

Shanoir: Software as a Service Environment to Manage Population Imaging Research Repositories

Christian Barillot, Elise Bannier, Olivier Commowick, Isabelle Corouge,
Justine Guillaumont, Yao Yao, Michael Kain

► **To cite this version:**

Christian Barillot, Elise Bannier, Olivier Commowick, Isabelle Corouge, Justine Guillaumont, et al..
Shanoir: Software as a Service Environment to Manage Population Imaging Research Repositories.
MICCAI Workshop on Management and Processing of images for Population ImagiNG, Oct 2015,
Munich, Germany. pp.23-30. <inserm-01244551>

HAL Id: inserm-01244551

<http://www.hal.inserm.fr/inserm-01244551>

Submitted on 15 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Shanoir: Software as a Service Environment to Manage Population Imaging Research Repositories

Christian Barillot^{1,2,3,*}, Elise Banner¹⁻⁴, Olivier Commowick^{1,2,3}, Isabelle Corouge^{1,2,3},
Justine Guillaumont^{1,2,3}, Yao Yao^{1,2,3}, Michael Kain^{1,2,3}

1 Inria, VisAGeS Project-Team, F-35042 Rennes, France

2 INSERM, U746, F-35042 Rennes, France

3 University of Rennes I, CNRS, UMR 6074, IRISA, F-35042 Rennes, France

4 CHU Rennes, Department of Neuroradiology, F-35043 Rennes, France

Abstract. Some of the major concerns of researchers and clinicians involved in population imaging experiments are on one hand, to manage the huge quantity and diversity of produced data and, on the other hand, to be able to confront their experiments and the programs they develop with peers. In this context, we introduce Shanoir, a “Software as a Service” (SaaS) environment that offers cloud services for managing the information related to population imaging data production in the context of clinical neurosciences. We show how the produced images are accessible through the Shanoir Data Management System, and we describe some of the data repositories that are hosted and managed by the Shanoir environment in different contexts.

Keywords. Population imaging, database, data sharing, neuroinformatics, “Software as a Service” (SaaS), Cloud Computing, web application, Java, web services, Shared repositories, centralized resources

1 Introduction

Some of the major concerns of researchers and clinicians involved in population imaging experiments are, on one hand, to manage the huge quantity and diversity of produced data and, on the other hand, to be able to confront their experiments and the programs they develop with peers. In practice, researchers or clinicians in the neuroimaging domain are encouraged to set up large-scale experiments, but the lack of resources and capabilities to recruit locally subjects who meet specific inclusion criteria motivates the need for sharing the load in order to produce the relevant imaging data. For these reasons, making possible the pooling of experimental results, through the Internet and between collaborative centers, allows to recruit large subject populations and to widen the scientific achievement of the conducted experimental studies. Also, through distributed imaging databases, the search for similar results, the search for images containing singularities or transverse searches via data mining techniques could highlight possible regularities. Moreover, this will broaden also the possible panel of people involved in neuroimaging studies, while protecting the excellence of the supplied work.

In this context, the Shanoir[†] (SHaring NeuroImaging Resources) environment aims at establishing the conditions allowing, through the Internet, to share distributed information sources in neuroimaging, whether these sources are located in various centers of experi-

* Corresponding Author: Christian.Barillot@irisa.fr / Ph : +33 299847505 / Fax : +33 299847171

† Shanoir: <http://www.shanoir.org>

mentation, clinical departments of neurology, or research centers in cognitive neurosciences or image processing. This enables a large variety of users to diffuse, exchange or reach neuroimaging information with appropriate access means, in order to be able to retrieve information almost as easily as if the data were stored locally by means of the “cloud computing” Storage as a Service (SaaS) concept [1].

In this paper, we introduce the Shanoir software environment that offers services for managing the information related to neuroimaging data production in the context of clinical neurosciences. We show how the produced images are accessible through the Shanoir Data Management System. The paper is organized as follows. In section 2, we rapidly describe the software environment, and their extension for loading the data for querying the data, and for processing the data. Section 3 describes some population data repositories and section 4 provides a discussion on the use of these repositories and the potential evolutions.

