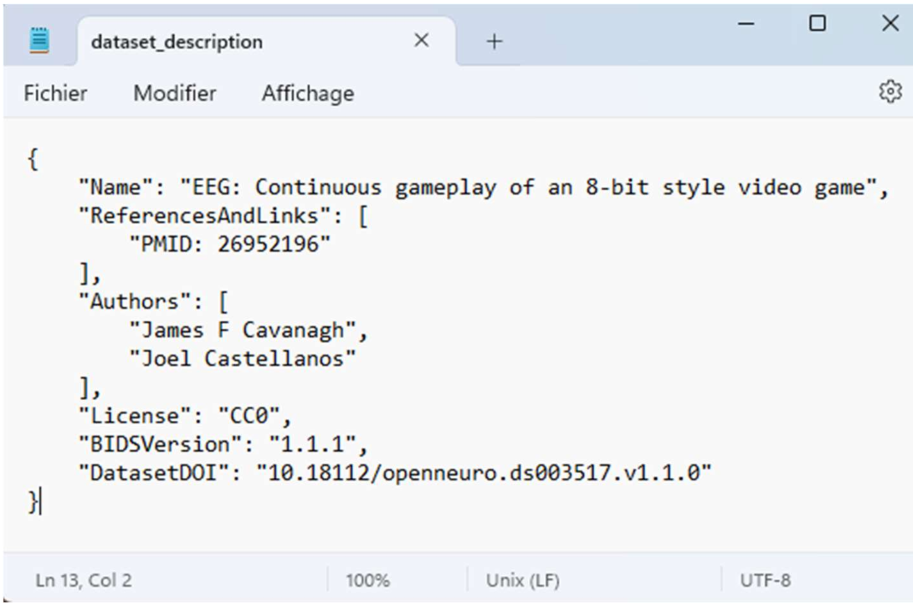
Author's source

Each of these folders may contain multiple different information based on the data they are storing. In our case, we will deal with two of them, the "eeg" and the "ieeg" folders. Inside them, we can find the EEG itself as well as other information that are not metadata like the EEG coordinates, the electrodes, the channels, the pictures, etc… Of course, there is a template that exists, we can't just throw random data in there or the folder won't be validated (Electroencephalography - Brain Imaging Data Structure V1.8.0., 2022).

2.3.4. Metadata in BIDS

In the medical field, the metadata of a document is as important as the data itself. The metadata is responsible to describe all the information that are around the data, it gives a general context to the data and is crucial when one want to compare cases that are alike. In BIDS, the metadata is stored in the .json and the .tsv files. The json files are text files that adopt the key-value principle, they are easy to read/write with code, which makes them useful as they can totally work with machines (Electroencephalography - Brain Imaging Data Structure V1.8.0., 2022). Here is an example of a BIDS dataset description in JSON:

Figure 4 : Example of a JSON file in BIDS

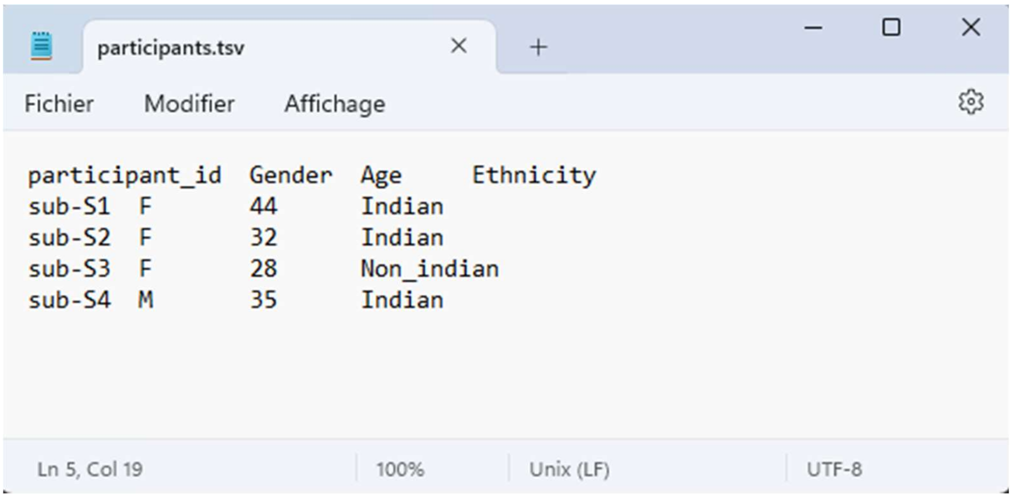



```
{
    "Name": "EEG: Continuous gameplay of an 8-bit style video game",
    "ReferencesAndLinks": [
        "PMID: 26952196"
    ],
    "Authors": [
        "James F Cavanagh",
        "Joel Castellanos"
    ],
    "License": "CC0",
    "BIDSVersion": "1.1.1",
    "DatasetDOI": "10.18112/openneuro.ds003517.v1.1.0"
}
```

Author's source

The TSV (Tab-separated values) files are basically text files with tabs that separate the different fields of the file. It is also quite easy to read/write files with different languages of code which also makes them useful when it comes to the medical field as they need to be analyzed and read by machines. Here is an example of a TSV file in a BIDS structure that define the different participants and add some useful information, as the separation is made by one tabulation, it is normal that the columns (in a notepad) are not aligned:

Figure 5 : Example of a TSV file in BIDS



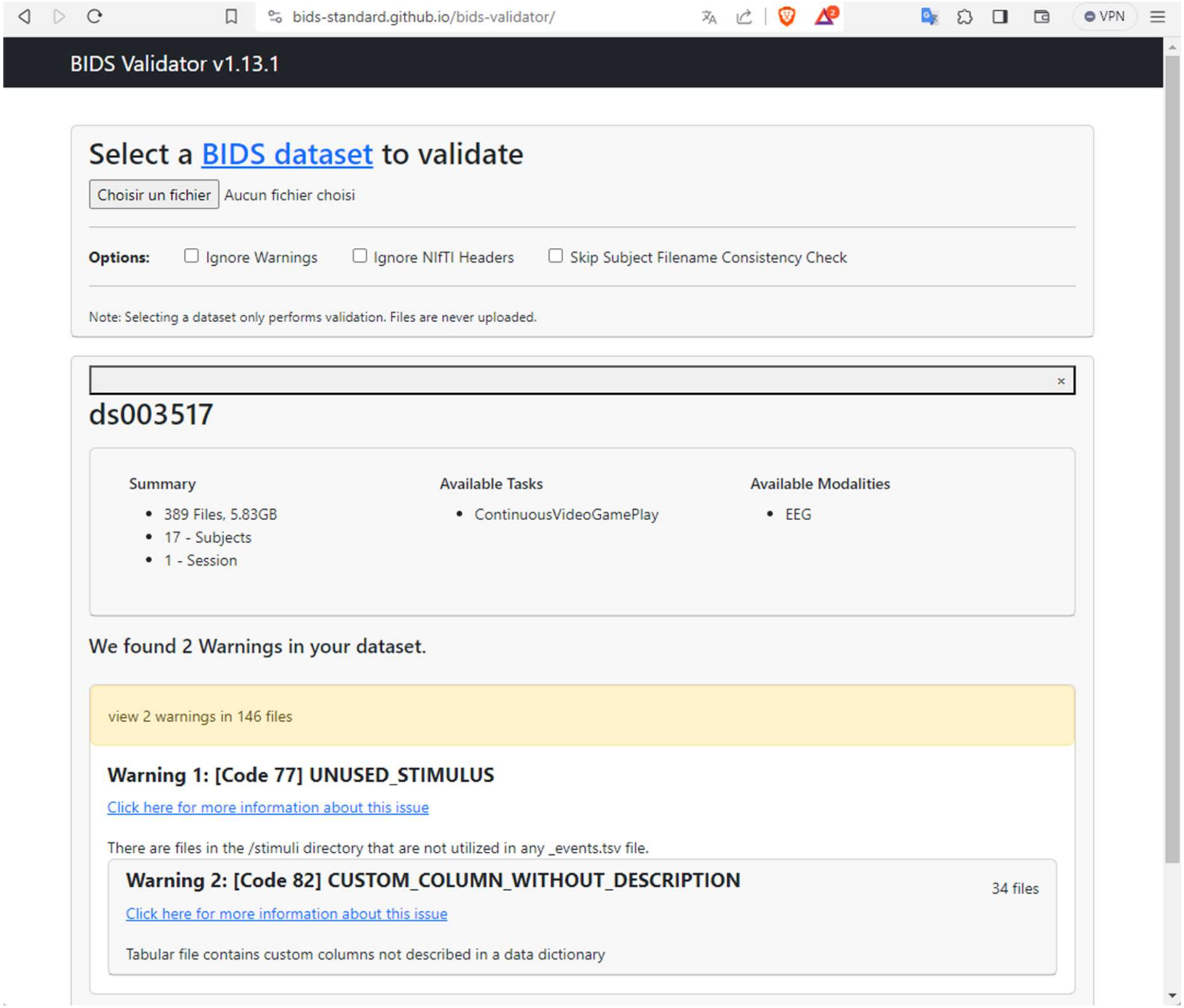
Author's source

The metadata in BIDS represents a big advantage and is the main reason why this standard is popular today. Another useful implementation is that the metadata is here inherited by all the lower levels. That means that we can specify the metadata once in a higher level and avoid any redundancy. BIDS also implements specific metadata for each imaging modality. For example, EEG metadata will have information that is irrelevant when compared to MRI or fMRI data. This makes the information more concise and precise with less blanks.

2.3.5. BIDS Validation and Tools

If we want to implement a standard the right way, we need to ensure that everybody is doing the same thing when it comes to organizing the database. To do so, BIDS have developed a tool that takes input a BIDS file and output the validation or the non-validation of the file. It also outputs the reason why and how the file could not be validated and provides some warnings to apply if we want to perform the validation the right way. The BIDS validator is open source and can be found on Git as a python library, as a web-app or directly included inside various software (Community, 2023). Here is an example of an online BIDS validator:

Figure 6 : BIDS validator visualization



Author's source

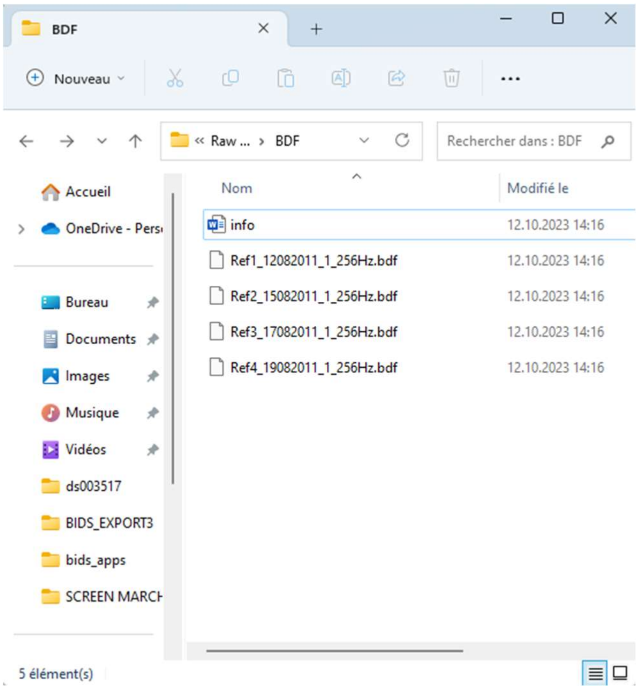
Apart from the validator, there are multiple tools available that allows us to perform data conversion. There are a lot of them, a good example that we might consider is called “EEG2BIDS” this extension will help us to transform raw EEG data into a BIDS-validated format. Because it is an open-source software, multiple formats are currently under development and some of them are available but not finished (BIDS, 2022). This is why it is important to be sure that we use the right tools developed and recognized by the right people before committing and implementing one. There is currently 40+ converters available on the official website. These come with the last update date as well as the expected input and the code language used.

BIDS extensions are also something to consider. Because the standard is open source, there are multiple extensions and tools that are still under development. We can find a list of them and their last update on the official website. If there is a tool that does not exist, it may still be under development on this page (Extending the BIDS Specification, 2022).

2.3.6. Benefits of Adopting BIDS

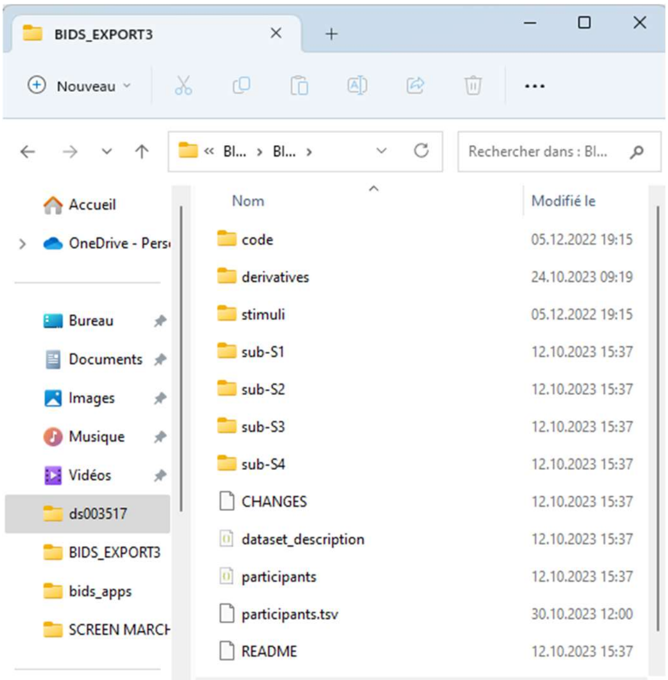
Today, the CHUV sorts its data in a non-homogeneous way. Each doctor organizes his or her data in his or her own way. The problem with this is that it is difficult to find data, projects or simply to analyze them. Introducing a structure like BIDS would restructure the entire database in a homogenous way so that every doctor and every person who needs to access this data can do so knowing exactly what they are going to come across when they venture into the database. As a comparison, on the left, there Is a typical structure adopted by the researchers; a simple word that contains all the information (sometimes an Excel spreadsheet or a simple Notepad) and the raw EEG data (more about this later). On the right, the same structure but organized in BIDS structure:

Figure 7 : Raw data example



Author's source

Figure 8 : BIDS folder organization example



Author's source

All the benefits come from a well-organized structure. It's easier to find a way around, to search or simply to analyze data when we have clear guidelines to follow. This saves a lot of time for researchers and doctors looking for specific or non-specific information. In addition, it will be easier for users to share and to collaborate with data by ensuring that it is understandable to everyone (anyone with a basic knowledge of the BIDS structure). The BIDS-validator also encourages people to insert data in a structured way and to respect the integrity of the data. As a result, the data is better annotated and more respected (if not, it won't be validated by the validator). It also facilitates the task of the people responsible for quality control by providing them with a simple structure and tools for analyzing it. Another advantage is that, as mentioned above, BIDS is an open-source standard. This includes several benefits, such as it is constantly being developed, both in terms of the structure itself and the extension tools. This also means that the data and the structure are organized in line with the times and are useful for the researcher. What's more, the data conversion tools, which are updated every week or so, make for simpler, more effective implementation when the organization is transferred (Gorgolewski, K. J., et al., 2016).

In conclusion, BIDS is a standard that provides many solutions to the problems related in the storage organization of the medical datasets. Being able to search for and analyze information in an unknown dataset with just a few clicks would be an ideal step forward for the CHUV.

2.3.7. Limitations and Challenges

There is no doubt that BIDS has several advantages that place it at the top of the list of the best standards to adopt today. However, it should be noted that no standard is perfect. In this section, we will list some of the limitations and difficulties that can be encountered with the BIDS standard.

The first difficulty is standard to every change of IT structure; in our case, it will be complicated for researchers used to working in a certain way to change their habits entirely. They will have to learn a new way of working and organizing data, which may mean lost time and frustration for some. And while BIDS provides clear guidelines, there’s still room for mistakes in implementation.

Another major difficulty will be converting the existing dataset into BIDS format. Although several tools and extensions exist, some remain unsupported and must therefore be sorted manually on a case-by-case basis. This results in many extra hours of work if old data is to be transferred. Due to the difficulty of understanding the current CHUV system, it is possible that the transfer of data could result in the loss or inaccuracy of some of it. Some documents, which are too old or contain too little information, may not even pass the basic validation of the BIDS validator, and may therefore be lost.

Additionally, another concern is the open-source nature of the structure. Its constant development is a double-edged sword. Adopting it means accepting that its structure will be adapted and changed, and that doctors will constantly have to adapt to new updates. Another shortcoming of open source in general is that its upgradability is dependent on its community. If its community shrinks, upgrades will also be reduced, and BIDS risks being deprecated.

To conclude, we can say that BIDS has some shortcomings linked to its open-source origin and to its implementation, which requires an effort on the part of doctors to learn and set up this organization. However, it should be noted that the standard is constantly evolving and that its structure will be further improved over time. If we now look at its implementation within the CHUV, in this case it is not a question of replacing an organizational strategy but of implementing it, as no organization is currently in place. This implementation of BIDS will certainly pose a few constraints, but it is necessary if the CHUV wants to be able to carry out research and analysis in its database.

2.3.8. BIDS EEG Structure and Components:

As mentioned above, BIDS has a highly regulated structure. It is difficult to generate an entire structure by hand. That is the reason why there are tools that can generate this for us. However, the BIDS structure was not created to be as efficient as possible, but to resemble as closely as possible the way a doctor would have organized it if he'd had to organize it himself. The way in which this standard is organized is therefore very close to human thinking (Gorgolewski, K. J., et al., 2016).

If we take a closer look at the structure, the first three files (highlighted in green below) are, in terms of the EEG structure, always the same (detailed description in chapter 2.1.3). The part that is not highlighted corresponds to the name of the folder indicating several items of information such as the patient index, the session, the task performed, the acquisition (called "acq") and the run index (Electroencephalography - Brain Imaging Data Structure V1.8.0., 2022). The part highlighted in yellow corresponds to the file itself and its extension. The contents of the "eeg/" folder may vary from study to study and may be complex. Here's what it might look like if it were made as complex as possible:

Figure 9 : BIDS structure for EEG

```
sub-<label>/
   [ses-<label>/]
      eeg/
         sub-<label>[_ses-<label>]_task-<label>[_acq-<label>][_run-<index>]_eeg.<extension>
         sub-<label>[_ses-<label>]_task-<label>[_acq-<label>][_run-<index>]_eeg.json
         sub-<label>[_ses-<label>]_task-<label>[_acq-<label>][_run-<index>]_events.json
         sub-<label>[_ses-<label>]_task-<label>[_acq-<label>][_run-<index>]_events.tsv
         sub-<label>[_ses-<label>]_task-<label>[_acq-<label>][_run-<index>][_recording-<label>]_physio.json
         sub-<label>[_ses-<label>]_task-<label>[_acq-<label>][_run-<index>][_recording-<label>]_physio.tsv.gz
         sub-<label>[_ses-<label>]_task-<label>[_acq-<label>][_run-<index>][_recording-<label>]_stim.json
         sub-<label>[_ses-<label>]_task-<label>[_acq-<label>][_run-<index>][_recording-<label>]_stim.tsv.gz
```

Author's source

This template is dedicated to the data recorded during an EEG. The first two types, "_eeg" and "_events", are mandatory and very often present. The next two, "_physio" and "_stim", are mandatory when using continuous recording files. They are much rarer to find. In terms of extensions, ".tsv.gz" corresponds to a gzipped and tab-delimited file. The file can be zipped if its volume is too large. If we have a smaller file, we recommend using only the ".tsv" extension (Electroencephalography - Brain Imaging Data Structure V1.8.0., 2022).

There are four other types of files that can be found in this folder. The first two, "_channels.tsv" and "_electrodes.tsv", are very similar because they both define information about the electrodes. If the latter exists, it is compulsory to add the file "coordsystem.json" with the more precise information about the location. The last type would be called "_photo.jpg"; it is optional and may contains any photos of the landmarks (Gorgolewski, K. J., et al., 2016). Here is an example of the potential template that would be used:

Figure 10 : BIDS structure for EEG

```
sub-<label>/
    [ses-<label>/]
        eeg/
            sub-<label>[_ses-<label>]_task-<label>[_acq-<label>][_run-<index>]_channels.json
            sub-<label>[_ses-<label>]_task-<label>[_acq-<label>][_run-<index>]_channels.tsv
            sub-<label>[_ses-<label>][_acq-<label>][_space-<label>]_electrodes.json
            sub-<label>[_ses-<label>][_acq-<label>][_space-<label>]_electrodes.tsv
            sub-<label>[_ses-<label>][_acq-<label>][_space-<label>]_coordsystem.json
            sub-<label>[_ses-<label>][_acq-<label>]_photo.jpg
            sub-<label>[_ses-<label>][_acq-<label>]_photo.png
            sub-<label>[_ses-<label>][_acq-<label>]_photo.tif
```

Author's source

This structure and the file names must be applied to pass the BIDS validation. A good structure and conventional file naming leads to a better general understanding of the data and makes it much more valuable. Bear in mind that the structure seen so far represents everything that is feasible and accepted with BIDS. On average, an "eeg/" folder contains 4 to 5 files (Electroencephalography - Brain Imaging Data Structure V1.8.0., 2022).

2.3.9. Metadata in EEG BIDS:

As a reminder, the metadata corresponds to all the data adding context and information to the basic measurement (in this case the Electroencephalogram). The BIDS standard proposes a hereditary implementation of metadata. This means that it is possible to place a metadata file in the root folder. If this metadata is specified in lower folders, it will overwrite the metadata that is higher in the hierarchy. In the BIDS standard, metadata is stored in Json or tsv files called "sidecar". In an "eeg/" folder, the metadata linked to the result are inside the "eeg.json" file (Electroencephalography - Brain Imaging Data Structure V1.8.0., 2022). This file is highly regulated and must be structured in accordance with the following example:

Figure 11 : Mandatory fields in JSON file

```
{
   "TaskName": "Seeing stuff",
   "TaskDescription": "Subjects see various images for which phase, amplitude spectrum, and color vary continuously",
   "Instructions": "Your task is to detect images when they appear for the 2nd time, only then press the response button",
   "InstitutionName": "The world best university, 10 Beachfront Avenue, Papeete",
   "SamplingFrequency": 2400,
   "Manufacturer": "Brain Products",
   "ManufacturersModelName": "BrainAmp DC",
   "CapManufacturer": "EasyCap",
   "CapManufacturersModelName": "M1-ext",
   "EEGChannelCount": 87,
   "EOGChannelCount": 2,
   "ECGChannelCount": 1,
   "EMGChannelCount": 0,
   "MiscChannelCount": 0,
   "TriggerChannelCount": 1,
   "PowerLineFrequency": 50,
   "EEGPlacementScheme": "10 percent system",
   "EEGReference": "single electrode placed on FCz",
   "EEGGround": "placed on AFz",
   "SoftwareFilters":{
      "Anti-aliasing filter":{
         "half-amplitude cutoff (Hz)": 500,
         "Roll-off": "6dB/Octave"
      }
   },
   "HardwareFilters":{
      "ADC's decimation filter (hardware bandwidth limit)":{
         "-3dB cutoff point (Hz)": 480,
         "Filter order sinc response": 5
      }
   },
   "RecordingDuration": 600,
   "RecordingType": "continuous"
}
```

Author's source

The keys highlighted in yellow correspond to the mandatory fields for validation in BIDS format. The other values are recommended to enable a more precise search and analysis of the data. Obviously, the doctor will not be asked to fill in the information directly in this file, but in existing programs that will complete it automatically (Electroencephalography - Brain Imaging Data Structure V1.8.0., 2022).

2.4. DICOM

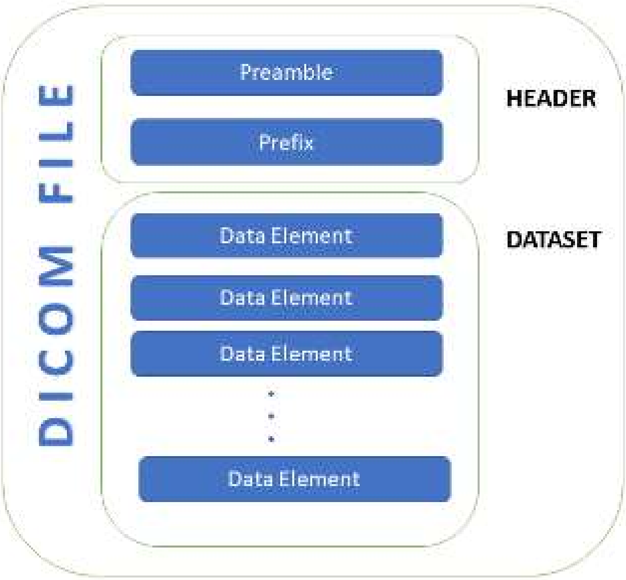
2.4.1. Introduction

DICOM (Digital Imaging Communication in Medicine) is a standard used worldwide. Implemented in 1993, it has proven its usefulness in the medical world (Mildenberger, Eichelberg, & Martin, 2001). Hundreds of machines are compatible with the standard, and millions of medical images have since been generated. It was created to make doctors' work with their patients more fluid and digital. It is, of course, the main competitor to the BIDS structure, and it is therefore necessary to make a comparison to confirm whether to adopt it. But before that, we will have an overview of this standard with topics on how it is structured, the metadata, the implementation, the advantages, and disadvantages and finally a conclusion to understand better the DICOM standard to better compare and analyze it afterwards (Weston, 2021).

2.4.2. DICOM file format

If the BIDS standard has a folder structure to identify and store its data, in DICOM, everything is under the one and unique file. This file can be separated in two parts; the Header, that contains metadata about the patient, the study and other information like the procedure, details and so on. The other part is the Image Data that contains one or multiple images. Unlike the typical data format, we are used to (JPG or PNG), the DICOM has his unique format that basically has two colors black represented as a 0 and white represented as a 1. Now imagine a huge grid containing only 0 and 1, this is exactly what is inside a DICOM image file (Graham, R., Perriss, R., & Scarsbrook, A., 2005). Here is what a typical DICOM file structure looks like:

Figure 12 : DICOM file structure



Source: https://www.vladsiv.com/dicom-file-format-basics/

2.4.3. DICOM data elements

Data elements are like the Lego bricks of DICOM. Each brick provides a specific information about patient, study, or image. The way this information is structured and identified allows for efficient storage, retrieval, and communication of medical imaging data (DICOM, 2023). Here is a brief explanation of the most important data elements:

- **The Tags:** It is a unique identifier that is assigned to each data element. It is defined as two numbers separated by a comma and tells us what kind of information the data element holds. For example, the tag “(0008, 0020)” represents the study date. A list of every tag can be found on internet on multiple websites. When we want to search through several documents, there are tools that will look and compare the search with the tags.

