Grant agreement no. 211714

neuGRID

A GRID-BASED e-INFRASTRUCTURE FOR DATA ARCHIVING/ COMMUNICATION AND COMPUTATIONALLY INTENSIVE APPLICATIONS IN THE MEDICAL SCIENCES

 Combination of Collaborative Project and Coordination and Support Action

Objective INFRA-2007-1.2.2 - Deployment of e-Infrastructures for scientific communities

Deliverable reference number and title: D9.2 Final User Requirements Specification

Due date of deliverable: Month 26

Actual submission date: 8th January 2010

Start date of project: February 1st 2008   Duration: 36 months

Organisation name of lead contractor for this deliverable: **P3** University of the West of England, Bristol UK

Revision: Version 2.7 – Final Release.

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission	
RE	Restricted to a group specified by the consortium (including the	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of Contents

Document Revision History

Version	Reviewers	Date
1.0 – 1.3	All WP9 and project partners.	4/2/2009
1.4	Ronald van Schijndel, Ingrid Sluimer, Keith Cover (VUmc), Eva Örndahl (KI), Alberto Redolfi, Marco Lorenzi and Anna Caroli (FBF)	10/3/2009
1.5	Alex Zijdenbos (Prodema), Richard McClatchey (UWE) and David Manset (MAAT)	24/3/2009
1.6	Giovanni Frisoni (FBF) and D9.1 delivered to the Commission	26/3/2009
Revision Phase		
2.0 – 2.4	All WP9 and project partners.	9/12/2009
2.5	Ronald van Schijndel, Ingrid Sluimer, Keith Cover (VUmc), Eva Örndahl (KI), Alberto Redolfi, Marco Lorenzi and Anna Caroli (FBF)	17/12/2009
2.6	Alex Zijdenbos (Prodema), Richard McClatchey (UWE), Corinna Porteri, Elena Togni (FBF) and David Manset (MAAT)	24/12/2009
2.7	Giovanni Frisoni (FBF) and D9.2 delivered to the Commission	6/1/2010

Purpose and Intended Audience of this Document

D9.2 The Final User Requirements Specification (URS) forms a public deliverable and documents the requirements of the user communities in the project. This is written to be of use to clinical researchers, system designers, developers and maintainers from within and outside of the neuGRID consortium. If you wish to learn more about the project at a high-level then Sections 1, 3 and 4 will provide you with this. If on the other hand you are interested in the more detailed aims and requirements of the neuGRID infrastructure, then Sections 5 and 6 cover these aspects.

Preface to the Revision of D9.1

The requirements gathering process in neuGRID has experienced sustained and excellent support from the clinical researcher community. The first series of requirements elicitation sessions were completed and were most productive in bridging the gap between system developers and clinical researchers. Meetings focused initially on the description of high-level stories and usage patterns that would later be used to cross-check system functionality during final system testing. As these were produced a range of use-cases were created and then prioritised. This provided a clear framework on which more detailed individual requirements could be based. It has been of benefit in terms of describing the project and ensuring that important components are not overlooked. This also led to a clear hierarchical conceptual framework being identified that linked high-level stories to more finely grained use-cases and to individual users requirements. The primary focus of this document has been on the production of easily understandable models that are meaningful to both clinical researchers and software developers. The verification, prioritization and refinement of the constructed models has greatly benefited from the identified stakeholders at FBF, VUmc and KI.

Through this work an important prioritization of the services that the neuGRID platform will offer to the final users has been collected, agreed and documented. A list of pipelines and capabilities coming directly from the neuroimaging community, represented by three of the major neuroimaging centres in the world VUmc, Karolinska and IRCCS-FBF, has been studied and evaluated. This survey makes it possible to draw a safe path to ensure an effective development of the neuGRID platform. WP9 has responded to all requests for information and will continue to do so as the project moves towards completion.

It was originally planned that the final revision of this document would take place between months 22 and 26 of the project. It became clear however, that this process needed to be brought forward so that system developers had the most complete information on which to base their efforts. With this in mind work began in month 18 and the delivery of the final requirements specification was moved forward to month 23. Given that this means the user requirements analysis process will conclude earlier that originally planned, WP9 will continue its work of bridging the gap between developers and the research community until month 26 as originally intended.

The revision process initially focused on gathering feedback from developers and WP leaders regarding what was documented in D9.1. It was felt that it was important to interpret the initial user requirements in the context of what was technically feasible. Developers were also able to ask for further information and clarification where they were unsure what users had

meant or needed greater detail on specific requirements. For example, some additional work was carried out to determine the requirements that related to authorisation mechanisms and policies. This phase was beneficial to developers in that it allowed them to assess their initial designs and emerging prototype system components against what users had given priority. The process was two-way in that it also raised some important questions for end users and gave them the chance to clarify their needs. This fed into the final revision of the user requirements specification.

The initial requirements that were gathered during the preparation of D9.1 were re-evaluated in the light of subsequent developments in the project. Each requirement was analysed in turn and an assessment was made regarding its clarity and the appropriateness of the level of priority that it had been assigned. An effort was also made to ensure that the priorities assigned to requirements were, wherever possible internally consistent with each other. Following this, the final stage in the revision process was the identification of the functionality that each group of end user (basic, intermediate and advanced) could expect as a minimum from the final neuGRID platform. A usage scenario was identified for each user group that will allow the functionality of the system to be exercised and therefore validated during system integration testing by WP11. The project should use these scenarios together with the final requirements specification to measure and verify that the users' requirements are being fully addressed. To this end, it is recommended that a thorough evaluation is made with respect to each technical workpackage at Month 24.

A second round of visits to each of the clinical sites (FBF, VUmc and KI) was originally planned. Given that the requirements revision task was brought forward by three months however, this series of meetings was discussed and found to be unnecessary at that stage of the project. Instead of this, the user requirements team will where possible, take part in presenting the prototype neuGRID infrastructure to end users. This will allow them to benefit from the information and questions that developers gather during the analysis and prototyping of system components. Where prototypes have been produced, they can be used to validate the requirements that have been gathered thus far and provide useful feedback to developers. It is felt that this will encourage the translation of the final User Requirements Specification into a successful neuGRID infrastructure that addresses the essential requirements of users.

Executive Summary

The aim of the neuGRID project is to provide a user-friendly grid-based e-infrastructure and a set of generalised services, which will enable the European neuroscience community to carry out research that is necessary for the study of degenerative brain diseases. In order to achieve this goal, clinical researchers and computer scientists need to work together closely in order to determine the features that the infrastructure will provide to end users. This is challenging because the two communities are complex, have different terminologies and ways of working. Workpackage nine (WP9) was designed specifically to bridge the gap between the various stakeholders through a range of face-to-face meetings, telephone conferences and other activities in order to produce an agreed User Requirements Specification document that will drive the technical design and implementation phases of the project.

Key Objectives of WP9:

- **Conceptualisation**: to establish a common language and models among users, developers and the system deployment teams.

- **Elicitation**: to gather the end-user and developer requirements which are essential for the delivered software product to fulfil the clinical goals, the developers to understand the use-cases in which the software will be used, to understand constraints posed by legacy application and data.

- **Abstraction**: to represent the elicited and agreed requirements in the established conceptual framework.

- **Documentation**: to deliver a User Requirements Specification, that is accurate, reliable, complete and consistent. It will define functional, non-functional requirements and technical specifications known at this stage and their relationship to project objectives.

The requirements gathering process in neuGRID has benefited from enthusiastic support from the clinical researcher community. The requirements team organised elicitation sessions at FBF in Brescia, VUmc in Amsterdam and KI in Stockholm during the initial months of the project (please see Table 1 for further details.) The main purpose of these meetings was to work with researchers in order to identify the features that were necessary during their day to day work. This involved visiting research facilities and interacting with as many of the clinical researchers as possible in order to hear their views. During each of the visits researchers presented their work and methods of analysis. By bringing computer scientists and clinical researchers together in this way, a common understanding of the problem domain was reached. The series of planned meetings have now been completed and have been most productive in bridging the gap between system developers and clinical researchers. Initially meetings focused on the identification of some high-level stories and usage patterns. As these developed a range of use-cases were created and then prioritised. This provided a clear framework on which more detailed individual requirements could be based.

Date	Location	Content	Attendance
2008-02-04	Fatebenefratelli – Brescia, Italy	Initial series of requirements meetings and technical brainstorming.	ALL
2008-03-15	Karolinska Institute – Stockholm, Sweden	Second series of requirements meetings.	FBF, UWE, MAAT, PRODEMA, KI
2008-05-15	VUmc – Amsterdam, The Netherlands	Third series of requirements meetings and in person technical brainstorming.	FBF, UWE, MAAT, PRODEMA, VUmc
2008-09-02	Fatebenefratelli – Brescia, Italy	Fourth series of requirements meetings.	FBF, UWE, MAAT, PRODEMA, VUmc

Table 1: Requirements meetings held during the initial months of the project.

A major output of WP9 was the development and modelling of a group of stories which illustrate the end-to-end use of the neuGRID infrastructure. This has been of benefit in terms of describing the project and ensuring that important components are not overlooked. This led to a clear hierarchical conceptual framework being identified that linked high-level stories to more finely grained use-cases and to individual users requirements. This approach and the structure of the final user requirements specification were discussed and agreed to by project partners during face-to-face meetings. It was decided that the primary focus should be on the production of easily understandable models that are meaningful to both clinical researchers and software developers. The verification, prioritization and refinement of the individual requirements and system models has greatly benefited from the identified stakeholders at FBF, VUmc and KI.

1. Introduction

Analysis of the project scope and context (and associated users' requirements) is seen as an essential component of neuGRID that will ensure common understanding between the clinicians and those responsible for IT research and development from the outset of the development stage. It guides the development process involving multiple partners and will assist the test phase of the delivered components. *D9.2 Final User Requirements Specification* describes the requirements that need to be met for the project to achieve the goals described in the project proposal. It documents the scope of the project by reflecting the interests of all the major actors. This document establishes a hierarchical set of requirements which takes into account core project goals, the participating clinicians' views as well as constraints which ensure that focus on innovation in the promised areas is maintained from the outset.

The major outputs from this deliverable are the following:

(i) An agreed form of expression ("language") of the concepts, including but not limited to textual and diagrammatic models.

(ii) The set of prioritized functional and non-functional specifications reflecting on the requirements of both the end-users and developers, expressed using the agreed conceptual framework.

(iii) A commitment to and a plan for reviewing the User Requirements Specification (URS) document as the project progresses.

A hierarchical conceptual framework has been created that links stories to more finely grained use-cases and to the users requirements. It was decided that the needs of the project place the main emphasis on producing easily understandable models that are clear to both researchers and software developers. This process began during the initial meetings with clinical researchers and discussions led to a set of stories being identified that spanned the problem domain and allowed use-cases to be grouped into areas of common purpose. Each story was modelled and thoroughly analysed to define the group of use-cases that were present in it. At this point the requirements team went through several cycles of review with clinical researchers, which resulted in a final frozen set of stories and use-cases that were agreed by

all the project partners. Figure 1 shows the importance of review loops during the requirements gathering and within the wider system engineering process.
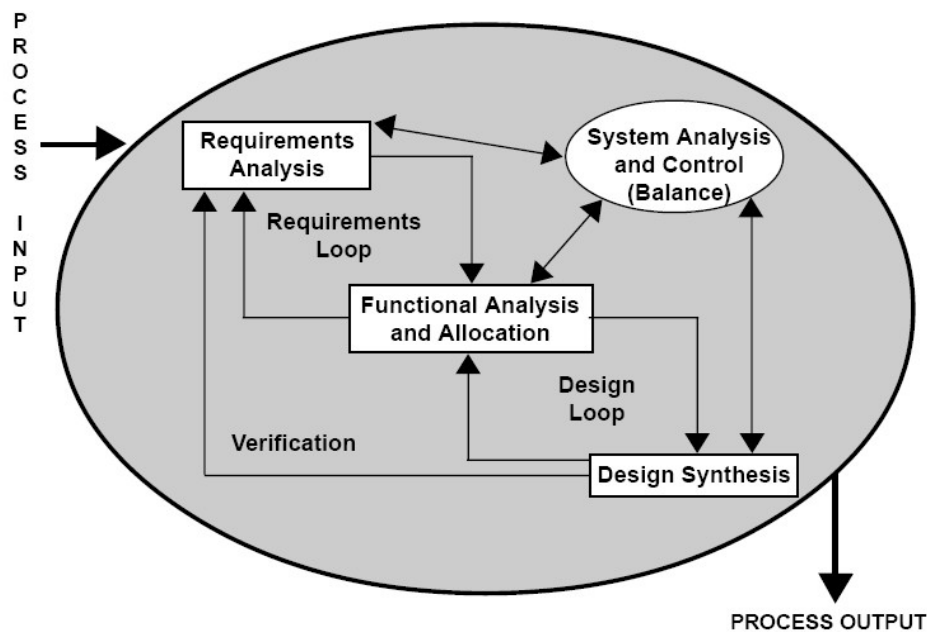

Figure 1: The Requirements Engineering Process, Image taken from [1].

The next stage in the requirements gathering process was to make some initial decisions regarding the priority of use-cases. It was clear given the scope and budget of neuGRID that not all of the use-cases would be achievable during the course of the project. With this in mind it was decided that use-cases should be prioritized using a variant of the MoSCoW technique [2]. In this prioritization framework requirements are assigned one of the following levels of priority:

- M- MUST have this.
- S - SHOULD have this if at all possible.
- C - COULD have this if it does not affect anything else.
- W -WON'T have this time but WOULD like in the future.

The neuGRID prioritization framework is almost the same as this, but for clarity and brevity it was decided use-cases that definitely would not be implemented should be removed from the specification. Use-cases that were not included in the final specification are recorded in the minutes of the requirements meetings and will therefore remain accessible as the project continues. It was also thought sensible to prioritize use-cases at this stage and in this way because they would be used during the final requirements gathering stage as a foundation upon which individual user requirements would build. The following levels of priority were assigned to use-cases in the project:

- Essential (E): Those which are absolutely vital to the production of a functional infrastructure.

- Desirable (D): Those that whilst not vital, would provide important functionality to users.
- Optional (O): Those that might be useful but don't fit into the previous two categories and will probably be the last to be implemented if time / budget allows.

The prioritization of the use-cases was useful in stimulating debate and helping developers get acquainted with the requirements process that was underway. This also led to some additional use-cases being identified and some even being removed. Once the final set of use-cases and associated priorities had been agreed, they were frozen and the concentration was placed on the gathering of individual user requirements. The project as a whole felt that clinical researchers should play an active part in writing this part of the specification. Therefore an initial draft was circulated by the workpackage leader and this was then built upon by the end-users at FBF, KI and VUmc. This process was based on individual use-cases as a means of focusing in greater detail on smaller aspects of the system. For each uses-case the relevant user requirements were identified and described. The clinical researchers responded well to this task and produced a very comprehensive list of functional and in some cases even non-functional requirements which formed the basis of D9.1.

In preparation for this deliverable, the revision process gathered feedback from developers and WP leaders regarding what was initially captured in D9.1. The interpretation of the user requirements in the light of what was technically feasible was seen as a vital step in this process. Developers in some cases asked for clarification of what users had meant or requested greater detail on specific requirements. In this way the requirements that were gathered during the preparation of D9.1 were thoroughly re-evaluated in the light of subsequent developments in the project. Each requirement was analysed and a determination was made regarding the level of priority that it had been assigned. At this stage every effort was taken to ensure consistency between that the priorities assigned to individual requirements. The final stage in the revision process was the identification of the functionality that each group of end user (basic, intermediate, advanced and pipeline developer) can expect from the final neuGRID platform. A usage scenario was developed for each user group that will allow the functionality of the system to be fully validated. The project may use these scenarios together with the final requirements specification to measure and verify that the users' requirements are being fully addressed.

2. Glossary of Terms Used

This section lists and briefly describes some clinical and technical terms that are subsequently used in the following sections of the document.

3		
	3D parametric surface mesh models	Collection of vertices, edges and faces that defines the shape of a polyhedral object in 3D computer graphics and solid modeling.
A		
	Actor	An indicative group of users or stakeholders in a system.
	Acquisition centre	A Medical facility staffed with a clinical investigator (MD) qualified for performing clinical research. The centre/site where the medical images

		and/or medical parameters are acquired.
	AD	Alzheimer Disease
	Administrator	A role in a computer system, which has complete privileges to perform any action without restriction.
	ADNI Protocol	Set of roles which define the acquisition of the MR imaging sequences according to the Alzheimer Disease Neuroimaging Initiative (ADNI).
	Algorithm	A step-by-step problem-solving procedure.
	Anonymization	The process removing or obfuscating information from data which could be used to identify the concerned person or source.
	Artifacts	Artifacts are misrepresentations of tissue structures seen in medical images produced by modalities such as Computed Tomography, and Magnetic Resonance Imaging. These artifacts are caused by a variety of mechanisms, such as the underlying physics of the energy-tissue interaction, data acquisition errors (mostly from patient motion) and reconstruction algorithm's inability to represent the anatomy.
	Authentication	A security measure designed to protect a communications system against acceptance of a fraudulent transmission or simulation by establishing the validity of a transmission, message, or originator.
	Authorisation	This facilitates fine-grain control of privileged operations, such as accessing restricted areas of the operating system and self-restricted parts of the neuGRID applications.
B		
	Bandwidth	The maximum throughput, in bits per second, of a physical communication path in a digital communication system.
C		
	Clinical Biological Data	Data or measurements collected from clinical biological sources, which are commonly stored in files or databases.
	Conceptualisation	To establish a common language and models among users, developers and the system deployment teams.
	Cortical Thickness	Cortical Thickness refers to the quantitative measurement of the thickness of the human cerebral cortex.
	Core Lab	A centre that collects data from various acquisition centres and checks for commitment to a given scan protocol, image quality and completeness of the acquired images. The DACS (Data Archiving and Computing Site) is synonymous with this concept.
	CSF	Cerebrospinal Fluid
	CTR	Normal elderly control subjects
D		
	DACS	Data Archiving and Computing Site
	Data Model	An abstract model that describes how data is represented, stored and accessed.
	Data Registration	The process of inputting new data to a data store
	Data Store	A repository where data is stored.
	Deliverable(s)	A document (or a set of documents) necessary to govern and monitor a European project
	Dependencies	In workflows or pipelines, dependencies refer to the tasks which provide input data to a specific task.
	DICOM	The Digital Imaging and Communications in Medicine (DICOM) is a standard for distributing and viewing almost any kind of medical image.

