Shanoir: Applying the Software as a Service Distribution Model to Manage Brain Imaging Research Repositories

Christian Barillot, Elise Bannier, Olivier Commowick, Isabelle Corouge, Anthony Baire, Ines Fackfack, Justine Guillaumont, Yao Yao, Michael Kain

▶ **To cite this version:**

Shanoir: Applying the Software as a Service Distribution Model to Manage Brain Imaging Research Repositories

Christian Barillot[1,2,3,*], Elise Bannier[1-4], Olivier Commowick[1,2,3], Isabelle Corouge[1,2,3], Anthony Baire[1], Ines Fachfach[1,2,3], Justine Guillaumont[1,2,3], Yao Yao[1,2,3team], Michael Kain[1,2,3]

1 Inria, F-35042 Rennes, France
2 INSERM, Visages 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

Two of the major concerns of researchers and clinicians performing neuroimaging experiments are managing the huge quantity and diversity of data and the ability to compare their experiments and the programs they develop with those of their peers. In this context, we introduce Shanoir, which uses a type of cloud computing known as software as a service (SaaS) to manage neuroimaging data used in the clinical neurosciences. Thanks to a formal model of medical imaging data (an ontology), Shanoir provides an open source neuroinformatics environment designed to structure, manage, archive, visualize and share neuroimaging data with an emphasis on managing multi-institutional, collaborative research projects. This article covers how images are accessed through the Shanoir Data Management System and describes the data repositories that are hosted and managed by the Shanoir environment in different contexts.

Keywords

Neuroimaging, database, data sharing, neuroinformatics, software as a service (SaaS), cloud computing, web application, Java, web services, shared repositories, centralized resources

Abbreviations

DICOM	Digital Imaging and Communications in Medicine
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
IRC	Imaging resource center
J2EE	Java Platform Enterprise Edition
JAX-WS	Java API for XML web services
JWS	Java Web Start
NIfTI	Neuroimaging Informatics Technology Initiative
OFSEP	Observatoire Français de la Sclérose en Plaques (French Multiple Sclerosis Observatory)
OWL	Web Ontology Language
PACS	Picture archiving and communication system
PI	Principal investigator
SaaS	Software as a service
Shanoir	Sharing Neuroimaging Resources
SOAP	Simple Object Access Protocol
WSDL	Web Service Description Language

[*]Corresponding author: Christian.Barillot@irisa.fr / Phone: +33 2 9984 7100 / Fax: +33 2 9984 7171

1 Introduction

1.1 Context

Two of the major concerns for researchers and clinicians performing neuroimaging experiments are managing the huge quantity and diversity of data and the ability to compare their experiments and the programs they develop with those of their peers. In practice, researchers and clinicians in the neuroimaging field are encouraged to set up large-scale experiments, but the inability to recruit sufficient local subjects who meet specific criteria results in the need for cooperation to gather the relevant imaging data. Pooling experimental results via the Internet and cooperative efforts by centers provide larger and more specific subject populations that expand the scope and value of scientific research.

Searches on distributed neuroimaging databases for similar results and images containing singularities (quirk, peculiarities, …) or the use of data mining techniques may highlight possible similarities. Such efforts also broaden the possible panel of people involved in neuroimaging studies while maintaining the quality of the work. Indeed, the explosion of data generated by the neurosciences community in the early 90s has resulted in the need for innovative techniques for data and knowledge sharing and reuse (Mazziotta et al., 1995; Roland and Zilles, 1994; Shepherd et al., 1998). This has led to the emergence of large-scale projects on the human brain. A recent objective added to these initial issues is the application of data analysis and data processing software to various data repository systems for knowledge discovery and data mining, including its more recent extension to merging imaging and genetic data (Hibar et al., 2015). In parallel the development of web applications has stimulated the interest of researchers and clinicians in distributed databases and information sharing.

1.2 Background

It is now commonly accepted in the neuroimaging community that sharing data and image processing services will play a crucial role in translational research (Barillot et al., 2003; Keator et al., 2013; Poldrack and Gorgolewski, 2014; Poline et al., 2012; Van Horn and Gazzaniga, 2013; Walport and Brest, 2011). Research funding agencies now clearly identify the sharing of scientific resources (data processing) as a top priority. International organizations such as the International Neuroinformatics Coordinating Facility (INCF[†]) are now dedicated to promoting the field of neuroinformatics (Book et al., 2013; Kennedy et al., 2015). Sharing data and image processing services for translational research is needed for:

1) the integration of large data sets for population-wide studies and *construction of imaging cohorts* (Barillot et al., 2006; Evans and Brain Development Cooperative, 2006; Hall et al., 2012; Jack et al., 2008; Marcus et al., 2013; Shepherd et al., 1998; Van Essen et al., 2013; Van Horn et al., 2001; Weiner et al., 2012),
2) the validation of image processing tools on reference datasets for *validation and quality control of image processing procedures* (Menze et al., 2015; Styner et al., 2008),
3) the reuse of image processing pipeline on different sets of data and different peers for *sharing processing tools* (Bellec et al., 2012; Dinov et al., 2010; Glatard et al., 2015; Gorgolewski et al., 2011; Keator et al., 2013; Keator et al., 2009; Ooi et al., 2009), and
4) the validation of research results based on proofed control statistical analysis of images for *validation and quality control of experimental research* (Button et al., 2013; Carp, 2012; Ioannidis, 2014; Ioannidis et al., 2014).

This is particularly significant in the field of neuroimaging as several large recent multicenter initiatives have shown. These include Evans et al. (Evans and Brain Development Cooperative, 2006), which performed a study using magnetic resonance imaging (MRI) of normal brain maturation from birth to adulthood in approximately 500 children with behavior disorders, and the Alzheimer's Disease Neuroimaging Initiative (ADNI), which has assembled a very large variety of images for its work (Weiner et al., 2012). The Human Connectome Project (HCP), which worked with 1,200 healthy volunteers to investigate brain connectivity in the normal brain (Marcus et al., 2013; Van Essen et al., 2013), is another well-known example illustrating the importance of aggregate imaging data and relating data warehouses to image processing resources.

[†] http://www.incf.org/

To provide archiving solutions for large or various multicenter projects, several architectures have already been proposed. The Biomedical Informatics Research Network (BIRN) has been a pioneer in launching brain imaging solutions (Ashish et al., 2010; Gupta et al., 2003; Keator et al., 2008; Keator et al., 2009). Another early initiative, the FMRIDC project sought to share task-based fMRI imaging data (Van Horn and Gazzaniga, 2013). @NeurIST set up a dedicated solution (funded by an Integrated European Project) to support research and treatment of cerebral aneurysms using heterogeneous data, computing and complex processing services (Benkner et al., 2010). The LORIS/CBRAIN project is an initiative to develop a pan-Canadian platform for distributed processing, analysis, exchange and visualization of brain imaging data (Das et al., 2011; Sherif et al., 2014). Finally, other generic data management systems have been proposed to offer shared solutions for managing multicenter studies. These include the Extensible Neuroimaging Archive Toolkit (XNAT) (Marcus et al., 2007), which has been successful due to its integration in the management of large projects (Marcus et al., 2013) and ability to communicate with data management servers via dedicated REST web services, and the Collaborative Informatics and Neuroimaging Suite (COINS), which provides a web-based neuroimaging and neuropsychology software suite (Scott et al., 2011). Though, the extensibility of these platforms is part of the motivations, none of them are built on top of a formal semantic model or ontology that can guarantee the sustainability of any evolution of the original da-ta scheme.

1.3   Significance

In this context, the Sharing Neuroimaging Resources (Shanoir) environment enables sharing between distributed sources of neuroimaging information over the Internet, whether the sources are located in various centers of experimentation, clinical departments of neurology, or research centers in cognitive neurosciences or image processing. A large variety of users can thus share, exchange and have controlled access to neuroimaging information using the software as a service (SaaS) type of cloud computing (Rimal et al., 2009) almost as easily as if the data were stored locally.

