Universidade de Aveiro
2013/14

Departamento de Eletrónica, Telecomunicações e Informática

Tiago Miguel Faria de Sousa

NEArBy: normalização lexical na pesquisa de imagens cerebrais com atlas

NEArBy: lexical normalization for Atlas enabled CBR System for Neuroimaging

Universidade de Aveiro
2013/14

Departamento de Eletrónica,
Telecomunicações e Informática

Tiago Miguel Faria de Sousa

NEArBy: normalização lexical na pesquisa de imagens cerebrais com atlas

NEArBy: lexical normalization for Atlas enabled CBR System for Neuroimaging

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia de Computadores e Telemática, realizada sob a orientação científica do Doutor José Maria Amaral Fernandes, Professor Auxiliar do Departamento de Engenharia, Telecomunicações e Informática da Universidade de Aveiro e co-orientação científica do Doutor Augusto Marques Ferreira da Silva, Professor Auxiliar do Departamento de Engenharia, Telecomunicações e Informática da Universidade de Aveiro

o júri

Presidente / president	Prof. Doutor António Manuel Melo de Sousa Pereira Professor Catedrático do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro
Vogais / examiners committee	Prof. Doutor João Paulo Silva Cunha Professor Associado com Agregação, Departamento de Engenharia Eletrotécnica e de Computadores, Faculdade de Engenharia, Universidade do Porto
	Prof. Doutor José Maria Amaral Fernandes Professor Auxiliar, do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro

agradecimentos

A realização desta tese de mestrado apenas foi possível graças a importantes apoios e incentivos, aos quais estou eternamente grato.

Em primeiro lugar quero deixar aqui bem presente os meus maiores agradecimentos às pessoas que me orientaram durante este percurso, salientando o Professor Doutor José Maria Amaral Fernandes e o Professor Doutor Augusto Marques Ferreira da Silva. Ambos transmitiram-me a confiança, o apoio, a orientação e a independência necessária para a produção de um trabalho sólido aliado a algumas ideias disruptivas.

Em segundo lugar, e porque a vida não é feita exclusivamente de trabalho, queria agradecer a todos os meus amigos e colegas que me apoiaram durante estes cinco anos proporcionando experiências e momentos que muito dificilmente viveria se eles não estivessem presentes.

Por fim, mas provavelmente o mais importante de todos, queria agradecer a toda a minha família por todos os esforços e pelo apoio incansável que me deram desde o meu primeiro ano de escolaridade. Dedico-lhes a todos eles, aos presentes bem como aos que já partiram, este trabalho pois sem a sua ajuda nunca conseguiria atingir o sucesso.

palavras-chave Neuroimagem, Atlas Cerebrais, DICOM, Dicoogle, Recuperação de dados por conteúdo, *Cloud*

resumo

Os atlas cerebrais têm sido usados como referências espaciais para classificar e catalogar informação topológica estrutural ou funcional, de imagens do cérebro. A informação semântica obtida a partir dos dados existentes da imagem é mapeada espacialmente de acordo com os descritores do atlas.

Porém o processo de classificação e catalogação de imagens cerebrais usando um atlas é um processo moroso e entediante, dependente na maior parte das situações de observação e validação humana. Para além disso, mesmo quando disponível, é um processo de difícil utilização, nomeadamente quando se recorrem a serviços de consulta e recuperação de informação em repositórios de imagens médicas (p.e, PACS com base DICOM).

Neste trabalho propomos a NEArBy, uma solução baseada em *cloud* que oferece serviços de consulta e recuperação de informação com base na semântica de atlas cerebrais, facilmente integrada em repositórios de imagem existentes, baseados em DICOM. Fazendo uso de uma interface *Web*, a NEArBy para além de suportar as pesquisas habituais de consulta e recuperação sobre os dados DICOM, permite ainda o uso de termos suportados em dicionários de altas cerebrais.

Para automatizar a catalogação semântica das imagens cerebrais recorremos a métodos externos para identificar características espaciais relevantes que são posteriormente catalogadas usando um atlas cerebral *standard*. Sendo o DICOM um *standard* baseado em *tags*, as características identificadas são embebidas em campos privados do ficheiro DICOM sob a forma de descritores NEArBy JSON usando vocabulário suportado em NeuroLex. Estes descritores codificam o mapeamento entre o tipo de característica, a localização espacial no atlas e as respetivas *tags* dos atlas. A codificação das *tags* de acordo com o formato JSON é bastante adequado para a indexação por parte de ferramentas de consulta e recuperação de imagem médica, como é o caso do Dicoogle, permitindo consultas baseadas quer nas habituais *tags* DICOM bem como nas novas *tags* incluídas pelo *middleware* NEArBy.

A NEArBy fornece assim uma nova forma de realizar consultas, que não são centradas no paciente, sobre repositório de imagens neurológicas usando informação técnica e topológica baseada em atlas. Ao longo da dissertação é apresentada a potencial utilização deste projeto num conjunto de imagens obtidas por ressonância magnética funcional (IRMf), utilizando a interface web do utilizador para formular as consultas em critérios relacionados com o atlas e aceder aos resultados daí recuperados.

keywords Neuroimaging, Brain Atlas, DICOM, Dicoogle, Content Based Retrieval, Cloud

abstract

Brain atlases have been used as spatial references to classify and tag either structural or functional topological information from brain images. Semantic information obtained from the existing image data is thus spatially mapped according the atlas descriptors. However the process of classifying and tagging brain images using an atlas is often tedious and mostly dependent on human observation and validation. At the same time, even when available, it is often difficult to use, particularly when using standard query and retrieve services in modern imaging repositories (e.g. DICOM based PACS).

In this work we propose NEArBy, a cloud based solution that provides query and retrieve services based on brain atlas semantics that can be easily integrated in existing DICOM based imaging repositories. Using a web interface, NEArBy supports not only typical DICOM query retrieve searches but also query tokens matching the brain atlas dictionary.

To automate the semantic tagging of the brain images we rely on external methods to identify relevant spatial features that are later labelled using standard brain atlas. Being DICOM a tag based standard, atlas related tags are then privately embedded into DICOM files as NEArBy JSON descriptors using lexicon as proposed in NeuroLex. These descriptors encode the mapping between feature type, spatial location in the atlas and the respective atlas tag. JSON encoded tags are also suitable for indexing by a medical imaging Q/R tool such as Dicoogle allowing queries based both on standard DICOM tags and specifically on atlas related tokens included by NEArBy middleware.

NEArBy provides a new way to perform non- patient centric queries over neuro-imaging repositories using technical and atlas based topological information.

During this dissertation, the NEArBy potential usage is illustrated over a set of functional magnetic resonance imaging (fMRI) datasets using the web user interface to formulate the queries with atlas related criteria and access the retrieved results.

Table of Contents

List of Figures

List of abbreviations and acronyms

AAL	Automated Anatomical Labelling
ACR	American College of Radiology
ADNI	Alzheimer’s Disease Neuroimaging Initiative
AFNI	Analysis of Functional NeuroImages
ANT	Advanced Normalization Tools
API	Application Programming Interface
BET	Brain Extraction Tool
BIRN	Biomedical Informatics Research Network
BLAST	Basic Local Alignment Search Tool
BOLD	Blood Oxygen Level Dependent
CBR	Content-Based Retrieval
CT	Computed Tomography
DB	Database
DG	DICOM Generator
DICOM	Digital Imaging and Communications in Medicine
DTI	Diffusion Tensor Imaging
FINDLab	Functional Imaging in Neuropsychiatric Disorders Lab
FLIRT	FMRIB’s Linear Image Registration Tool
fMRI	Functional Magnetic Resonance Imaging
FMRIB	Functional MRI of the Brain
FNIRT	FMRIB’s Non-Linear Image Registration Tool
FSL	FMRIB Software Library
GPL	General Public License
HBN	Human Brain Network
HBP	Human Brain Project
HGP	Human Genome Project
HTTP	Hypertext Transfer Protocol
IaaS	Infrastructure as a Service
iAtlas	Atlas Interface
ICBM	International Consortium for Brain Mapping
ID	Identifier
INCF	International Neuroimaging Coordinating Facility

IOD	Information Object Definition
IRB	Institute for Research in Biomedicine
IRP	Image Related Processes
JDK	JAVA Development Kit
JS	JavaScript
JSON	JavaScript Object Notation
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
NDAR	National Database for Autism Research
NEMA	National Electrical Manufacturers Association
NIF	Neuroscience Information Framework
NIFTI	Neuroimaging Informatics Technology Initiative
NIGMS	National Institute of General Medicine Sciences
NIH	National Institute of Health
NIHBNR	National Institutes of Health Blueprint for Neuroscience Research
Nipype	Neuroimaging in Python Pipelines and Interfaces
NPC	NEArBy Processing Core
OBART	Online Brain Atlas Reconciliation Tool
P2P	Peer-to-Peer
PACS	Picture Archiving and Communication System
PET	Positron Emission Tomography
PF	Plain Film
PUBMED	Public/Publisher MEDLINE
QR	Query Retrieval
RDBMS	Relational Database Management System
REST	Representational State Transfer
ROI	Region Of Interest
SDK	Software Development Kit
SOAP	Simple Object Access Protocol
SPM	Statistical Parametric Mapping
SQL	Structured Query Language
T&T	Talairach & Tournoux
THOR	Technology by Highly Oriented Research
TLV	Tag-Length-Value
URL	Uniform Resource Locator
VR	Value Representation

XML	Extensible Markup Language
XNAT	Extensible Neuroimaging Archive Toolkit

1. Introduction

The analysis of brain imaging datasets by neuroscientists requires the knowledge of the brain structure in order to classify the regions of interest and consequently extract useful information within the neuroscience context. This implies the thorough knowledge of the brain structure and/or often the support of a brain atlas when performing the analysis. The information extraction process is typically considered a high- level task that is indeed the top layer of a complex workflow that is technically grounded upon several image processing and pattern recognition steps. A dual perspective can then be broadly accommodated within the typical neuroimaging analytical workflow: an utilitarian perspective of pure information extraction taking for granted the adequacy of the lower level technical tools and a technical perspective focused on image handling and processing tools. While the former is most often concerned with the proper assessment of brain structure and/or function in clinical or research contexts the latter has to address all the technical issues of each and every processing tool and the ways that they can be assembled together leading to coherent workflows to the ultimate users.

1.1. Sharing information

In recent years the human genome project (Lander et al., 2001) was a clear example of the relevance of creating common and annotated data repositories to support a multi-centre research to cope with the inherent complexity of the human genome. This implied establishing common protocols (Cuticchia et al., 1993), standards (Shea, 2006) and tools, such as Basic Local Alignment Search Tool (BLAST) (Madden, 1990) among participating groups allowing that the information acquired, produced and deduced was useful for the common goal – the human genome. As a result now is possible to navigate and query the genome related information in services like PubMED to support both research and clinique (Roberts, 2001).

In the neuroscience field, the first steps are still underway and major projects like Human Connectome (Toga et al., 2012) and Human Brain Project (Kandel et al., 2013) were launched in the last decade. The overall meta objective is to gather and integrate multimodal information that now is possible to acquire using brain imaging techniques like diffusion tensor imaging (DTI) among other (Toga et al., 2006). However, in contrast with the Genome project, there are many diverse hurdles.

The increasing number of brain imaging studies and the quest for results dissemination and reproducibility of analytical outcomes raises several challenges for the efficient secondary usage of clinical or research imaging information (Van Horn & Toga, 2014). Most of the primary studies are carried away without propagation of imaging findings or semantic content to the actual dataset repository. Posterior data mining within standard PACS environments has to circumvent the fact that common query and retrieval (Q/R) services are restricted to demographic data related to the patient centric DICOM information model. So far, topological or semantic content is far from being seamless integrated with image data unless image database systems are specifically conceived (Marcus et al., 2007).

Besides the need of creating a culture of sharing datasets, sharing protocols, setups and processing workflows, there are several transversal problems:

- The use of different processing workflows and frameworks
- Establishing a common lexicon
- Annotate the datasets with "all" information in order to be useful in multi centre/ patient studies

Still, not even the whole community of neuroscience surrendered to the technology and data analysis is often made on a manual way. The analysis of a brain image from the neuroscientist requires the knowledge of the brain structure in order to classify the regions of interest and consequently extract useful information within the neuroscience context (Bohland et al., 2009). This implies the previous knowledge of the brain structure and/or often the support of a brain atlas when performing the analysis.

1.2. The role of the brain atlases:

In this context, brain atlases play a very important role in the data analysis. Given the fact that the atlas provides a way to map a particular brain structure it can be used to relate the information extracted from the image data with the structural or functional information.

Within its proper spatial context, an atlas provides a volumetric segmentation of anatomical structures where the different brain regions are properly labelled and can be useful as a reference source for feature annotation providing the basis for better assessment of brain structure and or function (Devlin & Poldrack, 2007). In practice an atlas can also establish a brain spatial reference – a template – for alignment and interpretation of the brain scans allowing not only a quicker labelling of image datasets but also allow a (sometimes coarse) spatial inter and intra subject comparison between different datasets and features extracted from brain imaging data (Devlin & Poldrack, 2007).

This relationship is commonly done manually when a neuroscientist surveys the brain image and often uses an atlas to map the different areas of the brain (Bohland et al., 2009). The information existing in the image data is merely information provided by the scanner. However, the digital repositories contain more information than the image data itself from the scan. This information, usually related with the acquisition process, has been proven to be extremely useful to provide solutions for tag based image retrieval (Valente, Costa & Silva, 2013).However, the atlas do not solve all the problems and when in addressing sets of datasets with same or different subjects, even when there is already a strong question or hypothesis to lead the analysis, the query and retrieve process can be a daunting task due to different brain topologies, the inter and intra personal variability during acquisition process just to name a few. In the recent interest (Van Horn & Toga,

2014) in building shared brain imaging repositories these issues gain more relevance in order to make them into useful tools and not only a repository of digital information that cannot be easily access and/or queried.

1.3. NEArBY

The initial orientation of this thesis was to extend an existing solution by João Lemos (NEArBy: Neuroimaging Atlas Based Q/R) with a Semantic Enabled mechanism for Neuro-Imaging repositories.NEArBy project tried to propose a solution based on tagging the datasets with relevant morphological information (from morphological atlas used by experts) that later on can be searched using neuroscience domain language extracted from the atlas. By using atlas labels it is possible to use standard QR solutions to perform structural queries over the repository - and not only based on imaging specific properties – allowing to find the area or nearby areas, hence the name NEArBy. The current solution relies on using DICOM (Pianykh, 2008) as a storage medium and a DICOM indexer (Dicoogle - http://www.dicoogle.com/) (Costa et al., 2011) for supporting the Query and Retrieve (Q/R) over the existing information. Nearby proposed an autonomous system that is able of merging two common approaches in the field of neuroimaging to support neuroscience research:

- Tag based Image Retrieval - nowadays the technical support for QR of datasets namely in Digital Imaging and Communications in Medicine (DICOM) based PACS ensure good options to quickly retrieve classified/tagged information.
- Atlas based Neuroimaging - using reference maps/atlases to classify and extract topological information is popular in Neuroimaging and, in most of the times, is the quickest way of relating findings in data to topological related information either structural and/or functional.

1.4. Synopsis of the thesis

Through this dissertation you may find the evolutions introduced to the previous version of NEArBy, which allowed the inclusion of far more suitable search engines for experts in the field of neuroscience. In order to develop a system with the features that we promote, it was necessary to extend the following aspects:

- NEArBy original architecture and deploy it in the cloud: Proposing and implementing a modular and flexible architecture to support the initial NEArBy proof of concept. Besides establishing clear roles for each node in the system we specified and implemented normalized interfaces based on webservices and REST interfaces. The roles include atlas labelling service, feature extraction service and preprocessing.
- Introduced normalized labelling based on the accepted NeuroLex lexicon (Larson, 2012) and existing available atlas services. As a proof of concepts labelling services based on AAL and Broadman atlas were implemented.

- Introduced a web-based visualizer of NIFTI files.

From this point onwards NEArBy will refer to the new extended version unless stated otherwise.

1.5. Thesis structure

In chapter 2, we present an overview of the State of the Art. The focus will be on the brain imaging / eScience repositories, the role of the templates and atlases, the spatial registration / normalization, the available mechanisms for handling and QR image repositories and, to conclude, we discuss the need to establish a common nomenclature in the neuroscience field. In chapter 3, NEArBy is presented, a turnkey solution that unifies all the aforementioned issues. After making a brief presentation of the reasons that led us to make this proposal is given a global perspective of the system architecture, discussing the role played by each of the components that make up the system. Through this chapter the solutions adopted are grounded, stating the reason we opted for a deployment type over another. To conclude, we make a presentation of the overall workflow, demonstrating how the components interact in order to solve the raised problem. In chapter 4, we tried to find a case where the application of our system may introduce an added value. For this purpose, we studied the possibility of automating a process closely related with the fMRI. In order to contextualize the reader, we make a brief introduction of the most relevant aspects of the fMRI, such as the activations/deactivations. In addition we describe some of the available mechanisms for the fMRI pre-processing. Finally, we demonstrate the practical application of our system in this field, presenting the obtained results. In chapter 5, it will be presented a brief overview of the work as well as the main conclusions. We also discuss a possible framework extension for future developments.

2. State of the Art

In this chapter, we present an overview of the State of the Art. Initially the focus will be on the brain imaging / eScience repositories emphasizing the need to establish standards for storage and processing of medical images. We also addressed the role of the templates and atlases in the spatial registration / normalization. To conclude we will present the available mechanisms for handling and QR image repositories and discuss the need to establish a common nomenclature in the neuroscience field.

2.1. Brain imaging / eScience Repositories

One of the biggest advances witnessed in recent years concerns the sharing of medical data, making use of the so-called eScience Repositories (Nicholas et al., 2012). The data sharing efforts increasingly contribute to the acceleration of scientific discovery. In fact, many of the major new discoveries in the genetics of schizophrenia, diabetes, obesity and other metabolic traits have been possible only through collaborative data sharing (Ripke et al., 2011; Speliotes et al., 2010).In the specific case of Neuroimaging, repositories are growing in an astonishing way, accumulating in distributed domain-specific databases, rather than in a small number or even in a unique central repository, without any standardization at the data access level.

Moreover there is currently no universally accepted integrated access mechanism nor formats for the encoding of critically important (meta)data in order to make use of it (Keator et al., 2013).There are neuroimaging repositories that require some form of credentials to access the data and other that have completely open-access approach. For the first there are some examples like Institute for Research in Biomedicine, IRB, the National Database for Autism Research (NDAR) which contains data from over 6000 subjects and the Alzheimer's Disease Neuroimaging Initiative (ADNI) which contains imaging data from over 800 subjects. XNAT Central (central.xnat.org) is a good example for the open-access neuroimaging repositories, which includes over 3000 subjects stored in the XNAT database (Marcus et al., 2007). Other examples are the 1000 Functional Connectomes project (http://fcon_1000.projects.nitrc.org/) (Biswal et al. 2010) which at the present time, contains over 1000 subjects, the relatively new OpenFMRI (Poldrack et al., 2013) repository which contains imaging data from over 200 subjects, and the BIRN data repository which includes large cohorts of both mouse and human imaging data stored in the BIRN Human Imaging Database (Florescu et al., 1996; Ozyurt et al., 2010).

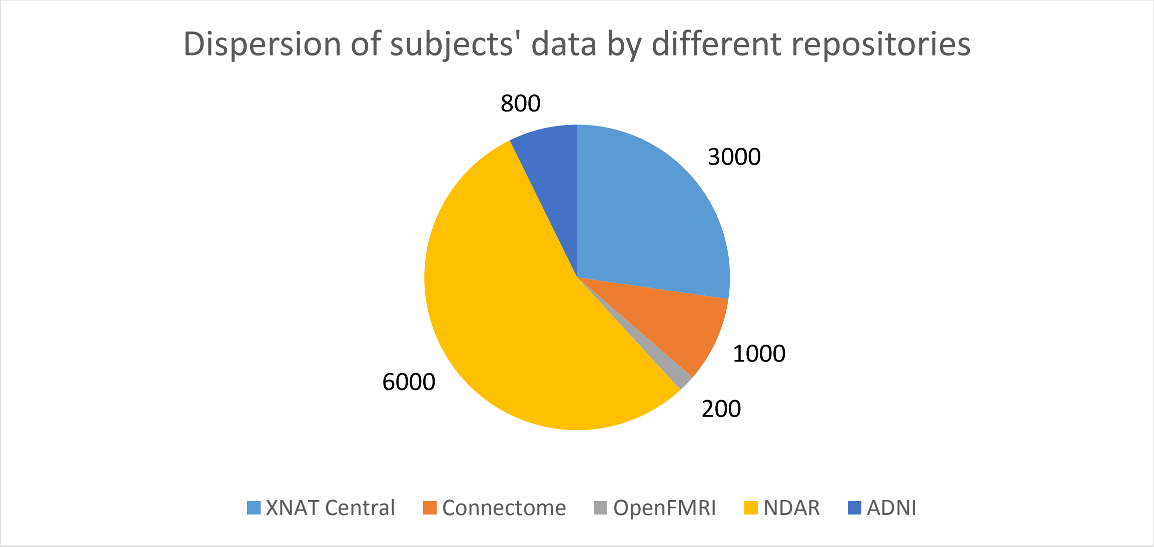


Figure 1 - Dispersion of data based on available data on neuroimaging. Supported on the (approximate) numbers reported by each repository we built this graphical representation - comprising over 11000 cases/subjects.

Several organizations are trying to address the data dispersion found in different neuroimaging repositories (Figure 1). This data is growing continuously mainly because neuroimagers are now collecting more information in a few days than was collected in over an entire year just a decade ago. Due to that, it is now acceptable to say that human neuroimaging is now considered a “big data” science (Van Horn & Toga, 2014).

Despite the substantial growth in data that is being placed in different repositories of public access, there are still plenty of issues with still no concrete answers: how to manipulate, store, analyse and share this data (Van Horn & Toga, 2014)?

A good example of an entity that has made a huge contribution in this area is the Biomedical Informatics Research Network (BIRN) (http://www.birncommunity.org/). The BIRN is an American initiative to advance biomedical research through data sharing and online collaboration. This entity was funded by the National Institute of General Medicine Sciences (NIGMS), a component of the US National Institute of Health (NIH), and it provides data-sharing infrastructure, software tools, strategies and advisory services.

Together with the International Neuroimaging Coordinating Facility (INCF) (http://www.incf.org/), the Derived Data Working Group (Keator et al., 2013) (an open –access group sponsored by the BIRN) has been directing efforts on developing models and tools for facilitating the structured interchange of neuroimaging metadata exchange (Keator et al., 2013)

Among the wide range of tools, one of which form part of the BIRN catalog, is the Extensible Neuroimaging Archive Toolkit (XNAT) (Marcus et at., 2007), an open source software platform designed to facilitate management and exploration of neuroimaging and related data. XNAT provides a web-based data management for diverse data sets. Making use of data structures for persistence both to define data types and to automatically generate the data store, XNAT defines relationships between different data types thus enabling complex queries. In addition to that, it provides the user with visualization tools, methods for data validation and integrity-checking, the ability to export data into a number of convenient formats and security features. Having at our disposal an aggregation tool as XNAT, it is reasonable to create full-services for queries over public data exploring the panoply of metadata that can be extracted from the datasets. This metadata consists of descriptive elements associated with data that provide additional clarity regarding acquisition parameters, experimental conditions, analysis procedures, and any other formation about the experiment or analyses that help the understanding of the data (Keator et al., 2013). However there is still no standard format that allows sharing data and metadata in a structured and consistent or tools to allow performing queries on the different existing databases especially in case of small laboratories or even sole proprietorships (Poline et al., 2012).

Also efforts like the Neuroscience Information Framework (NIF) (http://www.neuinfo.org/) (Gupta et al., 2008) – now supported by NIH, proposed a solution where they intend to provide "a dynamic inventory of Web-based neuroscience resources" including tools, standards for data annotation and ontology services accessible to publically available NIF interfaces for both human users and web enabled application.

2.2. Brain coordinates systems and Atlas

Neuroimaging techniques evolved over the last 20 years have allowed neuroscientists to re-visit the issue of mapping the human brain, such that a modern brain atlas is now expressed as a digital database that can capture the spatio-temporal distribution of a multitude of physiological and anatomical metrics. Together with the availability of better imaging techniques, brain mapping methods and analytical strategies has the potential to revolutionize the concept of the brain atlas (Toga et al., 2013).

The main goal of the brain atlas and templates is the use of a standardized 3D coordinate frame for data analysis and reporting of findings from neuroimaging experiments. Thanks to that it is feasible to compare and/or combine brain-mapping findings from different imaging modalities and laboratories around the world (Evans et al., 2012).