- **The Value Representation (VR):** To ensure that the right information is in the right field, DICOM has a Value Representation that is responsible for denoting the data type and format of a data element’s value. For example, the VR of the date is DA, the VR for a unique identifier is UI. If a field has the value "20230101" this can be January 1, 2023, or a unique identifier. The VR is here to increase the data consistency as well as data validation by checking that it is correct.

- **The Value length (VL)**: This section simply specifies the length of the data element’s value.

- **The Value Field**: Contain the actual value of the data element.

2.4.4. Hierarchical structure

The structure that must be applied when using DICOM is hierarchical. It is organized according to the following structure:

- **Patient**: The top of the hierarchy. The person to whom the data belongs. One patient can have multiple studies.
- **Study**: Set of images that correspond to a specific project or study. One study can contain multiple Series.
- **Serie**: It is a part of the set of images that correspond to the same modality (e.g., EEG, MRI)
- **Image**: Single image in the Serie.

Here is an example of the structure that can be found in a database. This example was found in a public database of the cancer imaging archive and adapted to make fake patient data by Weston, A., PhD. (2021, December 16).

Figure 13 : DICOM structure visualization

```
.
└── 5553226
    └── 1.2.826.0.1.3680043.2.1125.1.38381854871216336385978062044218957
        └── 1.2.826.0.1.3680043.2.1125.1.68878959984837726447916707551399667
            ├── 000000.dcm
            ├── 000001.dcm
            ├── 000002.dcm
            ├── 000003.dcm
            ├── 000004.dcm
```

Source: https://www.dimitripianeta.fr/documents/traitements/comprendre%20les%20m%C3%A9tadonn%C3%A9es%20.pdf

As we can see, it's hard for a human to understand this content immediately.

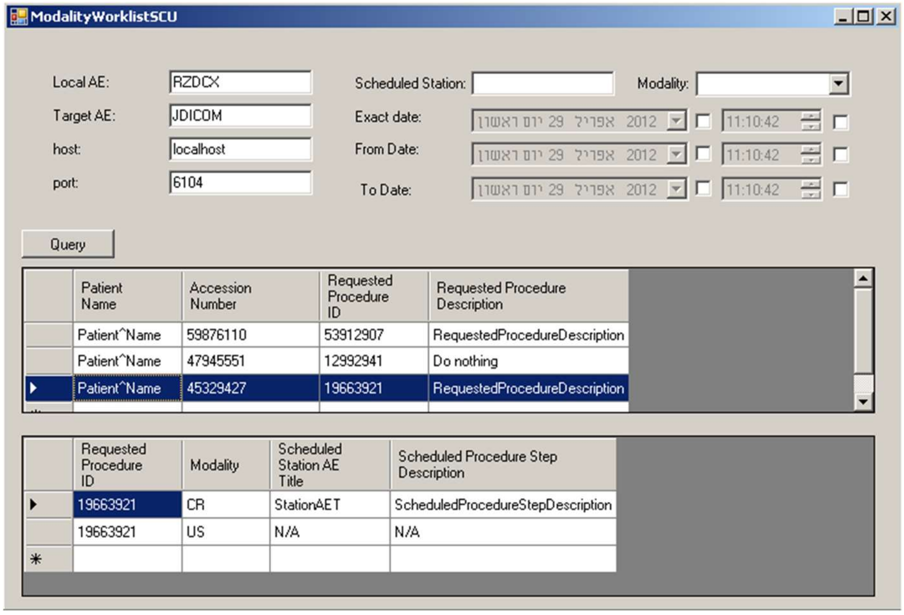
2.4.5. DICOM Services

As we saw above, the DICOM standard is not made for a human to understand, create and search manually the data. Therefore, DICOM has a set of standardized operations and procedures that allow different software and devices to communicate and exchange data. These services are the core of DICOM as they enable various imaging equipment to interact and communicate between them (Ae, 2022). Here are few examples of the most important DICOM services:

- **C-STORE**: This command is used when an object needs to be transferred from one Application Entity (AE) to another. For example, when an MRI scanner needs to send the acquired data to a PACS (Picture Archiving and Communication System) for storage.
- **C-FIND**: This command allows the researcher to query for a specific AE based on given criteria, for example if he wants to find the AE for M. Shaufat the 23 May 1989 at 18h30.
- **C-GET**: Usually used right after the C-FIND command, this one is used to retrieve an AE from the database and display it.
- **MWL**: The Modality worklist is used to get the list of scheduled procedures of a specific machine. So, the technologists know which patients are scheduled and their procedure details (Mildenberger, Eichelberg, & Martin, 2001).
- **MPPS**: This command is used to get the status of the machine for a given procedure. Retrieves the report with the start, the progress, and the completion of a specific procedure (Mildenberger, Eichelberg, & Martin, 2001).

To show a concrete example, here is an image of a MWL command with the filtering on top and the results of the machine at the bottom:

Figure 14 : DICOM MWL Visualization



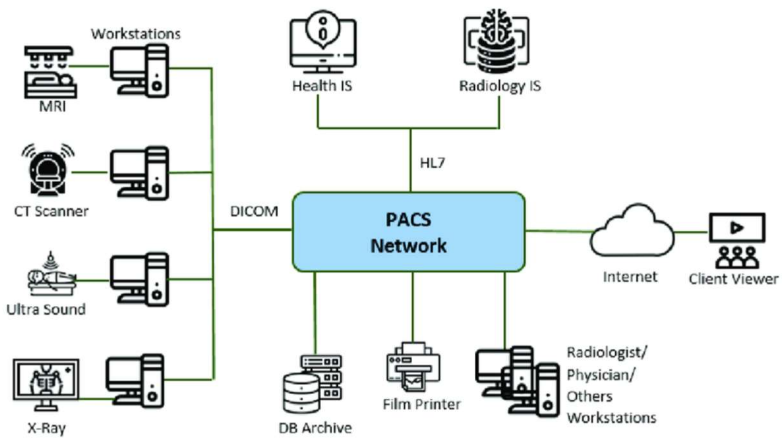
Source: https://dicomiseasy.blogspot.com/2012/04/dicom-modality-worklist.html

There are more of these commands, but we now have a good idea of how it works. The machine communication is a big plus with the adoption of the DICOM standard. Each services simplifies the workflow of the imaging procedure while ensuring that the transcription between the machines is accurate, efficient, and secure.

2.4.6. Picture Archiving and Communication System (PACS)

The PACS is a software responsible is a critical component responsible to manage the storage, retrieval, management, sharing and the presentation of medical images. PACS is a software that uses the DICOM standard to manage the overall data imaging of a hospital, research facility, laboratory and so on. There are multiple different PACS solutions that are available. Most of them are under a license cost but there are some that are open-source, custom made and free to use. A PACS system is divided in multiple components; the imaging modalities (MRI scanner, CT, X-ray) that outputs the medical image. A Secure network that will transmit the medical image around. Workstations for viewing the images and an archive section (the database) (Yan, Yu, & Ma, 2018). Here is a scheme that shows where a PACS system is located:

Figure 15 : PACS network schema example



Source: https://www.mdpi.com/1099-4300/24/2/176

As we can see, the PACS network is in the center of every medical imaging interaction and is here to manage everything from one single software. As we can imagine, setting up a PACS solution is challenging and complex. As it is highly critical, it does require a lot of planning, resources, and effort.

The reason why we are talking about PACS is that multiple hospitals and research facilities do have one and made the effort to integrate it for their management system. This is the main reason why the DICOM standard is still used today and will probably be used for the next decade if not more. While the CHUV does not have one, it is still possible to implement the DICOM standard to organize their EEG data, and this is the next chapter's topic.

2.4.7. DICOM and EEG data

DICOM was made to work with data imaging. It is efficient when it comes to handle the image that goes out of an MRI, automatically transferring it into a given database and adding all the necessary information (Yan, Yu, & Ma, 2018). But is it possible to apply this with electroencephalography research? Yes, and it may be a little bit more difficult as the EEG machine does not directly output the data into DICOM format, but it is still possible to achieve this. If the company already has a PACS, the conversion and the storage is made easy to do and does not require an extended effort (Lang, C., et al., 2023). But it is not the case of the CHUV that works with another system to store and manage their medical imaging data. Therefore, to be able to setup a DICOM system in the CHUV, we could for example transform the raw EEG data (the .bdf and .edf files seen in the BIDS chapter) and transform it into a DICOM readable files using a converter (this one can be an application or a code). Then, we could use this result and store it inside a normal folder database by organizing like the following example:

/DICOM

/Patient01

/EEGSession1

/EEGSession2

/EEGSession3

/Patient02

/EEGSession1

Then, we could create and develop a way to add the metadata directly to the DICOM folder and to store additional metadata to it. This could also allow us to create another tool to search efficiently for a specific criterion like every patient that is 50 or more.

2.4.8. DICOM limitations and challenges

DICOM is a standard that is invaluable for the medical imaging field and could also be a solution for our problem of organizing the EEG data of the CHUV. But like every standard, there are flaws and difficulties that we might encounter if we choose to use it. This chapter is here to understand the risks and to be aware of the problems we might encounter.

The first thing is that the DICOM standard is complex and requires a lot of understanding to work with in the first place. If we choose to go with this solution, we might struggle to teach the researchers how to use it properly as there are more than 2000 different attributes that can be added to a DICOM file (Paessler, 2023). Implementing the DICOM standard also can be challenging to implement inside a structure like the CHUV as they don't have a PACS system in place for EEG. And because it's impossible for us to implement this type of system, we might have some issues with the storage; Because there are no tools that currently organize the DICOM into a folder structure, we might need to create it. So, the development of such a tool will be a considerable effort (if we need to develop a whole GUI with every metadata possibility), there is a risk that the database could have some issues with the organization (which is already the case, then the job we did could be useless if not worse).

In conclusion, while the DICOM standard has significantly improved the storage, transmission, and interoperability of medical images, it comes with its own set of challenges. The absence of a PACS system can further complicate matters, especially when dealing with large datasets or aiming for efficient data management and retrieval.

3. Standard Evaluation

3.1. Evaluation Criteria

To evaluate a standard in the best possible way, we need to understand and relate points of comparison to the project. In our case, these are the points I'm going to use as a basis for comparing the two standards with a given weight and a given grade. The final score will be calculated based on these two factors described below. But first, here is the list of the different criteria:

- **Purpose**: weight = 10
  Every standard was made to respond to a specific problematic, and the problematic can vary. In our case, we want a standard that is developed for research purposes, this will ensure that it is convenient and made for the purpose of this project. If the standard is initially made to do something else, it can be difficult, and we may have to find a lot of workarounds to fulfill our need. This is the reason why this topic has a high weighting.

- **Main users**: weight = 9
  Same principle as above; The tool must be designed to match the use that we will use for this project. For the CHUV, the main users will be researchers.

- **Data structure**: weight = 7
  The data needs to be stored in a manner that is easily recognizable and makes it easy for the researcher to do some analysis on it.

- **Metadata**: weight = 9
  The metadata must be easily findable for the tool and for the researcher that wants to have information on the study. If possible, the metadata needs to be secured and have the possibility to be removed from the original file if we want to make it anonymous.

- **Flexibility and extension**: weight = 7
  For our project, we want to be able to add extensions and various tools to help us to convert and validate data easily. Not too flexible though, because we might make some mistakes and have some incomplete data if we do so.

- **Interoperability**: weight = 5
  In our case, it’s not mandatory as we do not need to output the EEG devices directly into the format. It would be interesting to have this if the CHUV is adopting the standard entirely, but this would be another project. However, it is still a criterion that would be interesting to have in the comparison.

- **Anonymization and Privacy**: weight = 6
  As we will work with sensitive data, we do need a secured standard. Even if the data will stay inside the CHUV database, it is still a criterion to evaluate if we compare two standards.

- **Storage space**: weight = 7
  As all the data will be restructured, it is important that the metadata that will be added to the raw data does not cause the addition of too much space in the database.

- **Community support**: weight = 8
  This point is important for a standard. As we want to have a maximum of people adopting it and it is generally important to have someone that is available whenever we have a question. As most of the standards are open source, the diminution of active community means less updates and development.

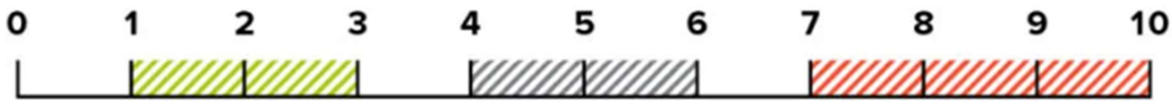
- **Complexity and implementation**: weight = 9
  It is important for a standard to have a smooth adoption whenever someone is willing to change. If the standard is too complex, less people will adopt it and it could be a major loss of time for the people working with it. If it is too complex, one might make a mistake and create a failure in the data storage.

- **Modality coverage**: weight = 5
  This one would be important if we want a complete adoption of the standard. In our case, we don’t need to have every modality possible. It is, however, a good criterion to evaluate the scope that is possible to have.

- **Limitations**: weight = 8
  Standards are not perfect; this criterion rates the overall problematics that could happen when adopting one. The higher the note, the less impactful is its limitations.

Grade meaning:

Figure 16 : Grade



Author's source

0 - 1 Nonexistent or not applicable for this project.

1 - 3 Hardly applicable or does require too much effort to set up.

4 - 6 Does have it but requires some setup or workarounds to make it work, not ideal

7 - 10 Ideal solution, does fit the project and can be implemented.

Weight meaning:

0 - 1 Does not have any impact on the project.

1 - 3 The impact on the project is minimal.

4 - 6 Does have an impact on the project, need to be considered.

7 - 10 Have a big / critical impact on the project, must be chosen meticulously.

3.2. BIDS Evaluation

Before evaluating the standard, we need to rank the criteria with their importance regarding this specific CHUV project. The ranking with a note on 1 to 10 is based on the explanation above. The higher the weight, the higher the importance of the criteria. Then, a note is applied to each criterion as well as a little explanation regarding the note. In the end, we can calculate a score based on the results that we have found here by multiplying the note with the weight for each criterion, making the sum of everything and dividing the result by the weight multiplied by the perfect score. We then only need to multiply the result by 100 to have the matching score in percentage. We can then compare the results to choose the best of them. Here is the table evaluation for the BIDS standard:

Table 1 : BIDS standard evaluation

	BIDS	Weighting	Note	Explanation
Purpose	Research-focused	10	10	Does match the project needs pretty good
Main Users	Neuroscientists and researchers	9	9	Highly active community that develops tools and extensions everyday.
Metadata Handling	Separate JSON and TSV files	9	9	Metadata is easy to find, can be removed. Hereditary principle to avoid repetition.
Complexity	Initial learning curve but standardized	9	7	Need to understand how the basic tools work. Easy to manage afterwards.
Community Support	Growing tool and pipeline support in research	8	8	Highly active community that develops tools and extensions everyday.
Limitations	Manual conversion required, not suitable for all modalities without	8	6	Manual conversion can take time, some pipelines are under dev. Extensions can be found for some
Data Structure	File and folder organization	7	9	Folder organization, easy to comprehend and to search information inside
Flexibility + extension	Allows various extensions	7	8	New technology so not compatible with everything. Need to install an extension.
Fill size and storage	Optimized for efficient storage	7	9	Good optimization storage, the files are genreally pretty small.
Anonymization and Privacy	Structured for de-identification	6	9	The metadata of the patient can be removed + tools that validate the de-identification
Interoperability	Growing acceptance in research community	5	6	Needs to be done manually with extensions that can transform raw data to BIDS.
Modality Coverage	Focused on neuroimaging with extensions for others	5	7	Some extensions are still under development. Good coverage overall.

Author's source

The final score for the BIDS standard is 82%. That means that the BIDS standard matches well with our project.

3.3. DICOM Evaluation

To evaluate the DICOM standard, we then apply the same criteria with the same weighting, but we change the results with the other standard. We then re-evaluate each criterion and we apply the same calculation to find the final score. Here is the second table for the DICOM standard:

Table 2 : DICOM standard evaluation

	DICOM	Weighting	Note	Explanation
Primary Purpose	Clinical use	10	4	The project is more research oriented. The DICOM clinical use is not interesting here.
Main Users	Radiologists and clinicians	9	5	The main users on the standard are not the ones that will use our tool.
Metadata Handling	Embedded in Image files	9	6	Metadata is kind of hard to find as it is inside each file. Lot of repeatability
Complexity	Can be challenging due to vendor variations + complex	9	4	Cannot be fully adapted to the needs of the hospital, complex to implement
Community Support	clinical device support	8	8	The active community is smaller than BIDS. But better device support.
Limitations	Proprietary changes by vendors, data loss due to conversion	8	8	The limitations are not the real problem in this case, complexity is.
Data Structure	Patient > Study > Series > Image	7	6	The structure and the naming convention is hard to comprehend by only looking at it.
Flexibility + extension	Rigid structure	7	7	The rigid structure of DICOM means that you have less mistakes. But too much work.
Fill size and storage	Large files due to metadata	7	6	Files are larger, it takes more time to search inside the data + takes more space
Anonymization and Privacy	Potential for identifiable information in headers	6	6	The Data cannot be shared as the patient information is in the file's header.
Interoperability	Universal standard in clinical settings	5	8	Pretty good implementation in the medical field. Known workflow implementation
Modality Coverage	Wide range of medical imaging modalities	5	9	Covers a lot of modalities as it is here since more than 20 years. Good coverage.

Author's source

The DICOM standard is having a matching score of 61.8% for our project.

3.4. Conclusion

In the process of evaluating standards for the reorganization of CHUV's EEG database, we have made a deep analysis of both BIDS and DICOM. Each standard was examined through multiple angles, considering various criteria weighted according to their relevance to the project's demands. BIDS, with its novelties and his consideration to the research field, resonates closely with the project's aim to provide a research environment. It is flexible and yet has a structured approach to data and has strong community support and it is relatively easy to implement, had an impressive compatibility score of 82%. This indicates a big alignment with the project's objectives, boosting the research utility and efficiency.

On the other side, DICOM, the veteran in the field, has shown good in in established practices and universal recognition within clinical settings. However, its focus on a field of medical imaging more than the EEG field and is complex for the customization and data management. It is a less ideal fit, reflected by a 61.8% compatibility score. While DICOM has an infrastructural solidity and a widespread adoption is undeniable, the specificity and evolving nature of the CHUV project call for a standard that prioritizes agility and research-centric functionalities, aspects where BIDS holds a better solution.

BIDS emerges as the more suitable standard for this project, promising an environment where research data is not only accessible but also axed on growth and innovation. The choice of BIDS aims for a forward-thinking approach, one that anticipates the needs of researchers and adapts to the ever-evolving landscape of medical data analysis. This decision encapsulates our need to create a database that is not just a repository of information but a dynamic toolkit, that can be queried and filtered for our need and pre-processed to be analyzed by the CHUV researchers.

4. Evaluation and technology choice

4.1. Introduction

Before we start selecting the technology we'll use, we need to understand what the scope is, from the initial raw EEG data to the end point where we can analyze the research data. This introduction will explain how we are going to process to decide of the technology we will use.

We will begin with raw EEG data, which we will transform by adding necessary details (metadata) to it. This data must then be converted into a structured format known as BIDS. Once in BIDS format, we'll check the data to make sure it fits our standards. If it passes the check, it gets stored in a database; if not, we must fix it according to the feedback provided. An essential part of our work will be developing a way to search through the stored data easily. We'll need to decide whether to build our own search tool or use an existing one (if it exists), and then how to best set it up to meet our needs. This could be a web-based service, a desktop program, or a desktop program on the web-based service.

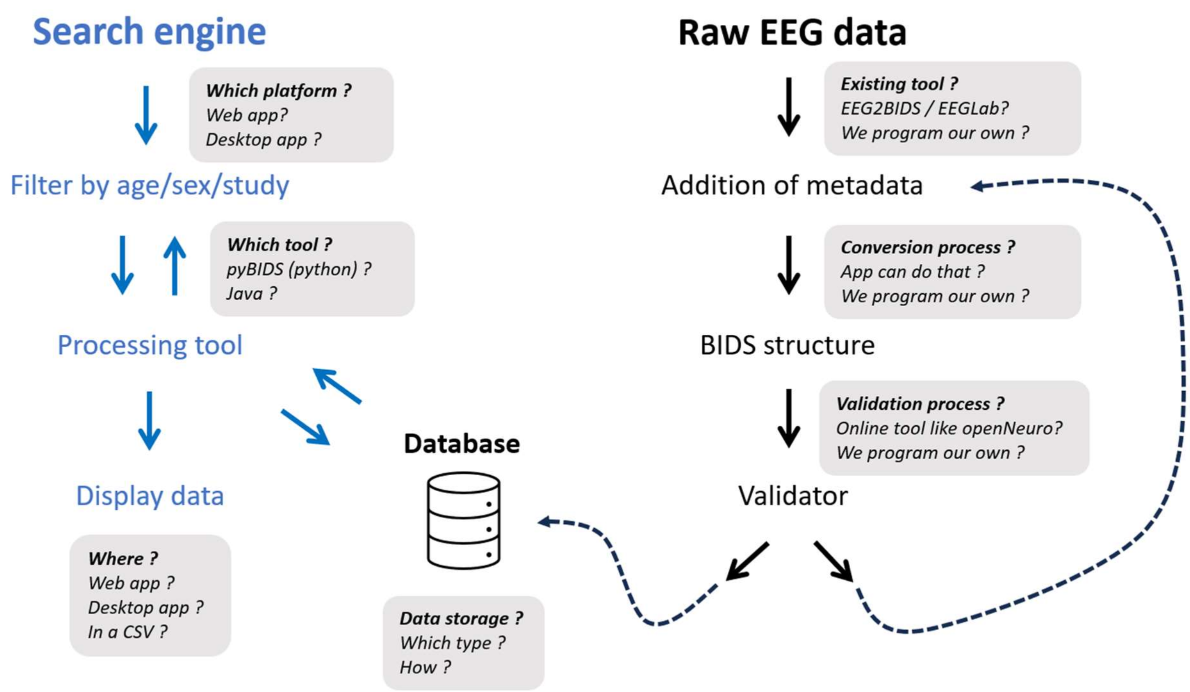
Throughout this chapter, we will answer a series of questions that will guide us in choosing the right tools and technologies. We'll investigate different software options for transforming data, adding metadata, and ensuring everything matches the BIDS format. Then we'll decide if we should use the existing CHUV database or find a better alternative. Finally, we'll choose the programming language and development approach for our search tool.

The decisions we make are not just theoretical; they are important for the successful implementation of our solution. We will choose carefully with specific criteria and a comparison with the CHUV's needs, as they are the future users of these tools. With a clear plan and the right tools in hand, we will be ready to move into the implementation phase.

4.2. Action Map

Now that we know which standard to take and before choosing the different technologies, we need to have a good understanding of how we are going to work, how we are going to proceed from the raw data to the final BIDS research and analysis. Here is a map indicating the process as well as the questions that we will need to answer in this chapter:

Figure 17 : Action Map Schema



Author's source

Here is what we know; we will receive raw EEG data that we will need to modify by adding some information to it (metadata). Then, we will need to transform this into the BIDS structure that we have chosen above and then validate it. If the data is validated, it can be stored inside the database. If not, the data needs to be modified according to the recommendations (so we go back to the data modification). We also must develop a tool to be able to search inside the database. We will need to enter some filters, grab some data inside the database and then display them somewhere.

Now that we know the process, some questions arise. We need to focus on data transformation first (how to transform raw data into BIDS basic structure). For that, multiple applications do exist and need to be reviewed on whether they could help us with our task or not. Depending on the first choice, the second one is how we can add the metadata into the BIDS structure. This can also be done by applications but can also be done by hand or automatically. So, we will need to choose but because it is directly linked to the first question, we will answer both questions in one chapter. Then, we will need to validate the data that we have transformed to be sure that it corresponds to the BIDS format.

There are a lot of extensions and tools that exist for that, it won't be a problem to implement but we still must choose and evaluate them to choose the right one.

Then, we need to store this data inside a database. For our project at CHUV, understanding the compatibility of these databases is critical. The existing database at CHUV, primarily based on file storage, aligns well with the requirements of the BIDS standard, offering a compatible and efficient solution for managing neuroimaging data. This file storage system provides a straightforward and adaptable framework, making it a preferable choice for BIDS implementation. In contrast, other database types, which are not currently implemented in the neuroscience area of the CHUV, such as relational or object-oriented databases, present complexities in integration and adaptation for BIDS. Their structures, often designed for different kinds of data handling, address multiple challenges in terms of customization with the specific needs of BIDS-based EEG data management. Therefore, choosing CHUV's existing file storage database emerges as the most pragmatic and effective approach for this project.

Finally, we have the search engine to implement. We need to check whether there is an existing tool that does the job or not. If not, and therefore that we make the conclusion that it is better to develop our own, we need to choose whether it's better to do a web application that would run on a local server or if we develop a desktop application that runs locally and analyze the local database. Once we know this, we will debate on the code language for both front and back end.

