Euro BioImaging
Preparatory Phase II Project

D8.4 Procedure for enhanced process interoperability for the existing and future technology portfolio.

Project N.	688945
Project Title	Euro-BioImaging Preparatory Phase II
Project Acronym	EuBI PPII
Associated Work Package	WP8
Associated Task	Task 8.1
Lead Beneficiary (short name)	CRG, FISABIO
Nature	Report
Dissemination Level	Public
Estimated Delivery Date (Grant Agreement, Annex I)	31/12/2016
Actual Delivery Date	16/02/2017
Task leader	Timo Zimmermann
Contributors	Timo Zimmermann, Maria de la Iglesia-Vaya

Funded by the Horizon 2020 Framework Program of the European Union

Abstract

This document describes process interoperability requirements between different components of Euro-BioImaging, specifically between established technologies and new technologies in the portfolio. The different levels and characteristics of processes that require interoperability are defined, available tools are listed and a procedure is proposed to enhance interoperability for these cases.

Table of Contents

"Interoperability is a characteristic of a product or system, whose interfaces are completely understood, to work with other products or systems, present or future, in either implementation or access, without any restrictions."[1]

1. Introduction

For Euro-BioImaging (EuBI) to function efficiently as an infrastructure, its different components need to be able to work together and exchange user-project based information and possibly even other aspects of the user project. This need is met by applying interoperability procedures to the operation of EuBI.
The concept of interoperability exists in many different fields including computing, telecommunications, European politics (European Interoperability Framework), the medical industry, international transport networks and military operations. It is supplemented by many common sense approaches that are used without a formal definition whenever two different entities want to cooperate efficiently.
Because of EuBI being a biological and medical imaging infrastructure, interoperability concepts have to be integrated into several of its components. For this the computing and software based interoperability concept of exchangeability of data between different operating systems and softwares (data level) will be very important, but interoperability will also need to be addressed at the operational level.
Preparatory Phase II Work Package 8 deals with the identification and implementation of new technologies into EuBI and this is a clear instance where interoperability between established technologies and structures and the newly implemented ones will be needed. The procedures for this will be outlined in this deliverable. The concepts described here do however also have relevance for the general operation of such a heterogeneous infrastructure as EuBI, which combines the fields of biological and medical imaging which in themselves consist of many subfields.

2. Interoperability concepts

For describing the procedures below, it is helpful to define the following concepts of the degree of interoperability (see Figure 1). Many of the principles below are using data as examples as interoperability at the data level is an essential factor:

I. **(Limited) compatibility:** Some components or systems of a structure share interfaces with other components of the same structure, but not with all of them. This is very often the case for commercial imaging systems and software as there is interest by companies to connect their product range through proprietary interfaces that enhance their ability to connect different own products as part of an application suite while excluding competitor products with similar capabilities. Partial compatibility is often

[1] Taken from http://interoperability-definition.info/en/ of the French speaking Libre Software Users' Association (AFUL)

already provided by the company products as it is not useful to offer a hermetic system in a connected environment, but the compatibility is often limited to parts of the metadata and in the worst case to the bare image data. Optional compressed data formats for export may even introduce information loss at the data level at the first export step although in most cases loss-free export of the data (but not the meta-data!) is supported. This strategy may be partially an intentional competition strategy by a company, but economic product development also intrinsically limits the support of multiple (and continuously changing) data formats in-house. Full compatibility (transfer of all meta-data) is often provided by a service provider at a layer below the initial data generation, often in the form of a cooperation with the data provider to get access to non-open data formats, as it is in their interest to use the data for further operations that define their product offer. Inside this environment multiple proprietary formats can be read in and partially exported into, although it doesn't make sense for all meta-data and would be potentially harmful, as will be explained in the next section. In open source communities, many data formats are supported[2], these days often in cooperation with the data providers who provide internal information on their closed data formats. This is a benefit to the companies as the issue of compatibility is outsourced without committing internal resources.

II. **De facto standard:** In many fields, a proprietary solution turns over time into a de facto standard, either by being the first and without competition for a long time or by beating out competing solutions to a point that all entities in the field at least provide compatibility to this solution, in addition to their generic formats. Microsoft (MS) Office data formats are a good example of such de facto standards. MS Word's .doc format for text processing has left behind initially quite comparable product formats (e.g. WordStar, WordPerfect) and is supported as frequently as the open standard Rich Text Format (RTF) which was developed for text processing interoperability.

