MJSAT

Malaysian Journal of Science and Advanced Technology

journal homepage: https://mjsat.com.my/

Neuroscience Data Formats, Models, Repositories and Analytics: A Review

Sze Wei Fong[1], Nurfaten Hamzah[1,2], Nurul Hashimah Ahamed Hassain Malim[2,4*], Jafri Malin Abdullah[1,2,3]

[1]*Department of Neurosciences, School of Medical Sciences, Universiti Sains Malaysia Health Campus, 16150 Kubang Kerian, Kelantan*

[2]*Brain and Behaviour Cluster, School of Medical Sciences, Health Campus, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia*

[3]*Department of Neurosciences & Brain Behaviour Cluster, Hospital Universiti Sains Malaysia, Health Campus, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia*

[4]*School of Computer Sciences, Universiti Sains Malaysia, 11800 Gelugor, Pulau Pinang, Malaysia*.

KEYWORDS

Neuroscience
Data model
Data format
Data repository
Data analytic

ARTICLE HISTORY

Received 4 April 2023
Received in revised form 14 June 2023
Accepted 15 June 2023
Available online 4 July 2023

ABSTRACT

As neurotechnologies have gotten better, a lot of neuroscientific research has been done using these new technologies. Even though labs all over the world produce a lot of neuro-data, most of this data has not been shared to help people from different fields understand neuroscience. The neuro-data sharing is essential because it accelerates the pace of discovery in neuroscience. Effective data sharing will depend on the standardized use of file or data formats, highly reusable data analytics tools, and data storage formats. In this review paper, we review the four domains (data format, data model, data repository, and data analytics) that are currently in use in the neuroscience community. In the end, we are discussing several challenges associated with data sharing.

1. INTRODUCTION

With the passing of time, it's impossible to deny that technology has changed every part of human life. Every industry's early leaders have been focused on using these new technologies to boost their own fields. In the field of neuroscience, which works well with information technology, this is by no means an exception. Many powerful and complex neuromodalities have been developed or are in the process of being developed to aid in the understanding of living creatures. Neuroscience is known to be a complex research field as it does not focus solely on one single perspective but involves an integrated analysis of information from multiple disciplines such as structural, functional, behavioral, cognitive, genomic, proteomic, and/or other related sub-disciplines. As such, data sharing has increasingly become an important concern for the neuroscience community.

Data sharing initiatives have a positive impact on the acceleration of scientific discovery [1 - 3]. It hastens scientific progress by assembling massive data sets from several sources. Sharing data also enables validation and verification of scientific findings, which supports open science and strengthens public confidence in scientific research [3]. Data sharing also allows for larger sample sizes and replication of results and analysis, resulting in many benefits for neuroimaging research [4]. Furthermore, data sharing can increase the chances of conducting research in developing countries at lower costs by reusing datasets [5] and can help researchers conduct test-retest studies when more

*Corresponding author:
E-mail address: Nurul Hashimah Ahamed Hassain Malim <nurulhashimah@usm.my>.
https://doi.org/10.56532/mjsat.v3i3.155
2785-8901/

neuroimaging public databases are available, thus improving the reproducibility of study findings [6].

The availability of shared neuroimaging data that is FAIR-compliant, such as magnetic resonance imaging (MRI), electroencephalography (EEG), magnetoencephalography (MEG), and positron emission tomography (PET), has significantly improved during the past ten years [2, 7]. Even though sharing data is important, it is hard for the neuroscience community to meet this highly requested need. This is because neuroscience research involves a lot of complicated experiments with a lot of different types of data from many different modalities [8]. These various data types are frequently highly specialized for specific domains and are technologically and scientifically designed to match specific neuromodalities and data types. As mentioned by Rübel et al. [9], neuroscientists are often required to work with dozens of different formats even in a single experiment, one for every single recording modality and/or analysis. Standards for such a great quantity of data are not described profoundly, and, in many instances, these data are only accessible through certain proprietary software programmes. Thus, the efficiency of data sharing and analysis might be hindered, and, in the worst-case scenario, it might lead to data inaccuracy and misinterpretation.

In this review paper, we review various types of neuroscience data formats, models, repositories, and analytics. The need for the standardization of neuroscience data formats and models, standardizing data storage formats, and developing highly reusable data analytics tools will be critical prerequisites for effective data sharing. In Malaysia, there is currently no neuroscience data repository or data sharing. This review will also be a great help for Malaysian researchers to come up with a neuroscience data repository and data sharing.

2. NEUROSCIENCE DATA FORMAT

The best format for neuroscience data is one that is "simple, efficient, flexible, and contains full information about the stimulation and simulator from which the data came" [10]. A shared, open, and standardized file format that is adaptable enough to represent various types of data along with metadata may boost both the development of community-based tools and data sharing between various labs [11]. Still, making or standardizing a consensus data format for sharing data is a very hard and tough task [12]. For this kind of standardization to happen, there are a lot of complicated requirements that go beyond the simple and common requirements of data formats that are specific to a mode, like video or image formats. This standard data format needs to have the competency to support researchers in managing and organizing complex data collections acquired from different types of modalities and stimulators (e.g., eye and/or motion tracking, neurological imaging recordings, etc.) [9].

2.1 Data Format Standardization

The growing understanding of how crucial it is to share data has made it possible for various data or file formats (see Table 1).

