COMMIT/

CHALLENGES FOR SERVICES INTEGRATION INTO SCIENCE GATEWAYS: THE LOCAL STORY OF THE AMC

Silvia Delgado Olabarriaga

e-Science group
Dept of Epidemiology, Biostatistics and Bioinformatics
Academic Medical Center of the University of Amsterdam

www.ebioscience.amc.nl

COMMIT/

THE GOOD, THE BAD, AND THE UGLY

Silvia Delgado Olabarriaga

www.ebioscience.amc.nl

COMMIT/

THE “GOOD”:
THE COMMIT/ PROGRAM

http://www.commit-nl.nl/

use-inspired **fundamental ICT-research** program in well-being and well-working, in public safety, in science, in information services and search

2010-2016
173M€ (80M€ funded)
16 Projects
10 universities
5 tech institutes
>60 companies

COMMIT PROJECTS

INFINITI (Information retrieval for information services)
IUALL (Interaction for Universal Access)
SenseI (Sensor based Engagement for Improved Health)
Virtual worlds for well-being
SWELL (Smart Reasoning Systems for Well-being at Work and at Home)
SENSAFETY (Sensor Networks for Public Safety)
EWIDS (Very large wireless sensor networks for well-being)
ALLEGIO (Composable Embedded Systems for Healthcare)
METIS (Dependable Cooperative Systems for Public Safety)
THeCS (Trusted Healthcare Services)
TimeTrails (Spatiotemporal Data Warehouses for Trajectory Exploitation)
IV-e (e-Infrastructure Virtualization for e-Science Applications)
Data2Semantics (From Data to Semantics for Scientific Data Publishers)
e-BIOBANKING (e-Biobanking with Imaging for Healthcare
e-FOOD (e-Foodlab)

E-BIOBANKING WITH IMAGING FOR HEALTHCARE

http://www.commit-nl.nl/projects/e-biobanking-with-imaging-for-healthcare

Work packages

- e-BioCognition
- Molecular Imaging and Knowledge Management for Molecular Histology
- Front-end for Biomedical Data Analysis on e-Infrastructures
- Multiplex Imaging of Tissues
- Data Generation and Metadata Management for Biobanking with MRI and IMS
- BBMRI: Framework Selection for e-Biobanking: Identification and Evaluation of Core Services
- e-Biobanking for Inflammatory Bowel Diseases
- BBMRI: Mining of Distributed Biobanks: Coronory Heart Disease

COMMIT/ THE GOOD

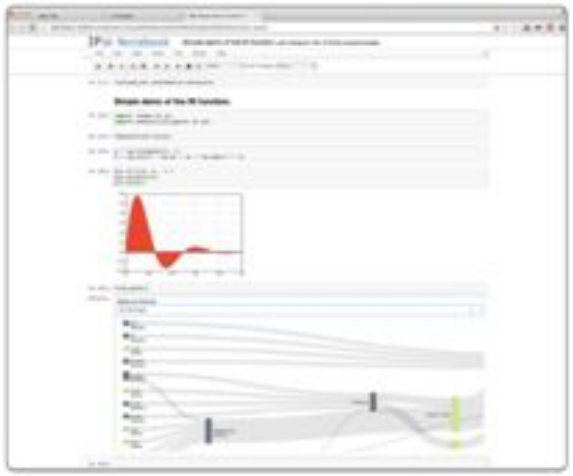
THE BIG FUTURE OF DATA
60 DEMOS

http://www.commit-nl.nl/the-national-event-the-big-future-of-data

Oct 2, 2014 - Amsterdam

http://www.data2semantics.org/

1. Tracking the use of data all the way



Data analysis and transformation are increasingly important activities in both scientific research (e.g. climatology) and other fields (e.g. open government data). Unfortunately it is hard to assess the trustworthiness and quality of the results without knowledge of what data the outcome was based on, and through what procedure the outcome was reached. This information about entities, activities and people involved in using data is called *data provenance*.

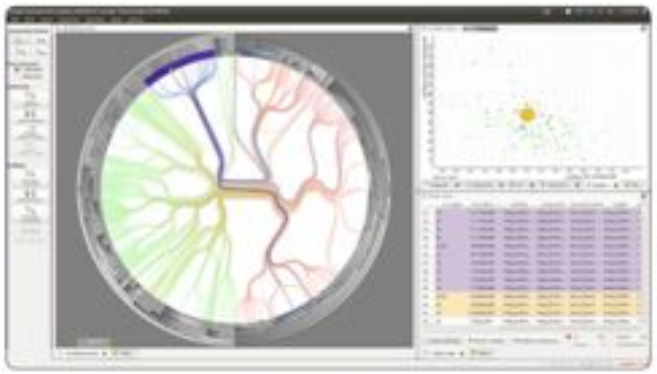
Our demo shows the integration of data provenance tracking and visualization in an existing, popular data science environment. The demo is an application of our work based on the PROV W3C standard, provenance visualization and tracking.

Our work allows for fine-grained tracing of conclusions in scientific papers to intermediate results, other publications, across applications and source data.

PROV-O-Viz™

https://github.com/Data2Semantics/provoviz

8. Finding new drugs by visualizing the effect of their ingredients



Finding new drugs to cure diseases is a hard task. This is because the chemicals in the drug interact in a very complex way with the cells and proteins in the human body. Visualizing this complex network of interactions is important to improve the development of new drugs.

Our demo shows how the interaction between the chemicals in a drug and the proteins in the body can be interactively explored in a rapid way. This rapid interaction makes it possible to get answers while you think, as opposed to waiting for answers, which breaks the train of thought.