	Download	The process of copying remote data to a local data store.
	DTI	Diffusion Tensor Imaging.
E		
	Ethical Compliance	The management of compliance with the ethical framework that has been adopted by the project.
	EDSS	Expanded Disability Status Scale.
F		
	Face Scrambling	An anonymization technique which obfuscates the facial features of an image.
	FBF	Fatebenefratelli.
	Field Inhomogeneities	A disturbance of the field homogeneity, because of magnetic material, technical problems or scanning at the edge of the field.
	FLIRT	Automated linear (affine) registration program.
	fMRI	Functional Magnetic Resonance Imaging (fMRI) is a type of specialized MRI scan. It measures the haemodynamic response related to neural activity in the brain or spinal cord of humans or other animals. It is one of the most recently developed forms of neuroimaging.
	FSL	FSL is a library of analysis tools for fMRI, MRI and DTI brain imaging data.
	Functional Assessment Questionnaire Total Score	Psychiatric Rating Scales for dementia evaluation.
G		
	GDScale Total Score	Psychiatric Rating Scales for dementia evaluation.
	Global CDR	Psychiatric Rating Scales for dementia evaluation.
	Gradwarp	A system specific correction of image geometry distortion due to gradient non-linearity.
	Grid Computing	A form of distributed computing, where the system is created by forming a virtual organization over geographically distributed heterogeneous clusters. Grids can be both data centric and computation centric. In a data centric Grid, geographically distributed heterogeneous data sources are linked, and users can access and use data irrespective of location in the Grid. A Compute Grid is a Grid which unifies the processing capabilities distributed in heterogeneous sites.
	GUI (Graphical User Interface)	A graphical user interface (GUI) is a human-computer interface that uses windows, icons and menus and which can be manipulated by a mouse.
I		
	Image	In the context of neuGRID, an image is a MRI brain scan.
	Image Acquisition	The process of acquiring a scan from a patient.
	Image Scrambling	The process of removing or obfuscating features from an image, in order to anonymize it.
	Interleaved scanned series	A typical MRI sequential collection of raw data from a multiple excitations approach.
	Inter-Slice Movement	Artifacts consisting in the misalignment between two or more slices within a stack and/or movement within a slice.
K		
	KI	Karolinska Institutet.
L		

	Linux	An Open Source Computer operating system, similar to Microsoft Windows, Mac OSX, Unix etc.
	LONI	The Laboratory of Neuroimaging (LONI) is a research centre dedicated to studying the relationship between brain structure and function using image data. It is based at the University of California, Los Angeles.
	LORIS	The LORIS system (On-line Research Imaging System, formerly NeuBase) was originally implemented for the collection, management, and processing of the imaging data acquired in a multi-centre Alzheimer's Disease project (AddNeuroMed).
M		
	Metadata	Information about data which may include, Acquisition Plane, Acquisition type and Field strength.
	Meta-model	In terms of software engineering, this is the development of the models and theories that are useful for modeling a predefined class of problems.
	MCI	Mild Cognitive Impairment
	MEG	MagnetoElectroEncephalography
	MMSE Total Score	The mini-mental state examination (MMSE) is a brief 30-point questionnaire test that is used to screen for cognitive impairment. It is commonly used in medicine to screen for dementia. (source: http://encyclopedia.thefreedictionary.com/Mini+Mental+State+Examination)
	MPRAGE	Magnetization-prepared rapid acquisition gradient-echo.
	Modality	Modality is used to describe the various classes of imaging devices used to image the internal structures of object. The modality is mostly differentiated by the physics used to create the image. For example Magnetic Resonance. and Computed Tomography are different modalities (source http://www.angelfire.com/co2/whatdicom/yong.html.) This includes the various types of equipment or probes that are used to acquire images of the body.
	Modified Hachinski Total Score	Psychiatric Rating Scales for Dementia evaluation.
	MRI	Magnetic resonance imaging (MRI) is a medical imaging technique most commonly used in radiology to visualize the structure and function of the body. MR imaging uses a powerful magnetic field to align the nuclear magnetization of (usually) hydrogen atoms in water in the body. Radiofrequency fields are used to systematically alter the alignment of this magnetization, causing the hydrogen nuclei to produce a rotating magnetic field detectable by the scanner. This signal can be manipulated by additional magnetic fields to build up enough information to construct an image of the body.
N		
	NIFTI	Neuroimaging Informatics Technology Initiative
	Non-uniformity Correction	Non-uniformity Correction: A mathematical method for the automatic procedure that reduces residual intensity non-uniformity due to the wave or the dielectric effect (source:http://www.fields.utoronto.ca/programs/scientific/08-09/mathoncology/courses/3_FreqTransforms.pdf .)
	NPI-Q Total Score	Psychiatric Rating Scales for Dementia evaluation.
P		

	PD	Proton Density MRI Acquisition.
	PDF	The Portable Document Format (PDF) is a popular way to store and transmit electronic documents.
	PET	Positron Emission Tomography. PET is a nuclear medicine medical imaging technique which produces a three-dimensional image or map of functional processes in the body.
	Platform	In relation to hardware, platform often describes the set of hardware components that make up the computer itself, that the software is written to target (often just described as "written for an architecture.")
	Platform-independence	A property of a system, where the system is not tightly coupled with a specific platform.
	Pipeline	See definition for workflow.
	Pre-processing	Steps that are put in action in order to correct image artifacts.
	Protocol	A set of rules which is used by computers to communicate with each other across a network.
	Provenance	The maintenance of the history of workflow specifications and their evolution between different stages.
Q		
	Quality Control	The process of ensuring that a certain system or product meets user requirements.
	Querying Language	A computer language used to query data or information from a data store.
	Quota	An allotment of a certain share from a resource. (i.e. Disk quota, bandwidth quota.)
R		
	Realigning correction	Algorithm for Transformation from Native space to Stereotactic space and vice versa.
	Registration	In computer vision, sets of data acquired by sampling the same scene or object at different times, or from different perspectives, will be in different coordinate systems. Image registration is the process of transforming the different sets of data into one coordinate system. Registration is necessary in order to be able to compare or integrate the data obtained from different measurements. Medical image registration (e.g. for data of the same patient taken at different points in time) often additionally involves elastic (or nonrigid) registration to cope with elastic deformations of the body parts that are imaged. Nonrigid registration of medical images can also be used to register a patient's data to an anatomical atlas, such as the Talairach atlas for neuroimaging.
	Research Set	A set of brain scans which will be used as input to a neuro-imaging pipeline.
	Raw Data	Data that is used as the initial input to workflows for processing.
	Raw Files	Files which are stored in the default neuGRID format.
	Raw Output	The output as it comes directly from a given workflow, before further analysis has been carried out.
S		
	Security	This aims to protect information and information systems from unauthorized access, use, disclosure, disruption, modification or destruction.
	Sensitive individual data	Any information regarding a physical or mental health condition is considered to be sensitive.
	Service	An independent, self contained module in a Service oriented Architecture. It

		provides a single functionality, which is exported via standardized interfaces (WSDL) and communicates via standardized communication protocols, mainly SOAP.
	Sequence type	All the different MRI acquisition protocol (e.g: T13D, T2, PD, DTI, PET, PIB, fMRI and others.)
	Signal Noise Ratio	Signal-to-noise ratio compares the level of a desired signal to the level of background noise. The higher the ratio, the less obtrusive the background noise is.
	Slice timing correction	Algorithm that will correct for the differences in each slice's acquisition time (e.g: BVQX v1.6.)
	System Maintenance	The modification of a system to correct faults, to improve performance, or to adapt the system to a changed environment or changed requirements.
	Service Oriented Architecture (SOA)	Service Oriented Architecture is an architecture which uses loosely coupled ad-hoc collection of independent services. Each service is self contained and provides a specific piece of functionality. This architecture is popular in large-scale distributed systems, primarily because it is robust, scalable, extensible and potentially resistant to failure.
	SOAP	Simple Object Access Protocol, a standardized means of inter-communications between services and clients in a SOA.
	Source Code	The human readable logic of a computer program.
	Story	A high-level model of several user requirements.
	Surface rendering tool	Algorithm that use a three-dimensional representation of geometric data.
T		
	T1	Spin-lattice relaxation time, known as T_1, is a time constant in Nuclear Magnetic Resonance and Magnetic Resonance Imaging. T_1 characterizes the rate at which the longitudinal M_z component of the magnetization vector recovers.
	T2	Spin-spin relaxation time, known as T_2, is a time constant in Nuclear Magnetic Resonance and Magnetic Resonance Imaging. T_2 characterizes the rate at which the M_{xy} component of the magnetization vector decays in the transverse magnetic plane.
	TBSS	Tract-Based Spatial Statistics, a voxel-wise analysis of multi-subject diffusion data.
	TE value	Time echo.
	TR value	Repetition Time.
U		
	Unwarping correction	Pre-processing step in which there is an estimate and correction of the "static" deformation field, yielding an unwarped (to some true geometry) version of the MRI acquired time series.
	Upload	The process of copying data from a local data store to a remote data store.
	User Collections	Lists of images collected by the neuGRID users.
	Use-case	A use case describes what can be done with a system. This technique is often used to capture a system's functional requirements through the description of a set of usage scenarios.
	User-friendly	Easily operated and understood by means of a straightforward guide in jargon-free language.
V		
	VBM	Voxel-based Morphometry. VBM is a neuroimaging analysis technique that

		allows investigation of focal differences in brain volume.
	Voxel	A volume element, representing a value on a regular grid in three dimensional space. This is analogous to a pixel, which represents 2D image data.
	Visual inspection process	A qualitative assessment of the results derived from a MRI pipeline analysis.
	Vumc	VU medical centre.
W		
	Webinar	A webinar is a collaborative meeting, analogous to a seminar, where the participants attend from remote locations linked by the Internet.
	Workflow	The defined series of tasks within an organization to produce a final outcome.
	Workpackage	Subset of a project that can be assigned to a specific party for execution.
	WSDL	Web Service Description Language, the standardized means of describing and exporting service functionality in a SOA.

3. The Actors In neuGRID

A key part in analyzing the requirements for any system is identifying the types of users that will make use of it in some way. This allows the requirements team to ensure that they do not miss out features that a small number of members within a wider user community may desire. By modeling the ways in which the Actors interact with the system that is being designed, a range of important conclusions can be drawn. Practically speaking, this may mean ensuring that representative members from each group of Actors are present during requirements elicitation sessions and that they review any specifications that are produced. This section briefly describes the Actors that have been identified in neuGRID and gives some profiles of projects members from within the neuGRID consortium that are members of these groups.

Research Leaders

Team leaders who need to monitor the progress, resource usage and perhaps distribute research studies to a research team.

Example Profile

Giovanni is a Neurologist and Vice-Scientific Director of IRCCS-Fatebenefratelli Hospital Brescia (Italy). His main research interests are focussed on the exploitation of intensive computational neuroanatomy algorithms in translational neuroscientific research and in the dissemination of new brain image analysis tools to clinical neuroscientists and clinical physicians. He works with his team to carry out research and communicate findings to the wider community through publications and other scholarly activities.

Example publications:

1. Frisoni GB et al., Neuroimaging tools to rate regional atrophy, subcortical cerebrovascular disease, and regional cerebral blood flow and metabolism: consensus

paper of the EADC, Journal of Neurology, Neurosurgery, and Psychiatry 2003;74:1371-1381.

2. Ashburner J, Frisoni GB, et al., Computer-assisted imaging to assess brain structure in healthy and diseased brains. Lancet Neurol 2003;2:79-88.

3. Frisoni GB, et al., Detection of grey matter loss in mild Alzheimer's disease with voxel based morphometry. Journal of Neurology, Neurosurgery, and Psychiatry. 2002; 73(6): 657–664.

4. Frisoni GB, et al., In vivo neuropathology of the hippocampal formation in AD: a radial mapping MR-based study. NeuroImage 2006;32(1):104-10.

5. Frisoni GB, et al., The topography of grey matter involvement in early and late onset Alzheimer's disease. Brain. 2007;130:720-30.

Researchers

Individual members of the research team who will use neuGRID during their day-to-day research work. These may interact with the system in different ways depending on their experience and the nature of the research that they are carrying out. Broadly speaking the following groups of users has been identified:

Basic User

This group represents users who have a certain level of computing expertise, but are mainly content to use software as it was installed and are not inclined to customize environments to their needs. They expect a reasonably straightforward user interface through which they can carry out their day to day tasks.

Example Profile:

Olof is a PhD student at K.I. with Professor Wahlund since 2007. His research area is the anatomy and volumetry of the frontal lobe. His main research project involves frontal lobe dementia, which can be investigated by the shrinking of various small structures in the brain such as the putamen and caudate. A typical day at SMILE (Stockholm Medical Imaging Laboratory and Education) for Olof involves using the Hermes system to manually trace the 3D outline of the brain structures of interest, sometimes importing more images into the system (the material consists of 600 patients being scanned at intervals of a year or so) to work on. Even though Olof has studied some "computer science", he knows very little of computer programming and more complex operations. He can navigate inside a Windows system (but not add a printer, for instance) and do some basic tasks on a Linux system (cd, ls -- grep is the limit of his knowledge). The Hermes system has GUIs with buttons (and a Unix platform which the average user needs not bother with, usually), which he handles expertly. Olof also knows how to run FSL and FreeSurfer, but cannot write scripts at all, on any platform.

Intermediate User

This user is similar to the Basic user but requires a little more flexibility in the way that they work and want to have more control over their environment. They may wish to extend

existing workflows or make some changes to settings or the way in which they are configured.

Example Profile:

Michela is a PhD student at IRCCS-FBF with Dr. Frisoni since 2004. Her research area is the control, pre-processing and post-processing of diffusion tensor images (DTI) with specific tools for the analysis of the weighted images. A typical day for Michela involves the usage of the FDT (FMRIB's Diffusion Toolbox that is part of FSL system) to perform scanner pre-processing (e.g. averaging of multiple acquisitions, removal of images affected by large artifacts). These initial steps are usually done manually by Michela. Then, in order to correct stretches and shears due to current distortion in the images she runs different command line utilities. A probabilistic diffusion model of the corrected data is generated and finally a probabilistic tractography map is outputted for each image. Michela is an end user that is able to run programs from the command line shell and knows how to write bash scripts or simple programs in a language such as Matlab, Perl or Python on a range of different platforms.

Advanced User

This group of users wants full control over their work environment. They may wish to construct new tools or adapt existing ones for other purposes. It is likely that such users have a high degree of experience and probably a good understanding of computing techniques. The flexibility to do what they want is paramount to this group of users and they do not wish to be constrained in their work by the system. They may also perform tasks that are covered by the Basic and Intermediate user roles from time to time.

Example Profile:

Researchers at VUmc performed volume changes over time for the brain and the hippocampi of MCI patients. For this reason the hippocampi where manually outlined at baseline. The mask for the hippocampi was converted into Analyze to combine them with the original images. For the brain volume change and the change in hippocampi volume the brains on follow up were registered to BL. The Fluid algorithm of the dementia research group of London (DRG) was separated from its surrounding GUI and executed on the hippocampi and the whole brain. This resulted in a relational comparison between the brain volume change, the hippocampi volume change and the MMSE values of the used data. To perform this analysis a number of scripts were used. Some existing programs of other research centre were slightly modified and used in a fashion that better matched the used data.

Pipeline Developers

The developers of new research pipelines need to integrate them within the system in order to provide facilities to researchers. These are very technical users and share many similarities with the Advanced User. Given the cutting-edge nature of their work, it is likely that they may go beyond this profile and may require access to development and debugging tools. They will also require a good degree of flexibility from the system.

Example Profile:

Alex is a researcher with a long track record in the development and validation of image processing algorithms and pipelines for the quantitative analysis of brain MRI. Typically the development of novel algorithms relies on rapid prototyping and testing cycles, in which algorithm or parameter changes are implemented, executed, and their results observed. This requires relatively low-level, "hands-on" access to the system, with the ability to rapidly modify modules in a pipeline and/or modify the pipeline itself, and execute immediate tests. For thorough testing and validation though, an algorithm or pipeline may need to be run on tens or hundreds (or even more) cases; and/or collection of scans may need to be processed using a range of parameter values in order to establish optimal parameter values. This latter case requires the ability to process large numbers of scans and/or a set of scans with a possibly large range of pipeline configurations.

Image / Data Input Managers

Managers and administrators that work to upload and manage the data stored within the system.

Example Profile:

Anna is a PhD student at IRCCS-FBF with Dr. Frisoni since 2004. She is a key figure in the neuGRID Data Archiving and Computation Centres (DACS). Her main task will be to ensure the correct uploading of both images and data from the data collecting sites (DCS). She will have to maintain contact with the "data input managers" in the other neuGRID core labs in order to adopt procedures for standard data handling. Before the upload of each data set she will perform a quality control procedure. A key aspect of the data input Manager is to organize the available data for use by the neuGRID community providing different levels of access and maintaining data integrity. She will ensure proper data management and the saving of local mirror copies of data. Finally, she has a deep knowledge of MySQL because the data management will be conducted through the LORIS relational database system.

The data managers at VUmc are collecting data from various sites. From each site firstly a dummy run is requested. This dummy run is checked for image quality and commitment to the scan protocol (both by the data manager and a Radiologist). After one or more rounds to establish the best acquisition parameters, scan parameters are frozen. After the successful dummy procedure the site can send images to VUmc. Each scan is checked whether it fulfils the parameters agreed on at the dummy run, whether the image quality is good enough, whether the required patient information (random codes) are in the file header and for other quality indicators. After these checks the data can further be anonymized and will be sent to an image archive.

System Administrators

Technical support operators are responsible for installing, monitoring and generally administrating the system.

Example Profile:

Marco is a graduate in Mathematics and has started his PhD at IRCCS-FBF with Dr. Frisoni from December 2009. Marco will maintain and operate the neuGRID computer system and its network. He is usually charged with installing, supporting, and maintaining servers or other computer systems. This entails a good knowledge of operating systems and applications, as well as hardware and software troubleshooting. An important thing is that he must also have a detailed knowledge of the purposes for which people use the neuGRID platform and most importantly, he has strong problem solving skills. Marco has already demonstrated a blend of technical skills and responsibility.

