CATI: A Large Distributed Infrastructure for the Neuroimaging of Cohorts

Grégory Operto, Marie Chupin, Bénédicte Batrancourt, Marie-Odile Habert, Olivier Colliot, Habib Benali, Cyril Poupon, Catherine Champseix, Christine Delmaire, Sullivan Marie, et al.

CATI: A Large Distributed Infrastructure for the Neuroimaging of Cohorts

Grégory Operto[a,b,c,*], Marie Chupin[a,c,d], Bénédicte Batrancourt[a,c,e], Marie-Odile Habert[a,f,g], Olivier Colliot[a,c,d], Habib Benali[a,g], Cyril Poupon[a,b], Catherine Champseix[a,b,c], Christine Delmaire[a,i], Sullivan Marie[a,f,g], Denis Rivière[a,b], Mélanie Pélégrini-Issac[a,g], Vincent Perlbarg[a,g,h], Régine Trebossen[a,b,e,j], Michel Bottlaender[a,b,j], Vincent Frouin[b], Antoine Grigis[b], Dimitri Papadopoulos Orfanos[b], Hugo Dary[a,c,d], Ludovic Fillon[a,c,d], Chabha Azouani[a,c,d], Ali Bouyahia[a,c,d], Clara Fischer[a,b,c], Lydie Edward[a,c], Mathilde Bouin[a,b,c], Urielle Thoprakarn[a,b,c], Jinpeng Li[a,b,e], Leila Makkaoui[a,b], Sylvain Poret[a,c,k], Carole Dufouil[l], Vincent Bouteloup[m], Gaël Chételat[a,n], Bruno Dubois[a,c,k], Stéphane Lehéricy[a,c,o], Jean-François Mangin[a,b], Yann Cointepas[a,b,e] and the CATI Consortium

[*] Corresponding author: gregory.operto@cea.fr - Tel. +33-1-69082693

[a] Centre pour l'Acquisition et le Traitement des Images (www.cati-neuroimaging.com), France

[b] NeuroSpin, I2BM, Commissariat à l'Energie Atomique, Saclay, France

[c] INSERM U1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France

[d] INRIA Paris-Rocquencourt, 75013, Paris, France

[e] Information Analysis and Management project of France Life Imaging (http://project.inria.fr/fli)

[f] Nuclear Medicine Department, Pitié-Salpêtrière University Hospital, AP-HP, Paris, France

[g] Sorbonne Universités, UPMC Univ Paris 06, Inserm U 1146, CNRS UMR 7371, Laboratoire d'Imagerie Biomédicale, Paris, France,

[h] IHU-A-ICM, Bioinformatics/Biostatistics Platform, F-75013, Paris, France

[i] Université Lille Nord de France, U1171 - Département de Neuroradiologie, Centre Hospitalier Régional Universitaire de Lille, F-59037, Lille, France

[j] Service Hospitalier Frédéric Joliot, I2BM, Commissariat à l'Energie Atomique, Orsay, France

[k] Institut de la Mémoire et de la Maladie d'Alzheimer (IM2A), Département de Neurologie, Hôpital de la Pitié-Salpêtrière, AP-HP, Paris, France

[l] INSERM U708, Neuroepidemiology, CIC-EC7 & Bordeaux University, Bordeaux, France

[m] INSERM U897, Clinical Investigation Center-Clinical Epidemiology-CIC-1401, Epidemiology and Biostatistics Center, Bordeaux School of Public Health, Bordeaux University, Bordeaux, France

[n] INSERM U1077 - Université de Caen Basse-Normandie, UMR-S1077, Ecole Pratique des Hautes Etudes, CHU de Caen, Caen, France

[o] Centre de NeuroImagerie de Recherche — CENIR, Département de Neuroradiologie, Hôpital de la Pitié-Salpêtrière, AP-HP, Paris, France

Abstract

This paper provides an overview of CATI, a platform dedicated to multicenter neuroimaging. Initiated by the French Alzheimer's plan (2008-2012), CATI is a research project called on to provide service to other projects like an industrial partner. Its core mission is to support the neuroimaging of large populations, providing concrete solutions to the increasing complexity involved in such projects by bringing together a service infrastructure, the know-how of its expert academic teams and a large-scale, harmonized network of imaging facilities. CATI aims to make data sharing across studies easier and promotes sharing as much as possible. In the last four years, CATI has assisted the clinical community by taking charge of 35 projects so far and has emerged as a recognized actor at the national and international levels.

Keywords

large-scale studies; multicenter protocols; neuroimaging biomarkers; data sharing; data mining

1. Introduction

The current and forthcoming challenges for the neuroimaging community involve sharing at a large scale and for different matters, including data, software and expertise. Hence, data sharing has been expanding, providing benefits, such as a global reduction of costs, enhanced reproducibility, transparency and improved research practices (Poldrack *et al.*, 2014; Poline *et al.*, 2012). Furthermore, the community is actively looking for large sample sizes for imaging and genetics studies (Sejnowski *et al.*, 2014; Ferguson *et al.*, 2014; Frisoni *et al.*, 2011). Multiple, long-term, longitudinal follow-up is also crucial for understanding aging, neurodegenerative and neurodevelopmental processes. An ever-growing number of databases are available to researchers (Poldrack *et al.*, 2014; Poline *et al.*, 2012). Each database typically grants access to a comprehensive dataset, which can include raw data, clinical data, quality assessments, pre-processing steps, post-processing outputs and their endpoints. The different studies that supplied the data to these databases were generally based on specific sets of hypotheses, dataset acquisition parameters, quality control policies, and database browsing tools, which can vary greatly from one to another.

Currently, multicenter neuroimaging projects have to recruit a large set of required advanced skills, i.e., imaging physics, screening and enrollment of subjects, image processing, statistics and clinical neuroscience (Van Horn *et al.*, 2014). Meanwhile, some other disciplines, such as high energy physics and astronomy, rely on large coordinated teams and mutualized extensive instruments (Sejnowski *et al.*, 2014, Ferguson *et al.*, 2014, Wallis *et al.*, 2013, Weinberg *et al.*, 1961). For neuroimaging to follow this path, it must make the leap towards “big science” by providing large-scaled, shared infrastructures to provide support to academic research projects.

In this context, the CATI project, initiated in 2010 by the French Alzheimer's plan (2008-2012), is a national platform designed to support large-scale, multicenter, neuroimaging studies with optimized imaging workflows, from the initial acquisition of images to the final delivery of analysis outputs, e.g., surrogate endpoints. In close collaboration with the French societies of radiology and nuclear medicine, CATI has harmonized imaging acquisition across a network of over 50 French sites (Table 1) that can be extended on demand. The platform is dedicated to serving the community as a "large instrument" for brain studies by bringing together a service infrastructure, the know-how of expert teams and a wide-ranging imaging network.

After four years, CATI is now assisting 35 projects, of which 16 are currently managed on a daily basis, and 19 are at the startup phase with officially granted funding. Additionally, the platform has provided cost estimates to 10 more projects currently seeking support. These projects cover pathologies such as

Alzheimer's disease and related disorders, Parkinson's, Huntington's, Lewy body, frontotemporal dementia, hypertension, primary progressive aphasia, hippocampal sclerosis, AIDS, bipolar disorder and normal brain aging. To date, CATI's database contains more than 10000 subjects for which at least one MR exam has been acquired, providing services including the following:

- implementation of standardized imaging protocols covering various modalities for a pool of MR, PET and SPECT scanners of various manufacturers,

- creation and monitoring of a harmonized acquisition network where the quality and the settings of each scanner are checked over time,

- image analysis using an extensive portfolio covering most imaging modalities,

- quality control performed by experts on all acquired data and on every output of the analysis portfolio,

- machine learning tools for inference of biomarkers.