2 Shanoir software environment

2.1 General description of the software environment

Shanoir is an open source software environment, with QPL licensing, designed to archive, structure, manage, visualize and share neuroimaging data with an emphasis on managing distributed collaborative research projects. It provides common features of neuroimaging data management systems along with research-oriented data organization and enhanced accessibility. Shanoir is based on a secured J2EE application running on a JBoss server, reachable via graphical interfaces in a browser or by third party programs via SOAP web services. It behaves as a repository of neuroimaging files coupled with a relational database holding meta-data (**Fig. 1**).

Shanoir uses semantics for concepts organization that are defined by ontology, called OntoNeuroLOG[‡] [2] [3]. OntoNeuroLOG reuses and extends the OntoNeuroBase ontology [4]. In Shanoir, the OWL-Lite implementation was manually derived from the OntoNeuroLOG initial expressive representation to Java classes. The data model based on this ontology is devoted to the neuroimaging field and is structured around research studies whereof involved patients have examinations, which either produce image acquisitions or clinical scores. Each image acquisition is composed of datasets represented by their acquisition parameters and image files. For security and regulation reasons, by default, the system only keeps anonymous data. Raw and derived (i.e. post-processed) image files can also be imported into the system from various sources (DICOM CDs, PACS, image files in NIFTI / Analyze format) using either online wizards, with completion of related metadata, command line tools or SOAP web services. For raw data, once de-identified during import, DICOM header's content is automatically extracted and inserted into the database by a customizable feature called “Study Card”.

Shanoir can also record any executed processing allowing retrieving workflows applied to a particular dataset along with the derived data. Clinical scores resulting from instrument assessments (e.g. neuropsychological tests) can be recorded and easily retrieved and exported in different formats (Excel, CSV, XML). Scores, image acquisitions and post-processed images are bound together, which makes relationship analysis possible. The instrument database is scalable and new measures can be added in order to meet specific project needs.

Using cross-data navigation and advanced search criteria, the user can quickly point to a subset of data to be downloaded. Client side applications have also been developed to illus-

[‡] OntoNeuroLOG: http://neurolog.i3s.unice.fr/public_namespace/ontology

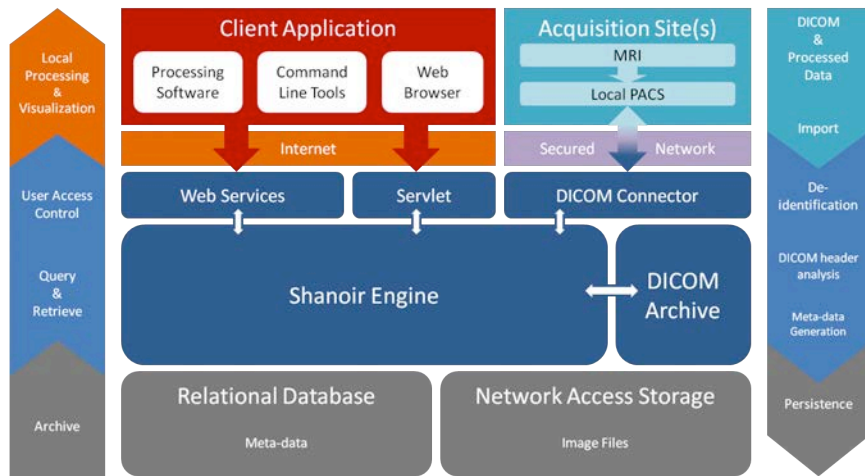


Fig. 1. Shanoir Software Architecture

trate how to locally access and exploit data through the available web services. With regards to security, the system requires authentication and user rights are tunable for each hosted studies. A study manager can thereby define the users allowed to see, download or import data into his/her study or simply make it public.

In practice, Shanoir serves neuroimaging researchers in organizing efficiently their studies while cooperating with other laboratories. By managing patient privacy, Shanoir allows the exploitation of clinical data in a research context. It is finally a handy solution to publish and share data with a broader community.

2.2 The Study Card and quality control concepts

Images can be imported in Shanoir from various sources: DICOM CDs, PACS (with DICOM Query & Retrieve), and image files (in NIfTI and Analyze format). Users are guided step by step through online forms to perform imports. In addition of archiving DICOM files, NIfTI copies are automatically generated and saved. This is convenient since the NIfTI format is better suited to local 3D image processing (such as registration, segmentation, statistical analysis, etc.) than the DICOM format.

