PaNdataODI

Lund

13-14 March 2013

Agenda - **Wednesday 13, 14:00-~17:30**

Plan for Meeting (Juan) 10 mins

Service Verification and Integration (Alastair) 10 mins

WP5 as a way to demonstrate the results of other WPs (Brian and Thorsten) 10 mins

Presentation of plans for each workpackage, ~30 mins each, answering the following questions:

- What will be delivered?
- When will it be delivered?
- What will it be used for? (Use Case)
- Who is involved?

WP3, Heinz Joseph (Umbrella & Affiliation DB)

WP4, Milan (and Tom)

WP5 Thorsten

WP6, Erica and Brian

WP7, Jean Francois

WP8, Bill

Agenda – Thursday 14, 9:00-12:30

From Yesterday- WP8, Bill

Session 1, Integration, (9:20)

- Integrating 3+4 (Tom and Antony)
- HJ's questions about Umbrella

Break ~10:00 and Demo of affiliation DB

- Integrating 6 into 3+4 (Erica to lead a discussion)
- Integrating 7 into 3+4 (Jean Francois to lead a discussion)
- Integrating 8 into 3+4 (Bill to lead a discussion)

How will we use WP5 as the demonstration of integration, Thorsten(11:00)

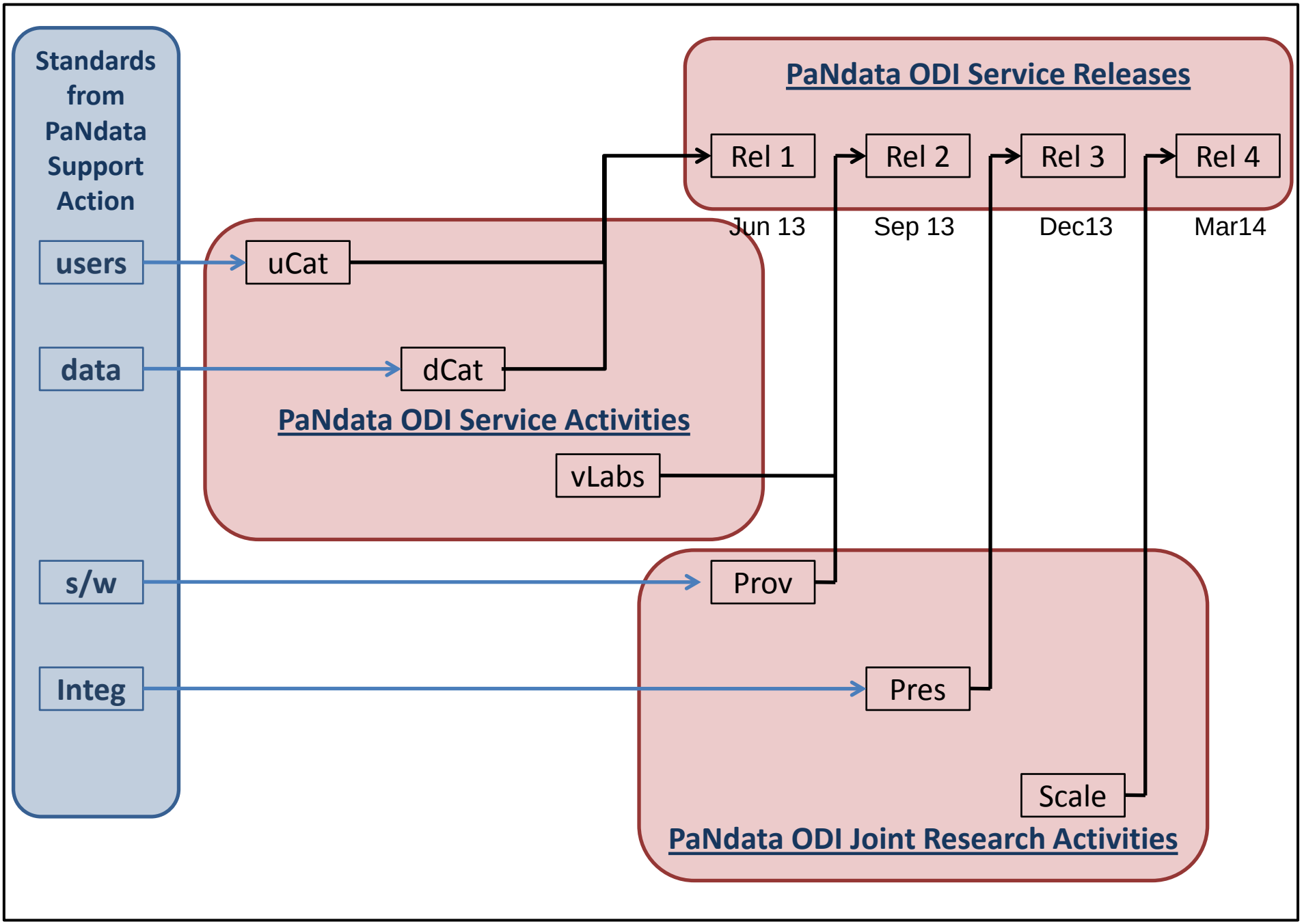
Recap of actions (~11:20)