4. Description of an End-To-End Example
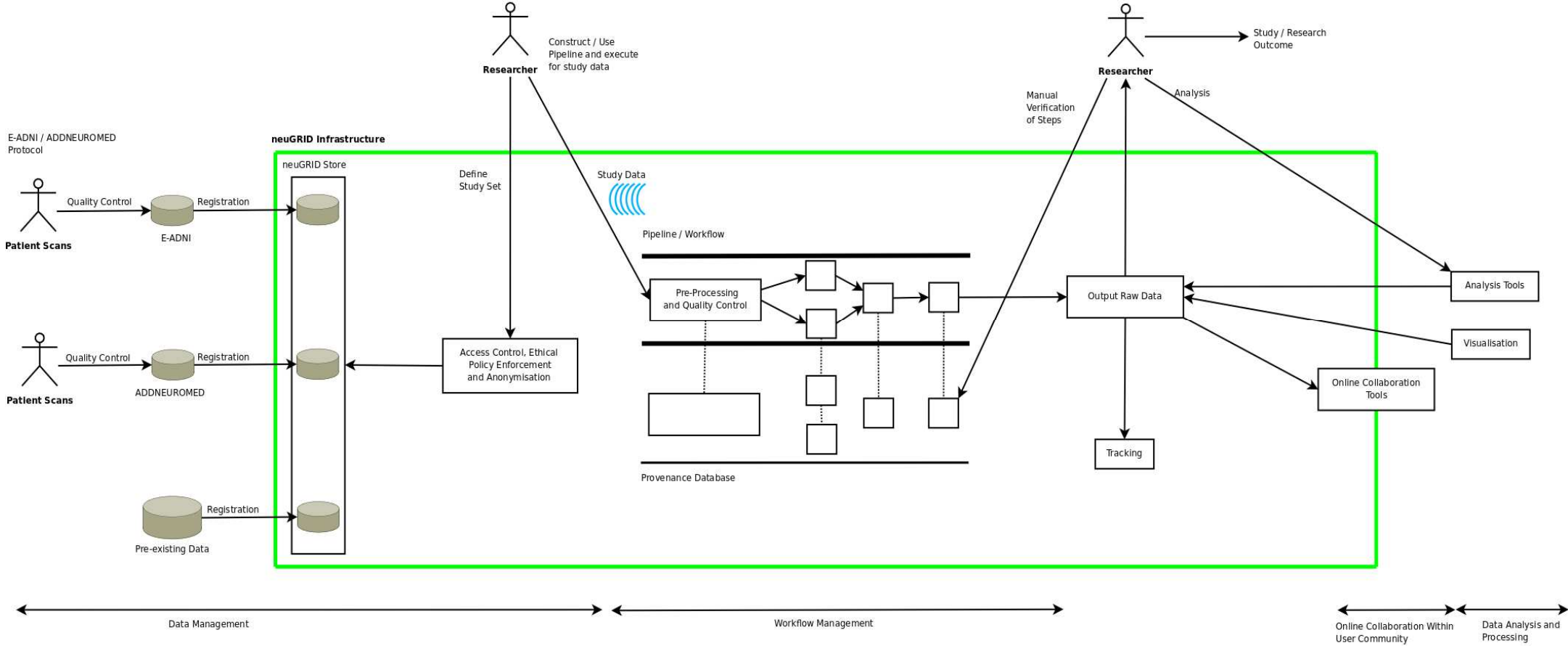
This section describes a potential end-to-end example of the use of the neuGRID platform. This sets the scene for the following section in which the requirements are specified, by identifying the key stages in the operation of the infrastructure. In neuGRID, users may pass through the following stages to carry out their analysis on a set of images:

1. Data registration into the neuGRID Store, data management and quality control.
2. Data access, querying and browsing.
3. Workflow development, execution and management.
4. Validation of results and workflows using the provenance data.
5. Sharing workflows, histories and results.
6. Visualization of the results.

The first stage in the analysis cycle is to register images in the neuGRID store that have been collected from the hospital data acquisition system or have been imported from other research projects. For example, a new acquisition centre may wish to make use of the neuGRID infrastructure to share data within the wider research community. Existing data is thus put through a process which enforces quality control, formatting and ethical compliance. Finally the data is integrated with the neuGRID standard data model, which enables other researchers to access it and carry out their research. As new data sets are acquired they go through an initial local quality control step before passing through the same system-wide quality control, formatting, ethical compliance and data model integration processes that the pre-existing data goes through.

The role of the second stage in the analysis process is to make the data browsable through automated querying tools. Therefore, an appropriate data access mechanism needs to be put in place. For example, a researcher may be interested in a rare form of a disease and wants to do a statistically meaningful analysis. Unfortunately the researcher's institution does not have sufficient images to make this possible. The user will interact with the system using the neuGRID store, to search for and to identify an appropriately large set of images from a group of hospitals that match the required criteria. At this stage access controls and ethical policies are fully enforced to protect sensitive data. The researcher then uses the system to submit the study set for analysis through a workflow.

Once the data has been imported into the neuGRID system and users are able to access and query it, they may like to carry out studies and data analyses to find results of interest. Workflow development is a methodology that can be used to represent user preferences for an automatic analysis of data and this is the third stage in the analysis chain. Users may create workflows and then execute them more quickly on the distributed resources provided by the Grid. The workflow development and execution is an important stage in the analysis life cycle in the neuGRID project. For example, a researcher may wish to run a comparative analysis using a study set of 3000 MRI scans stored in geographically distributed medical centres.

It is important that the results are generated in a timely fashion as the researcher may have a number of different studies to carry out that week. The user may as the available data grows interact with the system to choose a study set of perhaps 3000 images, selects the pipeline or workflow through which the analysis will take place and starts the analytical process. Users do not have to use the workflows and study samples that have been developed previously; they can also construct new workflows. For example, a new image analysis methodology may be developed and a researcher may wish to build a workflow to run it. Using an interactive creation tool the user can construct a workflow and specify some initial settings. The user

may also create a record which describes the workflow and gives other users information about its purpose and access controls. The system allows different versions of the workflow to be created, tested and released when they are ready for use by other researchers.

Simply creating and executing workflows is not enough on its own however. It is important that results, as and when required, should be reproduced and reconstructed using past information. The maintenance of the history of workflow specifications and their evolution between different stages is known as provenance and may help in the verification of results using audit trail information. For example, a workflow yields some surprising and possibly significant results. A researcher may wish to confirm that the results are accurate and identify any mistakes that may have been made. By analysing all the intermediary image sets and workflow execution logs the user is able to manually verify that the results were incorrect. It may be found that the error was due to a specific group of images interacting badly within the workflow. The user can then annotate the workflow so that other users are warned if they attempt a similar analysis.

Sometimes it may not be enough to reproduce the results. It may also be necessary to validate and, if required, reproduce the workflow that has been used to obtain the results. This makes users confident not only in the results that have been produced but also in the process that led them to generate these results. For example, a user may create a new workflow and run it on a test data set. At each stage in the execution of the workflow, the intermediary images or data are stored and a full provenance track is kept. After results have been produced, the user can examine the provenance to check that each stage of the analysis was completed correctly. The raw results can then be exported into the user's preferred analysis tool and the whole process can be added to the researcher's history for future reference. Initially the new workflow may produce some poor results during testing. The researcher therefore can inspect the logs of the workflow execution and locate the problem. The user can then interact with the system to make changes to the relevant settings and re-run the test study. This time the process may run correctly and meaningful results may be produced. Without the mechanism to validate workflows, it would not be possible to correct the process and generate accurate results. Therefore the validation of results and workflows are two significant requirements that should be addressed in the neuGRID system.

Once a workflow has been developed and verified, a user should be able to share it with other researchers in the field. The user may make the workflow available to a team or group of users from a partner institution or project. This will save time, effort and resources from other teams and they will not have to reinvent the pipelines which have been produced by their peers or partners. Users may also be able to share results and histories of their analysis processes. For example, a user might interact with the system to search existing studies and to compare, contrast and validate their results against research from other groups. This process helps the researcher to identify an error in their methodology and prevents them from making any embarrassing claims. A researcher could have carried out a similar study six months ago and may be worried that it too, might have been influenced by a similar error. The user can look up their research history and identify the appropriate study. The original process can be re-run on the original data set using the stored settings and pipeline configuration. This allows the researcher to confirm that the previous results were correct.

This abstract example is useful in describing the system components within neuGRID. It is also important however, to understand how more detailed examples of real research processes can enrich this conceptual framework. The remainder of this section will consider a real-

world use-case in greater detail and show how this is useful in identifying potential requirements.
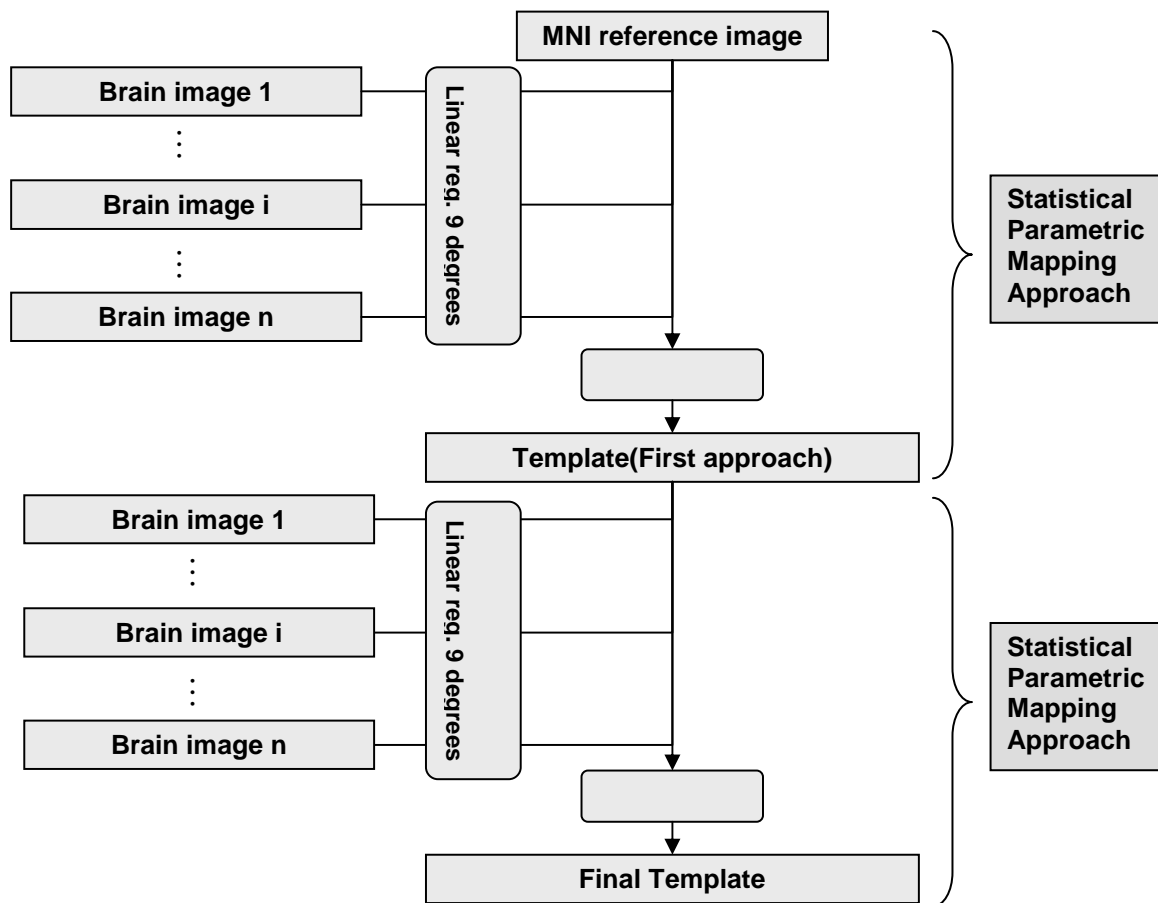

Figure 3: An example of creating a study specific template for VBM.

A Real Research Example

A VBM analysis of Alzheimer's and Frontal Lobe Dementia patients is calculated using a template based on the given patients groups. Before the pipeline is used, it is tested on the MPRAGE (magnetization-prepared rapid acquisition gradient-echo) scans of Alzheimer's and MCI patients of the ADNI data. When the pipeline appears to work correctly on the ADNI data, the pipeline is used for a set of Alzheimer's and Frontal Lobe Dementia patients that does not yet reside on neuGRID. At the end when the results looks promising the data and pipeline are shared with the neuGRID community. This process is shown in figure 3.

The pipeline used consists of two parts:

- The creation of the template (which is based on images of the populations that are compared; not the VBM process itself.)

- The VBM analysis

1. Creation of the template.

The template creation is done by registering/aligning all scans to the MNI template (a template that comes with FSL in the Nifti file format) using 9 degrees of freedom. All images

registered to this template are re-aligned together to create a new template. This new template is now used as template instead of the MNI template. A final template is build using the template of the first iteration.

2. The VBM analysis itself.

The VBM analysis is a regular VBM analysis on the data.

Steps necessary:

1.   Select a list of MCI and Alzheimer patients of the ADNI dataset.

2.   Register/Align (9 degrees of freedom) the selected scans to the MNI template which comes with FSL (in the Nifti file format).

3.   Re-aligned all registered images to create the template.

4.   Register/Align (9 degrees of freedom) the selected scans to the template generated in the first run and re-aligned the registered images to create the final template.

5.   Perform a VBM analysis through the Statistical Parametric Mapping Approach on the selected scans using the template generated to compare MCI with Alzheimer's.

6.   Validate whether the pipeline works correctly.

7.   Upload a new set of scans of Alzheimer's and Frontal Lobe Dementia patients to be used as a private date set to neuGRID and use this scans to generate a template of these new dataset (using step 2 up to 4) and run a VBM analysis on the scans.

8.   Make the uploaded dataset public.

9.   Make the generated pipeline public.

Indicative User Requirements:

- The user should be able to generate a pipeline.

- The user should be able to test a pipeline with existing data.

- The user should be able to test a pipeline on their data.

- A user has to be able to upload an initial template (e.g. MNI template in .mnc file format.)

- It should be possible to convert the original Dicom files (of ADNI) Into Nifti.

- It should be possible to use the registration program FLIRT (FSL) to register a set of images to a given template.

- It should be possible to perform image calculations (for example through the "avwmaths" program of FSL.)

- It should be possible to perform quality control on the registrations.

- It should be possible to realign all registrations into a new image.

- It should be possible to use a generated image (first approach template) as a new template for registration.

- It should be able to perform a VBM analysis on the data.

- The user should be able to make its data public to a given community.

- The user should be able to make its pipeline public to a given community.

It is clear that detailed examples like this are useful in capturing a well rounded set of requirements. During requirements meetings, such examples have been presented to the requirements team by clinical researchers and the implications for neuGRID discussed. This has been of great benefit in compiling the requirements specification that is considered in the following section.

5. The User Requirements

Each segment of the user requirements specification begins with a Story. The relevant use-cases that are contained within it are described and then broken down further to form individual user requirements. The numbering scheme allows the hierarchical relationships between Stories, Use-cases and Requirements to be easily traced. The high-level Stories are indicated by the S prefix and Use-cases are given the prefix U. Individual Requirements are denoted by the R prefix. The prioritization scheme focuses on Essential, Desirable and Optional requirements and is based on the variation of the MoSCoW technique [1] that was described in section 1.

Essential requirements are those which are absolutely vital to the production of a functional infrastructure. Desirable requirements are those that whilst not vital, would provide important functionality to users and a reasonable proportion of these should be implemented. Optional requirements are those that might be useful but don't fit into the previous two categories and will probably be the last to be implemented if time / budget allows. The individual use-cases and requirements have been prioritised using this scheme. The aim of this is to relate the priorities of finer-grained requirements within the context of the broader use-cases. This is not always easy to achieve and there are bound to be some conflicting demands. It was felt however, that this provides an insight into how users think about and assess the priority that should be given to the various components of neuGRID.

The requirements that were initially captured in D9.1 have been further refined and clarified through a revision process in which all neuGRID project partners have participated. Every effort has been made to ensure that there is internal consistency between the requirements that have been documented in this section. In some cases, it is possible however, that similar requirements may have been assigned different levels of priority because of the use-case

context in which they are placed. In this situation the highest level of overall priority should be taken, as this will likely best reflect the importance of the requirement within the project as a whole.

Where E = Essential D = Desirable O = Optional.