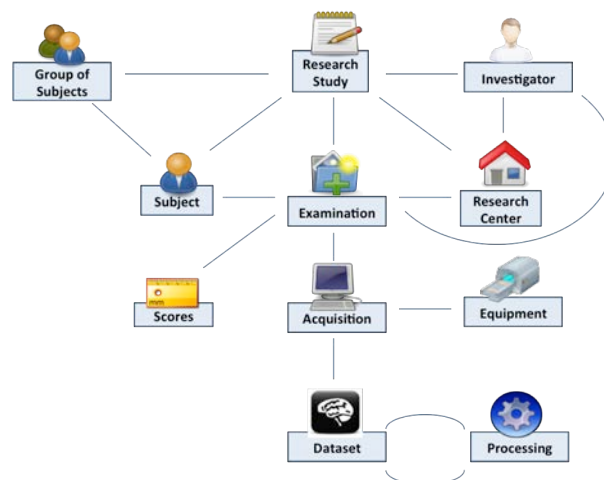


Fig. 2. Shanoir data organization

The Study Card concept

While being archived, the DICOM files are processed in two phases. The first phase de-identifies the images. The second one inserts into the database the metadata items generated from the DICOM header. This is achieved thanks to the “Study Card” concept. This concept allows the online meta-data wrapping between the local data to be imported (center, acquisition equipment...) and the semantic concepts of the research study the data will be assigned to. This allows the alignment between the actual DICOM metadata to the ontology, and also provides additional allocation of concepts to the stored images that is more related to the research study protocol (e.g. functional MRI, perfusion imaging, contrast agent, diffusion imaging...). The mechanism behind this feature is based on a user-predefined set of rules associating to particular acquisition equipment, and a particular data production site to the desired research study. Each rule determines the specific value of a metadata item according to the value(s) of one or more specific DICOM tag(s) (e.g. Series Description...). This greatly facilitates the consistent recording and alignment to the ontology of metadata for all the data of a research study without tedious workflow during the online import of images. Due to the simplicity of the process, no specific skills are pre-required to perform the import of data and it only takes a few minutes over the Internet. This “Study Card” concept also allows an automatic quality control of the data imported based on their metadata. For instance, a conformal statement can be attached to the imported data according to a match score to the Study Card rules.

2.3 The web portal

Shanoir provides a user-friendly secure web access and offers an intuitive workflow to ease the collection and retrieval of neuroimaging data from multiple sources (Fig. 3). On the home page, the user can access to the most frequent functionalities: Find and Download Datasets, Explore the Research Studies, Find Clinical Scores, and Import Data. On the top of all pages, the user always has a very complete navigation menu that leads to all services.

2.4 The interoperability

Interoperability is a very important concern for the Shanoir environment. For this purpose, Shanoir offers a web service interface that is open to all possible clients. This interface is already used by different external applications, developed either in C++, Java or Objective-C environments such as ShanoirUploader, medInria (<http://med.inria.fr>) and Shanoir.

SOAP for the integration of services.

Shanoir web services interfaces are based on the Simple Object Access Protocol (SOAP). Messages between client and server that are exchanged based on XML, with defined elements. As transport layer HTTP on base of TLS is used. The elements and services are described with the Web Service Description Language (WSDL). On base of this description client stubs can be automatically generated to simplify the connection of new clients. The web service layer is implemented with the Java API for XML web services.

“ShanoirUploader” for seamless integration of data.

“ShanoirUploader” is a Java desktop application that transfers data securely between a PACS and a Shanoir server instance (e.g. within a hospital). It uses a DICOM query/retrieve connection to search and download images from a local PACS. After retrieval, the DICOM files are locally anonymized (using either a built-in process or a custom one)

and then uploaded to the Shanoir server. The primary goals of that application are to enable mass data transfers between different remote server instances and therefore reduce the waiting time of the users, when importing data into Shanoir. Most of the time during import is spent with data transfers.

Apache Solr for metadata querying.

