



PaN-data ODI

Deliverable D2.2

Dissemination & Engagement

Grant Agreement Number	RI-283556
Project Title	PaN-data Open Data Infrastructure
Title of Deliverable	Dissemination & Engagement
Deliverable Number	D2.2
Lead Beneficiary	STFC
Deliverable Dissemination Level	Public
Deliverable Nature	Report
Contractual Delivery Date	01 Jan 2012 (Month 3)
Actual Delivery Date	10 Mar 2012

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

Abstract

The document describes the current and future engagement and dissemination activities focussing on selected events and actions.

Keyword list

PaN-data ODI, Knowledge exchange, Engagement, Dissemination.

Document approval

Approved for submission to EC by all partners on 15.03.2012

Revision history

Issue	Author(s)	Date	Description
1.0	Frank Schlünzen	10 Mar 2012	Complete version for approval
1.1	Rudolf Dimper	10 Mar 2012	Corrections
1.2		10 Mar 2012	Final

Table of contents

	Page
1 INTRODUCTION	4
2 ENGAGEMENT	4
2.1 ENGAGEMENT WITH OTHER FACILITIES.....	4
2.1.1 <i>In Europe</i>	4
2.1.2 <i>Beyond Europe</i>	4
2.2 ENGAGEMENT WITH OTHER PROJECTS.....	5
2.2.1 <i>PNI-HDRI</i>	5
2.2.2 <i>CRISP</i>	5
2.2.3 <i>EUDAT</i>	6
2.2.4 <i>ORCID</i>	7
2.2.5 <i>OpenAire</i>	7
2.3 ENGAGEMENT WITH APPLICATION DEVELOPERS AND DETECTOR SUPPLIERS	7
2.4 ENGAGEMENT WITH AND DISSEMINATION TO USERS.....	8
3 COMMUNICATION CHANNELS	8
4 IMPORTANT EVENTS	9

1 Introduction

This document provides a brief outlook on current and future dissemination activities. Due to important events, which were expected to have some impact on specific details on the engagement with some of the projects, which we believe are of particular relevance for the work within PaNdata ODI, we considered it more appropriate to delay the delivery of this document to warrant a consistent view. In particular a collaborative workshop of PaNdata and PNI-HDRI, and the EUDAT User Forum, which took place in Barcelona 7-8th of March 2012, was thought to affect the dissemination activities and communication channels to arrive at a more collaborative environment, and the event clearly had this impact. We will outline this further below.

2 Engagement

2.1 Engagement with other facilities

PaNdata ODI is a joint effort of 11 partners, the same partners which were already involved in the PaNdata Europe project, a pure support action to define standards and policies forming an essential corner stone of PaNdata ODI.

Meanwhile the **Swedish lightsource Max-Lab**, which operates several photon sources in Sweden and the **Jülich Center for Neutron Science (JCNS)**, which operates a number of experimental facilities at both FRM-II and ILL, joined the PaNdata collaboration. Although they are not directly beneficiaries, both institutions are invited to all PaNdata ODI meetings and are directly involved in developments, implementations and decisions being made.

2.1.1 In Europe

Within Europe, PaNdata ODI encompasses almost all major Neutron and Photon large RIs. Two exceptions are the **European XFEL**¹ (Eur.XFEL) and the **European Spallation Source**² (ESS). However, both facilities will only become operational in a few years from now. We nevertheless aim to keep the Eur.XFEL and the ESS informed of the developments of general concepts, to be able to meet their future requirements if needed. The Eur.XFEL is already participating in a number of activities, namely the efforts on federated identities and the CRISP and PaNdata harmonization meetings. With ESS we made initial contacts and intend to invite both Eur.XFEL and ESS to selected future PaNdata events.

2.1.2 Beyond Europe

We initiated a small FP7 project (PaNGEA) to foster co-operations with Neutron and Photon facilities in the US and Australia. We initially contacted four such laboratories, which had already small on-going collaborations with PaNdata on specific topics, namely **ANSTO**³ and the **Australian Synchrotron**⁴ in Australia and **ORNL**⁵ an **SSRL/SLAC**⁶ in the US.

