



Universiteit Utrecht

UBD/ITS/ResearchIT
Datamanagement



In support of large-scale research



Ton Smeele

Chapel Hill, June 10 iRODS User Group meeting

Agenda

1. Context: YOUth cohort study and the Yoda project
2. Yoda architecture
3. Research data intake process
4. Group management portal
5. Q&A

Cohort study YOUth

Individual development: Why some children thrive, and others don't.

- Large-scale longitudinal research on child development.
- This study focuses on the relationship between brain development, behavior and environmental factors.
- 8 Dutch institutes collaborate to cover the required disciplines:
 - Campus level: UMCU-Hersencentrum and UU-FSW.
 - Consortium level: Utrecht, Amsterdam, Leiden, Nijmegen, Groningen and Rotterdam.

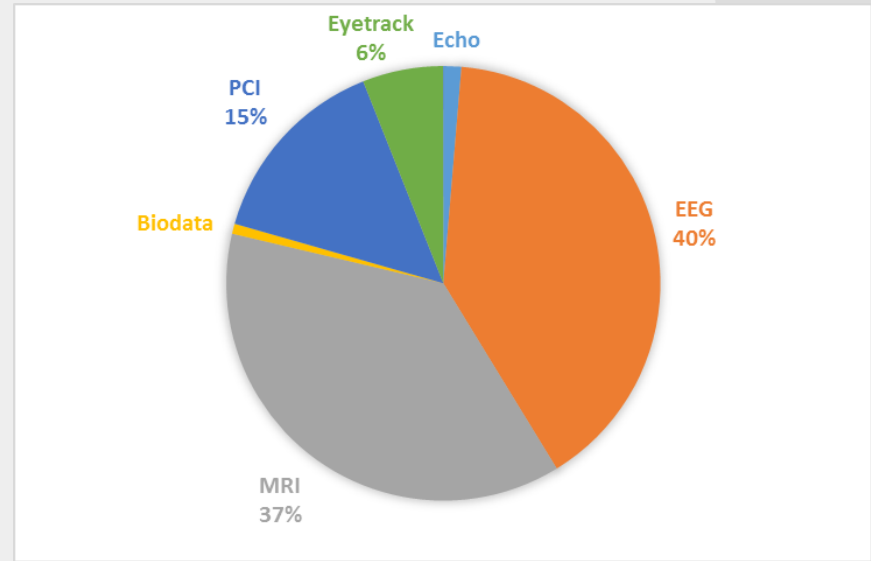
Human subject research

Data is collected on approximately 6000 subjects:

- Baby cohort: 3.000 children age between -7 months and 7 years old.
- Adolescents cohort: 3.000 children age between 8 and 16 years old.
- The study started May 2015 and continues for at least 8 years.

— Unique data

- Prenatal 3d echo
- EEG-measurements
- (f)MRI-scans
- Bio-data
- Parent Child Interaction (video)
- Eye-tracker / tasks
- Cognitive tests
- Questionnaires (demography, food, ...)
- Lab journals

