

# **CERN Workshop on Innovations in Scholarly Communication (OAI8)**

Wednesday 19 June 2013 - Friday 21 June 2013

University of Geneva



**UNIVERSITÉ  
DE GENÈVE**



## **Book of Abstracts**



# Contents

|  |   |
|--|---|
| Altmetrics in the Wild - Alternative Impact Measurement for Scientific Publications . . .  | 1 |
| BG1 - Gold OA Infrastructure (Room 1140) . . . . .   | 1 |
| BG2 - Open Annotations (Room 1150) . . . . .   | 1 |
| BG3 - Altmetrics (Room R160) . . . . .   | 2 |
| BG4 - Open Access Policy Developments (Room R170) . . . . .  | 2 |
| BG5 - How to make your university into a monograph publisher? (Room 1130) . . . . .  | 2 |
| BG6 - Reusing Open Acces materials - a Wikimedia perspective (Room R150) . . . . .   | 3 |
| BG6 - Using Wikipedia's popularity to share research . . . . .   | 3 |
| CERIF for Datasets (C4D) . . . . .   | 3 |
| COnnecting REpositories (CORE) - the current state of aggregating Open Access content  | 4 |
| Citation Finder. A tool for enhancing bibliographic research by extracting references from<br>unstructured scholarly works . . . . . | 4 |
| DML-CZ, Czech Digital Mathematics Library . . . . .  | 4 |
| Empowering Development: Why Open is Right for Development . . . . .  | 5 |
| EuDML, the European Digital Mathematics Library . . . . .  | 5 |
| Guidelines towards implementing Open Access policies . . . . .   | 6 |
| Humanities Session: OA Research Monographs in HSS: Opportunities & Challenges . . .  | 6 |
| Humanities Session: The Humanities in and for the Digital Age . . . . .  | 7 |
| Hydra: open, flexible workflows and community for open access to digital collections . .   | 7 |
| Jisc Gateway for Higher Education . . . . .  | 7 |
| Keynote . . . . .  | 8 |
| Make Knowledge Free [Flooved] . . . . .  | 8 |
| MedOANet Open Access Tracker . . . . .   | 8 |
| Metrics Session: An overview of scholarly impact metrics . . . . .   | 9 |

|  |    |
|--|----|
| Metrics Session: Assessing the transparency of peer review in (Open Access) journals . . .                 | 9  |
| Metrics Session: Discussions of scholarly articles online: who, why and where . . . . .                    | 9  |
| National and institutional open access advocacy campaigns to reach out to research communities . . . . .   | 10 |
| OPENAIREPLUS: 2nd Generation of Open Access Infrastructure for Research in Europe                          | 10 |
| OPENAIREPLUS: Supporting Repository Interoperability through Guidelines . . . . .                          | 11 |
| Open Access for some, Access and Open for all others . . . . .   | 11 |
| Open Access in Arab States- . . . . .  | 12 |
| Open Research Core (ROC) . . . . .   | 12 |
| Palgrave Pivot - Breaking Boundaries in Scholarly Publishing . . . . .                                     | 12 |
| Panel Session: Gold OA Infrastructure . . . . .  | 13 |
| Pundit. Augment Digital Libraries contents with semantically structured annotations . .                    | 13 |
| QUT Research Data Finder: An institutional metadata repository . . . . .                                   | 13 |
| Re-writing of Scholarly Communication . . . . .  | 14 |
| Res. Data Session: Interoperability of Research Data . . . . .   | 14 |
| Res. Data Session: Quality and curation of Research Data . . . . .   | 14 |
| Res. Data Session: Research Data Policies: Seachange or Zeitgeist? . . . . .                               | 15 |
| Res. Data Session: Working with large data sets . . . . .  | 15 |
| Semantic Linkages in Research Information Systems as a New Data Source for Scientometric Studies . . . . . | 15 |
| Semantic Session: Detecting knowledge-level claims in research articles . . . . .                          | 16 |
| Semantic Session: Semantic indexing in PubMed . . . . .  | 16 |
| Semantic Session: Transformation of keyword indexed collections into semantic repositories . . . . .       | 17 |
| Small Data, or: Bridging the Gap Between Smart and Dumb Research Repositories . . . .                      | 17 |
| T1 - Metrics (Room R160) . . . . .   | 17 |
| T2 - Metadata for the Research Lifecycle (Room M6289) . . . . .  | 18 |
| T3 - Metadata: from Records to Graphs (Room R150) . . . . .  | 18 |
| T4 - OJS, beyond editorial tradition (Room 1150) . . . . .   | 18 |
| T5 - The NISO/OAI ResourceSync Synchronization Framework (Room 1130) . . . . .                             | 19 |
| T6 - Open Access Café 2013 (Room 1140) . . . . .   | 20 |

|   |    |
|---|----|
| Tech Session: How semantic representations can support scholarly communication . . .  | 20 |
| Tech Session: Naming on the Web: What scholars should want, and what they can have  | 20 |
| Tech Session: W3C Open Annotation effort: Status and Use Cases . . . . .  | 21 |
| The National Bibliography Number Italia (NBN:IT) Project. A persistent identifier supporting national legal deposit for digital resources . . . . . | 21 |
| The Social Science Open Access Repository (SSOAR) as an element of a greater research information structure . . . . .                               | 21 |
| Towards a common map of academia . . . . .  | 22 |
| UAM institutional repository: use by scientists from the most important university field of research . . . . .                                      | 22 |
| VIVO - IR population and integration . . . . .  | 23 |
| Working towards a global open access repositories network for a networked scholarship   | 23 |
| ZENODO - A new innovative service for sharing all research outputs. . . . .   | 23 |
| bwFLA - a Practical Approach to Functional Access Strategies . . . . .  | 24 |
| re3data.org - A Registry of Research Data Repositories . . . . .  | 24 |



## Altmetrics in the Wild - Alternative Impact Measurement for Scientific Publications

Daniel Beucke<sup>1</sup> ; Ulrich Herb<sup>2</sup>

<sup>1</sup> *SUB Göttingen*

<sup>2</sup> *Digital Publishing, Open Access*

**Corresponding Author(s):** beucke@sub.uni-goettingen.de

The reputation of a scientist is to a large part determined by the visible dissemination and scientific repercussions of his publications, condensed as impact. Citation-based impact factors retain importance as the main instrument for measuring impact of publications.

Nevertheless the impact of a scientific publication is a multidimensional construct and cannot be determined by a single indicator. More and more research is being carried out in the web, be it to access data, to connect with other researchers or to search for or disseminate publications. In particular for the latter tools like blogs, social networks, feeds or research-affiliated platforms gain importance, an area that was traditionally reserved for formal publication providers like publishers. These new platforms for example allow the enrichment of publications through comments, citations or references.

With this more and more commonly use of the web for the dissemination of scientific output opportunities for impact measurement arise. So we perceive that in addition to the measurement of usage data the observation and analysis of the emergence of scientific publications in the above mentioned platforms can determine the impact of the work. This includes citations, references and comments and can be best described with the term "altmetrics".

### Breakout Groups / 28

#### BG1 - Gold OA Infrastructure (Room 1140)

Aim: to hear the perspectives of institutions, funders, platforms and publishers, in order to provide information on what infrastructure is likely to be needed for Gold OA.

Likely topics will include micropayments, certification, consortia payment and management models, discovery and tracking of OA material.

