



European Data  
Grant agreement number: RI-283304

### **D3.4.1: First EUDAT Conference Proceedings**

Author(s)	Nagham Salman
Status	Final
Version	v1.0
Date	14/02/2013

Abstract:

This report about the first EUDAT Conference describes how the conference was organized and run, and gives details of the conference programme and discussions that were held during the event, as well as the conclusions that were reached as a result of the activities held during the event. The report also mentions considerations to be taken into account when planning the next EUDAT conference.

Document identifier: EUDAT-DEL-WP3-D3.4.1	
Deliverable lead	BSC
Related work package	WP3
Author(s)	Nagham Salman (BSC)
Contributor(s)	Adam Carter (EPCC), Riina Salmivalli (CSC), Genet Edmondson (SNIC), David Vicente (BSC), Giuseppe Fiameni (CINECA), Shaun de Witt (STFC), Damien Lecarpentier (CSC)
Due date of deliverable	31/12/2012
Actual submission date	06/02/2013
Reviewed by	Peter Wittenburg, Alison Kennedy
Approved by	PMO
Dissemination level	PUBLIC
Website	www.eudat.eu
Call	FP7-INFRA-2011-1.2.2
Project number	283304
Instrument	CP-CSA
Start date of project	01/10/2011
Duration	36 months

*Disclaimer:* The content of the document herein is the sole responsibility of the publishers and it does not necessarily represent the views expressed by the European Commission or its services.

While the information contained in the document is believed to be accurate, the author(s) or any other participant in the EUDAT Consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the EUDAT Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the EUDAT Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

## TABLE OF CONTENT

<b>EXECUTIVE SUMMARY .....</b>	<b>4</b>
<b>1. ORGANISING AND RUNNING THE CONFERENCE .....</b>	<b>5</b>
1.1. Initial organization .....	5
1.2. Advertising.....	5
1.3. Attendees .....	6
<b>2. CONFERENCE PROGRAMME .....</b>	<b>8</b>
2.1. First day: Training track sessions .....	8
2.1.1. Track One: The technical track.....	8
2.1.2. Track Two: General training about working with digital data .....	10
2.1.3. Track Three: Communication sessions .....	11
2.2. Second Day: Plenary, parallel and poster sessions .....	11
2.2.1. Plenary sessions .....	11
2.2.2. Parallel sessions .....	13
2.2.3. Lightning talks .....	14
2.2.4. Poster session .....	15
2.2.5. Social dinner .....	15
2.3. Third day: Plenary, parallel and concluding sessions.....	16
2.3.1. Plenary sessions .....	16
2.3.2. Parallel sessions .....	16
2.3.3. Wrap-up and conclusion.....	17
<b>3. CONCLUSIONS .....</b>	<b>18</b>
3.1. Feedback from attendees .....	18
3.2. Overall conclusions .....	19
<b>ANNEX A. GLOSSARY .....</b>	<b>20</b>
<b>ANNEX B. PROMOTIONAL MATERIALS .....</b>	<b>22</b>

## LIST OF FIGURES

Figure 1: Plenary session .....	5
Figure 2: Conference flyer .....	6
Figure 3: Distribution of attendees .....	7
Figure 4: Training session .....	8
Figure 7: Feedback .....	18

## LIST OF TABLES

Table 1: Lightning talks .....	14
Table 2: Posters.....	15

## EXECUTIVE SUMMARY

The First EUDAT Conference was held on 22-24 October 2012 in Barcelona and organised by WP3 (Dissemination, Training and Outreach) on behalf of the EUDAT project. On a general level, the purpose of the conference was to provide a forum for both data infrastructure providers and researchers who use data facilities to meet and address the current issues involving data infrastructures. More specifically, the conference aimed to show researchers how they could benefit from using the services that EUDAT is developing and how to actually use those services, while also working with people who are running data infrastructures that will interact with EUDAT's services and looking at managing the interactions between such infrastructures and the EUDAT services.

The first day of the conference was dedicated to a range of training sessions, while the second and third days of the conference consisted of plenary sessions in the morning, followed by parallel sessions of talks in the afternoon. Additional shorter talks and poster sessions were held in the afternoon and evening of the second day prior to the conference dinner. The programme on the second day of the conference focused on giving an overview of the data infrastructure scene, primarily from a European perspective, with an introductory plenary session that discussed the research, policy and service-provider perspectives related to data infrastructures. Subsequent parallel sessions provided a unique opportunity for users of research data to learn about the services being developed by EUDAT and the potential benefits of using these services, with concrete demonstrations from research communities and data centres involved in the project.

Later in the day there were sessions for lightning talks and posters – these were designed to give attendees a broad view of the work that is being undertaken in Europe in this field. The lightning talks were divided into two categories: the first consisted of initiatives within particular research communities and disciplines describing their progress, and the second addressed more general issues related to the use, operation, and management of generic data infrastructures. At the poster session, 18 groups presented their research work or projects.

The programme for the final day was more global in scope and concentrated on issues relevant for truly world-wide data infrastructures, with plenary sessions in the morning by international speakers addressing the issue from the familiar angles of research, policy and service provision. In the afternoon, this theme was continued in a range of parallel sessions discussing the latest developments for achieving greater interoperability between international data infrastructures, the sustainability of data infrastructures, and the views and expectations of users in relation to the EUDAT project.

The conference enabled over two hundred people (from Europe and also further afield) who are involved with digital data (both researchers and providers of data services) to meet and discuss the development of the EUDAT European Collaborative Data Infrastructure, and more general world-wide issues relating to sharing data globally. Attendees were satisfied with the conference organization and sessions, and we look forward to the continuing collaboration at the next EUDAT conference in Rome later this year. We will keep the set-up similar to that of the first conference – with a mix of plenary and parallel sessions – as this proved to be very successful, however a stronger emphasis will be put on training, sustainability, and partnerships with other stakeholders.