Open PHACTS
Open Pharmacological Space

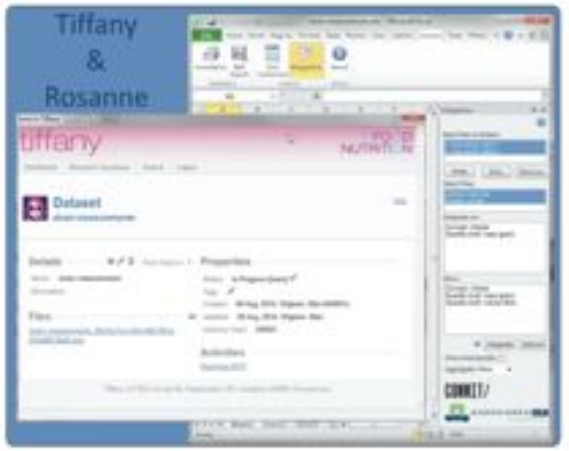
http://www.openphacts.org/

synerscope

http://www.synerscope.com/

33. How to get more out of food research

Effective food research requires that data and methods are shared. At present, careful data management is considered as a burden rather than a tool for good science. As a result, data can no longer be found or interpreted once time has passed. Furthermore, potential synergies slip through the net and costly duplications and mistakes occur. We have developed two tools for the easy management of food research data. The first tool, Tiffany, helps researchers to document their data in such a way that others can easily trace, understand and reproduce the research process. The researchers can use a second tool, Rosanne, to annotate their data in order to further improve search and reuse. Together, Tiffany and Rosanne increase the chances of successful valorisation of food research.



http://www.commit-nl.nl/sites/default/files/Demonstratortool%20eFood.pdf

Semantics in food research: **Tiffany and Rosanne**
https://www.youtube.com/watch?v=cQIdHTwPL1Y

32. Web-based tools for handling biomedical Big Data

Biomedical research is facing Big Data challenges. At present however, researchers don't have user-friendly IT tools to handle these data. To solve this problem, Science Gateways are developed. Science Gateways are built as easy-to-use, web-based and scalable tools that manage and integrate data, methods and infrastructure for scientific research.

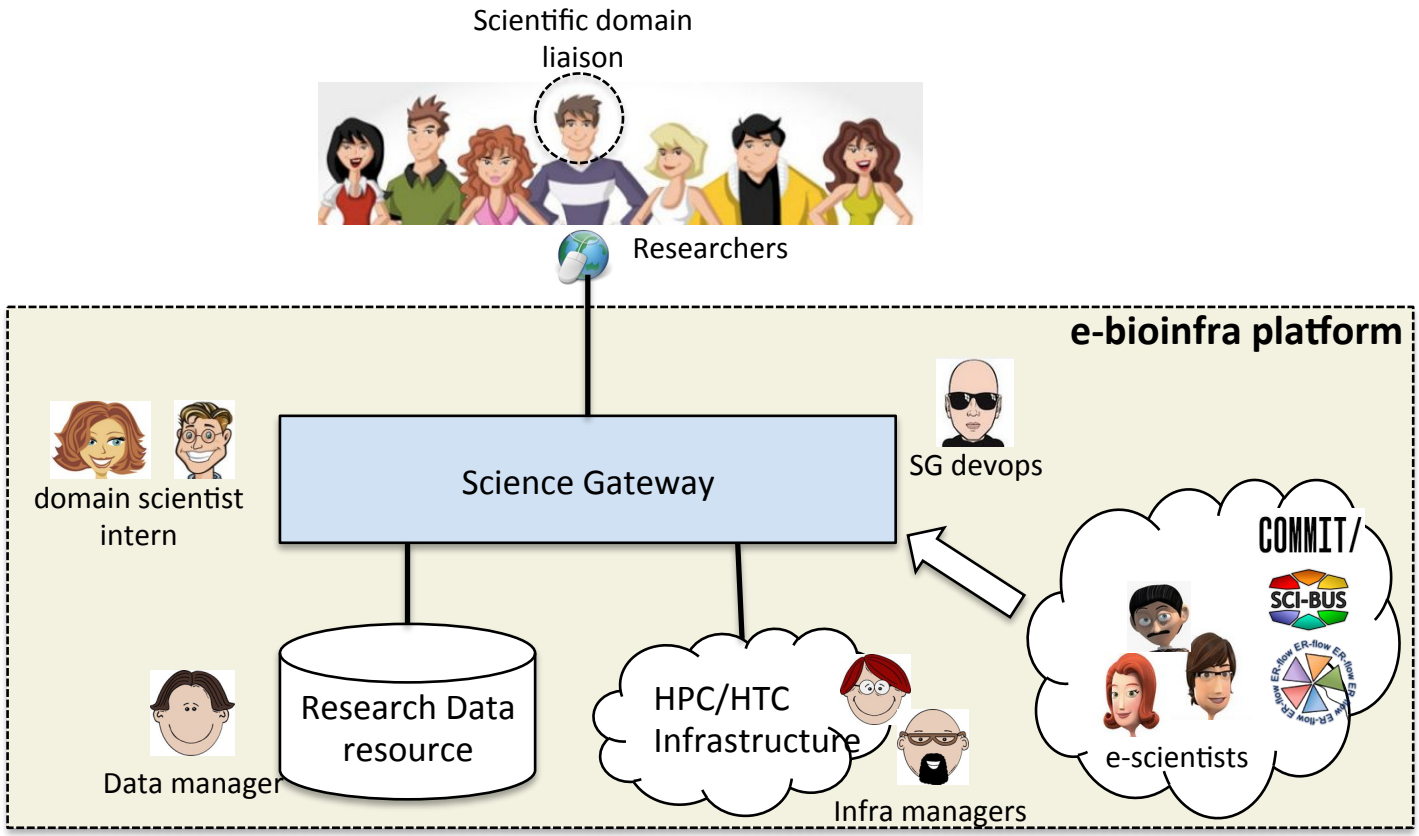
The AMC Science Gateways for biomedical research enable scientists to run large-scale data analysis easily and efficiently from web interfaces. From the gateway interface the scientist can start, monitor and inspect results of the analysis and collaborate with other researchers. The details of the underlying data and computing infrastructures are totally hidden from the researchers, so that they can focus on their research.