¹ <http://www.xfel.eu>

² <http://ess-scandinavia.eu/>

³ <http://www.ansto.gov.au/>

⁴ <http://www.synchrotron.org.au/>

The aim of the project is to identify potential areas of co-operation in data management, real time analysis and visualization. Of particular interest are issues related to interoperability of data infrastructures and harmonization of data policies and standards. The Australian facilities for example use a distributed data catalogue (TARDIS⁷). It could be an interesting issue, what requirements and measures have to be taken to interface between the PaNdata implementation and Tardis. A similar issue comes from LCLS at SLAC which use an iRODS⁸ based system. Interoperability with iRODS is hence another interesting issue potentially being tackled in co-operation with EUDAT (see below).

Regardless of the success of the grant application for the PaNGEA project, we will continue knowledge exchange and co-operation with our Australian and US American colleagues, and invited them for the next PaNdata workshop in fall 2012, in conjunction with the NOBUGS⁹ conference. DAITF and ICRO are additional platforms we intend to utilize for future engagements (see below).

2.2 Engagement with other projects

We are continuing to create and broaden collaborations and knowledge exchange with several projects based on the contacts established during the PaNdata Europe project. This means in particular to present the PaNdata ODI project to the corresponding communities, to identify opportunities for joint investigations and to intensify existing collaborations.

2.2.1 PNI-HDRI

PaNdata and PNI-HDRI¹⁰ have been collaborating on selected topics already, in particular on standardization and policies. There are however more themes of mutual interest. Both projects pursue developments on GPU-accelerated data analysis and visualization, workflows or analysis frameworks. Detector development and fast data acquisition are not part of the work programme of PaNdata ODI, but naturally of high relevance at least for all photon sources. On the other hand software and data catalogues developed by PaNdata ODI are very interesting for PNI-HDRI as well, though not covered by any of their work packages. It hence made sense to intensify and broaden the co-operation between PNI-HDRI and PaNdata ODI with a joint workshop¹¹ which recently took place at DESY in Hamburg. The co-operation will certainly be continued on several organisational and technical topics in the future.

2.2.2 CRISP

CRISP¹², an FP7 supported cluster of ESFRI projects, contains 4 work packages on topics related to data and user management. The work packages are somewhat complementary to PaNdata ODI. To avoid duplication of work and to benefit as much as possible from the complementarity, both projects get together on a bi-annual basis to exchange information on developments and

⁵ <http://www.ornl.gov/>

⁶ <http://www-ssrl.slac.stanford.edu/>

⁷ <http://code.google.com/p/mytardis/>

⁸ <https://www.irods.org>

⁹ <http://nobugs2012.org/>

¹⁰ <http://www.pni-hdri.de>

¹¹ <https://indico.desy.de/event/5517>

¹² <http://www.crisp-fp7.eu/>

strategies. The last event in this series of joint workshops was held at DESY¹³ on December 8th 2011, and involved also Biostruct-X and PNI-HDRI. One of the core themes of the meeting was focussing on user management and federated identities. The same topic is also discussed in a series of EiroForum initiated workshops on Federated identity system for scientific collaborations, with very active participation by PaNdata ODI. The 3rd workshop was held recently in Taipei¹⁴, the previous one organized by PaNdata partner STFC at RAL in November 2011¹⁵. PaNdata is contributing to the requirements paper to be presented at TERENA Networking Conference (TNC2012) in May 2012. PaNdata will continue to participate and contribute to these activities. PaNdata ODI will also participate in the upcoming two events, the CRISP annual meeting in April 2012 and the 4th workshop on federated identity systems in May 2012.

2.2.3 EUDAT

We started to liaise with EUDAT¹⁶ at the e-IRG concertation meeting in Lyon and the e-IRG workshop in Poznan in 2011. D.Corney presented the EUDAT project at the PaNdata ODI kick-off meeting at STFC in December 2011, and PaNdata ODI took part in the EUDAT requirements interviews as well as the EUDAT User Forum¹⁷ very recently. At the EUDAT User Forum J.Bicarregui presented the PaNdata ODI perspective on data infrastructure and outlined the potential areas of co-operation between EUDAT, PaNdata ODI and other participating projects.

Currently, EUDAT focuses on safe replication and persistent identifier (PID) as the first services to be implemented. While data replication is not particularly high on our list of priorities, persistent identifiers for data, software or instruments is an important element of PaNdata ODI's provenance strategy. Naturally, making data citable and use these citations in peer reviewed articles must involve publishers as well as organizations handling the PIDs. EUDAT is presumably supporting more than one handle system like EPIC¹⁸, DataCite¹⁹ or plain DOIs²⁰, which fits well with the PaNdata aims. More importantly, EUDAT is co-operating already with DataCite and this channel can be utilized to communicate requirements and developments.