In this paper, we introduce the Shanoir software environment for managing neuroimaging data production in the context of clinical neurosciences and show how the images are accessible through the Shanoir Data Management System. Shanoir is an open source neuroinformatics environment designed to structure, manage, archive, visualize and share neuroimaging data with an emphasis on multi-institutional, collaborative research projects. The software offers features commonly found in neuroimaging data management systems along with research-oriented data organization capabilities and enhanced accessibility. It also provides user-friendly secure web access and an intuitive workflow that facilitates the collection and retrieval of neuroimaging data from multiple sources.

In Section 2, we provide a brief overview of the software environment including its core (web portal, Study Card and quality control) and extensions for loading, querying and processing data. Section 3 describes the data repositories while Section 4 covers the use of these repositories and potential evolution.

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 collaborative research projects. It includes the common features of neuroimaging data management systems along with research-oriented data organization and enhanced accessibility. Shanoir is based on a secure J2EE application running on a JBoss server which is accessed via graphical interfaces in a browser or by third-party programs via web services using Simple Object Access Protocol (SOAP). It behaves like a repository of neuroimaging files coupled with a relational database containing metadata (Figure 1).

Shanoir uses semantics for structuring the concepts as defined by the OntoNeuroLOG‡ ontology (Michel et al., 2010) (Temal et al., 2008). OntoNeuroLOG reuses and extends the OntoNeuroBase ontology defined earlier (Barillot et al., 2006) (see Figure 2). Both were designed using the

---

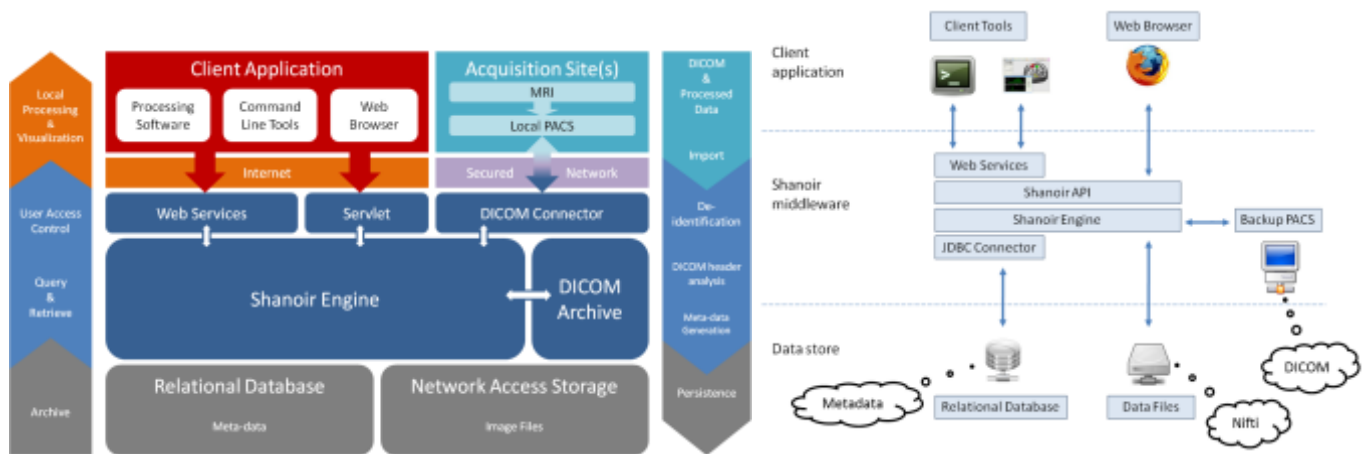‡ OntoNeuroLOG: http://neurolog.i3s.unice.fr/public_namespace/ontology

Figure 1: Shanoir Software Architecture

methodological framework (Temal et al., 2008) of the foundational Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (Masolo et al., 2003) and a number of core ontologies that provide generic, basic and minimal concepts and relationships in specific fields such as artefacts, participant roles, information and discourse acts. 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 dedicated to the neuroimaging field and is structured around research studies in which patients are examined to produce image acquisitions or clinical scores. Each image acquisition is composed of datasets represented by acquisition parameters and image files. For security and legal reasons, all data on the system is anonymous by default, this can be customized with specific algorithm (e.g. defacing is not currently implemented but can easily be embedded in a specific anonymizer that Shanoir will call).

Raw as well as derived (i.e., post-processed) image files can also be imported into the system using medical imaging technology (e.g., media based on the Digital Imaging and Communications in Medicine (DICOM) standard, picture archiving and communication system (PACS) or image files in the Neuroimaging Informatics Technology Initiative (NIfTI)/Analyze-style data format) using online
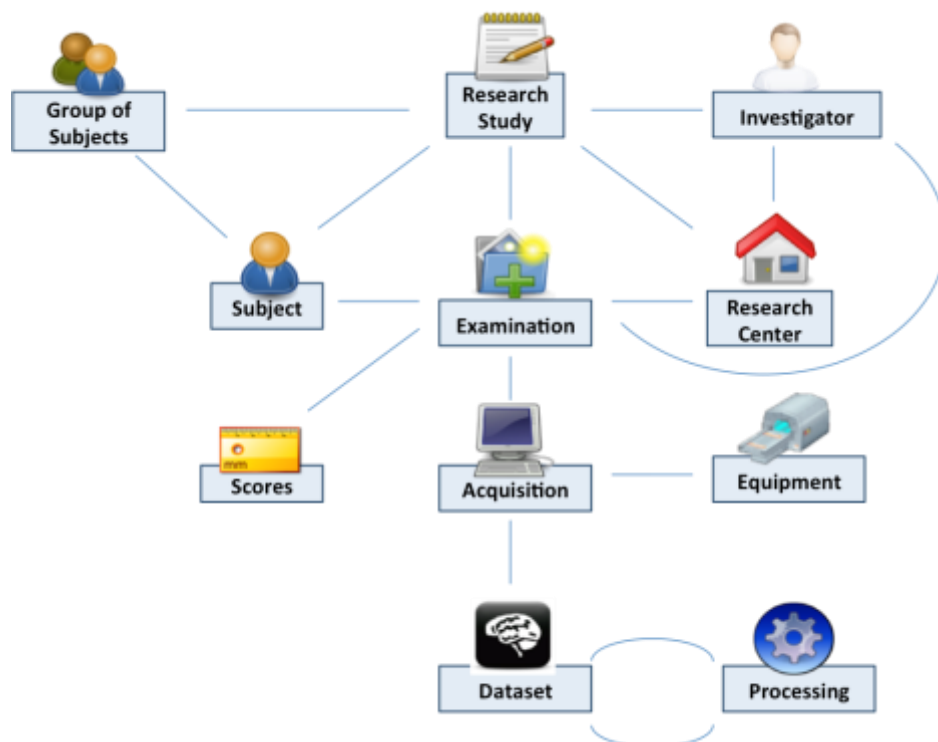

Figure 2: Shanoir data organization

wizards, which complete related metadata, command line tools or SOAP web services. Once identity information has been removed from raw data during the importation process, the DICOM header content is automatically extracted, enriched and inserted into the database with the customizable "Study Card" feature. Shanoir can also record any execution process for retrieval of workflows applied to a particular dataset along with the derived data.

Clinical scores from instrument assessments (e.g., neuropsychological tests) can be recorded and easily retrieved and exported in different formats (Excel, CSV, XML). The instrument database is scalable and new measures can be added in order to meet specific project needs (Figure 3). Scores, image acquisitions and post-processed images are bound together so that relationships can be analyzed. Using cross-data navigation and advanced search criteria, the user can quickly indicate a subset of data for download. Client-side applications have also been developed to locally access and exploit data though web services. The security features of the system require authentication with user rights set for each study. A study manager can 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 by efficiently organizing their studies while cooperating with other laboratories. By managing patient privacy, Shanoir offers the possibility of using clinical data in a research context. Lastly, it is a handy solution for publishing and sharing data with a broader community.