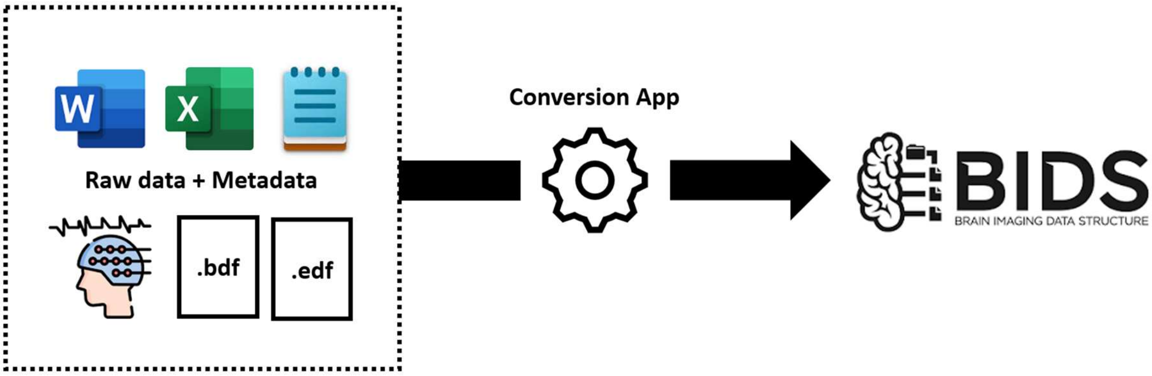
Once we have an answer to all these questions, we will start the implementation process (and the development if we need to). While they are theoretical, all these questions are important and will play a key role in the implementation of the solution. To be sure that we have made the right decision, we are going to rely on both the choice of the comparison (with defined criteria that we will create) and the confirmation of the CHUV that will use these tools (and might prefer one over another).

4.3. conversion process

4.3.1. Introduction

The conversion process will therefore represent the transformation of raw data to the data ready to be validated. We are also going to evaluate their ability to add metadata because these two themes are often included inside the same application.

Figure 18 : Conversion process schema



Author's source

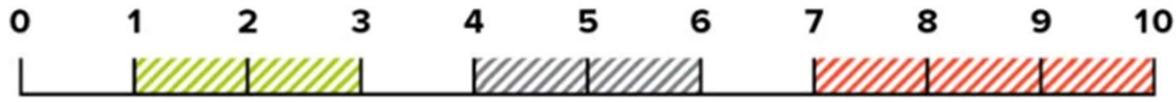
If we look at the possibilities available to us, there are only three of them. The first is EEG2BIDS, which, as its name suggests, transforms .bds files (a type of raw data) into BIDS files. It can also be used to add metadata files (manual procedure). The second is EEGLab, the largest system to date developed for transforming and using EEGs. It allows semi-automatic transformation of raw data into BIDS format. It can also be used for data pre-processing and analysis. The last one is much more manual and has no ready-made applications. This is MNE-BIDS, a tool developed using Python that allows users to manually transform data into BIDS format. After analyzing and comparing each of these methods, we'll draw a conclusion so that we can keep just one.

4.3.2. Evaluation criteria

To evaluate the conversion process tools and give them a score, we will base ourselves on some criteria and will rank them from 0 to 10. We did not implement a weight factor for this comparison as the criterion have an almost equal importance in the decision.

Grade meaning:

Figure 19 : Grade



Author's source

0 – 1 Nonexistent or not applicable for this project.

1 - 3 Hardly applicable or does require too much effort to set up.

4 - 6 Does have it but requires some setup or workarounds to make it work, not ideal

7 - 10 Ideal solution, does fit the project and can be implemented.

Criteria list:

- **Ease of installation and setup:** On one hand, by judging the difficulty of installation (for example, one-click vs compiling from the source) and on the other, the need to install additional dependencies or libraries to use the tool.

- **User interface and GUI**: Includes ease of use, understanding for a new user, and the available documentation. The interface should be user-friendly.

- **Addition of metadata**: How metadata can be added in a way that is organized, the user's margin of error when completing. Can multiple users organize in the same way?

- **Speed of conversion**: From raw data to BIDS organization, how quickly can this be done, assuming that the user is experienced in the field?"

- **File formats:** ability to integrate and transform several file formats such as .bdf, .edf, BrainVision, etc...

- **Error handling:** how the program handles BIDS containing problems and how it displays them. Are errors easy to understand and clearly explained by the software?

- **Integration of other tools**: Is it possible to integrate other EEG analysis tools into the software? Is the software flexible with the implementation of other software/workflows? Can another script be run from the application?

- **Community and support**: Availability and responsiveness of the various communication media (email, forums, etc.). Development activity (updates, bug fixes, etc.) and the abundance of online resources on this subject (tutorials, videos, Q&A).

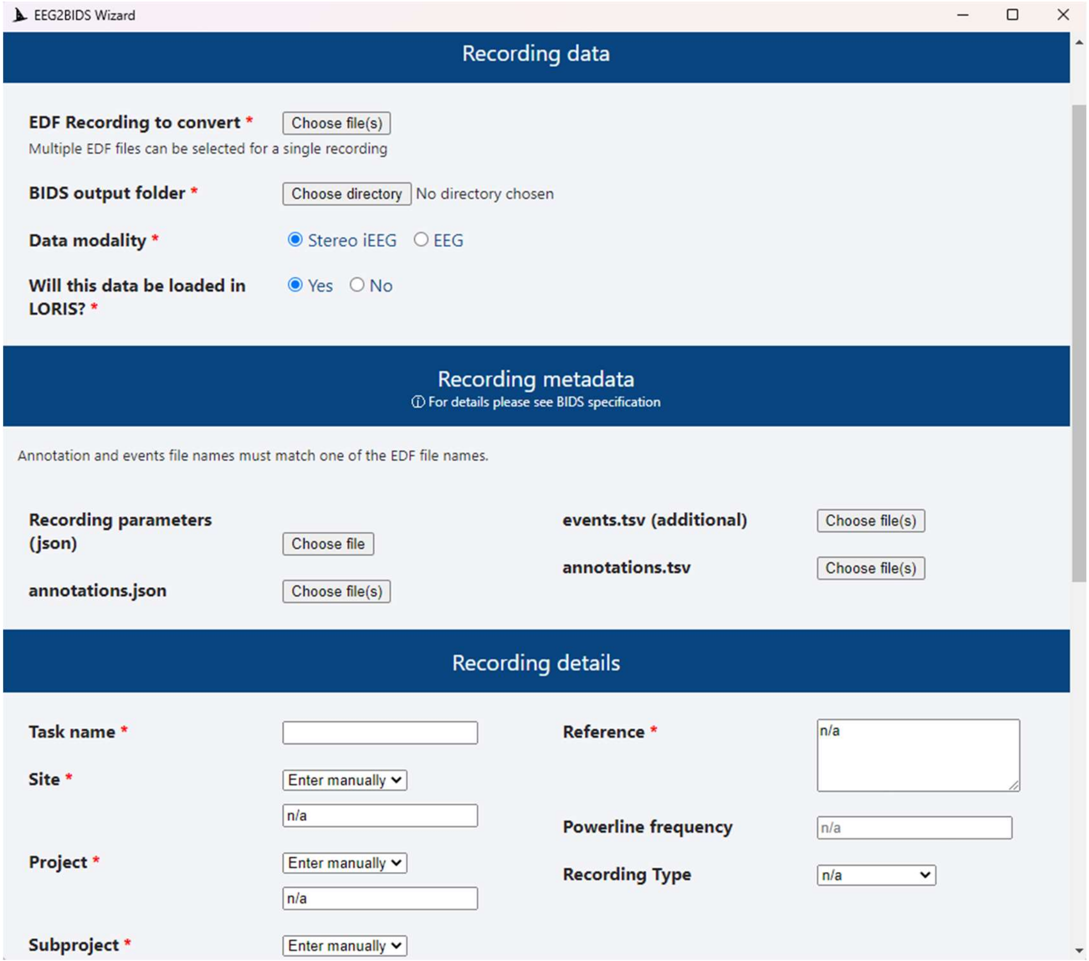
- **Integration of the BIDS validator:** does the tool contain a means of assessing the quality of the BIDS file, does it allow understandable errors to be returned and/or warnings to be given so that the error in the dataset can be understood and changed?

Every one of these criteria will be noted from one to ten based on their compatibility with the project. Then, the notes will allow us to create a compatibility score to determine which one is better.

4.3.3. Evaluation of EEG2BIDS

EEG2BIDS is GUI interface that allows us to transform EEG and iEEG data into the BIDS format. The app can be installed easily from git with a simple installer (Aces, 2023). It has a functionality that allows us to work directly with a LORIS (longitudinal online research and imaging system) application but because this tool does require a dedicated server and the CHUV IT cannot provide one, it cannot be considered for this project (Lo, 2023). The good side of this application is that the GUI is easy to use, there is not much to know beforehand regarding the use of the application itself and it's made to work with EEG and iEEG. The GUI is also nicely designed and has an implemented BIDS validator that tells us the warnings and the errors that the BIDS dataset contains. The downside of it is that we cannot create a study with it, the participants need to be recorded one by one and added to a folder manually. Another down part is that we need to create the BIDS almost manually, which can result in different annotations and therefore create some confusion in the organization and the naming process. Also, the application is not flexible with the file format as it only accepts .edf files and is mainly made to work with LORIS application. Here is an image of the application while transforming the raw data into the BIDS format:

Figure 20 : EEG2BIDS Visualization



Author's source

Now here is the table with the criterion, the grading on 1 to 10, and a small explanation on the decision of the grading. Note that the evaluation is not only based on the criteria itself, but also on the compatibility with our project.

Table 3 : EEG2BIDS conversion evaluation

	EEG2BIDS	Grade	Explanation
Installation and Setup	Easy to install and setup	9	No requirements, simple .exe installation, just considered as a dangerous application by windows.
User interface and GUI	Very Good user interface and GUI	10	The GUI itself is easy to understand, well made, design and simple
Addition of metadata	Not really good	4	Can add some of metadata, but not flexible. No ability to add more, too few of them can be added.
Speed of conversion	Good	7	If you know what you're doing, there are few metadata to complete and everything is in the same page.
File formats accepted	really bad	1	Only one file format accepted (.edf) and it is not the one that is most used in the CHUV.
Error handling	Good	7	The third page tells you what is good, the errors and what need to be changed. Does not go into small details.
Integration of other tools	Bad	3	The tool was made to work with one application which is called "LORIS". No other extensions accepted.
Community and support	Good	6	The last git update of the tool was made two months ago. The community is still active but not many people use it.
BIDS validator integration	Works but not perfect	5	The tool works fine, you know which files are good and which are wrong. But no more info/warnings.

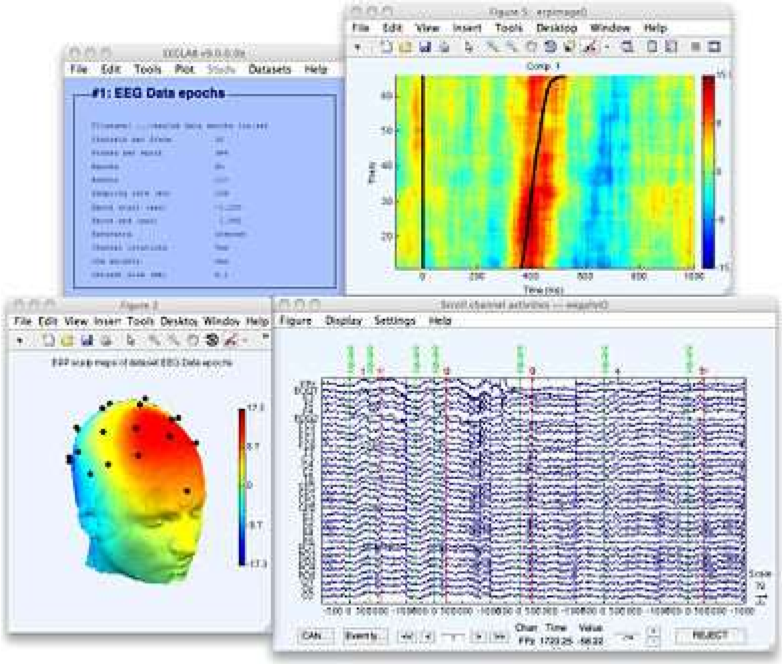
Author's source

The percentage of compatibility that EEG2BIDS has with our project, based on the results of the table above is 58%, which tells us that EEG2BIDS is good but is probably not the answer to our project.

4.3.4. Evaluation of EEGLAB

EEGLAB is the most used tool to transform EEGs into BIDS format. More than the transformation, we can also do the pre-processing and the analysis of the EEG itself. The tool is open-source and has the capability of having the possibility to add extensions from a catalogue that other developers have created. The catalogue of tools is entirely free and can be installed in one-click by the user. These options are the main reason why this tool is still up-to-date and still today, some extensions are maintained, modified, and even created. Another advantage is that we can import and transform almost every form of raw data we can find, and the addition of metadata is well guided, simple and has many possibilities (SCCN, 2023). Here is an example of EEGLAB in use:

Figure 21 : EEGLAB Presentation



Source: https://sccn.ucsd.edu/eeglab/index.php

EEGLAB is not easy to use if one does not have the medicine vocabulary and if it is the first usage. But the many tutorials, explanations, and forums that the community has created make the learning part enjoyable. The tool can be installed two different ways. One is from the MatLab software as an extension. This one is great because it allows us to be connected to the community and therefore to install the extensions and plugins, but it comes with the cost of having a MatLab license (1000.- per year per license, according to the CHUV IT team. This times 90 researchers). The other one is the desktop version that uses a MatLab runtime but does not require a license. It comes with the downside of not having the possibility of installing external extensions, but if we compare the potential gain of not having to pay the licenses, it is a really good alternative (SCCN, 2023).

Table 4 : EEGLAB conversion evaluation

	EEGLAB	Grade	Explanation
Installation and Setup	Easy to install and setup	8	No requirements, simple .zip installation, just need to install the mathlab runtime aswell (wizard helps you)
User interface and GUI	Good user interface and GUI	7	The GUI is understandable, there are many options but they are well organized. Need to have basic vocabulary.
Addition of metadata	Really good	10	You can add a lot of metadata, the tool guides you so two people would organize them the same way.
Speed of conversion	Good	7	If you know what you're doing, it goes pretty fast. You can add a lot of metadata, may loose time if no experience.
File formats accepted	Really good	10	There are a lot of file formats accepted. Almost every EEG machine output can be imported into the application
Error handling	Good	7	There is a console that is always running on the back where you can see every error as detailed as possible.
Integration of other tools	Very good	9	The tool was made to work with extensions and other tools, possible to run scripts and install your own tool.
Community and support	Very Good	9	The project has a good community that is really active. There is a wiki, tutorials and frequent updates
BIDS validator integration	not Integrated	1	Not integrated with the desktop version, if you upgrade it to the mathlab version, it would be an 8.

Author's source

With the score of 75,6%, the EEGLAB tool is convenient for our project and may be the solution to fulfill our needs.

4.3.5. Evaluation of MNE-BIDS

MNE-BIDS is a python package that allows the user to write and read BIDS datasets. As it is a python tool, it obviously requires understanding python language and to be familiar with coding in general. If we have these criteria and we want a solution totally customizable and free to use, this may be the right tool (MNE-BIDS, 2023). For our project, we may consider it as it is maybe the right solution if the other ones can't be installed or do not fit the CHUV requirements. As it is fully customizable, it will always be the best solution with the cost of someone that needs to code an entire GUI. If MNE-BIDS is in this list, it is also because there is an active community around and a well-made documentation that explains how to use and exploit this package. Here is an example of how to convert the data to the BIDS format (there are multiple steps before that, this is only the final command:

Figure 22 : MNE-BIDS Code example

```
# Load the data from "2 minutes eyes closed rest"
edf_path = eegbci.load_data(subject=subject, runs=run)[0]
raw = mne.io.read_raw_edf(edf_path, preload=False)
raw.info["line_freq"] = 50  # specify power line frequency as required by BIDS
```

Source: https://mne.tools/mne-bids/stable/auto_examples/convert_eeg_to_bids.html

Even if the documentation is well-written, this implementation would still require a lot of skill and a lot of work to be adapted to a useful GUI. This solution could also complete another one if we arrive at the conclusion that the other tools are not working or if there is a custom need from the CHUV that needs to be developed.

Table 5 : MNE-BIDS conversion evaluation

	MNE-BIDS	Grade	Explanation
Installation and Setup	Easy to install not to setup	4	It's a python library, so you need to have something that can use python first, then you can install it.
User interface and GUI	No user interface, no GUI	1	As it is a python library, there is no GUI available
Addition of metadata	Duable	6	You can add a lot of metadata, good guides online but still requires to code.
Speed of conversion	Good	7	If you know what you're doing, it goes pretty fast. You can add a lot of metadata, may loose time if no
File formats accepted	Duable	6	There are a lot of file formats accepted but the conversion needs an understanding of code
Error handling	May be hard	4	You can see the error in the terminal but it is sometimes pretty hard to understand.
Integration of other tools	Not ideal but can do	5	You can install other depedencies and other packages, but you may need to code the link between them.
Community and support	Good	7	The project has a good community that is really active. There is a wiki, tutorials and frequent updates
BIDS validator integration	Can be Integrated	5	As you need to code to get MNE-BIDS, you can include an other library for the validation. Still need some code

Author's source

With an overall compatibility score of 50%, the MNE-BIDS tools are not the best solution here. However, with enough time and competencies, we can code an application GUI with its structure that would be 100% customizable. This would require a lot of time and skill.

4.3.6. Conclusion

In conclusion, transforming raw EEG data into a format ready for validation is a vital process in our project. We've examined three tools: EEG2BIDS, EEGLAB, and MNE-BIDS, each with its strengths and weaknesses.

EEG2BIDS is user-friendly and directly interfaces with LORIS, but it's limited in file format support and requires manual input, which might lead to inconsistency. Its compatibility with our project is moderate as we do not plan to implement the LORIS system. Even if the tool can be used without this implementation, the EEG2BIDS application does not accept as many raw data types as we need to. This make the EEG2BIDS complex to implement for a CHUV application.

EEGLAB stands out as a complex and useful tool. It allows extensive data transformation and analysis and supports a wide variety of file formats. Although it requires a license for full functionality, it is totally usable without it and allows us to transform our raw data into a BIDS format without any issue. It is also flexible because we can execute scripts for our datasets within the application. Its strong community support and extensive resources make it a strong candidate, with a high compatibility score for our needs.

MNE-BIDS offers full customization but demands a high level of coding expertise and significant development time to create a user-friendly interface. While its adaptability is a plus, the coding process to obtain a comfortable solution comes short on immediate practicality.

Considering ease of use, functionality, and compatibility with our project requirements, EEGLAB seems to be the most promising choice. It is user-friendly and has comprehensive yet complex features, suggesting it is a good tool for efficiently converting EEG data into BIDS format.

4.4. Search tool

4.4.1. Introduction

In this part of our discussion, we're going to explore a tool that helps us dig into a BIDS database to find specific information. This search tool is crucial because it lets us quickly and easily locate the datasets that contain information like a participant's age, or the type of stimuli used in their brain imaging tests. Understanding how to pull out this information from the structured setup of a BIDS database will be our focus here. We'll break down what a search tool needs to find, why we need it, and how it navigates the complex layers of data we're dealing with.

Because the BIDS database is well-organized, with everything labeled and, in its place, it is easier to make our search tool to know where to look. We'll also touch upon the platforms that can host these vast amounts of data and discuss why we've decided to build our own solution: A desktop application created to meet the unique needs of handling sensitive data within the secure folder database of the CHUV.

To demonstrate better where the information is stored, here is a list of what we need to find according to the CHUV researcher Chrysa Retsa:

- **The age, sex, and nationality of a patient**
  C://BIDS Database > Dataset > participant.tsv file

- **The test and dataset information, stimuli**
  C://BIDS Database > Dataset > README.txt file

- **The frequency, event, and the number of channels**
  C://BIDS Database > Dataset > sub-xx > eeg > .json file

The goal of the search tool will be to scrape this information according to the filters selected and showing the dataset that match the requirements with useful information like the link of where it is stored, a little description of the dataset and the dataset name.

4.4.2. Choosing the right platform

When it comes to storing all our important brain imaging data, we need a safe place for it as we will store sensitive patient data. We considered OpenNeuro, a place online where researchers share brain data. It's user-friendly and great for finding and storing information, but there's a catch. Before we can move our data in, we must remove all personal details from it, and store only the information that is important for the research (Open Neuro, 2023). Hospitals and researchers find this process a bit too much as it requires to do a mountain of paperwork just to keep our things in a storage unit. We have asked an OpenNeuro expert regarding the possibility to create or implement an OpenNeuro site in a local server but unfortunately it is not possible. There are multiple websites that are created for brain data sharing but none of them are designed for a more private use.

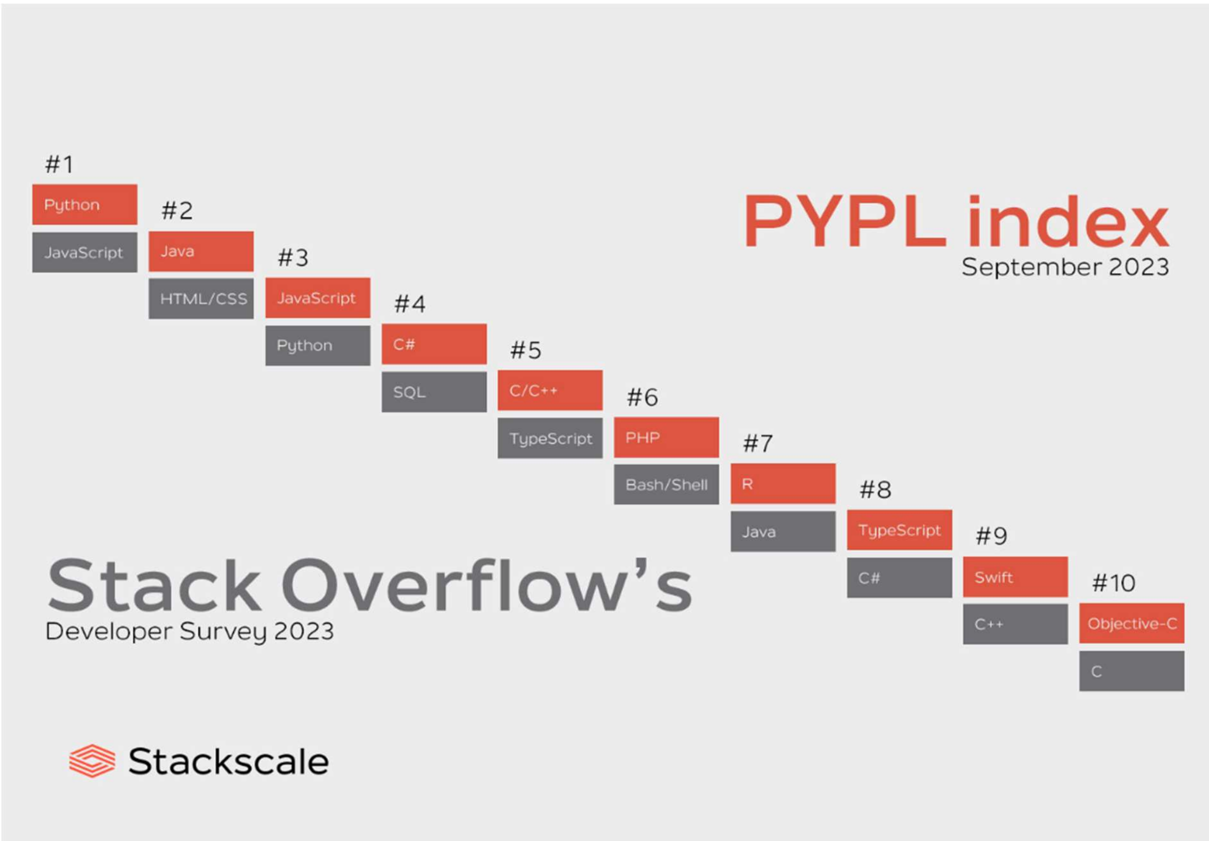
Because of this problem, we have decided to build our own application. A tool that would be deployed inside the CHUV computers, that would work with their local database directly, a tool that could search for the brain imaging data that we would have store with the EEGLAB tool. We're going for a desktop application as it does not require an internet connection for it to work. This way, our data stays put, safe and sound within the hospital's or research institute's walls without having to worry about external security.

We know the internet can be risky for private information, which is why our desktop application will be given out like a secure key, directly from the CHUV internal servers. This means researchers can easily 'pick up' the tool from within the safe bounds of our network or get it installed by the IT team, and start using it with confidence, knowing that the sensitive data won't be exposed to the world. We also considered developing the application on the local web server, this would have been a good idea because the researchers wouldn't have to worry about the application installation. But unfortunately, after discussing with the IT team, they refused to implement a local web-application as they don't have procedures to accept and install this type of application.

4.4.3. Tool and language selection

We decided to make a custom application, a tool that we will create, designed just for our research needs. But to make this tool, we needed to pick the right materials. In the world of software, this means selecting the programming languages for both the backend (the code logic) and the frontend (the code to display), that will help us build our application just the way we need it. The language is important as it a fundamental for the structure of the tool. When choosing a language for our code, it's important to stay as simple as possible. For example, if there are plugins that we need to use that are coded in a certain language, we still can code in another one, but it will complexify everything. Also, if we think about extending our code, it's important to choose a language that is known by most of the programmers. Again, an example, if we choose to code in Ruby or in Objective-C, there are some chances that the person that will be in charge to complete our code has never seen this language and, therefore, will struggle to develop it. For the backend, according to the Popularity of Programming Language index (PYPL index), and the Stack Overflow's survey, here are the most used languages programs in 2023:

Figure 23 : PYPL and Stack Overflow's language use index

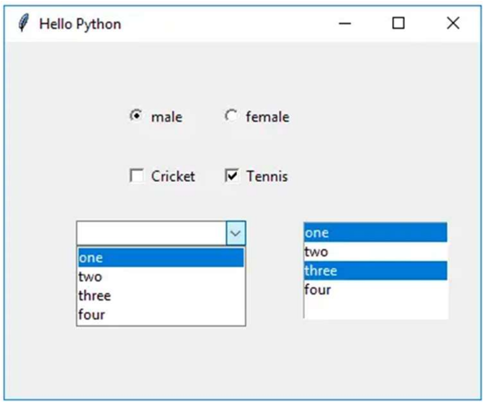


Source: https://www.stackscale.com/blog/most-popular-programming-languages/

For all these reasons, we picked Python, a computer language that's first in the PYPL index and 3rd in the Stack Overflow's survey. Python lets us use something called PyBIDS, a set of instructions already written that understand the way brain imaging data is organized. If we had gone with other languages like C# or Java, we'd have to write that tool from scratch or connect it with unusual ways, which would take more time and might not work as well. Python also functions well with other tools we're using, like MNE-BIDS, which helps us understand and analyze our brain data better.