Contributors:

- Pierre Mounier (OpenEdition) - France
- Johannes Fournier (DFG) - Germany
- Martin Rasmussen (Copernicus) - Germany
- Paul Ayris (UCL) - UK

### Breakout Groups / 29

#### BG2 - Open Annotations (Room 1150)

This breakout would focus on the use of annotation to help transform scholarly communications. Ideas include:

- Annotation as a form of peer review overlay mechanism

- Annotation of data to support eg discovery
- Annotation activity as an altmetric

#### **Breakout Groups / 30**

### **BG3 - Altmetrics (Room R160)**

**Aim:** To explore and discuss current developments in the context of alternative metrics. The session will focus on issues as methodology, data quality, stability and reliability.

**Possible outcomes:** Recommendations to improve data quality, acceptance and reliability of altmetrics.

#### **Breakout Groups / 31**

### **BG4 - Open Access Policy Developments (Room R170)**

**Aim:** to share and discuss recent policy developments, including the Executive Directive on public access in the U.S. and Horizon 2020 in Europe. The session will examine how policy developments align (and how they don't), and what we can do to work together globally to achieve the adoption of OA policies.

**Contributors:**

- European speaker : Iryna Kuchma & Victoria Tsoukala

#### **Breakout Groups / 32**

### **BG5 - How to make your university into a monograph publisher? (Room 1130)**

The traditional monograph market is becoming increasingly dysfunctional; failing to meet the needs of researchers in the arts and humanities. Price increases and library budget restrictions are limiting the distribution of newly published monographs while making it harder for authors (especially young, previously unpublished researchers) to get published.

New technology and business models allow us to consider alternatives to the traditional modes of scholarly book publishing. Increasingly academic institutions are finding both the need and the means by which they can take on book publishing activities directly.

**Aims:**

To identify and share experiences of scholarly publication activities within academic institutions and libraries.

To consider how these models might be adopted, adapted and developed by others to provide innovative and viable OA book publishing solutions across HSS and STM scholarly, textbook and open course-ware environments.

This breakout session will look at some of the models that are being used to revolutionise monograph publishing. Participants will be encouraged to share their own experiences of book publishing activity at their institutions.



**Breakout Groups / 67**

## **BG6 - Reusing Open Access materials - a Wikimedia perspective (Room R150)**

Content :

- Introduction to Wikimedia, its mission and how it is aligned with Open Access
- Using Wikipedia's popularity to share research
- Publishing to and from Wikipedia
- Reuse of Open Access materials on Wikimedia projects
- Discussion of current barriers to reuse, beyond licensing issues

Aims:

To share and discuss examples of reuse of Open Access materials, to identify barriers to reuse and to highlight some approaches to overcoming them.

Contributors :

- Daniel Mietchen
- Lane Raspberry
- Samuel Klein

Possible outcomes :

A set of recommendations on how to publish in a way that facilitates reuse

Comments :

The session will be planned at: [http://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Open\\_Access/Wikimedia\\_at\\_OAI8](http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Open_Access/Wikimedia_at_OAI8)  
- and you can edit it.

33

## **BG6 - Using Wikipedia's popularity to share research**

65

### **CERIF for Datasets (C4D)**

William Nixon<sup>1</sup>

<sup>1</sup> *University of Glasgow*

**Corresponding Author(s):** [w.j.nixon@lib.gla.ac.uk](mailto:w.j.nixon@lib.gla.ac.uk)

The overall aim of C4D is to develop a framework for incorporating metadata into CERIF such that research organisations and researchers can better discover and make use of existing and future research datasets, wherever they may be held. A demonstrator was built and this approach verified at the three partner HEIs, each within their own research administration infrastructure.

Many funders state that research organisations are expected to publish metadata to enable their results to be found, and also to support the data curation lifecycle, as part of the conditions of their award. The tools and ideas we have delivered will facilitate this process, and increase community engagement with the process of metadata production and publication.

The poster will present CERIF for Datasets and provide examples of workflows and integrations by universities, funders, and publishers to encourage re-use of ideas and uptake of standards for research data metadata in UK research organisations. This poster aligns with the Research Data theme of OAI8.

47

## Connecting REpositories (CORE) - the current state of aggregating Open Access content

**Author(s):** Zdenek Zdrahal<sup>1</sup>

**Co-author(s):** Petr Knoth<sup>1</sup>

<sup>1</sup> *The Open University*

**Corresponding Author(s):** zd.zdrahal@gmail.com

The CORE (COncecting REpositories) system has been developed to provide the technical infrastructure for aggregating both green and gold Open Access content from institutional repositories and journals in the UK and globally, building a layer of services on top of this aggregation. The CORE technology currently offers 5 applications: The CORE Portal, the CORE Mobile application (for Android and iOS), CORE API, Repository Analytics and the CORE (recommendation) Plugin. The system is unique in its aim to provide a single access point to Open Access research outputs at three access levels (raw data, transactional and analytical) serving different groups of stakeholders including from developers, libraries, repository managers, academics, researchers, life-long learners, funders and business intelligence. CORE goes beyond the services offered by commercial academic search engines (Google Scholar, MS Academic search) as well as metadata search systems offered by international library services. CORE currently aggregates about 10 million research articles from 280+ repositories, its Linked Open Data repository contains over 100 million RDF triples. The CORE systems receives over 170,000 hits every month and has been integrated into various library systems including Open Research Online and the system of the European Library.

38

## Citation Finder. A tool for enhancing bibliographic research by extracting references from unstructured scholarly works

**Author(s):** Filippo Chiocchetti<sup>1</sup>

**Co-author(s):** Vieri Emiliani<sup>2</sup>

<sup>1</sup> *RepubLit Srl*

<sup>2</sup> *Intext Srl*

**Corresponding Author(s):** filippo.chiocchetti@virgilio.it

Citation Finder is a tool for text mining developed by RepubLit Ltd, a software company established as a spin-off of the University of Eastern Piedmont (Vercelli, Italy). It aims at extracting bibliographic references from unstructured, full-text scientific publications (journal articles and monographs). During early development phase, we made use of the Open Library's knowledge base to univocally identify the extracted references.

Citation Finder has been especially trained on documents belonging to the HSS (Humanities and Social Sciences) field, where citations are very difficult to extract automatically, due to the peculiar standards adopted by scholars in that field (citations are not isolated from the text and set in a reference list at the end of the document, but are placed in footnotes and frequently encapsulated within a discourse, no indicators being available to mark-up where a reference starts and where it ends).

While our approach is to serve the publishing industry at large, we – as scholars ourselves, prior to become entrepreneurs – firmly believe in Open Access. Our aim is to illustrate our achievements to the OA Community and to explore new ways to fully leverage the wealth of bibliographic data hidden in the OA full-text archives.

57

## DML-CZ, Czech Digital Mathematics Library

**Author(s):** Vlastimil Krejcir<sup>1</sup>

**Co-author(s):** Michal Ruzicka<sup>2</sup>

<sup>1</sup> *M*

<sup>2</sup> *Masaryk University, Faculty of Informatics, Czech Republic*

**Corresponding Author(s):** krejcir@ics.muni.cz

Czech Digital Mathematics Library (DML-CZ, <http://dml.cz/>) is a project for digitizing relevant mathematical literature which has been published throughout history in the Czech lands. The DML-CZ primarily contains professional journals, proceedings and monograph collections published by Czech institutions and is continuously growing (ingesting periodically published content).

The poster briefly introduces the DML-CZ project itself, describes the digitization workflow, software tools and new technologies developed to help to create and enrich the content of the DML-CZ. Key features of the DML-CZ web user interface, built on the DSpace system, and the connection via OAI-PMH to the EuDML (European Digital Mathematics Library, <http://eudml.org/>) is presented too.

