An Interoperability Analysis of Virtual Imaging Laboratories for Marker Discovery in Neurodegenerative Diseases

2011-03-23 ISGC Conference
Academia Sinica – Taipei, Taiwan

David MANSET
outGRID Technical Coordinator
Imaging Markers - Background

What are markers used for?
- To support physicians in diagnosing diseases,
  - Prognosing may become a reality!
- To measure disease evolution,
- To assess treatment(s)/drug(s) efficacy,
  - Thus supporting pharma industries in drug developments,
- To further understand diseases and brain anatomy and functions

How do such markers materialize?
- Data mining Algorithms and Pipelines of Algorithms
- Heterogeneous Algorithms and Pipelines toolkits
  - I.e. FSL, MRIcron, FreeSurfer, MNI/BIC, LONI, SPM etc
- Computing and data intensive mining operations
Imaging Markers for Alzheimer’s
Gray Matter Loss

Isolated Memory Problems
Early Disability
Consolidated Disability
Imaging Markers for Alzheimer’s Nerve Fiber Loss
Recipe to Develop Markers for Alzheimer’s:

1. Large Databases (‘0,000s)

- US ADNI: ≈10,000 images
- Australian ADNI (AIBL): ≈4,000 images
- Japanese ADNI: (≈3,000 images)
- European ADNI (IMI Pharma-cog): (≈1,500 images)
Recipe to Develop Markers for Alzheimer’s:

2. Sophisticated Algorithms for Image Analysis
Recipe to Develop Markers for Alzheimer’s: 3. Sophisticated Statistical Models

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Gray matter change in</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy elders</td>
<td><img src="https://via.placeholder.com/150" alt="Graph" /></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very early Alzheimer’s</td>
<td><img src="https://via.placeholder.com/150" alt="Graph" /></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
neuGRID, A Grid-based Neuroscience Gateway
neuGRID Infrastructure integrates EGI Grid resources

neuGRID Infrastructure

Thousands of CPUs
Petabytes of storage
Strong liaison with BiomedVO/LSVRC

Scalable
Robust
Distributed
Grid
SOA
Workflow

neuGRID DACS are connected to GEANT2 Network

DACS1
DACS2
DACS3

20 Mb/s
100 Mb/s
100 Mb/s
1 Gb/s

20 Mb/s

Thousands of CPUs
Petabytes of storage
Strong liaison with BiomedVO/LSVRC

New Markers
Corelab
Pipelining

EXPLOITATION
2010 ONWARDS

neuGRID Infrastructure

Scalable Robust Distributed Grid SOA Workflow
Prototype Web Portal Concept

CoreLab
- Data Acq. (Imaging + Clinical Variables)
- Data Quality Control
- Data Anonymization

Pipelining
- Pipeline Authoring
- Pipeline Optimization
- Pipeline Enactment
- Multiple Toolkits (CIVET, BrainVISA, FreeSurfer, ITK/VTK, R etc)

Collaboration
- Document Sharing
- Pipeline Sharing
- Provenance Sharing
CIVET Pipeline
Pipeline Description

Non-uniformity correction, skull masking and tissue classification

Cortex masking and surface extraction

Gyrification index, resampling of surface and cortical thickness

- Various software dependencies (i.e. R, MINC, BIC etc)

heterogeneous distribution of pathological changes specific atrophy w.r.t. thickness of the cortical mantle

Alzheimer's characterized by heterogeneous distribution of pathological changes through the brain. One marker for the disease-specific atrophy is the thickness of the cortical mantle.
To promote interoperability among three e-infrastructures for computational neuroscience to converge into one unique worldwide facility.

**Related Initiatives Worldwide**

- **Home to US-ADNI/IDA**
  - LONI Pipeline Cluster Facility
  - NIH-funded
  - [http://www.loni.ucla.edu/](http://www.loni.ucla.edu/)

- **Access to large image datasets**
  - (future home to EU-ADNI)
  - Image Analysis Pipelines Grid Facilities
  - [http://www.neuGRID.eu/](http://www.neuGRID.eu/)

- **outGRID Facts**
  - FP7-funded
  - Started 11/09
  - 24 Months

---

**NATIONAL INSTITUTES OF HEALTH**

**CIVET Pipeline**

**HPC Facilities**

**canarie**

Canada’s Advanced Research and Innovation Network
Le réseau évolutif de recherche et d’innovation du Canada.
• Define activities (including research and development) and outline technical specifications for interoperability
• Foster, within the project’s lifetime, the maximum possible degree of interoperability allowed by this Support Action among the 3 infrastructures
• Lay the foundations for a larger research and development effort aimed to achieve full interoperability among the three infrastructures
In outGRID, interoperability is defined at two levels (DoW section B1.1.2)

- **Syntactic Interoperability** is required for infrastructures to communicate and understand each other.
  - If such systems are capable of communicating and exchanging data, they are syntactically interoperable.
  - For communicating data, specified data formats, communication protocols, interfaces descriptions and the like are fundamental.
    - E.g. XML or SQL standards provide syntactical interoperability.