Table 1 List of the acquisition center members of the imaging national network harmonized by CATI

Acquisition center name	Location	Modality		
CHU Amiens Hôpital Nord	Amiens	MRI		
Institut Faire Faces - CHU Hôpital Sud	Amiens	MRI / PET		

CHU Angers	Angers	MRI		
CHU Besançon Hôpital Jean Minjoz	Besançon	MRI / PET		
CHU Avicenne - Bobigny	Bobigny	MRI / PET		
CHU Bordeaux Hôpital Pellegrin	Bordeaux	MRI / PET		
CH Boulogne-sur-Mer	Boulogne-sur-Mer	MRI		
CHU Hôpital de la Cavale Blanche	Brest	MRI / PET		
Centre Cycéron Caen	Caen	MRI		
CH Cambrai	Cambrai	MRI		
CHU Hôpital Gabriel Montpied	Clermont-Ferrand	MRI		
Centre Jean Perrin	Clermont-Ferrand	PET		
Hôpital Henri Mondor	Créteil	MRI / PET		
CHU Dijon	Dijon	MRI		
Centre Georges-François Leclerc	Dijon	PET		
CHU Douai	Douai	MRI		
CIMD Dunkerque	Dunkerque	MRI		
Centre Hospitalier du Val d'Ariège	Foix	MRI		
CHU Grenoble	Grenoble	MRI / PET / SPECT		
CHU Michallon	Grenoble	MRI		
Hôpital Lariboisière - Paris	Paris	MRI		
CH Lens	Lens	MRI		
CHRU Lille Hôpital Salengro	Lille	MRI / PET		
Clinique du Bois Lille	Lille	MRI		
CHU Limoges	Limoges	MRI / PET / SPECT		
CH Lyon Sud	Lyon	MRI		
CERMEP	Lyon	MRI / PET		
Hôpital Pierre Wertheimer - CH Lyon Est	Lyon	MRI/ SPECT		
Hôpital La Timone - Marseille	Marseille	MRI / PET		
CH Sambres Aversions - Maubeuge	Maubeuge	MRI		
CHPG Monaco	Monaco	PET/SPECT		
CHU Gui de Chauliac - Montpellier	Montpellier	MRI / PET		
Hôpital Brabois	Nancy	MRI		
Hôpital Central - Nancy	Nancy	MRI / PET		
Polyclinique de l'Atlantique - Nantes	Nantes	MRI		
Hôpital Nord Laennec - Nantes	Nantes	MRI / PET		
Hôpital Pasteur - Nice	Nice	MRI		
Hôpital de l'Archet - Nice	Nice	PET		
Centre Antoine Lacassagne - Nice	Nice	PET		
CHU Carémeau - Nîmes	Nîmes	PET		
CENIR ICM - Paris	Paris	MRI		

CHU Pitié-Salpêtrière - Paris	Paris	MRI / PET
Hôpital Sainte-Anne - Paris	Paris	MRI
Centre Cardiologique du Nord	Saint-Denis	MRI / PET
CHU Poitiers	Poitiers	MRI / PET / SPECT
Hôpital Maison Blanche - Reims	Reims	MRI
Institut Jean Godinot - Reims	Reims	PET
Centre Eugène Marquis - Rennes	Rennes	PET
Plateforme Neuroinfo CHU Rennes	Rennes	MRI

Table 1 (continued)

Acquisition center name	Location	Modality
CH Roubaix	Roubaix	MRI
Hôpital Charles Nicolle - CHU Rouen	Rouen	MRI
Centre Henri Becquerel - Rouen	Rouen	PET
CHU Saint-Etienne	Saint-Etienne	MRI / PET
Institut de Physique Biologique	Strasbourg	MRI
CHRU Hautepierre - Strasbourg	Strasbourg	MRI / PET / SPECT
CH Bigorre - Tarbes	Tarbes	MRI
CHU Purpan - Toulouse	Toulouse	MRI / PET
Hôpital Pierre-Paul Riquet - Toulouse	Toulouse	MRI
Hôpital Bretonneau - CHRU Tours	Tours	MRI / PET / SPECT
CH Tourcoing	Tourcoing	MRI
Cabinet de radiologie des Dentellières - Valenciennes	Valenciennes	MRI

2. The internal data flow of CATI

2.1 A distributed workflow

All studies managed by CATI follow the same typical workflow (Figure 1). Prior to any acquisition, each center composing the imaging network (Figure 1-1) undergoes a setup procedure, not detailed here, to optimize data harmonization across all centers (Table 1). This procedure is specific to the imaging modality: MR, PET or SPECT. The procedure includes follow-up visits to maintain harmonization over time. Once all of the centers have been through the opening procedure, patient inclusion begins and subjects are scanned over the entire network, following the acquisition protocols prescribed by CATI.

Data anonymization and secure transfers (Figure 1-2) are under the control of an external partner from every acquisition site to a central storage facility located in the Brain and Spine Institute (Institut du Cerveau et de la Moelle épinière), Pitié-Salpêtrière hospital. Once data are collected, clinical research associates perform quality checks of all raw MRI acquisitions (Figure 1-3) using a dedicated software platform that was specifically developed by CATI engineers for the CATI protocol and can be adapted easily to new protocols. This platform allows for checking for protocol consistency (MRI scanner, software version, reception coil, sequences acquired, order of the sequences), comparing parameters with those set at the beginning at each center for each study and can convert the raw DICOM image to the NIfTI research format. Each sequence of the protocol is then assessed through its own specific documented series of qualitative and quantitative evaluation indices, aiming at characterizing the acquisition slab positioning, movement, spikes and other artifacts and their localization, overall quality of the image through contrast, noise and intensity non-uniformity evaluation and other parameters (Figure 2).

Passing this first quality check is a green light for further analysis. Because CATI's expert teams are located in five distinct research labs, the validated data are then duplicated at all analysis sites. The dataset is first dropped into a push zone (Figure 1-4) under periodic and automatic sanity checks. Then, the dataset is automatically transferred to a secure directory named catishared. This repository is duplicated across multiple sites (Figure 1-5) so that each of the analysis labs has access to an exact copy, with periodic synchronization with the master repository. The contents of this directory are stored in a database called catidb, allowing users to query the data. The various controls (Figure 1-6) performed over this storage pool are detailed in section 2.3.