Table 1. Summary of the neuroscience data format

Types of data	Data format
Poly-graphic recording data	Extensible Biosignal (EBS) European Data Format 'Plus' (EDF+) General Data Format (GDF)
Cellular-related electrophysiological recording data	Kwik Svoboda Lab File Format Orca
Electroencephalography (EEG) data	Multiscale Electrophysiology File Format (MEF)
Electrocortigraphy (ECoG) data	Lawrence Berkeley National Laboratory (LBNL BRAINformat)
Simulation experiment data	Neuroscience Simulation Data Format (NSDF)
Neurophysiology data	Neurodata without Borders: Neurophysiology (NWB:N)

Kemp et al. [13] were one of the first groups to try to make a simple but standard format for polygraphic recording data. This format was later used in seven laboratories to share sleep-wake recording data. Since then, several studies have started to come up with different standard data formats that will make it easier to share neuro-data. For instance, the EBS file format [14], the EDF+ [15], as well as the GDF [16], have succeeded the aforementioned format in the field of electroencephalograms and other related biomedical signals.

In the last 10 years, some of the most important neuro-data formats have been created to create standard data formats and make it easier for large-scale neuroscience data to be shared between user groups. They adopted different characteristics and have unique advantages over one another, even for similar data categories. For instance, data formats such as Kwik [17], Svoboda Lab File Format [18], and Orca [19] were created for cellular-related electrophysiological recording data.

Kwik is useful for sorting and fine-tuning spike-recording signals collected from electrophysiological modalities. It is especially optimized for automatically sorting multi- or high-channel electrophysiology signals. It adapted a modular design and a user-friendly approach that gave rise to its extensibility and made it relatively flexible [17, 19]. The Svoboda file format has features of extensibility and flexibility as well. One difference is that the Svoboda format makes use of the "session object" approach, in which the processed data and experimental metadata descriptions are computed in hash table format. The hash table was preferred because searching was greatly facilitated, and the code was simplified to promote an easier understanding of the data collections in the key-value pairs [8–19]. Orca, on the other hand, is a data format designed by the Allen Institute for internal data sharing between the labs in the institute. Utilizing an object-oriented

design, this format promotes high levels of modularity and extensibility, as well as backward compatibility [19].

Other data formats have also been designed to account for different types of data to match different user groups or user preferences. The MEF [20] is mostly used for large-scale electroencephalography data. It has features like data compression, data encryption to keep private information safe, and data redundancy to make sure the data is correct. The "manage object" idea was used to make the LBNL BRAIN format [21] for ECoG signals. It has a semantic component that makes it easy for users to manage the data in a way that makes sense and is interactive for the applications. It was developed using a modular approach and was optimized for data reuse. Besides, the NSDF [10] is another type of data format that was developed for a broad range of neuronal stimulation outcomes, ranging from models such as multiscale electrical and/or chemical signals, single-neuron recording data, and abstract neural networks. This format excels in that it is simulator-independent, allowing various kinds of stand-alone tools to manage, analyze, and visualize data stored in NSDF format [10].

Even though there are a lot of different data formats, there is still no one format that is the most widely used and accepted standard in the neuroscience community. This is because the most popular data formats, like Kwik, Orca, LBNL BRAINformat, and some of the other formats listed above, are mostly used by certain subgroups of the neuroscience community and domain. Hence, they might not be generic enough to support the wide range of neuroscience research data. Svoboda Data Format is not very user-friendly and is hard for beginners to learn because it is designed to hold a lot of information in a small space [19]. Similarly, the NSDF format was limited in terms of its structure for spatial-related data, which it does not support well but is becoming increasingly important in neuroscience research with various neuroimaging modalities [10].

NWB:N [22], one of the newest and ongoing data formats, was started as part of the Neurodata Without Boarders project to try to fix some of the problems with the other data formats. The NWB:N format is a successor to the previous "Orca" data format, and some of its structure is inspired by formats such as Kwik, Svoboda, and LBNL BRAINformat. NWB:N is designed for a broad range of neurophysiological data, such as neuronal activities (i.e., intra- and extracellular electrophysiology, two-photon imaging, etc.) and cell-based neurophysiological data. It adopted and supported a modular design and facilitated data plotting and comprehension [22], like LBNL BRAINformat did, but it excels in terms of wider data type coverage and providing a source of information that the former did not.

Nevertheless, NWB:N has room for improvement to overcome its shortcomings. One of the very important aspects that is lacking in the NWB:N format is file or data validation [22]. It is highly valuable to take data and file validation into account while designing the standard data format. As the collected data are the core resources for further analysis and interpretation, a validation process that covers different types of validation is highly desirable. Such a need for a validation tool should not only be considered by the NWB team but also be improved by other existing file formats, as well as be highly considered by future initiatives while designing new standard file formats.

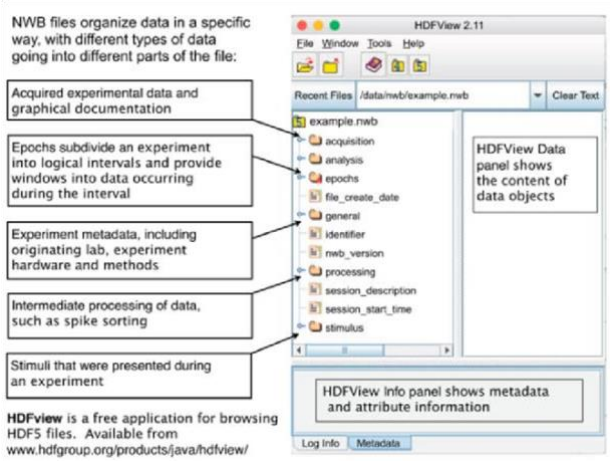

Fig. 1. Layout of NWB File in HDFview [22]

2.2 Application-Programming Interface: Data Conversion

Apart from designing and developing standard file formats, an alternative approach to facilitate neuro-data sharing is via data conversion by defining and involving an application programming interface (API). The German-Note (G-Note) common interface called Neuroshare API [23] is one early attempt to access various neurophysiological data files and formats. It is set up as a library that lets neurophysiological applications, such as data visualization programs, spike sorting application programs, and other neuroscience analysis software, directly access and extract data in multiple vendor formats using specific proprietary libraries. Nevertheless, the Neuroshare API suffers from several shortfalls. First of all, it is highly platform-specific, meaning it can only be accessed via the Windows operating system and is heavily based on vendor-specific formats. It does provide descriptions of how data is retrieved, but it does not promote standardization of data storage methods. Furthermore, the proprietary-specific libraries were closed-source, which meant that they required the purchase of corresponding manufacturers' or commercial hardware and/or software before they could be used [23].