## 1. ORGANISING AND RUNNING THE CONFERENCE

### 1.1. Initial organization

The First EUDAT Conference was held in Barcelona during October 2012. The aim of the conference was to bring together data infrastructure providers and practitioners from across the globe to discuss current data infrastructure challenges and solutions, with a focus on interoperability, cross-disciplinarily and cross-border collaboration.

Planning for the event began in February 2012. To ensure the high quality of the conference, WP3 – along with the PMO and the conference committee – began by creating a work plan that defined the tasks that needed to be undertaken. Groups were set up to implement the various parts of the plan as follows.

- A committee was created to prepare the technical programme for the conference. This committee was also given overall responsibility for supervising the organization of the conference.
- An organizational team was created to handle all the logistics related to the event.

As the conference was being hosted by BSC, the organizational team proposed holding the conference at the AXA Convention Centre in Barcelona (see Figure 1), as it has excellent facilities, and also great connections for both local and international travel.



Figure 1: Plenary session

### 1.2. Advertising

The conference was advertised on the EUDAT website (via its own dedicated webpage) and also on websites belonging to EUDAT's partner organisations and those of the research communities associated with EUDAT at the time. WP3 also promoted the conference through articles, newsletters, social networks and other websites (for example, iSGTW and e-science-talk).

A small one-page flyer containing details of the conference programme (see Figure 2) was created and distributed at earlier events to advertise the conference.



Figure 2: Conference flyer

We decided that it would be helpful for conference attendees to have a handy reference document that they could use during the conference. To that end, we put together a comprehensive conference booklet that included:

- A welcome by the project coordinator, project manager and scientific coordinator,
- The general programme for the conference,
- Details about the training sessions on offer,
- Details of the parallel sessions,
- Details about the lightning talks,
- Summaries of the posters being presented, and
- Information about the speakers at the conference.

This booklet was distributed to all attendees at the start of the conference.

WP3 designed promotional materials that were distributed at the conference. All conference attendees were given a conference bag containing a plastic folder with copies of the conference flyer and booklet and a map showing the layout of the rooms for the conference sessions, along with the EUDAT advertising materials: a USB stick, lanyard, pen, notebook, and a thermally-insulated mug (see Annex B for some illustrations).

### 1.3. Attendees

More than 200 people from all over the world attended the First EUDAT conference (see Figure 3). Although most of the attendees were from Europe, we were very pleased to welcome people from as far afield as Australia and the United States.

The majority of the attendees at the First EUDAT Conference came from scientific research communities. Those present included representatives from 23 European research communities in a wide range of fields:

- Biomedical sciences – VPH, ECRIN, INCF, Neugrid, DiXA and Biomedbridges,
- Environmental sciences – EPOS, ENES, LifeWatch, VERCE, GEO-SEAS, ODIP, ESPAS, BioVel, DRIHM, EISCAT, Euro-VO and ENVRI,
- Humanities and social sciences – CLARIN, GESIS and DASISH, and
- Physical science – DESY and CRISP.

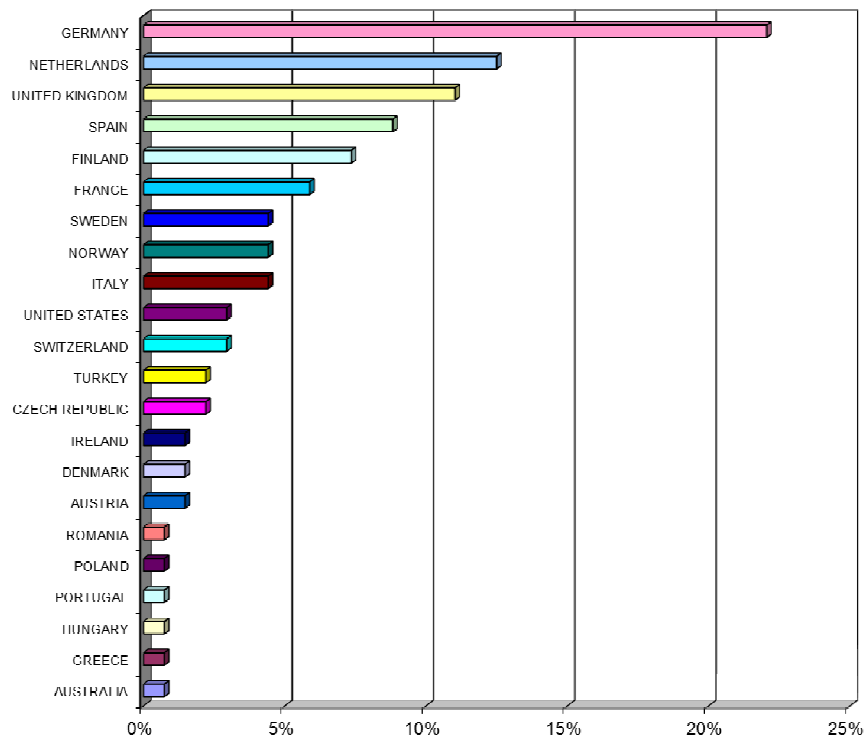


Figure 3: Distribution of attendees

Representatives from these research communities contributed to the discussions at the conference both by expressing their views and expectations in relation to EUDAT, and also by presenting data solutions that are being used within their own communities. Representatives from the four ESFRI “clusters” – ENVRI, DASISH, BioMedBridges and CRISP – were invited to host a session presenting the common solutions being developed in their fields, as well as the challenges they are working to overcome.