www.ebioscience.amc.nl/gateways

COMMIT/

THE “UGLY”:
SCIENCE GATEWAYS STORY FROM THE AMC

AMC SCIENCE GATEWAYS



EVOLUTION OF AMC SCIENCE GATEWAYS

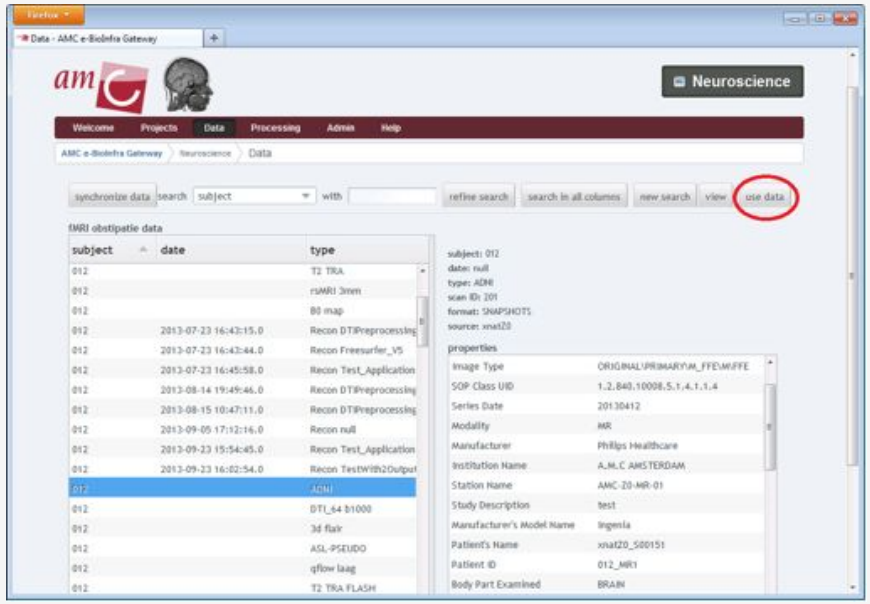
	1ST GENERATION	2ND GENERATION	3RD GENERATION	4TH GENERATION	5TH GENERATION
FUNCTIONS	Start	Upload/ Download files, Start	Upload/Download files, Start, See history, Monitor	Filter, Start, Resume, Abort, See provenance, Download result	Search, Start, Resume, Abort, See provenance, Download and visualize result, Tag and basket, send messages with attachments, notification center
APPLICATIONS	DTI Atlas	DTI Atlas	Freesurfer, DTI pre-processing, DTI Atlas, FSL BedpostX, BLAST, Genome Compare, T/B cell variation, Logistic Regression	Freesurfer, DTI pre-processing, FSL BedpostX, Tracula, Autodock Vina	Freesurfer, DTI pre-processing, FSL BedpostX, Tracula
UI	JSP	JSP	JSP Spring framework	Vaadin, JSP Liferay portlets	Bootstrap (responsive), AngularJS Play framework
INFORMATION SYSTEM	-	User, experiments MySQL	User, roles, experiments, data, history, applications MySQL	User, roles, experiments, data, projects, metadata, provenance, applications, resources MySQL	User, roles, experiments, data, workspace, metadata, provenance, applications, resources, tags, notifications, messages MongoDB
DATA INFRASTRUCTURE	Grid files	Local and grid files	FTP server Grid files	Metadata XNAT and WebDAV server Grid files	Metadata XNAT and WebDAV server Grid files
SCIENTIFIC WORKFLOW	Script	MOTEUR	MOTEUR DIANE	WS-PGRADE/gUSE DCI-Bridge	DIRAC Pumpkin
COMPUTING INFRASTRUCTURE	Grid	Grid	Grid	Grid, local cluster	Grid, local cluster and cloud

EXAMPLE: SCIENCE GATEWAYS FOR NEUROIMAGING DATA ANALYSIS

2011-2013

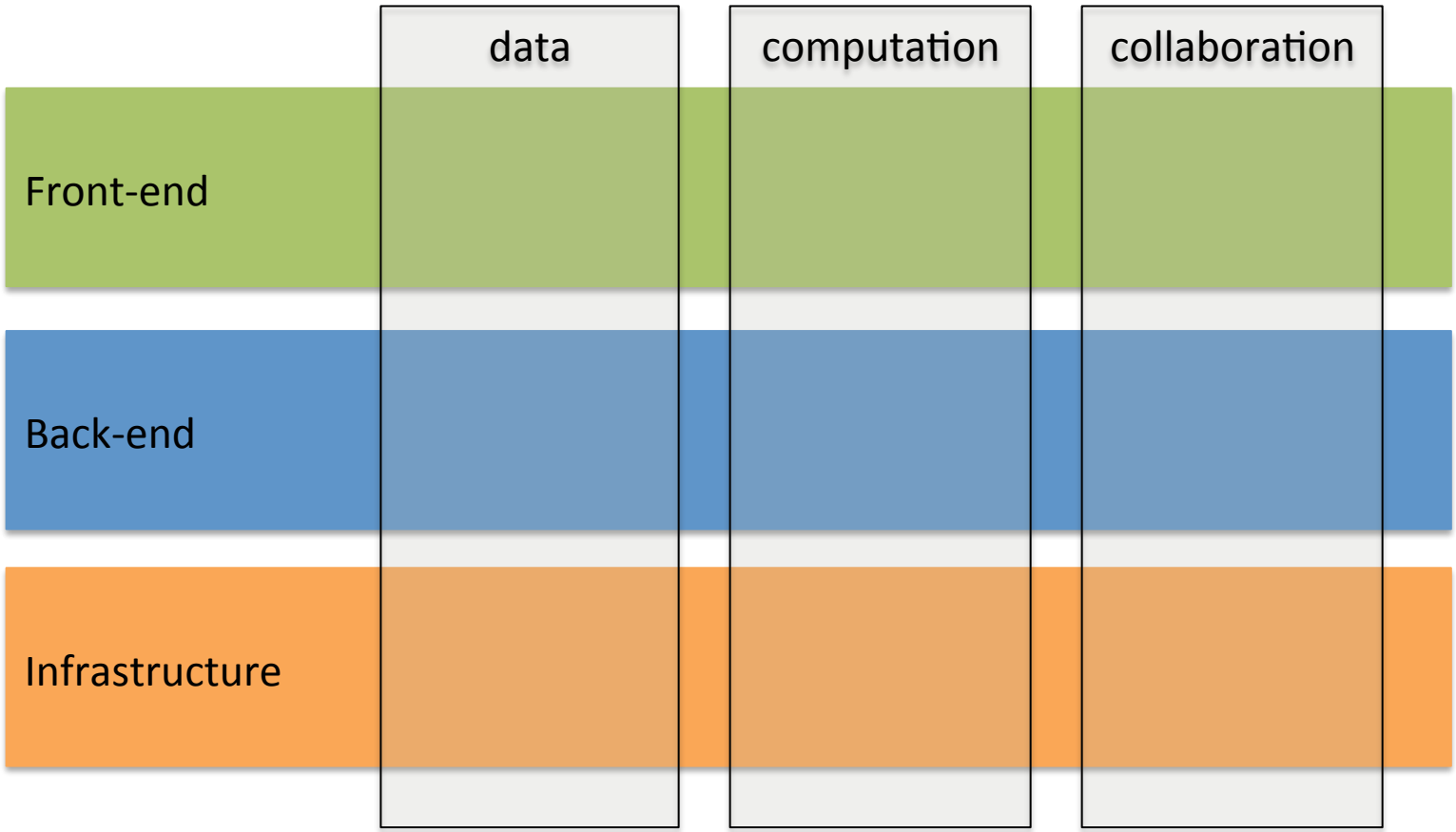
- 300 GB of input data, 1300 GB of output data
- 97,000 computation hours (~11 years)