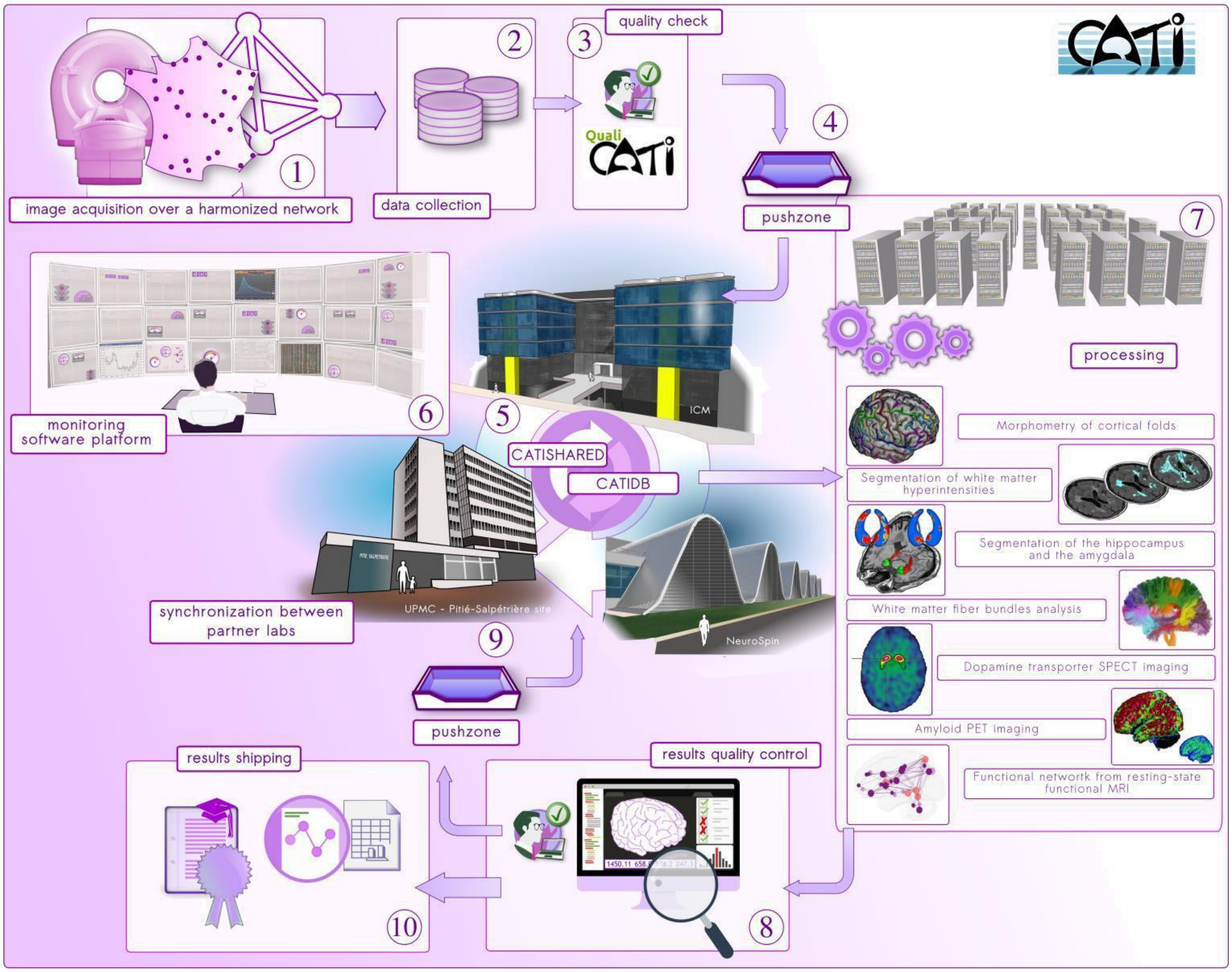

Figure 1. The data flow used in CATI (detailed step-by-step in section 2.1)

This system allows the different sites of the platform to gain access to the data and to analyze the data with any technique. The most intensive computations are performed using a dedicated 480-core cluster (Figure 1-7). The setup of parallelized processes on these computing resources is performed using the

BrainVisa software. The Python library Soma-Workflow forms the interface between BrainVisa and the cluster (Laguitton *et al.*, 2011). Once the processing jobs are completed, each expert team reviews the returned outputs following a detailed modality-specific procedure and assigns quality scores (Figure 1-8). The experts have support from dedicated tools that fire alerts for suspicious cases and seamless viewers (Operto *et al.*, 2015). The typical procedure for this review begins with generating summarized representations of the processing outputs, referred to as snapshots. SnapBase is a tool that was developed for this purpose: in a fully automated scheme, it loads datasets, selects relevant sets of slices or points of view, controls rendering parameters (e.g., colormap) according to expert specifications, captures an image, and saves the image in the database. The process is iterated with no need for supervision other than the initial selection of a dataset. The expert then reviews these snapshots using a dedicated interface called SnapCheck. When the snapshots are displayed for visual inspection, a series of automatic tests can drive the user's attention to subjects with abnormal features estimated from image volumes. This process takes advantage of the substantial amount of data that have already been processed and quality-checked by CATI by training learning algorithms that are then used to classify subjects prior to expert inspection. The user reviews subjects sequentially, assigning each of them a global quality score, ticking (if needed) specific observed issues, and adding detailed comments. The reports are then stored in the database. The interface is also connected to Anatomist visualization software for finer control, if necessary.

These verified outputs (image volumes, snapshots, measures, quality scores) are then ready to be integrated in catidb along with selected endpoints that are extracted from the output files. Some analysis pipelines require outputs from others: by adding them to the synchronized central system (Figure 1-9), they become available and can be reused by all teams. For instance, in this system, the diffusion-imaging expert team can routinely rely on quality-checked, grey-white matter masks produced by the T1-imaging experts.

The final step of this production pipeline consists of the collection of the final endpoints (detailed in section 3). This is performed using simple queries on catidb. The generated tables are delivered to the principal investigators (PI) in charge of the study (Figure 1-10).

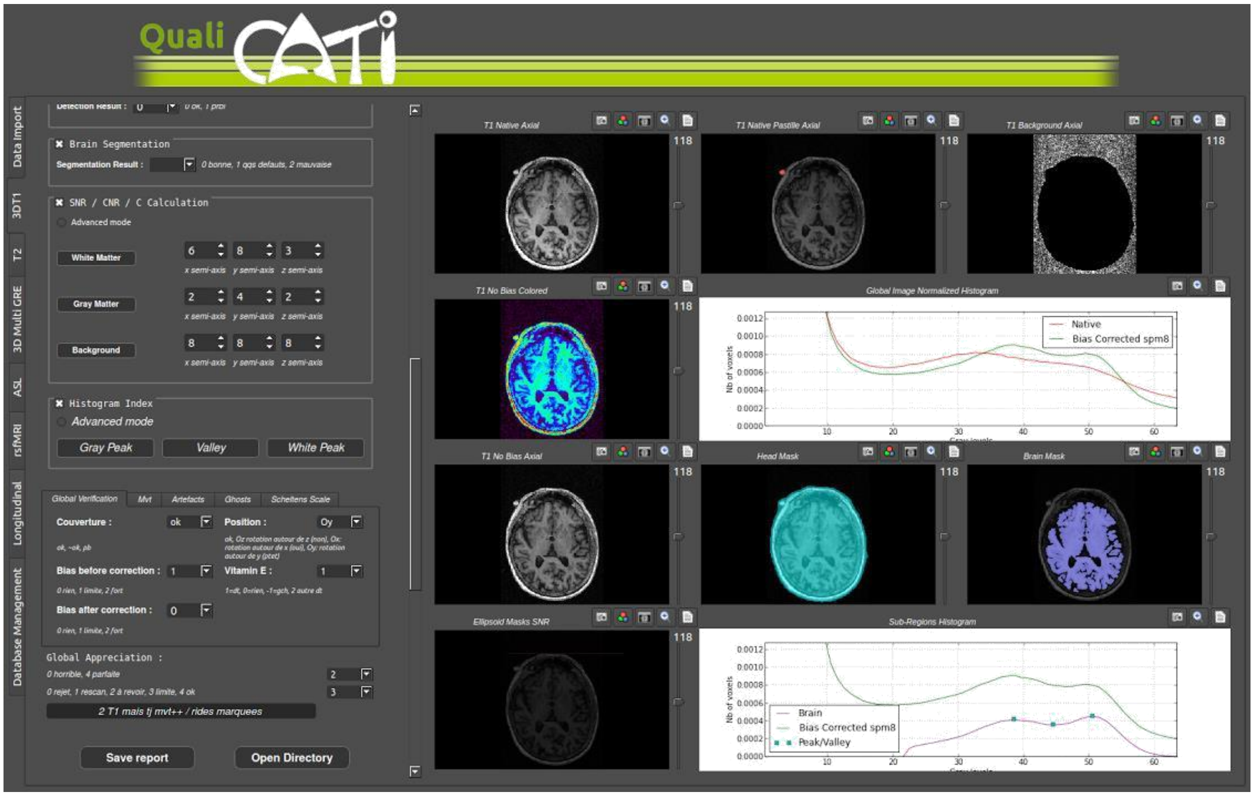

Figure 2. General interface of QualiCATI - the presented tab is dedicated to quality checks of T1-weighted images and allows verification of various control parameters or artifacts such as movement or spikes.