## 2. CONFERENCE PROGRAMME

The conference programme was spread over three days. The first day consisted of training sessions that were run by EUDAT, and there was also a communication training session offered by the e-ScienceTalk team. The second and third days consisted of plenary and parallel sessions, with additional short talks and poster sessions being held prior to the conference dinner on the second day. The final day of the conference finished with a wrap-up session and concluding remarks by Peter Wittenburg who is EUDAT's scientific coordinator.

### 2.1. First day: Training track sessions

The training day at the EUDAT conference comprised two parallel training tracks focusing on digital data issues and a communication training track offered by the e-ScienceTalk dissemination team. One of the data training tracks was geared towards technical issues for people who are responsible for data infrastructures that will interface with EUDAT's data services, and the other data track addressed more general issues for researchers working with large quantities of digital data.

The slides from the talks in these training sessions are available online at [www.eudat.eu](http://www.eudat.eu) and the following sections give a brief summary of each of the talks presented in the different tracks.



Figure 4: Training session

#### 2.1.1. Track One: The technical track

The technical training track consisted of a series of eight talks delivered in workshop style where discussion amongst the participants during the sessions was encouraged. The talks were as follows.

##### **Policies and Processes in iRODS – Adil Hasan, University of Liverpool**

This talk provided an introduction to policies and processes, and also to the iRODS system. iRODS provides an approach to policy-based data management, while policy-based data management provides flexibility in managing data and interacting with other data management systems. iRODS also makes the connection between policy and service level agreements more apparent, provided the enforced policy can be unambiguously measured.

##### **Policy rules: Concrete implementations – Jean-Yves Nief, Yonny Cardenas, CNRS**

This talk gave an overview of the use of iRODS in production at CC-IN2P3 – addressing the questions “Why are we using it?” and “Who is using it?” – and then went on to discuss future prospects for using iRODS at CC-IN2P3. Various examples of rules’ policies in iRODS were presented – these covered

resource monitoring, biomedical applications, arts and humanities applications, mass storage system management, access rights, and pitfalls.

**Policy rules: Concrete implementation (2): MPI-TLA iRODS Implementation – Willem Elbers, MPI-TLA**

This talk focused on the use of iRODS in CLARIN and MPI-TLA. Examples of data management policies were provided, and the current policy implementation was described. Logical replication with iRODS was also illustrated. iRODS was shown to be suitable as a logical replication layer in a small scale federation, because it is easy to perform multiple tasks from a single policy and easy to implement custom behaviour, such as SAM-FS interaction.

**Persistent Identifiers: Solving a number of problems through a simplistic mechanism – Tobias Weigel, Deutsches Klimarechenzentrum**

This talk described what a Persistent Identifier (PID) is, why EUDAT cares about PIDs, and how PIDs can be used. PIDs are a generic tool with sharply defined boundaries, while handles provide additional useful features and can support a large number of use cases. This talk also gave a brief introduction to the way that EUDAT is using the Handle System and EPIC.

**The Handle System and its Usage – Larry Lannom, CNRI**

This talk provided a technical overview of the Handle system. The talk covered the role of identifier resolution systems in information management on networks, the Handle System, Handle strings, and how these are resolved to typed data. Processes for handle resolution were described along with the role of the Handle client. Template handles, offline signatures, multiple resolution, content negotiation, and the use of namespace include were also discussed.

**EPIC Web Service for PID Management – Pieter van Beek, SARA**

This talk introduced EPIC, which is a consortium of four computer centres actively involved in governance bodies and working on its own governance structure. The talk also covered the EPIC web service, which is a RESTful web service providing, amongst other things, configurable server-side PID generation. During the talk, there was a demonstration of using the API.

**DataCite Services – Sebastian Peters, German National Library of Science and Technology**

DataCite is a global consortium of institutions focused on improving the scholarly infrastructure around datasets and other non-textual information. The consortium provides standards, workflows and best-practices, and works with datacentres and other organisations that hold data. The talk covered DOI and Metadata Management, data harvesting (including OAI-PMH, Dublin Core and the DataCite Schema), along with searching and retrieving metadata. Resolution statistics were also presented.

**The EPIC API: EUDAT EPIC iRODS Integration – Willem Elbers, MPI-TLA**

This talk presented the work that has been going on to integrate EPIC and iRODS in the context of EUDAT. The talk described an iRODS module offering a Python wrapper, iRODS integration, an iRODS rule-base and iRODS system hooks. Then a typical workflow was presented. The software that was described uses a flexible implementation based on linked lists. Current limitations, and pointers about how other actions could be implemented, were also provided.

### **2.1.2. Track Two: General training about working with digital data**

The second training track was a more general track targeted at all researchers. In this track a range of subjects were covered. The training talks were related to big data, the integration of data from distributed sources, and data management.

#### **Big Data – Adam Carter, EPCC, The University of Edinburgh**

This talk gave introduced Big Data including explaining what the term means, the challenges presented by very large data sets, and the opportunities offered by big data. It was argued that we could do a lot more with our data than we currently do. For example, combining data from other researchers and fields creates new opportunities. Using data infrastructures and services provided by others also allows for a division of labour and specialisation. It was concluded that data should be valued as much as the resulting research papers.

