

FROM RESEARCH TO INDUSTRY



## IMAGEN COHORT

A semantic web experience with  
the Imagen data

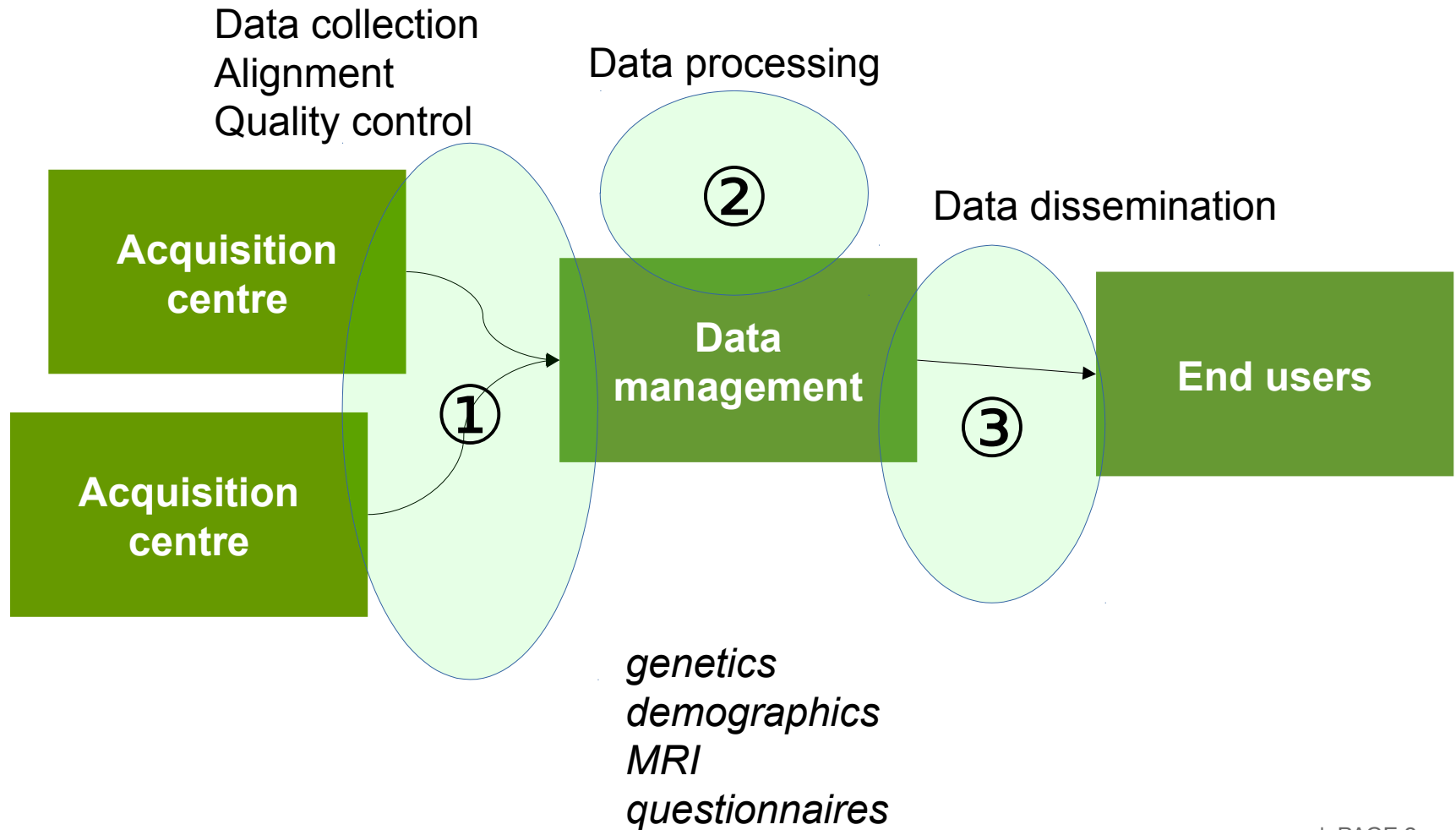


Gabriel Robert, KCL  
Dimitri Papadopoulos, CEA

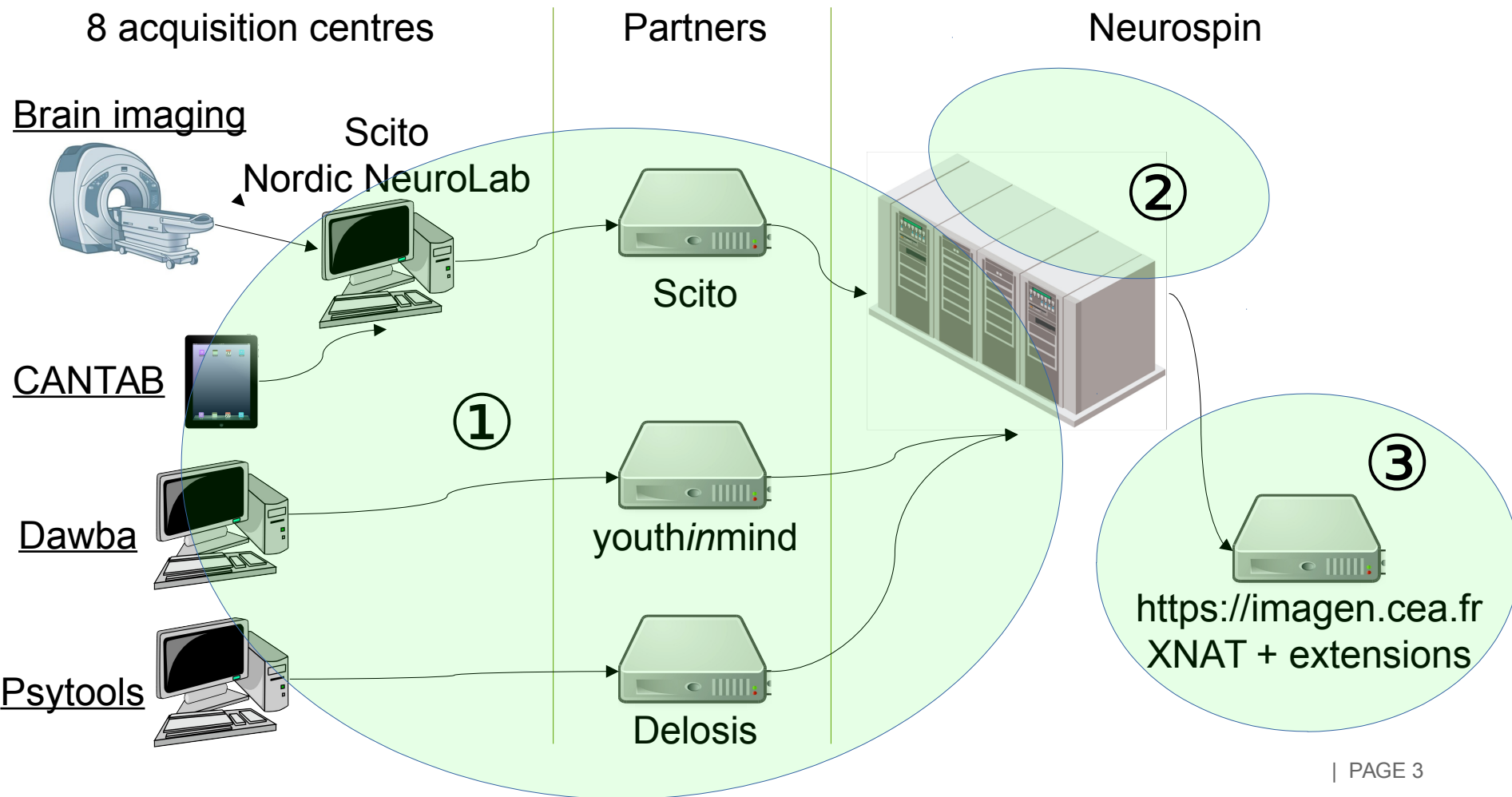
30 MARCH 2015

[www.cea.fr](http://www.cea.fr)

## Data management overview



2007-2012, 6<sup>th</sup> Framework Programme, 14 and 16 years of age



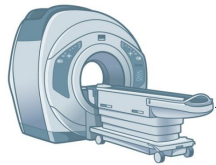
2007-2012, 6<sup>th</sup> Framework Programme, 14 and 16 years of age