2.2.1. Brain coordinate space

Works like the one developed by Jean Talairach (Talairach et al., 1967), where relevant to establish the need of a common 3D coordinate space to assist deep-brain surgical techniques that later on supported the definition of a printed atlas for guidance of deep-brain stereotactic procedures – the well-known Talairach and Tournoux Atlas (Talairach & Tournoux, 1988). The earliest application of Talairach space for brain mapping was by Fox who used it to map the 3D coordinates of activation foci from PET experiments in different individuals (Fox et al., 1985).

A brain coordinate system objective is to establish a map of any location in the brain to a given coordinate independent from individual differences in the size and overall shape of the brain.

There are two main brain coordinate systems in use today: the Talairach space and the MNI/ICBM space.

The Talairach space was first created by neurosurgeons Jean Talairach and Gabor Szikla in the 1967, to define a standardized grid for neurosurgery localization. This space was then used to describe the global features of the brain in the 1988 Talairach Atlas.

The Talairach space carries with it a Cartesian reference frame (Figure 2), and all measurements (positions, distances, sizes, angles, and shapes) are made in this space. The most common usage is to report locations in the brain with x-y-z Talairach coordinates.

It is ideal for global spatial normalization, and this brain is used by the majority of brain mapping centers as the standard for spatial normalization. A brain image that conforms to the global spatial features of this standard brain is said to be Talairach spatially normalized and registered in Talairach space.

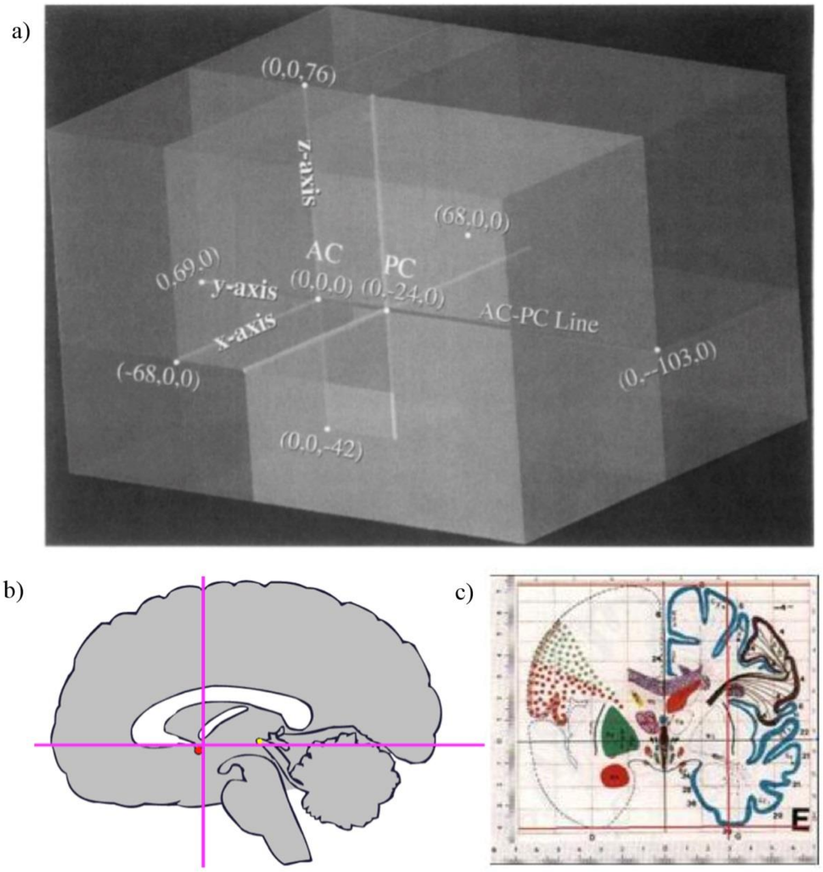


Figure 2 – a) The twelve regions of the 1988 Talairach atlas, grouped within the bounding box of the brain. Talairach coordinates are given for the AC and PC (adapted from Bankman, 2000). b) The diagram shows the position of the AC (red dot) on a midsagittal view (adapted from Rorden, 2002). c) Talairach Atlas, through which normalizations are performed so that neurologists may later make a uniform transmission of results in investigation studies (adapted from http://www.haogongju.net/art/2589297).

The MNI/ICBM space has been widely adopted in the last decade, an alternative to “Talairach space”. The MNI/ICBM reference is based on the initial MNI305 dataset (Evans et al., 1992) – MNI stands for Montreal Neurological Institute and ICBM for International Consortium for Brain Mapping. The MNI305 dataset (Figure 3), was created by first aligning a set of 305 structural MRI images to the Talairach atlas using landmark-based registration, creating a mean of those images, and then realigning each image to that mean image using a nine-parameter affine registration (Evans et al., 1993). Subsequently, another template, known as ICBM-152, was developed by registering a set of higher-resolution images to the MNI305 template. Versions of the ICBM-152 template are included with many of the major neuroimaging software packages. It is important to note that there are slight differences between the MNI305 and ICBM-152 templates, such that the resulting images may differ in both size and positioning depending upon which template is used (Lancaster et al., 2007).

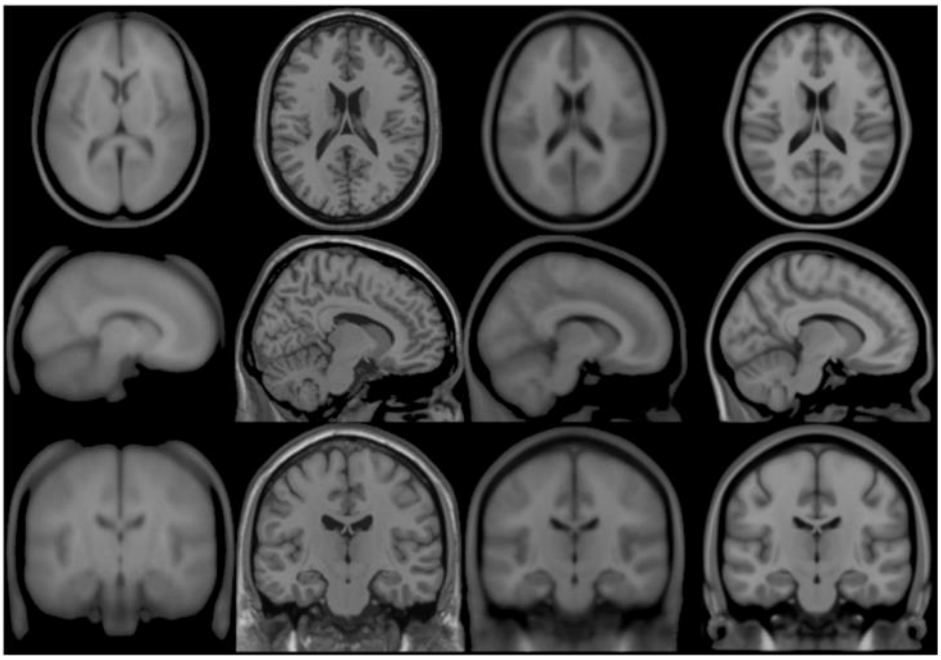


Figure 3 - Some phases of the evolution of the MNI template: the MNI305, Colin27, MNI152 (a.k.a. ICBM152), 40th Generation MNI152 and the ICBM 452, respectively from left to right.

The Talairach atlas and the stereotaxic methodology had a number of drawbacks. The atlas was generated from two series of sections from a single 60-year old female brain. One half was sectioned in the sagittal plane and the other in the coronal plane. The transverse images in the atlas were manually approximated from the information obtained in the sagittal and coronal planes. Left-right hemispheric asymmetry was ignored. This presented no problem for the original purpose of the atlas in guiding deep brain surgery, but it causes problems when employed for cortical analyses. As normal brains exhibit a left-right asymmetry there are many areas within the atlas itself that are not self-consistent between the three-views, eg.: area 44 and area 9. This is just one example of several problems that this template presents.

With regard to spatial normalization, a major problem is that there is no MRI scan available for the individual on whom the atlas is based, so an accurate MRI template cannot be created. This means that normalization to the template requires the identification of anatomical landmarks that are then used to guide the normalization. As described in the “Handbook of Functional MRI Data Analysis” (Poldrack et al., 2011), such landmark-based normalization has generally been rejected in favour of automated registration to image-based templates.

It was mainly thanks to these vulnerabilities, MNI template, more precisely, the ICBM-152 is currently the most common reference space used as an MRI-based template to support automated registration in brain imaging.

2.2.2. Brain atlas and templates

An atlas provides a guide to the location of anatomical characteristics in a coordinate space. (Poldrack et al., 2011). Brain atlases, regardless of the several existing ones, present a “physical” segmentation of the brain structure (Figure 4) – a natural physical reference for a kind of brain related information. Thus, atlases can be used as an element for localization of topological structures and interpretation of results. (Heckemann, Hajnal, Aljabar, Rueckert, & Hammers, 2006; Poldrack, Mumford, & Nichols, 2011; Thompson et al., 2000; Toga, 1998). The Anatomical Automated Labelling (AAL) (Tzourio-Mazoyer et al., 2002) and Brodmann Area are provided as part of MRIcro (http://www.mccauslandcenter.sc.edu/mricro/) that are among the most popular atlases.

However selecting an atlas depends on the purpose of the anatomical analysis, mainly the level of detail in specific areas like the basal ganglia (Abbas et al., 2011; Bardinet et al.) and the expected functional reasoning behind each of the different atlas areas (Shirer, Ryali, Rykhlevskaia, Menon, & Greicius, 2011). Within its proper spatial context an atlas provides a volumetric segmentation of anatomical structures where the different brain regions are properly labelled and can be useful as a reference source for feature annotation providing the basis for better assessment of brain structure and or function (Devlin and Poldrack, 2007). It can support a quantitative characterization of the (i) normal variability in those metrics across a population, (ii) the relationship between those metrics and behavioral performance and (iii) the detection of subtle changes associated with disease, gender or demographics (Evans et al., 2012).

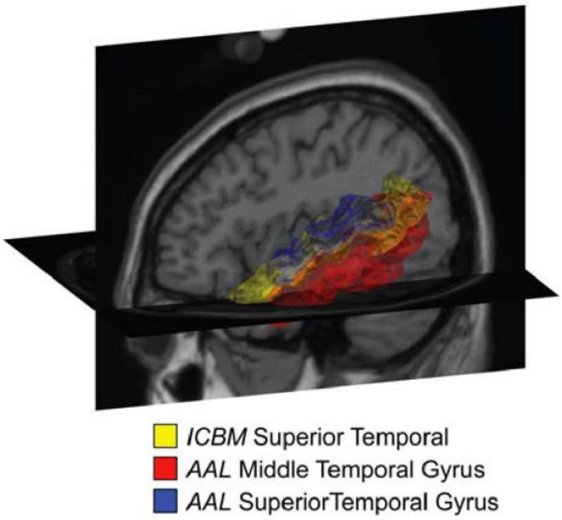

🟨 *ICBM* Superior Temporal
🟥 *AAL* Middle Temporal Gyrus
🟦 *AAL* SuperiorTemporal Gyrus

Figure 4 – Segmentation of three anatomical regions according two different atlases (adapted from Bohland et al., 2009).

In practice an atlas also establishes a brain spatial reference – a template – for alignment and interpretation of the brain scans allowing not only a quicker labelling of image datasets but also allow a (sometimes coarse) spatial inter and intra subject comparison between different datasets and features extracted from brain imaging data. A template is an image that is representative of the atlas and provides a target to which individual images can be aligned. It can comprise an image from a single individual or an average of a number of individuals. Whereas atlases are useful for localization of activation and interpretation of results, templates play a central role in the spatial normalization of MRI data.

Taking into account the brain atlas concordance problem, an issue that stems from difficulties and differences in the description of brain structures, and that presents certain obstacles for the neuroscience research community (Bohland et al., 2009), in some projects it is common to include various atlases in order to improve the achieved accuracy.

2.2.2.1. Talairach atlas

The best-known brain atlas is the one created by Talairach (1967) and subsequently updated by Talairach & Tournoux (1988). This atlas (Figure 5) provided a set of sagittal, coronal, and axial sections that were labelled by anatomical structure and Broadmann’s areas.

Once data have been normalized according to Talairach’s procedure, the atlas provides a seemingly simple way to determine the anatomical location at any particular location.

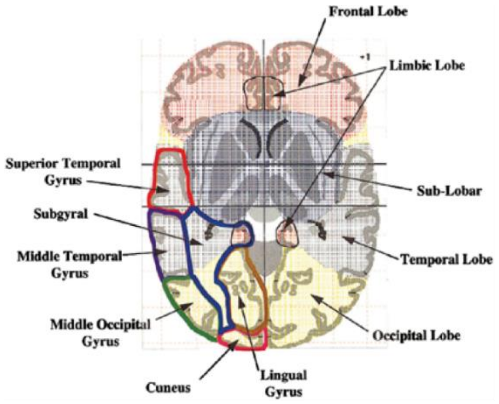


Figure 5 – The Talairach Atlas, with its associated labels (adapted from Johnson et al., 2005).

An entity that established efforts to make the access to this atlas to the all community, enabling to conduct research over it, was the THOR (Technology by Highly Oriented Research) Center (http://cogsys.imm.dtu.dk/thor/).

The THOR Center for Neuroinformatics Human Brain Project Repository, was established April 1998 at the Department of Mathematical Modelling, Technical University of Denmark. Besides pursuing independent research goals, the THOR Center hosts a number of related projects concerning neural networks, functional neuroimaging, etc. In what concerns the neuroimaging, they developed the Brede Database, a neuroinformatics database (http://neuro.imm.dtu.dk/services/brededatabase/). It contains taxonomies of brain functions and brain regions as well as results from neuroimaging experiments. Results from meta-analysis of the data are presented, and search engines can, e.g., retrieve nearby brain regions based on the given Talairach coordinates.

2.2.2.2. Automated Anatomical Labelling atlas

Anatomical Automatic Labeling (AAL) (Tzourio-Mazoyer et al., 2002) is a package for the anatomical labeling of functional brain mapping experiments. It is an in-house package made by Neurofunctional Imaging Group (GIN, UMR6095, CYCERON, Caen, France), which is available to the scientific community as a copyright freeware under the terms of the General Public Licence (GNU).

This project began in the nineties with the construction of a set of rules to be used for the anatomical parcellation of the brain according to major sulci and gyri. These rules were applied to build an anatomical parcellation of the spatially normalized single subject high resolution T1 volume provided by the Montreal Neurological Institute (MNI) (Collins et al., 1998).

It is typically used in functional neuroimaging-based research to obtain neuroanatomical labels (Figure 6) for the locations in 3-dimensional space where the measurements of some aspect of brain function were captured.

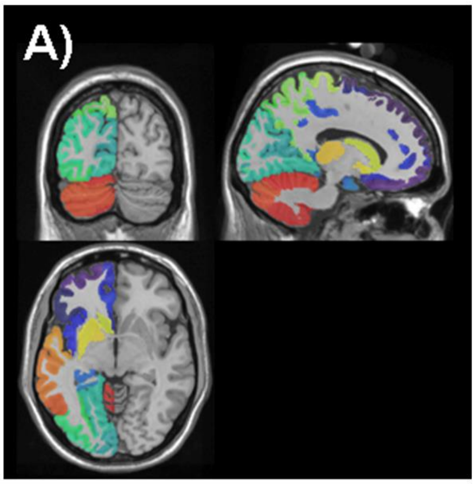


Figure 6 – Sections through the AAL parcellation of the test brain with different colors indicating different parcels (adapted from Bohland et al., 2009).

2.2.2.3. Functional atlas

Although atlases are usually associated with brain structure there are works that focus on building functional atlas of the brain i.e. map specific function to parts of the brain (Figure 7). Although this is the genesis of already existing brain atlas e.g. Broadmann, with new technologies like fMRI some new atlas are being proposed.

Examples of entities that have produced advances in this area is the case of the National Institutes of Health Blueprint for Neuroscience Research (NIHBNR) and the Functional Imaging in Neuropsychiatric Disorders Lab (FINDLab) from the Stanford University.

The NIHBNR is investing in the Functional ROI Atlas, an effort to provide a set of quasi-probabilistic atlases for established "functional ROIs" in the human neuroimaging literature.

Shirer et al. (Shirer et al., 2011) proposed a functional atlas which characterizes the underlying brain networks that support cognitive and emotional processing using magnetic resonance imaging (MRI) measures of functional and structural connectivity. The objective is to establish a reference based on the performance in healthy subjects, to examine how these networks are altered in several neuropsychiatric disorders including Alzheimer's disease and other dementias, depression, anxiety, coma, and chronic pain.

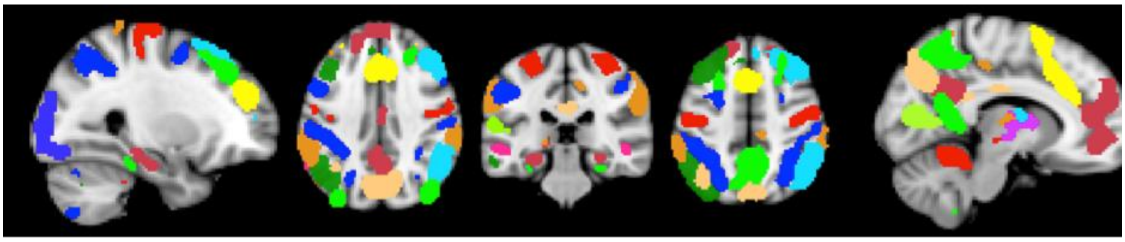


Figure 7 – All functional regions of interest according to the FINDLab (adapted from http://findlab.stanford.edu/functional_ROIs.html). (Shirer et al., 2011)

2.3. Spatial registration / Normalization

Any study that involves comparison between brain structure, of the same subject or from different subjects, differs in size and shape. In addition to that, if we add the fact that this information could be generated from different scans, it is critical that data be integrated across individuals. However, individual brains are highly variable in their size and shape, which requires that they first be transformed so that they are aligned with one another. The process of spatially transforming data into a common space for analysis is known as intersubject registration or spatial normalization (Ashburner and Friston, 1999).

The objective of registration is to map relevant features in the input image that can be easily mapped into the target image. When the target space is an atlas space, this registration is often referred as normalization (i.e. spatial co-registration with the atlas space). Through normalization relevant features in the input image can be easily mapped into the atlas space and labelled afterwards. One goal of spatial normalization is to map different human brain scans so one location in one brain scan corresponds to the same location in another brain scan (Poldrack et al., 2011).

After spatial normalization, the input data is resampled using linear and/or non-linear methods to interpolate it to the same space of the target image - producing a transformed dataset with the same orientation and referential usually presenting the same number and size of voxels for easier comparison (Poldrack et al., 2011).

With neuroimaging data there many initial steps are required both to organize and to prepare the data for analysis. With neuroimaging data, there are a series of image processing steps (Figure 8) required before any statistics are performed. These steps can include image registrations, transformations and filtering operations (Keator et al., 2009 and Astrakas et al., 2010). These pre-processing steps are common to all neuroimaging data modalities, be it Positron Emission Tomography (PET) or structural and functional Magnetic Resonance Imaging (MRI).

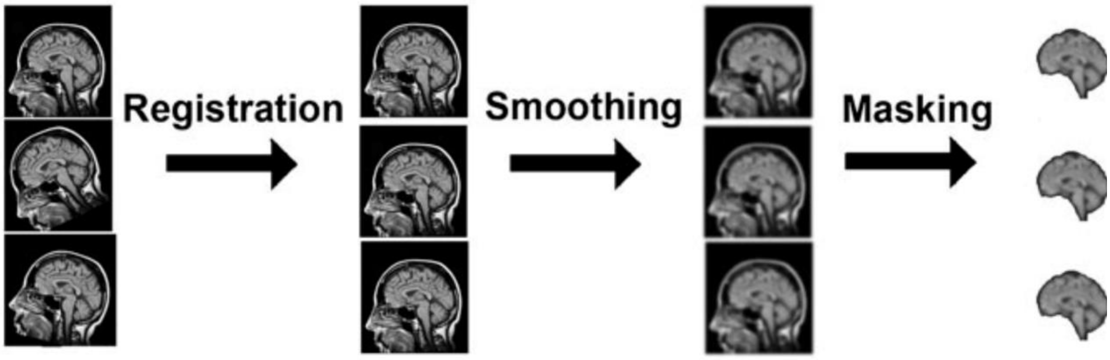


Figure 8 - Spatial normalization of brain images. Registration, smoothing and masking are described as pre-processing of data (adapted from Astrakas et al., 2010).

As we mentioned before, in order to perform the normalization a common space is required (Figure 9), i.e.,a template t is required, for example the 'MNI space' or the 'T&T Space'. According to Devlin and Poldrack (Devlin & Poldrack, 2007), it is recommended the adoption of the first rather than the second, because of the several well-known limitations that was mentioned before.

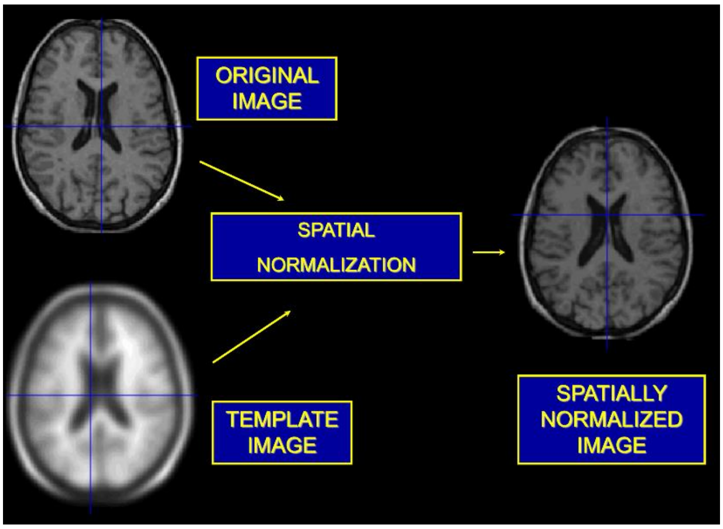


Figure 9 - Spatial normalization according to a template image (adapted from http://neurometrika.org/sites/default/files/uploads/images/2011JUN%20SPM/7%20Spatial%20Normalization.pdf).

The registration to the MNI space can be accomplished using various templates (Figure 3) and a variety of registration algorithms, including linear (for example, FLIRT (Bohland et al., 2009)) or non-linear transformations (for example, SPM software (Evans et al., 2012)).

Likewise, the registration to the T&T (Talairach & Tournoux) space can be done by making use of multiple templates and multiple registration algorithms (for example, AFNI (Evans et al., 2012)).

Despite this range of spaces and registration algorithms, some studies (Van Essen et al., 2007) showing that there is some discrepancy between each of the used methodologies (Figure 10).

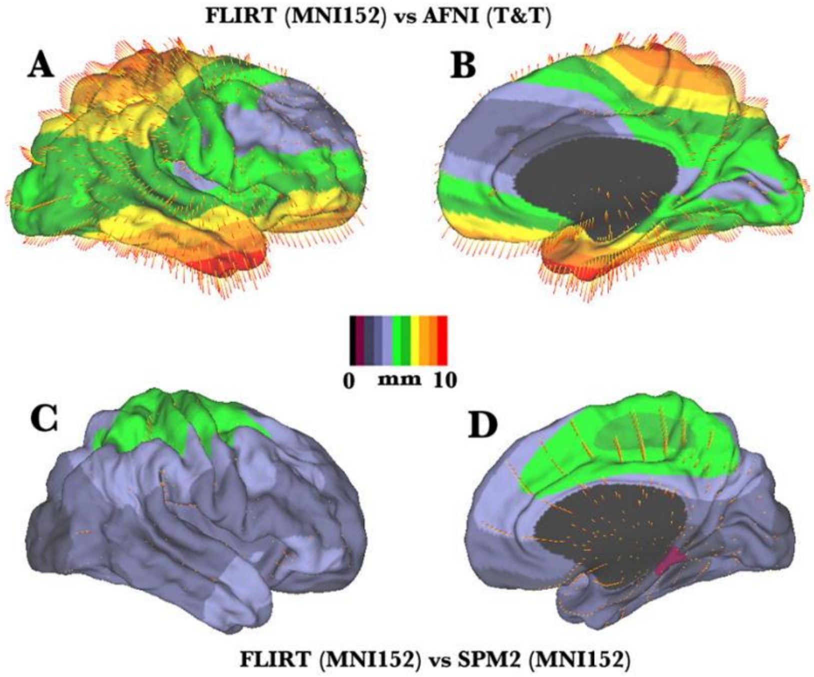


Figure 10 - Example of quantitative maps of the difference between stereotaxic spaces and registration processes (adapted from Van Essen et al., 2007).

Focusing on the MNI coordinate space, there are several tools namely SPM's and FSL's solutions that rely on the ICBM-152 brain template. In the next subsection we will present in more detail the FSL example.

