



PaN-data ODI

D2.3

First Open Workshop

Grant Agreement Number	RI-283556
Project Title	PaN-data Open Data Infrastructure
Lead Beneficiary	STFC
Dissemination Level	Public
Nature	Report

The PaN-data ODI project is partly funded by the European Commission under the 7th Framework Programme, Information Society Technologies, Research Infrastructures.

Table of contents

Deliverable D2.3	3
Extension of the description of work.....	3
Workshops and meetings directly related to PaNdata work packages	3
Meetings and activities related to identity systems (WP3)	3
FIM4R workshops	5
Terena AAA paper	5
Further activities	5
Meetings and Workshop related to standardization of (meta-)data (WP4-7).....	5
Joint PNI-HDRI and PaNdata ODI workshop	7
Contribution to and (co-)organization of community events	7
SRI 2012 and ECM 27.....	7
NOBUGS 2012	8
Joint Photon Science and European XFEL User meeting.....	9
NMI-3 general assembly.....	9
RDA, EUDAT and DataCite activities	9
Co-operations with industry.....	10
Events organized by PaNdata partners	11

Deliverable D2.3

D2.3 – the first Open PaNdata ODI workshop – was originally intended as a single workshop to provide a suitable platform for knowledge exchange and interaction with user communities, developers and Photon and Neutron RIs as service providers. Another goal was to intensify the knowledge exchange with related projects like EUDAT and FIM4R. EUDAT has been largely superseded by the activities of the Research Data Alliance, which promotes knowledge exchange and exploitation of synergies on a much broader scope.

Extension of the description of work

In contrast to the original plan we decided to slightly modify the description of work and seek for means to achieve a higher impact in particular with respect to communication with user and developer communities by going much more into community events rather than convening a single PaNdata ODI workshop. This includes co-organization of major events as well as topic specific meetings and satellite workshops. The details will be outlined below.

In addition, the recently established Research Data Alliance (RDA) provides a promising platform on knowledge exchange with all relevant projects in that area. Rather than interfering with the efforts of RDA to establish a common platform we decided to devote more efforts into EUDAT and RDA driven initiatives.

Workshops and meetings directly related to PaNdata work packages

PaNdata ODI contributes to a large number of topical events in various areas ranging from standardization to identity management. We will briefly describe a few of the more relevant activities.

Meetings and activities related to identity systems (WP3)

There are a number of PaNdata ODI goals, which require a high level of harmonization or standardization with projects or communities outside the PaNdata consortium. The core elements of the PaNdata Open Data Infrastructure are hence intensively discussed with and disseminated to the communities and projects. One of these core elements is the pan-European identity system.

PaNdata ODI aims to establish a common, unique identity for all its users. This identity system, the so called Umbrella, is intended to provide additional services to both users and facilities, like a harmonized proposal submission procedure, educational elements or a common facility database. The facility database is a major undertaking, since it requires establishing a common, multi-lingual schema and merging of literally ten thousand of entries. However, once established maintenance of the database will become much easier and more efficient since the effort is divided through all partners.

Umbrella has been thoroughly tested not only by facility staff but also by a significant number of users providing valuable feedback to the developers. To promote implementation and deployment two teams have been organized, one concentrating more on management issue, the other on in-depth technical aspects. The teams hold regular telephone conferences and meetings. Since the topic is of high interest not only for the RIs organized in PaNdata, a number of additional RIs and projects

are participating in the effort, such as EMBL¹ as the Biostruct-X² project leader, members of the NMI3 and CALIPSO projects, GSI/FAIR³ and the European XFEL⁴.

Umbrella has been presented at several meetings and conferences, in particular Biostruct-X and Calipso⁵ meetings, the IUCR⁶, the series of workshops on Federated Identity Management (FIM) for Research Collaborations⁷ initiated by EiroForum⁸ and more general events of EUDAT⁹ or e-IRG¹⁰. There has been a meeting between WP3 (Umbrella) and ESUO, the synchrotron user organization. This has led to an explicit recommendation by ESUO of Umbrella as the official identification tool at the European synchrotron facilities.