2.2 Study Card and quality control concepts

Images can be imported in Shanoir from various sources: DICOM media, PACS (with DICOM Query & Retrieve), and 3D/4D image files (in NIfTI/Analyze format). Users are guided step-by-step through online forms to perform imports. In addition to archiving DICOM files, NIfTI copies are automatically generated and saved. This is convenient since the NIfTI format is better suited to perform image processing (such as registration, segmentation, statistical analysis, etc.) than the DICOM format.

Figure 3: Shanoir "instrument" database can be used for attaching clinical scores to images (e.g., EDSS score in MS). An instrument can be any record were an alphanumerical value can be attached.
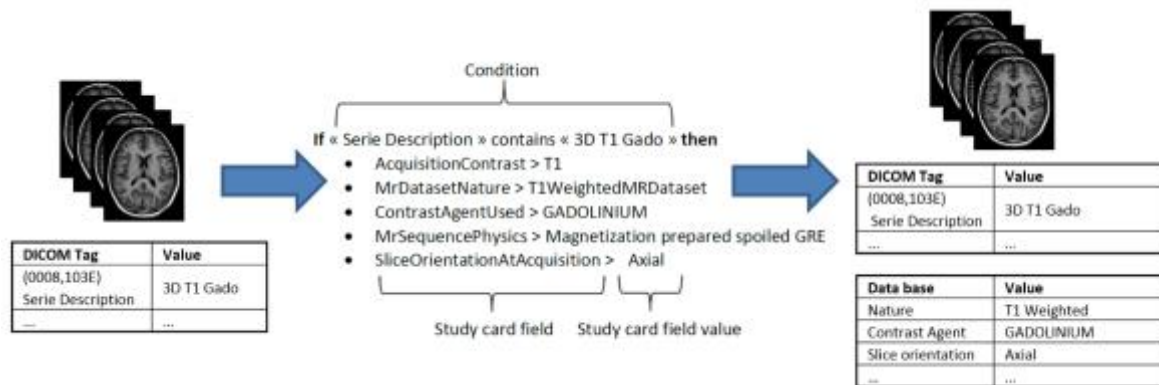
Figure 4: Example of a "StudyCardRule" for a 3D T1-Gd sequence

2.2.1 The Study Card

During archiving, the DICOM files are processed in two phases. The first phase de-identifies the images. The second phase populates the database with the new metadata items generated from the DICOM header and enriched with the Study Card, which enables online metadata wrapping between the local data to be imported (center, acquisition equipment, etc.) and the semantic concepts of the research study to which the data will be assigned. The actual DICOM metadata can thus be aligned with the ontology, and also provides additional allocation of concepts to the stored images that are more closely 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 set of rules that the user predefines to associate specific acquisition equipment and a specific 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, see Figure 4). This greatly facilitates the consistent recording and alignment to the ontology of metadata for all data in a research study without the need for tedious workflow during the online import of images. Due to the simplicity of the process, no specific skills are required to perform data import and it only takes a few minutes over the Internet. The "Study Card" concept makes possible an automatic quality control of the imported data using their metadata. For instance, a conformal statement can be attached to the imported data according to a match score to the Study Card rules.

2.2.2 Quality Assessment

Shanoir's next major functionality concerns the quality check of the images for conformity of the imported data with the predefined study protocol and ensure the integrity of the archived data. We have identified three levels of control:

- Study Protocol: controls the time interval between examinations (expected visits) as defined by the principal investigator (PI) of the study;
- Acquisition Protocol: controls the presence of all the sequences of the imaging protocol as defined by the PI of the study;
- Raw data:
  - The software automatically checks the range of parameters for a given protocol, experimental center or acquisition scanner as defined in the Study Card by the PI's technical representative;
  - Visual inspection of the image quality and integrity can be reported and assigned to the imported data; however, the mechanism to detect the visual quality is not yet integrated in the Shanoir environment.

In the next release of Shanoir, quality assessment will be present as flags (flawless, acceptable or inadmissible) in the database.

This QA capability do not address the control of image formation as for instance to check image artefacts (bias, motions, ghosting, etc…). This category of QA can be implemented in a dedicated image visualization and processing tools that interoperate with Shanoir through the dedicated web services.
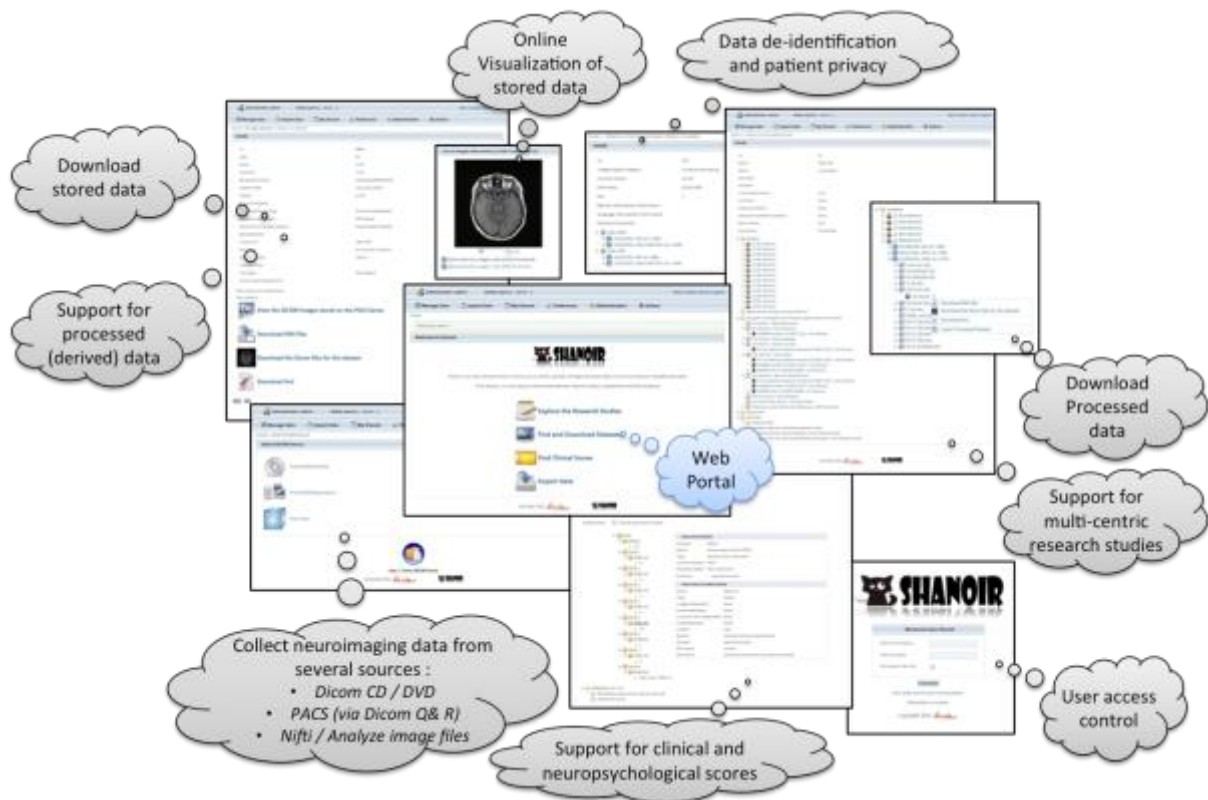
Figure 5: Shanoir web portal summary of the main functionalities

2.3 Web portal

Shanoir provides user-friendly secure web access and offers an intuitive workflow to facilitate the collection and retrieval of neuroimaging data from multiple sources (Figure 5). On the home page, the user has direct access to the most frequent functionalities: Find and Download Datasets, Explore the Research Studies, Find Clinical Scores, and Import Data (Figure 6). On the top of all pages, the user always has a very complete navigation menu that leads to all services.