S1.   Data Registration into the neuGRID Store, Management and Quality Control:

A new acquisition centre wishes to make use of the neuGRID infrastructure to share data within the wider research community. Existing data is put through a process which enforces quality control, formatting and ethical compliance. Finally the data is integrated with the neuGRID standard data model, which enables other researchers to access it and carry out their research. As new data sets are acquired they go through an initial local quality control step before passing through the same system-wide quality control, formatting, ethical compliance and data model integration processes that the pre-existing data goes through.
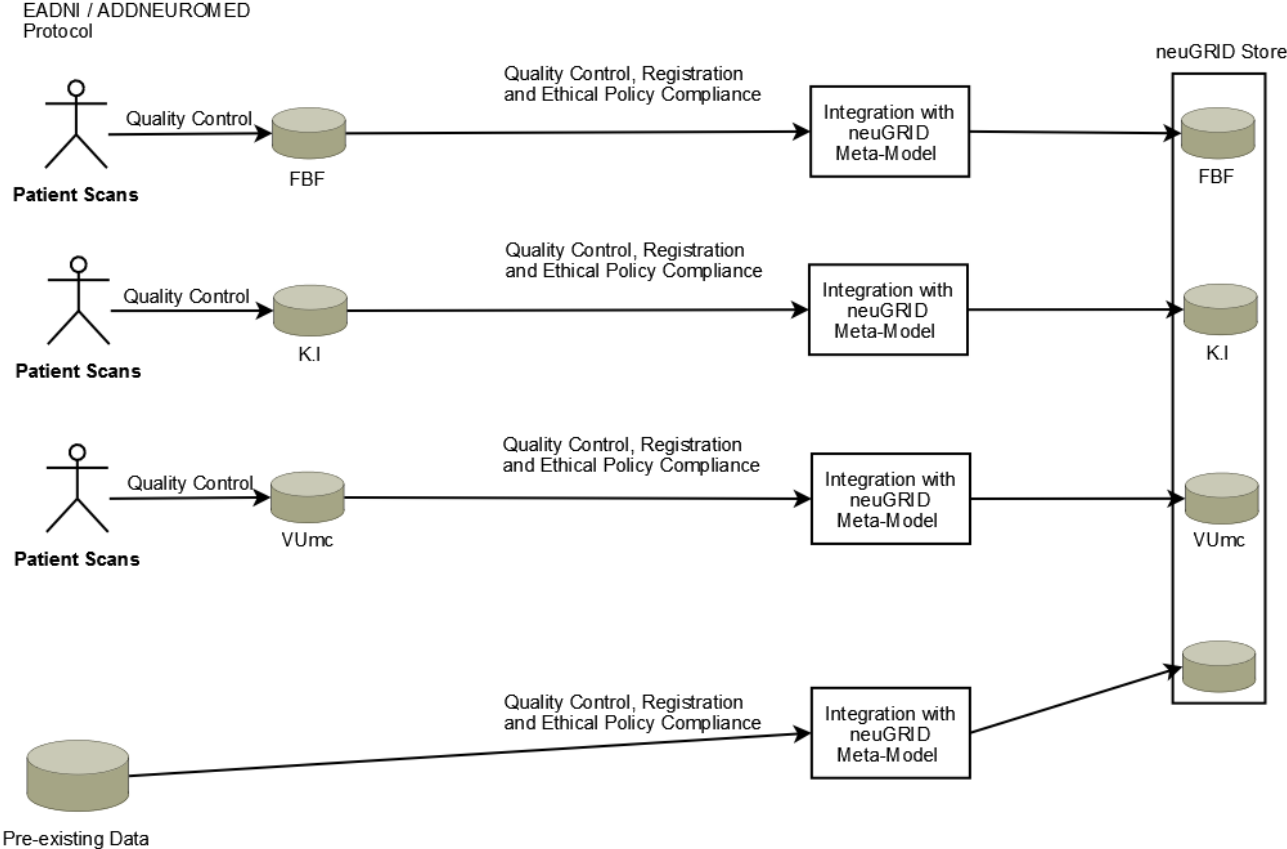
Indicative Use-cases:

U1.1 Perform quality control, ethical compliance (including appropriate anonymisation) and upload the new data sets into the system. E

User Requirements:

R1.1.1	An interface is required for the upload of images and data sets into data stores. This should allow images to be imported into a "storage area"	D

	(perhaps a drag and drop interface or list of file names.)	
R1.1.2	A basic QC viewer which allows comparison between different sets of images. It should be possible to show a DICOM Dump of at least one image from each series to check for any information that has leaked through the anonymization steps.	E
R1.1.3	The ability to record the outcome of manual QC validation.	D
R1.1.4	A tool to delete images of inferior quality from a set.	E
R1.1.5	Provide software to those uploading data into neuGRID that enables the anonymization of data sets. The ability to easily anonymize the principal image fields defined by neuGRID ethical guidelines as approved by the independent ethics committee set up for the neuGRID project (if they are not already treated in some previous steps) ensuring that no identifiable patient information crosses the network (Images Scrambling and anonymization.)	E
R1.1.6	The ability to adapt to new ethical policies would be desirable. The system should allow new anonymization methods to be applied as privacy standards evolve.	D
R1.1.7	Logs should be kept of what was uploaded and by whom. A tool to save the set (list of files) which will be uploaded (in case the upload is delayed or interrupted for some reason) would be useful.	E
R1.1.8	A means of preventing duplicate data upload.	D
R1.1.9	The ability to visualize image(s) metadata (Acquisition Plane, Acquisition type and Field strength.)	D
R1.1.10	The ability to visualize image field inhomogeneities, subject position and artifacts.	D
R1.1.11	The possibility to perform corrective steps on images.	D
R1.1.12	Security and authentication of users should be enforced before images can be uploaded.	E
R1.1.13	Documentation should be provided that defines quality control and ethical compliance.	E
R1.1.14	Quality control should be done automatically where possible (number of images in series ranges of TE and TR values, pixel sizes, used coils etc.)	D
R1.1.15	It should be possible to do some manual quality control: visual inspection on Signal Noise Ratio, movement artifacts, inter-slice movement (for interleaved scanned series) etc. to assist the visual inspection process, an orthogonal view should be provided so that checks can be made for missing slices and artifacts between the slices.	D
R1.1.16	Something similar to the Linux/Unix strings command should be executed on at least one image in each series, to check for hidden patient information.	E
R1.1.17	A surface rendering tool should be made available and used to show the effect of any face scrambling algorithms that have been applied to images.	O
R1.1.18	It should be possible to trace back data on neuGRID to the original information source (perhaps at the core labs.)	D
R1.1.19	When data is uploaded into the neuGRID storage area access restrictions should be specified.	E

U1.2 For pre-existing data, perform quality control, ethical compliance (including appropriate anonymisation), format standardisation and upload the data sets into the system. E

User Requirements

R1.2.1	An interface is required for the upload of images and data sets into data stores. This should allow images to be imported into a "storage area" (perhaps a drag and drop interface or list of file names.)	D
R1.2.2	A Basic QC viewer which allows comparison between different sets of images is necessary.	E
R1.2.3	A tool for deleting images of inferior quality from a set is required.	E
R1.2.4	Provide software to enable the anonymization of data sets. The ability to easily anonymize the principal image fields as defined by neuGRID ethical guidelines approved by the independent ethics committee set up for the neuGRID project (if they are not already treated in some previous steps) ensuring that no identifiable patient information crosses the network (Images Scrambling and anonymization.)	E
R1.2.5	Logs should be kept of what was uploaded and by whom. A tool to save the set (list of files) which will be uploaded (in case the upload is delayed or interrupted for some reason) would be useful.	E
R1.2.6	A means of preventing duplicate data upload.	D
R1.2.7	The ability to visualize image(s) metadata (Acquisition Plane, Acquisition type and Field strength.)	D
R1.2.8	The ability to visualize image field inhomogeneities, subject position, artifacts.	D
R1.2.9	The possibility to perform corrective steps on images. Perform specific corrective steps for each kind of acquisition: Gradwarp and Non-uniformity correction for MRI images. Realigning, unwarping and slice timing correction for fMRI images. Pre-processing steps for PET images with particular attention to the ADNI protocols. (See http://www.loni.ucla.edu/ADNI/Data/ADNI_Data.shtml for more information.)	D
R1.2.10	Security and authentication of users before images can be uploaded or managed. There should be a certificate-based system to identify users and to perform access control.	E
R1.2.11	Documentation should be provided for performing appropriate quality control and ethical compliance on data sets.	D
R1.2.12	Provide software for format conversion.	D
R1.2.13	The system should allow new anonymization methods to be applied as privacy standard evolve.	D

U1.3 Standardize all uploaded data to comply with the neuGRID data model. E

User Requirements

R1.3.1	Define a set of image data rules for neuGRID. A checklist should be provided which itemizes the image parameters that need to be removed or added to the set to make it comply with the neuGRID standard data model (e.g. date of birth was removed from image X.)	E
R1.3.2	A form/tool to allow complementary information to be written to the uploaded set, perhaps with a reporting mechanism.	D
R1.3.3	The ability to remove parameters not included in neuGRID standard.	E

R1.3.4	An image data archive tool which allows the suitability and quality of images to be determined before they are uploaded into the neuGRID data store.	D
R1.3.5	Quality control needs to be done both locally (each centre should only upload high quality/usable images) and centrally (all uploaded images should undergo quality control with a unified criteria).	D
R1.3.6	Provide software to check if uploaded data sets conform to the neuGRID data model, and provide tools for conversion if required. There should be tools to convert a given data set to the neuGRID data model. This could be supplied to the core labs, neuGRID users or both.	D

U1.4 Manage stored data. E

U1.4.1. Check and control system capacity. E

User Requirements

R1.4.1.1.	The ability to control and manage the system through a simple graphical interface.	D
R1.4.1.2.	Provide tools and software to monitor system storage capacities and user quotas.	E
R1.4.1.2.1	Set quotas.	E
R1.4.1.2.2	Edit quotas.	E
R1.4.1.2.3	Delete quotas.	E
R1.4.1.2.4	Interact with users when storage reaches quota limits (possibly e-mail users with warnings.)	D

U1.4.2. Back up data. E

User Requirements

R1.4.2.1	Manage backup data	E
R1.4.2.2.	Provide a means to backup data storage resources	E
R1.4.2.3	The ability to suggest to users that they save data that has not yet been backed up, in an iterative way.	D

U1.4.3. Perform system maintenance. E

User Requirements

R1.4.3.1.	The possibility to follow a step-by-step predefined GUI-based wizard for the performance of system maintenance.	D
R1.4.3.2.	Provide a manual for performing system maintenance	D
R1.4.3.3.	A means of communicating periods of service downtime to users	D
R1.4.3.4.	Mechanisms for recovering from system failure should be provided.	D
R1.4.3.5.	A maintenance mode with the ability to take the system off-line for a period	D
R1.4.3.6.	A platform dashboard could be provided to give an overall picture of status of the infrastructure at any given point in time.	D

U1.4.4. Query the stored data. E

User Requirements

R1.4.4.1.	The ability to search for images based on subject and image-related criteria including: type of illness, date of birth, age at time of scanning and other fields.	E
R1.4.4.2.	View images, form image collections (user collections) and download images in several file formats.	D
R1.4.4.3	The possibility to use two different research modalities: BASIC (Subject_ID, Sex, Age, Modality, Series description) or ADVANCED (Diagnosis, MMSE Score, GD Scale Score, TE, TR, Slice thickness and more) with different fields / levels of search.	D
R1.4.4.4	The ability to store and manage user defined data collections.	D
R1.4.4.5	The querying facility should be user friendly and the querying interface should be operable by both technical and non-technical users.	E

U1.4.5. Search and view records of the quality control, anonymisation and format conversion processes that have been applied to data sets as they are entered into the system. E

User Requirements

R1.4.5.1.	Search QC records for images that pass a given set of QC parameters.	D
R1.4.5.2.	View the QC records, sorted after parameter of choice (not just QC parameters.)	D
R1.4.5.3.	Search and view the record of format conversions that have been applied to an image/image set at upload.	D
R1.4.5.4.	The ability to display the QC results directly to the users with the subject image.	D
R1.4.5.5.	The anonymization process should not be visible to the final user. This step could be done within the neuGRID consortium and should not be accessible (except for special privileges) by the end users of neuGRID.	E
R1.4.5.6.	neuGRID system images should be uploaded and stored as DICOM files. The image conversion process is something that needs to be done during the execution of pipelines and consequently, is something that could be checked by the user that uploaded the original image or a data input manager. In the case there isn't a DICOM definition for a given type of image (e.g. MEG images), data could be uploaded in the original file format (but should be fully anonymized). But there is no guarantee that all the workflows will work on it.	D
R1.4.5.7.	Provide provenance information related to modifications made to a data set. Provenance information may include modifications made for quality control, ethical compliance, anonymization and any format conversions that were necessary and related information.	E

U1.4.6.   Handle potentially corrupted data sets.   E

User Requirements

R1.4.6.1.	A copy of the initial data should be kept safely.	E
R1.4.6.2.	No seriously corrupted or unusable data should remain in the neuGRID data store.	E
R1.4.6.3.	Provide tools to detect corrupted data sets, and to recover them as required.	D
R.1.4.6.4.	Once data has been made available to users, ensure that it remains unaltered (with the exception of legal requirements) even if it has some degree of corruption. Any improvements to the data are handled by making a new version available to users while still keeping the old version available to users. It should be clear to users which is the most up to date version.	D
R1.4.6.5	Delete corrupted data	E
R1.4.6.6	Locate origin of corrupted data and handle the possibility of systemic problems	D

U1.4.7. Remove data from the system. E

User Requirements

R1.4.7.1.	Select data sets or groups (perhaps using the facility that is requested in R1.4.8.3) to be removed as defined within authorization levels.	E
R1.4.7.2.	An automatic report of removal to be sent to the uploaded site contact email.	D
R1.4.7.3.	Provide functionality to delete data sets from the data store.	E
R1.4.7.4.	Provide functionality to delete data sets from the user collections (R1.4.4.4.)	D

U1.4.8. Track / determine the history of a piece of data. E

User Requirements

R1.4.8.1	Select a data set that fits a certain criteria.	D
R1.4.8.2	Generate an itemized list of when and by whom a data file in the set has been used in a workflow.	D
R1.4.8.3	Generate a list of which workflows have been applied to a data set, and sort sets into groups (group A had workflow X used on them, group B workflow Y, group C workflow X and Y.)	D
R1.4.8.4	To capture information on which kind of studies and analysis data has been used.	D
R1.4.8.5	To record a group of clinical results in which a particular piece of information was used during analysis.	D
R1.4.8.6	The possibility to define the roles of specific data in the AD pathology.	D
R1.4.8.7	The possibility to create a list of publications and view where neuGRID has played a part in the research.	D
R1.4.8.8	Data set specific provenance data should contain information related to the	E

	history of the piece of data.	
R1.4.8.9	Allow a user to opt out of making tracking information public to other users for a given period	E

U1.5.   Control the security of the stored data. E

U1.5.1.   Implement new and edit existing access control strategies. E

User Requirements

R1.5.1.1.	Sort data sets into groups to which a certain access control is set.	E
R1.5.1.2	Edit the access control of a group of data (see R1.5.2.2.)	E
R1.5.1.3.	Manage, sort and edit access controls for a named individual/ or group of researchers. Provide tools to administrators to define user specific access control policies (at the project and individual user levels.)	E
R1.5.1.4	A supervisor or responsible person should define both the access level and the policies that pertain to gaining access to the data stored inside neuGRID.	D
R1.5.1.5.	Provide secure access to data storage resources.	E
R1.5.1.6.	There should be a possibility to give individual users special access to a certain data set.	D

U1.5.2.   Configure a set of ethical rules that relate to and govern the use of particular data sets. E

User Requirements

| R1.5.2.1.	Compose specific data use agreements in writing which accompany a chosen data set/group.	E
R1.5.2.2.	Unless the relevant conditions are agreed to in writing by a user, a set that is subject to specific usage rules cannot be used.	E
R1.5.2.3.	Log the users who use such a set (see R1.4.8.2) with a flag that they have agreed to be bound by the data use agreement.	D
R1.5.2.4.	Define different agreements and rules regarding access for public, academic, research, and industrial neuGRID users.	D
R1.5.2.5.	All users should accurately provide requested information regarding who will use neuGRID data and the analyses that are planned.	D
R1.5.2.6.	All users should be requested to cite neuGRID as the source of their results in published work.	E
R1.5.2.7.	Provide tools for the configuration of data usage controls that can be applied to ensure that stored data sets are used appropriately by users.	D
R1.5.2.8.	There should be a description of the ethical usage requirements (e.g. informed consent should be used for this data set) for a given data set.	E

U1.5.3. Temporarily upload a private dataset to neuGRID.   E

User Requirements

R1.5.3.1	A user may wish to check whether an existing workflow also works for a private dataset which holds a rare subset of patients MRI scans. A way of temporarily uploading the private dataset to neuGRID may be desired. The temporary dataset should be accessible for a given period of time and then be removed from the system	D
R1.5.3.2	The dataset in R1.5.3.1 should be accessible to existing pipelines, uploaded batch scripts (e.g. bash) or Linux executables.	D

S2.   Data Access:

A researcher is interested in a rare form of a disease and wants to be able to do a statistically meaningful analysis. Unfortunately the researcher's institution doesn't have enough images to make this possible. The user interacts with the system using the neuGRID store to search for and identify an appropriately large set of images from the group of projects which neuGRID has access to. At this stage access controls and ethical policies are fully enforced to protect sensitive data. The researcher then uses the system to submit the study set for analysis through a workflow.
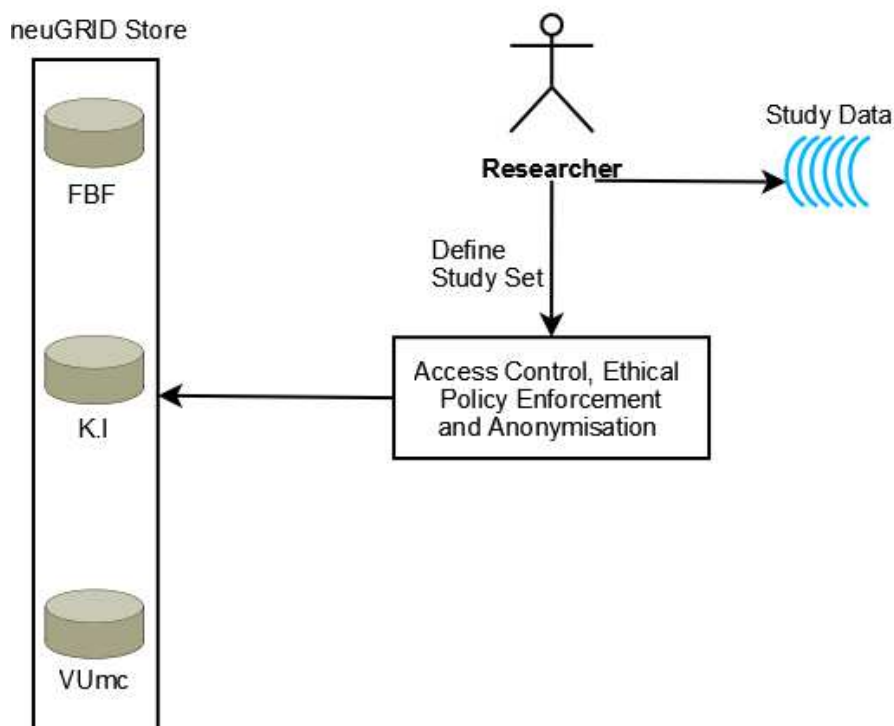
Indicative Use-cases:

U2.1   Authenticate a user and enforce access control / ethical policies. E

User Requirements

R2.1.1.	A neuGRID user interface should be provided. A signed usage agreement could be put in place at this stage. All neuGRID users should fill in an on-line form and neuGRID staff will provide them with a specific user ID and password.	E