2.3.1. The FSL example

One of the most popular solutions for spatial normalization is FSL (the FMRIB Software Library). FSL (Jenkinson et al., 2012) is now a mature a package and contains a comprehensive library of analysis tools for functional, structural and diffusion MRI brain imaging data, written mainly by members of the Analysis Group, FMRIB, Oxford.

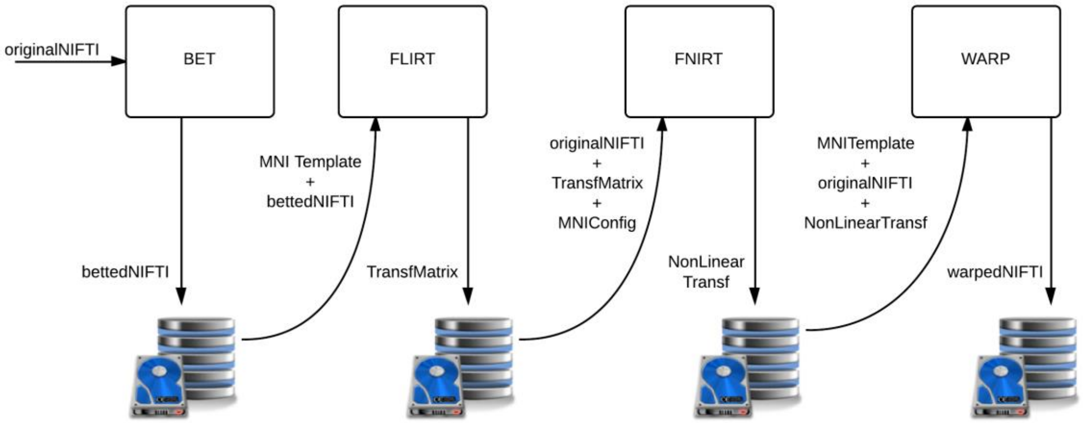

originalNIFTI
BET
FLIRT
FNIRT
WARP
MNI Template + bettedNIFTI
originalNIFTI + TransfMatrix + MNIConfig
MNITemplate + originalNIFTI + NonLinearTransf
bettedNIFTI
TransfMatrix
NonLinear Transf
warpedNIFTI

Figure 11 - FSL Spatial Normalization Sequence. This flow corresponds to the graphical representation of the sequence presented on the website: 'http://nuclear-imaging.info/site_content/2011/03/30/normalization-in-fsl/'. The original file (NIFTI) is submitted to an operation responsible for the extraction of brain (BET). Following that, the resulting file together with the MNI Template is submitted to a second operation responsible for the Linear Image Registration (FLIRT). Even with a good approximation provided by FLIRT, it is advisable to apply another nonlinear operation (FNIRT) in order to obtain more accurate results. To conclude, the resulting file from the FNIRT operation together with the original file (NIFTI) and the MNI Template are submitted to a warp (WARP) operation, originating the standard file that can be later analyzed.

The FSL normalization process is well defined and consists of the following sequence (Figure 11):

- Brain Extraction (BET):
- FLIRT (FMRIB's Linear Image Registration Tool):
- FNIRT (FMRIB's Non-Linear Registration Tool):
- WARP:

Some pre-processing streams include the step of brain extraction (also known as skull-stripping or BET). Removal of the skull and other non-brain tissue can be performed manually, but the procedure is very time consuming. Fortunately, a number of automated methods have been developed to perform brain extraction namely BET provided by FSL (Jenkinson et al., 2012). It should be noted that the brain extraction problem is of greater importance (and greater difficulty) for anatomical images (where the scalp and other tissues outside the brain have very bright signals) than for functional MRI data (where tissues outside the brain rarely exhibit bright signal). It removes non-brain tissues with highly variable contrast and geometry (e.g. scalp, marrow, etc.), and works with a wide range of MRI sequences (T1, T2, etc.) and resolutions (Figure 12).

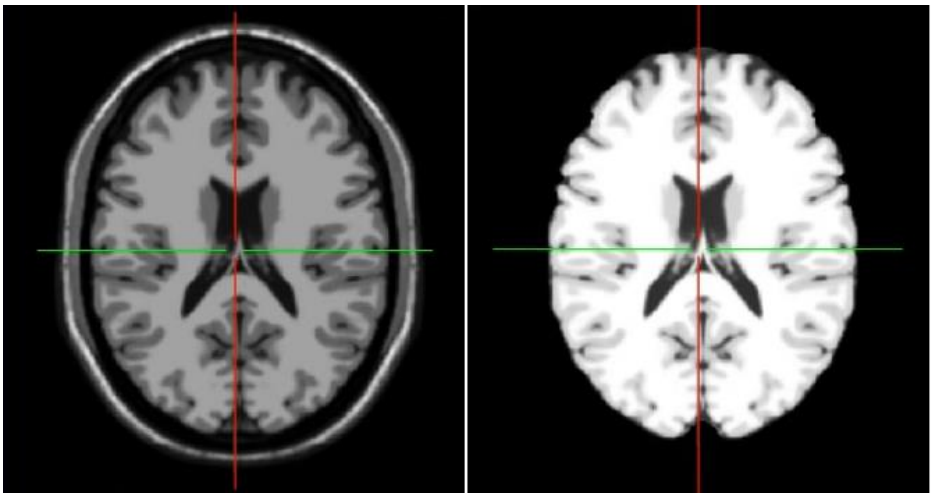


Figure 12 - Brain Extraction Example. On the left side we have the original image and on the right side the resulting image of the BET. The great difference is that the image on the right shows only the brain tissue.

Source: ‘http://sourceforge.net/apps/mediawiki/gimias/nfs/project/g/gi/gimias/8/89/BrainExtractionCLP.png’

Registration algorithms can be divided into linear and non-linear depending on the type of deformations they permit. FLIRT (Jenkinson and Smith, 2001) is an example of software that performs linear registration, meaning that it will translate, rotate, zoom and shear one image to match it with another (Figure 13).

Sometimes the differences between subjects are such that a linear transform is not sufficient to achieve good registration. The local deformations permitted by a non-linear method may then do a better job. The FNIRT tool (Andersson et al., 2010) includes the typical steps required for non-linear registration (Figure 13).

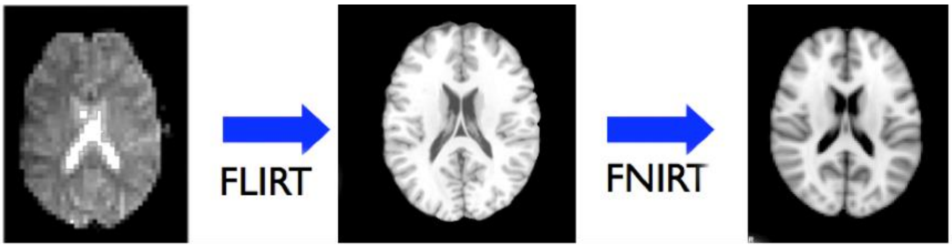


Figure 13 - Linear and Non-Linear Transformations. On the left side we have the resulting image from the BET operation. At the center we have the resulting image from the FLIRT operation, applied to the resulting image of BET. It corresponds to the linear transformation. On the right side we have the non-linear transformation, corresponding to the resulting image from the FNIRT operation (adapted from the FSL Course: ‘http://fsl.fmrib.ox.ac.uk/fslcourse/lectures/reg.pdf’).

The WARP is used to apply the warps or deformation field estimated by FNIRT (or some other software) to some image. The ultimate goal of this transformation is to obtain an image in a common space.

2.4. Handling and QR image repositories

One of the highlights of joining QR mechanisms and image repositories is the possibility of conduct research in a simple way, over datasets that have useful information. To achieve this result, it is necessary to use both specific files for storage as well as advanced search tools that enable the extraction of data from those same files. Having a system with these characteristics, it is then possible to provide a platform where the users can build queries that are natural in neuroscience.

Handling of the medical images, besides being a concern for neuroimaging repositories is also a concern for the industry when dealing with medical imaging repositories in a healthcare information system. In the vast majority of the cases a Picture Archiving and Communication System (PACS) is the solution that provides the access and economical storage of images from multiple modalities .

2.4.1. PACS

Pictures Archiving and Communication System (PACS) are medical systems consisting of necessary hardware and software components designed and used to run digital medical imaging procedures (Pianykh, 2008). They have been widely deployed in healthcare institutions, and they now constitute a normal commodity for practitioners. These systems are responsible for the acquisition, storage, retrieval management, communication, distribution and presentation of medical imaging (Association; Hood & Scott, 2006; Rouse, June 2010). It represents the natural evolution from working with digital modalities (e.g. CT, US, MRI, CR) towards a global digital environment where the film-based activities are progressively being replaced by their digital counterpart (Osteaux, M. et al., 1996).

With the existence of a digital system able to transfer images, display the images across multiple workstations and a more organized management, information retrieval is potentially facilitated. This system makes the entire process of medical imaging analysis and reporting more efficient (Hood & Scott, 2006).

Although a PACS is quite expensive to deploy, the advantage of making more efficient the process of delivering image related data can reduce the time to obtain proper results and reduce the underlying costs of data transport t (Feng, 2011; Hood & Scott, 2006). But for this system to be able to support all the features it is necessary to assemble a set of costly hardware and software components. A PACS consists of three major components (Figure 14):

- The imaging modalities such as X-ray plain film (PF), computed tomography (CT) and magnetic resonance imaging (MRI)
- Workstations for interpreting and reviewing images
- Archives for the storage and retrieval of images and reports.

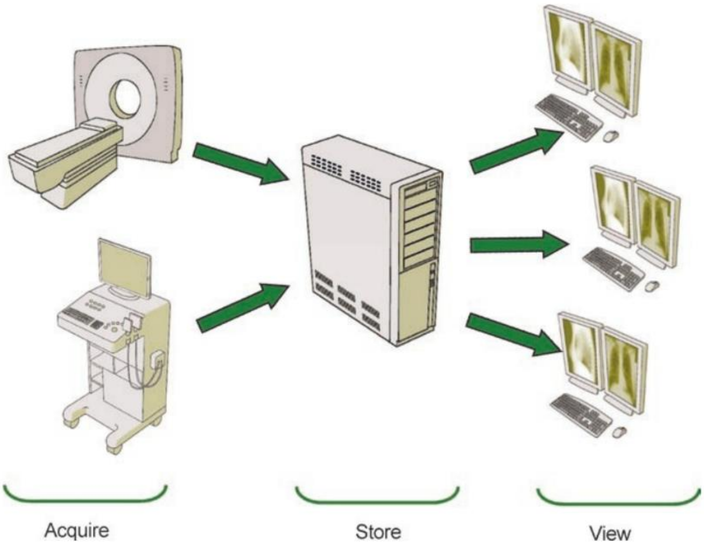


Figure 14 - PACS example architecture (adapted from Oosterwijk, 2005). In the left part of the figure are presented the devices responsible for acquiring medical images. These medical images are then stored on local servers or even in the cloud. Subsequently, queries may occur in different ways, such as in a graphical way through a computer or even through a tablet.

All these components are interconnected by a data network (Hovenga & Kidd, 2010; Pianykh, 2008).

The universal format for PACS image storage and transfer is based on the DICOM (Digital Imaging and Communications in Medicine) standard. It ensures the interoperability between the different system components. Imaging studies resulting from the patient scans are stored and transferred in DICOM format and can be handled whenever and wherever the network client software is required to operate.

2.4.2. DICOM

Digital Imaging and Communications in Medicine (DICOM) is the most universal and fundamental standard in digital medical imaging, providing all necessary tools for the diagnostically accurate representation and processing of medical imaging data (Pianykh, 2008).

This standard has emerged in order to solve a problem that was lasting for too much time, the exchanging of images between different kinds of equipment. Some efforts were made in the late eighties by groups from the ACR (American College of Radiology) and the NEMA (National Electrical Manufacturers Association), but the medical imaging community had to wait until 1993 to have a real usable standard available (Gibaud, 2008; Kagadis & Langer, 2011).

The DICOM is more than just an image or a file format, it enables the association of metadata that can cover all functional aspects of digital medical imaging, allowing the universal interoperability between medical imaging equipment (Bidgood Jr et al., 2011). It covers a good amount of different modalities, such as, Computed Tomography (CT), Magnetic Resonance (MR) and Positron Emission Tomography (PET), enabling the aggregation of imaging studies of a patient, resulting from the scan of one of the modalities, associating that information to the patient data.

This standard is actually the international standard in the field of biomedical imaging and played an important role in the emergence of multivendor technical solutions for Picture Archiving and Communication Systems (PACS). The adoption of this standard provided a good solution for the integration with other medical systems, especially the Hospital Information Systems and the Radiology Information Systems (Gibaud, 2008).

In practice the DICOM standard itself can be considered as a multi-part document, composed by eighteen different parts (Figure 15).

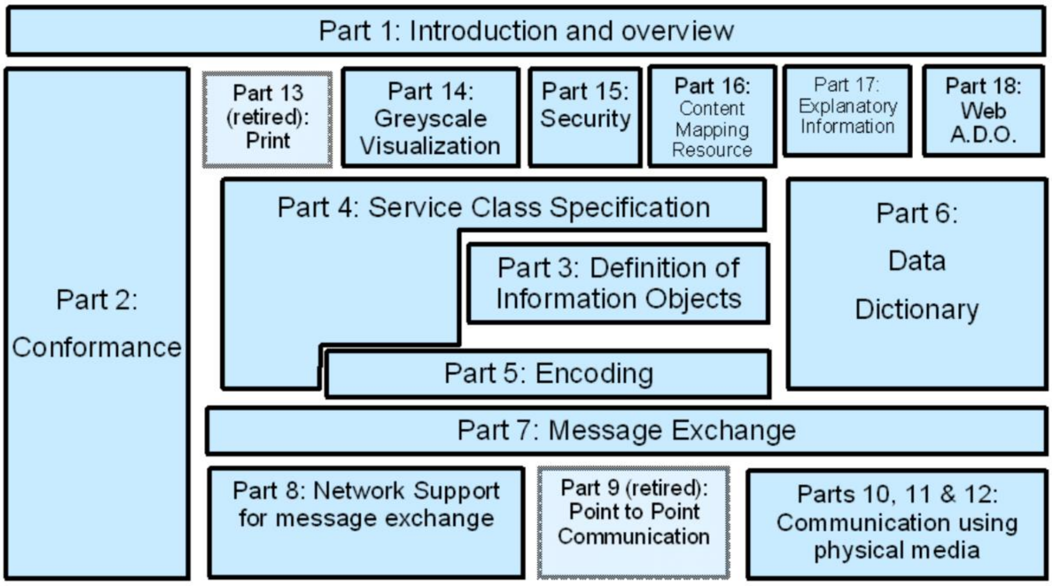
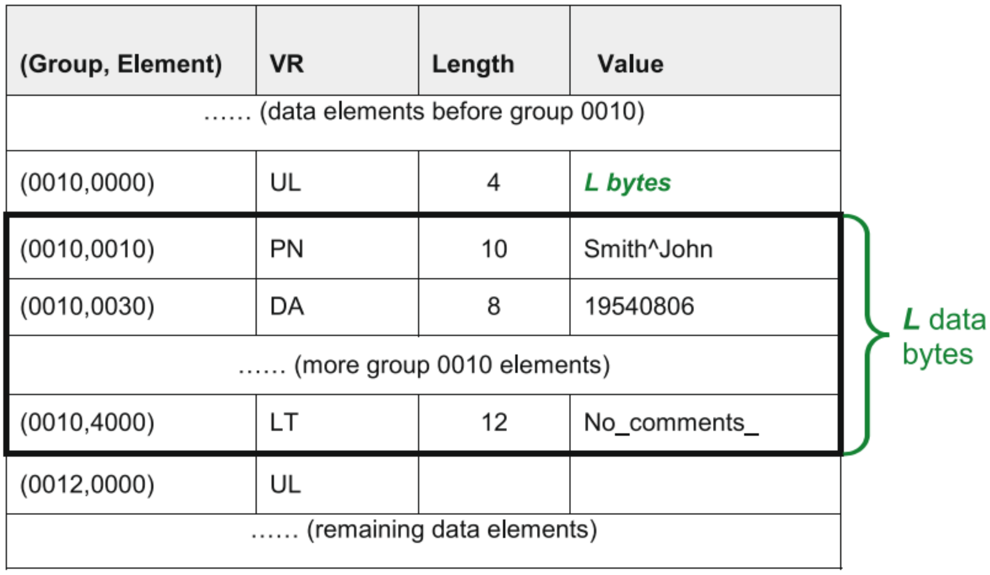

Part 1: Introduction and overview
Part 2: Conformance
Part 13 (retired): Print
Part 14: Greyscale Visualization
Part 15: Security
Part 16: Content Mapping Resource
Part 17: Explanatory Information
Part 18: Web A.D.O.
Part 4: Service Class Specification
Part 3: Definition of Information Objects
Part 5: Encoding
Part 6: Data Dictionary
Part 7: Message Exchange
Part 8: Network Support for message exchange
Part 9 (retired): Point to Point Communication
Parts 10, 11 & 12: Communication using physical media

Figure 15 - Decomposition of a DICOM into parts (adapted from Gibaud, 2008).

Among all the parts shown in the Figure 15, the most relevant are Parts 2, 3, 4, 5 and 16. If you want to further deepen your knowledge with respect of each of these parts, we recommend reading the article "The DICOM Standard: A Brief Review", written by Bernanrd Gibaud.

If we adopt a higher-level view, all parts that make up a DICOM image file can be grouped into two major parts: the header and the pixel data. The first one contains information about the features of the study acquisition procedure, including the imaging equipment that performed the scan and the characteristics of the modality. This part also contains the general information about the patient, such as, his name, the position in time of the scan, and other elements. In a broader sense, this part could be considered as meta-information elements associated to the pixel data. The pixel data corresponds to the acquired image itself, representing the sequence of bits that compose the image.

The DICOM standard uses the binary syntax, based on the transfer of (Tag-Length-Value) triplets, i.e., any elementary Data Element it is represented by a binary sequence followed by a Tag, e.g., Data Eleme*nt (0028, 0010) Rows* represents the number of rows in an image (Figure 16). One of the disadvantages of these objects is that they cannot be edited like, e.g. XML documents. To perform modifications and readings it is necessary to use a parser that decodes the TLV triplets in order to put the information in human-readable form (Gibaud, 2008).



(Group, Element)	VR	Length	Value
…… (data elements before group 0010)			
(0010,0000)	UL	4	***L bytes***
(0010,0010)	PN	10	Smith^John
(0010,0030)	DA	8	19540806
…… (more group 0010 elements)			
(0010,4000)	LT	12	No_comments_
(0012,0000)	UL		
…… (remaining data elements)			

L data bytes

Figure 16 - Example of a group of Data Elements (adapted from Pianykh, 2008). Each element is represented by its group number and element number.

These Data Elements can be grouped together to form a set, called in DICOM an Information Object Definition (IOD) (Pianykh, 2008). This set corresponds to the set of elements that shall or may be transmitted. They can be seen as a set of attributes of a class that is designed to describe an object (Deserno, 2011; Pianykh, 2008).

The creation of each and every medical image means instantiating a new IOD (NEMA, 2011a), with all of its fields empty. This object defines the different types of imaging studies that DICOM supports (Kagadis & Langer, 2011).

Despite the great freedom that is granted for the insertion of data, in order to ensure the consistency of the attributes name, there is the DICOM Data Dictionary (Part 6 of Figure 15), composed by a list of standard data items used in digital medical imaging. This list of items is divided into groups where each group has a set of elements. Thus, each element is represented by a tuple (group, element) that represents the attribute, i.e., the DICOM Data Element. A detail of this implementation is that all existing groups in the standard are even numbers (Clunie, 2000; Deserno, 2011), and all the attributes must be formatted with one of 27 possible values – Values Representation (VR) (Pianykh, 2008; Deserno, 2011)

Another peculiarity of the DICOM standard is that includes both public and private tags, being the second nonstandard tags (Kagadis & Langer, 2011). The private tags have the same structure of the public, but instead of an even number their group is identified by an odd value. It is this parity that allows the differentiation between these two types of tags (Clunie, 2000; Deserno, 2011).

Private tags play an important role in the medical images because they allow the insertion of proprietary data supporting the manufacturer’s needs to carry specific information about their systems and are often used for maintenance purposes (Kagadis & Langer, 2011). Despite being embedded in the DICOM file, in most cases they are ignored by the most common image viewing software tools. By default, unrecognizable tags should simply be ignored during the parsing of the DICOM.

Despite the eventual advantages of using private tags as previously shown, in practice they can cause a great confusion between different DICOM manufacturers, since they can have distinct meanings, leading to a misinterpretation of the tag information by the DICOM parser. So, the DICOM standard tries to prevent the occurrence of this tags and they advise all the community to reserve some private tags as private creator tags enabling the creation of private dictionaries (Pianykh, 2008).

2.4.3. Advanced QR

In order to conduct research in an easy manner on data, it is necessary to have at our disposal an advanced search engine that should provide a DICOM Query/Retrieve service. This search engine must be able to perform queries over the DICOM repository that extend the rather limited and confined query and retrieve services defined by the DICOM standard.

There are currently few tools that provide these features and among them we can highlight the Dicoogle (Figure 17) and XNAT (Figure 18). Both projects share the most basic concepts to supports the QR mechanisms. It assumes the existence of a repository for the storage of data on which a search engine is applied, responsible for responding to queries made by the user, returning the data that fits with the performed research.

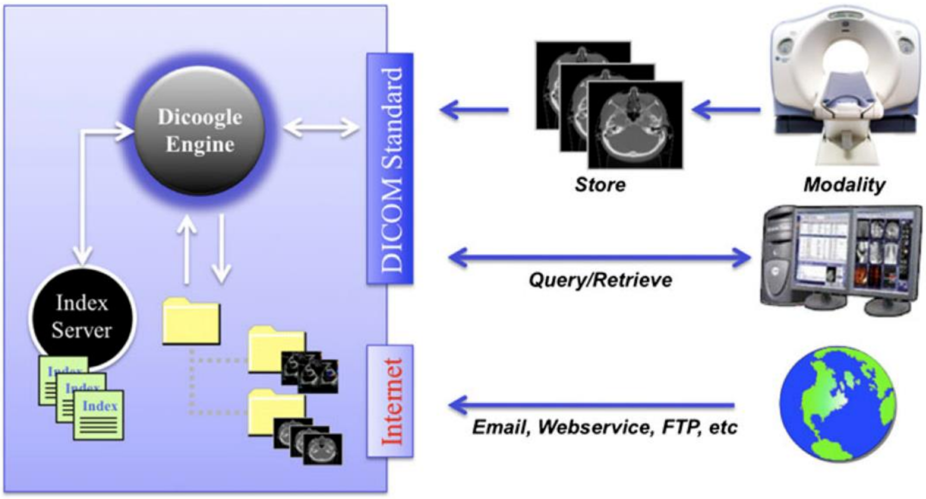


Figure 17 - Dicoogle components and interfaces (adapted from Costa et al., 2011). Medical images from different modalities can be indexed in the Dicoogle Engine and this indexation used in queries to the service provided by Dicoogle. Dicoogle can be used locally where the Dicoogle Engine is running or through Web Services.

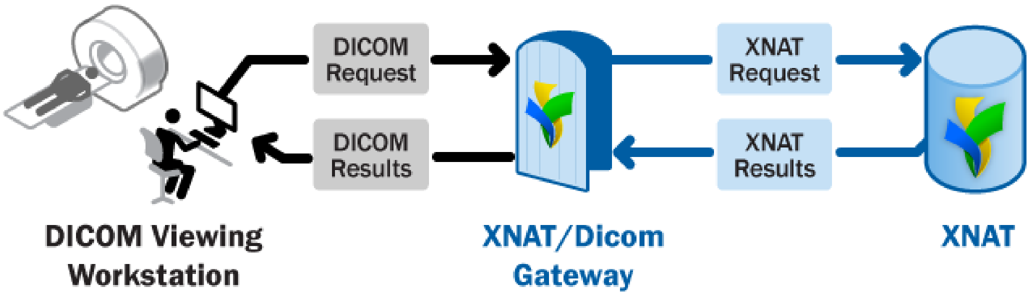


Figure 18 - XNAT DICOM Gateway (adapted from the website 'https://wiki.xnat.org/display/XNAT/XNAT-DICOM+Gateway'). Medical images, in DICOM format, that are stored in the XNAT can be obtained through the XNAT/DICOM Gateway.

2.4.3.1. Dicoogle

The Dicoogle is a PACS archive supported by a document-based indexing system and by peer-to-peer (P2P) protocols. It replaces the traditional database storage (RDBMS) by a documental organization, what permits the gathering and indexing data from file-based repositories, which allows searching the archive through free text queries. As a direct result of this strategy, more information can be extracted from medical imaging repositories, which clearly increases flexibility when compared with current query and retrieval DICOM services (Costa et al., 2011).