#### **Simple Store: An Overview of a Potential New EUDAT Service – Erwin Laure & Ilja Livenson, PDC/SNIC**

A Simple Store is one of the services proposed by EUDAT. The details of the service have not yet been fully defined, so this session presented the ideas behind the service and the results of an evaluation of a candidate technology, Invenio. The talk covered the problems with current storage systems, along with aspects of existing related services (such as Dropbox and YouTube), and presented the requirements that have been gathered for the service. Other candidate technologies were also presented.

#### **The Benefits of Using PIDs – Larry Lannom, CNRI**

This talk described the benefits of using Persistent Identifiers (PIDs). The PID concept was presented as the neck of an hourglass model for information management on networks, which separates the data sources from the value-added services that can make use of the data. The talk also described some disadvantages of using PIDs and offered guidance as to when they should be used. The requirements for a useful PID system were then described, and the Handle System was presented as an example of such a system.

#### **Metadata and EUDAT – Shaun de Witt, STFC**

This talk addressed the questions: “What is metadata and why do we need it?”, “What makes good metadata?”, and “Where do you put your metadata?”. The talk then went on to describe EUDAT’s efforts to address these issues. The definition of metadata as “data about data”, whilst accurate, hides much of the complexity. It was argued that communities need to spend more time understanding their metadata, and that cross-disciplinary access will mean that even more thought will be required.

#### **Moving Data: Overview of EUDAT data staging capabilities – Claudio Cacciari & Giuseppe Fiameni, CINECA**

The process by which data can be moved between storage facilities (such as those in the EUDAT infrastructure) and computing nodes (including HPC resources) was described. Requirements were discussed and various aspects of the process were described, including how users could interact with the system. Opportunities for integration with PRACE and EGI were also described.

#### **Sharing Data and Protecting Data – Adam Carter, EPCC**

Security of data is arguably a prerequisite for data sharing. An Authentication and Authorisation Infrastructure (AAI) can help to protect data and to share it securely. The talk covered ways of achieving

distributed AAI, the aspects of such an infrastructure that would be visible to an end user, some AAI terminology and a brief description of EUDATs planned approach for an AAI.

### **Workflows – Martin Steghöfer, UPF**

This talk presented the general idea of workflows and offered a general introduction to Taverna to help the audience understand the concepts behind workflows. The talk went on to describe what has been done with Taverna and Gimias in the context of VPHShare (where workflows are a very important foundation of the project).

### **Data Streaming – Amy Krause, EPCC, The University of Edinburgh**

Data Streaming and its relationship to Data Intensive Computing were described. The talk described what data streaming is, and why it might be used. Streaming has the potential to allow large-scale data handling and can exploit very large-scale parallelism and distribution. It was stressed that data flow descriptions are important for describing methods and composing tasks. The DISPEL language was presented as a means to construct data-flow graphs, generate data-flow patterns and discuss data-flow engineering.

### **New Risks and Solutions for Collaborative Data Security – Urpo Kaila, CSC**

This talk described what data security really involves, and also highlighted some misconceptions of what security entails. The talk described long-standing risks with collaborative data, and described existing solutions for dealing with these. The talk also described collaborative data infrastructures, pointed out that cloud solutions present new security risks, and explored strategies to address these risks.

#### **2.1.3. Track Three: Communication sessions**

We were fortunate to be able to collaborate with the e-ScienceTalk team so that a third training track (focusing on communication) was also available on the first day of the conference. In this track, the e-ScienceTalk team offered training and advice in a number of areas related to communications: media outreach, science writing, communications strategy, measuring the impact of communication and creating virtual worlds.

## **2.2. Second Day: Plenary, parallel and poster sessions**

The schedule on the second day of the conference was busy with plenary and parallel sessions, lightning talks and a poster session, as well as the conference dinner in the evening.

### **2.2.1. Plenary sessions**

There were two plenary sessions in the morning that included the following talks and presentations.

#### **Addressing the new data challenges – cross-disciplinary initiatives and open science**

The *first plenary session* focused on how to address the new data challenges that will arise in the near future for the cross-disciplinary initiatives and open science. During this session there were three talks:

- “Data sharing and research excellence in astronomy and beyond: Synergies and tensions” by Professor Bernard F Schutz from MPG,
- “Scientific data infrastructures: Towards Horizon 2020” by Kostas Glinos from EC, and
- “EUDAT: Towards a collaborative data infrastructure” by Kimmo Koski, Director of CSC and EUDAT Coordinator.

Professor Bernard Schutz from the Max Planck Institute for Gravitational Physics in Potsdam, Germany, introduced LIGO, the international network of gravitational wave detectors. He explained the synergies and tensions involved in data sharing and how it provides challenges, but also opportunities, for new ways of interacting. Professor Schutz also raised the important topics of the issues related to open data and open access to publications, and the lack of international standards for open data – which are urgently needed for data sharing to evolve.

Kostas Glinos, Head of e-Infrastructures at DG-CONNECT, explained how the European Commission has given funding of around 100 million Euros for data projects in recent years. The main objective of this funding has been to provide end-to-end services for real users. The EC is also looking to break down the barriers between disciplines and to offer multidisciplinary services that are as broad as possible.

Finally, Kimmo Koski, the Coordinator of the EUDAT project and Managing Director of CSC, explained the first year of the EUDAT project and how it is establishing a collaborative data infrastructure. He also remarked on the importance of the users in the EUDAT CDI, and pointed out that EUDAT itself includes communities from life sciences (VPH), earth sciences (EPOS, LifeWatch and ENES) and humanities (CLARIN) as some of the partners in the project. The EUDAT project is working to put different services (namely data staging, safe replication, simple storage, authentication and metadata) into production to fulfil the common requirements of these research communities.