III. **Interoperability:** Full interoperability is achieved by all components of a structure being connected through a joint interface, an open standard. This provides the highest flexibility inside the structure.
In the biological imaging field the Open Microscopy Environment (OME) has established such an open standard, the OME-TIFF. Like many of the proprietary microscopy data formats it is based on the Tagged Interchange File Format (TIFF, TIF), but different from the proprietary formats it follows an open standard for all possible metadata that allows the efficient exchange of image-associated information between different platforms and between databases.
In the medical imaging field, the sharing and reusing of data (within or between labs) is difficult if not impossible and complicates the application of automatic pipelines and quality assurance protocols. To solve this problem, the Brain Imaging Data Structure (BIDS) initiative has developed a standard for organizing and describing MRI datasets (see also Annex 1). The BIDS standard uses file formats compatible with existing software, unifies the majority of practices already common in the field, and captures the

[2] See the Open Microscopy Environment's (OME) Bio-Formats (http://www.openmicroscopy.org/site/support/bio-formats5.3/)

metadata necessary for most common data processing operations. In DICOM or other scanner specific files, the BIDS standard requires users to provide additional meta-information. NIfTI is one of the file format chosen because it is the largest common denominator across Medical imaging software.

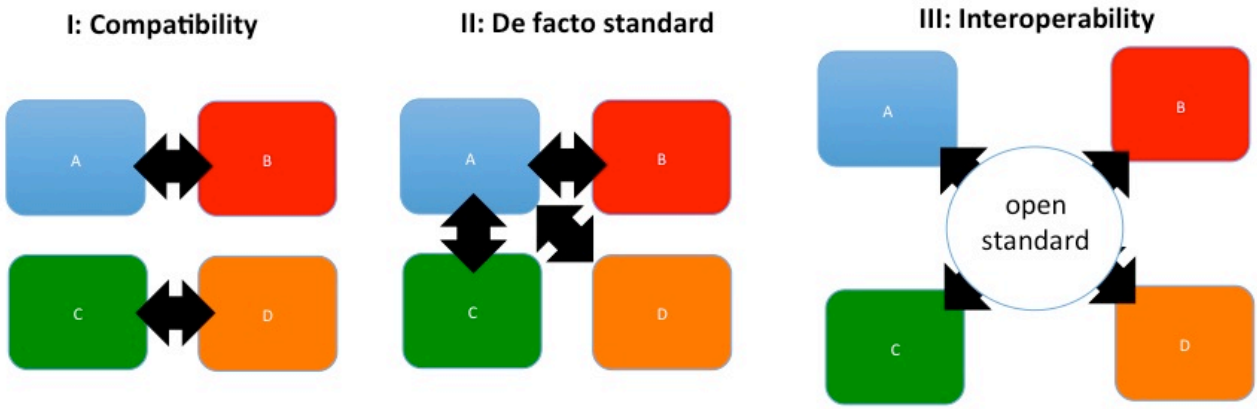

Figure 1: General interoperability concepts (adapted from the AFUL website http://interoperability-definition.info/en/)

Concepts I and III can be found in the EuBI infrastructure, often coexisting for the same operations. Concept II is not represented due to the plurality of participating entities (research labs, technology companies) of the imaging fields and the preferential data flow in imaging from data generating instruments to data processing applications and necessary limitations in the backflow.

3. EuBI interoperability requirements

3.1 Process interoperability requirements

Inside the EuBI infrastructure, process interoperability should allow the following:

A. Connect different technologies inside one project
B. Integrate data from new technologies
C. Integrate data from different imaging modalities
D. Integrate data from different imaging communities
E. Integrate data from a pan-European, distributed infrastructure (language, cultural, data management legal and ethical issues…)
F. Integrate with non-imaging data (this extends beyond the infrastructure)

Inside this deliverable, the focus will be set on points A-C. Points C-D are active areas of WP6 (Delivering Usable Data Resources for EuBI), more specifically the development of the EuBI Image Data Resource. Point F is an activity along the requirements of the Coordinated Research

Infrastructures Building Enduring Life-science Services (CORBEL) initiative, specifically CORBEL workpackage 6.

3.2 Technical considerations

Different from e.g. software platforms, interoperability for an infrastructure like EuBI does not imply omnidirectional flow between components, neither at the data nor at the operational or experimental level.