- **Semantic Interoperability**, refers to the understanding and treatment of exchanged data.
  - Two systems that are able to process exchanged data and produce meaningful results out of it are considered semantically interoperable.
LONI Infrastructure (US)

- **Pipeline Server**
- **Access Point**
- **User**
- **Pipeline Client**
- **Study Module**
- **Pipeline Applet**

**Study Module**

- **User**
- **XML**

**Pipeline Server**

- **Submit**
- **cranium**
- **Network Usage**
  - 56.0 MB/s
- **Cluster Usage**
  - 67.6% used

**HPC Plugin**
- **GRID**
- **Query / Store**

**Provenance**
- **Files**
- **Query / Store**
- **Clusters / PCs**
- **P2P Plugin**
- **Execute / Manage**

**Storage Resources**
- **Drmaa/JGDi – Resources Management**

**Computing Resources**
- **LONI Infrastructure (US)**
- **Pipeline Server**
- **Pipeline Applet**

- **Pipeline Client**
- **Study Module**

**Use XML**

- **Dedicated HPC cluster** facility at UCLA/LONI,
- **306-node, dual-processor** SUN Microsystems V20z cluster. Each V20z node has **dual 64-bit 2.4 GHz AMD Opteron 250 CPUs with 8 GB of memory**
- **64-node** Dell development cluster, with each node using **dual 64-bit 3.6 GHz Intel EM64T processors and 4 GB of memory**
- **64-processor SGI Origin 3800 SMP supercomputer with 32 GB of memory**

**http://www.loni.ucla.edu/**
CBRAIN Infrastructure (CA)

- **Shared HPC resources** from public facilities
- **7 HPC Centers** sharing computing resources, 6 across Canada and 1 in Europe
- **13’500 CPU Cores**

http://cbrain.mcgill.ca/
neuGRID Infrastructure (EU)

- Dedicated Grid nodes combined with public computing resources from EGI
- EGI LSGC VRC
- 10,000 CPU Cores

http://www.neuGRID.eu/
Interoperability Synthesis

40% Compatibility
- Syntactic Interoperability
- Semantic Interoperability
- outGRID Interoperability

10% Compatibility
- Syntactic Interoperability
- Semantic Interoperability
- Today

Syntactic

Semantic
A WORLDWIDE E-INFRASTRUCTURE FOR COMPUTATIONAL NEUROSCIENTISTS

Technical Impact and Achievements

2011-03-23 ISGC Conference
Academia Sinica – Taipei, Taiwan

David MANSET
outGRID Technical Coordinator
Interoperability Demonstrator

International Web Portal

Shared Workflow Authoring Space

oG Exchange Point

50% DONE
- Workflow submission,
- SSO integration
- DPS integration

Data Semantic Specifics
Workflow Semantic Specifics

oG Exchange Point

100% DONE
- DPS reengineering,
- SSO integration

Data Semantic Specifics
Workflow Semantic Specifics

oG Exchange Point

95% DONE
- UDDI connection
- DPS integration

Data Semantic Specifics
Workflow Semantic Specifics

CBRAIN

LONI

neuGRID

Vertical Integration Approach

100% DONE - DPS reengineering, SSO integration
95% DONE - UDDI connection, DPS integration
50% DONE - Workflow submission, SSO integration, DPS integration

Sound Security
Super Infrastructure
Continental Access for Communities

LONI oGEP
McGILL oGEP
CBrain oGEP
neuGRID oGEP
UCLA
FBF

outGRID

connected cities
current connection
future expansion
anticipated expansion

Map showing connections between LONI, McGill, UCLA, FBF, and neuGRID through outGRID.
neuGRID / LONI Pipeline Integration

- LONI users can access neuGRID seamlessly and vice versa.
- Users can cumulate portfolios from both neuGRID and LONI.
- Complex workflows/pipelines can be designed across infrastructures.
- LONI IDA data collections can be automatically retrieved and processed in neuGRID.
**Interoperability Impact**

