

Building bridges for research data sharing

FEATURE | OCTOBER 8, 2014 | BY ANDREW PURCELL

Last month, *iSGTW* attended [the EUDAT Third Conference](#) in Amsterdam, The Netherlands. [EUDAT](#) is facilitating efficient research by contributing towards a collaborative data infrastructure in Europe. Its vision is to enable European researchers and practitioners from any research discipline to preserve, find, access, and process data in a trusted environment. EUDAT seeks to achieve this by supporting a collaborative data infrastructure, which is conceived as a network of collaborating, cooperating centers. This will combine the richness of numerous community-specific data repositories with the permanence and persistence of some of Europe's largest scientific data centers. EUDAT has received funding under [the European Commission's FP7 scheme](#) and aims to provide a solution that will be affordable, trustworthy, robust, persistent, and easy to use.

B2 or not B2?

"There's a real need for these kinds of activities," says Kimmo Koski, managing director of [Finland's IT Center for Science \(CSC\)](#), which is the organization coordinating the EUDAT project. "Discussions at this week's other events — [especially the RDA meeting](#) — have clearly shown this."

He continues: "Of course, data access and reuse is important for bolstering European competitiveness, but [the RDA](#) has also helped us to think more globally, too. The longer we wait, the more complex it becomes to solve the challenges around building the collaborative data infrastructure. The time has come to do this now: Today's global economic situation highlights the need for broad collaboration to solve these challenges."

Much discussion at the EUDAT conference centered on the five key data sharing and preservation services the organization has launched to date. These are [B2SAFE](#), which enables safe replication of research data; [B2SHARE](#), which is designed for storing and sharing long-tail research data; [B2FIND](#), which facilitates discovery of research data; [B2STAGE](#), which is all about getting data to computational resources; and [B2DROP](#), which enables researchers to exchange and synchronise data among collaborators.

What to share, what to store

During last month's event, Robert Baxter of [the Edinburgh Parallel Computing Centre in the UK](#) reported on a survey conducted by EUDAT among researchers. This survey, which is due to be published on the EUDAT website later this month, shows that open access is generally viewed positively among research communities. Despite this, Baxter says that researchers are concerned about some of the legal aspects related to sharing and want to have training on this.

In addition to teaching researchers what data they can and cannot legally share, we also need to educate them about what does and doesn't need to be stored, argues Yannick Legré, director of [EGI.eu](#), which coordinates and manages [the European Grid Infrastructure \(EGI\)](#). For example, several participants at last month's event pointed out that, depending on the time period for which some basic genomic data needs to be stored, it may actually be more cost effective to simply throw away the data and resequence the genome when needed again.

Big data, big business, and big collaborations

Carl-Christian Bühr, [member of the cabinet of Neelie Kroes, vice-president of the European Commission \(EC\) responsible for the Digital Agenda](#), gave one of two keynote speeches in the opening session of the EUDAT conference. During this, he emphasized the importance of citizen science in today's society and argued that the desire of businesses to capitalize on 'big data' shows that they are keen to adopt more scientific methodologies in their work.

Edward Seidel, director of [the US National Center for Supercomputing Applications](#), gave the other keynote talk in this session. He argues that much larger scientific collaborations are now possible than ever before in history. "There's a dramatic change in the culture of doing science," he says. "It's not just about the 'big science' instruments, it's also about 'the long tail' of science." As such, it is vital that we support the sharing of research data, claims Seidel, who spoke at length about efforts in the US to establish [a National Data Service](#) with the aim of making it easier to find, use, and publish research data. He also highlighted the key role data sharing has to play in supporting reproducibility of scientific results, accelerating discovery, enabling deeper interdisciplinarity, improving public dissemination of publicly funded research results, and in supporting economic development.

David Rosenthal also gave a keynote speech at last month's event. Rosenthal started the [LOCKSS \(Lots Of Copies Keep Stuff Safe\) program](#) at [Stanford University Libraries](#) in California, US. The LOCKSS program is an open-source, library-led digital preservation system that applies the traditional purchase-and-own library model to electronic materials. The LOCKSS system enables librarians at each institution to take custody of and preserve access to the e-content to which they subscribe, restoring the print purchase model with which librarians are familiar. Rosenthal gave a fascinating talk about some of the challenges associated with long-term data preservation and argued that people's expectations are often far out of line with reality. "It isn't possible to preserve nearly as much as people assume is already being preserved, nearly as reliably as they assume it is already being done," he writes on [his blog](#). "This mismatch is going to increase. People don't expect more resources yet they do expect a lot more data. They expect that the technology will get a lot cheaper but the experts no longer believe it will."

A question of semantics

Alongside the EUDAT conference, [the European Ontology Network \(EUON\)](#) also held its first workshop. Read more about this new exciting initiative in our recent feature article: ['Shaping the future of data sharing with EUON'](#): "EUON's mission is networking," explains co-chair James Malone, a lead ontologist at [the European Bioinformatics Institute \(EMBL-EBI\)](#). "The idea is to bring together people who have expertise in semantics and ontologies and to connect them with people who don't."

In the closing session of the event, Peter Wittenburg, EUDAT scientific advisor, highlighted the importance of the areas being investigated by EUON. "We get so much data from so many areas today," he says. "Semantic interoperability is a big challenge." Damien Lecarpentier, the EUDAT project manager, who is also based at CSC, agrees on the importance of EUON: "It has really taken off," he says. "It works on areas that are of interest to EUDAT and brings new blood to the consortium."

What services may come

"We are working on services in the areas of workflows and semantics," explains Lecarpentier. "However, most of our current work is really focused on enhancement of existing core services." He continues: "EUDAT is very technical and has been oriented towards creating services from the very beginning. Our priority was to build a few services and get them running as soon as possible. Technically, we've made very good progress, but now the priority is to work on the policy aspects of the collaborative data infrastructure. This means bringing all community repositories and data centers together to support a single set of guidelines covering the legal aspects of data reuse and



Panelists discuss sustainability challenges for data and e-infrastructures. Image courtesy Mike Priddy.

combination, as well as strengthening the links and agreements between the sites within the collaborative data infrastructure. We will also focus our attention on the adoption of suitable business models for adopting, supporting and sustaining cost-effective common services."

Lecarpentier also highlights the importance of trust in ensuring EUDAT's success: "It's very important to think carefully about how you build the services," he says. "We've designed them jointly with the user communities right from the beginning, which is fundamental to establishing trust." Koski concurs, saying: "Trust between research infrastructures and e-infrastructure providers is key. User communities need to trust the people working in IT and IT folks need to trust the users."

Average:

Your rating: None Average: 4.3 (3 votes)

About the Author »**Andrew Purcell**

Editor

Andrew Purcell is the editor of iSGTW and is based at CERN, near Geneva.

RELATED TERMS: [Amsterdam](#) [data](#) [data sharing](#) [EGI](#) [EUDAT](#) [EUON](#) [Europe](#) [open data](#) [open science](#) [Research](#) [data management systems](#) [standards](#) [information services](#) [workflow management systems](#) [portals](#), [science gateways](#), and [hubs](#)

Comments

[ADD NEW COMMENT](#)**Post new comment****Subject:****Comment: ***

[Input format](#)

By submitting this form, you accept the [Mollom privacy policy](#).