Types of data flows:

Image-to-Analysis (preferential): As already mentioned in the concept descriptions of the preceding section, there is a preferential flow of data and information from the data generating instrumentation to processing solutions. For proper further analysis all meta-data must be accessible to the software(s) executing this task.
The image data rarely need to be sent back to the acquisition software and this is generally not possible by reimporting the generic format after it was written into by a third-party software. Instrument specific meta-data can often be used to reconstruct instrument settings and the technology providers have to avoid that these settings are manipulated as this may damage the instrument.
Data flow is therefore partially a one-way street, but this is not a limitation for interoperability as the reverse flow in these cases makes no sense. The current open standard used in biological imaging applications (the OME Data model) acknowledges this distinction by defining three types of metadata for OME TIFF (see Figure 2):

- Core Metadata that contain the image-relevant information.
- Original Metadata that contain file-format specific information with different nomenclatures.
- OME Metadata contains the two above categories converted to the OME data model. This standardizes microscopy metadata and makes it accessible for database integration and for import into processing and analysis software packages.

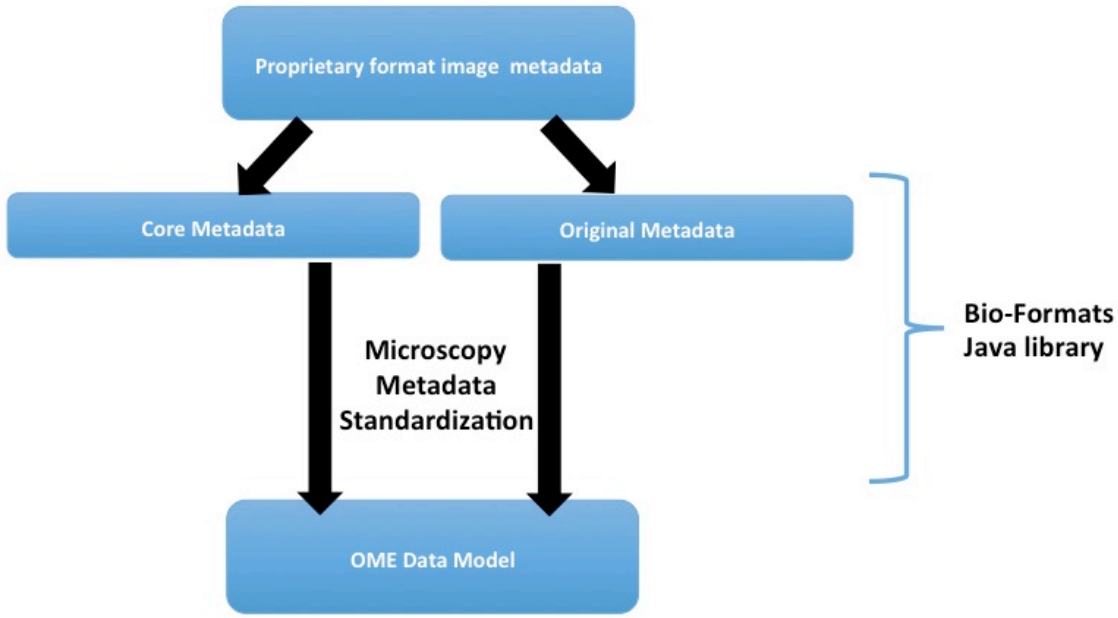

Figure 2: The OME data model for metadata conversion

Feed-back: Specific information on e.g. structure identification and image-data based decision making is sent back from an external software to the acquisition instrument. Image acquisition is then performed taking into account this information. This flow of information is essential for automated intelligent imaging strategies and high content data generation. This communication is executed through a specific interface provided by the acquisition software and used by the analysis software or alternatively by providing instructions in a file that can be imported for automated multidimensional imaging by the acquisition software.

Feed forward: In sequential workflows for correlative imaging, sample information data (position, dimension of the structure of interest) are passed on from one technology to another, either directly or indirectly after passing through additional processing or analysis platforms. In this case, interfaces must be shared between the different technologies. Depending on the imaging technologies and the sample requirements the technology sequence could be interchangeable, but for many correlative approaches (e.g. correlative light and electron microscopy), the sequence is clearly defined.