Shanoir integrates the enterprise search platform, Apache Solr (<http://lucene.apache.org/Solr/>), to provide the users a vast array of advanced features such as near real-time indexing and queries, full-text search, faceted navigation, autosuggestion and autocomplete. One of the most important features of Solr search is the faceted navigation. Facets correspond to properties of the Solr information elements. They are derived by analysis of the pre-existing meta-data that are related to the ontology model used by Shanoir. All the metadata are indexed in a JBoss server that hosts the Solr servlets. A custom security post-filter has been also developed and implemented in Shanoir for user access control. This filter retrieves user identification and access rights in Shanoir and interacts with the Solr server to show pertinent results that the user is allowed to access.

MedInria for image processing.

Shanoir web services may also be queried from standalone C++/Qt applications through the QtShanoir library (<http://qtshanoir.gforge.inria.fr>). QtShanoir uses the SOAP-based web services provided by a Shanoir server to get and display studies, patients, and data with their associated metadata. In QtShanoir, a set of Qt widgets are defined that can be embedded in any Qt application. This library was used to implement a Shanoir query plugin inside the medInria visualization and processing software. This implementation allows for the interrogation and the download of image data from Shanoir, to process it within medInria using the available processing tools and then upload back the processing results to the Shanoir server with the correct metadata values (**Erreur ! Source du renvoi introuvable.**).

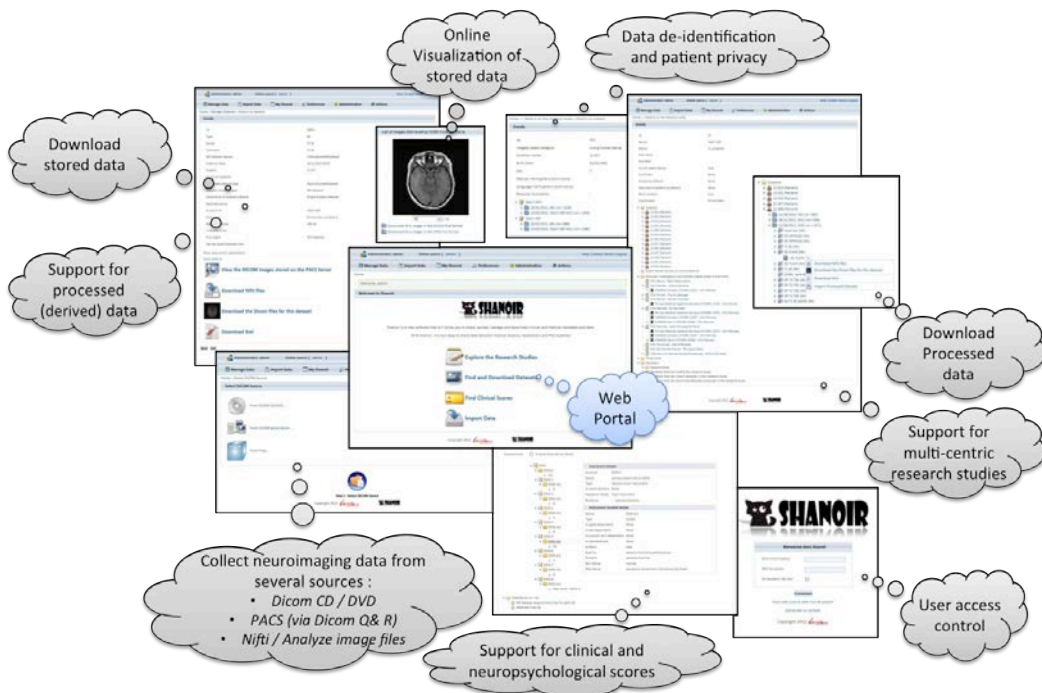


Fig. 3: Shanoir web portal summary of the main functionalities

3 Data repositories

Each Shanoir repository has an administrator that manages the access rights of the repository. Then, each user has to request an account through the dedicated web form and specify which study he/she wants to access, who is his/her contact, what will be his/her role concerning this study, and what level of expertise/access is needed (guest, user, expert, admin)... According to the information provided, the Shanoir administrator of the repository grants (or not) the user access to the system. The access to a specific study is granted by the person responsible of this study (i.e. the PI of the research study or its official representative). Depending on these settings, the new user will be able to see, download, and import datasets or even to modify the study parameters. The corresponding rights are set for a limited time and must go through a renewal process on purpose. If requested, the user can receive a report by email each time data are imported in his/her study.