2014-



- 27 users
- 1500+ experiments

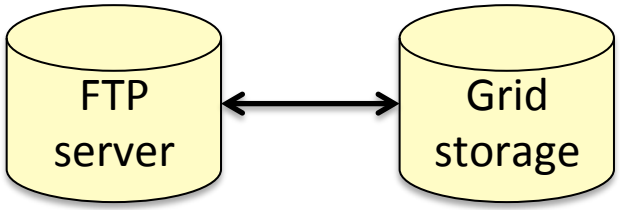
COMPONENTS OF A SCIENCE GATEWAY



1ST GENERATION: FILES ON THE GRID

2ND GENERATION: FILES THROUGH THE GATEWAY

3RD GENERATION: FILES HISTORY



Experiment Information

Overview

Experiment ID	744
Experiment Name	sm-first-predti
Submitted	2011-09-02 10:46:22.0
Experiment Type	predti
Owner	Jan de Vries
Result Location (might be partial)	Click to redirect
Workflow ID	workflow-6015ef75
Data in this Experiment	TOTAL: 1 1492.zip
Status Detail	

Status History

Refresh Status

Timestamp	Status
2011-09-05 11:52:07.0	finished
2011-09-02 11:25:14.0	running
2011-09-02 11:25:04.0	running
2011-09-02 11:24:48.0	running
2011-09-02 10:46:23.0	new

Data History

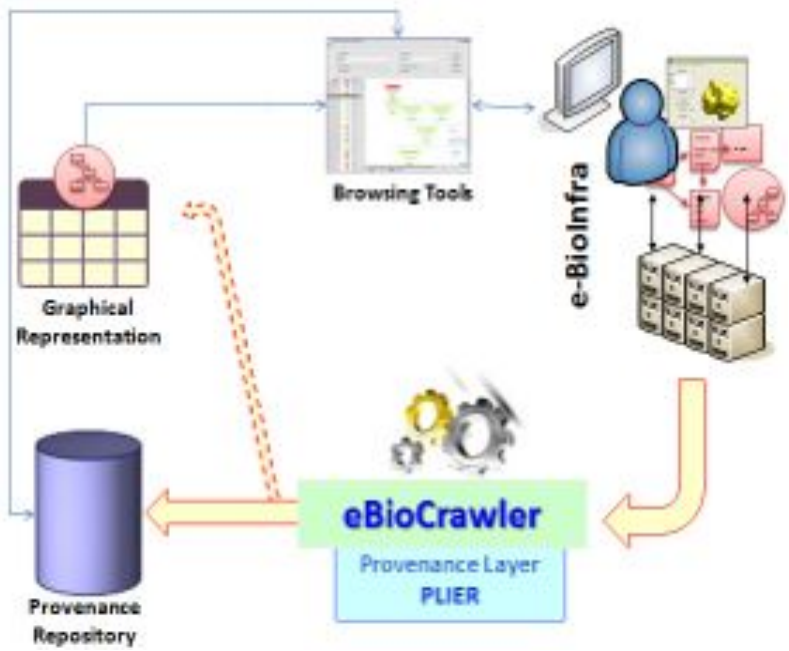
Jan de Vries's Data History

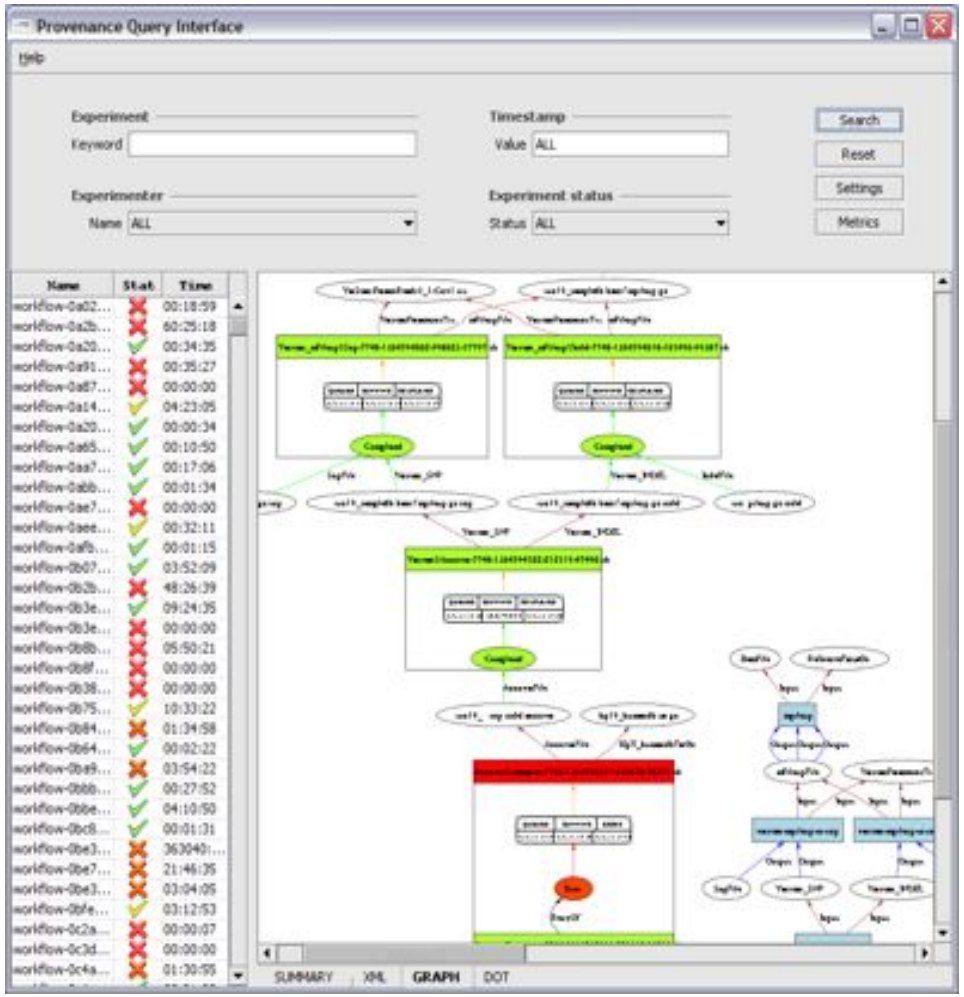
Note: Available only for the experiments started after 15/03/2011.