Operational flows:

Correlative Imaging: This is the experimental and sample-based equivalent of the feed forward data flow. At the level of sample handling and preparation however, no open standard can be implemented to connect different components. Similar to the interoperability concept in military operations, here it is important that different components establish and execute defined procedures to coordinate their specific activities. For sample-invasive approaches (common in biological imaging, e.g. sample fixation) the technology and handling sequence is mainly defined, for non-invasive approaches (common in medical imaging of patients or animal

models) the sequence may be changeable, but for every technology sequence a separate and clear protocol must be established. As a workflow, correlative imaging needs to be distinguished from multimodal imaging. For multimodal imaging (executed within the same examination) interoperability is required at the data level.

Non-correlative approaches: Here, different technologies can be combined inside projects mainly by interoperability at the data level using the above described data flow models.

Sample-to-Image-to-Analysis: For completeness one has to also consider that in most projects there exists a feedback loop between the analysis of acquired images and the preparation of subsequent samples. This is however at the level of the project/experiment and will not be considered as having interoperability requirements beyond the efficient connection of acquisition and analysis components.

All of the above described data and operational flows benefit from process interoperability that allows an efficient exchange of data, information and activities across the infrastructure. At the data level, a unifying open image data and metadata standard is the most important feature to allow interoperability inside the infrastructure.
At the level of the EuBI infrastructure, interoperability is most relevant when a project activity extends beyond one node. Of the workflows described above, this is mainly the case for the Feed forward (at the data level) and the Correlative Imaging (at the operational level) activities, but can also be the case for Image-to-Analysis, in case this involves different providers.