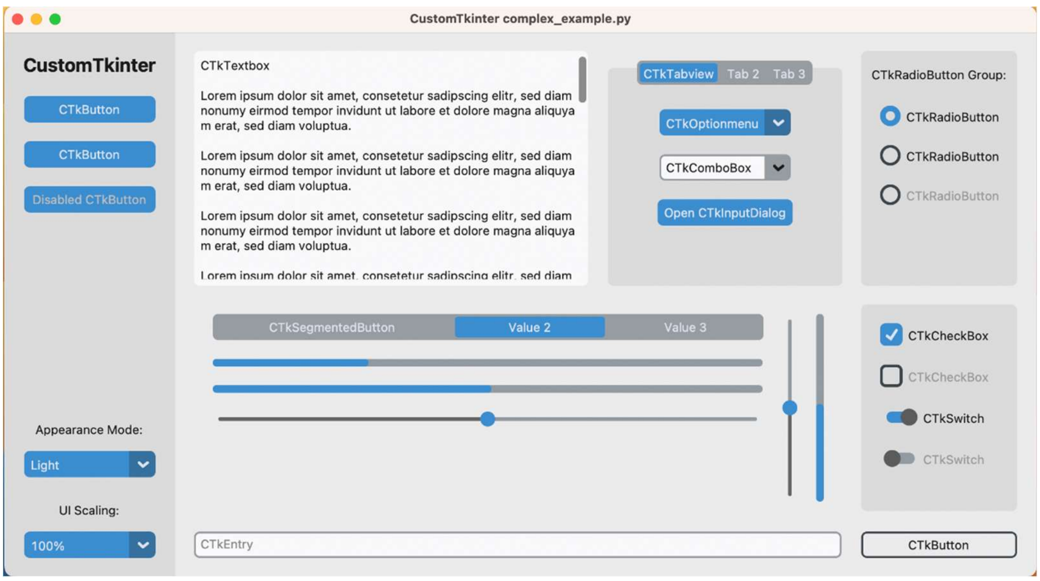
Now for the front end, we choose a library that is easy to implement as it works with the same language as the back end. This extension is based on an old one that was renovated to the actual standards. It is called Ctkinter and it can be used the exact same way as Tkinter. Here is an example of both to demonstrate what they are looking like:

Figure 24 : TKinter Visualization



Source: https://www.tutorialsteacher.com/python/create-gui-using-tkinter-python

Figure 25 : CTkinter Visualization



Source: https://github.com/TomSchimansky/CustomTkinter/blob/master/documentation_images/complex_example

5. Implementation and Testing

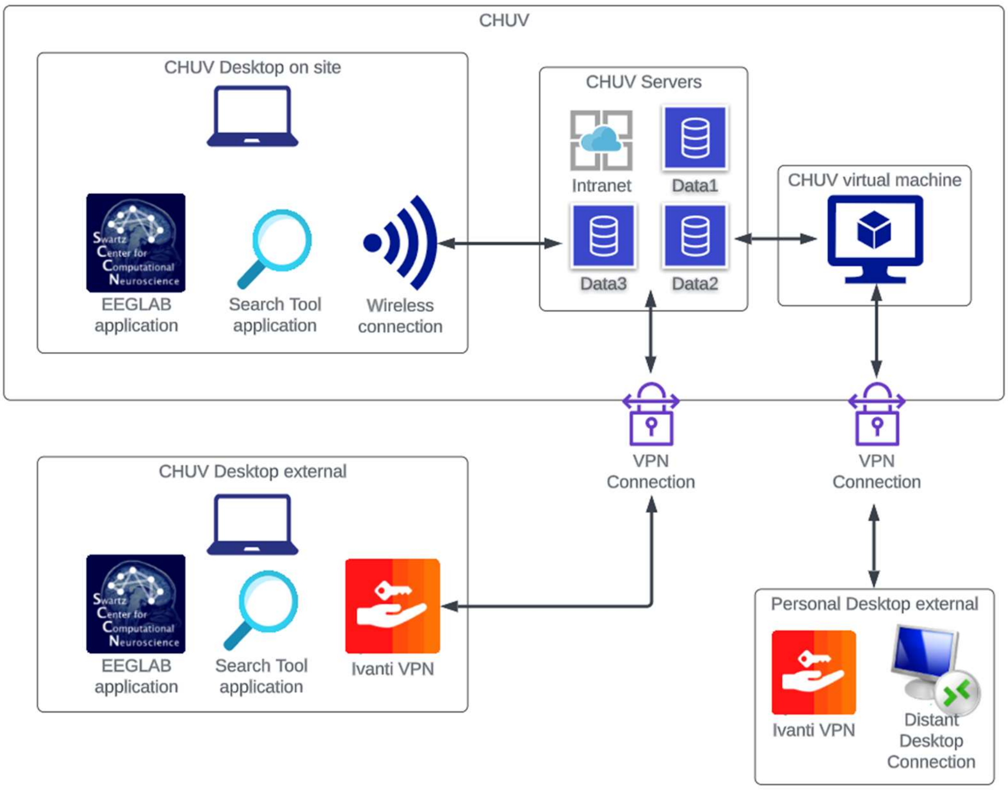
5.1. Overview of the CHUV Environment

Before implementing and developing the two applications, we need to dig a little deeper into the CHUV Environment. Every company has different ways to work, different organizations, different methods, and infrastructures. This chapter is dedicated to understanding these differences and align our application and way to work with the CHUV. This part is crucial if we want to resolve the problematic the best way possible with as little inconvenience as possible.

In this section, we will talk about multiple important points regarding the pre-implementation, the implementation itself and the testing of this solution. After that, we'll dive into the documentation part as one does need to know how to debug and develop further these tools. Finally, we'll review the challenges and their solutions as well as the feedback from the researchers on the use of these tools.

Before implementing the two applications, we need to know and understand the actual architecture that the CHUV has. To have a better visualization of the situation, we designed a diagram with its description to better represent how the systems are working together:

Figure 26 : CHUV Architecture Diagram



Author's source

Systems Definition:

- **CHUV**: Represent the CHUV site.
- **CHUV servers**: The servers that contains the three file databases and the intranet server.
- **CHUV Desktop on site**: The desktop that the researchers have. They have a unique identifier, and the IT team can install both of our applications on it. Can access the CHUV servers with a wireless connection.
- **CHUV Desktop external**: The desktop that the researchers have. They need to have a VPN connection to have access to the CHUV servers. They can have our application used even externally.
- **CHUV Virtual machine**: The virtual machine that we have access to as an external user that does not have a personal machine. We can't install our applications on it.
- **Personal Desktop external**: The desktop that we have access to. We do need Ivanti VPN, and the Distant Desktop connection application to connect to the CHUV virtual machine. Our applications can't be used from here as we do not have a direct access to the CHUV database.

5.2. Integration of EEGLAB

5.2.1. Introduction

The implementation of the EEGLAB tool is simple as it is available directly on their website as a .zip file that contains a .exe installer (SCCN, 2023). There are still prerequisites that both the computer and the user need to have for the installation and to be able to use it. These are the following.

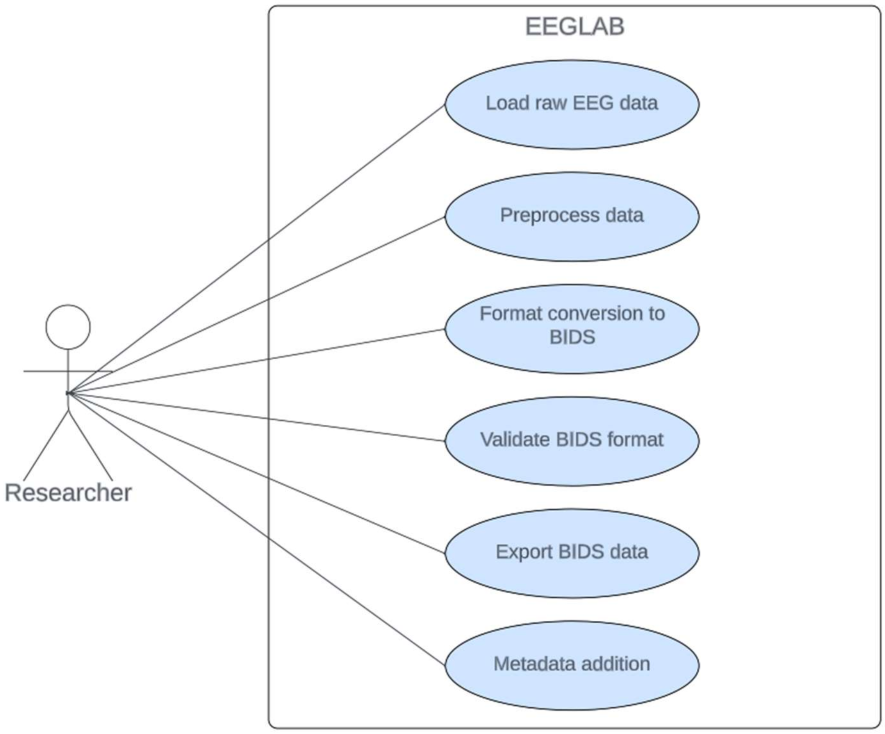
1. The user does need to have access to a personal computer known as a HOS number (unique ID). Without this, the IT team can't install the application and the user can't install it by himself.
2. The user needs to contact the IT team to have its application installed. This is because the security system in place at the CHUV is quite strong, this fact comes with the drawback that the user cannot install anything and everything on their machine.
3. The user needs to have the right to download two applications on his machine: the EEGLAB application as well as the MATLAB runtime (installed automatically with it).
4. In the case of BIDS conversion, the user needs to have a write access to the BIDS database.

If we do have all of these, we can go on and install, use, and work with this application. If we do not have one of these, this can restraint the intended use of the EEGLAB tool.

It is important to understand that this tool will change the usual workflow the researchers are used to. They will need to adapt and learn the new structured way of storing the data and this can be frustrating to do for some of them. Because of this frustration, the researchers can do the things the wrong way or use the tool not properly. To resolve this issue, it is important for the CHUV to train the users, so they know how to use the tool. As we are not part of the CHUV, we cannot ensure that the users will get proper training. However, we can provide them with a guide so that they can share this knowledge with their colleagues. This guide is available on the chapter 4-8 of the Appendix.

To better represent the actions that will matter for the implementation of EEGLAB, here is a Use case diagram and its definition:

Figure 27 : EEG Use case diagram.



Author's source

Use Cases definition:

- **Load Raw EEG Data:** Researchers load raw EEG data into EEGLAB.
- **Preprocess Data**: Filtering, artifact removal and other necessary preprocessing before the data can be converted into BIDS format.
- **Format conversion**: Convert the preprocessed EEG data into BIDS format using EEGLAB.
- **Validate BIDS format**: Ensuring that the converted data adheres to the BIDS format.

- **Export BIDS format**: Researchers can export the BIDS-formatted data for storage.
- **Metadata addition**: Researchers can create or update metadata and add it to the created document as required by the BIDS standard.

5.2.2. Implementation

To implement a non-official application that was never installed before on a CHUV-secured computer, we, as a person who is not part of the CHUV-IT team, need to send to the team a user-guide with two parts; the first one is the installation part which is in the chapter 2 of the Appendix III, and then there is the minimal test part, to ensure that the application works on the machine, which is located at the chapters 4-8 of the Appendix III. Then, the IT team needs to create a report based on the test, the results obtained and why the application should be installed and used in the CHUV. Finally, once the application is recognized by the hospital, it’s added to their catalogue and the researchers can ask for a simplified installation by giving their name, HOS number and why they need to use the EEGLAB tool.

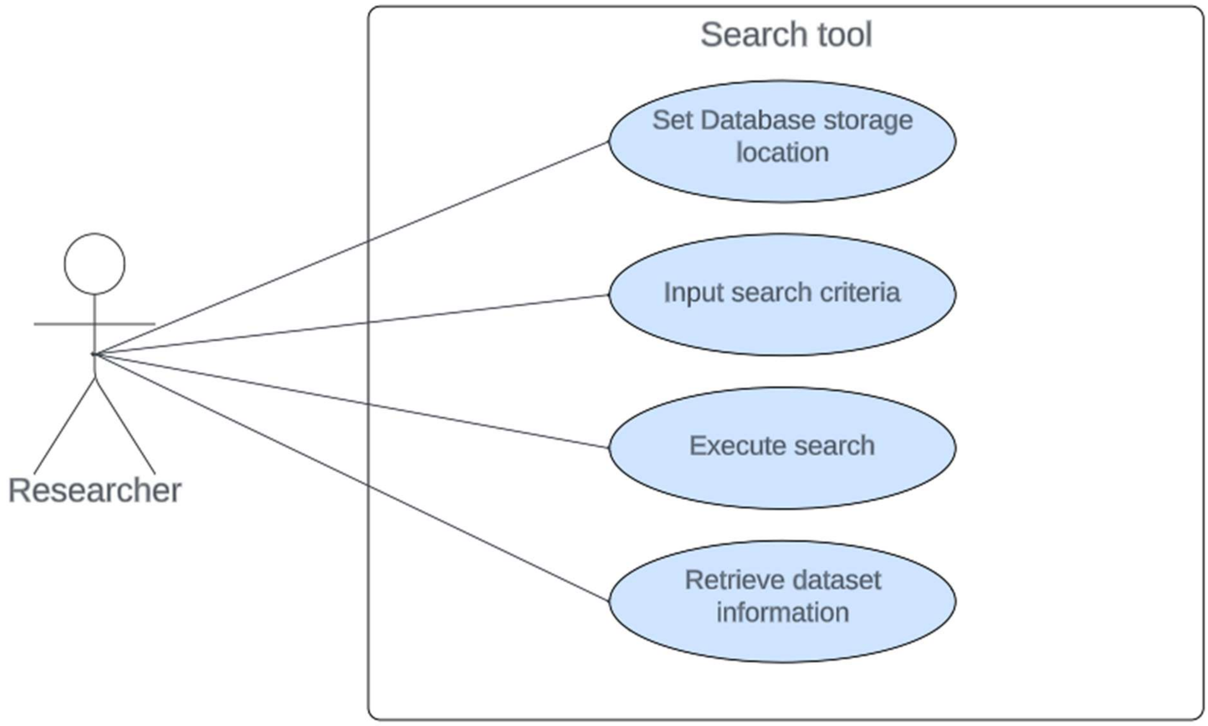
To contact and initiate the IT-review process, we need to first, have access to the CHUV intranet, which can be done by installing a VPN called Ivanti and connecting to a specific web page with the Distant desktop connection application that is installed by default on the windows computers. Then, we arrive on an ERP where we need to launch the CHUV’s intranet called “TRIBU” and access a file called “Formulaire DSI”. From here, we can access the “Installation d’un logiciel non-standard” page that is located under the “Demande de logiciels et matériel”. Once this is done, we need to fill and send the form to the IT team. If the procedures took time to get (more about this in the next chapter), the implementation itself is not complicated to do.

5.3. Development and Integration of the Search Tool

5.3.1. Introduction

The implementation of the search tool application is quite different because for this one, we need to develop our own solution. This changes the fact that the CHUV needs to accept to install an application that was never seen before, has no reviews and has never been tried in production. All these changes means that we need to document our solution as best as we could so that the IT team can make a report and controls that are solid enough to ensure that there won't be a major issue with the installation of this application. Before developing a solution like this one, we need to have the assurance on some questions that we had; After some telephones calls, we now have the assurance that they are working with an internal folder server that is comparable to a USB key or a Disc access on our computer. To have a better overview of what our tool must be capable of, here is a use case diagram and its definition under:

Figure 28 : Search tool use case diagram



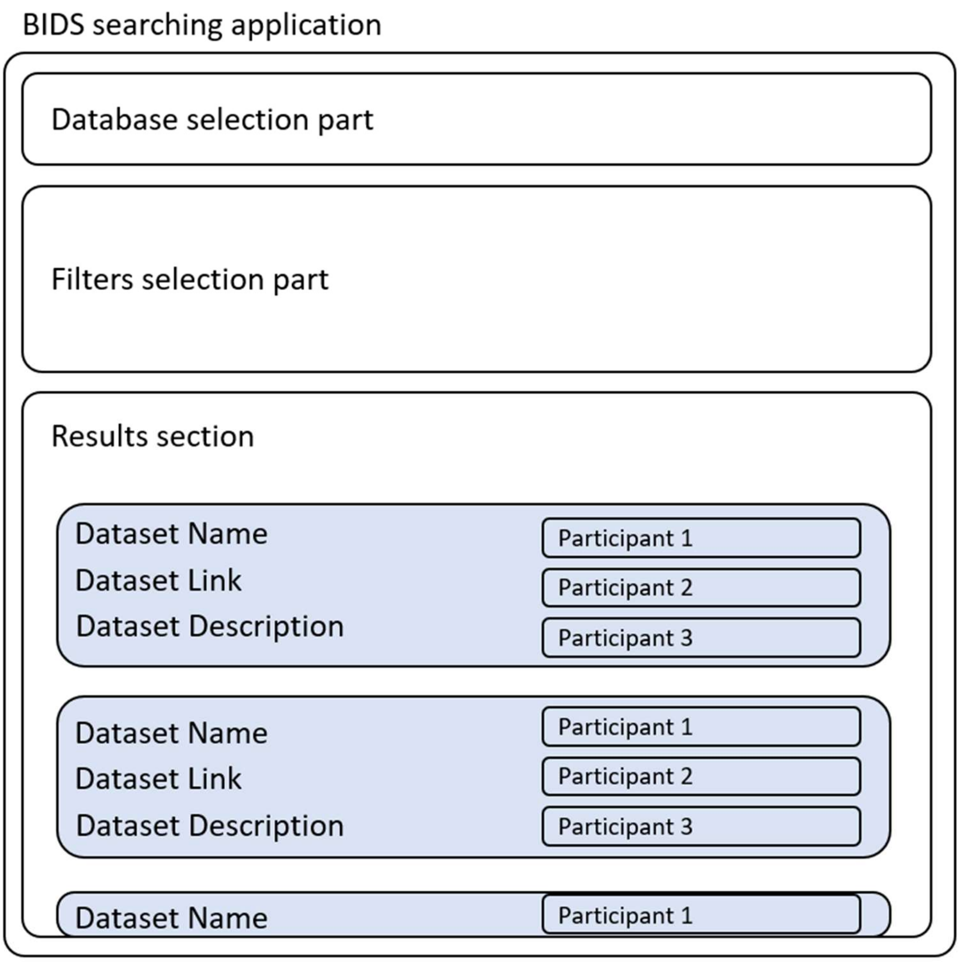
Author's source

Use case definition:

- **Set database storage location**: Researchers set the storage location of the BIDS dataset at the beginning of their session.
- **Input search criteria**: Researchers input different criteria such as age, sex, dataset name and stimuli.
- **Execute search**: The tool searches the BIDS database based on the specified criteria.
- **Retrieve dataset information**: The tool retrieves and displays a list of datasets that match the search criteria, including link to the storage location, the dataset name, the dataset description, and the filtered participant information if any.

Now that we have this information, we know what to do, where it will be stored and how to search based on our knowledge of the BIDS standard. We now must design the UI to accept all these actions the simplest way possible. Because we have few parameters and results to display on the screen, we proposed a structure that looks like the following:

Figure 29 : Searching application structure proposition.



Author's source

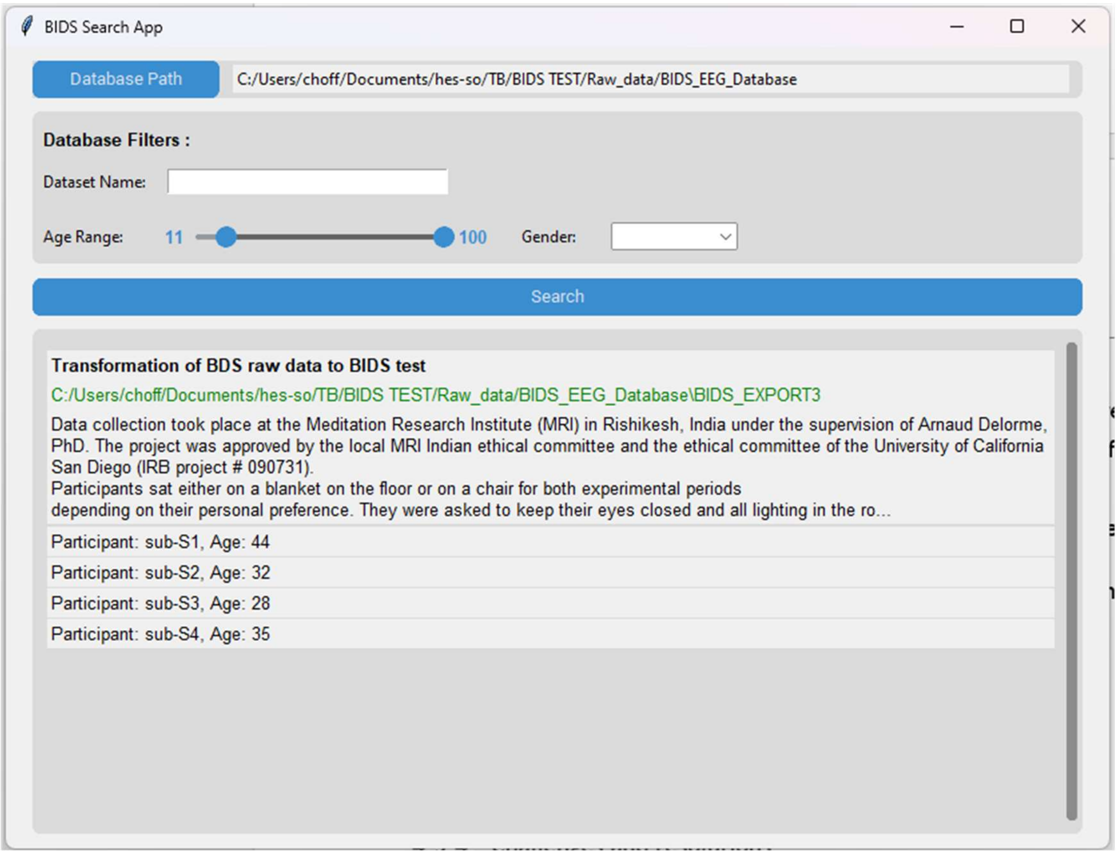
Here, we would have a Database selection part that would contain the selection of the database where we want to perform the query. This was mandatory to have because at the time we need to code our application, we do not know where the database will be situated, and the tool needs to be tested and validated on multiple different computers with multiple different database places. This section is here to select this link.

The next section is the Filters selection. It has a purpose to select the different filters based on the different criteria. Fill in the different filters, press the "Search" button and the results will automatically be displayed on the section under.

The result section will contain none to multiple datasets based on the filters and the datasets available. For each dataset that matches the filters, it will display the name of the research (can be found in the dataset_description.json file), the dataset link (can be retrieved if the dataset needs to be displayed), and the description will be found in the README.md file that could be mandatory to have the stimuli, the event and a description of the task. Because not all the participants fit the criteria, it's also important to display them, so we added them to the result list.

After some coding, here is the result obtained based on the schema just above:

Figure 30 : BIDS Search app result



Author's source

As we can see, the result looks a lot like the schema we did above. There are two more things; the Search button that we have placed in the center of the screen and another button on the top left to search for the database directly with the file system of your computer. Also, we have moved the participant list on the bottom of the dataset as the list is sometimes too long to put it on the right.

5.3.2. Implementation

Installing and implement the Search application in the CHUV, it's quite the same process as the one we have done to implement the EEGLAB tool. Here is how to do that:

First, as we have our code developed on python, need to release it somewhere so that the CHUV can download and use it. For that, we choose to release it on GitHub as a .zip file that outputs a .exe that directly launches the application when open. As we are not recognized authors by windows, there is a warning message when the app is launched for the first time. We only need to click on the "Execute anyways" button to launch it.

Then, to implement a non-official application that was never installed before on a CHUV-secured computer, we, because we are not part of the CHUV-IT team, need to send to the team a user-guide with two parts; the first one is the installation part which is in the chapters 3 of the Appendix III, and then there is the test part, to ensure that the application works on the machine, which is located at the chapters 9 of the Appendix III. Then, the IT team needs to create a report based on the test, the results obtained and why the application should be installed and used in the CHUV. Finally, once the application is recognized by the hospital, it's added to their catalogue and the researchers can ask for a simplified installation by giving their name, HOS number and why they need to use the Search tool for BIDS application.

5.4. Testing

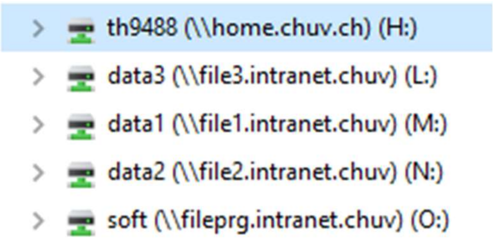
Because we do not have access to the CHUV database, manual tests were conducted to try to simulate as best as possible the EEGLAB tool as well as the search tool we developed. The manual tests were done on a different computer to ensure that there are no issues related to the export of these programs. To simulate the CHUV database as best as possible, we must know what we have at our disposal, which information do we have, and try to simulate them as closely as possible.

The first thing we need to replicate is the folder database they have. To do so, we have done a test on a USB Key that had write, read and sometimes no access to generated files. For the EEGLAB application, we needed to be sure that the application was not modifying the files that had no/read access. For the search application, we needed to be sure that it could read the files with reading access but be blocked when it comes to reading files with restricted access. Here is a comparison of what we know and what we have:

What the CHUV has

The chuv has a database structure like the following:

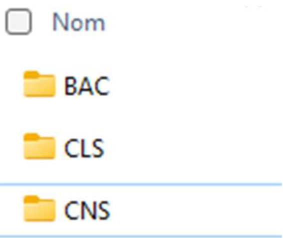
Figure 31 : CHUV database structure



Author's source

The database containing the patient information is a simple folder structure. We do have access to some and not to others.