**Preliminary Impact:**
unified space for developing new biomarkers

---

<table>
<thead>
<tr>
<th>Algorithms/Pipelines</th>
<th>Data</th>
</tr>
</thead>
</table>

---

### Image processing algorithms

- **neuGRID:** For structural MR imaging analysis
- **LONI:** For structural, functional and diffusion imaging analysis
- **CBRAIN:** For structural MR imaging analysis

### Statistical tools

- **neuGRID:** “R”
- **LONI:** 12 different tools
- **CBRAIN:** Integrated voxel-based statistics and voxel wise-linear models

---

### Graphical user interface

- **neuGRID:** Secure Web portal LifeRay technology
- **LONI:** Pipeline Server interface installed on local computer
- **CBRAIN:** Secure Web portal, HTML 5 and WebGL 3D visualization capabilities

---

<table>
<thead>
<tr>
<th>INFRA</th>
<th>DATASET</th>
<th>IMAGES</th>
<th>ACCESSIBILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>neuGRID</strong></td>
<td>ADNI1</td>
<td>8000 structural MR @1.5T, 4000 structural MR @3T, 1300 FDG PET</td>
<td>Public</td>
</tr>
<tr>
<td></td>
<td>ADNI GO</td>
<td>Clinical variables in the range of 10K, ADNI GO/ADNI2: only 3T MR, PET amyloid imaging, rest fMRI in Siemens, diffusion MR in GE, expand to subjects w/ earlier symptoms.</td>
<td>Public</td>
</tr>
<tr>
<td></td>
<td>ADNI 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Australian ADNI (AIBL)</strong></td>
<td></td>
<td>285 individuals: structural MR and PIB PET imaging, npsy scores.</td>
<td>Public</td>
</tr>
<tr>
<td><strong>ICBM</strong></td>
<td>Structural MRI, fMRI, PET and FDG PET brain atlases</td>
<td></td>
<td>Public</td>
</tr>
<tr>
<td><strong>neuGRID</strong></td>
<td>NIH pediatric MRI</td>
<td>Serial structural MR, spectroscopy, DTI and clinical/behavioral data from 500 healthy children ages newborn to young adult.</td>
<td>Public</td>
</tr>
<tr>
<td><strong>CBRAIN</strong></td>
<td>AddNeuro Med</td>
<td>Serial, multicentre, 1.5T structural MR scans of 250 healthy elders, scanned, 250 Alzheimer's, 250 MCI patients.</td>
<td>Proprietary</td>
</tr>
</tbody>
</table>
STATISTICS

TO BE PROCESSED

1) Mean thickness of every subjects against time. Thickness should decrease with age and disease progression

2) Mean thickness of every subjects against gender. Male thickness should be thicker than women

3) Computation of Principal Component analysis to explain variability (this can be done plotting the 2 main components against the age and gender to assess if there is “age-related atrophy effect” or “gender-related effect”)

4) “Massively Univariate Statistics” to assess that subjects fit the linear model through a breakdown by diagnosis.
**Objective:** Validating interoperability demonstrator and architecture,

- “1 variable” approach, i.e. “same application with different data”,

- Developing a “super workflow” involving all 3 infrastructures,

- CIVET Pipeline to be run in all 3 infrastructures,

- Different datasets to be processed in each infrastructure
  - CBRAIN: ICBM
  - LONI: US-ADNI
  - neuGRID: EU-ADNI AddNeuroMed

**Technical Objective**

Grand challenge in numbers...

→ 11’000 MR Scans
→ 100’000 CPU Hours
→ 1.5 TB of Scientific Data

→ 2 weeks runtime
Objective: Interconnecting major international/national initiatives and having them collaborating towards one common goal.

- Neuroscience Infrastructures
  - CBRAIN CA
  - LONI US
  - neuGRID EU

- DCI Infrastructures
  - HPC CA
  - EGI EU
  - SHIWA EU

- Research Infrastructures
  - GéANT EU
  - CANet CA
  - Internet2.0 US

Gateway to other initiatives
Conclusions and Future Plans
“Perfect background for our push to include this project in the G8 summit end-May in Deauville”
Global Timing

We are here

1st Prototype Demonstrator “Any source, any sink”

Interoperability Workshop @ EGI User Forum

LINGA Challenge Launch
From Research Infrastructures to Translational Medicine

Imaging markers developments, debugging and assessment

Markers Provision, Decision Support and Translational Medicine

Research Infrastructures Interoperability

THANK YOU