Dicoogle is the component that supports NEArBy by indexing the files with focus on the metadata produced during the processing workflow. Its key advantage is that there is no formal distinction between indexing standard DICOM tags and private tags namely those from NEArBy. To make this feature transparent to the eyes of the end user, it is necessary to conduct a preliminary setup so that the Dicoogle support the indexing of the NEArBy tags. These tags are allocated in the private fields of DICOM, allowing the inclusion of non-normalized DICOM information as text based semantic content. In this case the atlas labels that can also be used as part of the query strings.

This search engine is fitted with its own http-based query interface (REST services). These interfaces allow different query methods depending on the specific query parameters in the URL address:

- Query for a specific patient name within DICOM standard patient name tag: <http_address>/dim?PatientName:*xpto*
- Query for a specific brain region, according to NeuroLex lexicon, within NEArBy DICOM private tag defined by the name "neurolexTag": <http_address>/dim?advq=NearBy.neurolexTag:*xpto*

All fields that make part of the indexed DICOM file are liable to be queried. Thus, the metadata formerly integrated in DICOM private tags enables the use of the atlas dictionary to support the queries. In a given search, Dicoogle seeks all the DICOM files in the repository searching a

NeuroLex tag equal to the query token. If successful, it will return all data relating to patients containing that tag. Otherwise, it will return an empty list.

In addition to that, this search engine supports a plugin-based architecture that enables the development of new pieces of software that can be easily integrated. Thus, it's possible to develop some applications that run over the Dicoogle search engine, adding new functionalities.

2.5. Towards a common nomenclature

With the advent of the World Wide Web – an ever-evolving, easy-to-access, shared information system – the need for a shared semantic framework for neuroscience has become critically important, even more if individual researchers and automated search agents are striving to access and utilize the most up-to-date information.

For the sharing of data to be useful, the data must not only be stored in an organized fashion, but must alse be tagged with metadata that captures contextual information about the data using terms that are unambiguously defined. Unambiguous definitions of terms are necessary for researchers in order to produce meaningful results when combining data from disparate sources.

Although there are atlas with specific nomenclature, due to several reasons (e.g. from abbreviations, languages, atlas genesis, atlas resolution, etc.) it is often difficult without expertize to establish maps between information retrieve based on different atlas or other brain related information. With the increased need for building shared repositories, besides solving technical integration issues (previous section) there is a need for establishing a common nomenclature/ontology that enables proper associations of brain related information from different sources into a common reference.

To address this need, Neuroscience Information Framework (NIF) has created NeuroLex, (http://neurolex.org/wiki/Main_Page) a comprehensive lexicon of common neuroscience terminology woven into an ontologically consistent, unified representation of the biomedical domains typically used to describe neuroscience data.

To solve this problem, NIF proposed NeuroLex. The Human Brain Network (HBN) (http://www.thehumanbrain.info/) proposed other attempt similar to NeuroLex. Considering the obvious need for a stable neuroanatomical nomenclature, they have decided to use the abbreviation system of Paxinos and colleagues (Paxinos, 2007), from Neuroscience Research Australia, because it is consistently applied in companion atlases of various animals and used by many experts in the field of neuroscience.

2.5.1. NIF and NeuroLex

The Neuroscience Information Framework (NIF), an initiative of the National Institutes of Health (NIH) is a reputable repository of neuroscience resources such as data, materials or even tools that can be used through the NIF web services. The main goal of this project is to enable access to public research data and tools worldwide through an open source, networked environment.

Established in 2004, the NIH Blueprint for Neuroscience Research brings the 16 NIH Institutes, Centers and Offices that support neuroscience research into a collaborative framework to coordinate their ongoing efforts and to plan new crosscutting initiatives. Working together, representatives from the partner Institutes, Centers, and Offices identify pervasive challenges in neuroscience and any technological barriers to solving them. Early in their deliberations Blueprint representatives recognized that a framework for identifying, locating, relating, accessing, integrating, and analysing information from the neuroscience research enterprise is critical to enhancing cooperative activities in the neurosciences. A Broad Agency Announcement was issued, and in 2005, the Blueprint began support for a new initiative known as the "Neuroscience Information Framework" (NIF).

The idea of NIF is that while scientific repositories do have a multiplicity of interfaces, there should be a uniform way to conduct research on them. This concept of standardizing access has been extended to services so that developers can take advantage of the work done at NIF gaining access to all of the data available through the NIF interface.

When data is made public via NIF, it also becomes immediately available via web services. These RESTful web services can be thought of as programming functions that can be built into other applications. Currently, the data can be queried and pulled as an XML feed and several other sites are now pulling NIF data via services.

NIF provides a range of products and services that can be integrated into software tools and databases to enhance the produced applications, such as the Discovery Portal, NeuroLex, NIF Digest, NIFSTD Ontology, etc. We recommend consulting the NIF service catalog for more information on each.

Focusing on the NeuroLex service, it is a dynamic lexicon of neuroscience terms. The NeuroLex is being constructed to help improve the way that neuroscientists communicate about their data, so that information systems like the NIF can find data more easily and provide more powerful means of integrating data that occur across distributed resources. One of the big roadblocks to data integration in neuroscience is the inconsistent use of terminology in databases and other resources like the literature. When we use the same terms to mean different things, we cannot easily ask questions that span across multiple resources.

This platform provides an auto-complete service that is of enormous importance for our project. Using this service, we can assist users in their research, giving recommendations on the terms to look, according to the standard lexicon, as they go entering their data.

2.5.2. OBART

As we saw in section (2.2.2), that depending on the atlas used for the same spatial coordinates of the brain we can have different labels. For this reason, it makes sense to look for a vocabulary to map each of these labels into a common nomenclature like the NeuroLex (2.5.1) – as accepted in the neuroscience field.

The Online Brain Atlas Reconciliation Tool (OBART) (Bohland et al., 2014) aims to provide a quantitative solution to the so-called neuroanatomical nomenclature problem by comparing overlapped areas between regions defined as spatial entities in different digital human brain atlases. In addition to making comparisons between different atlases, the OBART still provides a standardization level of nomenclature, associating to each of the atlas' labels the respective value according to the NeuroLex lexicon.

In a broader sense, this tool can be considered as a bridge between the different names used by Atlas and the language used by the neuroscientists.

3. NEArBy

In this chapter, we describe NEArBy a software solution to automate a process of mapping features extracted from neuroimaging datasets to topological information contained in brain atlas, allowing to perform semantic enabled QR of brain imaging repositories. The cornerstone of this implementation relies on a set of pre-processing and analysis tools to obtain a set of features extracted after normalizing the target imaging study with the atlases.

In the next sections, we will present the developed solution that attempts to address the problems raised until now.

3.1. Rationale

For many years mapping lesions or activations into specific brain areas has been done in a crafty way, depending only on the expertise of humans. This mapping is typically done making use of brain atlas, as a reference to label the image findings in order to associate them with specific areas or functions.

The atlas plays a dual role in the cataloguing process. In addition to being a tagged spatial reference, it can be seen as a brain volumetric segmentation tool of relevant structures. As already mentioned several times, this is a very time consuming and tedious process, highlighting the opportunity to automate this entire process using computational means. Therefore, NEArBy appears as a possible solution to this problem, providing an automated way to catalogue brain activations and henceforth to setup query retrieve services from imaging repositories based on atlas topological information. To make this possible, NEArBy exploits the duality of the atlas:

- By using the atlas as a common spatial reference it allows direct spatial comparison between different normalized datasets features;
- By using the atlas as a dictionary to support automated labelling of brain feature sets it allows proper label cataloging and subsequent query and retrieve of datasets using textual criteria.

3.2. NEArBy architecture

In an attempt to address the issues raised in section 3.1, we have created a modular solution capable of responding to such problems in an efficient and easily scalable way. This solution is described by the NEArBy architecture diagram presented in Figure 19.

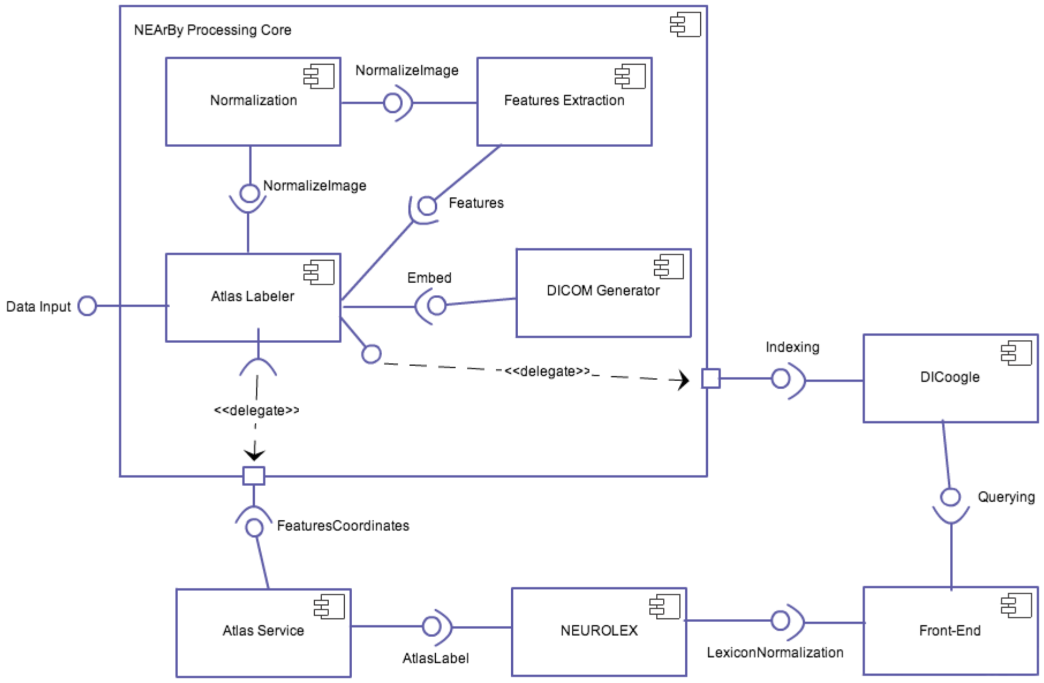


Figure 19 - NEArBy Architecture Diagram. The NEArBy Processing Core is where the NEArBy main workflows (pre-processing [normalization], feature extraction, dataset labelling [Atlas Labeler] and the DICOM generation) are coordinated using resources from Atlas Service for labelling datasets and Dicoogle to support indexation and QR. The Atlas Service use NeuroLex services in order to perform the lexical normalization. All this components are coordinated by the Front-End that is responsible for the direct interaction with the final user.

As it is evident in the architecture diagram, the system was decomposed into key components with each of them playing a distinct role.

The NEArBy Processing Core (NPC) can be seen as the orchestrator of the whole system, since it is who sets the execution order of the components responsible for the data processing. In addition to the components responsible for processing, we have a component responsible for determining the atlases' labels for the given coordinates, another component that performs lexical normalization in accordance with the NeuroLex and yet another component responsible for mediating the communication with the Dicoogle. All these components are managed at a higher level by the Front-End, a Web Interface that provides to the end user, all the offered services.

To succeed in the development of an architecture and implementation of a software solution that will provide the feature space enabling of the Query-Retrieve process, we rely on:

- A standard brain reference space (MNI) to normalize the original dataset into a common referential. This implies that NEArBy will store normalized information in the MNI space.
- Feature extraction tools, more precisely, to extract relevant topological neuroscience information from the normalized dataset.
- A Standard atlas to label relevant features.
- DICOM ID to store both the image and data relevant tags. This identifier (ID), corresponds to the patient identifier and is the only reference that allows relating the

dataset to other information other than the image itself. It is the responsibility of the data provider/user to know how to handle this id to retrieve the needed information.

- A DICOM indexation solution to provide the CBR based on actual image features and on atlas meta-labels. This allows to perform structural queries (topology)
- NeuroLex to normalize the lexicon, making the search terms match the terms of the neuroscience space.

3.3. Workflow

The entire operation of the NEArBy rests on two main data flows that are supported by its components:

- the flow responsible for the whole data processing, which culminates in the indexing of DICOM file that already contains the metadata, in the used search engine, the Dicoogle.
- the flow of querying that aims to obtain the indexed data according to the adopted search parameters.

3.3.1. Indexing

The NEArBy Processing Core (NPC), after detecting the original data file within the session folder, triggers the entire indexing flow, presented in Figure 20.

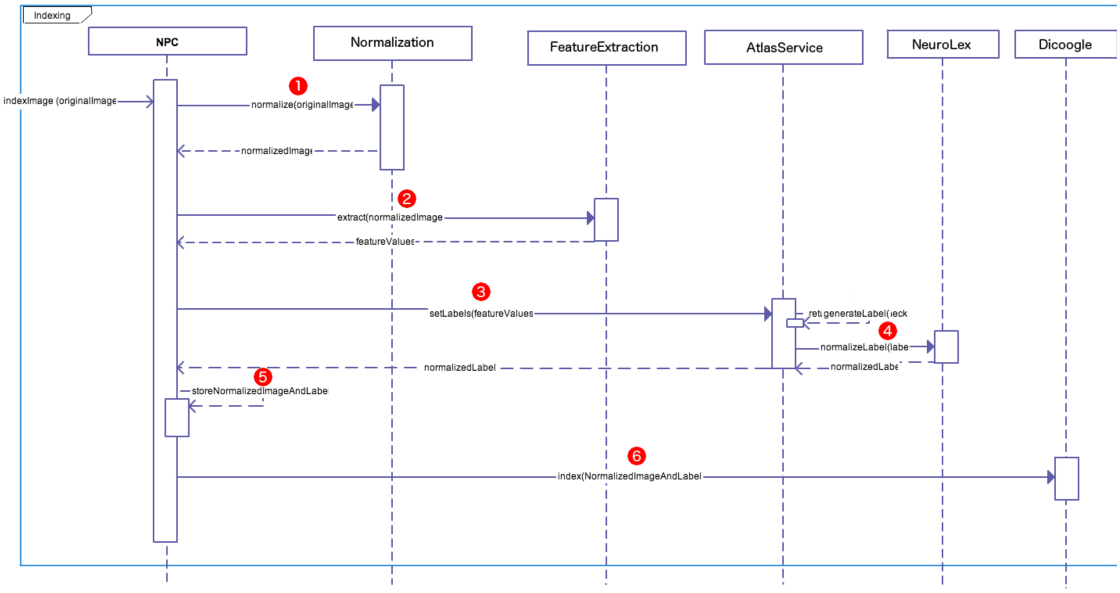


Figure 20 - NEArBy Indexing Flow. In this diagram are represented all methods that are invoked so that the entire indexing flow be performed. The most important actions are numbered and correspond to the spatial normalization (1), the extraction of the most important features (2), the association of the atlas labels (3), the lexical normalization according to NeuroLex (4), the embedding of data into the DICOM (5) and the Dicoogle indexing (6).

Initially, the file is submitted to a pre-processing step, known as normalization (1). Then the the dataflow proceeds with the feature extraction phase adapted to each particular study in (2). These features are then cataloged according to the labels of the existing atlases (3). In order to achieve standardization at the lexical level, a nomenclature normalization step is performed in accordance with NeuroLex (4).

All these metadata are then stored in a DICOM file along with the normalized data file resulting from the normalization step (5). The resulting file will then be indexed on the Dicoogle engine (6).

Since this is a data-driven architecture, one of the most important elements are the files that are exchanged between each module, during the indexing flow. These files are presented in Figure 21.

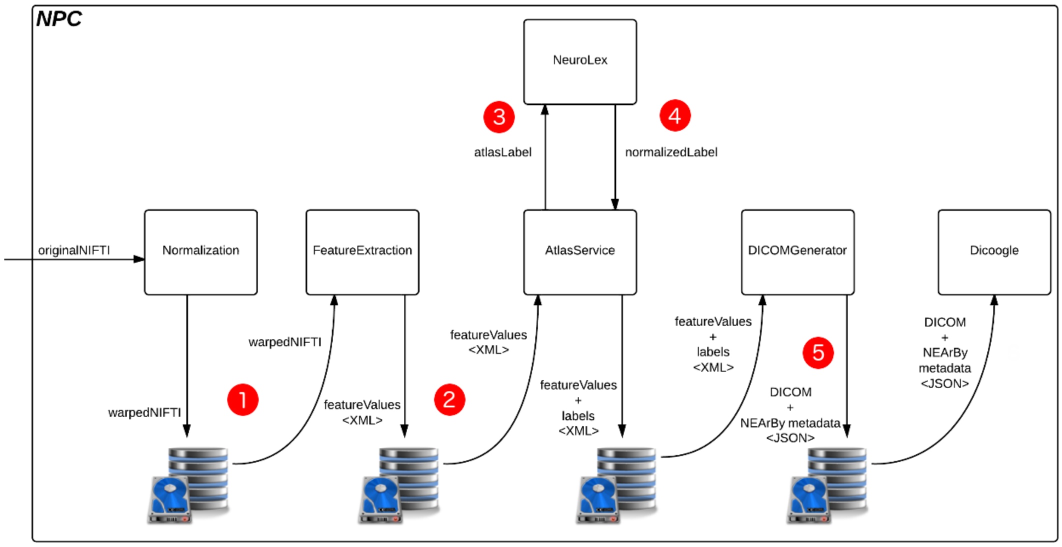


Figure 21 - NEArBy Indexing Flow. In contrast to Figure 20, in this diagram the focus is on the components and data sets - the files that are stored and exchanged between the individual modules. The more important files (numbered) are exchanged between the modules: the normalized file (1), the XML file with the extracted features (2), the features with tags (3), the tags normalized according to NeuroLex (4), the DICOM file with the metadata (5).

3.3.2. Querying

In order to benefit from the lexical standardization performed during indexing, the querying flow, presented in Figure 22, begins similarly with the same standardization, but now applied to the search terms entered by the end user.

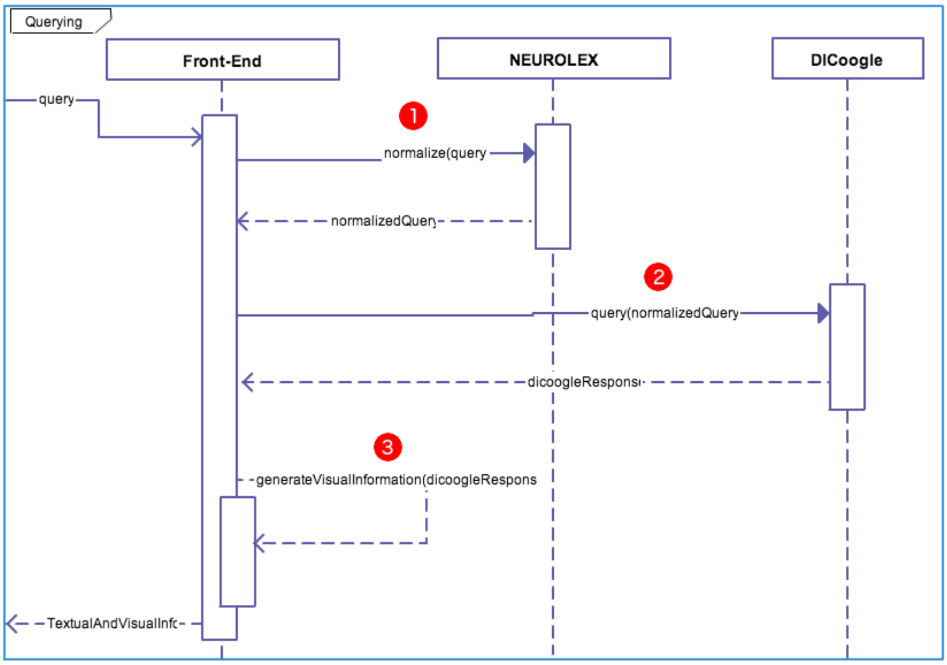


Figure 22 - NEArBy Querying Flow. The most important steps are the input search normalization (1), the Dicoogle search (2) and the processing of the returned data (3).

This mechanism is implemented with the help of NIF autocomplete service (1), providing the user with suggestions for terms that are likely to be searched. In the next step, these terms are passed to the services provided by DICoogle (2), and the returned data is pre-treated before being presented to the user through the Web Site Interface (3).

3.4. NEArBy components and processes

The execution of the two flows (section 3.3), is based on the modules presented in the architecture diagram (Figure 19). Some of these components may be grouped allowing the execution of tasks of higher level of complexity. Thus we have the group of Image Related Processes (IRP), which is composed by the Normalization, the Features Extraction and the DICOM Generator modules; the Label Generation Process, which includes the Atlas Service module; and the Indexing and QR that comprises the Dicoogle. In addition to that, the NEArBy provides a Web interface for performing the user-friendly searches supported on the Q/R engine. To conclude, although not evident in the architecture diagram, all the files that are exchanged between each process or module, are stored in a cloud based a file system. We will then present a more detailed description of each of these blocks.

3.4.1. Image Related Processes

The Processing Services are responsible for the Image Related Processes (IRP). These include Normalization to register input data into a common referential, Feature Extraction to localize the relevant features in the images and the DICOM Generator that attaches atlas based tags to the normalized input datasets.

The first step in NEArBy is to obtain the main features from the input image - these may depend on the modality and on its intended use. In case the original datasets are not in the atlas reference space, normalization (i.e. spatial co-registration with the atlas space) is performed so that relevant features in the input image can be easily mapped into the atlas space and labelled afterwards. In our solution we rely on FSL to perform the normalization using the ICBM-152 brain template in MNI (Montreal Neurological Institute) coordinate space.

NEArBy can be configured to accept several feature extraction methods as long they provide a Web Service interface compliant with NEArBy conventions that, given an input image and a set of parameters, returns the locations of features and associated information - typically some kind of value at that same location in the input data.

The DICOM Generator (DG) produces DICOM files combining both feature metadata and the image data from the input dataset. NEArBy creates for each image in the dataset a DICOM file with embedded NEArBy specific private tags where all the features and respective atlas labels are placed. The DG relies on the fact that the DICOM file format is based on sequences of Tag-Length-Value (TLV) triplets. For each atlas a different TLV is established, allowing the resulting DICOM dataset to simultaneously include labels according to several brain atlas.

3.4.2. Label Generation Process

The atlas services are responsible for labeling given locations using the atlas information. An atlas service, taking into account the chosen atlas coordinate system should return the atlas related label. Besides atlas labels, the atlas should use NeuroLex as the reference lexicon and each atlas service is responsible to establish a map between the atlas labels and the NeuroLex nomenclature. When a request is made to label a given location the atlas services return not only the atlas specific label but also the NeuroLex equivalent.

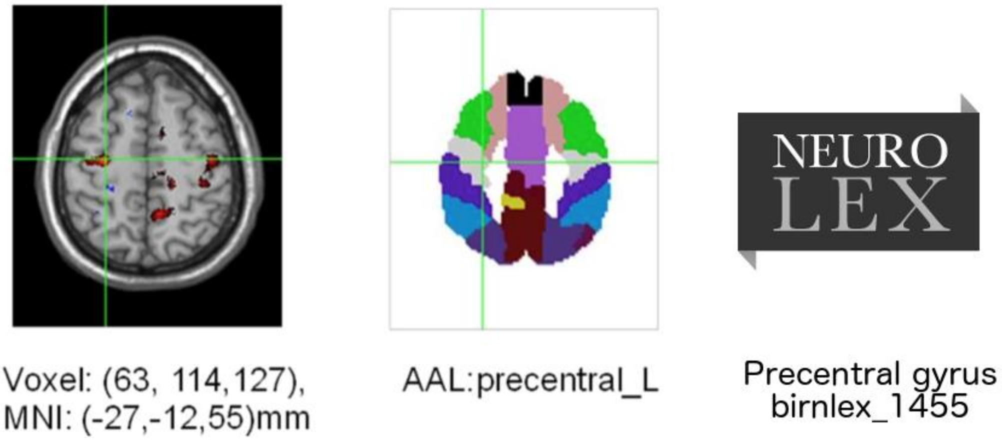


Figure 23 - Labelling and lexicon normalization for the given coordinates. The spatial coordinates (left image) are converted in a label according to the AAL atlas (center image). The atlas label is then translated to a universally accepted language, NeuroLex (right image).

For instance for a coordinate (-27,-12,55) in the MNI space, an atlas service using the Automated Anatomical Labeling (AAL) atlas, will return the 'precentral_L' and the 'Precentral gyrus' (birnlex_14551 id from NeuroLex) (Figure 23).

Currently, we have atlas services implementations based on three different atlases. Two of the most recognized structural atlas, the Automated Anatomical Labelling (AAL) (Bohland et al., 2009) and the Talairach atlas based on Brede online service (http://neuro.imm.dtu.dk/services/brededatabase/) and the FINDLab functional atlas (http://findlab.stanford.edu/functional_ROIs.html).