Plenary 5 / 18

## Empowering Development: Why Open is Right for Development

The world we live in today is more connected than ever before. Technology is the great enabler, with citizens around the world accessing information on their mobile phones, smart phones, tablets, and PCs. The World Bank has responded to this change by adopting an “Open Agenda,” a sea change in how it makes available its data, information, and knowledge. In the last three years, the World Bank has implemented an Access to Information policy, an Open Access Policy, Creative Commons licensing, and has “opened” statistical databases and other datasets that were previously not available to the public or available via subscription only. At the World Bank we believe that openness drives transparency, which leads to greater accountability, which—in turn—leads to better development results. This presentation reviews the World Bank’s Open Agenda and how it is driving transparency, accountability, and—most important—helping to deliver better development results.

56

## EuDML, the European Digital Mathematics Library

**Author(s):** Aleksander Nowinski<sup>1</sup> ; Thierry Bouche<sup>2</sup>

**Co-author(s):** José Borbinha<sup>3</sup> ; Petr Sojka<sup>4</sup> ; Vlastimil Krejcir<sup>5</sup>

<sup>1</sup> *Interdisciplinary Center for Mathematical and Computational Modelling University of Warsaw, Poland*

<sup>2</sup> *Institut Fourier (UMR 5582) & Cellule Mathdoc (UMS 5638), Université Joseph-Fourier, France*

<sup>3</sup> *INESC-ID, Portugal*

<sup>4</sup> *Masaryk University, Faculty of Informatics, Czech Republic*

<sup>5</sup> *M*

**Corresponding Author(s):** krejcir@ics.muni.cz

EuDML, the European Digital Mathematics Library <http://eudml.org>, is a project that has build a new multilingual service for searching and browsing the content of existing European mathematical portals. It is based on a rich aggregating metadata and full texts of heterogeneous and multilingual collections of digitised (math OCR including formulae) and born digital content (peer-reviewed articles, books, etc.) of mathematical content. The service merge and enhances the information about each document from each collection, and also match documents and references across the entire combined library. Entities such as authors, bibliographic references and mathematical concepts (mathematical subject classifications, MSCs) are linked to matching items in the collections; similar mechanisms are being provided as public web-services so that end-users or other external services are able to discover and link to EuDML items, or find semantically similar items. This way, EuDML has a chance to become a new major international player in the emerging landscape of scientific information discovery services, enabled for reuse in new added value chains – it might become half step towards a WDML, World-wide Digital Mathematics Library, or cornerstone of Linked Data of math content.

45

## Guidelines towards implementing Open Access policies

Clara Boavida<sup>1</sup> ; Eloy Rodrigues<sup>2</sup> ; Vasso Kalaitzi<sup>3</sup> ; Victoria Tsoukala<sup>3</sup>

<sup>1</sup> *UMinho*

<sup>2</sup> *Universidade do Minho*

<sup>3</sup> *EKT/NHRF*

**Corresponding Author(s):** [claraboavida@sdum.uminho.pt](mailto:claraboavida@sdum.uminho.pt)

The purpose of this poster is to present the ‘Guidelines towards implementing Open Access policies’ developed by the MedOANet project. These guidelines will serve as a basis for the policy-makers and other key stakeholders in six Mediterranean countries (France, Greece, Italy, Portugal, Spain and Turkey).

The MedOANet (Mediterranean Open Access Network) addresses the necessity for coordination strategies and policies in OA to scientific information in Europe. Taking into consideration the European Commission’s recommendations, the guidelines towards implementing OA policies are directed to top policy-makers and policy stakeholders (e.g. national science councils, research funding bodies, ministries, university administration, etc.) in view of facilitating the processes of development and coordination of OA in the six aforementioned Mediterranean countries.

The guidelines include: an overview of the European Commission’s recommendations; the current state of OA and policy developments in the six countries; effective practices and policies in other European countries; the approach to be followed for the successful implementation of policies nationally and institutionally; model policies and implementation plans that can be used as basis for customization at national and/or institutional level, and other resources and information sources.

Plenary 5 / 16

## Humanities Session: OA Research Monographs in HSS: Opportunities & Challenges

HSS scholars have been generally slower and, arguably, more resistant to OA publications than those in other disciplines. This presentation will look at some of the differences in publishing requirements between scientific and HSS scholars, alternative publishing models being developed to address these, and some of the opportunities OA publishing present for HSS disciplines. Specific attention will be given to monograph publication throughout the presentation.

## Plenary 5 / 17

**Humanities Session: The Humanities in and for the Digital Age**

The spread of digital technologies has presented scholars in the humanities with some extraordinary opportunities, as well as a few challenges, not least for their modes of communicating with one another. This talk will explore some of the changes taking place in the humanities today and their implications for scholars and their institutions. How are humanists' ways of thinking about scholarly communication changing as we do more and more of our work on digital platforms?

61

**Hydra: open, flexible workflows and community for open access to digital collections**Chris Awre<sup>1</sup><sup>1</sup> *University of Hull*

The Hydra project (<http://projecthydra.org>) originated in a collaborative effort between the University of Hull, University of Virginia and Stanford University to create flexible workflows and infrastructure to support the management of different types of digital content. Using the Fedora system as a basis, Hydra enables the implementation of multiple workflows over the core body of content, each designed to meet the needs of the content being processed (separate Hydra 'heads').

A core use case for a number of members of the community is open access to research publications, and research data. The ability to offer different workflows allows different types of content to be managed alongside each other in a way that suits that content albeit processed and/or displayed to suit local needs. Whilst focusing on open access, Hydra also provides granular access control to manage embargoes and other needs for controlled access to meet policy requirements.

This poster will describe the Hydra project and software, and detail the flexibility built in to the system that allows digital collections to be processed and presented so as to facilitate scholarly communication in a way that best suits that content and the audience using it.

55

**Jisc Gateway for Higher Education**Mark MacGillivray<sup>1</sup><sup>1</sup> C**Corresponding Author(s):** [mark@cottagelabs.com](mailto:mark@cottagelabs.com)

G4HE is a JISC-funded project which aims to engage with the BIS-funded RCUK Gateway to Research (GtR) initiative to improve the information exchange between Higher Education Institutions and the Research Councils.

The project is developing tools and interfaces, based on validated use cases shown to have specific and demonstrable value to HEIs, to allow both human and machine access to research grant meta-data.

We also intend to recommend ways in which the metadata could better be shared between institutions and research councils, and ways in which the dataset could be improved by either addition, cleaning or enhancement.

Use case and data quality analysis are nearly complete, and we will soon begin development towards the use cases most suited to the current state of the data; these (paraphrased) cases are:

- Generate reports about key collaborations within an institutional grant portfolio, to identify new institutions with whom to build partnerships
- Benchmark funding awarded between groups of researchers at different institutions for performance evaluation
- Visualise collaborations in particular fields and identify institutions doing similar work with whom collaborations are not yet in place.

We would like the opportunity to demonstrate work to date via presentation of our project poster at OAI8

34

## Keynote

63

## Make Knowledge Free [Flooved]

Nicolas Philippe<sup>1</sup>

<sup>1</sup> *Flooved*

Flooved breaks down the cost of the dissemination of knowledge by making studying resources free for undergraduates across the world. This is particularly important for students at universities that do not have a huge budget to put into their libraries, both in the UK and in emerging countries.