2.4 Interoperability

Interoperability is a very important concern for the Shanoir environment. Shanoir offers web services interface that is open to a large variety of clients. We already offer several dedicated interface that are already in used by different external applications. Hereafter, we described four of these external services that are currently available and run independently to each other: ShanoirUploader, QtShanoir, medInria and iShanoir developed either in C++, Java or Objective-C environments.
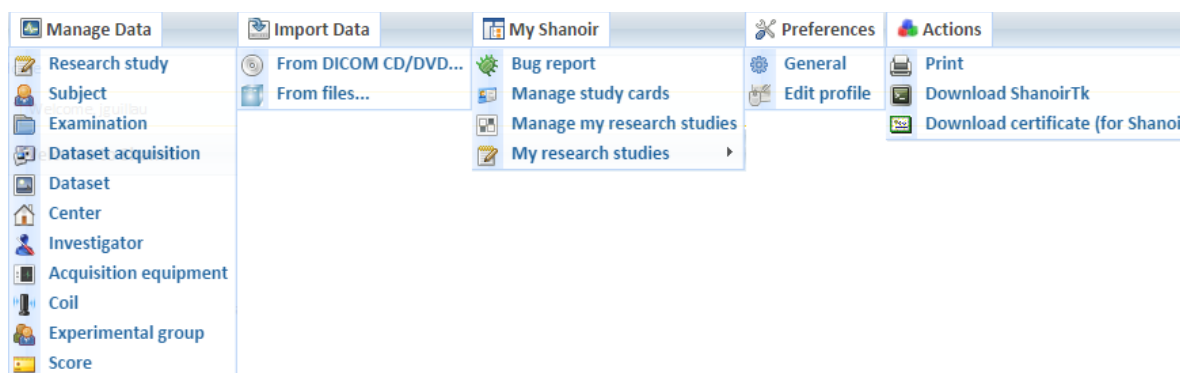

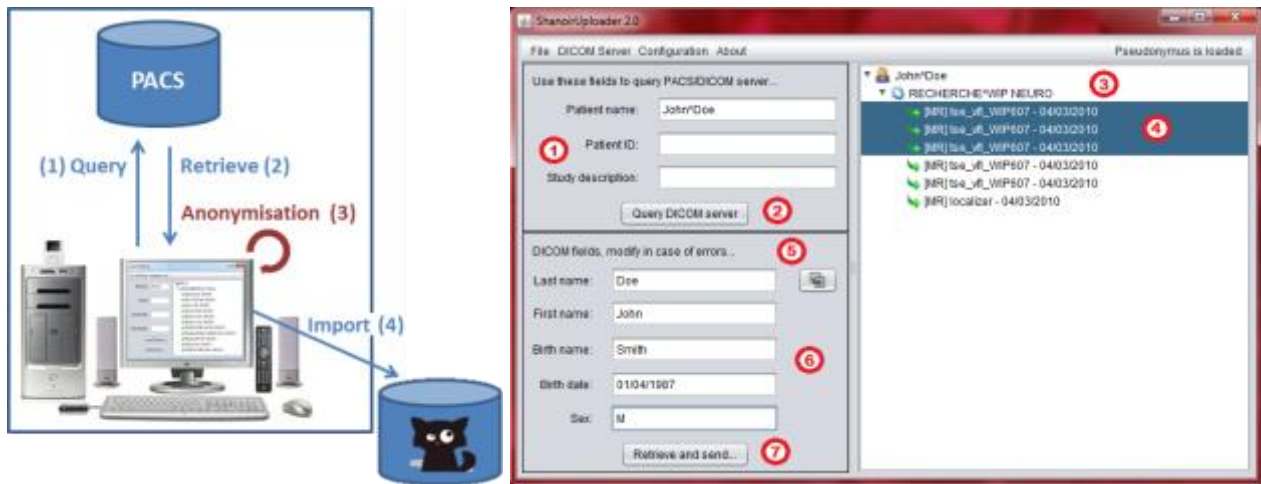**Figure 6:** Shanoir menu organization (on top of all web pages)

Figure 7: Shanoir Uploader architecture for secure transfer of local PACS data to a Shanoir Server (*left*) and user interface (*right*)

2.4.1 SOAP for integration of services

The Shanoir web services interface is based on the Simple Object Access Protocol (SOAP). Messages between clients and the server are exchanged using Extensible Markup Language (XML) with well-defined elements. The Hypertext Transfer Protocol (HTTP) is used with Transport Layer Security (TLS). Elements and services are described with the Web Service Description Language (WSDL). Based on 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 (JAX-WS). Shanoir offers numerous dedicated web services:

- "EntityCreator": creates new entities, such as creating a new subject in the database
- "CredentialTester": validates if username and password are correct
- "Downloader": downloading files/datasets on base of dataset IDs
- "CenterFinder": find center(s) based on different search criteria, i.e. study or investigator
- "DatasetAcquisitionFinder": find acquisition(s) based on IDs or examinations
- "DatasetFinder": find dataset(s) based on multiple search criteria/filters
- "DatasetProcessingFinder": find dataset processing(s) based on IDs
- "ExaminationFinder": find examination(s) based on multiple search criteria/filters
- "ExperimentalGroupOfSubjectsFinder": find group of subjects based on multiple criteria
- "InvestigatorFinder": find investigator(s) based on IDs or centers
- "MrDatasetFinder": find MR dataset(s) based on multiple search criteria/filters
- "StudyFinder": find study/-ies based on multiple search criteria/filters
- "SubjectFinder": find subject(s) based on IDs with multiple filters
- "DatasetImporter": import dataset files to already existing entities in the database
- "ReferenceLister": shows list of reference strings stored in the database
- "FileUploader": upload files in local archive for later import, used by ShanoirUploader

2.4.2 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 an hospital). It offers both a direct DICOM query/retrieve connection to search and download images from a local PACS and a DICOM CD upload facility. After retrieval, the DICOM files are locally anonymized and then uploaded to the Shanoir server (the anonymization algorithm can be customized according to specific operational/regulation constraints). The primary goals of the application are to enable mass data transfers between different remote server instances and reduce user waiting time when importing data into Shanoir. Most of the import time thus involves data transfer.

"ShanoirUploader" requires a local Java installation. For a simpler distribution and installation of the software, Java Web Start (JWS) can be used. The application can be installed with a simple web link that is opened in a web browser. Java takes care of the installation and, later, of automatic updates. Internal components are based on Java Swing for the graphical user interface (Figure 7), *dcm4chee*[§] libraries to connect with a PACS and Java, and WebServices (JAX-WS) to transfer data to a Shanoir server.

2.4.3   Apache Solr for metadata querying

Shanoir integrates the open source enterprise search platform Apache Solr (http://lucene.apache.org/solr/), which provides users with a vast array of advanced features such as near real-time indexing and queries, full-text searching, faceted navigation, autosuggestion and autocomplete.

One of the most important features of the Solr search is the faceted navigation. Facets correspond to properties of the Solr information elements and are derived by analyzing pre-existing metadata that are related to the ontology model used by Shanoir.

Shanoir users can access all metadata with a simple Solr search bar. After entering at least one character, a user will be automatically guided to complete his search. Data are sorted by categories and dynamically displayed once a facet is chosen. By clicking on Solr data results, users access all the additional information available in Shanoir corresponding to their search, and then use these queries for local downloading (Figure 8).