Figure 33 : CHUV Folder Structure



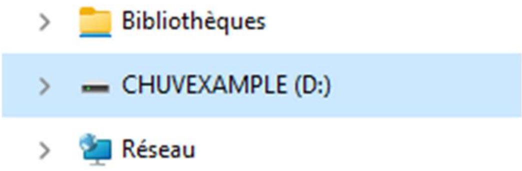
Author's source

Then we have the raw EEG data, as we do not have access to these files, here is what we know:

Our simulation

We have tried to replicate it by mounting an external database to our computer:

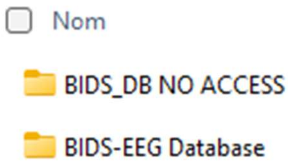
Figure 32 : Simulated database structure



Author's source

We simulated this folder structure by implementing two different databases. We do have access to the first one, not to the second.

Figure 34 : Simulated Folder Structure

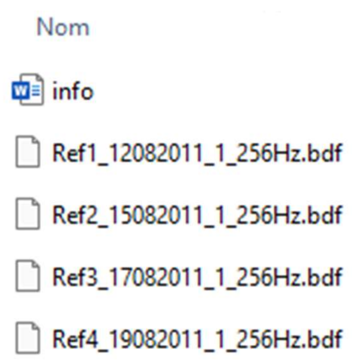


Author's source

The Data we found is publicly available online, the only difference is that the data is de-personified (there is no personal data):

Metadata is stored how the researcher want. It can be a Word, Excel, etc. The raw data supports .bdf or .edf extension.

Figure 35 : Raw data Simulation



Source: This data is part of a larger dataset available here: https://nemar.org/dataexplorer/detail?dataset_id=ds003061

We are now able to test our two applications manually by simply following the user guide. If the results are the same than the images, we can consider that the application works successfully.

5.5. Documentation

Having a well-explained and complete documentation is crucial for a project of this type as the application is meant to be maintained and extend. This way, the next person who will oversee this project will not lose much time on the code organization and the overall comprehension of the tool. The main one is a user guide that is developed for the researcher and the IT team. The file contains every detail regarding the installation process and on how to test the tool for both the EEG2LAB and the BIDS Searching tool.

The explanation is made in a step-by-step way that includes, for every step, an image, and an explication on what the user should do. The user guide is designed to be understood by someone that has no experience in the development or in the medical field. Apart from these chapters, there is also an advanced guide for using EEGLAB where we explain how to add metadata to a BIDS dataset. This chapter will be useful for the researchers that need to have an advanced comprehension of the tool and will transform the raw EEG data to a complex BIDS format.

6. Management

6.1. Work methodology

6.1.1. SCRUM

In software development, SCRUM represents a framework used to structure the development of an application or a technology. Regarding this project, we have tried to apply this methodology to have an organized technique. It is an agile methodology, meaning that it is flexible for the variability of events that can occur during a project (number of workers variability, response to change, unpredicted events, and so on). The SCRUM methodology provides a structure but doesn't dictate every move one should make; it's more about providing guidance and a set of principles to follow (SCRUM, 2023).

There are three primary roles within a SCRUM team:

- **The SCRUM Master**: It is the manager of the team, he helps to facilitate the process, resolve the potential issues, and guides the team to follow the SCRUM values and practices.

- **The project owner**: He is the person responsible for following and maximizing the results of the project. He also needs to manage the product backlog and ensure that everyone is following the priorities.

- **The Development team**: A cross-functional group responsible for developing a potentially releasable product at the end of each sprint.

There are also SCRUM events to know when we want to work with this methodology (SCRUM, 2023). The most important ones to know are the following:

- **The Sprint**: Event that is organized every 2 to 4 weeks, where everyone review what point could be considered as “done” and discuss about the sprint planning to come.

- **Daily scrum**: Everyday, the team should create a plan of what to do for the next 24 hours.

- **Sprint review**: Event where the team will investigate the product backlog and can also change it accordingly to the project.

To better understand the principles of this framework, his transparency and the focus, there are two files that are mandatory to have. The first one is the product backlog. It is a file that contains a list with everything that is needed for the project. All requirements that are known for the project are in this file. We also have the Sprint backlog, that contains a set of the product backlog items selected for the sprint. Plus, a plan that shows where the team is from the delivery of the sprint goals.

6.1.2. SCRUM adaptation for this project

Since there is only one person in the development and in the organization, it is necessary to adapt the SCRUM methodology and make things clear. The roles of SCRUM master and developer are performed by one and the same person. The product owner (PO) is provided by two representatives working with the CHUV.

During the first meeting, the initial product backlog was validated by everyone, and it was planned to separate the project into four different sprints of two weeks starting every two Tuesday. The product backlog created could be separated in different themes:

- **State of the art**: The learning part of what is an electroencephalogram, what are the BIDS and the DICOM standards and what are they used for.
- **BIDS and DICOM comparison:** Comparing and evaluating the two standards.
- **Tool development**: Developing the research tool and manage the implementation and testing part.
- **Report creation**: Creation and redaction of this report.
- **Application Improvements**: Potential improvements of the application.

The product backlog can be found in its integrity in the Appendix II.

Other than this, we have created and documented a logbook with the hours spend on this project. This document can be found in the Appendix I.

6.2. Meetings and communication

To facilitate organization and manage administrative tasks, communication was primarily done via email. This medium was chosen for its convenience and ability to provide a written record of the exchanges. In instances where matters were too complex or required immediate attention, telephone conversations were the preferred mode of communication. This choice was due to the efficiency and often clearer understanding that verbal communication can provide.

Bi-weekly sprint meetings were conducted using the Teams platform. Invitations to these meetings were given through email, ensuring timely and efficient coordination among team members. This approach to meetings and communication was integral to maintaining a smooth workflow and ensuring that all team members were aligned and informed about ongoing tasks and developments.

7. Conclusion

7.1. Results Obtained

The objective of selecting the most appropriate standard for the reorganization of CHUV's EEG database was a challenging process. After an analysis of both BIDS and DICOM, with various criteria relevant to the project's demands, BIDS emerged as the best choice. Its flexibility, structured approach to data, and strong community support, with the research-oriented objectives of CHUV, resulted in its selection. Also, BIDS achieved an impressive compatibility score of 82%, indicating a significant alignment with the project’s goals.

As for the conversion process, the evaluation of EEG2BIDS, EEGLAB, and MNE-BIDS tools highlighted distinct advantages and limitations of each. EEG2BIDS, while user-friendly and directly interfacing with LORIS, was limited in its file format support, making it less ideal for our purpose. On the other hand, EEGLAB was better for its extensive data transformation capabilities, support for a wide variety of file formats, complexity, and user-friendliness. Even with the absence of MATLAB licenses at CHUV, which initially posed a challenge, the desktop version of EEGLAB provided a feasible solution that does not include costly licenses and still catered to our project needs. MNE-BIDS, with its full customization capacity, was set aside due to its intensive coding demands and lack of immediate practicality.

Regarding the search tool, we decided to develop a custom solution using Python and CTkinter. This approach allowed for greater flexibility, enabling us to customize the tool precisely to the needs of CHUV without the problem of licensing fees. By utilizing the pyBIDS library for the backend and CTkinter for the frontend, we were able to craft a tool that not only met the project requirements but also provided room for future expansions and customizations.

Looking at the potential impact of these implementations on CHUV, the benefits, though still in the process phase, are expected to be substantial. The introduction of a searchable database and the integration of tools like EEGLAB assure to revolutionize the way EEG data is analyzed and processed. Our search tool, specifically designed for CHUV's requirements, could significantly enhance the efficiency of data retrieval and management, enabling researchers to filter and access data more easily.

In conclusion, while the full implementation and integration at CHUV are ongoing, we are close from transforming EEG data management. The methodologies and tools selected and developed in this project are not just solutions to current challenges but also foundations for future storage management in neuroimaging metadata management.

7.2. Future improvements

Our search tool can be more than just a basic tool for looking up information. It's got the potential to be a multitool, one that not only finds what we need quickly but also makes sure everything is in order. Right now, the tool allows the researcher to filter details like age, gender, name, and description. But we can give it some upgrades. The first one would be new ways to filter datasets, like filtering by the record frequency of the EEG or by the number of electrodes that were used during the examination.

The tool could also make sure our data is correct before we even start searching. This is possible by adding features that validate the data, ensuring that necessary files like README are present and not empty, or making sure our dataset descriptions make sense. There are tools that could be used for this like the BIDS validator made by OpenNeuro but does require to send the data to an internet server that is not ours, or we can customize and create our own to make sure that the CHUV information are relevant and ordered the same way (a stricter way than the BIDS validator itself that looks the overall problems).

Then there's the part about working with EEGLAB. Our tool could be a bridge where we could send information (like a specific dataset) back and forth with just a click, saving us the effort of doing it manually. It's a lot more complex to set up, but with code and enough time, it can be done.

Because the sharing part is also important in the research field in medicine, it also could be interesting to be able to share some filters result easily by clicking and sending to another researcher. This could be easy to set up but could cause some security issues.

Finally, we'll make sure that every data set is unique, with its own name that can't be mixed up with another, which is as crucial as labeling samples correctly in a lab. And to be sure that the researchers don't make a mistake when saving data, we also could automate this part to label and validate the BIDS that we want to save in the database.

7.3. Difficulties Encountered

There were two challenges when coming to the implementation of EEGLAB. The first one is that the CHUV does not provide a MathLab license to the researchers. Therefore, this license that costs around 1000.- per year per person is forcing us to come up with a free solution. This free solution is the desktop version of the tool that is installable directly from the official website. But it comes at the cost of not having the possibility to implement external extensions and tools. Which is not a problem for the problematic of this thesis but is less complete than the MathLab version. The good thing about this is that it is less hard to validate as it does not have the risk of installing a bad tool from the public library.

The second issue that we encountered during the implementation is that we needed to have access to the intranet to start the submit process and to confirm that our tool could work on their system. This was a long procedure as we need to have a contract with the CHUV to have access to the data. We also need some extra access to be able to connect with the VPN, these two things have taken one month and one week in total. In the effort to make this task faster, we were in direct contact with the IT team and some managers multiple times to ensure that we had access to their system and to submit the application installation process as fast as we could.

REFERENCES

Aces. (2023). *GitHub - aces/EEG2BIDS: EEG2BIDS Wizard: a tool for converting raw EEG and iEEG data into the BIDS standard data structure, prepared for LORIS (Longitudinal Online Research and Imaging System)*. Retrieved from GitHub: https://github.com/aces/EEG2BIDS/tree/main

Ae, B. (2022, 10 7). *DICOM Operations*. Retrieved from PACS Boot Camp: https://pacsbootcamp.com/dicom-operations/

Angus-Leppan, H. (2007). *Seizures and adverse events during routine scalp electroencephalography: A clinical and EEG analysis of 1000 records*. Retrieved from Clinical Neurophysiology, 118(1), 22–30.: https://doi.org/10.1016/j.clinph.2006.08.014

Azabou, É., Navarro, V., Kubis, N., Gavaret, M., Heming, N., Cariou, A., . . . Sharshar, T. (2018). *Value and mechanisms of EEG reactivity in the prognosis of patients with impaired consciousness: a systematic review*. Retrieved from Critical Care: https://doi.org/10.1186/s13054-018-2104-z

BIDS. (2022). *Benefits and Converters*. Retrieved from Brain Imaging Data Structure: https://bids.neuroimaging.io/benefits

BIDS. (2023, 9 26). *About BIDS*. Retrieved from Brain Imaging Data Structure: https://bids.neuroimaging.io/

BIDS. (2023, 9 26). *Governance*. Retrieved from Brain Imaging Data Structure: https://bids.neuroimaging.io/governance.html

BIDS. (2023, 9 26). *Meet the BIDS Steering group*. Retrieved from Brain Imaging Data Structure: https://bids.neuroimaging.io/2019/12/31/meet-the-bids-steering-group.html

Community, B. (2023). *BIDS-validator*. Retrieved from npm: https://www.npmjs.com/package/bids-validator

DICOM. (2023). *Dicom Tags*. Retrieved from Dicom Library: https://www.dicomlibrary.com/dicom/dicom-tags/

Electroencephalography - Brain Imaging Data Structure V1.8.0. (2022). Retrieved from BIDS-specification: https://bids-specification.readthedocs.io/en/stable/04-modality-specific-files/03-electroencephalography.html

Extending the BIDS Specification. (2022). Retrieved from bids-specification: https://bids-specification.readthedocs.io/en/v1.2.1/06-extensions.html

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., . . . Nichols, B. (2016). *The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments*. Retrieved from Scientific Data, 3(1): https://doi.org/10.1038/sdata.2016.44

Graham, R., Perriss, R., & Scarsbrook, A. (2005). DICOM demystified: A review of digital file formats and their use in radiological practice. Dans *Clinical Radiology* (pp. 1133-1140). https://doi.org/10.1016/j.crad.2005.07.003.

ISMS. (2023). *Standards for the medical industry*. Retrieved from ISMS: https://www.isms.online/sectors/iso-standards-for-the-medical-industry/

ISMS. (2023). *Technical Standard*. Retrieved from ISMS.online: https://www.isms.online/glossary/technical-standard/

Kirschstein, T., & Köhling, R. (2009). *What is the Source of the EEG?* Retrieved from Clinical Eeg and Neuroscience, 40(3), 146-149.: https://doi.org/10.1177/155005940904000305

Lachaux, J., Rudrauf, D., & Kahane, P. (2003). *Intracranial EEG and human brain mapping*. Retrieved from Journal of Physiology-paris, 97(4-6), 613-628.: https://doi.org/10.1016/j.jphysparis.2004.01.018

Lang, C., Winkler, S., Koren, J., Huber, M., Kluge, T., & Baumgartner, C. (2023). *DICOM® integrated EEG data: A first clinical implementation of the new DICOM standard for neurophysiology data*. Retrieved from Clinical Neurophysiology, 155, 107-112.: https://doi.org/10.1016/j.clinph.2023.07.008

Lenartowicz, A., & Loo, S. K. (2014). *Use of EEG to diagnose ADHD*. Retrieved from Current Psychiatry Reports: https://doi.org/10.1007/s11920-014-0498-0

Lo, D. (2023). *Features*. Retrieved from LORIS: https://loris.ca/features.html

Mildenberger, P., Eichelberg, M., & Martin, E. (2001, 09 15). Introduction to the DICOM standard. University Hospital Mainz, Langenbeckstrasse 1, 55131 Mainz, Germany.

MNE-BIDS. (2023). *MNE-BIDS 0.13 documentation*. Retrieved from MNE tools: https://mne.tools/mne-bids/stable/index.html

Open Neuro. (2023). *User Guide*. Retrieved from Open Neuro: https://docs.openneuro.org/user_guide.html

Paessler. (2023). *What is DICOM? Definition and Details.* Retrieved from Paessler, the monitoring experts: https://www.paessler.com/it-explained/dicom#:~:text=space%20and%20materials.-,What%20are%20the%20disadvantages%20of%20using%20DICOM%3F,or%20is%20incomplete%20or%20inaccurate.

SCCN. (2023). *Compiled EEGLAB.* Retrieved from eeglab: https://eeglab.org/others/Compiled_EEGLAB.html

SCCN. (2023). *What is EEGLAB?* Retrieved from EEGLAB.: https://sccn.ucsd.edu/eeglab/index.php

SCRUM. (2023). *Introduction to srum events.* Retrieved from scrum: https://www.scrum.org/resources/introduction-scrum-events

SCRUM. (2023). *What is Scrum?* Retrieved from scrum: https://www.scrum.org/resources/what-scrum-module

Springer, B. H. (2023, 9 28). *Standards and System Integration in Digital Medicine.* Retrieved from https://link.springer.com/chapter/10.1007/978-3-540-74571-6_14: https://doi.org/10.1007/978-3-540-74571-6_14

Weston, A. P. (2021, 12 16). *Understanding DICOM.* Retrieved from Towards Data Science: https://towardsdatascience.com/understanding-dicom-bce665e62b72

Yan, L., Yu, Y., & Ma, H. (2018). DICOM standard and its application in PACS Systems. Tianmen Medical College Affiliated Hospital, Hubei, China.

Appendix I : Logbook

Date	Hours	Product Backlog Number		
19.sept	4,00	Sprint Organization and Bachelor Initialization		
20.sept	3,00	US nr: 1		
21.sept	5,00	US nr: 1		
22.sept	2,00	US nr: 1	Nb of hours ini:	22,00
23.sept	2,00	US nr: 1	% total of the thesis:	6,2%
25.sept	6,00	US nr: 1		
26.sept	7,00	First meeting with the CHUV and sprint review		
27.sept	4,00	SPRINT 1 : US nr 1,2,3,4,5,11		
28.sept	8,00	US nr: 1		
29.sept	8,00	US nr: 2, 11		
02.oct	4,00	US nr: 2,3,11		
03.oct	7,00	US nr: 2,4,11		
04.oct	6,00	US nr: 2,4,5,11		
05.oct	8,00	US nr: 4,5,11	Nb of hours sp1:	67,00
06.oct	7,00	US nr: 4,11	% total of the thesis:	18,93%
09.oct	8,00	US nr: 1,2,3,4,5,11		
oct.23	10,00	Sprint meeting with the CHUV and sprint review		
11.oct	8,00	SPRINT 2 : US nr 6,7,11		
12.oct	8,00	US nr: 5,6,11		
13.oct	7,00	US nr: 6,11		
16.oct	3,00	US nr: 2,6,11		
17.oct	5,00	US nr: 11		
18.oct	4,00	US nr: 11		
19.oct	6,00	US nr: 6,7		
20.oct	8,00	US nr: 6,7		
21.oct	3,00	US nr: 6,7,11	Nb of hours sp2:	72,00
22.oct	2,00	US nr: 6,11	% total of the thesis:	20,34%
23.oct	8,00	US nr: 6,7,11		
24.oct	9,00	Sprint meeting with the CHUV and sprint review		
25.oct	9,00	SPRINT 3 : US nr 8,9,10,11		
26.oct	10,00	US nr: 8		
27.oct	8,00	US nr: 8		
30.oct	12,00	US nr: 8		
31.oct	8,00	US nr: 8		
01.nov	8,00	US nr: 8,9,10		
02.nov	7,00	US nr: 8,9,10	Nb of hours sp3:	90,00
03.nov	11,00	US nr: 8,9,10	% total of the thesis:	25,42%
06.nov	8,00	US nr: 8,9,10,11		
07.nov	13,00	Sprint meeting with the CHUV and sprint review		
08.nov	12,00	SPRINT 4 : US nr 8,9,10,11		
09.nov	7,00	US nr: 8,9,10		
10.nov	8,00	US nr: 8,9,10		
11.nov	5,00	US nr: 11		
12.nov	10,00	US nr: 11		
13.nov	12,00	US nr: 11		
14.nov	17,00	US nr: 11	Nb of hours sp4:	103,00
15.nov	14,00	US nr: 11	% total of the thesis:	29,10%
16.nov	5,00	US nr: 11		
total:	354,00			

Appendix II : Product Backlog

US Nr.	Thème	As an/a	I want to ...	So That	Acceptance Criteria	Priority	Status	Story Points	Sprint	MoSCoW
1	State of the Art	Student	Setup the work environment in order to be ready to begin the Bachelor thesis	I can begin to work without worring about the administration part	Beeing ready to write, to share and to start a sprint	1000		1	1	Must Have
2	State of the Art	Student	Learn about existing standards in brain imaging, e.g. DICOM, BIDS, etc	I understand this standard	I can explain the existing standards in brain imaging	950		2	1	Must Have
3	State of the Art	Student	Learn about existing metacata guidelines for brain imaging	I understand the guidelines	I can explain the existing metadata guidelines for brain imaging	900		3	1	Must Have
4	State of the Art	Student	Learn about the existing tools for brain imaging metadata management	I understand these tools	I have a good understanding of the existing tools for Brain imaging metadata management	850		3	1	Must Have
5	Decision	Student	Choosing the appropriate metadata standards and guidelines	I can implement it further	I can explain why i choose the metadata and i am ready to implement it	800		2	1	Must Have
6	Justification	Student	Elaborating a metodology for metadata and eeg organization	I have a better understanding on the strategy and the metodology	I know how i will manage the implementation	750		4	2	Must Have
7	Decision	Student	Choosing the technologies for the implementation of the metadata system and search tool	I am ready to implement them and start the development part	I can explain why i choose these technologies and i know how i will use them	700		3	2	Must Have
8	Implementation	Student	Proof of concept with the metadata management and the search tool	I can search among a catalog of images and metadata	Develop a program that stores the given metadata. You can search, find and display the metadata	650		12	3	Must Have
9	Testing	Student	Test the proof of concept with the SENSE datasets	I can evaluate if there are some issues with the tools	All tests passed	600		3	3/4	Must Have
10	Testing	Student	Evaluate and correct the potential errors of the testing theme	i can avoid the potential crash/bugs	All tests passed	550		3	3/4	Must Have
11	Documentation	Student	Write the bachelor thesis	I can share my results with the SENSE team and the people responsible to manage this bachelor thesis	Complete word document + annexes	500		12	1/2/3/4	Must Have
12	Implementation	Student	Have access to the CHUV Data	I can have a better understanding of how to implement the metadata search tool	Having access to the CHUV data	675		3	4	Should have

Appendix III : User Guide

User guide

BIDS Application

Théo Choffat

User Guide

Made by: Théo Choffat, HES-SO student.

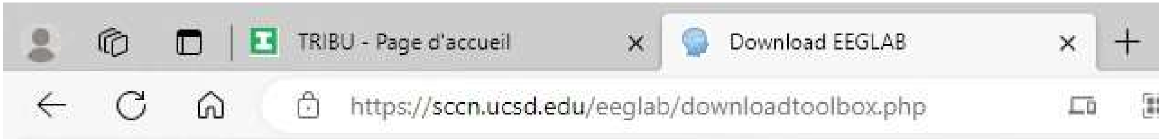
Context

This application was made through a bachelor thesis project. Thanks to the help of a student from the HES-SO Valais in business informatics, this application allows users to organize and search through electroencephalography (EEG) datasets.

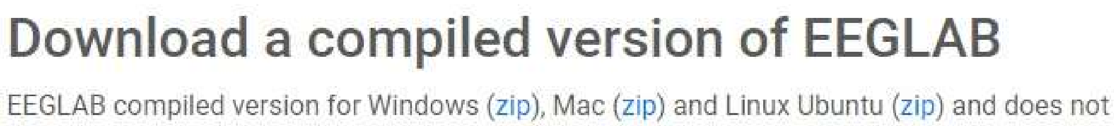
EEGLAB installation

To install the application that will be responsible for transforming the raw EEG-data into a BIDS (Brain Imaging Data Structure):

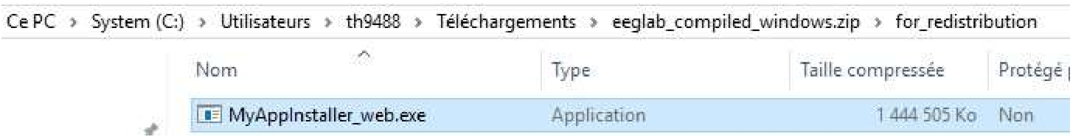
First, go to their official website under the download tab: https://sccn.ucsd.edu/eeglab/downloadtoolbox.php

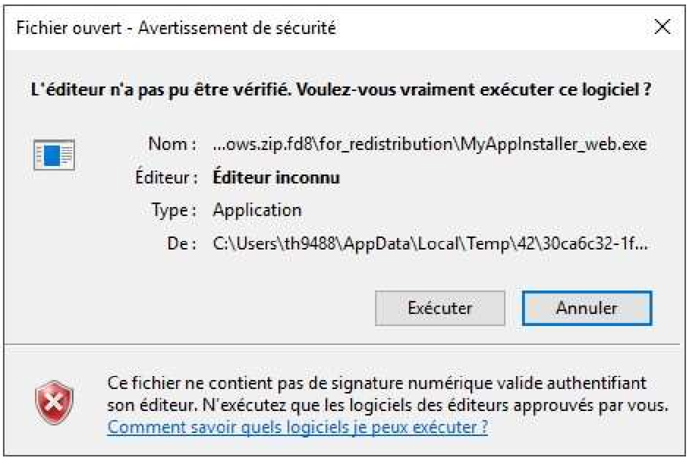


Then, click on the zip file to download the installation file.