Number of entries in this list: 21

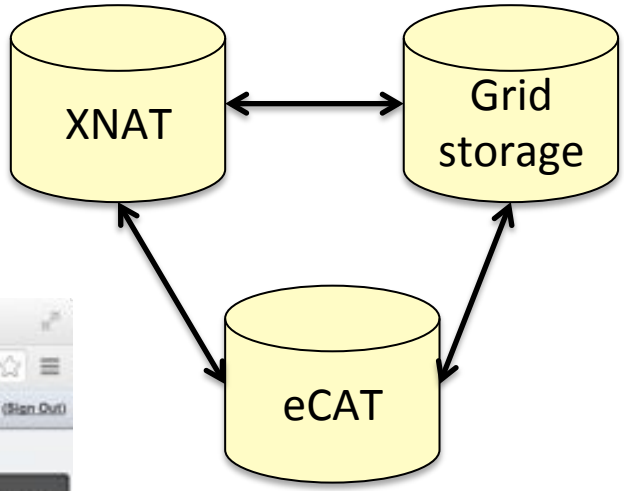
Name	Associated Experiment(s)
1492.zip	sm-first-predti(predti) test_predti(predti) test sql 2(predti) test3 null result(predti) test2 null result(predti) test null result(predti) 1492 and 1500(predti)
1500.zip	test_predti(predti) test sql 2(predti) test sql(predti) test3 null result(predti) test2 null result(predti) test null result(predti) 1492 and 1500(predti)
dti-sample-01.mat	hg(ctlatlas)
dti-sample-02.mat	hg(ctlatlas)
freesurfer-sample-firstfile/inseries-00001.tar.gz	mix(freesurfer)

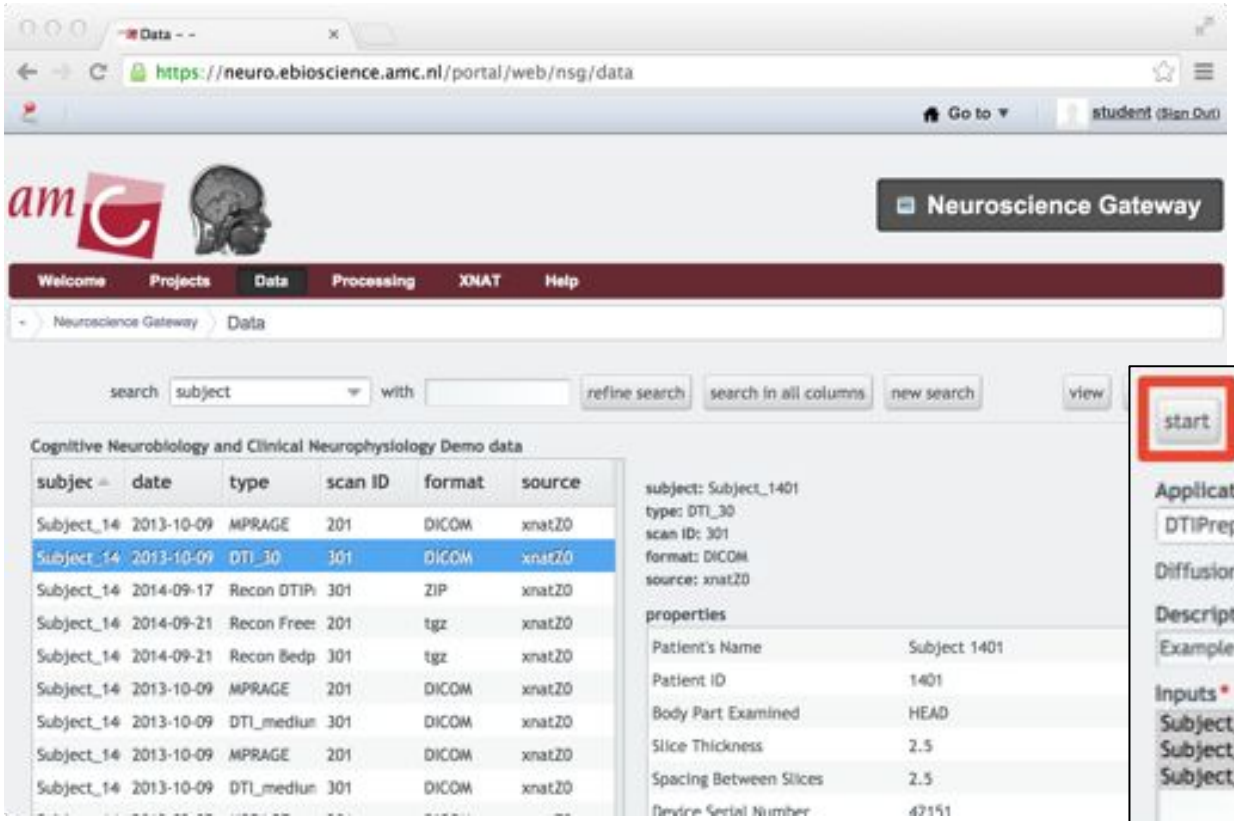
3RD GENERATION: PROVENANCE (OPM)