Figure 8: Example of Apache Solr search of the Shanoir server
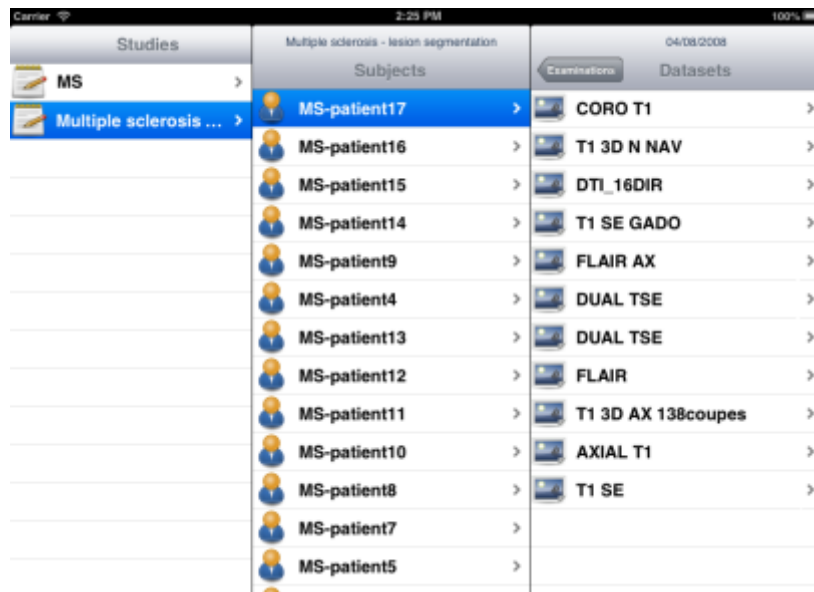

[§] http://www.dcm4che.org

Figure 9: Example of storyboard interfaces under the iShanoir iOS mobile application connected to a Shanoir server

All metadata are indexed in a JBoss server that hosts the Solr servlets. A custom security post-filter has also been developed and implemented in Shanoir to control user access. This filter retrieves user identification and access rights in Shanoir and interacts with the Solr server to show relevant results that the user is allowed to access.

2.4.4   iShanoir for mobile data access

An iOS application, iShanoir, has been developed for iPhones and iPads. It opens a secure connection with a Shanoir server and enables the user to access data stored on a Shanoir server. With iShanoir, the user can navigate within the Shanoir data tree structure on the server. After data is selected from the mobile app, the images can be downloaded to the local device, displayed and analyzed with any local DICOM viewer or through cloud services (i.e. Dropbox, iCloud, Google or OneDrive).

The iShanoir application has been developed with Xcode and implemented in Objective-C. For the graphical user interface, two storyboards have been developed to fit the different display sizes between iPhones and iPads (Figure 9). It uses the following iOS frameworks: Foundation, CoreFoundation, UIKit and CFNetwork. For implementation of the SOAP web services client, the WSDL2ObjC utility has been used as it offers a client stub code generation based on the server WSDL document.

2.4.5   Qt Shanoir for image processing

Shanoir web services may also be queried from standalone C++/Qt applications through the QtShanoir library[**], which uses SOAP web services provided by a Shanoir server to access 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. The library was used to implement a Shanoir query plugin inside the medInria visualization and processing software[††] for interrogation and downloading of image data from Shanoir for processing within medInria for example using the available processing tools and then upload the processing results back to the Shanoir server with the correct metadata values (Figure 10).

2.5   Distribution of Shanoir

The Shanoir server can be freely downloaded on request. It is currently deployed using Docker containers running on a Linux kernel. Linux containers are implemented using namespaces for locating each type of resource. Dockers are tools for managing lightweight method of virtualization (named containers) on Linux that are lighter than traditional virtual machines. The host and guest systems share the same kernel. The kernel is responsible for host↔guest and guest↔guest isolation (the result of

---

[**] QtShanoir library: http://qtshanoir.gforge.inria.fr
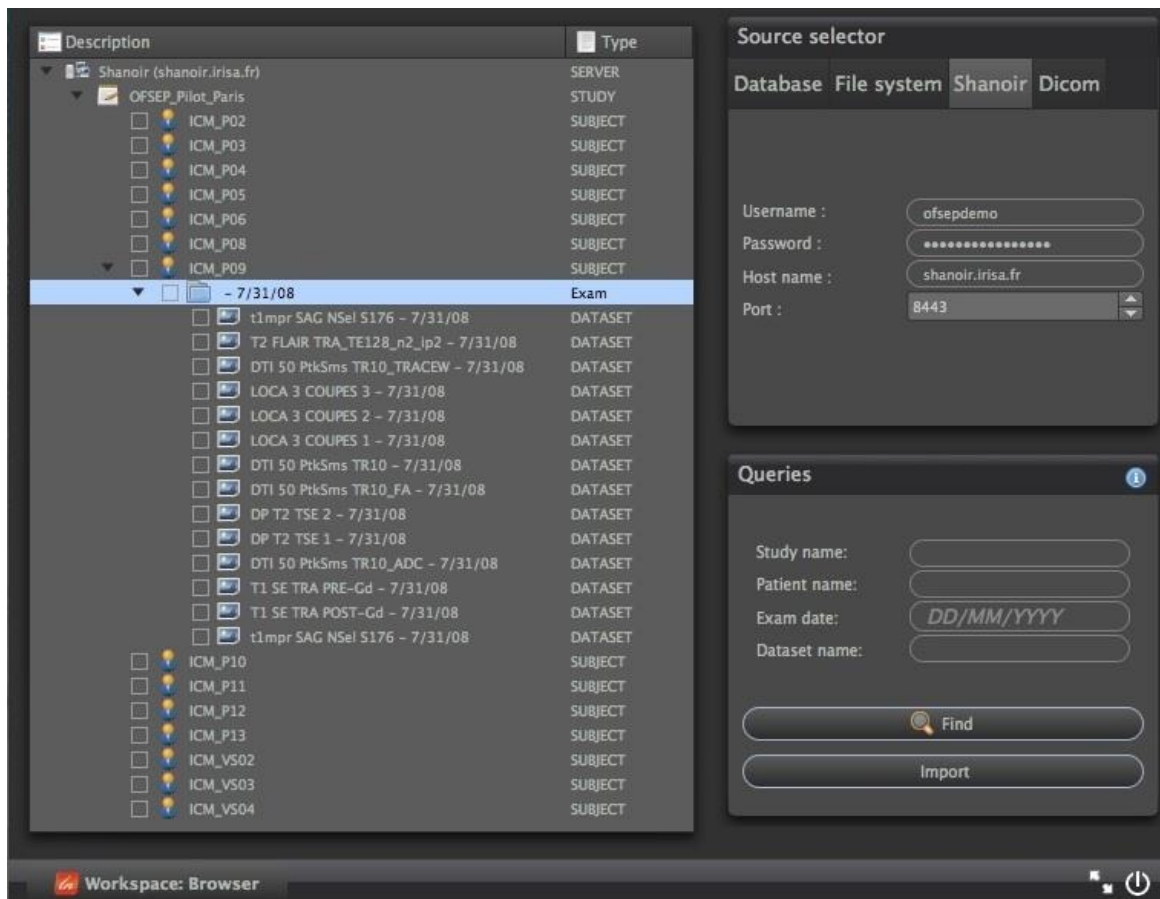
[††] medInria: http://med.inria.fr

Figure 10: Example of a Shanoir query service within the medInria environment by using QtShanoir web services

system calls depends on the container in which the calling process is running). As described in Figure 11, a minimal Shanoir deployment consists of four servers running in at least four separate containers:

- "*shanoir_container*": The actual Shanoir server. It relies on a *mysql* container (for the database) and on the PACS container (for archiving DICOM data),
- "*pacs_container*": The DICOM PACS server, currently managed by *dcm4chee*[‡‡],
- "*mysql_container*": The database server that is hosting two databases: *shanoirdb* and *pacsdb,*
- "*nginx _container*": The web frontend server based on a *nginx*[§§] HTTP server configured as a reverse-proxy for reaching the Shanoir server. It is the only server that is publicly reachable. It provides TLS encryption and security filtering,
- "*smtpsink _container*": An optional SMTP server for outgoing e-mails.