Later, a well-known API called Neuroscience Electrophysiology Objects (NEO) succeeded the Neuroshare API. NEO [23] is an object-oriented API intended to promote software interoperability for neurophysiological datasets. It is a cross-platform open-source package with a Python implementation that handles data conversion for most of the electrophysiology-related proprietary formats. Data representation was conceptually separated from data analysis and visualization. The core function of this API is deliberately focused on data representation to promote a lightweight interface, while software was constructed upon NEO for analyzing and visualizing neurophysiology data. NEO is a relatively powerful tool for data conversion, but there is still room for improvement. As NEO is a Python-only API implementation, expanding its implementation to other languages apart from Python could be one of its more improveable aspects.

3. DATA MODEL

3.1 Data Model Representation

Brain Imaging Data Structure (BIDS) [12], the first International Neuroinformatics Coordinating Facility (INCF) endorsed standard, is a data model or structure for organizing and managing neuroimaging data. BIDS uses simple language, but it also has a clear structure for organizing and managing data. It adopted a rather simple folder structure to organize neuroimaging data. Initially, it was developed as a standard to organize raw magnetic resonance imaging (MRI) data [12]. It has then extended its coverage to include additional data types such as magnetoencephalography (MEG) and EEG [24, 25, 26]. With a consistent hierarchical directory and naming system, filenames in BIDS are formed by starting with a key-value sequence and terminating with a related file type, in which the key, key values, and file type are chosen and predetermined by the users [12]. Such a simple system for organizing files and data could allow the user to locate and manage the data and files effortlessly.

The most recent standard that INCF has endorsed for the year 2020 is the Neuroscience Information Exchange (NIX) data model [27]. It is also a relatively simple and generic data model for neuroscience data storage. It allows the user to deposit a fully annotated neuroscience dataset by storing the neurodata along with its metadata in an identical container. It has high flexibility in the way it can manage various types of data.

3.2 Metadata model: Open Metadata Markup Language (odML)

In neuroscience studies, a lot of data, both processed and not processed, was made and changed. Not only do these data include the main data, like recordings and signals from the neuromodalities, but they also include datasets that describe the experimental conditions. These datasets are called metadata [28]. Metadata is information about the parameters of an experiment, how it was recoded, and other procedures and information about the experiment [29]. In a simplistic way of saying it, metadata is the data of data. The importance of these metadata should not be neglected because replication and reconstruction of research procedures and experimental analysis are highly dependent on these details. Despite their importance, these metadata are usually not organized in an orderly manner and are often lost during the experiment, which has become an open issue revolving around the neuroscience community. As such, implementing a unified or standard data model for managing and handling these metadata in an understandable and concise manner has been highly recommended.

Open Metadata Markup Language (odML) is a software framework that may be used to handle metadata for neuroscientific research [30, 31]. odML is a standard for storing metadata in a structured, machine- and human-readable manner. It offers a common schema (with implementations in XML, JSON, and YAML) to integrate metadata from diverse sources without restricting the content of the metadata. In addition, by offering metadata terminologies, odML promotes and supports standardization. As a metadata model, odML [28] can be thought of as a flexible structure that makes it easy to automatically collect and exchange metadata in a way that is complete, well-organized, and easy for machines to understand. With this, it allows for the uniform organization of metadata coming from many sources and the recording of that data in a standard, interoperable format. Because metadata can be organized and made available to all members of a scientific project in a consistent manner, providing it in such a standardized format along with an experiment's data files makes it easier for members of the project to collaborate [31, 32 It also supports the accuracy and consistency of data analysis through standardized and formalized access to the available metadata. It sets up the metadata in the form of key-value pairs based on a tree-like structure. Still, the developers have thought about the fact that tree-like structures can sometimes lead to organizations that are redundant and hard to understand. Because of this, they have added and combined things that aren't part of the tree-like structure. These elements were used to set up relationships. For instance, users could set up relationships for stimuli that happened more than once by defining the stimulus and then linking it to all the related datasets [28].

Additionally, as this model suggested, the format and content were clearly distinct, with the format itself not defining the keys or values. This made the model more flexible because it made it possible to store all available metadata right away without having to send new keys to the ontology [28]. odML has set up a number of standard odML terminologies. These terminologies were not made to force standards on the ontologies but to give users a choice. One could always go against the given terminology if it doesn't fit their personal tastes or goes against the standard terminology in their communities or labs. Even though consistency would be at risk with this approach, users would have to do most of the work to figure out if it was valid [28].

3.3 Neuronal Model: Neural Open Markup Language (NeuroML)

The complexity of neuronal functions has correspondingly induced complexity in describing, explaining, and managing the relevant neuronal models in neuroscience studies. As such, it is much needed to develop methodologies or implementations to promote collaboration among neuroscientists in the modelling process. In this sense, software applications that support and facilitate the discussion and exchange of these models are high on the list of necessities. The implementation of mark-up language is an approach for better describing the neuronal system models, connecting or linking databases of models, and formatting the neuronal data models in a form that is more compatible with the simulation software programme or applications [33].