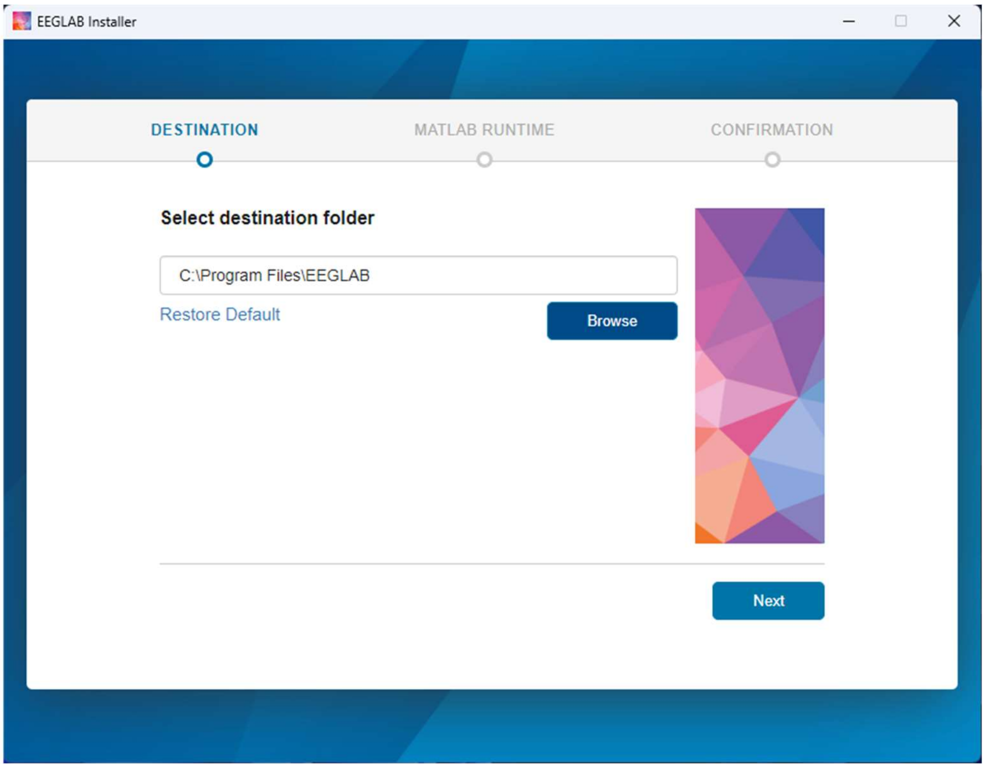


Unzip the file, go inside the “for_redistribution” folder and double click the “MyAppInstaller.exe”

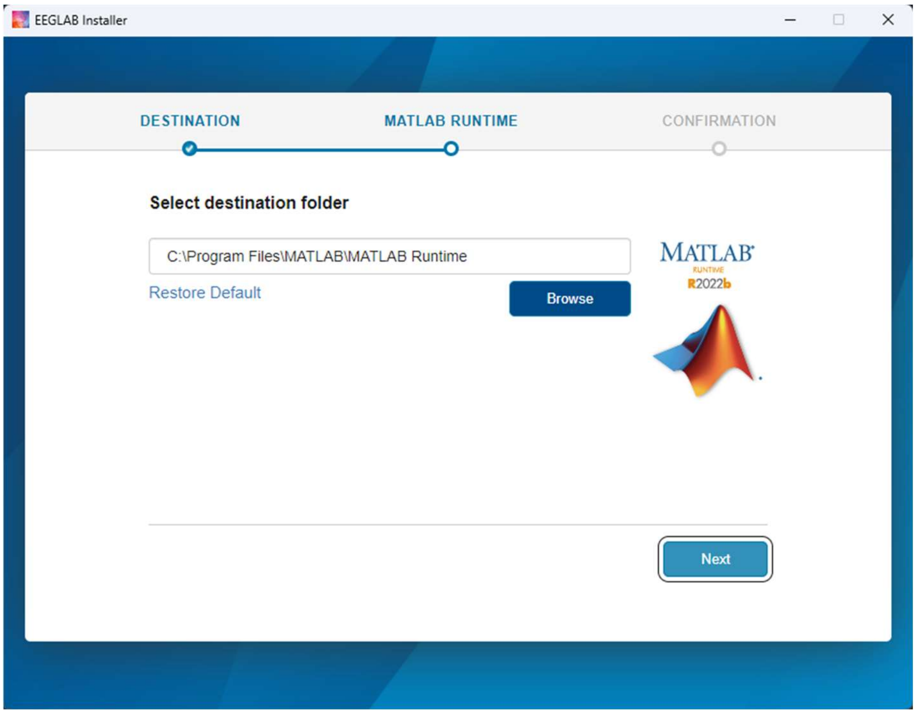




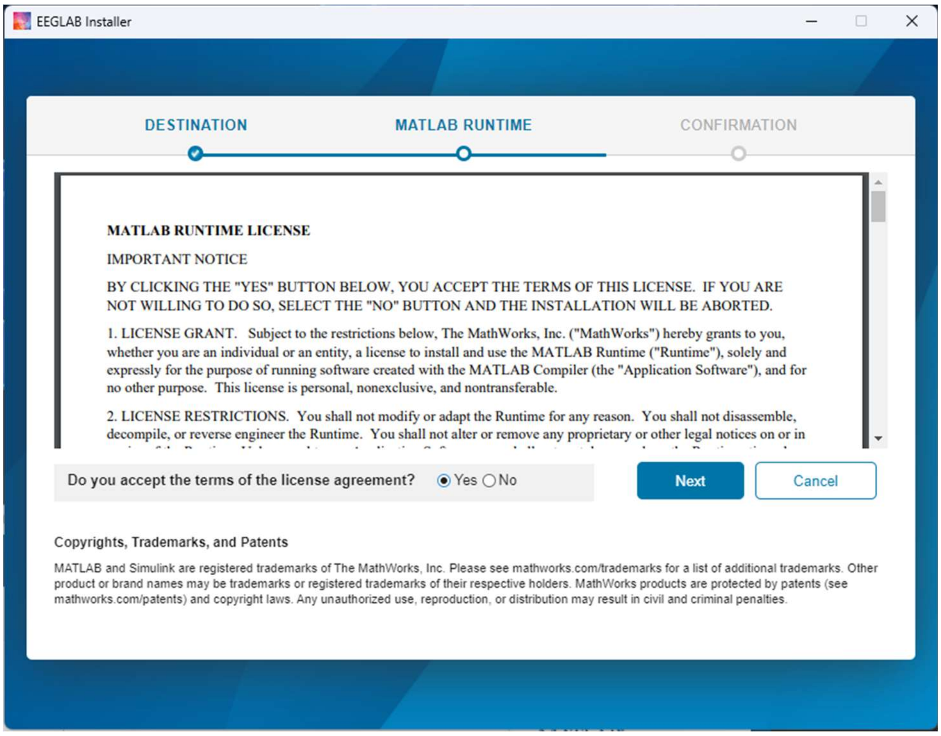
Click on “Exécuter”, the installation wizard should open, select the destination and click “Next”:



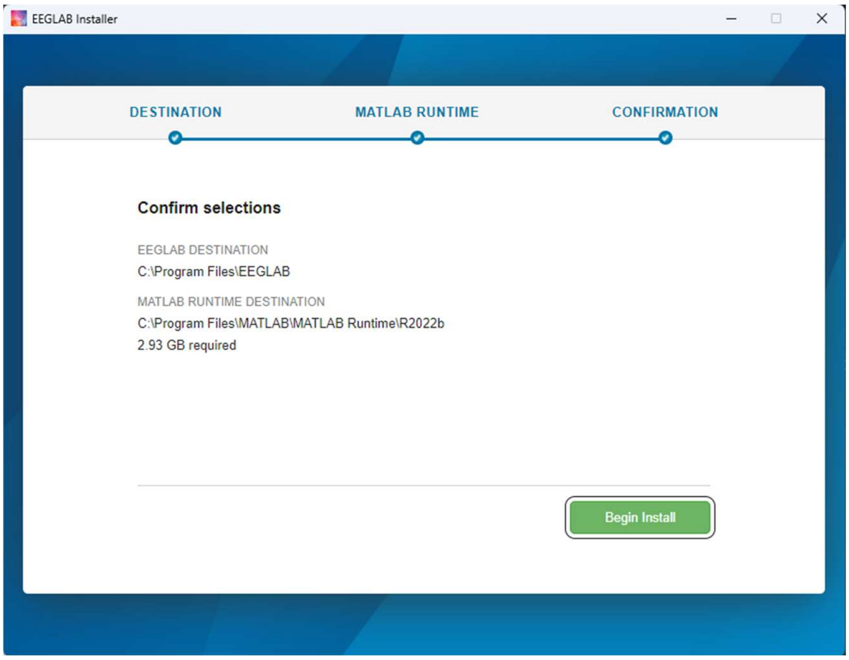
Then, we will be asked to install the MatLab runtime that is responsible to run the application:



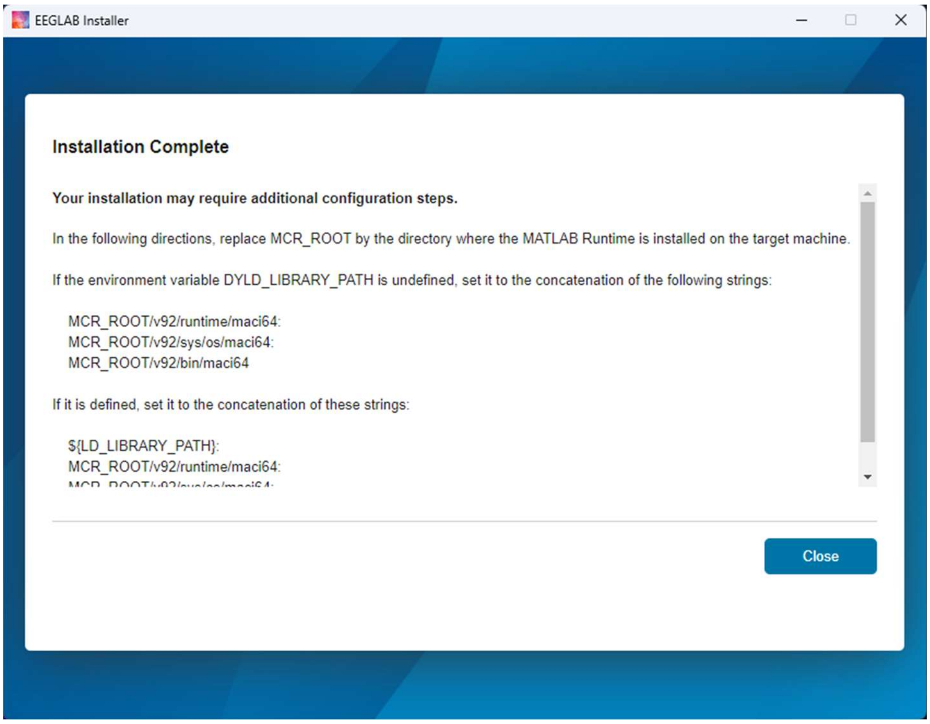
On the license agreement, click "yes" then "Next"



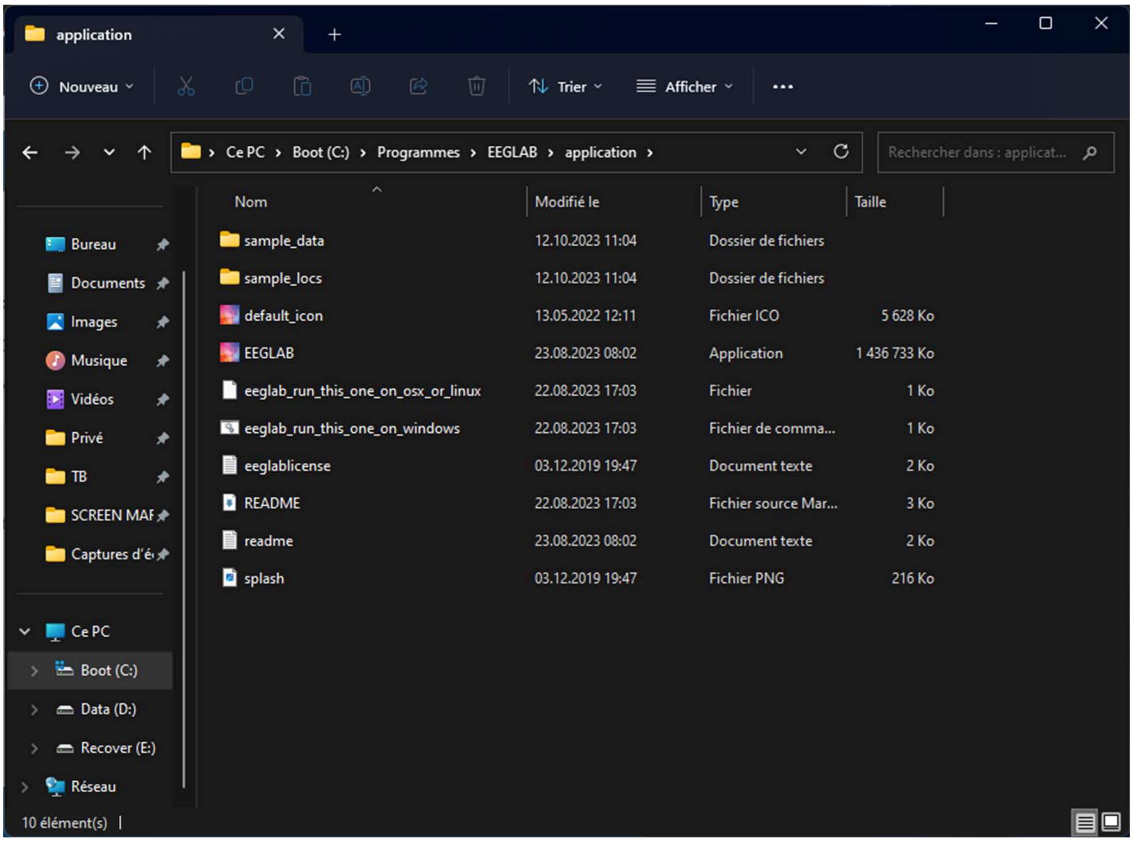
We can now start the installation by clicking on the "Begin Install" button:



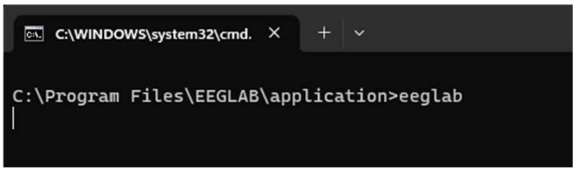
We can click on the "Close" button once the installation is complete.



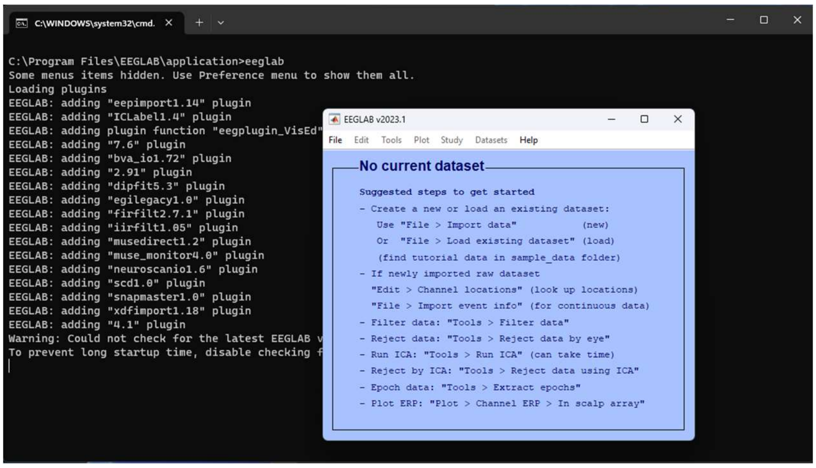
Once the installation is complete, we can go where the application is installed and click on the command file that is called "eeglab_run_this_one_on_windows" to start the application.



A command prompt should open with this command:



After a few seconds, the application should start normally. This is how it looks like:

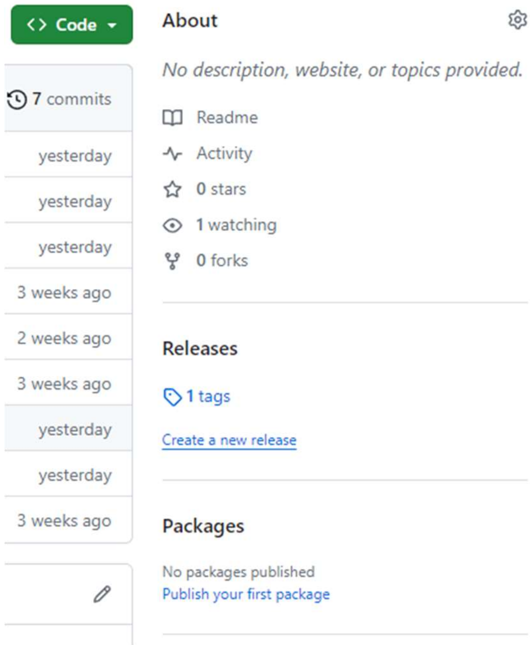


EEG-BIDS Research Tool Installation

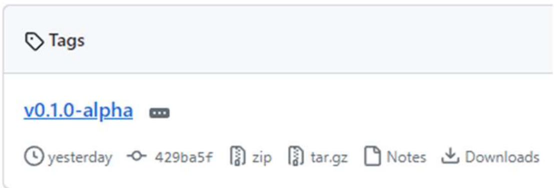
This tool was developed by Theo Choffat for CHUV researchers. The tool allows researchers to find the BIDS folder they need with the help of a search engine that displays the datasets available.

To install the application on the desktop, access this GitHub Page:
https://github.com/TheoChoffat/bids_apps

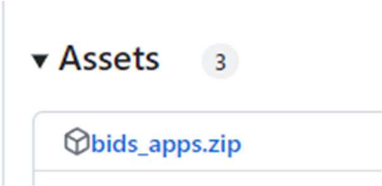
Once we are on the page, click on the "Release tags" on the right side of the page:



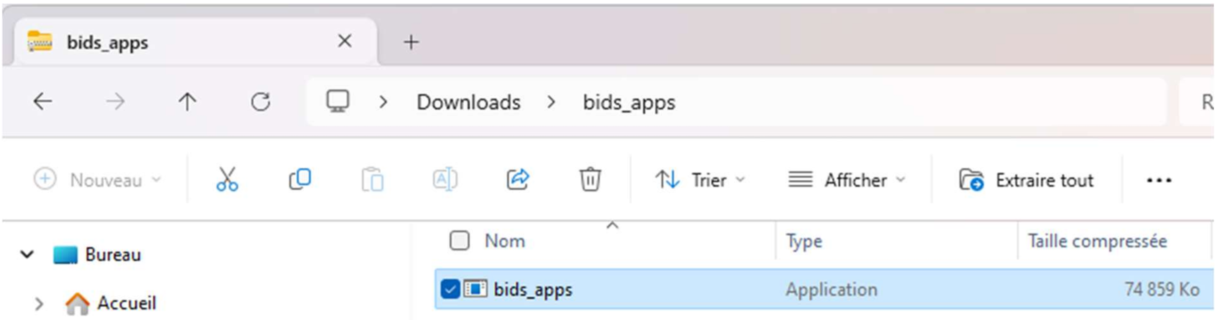
Then, select the latest release available (here is the v0.1.0-alpha):



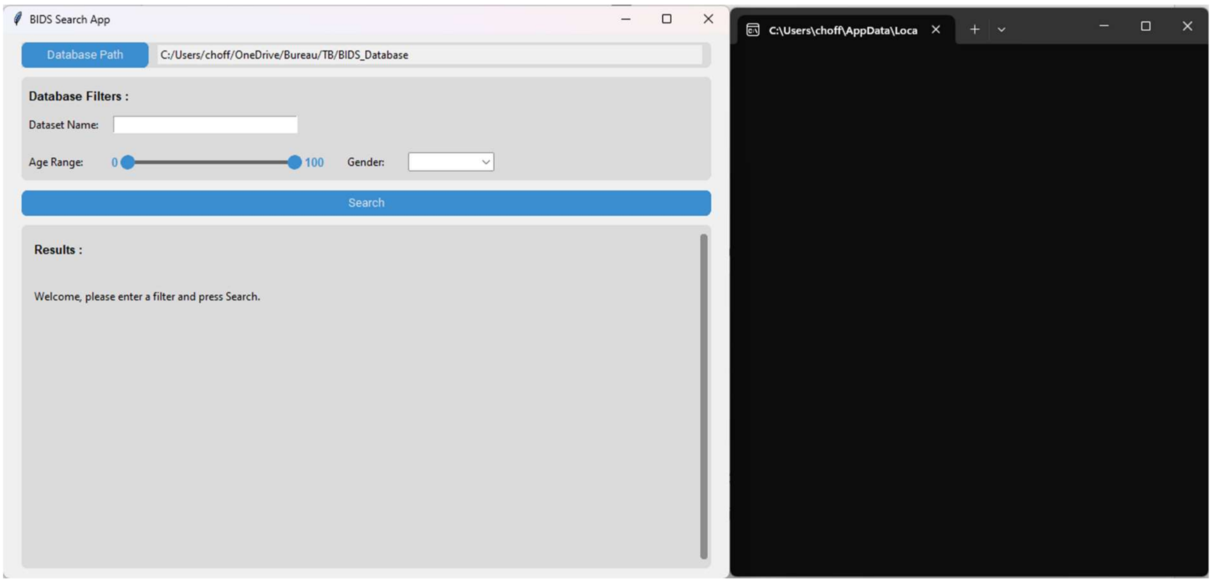
Under Assets, click and download the “bids_apps.zip” file:



We can then uncompress the .zip file and double click on the “bids_apps” file.



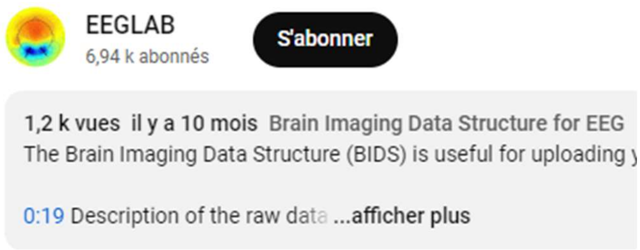
Since we are not a recognized developer by Windows, this popup will show, click on “plus d’information” and then “Exécuter quand même”. The application should start with a command prompt on the back. Here is what it should look like:



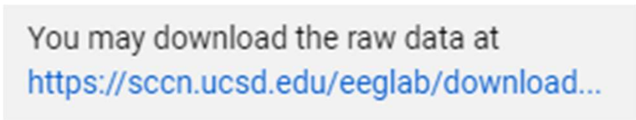
EEGLAB Basic Testing

To test the application, we will need some raw data to work with. An EEGLAB researcher did a tutorial on YouTube with a raw dataset without sensitive information that can be installed by clicking on this link: https://www.youtube.com/watch?v=LIiRj8-BVk4&t=296s

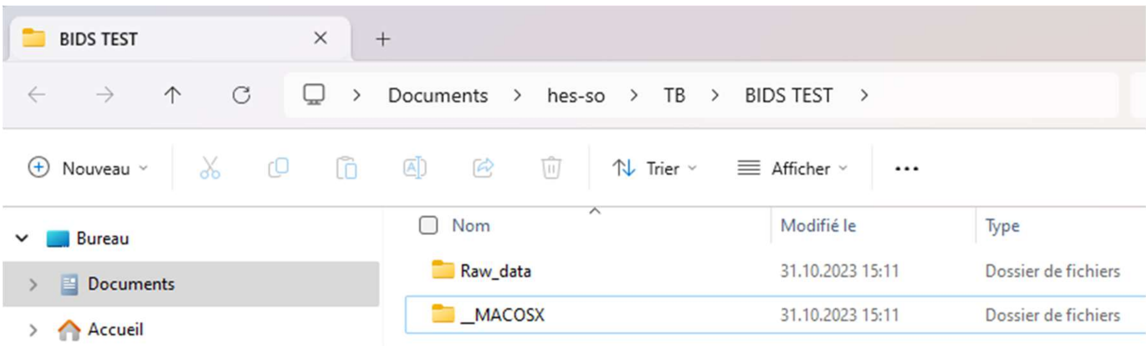
Then, click under the video on the "…afficher plus" button:



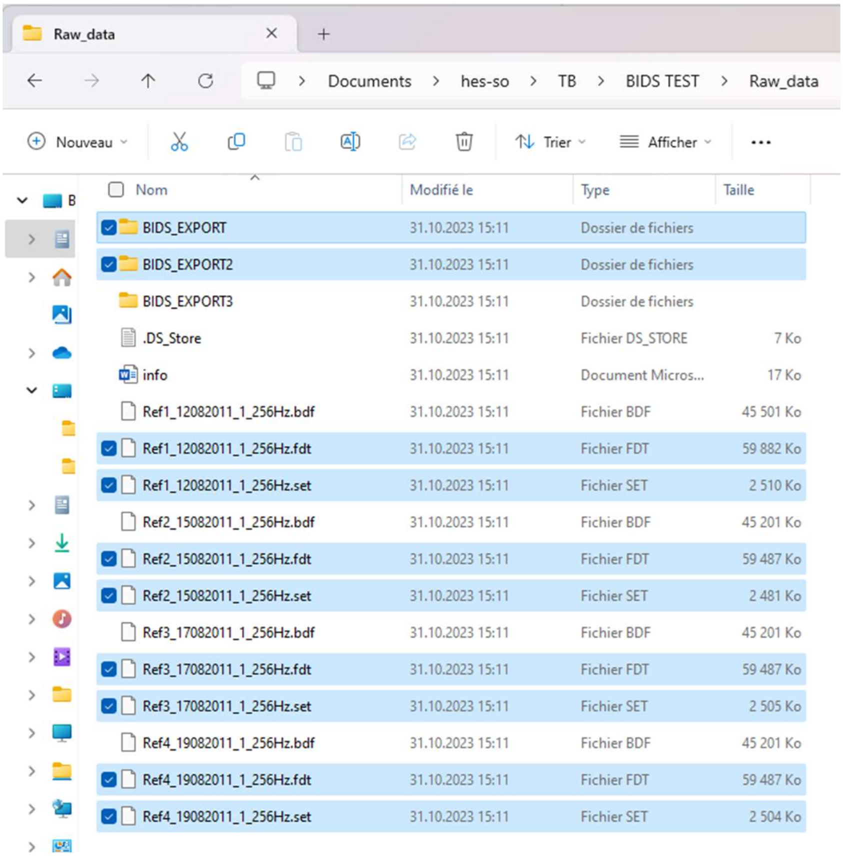
Under the description, click on the dataset example:



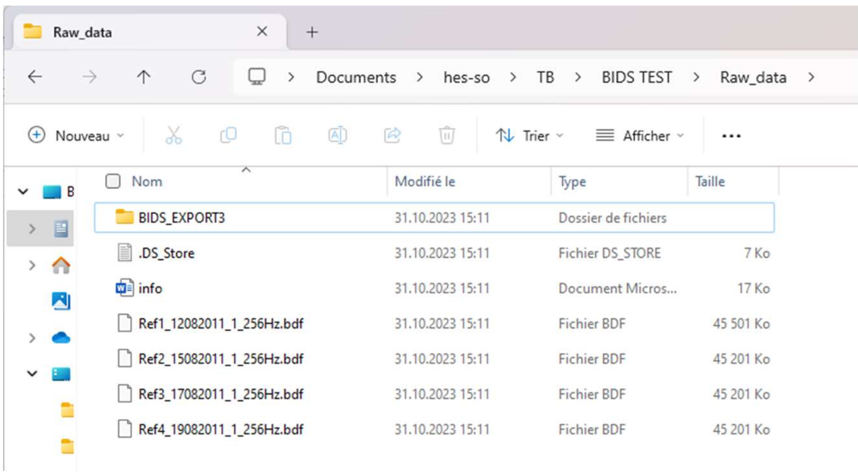
A zip file will be downloaded. Unzip it in a folder **where we have the write+read access:**