3.4.3. Indexing and QR

The indexing and QR engine has the role of first indexing all the metadata added by the NEArBy Processing Core (NPC) to the input image dataset and support the QR on image metadata including not only the primary image properties but also the embedded atlas metadata. The current NEArBy implementation relies on Dicoogle, an open source project that, using per-to-peer mechanisms, provides standard archiving and communication services for Medical Imaging Repositories as well as a "googlish" service for image query and retrieval. This is accomplished by the definition of a tag dictionary that establishes the Indexing System domain.

3.4.4. Web Interface

The Web interface provides the interface for performing the user-friendly searches supported on the Q/R engine. This interface uses the NeuroLex REST service to implement an autocomplete feature to assist the user in forming his queries on the client side. The query results are displayed in textual and graphical form. Although the Q/R services are non-patient centric we followed the DICOM information model to present the retrieval results. Thus the presentation style is patient-centric where the information from each patient is shown after the parsing of the associated DICOM files.

3.4.5. Storage

NEArBy assumes a cloud storage based solution for intermediate data being generated during the IRP and LGP. Using the cloud storage abstraction avoids the dependency on the specificities of NEArBy data clients and providers. The current implementation uses *Dropbox*, a cloud storage solution, as a shared file system, but any other cloud storage solution can be used - it only would imply configuring the cloud reference where the files are stored. So far the cloud resources are only committed to the processing workflow and intermediate storage. However, it is feasible in a short term to make the primary imaging repository available within the cloud. In other terms, a complete PACS solution is likely to be deployed over the cloud as long as it stands out as an efficient and cost-effective solution for an ever-growing end-user community.

3.5. NEArBy Implementation Choices and Details

Given the large amount of data produced that need to be stored and processed, it becomes imperative to create accurate and reliable software systems. Since the processing of such data has some complexity, the decomposition of the system into components arises naturally as the best solution to attack the problem.

A software project with this scale offers many features and develops a dense network of interdependences among its components, which needs to be managed efficiently.

Moreover, if we add the fact that the files to be processed can be large, in some cases presenting some GBs, being their processing computationally onerous, it is necessary to find a solution that circumvents these adversities.

One of the first options was to design and implement NEArBy to explore the benefits of inclusion of services in the cloud. The system architecture, presented in Figure 19, allows that as it is completely modular with very specific and clear role for each of the components. At the same time the definition of two workflows with clear interaction eased the tasks of defining and proposing normalized interfaces for each module that could be easily be implemented and deployed in the cloud. This means that each component can be installed on different physical machines, as shown in Figure 24.

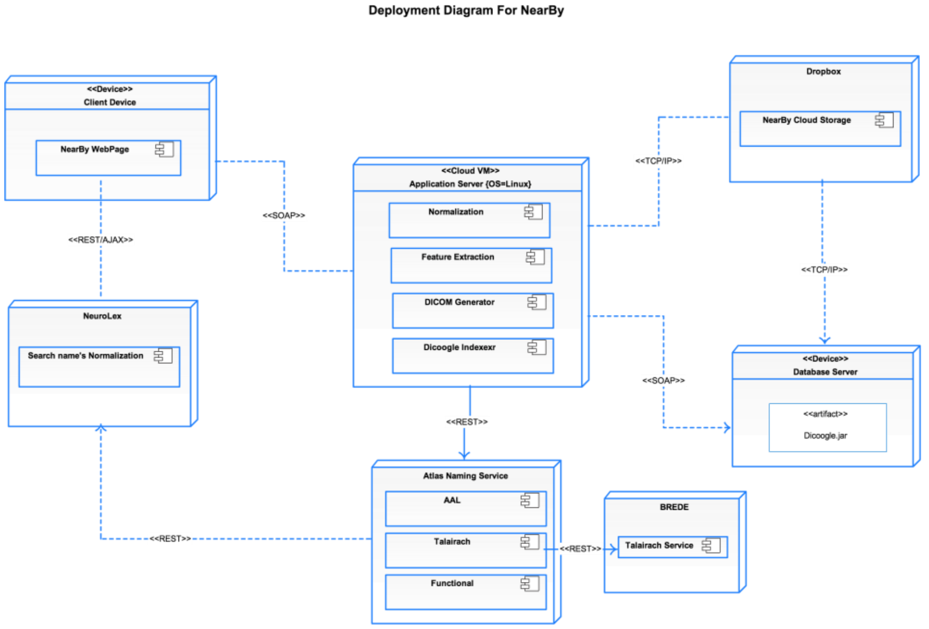


Figure 24 - NEArBy Deployment Diagram. The NEArBy modular structure allows each module to be deployed in different physical machines without compromising interaction – these are supported on standards REST and SOAP. The use of the cloud as storage solution based on Dropbox is an example as it provides a transparent storage of intermediate files produced by the system wide flows.

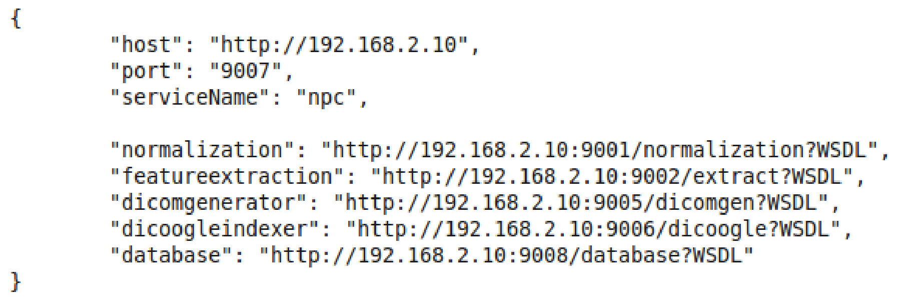
3.5.1. Modules configuration and interface

The entire implementation used the model of Infrastructure as a Service (IaaS), where in each processing node is running an application server, Glassfish (https://glassfish.java.net/), providing services to third parties through the use of SOAP Web Services and REST. Thus, we have created an open-source framework, where third parties can consume the NEArBy services to make their own workflows.

3.5.1.1. Module Configuration

To turn the system even more flexible, each of the processing nodes has a JSON configuration file where we can specify the URL, the service name and the port number, as shown in Figure 25, where the service will run. These configurations are loaded when the service starts. In addition to that, we can define the location of the other services that will be consumed by each processing node.

For example, in Figure 25, we have the configurations of the NPC processing node. As we can see, this node will publish its services on the address: http://192.168.2.10:9007/npc. This node will consume the services provided by the normalization, "featureextraction", "dicomgenerator", "dicoogleindexer" and database nodes.



```
{
        "host": "http://192.168.2.10",
        "port": "9007",
        "serviceName": "npc",

        "normalization": "http://192.168.2.10:9001/normalization?WSDL",
        "featureextraction": "http://192.168.2.10:9002/extract?WSDL",
        "dicomgenerator": "http://192.168.2.10:9005/dicomgen?WSDL",
        "dicoogleindexer": "http://192.168.2.10:9006/dicoogle?WSDL",
        "database": "http://192.168.2.10:9008/database?WSDL"
}
```

Figure 25 - NPC WebService Configuration. In NPC configuration file for a given service host it should be defined the URL of the host, the port number and the service name. These three elements will determine the location where the service will run. In addition to that, each service should declare the services that it will consume. In this case, the NPC will consume the services provided by the normalization, the "featureextraction", the "dicomgenerator", the "dicoogleindexer" and the database modules.

3.5.1.2. Application Programming Interface

To allow third parties to consume the NEArBy services to make their own workflows, the NEArBy provides a well specified Web Application Programming Interface (API).

With this API, third parties consumers are able to index and query neuroimaging data, enabling the development of interoperable client applications capable of consuming resources available on the NEArBy central repository.

`Normalization (WebServiceInterfaceNormalization):`

The normalization method is responsible to perform the spatial normalization according to FSL. As result, will be stored inside the session's folder, the files resulting from each stage of normalization (BET, FLIRT, FNIRT, WARP).

```
normalization(String niftiFile, int sessionID);
```

```
niftiFile - name of the NIFTI file to be normalized

sessionID - identifier of the Session
```

FeatureExtraction (WebServiceInterfaceFX):

This method will perform the feature extraction, in our case, the extraction of maxima and minima, according to the 3dExtrema algorithm. As result, a XML file is generated which contains all the values produced by the 3dExtrema. These values correspond to the brain activations, and each of them is associated with its spatial coordinates and the respective intensity value.

```
extract(String warpOutput, int numV, int windowSize, int sessionID)

    warpOutput - name of the normalized NIFTI file

    numV & windowSize - input parameters of 3dExtrema. Number of Voxels and the
    Window Size of the search.

    sessionID - identifier of the Session
```

DICOMGenerator (WebServiceInterfaceDICOMGenerator):

This method will generate a new DICOM file, according to the input parameters. As result, will be created a new DICOM file inside the session folder, containing all the fields defined by the user.

```
createDICOM(int sessionID, String dicomName, String patientID, String modality,
String studyDate, String institutionName, String imageToEmbed)

    sessionID - identifier of the Session

    dicomName - name of the output DICOM file

    patientID - identifier of the patient (DICOM Field)

    modality - (DICOM Field)

    studyDate - date on which the study was conducted (DICOM Field)

    institutionName - institution that conducted the study (DICOM Field)

    imageToEmbed - name of the image (NIFTI) file to embed in the DICOM
```

DICOMGenerator (WebServiceInterfaceDICOMGenerator):

This method is responsible for embed the generated values along the NEArBy’s workflow on the DICOM file. As result, will be created a new DICOM file inside session folder, containing these new values.

```
embedDICOM(String inDICOM, String outDICOM, String XMLFile, int sessionID)

    inDICOM - name of the input DICOM

    outDICOM - name of the output DICOM, that has already integrated the values
    generated by the NEArBy’s workflow

    XMLFile - name of the XML file that contains the activation values, to be
    embedded in the DICOM file.

    sessionID - identifier of the Session
```

AtlasService (WebServiceInterfaceAtlas):

For the coordinates passed as a parameter, we will get all the labels provided by Atlas plugins. As a result we will get a JSON data structure that contains the tags associated to these coordinates, and the Atlas' name, which generate each of the tags.

```
getLabels(int x, int y, int z)
       x, y, z - spatial coordinates
```

DICoogleIndexer (WebServiceInterfaceDicoogle):

This method is responsible for indexing the DICOM file on the Dicoogle.

```
index(String DICOM, int sessionID)
       name of the DICOM to be indexed
       identifier of the session
```

NEArByProcessingCore (WebServiceInterfaceProcessingCore):

This method is responsible for orchestrating the entire indexing flow performed by NEArBy. In this situation, we assume that there is already a DICOM file in which will be placed the data generated by the workflow.

```
indexing(String originalNIFTI, String originalDICOM, int numV, int windowSize)
       originalNIFTI - name of the NIFTI file to be normalized
       originalDICOM - name of the DICOM that contains the information associated
       to the patient numNV & windowSize - input parameters of 3dExtrema
```

NEArByProcessingCore (WebServiceInterfaceProcessingCore):

This method is responsible for orchestrating the entire indexing flow performed by NEArBy. In this situation, we assume that a new DICOM file will be created, in which will be placed the data generated by the workflow.

```
indexingWithoutDICOM(String originalNIFTI, String nameDICOM, String patientID,
String modality, String studydate, String institutionName, String imageToEmbed,
int numV, int windowSize)
       originalNIFTI - name of the NIFTI to be normalized
       nameDICOM - name of the DICOM that will be created
       patientID, modality, studydate, institutionName, imageToEmbed - fields of
       DICOM
       numV, windowSize - input parameters of 3dExtrema
```

3.5.2. Session Folder

Although not the main focus of the project, it was crucial to incorporate in the design of the final solution some mechanisms to deal with communication delays that may be introduced by the network. For that reason, NEArBy was conceived as a data-driven system, where each module implements a listener that is triggered when specific needed resources are available i.e. file appears in the session folder.

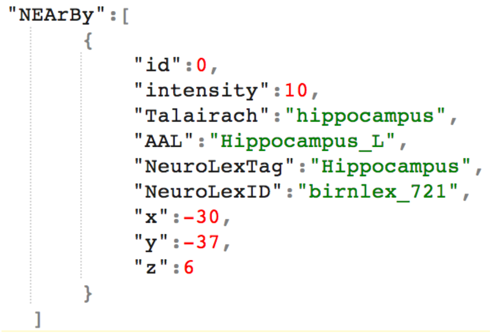
The session folder is an important concept in NEArBy. This folder corresponds to a repository of all data and metadata being generated during the Indexing flow (Figure 20). The name assigned to this folder is the result of the concatenation of the word “Session” and the ID that is returned from the database, which is auto incremented, when we add a new entry in the data base. Thus, we succeeded in creating an automatic way to naming folders that has no name collisions.

The session folder is stored in the cloud, allowing direct access to data and metadata at any time, since we have internet access. At this time we use the Dropbox as our cloud storage entity, but any other cloud storage system, or even local storage, can be used. To do that it is just necessary to change the storage path defined in the NPC.

The current implementation relies on mechanisms offered by the Java 7 (JDK 7). It uses the underlying file system functionalities to watch the file system for changes. Now, we can watch for events like creation, deletion, modification, and get involved with our own actions. Therefore, we managed to work around the issues raised, creating a lightweight implementation, not subjected to pooling.

3.5.3. iAtlas

With regard to the IRP, all atlas specific metadata is encoded in JSON as show in Figure 26. The selection of JSON, a text based codification, will ease both the indexing and query-retrieve (QR) processes, as we will explain shortly.



```
"NEArBy":[
    {
        "id":0,
        "intensity":10,
        "Talairach":"hippocampus",
        "AAL":"Hippocampus_L",
        "NeuroLexTag":"Hippocampus",
        "NeuroLexID":"birnlex_721",
        "x":-30,
        "y":-37,
        "z":6
    }
]
```

Figure 26 - NEArBy Metadata. Each feature is identified by its “id”, that is an auto-incremented attribute. In addition to that, the feature has some other attributes such as the intensity value (the extrema value), the different atlas labels, the NeuroLex tag and identifier, and finally the spatial coordinates (x, y, z). In the example we have labels for position “0” with coordinates (-30,-37,6) for the atlas “talairach” and “AAL” and respective Neurolex tag and id i.e. “hippocampus”, “Hippocampus_L”, “Hippocampus” and “birnlex_721” respectively.

The data structure presented in Figure 26 can be augmented with more tags. To do that, any of the atlas services should implement the iAtlas interface (Figure 27), a JAVA interface developed to enable us to implement a plugin-based architecture.

Each atlas is considered as a plugin. This way, we can associate multiple atlases to our workflow, even during the workflow. To achieve this implementation, we used JAVA Reflection, that allows add code snippets each time we begin a new run.

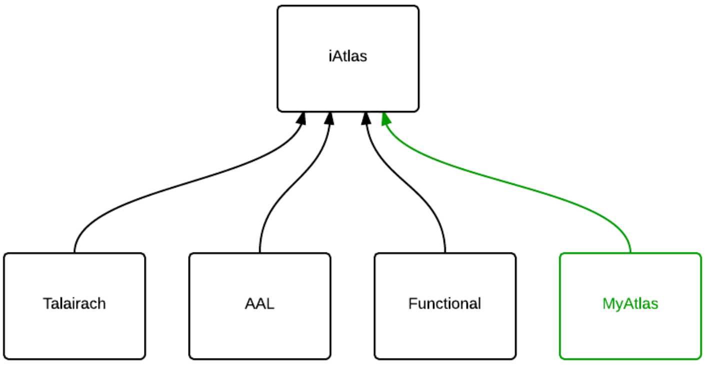


Figure 27 - NEArBy iAtlas, plugin architecture. Each atlas that implements the iAtlas interface can be added to the NEArBy system as it is a new plugin. Thus, when performing a new indexing flow, each extracted feature is labeled with the values returned by each atlas.

To add a new atlas to the NEArBy workflow, we have to do the following steps:

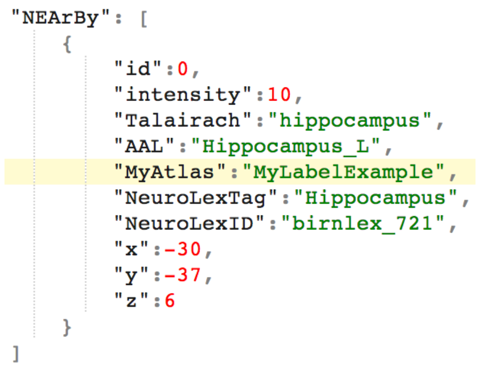
Create a new JAVA class, for example ‘MyAtlas’, that implements the iAtlas interface (Figure 28).

```
public class MyAtlas implements iAtlas
{

    @Override
    public String atlasName()
    {
        return "MyAtlas";
    }

    @Override
    public String getLabel(int x, int y, int z)
    {
        String label = "";

        // Your processing code here:
        label = “MyLabelExample”;

        return label;
    }
}
```

Figure 28 - New atlas code snippet. Each atlas should implement two methods, the atlasName and the getLabel. The atlasName method is responsible for return the atlas identifier that posteriorly will be added to the metadata structure (Figure 29). The getLabel method will receive the spatial coordinates of a given brain location and will return the corresponding label.

The method atlasName should retrieve the atlas name that will be used as the atlas identifier in the labelling process. All atlas service must provide a getLabel method that receives the coordinates of a given brain location – in Neaby in ICBM space – and returns the equivalent label. In the example it will retrieve always “MyLabelExample”

Putting the ‘MyAtlas.class’ inside the plugin folder when a new run starts, the NEArBy will add the new label to the data structure (Figure 29). Note that an entry for the “MyAtlas” appears and with the label “MyLabelExample”.



```
"NEArBy": [
    {
        "id":0,
        "intensity":10,
        "Talairach":"hippocampus",
        "AAL":"Hippocampus_L",
        "MyAtlas":"MyLabelExample",
        "NeuroLexTag":"Hippocampus",
        "NeuroLexID":"birnlex_721",
        "x":-30,
        "y":-37,
        "z":6
    }
]
```

Figure 29 - NEArBy Metadata, including the ‘MyAtlas’. Assuming that the ‘MyAtlas’ (Figure 28) was added to the NEArBy system, the metadata structure will now display a new entry. As we can see, this new entry is identified by the atlas name, returned by the atlasName method, and by the corresponding label, returned by the getLabel method.

3.5.4. DICOM Generation

For creating the DICOM dataset containing the embedded metadata (Figure 26) as private tags, we used dcm4che (https://github.com/dcm4che/dcm4che), an open source clinical image and object management. The resulting DICOM repository remains transparently available for the standard image processing and visualization tools that usually ignore private tag information. Private tagging is only meaningful within the NEArBy framework and is an implementation strategy completely hidden from the end-user that is assumed be familiar with the neuro lexicon.

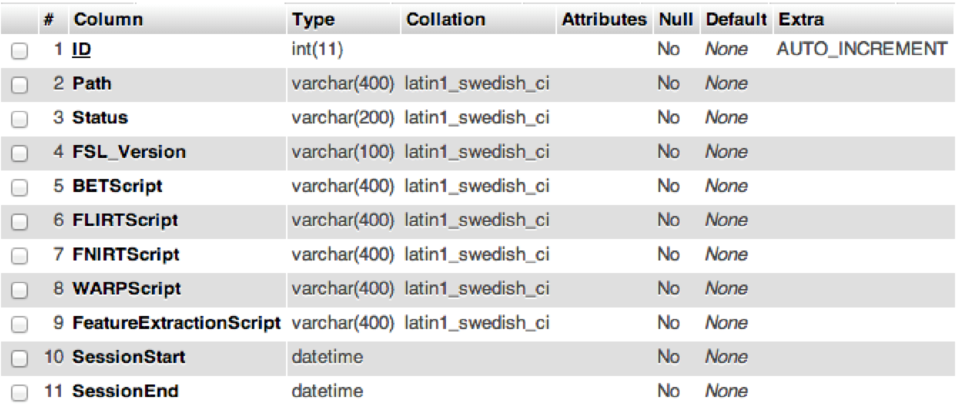
3.5.5. Dicoogle Plugin

The DICOM files embedded already with the NEArBy metadata will be then indexed in Dicoogle. To do that, we made use of the plugin-based architecture of Dicoogle, developing a plugin responsible either by the indexing process as well as the query/retrieval services. This plugin uses MongoDB as a basis to store the indexed data. This database is document-oriented, being specifically effective with JSON documents. This is the main reason for the option of storing the metadata in accordance with the JSON format (Figure 26), since the queries can be performed on a native form, without any previous parsing being required.

The Dicoogle plugin, an adaptation of the *dicoogle-mongo-plugin* (https://github.com/bioinformatics-ua/dicoogle-mongo-plugin), was developed in JAVA, implementing the IndexerInterface and the QueryInterface provided by Dicoogle SDK. The resulting .jar is then copied to the plugins folder of the Dicoogle.

3.5.6. Reproducibility

In order to allow the reproducibility of studies, a very important element in this area, we have a SQL Database in accordance with the presented structure in Figure 30. In each NEArBy run, it will be stored all the information associated to the session.



	#	Column	Type	Collation	Attributes	Null	Default	Extra
□	1	**ID**	int(11)			No	*None*	AUTO_INCREMENT
□	2	**Path**	varchar(400)	latin1_swedish_ci		No	*None*	
□	3	**Status**	varchar(200)	latin1_swedish_ci		No	*None*	
□	4	**FSL_Version**	varchar(100)	latin1_swedish_ci		No	*None*	
□	5	**BETScript**	varchar(400)	latin1_swedish_ci		No	*None*	
□	6	**FLIRTScript**	varchar(400)	latin1_swedish_ci		No	*None*	
□	7	**FNIRTScript**	varchar(400)	latin1_swedish_ci		No	*None*	
□	8	**WARPScript**	varchar(400)	latin1_swedish_ci		No	*None*	
□	9	**FeatureExtractionScript**	varchar(400)	latin1_swedish_ci		No	*None*	
□	10	**SessionStart**	datetime			No	*None*	
□	11	**SessionEnd**	datetime			No	*None*	

Figure 30 - Database structure to support the reproducibility. This structure is composed by the session identifier (id), the path of the session folder, the current execution state of the indexing flow (Status), the version of the used FSL, the full path to the scripts used in the Normalization step (BETScript, FLIRTScript, FNIRTScript, WARPScript), the full path to the script used in the feature extraction step (FeatureExtractionScript) and the start date and completion date of the indexing flow.

Assuming that we would start a new session, for example, the session 1. The table would be populated with the following data:

ID	1
Path	/home/user/Dropbox/Public/Nearby/Session_1
Status	END_INDEXING
FSL_Version	fsl4.1
BETScript	./bet.sh /home/user/Dropbox/Public/Nearby/Session_1/sample.nii /home/user/Dropbox/Public/Nearby/Session_1/betOutput
FLIRTScript	./flirt.sh /home/user/Dropbox/Public/Nearby/Session_1/betOutput.nii.gz /home/user/Dropbox/Public/Nearby/Session_1/flirt.mat /home/user/Dropbox/Public/Nearby/Session_1/flirtOutput
FNIRTScript	./fnirt.sh /home/user/Dropbox/Public/Nearby/Session_1/betOutput.nii.gz /home/user/Dropbox/Public/Nearby/Session_1/flirt.mat /home/user/Dropbox/Public/Nearby/Session_1/fnirtOutput /home/user/Dropbox/Public/Nearby/Session_1/fnirtOutput.img
WARPScript	./warp.sh /home/user/Dropbox/Public/Nearby/Session_1/betOutput.nii.gz /home/user/Dropbox/Public/Nearby/Session_1/fnirtOutput.nii.gz /home/user/Dropbox/Public/Nearby/Session_1/warpOutput
FeatureExtractionScript	./3dExtrema /home/user/Dropbox/Public/Nearby/Session_1/warpOutput.nii.gz 7 100 /home/user/Dropbox/Public/Nearby/Session_1/output.xml
SessionStart	2014-06-29 10:49:05
SessionEnd	2014-06-29 11:01:14

The ‘Path’ corresponds to the full path of the Session folder, where all files will be stored.

The ‘Status’ corresponds to the actual state of execution. This field could assume different values and it is used by the Website to display a progress bar that gives feedback to the user about the state of the session. All states can be consulted in the Attachments (section 7.2).

The ‘FSL_Version’ corresponds to the FSL version. It is important so that when someone redoes the execution knows which version was used.

The ‘BETScript’, ‘FLIRTScript’, ‘FNIRTScript’ and ‘WARPScript’ correspond to the full path of the scripts that are used in the Normalization step. These scripts correspond to a wrapper above the original FSL scripts.