The main goal of the PaNdata ODI project is to establish a common, federated, open data infrastructure. This involves a number of issues ranging from federated identity management to meta-data standards or persistent identifier for data, instruments and related publications. Quite a number of projects and initiatives are working in closely related topics. Seeking for synergies, establishing co-operations and exchanging knowledge is therefore particularly important to arrive at solutions facilitating integration into a global data infrastructure.

Projects like CALIPSO or Biostruct-X aim for example to ease the access to the facilities through transnational access support or harmonized proposal submission systems for protein crystallography (PX) beamlines. Implementation of Umbrella at the synchrotron facilities is now an explicit part of the CALIPSO project. Both projects consequently need to tackle identity management issues, and the Umbrella also intends to incorporate certain parts of the proposal submission, but not restricted to PX instrument. To arrive at common solution satisfying the needs of the user communities, intense consultations between these projects and PaNdata ODI have been established and Biostruct-X is meanwhile participating in the Umbrella management team to exploit synergies in this area of common interest.

One platform to discuss directions in identity management and the implementation of a common system are the series of so called harmonization meetings. Four of these meetings^{11 12 13} have been organized by PaNData ODI partners, the latest jointly between the European XFEL and PSI. The meeting took place in Hamburg at the Eur.XFEL site and was visited by 35 participants from various RIs¹⁴. At the meeting important security and organizational issues have been intensely discussed. As a result the participants agreed on major standardization issues and ILL will draft a Memorandum of Understanding to provide a solid basis for a sustainable operation of the IdM infrastructure.

¹ European Molecular Biology Laboratory: <http://www.embl-hamburg.de>

² <http://www.biostruct-x.eu/>

³ Facility for Antiproton and Ion Research: <http://www.fair-center.de/>

⁴ European X-Ray Free Electron Laser: <http://www.xfel.eu>

⁵ Calipso: Coordinated Access to Lightsources to Promote Standards and Optimization

⁶ International Union of Crystallography: <http://www.iucr.org/>

⁷ <https://cdsweb.cern.ch/record/1442597>

⁸ <http://www.eiroforum.org/>

⁹ <http://www.eudat.eu/>

¹⁰ e-Infrastructure Reflection Group: <http://www.e-irg.eu/>

¹¹ <http://indico.psi.ch/conferenceDisplay.py?confId=1039>

¹² <https://indico.desy.de/conferenceDisplay.py?confId=5061>

¹³ <http://indico.psi.ch/conferenceDisplay.py?confId=1752>

¹⁴ <http://indico.psi.ch/conferenceDisplay.py?confId=2159>

FIM4R workshops

Federated identity management is also a crucial topic heavily discussed in the FIM4R workshops. Within the last two years there have been four workshops, the most recent one in Nymegen¹⁵. An important outcome of these activities is a paper ‘Federated Identity Management for Research Collaborations’¹⁶.

Terena AAA paper

An interesting study has been performed by Terena “Advancing Technologies and Federating Communities¹⁷” with significant contributions from PaNdata, such as a use case “Accessing Experiments and Data” from the Photon and Neutron Community as well as The Umbrella Project as an example of a AAI.

Further activities

Beyond these activities FIM has been intensely discussed with projects like Eur.XFEL, CRISP¹⁸ and PNI-HDRI¹⁹ or LSDMA²⁰ and the Umbrella developments are regularly tested and investigated by users from the scientific communities. The upcoming FIM4R²¹ workshop will be hosted and organized by PaNdata partner PSI.

Project moonshot²² is developing tools which permit integration of the Umbrella in non-web based analysis workflows and could provide access to compute resources through the Umbrella credentials. PaNdata ODI is currently investigating the set of tools with Project Moonshot.

Identity management and AAI is also an important topic of the EUDAT project.²³ PaNdata has provided feedback on the requirements of the Photon and Neutron user communities to EUDAT through participation in interviews, workshops and conferences. Likewise, PaNdata has submitted very detailed comments on the e-IRG blueprint²⁴ on data infrastructures to e-IRG reflecting the particular view of our user communities.