2.2 Focus on quality assessment

This production workflow is supported by extensive efforts to ensure data quality. The data flow is marked with repeated checkpoints, among which at least two involve visual control by experts, first on the acquired raw data and later on the processing outputs (identified in Figure 1 by icons with a green check mark).

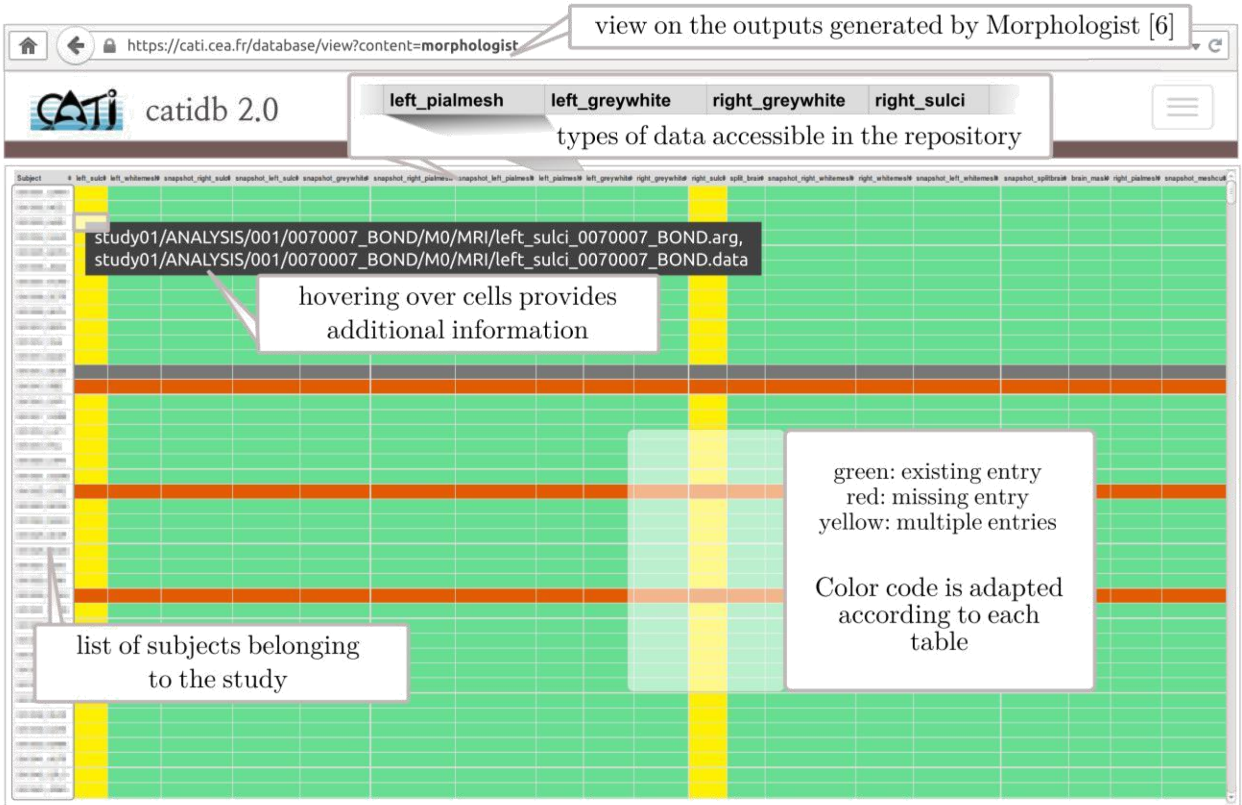

Figure 3. Example of a control table providing a quick overview of existing items in the database. Red cells indicate erroneous data that require immediate corrective action

Moreover, an integrated control station facilitates the monitoring of this data flow. Various verification tools are embedded in the system. Control tables (Figure 3) allow verification of the existing files over the entire system. Cells are color-coded so that invalid cases are quickly identified over whole studies. Various counters also help to keep track of any data added to the system. Additionally, daily checks are automatically performed, and their results sent to a private web server and by e-mail to the concerned users (Figure 4). These checks cover the different core functions of the system (e.g., general status, database, synchronization). The results are used to trigger maintenance operations and to give users a go/no go before further processing.

The system is automatically and periodically supplied with new data, and CATI has built in necessary verification systems, now operated by dedicated teams, to ensure the seamless running of the workflows and their security.

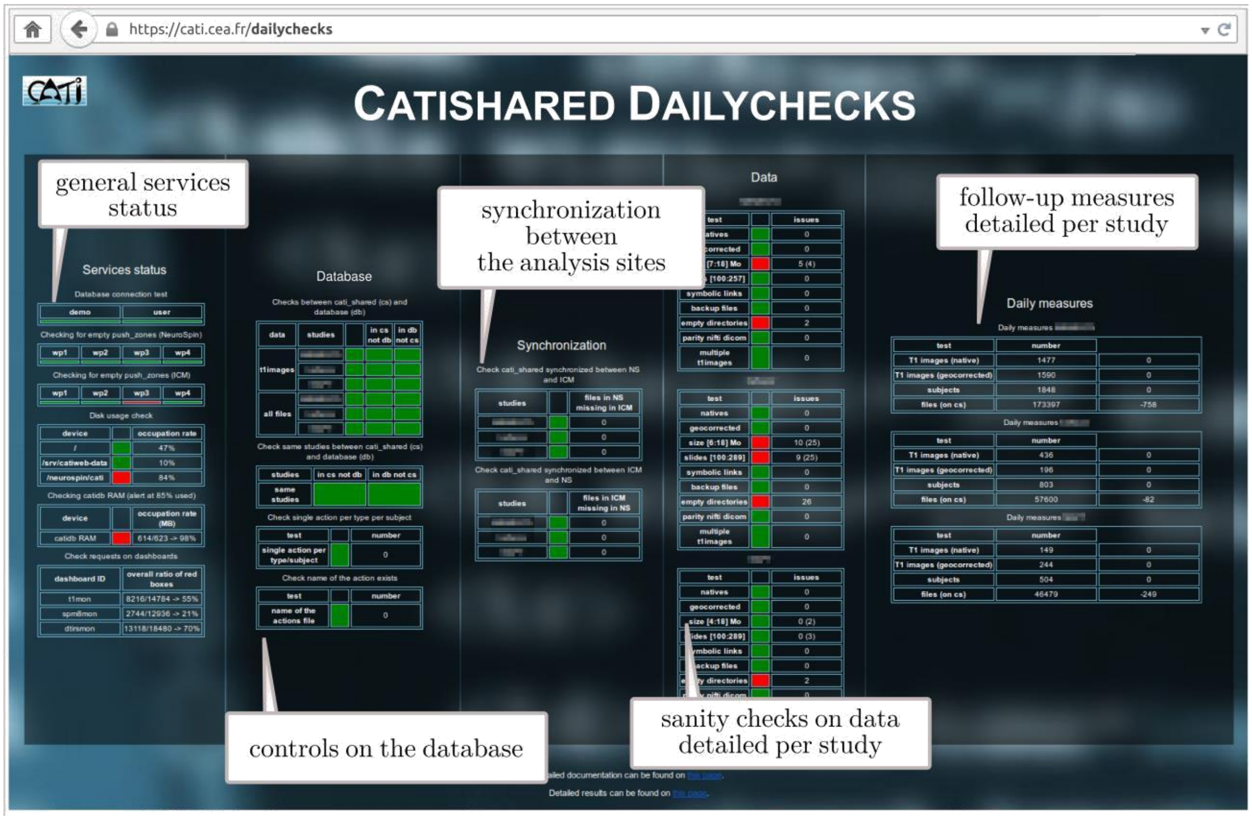

Figure 4. Fully automated, daily control procedures provide users with feedback on the status of services. Red cells drive the expert's attention on local failures and allow to fire specific maintenance operations