NeuroML [33, 34, 35] is a model language for neuroscientific concepts that is based on Extensible Markup Language (XML). Using the standard description language, XML, has made it easier to work with other systems, made neuronal data models easier to access, and made it easier to reuse neuronal models. NeuroML defines, describes, stores, and exchanges detailed neuronal models in a standalone format. This model language is stimulator-independent, meaning that neuronal data models stored in NeuroML format can, in turn, be used and handled across multiple different simulators or applications. For example, a neurological channel model that was implemented in NeuroML, without regard to the original simulator used, can be converted to various kinds of formats for later inspection and analysis [34].

Table 2 shows the summary of the neuroscience data model.

Table 2. Summary of the neuroscience data model

	Neuroscience data model
BIDS	The INCF has approved the use of the common data model known as BIDS to organize and manage MRI data. Later, it was extended to include MEG and EEG data. To provide the community with an easy-to-use system for organizing neuroimaging data, BIDS was created.
NIX	NIX specifies a data model for annotated scientific datasets, or data combined with metadata, as well as a related file format based on HDF5 for storing and sharing such datasets. This format was created specifically for the metadata-assisted storage of electrophysiology and other neuroscience data.
odML	odML is a format for storing metadata in an organized human- and machine-readable way and a software approach to managing neuroscientific metadata.
NeuroML	NeuroML is a model language for concepts from the field of neuroscience that is based on XML. With the help of the standard description language, XML, it is now simpler to collaborate with other systems, access neuronal data models, and reuse neuronal models. The detailed neural models are defined, described, stored, and shared using NeuroML in a stand-alone format.

4. DATA REPOSITORY

The availability and accessibility of data repositories in neuroscience communities is also one of the very important factors that influences and facilitates data sharing behaviour among neuroscientists. Table 3 presents a summary of the available data repository for neuroscience data.

Table 3. Summary of the neuroscience data repository

Neuro-data repositories	Descriptions
NeuroMorpho.Org	It provides neuronal reconstruction data and associated metadata.
NeuroVault	It is a place where researchers can publicly store and share unrestricted statistical maps, parcellations, and atlases produced by MRI and PET studies.
NITRC-IR	It is a platform for depositing and sharing MRI data in DICOM and NIfTI formats.
OpenNeuro	It is an open platform for validating and sharing BIDS-compliant MRI, PET, MEG, EEG, iEEG, and ECoG data.

NeuroMorpho.Org [36, 37, 38], which was started in 2006, is one of the largest web-accessible repositories for publicly shared digitally reconstructed neuromorphological descriptions. This implies that other neuroscientists can digitally reconstruct and use previously stored neuromorphology. Due to the fact that neuronal reconstructions are different in terms of the field of the original research studies (e.g., electrophysiological, physiological, or anatomical) and the data-related specifications (e.g., data format, file size, and resolution of the recoding), NeuroMorpho.Org has tried to come up with a design that can accommodate all of these aspects of neural morphology studies. This repository associates three types of data [36]. First, metadata related to the reconstruction, such as the data source, research subjects' data, and experimental methodology, were extracted. Second is the data file itself, both the raw and standardized versions. The third type of data is the information related to the twenty-one morphometric calculations, which were used to calculate the shape variation of each neuromorphology reconstruction. NeuroMorpho.Org has served as a repository that stores over 120,000 digital neuron reconstructions of neuronal and glial morphology data [39].

NeuroVault [40, 41] is another place where researchers can store and share statistical brain maps, neuroimaging images, and metadata. Since it's a web-based platform, users don't have to install any additional software to use it. Also, one benefit of this repository is that it creates a permanent link for the data or images that are uploaded. This makes it easy for researchers to share the stored data by accessing the link. One thing to keep in mind is that neither the developer nor the NeuroVault platform supported the meta-analysis method [40]. This is neither good nor bad and pretty much depends on user preferences because some users might wish to have the meta-analysis processes done within the same platform, while others might opt for other well-suited or prevalent analysis software.

Neuroimaging Informatics Tools and Resources The Clearinghouse Image Repository (NITRC-IR) [42] is another available neuroscience repository. It serves as an image database platform for depositing and sharing MRI data collected in different states, such as resting, diffusion, and structural states. It is based on the eXtensible Neuroimaging Archive Toolkit (XNAT), which is a software platform that was developed for managing neuroimaging data. One limitation of NITRC-IR is that it does not host neuroimaging data from other domains except MRI, although XNAT is capable and supportive of doing so [42].

OpenNeuro [43, 2], launched in 2018, is also a free data repository available to the neuroscience community. OpenNeuro is the successor to OpenfMRI [44], which was launched in 2011 and is primarily designed for task-based functional neuroimaging data. It stores raw and processed whole-brain functional magnetic resonance imaging (fMRI) datasets as well as related metadata. Succeeded by OpenNeuro, it is now based on the BIDS, allowing the researchers to store and share various kinds of neuroimaging data collected from different neuromodalities such as MRI, PET, MEG, EEG, intracranial electroencephalography (iEEG), and EcoG [42, 2]. Its wide coverage for depositing different neuroimaging data appears to be a good complement to other repositories mentioned, such as NITRC-IR.

As an idea for the future, linking these databases is one way to make it easier to share data. The abovementioned repositories serve their purposes by providing a platform for the community to store and manage neuroscience data. However, these databases are sometimes exclusive to certain data types; for example, NeurorMorpho.Org only caters to neuromorphological data, and NITRC-IR only serves MRI data types. If it's possible, it would be a good idea to connect these repositories or give users an easy way to quickly access different databases in the future. This would give users a single place to access all the different kinds of data stored in these repositories.

5. DATA ANALYTIC

In the domain of neuroimaging, there are three prominent software packages that are often used for data analysis: Statistical Parametric Mapping (SPM), FMRIB Software Library (FSL), and Analysis of Functional NeuroImages (AFNI).