3.1 The Shanoir@Neurinfo Repository

Started in 2009, the Neurinfo research facility[§] promotes translational clinical research and supports the development of clinical research, technological and methodological activities. It offers resources for in vivo human imaging acquisition, image data analysis and image data management. A large community of users, both clinicians and scientists, uses these resources as part of local, national or international imaging based research projects.

All the data produced at Neurinfo for academic or clinical research purposes are managed through a dedicated Shanoir@Neurinfo repository (**Fig. 4**) administered by the facility technical staff. The Shanoir@Neurinfo server also hosts data from multi-sites imaging studies. In total, more than 1To data from 31 centers and 37 scanners are archived within this repository (see Left Table in **Fig. 4** for details).

On the daily practice, DICOM data are imported by a technician from either a local PACS, a CD/DVD or a disk drive containing the DICOMDIR at its root and the DICOM files. The clinical studies conducted at Neurinfo concern the whole-body (brain, spine, heart, lung, pelvis, vasculature...) with a major focus on brain anatomy and function in normal control and pathological populations. Out of the 60 or so ongoing research studies at the Neurinfo platform, 75% relates to brain imaging, 15% to abdominal imaging and

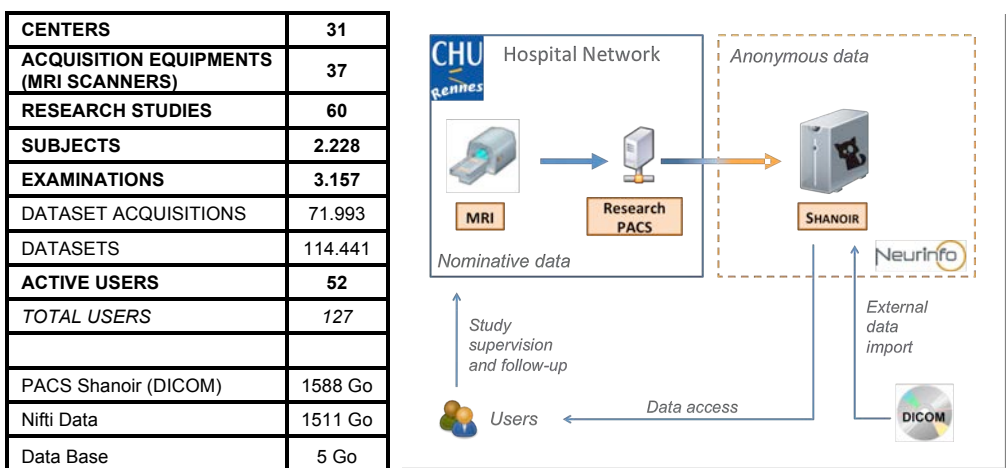


Fig. 4. The Shanoir@Neurinfo repository: Current Global Statistics (*left*) and Service Infrastructure (*right*)

[§] <http://www.neurinfo.org/>

10% to heart imaging. Among the neuro-imaging clinical studies, multiple sclerosis, dementia, tumors, stroke and mood disorders are the most investigated pathologies.

The general policy for the Shanoir@Neurinfo repository for dissemination of the data related to one study is decided upon beforehand with the principal investigator, in agreement with the informed consent form approved by the ethics committee and signed by the participant. Any opening of the data to third parties is submitted to the approval of the principal investigator prior to allow the (complete or partial) access to a third party user. Nonetheless, to ensure dissemination and the best usage of data acquired from public funding, the Neurinfo team strongly encourages investigators to share their data, which is usually agreed after an embargo period.

3.2 The Shanoir@OFSEP Repository

The OFSEP project** was selected in response to the national call for projects “Cohorts 2010” as part of the “Investments for the Future” program. This is a collaborative project involving over 40 French expert MS centers. The aim of this project is to build and maintain a nationwide cohort of patients with Multiple Sclerosis (MS), and enrich the clinical data with biological samples, socio-economic data and neuro-images.

A dedicated imaging working group is in charge of acquiring, processing, integrating imaging and derived imaging data into a shared Imaging Resource Centre (IRC), and make this IRC inter-operate with the clinical databases. The consistent assessment of MRI-based measurements at a large scale require robust and efficient image processing pipelines. A further goal of this project is to establish an information technology (IT) infrastructure enabling audited access to imaging data, as well as a “virtual laboratory” environment supporting the distributed, synergistic development, validation, and deployment of specialized image analysis procedures, developed by different national and international research centers. To ensure an easy access to the imaging data and allow modifications, queries, annotations and access control, the Shanoir environment has been selected. It also ensures interoperability and data management related to the imaging part of this cohort (the clinical part is managed by the EDMUS†† system).