3. How to access the CATI data

One of the challenges of CATI was to manage data coming from different projects without imposing a common data organization. Many organization concepts can be aligned on the same schema across studies (subject, center, image type, modality). However, new studies regularly contain specific concepts missing in the original schema. This challenge raised the need for an infrastructure that is able to store and query singular concepts (specific to a single study) and to allow easy update of the schema to include new common concepts. To address these needs, catidb built its own infrastructure based on well-established technologies (such as Postgresql, Python, Javascript, and HTML5). CATI has a community of tens of developers with good knowledge of the Python language therefore this language plays a central role in this infrastructure, and it is used to make homogeneous interfaces with various neuroimaging tools (Cointepas *et al.*, 2010).

3.1 Data access rights and security

Although CATI is able to host open data, it was initially designed to host confidential data. Therefore, security is an important aspect of its infrastructure. All access is encrypted, and it is mandatory to have an account to gain access to any neuroimaging data. Once the account is created and validated, the user can ask for access to one or more datasets for a research project. CATI does not make the decision of making the data available, it sends the user requests to the PI of the targeted study and waits for their decision to allow or refuse data access. If a study wants public access, it is possible to ask CATI to make the data available automatically to all registered users (if legal and ethical rules are respected). There is no such open data study hosted on the system at the time of this writing. The two networking protocols used are HTTPS for web access (including small data download) and SFTP for downloading imaging data. Access is controlled by logins and passwords that are stored in a single LDAP directory and on a separate machine not accessible via the internet.

Users are only granted read access to the data. Only the CATI team has write access to control incoming data. The main method for collecting incoming data is a dedicated tool with a web interface, installed at the acquisition centers, that securely sends anonymous data. Data can also be sent using SFTP or even using mail and DVDs when necessary.

3.2 Technical overview

For security reasons, we chose to use two different servers for the public website (http://cati-neuroimaging.com) and the database website (https://cati.cea.fr). The database website is easily accessible from the public website. In this section, we provide an overview of the architecture of the database server. There are three main components that can be used to access the data:

REST API: The database system is not directly exposed to the user. Instead, one must use an API which performs HTTP requests (GET, PUT, POST or DELETE) and JSON encoded data exchanges. Such an API offers various advantages.

- Hides all of the complexity of the database and the various optimization strategies employed for efficient queries.
- Shares all of the authentication and authorization implementation with the database website.
- Relies on standard, well-established technologies that are available in many (if not all) existing and potential user environments. Thus, developing new client software is only a matter of understanding the API; there are no difficulties due to technical frontiers.

Website: The database website (https://cati.cea.fr) allows a novice user to discover the data by browsing the various studies and their data. It also lets the user track the progress of acquisitions and processing using dashboards (Figure 3). The database website relies on the REST API for accessing data. Therefore, it is possible to change the underlying database system (which has already happened twice in CATI's history) without re-implementing the website.

Virtual file system: Considering the volumes of managed data, easy solutions for downloading gigabytes of data and thousands of files are essential. To this end, files and directories are exposed through the SSH File Transfer Protocol (SFTP). This protocol is supported by client software on most systems, some of which are freely available and provide a graphical interface that can be used by non-experts. Advanced users can find command line tools and libraries to access the data from scripts or dedicated software. Standard SFTP software conveniently allows resuming an interrupted download without resending all of the data. However, a virtual file system is employed in order to expose only a subpart of the files following a selection made by the user (e.g. on the website) and to manage files access rights using the REST API. This results in a directory whose content is entirely controlled by a server program developed by CATI.

Figure 5 shows the technical organization of the catidb server. All user access occurs via either HTTPS (using Apache) or SFTP (using OpenSSH). User authentication is managed by LDAP (http://www.openldap.org/). The main imaging data are stored in the file system, and all other data (concerning studies, subjects, acquisition centers) are stored in a PostgreSQL (http://www.postgresql.org/) database. REST API and the website are developed using Pyramid (http://www.pylonsproject.org/), and the virtual file system is Python software based on FUSE technology (http://fuse.sourceforge.net/).

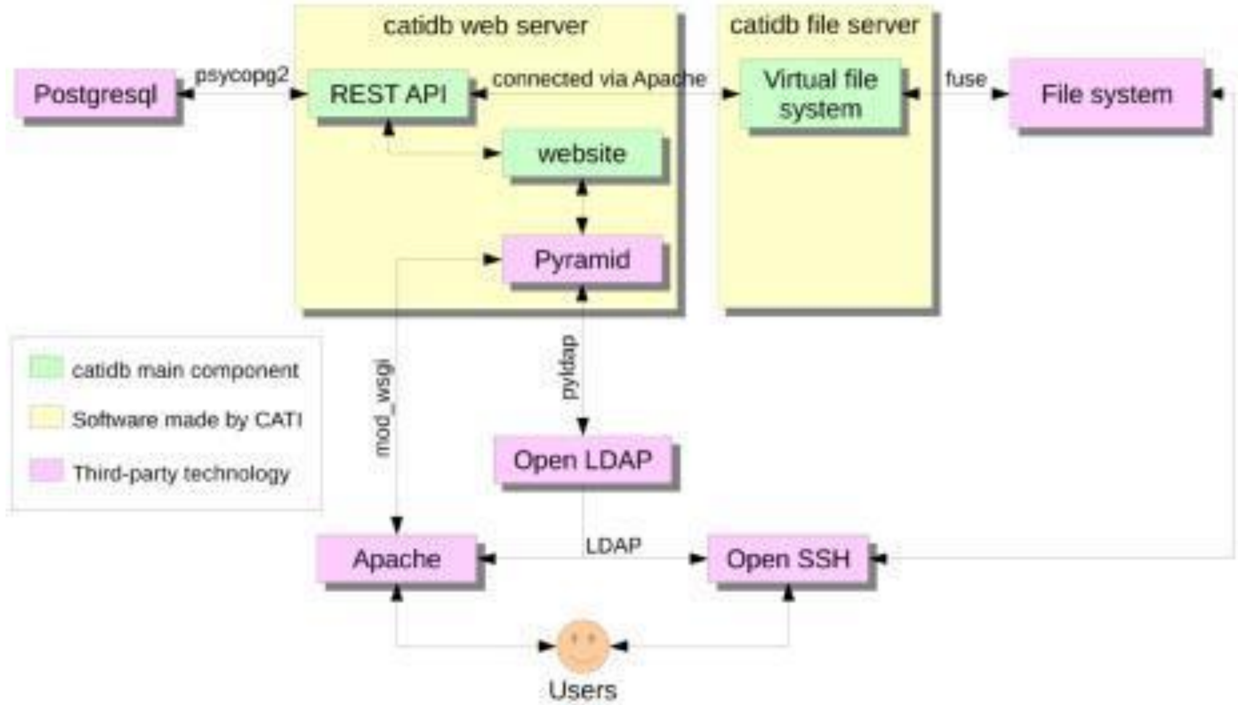

Figure 5. Illustration of how the three main catidb components (in green) are connected to users using well-established technologies (in pink)

3.3 The REST API

It is beyond the scope of this article to give a full reference of the web services API (the REST API). This section will give an overview of this API and its principles. The API is built on two main concepts :

Studies: All data are organized by study and access rights are mainly based on studies. Therefore only a few services do not require a study name in parameters.