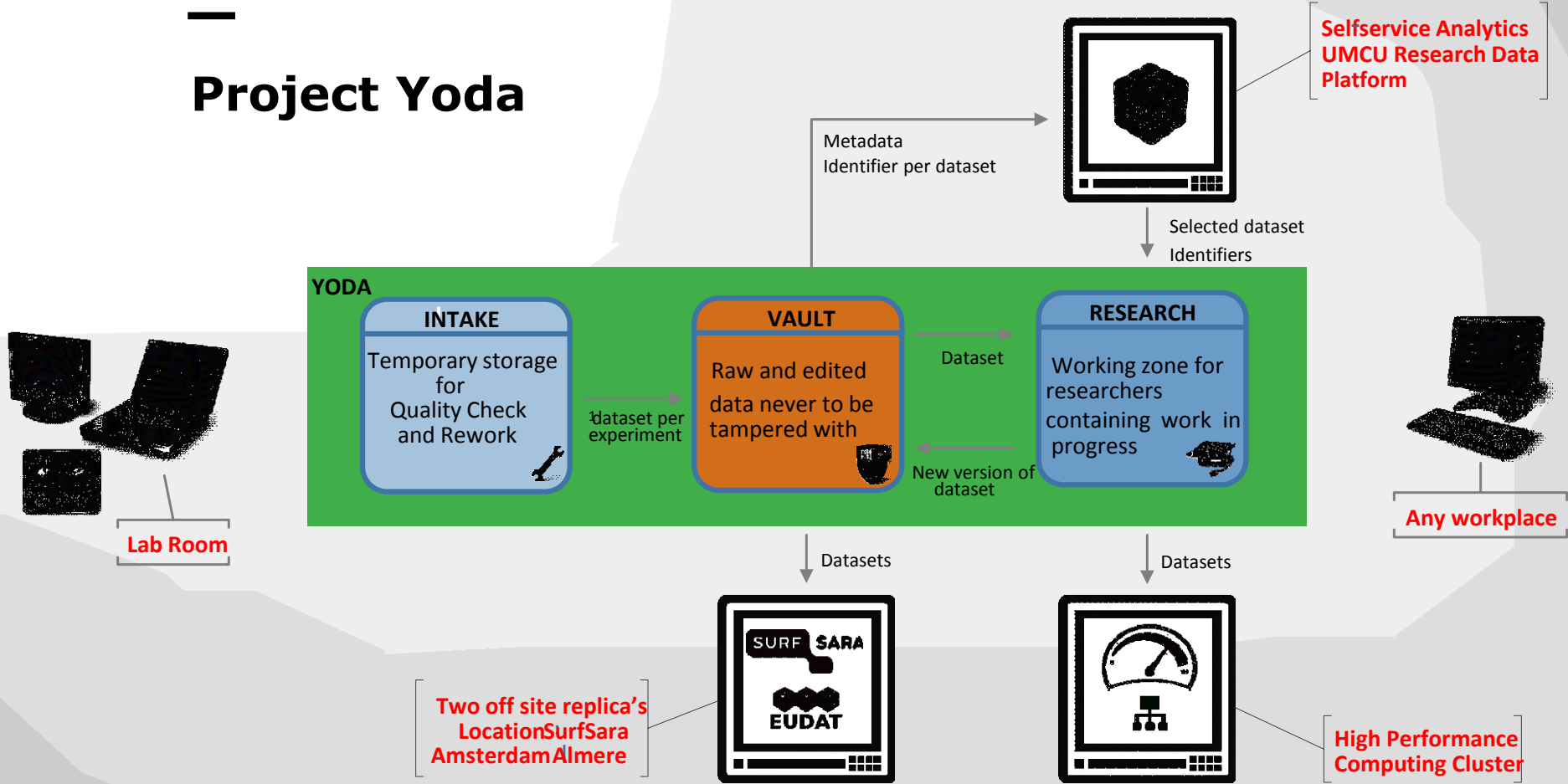


Total data size expected after 10 years: 1 PB

Our retention and dissemination requirements

- **Transparent:** Data must be FAIR (Findable, Accessible, Interoperable and Re-usable).
- **Shared:** UMCU, UU as well as third parties must be able to collaborate and share the data.
- **Secure:** Data must be protected against theft, loss and other damage.
- **Integrity:** Data must be protected against change. It is stored using transparent and tracable methods.
- **Durable:** During research and especially after research is completed.