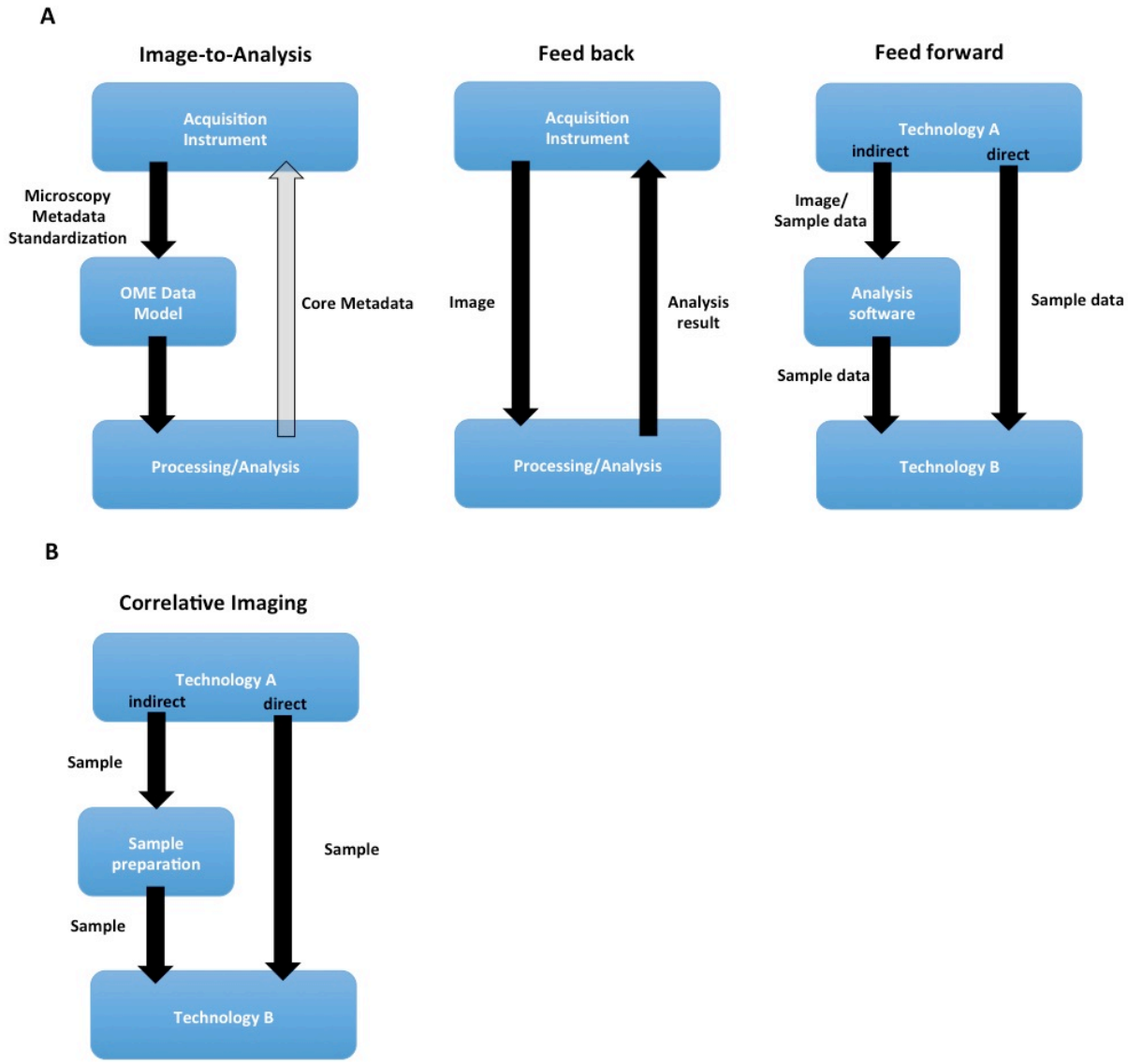

Figure 3: Schematics of the described data and operational flow concepts. A: Data flows B: Operational flow for correlative imaging

4. Tools for Interoperability

To deal with the requirements described above, the following tools are available to provide interoperability between different activities and processes:

- **Standards and ontologies:** This can be implemented at the data[3] and at the operational level. EuBI will need to identify the existing most suitable ones across all fields and technologies and implement their use across the Hub and all EuBI nodes. This is an activity of the currently ongoing hub creation and is also required for a consistent and

[3] See WP6 Task 4, D6.3

streamlined operation of the EuBI Web Acess Portal (WAP) and WAP-associated databases[4].

- **Interoperable metadata:** This can be achieved inside EuBI by using standardized interfaces for the biological and medical imaging fields as described above. There are ongoing efforts to make the thus generated standardized image information available to the EuBI users and the data and metadata level is the key to interactions with other infrastructures or initiatives[5].
- **Identification of gaps and provision of solutions:** An EuBI "Interoperability Platform", or expert group, that could be linked or not to the Technology Watch Panel (TWP)[6] could help in regularly adapting and fine-tuning the interoperability protocols. This is a concept that exists in the life science information infrastructure ELIXIR, which has interoperability at the heart of its function. An interoperability platform is very suited to ensure the proper integration of new technologies.
- **Training:** This is necessary for the correct implementation of common standards and ontologies inside the infrastructure. In the case of new technologies it can be included in the training for their operation.[7]

5. Procedure for interoperability of established and new technologies

With the requirements and tools defined, the following procedure to ensure interoperability of a new technology with established technologies is proposed[8]:

- Upon introduction of a new technology into the portfolio, the EuBI Hub assesses together with the TWP[9] the implications for interoperability this will have for the infrastructure. Ways how to best connect data and workflows with the existing infrastructure are defined. Compliance with the WAP workflows for project application, project reporting, statistics and and quality control is established to ensure compatibility with the WAP operation as applied to the already existing technologies. These steps are taken as soon as the technology is included and should ideally be concluded by the next call for nodes that features the technology.
- A EuBI Hub representative liaises with the technology providers to ensure that compatibility with the supported standard open biological or medical data interface is

[4] See WP5 (Technical preparation for user access)
[5] See CORBEL WP6
[6] Previously called Technology Watch Board, TWB (D7.2 and D7.3). To be further defined in Deliverable D3.3 "Detailed description of the EuBI ERIC executive management (description of all positions, reporting lines, responsibilities and operational procedures in the EuBi ERIC)"
[7] See WP8 Deliverable 8.1, WP7 Deliverable 7.3
[8] Prerequisite for this procedure: EuBI must have selected a standardized open data interface for the biological and the medical imaging field each that provides data level interoperability inside the two parts of the infrastructure.
[9] Or (in case of its creation) the interoperability platform

provided, preferentially by using a generally supported standardised library (updated for the new technology) to write and read data. In case of commercially available technology, this is done together with the EuBI industry board and the technology providing companies. Ways to efficiently represent technology specific metadata in the open data format are defined.

- The standards and ontologies used in EuBI are updated to include the concepts of the new technology. This is done for the concepts used in the WAP to ensure they fit to the established standards and databases as only a technology that can be efficiently and seamlessly accessed through the EuBI WAP can be considered integrated into the EuBI operation. To ensure consistency at the data level, the ontologies used in the EuBI Image Data Resource (IDR, previously called Image Data Repository) are updated as needed.
- Established technologies that can be connected for correlative approaches are identified. At the latest at the beginning of the new technology offer, the new and established technology providing nodes meet to define strategies how to efficiently combine their offers inside potential user projects.
- In the adaptation of the EuBI training offer to the new technology, the correct standards and ontologies are included.