SPM (www.fil.ion.ucl.ac.uk/spm) is a free statistical analysis software written in MATLAB that is used to analyze and identify specific regional effects and brain activities from different neuroimaging recordings, such as fMRI, EEG, MEG, single-photon emission computed tomography (SPECT), and positron emission tomography (PET) [45, 46]. It is especially prevalent to examine functional neuroanatomy and relative brain activity changes. It used a voxel-based approach to infer the brain's specific region of interest in response to the experimental stimulus and factors, and then mapped the recorded activities to specific brain structures and anatomical space [47].

FSL (www.fmrib.ox.ac.uk/fsl/), written in C, is a relatively comprehensive software for processing and analyzing different MRI-related brain imaging data, including structural or anatomical MRI images, functional MRI recordings, and diffusion-weighted MRI images. FSL is capable of combining and integrating various types of data into a single, consolidated analysis because it created and implemented the Bayesian framework [48].

Similarly, AFNI (http://afni.nimh.nih.gov/afni/), written in C++, is a well-known and freely available software tool for analysing structural and functional MRI image data. Like SPM, it is capable of normalizing the images according to Talairach coordination and mapping the neuronal activation onto the anatomical images, as well as performing default pre-processing such as realignment based on voxel time series, normalization of structural volume, and smoothing the volume for further statistical inferences [49].

There is also other software for analysis that was made for the general purpose of processing and analysing data in neuroscience studies. Functional Imaging Analysis Software, Computational Olio (FIASCO) is another general-purpose tool that provides pre-processing functions such as detrending and motion correction, fits linear models to the data, and displays or visualizes images. It provides greater flexibility to the users by allowing them to customize their analyses by writing their own procedures [49]. BrainVoyager QX is another powerful and complete software package for processing and analysing neuroimaging data. It was written in C++. It started out as a tool for analysing anatomical and functional MRI neuroimaging data, but it has since become a multi-model analysis software package that can handle a wide range of neuro-data from different modalities, such as diffusion tensor imaging (DTI), EEG, MEG, and transcranial magnetic simulation (TMS) [50].

Aside from the above-mentioned software packages, which cover a wide range of data analytics tasks, there are other software packages that do more specific processing and analysis tasks. BrainVisa [51] is a software programme developed for image processing. It allows users to perform sequences of actions with command lines using a simple control panel. Various tools or toolboxes have been embedded in BrainVisa to provide additional processing features for visualizing and analyzing multi-modal neurodata collected from different neuromodalities. Other processing modules as well as data formats can be easily added to the BrainVisa platform, allowing users to efficiently manage their data [51]. VoxBo [49], a programme designed for MRI data analysis, has embedded functions for standard preprocessing such as normalization, smoothing, and motion correction. The analysis approach of this software is based on a univariate general linear model, disregarding other types of analyses, which may be one of its drawbacks compared to other software.

SPM, FSL, and AFNI, on the other hand, have good graphical user interfaces (GUI). FIASCO, on the other hand, does not have an embedded GUI and instead uses command-line operations. Large GUIs may make programmes easier to use, but some might say that this could make them less flexible and make it harder for users to customise their analysis [49]. Besides, SPM is based on MATLAB, while FSL, AFNI, and BrainVoyager QX are examples that are based on the C and C++ programming languages. Some have argued that MATLAB is not strong or robust enough to meet the complexity of neuroimaging and is not supportive enough to match the code size of the existing neurotechnology-related industrial level, and that C languages are not supportive enough for rapid development [49], and thus some have recommended the Python programming language instead.

When comparing these prominent data analysis software programmes, it is undeniable that they all have advantages and drawbacks over one another. Hence, it is a matter of preferences for the user to opt for which software tool to use to perform the data analysis, not to mention that users are not limited to one independent tool but are free to optimize two or more analytic tools in their studies (see Table 4). Nevertheless, one technical issue that requires attention is the necessity to establish a common representation across different software packages. This is due to the fact that, while popular software packages perform similar analysis on neuroimaging data, the terminologies and parameters employed are sometimes inconsistent and do not refer to the same concept [52]. As such, these software packages do accomplish their role of supporting the analysis process, but in terms of the intention to promote data sharing, such inconsistency would hinder the users' ability to compare the analyzed data across software. Defining or developing a unified descriptive standard would be one of the main focuses for the developers when designing a new package or upgrading the existing version.

Table 4. Summary of neuro-data analytics

Software	Descriptions
SPM	Free statistical analysis software Written in MATLAB To analyze brain activities form fMRI, EEG, MEG, SPECT and PET data
FSL	It is freely available for non-commercial use Written in C language Process and analyse MRI, fMRI and DTI brain imaging data
AFNI	Free software Written in C++ Analyze structural and functional MRI data
FIASCO	Free software To analyzee fMRI data
BrainVoyager QX	Written in C++ To analyse MRI, fMRI, DTI, EEG, MEG and TMS data
BrainVisa	Neuroimaging software platform for mass data analysis Morphologist: brain segmentation and sulcal morphometry Processing tools and toolboxes Interactive 3D neuroimaging data visualization

6. DISCUSSION AND CONCLUSION

Given the above discussion on data formats, data models, data repositories, and data analytics in their respective sections, one should acknowledge that these domains should not be viewed in segregation but rather that they all work interdependently to facilitate the sharing of data with the neuroscience community. As such, international coordination or initiatives hold a significant role with regard to the development of these domains. One of the most visible initiatives facilitating such development is the INCF [19]. Founded in 2005, the INCF is an international organization with many divisions around the globe that aims to coordinate the cooperation of interdisciplines between neuroscience and information science. INCF is focusing on four main areas of development: digital brain atlasing, ontologies for neural structures, multi-scale modelling, and standards for data sharing. It is working towards developing and reviewing standard data formats, advancing the data management model, digitalize neurological data and information, and creating common ontologies for the communities.