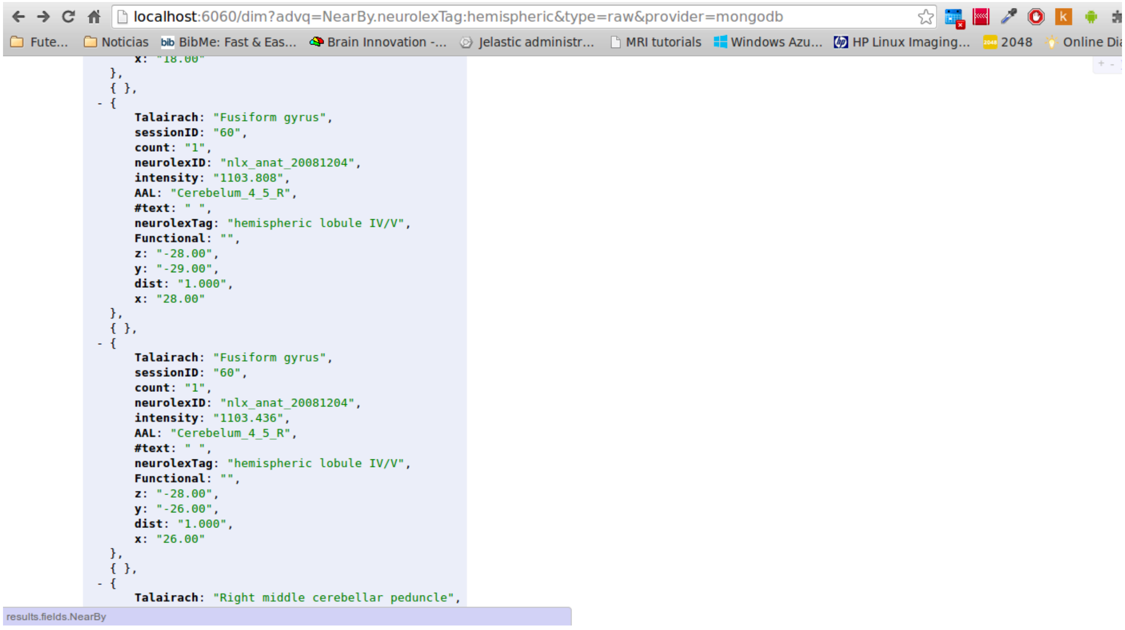
The ‘FeatureExtractionScript’ corresponds to the full path of the script that is used in the Feature Extraction step. This script is responsible to execute the 3dExtrema program.

To conclude, the ‘SessionStart’ and ‘SessionEnd’ are important elements since they monitor when the session was executed and how long it lasted.

Thus, with all these data together it is possible to reproduce an entire experiment or study, either by researcher or by someone else working independently. This reproducibility is very important because in this way it is possible to prove the veracity of the obtained results.

3.5.7. NIFTI Viewer

In what concerns the query flow, once the user chooses a term, the system makes use of the services provided by the Dicoogle, returning all entries that contain the search term. These results are returned according to a JSON data structure as shown in Figure 31.



localhost:6060/dim?advq=NearBy.neurolexTag:hemispheric&type=raw&provider=mongodb

```
        x: "18.00"
    },
    { },
  - {
        Talairach: "Fusiform gyrus",
        sessionID: "60",
        count: "1",
        neurolexID: "nlx_anat_20081204",
        intensity: "1103.808",
        AAL: "Cerebelum_4_5_R",
        #text: " ",
        neurolexTag: "hemispheric lobule IV/V",
        Functional: "",
        z: "-28.00",
        y: "-29.00",
        dist: "1.000",
        x: "28.00"
    },
    { },
  - {
        Talairach: "Fusiform gyrus",
        sessionID: "60",
        count: "1",
        neurolexID: "nlx_anat_20081204",
        intensity: "1103.436",
        AAL: "Cerebelum_4_5_R",
        #text: " ",
        neurolexTag: "hemispheric lobule IV/V",
        Functional: "",
        z: "-28.00",
        y: "-26.00",
        dist: "1.000",
        x: "26.00"
    },
    { },
  - {
        Talairach: "Right middle cerebellar peduncle",
```

results.fields.NearBy

Figure 31 - Returned results from Dicoogle Web Services. As we can see, each of the returned values corresponds to the metadata structure presented in Figure 26.

As we can see in the address bar (Figure 31), the query made will return all results that have a NeuroLex tag that contains the ‘hemispheric’ term.

In addition to the textual information presented in Figure 31, we decided to provide the user with the capability to preview the images (Figure 32) from the query result using the Neurosynth Viewer (http://neurosynth.org/), a CoffeeScript/JS library for visualization of functional MRI data (https://github.com/neurosynth/nsviewer).

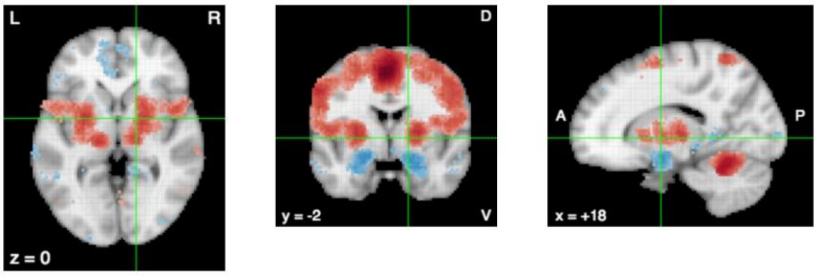


Figure 32 - NeuroSynth Javascript library. (adapted from http://neurosynth.org/). This library is a result of a project supported by Poldrack and allow the visualization of the axial, coronal and sagittal of brain.

3.6. Extending NEArBy with a new flow

One important objective of NEArBy was to create an open-source framework that could be reused by third parties, creating their own workflows, according to their needs.

The indexing flow (Figure 20) is only one example of a flow that can be built based on the modules shown in the system architecture diagram (Figure 19).

Those responsible for the creation of a new flow may well remove some modules or even add others, respecting only the data-driven methodology that we detailed above.

Imagining a situation where we had a repository where data are already normalized, i.e., it was not necessary to perform the normalization step. To seamlessly integrate the data of this repository in the NEArBy, is only necessary to perform the remaining steps of the indexing flow (Figure 33), the feature extraction, the DICOM generator and the Dicoogle indexer.

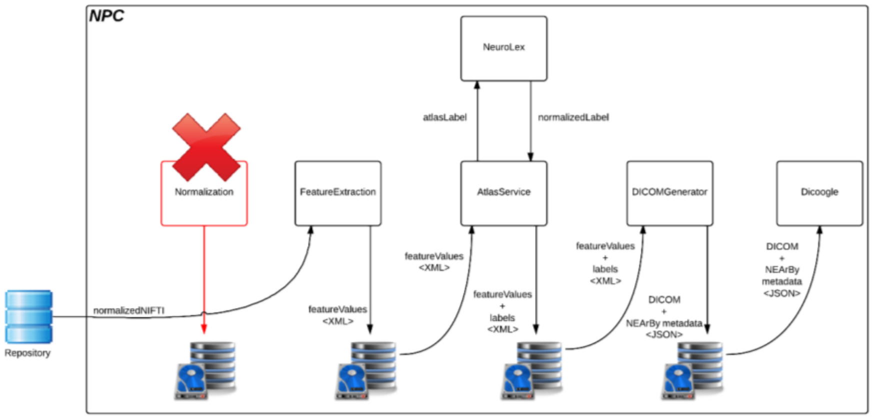


Figure 33 - Creation of a new flow. In this case, the main difference between this flow and the NEArBy flow (Figure 21) is the removal of the Normalization module. The images come directly from a repository of standardized data.

As mentioned previously, the first step is to create the session folder, where all intermediate files will be stored. Since the methodology used was data-driven, each module of the flow will be triggered by the occurrence of pre-specified file names. In this particular case, within the session folder we would put the file already normalized, with the same name as passed as parameter (‘warpOutput’) to the extract method of the API, because this is the name that the feature extraction module is waiting to be activated. All the rest of the flow would run automatically, because each module already generates the file name that the following module is waiting.

Supposing that after the extraction stage we decided to include a new stage in the flow. Assuming that we would continue with the data-driven approach, and considering that this new module was a black box, it would consume the data from the extraction stage and generate a file with data and/or metadata with the same name that the DICOM Generator module is waiting. Thus, we would be introducing a new module in the flow that would be virtually transparent to the overall system (Figure 34).

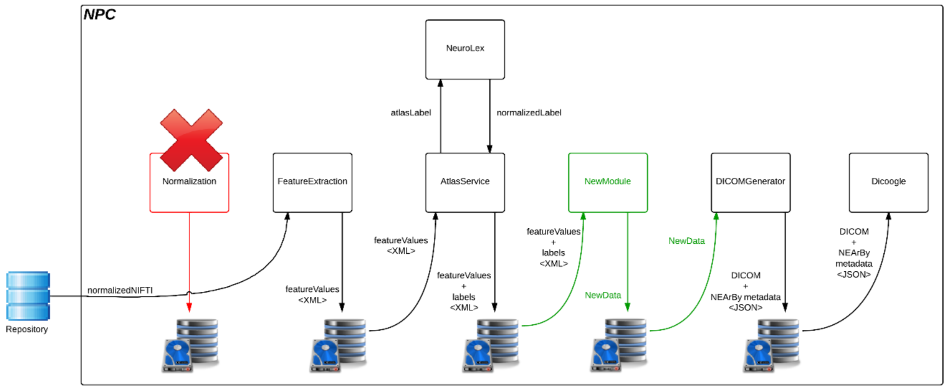


Figure 34 - Insertion of a new module in the flow. Compared with Figure 33, it is evident the presence of a new module, colored in green. This new module is responsible for the processing of data generated by the AtlasService and the generation of new data that will be used by the DICOMGenerator module.

With these examples we can check how flexible and generic is the framework presented in the NEArBy and how easy is to create new streams. At the programmatic level, the changes that are needed are few and they are presented in Figure 35.

```
 int sessionID = -1;

// Create new Session:
sessionID = DataBase.createSession();
DataBase.setSessionStatus(sessionID, "BEGIN");


String home = System.getProperty("user.home");
String directory = home + "/Dropbox/Public/Nearby/";

File f = new File(directory + "Session_" + sessionID);
if(!f.exists())
    f.mkdir();      // Create directory

// Set Session Directory:
DataBase.setSessionDirectory(sessionID, directory + "Session_" + sessionID);

// Set FSL Version:
DataBase.setFSLVersion(sessionID, "fsl4.1");

// +++++++++++++++ WORKFLOW ++++++++++++++
// ----- Feature Extraction -----
FeatureExtraction.extract(directory + "featureExtractionOutput.nii.gz", numNV,
windowSize, sessionID);

 // ----- The New Module -----
 NewModule.doSomething("featureExtractionOutput.nii.gz",
"DICOMGeneratorInput", sessionID);


// ----- DICOM Generator -----
DICOMGenerator.embed("originalDICOM.dcm", "outDICOM.dcm",
"DICOMGeneratorInput", sessionID);

// ----- DICoogle Indexer -----
DICoogleIndexer.index("NEArBy.dcm", sessionID);
// ++++++++++++++++++++++++++++++++++++++++++
```

Figure 35 - Create a new flow, inserting a new module. This code snippet corresponds to the reorganization of the indexing flow to reproduce the graphical result shown in Figure 34.

3.7. Related Work

As in several other areas of modern biology, neuroscientists are faced with an embarrassment of riches: an explosion of experimental data that overwhelms the information-carrying capacity of traditional publication mechanisms. In a similar way to what was done a few years ago with the Human Genome Project (HGP)[1], several efforts are now aimed at the creation and development of solutions capable of analysing the collected data from brain imaging, as is the case of the Human Brain Project (HBP)[2], supported by the European Commission.

Taking advantage of this wave of growth and existing financing, several projects followed this path, trying to find solutions to some of the issues raised here already.

There are similar projects that use multi-atlases to produce automated labelling of human brain images. The Mindboggle (Klein et al., 2005) software provides confidence measures for labels based on probabilistic assignment of labels and could be applied to large databases of brain images (Bohland et al., 2009). However, Mindboggle although more accurate that simple atlas labelling, relies on computational expensive software solutions such as FreeSurfer (Dale et al., 1999) and ANT's (Chen, 2001) to obtain segmentation information and labelling information - this represents an added complexity in comparison to the NEArBy solution.

NEArBy solution, when using NeuroLex as a normalized lexicon is not new as is starting to be widespread namely in projects like 'The Online Brain Atlas Reconciliation Tool' (OBART) (Bohland et al., 2014) to "reconciliate" different atlas nomenclature. In contrast with OBART, NEArBy focus on the full automated multi-atlas labelling of datasets for QR and not labelling specific individual regions of interest (ROI). Both approaches are in fact complementary and the integration of OBART labelling solution as an external atlas service in NEArBy could improve the quality of the labelling and therefore of the QR.

With respect to the pre-processing of the data, it is important to emphasize the work carried out by the Nipype group. Nipype (Gorgolewski, 2011) is a Python project that provides a uniform interface to existing neuroimaging software and facilitates interaction between these packages within a single workflow. It provides an environment that encourages interactive exploration of algorithms from different packages (e.g., SMP, FSL, FreeSurfer, Camino, MRtrix, MNE, AFNI, Slicer), eases the design of workflows within and between packages, and reduces the learning curve necessary to use different packages. In our case, we have developed our own workflow because in addition to a unique tool for pre-processing, which is the FSL (used to perform space normalization), we still need to assemble a set of modules to put into practice the hypothesis that we have proposed. It is for this reason that we directly use the FSL tool instead of NiPype.

[1] http://www.genome.gov/10001772

[2] https://www.humanbrainproject.eu/

In this context, XNAT (Marcus et al., 2007) is a mandatory reference as open source software platform with the goal to facilitate common management and productivity tasks for neuroimaging. XNAT can handle DICOM images, providing an online viewer to display not only DICOM files but also neuroimaging formats as Analyze. This platform supports several modalities as MRI, CT and PET. XNAT also provides a web interface to store, retrieve, navigate and to perform queries over the data (Health, 2013; Marcus, Olsen, Ramaratnam & Buckner, 2007).

Another tool from the XNAT is the XNAT- Desktop that allows the storing of text metadata in managed files, allowing local searches based on a tagging system. However, it still depends on an external source for tags and relies on XNAT engine in contrast with the automated atlas based tagging in NEArBy.

It is a fact that XNAT provides storage and a Q/R environment for neuroimaging studies. However our approach is rather distinct in the sense that we rely on atlas data structures to build the specific and flexible query tokens with which direct "interrogations" to the DICOM based repositories will be feasible. The forthcoming chapters will enlighten this approach.

Another entity that is engaged in this line of research is the Neuroinformatics Research Group that is working on the exploration and research of large data sets. The goal behind this project intends to enables users to store, retrieve and query data using for that data structures. ConnectomeDB is a tool in progress to support data gathered as part of the Human Connectome Project (Toga et al., 2012). This project aims to understand the functional relationship of the different regions of the human brain.

To conclude, when we talk about fMRI analysis, another entity that cannot be overlooked is the Brain Image Analysis Research Group from Carnegie Mellon University (CMU) (Langs et al., 2012). Although not identical to this project, they carried out other projects where we can find some synergies. They are trying to answer some Neuroscience questions, such as "How does the human brain use neural activity to represent the meanings of common words?", making use of statistical machine learning algorithms to analyse fMRI data. Particularly, they are interested in algorithms that can learn to identify and track the cognitive processes that give rise to observed fMRI data. One of the developed solutions is capable of decode which candidate word a person is thinking about, based only on the neural activity captured in their fMRI data. The program was trained using data from other people, indicating that our different brains encode word meanings in quite similar ways. This is an important statement for our project since in the NEArBy project, we are interested in doing the mapping and cataloging of brain activations in an automated way. And as it begins to become evident, the addition of such a project described above to what we are proposing, brings with it an increased remarkable value.

4. The case of functional MRI

Functional magnetic resonance imaging (fMRI) is an imaging technique, which is primarily used to perform brain activation studies by measuring neural activity (S. Ogawa et al.).

. In neuroscience, fMRI has been an important tool for the investigation of functional areas that govern different mental processes, including memory formation, language, pain, learning and emotion (Huettel et al. 2004, Cabeza et al. 2000).

It is remarkable the extraordinary growth in the use of fMRI. A decade ago, fMRI was only performed at an handful of institutions and only a few papers had been published. Now, hundreds of laboratories publish thousands of studies annually (Huettel, Song & McCarthy, 2008).. This growth is easy to see from the plot of the number of papers that mention the technique in the PubMed database of biomedical literature (Poldrack et al., 2011), shown in Figure 36.

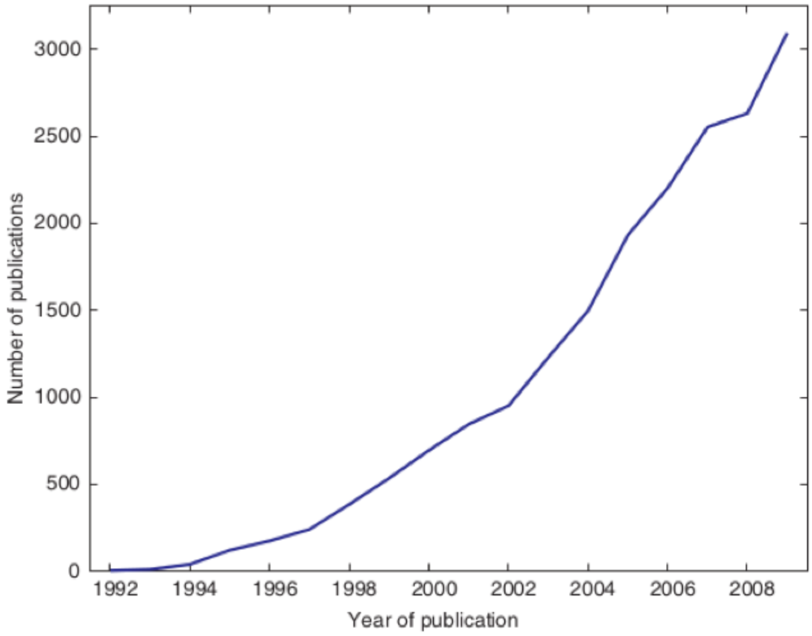


Figure 36 - Publications about fMRI per year, Poldrack et al., 2011. As it is evident from the figure, the number of publications that mention fMRI have increased significantly in recent years, attracting the interest of several researchers for its study.

In the current year (2014), if we search for fMRI in only on PubMed, we get 7570 results.

Functional Magnetic Resonance Imaging (fMRI) has provided exciting new opportunities to study topics that had long seemed out of reach of rigorous scientific investigation (Ashby, 2011).

Although the researchers are interested in using this new technology, there are still quite a few challenges to be overcome, such as the analysis of the data that are collected. Currently, an fMRI experiment produces massive amounts of highly complex data, which from the computational point of view triggers a set of challenges that try to respond to the treatment of neuroscientific data.

4.1. Activation/deactivation

One of the main contributions of fMRI is helping to map activation and/or deactivation of brain hemodynamic functions to specific brain neural activity by relating changes in the blood flow. This is possible because neural activity produces an increase in blood flow richer in oxyhemoglobin to compensate the increase in oxygen consumption. This change in oxyhemoglobin is called the Blood Oxygen Level Dependent response or BOLD effect (Ogawa et al., 1990) and induces magnetic variation observable in fMRI sequences.

One way of inferring which BOLD variations are associated with a given task is performing the Statistical Parametric Mapping (SPM). SPM is the widely used method for assessing statistical significance of neural correlates in the brain. It is a voxel-based univariate approach based on the Generalized Linear Model (Friston et al., 1995, Meyers et al, 1997) that relies on the assumption that the data from a particular voxel correspond to the time variation of the BOLD on the same physical location and, if somehow this voxel is related to a given condition known during the acquisition (e.g. protocol related), it will be statistically different from the unrelated locations in the brain cortex..

SPM based analysis produces brain volumes where each voxel values represents the statistical significance under the null hypothesis that the voxel are statistically unrelated to a given condition. Very commonly, the distribution for the activation is assumed to be Student's T distribution producing SPM maps called T-maps. The common use of SPM based analysis stems from the simplicity of using a univariate statistical test at a voxel level.The BOLD activation / de activations refer to regions of interest (ROI) analysis where we can identify the cluster of adjacent voxels related (i.e. activation) or inversely unrelated (i.e. deactivation). Most of the times these cluster are represented by their extrema/peaks (Poldrack, 2007).

4.2. fMRI pre-processing

Prior to any BOLD analysis some pre-processing is needed. For example, subject motion during the scan accounts for an important part of the unwanted variance in voxel time series analysis. For that reason there are a series of image processing steps required to overcome this issue before any statistics are performed. These steps usually include image registrations, transformations and filtering operations and are non-trivial in their section and implementation (Keator et al., 2009).

These spatial pre-processing steps are performed by the Normalization module presented in Figure 22 (Architecture Diagram). This module relies on FSL to perform the spatial normalization.

After spatial pre-processing it is possible to perform the statistical parametric mapping or other analysis in order to identify relevant activations/deactivations. The SPM processing is out scope of NEArBy as it implies dealing other meta information like protocol design and conditions that are out of scope of the current work and by themselves provide a complex tasks for any information system approach especially due to the unbounded protocol and experiments possible.

The NEArBy, when the SPM T-maps are available, performs the feature detection – in case of the fMRI the extrema on the T-maps using the AFNI program 3dExtrema, that finds local extrema (minima or maxima) of the input dataset values for each sub-brick of the input dataset. The extrema may be determined either for each volume, or for each individual slice. Only those voxels whose corresponding intensity value is greater than the user specified data threshold will be considered.

4.3. Our fMRI test case

We applied the NEArBY workflow to support the study of fMRI. In our scenario, we have a data repository for fMRI exploration containing anonymized labeled SPM activation maps. Our intention is to explore activation on specific areas, and for example, explore similar cases (Figure 37).

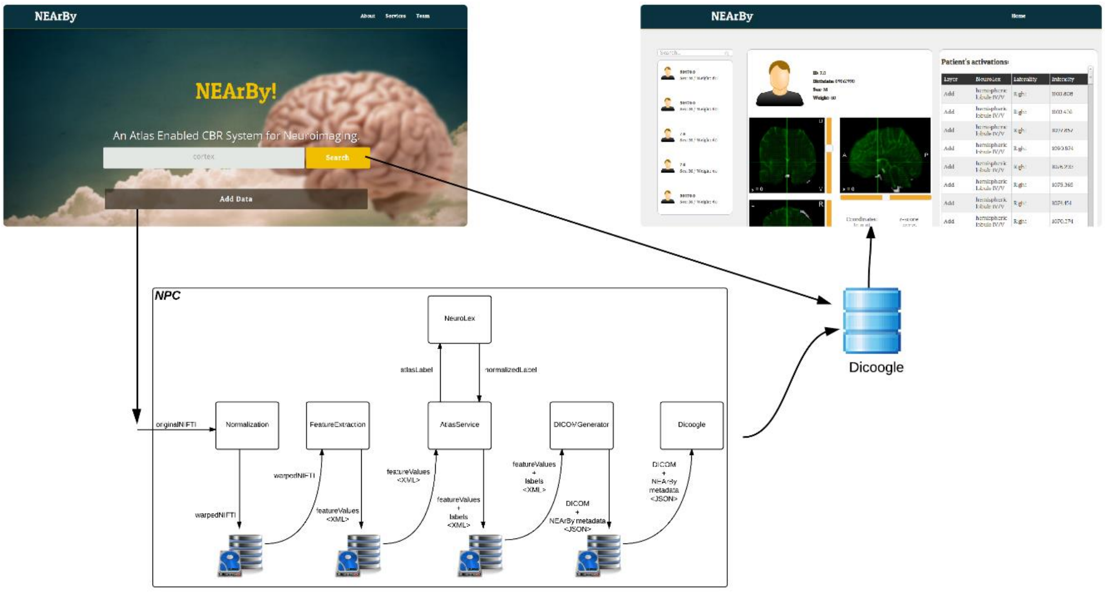


Figure 37 - NEArBy website (built over the LEGEND: Free Responsive One Page Template [http://www.dzyngiri.com/legend-free-responsive-one-page-template/]) Storyboard. The website supports two main flows: Indexing and Querying. Through the Website interface, the user can add new data to the system that will later be indexed in Dicoogle. The user can also conduct research on these same indexed data. The returned data from this research are presented graphically (screen crop in the upper right corner).

With respect to the detection of activations, in the current version, NEArBy has an *extrema* detection service that returns *extrema* from the input image - suitable in typical analysis of functional modalities such as fMRI SMP maps. The service is a wrapper of the AFNI program *3dExtrema* that finds local *extrema* (minima or maxima) of the input dataset.

We developed a Web Interface, presented in Figure 38, where the user can find a facilitated way to perform the two indexing and querying flows. If the user wishes to perform searches on the data already indexed, simply enter the search terms into the query box. On the other hand, if the user wishes to add new data to the system, simply click the “Add Data” button.