Meetings and Workshop related to standardization of (meta-)data (WP4-7)

The standardization efforts in PaNdata focus mainly on the data formats, metadata schemes and defined vocabularies. HDF5 has been proposed by the EC to serve as the ISO standard for all binary data. PaNdata has adopted this proposal and selected NeXus as their community wide standard. NeXus is fully HDF5 compliant, but with a standardized meta-data scheme and controlled vocabularies. So NeXus is not really a new data format, but more of a convention how to organize data and meta-data in a HDF5 container, which greatly facilitates exchange and concurrent use of data from different scientific fields or experiments. The NeXus structure enables the automatic processing of the meta-data and ingestion of the data into data catalogues. For example, a NeXus

¹⁵ <https://indico.cern.ch/conferenceDisplay.py?confId=191892>

¹⁶ <https://cdsweb.cern.ch/record/1442597>

¹⁷ <https://confluence.terena.org/download/attachments/30474266/2012-AAA-Study-report-final.pdf?version=1&modificationDate=1355503760046>

¹⁸ Cluster of Research Infrastructures for Synergies in Physics: <http://www.crisp-fp7.eu>

¹⁹ Photon and Neutron Infrastructures - High Data Rate Initiative: <http://www.pni-hdri.de>

²⁰ Large Scale Datamanagement and Analysis: <http://www.helmholtz-lsdma.de>

²¹ Federated Identity Management for Research Collaborations

²² <http://www.project-moonshot.org/>

²³ <http://www.eudat.eu/>

²⁴ <http://www.e-irg.eu/publications/blue-papers.html>

ingestor is currently being developed for ICAT²⁵ and the ICAT scheme, which is fully compliant with the Dublin Core standard²⁶, and will be further adopted to support community requirements and the NeXus data model.

Standardization requires naturally the involvement of many disjoint communities, like developers, vendors, users and IT-expert. PaNdata partners aim to continuously contribute and drive the process. One core platform to engage with the NeXus community is the Nexus International Advisory Committee (NIAC)²⁷. Some PaNdata partners have been participating in the NIAC almost since its existence. The PaNdata ODI efforts have led to a stronger and broader representation in the NIAC. At the latest NIAC meeting, H. Bernstein representing imgCIF has joined the NIAC. CIF is another well-established standard widely used in the field of protein crystallography. To achieve interoperability between NeXus and CIF is hence an important issue tackled by the NIAC and the IUCr, and is well progressing²⁸. More information can also be found on the IUCr forum on NeXus HDF5 CIF convergence²⁹.

High Speed data recording is becoming more and more important. Free Electron Laser facilities aiming to resolve processes at a femto-second scale strongly depend on the ability to record images at a MHz rate, but also synchrotrons, FAIR and even some Neutron facilities like ESS face the same problem with emerging new detector generations. To cope with these challenges PaNdata is co-operating with developers and vendors. There is for example an intense co-operation with Dectris³⁰, one of the leading detector companies, the NIAC and HDF5.org to enable detectors writing NeXus/HDF5 natively at the speed required. A workshop has been organized by PSI to support this process and DESY has developed in co-operation with the PNI-HDRI project an implementation of the NeXus API capable to deal with such data rates. Work on compression algorithms and methods to augment HDF5 with pluggable compression and image filtering modules is on-going.

The recent NeXus developments have been presented at the NIAC and NeXus code camp and the NOBUGS conference in September 2012.

²⁵ <http://www.icatproject.org/>

²⁶ <http://dublincore.org/>

²⁷ <http://wiki.nexusformat.org/NIAC>

²⁸ http://wiki.nexusformat.org/NIAC2012#Meeting_Minutes

²⁹ <http://forums.iucr.org/viewforum.php?f=31>

³⁰ <https://www.dectris.com/>

Joint PNI-HDRI and PaNdata ODI workshop



The first Joint PNI-HDRI and PaNdata workshop took place in Hamburg beginning of 2012 and was rather successful in knowledge exchange and creation of collaborations on specific aspects of application development. Due to the success the projects decided to continue the initiated series of workshops. The upcoming workshop concentrates more on technical computing aspects like GPU and high performance computing, and topics related to large scale data management, visualization and analysis, also inviting external experts. The workshop hence concentrates on issues related to work packages 5 (Virtual Labs) and 7 (Scalability).