Project Yoda

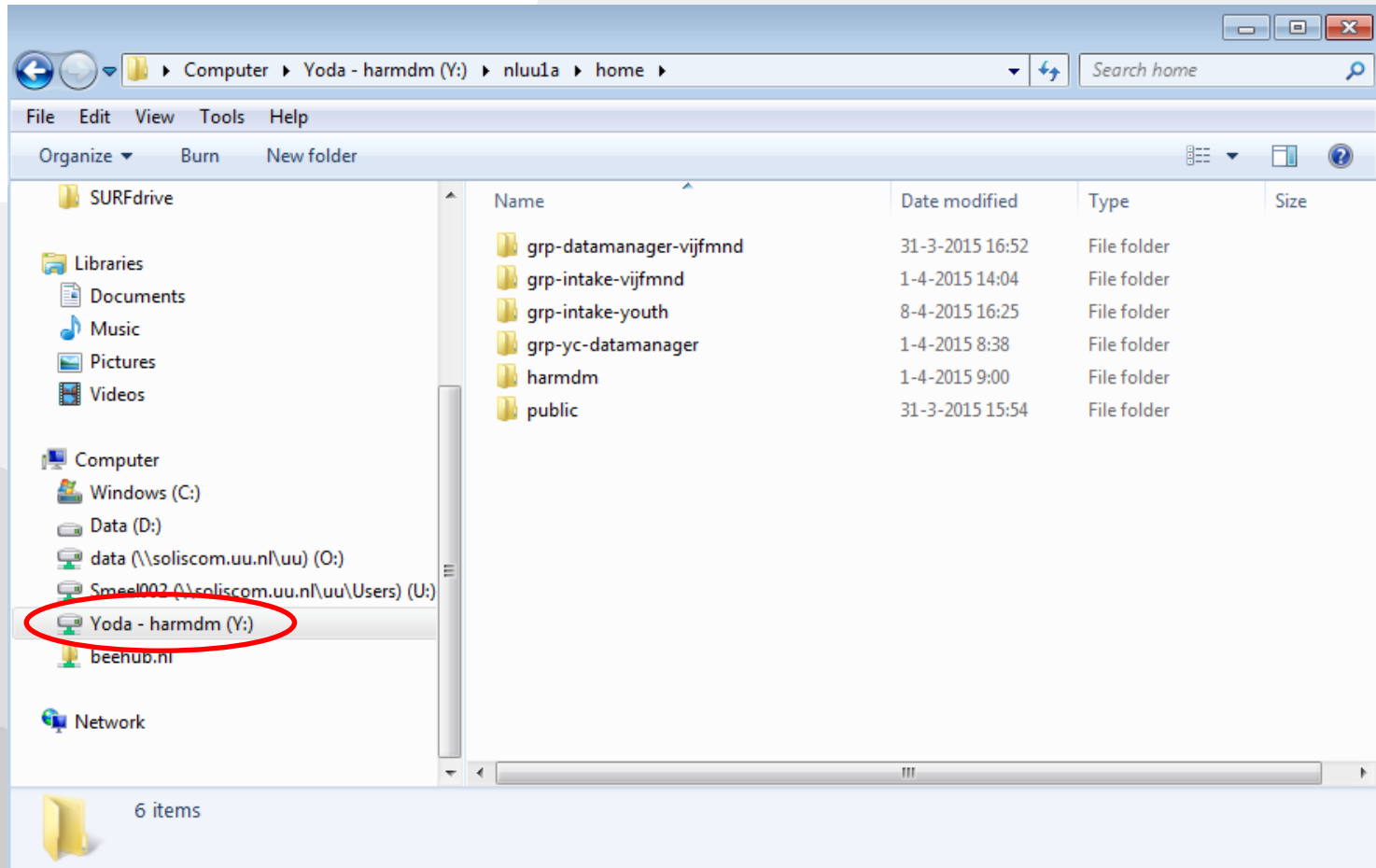


Yoda (YOUth data) is an iRODS based infrastructure to safely store, manage and provision research data sets