### **Developing common solutions through cluster initiatives**

The *second plenary session* focused on the different initiatives behind the ESFRI clusters, and consisted of four talks:

- “BioMedBridges: Constructing data and service bridges in the life sciences” by Stephanie Suhr, EBI,
- “ENVRI: Operating an environmental RI” by Wouter Los, UvA,
- “CRISP: Cluster of Research Infrastructures for Synergies in Physics” by Laurence Field, CERN, and
- “DASISH: Weaving the SSH infrastructure” by Daan Broeder, MPI-PL, and Johan Fihn, SND.

Stephanie Suhr of the BiomedBridges project described how the project is bringing together ten emerging Research Infrastructures (RIs) in the Biological and Medical Sciences on the ESFRI roadmap. This constitutes a large number of RIs and partners (as each of the RIs has up to 50 partners) – these are all coordinated in the project by the European Bioinformatics Institute (EBI). Much of the data in the biomedical sciences has ethical, legal or societal implications, which is one of the main reasons why it is not possible to share this data. However BioMedBridges aims to add value to the existing biomedical data by linking it, while taking into account the ethical, legal and regulatory issues resulting from sharing this kind of data. A case study was mentioned where BioMedBridges is linking data on diabetes and on obesity-related large-scale datasets in mice and humans to bridge the phenotype gap. This case study is an example of how the BiomedBridges project is adding value to existing data by bringing together data that was previously held by separate RIs.

Wouter Los from ENVRI explained the diversity of data that is produced by the different research infrastructures within the ESFRI cluster for environmental science. ENVRI’s main goal is to enable multidisciplinary scientists to access and study data from multiple domains. The data that ENVRI works with ranges from readings from marine sensors and seismic data to DNA sequences for biodiversity. ENVRI is looking for common services to make it possible to share data between infrastructures while taking into account the specific characteristics of the data. Consequently ENVRI sees EUDAT as a suitable

actor to tackle generic services (for example, the real-time data handling required for the data from radar sensors) that are relevant to the environmental science community.

Laurence Field from CERN talked about the Cluster of Research Infrastructures for Synergies in Physics (CRISP). CRISP unites 11 Research Infrastructures (RIs) across different research domains, from experimental facilities like CERN to analytical facilities like ESRF and ending with instruments such as the Square Kilometer Array telescope. Essentially this means that CRISP is dealing with images from the very small to the very big: from sub-atomic particles to the whole Galaxy. Data handling for these RIs converges in three main areas: use cases, policies and technology. These areas could be transformed into three main items of collaboration: identity management, data archiving and preservation, and data access policies.

Daan Broeder from TLA-MPI and Johan Fihn from the Swedish National Data Service introduced the DASISH project. This project brings together five ESFRI infrastructures in the area of social sciences and humanities by focusing on common activities across disciplines and infrastructures. The aim of the project is to provide solutions to common problems that will also strengthen international collaboration. DASISH is working to enhance the visibility and reusability of data, tools and services between the infrastructures, with the main goal of enriching the final data by enhancing its value. They foresee a need to create a common data infrastructure, but such a CDI would need to take into account the multiple policies required by the different disciplines.

### **2.2.2. Parallel sessions**

Three parallel sessions were held in the afternoon of the second day where the following topics were addressed.

#### **E-Infrastructure Collaboration**

This session focused on giving an update on the status of the major European e-infrastructures initiatives. The talks were given by Maurice Bouwhuis for EUDAT in the area of data infrastructures, Steven Newhouse for EGI in the area of grid infrastructures, Thomas Eickermann for PRACE in the area of high performance computing, Josep Martrat for Helix Nebula on representing the cloud infrastructures, and Tom Fryer for GÉANT representing the European communications and network infrastructure. The different speakers provided updates on the current status of their infrastructures and also explained the possible collaborations between the infrastructures to provide better services to the final users.

#### **EUDAT Services: Safe Replication and Data Staging**

In this session, the preliminary EUDAT services for Safe Replication and Data Staging were presented to the scientific research communities. This was followed by an opportunity for interaction with the research communities to collect comments and feedbacks. Both EUDAT services were successfully demonstrated, gaining the attention of the attendees. Implementation details and further extensions were discussed in depth to ensure that the current communities clearly understand how to employ EUDAT services in their existing solutions. The questions that were raised referred to issues such as the capability of the services to accommodate further requirements, the stability and efficiency of the services, the usability of the interfaces that are provided, and the level of compliance with existing standards.

#### **EUDAT Services: Metadata**

This talk gave a high level overview of what metadata means, namely that it is not just 'data about data' but also includes information relating to file structure, data usage restrictions and ownership. What

makes ‘good’ metadata was also discussed. In particular, it was noted that the use of well-defined ontologies and the usage of existing standards (where applicable) increases the reusability of data and simplifies searches by other communities. The problem of developing a metadata tool to work across different scientific domains was also discussed, and the preliminary thoughts of the EUDAT metadata task force were presented. Ross Wilkinson from the ANDS then gave a talk about how the ANDS is approaching the problem of cross disciplinary metadata. That was followed by some debate on the detailed plans of the EUDAT metadata project, particularly with respect to the handling of different schemas, ontology mapping and the language of records.

### 2.2.3. Lightning talks

In this first EUDAT conference we also wanted to encourage interesting data-related projects and initiatives to present their work – this was done through the lightning talks and poster sessions.

The lightning talks session included presentations from 14 initiatives and projects as shown in Table 1.