Actions[1]**:** an action is attached to any file that is pushed in catidb. This action is a JSON object that allows the identification of the type of data contained in the file and may contain some metadata. For instance, the following action (in a simplified version for clarity reasons) is produced and used when a T1 MR image volume is acquired and integrated in the system. It contains metadata such as subject code, acquisition center and time point and two file names corresponding to the two stored versions of the image volume in both DICOM and NIfTI formats.

[1] The *action* term is associated to the *action* of integrating a new dataset in the system.

```
{ "action_name": "mri_3dt1",
"center_code": "029",
"subject_code':"0290015_LAJF",
"time_point": "M000",
  "t1_dicom":"memento/029/fake_subject_code/M000/3DT1/fake_subject_code_M000_3DT1_S002.tar.gz", "t1_nifti":
    "memento/029/fake_subject_code/M000/3DT1/fake_subject_code_M000_3DT1_S002.nii.gz"
}
```

The content of the actions is used to populate catidb database. Therefore, actions can contain information extracted from the files. As an example, actions used to push some post-processing outputs contain not only the names of the files generated by the software but also all the measures that CATI makes available via catidb.

All web services of the API can be reached by issuing HTTPS requests to an URL beginning with `https://cati.cea.fr/api/1/` (where 1 is the version of the API). The body of the exchanged data is in JSON format both in these requests and in the answers from the server. For instance `https://cati.cea.fr/api/1/studies` (the base URL of the API will be omitted further in this section, as in `/studies`) will return an array of studies, each study being an object containing the study attributes.

Using the service `/login` is a first mandatory step in order to be identified and be able to use the API. All data related services include the name of the study, as in `/memento/paths`, which will return the paths of all the files in the Memento study. For every service returning arrays of objects, parameters can be added to filter the result: `_limit=10` will limit the size of the array to 10 elements, `_offset=50` will ignore the first 50 objects, `subject_code=james_bond` will only return objects having "`james_bond`" as value for the attribute named "`subject_code`". Using the `_count` parameter will get only the size of the returned array. For example, `/memento/paths?subjectcode=james_bond&_count` returns the number of files concerning the subject whose code is "`james_bond`".

Since Python is widely used by CATI, we also developed a Python API that directly uses the REST API but hides all the difficulties of managing authentication (Figure 6).

```
In [1]:  from catidb.api import get_catidb
         catidb = get_catidb()


         # Count the number of freesurfer analysis for study memento
         catidb.actions('memento','freesurfer', _count=True)

Out[1]:  2077

In [11]: # Get the left and right precentral measure from freesurfer subject
         # 'fake_subject_code' at time point M000 in study Memento
         catidb.actions('memento','freesurfer', subject_code='fake_subject_code',
                        time_point='M000', _fields='left_precentral,right_precentral')

Out[11]: [{u'left_precentral': 2.543, u'right_precentral': 2.472}]
```

Figure 6. Example of use of the Python API. It allows to connect to catidb by using previously stored login and password and handles authentication issues (e.g. reconnects when the session expires). The Python API is exposing the same features as the REST API under the form of a classical Python module.

All catidb interactions with users and CATI members go through the API (with the exception of pushing data files that is done with sFTP). Even the database website is not accessing directly the database but is only relying on the API (Figure 5).

4. CATI's portfolio

CATI's portfolio includes offers for numerous MRI, PET and SPECT modalities covering acquisition and processing.

4.1 Managed image modalities

- anatomical MRI: 3D T1, T2 FLAIR, T2* GRE, T2 TSE
- resting-state fMRI
- ASL MRI
- diffusion MRI
- nuclear imaging: PET-FDG, PET-amyloid, 123I-FP-CIT SPECT

4.2 Image analysis portfolio

The image analysis portfolio covers the various managed modalities. The methods involved in this portfolio are applied to the data from every subject to produce the endpoints that are eventually delivered to the PI of each study. The process consists of both standards from the literature and methods that are developed within CATI's expert teams. The delivered endpoints are primarily chosen for their potential as surrogate markers in Alzheimer's disease and related disorders. These endpoints can be extended on demand.

The methods developed by founder labs include the following:

- hippocampus volumetry (Chupin *et al.*, 2009; Colliot *et al.*, 2008),
- estimation of cortical folds openings, shape analysis, gyrification indices (Mangin *et al.*, 2010; Mangin *et al.*, 2004),
- segmentation of white matter hyperintensities (Samaille *et al.*, 2013),
- integration of main functional networks,
- tractography, profile analysis on white matter fiber bundles (Guevara *et al.*, 2012),
- voxel-wise and ROI-based analysis of FDG-PET or amyloid-PET data in the normalized and the subject's native space,
- partial volume effect and cortical atrophy corrections in PET data analysis,
- calculation of binding potentials for SPECT dopamine transporter imaging (Martini *et al.*, 2014)

The portfolio also includes the following well-established methods:

- global volumetry and voxel-based morphometry using SPM software (both versions 8 and 12) (Penny *et al.*, 2011),
- estimation of cortical thickness using FreeSurfer 5.3 software (Fischl *et al.*, 2012)

5. Conclusion and outlook

5.1 Towards a long-term national structure

CATI was initiated in 2010 by the French Alzheimer's plan (2008-2012) with the main objective to build the imaging core of the Memento cohort, a French multicenter national prospective study including 2300 subjects having cognitive symptoms ranging from isolated cognitive complaints to mild cognitive impairment, recruited from memory clinics and followed-up over four years. MRI and FDG-PET acquisition were planned at inclusion and then every two years, and the funding for three time steps is still secured. Additionally, amyloid and ASL imaging are currently being performed on two subsamples of 800 subjects (namely the Memento-Amyging and Memento-Vascod ancillary studies). Several institutions of higher education and research are involved in CATI (CEA, UPMC, ICM, AP-HP, CNRS, INRIA), which aggregates a large variety of expertise.

The CATI project has been extended by the French Neurodegenerative Diseases plan (2014-2019). The French government aims to secure its sustainability, possibly through the creation of a structure associating public and private partners. Its long-term missions include operating and extending the current imaging network, managing the national neuroimaging database, supporting multi-center projects, providing impetus for research on imaging biomarkers and bolstering clinical tools. In addition, CATI is now the imaging core of several European projects, which will extend the harmonized network and collect additional expertise.

5.2 Platform interoperability, open science and reproducible research

The managed projects are currently in their acquisition and curation stages. Each project can choose its own sharing policy, but CATI and financial incentives encourage the project PIs to extend sharing as much as possible. The underlying infrastructure combining components, such as the harmonized imaging network, the extended analysis portfolio and the controlled data flow, ensure consistency between datasets, providing key items to increase the community capacity to perform cross-study meta-analyses.

In addition to its core missions, CATI is also involved in joint efforts to support platform interoperability, the opening and sharing of data, tools and code, and the promotion of open science and reproducible research. To achieve these goals, CATI is involved in several national projects e.g., the Information Analysis and Management project (IAM http://project.inria.fr/fli), which is a transversal node of the France Life Imaging infrastructure (FLI), and international e.g., neuGRID4you (N4U http://neugrid4you.eu).