35

## MedOANet Open Access Tracker

Ilaria Fava<sup>1</sup> ; Paola Gargiulo<sup>1</sup>

<sup>1</sup> *CINECA*

**Corresponding Author(s):** i.fava@ceneca.it

The Open Access Tracker is an online tool that helps document the growth of open access in six Mediterranean Countries, Greece, Turkey, Italy, France, Spain, Portugal (<http://www.medoanet.eu/browse>). It has been developed within the project MedOANet ([www.medoanet.eu](http://www.medoanet.eu)), a two-year project funded under the Science in Society Programme of the EC FP7. The OA Tracker provides a real time overview of the state of the art and contributes to evidence-based advocacy and policy-making. The OA Tracker brings together information on journals, repositories, institutional policies, funder's policies and publishers' self-archiving policies, representing Open Access activities in the six countries based on the information and data registered in the selected authoritative International sources. The

OA Tracker as an interoperable tool, also allows updated information to be exported to international authoritative sources, and likely to be expanded to other kind of resources. The Tracker aims at encouraging the stakeholder community to register its open access resources with these established services, in order to increase the accuracy of information regarding the state of open access in the six countries. Such a tool could be adopted in another geographic areas to advocate OA venues and policies.

Plenary 2 / 3

## **Metrics Session: An overview of scholarly impact metrics**

The interest in developing scholarly impact metrics is frequently justified by the need to objectively prioritize scarce resources and to better manage scholarly productivity. However, the study of scholarly communication in general, including scholarly impact metrics, has significant relevance to a number of other scientific domains such as computational social science, social network analysis, web science, and complex systems. In this presentation I will provide an overview of established scholarly impact metrics, grounding each in their respective scientific traditions and backgrounds. Changes in scholarly communication patterns, including the move to online environments and the increasing use of social media, have recently prompted a Cambrian explosion of new impact metrics derived from new data sources. These metrics may reflect previously unexplored facets of scholarly communication and impact, and may thus yield a more complete picture of scholarly communication. In my presentation I will provide an overview of these new metrics, and identify the opportunities as well as challenges that they present.

Plenary 2 / 6

## **Metrics Session: Assessing the transparency of peer review in (Open Access) journals**

The peer review system is a standard control mechanism in scholarly publishing. Despite its centrality in science, peer-review typically takes place behind closed curtains. For both established and Open Access journals this obscurity may give room to substandard peer-reviews. I contend that transparency concerning the peer-review process at academic journals can be viewed as an indicator of the quality of the peer review. I present results of three studies of the validity and reliability of a straightforward online assessment of transparency of the peer review process at (OA) journals. The assessment can be readily used by different stakeholders (publishers, researchers, librarians, and funders) and entails a list of criteria for transparency concerning the peer review process (e.g., clarity on scope of the journal, rejection rates, decision makers, criteria used by reviewers, publication ethics). Results show good validity and sufficient reliability to use the tool to determine which journals meet common standards of transparent peer-review.

Plenary 2 / 4

## **Metrics Session: Discussions of scholarly articles online: who, why and where**

The field of altmetrics is based on the online activity around articles. How much of this exists, where is it happening and who is involved? Altmetric.com has been tracking this data for publishers,

funderson and institutions since July 2011 and in this talk we'll make some general observations and highlight interesting trends.

48

## National and institutional open access advocacy campaigns to reach out to research communities

Iryna Kuchma<sup>1</sup>

<sup>1</sup> E

**Corresponding Author(s):** iryna.kuchma@eifl.net

The poster presents case studies of national and institutional open access (OA) advocacy campaigns to reach out to research communities in Africa and Eastern Europe released by EIFL-OA Programme. They illustrate key achievements, activities, advocacy strategies, tactics and tools, success stories and lessons learnt, thus libraries everywhere can benefit. Some highlights include: students and OA: OA from the perspective of young researchers (Lithuania); promoting OA publishing: promoting OA through the implementation of Open Journal Systems for the KNUST research community (Ghana); advocacy of OA: Belarusian national OA website as a focal point for advocacy campaigns to reach out to research communities, Knowledge without Boundaries: Advocacy campaign in Kenya for OA and institutional repositories; Promotion of OA at the University of Belgrade (Serbia), OA Advocacy at Sudan Higher Education Institutions National Level; OA policies adoption and implementation: Promoting the adoption of campus-based, faculty-driven OA policies in Uganda, Capacity building for the adoption and implementation of OA policies and repositories in Zimbabwe, Jomo Kenyatta University of Agriculture and Technology OA Institutional Repository (Kenya), and OA – Optimizing Scholarly Communication in Moldova. The EIFL-OA Programme is supported by the Information Programme, Open Society Foundations.

59

## OPENAIREPLUS: 2nd Generation of Open Access Infrastructure for Research in Europe

Najla Rettberg<sup>1</sup>

<sup>1</sup> *Göttingen State and University Library*

The main goal of OpenAIREplus is to create a robust, participatory service for the cross-linking of peer-reviewed scientific publications and associated datasets. The project will establish an e-Infrastructure to harvest, enrich and store the metadata of Open Access scientific datasets and Innovative underlying technical structures will be deployed to support the management of and inter-linking between associated scientific data.

More specifically, it aims at growing a richer graph of data, covering material from all research disciplines and further countries, and including projects from national funding schemes, non-peer-reviewed publications, and research datasets. Ultimately, OpenAIRE will extend its service to push beyond the realm of publications within FP7 and to widen the publication scope, but also to link publications to associated research data. This interlinking and reuse of research is seen as crucial to growth and innovation as Europe heads towards Horizon2020 and supporting an infrastructure for Open Science, for which Interoperability is the key.

As substantial technical infrastructures continue to develop, it is clear that this growth has to be complemented by networking activities that give out clear, simple messages, and support the very people the system is built for: scholars and researchers.



62

## OPENAIREPLUS: Supporting Repository Interoperability through Guidelines

Najla Rettberg<sup>1</sup><sup>1</sup> *Georg-August University of Göttingen*

Supporting the open access policy of the European Commission, OpenAIRE is moving from a publication infrastructure to a more comprehensive infrastructure that covers all types of scientific output, funded by the

European Commission, and widening to other European funding streams. It harvests content from a range of

European repositories, and ensures raised visibility of valuable open access content, as well as links to project and funding information. In order to ensure interoperability from these research infrastructures, a common approach is need to adhere to existing and future guidelines.

In this context, an integrated suite of guidelines have been developed. The poster will briefly outline the OpenAIRE Guidelines: Guidelines for Data Archive Managers, for Literature Repository Managers and for CRIS Managers.

By implementing all three sets of the OpenAIRE Guidelines, repository managers will be able to enable authors who deposit publications in their repository to fulfill the EC Open Access requirements, as well as the requirements of other (national or international) funders with whom OpenAIRE cooperates.

In addition it will allow the OpenAIRE infrastructure to add value-added services such as discoverability and linking, and creation of enhanced publications.

41

## Open Access for some, Access and Open for all others

John W. Miescher<sup>1</sup><sup>1</sup> *Bizgraphic*

All others are the users of your output and they want:

- easy finding on the web
- easy re-finding in their knowledge base
- easy reading and navigation
- easy discovery of relevant information
- easy extraction of texts and key data
- easy citation copying and tracking

What scholars, researchers and other users really expect:

To quickly find, re-find and review information - to be able to build their own knowledge base and to facilitate their research work by organising citations and references from stored documents with fast-open links.

They want “well-behaved” documents that are user-friendly for reading AND library-friendly with embedded meta data for easy classification.

And they need tools that help them to organizes the mass of downloaded, virtual and physical books and documents

There is an app for that

Content providers' challenge: User orientation

Your content: Provide what users want most both in terms of content and usability.

User-friendly ...

Library-friendly...

as publisher: Please make sure to embed descriptive metadata in every PDF file you publish offer free tools: give away the free digi-libris Reader together with your content.

Visual and short description of the app.