3   Data repositories

Each Shanoir repository has an administrator that manages the access rights of the repository. Each user requests an account through a web-based form and specifies which study he/she wants to access, contact, role in the study, required level of expertise/access (guest, user, expert, admin), etc. According to the information provided, the Shanoir administrator of the repository determines whether the user can access the system. Access to a specific study is granted by the person responsible for the study (i.e., the PI of the research study or the official representative). Depending on these settings, the new user will be able to see, download, and import datasets or even modify the study parameters. The corresponding rights are set for a limited time and must be renewed regularly. If requested, the user can receive a report by e-mail each time data are imported into the study.

[‡‡] http://www.dcm4che.org/
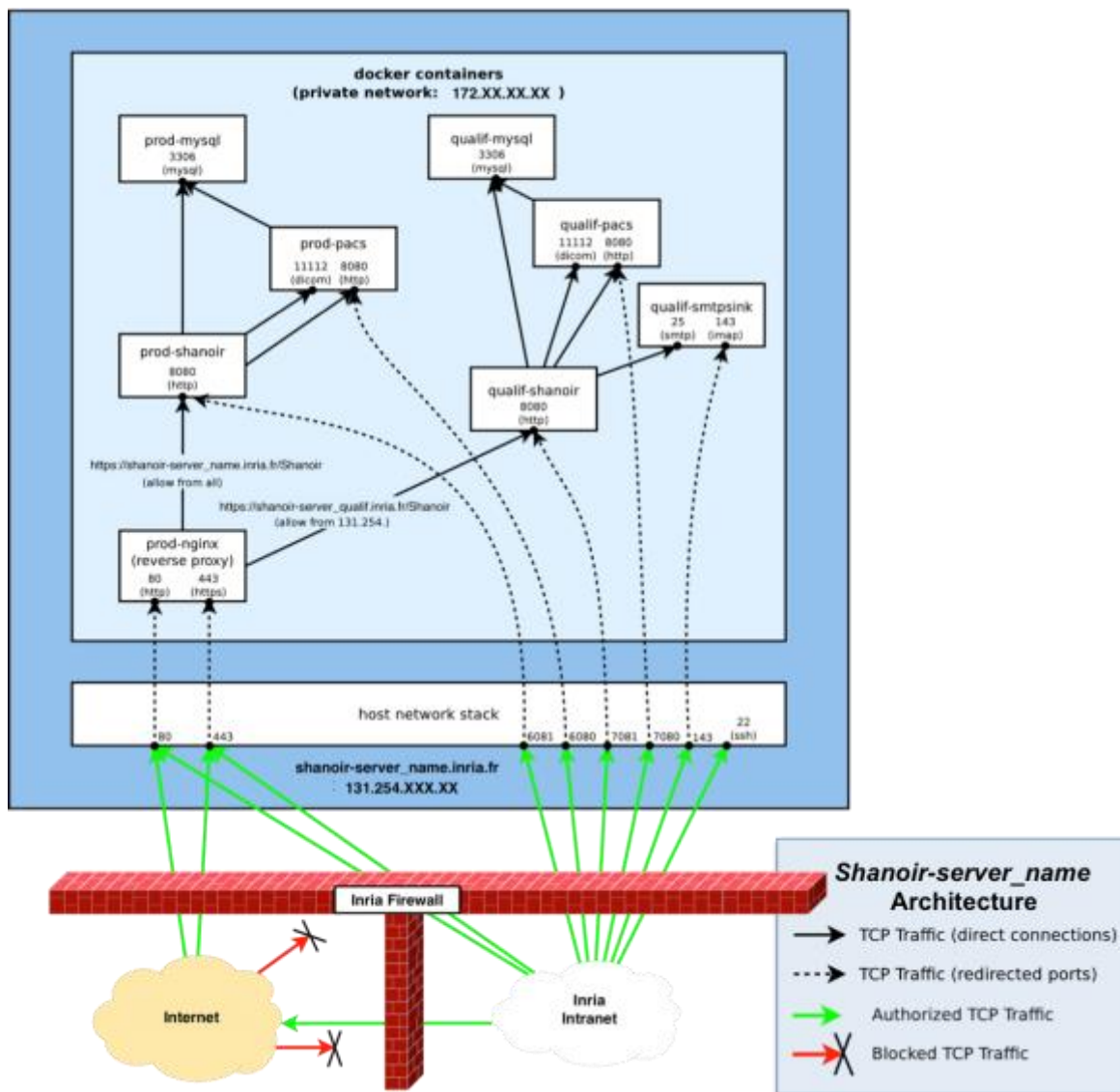[§§] http://wiki.nginx.org/

Figure 11: Schematic diagram of a Shanoir server. There is one set of containers for production: *prod-Shanoir*, *prod-pacs*, *prod-mysql* and *prod-nginx*, and one set for qualification: *qualif-shanoir*, *qualif-pacs*, *qualif-mysql* and *qualif-nginx* (not represented here). The arrows represent relationships between containers

3.2   The Shanoir@Neurinfo Repository

Started in 2009, the Neurinfo MRI 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 the resources as part of local, national and international imaging-based research projects.

All data produced at Neurinfo for academic or clinical research purposes are managed through a dedicated Shanoir@Neurinfo repository (Figure 12) administered by the facility's staff. The Shanoir@Neurinfo server also hosts data from imaging studies at multiple sites. In total, around 2To of data from 42 centers and 50 MR scanners are archived at this repository. The amount of data increases by 30Gb per month (see table in Figure 12).

In daily practice, DICOM data are imported by a technician from either a local PACS, a CD/DVD or a disk drive containing the DICOMDIR in its root directory and the DICOM files.

[***] http://www.neurinfo.org/

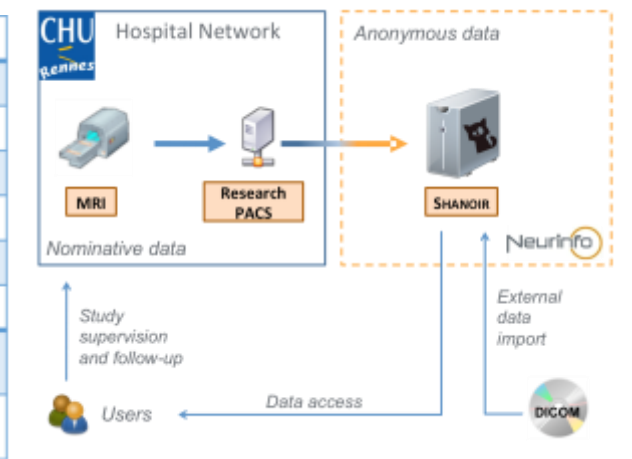| Last 2 years ←	09/2013	09/2014	09/2015	05/2016
Users (active)	(53)	118 (52)	178 (66)	206 (99)
Centers	27	31	42	50
Studies	53	60	70	121
Subjects	1706	2228	2833	3243
Examinations	2268	3157	4005	4531
Datasets	-	114 441	151582	176356
Raw data (Dicom)	1 126 GB (1.1 TB)	1 434 GB (1.4 TB)	1 588 GB (1.5 TB)	2 060 GB (2.1 TB)
Processed data (Nifti)	989 GB (0.9 TB)	1 331 GB (1.3 TB)	1 758 GB (1.7 TB)	1 841 GB (1.8 TB)

Figure 12: Evolution of the Shanoir@Neurinfo repository Global Statistics (*left*) and Service Infrastructure (*right*)

The clinical studies stored on the Shanoir@Neurinfo server concern the whole body (brain, spine, heart, lung, pelvis, vasculature, etc.) with a major focus on brain anatomy and function in normal control and pathological populations. Out of the 70 or so ongoing research studies on the Neurinfo platform, 75% relates to brain imaging, 15% concern abdominal imaging and 10% concern heart imaging. Among the neuro-imaging clinical studies, multiple sclerosis, dementia, tumors, stroke and mood disorders are the most investigated pathologies.

