



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: D3.1, Data dictionary of the unified neuGRID
database.**

Due date of deliverable: month 03

Actual submission date: 2009/07/24

Start date of project: February 2nd 2008 Duration: 36 months

Organisation name of lead contractor for this deliverable: Provincia Lombardo Veneta Ordine
Ospedaliero di San Giovanni di Dio - Fatebenefratelli

Revision: 2nd Version

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 Services)	
RE	Restricted to a group specified by the consortium (including the Commission	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of contents

Executive summary	3
1. Introduction	3
2. Methodological approach	3
3. Activity carried out and results	3
4. Conclusions	4
5. Bibliographical references (if applicable).....	4
6. Publications resulting from work carried out in the project	4
APPENDIX A: NeuGRID Data Dictionary.....	4

Executive summary

A specific aim of the neuGRID project is to offer researchers access to “ADNI-compatible” data acquired using the ADNI (Alzheimer’s Disease Neuroimaging Initiative) protocol, and for them to be able to use neuGRID for their scientific studies. For this purpose, a unified data dictionary is required that describes in detail the data and variables that will be available for analyses through neuGRID.

1. Introduction

Deliverable 3.1, “Data dictionary of the unified neuGRID database” is a data dictionary describing the variables and data which will be available for research conducted through neuGRID. Specifically, this data dictionary will unify the data descriptions of data acquired using the ADNI and AddNeuroMed (ANM) protocols, and in essence results in an ontological mapping between the two projects. This data dictionary will be used as a reference for researchers, as well as made available through the neuGRID on-line interface so as to allow users to select variables, formulate queries, and conduct analyses using these variables. The specific objective of this deliverable is the neuGRID data dictionary document.

2. Methodological approach

The generation of the unified data dictionary consists of the following stages:

1. Collection of the ADNI and AddNeuroMed acquisition protocols.
The data collected from the two different datasets were compared in-depth in order to make cross-studies and aggregate analyses possible.
2. Generation of unified variable nomenclature and variable descriptions, necessary to make the core structure of the neuGRID dataset accessible to all participants via GUI (Graphical User Interface).
3. Analysis of the comparability between the ADNI and AddNeuroMed variables:
 - a. Identify subsets of variables with complete equivalence (identical variables that exist in both datasets)
 - b. Identify subsets of variables which require ontological mapping (partially identical variables that may be linked as a result of a careful assessment of differences)
 - c. Identify subsets of variables which are unique to each protocol and cannot be mapped (totally different variables without any counterpart in the opposite dataset)
4. Establishment of explicit variable mappings between ADNI and AddNeuroMed in order to set up an *a posteriori* commonality database.

3. Activity carried out and results

Partners P2 NE and C01 FBF have collected, merged, and structured the ADNI and ANM acquisition protocols, resulting in the Data Dictionary shown in Appendix A. The major steps undertaken to realize the ADNI-ANM Data Dictionary can be summarized as follows:

- 1) Survey of the official ADNI and ANM protocols, in order to understand what data have been considered and how they have been managed;
- 2) Information gathering for structural and functional imaging methodologies adopted by ADNI;

- 3) Information gathering for tested biomarkers;
- 4) Information gathering for clinical data;
- 5) Collection and analysis of "Cognitive and Neuropsychological Test batteries" protocols;
- 6) Collection and analysis of "Clinical Repository forms" protocols;
- 7) Study of data collected and tools used by ADNI to extract information and to find relevant evidences.

Afterward, the resulting information was classified in 57 main ontological fields. This was done in order to visualize, in a comprehensive and rigorous way, the pattern of the ADNI database and to define the neuGRID data dictionary backbone.

Following the creation of the ADNI-only Data Dictionary as described here, the available ANM variables were added to the dictionary alongside the ADNI variables to facilitate comparison. Where possible the ANM variables, considerably fewer than the ADNI variables, were mapped to equivalent ADNI variables. When no equivalence existed, both variables were kept.

The data dictionary has been structured into a colour-coded, 9-column table, with the Field/Variable Category Name as the first column, followed by 4 columns for the ADNI data, and 4 columns for the ANM data, each set of 4 columns reflecting:

- Variable name: Variable acronym (for specific queries);
- Description: brief description of the variable;
- Type: Type of variable (Numerical, Text-String, Alfa-numeric, Date, Timestamp);
- Value: Value associated with the variable.

The colour-coding identifies the common variables, those unique to ADNI, and those unique to ANM. A few variables were identified which may violate ethical and privacy considerations; these were marked, to be assessed at a later point in time.

The resulting Data Dictionary contains more than 1900 variables in a hierarchically ordered structure which lists the most relevant factors, relationship and constraints between the variables in the ADNI and ANM data sets.

4. Conclusions

A complete, merged Data Dictionary containing ADNI and AddNeuroMed variables was constructed. This completes Deliverable 3.1.

5. Bibliographical references (if applicable)

- [ADNI Protocol \(Protocol # ADNI-024\)](#)
- [ADNI Procedures Manual](#) (version 1.2)
- [ADNI MRI Technical Procedures Manual](#) (version 1.0)
- [ADNI PET Technical Manual](#) (version 9.2)
- ADNI Cognitive Testing Worksheet.
- Additional information were collected from ADNI web site: <http://www.loni.ucla.edu/ADNI>

6. Publications resulting from work carried out in the project

TBD

APPENDIX A: NeuGRID Data Dictionary