6. Conclusion

To efficiently function as such a diverse infrastructure and to be truly more than the sum of its parts, EuBI will need interoperability not only with new technologies (the focus of this workpackage), but between all parts and between the biological and medical imaging fields. Concerning interoperability requirements, the EuBI infrastructure does not exist in a vacuum. Many other scientific infrastructures face similar challenges and existing considerations from them can also be applied to Euro-BioImaging. At the data level there is significant overlap with the shared efforts in the CORBEL and BioMedBridges[10] initiatives, and with the ELIXIR information infrastructure.

[10] BioMedBridges WP4 Technical integration

Annex 1:

Open data format examples for medical imaging:

The development of magnetic resonance imaging (MRI) techniques has defined modern Medical Imaging. Using techniques such as functional MRI and diffusion weighted imaging have allowed significant advances in research for non-invasive studies in the human body.
The sharing and reusing of data (within or between labs) is difficult if not impossible and complicates the application of automatic pipelines and quality assurance protocols.
To solve this problem, the Brain Imaging Data Structure (BIDS), have developed a standard for organizing and describing MRI datasets. The BIDS standard uses file formats compatible with existing software, unifies the majority of practices already common in the field, and captures the metadata necessary for most common data processing operations.
NIfTI is one of the file formats chosen because it is the largest common denominator across medical imaging software packages. in DICOM or other scanner specific files, the BIDS standard requires users to provide additional meta Information. The JSON file format (with the same filename as the .nii.gz file, but with a .json extension) is chosen for this proposal.

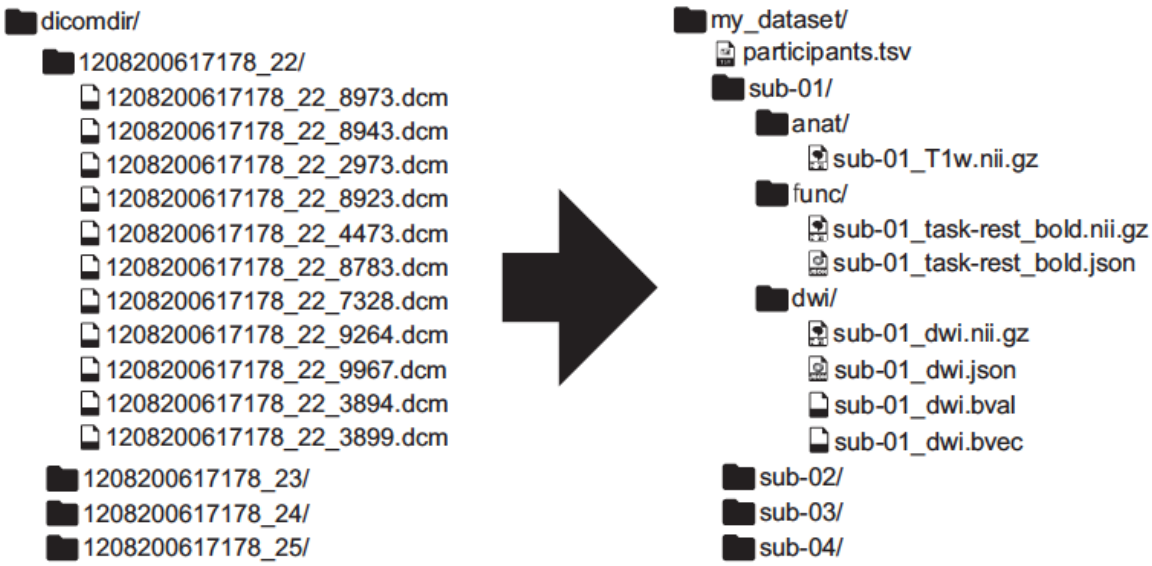

Figure 1: Illustration of a Brain Imaging Data structure (BIDS) structured data (1). BIDS is a format for standardizing and describing outputs of neuroimaging experiments (DICOM left) in a way that is intuitive to understand and easy to use with existing analysis tools (right). This structure is extensible to other kinds of studies and medical imaging modalities. (Figure extracted from [1])

[1] Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., ... & Handwerker, D. A. (2016). The brain imaging data structure, a format for organizing and

describing outputs of neuroimaging experiments. *Scientific Data*, *3*, 160044. (www.nature.com/scientificdata)