Contribution to and (co-)organization of community events

There were or will be four major events organized by PaNdata partners which we considered particularly suited to promote PaNdata ODI. Each of these events has a major impact in user communities as well as Photon and Neutron RIs.

SRI 2012 and ECM 27



In July 2012 the 11th International Conference on Synchrotron Radiation Instrumentation³¹ took place in Lyon, by far the largest event of this kind, organized by PaNata partners ESRF and Soleil with a number of satellite workshops at PSI, DESY or the European XFEL. One particularly noteworthy activity at the conference was the information gathering on data management.

Rudolf Dimper (ESRF) and Philippe Martinez (Synchrotron SOLEIL) organized a round table on large data volume management³². The outcome of the round table discussion was presented³³ at the Workshop on Data Diffraction Deposition³⁴ (DDD) at ECM27³⁵, the European Crystallography Meeting

³¹ SRI2012: <http://www.sri2012.org>

³² Programme: <http://www.lepublicsystemepco.com/files/modules/freezones/ProgrammeSRI2012-Web2.pdf>

³³ http://www.iucr.org/data/assets/pdf_file/0008/69479/Data-Mgt-RT.pdf

³⁴ <http://www.iucr.org/resources/data/dddwg/bergen-workshop>

in Bergen/Norway as a supplement to the talk by Heinz Weyer presenting PaNdata and related projects³⁶. Erica Yang and Brian Matthews were completing the PaNdata ODI presentations with a talk on *Linking raw experimental data with scientific workflow and software repository*³⁷. The presentations revived the discussion on data deposition and open access to scientific data in the IUCr, and the PaNdata ODI policy framework and the implementation with ICAT at ISIS was featured in the workshop report as *“An exemplar of good practice demonstrating access to raw data is at the ISIS UK neutron source”*³⁸. This workshop offered an ideal occasion to exchange ideas and requirements with between facility staff and users at the large facilities represented by PaNdata.

The workshop report lists some actions and recommendations. It suggests in particular *“to encourage and recommend to the IUCr Executive Committee that authors should provide a permanent and prominent link from an article to the raw data sets underpinning a journal publication with a view to making this a formal requirement on authors at such time as the community has adopted raw data deposition as a routine procedure”* and emphasized the *“urgent need to be clear about the metadata required for the various IUCr Commissions and their experimental raw data”*.



This is an important step towards an open data infrastructure for one of our particularly prominent user communities. The deposition of scientific data, linked to persistent identifier cited in a publication, standardization of meta-data with the idea to make data re-usable and accessible is exactly the type of infrastructure PaNdata ODI aims to provide.

NOBUGS 2012



Another particularly important event was NOBUGS 2012³⁹ which was organized by STFC and DLS and took place at RAL Sept. 2012. NOBUGS is the only event for photon science applications where developers and users from various scientific areas come together. It was hence an obvious choice to

³⁵ <http://ecm27.ecanews.org/>

³⁶ <http://pan-data.eu/sites/pan-data.eu/files/03-BergenECM272012fb-Weyer.pdf>

³⁷ <http://pan-data.eu/sites/pan-data.eu/files/04-Linking-Raw-Data-with-Scientific-Workflow-PanData-ODI-Early-Experience.pdf>

³⁸ <http://forums.iucr.org/viewtopic.php?f=21&t=102>

³⁹ <http://nobugs2012.org/>

attach a significant part of the PaNdata ODI activities to the NOBUGS conference, there is hardly a better opportunity to disseminate and exchange knowledge with developers and user communities.

The NOBUGS conference was hence accompanied by a number of PaNdata satellites, like Umbrella and ICAT workshops as well as a NeXus Code camp and NIAC meeting. PaNdata work was presented in several of the NOBUGS talk like the data management reports from DESY⁴⁰ and ALBA⁴¹, a status report for NeXus⁴², the data catalogue requirements analysis⁴³ and the practice of data citation at ISIS⁴⁴. Several of the talks were again focusing on data analysis frameworks and workflows further emphasizing the need for an integrated data analysis framework. More information about the satellite workshop can be found on the PaNdata web and the NOBUS2012 site.