**Table 1: Lightning talks**

Project/ Initiative	Title	Speaker
VERCE	An e-Science Environment for the Virtual Earthquake and Seismology Research Community	Alberto Micheleni
GEO SEAS	A Pan-European Infrastructure for Management of Marine and Ocean Geological and Geophysical Data	Helen Glaves
ESPAS	A Near Earth Space Data Infrastructure for e-Science	Natalia Manola
VPH-SHARE	Virtual Physiological Human: Sharing for Health Care – A Research Environment	Peter Coveney
N4U	The European Science Gateway to Neuroimaging	David Manset
DRIHM	A Distributed Research Infrastructure for Hydro-Meteorology Study	Han van Veldhuizen
ODIP	Ocean Data Interoperability Platform	Helen Glaves
eduGAIN	Interconnecting Federations to Link Services and Users Worldwide	Mikael Linden
OPENAIRE+	Linking Data to Publications	Natalia Manola
SIM4RDM	Enabling Researchers to Effectively Utilise the Emerging Data Infrastructures	Matthew Dowey
KNOWLEDGE EXCHANGE	Actions to Take on Sharing and Re-using Research Data – The Knowledge Exchange Perspective	Stefan Winkler-Nees
ODE	Drivers and Barriers in Data Sharing	Hans Pfeiffenberger
e-FISCAL	Cost Assessment of Computing Infrastructures: Challenges, Approaches and Case Studies	Matti Heikkurinen
e-IRG	A Blue Paper on Data Management	Norbert Meye

Two categories of lightning talks were scheduled: the first consisted of initiatives within particular research communities and disciplines, in the areas of marine biology (GEO-SEAS, ODIP), hydro-meteorology (DRIHM2US), biomedical science (VPH-Share, N4U), and earth sciences (ESPAS, VERCE). The second category addressed more general issues, such as linking data to publications (OpenAIRE+),

sharing data (Knowledge Exchange, ODE), infrastructure usage (SIM4RDM) and costs (e-Fiscal), as well as federated identity management (eduGAIN) and policy developments (e-IRG).

#### 2.2.4. Poster session

Eighteen posters were presented during the poster session, with each presenter being allowed five minutes for their poster. Details of the poster topics are shown in Table 2.

Table 2: Posters

Title	Presenter
Improvements to data transfer facilities in PRACE	Frank Scheiner
Effecting Culture Change with Research Data MANTRA	Stuart MacDonald
The Archive and Data Management Training Center: An Infrastructure for the Social Sciences	Alexia Katsanidou
Cost Study of the European Computing Infrastructure: Comparing the HPC, HTC & Cloud offer	Kashif Iqbal
The Archive and Data Management Training Centre: An Infrastructure for the Social Sciences	Alexia Katsanidou
Data coding and harmonization: How CharmStats is transforming social science data	Kristi Winters
Scientific Large Scale Data Management and Analysis	Christopher Jung
Radieschen – A Framework for a multi-disciplinary Research Data Infrastructure	Leonie Schaefer
Coastal European Water oil Pollution, Satellite Data Integration with local analysis	Jose Manuel Redondo
High Performance Networking in Europe – An overview of SURFnet's European lightpath collaborations	Nicole Gregoire
The European Marine Biological Resource Centre: data and e-infrastructure	Chuck Cook
Web-based Direct Data Transfer Service	Jedrzey Rybicki
Building blocks of the PaNdata Open Data Infrastructure	Kashif Iqbal
Initiative for Globus in Europe	Ioan Lucian Muntean
The XSEDE User Portal File Manager: A Multi-Protocol Solution for Distributed Data Management	Rion Dooley
EGI-The European Grid Infrastructure	Sara Coelho
Legal challenges of cross-border data transfer – implications for research of the 2012 draft EU regulation on data protection and privacy	Roland Krause
ICORDI – International Collaboration On Research Data Infrastructures	David Vicente

#### 2.2.5. Social dinner

The second day of the conference concluded with the conference dinner (which was held at La Font del Lleó in Barcelona). This provided an opportunity for attendees to get to know each other further and to continue informal discussions.

## 2.3. Third day: Plenary, parallel and concluding sessions

The last day of the conference began with a series of plenary sessions, which were followed by a range of parallel sessions. The conference then concluded with a “Wrap up and conclusions” session.

### 2.3.1. Plenary sessions

The plenary sessions on the third day of the conference were focused on the topic: Towards global data infrastructure components. Four plenary talks were given:

- “An Open Architecture for Managing Information in the Internet” by Robert Kahn (CNRI),
- “Data – It’s one world” by Walter Stewart (Research Data Co-ordinator, Research Data Canada),
- “Global platforms to facilitate research data sharing and exchange: Introducing iCORDI and the research Data Alliance (RDA)” by Leif Laaksonen (CSC), and
- “Funder’s perspective on the Research Data Alliance” by Carlos Morais-Pires (European Commission), Alan Blatecky (US National Science Foundation) and Ross Wilkinson (Australian National Data Service).

Robert Kahn from CNRI (Corporation for National Research Initiatives) explained that some level of abstraction is necessary for long-lived systems, since the underlying technologies will surely change during the lifetime of such systems. He emphasised that contextual information may be needed across heterogeneous information systems. He also pointed out that interoperability is critical, but requires effective security for wide-spread application, and that decentralization of administration is advisable provided there is effective coordination. He concluded that the Digital Object Architecture is an appropriate choice to satisfy these objectives.

Walter Stewart from Research Data Canada explained how “data is one world”, and he also discussed the problems and difficulties facing this one world of data and the need for that world to have an infrastructure. He remarked that the challenge is to build an infrastructure that will allow for the open movement of data. He also highlighted the next steps that the Research Data Strategy Working Group will be taking.