Depending on the specific nature of the research study, typical neuro-imaging protocols include structural imaging, functional BOLD MRI, Arterial Spin Labeling perfusion imaging, diffusion imaging, relaxometry sequences, pre- and post-gadolinium T1w sequences or vascular sequences. The following studies are examples of research carried out with the Shanoir@Neurinfo service as well as the types of data that are managed by the Shanoir@Neurinfo repository.

Along with the MRI raw data, post-processed images can be stored for each dataset. For example, in a study on functional motor activation, the motor areas were delineated by a trained radiologist and associated with each 3D T1w image. MS lesion segmentation masks can also be attached to the examination. In addition to image data, clinical scores can also be stored for each subject in the repository. Several MS clinical studies collect measurements such as the number of T2 new lesions, number of T1w Gd enhancing lesions or clinical scores such as EDSS. These measurements are also included in the search engine and consequently easily accessible through requests. For more advanced clinical follow-ups, Shanoir can easily be interfaced with existing databases.

The general policy for the Shanoir@Neurinfo repository for dissemination of data related to a particular study is decided upon beforehand with the PI in compliance 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 PI prior to allowing (complete or partial) access to a third party user. Nonetheless, to ensure dissemination and the best use of data acquired from public funding, the Neurinfo team strongly encourages investigators to share their data, which is usually done after an embargo period.

3.3 Shanoir@OFSEP Repository

The French Multiple Sclerosis Observatory (OFSEP)[†††], a major epidemiological tool on multiple sclerosis (MS) for the scientific community, was selected after a call for projects for Cohorts 2010, funded by France's Investment in the Future Program. It is a collaborative project involving over 40 MS research centers in France. The aim of the project is to build and maintain a nationwide cohort of patients with MS, and enrich the clinical data with biological samples, socio-economic data and neuro-images.

[†††] The OFSEP MS Cohort observatory: http://www.ofsep.org/en/

A dedicated imaging working group is in charge of acquiring, processing and integrating imaging and derived imaging data into a shared imaging resource center (IRC), and ensuring that the IRC is integrated with clinical databases. The consistent assessment of MRI-based measurements on a large scale requires robust and efficient image processing pipelines. A further goal of this project is to establish an information technology 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 easy access to the imaging data and allow modifications, queries, annotations and access control, the Shanoir environment has been selected. It will also ensure interoperability and data management related to the imaging aspect of the cohort (the clinical part is managed by the EDMUS[‡‡‡] system). For this purpose, we have set up a specific Shanoir@OFSEP image repository that is currently in its pilot phase.

Begun in 2012, the Shanoir@OFSEP server was installed to store the imaging data of the OFSEP cohort, which will study the neuroimaging data of 40,000 MS patients over the next 10 years. A consensus has emerged concerning the acquisition protocol, which requires (at least) one brain MRI every three years, one spinal MRI every six years, i.e., 200,000 MRIs over 10 years. The Shanoir@OFSEP database will grow during this period and beyond (Cotton et al., 2015).

Since OFSEP is a nationwide project covering many patients, many IRCs and many different kinds of MRI acquisition equipment, a national repository with nationwide access and uniform measures was therefore needed. The OFSEP imaging working group is continuously gathering new acquisition centers volunteering to take part to the cohort. In Shanoir@OFSEP, there are currently about 30 IRCs which include 31 pieces of MRI acquisition equipment representing 14 different MR scanner models from three MR manufacturers (Siemens, Philips and GE). All the centers are importing data in one main study called the "Mother Cohort". Each center follows the OFSEP protocol, which will be checked through the quality control module as described in section 2.2.2. If necessary, derived imaging data can then be imported back to the server in order to refer to potential post-processing information and MS specific imaging biomarkers to make them available for authorized users.

Currently the Shanoir@OFSEP repository is hosting five studies: the "Mother Cohort" (200,000 MRIs planned over the next 10 years) as well as four MS imaging clinical research projects. More of these "OFSEP-labeled" clinical research projects or nested cohorts will be integrated in coming 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 PI 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 OFSEP application.

4 Conclusion and perspectives

The Shanoir software as a service manages the sharing of distributed information sources in neuroimaging over the Internet, whether these resources are located in centers of experimentation, clinical departments in neurology, or research centers in cognitive neurosciences or image processing. Through the description of two repositories that administer 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 at their local hospital, research lab or company. Through the description of the Shanoir software environment, we have illustrated how neuroimaging data can be structured, managed, archived, visualized and shared.

In the medium term we plan to integrate Shanoir's resources and services with the open community through the French National Infrastructure's "France Life Imaging" (FLI)[§§§], and more specifically the "Information Analysis and Management" (IAM) node that is dedicated to providing large scale IT infrastructure for in vivo imaging. For this purpose, the FLI-IAM node will build and operate an

---

[‡‡‡] EDMUS: http://www.edmus.org

[§§§] France Life Imaging https://www.francelifeimaging.fr) with the IAM node (https://project.inria.fr/fli/) is a national infrastructure for in vivo imaging.

infrastructure to store, manage and process in vivo imaging data from human or pre-clinical procedures. The main achievements of the IAM node will consist of a versatile software platform composed of several sub-components that will connect hardware and software facilities to build:

- an archiving and management infrastructure of in vivo images as well as provide solutions to process and manage the acquired data through dedicated software and hardware solutions;
- versatile image analysis and data management solutions for in vivo imaging to facilitate interoperability between production sites and users and provide heterogeneous and distributed storage solutions for raw and metadata indexing (e.g. through the use of semantic models).

As such, we are under integrating Shanoir as one of the data management solutions of the FLI-IAM facilities along with a collection of companion data management software platforms such as CATI-DB (http://www.cati-neuroimaging.com) and ArchiMed, or a collection of processing clients or high-performance computing workflow facilities such as medInria, BrainVisa (brainvisa.info) and the VIP platform (http://www.creatis.insa-lyon.fr/vip). For this purpose, within FLI-IAM, we are setting up the "*glue*" between these platforms that will make it possible to connect and interoperate between them. In addition, through FLI-IAM, we will provide the necessary information for additional resources to join the FLI-IAM infrastructure by defining the basic conformal statement that will make the technology and scalability of FLI-IAM possible.

Nonetheless, as described in the introduction, there are a lot of similar initiatives going on recently in the medical imaging research field, like Human Brain Project, ADNI, XNAT-based solutions, etc. For these initiatives, as well as for Shanoir, the goal is to share the data at a large extent. This cannot be done without a significant additional effort on standardization in the field and on interoperability between software platforms addressing similar services. This is what motivates the integration of Shanoir in the French FLI-IAM e-infrastrucre initiative. This challenge of tomorrow is to continue this effort at the international level.

Acknowledgements

5   References

Ashish, N., Ambite, J.L., Muslea, M., Turner, J.A., 2010. Neuroscience data integration through mediation: An (F)BIRN case study. Frontiers in Neuroinformatics 4, 12 Original Research.

Barillot, C., Amsaleg , L., Aubry , F., Bazin , J.-P., Benali, H., Cointepas, Y., . . . Simon, E., 2003. Neurobase: Management of distributed knowledge and data bases in neuroimaging. Human Brain Mapping. Academic Press, New-York, NY, p. 726.

Barillot, C., Benali, H., Dojat, M., Gaignard, A., Gibaud, B., Kinkingnehun, S., . . . Temal, L., 2006. Federating Distributed and Heterogeneous Information Sources in Neuroimaging: The NeuroBase Project. Stud Health Technol Inform 120, 3-13.

Bellec, P., Lavoie-Courchesne, S., Dickinson, P., Lerch, J.P., Zijdenbos, A.P., Evans, A.C., 2012. The pipeline system for Octave and Matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows. Front Neuroinform 6, 7.

Benkner, S., Arbona, A., Berti, G., Chiarini, A., Dunlop, R., Engelbrecht, G., . . . Wood, S., 2010. @neurIST: infrastructure for advanced disease management through integration of heterogeneous data, computing, and complex processing services. IEEE Trans Inf Technol Biomed 14, 1365-1377.