Despite the fact that data sharing is greatly simplified in the modern era, there are a few associated challenges to be concerned about. The first challenge is quality control. According to Poldrack and Gorgolewski [7], in the process of data sharing, researchers are greatly encouraged to share beneficial neuro-data and to re-use those data to promote further understanding of human neurology and its functions. However, the quality and integrity of the data collected were not safeguarded. One could not accurately determine the condition in which the data were obtained, nor could one determine whether the data acquisition procedures were in fact consistent with their publication. In this regard, optimizing the tools or taking the initiative to perform some degree of quality control was highly recommended to ensure the data being shared met the minimum quality standards.

Another extremely important domain to consider during the data sharing procedure is ethics. Privacy and confidentiality issues have often been raised during data sharing [53]. This is because in studies related to neuroimaging, there is a possibility that the subjects could be identified through reconstructing the individual's facial features or structural features. A "de-facing" technique is often used to mitigate such a risk, but there is still a risk of subject identification via other confidential information such as age and other related demographical details [53]. With regard to this, setting higher restrictions on highly confidential information would be a positive way to protect subjects' privacy throughout the course of data sharing.

In addition, due to the progressively complex nature of data analysis, it would become increasingly complicated to perform analysis replication without access to the original analysis code [53]. As such, apart from sharing the neuro-data collected from the study, sharing related analysis and processing code utilize in the respective study is also a necessity. Having the researcher submit the analysis code used during the process of data analysis along with their data would greatly benefit the neuroscience community, as it would diminish the difficulty of reproducing the analysis from scratch.

Along with solving the problems listed above, creating a standard neuro-data format and model for each type of neuro-data, and making data repositories and analysis tools better, the neuroscience communities could share a lot of data, which would undoubtedly help people understand neuroscience better. Nonetheless, such progress is the result of the combined efforts of every member of the community.

ACKNOWLEDGEMENT

We would like to express our gratitude to all of USM's staff, especially lecturers and staff from the School of Medical Sciences and the School of Computer Sciences, USM for their guidance and contribution in finishing this manuscript.

AUTHOR CONTRIBUTIONS

Conception and design: Fong Sze Wei, Nurul Hashimah Ahamed Hassain Malim. Drafting of the manuscript: Fong Sze Wei, Nurfaten Hamzah. Critical revision of the manuscript: All authors. All authors read and approved the final version of the manuscript.

REFERENCES

[1] D. B. Keator et al., "Towards structured sharing of raw and derived neuroimaging data across existing resources," NeuroImage, vol. 82, pp. 647–661, 2013. doi:10.1016/j.neuroimage.2013.05.094

[2] C. J. Markiewicz et al., "The OpenNeuro resource for sharing of Neuroscience Data," eLife, vol. 10, 2021. doi:10.7554/elife.71774

[3] A. S. Jwa and R. A. Poldrack, "The spectrum of data sharing policies in neuroimaging data repositories," Human Brain Mapping, vol. 43, no. 8, pp. 2707–2721, 2022. doi:10.1002/hbm.25803

[4] K. Rootes-Murdy et al., "Federated analysis of Neuroimaging Data: A review of the field," Neuroinformatics, vol. 20, no. 2, pp. 377–390, 2021. doi:10.1007/s12021-021-09550-7

[5] Donaldson, Devan Ray, and Joshua Wolfgang Koepke. "A Focus Groups Study on Data Sharing and Research Data Management." Scientific Data 9, no. 1, 2022. https://doi.org/10.1038/s41597-022-01428-w.

[6] P. Gao et al., "A Chinese multi-modal neuroimaging data release for increasing diversity of human brain mapping," Scientific Data, vol. 9, no. 1, 2022. doi:10.1038/s41597-022-01413-3

[7] R. A. Poldrack, and K. Gorgolewski, "Making big data open: Data sharing in neuroimaging" Nature Neuroscience, vol. 17, no. 11, pp. 1510-1517, Oct 2014, doi: 10.1038/nn.3818

[8] W. Thomas, P. Bauer, M. Denker, S. Grün, M. Hanke, J. K., Steffen Oeltze-Jafra, et al. "NFDI-Neuro: Building a Community for Neuroscience Research Data Management in Germany." Neuroforum, 2021. https://doi.org/10.1515/nf-2020-0036.

[9] O. Rübel, Prabhat, P. Denes, D. Conant, E. Chang, and K. Bouchard. "BRAINformat: A Data Standardization Framework for Neuroscience Data" Biorxiv, pp. 1-23, Aug. 2015, doi:10.1101/024521

[10] S. Ray, C. Chintaluri, U. S. Bhalla, and D. K. Wójcik, "NSDF: Neuroscience Simulation Data Format," Neuroinformatics, vol.14, pp. 147-167, Nov. 2015, doi:10.1007/s12021-015-9282-5

[11] A. Stoewer, K. C. Benda Jan, W. Thomas, and G. Jan. "File Format and Library for Neuroscience Data and Metadata." Frontiers in Neuroinformatics 8, 2014. https://doi.org/10.3389/conf.fninf.2014.18.00027.