Leif Laaksonen from CSC presented global platforms to facilitate research data sharing and exchange, and he introduced the iCORDI project along with its goals and activities. Laaksonen also introduced the Research Data Alliance (RDA) and the current status of its work.

Carlos Morais-Pires from the European Commission, Alan Blatecky from the US National Science Foundation and Ross Wilkinson from the Australian National Data Service explained what motivates the organisations that fund research and development for global data infrastructures. They also discussed the threats to establishing a Global Data Research Infrastructure, the expectations of the Research Data Alliance and its principles.

### 2.3.2. Parallel sessions

The morning of the final day concluded with three parallel sessions.

#### Research Data Alliance

The current status of the RDA was presented in this session. This included aspects of the RDA such as the approach that it uses, its current development status and organizational structure, and the constitution of its working groups. The RDA’s future plans were also outlined.

### **EUDAT User Forum**

This session focussed on discussions about how to integrate new communities into EUDAT, and how to identify the requirements of those communities in relation to data services. The EUDAT task forces were presented and the levels of collaboration were explained. Two communities that have begun to engage with EUDAT (INCF and DiXa) presented their expectations and views on the project, and the next collaboration steps were discussed.

### **Sustainability and Funding Models**

This session discussed the sustainability of data infrastructures and presented the current thinking on the long-term future of EUDAT (as documented in the first EUDAT Sustainability Plan). The session was highly interactive and participants showed great interest in the topics that were discussed. A final straw poll indicated that a sizeable fraction of the audience would be interested in continuing their engagement in sustainability discussions through some other forum.

### **2.3.3. Wrap-up and conclusion**

A final wrap-up session was held to conclude the conference. This was run by the EUDAT Scientific Coordinator, Peter Wittenburg, who emphasised the importance of such events in providing opportunities to interact with each other in person, and in bringing together data practitioners, service providers, and research communities to discuss data infrastructure issues. Wittenburg reiterated the goals of EUDAT and highlighted the achievements of the project during its first year. He reminded us of the long road that still lies ahead of us in the quest to establish a genuine Collaborative Data Infrastructure in Europe and to develop data infrastructures for research data on a global scale.

### 3. CONCLUSIONS

#### 3.1. Feedback from attendees

In order to help with planning for the next EUDAT conference, a feedback form was prepared and sent by e-mail to all conference attendees shortly after the conference.

Only about ten percent of the conference attendees sent back responses to the conference feedback questionnaire. Nevertheless the feedback they provided gave useful information for improving the next EUDAT conference.

The questions on the form were divided into three categories:

- 1) Ranking and evaluating different conference sessions,
- 2) Evaluating the training sessions, and
- 3) Giving suggestions for improvements.

The form allowed the respondents to give a rating from 1 (poor) to 5 (excellent) or to enter comments. Figure 5 shows the average evaluations for different aspects of the conference.

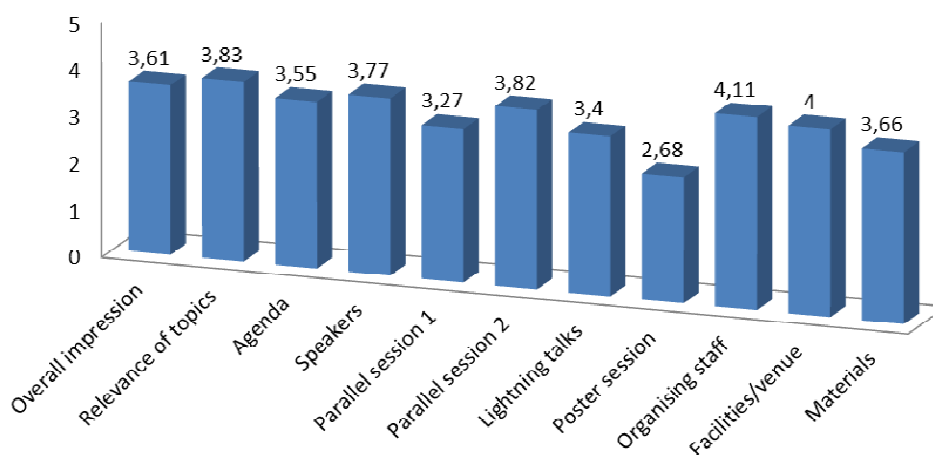


Figure 5: Feedback

The average rating for the conference (including all the different sections of the questionnaire) was 3.28 which is good, but there were differences in how satisfied people were with the different sessions. Respondents were most satisfied with the relevance of the topics generally - which received a rating of 3.83 – and the parallel sessions of the second day (when the topics were: RDA, EUDAT User Forum and Sustainability and funding models) – which received a rating of 3.83.

The majority of the people who attended training sessions were content with the training that was offered and found it to be interesting and pitched at the right level. They also noted that the materials were presented clearly and that the organization of the training sessions was good.

The respondents found the poster session to be the weakest part of the programme (rating 2.68). Even though the conference participants found the poster session interesting, the fact that the individual posters were being presented in a hall where presentations were being given on stage at the same time made it difficult to follow the poster presentations due to the noise from the stage.

The overall organization of the conference was successful according to the respondents. The highest ratings were given for support from the organizing staff (rating 4.11), for the background and information documents (rating 3.66) and for the conference venue and facilities (rating 4.11).

What the respondents would like to see in the next EUDAT Conference is an update about the RDA, more use cases, and discussions about what other projects can do with EUDAT in practise. People also wanted the opportunity for more audience involvement (through scheduling longer periods for discussions) and wanted more technical information to be available at the next conference. The latter could be achieved either by running a larger number of technical sessions during the conference itself or by holding a technically-oriented pre-conference.