Session 2, Management and Dissemination, 11:30-12:30

- WP2 (Frank)
- WP1 (Juan and Denise)

Aims of the meeting

- Consolidate plan for software development and delivery for the rest of the project
- Refine and coordinate plans for each workpackage
- Understand how we will integrate the outputs of the WPs
- Clarify the role of VLabs as demonstrators
  - Can refine DoW if necessary


Standards from PaNdata Support Action
users
data
s/w
Integ
PaNdata ODI Service Releases
Rel 1
Rel 2
Rel 3
Rel 4
Jun 13
Sep 13
Dec13
Mar14
uCat
dCat
PaNdata ODI Service Activities
vLabs
Prov
Pres
Scale
PaNdata ODI Joint Research Activities

Service Verifications

- Alistair....

Case Studies as a demonstration of Integration

WP6 has considered a number of scenarios

- SANS2d (ISIS)
  - Co-ordinating and automating “near to experiment” processes
- TwinMic x-ray spectromicroscopy beam line (ELETTRA)
  - Co-ordinating multiple experimental runs
- Tomography beam lines (DLS)
  - Managing high volumes of data
- Express services (ISIS)
  - Automating “standardised” experiments.
- Publication linking (ISIS)
  - Tracing research outputs for impact analysis

- Need to choose those which will be developed and demonstrate them in WP5
  - Who does what?

WP1 – Management and administration update

- Effort Reporting
  - for P2 to date (1 October to 28 February 2013).
  - To make it easy, just email me your person month effort per work package!

  Any large discrepancies in effort – give a short note of explanation.
- P1 Form C submission.
  - We have received some queries back from the Project Officer and a response with our corrections is almost complete.
  - Thank you all for being so accurate on your cost statement submissions which has meant fewer queries ☺

Other matters

- PaNdata Europe
  - P2 money has now been received at STFC and will be distributed shortly. I will advise all when this has been completed.
- Subcontracting
  - there will need to be some paperwork to allow this
- Next F2F meeting
  - please agree date and venue ☺

Other things

- Extending the Collaboration
  - ICAT, Umbrella, MoUs
  - Research Data Alliance WG?
- Horizon 2020
  - first call January 2014 – a topic for next meeting
- Date and Place of next meeting
  - June or September?
  - Where
- Second Review
  - November
  - Should we also prepare a paper on PaNdataODI?

Actions

- Submitted deliverables to mailing list
- Deliverables 7.x, 6.x, 5.2,
- Action for P1 review,
- Financial reporting process – distribute timetable for P2 reports
- Prepare paper(s) on project. (Juan/Frank)
- RDA – PaNData/FedId/MD catalogue (feedback to HJ before Wednesday)
- Next meeting doodle

Objectives

Objective – Users (WP3)

To deploy, operate and evaluate a system for pan-European **user identification across the participating facilities** and implement common processes for the joint maintenance of that system.

Objective – Data (WP4)

To deploy, operate and evaluate a generic catalogue of scientific data across the participating facilities and promote its integration with other catalogues beyond the project.

Objective – Demonstration (WP5)

To deploy and operate the services and technology developed in the project in **virtual laboratories for three specific techniques** providing a set of integrated end-to-end data services.

Objective – Provenance (WP6)

To research and develop a conceptual framework, defined as a **metadata model, which can record the analysis process**, and to provide a software infrastructure which implements that model to **record analysis steps** hence enabling the **tracing of the derivation of analysed data outputs**.