In addition to PIDs, a few more areas of potential (long term) co-operations have been identified:

- **Interoperability:** EUDAT will focus primarily on iRODS. PaNdata will presumably use ICAT, but some of our US partners are also running an iRODS instance. Interfacing between iRODS and ICAT is hence an important issue.
- **Workflows:** PaNdata ODI and PNI-HDRI are both developing analysis frameworks and implementation of workflows for automatic and remote data analysis is an aspect of the Virtual Labs work package. PaNdata ODI and PNI-HDRI will align their developments. However, EUDAT aims to deploy a generic, flexible workflow system (or execution framework), which could form a valuable basis for PaNdata ODI and PNI-HDRI developments.
- **AAI:** EUDAT aims to provide an AA Infrastructure based on Shibboleth and Moonshot. That fits exactly with the current PaNdata developments (Umbrella). As soon as EUDAT succeeds to create the global AAI, PaNdata will easily be able to join the federation. We will

¹³ <https://indico.desy.de/event/5061>

¹⁴ <http://indico.cern.ch/conferenceDisplay.py?confId=177418>

¹⁵ <https://indico.cern.ch/event/157486>

¹⁶ <http://www.eudat.eu/>

¹⁷ <http://www.eudat.eu/news/eudat-user-forum-barcelona-7-8-march-2012>

¹⁸ <http://www.pidconsortium.eu>

¹⁹ <http://www.datacite.org>

²⁰ <http://www.doi.org>

follow the EUDAT developments and contribute our experiences with Shibboleth and Moonshot.

It was agreed at the EUDAT user forum to create suitable communication channels to inform about relevant events and developments. Selected technical events (like for example an upcoming iRODS workshop) will be open for all EUDAT partners and interested projects. This will become an important channel for PaNdata's dissemination and engagement activities.

Following the EUDAT user forum, PaNdata was invited to participate in the DAITF workshop and ICRI 2012 in Copenhagen in March 2012²¹. This will provide a great opportunity to present PaNdata's perspective to an international consortium, and to influence the data infrastructure strategies on a global scale.

2.2.4 ORCID

Initial contacts to ORCID²² were made at the e-IRG concertation meeting in Lyon and the e-IRG workshop in Poznan. ORCID was invited to present the project at the PaNdata ODI kickoff meeting, where it was agreed to keep informed on various aspects related to identities. The Researcher ID being established by ORCID is a highly relevant topic for PaNdata's provenance plans, and creating unique correspondence between researcher IDs and the personal IDs provided by the Umbrella system could be quite beneficial for researchers, users and facilities. ORCID also provides a channel to receive and send feedback to publishers, which will be helpful to establish PID systems.

2.2.5 OpenAire

OpenAire+²³ aims to provide an Open Access Infrastructure for Europe. Though it clearly focuses on publications and funding, it permits to establish links between data and publications as well as to register open access data in so called orphan repositories. We have been in contact with OpenAire and the project was also presented at the PaNdata ODI kickoff meeting. We will continue knowledge exchange and mutual participation in future events.

2.3 **Engagement with application developers and detector suppliers**

It has generally been agreed to use NeXus/HDF5 as *the* standard data format, not only within PaNdata ODI but also for example by PNI-HDRI. This widely adopted format permits to promote the standard to detector vendors and to foster NeXus compliant application development.

As a first success, a popular detector company (Dectris) has agreed to support NeXus natively with the next generation of detectors²⁴. The NeXus API has been adjusted to cope with the requirements of the detector producer. We are currently in discussion with other companies as well.

PaNdata is also co-operating with HDF5.org on feature enhancements of HDF5, which forms the basis for NeXus and has been proposed as an ISO standard by the European Commission. Contribution to HDF5 developments will hence be highly beneficial beyond the PaNdata project.

²¹ <http://www.ereg.me/ehome/index.php?eventid=31679&>

²² <http://about.orcid.org/>

²³ <http://www.openaire.eu/>

²⁴ http://wiki.nexusformat.org/NIAC2011_CodeCamp

PaNdata is developing a software catalogue which meanwhile contains a significant fraction of the applications in use at Photon and Neutron sources. The software catalogue also facilitates establishing interest groups, which can be used to disseminate and promote applications and NeXus developments.

PaNdata is strongly contributing to NeXus developments and will continue its work with the NeXus International Advisory Committee (NIAC).