### **3.2. Overall conclusions**

The conference lived up to its aims of enabling data infrastructure providers and practitioners to discuss current data infrastructure challenges and solutions. The conference provided a good opportunity for over two hundred representatives from leading research institutions, projects and initiatives in the field of research data infrastructure to meet each other.

The participation of so many research communities (23 communities were represented) from a wide range of scientific fields was a success. It proved that EUDAT is seen as an interesting initiative that has the potential to provide these research communities with viable solutions to some of their data infrastructure needs. We will continue discussions with these communities – in particular, we plan to invite them to the next EUDAT User Forum in March 2013, which will be an ideal forum for engaging further and planning concrete collaboration activities.

The participation of the four ESFRI clusters was also very interesting, and collaboration between EUDAT and these four initiatives will be pursued. ENVRI, DASISH, CRISP, and BioMedBridges are working on identifying and developing common solutions in the domains of environmental sciences, social science and humanities, physical sciences, and biomedical sciences. Their work is very important to EUDAT because it helps us to achieve a greater understanding of the data infrastructure requirements that are shared across different disciplines pertaining to a generic field.

The international scope of the event, which included many participants from the United States and Australia, was also a key note for the conference. In various regions of the globe, there is now a shared understanding that data solutions must be global and that the development of an integrated and interoperable data domain can only be achieved through increasing global collaboration. EUDAT welcomes the Research Data Alliance (RDA) initiative, and will contribute to its efforts to foster interoperability at a global level and to remove barriers to sharing and (re)using data.

EUDAT is planning to hold a second conference in Rome during the autumn of 2013. The set-up will be similar to that of the first conference – consisting of a mix of plenary and parallel sessions – as this proved to be very successful. Stronger emphasis will be put on training, sustainability, and partnerships with other stakeholders, including those from the industrial sector. Supporting each other in developing common solutions, achieving greater technical interoperability between different data infrastructures, and clarifying the roles and responsibilities of each actor in the emerging CDI, will be some of the key objectives of the second conference.

**ANNEX A. GLOSSARY**

ANDS	Australian National Data Service
API	Application programming interface
BioVel	Biodiversity Virtual e-Laboratory
CC-IN2P3	Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules
CDI	Collaborative Data Infrastructure
CERN	European Organization for Nuclear Research
CLARIN	Common Language Resources and technology Initiative. An ESFRI project in the Social Sciences and Humanities domain.
CNRI	Corporation for National Research Initiatives
CNRS	Centre national de la recherche scientifique
CSC	CSC - IT Center for Science Ltd.
CRISP	Cluster of Research Infrastructures for Synergies in Physics
DASISH	DAta Service Infrastructure for the Social Sciences and Humanities
DG-CONNECT	European Commission Directorate General for Communications Networks, Content and Technology
DAITF	Data Access and Interoperability Task Force
DESY	The Deutsches Elektronen-Synchrotron
DiXa	Data Infrastructure for Chemical Safety
DRIHM	Distributed Research Infrastructure for Hydro-Meteorology
EC	European Commission
EBI	European Bioinformatics Institute
EGI	European Grid Infrastructure
EISCAT	European Incoherent Scatter Scientific Association
ENVRI	Common Operations of Environmental Research Infrastructures
ENES	European Network for Earth System Modelling
EPIC	European Persistent Identifier Consortium
EPOS	European Plate Observing System. An ESFRI project in the Environmental Sciences domain.
ECRIN	European Clinical Research Infrastructure Network
ESFRI	European Strategy Forum on Research Infrastructures
ESPAS	Near earth space data e-infrastructure for e-science

EURO-VO	EURO-VO (The European Virtual Observatory) project has been funded by the EU under the FP6 and the FP7.
GÉANT	The fast and reliable pan-European communications infrastructure serving Europe's research and education community
Geo-Seas	The Geo-Seas project is an Integrated Infrastructure Initiative (I3) of the Research Infrastructures programme within EU FP7.
iCORDI	International Collaboration on Research Data Infrastructure
iRODS	The integrated Rule-Oriented Data System
iSGTW	International Science Grid this week
INCF	International Neuroinformatics Coordinating Facility.
LifeWatch	E-Science and Technology Infrastructure for Biodiversity Data and Observatories. An ESFRI project in the Environmental Sciences domain.
MPG	Max Planck Society
MPI-TLA	Max Planck Institute for Psycholinguistics - The Language Archive
OAI-PMH	Open Archives Initiative - Protocol for Metadata Harvesting
ODIP	The Ocean Data Interoperability Platform
PDC	Center for Parallel Computers
PID	Persistent Identifier
PMO	Project Management Office
PRACE	Partnership for Advanced Computing in Europe
RDA	Research Data Alliance
REST	Representational State Transfer
SAM-FS	Storage and Archive Management File System
SARA	Stichting Academisch Rekencentrum Amsterdam
SND	Swedish National Data Service
SNIC	Swedish National Infrastructure for Computing
STFC	Science and Technology Facilities Council
UEDIN-EPCC	The University of Edinburgh - Edinburgh Parallel Computing Centre
UPF	Universitat de Pompeu Fabra
VERCE	Virtual Earthquake and seismology Research Community in Europe e-science environment
VPH	Virtual Physiological Human

## ANNEX B. PROMOTIONAL MATERIALS

We wanted people to go away from the conference with some items bearing the EUDAT logo that would then be used regularly in the workplace (rather than being kept at home) and thus act as advertising for EUDAT in general and for the EUDAT conferences in future. A range of such items were produced with versions of the EUDAT logo or with the name “EUDAT” (as shown below) and were distributed to all conference attendees.