49

## Open Access in Arab States-

Mandy Taha Abdou<sup>1</sup>

<sup>1</sup> *L*

**Corresponding Author(s):** mandy.taha@bibalex.org

The Poster will present an overview about the Open Access movement in Arab States, shedding the light on its historical back ground, present status and future perspectives.

53

## Open Research Core (ROC)

Chris Chapman<sup>1</sup>

<sup>1</sup> *Pentandra Research Solutions, Inc.*

**Corresponding Author(s):** chris@pentandra.com

The Open Research Core, or ROC for short, is a lightweight open-source framework for the development of domain-specific collaborative research systems. Using an innovative approach to research collaboration, the entire research process is preserved in what are termed “cases.” A case starts with a research question and ends with a conclusion, and keeps track not only of which data was used in analysis, but also how that data was used to answer the question. Although a research methodology called case study research makes use of case studies, cases in ROC are a broader concept that can support any type of hypothesis-driven research. ROC is built upon open web and knowledge management standards, including the semantic web, the Open Annotation Model and specifications from the Open Archives Initiative, and is designed to be used as a starting point upon which to build domain-specific research systems. ROC is currently in heavy development and has been piloted in the development of a large collaborative research system in the domain of generational history.

43

## Palgrave Pivot - Breaking Boundaries in Scholarly Publishing

Carrie Calder<sup>1</sup>

<sup>1</sup> *Palgrave Macmillan*

The evolution of the digital age has led to significant developments within the publishing industry. But what other changes should take place?

Palgrave Macmillan has undertaken a number of surveys to explore publishing consumption behaviours. Questioning over 1000 researchers across humanities and social science (HSS), the surveys

reveal insights into researchers' publishing requirements, as well as current publishing boundaries. In examining the results, we started to question the traditional - from the dominance of the widely-accepted formats of articles and monographs, to interdisciplinary research, publication times, pricing flexibility and beyond.

This presentation will discuss these issues in detail. It will demonstrate how listening to the community has brought about change at Palgrave Macmillan in the evolution of a new 'format' for academic output within HSS, Palgrave Pivot. This allows scholarly works to publish just 12 weeks after submission - getting research to the community faster than ever before.

In addition to this Palgrave Macmillan have recently announced that there will be an Open Access option across all publication formats - journal articles, monographs and Palgrave Pivot publications.

The presentation will encourage the conference delegates and all stakeholders to work collaboratively in breaking more of the boundaries of publishing and scholarly research together.

Plenary 6 / 21

## Panel Session: Gold OA Infrastructure

36

### **Pundit. Augment Digital Libraries contents with semantically structured annotations**

Alessio Piccioli<sup>1</sup> ; Christian Morbidoni<sup>2</sup> ; Francesca Di Donato<sup>3</sup> ; Simone Fonda<sup>1</sup>

<sup>1</sup> *Net7*

<sup>2</sup> *Università politecnica delle Marche*

<sup>3</sup> *Scuola Normale Superiore / Net7*

**Corresponding Author(s):** francesca.didonato@sns.it

Pundit (thepund.it) is a client-server web application that allows users to annotate web contents, collecting records in notebooks and sharing them with others. The main idea behind Pundit is to allow users to comment on or mark fragments of web pages and to create semantically structured data during the annotation process, thus enriching the Web of data. Annotations span from simple comments to links to the Web of data or controlled vocabularies created ad hoc, to fine granular cross-references and citations. Pundit stores the annotations on a server using RDF technologies that provides APIs that allow to use and create annotations even by machines, thus making it possible inference of new knowledge and allowing to make very complex queries. The ability to express semantically typed relations between resources not only allows users to express univocally the semantic of their annotations, but also facilitates the reuse of knowledge within other web applications. In the poster will be explained the annotation workflow, showing some real-world Pundit use cases in the field of the Digital Libraries. Thereafter, will be presented some experiences of reuse of the data annotated through Pundit by other apps which offer rich visualizations of data.

58

### **QUT Research Data Finder: An institutional metadata repository**

**Author(s):** Stephanie Bradbury<sup>1</sup>

**Co-author(s):** Gawri Edussuriya ; Philippa Broadley <sup>1</sup>

<sup>1</sup> QUT

QUT's new metadata repository (data registry), Research Data Finder, has been designed to promote the visibility and discoverability of QUT research datasets. Funded by the Australian National Data Service (ANDS), it will provide a qualitative snapshot of research data outputs created or collected by members of the QUT research community that are available via open or mediated or closed access.

51

## Re-writing of Scholarly Communication

Cristina Marques Gomes<sup>1</sup>

<sup>1</sup> *Universidade Federal do Estado do Rio de Janeiro - Brazil*

Traditionally, the tripod 'research/system/society' (RSS) has been the basis of the scholarly communication (SC) conception. Nevertheless, many technological transformations have been taking place in this field, as well as in society at large, over the last decades. The present research is focused on the emerging research problem: is the notion of SC related to the tripod RSS still consistent and / or sufficient to explain this phenomenon nowadays? If not, what other stratas should a contemporary concept of the SC embed? As a way to answer these questions, this study outlines a SC cartography from an holistic approach. The analysis is grounded on a praxis perspective - whereas 12 UE projects had been scrutinized through multiple qualitative and quantitative methodological procedures. What this thesis brings forth as fruits can be cited, in light of SC discussions and debates, the fact it adds the contributions of the reconstruction of the phenomenon's trajectory and, not confirming the main hypothesis of the thesis, points at the end, some of the basic principles that can guide a possible re-writing of scholarly communication.

Plenary 4 / 13

## Res. Data Session: Interoperability of Research Data

The advent of the new paradigm of science largely based on data analysis and mining is having a relevant impact on Scholarly Communication. In order to document their work scholars are starting to publish not only research papers but also to make available the experimental datasets and the tools used for achieving such results. The availability of these products facilitates the reproducibility of the results and open the way towards a more wider re-usability of the data for other scientific purposes. However, In order to fully exploit these new possibilities appropriate level data interoperability between the publisher and the consumed of the data have to be assured. This presentation introduces major issues of data interoperability between a data publisher and a data consumer and it discusses how emerging data infrastructures can help in minimizing them.

Plenary 4 / 14

## Res. Data Session: Quality and curation of Research Data

Everyone who wants data wants high-quality data, and curation processes are designed to improve data quality. Since we are all agreed on these things there should be little to discuss. But in practice we use 'quality' to mean different things, and our curation processes emphasise some quality dimensions at the expense of others. This is not always beneficial to the research process.

I will discuss what research tells us about data quality and how this should inform curation practice. Finally I will speculate on how metadata on quality could ease reliable automated data integration and hence promote data reuse.

#### Plenary 4 / 12

### Res. Data Session: Research Data Policies: Seachange or Zeitgeist?

What will be the practical implications of research data policies for the everyday life of researchers or institutional and commercial service portfolios? The question is not new: the German Research Foundation DFG addressed data management as part of the recommendations for safeguarding good scientific practice in 1998 [1], the National Institute of Health NIH in the US presented a draft statement on sharing research data in 2002 and concepts like "Data Deluge" [3] celebrate a 10th anniversary. Recently, the National Science Foundation NSF in the US introduced a requirement for data management plans [4], UK's Engineering and Physical Sciences Research Council EPSRC asks universities to deliver a data roadmap for implementation by 2015 [5] and institutions such as the University of Edinburgh start publishing local policies [6].

At the same time, disciplinary practice varies significantly. Some life science journals require mandatory data deposits with an article publication. Data journals as well as data publishing platforms are appearing. In other disciplines even mentioning the term 'data' causes suspicion. The presentation will analyze selected research data policies in the light of examples of current research practice.