Objective – Preservation (WP7)

To add to the PaNdata infrastructure extra capabilities oriented towards **long-term preservation** and to integrate these within selected virtual laboratories of the project to demonstrate benefits. These capabilities should, as for the developments in the provenance JRA, be integrated into the normal scientific lifecycle as far as possible. The conceptual foundations will be the **OAIS** standard and the **NeXus** file format.

Objective – Scalability (WP8)

To develop **a scalable data processing** framework, combining **parallel filesystems** with a parallelized standard data formats (pNexus pHDF5) to permit applications to make most efficient use of dedicated multi-core environments and to permit simultaneous ingest of data from various sources, while maintaining the possibility for real-time data processing.

Engagement (WP2)

Engagement with **other initiatives** and dissemination of project results, in particular to other research infrastructures.

WP1 Management

Objectives

To establish an effective and efficient collaboration between the partners...

To ensure that the project achieves its objectives ...

To report to the Commission as required...

Task 1.1: Set up mechanisms to run the project through the rest of its duration (M1–M2).

Task 1.2: Monitor progress of project activities and put in place appropriate corrective actions

Task 1.3: Organise general meetings of the project (kick-off and bi-annually thereafter).

Task 1.4: Report to EC on the technical and financial progress of the project (annually and at the end of the project).

D1.1: Project management structures, reporting, risk and quality ... procedures (M3)

D1.2: First annual management report (M12)

D1.3: Second annual management report (M24)

D1.4: Final management report (M30)

WP2 Dissemination

Objectives

Engagement with other initiatives and dissemination of project results, in particular to other research infrastructures.

Task 2.1. Establish an external web site as an extension to the existing website for the PaNdata collaboration (www.pandata.eu).

Task 2.2. Establish an interest group for project news items via community channels, informing them of project progress.

Task 2.3. Presentations to relevant international audiences at conferences, symposia, other project meetings etc.

Task 2.4. Provision of the open source software and appropriate documentation to potential partner bodies.

Task 2.5. Workshops to present the integrated systems to user and facility communities.

D2.1 : Project Website (M1) – November 2011

D2.2 : Dissemination plan (M3)

D2.3 : First Open Workshop (M15) – January 2013

D2.4 : Open Source software distribution procedure (M21)

D2.5 : Second Open Workshop (M27) - January 2014

WP3 User Catalogue and AAA **Service**

Objectives

To deploy, operate and evaluate a protocol for pan-European user identification across the participating facilities and implement common processes for the joint maintenance of that system.

Task1: Consultation on existing software components → recommendations for technologies to be implemented.

Task 2: Set up team includes representatives from the user office and/or IT staff of the partners.

Task 3: Specify an architecture which ... builds on the IRUVX "umbrella" concept.

Task 4: Implement ... the necessary local modifications (including trust management).

Task 5: Implement a standard affiliation database which is accessible for update and use by the participating facilities ...

- Introduce a central affiliation database according to the PaNdata de-facto standard.
- Provide an interface of the local WUO systems to this standard.
- Organise and support the migration of the local WUOs to this new affiliation database.

Task 6: Deploy the user management system at all participating facilities.

- A major factor will be the integration with the facility's bespoke user administration systems.
- The deployment will include setting up of an administration authority for the system.

Task 7: Evaluate the system within a subset of the collaborating facilities.

Task 8: Operate and report on the AAA trust system for the remainder of the project.

Task 9: Maintain communication with other user authentication systems (through Workpackage 2) ...

D3.1 : Specification of AAA infrastructure (M6) Apr 2012

D3.2 : Pilot deployment of initial AAA service infrastructure (M12) Nov 2012

D3.3 : Production deployment of AAA service infrastructure (M18) Apr 2013

D3.4 : Evaluation of initial AAA service infrastructure (M24) Nov 2014

WP4 Data Catalogue

Objectives To deploy, operate and evaluate a generic catalogue of scientific data across the participating facilities and promote its integration with other catalogues beyond the project

- develop generic software infrastructure to support interoperation of facility data catalogues
- deploy this software to establish a federated catalogue of data across the partners,
- provide data services based upon this generic framework which will enable users to deposit, search, visualise, and analyse data across the partners' data repositories,
- evaluate this service from the perspective of facility users,
- manage jointly the evolution of this software and the services based upon it,
- promote the take up of this technology and the services based upon it beyond the project

Task 4.1. Survey the features of existing implementations of metadata catalogues ...

Task 4.2. ... deploy the chosen metadata catalogue in the legacy context of the facilities.