R2.1.2.	Fill in neuGRID user data (institute, name etc) and store as "My profile" under "My account".	E
R2.1.3.	The possibility to become part of a group of researchers (start a new group, be invited to an existing one).	D
R2.1.4.	Sensitive individual data sets will generate specific access agreements to be signed (R1.5.2.2), which will be stored under "My account"	D
R2.1.5.	Provide a global security model, which enables individual researchers from collaborating institutes to access other institutes' data sets. There should be a certificate-based system to identify users and to perform access control.	E
R2.1.6.	Allow institutes to define local access control policies.	D
R2.1.7.	It is necessary to have access controls at the Project and Individual data set levels.	E

U2.2   Search for a group of images or data that matches a given criteria. E

User Requirements

R2.2.1.	Select a set of properties with which to generate data subsets from the database.	E
R2.2.2.	Generate feedback about the data sets (lists), which can be sorted by QC parameters, type of camera and other features.	D
R2.2.3.	Fine-tune the property set interactively.	D
R2.2.4.	Store the final property set under "My account > My search property sets."	D
R2.2.5.	Store the resulting data sets (lists of data generated by applying the property set on the data base) under "My account > My data sets."	D
R2.2.6.	Provide a global search utility which searches distributed neuGRID data stores based on a user defined criteria. Researchers should be able to search for a certain type of patients based on medical information as well as on imaging information. For instance patients with Mild Cognitive Impairment with a given range of MMSE values which have had a T1 MPR sequence with a pixel size smaller equal 1.5 mm in each direction. The range of fields that can be used for searching should include:	E
R2.2.6.1.	Subject information: Subject Id, Sex, Research Group, Age, Weight.	E
R2.2.6.2.	Project specific information.	D
R2.2.6.3.	Clinical assessment information: MMSE Total Score, GDScale Total Score, Global CDR, Modified Hachinski Total Score, NPI-Q Total Score and Functional Assessment Questionnaire Total Score.	D
R2.2.6.4.	Study information: Study date, Visit.	E
R2.2.6.5.	Image information: Original (Modality, Series Description, Acquisition type, Weighting, Slice Thickness, TE, TR, Acquisition Plane, Manufacturer, Field Strength) – Pre-processed (Series Description, Image File Type, Anatomic Structure, Tissue Type, Laterality, Registration/Space) – Post Processed (Series Description, Image File Type, Anatomic Structure, Tissue Type, Laterality, Registration/Space.)	E
R2.2.7.	Metadata will need to be stored for images in order to enable search facilities to be provided, this will identify images and the search will be performed on the metadata.	E

U2.3   Define and group the data that comprises a set for use in research. E

User Requirements

R2.3.1.	Combine labelled property sets into meta-sets.	E
R2.3.2.	Store the meta-sets under a label under "My account > My search property sets."	D
R2.3.3.	Store the resulting data sets (lists) under a label under "My account > My data sets."	D
R2.3.4.	Note prominently which property sets/meta-sets/data lists are bound by which ethical agreements.	D
R2.3.5.	Provide an interface which allows users to define groups of search results for research purposes.	E

U2.4   Visualize a research set. D

User Requirements

R2.4.1.	The ability to visualize in 3D, large amounts of post-processed data, e.g. registered whole brains, white-gray matter interface, segmented cortical grey matter, etc. This includes:	E
R2.4.1.1.	Clinical biological data (e.g.: Tau, Ab1-42, P-Tau 181P, Tau/Ab1-42 P-Tau181P/Ab1-42) regarding the group of patients considered in a specific study.	E
R2.4.1.2.	Imaging data (DTI, 3dT1, T2, PD, fMRI, PET and others) regarding the group of patients considered in a specific study.	E
R2.4.2.	The provision of a summary of a user's research sets in list form under "My account > My data sets."	D
R2.4.3.	View condensed lists of clinical biological data and the imaging data set properties (44 images with a 3 T camera, 1445 different patients in a total 1943 images) under "My account > My data sets."	D
R2.4.4.	The possibility to generate some descriptive statistics about the parameters that have been chosen using a basic statistical package that is integrated within the infrastructure.	D
R2.4.4.1.	Provide appropriate visualization tools that are integrated in the search utility, perhaps displaying thumbnails of images.	D
R2.4.4.2.	The user should be able to visualize data sets without downloading them.	D
R2.4.5.	An image viewer should be provided that provides a convenient browsing mechanism for users.	E

U2.5   Store a research set for future use. O

User Requirements

R2.5.1	A research set of images and clinical data should be saveable and reusable. Properties of the group will be stored and the resulting sets may be accessed through a saved data set list.	D
R2.5.2	Users should be able to view and download their own "User Collection" for local back up.	D
R2.5.3	Create a structured environment with directories and subdirectories where research results can be stored.	D

R2.5.4	Perform actions on stored datasets and images (moving, copying, deleting, renaming and adding new images.)	D
R2.5.5	The search utility should be able to export and save searches for future use.	O
R2.5.6	Saved searches should be easily accessible via an interface.	O
R2.5.7	It should be possible to store a query that was used to create a research set (this is the property set mentioned in R2.2.4.)	O

U2.6   Monitor data quality and allow users to give feedback regarding raw data in research sets. D

User Requirements

| R2.6.1	A quality viewer should be provided, together with information regarding the quality assessment that was made by the researcher that uploaded the raw data.	D
R2.6.2	A "Comment on this image" facility: other users' comments might be visible under a special link in the data list.	O
R2.6.3	It should be possible to share specific research sets with some predefined groups giving information about research methods, data type and other issues.	D
R2.6.4	The possibility to express a judgment about the quality of data could be useful. The judgment (e.g.: 4- Excellent; 3-Good; 2-Sufficient; 1-Bad) could be taken into account during the creation of a research set.	D
R2.6.5	Provide tools to determine and monitor data set quality.	D
R2.6.6	The interface through which saved searches and research sets are managed, should have the functionality to allow permitted users to post comments and give feedback on research sets of other users.	D
R2.6.7	The interface for saved searches will allow users to add or remove users from commenting on research sets.	O

U2.7   Annotate a research set with useful information regarding the data that is contained within. O

User Requirements

R2.7.1	Comment the set lists that are stored under "my data sets" (comments should be seen when set lists viewed in R2.4.2)	O
R2.7.2	The information should be of a high-level and will describe the type of a specific data user collection in an efficient way (for example: reporting the number of patients with AD, MCI and CTR and the Sequence type.)	O
R2.7.3	Set "annotation" needs to be strictly controlled: users should have the ability to submit annotations (e.g. comments on image quality, new measures.) but such annotations should be reviewed centrally and included only whenever they satisfy specific criteria. Finally, as measures are often protocol/scale-dependent the protocol/scale should be specified.	O
R2.7.4	An interface should be provided for the saving of searches and have the capability for users to provide annotations and metadata for saved research sets.	O

S3   Workflow Specification and Development:

A new image analysis methodology is developed and a researcher wishes to build a workflow to run it. Using an interactive creation tool the user constructs a workflow and specifies some initial settings. The user also creates a record which describes the workflow and gives other users information about its purpose and access controls. The system allows different versions of the workflow to be created, tested and released when they are ready for use by other researchers.
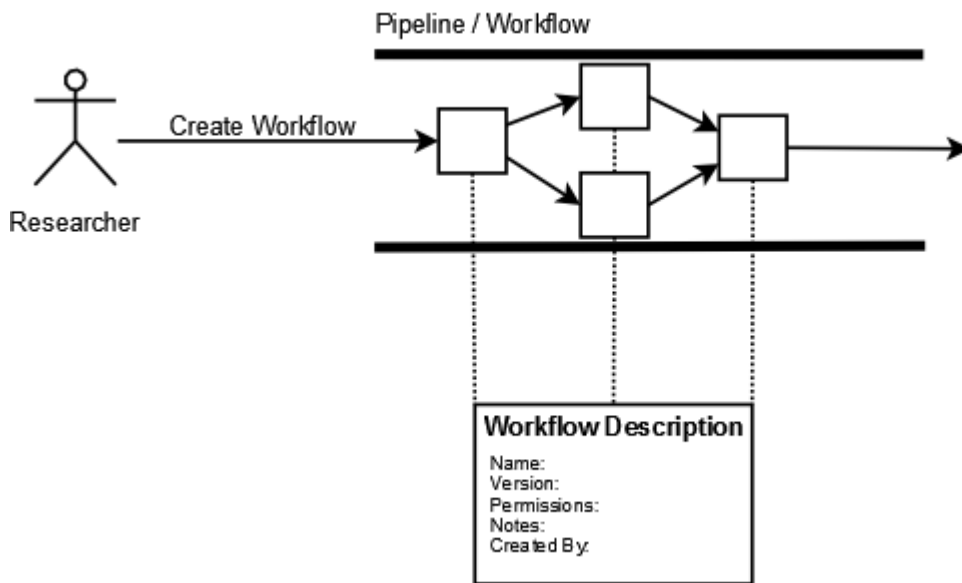
Indicative Use-cases:

U3.1   Construct, visualize, annotate and edit new workflows. E

User Requirements

R3.1.1	Select software packages from categories of algorithms (e.g. "statistical" or "brain stripping.")	D
R3.1.2	Construct a workflow by stringing together various algorithms and packages in a work area (in a drag and drop fashion), creating a series of connected boxes. Divisors, yes / no alternatives for branching workflows may also be available in a graphical toolkit. This should be as simple as possible using a combination of arrows and nodes within a Graphical interface.	E
R3.1.3	Add comments next to each box in the workflow.	D
R3.1.4	The possibility to divide the workflow into logical units (the first three steps are brain stripping and have a pink background, next are five volumetric steps with a yellow background) with labels describing them.	O
R3.1.5	Visualize the workflow as a schematic boxed flow diagram (a rough version could be viewed in the work area and should be exportable to other applications for use in papers.)	E
R3.1.6	Edit the workflow by moving boxes around. A warning system saying "block A does not generate output that enables running block B directly after it" would be helpful.	D
R3.1.7	A possibility to edit input parameters in each algorithm (maybe an execution	E

	crashes because it requires a "4" instead of "2" in a given sub-process.)	
R3.1.8	Save the workflow with a label under "My account > My workflows > Drafts."	E
R3.1.9	The provision of a command line scripting interface is necessary. It should be possible to upload a workflow as a Linux command script (e.g. bash) which calls a number of Linux executables residing on the grid or uploaded together with the script.	E
R3.1.10	The possibility to have a range of pre-configured atomic modules from which new workflows can be created or to be able to integrate new functions efficiently.	D
R3.1.11	The opportunity to have a functional test-bed to efficiently validate workflows that are in construction (using trial appropriate reference data set already uploaded remotely and an efficient validating execution interface.)	D
R3.1.12	The ability to do a "debug error procedure" in order to show different actions that a final user can take in order to debug any validation or execution errors that could be encountered while using the Pipeline.	D
R3.1.13	The ability to preserve the order execution and the dependencies of the pipeline workflow.	D
R3.1.14	The ability to upload workflows generated by the major workflow management systems that are in use today (e.g. the LONI pipeline, Scientific Kepler system and others.)	O
R3.1.15	The infrastructure should be platform-independent.	D
R3.1.16	The possibility to use images stored in the neuGRID store to run a local analysis (e.g. in case a user wishes to run an analysis on neuGRID images using software developed locally, which is not to be shared.)	D
R3.1.17	Provide a means of editing existing workflows.	E

U3.2   Work with draft workflows and use version control to manage them. D

User Requirements

R3.2.1	Open a workflow and edit it.	E
R3.2.2	When saving a previously existing draft workflow, automatically append version number and save under the workflow label under "My account > My workflows > Drafts" together with date edited. There should be a version control system for workflows that reside on neuGRID that is independent of their implementation (as a script file, program of graphical workflow.)	D
R3.2.3	The possibility to save draft personal modules and workflows inside the neuGRID system.	D
R3.2.4	The possibility to open, drag and drop draft modules into a workflow quickly and easily.	D
R3.2.5	The possibility of creating pipelines by assembling existing workflows.	D
R3.2.6	Provide a repository for workflows with version control management.	D
R3.2.7	Provide user friendly interfaces, integrated with the workflow authoring software to upload/download/update workflows to the workflow repository.	D
R3.2.8	Changes between different versions of the software should be documented.	D

U3.3   Visualise, annotate and edit existing workflows. E

User Requirements

R3.3.1	Locate an existing workflow from the database of generally available workflows by selecting categories of algorithms that are desired (this may generate a list of workflows.) The possibility to use some of the visualisation features that are provided by popular toolkits such as FSL, FREESURFER, SPM and MNI.	D
R3.3.2	Select desired/interesting workflows and save under "My account > My workflows > Published."	D
R3.3.3	Provide a tool by which users can visualize existing workflows as in R3.1.5.	E
R3.3.4	Provide users the functionality to add annotations or comments to workflows as in R3.1.3	D
R3.3.5	Provide users the capability to edit existing workflows as in R3.1.4 and R3.1.6.	E
R3.3.6	Save as in R3.1.8 and R3.2.2 (to "Drafts.")	D
R3.3.7	The opportunity to have a responsible person or group of people that maintain the main pipelines in use in the neuro-imaging field.	D
R3.3.8	The system should send email alerts to the final users when the workflow outputs are ready.	O
R3.3.9	There should be a way to attach a known bug list to a workflow.	D

U3.4   Upload new packages, algorithms or analysis software to system for use in workflows. E
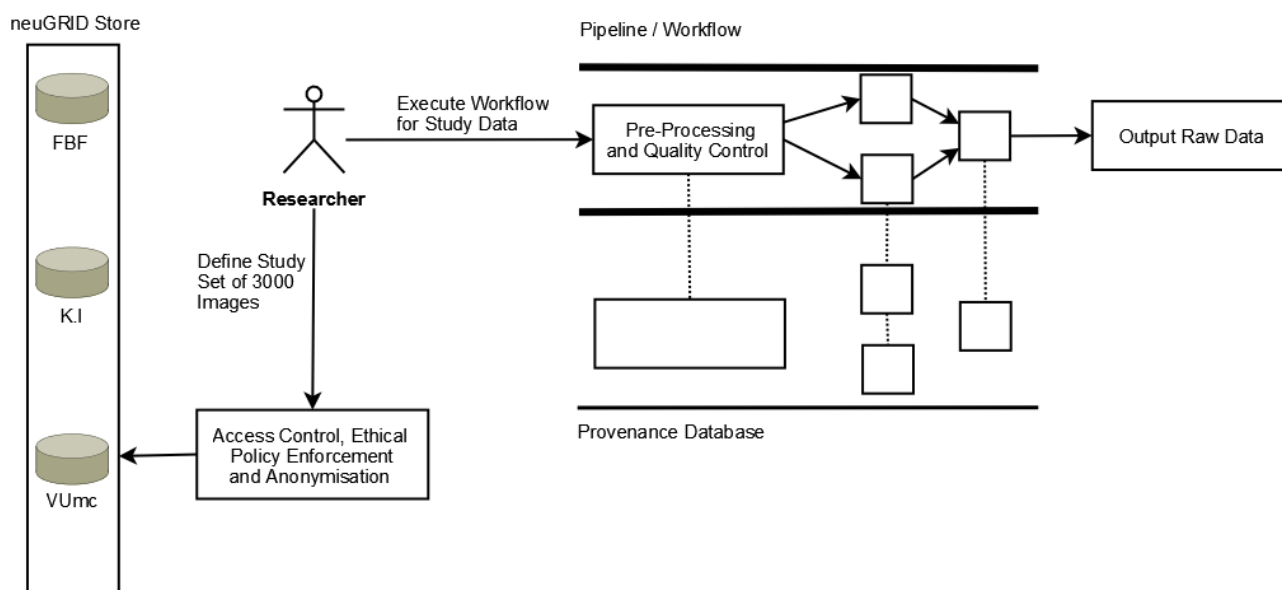
User Requirements

R3.4.1	Upload an algorithm or package (or draft) including source code; fill in what categories to store it under (R3.1.1.)	E
R3.4.2	Make a workflow accessible and set access rights, fill in label name, author name and institute, way to reference, terms of use (grant me a footnote, grant me authorship on any papers produced with the help of my flow) etc. See also Story 7.	E
R3.4.3	The possibility to download an algorithm/package, to be able to tweak it oneself, by editing code or settings.	D
R3.4.4	Save the tweaked algorithms under "My account > My algorithms."	O
R3.4.5	If an algorithm is uploaded with the same name as an already existing one, automatically append version number (and ask the uploading researcher to enter a comment on what has changed.)	D
R3.4.6	If a package name changes, include source code dependencies (builds on package X by adding Y.)	D
R3.4.7	The upload of new packages/algorithms should be controlled centrally.	E
R3.4.8	New tools that are to be uploaded into neuGRID should be rigorously tested and validated. All tools should be uploaded together with documentation, including a user guide, algorithm explanations and appropriate references.	E
R3.4.9	The neuGRID "workflow management system" should be able to support and interface with many common languages classically used in the neuro-imaging field (like PERL, C++, Matlab, bash script and Python.)	D
R3.4.10	Where possible maintain architectural compatibility with the latest imaging software.	D

R3.4.11	Upload temporarily personal packages/software for specific studies. It should be able to upload a workflow as a Linux command script (e.g. bash) which calls a number of Linux executables residing on the grid or uploaded together with the script.	O
R3.4.12	Provide users with an interface for uploading new software packages, algorithms and analysis software subject to appropriate validation, which may then be used in future workflows.	D
R3.4.13	Provide users with a means of browsing the library of existing algorithms, packages and analysis software that are available for use in their workflows.	E

S4   Workflow Execution and Management

A researcher wishes to run a comparative analysis using a study set of 3000 MRI scans stored in geographically distributed medical centres. It is important that the results are generated in a timely fashion as the researcher has a number of different studies to do that week. The user interacts with the system to choose a study set of 3000 images, selects the pipeline or workflow through which the analysis will take place and starts the analytical process.