[1] [http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/self\\_regulation\\_98.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/self_regulation_98.pdf)

[2] <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-02-035.html>

[3] [http://eprints.soton.ac.uk/257648/1/The\\_Data\\_Deluge.pdf](http://eprints.soton.ac.uk/257648/1/The_Data_Deluge.pdf)

[4] <http://www.nsf.gov/eng/general/dmp.jsp>

[5] <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx>

[6] <http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy>

#### Plenary 4 / 15

### Res. Data Session: Working with large data sets

Whether large or small, data sets need to be managed. Scale however makes standard data operations more challenging, especially when the data sets expand beyond the capacities of a single data centre. Consequently replication, migration and archival require optimised, sometimes domain specific solutions. Tim will illustrate the HEP approach by describing how CERN collected 100 petabytes of research data and how it organises storage and access by researchers across the globe. He will run through the data reduction and analysis chains up to paper production, and finish with the challenges of archiving these large data sets.

## Semantic Linkages in Research Information Systems as a New Data Source for Scientometric Studies

**Author(s):** Sergey Parinov<sup>1</sup>

**Co-author(s):** Mikhail Kogalovsky<sup>2</sup>

<sup>1</sup> *CEMI RAS*

<sup>2</sup> *Market Economy Institute of RAS*

**Corresponding Author(s):** sparinov@gmail.com

A modern semantic linkage technique allows scientists to express their knowledge, hypotheses and opinions about relationships between any pair of available information objects from a content of research information systems where the technique is implemented. Scientific ontologies and semantic vocabularies are used as taxonomy to classify in computer readable form known classes of scientific relationships that can exist between research objects. Making semantic linkages between research objects scientists provide some kind of professional tagging and folksonomy over research data and information space. Based on knowledge how semantic linkages are stored in research information systems we propose an abstract model of this data source. This model allows a better understanding what new indicators can be designed for scientometric studies. In the poster we also present some experiments with this new data source to build scientometric indicators.

**Plenary 3 / 10**

### Semantic Session: Detecting knowledge-level claims in research articles

The research article genre has developed a specific structure for the effective communication of scientific results. The title, the abstract, more or less standardized section types, etc. are all structural elements in the service of facilitating comprehension and searchability. These elements are apparent for the reader through formatting, and their markup makes it possible for search algorithms to take advantage of them in relevance ranking. In this presentation we propose another type of content element that can facilitate comprehension and search: knowledge-level claims. We define knowledge-level claims as discourse elements in articles that indicate the status of scientific propositions within the state of the art. Knowledge-level claims significantly contribute to comprehension, they are usually rhetorically salient, but traditionally they are not made prominent through formatting or markup. We show some applications where the automatic detection of knowledge-level claims has been used for enhancing both comprehension and search, and we indicate some further potential applications.

**Plenary 3 / 8**

### Semantic Session: Semantic indexing in PubMed

About 5,000 biomedical journals are indexed and included in the National Library of Medicine's MEDLINE bibliographic database, available through PubMed. MEDLINE is used internationally to provide access to the world's biomedical journal literature. PubMed supports both text searches and searches based on the indexing of the articles in reference to the Medical Subject Headings (MeSH) thesaurus.

We will review several aspects of semantic indexing in MEDLINE, including traditional MeSH indexing performed manually by human indexers, indexing to UMLS concepts, and indexing of relations or facts ("nano-publications"). We will discuss automatic approaches to MeSH indexing, as

well as indexing of specific entities, such as genes. This presentation also explores the relations between semantic indexing and the Semantic Web, as well as applications of semantic indexing, e.g., to literature-based discovery.

### Plenary 3 / 9

## **Semantic Session: Transformation of keyword indexed collections into semantic repositories**

In the information retrieval context, resource collections are frequently classified using simple knowledge models such as thesauri. However, the limited semantics provided restricts their search and browsing capabilities. This work shows a process that improves these capabilities through the conversion of the selected knowledge model into a domain ontology. The process has been tested with the European Urban Knowledge Network and the Urbamet thesauri. Additionally, Urbamet model has been used to create an atlas of urban related resources with advanced search capabilities.

### Plenary 3 / 11

## **Small Data, or: Bridging the Gap Between Smart and Dumb Research Repositories**

Scientific research mostly consists of many tiny niches, with many thousands of small data sets: a 'long tail' effect. So we have a 'Small Data' problem: how do we connect vastly different experimental results, so that they can be used by other scientists? Currently, there are many large, topically agnostic repositories, requiring little metadata or informatics support, which serve an archival need but provide little opportunity for allowing overarching analytics. On the other end of the scale, highly usable topical repositories require painstaking manual curation, which does not scale. This talk will present a proposal on bridging the chasm between these two approaches, to enable systems for interoperable results reporting. After presenting a general overview of some pertinent developments I'll focus on two use cases, in electrophysiology and geochemistry, where we will attempt to build a system that allows bridging the gap between such 'big and dumb' and 'small and smart' solutions.

### Tutorials / 22

## **T1 - Metrics (Room R160)**

Altmetrics is a hot buzzword. What does it mean? What's behind the buzz? What are the benefits and risks of including alternative metrics of research impact in our discovery and evaluation systems? What altmetrics tools exist today, what are their strengths and weaknesses, and where is the field going?

Join Heather Piwowar, cofounder of ImpactStory, for this tour of the altmetrics landscape. The session will be relevant to anyone who produces, publishes, or evaluates research: funders, university administrators, journal and repository leaders, and individual scholars.

If you can, bring a laptop to play along during the session. Do you have research products you'd like to experiment with? Prep a digital list of IDs (DOIs, PMIDs, URLs, researcher ORCIDs, etc) and we'll see what we can discover!

**Tutorials / 23****T2 - Metadata for the Research Lifecycle (Room M6289)**

The management and publication of research data is increasingly important, and relies heavily upon appropriate descriptive metadata. The tutorial will provide a walkthrough of the research data management life cycle, from planning through data acquisition to data publication and preservation. Each of these stages requires different tools and methods, for example DMPonline, DataStage, DataBank and DataCite. However, some things stay the same: for example, person, subject, method, instrument, project and funder are likely to be constant throughout the research lifecycle, and metadata describing these can be re-used in different contexts. If tools are connected appropriately, researchers could work in a dynamic web-based data management environment to which they could regularly return, confident that metadata that had already been entered would be available for reuse, and that datasets they describe would be securely backed up.

This tutorial is intended for data service developers and managers. Participants will be expected to adopt a researcher's perspective, and to identify aspects of research data management that can simplify the researcher's life. Overviews of available tools and methods for each stage in the research life cycle – planning, acquisition, publication and preservation – will be alternated with hands-on sessions employing existing tools for handling exemplary data and metadata, including the participants' own data. The use of ontologies and RDF metadata representations will be explored. Participants will be invited to suggest additional tools and use cases. Details will be provided in due course.

**Tutorials / 24****T3 - Metadata: from Records to Graphs (Room R150)**

The tutorial will introduce traditional as well as state of the art approaches to modeling information objects and the processing environments they are part of, both in traditional memory institutions and in the Linked Data web. The central issue is to understand what is changing with RDF-graph based approaches to metadata generation and management. We will do some hands on work on semantic annotation as part of the workshop using the Pundit tool (<http://thepundit.it>)

**Tutorials / 25****T4 - OJS, beyond editorial tradition (Room 1150)**

This tutorial aims to address the following matters:

- I. A view upon the software package,
- II. Setting and organising the team (software and teamwork bootstrapping), and exploring possible workflows,
- III. Data concerning issues.