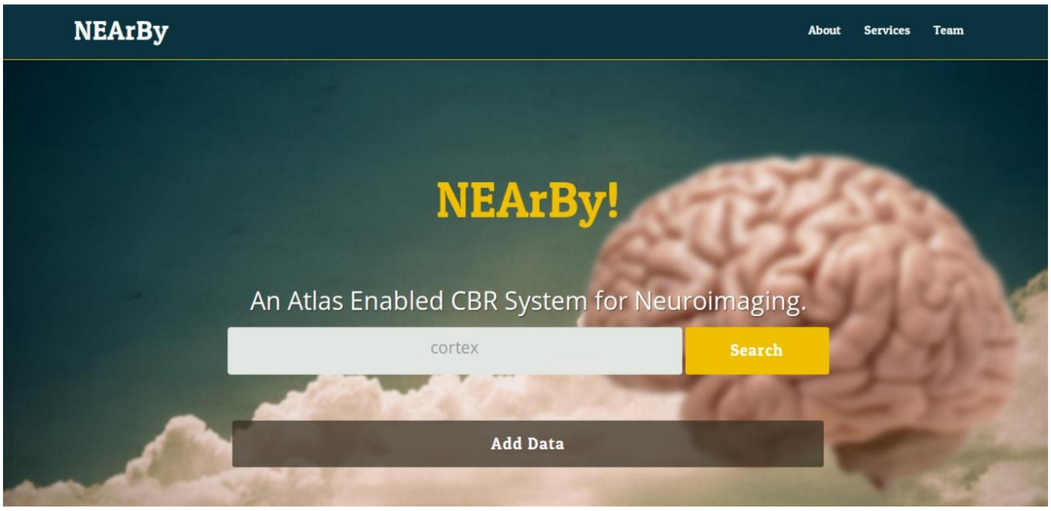
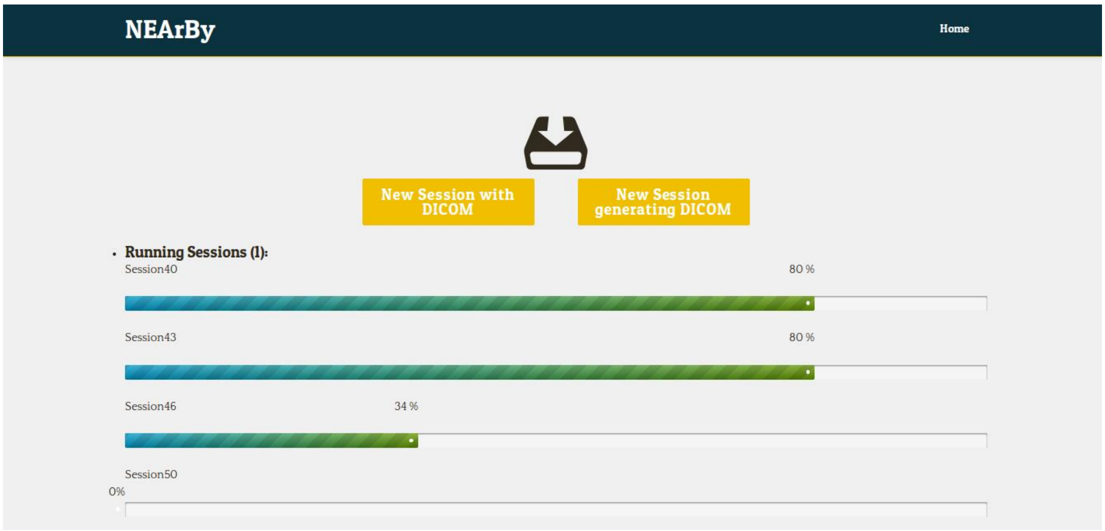


Figure 38 - NEArBy Website. The home page of the Website allows user to perform searches by entering new values in the field (contains the value of cortex) as a placeholder. It also allows introducing new data by clicking in "Add Data". The user can also read a brief description of the project (About), see what services are available (Services) and information about the NEArBy team (Team).

Regarding the process of addition of new data, the user has at his disposal the interface shown in Figure 39. In this area, we can have an easy access for adding new data, based on an existing DICOM file or even assuming that a new DICOM file will be generated. In addition, we are able to monitor in real time the evolution of the indexing process of different input data through progress bars.



- **Running Sessions (1):**

Session40 80 %

Session43 80 %

Session46 34 %

Session50 0%

Figure 39 - Add new data to the system. These new data may or may not be associated with a DICOM file. When the user want to add the NEArBy metadata to an existing DICOM, he clicks in the "New Session with DICOM" (Figure 40 A). When the user don't have a DICOM file, he clicks in the "New Session generating DICOM" (Figure 40 B). The user can still see which sessions are ongoing, by viewing the progress bars associated to each session.

Regarding the addition of new data, we have two distinct possibilities:

- We can state the name of the NIFTI file being processed and the DICOM file name where all the metadata resulting from the performed processing during the workflow will be stored. This option is present in Figure 40 A.

- We assume that there is not a DICOM file, and that it will be generated during the workflow. To respect the anonymization, we only request the user to enter the most basic data that will be part of the DICOM file. Moreover, once again, it is necessary to define the name of the NIFTI file being processed, but this time the metadata will be stored in the new generated DICOM file. This option is present in Figure 40 B.

In both cases it is necessary to specify two further values. The number of voxels and the size of the search window. These values will have a direct impact with regard to the detection of the number of activations. The smaller the number of voxels and the larger the size of the search window, most accurate will be the results. However, as in almost everything that is related to computing, this increased accuracy translates into a longer processing time. It is the responsibility of the user to define the values that best fit his study, dealing with this trade-off.

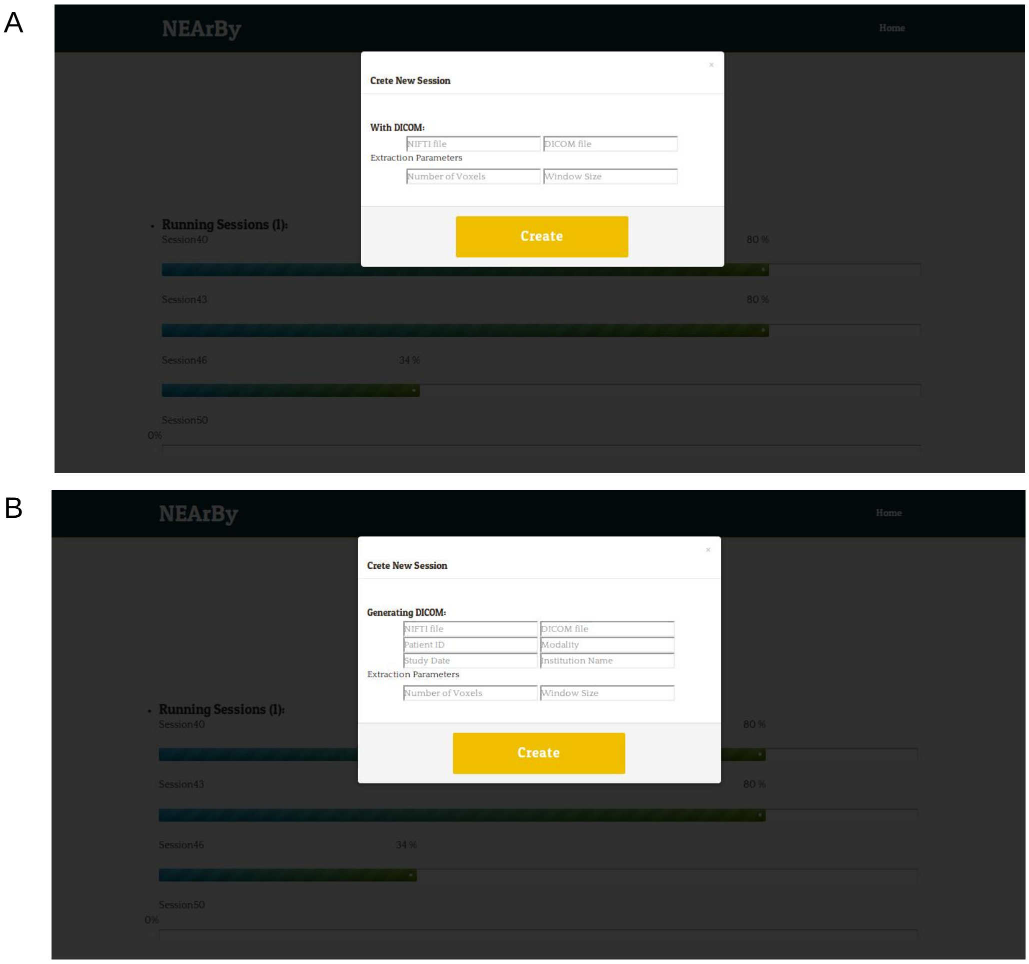
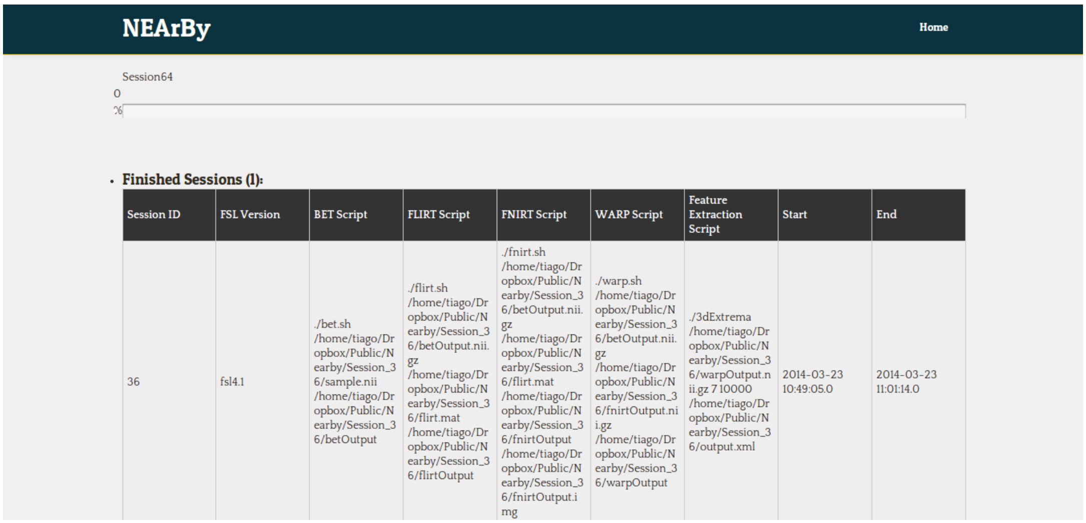


Figure 40 - Adding data taking at the outset a (A) DICOM file followed by the creation of a new DICOM file (B). In both cases it is necessary to define the name of the NIFTI file to be used and the name of the DICOM file to be used (A) or that will be generated (B). It is also required the introduction of the extraction parameters used in the feature extraction step. These values will shape the accuracy of the results. In the situation where the DICOM file is generated (B) it is necessary to define the Patient identifier (Patient ID), the used modality, the study date and the Institution name that conducted the study.

Whenever an indexing flow ends, a new entry is added to the finished session's table, as shown in Figure 42. In this table the user can check the most important steps taken during the flow, allowing the reproducibility of the data.



NEArBy

Home

Session64

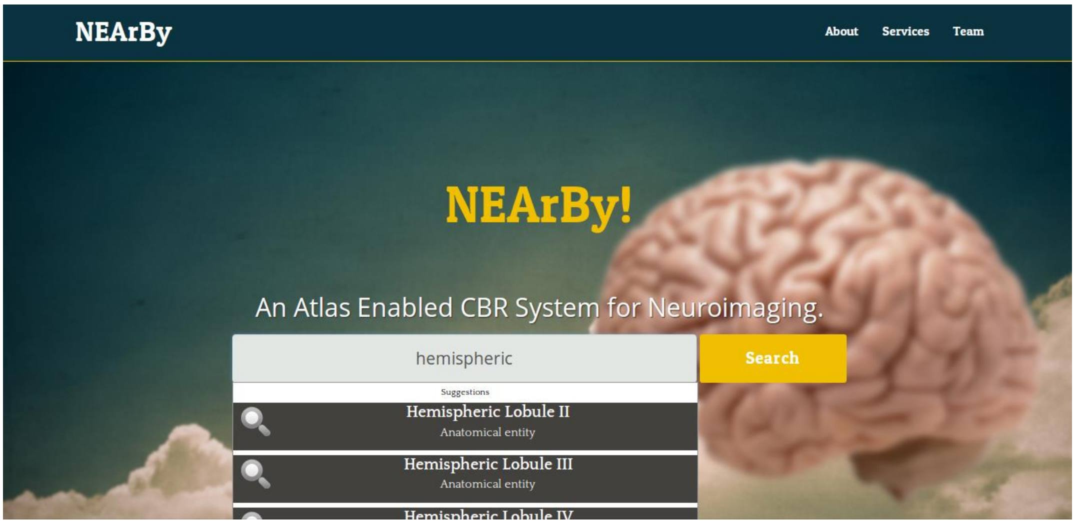
0 %

- **Finished Sessions (1):**

Session ID	FSL Version	BET Script	FLIRT Script	FNIRT Script	WARP Script	Feature Extraction Script	Start	End
36	fsl4.1	./bet.sh /home/tiago/Dropbox/Public/Nearby/Session_36/sample.nii /home/tiago/Dropbox/Public/Nearby/Session_36/betOutput	./flirt.sh /home/tiago/Dropbox/Public/Nearby/Session_36/betOutput.nii.gz /home/tiago/Dropbox/Public/Nearby/Session_36/flirt.mat /home/tiago/Dropbox/Public/Nearby/Session_36/flirtOutput	./fnirt.sh /home/tiago/Dropbox/Public/Nearby/Session_36/betOutput.nii.gz /home/tiago/Dropbox/Public/Nearby/Session_36/flirt.mat /home/tiago/Dropbox/Public/Nearby/Session_36/fnirtOutput /home/tiago/Dropbox/Public/Nearby/Session_36/fnirtOutput.img	./warp.sh /home/tiago/Dropbox/Public/Nearby/Session_36/betOutput.nii.gz /home/tiago/Dropbox/Public/Nearby/Session_36/fnirtOutput.nii.gz /home/tiago/Dropbox/Public/Nearby/Session_36/warpOutput	./3dExtrema /home/tiago/Dropbox/Public/Nearby/Session_36/warpOutput.nii.gz 7 10000 /home/tiago/Dropbox/Public/Nearby/Session_36/output.xml	2014-03-23 10:49:05.0	2014-03-23 11:01:14.0

Figure 41 - Session history. This information will enable other interested parties to be able to reproduce the same processing i.e. know which tools, parameters and sequence was used. These data result from viewing the data present in the structure of Figure 30.

Assuming that we already have many indexed files, then we can make searches on this data. Each time a new character is typed, the system makes suggestions about the terms that can be searchable, as shown in Figure 43.



NEArBy

About Services Team

NEArBy!

An Atlas Enabled CBR System for Neuroimaging.

hemispheric

Search

Suggestions

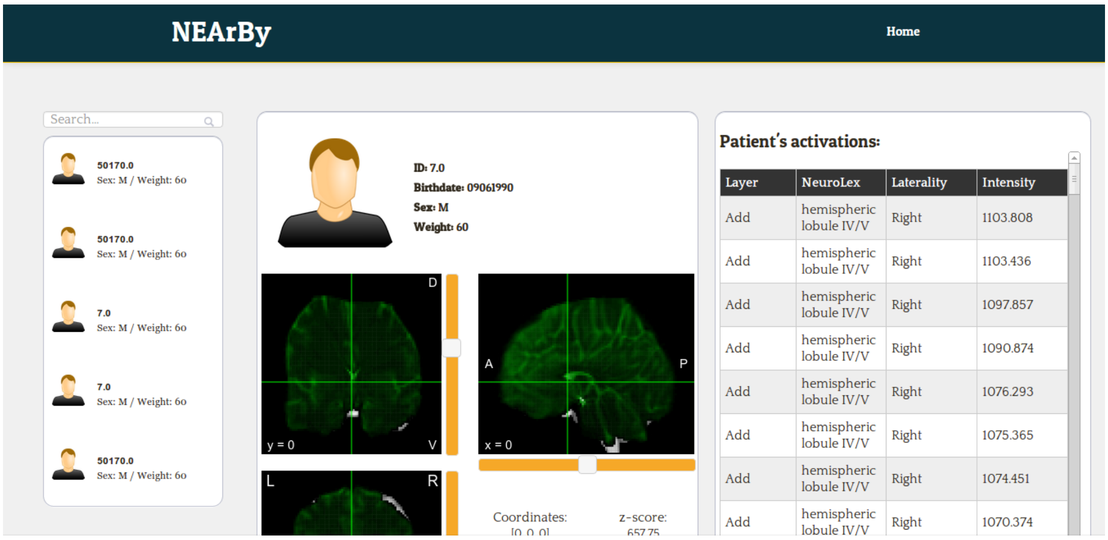
Hemispheric Lobule II
Anatomical entity

Hemispheric Lobule III
Anatomical entity

Hemispheric Lobule IV

Figure 42 - Suggestions according to NeuroLex lexicon. These suggestions are presented in real time whenever the user enters a new character, making use of the auto-complete service provided by NeuroLex.

The query result will be presented in a screen similar to Figure 44. This screen can be decomposed in three distinct sections. In the leftmost section are displayed all the cases that contain an activation in the same area as the search term. In the central section it is presented a graphical representation of the fMRI superimposed on a MNI template. This section changes each time the user selects a new case in the leftmost section. In the rightmost section are presented all detected activations for the case in question.



NEArBy

50170.0
Sex: M / Weight: 60

50170.0
Sex: M / Weight: 60

7.0
Sex: M / Weight: 60

7.0
Sex: M / Weight: 60

50170.0
Sex: M / Weight: 60

ID: 7.0
Birthdate: 09061990
Sex: M
Weight: 60

Coordinates:
[0, 0, 0]

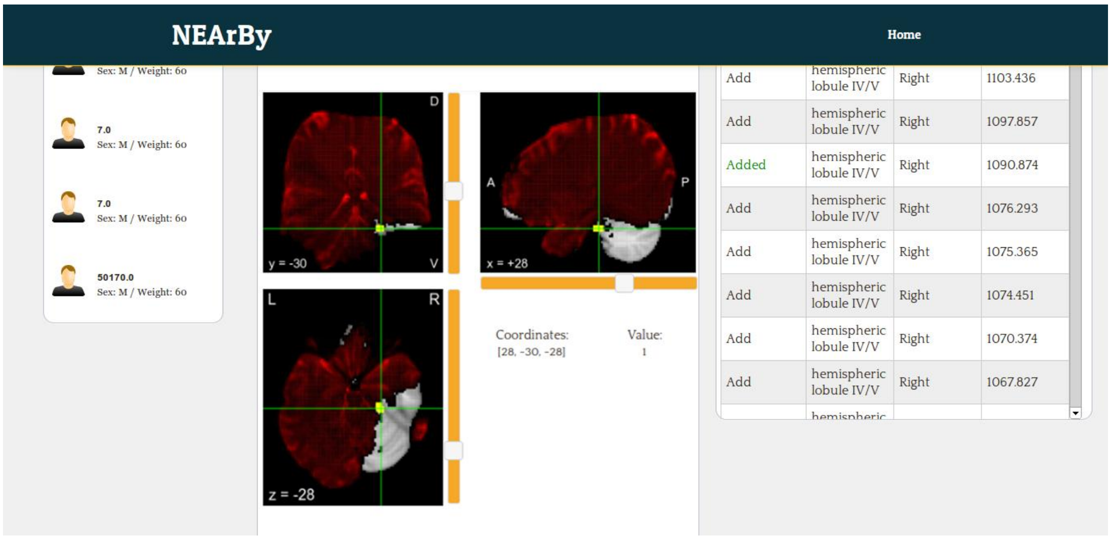
z-score:
657.75

Patient's activations:

Layer	NeuroLex	Laterality	Intensity
Add	hemispheric lobule IV/V	Right	1103.808
Add	hemispheric lobule IV/V	Right	1103.436
Add	hemispheric lobule IV/V	Right	1097.857
Add	hemispheric lobule IV/V	Right	1090.874
Add	hemispheric lobule IV/V	Right	1076.293
Add	hemispheric lobule IV/V	Right	1075.365
Add	hemispheric lobule IV/V	Right	1074.451
Add	hemispheric lobule IV/V	Right	1070.374

Figure 43 - Presentation of results in the NEArBy WebSite after parsing the returned values from DICoogle. In the leftmost section we can see all patients containing at least one activation in an area equal to that which was used as a search term. Clicking on one of these patients, the central and the right section will be changes according to the data of this patient. In the central section is presented the image resulting from the fMRI superimposed on a MNI template. In the right section is presented a table with all activations recorded for this patient.

Clicking on one of the activations presented in the table in the rightmost section, result in the addition of an overlay on the image presented in the central section, as shown in Figure 45.



NEArBy

Sex: M / Weight: 60

7.0
Sex: M / Weight: 60

7.0
Sex: M / Weight: 60

50170.0
Sex: M / Weight: 60

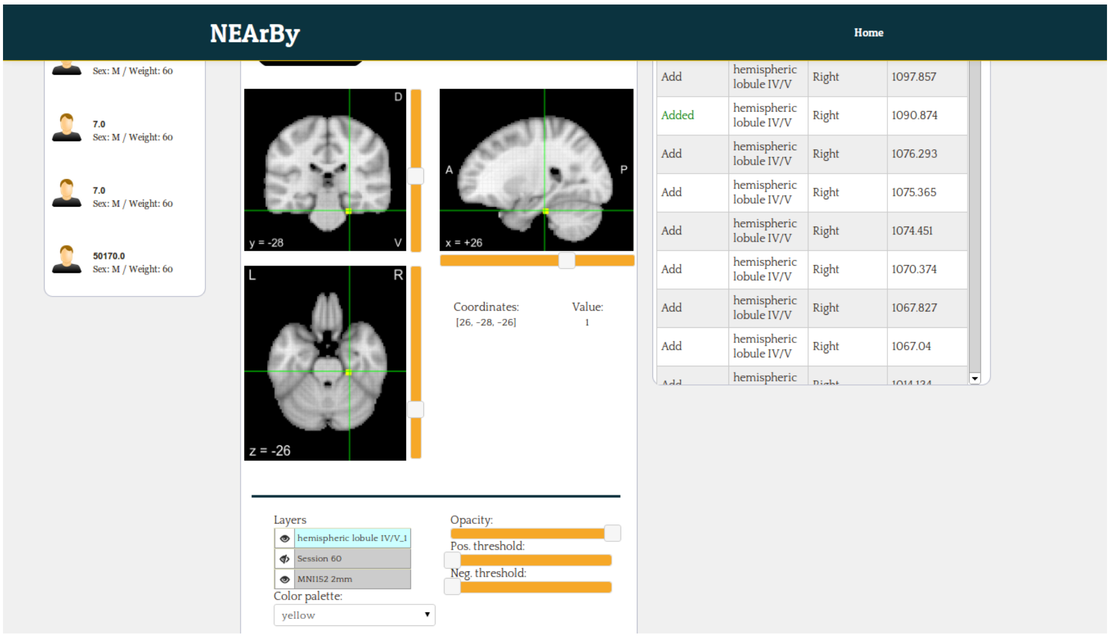
Coordinates:
[28, -30, -28]

Value:
1

Add	hemispheric lobule IV/V	Right	1103.436
Add	hemispheric lobule IV/V	Right	1097.857
Added	hemispheric lobule IV/V	Right	1090.874
Add	hemispheric lobule IV/V	Right	1076.293
Add	hemispheric lobule IV/V	Right	1075.365
Add	hemispheric lobule IV/V	Right	1074.451
Add	hemispheric lobule IV/V	Right	1070.374
Add	hemispheric lobule IV/V	Right	1067.827
	hemispheric		

Figure 44 - Visual feedback when an activation point is selected. In this case, was selected the “hemispheric lobule IV/V” and a new layer was superimposed on the image of the central section.

Below the image presented in the central section, the user can make the management in what concerns the overlays. In addition to be able to set the visibility of each of the overlays (Figure 46), it is still possible to define the order in which they appear as well as it is possible to change the color of each of the overlays.



Sex: M / Weight: 60

7.0
Sex: M / Weight: 60

7.0
Sex: M / Weight: 60

50170.0
Sex: M / Weight: 60

Add	hemispheric lobule IV/V	Right	1097.857
Added	hemispheric lobule IV/V	Right	1090.874
Add	hemispheric lobule IV/V	Right	1076.293
Add	hemispheric lobule IV/V	Right	1075.365
Add	hemispheric lobule IV/V	Right	1074.451
Add	hemispheric lobule IV/V	Right	1070.374
Add	hemispheric lobule IV/V	Right	1067.827
Add	hemispheric lobule IV/V	Right	1067.04
Add	hemispheric	Right	1014.134

Coordinates: [26, -28, -26]

Value: 1

Layers

- hemispheric lobule IV/V_1
- Session 60
- MNI152 2mm

Color palette: yellow

Opacity:

Pos. threshold:

Neg. threshold:

Figure 45 - Possibility to manage the visibility of the different Layers. If the user wants to get a better perspective of the physical location of activation, he can hide the fMRI image and display only the image of the MNI template. The management of the visibility of layers is taken through a section that lies underneath the central image. The hidden layers are identified by a risk superimposed over the eye.