Indicative Use-cases:

U4.1   Search for existing research sets or define new groups of images and other information to be processed using the workflow. D

User Requirement

R4.1.1	Find a previously defined research set by selecting a data set under "My account > My data sets" (see R2.4.2 and R2.2.5.)	D
R4.1.2	Generate a new research set as in R2.5.1.	D
R4.1.3	A means to search publicly available research sets.	O
R4.1.4	The ability to edit research set access rights.	D

R4.1.5	In order to optimize performance, the images that are used in an analysis should be present (if possible) locally in those nodes of the grid that don't have a high level of available bandwidth. This is due to the fact that the transfer of a large number of images on the network will greatly increase the time to get the final results. Clearly, this is particularly true for a centre like FBF which is characterized by a connectivity of 10 Mbps.	D
R4.1.6	High redundancy and data availability is necessary.	D
R4.1.7	The possibility to integrate information provided by images and metadata with the definition of mathematical variables like vectors, list and structures. Define basic operations (like indexing, push, pop and length count) in order to perform command line operations on these objects containing images of interest.	D
R4.1.8	Provide a global search utility which searches distributed neuGRID data stores based on a user defined criteria.	D
R4.1.9	Metadata should be present for each of the images in the system. This will identify images and allow them to be searched.	E
R4.1.10	A user should be able to download data sets, (if proper authentication has succeeded) subject to the usage agreements that are in place.	D
R4.1.11	Allow users to save a search result set and define it as a research set (see R2.5.1).	D
R4.1.12	Provide the capability of using saved data sets and research sets for input for workflows (see R4.2.1.)	D

U4.2   Run, monitor and control the execution of a workflow. This would involve perhaps the ability to cancel, edit and restart an execution. E

User Requirements

R4.2.1	Execute a workflow on a given data set in a step by step way or as a single batch of tasks that are processed in one run.	D
R4.2.2	Output from the individual processes within the workflow is output to a progress window; also when a new process is started (process name_1: <output from 1 such as "calculating chi-2"> -- process name_2: <output> and so on).	D
R4.2.3	When a process ends (prematurely or not) the user can add comments at the bottom of the window.	O
R4.2.4	A Graphical User Interface (GUI) should be provided that has buttons to start, stop and restart the workflow.	E
R4.2.5	The possibility to change the input parameters to a sub-process of a workflow (see R3.1.7.)	E
R4.2.6	The possibility to test a workflow on single images or subsets of the chosen data set (one could of course generate a new data set but that is probably not as practical.)	D
R4.2.7	The ability to create, visualize and edit complex workflows in a convenient way.	E
R4.2.8	Simple way to monitor workflow execution.	D
R4.2.9	The user should have the possibility to check and perform quality control on each intermediate output.	E
R4.2.10	The ability to cancel, restart and debug workflows.	E

R4.2.11	The ability to share workflows with other researchers in the system.	D
R4.2.12	The possibility to provide the user with sample images for any kind of scan modality (MRI, fMRI, PET and others) in order to test his/her own workflow (or parts of it) using them and saving time uploading their own images.	D
R4.2.13	Extend the workflow authoring environment to include basic execution functionality for:	D
R4.2.13.1	Starting the execution of a workflow.	D
R4.2.13.2	Providing an interface to monitor the status of a workflow.	D
R4.2.13.3	Provide ability to control the execution by cancelling or restarting the workflow.	D

U4.3   Search for and select the desired analysis pipeline from a set of existing workflows, edit settings if required and execute. E

User Requirements

R4.3.1	New workflow sharing should be controlled (only functioning and validated workflows should be uploaded and shared.)	O
R4.3.2	All workflows should be organized in a clear and efficient way in order to make their usage as convenient as possible.	D
R4.3.3	The presence of a facility that allows users to query for specific modules. The Search function should return results drawn from the module's name, author list, citations, tags, description, and parameter fields.	D
R4.3.4	Most modules could have two or three required metadata parameters and several optional parameters. The possibility to switch on these additional options simply clicking on the modules could be useful.	O
R4.3.5	Provide a service for users to upload workflows.	D
R4.3.6	Provide an interface to allow users to select pre-authored workflows and execute them with a new/existing research set.	E
R4.3.7	Provide the capability of editing an existing workflow, and executing it.	E

U4.4   Search the history of a given workflow to find a particular version of it for use in a specific piece of research. D

User Requirements

R4.4.1	The possibility to compare different versions of the same workflow.	D
R4.4.2	Each workflow is described by its components (viewed as in R3.1.5) highlighting the differences of each version and by its provenance (who built it, uploaded when, changed when), their new applications or improvements.	D
R4.4.3	The ability to "Unfold the history" of a workflow to see older versions of it. It would be useful if this also showed versions where no one has changed the workflow per se, but one of the packages/algorithms that it is comprised of has had an impact.	D
R4.4.4	An older version of a workflow could be retrieved for validating previous research or error testing (it may need to be rebuilt using stored settings.)	D
R4.4.5	Provide a service for uploading workflows.	D

R4.4.6	Provide the capability to annotate the history of a workflow.	D
R4.4.7	Provide an interface to search existing workflows and their respective histories.	D
R4.4.8	The possibility to use a workflow as it was on a given date by entering the date of interest.	O

U4.5   Store a history of each workflow execution, research set and settings. Allow user annotation of such information. D

User Requirements

| R4.5.1 | The progress window output from R4.2.2 could be saved as a file. It may have a header consisting of a description of the data set used and the settings made for each (named) algorithm in the workflow (this may also be saved separately as a "workflow setup", which lists all the parameters that were given to the workflow's algorithms.) It could end with some user generated comments as in R4.2.3. | D |
R4.5.2	The possibility to efficiently retrieve some standardized workflows that are used in daily routine tests and procedures by different labs.	D
R4.5.3	Provide capability to annotate history of a workflow.	D
R4.5.4	Provide an interface to search existing workflows and their respective history.	D

U4.6   Process raw output data by importing it into user specified analysis tools and toolkits. E

User Requirements

R4.6.1	Save the data set which has gone through the workflow with a label under "My account > My processed sets."	D
R4.6.2	Allow transformation of data to suit the needs of some analysis tools. Provide conversion tools for toolkits compatibility	O
R4.6.3	Build a range of common analysis tools into the infrastructure (but licensing may prevent this).	O
R4.6.4	Save analyzed data under "My account > My analyzed sets" with links connecting each analyzed set to the corresponding processed set (by actual linking or by naming convention).	D
R4.6.5	Save the workflow setup and the progress window output under each processed set.	O
R4.6.6	Define a drag and drop interface in order to send raw output data from one analysis into another workflow as input.	D
R4.6.7	Allow user to use data in their desired format.	O
R4.6.8	Provide notifications to users on the status of a workflow.	D
R4.6.9	Upon completion of a workflow, allow users to download raw data output.	E
R4.6.10	Provide the necessary functionality to export the raw output into the desired data analysis software.	E
R4.6.11	Enable some basic analyses using inbuilt statistical tools such as those provided by FSL.	D

S5.   Validation of Workflows:

A user creates a new workflow and runs a test data set using it. At each stage in the execution of the workflow, the intermediary images or data are stored and a full provenance track is kept. After results are produced, the user examines the provenance to check that each stage of the analysis was completed correctly. The raw results are then exported into the user's preferred analysis tool and the whole process is added to the researcher's history for future reference. Initially the new workflow produces some poor results during testing. The researcher therefore looks at the logs of the workflow execution and locates the problem. The user then interacts with the system to make changes to the relevant settings and re-runs the test study. This time the process runs correctly and meaningful results are produced.
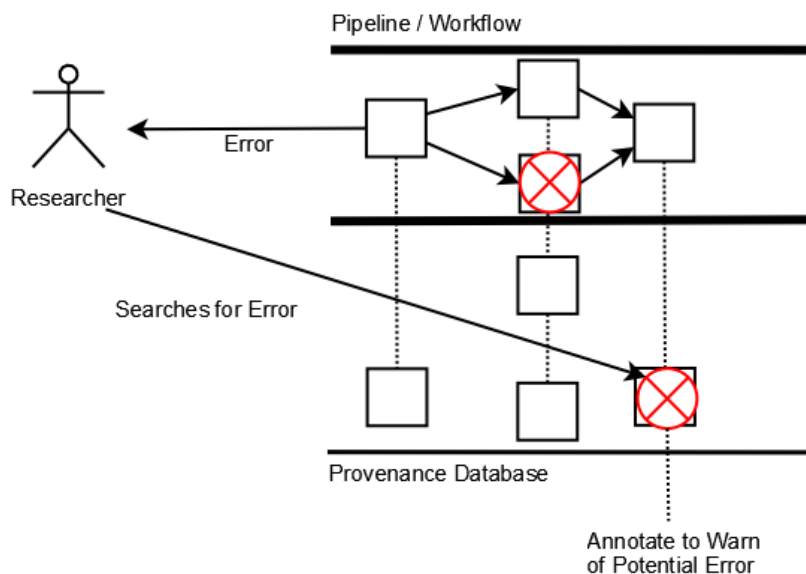
Indicative Use-cases:

U5.1   Validate a workflow using provenance data to locate points of failure in it. E

User Requirements

R5.1.1	Load the workflow into a variant of the work area in R3.1.2. The order of the boxes and layout of the workflow cannot be changed, but by clicking on each box the appropriate set of provenance data can be viewed: lists of images that can be put into the viewer (possibly to compare images, from different provenance sets and within sets) and numerical output data (chi-2 etc). Also the workflow setup can be viewed.	D
R5.1.2	To check for errors try to execute the workflow.	E
R5.1.3	If any errors are found it could be useful that a dialog box will pop up listing all the errors found in the workflow.	D
R5.1.4	During the validation of the workflow the outputted data should be visualized.	O
R5.1.5	Provide users with the capability to browse provenance data collected from execution of workflows.	E
R5.1.5.1	The interface should be user friendly, and allow for browsing of process by process provenance data.	D

R5.1.5.2	Provenance data should link to the intermediary output produced during execution of the workflow.	E
R5.1.6	There should be a way to report outliers and to be able to check intermediate data for such indicators.	D

U5.2 Search for an appropriate reference data set to automatically verify the output from a workflow and create a test set for a newly developed analysis workflow. D

User Requirements

R5.2.1	When someone has developed a workflow, at upload they can be asked to specify a reference data set to be associated with it. This reference set could then be found as a property of the workflow ("Test with reference data set?")	D
R5.2.2	Generate a new data set for testing old or new workflows.	O
R5.2.3	The possibility to add a reference data set to the workflow's properties, even for those who have not constructed the original workflow.	O
R5.2.4	It could be useful to be able to choose from a number of predefined reference data sets (for example: one characterized by 3D volumetric images, fMRI images, DTI images and PET images) comprising several images of reference.	D
R5.2.5	Provide a tool to users to browse and select reference data sets for execution with a workflow.	O
R5.2.6	Provide the user with a comparative analysis of the output produced to output in the reference data set.	D

U5.3 Report errors in workflow execution. E

User Requirements

R5.3.1	An error report button should be included within the R4.2.4 GUI. It should send an email to the appropriate place with information regarding workflow setup, workflow name and data set properties. It should also generate an error number for convenience and easy follow up.	D
R5.3.2	Some instances of a module could fail from time to time. In this case, it could be useful to have a viewer box in which all the failed instances of the module could be shown. With this information neuGRID users could diagnose the problems encountered during the execution of a workflow and hopefully solve them.	O
R5.3.3	Provide notification for critical events during an execution of a workflow.	E

U5.4 Annotate workflows with version information and a full change history. D

User Requirements

R5.4.1	Add a comment to a workflow which can be seen under "Unfold history" in R4.4.3.	O
R5.4.2	The possibility to make an analysis of the different usage patterns for each workflow that is available in the infrastructure. It would be useful to understand which data values are most commonly used by the scientific community and to analyze different types of acquisitions through different	O

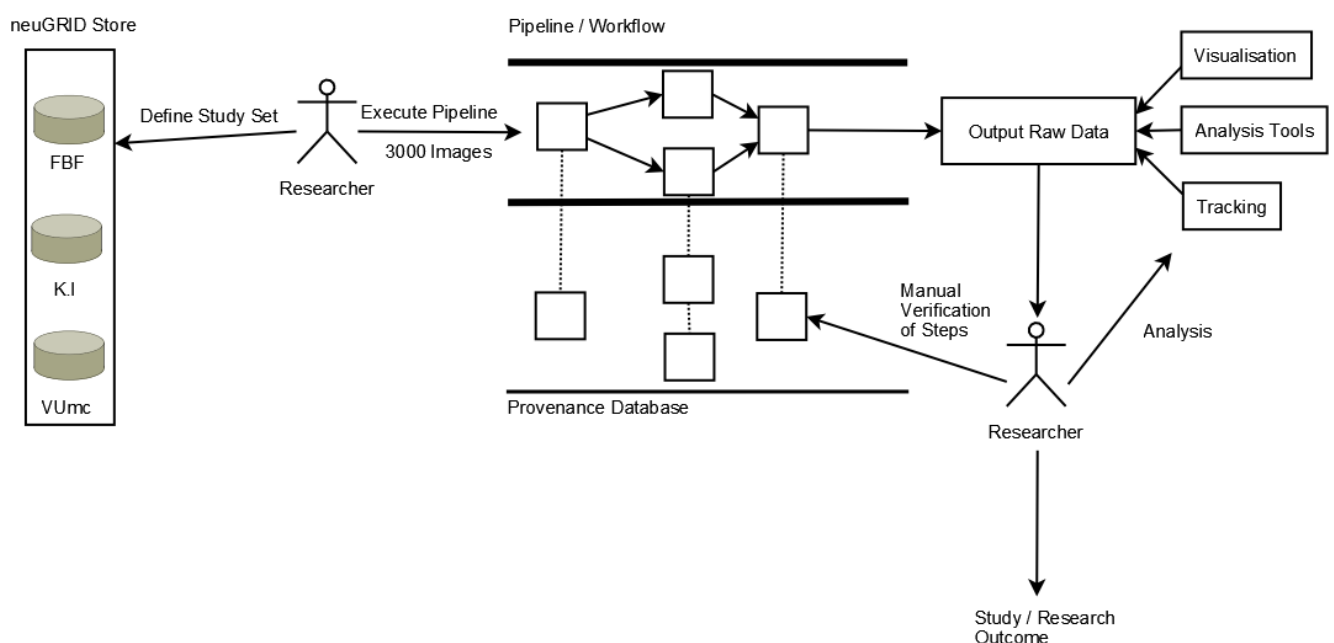|	workflows.	
R5.4.3	Provide a repository for workflows with version control management.	E
R5.4.4	Provide user friendly interfaces, integrated with the workflow authoring software to upload/download/update workflows to the workflow repository.	D
R5.4.5	The repository should have the functionality to add annotations from users about versions of the workflow. A description of the differences between versions should be provided.	D
R5.4.6	The repository should log and document historical changes to a workflow.	D

U5.5.   Annotate a workflow with information regarding the settings that are appropriate in different situations.   D

User Requirements

R5.5.1	See related requirements in U5.4.	D
R5.5.2	The possibility to summarize the most useful and appropriate parameters used in the workflows through synoptic tables.	O
R5.5.3	Ease of reference parameters.	O
R5.5.4	Provide capability to users to annotate workflows, providing information about settings of the workflow in different execution contexts.	D

S6.   Validation of Results using Provenance Data:

A workflow yields some surprising and possibly significant results. A researcher wishes to confirm that the results are accurate and identify any mistake that has been made. By analyzing all the intermediary image sets and workflow execution logs the user is able to manually verify that the results were incorrect. It is found that the error was due to a specific group of images interacting badly within the workflow. The user annotates the workflow so that other users are warned if they attempt a similar analysis.

Indicative Use-cases:

U6.1   Capture a complete provenance of workflow execution. E

User Requirements

R6.1.1	Store intermediary execution steps.	E
R6.1.1.1	The workflow processes and the workflow setup should be saved (see R4.5.1.)	D
R6.1.1.2	The intermediate, processed files (provenance data) are saved, according to a clearly defined structure. This may be in a format such as: Run number/Process number/files, e.g. Run5/Process2 (Brain strip)/file_no5. It is useful to be able to save the output from more than one run at a time, for comparison. There could be an upper limit of around 10 runs, and the output may be accessed through "My data sets > Provenance data."	D
R6.1.1.3	Provide explanation and details of any errors that occur and report possible causes.	E
R6.1.1.4	Send potential errors to the neuGRID administrators if the workflow resides on the neuGRID infrastructure.	D
R6.1.2	Keep a full record of all intermediary images and data.	E
R6.1.2.1	The data structure that is suggested in R6.1.1.2 should also include a summary of any numerical data that is produced (chi-2 etc.)	D
R6.1.2.2	Store error messages and be able to navigate through them.	E
R6.1.2.3	Post problems on a neuGRID technical forum.	O
R6.1.2.4	All intermediary data and related logs should be stored during workflow execution.	E
R6.1.2.5	Provenance data should be presented in a user friendly fashion.	D

U6.2   Carry out a manual verification of all the stages that have been processed during workflow execution using the data stored in the provenance database. E

User Requirements

R6.2.1	The possibility to import selected files from R6.1.1.2 into the appropriate step in a given workflow using the GUI in R4.2.4 and then to analyze the results with a range of external toolkits.	D
R6.2.2	The ability to take the output from a single step in a workflow and look at it through a viewer/full text output (see R5.1.1.)	D
R6.2.3	The possibility to re-execute single workflow functions that were previously carried out during processing using a simple command line interface.	E
R6.2.4	Provide the user with an interface to browse a completely executed workflow, process by process, and enable user to view all relevant intermediary output and logging information.	E

U6.3   Search the provenance database for interesting information. D

User Requirements

R6.3.1	The possibility to check image anomalies through a specific link.	O

R6.3.2	Compare the outputted raw files with information from saved workflows (if any exists.)	O
R6.3.3	Provide a querying interface to the provenance store.	D

U6.4   Perform statistical analysis on the provenance data.   O

User Requirements

R6.4.1	Check for additional abnormalities passed over in silence (weak field inhomogeneities, ringing artifacts etc.)	O
R6.4.2	Compare the results obtained with reference images.	D
R6.4.3	Allow a user to export/download provenance data to their computer system and perform statistical analysis on it subject to neuGRID usage policies.	D
R6.4.4	Results should be saved as a property of the provenance data set. Files may go into a directory structure such as: Run number/Process number/User-selected analysis set name/files.	D
R6.4.5	Provide any necessary format conversion tools.	O

U6.5   Annotate a workflow with information regarding potential errors and incompatibilities.   O

User Requirements

R6.5.1	As R5.4.1. The workflow comments should not be unstructured text inputs but sorted into categories (General, Errors, Inconsistencies and Comment made by <name>.)	O
R6.5.2	When an error occurs a red colour could be used to depict that the workflow has a problem.	O
R6.5.3	Provide a user with the capability to annotate an item in the provenance store.	O

U6.6   Search a list of common errors that are known to affect a given workflow.   D
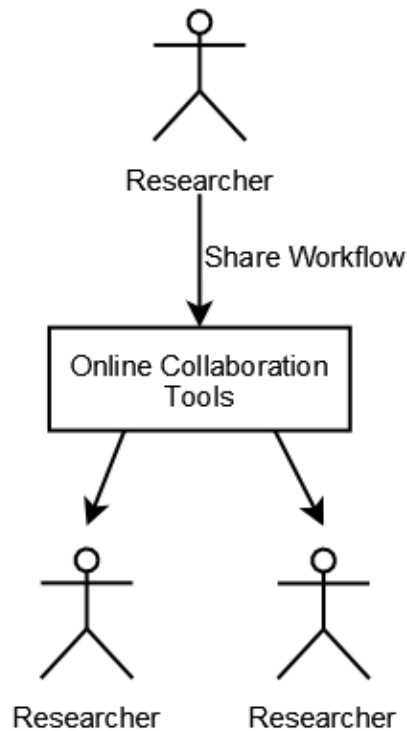
User Requirements

R6.6.1	Search and display workflow comments regarding errors. Also, automatically save and compile statistics on which errors crop up during the run of a certain workflow.	D
R6.6.2	See neuGRID technical on-line forum R6.1.2.3.	O
R6.6.3	Create a frequently asked question sections for each workflows.	O
R6.6.4	Provide the user with information about common errors that severely affect a workflow.	D

S7.   Online Collaboration:

Sharing Workflows
A new workflow has been developed and verified. A user decides that it might be useful to share it with other researchers in the field. The user makes the workflow available to a team from a partner institution in a given project. The other team is delighted as it saves them some

time and effort. The research that is produced acknowledges the contribution of the workflow it becomes an established research method more quickly than would have been possible otherwise.