“A view upon the software package” will answer to the common questions like: what is it, who is behind it, is it suitable for what we do, how long it will take me to set it up, is it simple to use, are there many working with the package?

All these questions will find a set of answers backed with practical work on the software itself starting from installation easy steps up to the moment when we will have a fully workable setup.

“Setting and organizing the team and exploring possible workflows” will be one of the central points of the tutorial because it is also the power behind the long run efficiency of the journal. Certain aspects modeled by the different interaction levels with the software will be explored and some possible workflows will be discussed among the participants.



“Data concerning issues” will be the signal that will mark the ending of the tutorial making all aware of the present actions with regard to data management and where is the value of some possible actions concerning preserving it.

The attendants do not need to have programming skills, but to understand the fundamentals of operating an Operating System - GNU/Linux, Mac OSX or Windows.

Need to have on your belt

Participation to this tutorial requires a laptop and depending on you OS of choice the following software packages already installed:

GNU/Linux (LAMP installed + PhpMyAdmin)

Mac OSX:

- MAMP (<http://www.mamp.info/en/index.html>),
- XAMMP (<http://www.apachefriends.org/en/xampp-macosx.html>).

Windows (you may choose between):

- EasyPHP (<http://www.easyphp.org/save-easyphp-latest.php>);
- XAMPP (<http://www.apachefriends.org/en/xampp-windows.html>).

Download the latest OJS package (Current Development Release) from [http://pkp.sfu.ca/ojs\\_download](http://pkp.sfu.ca/ojs_download). Unarchive the package in htdocs or www directory of your preferred package.

Having all these set up before the tutorial is mandatory.

## Tutorials / 26

### T5 - The NISO/OAI ResourceSync Synchronization Framework (Room 1130)

This tutorial will provide an overview and a practical introduction to ResourceSync, a framework to synchronize web resources that consists of multiple modular capabilities that a server can selectively implement to enable third party systems to remain synchronized with its evolving resources. All capabilities leverage document formats introduced by the widely adopted Sitemap protocol.

The editors of the ResourceSync specification are affiliated with the Los Alamos National Laboratory, Cornell University, Old Dominion University, and the University of Michigan. They have been involved in other interoperability specification efforts, including the OAI Protocol for Metadata Harvesting, OAI Object Reuse and Exchange, Memento, and Open Annotation. An international Technical Committee has supported the editors in compiling the draft specification.

The tutorial will:

- Motivate the ResourceSync approach by outlining several synchronization use cases including scholarly article repositories, linked data knowledge bases, and resource aggregators.
- Detail the nature of the various ResourceSync capabilities (Resource List, Resource Dump, Change List, Change Dump) that a server can implement.
- Show how support for these capabilities can be expressed (Capability List) and discovered.
- Describe the extensibility mechanism built into the framework that allows addressing specific needs, such as synchronizing from mirror sites, synchronizing resources by exposing patch information, and synchronizing both metadata and content described by that metadata.
- Provide details about the serialization format used to express ResourceSync capabilities, which is based on document formats introduced by the Sitemap protocol.
- Describe experiences developing general ResourceSync software libraries and particular support for an institutional repository platform.

**Intended Audience:**

The intended audience are people involved in both technical and management aspects of digital repositories or in the creation of value-add services across such repositories. The tutorial will assume a basic level of familiarity with fundamental web concepts (URI, resource, representation) and XML, but will be presented in a way that is accessible to people with a non-technical job description that have basic technical knowledge.

**Tutorials / 27****T6 - Open Access Café 2013 (Room 1140)**

Following the success of this session in the OAI7 Workshop, we are pleased to offer again the chance to learn more about Open Access topics that you would like to understand further. We will present a session where you can mingle with various experts, ask them questions in a café-style setting and discuss issues in a very informal atmosphere. There will be time during the session for you to explore 4 or 5 different topics in which you are particularly interested. We will engage experts from across the spectrum in order to cover technical-, cultural- and policy-related issues.

Come along to the Open Access Café, meet people, talk about Open Access – and drink coffee, of course!

**Plenary 1 / 0****Tech Session: How semantic representations can support scholarly communication**

One part of scientific communication is to effectively structure information so that it can be easily consumed. We spend a great deal of effort doing this for humans, however, given the amount of scientific information we are producing we should also spend some effort doing this for machines. In this talk, I review several semantic representations that are making it easier for machines to consume scholarly content. Moreover, I discuss how these representations can facilitate the decoupling of the journal and the ability to remix scholarly content.

**Plenary 1 / 2****Tech Session: Naming on the Web: What scholars should want, and what they can have**

It's alleged that when Zhou Enlai was asked what he thought of the French Revolution he replied: "It's too early to tell". The scholarly community could be forgiven for wishing they could say the same about the Web, but we don't have that luxury. The incentives for moving scholarship, *all* scholarship, onto the Web are enormous, and the penalties for failing to do so are rapidly increasing—for the current generation of students, it is increasingly true that "if it's not on the Web, it doesn't exist". Yet how can responsible scholarship depend on such a manifestly uncertain technology? My colleague Michael Sperberg-McQueen once said, in his role as technical advisor to the Model Editions Partnership: "[W]hen I advise people on building systems that will last for the time spans needed for cultural heritage data, I will advise them to build on some system whose design story holds up for more than a minute and a half before an inconsistency is introduced." In this talk I'll try to separate the *necessary* properties of *any* web-scale naming system from the *contingent* socio-technical realities

of the Web as it is today. I'll close by attempting to make clear exactly what trade-offs confront us as we try to move scholarly discourse onto the Web in a responsible manner.

## Plenary 1 / 1

### **Tech Session: W3C Open Annotation effort: Status and Use Cases**

The Open Annotation Data Model specifies an interoperable framework for creating associations between related resources - Annotations - using a methodology that conforms to the Architecture of the World Wide Web. Open Annotations can easily be shared between platforms, with sufficient richness of expression to satisfy complex requirements while remaining simple enough to also allow for the most common use cases, such as attaching a piece of text to a single web resource.

In this presentation we will discuss the data model, motivated by scholarly communication use cases for annotation. These include open peer review, nano-publications, personal note-taking, teaching and learning systems, resource organization and many others. We will also discuss the current status of the work, and the next steps in the standardization process.

42

### **The National Bibliography Number Italia (NBN:IT) Project. A persistent identifier supporting national legal deposit for digital resources**

Jordan PIŠČANC<sup>1</sup>

<sup>1</sup> U

**Corresponding Author(s):** piscanc@units.it

The association of a Persistent Identifier to a digital resource is a best practice helping in solving a reliable information resources retrieving on the web. At the present moment, different kind of technologies and standards such as DOI, Handle are establishing. However, within scientific and cultural communities, the main need is still focused on ensuring resources authenticity, provenance, and preservation in the long term period. For this reason, a third authoritative party such as the National Library is necessary, managing identifiers and process requirements for preservation aims. The NationalBibliographyNumber originated in the context of the role and the responsibilities of national libraries in the field of universal bibliographic control, oriented to the identification of resources to be preserved in the long period. In Italy NBN:IT is going to be linked to the legal deposit, and activity traditionally up to the national libraries guaranteeing its trustability. The project, coordinated by the national libraries of Florence, Rome, and Venice, by the Foundation Rinascimento Digitale and by the Conference of Rectors of the Italian Universities, completes the current offer of PI. Today we have assigned almost 10.000 NBN:IT. Plugins for the main repository software as DSpace, Eprints, Ojs were released.