Book, G.A., Anderson, B.M., Stevens, M.C., Glahn, D.C., Assaf, M., Pearlson, G.D., 2013. Neuroinformatics Database (NiDB)--a modular, portable database for the storage, analysis, and sharing of neuroimaging data. Neuroinformatics 11, 495-505.

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafo, M.R., 2013. Confidence and precision increase with high statistical power. Nat Rev Neurosci 14, 585-586.

Carp, J., 2012. The secret lives of experiments: methods reporting in the fMRI literature. NeuroImage 63, 289-300.

Cotton, F., Kremer, S., Hannoun, S., Vukusic , S., Dousset, V., The Imaging Working Group of OFSEP, O., 2015. OFSEP, a Nationwide Cohort of People with Multiple Sclerosis: Consensus minimal MRI protocol. Journal of Neuroradiology (in Press).

Das, S., Zijdenbos, A.P., Harlap, J., Vins, D., Evans, A.C., 2011. LORIS: a web-based data management system for multi-center studies. Front Neuroinform 5, 37.

Dinov, I., Lozev, K., Petrosyan, P., Liu, Z., Eggert, P., Pierce, J., . . . Toga, A., 2010. Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. PLoS One 5.

Evans, A.C., Brain Development Cooperative, G., 2006. The NIH MRI study of normal brain development. NeuroImage 30, 184-202.

Glatard, T., Rousseau, M.-E., Camarasu-Pop, S., Rioux, P., Sherif, T., Beck, N., . . . Evans, A.C., 2015. Interoperability between the CBRAIN and VIP web platforms for neuroimage analysis. Frontiers in Neuroinformatics Abstract.

Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S.S., 2011. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front Neuroinform 5, 13.

Gupta, A., Ludaescher, B., Martone, M., Rajasekar, A., 2003. BIRN-M: A Semantic Mediator for Solving Real-World Neuroscience Problems. ACM SIGMOD 2003. ACM, San Diego, CA., p. 678.

Hall, D., Huerta, M.F., McAuliffe, M.J., Farber, G.K., 2012. Sharing heterogeneous data: the national database for autism research. Neuroinformatics 10, 331-339.

Hibar, D.P., Stein, J.L., Renteria, M.E., Arias-Vasquez, A., Desrivieres, S., Jahanshad, N., . . . Medland, S.E., 2015. Common genetic variants influence human subcortical brain structures. Nature 520, 224-229.

Ioannidis, J.P., 2014. How to make more published research true. PLoS Med 11, e1001747.

Ioannidis, J.P., Munafo, M.R., Fusar-Poli, P., Nosek, B.A., David, S.P., 2014. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. Trends Cogn Sci 18, 235-241.

Jack, C.R., Jr., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., . . . Weiner, M.W., 2008. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging 27, 685-691.

Keator, D.B., Grethe, J.S., Marcus, D., Ozyurt, B., Gadde, S., Murphy, S., . . . Coordinating, B., 2008. A national human neuroimaging collaboratory enabled by the Biomedical Informatics Research Network (BIRN). IEEE Trans Inf Technol Biomed 12, 162-172.

Keator, D.B., Helmer, K., Steffener, J., Turner, J.A., Van Erp, T.G., Gadde, S., . . . Nichols, B.N., 2013. Towards structured sharing of raw and derived neuroimaging data across existing resources. NeuroImage 82, 647-661.

Keator, D.B., Wei, D., Gadde, S., Bockholt, J., Grethe, J.S., Marcus, D., . . . Ozyurt, I.B., 2009. Derived Data Storage and Exchange Workflow for Large-Scale Neuroimaging Analyses on the BIRN Grid. Frontiers in Neuroinformatics 3 Original Research Article.

Kennedy, D.N., Haselgrove, C., Riehl, J., Preuss, N., Buccigrossi, R., 2015. The Three NITRCs: A Guide to Neuroimaging Neuroinformatics Resources. Neuroinformatics 13, 383-386.

Marcus, D.S., Harms, M.P., Snyder, A.Z., Jenkinson, M., Wilson, J.A., Glasser, M.F., . . . Consortium, W.U.-M.H., 2013. Human Connectome Project informatics: quality control, database services, and data visualization. NeuroImage 80, 202-219.

Marcus, D.S., Olsen, T.R., Ramaratnam, M., Buckner, R.L., 2007. The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. Neuroinformatics 5, 11-34.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., 2003. WonderWeb Deliverable D18, Ontology Library (final). LOA-ISTC, CNR, Tech. Rep.

Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J., 1995. A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). NeuroImage 2, 89-101.

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., . . . Van Leemput, K., 2015. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging 34, 1993-2024.

Michel, F., Gaignard, A., Ahmad, F., Barillot, C., Batrancourt, B., Dojat, M., . . . Wali, B., 2010. Grid-wide neuroimaging data federation in the context of the NeuroLOG project. Proceedings of the HealthGrid, pp. 112-123.

Ooi, C., Bullmore, E.T., Wink, A.-M., Sendur, L., Barnes, A., Achard, S., . . . Suckling, J., 2009. CamBAfx: Workflow design, implementation and application for neuroimaging. Frontiers in Neuroinformatics 3 Original Research Article.

Poldrack, R.A., Gorgolewski, K.J., 2014. Making big data open: data sharing in neuroimaging. Nat Neurosci 17, 1510-1517.

Poline, J.B., Breeze, J.L., Ghosh, S., Gorgolewski, K., Halchenko, Y.O., Hanke, M., . . . Kennedy, D.N., 2012. Data sharing in neuroimaging research. Front Neuroinform 6, 9.

Rimal, B.P., Choi, E., Lumb, I., 2009. A taxonomy and survey of cloud computing systems. INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on. Ieee, pp. 44-51.

Roland, P.E., Zilles, K., 1994. Brain atlases - a new research tool. Trends in neurosciences 17, 458-467.

Scott, A., Courtney, W., Wood, D., de la Garza, R., Lane, S., King, M., . . . Calhoun, V.D., 2011. COINS: An Innovative Informatics and Neuroimaging Tool Suite Built for Large Heterogeneous Datasets. Front Neuroinform 5, 33.

Shepherd, G.M., Mirsky, J.S., Healy, M.D., Singer, M.S., Skoufos, E., Hines, M.S., . . . Miller, P.L., 1998. The Human Brain Project: neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data. Trends Neurosci 21, 460-468.

Sherif, T., Rioux, P., Rousseau, M.E., Kassis, N., Beck, N., Adalat, R., . . . Evans, A.C., 2014. CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research. Front Neuroinform 8, 54.

Styner, M., Lee, J., Chin, B., Chin, M.S., Commowick, O., Tran, H., . . . Warfield, S., 2008. 3D Segmentation in the Clinic: A Grand Challenge II: MS lesion segmentation. The Midas Journal 11.

Temal, L., Dojat, M., Kassel, G., Gibaud, B., 2008. Towards an ontology for sharing medical images and regions of interest in neuroimaging. J Biomed Inform 41, 766-778.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.U.-M.H., 2013. The WU-Minn Human Connectome Project: an overview. NeuroImage 80, 62-79.

Van Horn, J.D., Gazzaniga, M.S., 2013. Why share data? Lessons learned from the fMRIDC. NeuroImage 82, 677-682.

Van Horn, J.D., Grethe, J.S., Kostelec, P., Woodward, J.B., Aslam, J.A., Rus, D., . . . Gazzaniga, M.S., 2001. The Functional Magnetic Resonance Imaging Data Center (fMRIDC): the challenges and rewards of large-scale databasing of neuroimaging studies. Philos Trans R Soc Lond B Biol Sci 356, 1323-1339.

Walport, M., Brest, P., 2011. Sharing research data to improve public health. Lancet 377, 537-539.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., . . . Alzheimer's Disease Neuroimaging, I., 2012. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement 8, S1-68.