8 acquisition centres

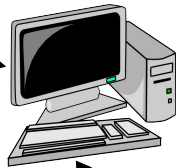
Partners

Neurospin

Brain imaging



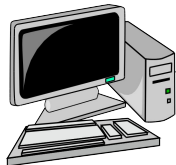
Scito  
Nordic NeuroLab



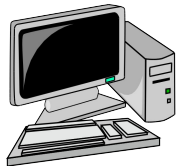
CANTAB



Dawba



Psytools



Scito



youthinmind



Delosis



<https://imagen.cea.fr>  
XNAT + extensions

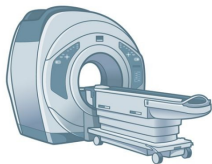
2013-..., 19 years of age

8 acquisition centres

Partners

Neurospin

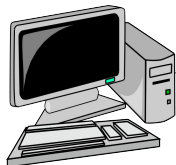
Brain imaging



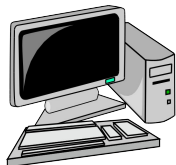
CANTAB



Dawba



Psytools



SFTP



youthinmind



Delosis



<https://imagen2.cea.fr>

CubicWeb  
Brainomics

## (Not so) subtle changes

- Subject identifier
  - Baseline (BL, 14 years) and Follow-up 1 (FU1, 16 years)
    - 080000123456
  - Follow-up 2 (FU2, 19 years of age)
    - 080000123456FU
    - Both versions actually found in FU2 data
- New people in acquisition centres
  - Standard operating procedures (SOPs) not followed precisely in FU2
- Unmaintained equipment in acquisition centres
- Core partners left
  - Lost access to software for data collection, alignment and quality control

## Lessons learned

- Use a **portal** for data collection: automatic **error detection early**
- Use standard software (free or commercial)
- Do not underestimate difficulty of data collection and curation
  
- Whatever you do, errors may be noticed late!
- **Data dissemination and data collection are linked**
- **Data may change over time**

## Web servers



### ■ XNAT

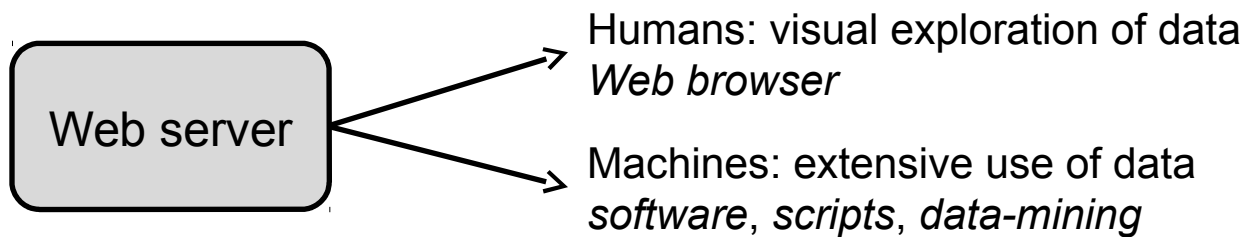
- No [XNAT](#) expertise in-house
- Long term maintenance of XNAT extensions
- [pyxnat](#)
- Performance issues

### ■ *Brainomics* project (Neurospin, Logilab, ...)

- Neuroimaging + genomics
- [CubicWeb](#) semantic web framework
- Specific Brainomics modules:
  - *medicalexp, neuroimaging, questionnaire, genomics, brainomics*



- Web of data – data can be processed by machines



- Query languages
  - Higher abstraction level than SQL
  - [SPARQL](#) Query Language for RDF – a W3C recommendation
  - Relation Query Language ([RQL](#)) – specific to CubicWeb

- URL publishing: associate URL to RQL queries
  - Reproducibility, the URL is a permanent link to specific data

**Any X WHERE S is Subject, S gender "male", X is Scan, X concerns S**

**<http://server.org/view?rql=Any+X+WHERE+S+is+Subject%2C+S+gender+%22male%22%2C+X+is+Scan%2C+X+concerns+S>**

- Performance
  - Reasonable query times with CubicWeb (seconds instead of minutes)
  - Compared to:
    - XNAT 1.4
    - In-house additions to XNAT

# EXAMPLES OF CUBICWEB USE

- Imagen is big data but not open: <https://imagen2.cea.fr/>
- Localizer is open data: <http://brainomics.cea.fr/localizer>