Joint Photon Science and European XFEL User meeting

In January the joint Photon Science and European XFEL User meeting in Hamburg was organized by the European XFEL and PaNdata partner DESY. With more than 800 registered participants it is one of the largest user community events. PaNdata ODI presented a poster (jointly with LSDMA) and pursued a survey to investigate users' needs, requirements and awareness of data management issues. We in particular approached user groups asking for the availability of open access data. So far the feedback was surprisingly positive, and just hindered by the problem of users to actually locate experimental data, which nicely illustrated the need for a sustainable data infrastructure. These efforts will be continued and summarized at a later stage.

On the occasion of the users meeting we gave some interviews to a master student in science sociology, who intends to use PaNdata ODI as a case study to illustrate the pathways for creation of virtual environments.

NMI-3 general assembly

The General Assembly NMI3-II 2013 conference in June 2013 will be organized by PaNdata partner HZB. The conference assembles all European Neutron and Muon RIs. Though not exactly a PaNdata ODI event, we will continue interactions with NMI-3 particularly in view of the authentication system and the user surveys.

RDA, EUDAT and DataCite activities

One of the crucial topics for any data infrastructure is the persistent identification (PID) of digital objects and the citability of the corresponding PIDs. Some PaNdata partners are actively working on this topic in co-operation or at least based on knowledge exchange with projects and initiatives like EUDAT, OpenAire+⁴⁵ or DataCite⁴⁶. For example, ILL is building up an infrastructure to assign DOIs to datasets and register them through DataCite.

⁴⁰ T Kracht, T Nunez, A Rothkirch, E Wintersberger: [Experiment Control and Data Acquisition at PETRA III, DESY](#)

⁴¹ C Pascual-Izarra et al: [Data management for the Beamlines at Alba](#)

⁴² M Koennecke et al: [The State of NeXus](#)

⁴³ M Prica: [Requirements for data catalogues within facilities](#)

⁴⁴ MD Wilson, BM Matthews, S Nagella, AJ Wilson: [Using DataCite DOIs for ISIS Neutron Source Data](#)

⁴⁵ <http://www.openaire.eu/>

⁴⁶ <http://datacite.org/>

The newly formed Research Data Alliance (RDA)⁴⁷ also aims to establish a number of working groups dealing with PIDs and related meta-data. RDA is explicitly not a platform to promote projects, but rather a bottom-up community or experts approach to derive, implement and establish standards or recommendations in form of RFCs. As such RDA provides the prime platform to exchange knowledge and ideas with individuals involved in related projects and service providers, in particularly EUDAT, iCORDI or ORCID to name a few. One major goal will hence to take up the RDA recommendations and align the PaNdata ODI implementations accordingly (were feasible). To facilitate and guarantee the information flow from RDA to PaNdata ODI and vice versa, PaNdata partners are participating in the management board (J.Bicarregui, STFC) and three of the RDA working groups, in particular on PID Information Types and Type registries. In addition, PaNdata has been participating and contributing to all (open) RDA and EUDAT events so far, and provided substantial feedback to EUDAT in interviews as well as to the e-IRG blueprint.

Co-operations with industry

As mentioned above we are seeking co-operations or exchange of technical knowledge with industrial partners. The focus is here the promotion of the standard data format and the improvement of system for high-speed data collection and recording.

The co-operation with a number of PaNdata partners, in particular DESY, STFC and PSI, and Dectris and the HDFgroup has already led to some results. Dectris is ready to enable their next-generation detectors to natively write NeXus/HDF5. To achieve the desired performance/bandwidth some adjustment to NeXus/HDF5 are required. The adjustments of the NeXus header is discussed and supported by the NIAC. Further acceleration of the HDF5 API is an on-going process between PaNdata partners and the HDFgroup. The HDFgroup will presumably receive funds to implement the desired accelerators, which would be beneficial not only for the PaNdata collaboration and detector vendors, but also for the entire HDF5 user community.