Indicative Use-cases:

U7.1   Control access to workflows and allow users to create and manage groups of collaborators with whom they wish to share workflows. E

U7.1.1   Publish a new stable workflow within a group or wider community. E

User Requirements

R7.1.1.1	A researcher on uploading / publishing a workflow should be able to define access permissions for individuals or groups.	E
R7.1.1.2	Provide a service where users can upload and share workflows.	E
R7.1.1.3	Authorization should identify users uniquely.	E
R7.1.1.4	A specific group member should be able to share a workflow with other members of that group.	D

U7.1.2   Publish a developmental workflow for testing and evaluation within a group or wider community. D

User Requirements

R7.1.2.1	As in R7.1.1.1 but the uploading researcher can also tag the workflow as	D

	under development, which will show up clearly in connection to the workflow name, when searched for and viewed (as in R3.1.5.)	
R7.1.2.2	Allow users to create virtual groups within the service.	O

U7.1.3   Leave feedback regarding the effectiveness of workflows. O

User Requirements

R7.1.3.1	See R6.5.1. Workflows under development may have more categories to comment under.	O
R7.1.3.2	Users should be able to provide feedback and comment on workflows that have been created by other users	O

U7.1.4   Provide information on the authors of a workflow and suitable references for referencing them in scholarly work. O

User Requirements

R7.1.4.1	See R3.4.2 for more information about capturing this information.	O
R7.1.4.2	Users should provide details about themselves in their account settings. Some of this information should be associated with the workflows they upload.	O

U7.1.5   Share an interesting workflow with a colleague. O

User Requirements

R7.1.5.1	See 3.4.2 for further details about how user data is collected.	O
R7.1.5.2	Users should be allowed to share a workflow with another specific user of the service, irrespective of groups.	D

U7.1.6   Reproduce the results of another research team. D

User Requirements

R7.1.6.1	To reproduce results exactly, one needs the workflow and the data set it was applied to (i.e. the search property set.) This could be accomplished by having research teams enter their publications into an article database in neuGRID. When entering the publication reference, the team could be asked to supply the names of the workflows that were used and a copy of the search property set. This could be made a requirement for access to the neuGRID project being granted.	O
R7.1.6.2	Provide users the capability to download a workflow, import to their workflow execution environment and compare with the results of previous executions.	D
R7.1.6.3	There should be a way to reuse a given dataset on a given workflow.	E

U7.1.7   Certification of workflows. D

User Requirements

R7.1.7.1	A policy is needed for who can certify workflows and the process by which certification takes place in neuGRID.	D
R7.1.7.2	Provide tools for certifying a workflow according to R7.1.7.1.	O
R7.1.7.3	An administrator should manage and control the certification process including requesting information regarding the data/software/workflows as needed.	D

U7.2   Request a given research community to develop a new workflow for a particular task or add a feature to an existing method.   O

User Requirements

R7.2.1	Supply contact details when uploading a flow as in R3.4.2.	D
R7.2.2	Provide a category of tags that can be added to a workflow and allow users to request new features.	O
R7.2.3	Share new workflow features with the research community according to the permissions of the various groups.	D
R7.2.4	After a certain period of exclusivity workflows of a given quality should be shared with the entire neuGRID community.	O
R7.2.5	Provide functionality within the service to enable users to request a workflow for a particular task from other users.	O

U7.3   Get assistance with the construction of a complex workflow from the wider research community.   O

User Requirements

R7.3.1	Provide a discussion forum within the system, in order for users to discuss and solicit advice from other users about construction of workflows. A category of tags could be created that can be added to a workflow and allow users to request assistance from more experienced researchers (this might ease the pressure on the comments section of the workflows.)	O
R7.3.2	A user guide is necessary.	D
R7.3.3	A technical glossary should be created.	D

U7.4   Request and interact with a consultant to construct a workflow. O

User Requirements

R7.4.1	Provide access to neuGRID-affiliated application specialists and consultants manning a built-in helpdesk.	O
R7.4.2	Organize a mailing list for workflow constructors so that important messages can be circulated.	D
R7.4.3	Provide functionality to interact with volunteer specialist users to construct new workflows.	O
R7.4.4	Specialist users may be given a special account, and may at their choice be listed for easy discovery.	O

U7.5   Identify weaknesses in workflows and act as a community to resolve them quickly.   O

User Requirements

R7.5.1	Provide a forum type capability to discuss specific workflows when problems arise (see R7.3.1.)	O
R7.5.2	Create a neuGRID community in which users can see which modules are the most used, the statistic concerning the different workflows, the efficiency or malfunction of these workflows, and other various topics of interest for the users.	O

U7.6   Rapidly deploy advanced techniques and use online collaboration for training purposes. O

User Requirements

R7.6.1	Organize some webinar meetings.	O
R7.6.2	Provide a modular service, so that new features can be added to enhance collaboration between users.	D

U7.7   Keep commercially or otherwise sensitive workflows private and secure. E

User Requirements

R7.7.1	See U1.5 for security related information.	E
R7.7.2	Identify different levels of security and confidentiality within the grid.	E
R7.7.3	Access can be restricted to one person only. Provide users with the capability to limit access to certain workflows.	E

Sharing Results and Histories

A user interacts with the system to search existing studies and to compare, contrast and validate their results against research from other groups. This process helps the researcher to identify an error in their methodology and prevents them from making any embarrassing claims. The researcher did a similar study six months ago and is worried that it too, might have been influenced by a similar error. The user looks up their research history and identifies the appropriate study. The original process can be re-run on the original data set using the stored settings and pipeline configuration. This allows the researcher to confirm that the previous results were correct.

Indicative Use-cases:

U7.8   Create groups of researchers with similar / overlapping interests.   O

User Requirements

R7.8.1	Add "Research interests" to the profile data entered in R2.1.2.	O
R7.8.2	Similar interests could be assessed during user registration through a simple and fast questionnaire as checklists or free, searchable text.	O
R7.8.3	Allow users to create virtual groups within the service.	O

U7.9 A user can search their and public histories for a specific workflow execution and allow it to be re-run on the original or new data. D

User Requirements

R7.9.1	The workflow setups from R4.5.1 can also be saved and searched under "My workflow setups." They could be coupled with the data set that was processed with these criteria (or the search property set that was used to generate the data.)	O
R7.9.2	A tag can be set specifying that the workflow setup is public. Then a search for a specific workflow execution can also include all public workflow setups (and their connected processed data sets/search property sets).	O
R7.9.3	Provide a simple query interface through which past executions can be discovered.	D
R7.9.4	A user should be able to download an archived workflow specification and select new/existing data set for processing.	D

U7.10 Allow records of common mistakes to be searched to improve the training of new researchers. O

User Requirements

R7.10.1	Create documentation or a FAQ like page for "Frequent errors & workflow mistakes." These might give tips on how to check that the output from block A can be used as input to block B in a workflow.	O
R7.10.2	Create a user comment database where researchers can note mistakes they made and how to avoid them.	O
R7.10.3	Log the error outputs and compile statistics on their frequency. The helpdesk (R7.4.1) could help connect the error outputs to the mistakes creating them.	O
R7.10.4	Save a certain amount of bad workflows executions that should be useful as examples for the new users of the neuGRID platform.	D
R7.10.5	Make a validation test on the main tools that neuGRID provides.	D
R7.10.6	There should be a way to store non-standard patients, typical examples etc for a given workflow in a separate store.	O

The remainder of this section considers the image processing and statistical analysis tools that are in frequent use by the research centres within the neuGRID project in greater detail. The purpose of this is to provide an increased level of detail regarding individual software packages and tools. It was generally felt that including these within the earlier requirements specification might complicate what is presented and so this dedicated segment has been created. In D6.1 it was proposed that the brain imaging tools broadly fall into the following categories:

Image Processing

This includes a library of image processing algorithms focused on manipulating the source images so as to ultimately extract features of the images which can be used in a variety of statistical analyses. Examples of this include, the spatial normalization and blurring operations necessary to perform so-called Voxel-Based Morphometry (VBM); the registration and surface extraction algorithms used in the estimation of cortical thickness (e.g., the CLASP

algorithm [8]); or the registration and voxel classification algorithms used in brain tissue identification. Some real-world examples include:

Example 1: FSL/TBSS

A recently released version of FSL, with the latest version of TBSS was installed. TBSS was then run on 61 scans, each of which was under 1MB in size. TBSS required CPU intensive calculations to be run on each pair of scans. Therefore, 3,721 jobs, which took about 30 minutes each, needed to be run. Combining the output of the 3,721 jobs was easily performed on a single machine after the completion of all the jobs.

Example 2: FLUID

For an Alzheimer's study, 180 pairs of MRI scans needed to be compared to detect how the shape of the brain changed over time. The specialized software Fluid, which was available as a Linux executable, was used to compare the pairs of scans. The FSL routines BET and FLIRT were used to pre-process the scans before Fluid. Each pair of scans took about 6 hours of CPU time to process. Each individual scan was about 23MB in size.

Statistical Analysis

This includes any statistical analyses performed on data, be they "raw" (unprocessed) source data or more likely data processed using the library of methods covered under "Image processing."

Statistical Analysis Example:

In order to locate structural changes within the hippocampal formation in AD patients of mild to moderate severity, several analysis steps are performed. First of all, the hippocampal formation has to be isolated by manually tracing on MRI coronal slices. Then 3D parametric surface mesh models are generated from the manually segmented hippocampal tracings. The models of each individual's hippocampi are analyzed to estimate the regional specificity of hippocampal volume loss in AD compared to controls. To assess hippocampal morphology, a medial curve is automatically defined as the 3D curve traced out by the centroid of the hippocampal boundary in each image slice.

The radial size of each hippocampus at each boundary point is assessed by automatically measuring the radial 3D distance from the surface points to the medial curve defined for individual's hippocampal surface model. Shorter radial distances are typically used as an index of atrophy. Atrophy maps are visualized on 3D models of the hippocampal formation. The percent change relative to control and the associated P value describing the significance of group differences are plotted onto the model surface at each point of the hippocampus using a colour code to produce statistical maps. Overall P values are computed for the maps of the left and right hippocampal formation using a permutation testing approach. Permutation methods measure the distribution of features in statistical maps that would be observed by accident if the subjects were randomly assigned to groups and provide a P value for the observed effects that is corrected for multiple comparisons.
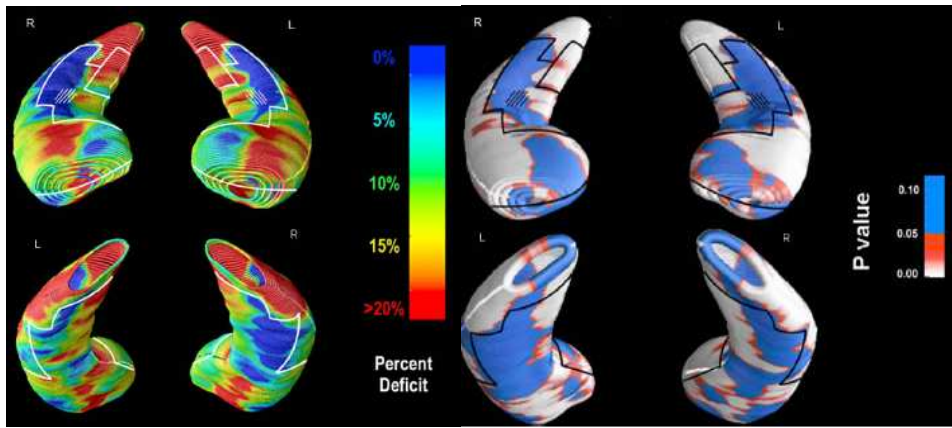
Figure 4: Topography distribution of atrophy in the hippocampal formation of AD compared to elderly controls. On the right is shown the hippocampal volume loss corresponded to a P value ranging from 0.10 to 0.00

The following table lists the pipeline tools that are in frequent use by the research centres within the neuGRID project.

Institute	PIPELINE TOOLS/ OPERATIONS	Analysis Tools

VUmc (See APPENDIX A)	• FSL Tools (FMRIB Software Library): FLIRT, FNIRT, FDT, FAST, BET, FEAT, Melodic, Siena, XSiena.
	
	• MNI (BIC Tools & Software): N3.
	
	• BIRN (Gradient Non-Linearity Distortion Correction): Gradient non-linearity.
	
	• DRG Fluid.
	
	• Generic:
	  • Image calculations (adding subtracting, multiplying etc.)
	  • Morphological operations on images
	  • File format conversion
		SPM (Statistical Parametric Mapping)

KI (See	CIVET Pipeline (Pipeline 3) (CIVET Pipeline), FSL, Brainvoyager, Matlab, AFNI, E-prime and	Hermes (Hermes Medical)

APPENDIX B)	Statistica.	(Pipeline 1 and Pipeline 2)
FBF (See APPENDIX C)	• FSL Tools: FMRIB's Diffusion Toolbox FDT 2.0, Melodic, BET Function, FLIRT, FNIRT, Siena, Melodic • MNI Tools: Display, Register, • Brainsuite • LONI functions (LONI Software Tools): Dual_warpe_warpcurve, Decoder_blend_all, mk_seg16bit, mk_gray, add_gray_to_inflated_LEFT1, add_gray_to_inflated_RIGHT1, pmap_apeVSctrl, make_UVL_*; 1st_script_tracer_avg_DIAG; 2nd_script_core_test_L_DIAG; 2nd_script_core_test_R_DIAG; Pmap_DistCore_DIAG • MRIcro (MRIcron) (visualization) • Quanta 6.1 • IdeALab Tools (IdeALab) • Image Conversion software: MRIconverter; dcm2nii; nii2ana and fslchfiletype. New Promising Tools FBF is evaluating: 3D Slicer, VTK, Freesurfer, MIPAV, NA-MIC Kit components, MED-INRIA, BrainVoyager, BrainMAP	SPSS, SPM, (Statistical Parametric Mapping) Matlab R2008a R

Common Pipeline Tools

Pipeline Tools	INSTITUTE
FSL Tools	VUmc, FBF
MNI tools	VUmc, KI, FBF
Generic Operations (Image conversion, calculations)	FBF, VUmc

Common Analysis tools

Analysis tool	INSTITUTE
SPM	VUmc, FBF

6. Non-Functional Requirements

Several non-functional requirements have been identified in the previous section. These relate especially to user interface and certain performance requirements. This section specifies the remaining non-functional requirements that have been collected.

Where E = Essential D = Desirable O = Optional

NR1   The ability to manage restricted bandwidth by submitting processing intensive standalone computing jobs to the closest high-performance grid node. Data should be archived and made available locally. In this way we will ensure that the time for image processing will be quick and avoid any problems during the sending of images from one node to another. Another important aspect could be to move the queries from one centre to another, rather than large quantities of data.   D

NR2   The system should be designed so that it can be scaled up when new centres join the infrastructure. Further functionality should where possible be incorporated into the image analysis processing architecture. neuGRID should be flexible enough to be able to evolve and support a range of capabilities in the future.   E

NR3 The system should, where possible, be capable of exporting data into a range of future analysis packages. D

NR4 From the user perspective, the neuGRID platform should be accessible through and compatible with all major operating systems (Mac OS, Unix/Linux and Windows.) E

NR5 An online help facility should be incorporated into the system. D

NR6 A service level agreement should be put in place at an appropriate level (perhaps 95%) and improved where possible. E

NR7 Users should be equipped with the best allocation of storage and compute resources that are possible. E

NR8 The system should be compatible with the g-Lite middleware. E

NR9 The system should be compatible with and be able to make use of grid resources that run a range of different middleware other than g-Lite. Components must not be employed that couple the system to any particular middleware or software package. D

NR10 The neuGRID infrastructure should have a sustainable post-project plan. E

NR11 The infrastructure must be fully compliant with Service Oriented Architecture principles and design methodologies. E

NR12 The medical services that are produced are to be generic and reusable. E

NR13 The look and feel of the user interface should where possible, follow common neurological research environments and users should be able to recognize functions and options that they usually use in the different analysis tools. As a consequence and where possible, commonly used analysis functions should be incorporated in neuGRID. D

NR14 Where users construct new pipelines themselves, an appropriate disclaimer should be put in force regarding potential errors. E

7. User Functionality and Validation Scenarios

This section is intended to bring the functional and non-functional requirements together and to produce a verifiable set of core features that will be made available to end users as the project develops. This will focus on the user groups that were identified in D9.1. These included the basic, intermediate, advanced and pipeline developer user roles. Alongside each set of core features, a scenario is described that will be used to evaluate the specified functionality during system testing by WP11.