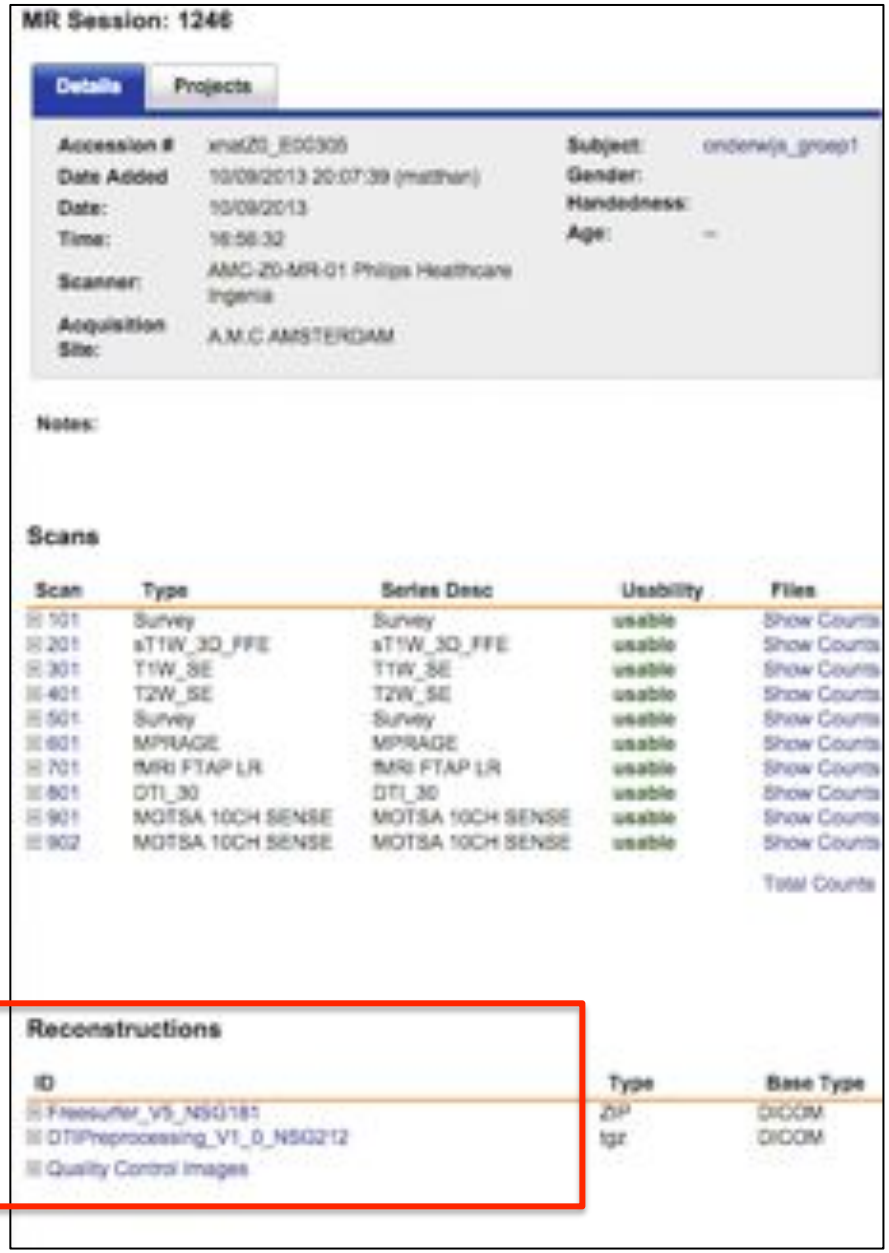


4TH GENERATION: DATA METADATA

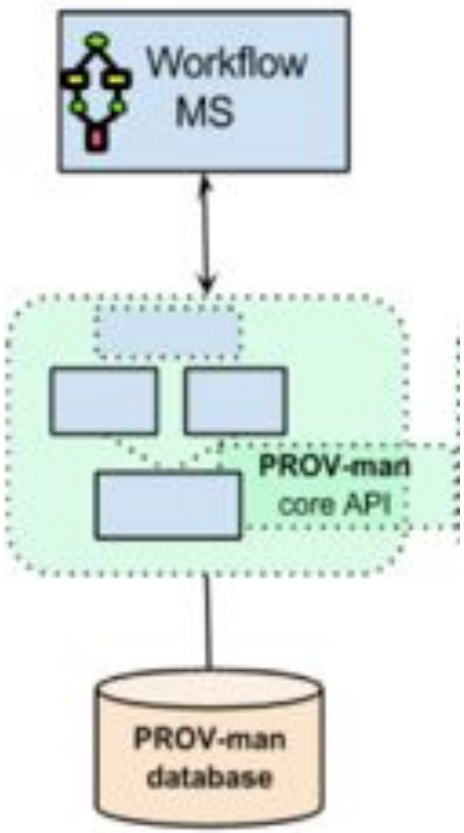