We are well aware that there are many unsolved problems in this area, such as the non-standardized storage formats, the dispersion of data among different repositories, the security of the data, etc., but the goal of this thesis is not the development of an entire software solution, but rather to address a functional solution that solves a particular problem as the case of the processing, analysis and sharing of fMRI data.

5. Conclusions and Future Work

The NEArBy cornerstone idea (chapter 3) is using an atlas both as spatial reference allowing a common reference space and as an “annotated” feature space – i.e. named structurally segmented areas of interest associated with a given function. The key step towards CBR relies on regarding the atlas as a tag provider and as natural candidate to provide primary semantic information that is seamless integrated into tag based image objects.

This work evolves and extends the previous version of NEArBy, which allowed the inclusion of a far more suitable search engines for experts in the field of neuroscience and, to our knowledge, our approach is unique in the sense that we are able to combine within the NEArBy workflow the automated atlas driven labelling process and a “googlish” non-patient centric QR service. The implementation relies on DICOM persistent objects as data and metadata placeholders enabling the use of any DICOM enabled tools to embed the atlas-mapped features as private tags. Thereafter the enriched metadata is made ready for indexing and QR services with a scope that goes well beyond the limited standard DICOM tokens.

Using DICOM (section 2.4.2) as a placeholder of image information and atlas related metadata through private tagging ensures that a wealth of DICOM solutions is able to process the NEArBy generated information. In the present implementation we used Dicoogle (section 2.4.3.1) indexing and QR functionalities to support a semantic enabled QR service over brain imaging datasets using atlas label and NeuroLex (section 2.5.1) lexicon conventions.

By design, the NEArBy cloud based implementation is flexible and modality agnostic. All interactions with external services (atlas services, feature extraction and processing services) are performed using standard web service interfaces (web services and REST) using standard encoding (JSON) that depend on three concepts: location, label and atlas, leaving to each external service all modality and atlas specific details.

As we saw before, in section 3.7, there are several projects that have some similar approaches to which we are proposing, but none of them is able to play all this lightweight workflow we are presenting.

5.1. Major contributions

In this section we summarize the major contributions of the research work presented in this thesis:

- Catalogue lesions or activations into specific brain areas in an automated way.
- Query and retrieval over imaging repositories based on atlas topological information.
- Using of a common nomenclature/ontology that enables relating brain related information from different source into a common reference.

- Creation of a highly scalable modular solution capable of being extended with new modules in order to solve specific problems closely related to the project presented in this thesis.
- Conception of a user-friendly interface specially adapted for requirements of neuroscientists.

5.2. Perspectives and future work

As future work, we plan to introduce a new layer abstraction on our service through the integration with services like PubMed that will allow finding scientific articles related to query results. Furthermore, we are currently assessing the possibility to integrate other lexicon services besides NeuroLex like imaging features, specific brain areas and pathologies/conditions.

With regard to the technical aspects, it would be interesting to migrate the website implementation to a more scalable platform, as is the case of the Play Framework[3] or NodeJS[4], instead of the use of Java Server Pages (JSP). These frameworks have the main advantages of being non-blocking I/O and specially designed for real-time communications. This would permit the installation of the NEArBy on a cloud system capable of responding in good time to a significantly large number of concurrent users.

Finally, once a tool like this (NEArBy) that we are presenting requires a high level of accuracy in relation to the results presented, as well as system reliability, it is crucial that the system be subjected to a more technical review as well as a comprehensive set of tests. Moreover to this work of technological nature, since this tool has a user interface for human users, it is also critical to conduct usability tests to see if what is being presented, really meets the needs of end-users such as neuroscientists. Although tools that are being used are validated by the community such as FSL and 3dExtrema, only after the submission of the system to all these tests, we could make this project move from a proof of concept to a tool of commercial use.

[3] http://www.playframework.com/

[4] http://nodejs.org/

6. References

Abbas, F. S., Chakravarty, M. M., Gilles, B., Vladimir, V. R., Fahd, A.-S., & Collins, D. L. (2011). Creation of Computerized 3D MRI-Integrated Atlases of the Human Basal Ganglia and Thalamus. *Frontiers in Systems Neuroscience, 5*. doi: 10.3389/fnsys.2011.00071

and fMRI studies. Journal of Cognitive Neuroscience, 12(1):1{47.

Andersson JLR, Jenkinson M, Smith S (2010) Non-linear registration, aka spatial normalisation.

Ashburner J. and Friston K.J. Nonlinear spatial normalization using basis functions. Human Brain Mapping, 7(4):254{266, 1999.

Ashby, F. Gregory. (2011) Statistical analysis of fMRI data. Cambridge, Mass.: MIT Press. Print.

Astrakas, L. G., & Argyropoulou, M. I. (2010). Shifting from region of interest (ROI) to voxel-based analysis in human brain mapping. *Pediatric Radiology*, *40*(12), 1857–67. doi:10.1007/s00247-010-1677-8

Bardinet, E., Bhattacharjee, M., Dormont, D., Pidoux, B., Malandain, G., Schüpbach, M., Yelnik, J. A three-dimensional histological atlas of the human basal ganglia. II. Atlas deformation strategy and evaluation in deep brain stimulation for Parkinson disease. *Journal of Neurosurgery, 110*(2), 208-219. doi: 10.3171/2008.3.17469

Bidgood Jr, W. D., Horii, S. C., Prior, F. W., & Van Syckle, D. E. Understanding and using DICOM, the data interchange standard for biomedical imaging. *J Am Med Inform Assoc, 4*(3), 199-212.

Biswal, B. B., Mennes, M., Zuo, X. N., Gohel, S., Kelly, C., Smith, S. M., Beckmann, C. F., Adelstein, J. S., Buckner, R. L., Colcombe, S., Dogonowski, A. M., Ernst, M., Fair, D., Hampson, M., Hoptman, M. J., Hyde, J. S., Kiviniemi, V. J., Kotter, R., Li, S. J., Lin, C. P., Lowe, M. J., Mackay, C., Madden, D. J., Madsen, K. H., Margulies, D. S., Mayberg, H. S., McMahon, K., Monk, C. S., Mostofsky, S. H., Nagel, B. J., Pekar, J. J., Peltier, S. J., Petersen, S. E., Riedl, V., Rombouts, S. A., Rypma, B., Schlaggar, B. L., Schmidt, S., Seidler, R. D., G, J. S., Sorg, C., Teng, G. J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X. C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y. F., Zhang, H. Y., Castellanos, F. X., Milham, M. P. (2010). Toward discovery science of human brain function. Proceedings of the National Academy of Sciences U S A.

Bohland J.W., B, Allen C.B, Mitra P.P. (2009) The Brain Atlas Concordance Problem: Quantitative Comparison of Anatomical Parcellations. PLoS ONE 4(9): e7200. doi: 10.1371/journal.pone.0007200

Bohland, J. W., Myers, E. M., Kim, E. (2014). An informatics approach to integrating genetic and neurological data in speech and language neuroscience. Neuroinformatics, 12(1), 39-62. Doi: 10.1007/s12021-013-9201-6

Cabeza R., Nyberg L. (2000) Imaging cognition II: An empirical review of 275 PET

Cabezas, M., Oliver, A., Lladó, X., Freixenet, J., & Cuadra, M. B. (2011). A review of atlas-based segmentation for magnetic resonance brain images. Computer Methods and Programs in Biomedicine, 104(3), e158–77. doi:10.1016/j.cmpb.2011.07.015

Chen, a C. (2001). New perspectives in EEG/MEG brain mapping and PET/fMRI neuroimaging of human pain. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *42*(2), 147–59.

Clunie, D. A. (2000). *DICOM Structured Reporting*: PixelMed Publishing.

Costa, C., Ferreira, C., Bastião, L., Ribeiro, L., Silva, A., Oliveira, J. L. (2011). Dicoogle – an open source peer-to-peer PACS. Journal of Digital Imaging, 24(5), 848-56. Doi:10.1007/s10278-010-9347-9

Cuticchia, A. J., Chipperfield, M. A., Porter, C. J., Kearns, W., & Pearson, P. L. (1993). Managing All Those Bytes: The Human Genome Project Presymptomatic Diagnosis: A First Step Toward Genetic Health Care.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Segmentation, I., Reconstruction, Cortical Surface-Based Analysis, *194,* 179–194.

Deserno, T. M. (2011). *Biomedical image processing*: Springer Berlin Heidelberg.

Devlin J.T., Poldrack R.A. (2007) In praise of tedious anatomy.

Devlin, J., & Poldrack, R. (2007). In praise of tedious anatomy. *Neuroimage, 37*(4), 1033-1041. doi: citeulike-article-id:1732009

Evans, A, Collins, D, Mills, S, Brown, E, Kelly, R, & Peters, T. 1993. 3D statistical neuroanatomical models from 305 MRI volumes. Nuclear Science Symposium and Medical Imaging Conference, 1993, 1993 IEEE Conference Record, January. vol. 3, 1813–17.

Evans, A. C., Janke, A. L., Collins, D. L., & Baillet, S. (2012). Brain templates and atlases. *NeuroImage*, *62*(2), 911–22. doi:10.1016/j.neuroimage.2012.01.024

Evans, A.C., Collins D.L., Milner B. (1992). An MRI-based stereotactic atlas from 250 young normal subjects, Journal Soc. Neurosci. Abstr. 18: 408

Feng, D. D. (2011). *Biomedical Information Technology*: Elsevier Science.

Florescu, D., Raschid, L., Valduriez, P., 1996. A methodology for query reformulation in CIS using semantic knowledge. International Journal of Cooperative Information Systems 5, 431 - 468.

Fox PT, Fox JM, Raichle ME, Burde RM (1985) The role of cerebral cortex in the generation of voluntary saccades: a positron emission tomographic study. J Neurophysiol 54:348-369

Friston, K. J., Holmes, A. P., Poline, J., Grasby, P., Williams, S., Frackowiak, R. S., & Turner, R. (1995). Analysis of fMRI time-series revisited. *Neuroimage, 2*(1), 45-53.

Gibaud, B. (2008). The DICOM Standard: A Brief Overview (pp. 229-238).

Gorgolewski K., Burns C.D., Madison C., Halchenki Y.O., Waskom M.L., Ghosh S.S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. Front. Neuroimform. 5:13

Health, N. I. o. (2013, 14 june 2013) Retrieved 18 may 2013, from http://www.humanconnectome.org/connectome/connectomeDB.html

Heckemann, R. A., Hajnal, J. V., Aljabar, P., Rueckert, D., & Hammers, A. (2006). Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage, 33*(1), 115-126. doi: http://dx.doi.org/10.1016/j.neuroimage.2006.05.061

Hood, M. N., & Scott, H. (2006). Introduction to Picture Archive and Communication Systems. *Journal of Radiology Nursing, 25*(3), 69-74.

Hovenga, E. J. S., & Kidd, R. (2010). *Health Informatics: An Overview*: IOS Press, Incorporated.

Huettel S.A., Song A.W., McCarthy G. (2004) Functional Magnetic Resonance Imaging. Sinauer.

Huettel, Scott A., Allen W. Song, and Gregory McCarthy. (2008) Functional magnetic resonance imaging. 2nd ed. Sunderland, Mass.: Sinauer Associates. Print.

Jenkinson M. and Smith S.M. A global optimisation method for robust affine registration of brain images. Medical Image Analysis, 5(2):143-156, 2001.

Jenkinson M., Beckmann C.F., Behrens T.E., Woolrich M.W., Smith S.M. FSL. NeuroImage, 62:782-90, 2012

Johnson, B. a, Dawes, M. a, Roache, J. D., Wells, L. T., Ait-Daoud, N., Mauldin, J. B., Fox, P. T. (2005). Acute intravenous low- and high-dose cocaine reduces quantitative global and regional cerebral blood flow in recently abstinent subjects with cocaine use disorder. *Journal of Cerebral Blood Flow and Metabolism : Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, *25*(7), 928–36. doi:10.1038/sj.jcbfm.9600093

Kagadis, G. C., & Langer, S. G. (2011). *Informatics in Medical Imaging*: Taylor & Francis Group.

Kandel, E. R., Markram, H., Matthews, P. M., Yuste, R., and Koch, C. (2013).Neuroscience thinks big (and collaboratively). Nat. Rev. Neurosci. 14, 659–664. doi: 10.1038/nrn3578

Keator DB, Helmer K, Steffener J, Turner JA, Van Erp TG, Gadde S, Ashish N, Burns GA, Nichols BN. Towards structured sharing of raw and derived neuroimaging data across existing resources. *Neuroimage*, 2013 Nov. PUBMED PMID: 23727024

Keator, D.B., Wei, D., Gadde, S., Bockholt, J., Grethe, J.S., Marcus, D., Aucoin, N., Ozyurt, I.B., (2009). Derived data storage and exchange workflow for large-scale neuroimaging analyses on the BIRN grid. Front. Neuroinform. 3, 30.

Klein, A., Mensh, B., Ghosh, S., Tourville, J., & Hirsch, J. (2005). Mindboggle: automated brain labeling with multiple atlases. *BMC Medical Imaging*, *5*, 7. doi:10.1186/1471-2342-5-7

Lancaster, JL, Tordesillas-Gutiérrez, D, Martinez, M, Salinas, F, Evans, A, Zilles, K, Mazziotta, JC, & Fox, PT. 2007. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. Hum Brain Mapp, 28(11), 1194–205.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Szustakowki, J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. doi:10.1038/35057062

Langs G., Rish I., Grosse-Wentrup M., Murphy B. (2012): Machine Learning and Interpretation in Neuroimaging, Lecture Notes in Computer Science, Vol. 7263, Springer Verlag.

Larson, Stephen David. Semantic and spatial multi-scale information models of the nervous system. La Jolla: University of California, San Diego, 2012. Print.

Madden, Tom. BLAST [Basic Local Alignment Search Tool].. Version 2.0. ed. Bethesda, MD: National Center for Biotechnology Information, 1990. Print.

Marcus D.S, Olsen T., Ramaratnam M., Buckner R.L. (2007) The Extensible Neuroimaging Archive Toolkit (XNAT): An informatics platform for managing, exploring, and sharing neuroimaging data.

Marcus D.S, Olsen T., Ramaratnam M., Buckner R.L. (2007) The Extensible Neuroimaging Archive Toolkit (XNAT): An informatics platform for managing, exploring, and sharing neuroimaging data.

Myers R.H., Montgomery D.C.(1997) A tutorial on generalized linear models. Journal of Quality Technology.

Nicholas, D., Rowlands, I., Watkinson, A., Brown, D., & Jamali, H. R. (2012). Digital repositories ten years on: what do scientific researchers think of them and how do they use them? *Learned Publishing*, *25*(3), 195–206. doi:10.1087/20120306

Ogawa S., Lee T. M., Ray A. R., Tank D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. In Proc. Natl. Acad. Sci. USA.

Oosterwijk H: Dicom Basics, third edition. OTech, Aubrey, 2005

Osteaux M., Van den Broeck R., Verhelle F., Mey J. (1996) Picture archiving and communication system (PACS): A progressive approach with small systems. European Journal of Radiology , Volume 22 , Issue 3 , 166 – 174

Ozyurt, I.B., Keator, D.B., Wei, D., Fennema-Notestine, C., Pease, K.R., Bockholt, J., Grethe, J.S., 2010. Federated web-accessible clinical data management within an extensible neuroimaging database. Neuroinformatics 8, 231-249.

Paxinos, George. Atlas of the developing mouse brain: at E17.5, PO, and P6. Amsterdam: Academic Press, 2007. Print.

Pianykh, Oleg S.. Digital imaging and communications in medicine (DICOM) a practical introduction and survival guide. Berlin: Springer, 2008. Print.

Poldrack RA, Barch DM, Mitchell JP, Wager TD, Wagner AD, Devlin JT, Cumba C, Milham MP. "Towards open sharing of task-based fMRI data: The OpenfMRI project.," *Frontiers in Neuroinformatics*, v.7, 2013, p. 12.

Poldrack, R. a. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, *2*(1), 67–70. doi:10.1093/scan/nsm006

Poldrack, Russell A., Jeanette A. Mumford, and Thomas E. Nichols.*Handbook of functional MRI data analysis*. New York: Cambridge University Press, 2011. Print.

Poline, J.B., Breeze, J.L., Ghosh, S., Gorgolewski, K., Halchenko, Y.O., Hanke, M., Haselgrove, C., Helmer, K.G., Keator, D.B., Marcus, D.S., Poldrack, R.A., Schwartz, Y., Ashburner, J., Kennedy, D.N., 2012. Data sharing in neuroimaging research. Front Neuroinform 6, 9.

Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, et al. Genome-wide association study identifies five new schizophrenia loci. Nature Genetics. 2011;43(10):969–976.

Roberts, R. J. (2001). PubMed Central: The GenBank of the published literature. Proceedings of the National Academy of Sciences of the United States of America, 98(2), 381–2. doi:10.1073/pnas.98.2.381

Shea, N. (2006). Representation in the genome and in other inheritance systems. *Biology & Philosophy*, *22*(3), 313–331. doi:10.1007/s10539-006-9046-6

Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD: Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cereb Cortex

Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., & Greicius, M. D. (2011). Decoding Subject- Driven Cognitive States with Whole-Brain Connectivity Patterns. *Cerebral Cortex*. doi: 10.1093/cercor/bhr099

Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature Genetics. 2010;42:937–948.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotactic Atlas of the Human Brain. Thieme, Stuttgart/New York.

Talairach, Jean; Szikla, G. (1967). Atlas of stereotactic concepts to the surgery of epilepsy.

Thompson, P., Mega, M., Narr, K., Sowell, E., Blanton, R., & Toga, A. (2000). Brain Image Analysis and Atlas Construction. In M. Fitzpatrick & M. Sonka (Eds.), *Handbook of Medical Imaging. Medical Image Processing and Analysis.* (Vol. 2, pp. 1063-1131): SPIE Press.

Toga Arthur W, Thompson PM, Mori S, Amunts K, Zilles K. Towards multimodal atlases of the human brain. Nat Rev Neurosci. 2006;7:952–66.

Toga, A. W. (1998). *Brain Warping*: Elsevier Science.

Toga, Arthur W. PhD; Clark, Kristi A. PhD; Thompson, Paul M. PhD; Shattuck, David W. PhD; Van Horn, John Darrell PhD, 2012. Mapping the Human Connectome, Neurosurgery Journal, July 2012 – Volume 71 – Issue 1 – p 1 – 5. DOI: 10.1227/NEU.0b013e318258e9ff

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labelling of activations in spm using a macroscopic anatomical parcellation of the MNI MRI single subject brain. Neuroimage 2002; 15: 273-289.

Valente, F., Costa, C., & Silva, A. (2013). Dicoogle, a PACS featuring profiled content based image retrieval. PloS One, 8(5), e61888. Doi:10.1371/journal.pone.0061888

Van Essen, D.C., Dierker, D., 2007. On navigating the human cerebral cortex: response to ‘in praise of tedious anatomy’. Neuroimage 37, 1050–1054.

Van Horn, J. D., & Toga, A. W. (2014). Human neuroimaging as a “Big Data” science. *Brain Imaging and Behavior*, *8*(2), 323–31. doi:10.1007/s11682-013-9255-y

7. Attachments

In this chapter we will present some details of the NEArBy, but we thought that they are not essential to be part of the main structure of the thesis, so we have created this section of attachments. In the next sections you will mainly find technical details as well as some additional data structures used in the design of the system.

7.1. Fully Functional System

In order to have a fully functional system, the installation of some extra software is required.

Regarding the normalization step, the installation of the FSL is required, in order to use the tools such as the BET, FLIRT, FNIRT and WARP.

In the feature extraction step, since the 3dExtrema is written in C++, it is necessary to compile the program to generate an executable compatible with the computer where the program will be executed.

To conclude, since the indexation plugin was developed over a non-relational database, MongoDB, its installation is required.

7.1.1. FSL Installation

The FMRIB Software Library (FSL) is compatible with the most common operational systems, Windows, Linux and Mac OS.

This software is available online, along with its installation steps, on the website: http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslInstallation.

Since being installed, it is necessary to change the simple shell script files bet.sh, flirt.sh, fnirt.sh and warp.sh, adding the path to the executable of the FSL.

7.1.2. 3dExtrema compilation

Inside the NEArBy_FeatureExtraction/nifticlib-2.0.0/examples folder, there is a shell script named ‘compile.sh’. Running this script, the program will be recompiled for the machine in use. There is only a small detail. In order to compile the program, it is necessary that the machine has the clang installed (http://clang.llvm.org/).

7.1.3. MongoDB Installation

Just like the FSL, MongoDB is compatible with the most common operational systems, Windows, Linux and Mac OS.

This software is available online, along with its installation steps, on the website: http://docs.mongodb.org/manual/installation/

7.2. NEArBy Status

In order to be able to perform real-time monitoring of the state of indexing flow, different states corresponding to different stages of execution were created.

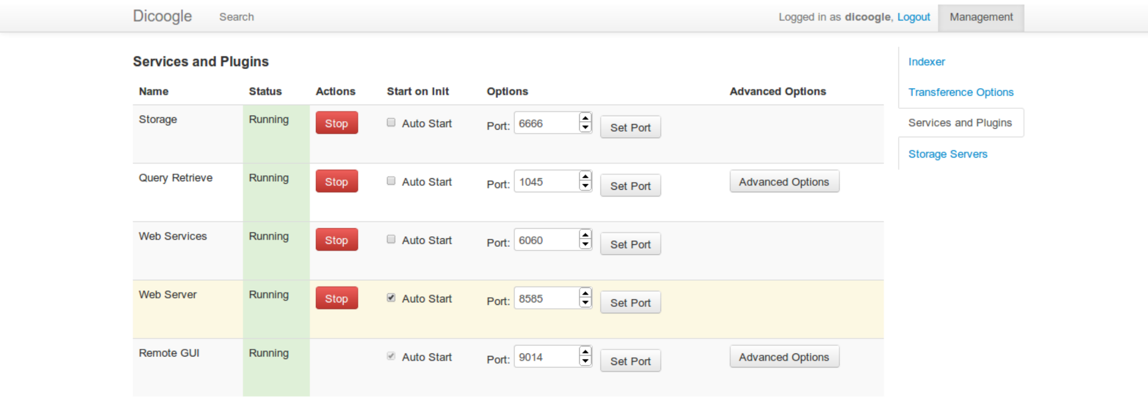
The NEArBy status are the following:

- BEGIN – When the session starts.
- START_BET – When the BET, from Normalization step, starts.
- END_BET – When the BET, from Normalization step, ends.
- START_FLIRT – When the FLIRT, from Normalization step, starts.
- END_FLIRT – When the FLIRT, from Normalization step, ends.
- START_FNIRT – When the FNIRT, from Normalization step, starts.
- END_FNIRT – When the FNIRT, from Normalization step, ends.
- START_WARP – When the WARP, from Normalization step, starts.
- END_WARP – When the WARP, from Normalization step, ends.
- START_3DEX – When the 3dExtrema, from FeatureExtraction step, starts.
- END_3DEX – When the 3dExtrema, from FeatureExtraction step, ends.
- START_DICOM_GEN – When the DICOM generation starts.
- END_DICOM_GEN – When the DICOM generation ends.
- START_INDEXING – When the Dicoogle indexing starts.
- END_INDEXING – When the Dicoogle indexing ends. Corresponds to the NEArBy's final stage.

7.3. Dicoogle Configurations

In order to correctly index the files on Dicoogle indexing engine is necessary that the Dicoogle beyond to be running, have connected their services (Storage service).

To obtain the indexed data, it is also necessary that the services "Query Retrieve" and the "Web Services" are also connected (Figure 46).



Services and Plugins

Name	Status	Actions	Start on Init	Options	Advanced Options
Storage	Running	Stop	□ Auto Start	Port: 6666 Set Port	
Query Retrieve	Running	Stop	□ Auto Start	Port: 1045 Set Port	Advanced Options
Web Services	Running	Stop	□ Auto Start	Port: 6060 Set Port	
Web Server	Running	Stop	☑ Auto Start	Port: 8585 Set Port	
Remote GUI	Running		☑ Auto Start	Port: 9014 Set Port	Advanced Options

Figure 46 - Configuration page of services and plugins provided by Dicoogle. On this page the user can connect or disconnet the services that are currently running on Dicoogle.