[12] K. J. Gorgolewski, T. Auer, V. D. Calhoun, R. C. Carddok, S. Das, E. P. Duff et al, "The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments," Sci Data, vol. 3, pp. 1-9, doi: 10.1038/sdata.2016.44

[13] B. Kemp, A. Värri, A. C. Rosa, K. D. Nielsen, and J. Gade, "A simple format for exchange of digitized polygraphic recordings," Electroencephalogr Clin Neurophysiol, vol. 82, no. 5, pp. 391-303, May. 1992, doi:10.1016/0013-4694(92)90009-7

[14] G. Hellmann, M. Kuhn, M. Prosch, and M. Spreng, "Extensible biosignal (EBS) file format: simple method for EEG data exchange," Electroencephalography and Clinical Neurophysiology, vol. 99, no. 5, pp. 426-431, Nov. 1996,

[15] B. Kemp and J. Olivan, "European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data," Clin Neurophysiol, vol. 114, no. 9, pp. 1755-1761, Sep. 2003, doi: 10.1016/s1388-2457(03)00123-8

[16] A. Schlögl, "GDF-a general dataformat for biosignals,"ArXiv Prepr. Cs0608052, 2006.

[17] C. Rossant, S. N. Kadir, D. F. M. Goodman, J. Schulman, M. Belluscio, G. Buzsaki, and K. D. Harris, "Spike sorting for large, dense electrode arrays," Nature Neuroscience, vol. 19, pp.634-641, Mar. 2016, doi:10.1101/015198, 2015

[18] Collaborative Research in Computational Neuroscience, "Svoboda Lab data format – General Information," 2014. [Online]. Available: https://crcns.org/files/data/alm-1/Svoboda_lab_data_format_general.pdf

[19] C. Friedsam, "Development of a new uniform file format for neuroscience data across the globe," M.S. thesis, Dept. Biotech., Harvard Univ., Cambridge, MA, USA, 2016.

[20] B. H. Brinkmann, M. R., Bower, K. A. Stengel, G. A. Worrell, and M. Stead, "Multiscale electrophysiology format: an open open-source electrophysiology format using data compression, encryption, and cyclic redundancy check," in Conf Proc IEEE Eng Med Biol Soc, 2010, pp. 7083-7086, doi:10.1109/IEMBS.2009.5332915

[21] O. Rübel, M. Dougherty, Prabhat, P. Denes, D. Conant, E. F. Chang, and K. Bouchard, "Methods for specifying scientific data standards and modeling relationships with applications to neuroscience," Frontiers in Neuroinformatics, vol. 10, pp. 1-16, Nov. 2016, doi:10.3389/fninf.2016.00048

[22] J. L. Teeters, K. Godfrey, R. Young, C. Dang, C. Friedsam, B. Wark, et al, "Neurodata Without Borders: Creating a common data format for neurophysiology," Cell Press, vol. 88, no. 4, pp. 629-634, Nov. 2015, https://doi.org/10.1016/j.neuron.2015.10.025

[23] S. Garcia, D. Guarino, F. Jiallet, T. Jennings, R. Pröpper, P. L. Rautenberg er al, "Neo: an object model for handling electrophysiology data in multiple formats," Front. Neuroinform, vol. 20, no. 8, pp. 1-10, Feb. 2014, doi: 10.3389/fninf.2014.00010

[24] P. Cyril R., S. Appelhoff, K.J. Gorgolewski, G. Flandin, C. Phillips, A. Delorme, and R. Oostenveld. "EEG-Bids, an Extension to the Brain Imaging Data Structure for Electroencephalography." Scientific Data 6, no. 1 (2019). https://doi.org/10.1038/s41597-019-0104-8.

[25] P., Cyril R., R. Martinez-Cancino, Dung Truong, S. Makeig, and A. Delorme. "From Bids-Formatted EEG Data to Sensor-Space Group Results: A Fully Reproducible Workflow with EEGLAB and Limo EEG." Frontiers in Neuroscience 14 2021.

[26] N. Guiomar, K. J. Gorgolewski, E. Bock, T. L. Brooks, G. Flandin, A. Gramfort, R. N. Henson, et al. "Meg-Bids, the Brain Imaging Data Structure Extended to Magnetoencephalography." Scientific Data 5, no. 1 2018. https://doi.org/10.1038/sdata.2018.110.

[27] M. Martone, R. Gerkin, R. Moucek, S. Das, W. Goscinski, J. Hellgren-Kotaleski et al, "NIX – Neuroscience information exchange format," Conference Abstract: Neuroinformatics, 2014. doi: 10.3389/conf.fninf.2014.18.00027

[28] J. Grewe, T. Wachtler, and J. Benda, "A bottom-up approach to data annotation in neurophysiology," Front Neuroinform, Aug. 2011, doi: 10.3389/fninf.2011.00016

[29] U., Hannes, A.-K. Kock-Schoppenhauer, N. Deppenwiese, R. Gött, J. Kern, M. Lablans, R. W. Majeed, et al. "Understanding the Nature of Metadata: Systematic Review." Journal of Medical Internet Research 24, no. 1 2022. https://doi.org/10.2196/25440.

[30] J. Sprenger, L. Zehl, J. Pick, M. Sonntag, J. Grewe, T. Wachtler, S. Grün, and M. Denker. "OdMLtables: A User-Friendly Approach for Managing Metadata of Neurophysiological Experiments." Frontiers in Neuroinformatics 13 2019. https://doi.org/10.3389/fninf.2019.00062.

[31] J. Grewe, T. Wachtler and J. Benda. "A bottom-up approach to data annotation in neurophysiology". Front. Neuroinform. Vol. 15, no. 16, 2011 doi: 10.3389/fninf.2011.00016

[32] L. Zehl , F. Jaillet, A. Stoewer, J. Grewe, A. Sobolev, T. Wachtler, et al. "Handling metadata in a neurophysiology laboratory". Front. Neuroinform. vol. 10, no. 26, 2016. doi: 10.3389/fninf.2016.00026

[33] N. H. Goddard, M. Hucka, F. Howell, H. Cornelis, K. Shankar, and D. Beeman, "Towards NeuroML: Model description methods for collaborative modeling in neuroscience," Philosophical Transactions of The Royal Society B Biological Sciences, vol. 365, pp. 1209-1228, Sep. 2001, doi: 10.1098/rstb.2001.0910

[34] S. M. Crook, J. A. Bednar, and R. C. Cannon, "Creating, documenting and sharing network models," Network Computation in Neural Systems, vol. 23, no. 4, Sep 2012, doi: 10.3109/0954898X.2012.722743

[35] G. Padraig, S. Crook, R. C. Cannon, M. L. Hines, G. O. Billings, M. Farinella, T. M. Morse, et al. "NeuroML: A Language for Describing Data Driven Models of Neurons and Networks with a High Degree of Biological Detail." PLoS Computational Biology 6, no. 6 2010. https://doi.org/10.1371/journal.pcbi.1000815.