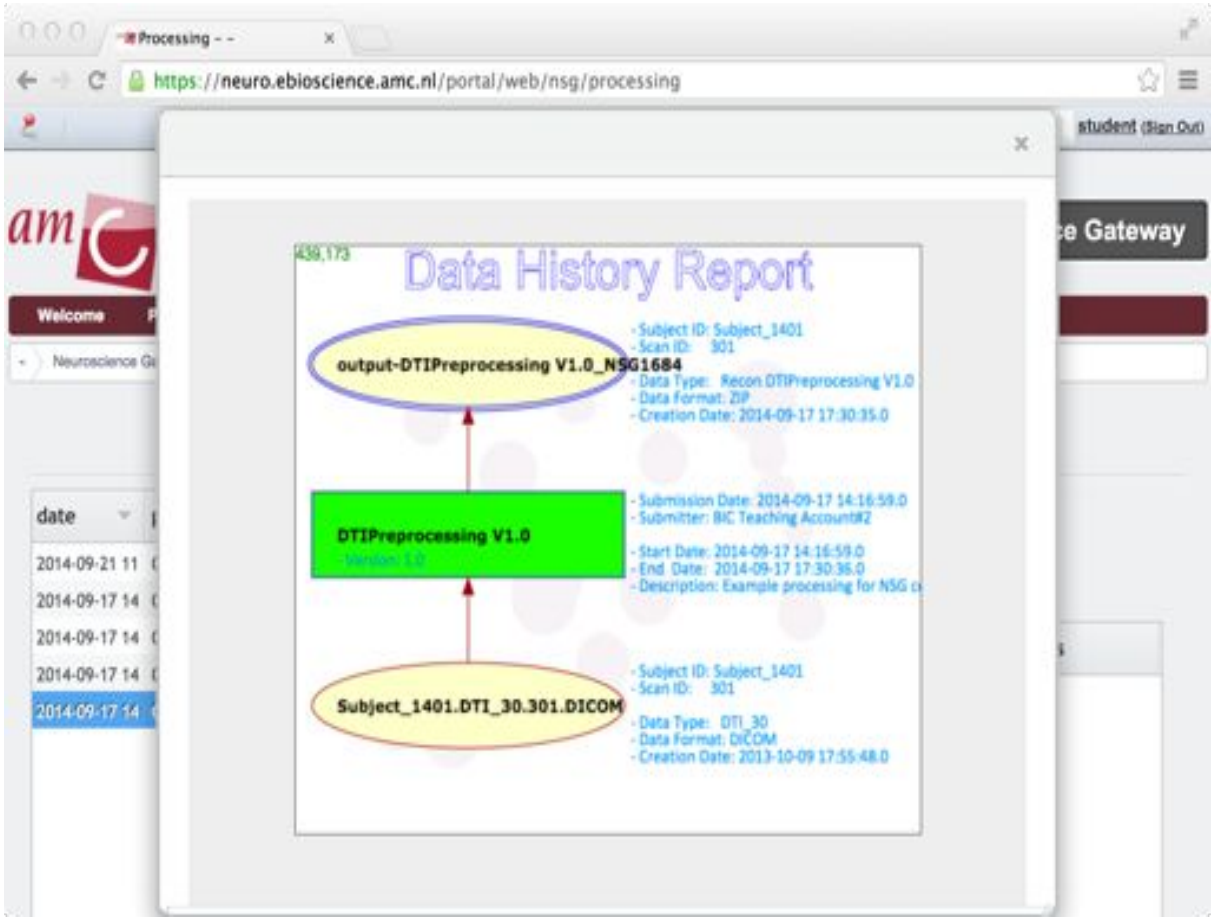


RESULTS ON XNAT

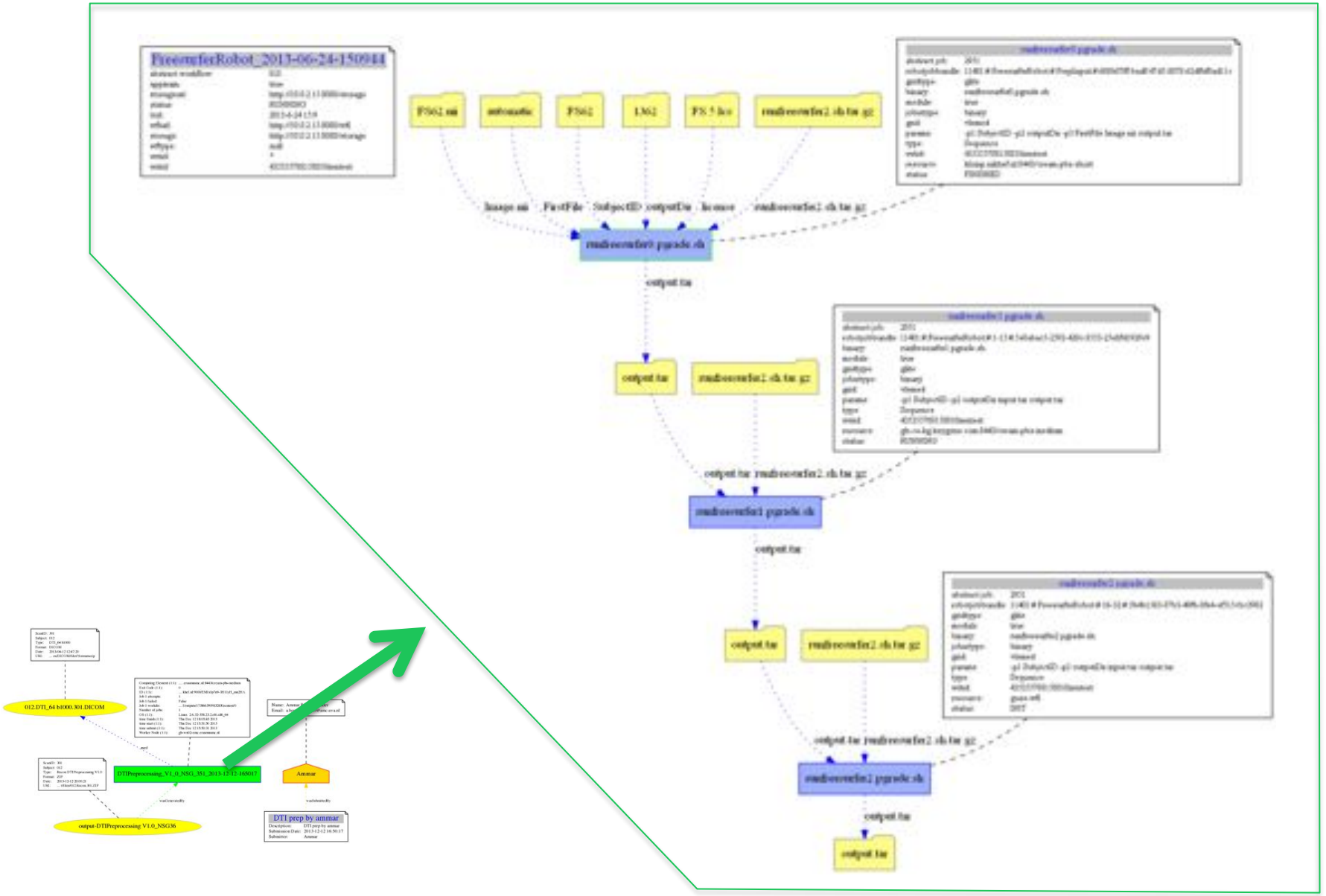


4TH GENERATION: PROVENANCE (PROV)