Another topic are low level co-operations with some storage system providers. DESY for example is testing new platforms from Netapp and IBM. The API, in particular NFS 4.1 are thoroughly tested against the current needs of the Photon and Neutron facilities, while keeping an idea on future requirements originating for example from Eur.XFEL, ESS⁴⁸ or SKA^{49 50}. Similar activities are pursued at other facilities also testing products like dCache⁵¹, pvfs⁵², Lustre⁵³, fhgfs⁵⁴ or hadoop⁵⁵. Although these activities are mostly bilateral, experiences and knowledge are exchanged between all partners and the vendors hopefully accelerating the process to arrive at stable high performance systems tailored for the needs of current and future data infrastructures.

⁴⁷ <http://rd-alliance.org/>

⁴⁸ <http://ess-scandinavia.eu/>

⁴⁹ <http://www.ska.ac.za/>

⁵⁰ <http://www.ska.gov.au/>

⁵¹ <http://www.dcache.org/>

⁵² <http://www.pvfs.org/>

⁵³ http://wiki.lustre.org/index.php/Main_Page

⁵⁴ <http://www.fhgfs.com/cms/>

⁵⁵ <http://hadoop.apache.org/>

Events organized by PaNdata partners

The following table summarizes the most important events (co-)organized by PaNdata ODI partners.

Date	Event	Org.	Location	Type	Materials
20-21.06.2013	General Assembly NMI3-II 2013	HZB	Berlin,DE	Open	Agenda
20-21.03.2013	5th Meeting on Federated Identity for Research Collaborations	PSI	Villigen,CH	Open	Agenda , Participants
13-14.03.2013	5th PaNdata ODI Project Meeting	MAX IV	Lund,SE	Internal	
12-13.03.2013	ICAT workshop in Lund	MAX IV	Lund,SE	Open	
04-05.03.2013	(parallel) Computing in Photon and Neutron Science Applications	DESY	Hamburg,DE	Open	Agenda
21-22.01.2013	4th PaNdata & CRISP Harmonisation meeting	Eur.XFEL & PSI	Hamburg,DE	Open	Agenda , Slides Participants
10-11.10.2012	European Common Affiliation Database Workshop	ESRF	Grenoble,FR	Open	Agenda
28.09.2012	4th PaNdata ODI Project Meeting	STFC	Didcot,UK	Internal	Agenda

PaNdata ODI open workshops and NOBUGS 2012

Date	Event	Org.	Location	Type	Materials
27.09.2012	Dawn satellite workshop	DLS & STFC	Didcot,UK	Open	Calendar , Programme
27.09.2012	Pan European Authentication workshop	DLS & STFC	Didcot,UK	Open	Programme
27.09.2012	ICAT satellite workshop	DLS & STFC	Didcot,UK	Open	Agenda & Materials , All slides (zip)
24-26.09.2012	NOBUGS 2012	DLS & STFC	Didcot,UK	Open	Programme
20-21.09.2012	NIAC2012	DLS &	Didcot,UK	Open	Participants &

		STFC			Minutes
18-19.09.2012	NeXus code camp	DLS & STFC	Didcot,UK	Open	Participants & Minutes
Date	Event	Org.	Location	Type	Materials
09-13.07.2012	SRI 2012	ESRF & Soleil	Lyon,FR	Open	Agenda
27-28.06.2012	3rd PaNdata ODI Project Meeting	Elettra	Trieste,IT	Internal	Slides & Agenda, Participants
13.06.2012	3rd PaN-Data & CRISP Harmonisation Meeting	PSI	Zurich,CH	Open	Participants, Slides & Agenda, all slides (zip)
20.03.2012	ICAT workshop	ILL	Grenoble,FR	Open	Slides, Participants & Agenda
27-28.02.2012	1st Joint PNI-HDRI & PaNdata workshop and 2nd PaNdata ODI Project Meeting	DESY	Hamburg,DE	Open	Agenda, Participants, Slides, Report
08.12.2012	2nd PaN-Data & CRISP Harmonisation Meeting	DESY	Hamburg,DE	Open	Agenda, Minutes, Participants, Slides
03.11.2011	PaNdata ODI kickoff meeting	STFC	Didcot,UK	Open	Slides, Participants
02-03.11.2011	2nd workshop on Federated identity system for scientific collaborations	STFC	Didcot,UK	Open	Agenda & Slides