Task 4.3. Provide remote API access to the individual catalogues and integrate to provide a single search capability across the collaborating facilities.

Task 4.4. Evaluate the performance of searching the metadata catalogue and retrieving data.

D4.1. Requirements analysis for common data catalogue (M9)

D4.2. Populated metadata catalogue with data from the virtual laboratories (M15)

D4.3. Deployment of cross-facility metadata searching (M21)

D4.4. Benchmark of performance of the metadata catalogue (M27)

WP5 Virtual Laboratories (Service)

Objectives

To deploy a set of integrated end-to-end user and data services supporting three specific techniques:

- Structural 'joint refinement' against X-ray & neutron powder diffraction data
- Simultaneous analysis of SAXS and SANS data for large scale structures
- Access to tomography data exemplified through paleontological samples

D5.1: Specific requirements for the virtual laboratories (M6) Apr 2012

D5.2: Deployment of Specification of the three virtual laboratories (incorporating any specific requirements software to support them) (M18) Apr 2013

D5.3: Report on the implementation of the three virtual laboratories (M30) Apr 2014

WP6 Provenance (JRA)

Objectives

To develop a conceptual framework, which can record and recall the data continuum, and especially the analysis process, and to provide a software infrastructure which implements that model to record analysis steps hence enabling the tracing of the derivation of analysed data outputs

Task 1: Requirements for Provenance

Task 2: Modelling the data continuum

Task 3: Ontologies for specific instruments/techniques

Task 4: Tool Support for the Data Continuum

Task 5: Tracing the Data Continuum

Task 6: Evaluation

D6.1: Model of the data continuum in Photon and Neutron Facilities (M12) Nov 2012

D6.2: Common ontology definition and definition of tools to support the use of provenance for Photon and Neutron Facilities (M18) Apr 2012

D6.3: Tools for building research objects in Photon and Neutron Facilities (M24) Nov 2013

D6.5: Evaluation report on provenance management in Photon and Neutron Facilities (M30) Apr 2014

WP7 Preservation (JRA)

Objectives

To incorporate models and tools oriented towards long-term data preservation into the PaNdata infrastructure, focussing on several aspects considered of benefit: an OAIS-based infrastructure; persistent identifiers; and certification of authenticity and integrity

Task 1. Baseline and OAIS application

Task 2. Persistent identifiers (for datasets)

Task 3. Representation information and archiving

RI for datasets, and AIPs (Archival Information Packages)

This will include software as a kind of representation information, and the need to preserve the software itself.

Task 4. Integrity of datasets

Mechanisms for maintaining and checking integrity of datasets. (for individual datasets (as preservation actions are performed) and for data holdings as a whole.

Task 5. Evaluation and reporting

D7.1 Implementation of persistent identifiers for PaNdata datasets (M15) Jan 2013

D7.2 Mechanisms and tools for representation information and archiving (M21) July 2013

D7.3 Mechanisms and tools for integrity of datasets(M27) Jan 2014

D7.4 Report on evaluation of preservation mechanisms (M30) Apr 2014

WP8 Scalability (JRA)

Objectives

To develop a scalable data processing framework combining parallel filesystems with a parallelized standard data format (pNexus pHDF5) to permit applications to make most efficient use of dedicated multi-core environments and to permit simultaneous ingest of data from various sources, while maintaining the possibility for real-time data processing.

Task 1: pNexus API (Develop a pHDF5 compliant Nexus API.)

Task 2: Investigate parallel file systems.

Task 3: Investigate implementations on specific file systems

MPI-I/O implementations and pHDF5/pNexus on an even smaller number of preselected file systems.

Task 4: Coupling of advanced (pre-)processing engines.

- Test the capability of the system to cope with multiple parallel data streams. This will contain for example explicit tests feeding a pHDF5-file consisting of a large number of individual images into a multi-core analysis engine.

Task 5: Demonstration.

D8.1: Definition of pHDF5 capable Nexus implementation (M9) - Software

D8.2: Evaluation of Parallel filesystems and MPI I/O implementations (M9) - Report

D8.3: Implementation of pNexus and MPI I/O on parallel filesystems (M21) - Prototype

D8.5: Examination of Distributed parallel filesystem (M21) - Report

D8.6: Demonstrate capabilities on selected applications (M21) - Demonstrator

D8.7: Evaluation of coupling of prototype to multi-core architectures (M30) - Report