Like many of other fields, neuroimaging is entering the era of big data. Therefore, the future will require interoperable service infrastructures, such as CATI, N4U, CBRAIN (http://mcin-cnim.ca/neuroimagingtechnologies/cbrain) and LONI (http://ida.loni.usc.edu). A super-arching organization in charge of globally synchronizing this network of platforms is needed to proceed with the advent of standard protocols and data sharing. A big data perspective is mandatory to generate the normative charts that will support the future clinical use of imaging biomarkers.

Acknowledgments

This work received funding from the French Fondation Alzheimer; the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n. 283562; the French program “Investissement d’Avenir” run by the ‘Agence Nationale pour la Recherche’; grant 'Infrastructure d’avenir en Biologie Santé - ANR-11-INBS-0006’ and IHU-A-ICM, Paris Institute of Translational neuroscience, ANR-10-IAIHU-06.

Appendix:

The members of the CATI Consortium are :

Executive committee: Jean-François Mangin, Stéphane Lehéricy, Bénédicte Batrancourt, Habib Benali, Catherine Champseix, Marie Chupin, Yann Cointepas, Olivier Colliot, Bruno Dubois, Marie-Odile Habert, Sullivan Marie, Grégory Operto, Mélanie Pélégrini-Issac, Cyril Poupon, Denis Riviere, Bertrand Thirion, Régine Trébossen
Overall project management: Catherine Champseix, Thomas Estienne, Karima Lamrid
MRI Acquisition: Marie Chupin, Cyril Poupon, Chabha Azouani, Ali Bouyahia, Hugo Dary, Sonia Djobeir, David Gay, Johanne Germain, Cedric Meurée, Amadou Tall, Urielle Thoprakarn, Kelly Martineau, Christine Delmaire, Stéphane Lehéricy, Mélanie Pélégrini-Issac, Vincent Perlbarg, Alexandre Vignaud
PET/SPECT Acquisition: Marie-Odile Habert, Régine Trebossen, Sullivan Marie, Hugo Bertin, Maxime Locatelli, Jean-Baptiste Martini, Michel Bottlaender, Yann Cointepas, Aurélie Kas
Database / Dataflow: Bénédicte Batrancourt, Yann Cointepas, Mathilde Bouin, Lydie Edward, Jinpeng Li, Sylvain Poret
Image Processing Factory: Jean-François Mangin, Marie Chupin, Grégory Operto, Mamadou Diallo, Ludovic Fillon, Clara Fischer, Malo Gaubert, Mickaël Labit, Clarisse Longo Dos Santos, Yann Cointepas, Mélanie Pélégrini-Issac, Vincent Perlbarg, Cyril Poupon, Denis Rivière

Research & Development: Habib Benali, Olivier Colliot, Enrica Cavedo, Mathieu Dubois, Takoua Kaaouana, Kamalakar Reddy, Marion Houot, Marie Chupin, Stanley Durrleman, Edouard Duchesnay, Jean-François Mangin, Alexandre Routier, Zhong Yi Sun
Former members: Xavier Badé, Nader Cheaib, Samuel David, François De Guio, Delphine Duclap, Jessica Erber, Soizic Laguitton, Leila Makkaoui, Matthieu Perrot, Cyrille Turpin, Nicolas Vibet

MEMENTO study core members: Carole Dufouil, Geneviève Chêne, Vincent Bouteloup, Helen Savarieau

MRI network: Jean-Marc Constans, Hervé Deramond, Corinne Guillaume, Danielle Lembach, Sophie Potier (Amiens); Jacques Guyard, Anne Pasco-Papon (Angers); Françoise Cattin (Besançon); Guillaume Bertrand, Patrice Cayacy (Bobigny); Michèle Allard, Bixente Dilharreguy, Vincent Dousset (Bordeaux); Thierry Stekelorom (Boulogne-sur-Mer); Douraied Ben Sallem, Elisabeth Bruyant (Brest); Nicolas Delcroix (Caen); Philippe Lavau (Cambrai); Jean-Marie Bonny, Betty Jean, Stéphane Pouponneau (Clermont-Ferrand); Pierre Brugières, Julien Lincot (Créteil); Sylvain Bornier, Frédéric Ricolfi (Dijon); Arnaud Beauguitte (Douai); Eric Berlemont (Dunkerque); Dominique Dubois (Foix); Alexandre Krainik, Johan Pietras, Irène Troprès (Grenoble); Denis Berteloot (Lens); Arnaud Aubertin, Isabelle Lambert (Lille-Loos); Christine Delmaire, Julien Dumont, Jean-Pierre Pruvot (Lille); Marie-Paule Boncoeur-Martel, Nelly Viremonet (Limoges); François Cotton, Lionel Henarejos, Danielle Ibarrola, Franck Lamberton, Guy Louis-Tisserand (Lyon); Nadine Girard (Marseille); Philippe Debastelier (Maubeuge); Giuliano Michelozzi (Monaco); Alain Bonafé, Thierry Chaptal, Emmanuelle Le Bars, Nicolas Menjot de Champfleur, Mélanie Pigot (Montpellier); Serges Bracard, Marc Braun, Emmanuelle Schmitt, Anna Simler (Nancy); Elisabeth Auffray-Calvier, François Xavier Bertrand (Nantes); Stéphane Chanalet, Sylvie Duverneuil, Lydiane Mondot (Nice); Eric Bardinet, Anne Bertrand, Antoine Burgos, Mélanie Didier, Didier Dormont, Frédéric Humbert, Stéphane Lehéricy, Stéphanie Lion, Kevin Nigaud, Catherine Oppenheim, (Paris); Rémy Guillevin (Poitiers); Laurent Pierot, Christophe Portefaix (Reims); Elise Bannier, Isabelle Corouge, Jean-Christophe Ferré, Jean-Yves Gauvrit, Catherine Guillemot, Agnès Pélerin (Rennes); Emmanuel Michelin (Roubaix); Yohann Cruypeninck, Emmanuel Gerardin (Rouen); Johann Girard Chéron, Philippe Larralde, Mathieu Rodallec (Saint-Denis); Claire Boutet, Isabelle Faillenot, Michel Paret (Saint-Etienne); Jean-Paul Armspach, Stéphane Krémer, Paulo Loureiro de Sousa (Strasbourg); Françoise Hugon (Tarbes); Fabrice Bonneville, Pierre Celsis, Hélène Gros-Dagnac, Nathalie Vayssière (Toulouse); Laurent Barantin, Jean-Philippe Cottier (Tours); Pierre Dobbelaere, Maxime Lalisse (Valenciennes)

PET imaging network: Pascal Bailly, Marc-Etienne Meyer (Amiens); Francis Bouchet, Audrey Beneteau, Pacome Fosse (Angers); Yolande Petegnief, Hatem Boulahdour (Besançon); Gabriel Pop, Marie-Joséphine Waryn (Bobigny); Frederic Lamare, Sandrine Fouchet, Michèle Allard (Bordeaux); Solène Querellou, David Bourhis (Brest) Antony Kelly, Corinne Millardet, Elodie Jouberton (Clermont-Ferrand); Inna Dygai-Cochet, Jean-Marc Vrigneaud (Dijon); Céline Gallazzini-Crepin, Ghislaine Reboulet, Emilien Genod (Grenoble); Claude Hossein-Foucher, Maximilien Vermandel, Elodie Cordier (Lille); Isabelle Quelven, Romain Turcat (Limoges); Franck Lavenne, Christian Scheiber, Perrine Tylski, Christian Scheiber (Lyon); Bardia Farman, Éric Guedj, Fanny Barthélémy (Marseille); Benjamin Serrano, Florent Hugonnet (Monaco); Delphine De Verbizier Lonjon, Marjolaine Fourcade, Michel Zanca (Montpellier); Laetitia Imbert, Antoine Verger, Véronique Roch (Nancy); Amandine