Started in 2012, the Shanoir@OFSEP server has been installed to store the imaging data of the OFSEP cohort. This cohort aims at studying neuroimaging data of 40.000 MS patients over the next 10 years. A consensus has emerged concerning the acquisition protocol that requires: a brain MRI every 3 years, a spinal-MRI every 6 years, that is to say 200.000 MRI over 10 years. Shanoir@OFSEP database will grow during this period and beyond [5].

Since OFSEP is a nationwide project gathering many patients, many IRC and much different MRI equipment, a federated repository with nationwide access and with thorough homogenization mechanism was therefore needed. The OFSEP imaging WG is continuously gathering new acquisition centers volunteering to take part to the cohort. In Shanoir@OFSEP, there are currently about 30 IRCs pooling 40 MRI acquisition equipment representing 12 MR scanner models from 3 MR constructors (Siemens, Philips, GE). All the centers are importing data in one main study called the “Mother Cohort”. If necessary, derived imaging data will be then imported back to the server in order to refer to potential post-processing information, MS specific imaging biomarkers making them available for others authorized users.

Currently Shanoir@OFSEP repository is hosting 5 studies: the “Mother Cohort” (200.000 MRI planned over the next 10 years) as well as 4 MS imaging clinical research projects. More of these “OFSEP-labeled” clinical research projects or nested cohorts will

** The OFSEP MS Cohort observatory: <http://www.ofsep.org/fr/l-observatoire/presentation-ofsep>

†† EDMUS: <http://www.edmus.org>

be integrated in the following years. Everyone can join the “Mother Cohort” study as long as they use the OFSEP protocol. One can also ask the OFSEP to contribute to the project through his study as soon as the principal investigator presents his research study subject to the OFSEP scientific committee that can grant (or not) the hosting. Data hosted on Shanoir@OFSEP, will remain confidential (private) throughout the duration of the study, but can be made available to all researchers through a specific application to OFSEP.

4 Conclusion and perspectives

The Shanoir Software as a Service environment has been presented. We have shown how this system manages to share distributed information sources in neuroimaging over the Internet, whether these resources are located in various centers of experimentation, clinical departments in neurology, or research centers in cognitive neurosciences or image processing. Through the description of two repositories that administrate a Shanoir environment (Neurinfo and OFSEP), we have illustrated how a large variety of users can diffuse, share or access neuroimaging information between peers almost as easily as if the data were stored on their local hospital, research labs or companies. Through the description of the Shanoir software environment, we have illustrated how neuroimaging data can be structured, managed, archived, visualized and shared with examples on multi-institutional, collaborative research projects.

5 Acknowledgements

This work has been supported by the “technological development program” of Inria, by the Brittany region council and the EU-Feder program for the Neurinfo platform, and by two grants provided by the French Government and handled by the “Agence Nationale de la Recherche,” within the framework of the “Investments for the Future” program, under the references ANR-10-COHO-002 (for OFSEP) and ANR-11-INBS-006 (for FLI).

6 REFERENCES

1. Rimal, B.P., E. Choi, and I. Lumb. *A taxonomy and survey of cloud computing systems*. in *INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on*. 2009. Ieee.
2. Michel, F., et al. *Grid-wide neuroimaging data federation in the context of the NeuroLOG project*. in *Proceedings of the HealthGrid*. 2010.
3. Temal, L., et al., *Towards an ontology for sharing medical images and regions of interest in neuroimaging*. *J Biomed Inform*, 2008. **41**(5): p. 766-78.
4. Barillot, C., et al., *Federating Distributed and Heterogeneous Information Sources in Neuroimaging: The NeuroBase Project*. *Stud Health Technol Inform*, 2006. **120**: p. 3-13.
5. Cotton, F., et al., *OFSEP, a Nationwide Cohort of People with Multiple Sclerosis: Consensus minimal MRI protocol*. *Journal of Neuroradiology*, 2015. (**in Press**).