Intake process

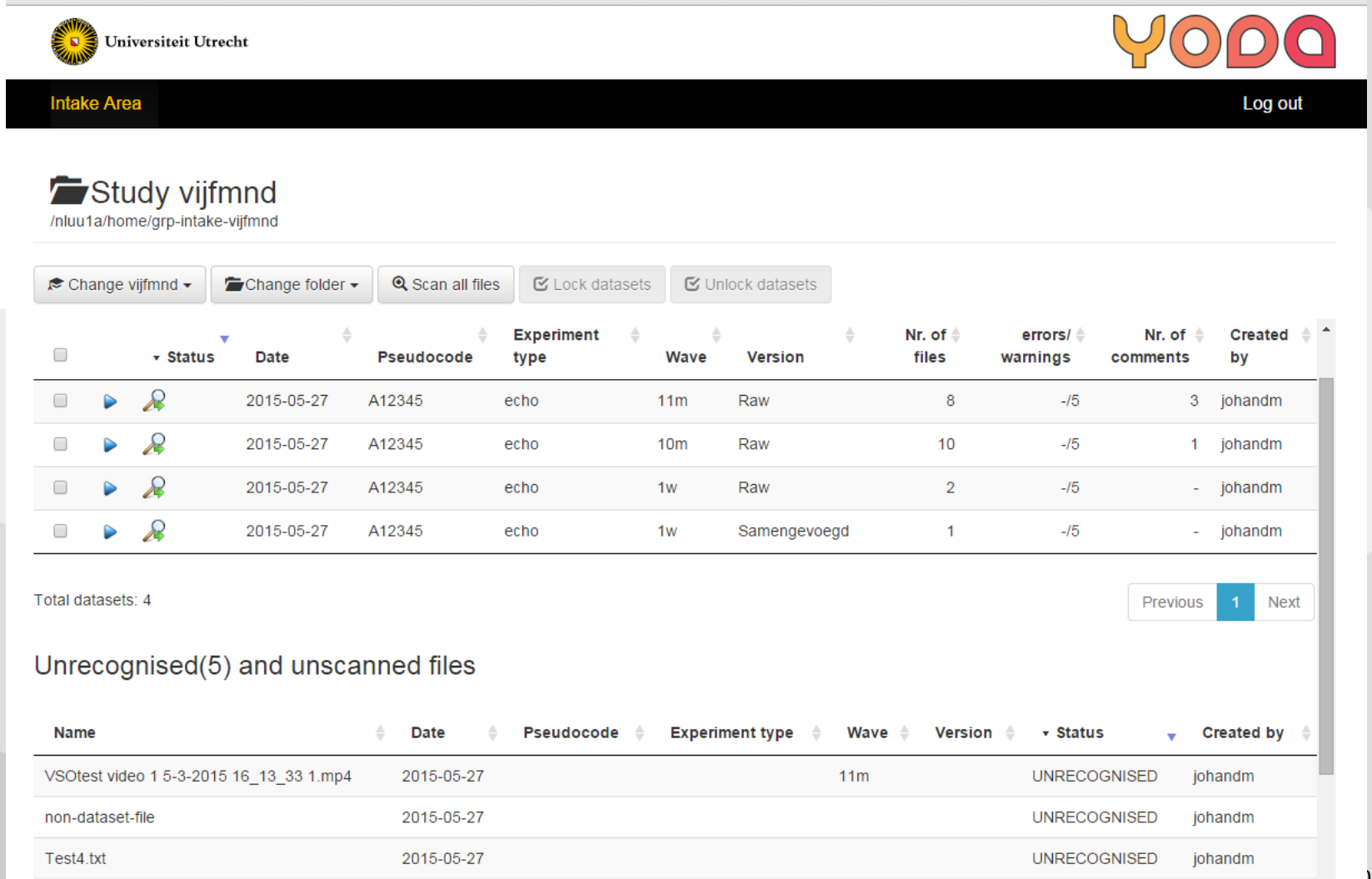
- **Intuitive user interface** for researcher and staff:
 - Web portal to support workflow and management of data
 - Familiar network disk concept to move data in/out iRODS
- Deploy iRODS rules to **unburden researcher and data manager:**
 - automated discovery of identifying attributes in names of subdirectories and files -> enrich data with metadata
 - Quality controls on incoming datasets and individual data files (completeness, formats)

Users drag-and-drop data via the Yoda network drive



currently webdavis -> irods-webdav

Results of data scan are shown in a web portal



The screenshot displays the YODA web portal interface. At the top, there is a navigation bar with the 'Intake Area' label and a 'Log out' button. The main content area shows the details for a study named 'Study vijfmd' located at '/nluu1a/home/grp-intake-vijfmd'. Below the study name, there are several action buttons: 'Change vijfmd', 'Change folder', 'Scan all files', 'Lock datasets', and 'Unlock datasets'. A table lists the datasets, with columns for checkboxes, status, date, pseudocode, experiment type, wave, version, number of files, errors/warnings, number of comments, and created by. The table contains four rows of data. Below the table, it indicates 'Total datasets: 4' and provides navigation buttons for 'Previous', '1', and 'Next'. A section titled 'Unrecognised(5) and unscanned files' contains a table with three rows of unrecognised files, including 'VSOtest video 1 5-3-2015 16_13_33 1.mp4', 'non-dataset-file', and 'Test4.txt'.