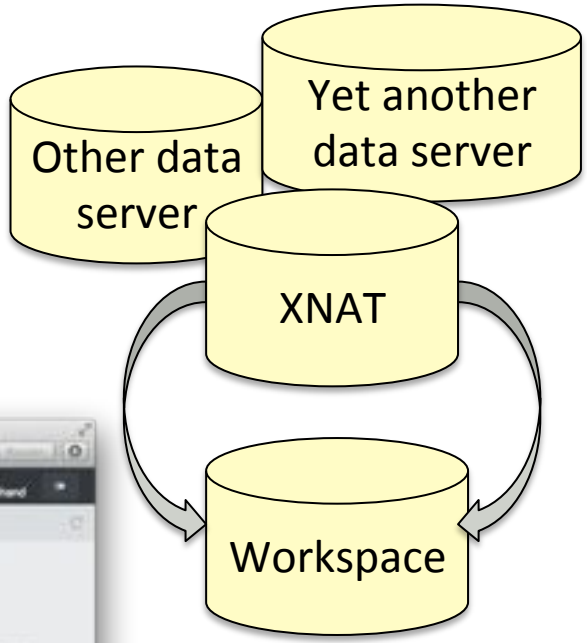


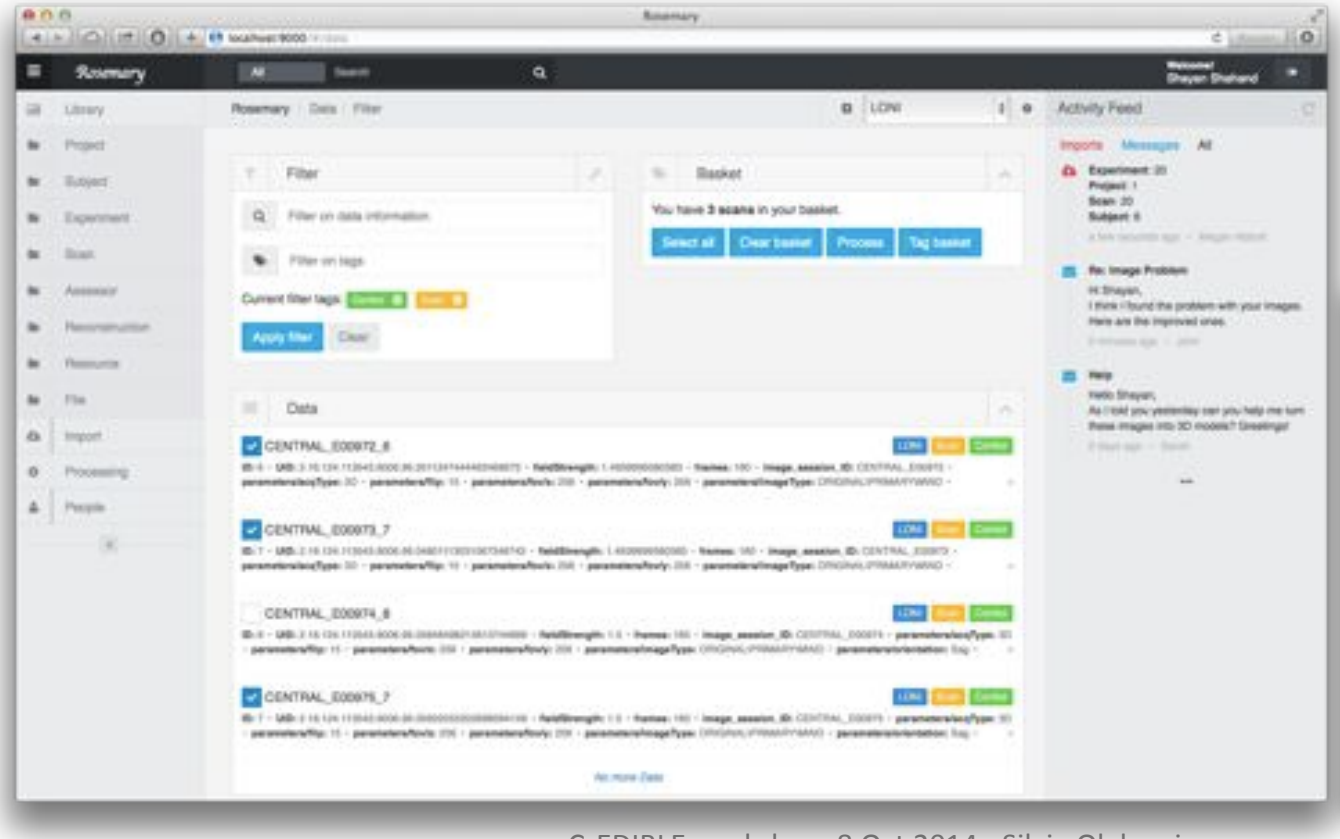


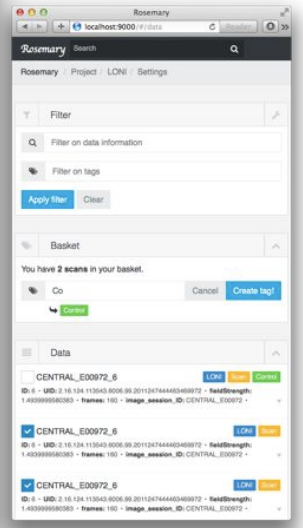
COMMIT/ 4TH GENERATION: PROVENANCE GRANULARITY?



5TH GENERATION: DATA METADATA







COMMIT/

THE BAD:

“THE DEVIL IS IN THE DETAILS”

CHALLENGES/DISCUSSIONS/MISSING

Collection of metadata
Location of data, metadata
How to annotate derived data
Persistent identifiers
Granularity of provenance
Retrieval of provenance
Archive vs. working data
Vocabularies
Security (authentication, authorization)
Tools
Evolution
Resistance /culture
Expertise
…

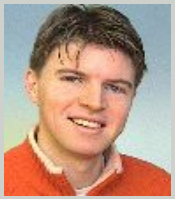
THANKS!















European Commission

SCI-BUS

ER-flow ER-flow ER-flow ER-flow ER-flow

e-infrastructure

COMMIT/

NWO Netherlands Organisation for Scientific Research