[36] M. Halavi, S. Polavaram, D. E. Donohue, G. Hamilton, J. Hoyt, K. P. Smith, and G. A. Ascoli, "NeuroMorpho.Org implementation of digital neuroscience: dense coverage and integration with the NIF," Neuroinformatics, vol. 6, no. 3, pp. 241-252, Sep 2008.

[37] G. A. Ascoli, E. D. Duncan, and M. Halavi. "Neuromorpho.Org: A Central Resource for Neuronal Morphologies." The Journal of Neuroscience 27, no. 35 2007: 9247–51. https://doi.org/10.1523/jneurosci.2055-07.2007.

[38] M. A. Akram, S. Nanda, P. Maraver, R. Armañanzas, and G. A. Ascoli. "An Open Repository for Single-Cell Reconstructions of the Brain Forest." Scientific Data 5, no. 1 2018. https://doi.org/10.1038/sdata.2018.6.

[39] B., Kayvan, M. A. Akram, and G. A. Ascoli. "An Open-Source Framework for Neuroscience Metadata Management Applied to Digital Reconstructions of Neuronal Morphology." Brain Informatics 7, no. 1 2020. https://doi.org/10.1186/s40708-020-00103-3.

[40] K. J. Gorgolewski, G. Varoquaux, G. Rivera, Y. Schwarz, S. S. Ghosh, C. Maumet et al, "NeuroVault.org: a webbased repository for collecting

and sharing unthresholded statistical maps of the human brain," Front. Neuroinform, Apr. 2015, https://doi.org/10.3389/fninf.2015.00008

[41] K. J. Gorgolewski, G. Varoquaux, G. Rivera, Y. Schwartz, V. V. Sochat, S. S. Ghosh, C. Maumet, et al. "NeuroVault.Org: A Repository for Sharing Unthresholded Statistical Maps, Parcellations, and Atlases of the Human Brain." NeuroImage 124 2016: 1242–44. https://doi.org/10.1016/j.neuroimage.2015.04.016.

[42] D. N. Kennedy, C. Haselgrove, J. Riehl, N. Preuss, and R. Buccigrossi, "The NITRC image repository," NeuroImage, vol. 124, pp. 1069–1073, Jun. 2015, doi:10.1016/j.neuroimage.2015.05.074

[43] K. J. Gorgolewski O. Esteban, G. Schaefer, B. Wandell, and R. Poldrack, "OpenNeuro – a free online platform for sharing and analysis of neuroimaging data" F1000Research, vol. 6, 2017, doi: 10.7490/F1000RESEARCH.1114354.1

[44] R. A. Poldrack, and K. J. Gorgolewski, "OpenfMRI: Open sharing of task fMRI data," Neuroimage, vol. 144, pp. 259-261, Jan. 2017, doi: 10.1016/j.neuroimage.2015.05.073

[45] M. Behroozi, and M. R. Daliri, "Software tools for the analysis of functional magnetic resonance imaging" Basic and Clinical Neuroscience, vol. 3, no. 5, pp. 71-83, Aug. 2012.

[46] L., Vladimir, J. Mattout, S. Kiebel, C. Phillips, R. Henson, J. Kilner, G. Barnes, et al. "EEG and MEG Data Analysis in SPM8." Computational Intelligence and Neuroscience 2011 2011: 1–32.

[47] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, "Statistical Parametric Mapping," in Statistical Parametric Mapping: The Analysis of Functional Brain Images, CA, USA: Elsevier, 2007, pp. 10-31.

[48] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith. "FSL," NeuroImage, vol. 62, no. 2, pp. 782–790, Sep. 2011, doi:10.1016/j.neuroimage.2011.09.015

[49] N. Lazar, "Analysis of Functional NeuroImages: AFNI" in The Statistical Analysis of Functional MRI Data, Georgia Athens, GA, USA: Springer, 2008, pp. 248-254.

[50] R. Goebel, F. Esposito, and E. Formisano, "Analysis of FIAC data with BrainVoyager QX: From single-subject to cortically aligned group GLM analysis and self-organizing group IC," Human Brain Mapping, vol. 27, no. 5, pp. 392-401, May. 2006, doi:10.1002/hbm.20249

[51] Y. Cointepas, J. F. Mangin, L. Garnero, J. B. Poline, and H. Benali, "BrainVISA: Software platform for visualization and analysis of multi-modality brain data," NeuroImage, vol. 13, no. 6, p. 98. Jun. 2001, doi:10.1016/s1053-8119(01)91441-7

[52] C. Maumet, , T. Auer, A. Bowring, G. Chen, S. Das, G. Flandin, S. Ghosh, et al. "Sharing Brain Mapping Statistical Results with the Neuroimaging Data Model." Scientific Data 3, no. 1 2016. https://doi.org/10.1038/sdata.2016.102.

[53] J. B. Poline, J. L. Breeze, S. S. Ghosh, K. Gorgolewski, Y. O. Halchenko, M. Hanke, et al. (2012). Data sharing in neuroimaging research. Frontiers in Neuroinformatics, 6. https://doi.org/10.3389/fninf. 2012.00009