37

### **The Social Science Open Access Repository (SSOAR) as an element of a greater research information structure**

Agathe Gebert<sup>1</sup>

<sup>1</sup> *GESIS***Corresponding Author(s):** [agathe.gebert@gesis.org](mailto:agathe.gebert@gesis.org)

SSOAR is one of the 4 big disciplinary Open Access Repositories in Germany covering the Social Sciences. It is maintained by the Specialized Information Unit of the Leibniz-Institute for the Social Sciences. Unlike in the STM-Sector preprints do not play a role in the social science discourse. Self-archiving of either pre- or postprints is thus not common practice in the social sciences. Therefore in order to obtain a sufficient amount of high quality publications that is appealing to scholars SSOAR acquires content through a broad network of partners. Among them academic medium enterprise publishing houses relevant for the Social Sciences who SSOAR offers services in exchange for Open-Access licenses. SSOAR follows the Green Road of Open Access and is an instrument of secondary publishing. On the long run SSOAR strives to be the central entrance into all socio-scientific literature (within the scope of SSOAR) that is freely available to enhance its worldwide visibility as well as its cross linking with other data and information available online.

64

## Towards a common map of academia

Lukasz Bolikowski<sup>1</sup><sup>1</sup> *U*

At Centre for Open Science, ICM UW, we are constructing a comprehensive map of academic community: a large-scale graph of semantic relations between people, documents, institutions, venues (journals, conferences, blogs), topics, funding sources, and more. The graph is being built using publicly available resources: web pages from Common Crawl, OAI-PMH compliant repositories, OA subset of PubMed Central, open databases such as DBLP and DOAJ, or metadata from the NPG Linked Data Platform. Our aim is *not* to create yet another service competing with Google Scholar, Microsoft Academic Search or Web of Science, but rather to build and maintain a common, publicly available resource that will be used by a wide spectrum of users: scientometrists, data researchers, and software developers (bibliographic databases, digital libraries, research information systems... and academic search engines).

52

## UAM institutional repository: use by scientists from the most important university field of research

Marisa Perez Aliende<sup>1</sup><sup>1</sup> *UAM***Corresponding Author(s):** [mp.aliende@uam.es](mailto:mp.aliende@uam.es)

According to the I-UGR, Spanish Universities Ranking by broad subject fields and scientific subjects, the Autonomous University of Madrid (UAM) for the 2007-2011 period head the second and best position by broad fields in Biological Sciences, and the third in Medicine and Physics. It's considered the second university of Spain in Biochemistry, Cell & Molecular Biology subject for the same period of time; however for a wider period of time such as 2002-2011 it's considered the best University of Spain researching on this subject. This information can also be obtained from the Incites, Subject Ranking UAM report for Biochemistry & Molecular Biology, being the subject with more documents web of science generated and most cited, followed by Cell Biology.

Two UAM departments, Biochemistry department and Molecular Biology department produce most of the research papers on Biochemistry, Cell & Molecular Biology. This poster is going to describe the typology of research papers generated by these two departments and ingested in the IR; if they

are among the most downloaded objects from the repository, trying to obtain a feedback from these researchers about the open access policy, the IR use, or their preferences between institutional or any other subject repository.

54

## VIVO - IR population and integration

Graham Triggs<sup>1</sup>

<sup>1</sup> S

**Corresponding Author(s):** graham@symplectic.co.uk

Making the population of institutional repositories and VIVO discovery systems much easier and faster, but also leveraging the power of both by making connections between them.

46

## Working towards a global open access repositories network for a networked scholarship

Alicia Lopez Medina<sup>1</sup> ; Katharina MUELLER<sup>1</sup>

<sup>1</sup> COAR

**Corresponding Author(s):** katharina.mueller@sub.uni-goettingen.de

The World Confederation of Open Access Repositories (COAR) is a young, fast growing association of repository initiatives and networks uniting and representing almost 100 institutions worldwide (Europe, Asia, Latin and Northern America). Our mission is to facilitate greater visibility and application of research outputs through global networks of open access repositories based on interoperability and international cooperation. Activities are guided by four strategic directions:

1. To promote the development of a global, sustainable network of open access repositories as key element of the evolving international scientific information infrastructure, including social sciences and humanities, in support of excellence in research and education
2. To provide support for the OA repository community from institutions, countries, regions and disciplines, through international cooperation and information sharing
3. To define and promote interoperability, standards, and infrastructure policies
4. To stimulate the development and take-up of user added-value services on top of the repositories

The poster presents COAR's aims, activities and projects in order to promote opportunities for collaboration and presents the benefits of membership in COAR for institutions especially, but not only in developing countries in terms of support, visibility and integration into the global knowledge network.

44

## ZENODO - A new innovative service for sharing all research outputs.

Christopher C. Erdmann<sup>1</sup> ; Lars Holm Nielsen<sup>2</sup>

<sup>1</sup> *Harvard-Smithsonian Center for Astrophysics*

<sup>2</sup> *CERN*

**Corresponding Author(s):** lars.holm.nielsen@cern.ch

This poster will present ZENODO, a new simple and innovative service that enables researchers, scientists, EU projects and institutions to share and showcase multidisciplinary research results (data and publications) that are not part of the existing institutional or subject-based repositories of the research communities.

ZENODO enables researchers, scientists, EU projects and institutions to 1) easily share the long tail of small research results in a wide variety of formats including text, spreadsheets, audio, video, and images across all fields of science 2) display and curate their research results and get credited by making the research results citable and integrate them into existing reporting lines to funding agencies like the European Commission. 3) easily access and reuse shared research results.

The poster will highlight use-cases of ZENODO by The John G. Wolbach Library at Harvard-Smithsonian Center for Astrophysics as well as the European Middleware Initiative (EU FP7 project). ZENODO is developed under the EU FP7 project OpenAIREplus (grant agreement no. 283595) by CERN based on Invenio.

66

## **bwFLA - a Practical Approach to Functional Access Strategies**

Annette Strauch<sup>1</sup>

<sup>1</sup> *University of Ulm*

This is about organization of knowledge today. Sustainability of data, process-oriented. bwFLA as an innovation tool and plays a vital role in scholarly communication right now re. the wider context of research data management and cloud-based services around the globe.

The project bwFLA (functional digital preservation).

bwFLA has developed a functional approach to Digital Preservation with an emphasis on workflows.

bwFLA deals with digital preservation and sharing of reliable data, preservation of computer systems and their emulators. System environments play an important role here.

60

## **re3data.org - A Registry of Research Data Repositories**

**Author(s):** Maxi Kindling<sup>1</sup>

**Co-author(s):** Frank Scholze ; Heinz Pampel ; Paul Vierkant

<sup>1</sup> *Humboldt University of Berlin*

The aim of re3data.org is to create a webbased registry of research data repositories. The re3data.org registry index data repositories to provide an overview on existing research data collections. In March 2013,

343 research data repositories were indexed in re3data.org. 188 of them are described using a comprehensive vocabulary (Vierkant et al., 2012).

The re3data.org registry makes it possible to identify a certified research data repository which allows for instance the storage of geoscience data under a Creative Commons license. Furthermore a system of icons has been developed to show crucial characteristics of a repository. The icon system helps users to identify a suitable repository for the storage of their data. In re3data.org researchers can clearly see the terms of access and use of each repository. re3data.org is a joint project of the Karlsruhe Institute of Technology (KIT), the German Research Centre for Geosciences (GFZ) and the Berlin School of Library and Information Science at the HumboldtUniversität zu Berlin.

Funded by the German Research Foundation (DFG) re3data.org cooperates with The German Initiative for Network Information (DINI).

The poster gives an overview of the project and current features of re3data.org