Study vijfmd
/nluu1a/home/grp-intake-vijfmd

Change vijfmd Change folder Scan all files Lock datasets Unlock datasets


<input type="checkbox"/>	Status	Date	Pseudocode	Experiment type	Wave	Version	Nr. of files	errors/warnings	Nr. of comments	Created by
<input type="checkbox"/>		2015-05-27	A12345	echo	11m	Raw	8	-/5	3	johandm
<input type="checkbox"/>		2015-05-27	A12345	echo	10m	Raw	10	-/5	1	johandm
<input type="checkbox"/>		2015-05-27	A12345	echo	1w	Raw	2	-/5	-	johandm
<input type="checkbox"/>		2015-05-27	A12345	echo	1w	Samengevoegd	1	-/5	-	johandm

Total datasets: 4

Previous 1 Next

Unrecognised(5) and unscanned files

Name	Date	Pseudocode	Experiment type	Wave	Version	Status	Created by
VSOtest video 1 5-3-2015 16_13_33 1.mp4	2015-05-27			11m		UNRECOGNISED	johandm
non-dataset-file	2015-05-27					UNRECOGNISED	johandm
Test4.txt	2015-05-27					UNRECOGNISED	johandm

 Study vijfmd

/nluu1a/home/grp-intake-vijfmd

Change vijfmd ▾

Change folder ▾

Scan all files

Lock datasets

Unlock datasets




<input type="checkbox"/>			2015-05-27	A12345	echo	11m	Raw	8	-/5	3 johandm
--------------------------	---	---	------------	--------	------	-----	-----	---	-----	-----------

Dataset

Path	/nluu1a/home/grp-intake-vijfmd/Gonneke/11m/echo/Y2015M03/D10
Scanned by	johandm (2015-05-29 09:06:39)

Warnings

Description

-  Expected at least 20 files of type '10000000.raw', found 2
-  Expected at least 20 files of type '10000000.index.tiff', found 2
-  Expected at least 5 files of type '10000000.vol', found 1
-  Expected at least 5 files of type '10000000.dcm', found 1
-  Expected at least 20 files of type '10000000.usr', found 2

Comments

User	Date	Comment
------	------	---------

Group management

- **Self service portal** to manage access to research group data:
 - Can manage membership of their own iRODS group
 - Real-time service to add a new iRODS user account (iRODS login possible only if a corresponding institute account exists)
- **Flexible delegation of selected privileges** to Yoda group managers:
 - Ability to add new groups
 - Ability to add a field of research category

Self service group membership management

[Yoda Portal](#)[Intake Area](#)[Group Manager](#)[Log out johandm](#)

Group Manager

Yoda groups

▶ System

▼ YOUTH

Intake

grp-datamanager-vijfmnd



grp-datamanager-youth



grp-intake-vijfmnd



grp-intake-youth



grp-vaultlezen-youth



▶ YOUTH Research Areas

Add group

Group properties

Category

YOUTH

Subcategory

Intake

Group name

grp- intake-vijfmnd

Group description


intake area for the short 5 months study done in 2015


Update

Group members


Search users



 a.p.m.smeete@uu.nl

 johandm (you)

 gonneke@uu.nl

 p.kemmener@umcu.nl

[Click here to add a new user to this group](#)

Change role

Remove

— Summary: Business benefits

- Datasets stored in a structured, integer way:
 - 1 dataset per version of an experiment, individual datasets each have a persistent identifier
 - (new versions of) Ingested datasets are enriched and quality-checked
 - Off-site replicas per dataset
- User-friendly access across institutions.
- Self-service: researcher is unburdened yet remains in control.
- Flexible data selection and provisioning using RDP/Yoda.



Universiteit Utrecht



In support of large-scale research

—

Thank you

Questions?

or contact me at a.p.m.smeele@uu.nl