Pallardy, Thomas Carlier (Nantes); Fabien Maurel, Micheline Razzouk-Cadet, Guillaume Nivaggioni, Pierre Malick Koulibaly, Jacques Darcourt, Estelle Occelli (Nice); Marie-Odile Habert, Navichka Jungalee, Aurélie Kas, Marine Soret, Valérie Causse-Lemercier (Paris); Christelle Gallais, Ludovic Blanchard, Melissa Collot, Rémy Perdrisot (Poitiers); Florence Lejeune, Sophie Laffont (Rennes); Pierre Vera, Sébastien Hapdey, Arthur Dumouchel, Pierrick Gouel (Rouen); Mathieu Queneau, Karim Farid, Laurence Merlin (Saint-Denis); Nathalie Prevot-Bitot, Marielle Decousus (Saint-Etienne); Izzie Jacques Namer, Fabrice Hubelé, Élise Enderlin (Strasbourg); Anne Hitzel, Pierre Payoux (Toulouse); Maria-Joao Santiago-Ribeiro, Catherine Pérault (Tours)

Method labs and platforms network: CHU Toulouse (Pierre Payoux); Clinical Imaging Core faCility – CI2C, CHRU Lille (Christine Delmaire, Julien Dumont, Renaud Lopes); INSERM U1077, Caen (Gaël Chételat); Neurinfo platform / IRISA VisAGeS research team, Rennes (Elise Bannier, Christian Barillot, Isabelle Corouge, Jean-Christophe Ferré); Plateforme d'imagerie in-vivo / ICube lab, Strasbourg (Paulo Loureiro de Sousa); IRMaGe platform / GIN lab, Grenoble (Alexandre Krainik, Laurent Lamalle, Irène Troprès, Jan Warnking); IR4M, UMR8081, Paris-Sud University – CNRS, Orsay (Ludovic de Rochefort); Plateforme Ibio / INCIA, Bordeaux (Michèle Allard, Bixente Dilharreguy); CERMEP, Lyon (Danielle Ibarrola, Franck Lamberton); MRI platform, ISCT, Toulouse (Hélène Gros-Dagnac, Nathalie Vayssière); MRI platform, I2FH, CHU Gui de Chauliac, Montpellier (Emmanuelle Le Bars, Nicolas Menjot de Champfleur); Unité Imagerie et Cerveau, Tours (Laurent Barantin); LaBRI – UMR 5800, Bordeaux (Pierrick Coupé); ASCLEPIOS, Sophia-Antipolis (Nicholas Ayache, Xavier Pennec); INT / CNRS UMR6168, Marseille (Olivier Coulon, Guillaume Auzias, Julien Lefèvre); MIRCen, Fontenay-aux-Roses (Thierry Delzescaux, Nicolas Souedet); ATHENA, Sofia-Antipolis (Rachid Deriche)

References

Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, Gramfort A, Thirion B and Varoquaux G (2014) Machine learning for neuroimaging with scikit-learn.*Front. Neuroinform*. **8**:14. doi: 10.3389/fninf.2014.00014

Chupin M, Hammers A, Liu RS, Colliot O, Burdett J, Bardinet E, Duncan JS, Garnero L, Lemieux L (2009) Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. *Neuroimage* 46:749-761.

Cointepas, Y., Geffroy, D., Souedet, N., Denghien, I., Rivière, D., Roses, F. (2010). The BrainVISA project: a shared software development infrastructure for biomedical imaging research. *Proceedings 16th OHBM*.

Colliot O, Chetelat G, Chupin M, Desgranges B, Magnin B, Benali H, Dubois B, Garnero L, Eustache F, Lehericy S (2008) Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. *Radiology* 248:194-201.

Ferguson, AR, Nielson, JL, Cragin, MH, Bandrowski, AE, & Martone, ME (2014). Big data from small data: data-sharing in the 'long tail' of neuroscience. *Nature neuroscience*, *17*(11), 1442-1447.

Fischl, B. (2012). FreeSurfer. *Neuroimage*, *62*(2), 774-781.

Frisoni, GB, Redolfi, A, Manset, D, Rousseau, MÉ, Toga, A, & Evans, AC (2011). Virtual imaging laboratories for marker discovery in neurodegenerative diseases. *Nature Reviews Neurology*, *7*(8), 429-438.

Guevara P, Duclap D, Poupon C, Marrakchi-Kacem L, Fillard P, Le Bihan D, Leboyer M, Houenou J, Mangin JF (2012) Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. *Neuroimage* 61:1083-1099.

Laguitton, S., Rivière, D., Vincent, T., Fischer, C., Geffroy, D., Souedet, N., Denghien, I., and Cointepas, Y. (2011). "Soma-workflow: a unified and simple interface to parallel computing resources," in MICCAI 2011 Conference, eds T. Peters, G. Fichtinger, and A. Martel (Toronto: Springer LNCS)

Mangin JF, Jouvent E, Cachia A (2010) In-vivo measurement of cortical morphology: means and meanings. *Curr Opin Neurol* 23:359-367.

Mangin JF, Riviere D, Cachia A, Duchesnay E, Cointepas Y, Papadopoulos-Orfanos D, Collins DL, Evans AC, Regis J (2004) Object-based morphometry of the cerebral cortex. *IEEE Transactions on Medical Imaging* 23:968-982.

Martini, JB, Habert, MO, Yeni, N, Giron, A, Gharrad, I, Hartmann, A, ... & Kas, A (2014). Large-scale validation of a computer-aided quantification for 123I-FP-CIT images. In *Society of Nuclear Medicine Annual Meeting Abstracts* (Vol. 55, No. Supplement 1, p. 2037).

Operto, G., Fischer, C., David, S., Bouin, M., Fillon, L., Champseix, C., Mangin, J-F., SnapBase/SnapCheck: Assisting Quality Control of Post-processing Results over Large Cohorts (2015), Human Brain Mapping Conference, 2015

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (Eds.). (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images: The Analysis of Functional Brain Images*. Academic Press.

Poldrack, R. A., & Gorgolewski, K. J. (2014). Making big data open: data sharing in neuroimaging. *Nature Neuroscience*, *17*(11), 1510-1517.

Poline J-B, Breeze JL, Ghosh S, Gorgolewski KF, Halchenko YO, Hanke M, Haselgrove C, Helmer KG, Keator DB, Marcus DS, Poldrack RA, Schwartz Y, Ashburner J and Kennedy DN (2012) Data sharing in neuroimaging research. *Front. Neuroinform*. **6**:9.doi: 10.3389/fninf.2012.00009

Samaille T, Fillon L, Cuingnet R, Jouvent E, Chabriat H, Dormont D, Colliot O, Chupin M (2013) Contrast-based fully automatic segmentation of white matter hyperintensities: method and validation. *PLoS One* 7:e48953.

Sejnowski, TJ, Churchland, PS, Movshon, JA (2014) Putting big data to good use in neuroscience, *Nature Neuroscience*, 17, 1440–1441

Van Horn, JD., & Toga, AW. (2014). Human neuroimaging as a "Big Data" science. *Brain imaging and behavior*, *8*(2), 323-331.

Wallis JC, Rolando E, Borgman CL (2013) If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. PLoS ONE 8(7): e67332. doi:10.1371/journal.pone.0067332