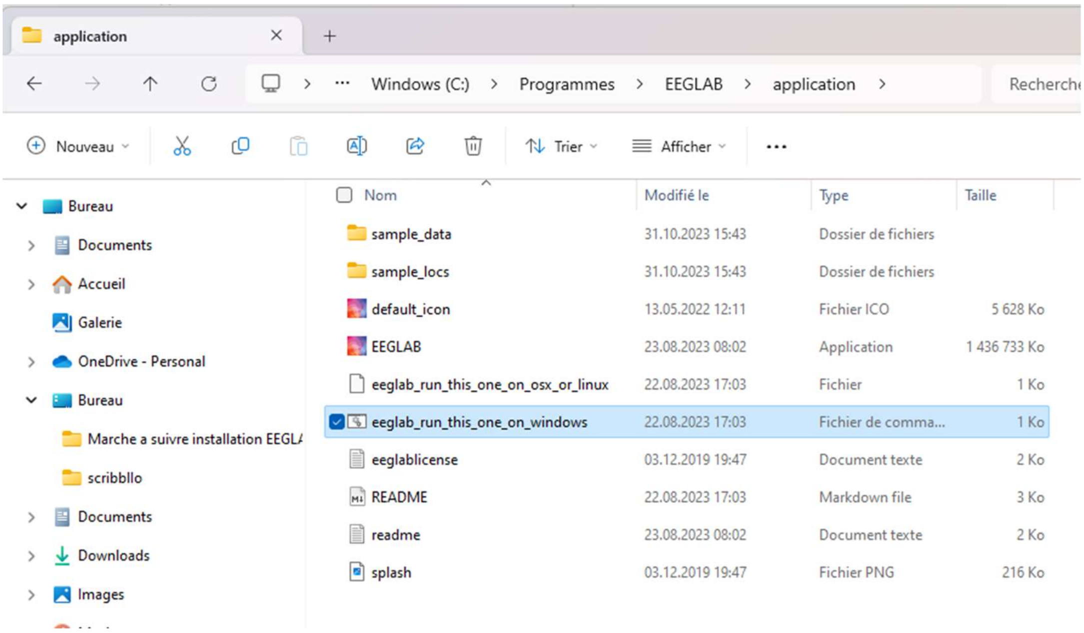
We can delete the "_MACOSX" version if we are not using it. Then, click on Raw_data. The folder should look like this:



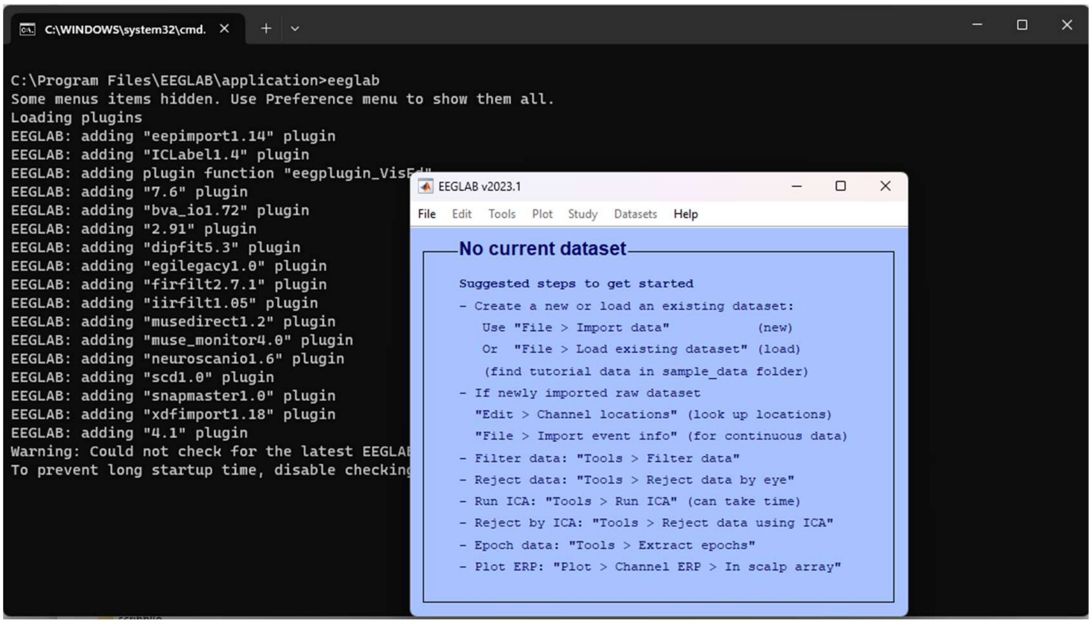
As this was used for a YouTube tutorial, we do not need all this information to test the EEGLAB application. We can therefore delete the ".set" and the ".fdt" files as well as the BIDS_EXPORT and the BIDS_EXPORT2 folders. Keep the ".bdf" files (these are a type of raw EEG data that were generated by the electroencephalography computer) and the BIDS_EXPORT3 folder, we will need them afterwards. The "Raw_data" folder should look like this:



Then, start the EEGLAB application by clicking on the “run_this_one_on_windows”:



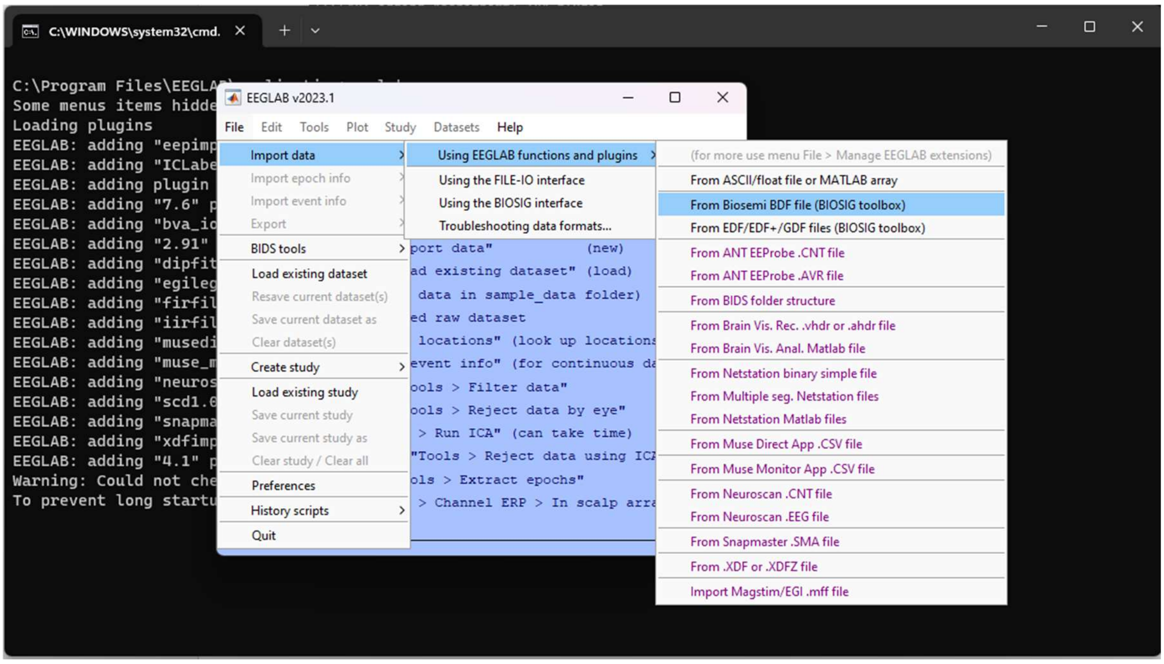
Wait until the application starts, this can take around one minute if it is the first time. Here is what we should have:



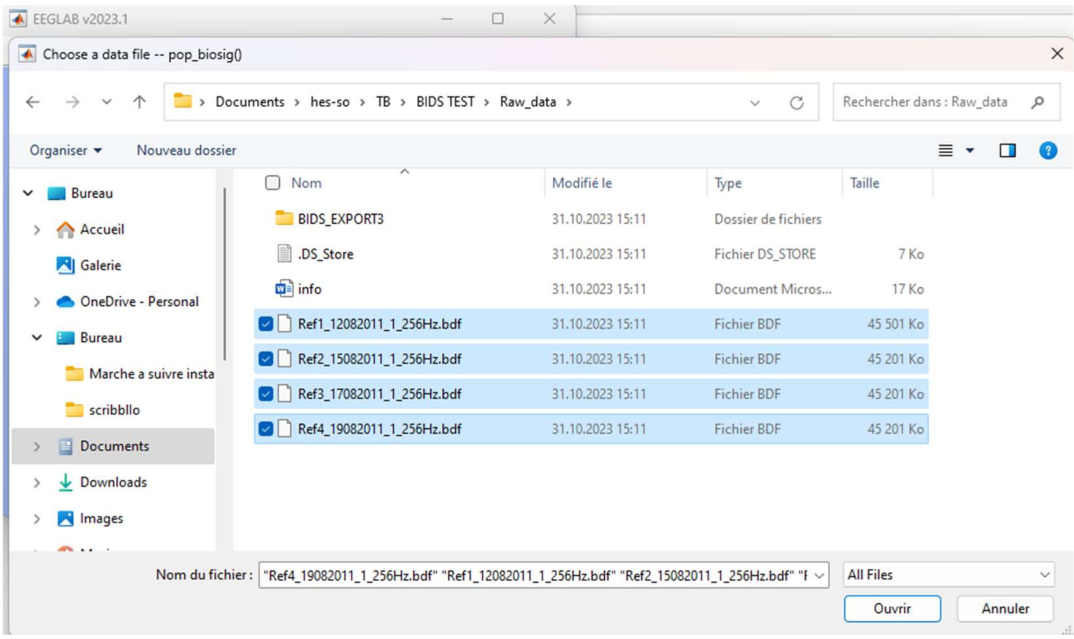
To test the application, simply continue this guide.

Raw data import

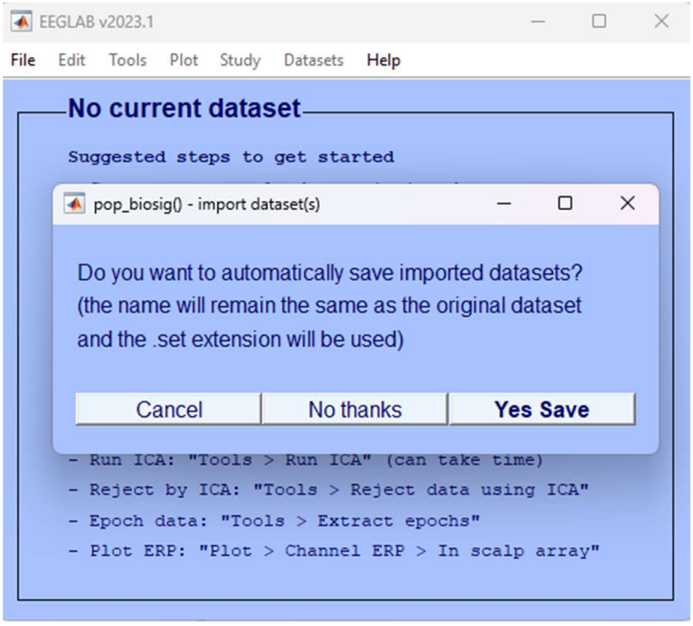
Now, we will try to transform the raw data (the ".bdf" files) into a BIDS structure. To do so, we need to import the different bdf files into this application. To import the files, click on File > Using EEGLAB functions and plugins > From Biosemi BDF file (BIOSIG toolbox) like so:



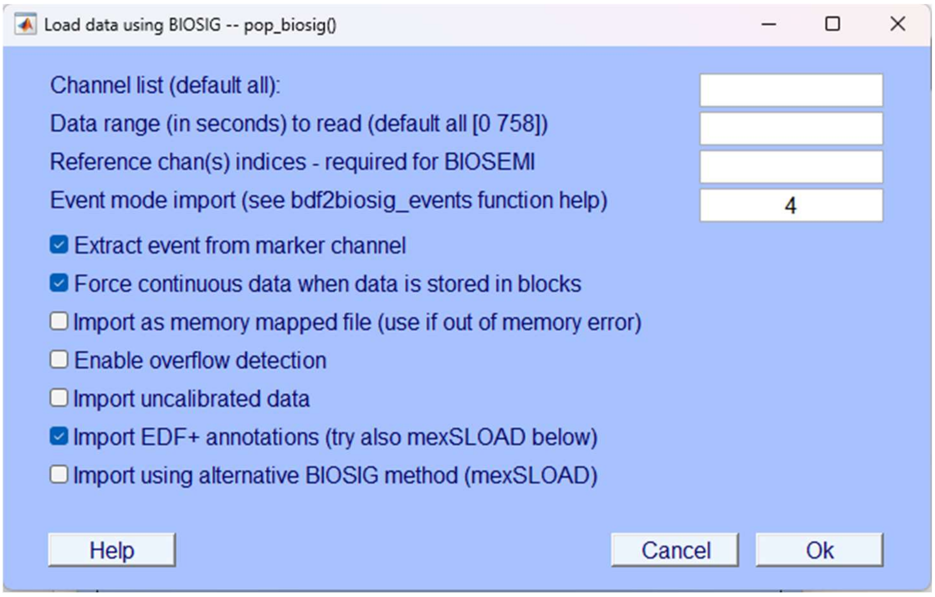
Then, select the 4 different".bdf" files that are in the "Raw_data" folder and click "Ouvrir":



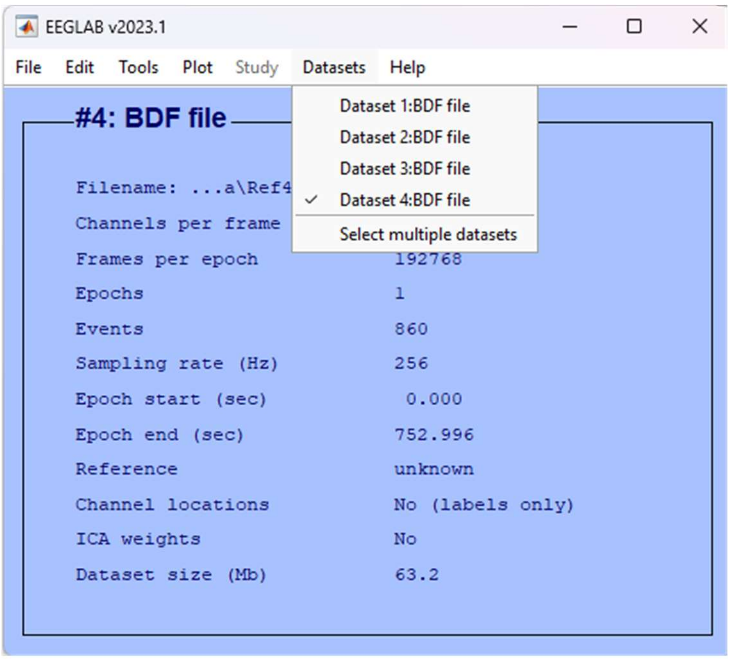
Then, a popup shows up asking us if we want to save the dataset, click “Yes Save”;



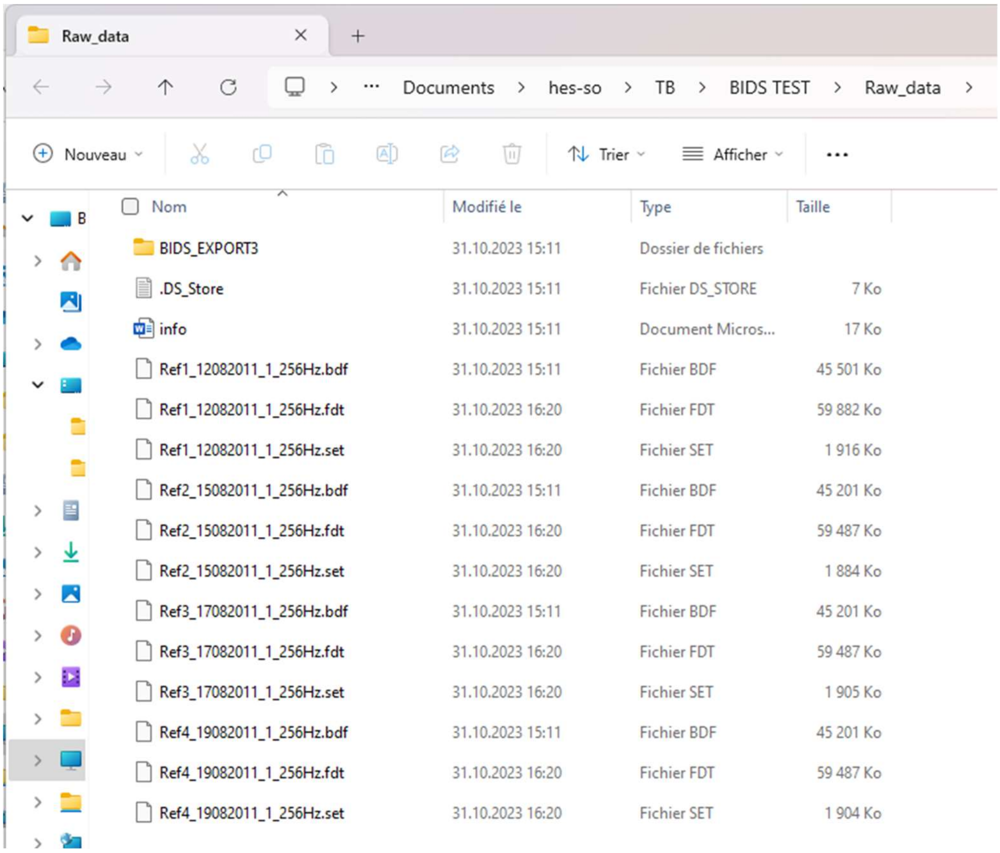
The researchers can then add some import information, for this example, just click the “Ok” button.



The four files should be imported inside the application. Check it by clicking on the Dataset tab:

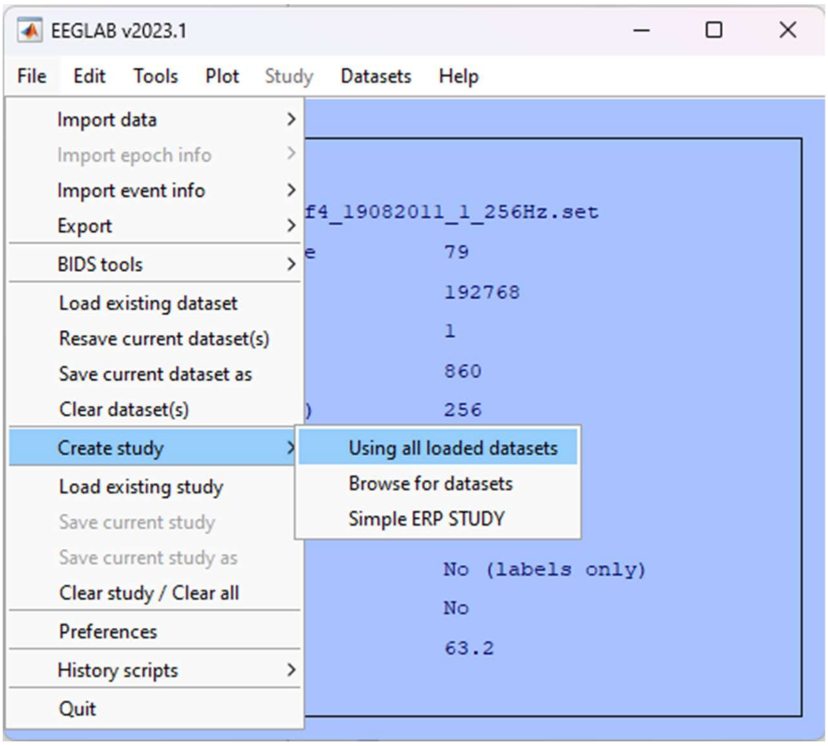


Normally, this just separated the .bdf files into 3 parts. we can check our “Raw_data” folder:

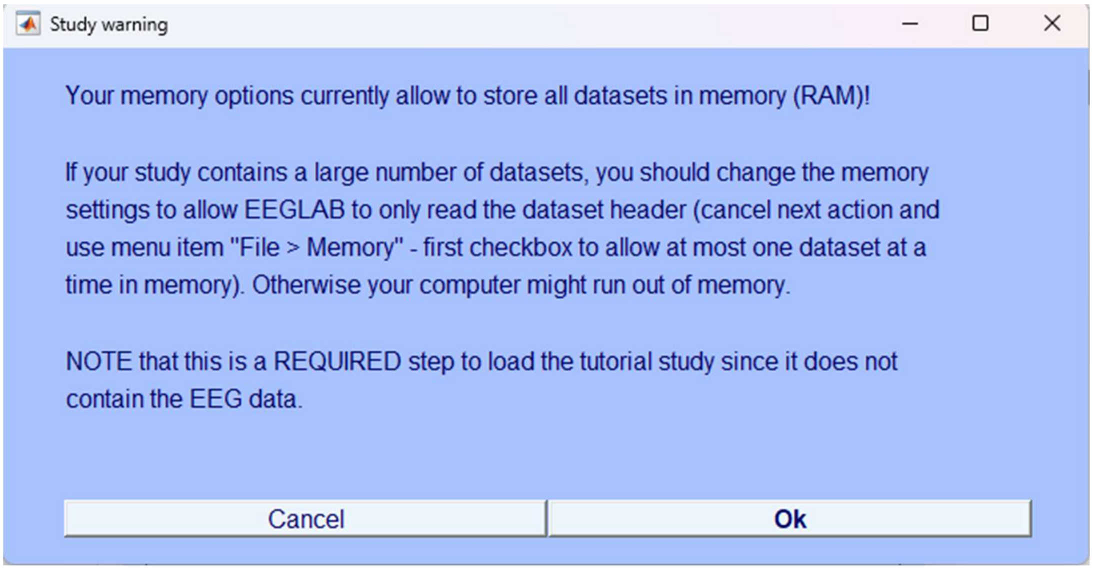


Dataset to Study transformation

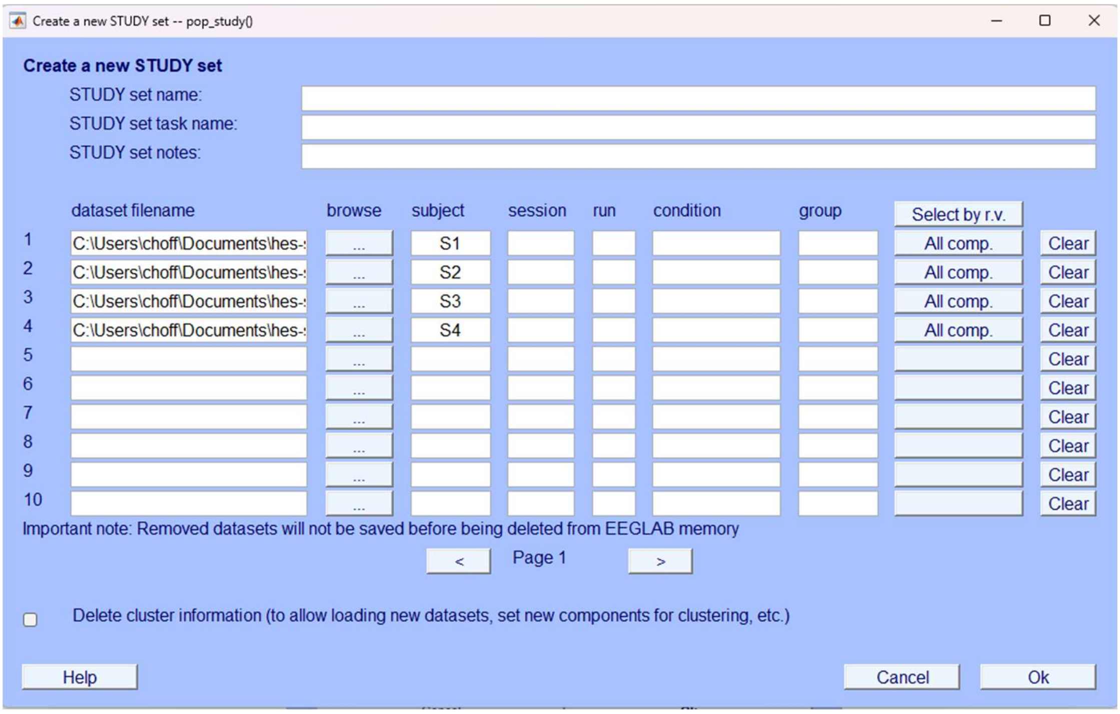
To transform this raw data into BIDS, we first need to create a study. Click Files > Create study > Using all loaded datasets:



There is then a warning about large datasets. If we have it, click on “Yes” as the dataset that we will use does not have a problem.



Then, a new window opens and shows how we can create a study. For this example, enter under the “subject” column the name: “S1”, “S2”, “S3” and “S4” and click on “Ok”:

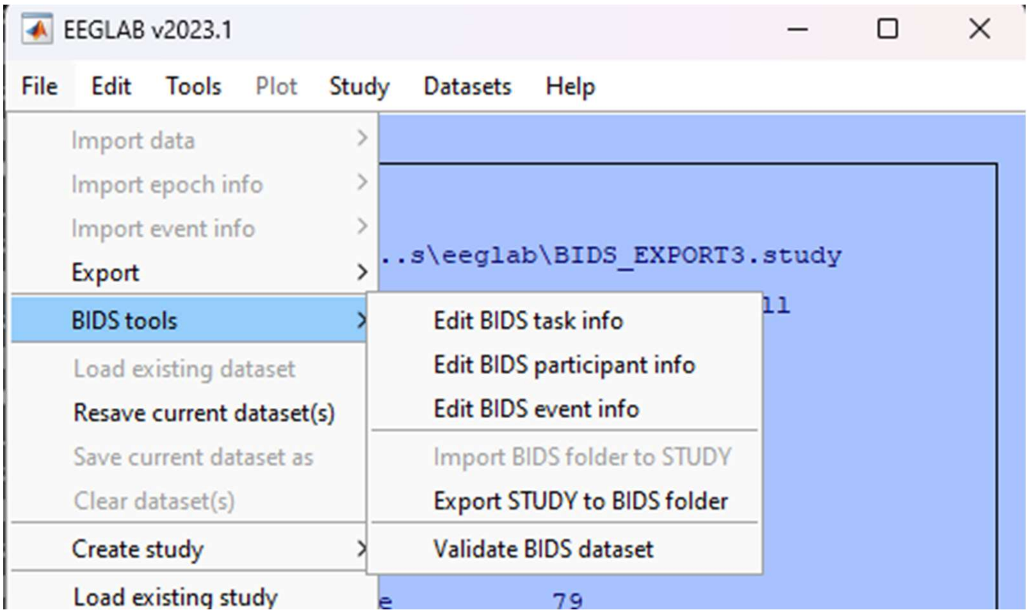


Metadata addition

We can add many metadata information in EEGLAB. Here is how we can do:

First, select the study we want to modify by importing it into the application. It can be done by clicking on the File > Load Existing Study if the study was previously created with the EEGLAB tool. If it is a regular BIDS file, import it by clicking on File > Import data > From BIDS folder.