7.1. The Basic User

The neuGRID Infrastructure will provide:

- The ability to select from and use a library of existing and validated research workflows.
- A search mechanism through which the raw data that is accessible by neuGRID can be found and grouped into a research set for analysis using a workflow.
- Access to distributed computing resources.
- A means of visualizing a workflow in a familiar graphical way.
- The storage of all the intermediary output from workflow executions and a means of accessing this data in order to confirm the results that are produced.
- The ability to execute a workflow on a given research set using the glueing service that allows access to a range of different distributed resources.
- A secure portal that allows users to manage their accounts and that enables them to interact with the infrastructure.
- The ability to export the results from a workflow execution in a generic format that can then be imported into a range of user defined statistical analysis packages.

Test Scenario:

As part of a research project a new PhD student wishes to run the FSl SienaX algorithm on a data set of sagittal MPRAGE scans.

Possible Stages in Scenario:
- The user securely logs into neuGRID.
- Searches for and selects an appropriate dataset based on its orientation.
- Searches for and chooses a version of FSL.
- Performs the analysis using grid resources.
- Gets back the brain volumes and brainmasks.
- Exports results into an analysis package.

7.2. The Intermediate User

In addition to the features that will be provided to Basic users, the neuGRID Infrastructure will:

- Allow researchers to control the parameters that are applied during workflow execution.
- Provide a workflow editing facility that allows users to edit existing workflows and to tailor them to their needs.
- Keep a record of the changes that are made to workflows and the settings that were applied during each execution.
- Share workflows with other neuGRID users.
- Have a graphical means of building new workflows in a drag and drop manner from a range of validated modules.

Test Scenario:

In Scenario 7.1 it is found that a number of the scans suffer from poor results during the brain extraction stage. The PhD student asks a more experienced fellow student (intermediate user)

to have a look at the problem. It is known that the BET algorithm (of FSL) has a number of parameters which may affect the result. The user tries to tweak the parameters using a mixture of experience and educated guesswork. The settings are tuned to specific scanner dependent parameters. The workflow is also edited to include site specific scanner BET parameters. The user checks whether a newer version of FSL gives better results. The intermediate user saves their improved workflow and passes it back to the original basic user. The intermediate user is still not fully satisfied with the results from the new workflow and so passes it on to a postdoc researcher who is working in the lab (advanced user) for further enhancement.

Possible Stages in Scenario:
- The user edits workflow parameters.
- Edits the workflow using a graphical drag and drop interface.
- Executes the workflow using different versions of FSL.
- Shares new workflow with basic and advanced users.

7.3. The Advanced User

In addition to the features that will be provided to Basic and Intermediate users, the neuGRID Infrastructure will:

- Manage the versioning and documentation of workflows as they evolve.
- Provide workflow debugging features through which new workflows can be validated.
- Enable users to select a test research set from a list of sets that have known properties and are useful for testing workflows.
- Enable workflow specification (scripting) and execution via the command line.

Test Scenario:

After evaluating the output from Scenario 7.2 the advanced user finds that BET often includes a significant amount of the neck in the brain extraction results. In order to prevent this from happening in the future, an algorithm is created to remove the neck. This is achieved through a registration using a template brain and an estimation of what part is the neck. This is then removed before the SienaX algorithm is called.

Possible Stages in Scenario:
- The user investigates and debugs the workflow using a number of test research sets.
- Via the command line, the user creates modified workflows using the new algorithm.
- Through the command line the user executes the new workflows and evaluates the results.
- The changes that are made to the workflow are captured and recorded.

7.4. The Pipeline Developer

In addition to the features that will be provided to Basic, Intermediate and Advanced users, the neuGRID Infrastructure will:

- Allow users to traverse error logs in order to determine the causes of workflow execution failures.

- Provide a quality control mechanism that helps users to understand why a workflow component fails on a given image.
- Enable pipeline developers to visualize complex workflows and thereby locate potential points of failure.
- A mechanism by which workflows can be validated and released for use by the wider research community.
- Subject to approval, support the development of new modules / algorithms and the submission of these for use in the infrastructure.

Test Scenario:

In order to make the approach that was discovered in Scenario 7.3 available as a graphical pipeline, a pipeline developer does some further work to ensure stability and packages it up in such a way that it can be submitted for inclusion within neuGRID. This allows it to be extended and re-used by basic and intermediate users. The neck stripping is brought into a graphical representation. It is also fully documented and supplied with a dataset which it can be tested against.

Possible Stages in Scenario:
- The developer looks at the error logs to confirm that the algorithm and new workflow from the advanced user is effective.
- Using visualization techniques the developer confirms that the new workflow is robust.
- The workflow is approved by the developer and submitted for inclusion in neuGRID along with a reference test set and appropriate documentation.
- neuGRID approve the workflow and it is made available to users of the infrastructure.

WP9 will play a role in confirming that the required user functionality is delivered by neuGRID. In order to achieve this the capabilities that have been detailed in this section will be arranged into sets that correspond to the delivery of system components during 2010. This will be done by the leader of WP9 (PB) in conjunction with the area leaders (AZ and DM) and technical supervisor (RM.) At each of the remaining face to face meetings, system functionalities will be rolled out to end users and evaluated in the context of D9.2. A User Manager (PB) will work with and represent the interests of users during this process.

8. Conclusion

The requirements revision process has gathered feedback from developers and WP leaders. Developers have been able to ask for further information and clarification where greater details regarding specific requirements were needed. The initial requirements that were gathered during the preparation of D9.1 have been individually evaluated in the light of subsequent developments in the project. An effort was also made to ensure that the priorities assigned to requirements were, wherever possible internally consistent with each other. The final stage in the revision process was the identification of the functionality that each group of end user (basic, intermediate and advanced) can expect as a minimum from the final neuGRID platform. A usage scenario has been identified for each user group and will be used to exercise system functionality during the system validation which will be carried out as part of the integration testing by WP11.

A second round of visits to each of the clinical sites (FBF, VUmc and KI) was originally planned. Given that the requirements revision task was brought forward by three months this series of meetings was found to be unnecessary. Instead of this, the user requirements team will where possible, take part in presenting the prototype neuGRID infrastructure to end users. This will allow them to benefit from the information and questions that developers gather during the analysis and prototyping of system components. Where prototypes have been produced, they can be used to validate the requirements that have been gathered thus far and provide useful feedback to developers. It is felt that this will encourage the translation of the final URS into a successful neuGRID infrastructure that addresses the essential requirements of users. To this end, a new role in the running of the project named "User Manager" has been created. The User Manager will work with end users to validate that their requirements are addressed by the final infrastructure.

Bibliographical References

[1] Systems Engineering Fundamentals. Defence Acquisition University Press, 2001, ISBN: 0160732905 .
[2] Clegg, Dai; Barker, Richard (2004-11-09). Case Method Fast-Track: A RAD Approach. Addison-Wesley. ISBN 978-0201624328.

APPENDIX A: VUmc Pipelines and Capabilities

VUmc are contributing to neuGRID in three main ways:

1. As a **data acquisition centre** Over the years, the VU university medical centre has built up a wealth of images, including MRI, CT, PET and MEG data, from a wide patient populations including Alzheimer's, other types of dementia, and multiple sclerosis. Supplemental data, such as medical history, MMSE, EDSS, and CSF samples etc., were also acquired. This data may be made available to neuGRID depending on the patient's informed consent and other considerations. The VUmc also participated as an acquisition centre in the EADNI pilot project.

2. As a **core lab** the Image Analysis Centre (IAC) within the VUmc performs core lab functions for various clinical trials, including pharmaceutical trials. As part of its core lab functions, the IAC co-ordinates and collects data from acquisition centres, quality controls the collected data, and can fully anonimize (including defacing of) the collected data.

3. As an **image processing lab** VUmc has an ongoing research line assessing existing software for analysing MRI scans of the human brain. Assessments include (1) which software package is the best for performing a particular segmentation or calculating a measure or biomarker, (2) the reliability of the software, (3) whether the software works correctly on data that was acquired under different circumstances than it was original designed for, such as different patient populations and/or other MRI sequences. It also develops/test combinations of existing software to perform extended measurement. For example, VUmc recently evaluated the "fluid" software (Dementia Research Group of London) for atrophy measures of the hippocampi instead of its routine use on the whole brain.

A summary of the pipelines in use at VUmc. While they do use other packages, (more detail is provided below) FSL appears to be the primary tool.

Researcher's Comments:

The most important packages for use in the near future in FSL are TBSS,
maybe FIRST (for segmentation) and FDT (other DTI tools) too.

Building blocks that can be combined into pipelines (this is how we usually work!):

- Image intensity homogenisation e.g. MNI N3.

- Geometric corrections e.g. BIRN Gradient non-linearity distortion correction.

  Registration:
- Linear e.g. FSL FLIRT.

- Nonlinear e.g. DRG Fluid, FSL FNIRT.

- VBM e.g. SPM.

- DTI Tracing e.g. FSL FDT.

Segmentation:

- Brain e.g. FSL BET.

- Grey / white matter e.g. FSL FAST.

- fMRI analysis e.g. FSL Melodic, FSL FEAT.

- Brain volume measurements e.g. FSL Siena, FSL SienaX.

File format conversions.

- Image calculations (adding, subtracting, multiplying etc.)

- Morphological operations on images.

Examples of pipelines we currently use (identically applied for multiple subjects):

* Brain volume measurements:
- [needs file conversions dicom/nifti]
- BET
- SienaX + Siena

* Non linear registration of brains:

- [needs file conversions dicom/nifti/mgh/minc]
- Gradient distortion correction
- N3
- BET
- Linear registration
- Fluid

* Non linear registration of hippocampi:

- [needs file conversions dicom/nifti]
- [needs file conversions for manually drawn ROI files]
- Extraction of subimages within ROIs
- Fluid

* VBM

- [needs file conversions dicom/nifti]
- Linear registration
- Nonlinear registration

- Segmentation (e.g. grey/white matter)
- Voxelwise calculations w.r.t. template/atlas/average

APPENDIX B: KI Pipelines

Examples of pipelines in use at KI SMILE:

Volumetric pipeline 1:

1. Data is moved into SMILE by
2. Downloading DICOM files from the internet
3. Making an import of DICOM images to the Hermes DICOM server (see below)
4. Sending data from the hospital's PACS system (patient database) to the Hermes DICOM server

Hermes is a commercial system (see www.hermesmedical.com) with its own software solutions and file format (InterFile). A DICOM server forms the image database and applications can be launched within the "GOLD" milieu. It is possible to develop lab-specific programs (in C) and turn them into local Gold applications.

5. The MR-data is subjected to structural analysis via the following steps:
6. Pre-processing with in-house Hermes application which reorients the brain and re-slices it
7. Registering the brain using 9 parameters in Hermes multimodality application
8. Performing brain extraction… (i.e. skull stripping)
9. Performing inhomogeneity correction…
10. Segmentating tissues…
11. Performing regional analysis…
12. …all with the help of in-house Hermes applications.

Note that all systems are inside the hospital's firewall. There is a special telerad connection between different hospitals in the Stockholm area which is used to transfer images between the hospitals' PACS systems.

Volumetric pipeline 2:

1. Images are continuously scanned and transferred in DICOM format via SCP from the camera (at another hospital) to an account in a Linux machine at SMILE.
2. The images are imported into Hermes
3. Register the brain in Hermes in cubic voxels
4. Start the program cut out and zoom in to make the brain bigger
5. Make new mean slices in multimodality (Process->Add slices) to average four slices into one
6. Use (Process->View) to show nine averaged slices at a time
7. Print out the views (with nine slices on each page)
8. Export the images in DICOM format from Hermes to a Linux machine
9. Use avwswapdim to change the axes (Hermes swaps axes around)
10. Use the freeware MRIcro to rescale the images (lessens "granularity")
11. Run BET in MRIcro to extract the brain
12. Use MRIcro to show the brain surface in 3D and compare it with the slice printouts to identify landmarks such as the frontal gyrus and the frontal/orbital cortex

13. Mark these landmarks by hand on the printout
14. Return to the Hermes system and use the in-house application Display MR from scaled to perform greyscale normalization
15. Step through the slices and draw ROI:s (region of interest) on the various gyri
16. Collect the ROI:s into a VOI (volume of interest) and save it
17. Run multimodality on both brain and VOI, using a personalized protocol with initial values
18. The VOI is displayed on the 3D image of the brain, check for consistency.

INNOMED project (SMILE part):

1. DICOM-data arrives on CD
2. The data is read into a Linux Ubuntu machine using rsync
3. The data is sorted through using various perl scripts to see that all parameters (date of birth etc) exist and that the images have been anonymized. A cross-check that all parameters are the same between visits is also made.
4. If everything is OK the data is uploaded to the DICOM archive database on the server outside KI's firewall (otherwise the responsible site is contacted).
5. The data is converted to MINC format.
6. A manual QC is made on the MINC images in the database, looking for among other things homogeneity, coverage and artifacts such as ringing and movement. If the images do not pass QC a rescan of the patient is requested.
7. The data is run through the CIVET pipeline, which uses perl scripts developed at McGill to do inhomogeneity correction, skull stripping etc.
8. After the processing is done, the server outside the firewall contains images, processed images and clinical data (also memory test results etc) for each scan.

At the moment we use the following programs sparingly, but have and will use them again:

- FSL.
- Brainvoyager.
- Matlab.
- AFNI.
- E-prime.
- Statistica.

APPENDIX C: FBF Pipelines

PIPELINES IN USE AT IRCCS - FBF			
PIPELINE NAME	MODULES USED	PRIORITY	NOTES
IMAGE CONVERTION; VISUALIZATION & REGISTRATION	MRIconverter (freeware: http://lcni.uoregon.edu/~jolinda/MRIConvert/) / dcm2nii (freeware: http://www.sph.sc.edu/comd/rorden/mricron/dcm2nii.html)./ MNI ad hoc functions (mnc2dcm, dcm2mnc, ana2mnc, mnc2ana, minc2nii, nii2minc, ana2dcm, dcm2ana) (GNU: http://packages.bic.mni.mcgill.ca/) / FSL tools (fslchfiletype) (GNU: http://www.fmrib.ox.ac.uk/fsl/) / 3D-SLICER / MIPAV/ MRIcron / MNI tools / ITK library.	Low	
PET-FDG IMAGE PROCESSING PIPELINE	SPM5 (Matlab) & homemade scripts (ppvspm.m; ppv_template.m; ppv_priors.m; ppv_complete.m; ppv_TPC; ppv_defaults.m, mask.m, normalize.m) (GNU: http://www.fil.ion.ucl.ac.uk/spm/software/)	High	
MRI IMAGE PROCESSING PIPELINE	SPM5 (Matlab) & homemade scripts (ppvspm.m; ppv_template.m; ppv_priors.m; ppv_complete.m; ppv_TPC; ppv_defaults.m, normalize.m)	High	
DARTEL	SPM5 (Matlab)	High	
VOXEL BASED MORPHOMETRY (VBM)	SPM5 (Matlab) (GNU: http://www.fil.ion.ucl.ac.uk/spm/software/)	High	
INDIPENDENT COMPONENT	FSL-MELODIC (GNU: http://www.fmrib.ox.ac.uk/fsl/) / GIFT	Medium	

ANALYSIS (ICA)	(Matlab) (GNU:http://icatb.sourceforge.net/)		
CORTICAL PATTERN MATCHING (CPM)	MRICro (freeware:http://www.sph.sc.edu/comd/rorden/mricro.html), SPM99/SPM2/homemade scripts (MatLab), DISPLAY 1.4.2 (freeware: http://packages.bic.mni.mcgill.ca/), MNI functions (mni2ana, register, classify, ana2mnc, myana2mnc, crop_mnc, crop_back.sh, mincmask, mincresemple) (GNU: http://packages.bic.mni.mcgill.ca/), BrainSuite (freeware: http://brainsuite.usc.edu/), LONI analysis tools (Dual_warpe_warpcurve, Decoder_blend_all, mk_seg16bit, mk_gray, add_gray_to_inflated_LEFT1, add_gray_to_inflated_RIGHT1, pmap_apeVSctrl) (Private-Licence)	High	
WMHs MAPPING (WHITE MAPPING HYPERINTENSITIES)	Quanta 6.1 & other IDeALab Tools (svcleanup, 1.2.chg_parityFL, chg_nameFL_ima, ima2img, chg_data-matchParity, LinCoreg3, wmt_replace, sv) (GNU & Private PV-WAVE Licence: http://neuroscience.ucdavis.edu/idealab/software/index.php), BET function (freeware: http://www.fmrib.ox.ac.uk/fsl/bet2/index.html)	Medium	
DTI (TRACTOGRAPHY AND DIFFUSION TENSOR)	FMRIB's Diffusion Toolbox - FDT v2.0 (FSL), MRIconverter	High	
RADIAL ATROPHY MAPPING (RAM)	MRICro, SPM2 & home made scripts (MatLab), Dx (freeware: http://www.opendx.org/download.html), Seg3D, MNI functions (mni2ana, register, classify, ana2mnc), LONI analysis tools (make_UVL_*; 1st_script_tracer_avg_DIAG; 2nd_script_core_test_L_DIAG; 2nd_script_core_test_R_DIAG; Pmap_DistCore_DIAG) (Private-Licence)	High	

HIPPOCAMPUS VOLUME	MNI functions (dcm2mnc, preproc, mincresample, mincinfo, mincreshape, autocrop, volume_extraction, manualfit, linfit), REGISTER 1.3.6 (GNU: http://packages.bic.mni.mcgill.ca/), DISPLAY 1.4.2 (GNU: http://packages.bic.mni.mcgill.ca/), SPSS 12.0 (Private Licence).	Medium	
TOTAL INTRACRANIAL VOLUME (TIV)	MNI functions (dcm2mnc; autocrop; mincinfo; mincreshape; mincresample), DISPLAY 1.4.2. ; MultiTracer SW.	Low	
CORTICAL THICKNESS ESTIMATION	CIVET Pipeline; Brain-Visa; Freesurfer	High	
STATISTICAL ANALYSIS	R; Matlab	High	