2.4 Engagement with and dissemination to users

All PaNdata facilities are constantly in close exchange with the user communities. The beamline scientists at the facilities are in discussion with users on a daily basis. The large majority of the user groups have collaborations to groups and users within the individual facilities. In addition, all facilities hold annual user meetings; offer schools and tutorial workshops on various research topics. All these communication channels are used to inform users about PaNdata developments, and receive feedback and requirements from user communities. In addition, we consider preparing a user survey to consolidate the requirements gathering and disseminate the PaNdata view on an open data infrastructure.

There is some information exchange with the User Organizations through Elisa/ESUO²⁵ and NMI-3²⁶. PaNdata is also discussing strategies with the upcoming project CALIPSO and Biostruct-X²⁷, which is also a more user oriented project. In particular, PaNdata advocated the Umbrella system, which might be suitable to support the proposal submission system and AAI of CALIPSO and Biostruct-X.

The AAI developed by PaNdata (Umbrella²⁸) has meanwhile reached the prototype phase. Test instances of the prototype are operational at various facilities with a full integration into the established user management systems. In addition, a number of central services are available for testing. To promote the AAI and to receive feedback on the usability and user friendliness of the infrastructure, the Umbrella has entered a “friendly user phase”. Several user communities distributed over Europe are currently testing the system and provide feedback on all aspects. The friendly user phase will be concluded at the CRISP annual meeting in April 2012 followed by a meeting with the user communities later on.

3 Communication channels

As outlined above, we utilize several established communication channels. In addition, the new PaNdata web site provides up to date news on events and developments. A news mailing list is currently being created in particular to communicate information to related projects like EUDAT, CRISP, PNI-HDRI to mention a few. In addition, we have been in contact with e-science talk²⁹ and hope to co-operate with the project on various dissemination activities.

²⁵ <http://www.esuo.org/>

²⁶ <http://nmi3.eu/>

²⁷ <http://www.biostruct-x.eu/>

²⁸ <http://umbrella.psi.ch>

²⁹ <http://www.e-sciencetalk.org/>

4 Important events

There are an ever increasing number of events with PaNdata participation, events organized by PaNdata ODI or PaNdata partners and events of closely related projects. On the PaNdata web site we aim to provide a fairly complete and newsworthy list of events. Below are listed a number of selected events of particular relevance for the PaNdata ODI project:

Date	Event / Reference	Location	Type ³⁰
~20.10.2012	EUDAT conference	Barcelona	c
24-26.09.2012	NOBUGS 2012 http://nobugs2012.org/	STFC / DLS Didcot	o
~27.09.2012	Open PaNdata workshop at RAL Satellite to NOBUGS 2012	STFC / DLS Didcot	o
09-13.07.2012	11 th International Conference on Synchrotron Radiation Instrumentation, http://www.sri2012.org/	Lyon	o
~13.07.2012	Face to face Umbrella meeting Satellite to SRI 2012	Lyon	o
09-12.07.2012	Open Repositories 2012 http://or2012.ed.ac.uk/	Edinburgh	p
06.2012	PaNdata ODI workshop	ELETTRA Trieste	o
21-22.06.2012	4 th workshop on federated identity systems	MPI Nimjegen	c
21-24.05.2012	Terena Networking Conference TNC 2012 https://tnc2012.terena.org/	Reykjavík	p
23-25.04.2012	CRISP annual meeting	Crete	c
21-23.03.2012	DAITF and ICRI 2012 conference http://www.ereg.me/ehome/index.php?eventid=31679&	Copenhagen	c
07-08.03.2012	EUDAT User Forum http://www.eudat.eu/news/eudat-user-forum-barcelona-7-8-march-2012	Barcelona	c
27.02.2012	3 rd workshop on federated identity systems for http://event.twgrid.org/isgc2012/idm.html	isgc2012 Taipei	c
27-28.02.2012	Joint PNI-HDRI and PaNdata workshop https://indico.desy.de/conferenceDisplay.py?confId=5517	DESY Hamburg	o
08.12.2011	2 nd PaNdata and CRISP Harmonization meeting https://indico.desy.de/conferenceDisplay.py?confId=5061	DESY Hamburg	o
04-05.11.2011	PaNdata Europe Final and PaNdata ODI kickoff meeting See http://wiki.pandata.eu/	STFC / DLS Didcot	o
02-03.11.2011	2 nd workshop on federated identity systems https://indico.cern.ch/conferenceDisplay.py?confId=157486	STFC Didcot	o

³⁰ o: organized by PaNdata partners;
c: with contribution (poster, talk) by PaNdata;
p: participating