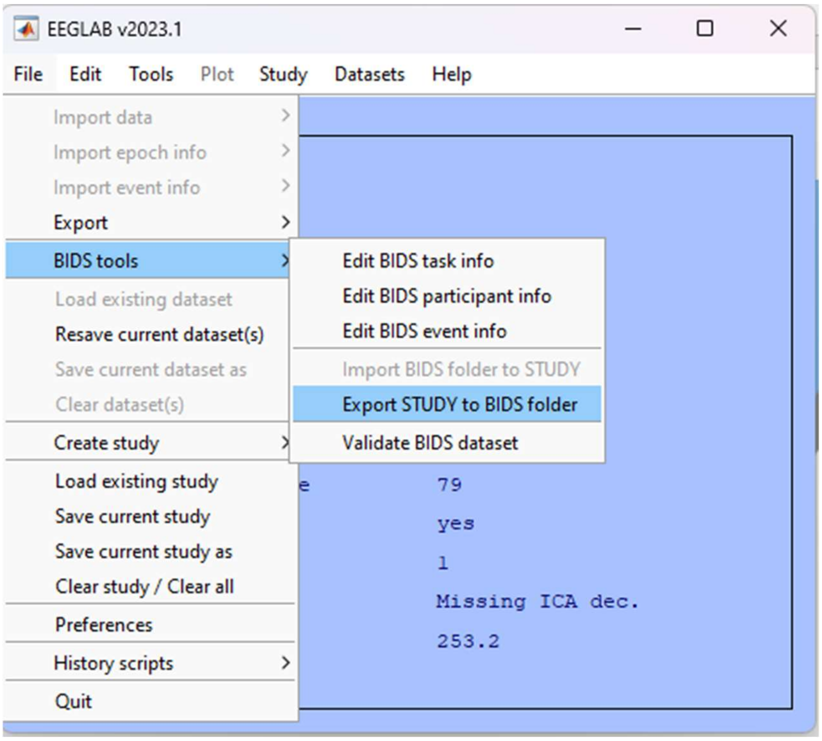
There are three different ways to add metadata that are under the same tab. After this, the researcher can either add information to the task, the participant, or the event.



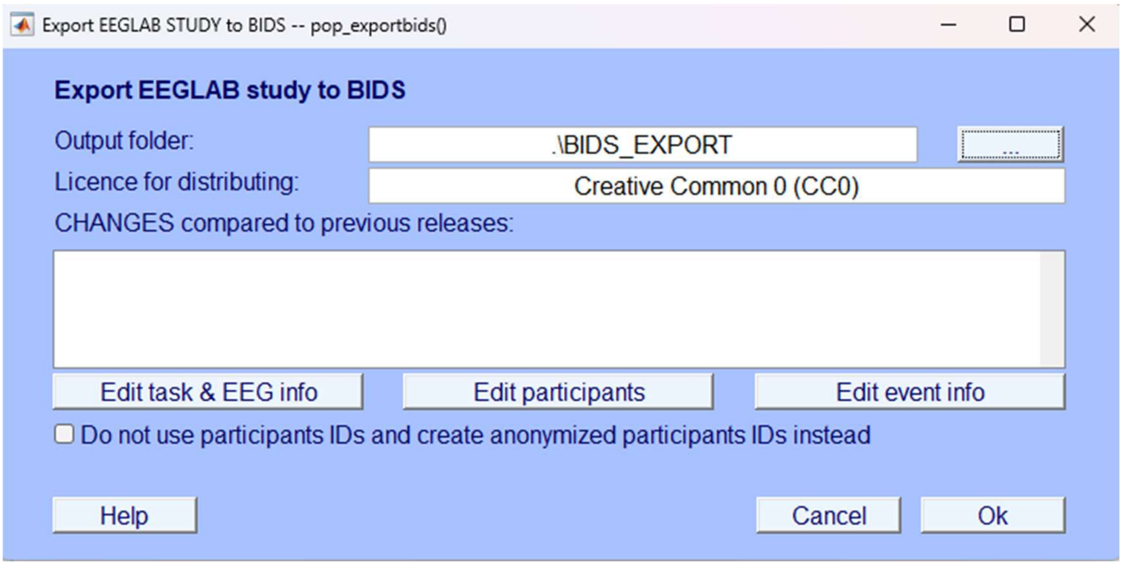
These three options will modify the Dataset. Therefore, it is mandatory that the researcher has written access to the dataset.

Export a STUDY to the BIDS format.

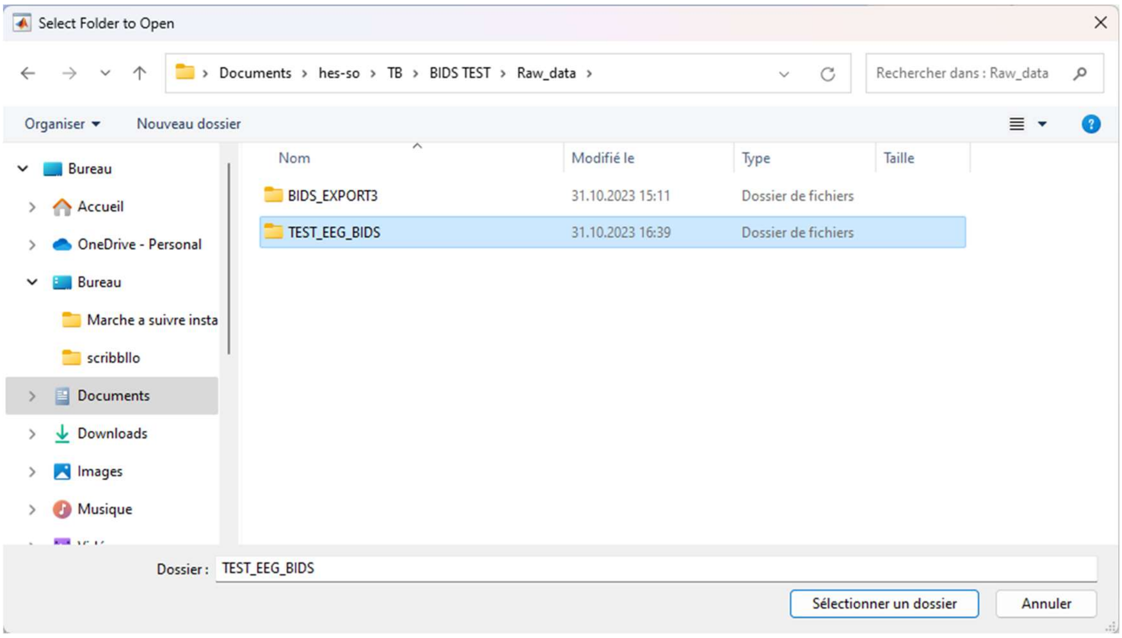
Now that the study is created, click on File > BIDS tools > Export STUDY to BIDS folder.



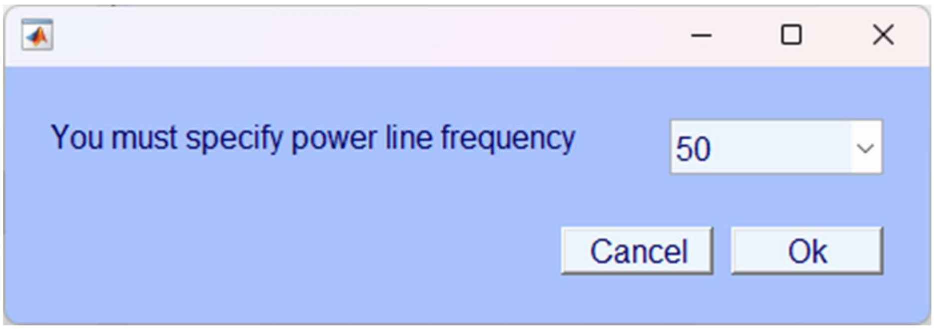
By default, the BIDS will be created under the program files where EEG is located, change this by clicking on the top right button:

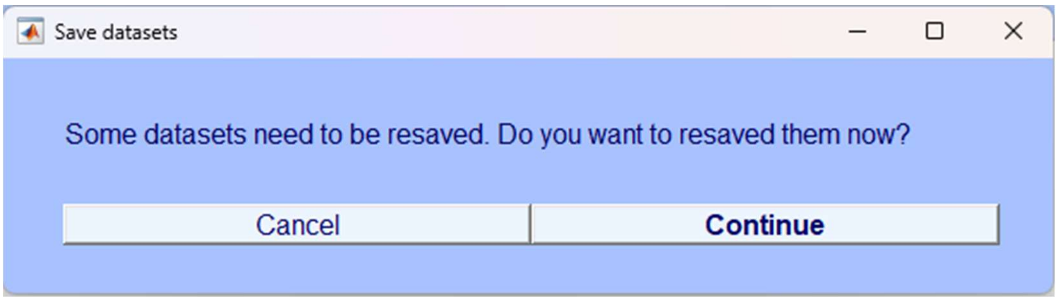


Then, go where we have our "Raw_data" folder and create a new folder (like the TEST one under) and click on "Sélectionner un dossier" then click "Ok":

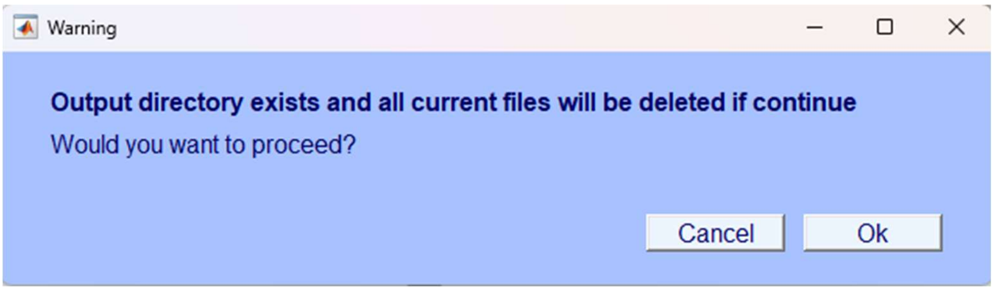


Then, we will put the line power, we can keep 50 and press "Ok" and "Continue".

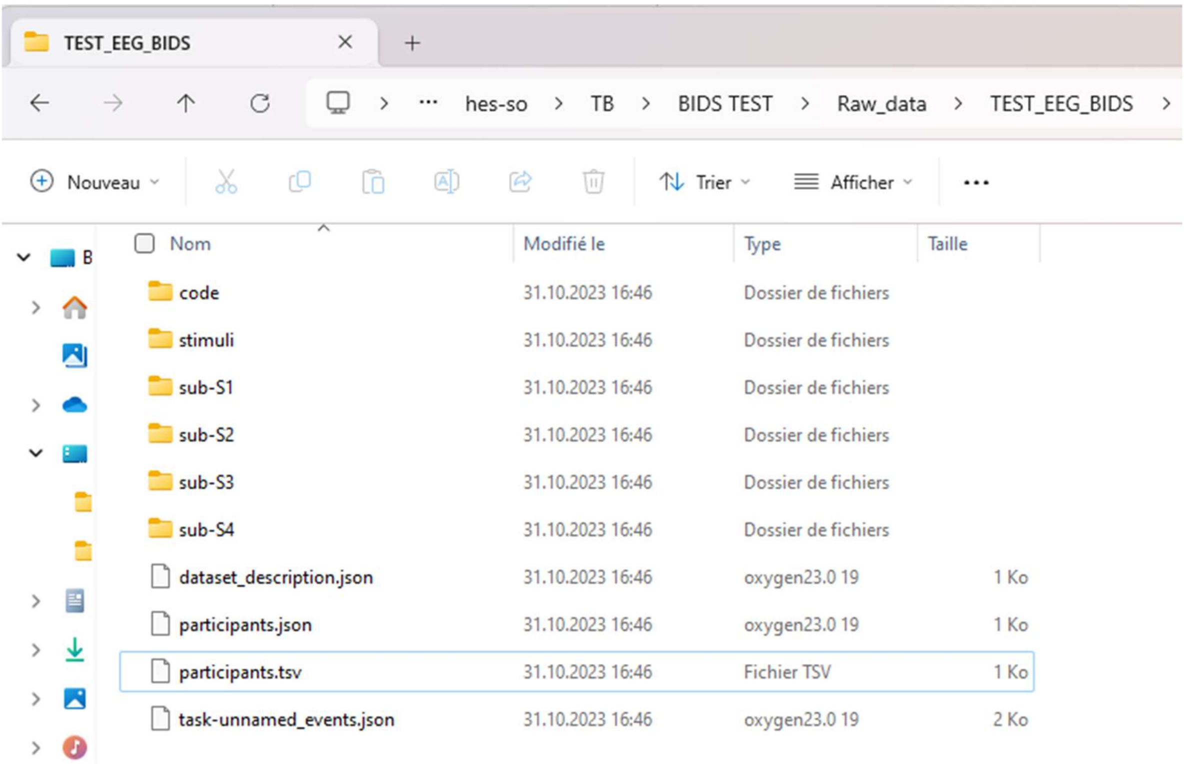




We may have this warning concerning the fact that we are saving data somewhere else. This is because the folder is different than the one where we stored the raw data (No worries the raw data will be copied into the new folder. Click "Ok".



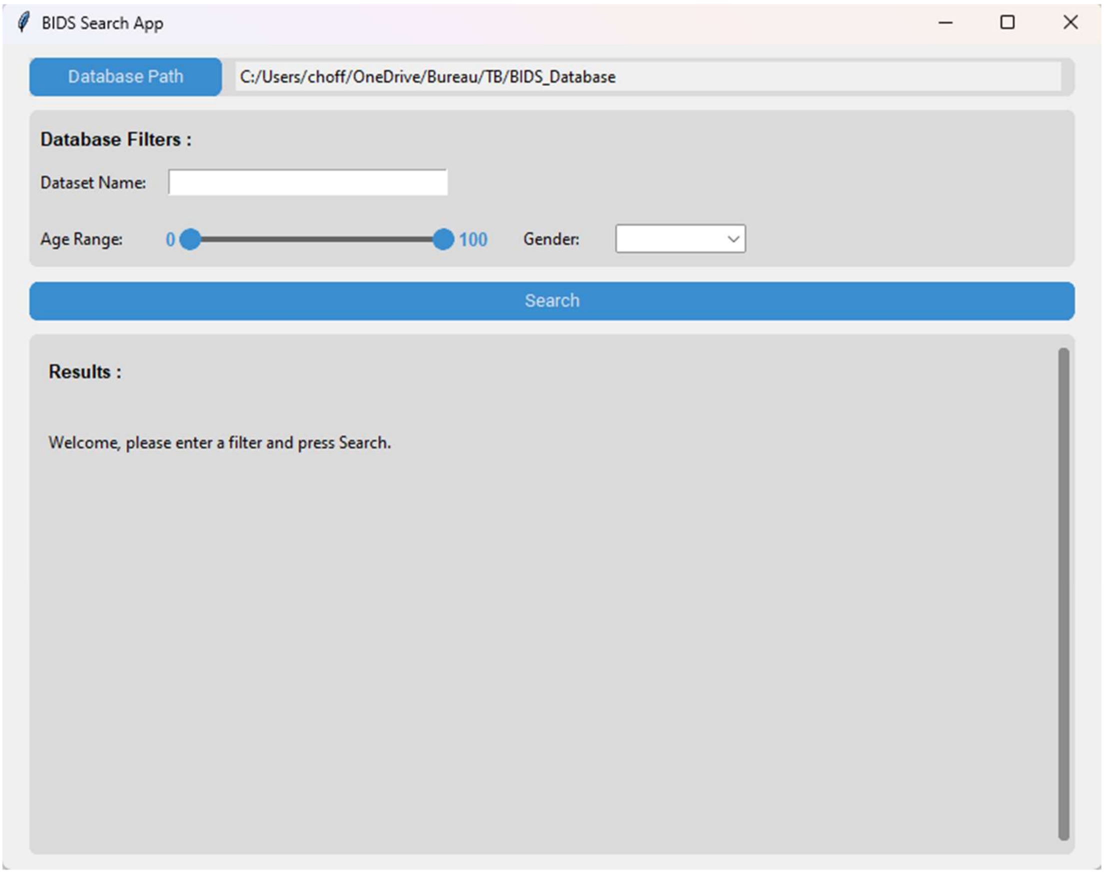
Now, normally, if the application works, we will have a BIDS dataset in the TEST folder that should have the following structure:



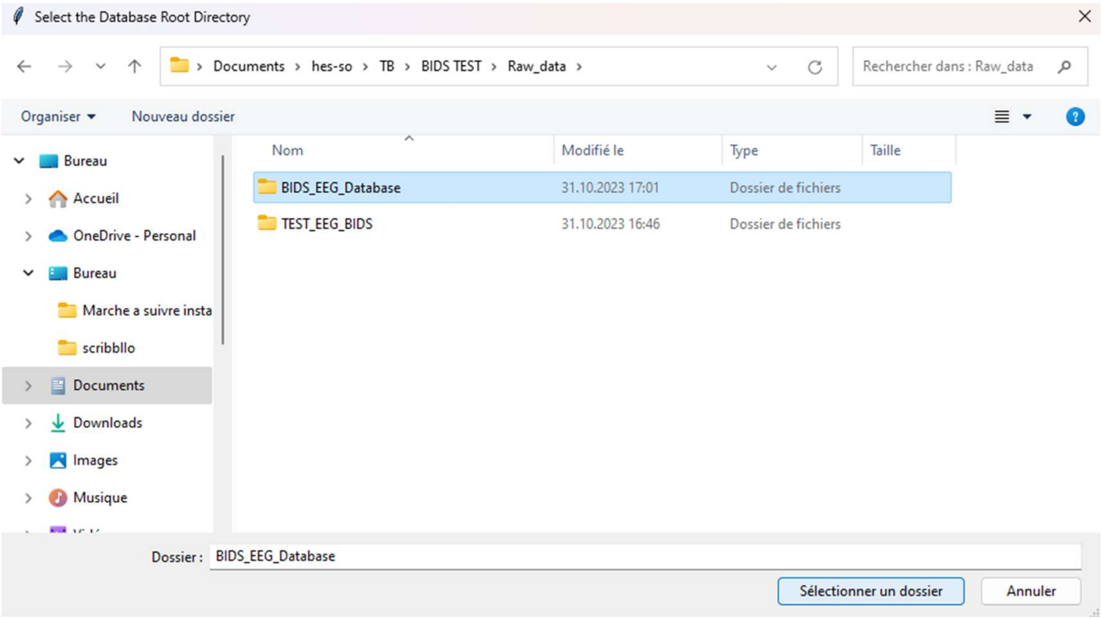
If we have these, the application basic test is complete, and the application can create a BIDS structure that will be helpful for the researchers.

BIDS-EEG Research Tool

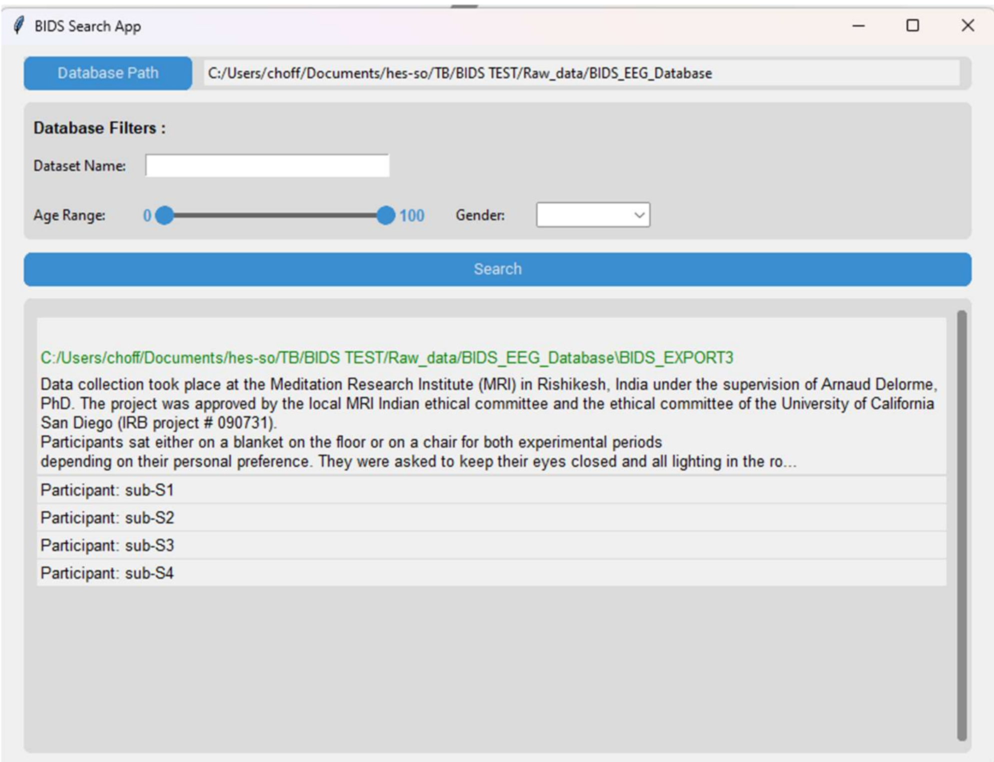
Launch the “bids_apps” application:



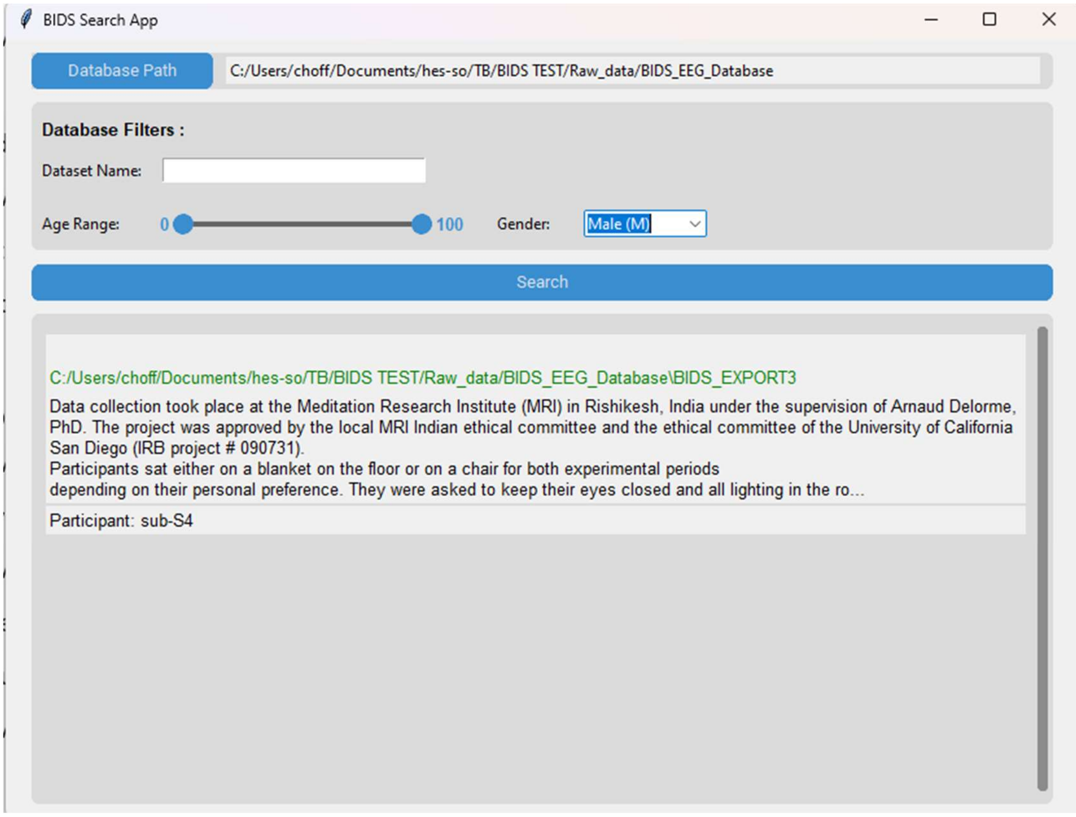
As the link to the database won't be the same, click on the "Database Path" on the top left of the screen and select the "BIDS_EEG_Database" we just created:



Now, press the blue "Search" in the center of the screen. We should have the same result as below:



As this is a simple BIDS example out of a YouTube tutorial, the Dataset does not have a name and their participants don't have their age. But we can try the "Gender" filter by applying the "Male" option, we should only have the Participant S4 left like this:



If we do have this result, the application is working properly and the researchers and the basic test for this BIDS-EEG Research tool is now complete.

If we have any questions or problems on the software installation, contact me with the following address:

Name: Théo Choffat

Telephone: +4179 178 70 55

Mail address: theo.choffat@students.hevs.ch (will be deleted at the end of 2023).

Declaration of author

I declare, by means of this document, that I have completed the attached Bachelor's thesis on my own, without any assistance other than that duly indicated in the references, and that I have used only the sources expressly mentioned. I will not provide any copies of this report to a third party without the joint authorization of the program coordinator and the professor in charge of supervising the bachelor’s thesis.

SIERRE, LE 20 NOVEMBRE 2023
Théo Choffat