



Open Data Foundation Europe 2008

UKDA, University of Essex, Colchester, UK

14-15 April, 2008

Meeting Summary

Participants:

Nikos Askitas	IZA, Germany
Stefan Bender	IAB, Germany
Louise Corti	UKDA, UK
Raymond Currie	RDC Network, Canada
Conny Dippold	FDZ-RV, Germany
Bryan Fitzpatrick	ODaF, UK
Jan Goebel	DIW Berlin, Germany
Rune Gloersen	SSB, Norway
Arofan Gregory	ODaF, USA
Pascal Heus	ODaF, USA
Chuck Humphrey	RDC Network, Canada
Jannik Jensen	DDA, Denmark
Herve Lhours	UKDA, UK
Ken Miller	UKDA, UK
Dana Mueller	IAB, Germany
Jostein Ryssevik	ODaF, Norway
Laurent Sesink	DANS, Netherlands
Kevin Schurer	UKDA, UK
Rene van Horik	DANS, Netherlands
Mary Vardigan	ICPSR, USA
Joachim Wackerow	GESIS/ZUMA, Germany
Andrew Westlake	SSC, UK
Matthew Woollard	UKDA, UK
Melanie Wright	UKDA, UK
Mathias Zenke	DESTATIS, Germany

This document does not capture all of the discussion which took place during the two days of meetings, but does attempt to record decisions made and actions which people have committed to as a result of discussions.



April 14th (Day 1)

The first order of business was to thank our hosts at the UK Data Archive who provided the facilities for the meeting. Introductions were performed, with each attendee giving a brief statement of their background, their interest in the organization and reasons for attending.

The agenda for the first day of the meeting covered two areas: an introduction to ODaF for those not familiar with the activities of the organization and a series of presentations made by participants. The topics covered can be seen in the agenda document, and the presentations are available on the ODaF website¹. Rather than paraphrase here all that was presented, please see these more complete resources. The coverage in these notes will be kept very light.

Many topics for further discussion did come up during the various presentations, however, and were addressed on the second day. These will be noted.

The morning's agenda was taken up with a general introduction of the history and reasons for establishing ODaF in 2006, with an explanation of the general view that various constituencies could be better served in their use of technology by communicating and establishing open-source tools and best practices in relation to the various open standards now extant in the fields of research and statistical data and metadata.

This was followed by a more specific report on the recent activities of the organization, including work on the ongoing projects and proposals ODaF is involved in:

DeXT²: This project is being driven by the UKDA, and is part of their participation in the DDI Foundation Tools Program. Prototypes of stand-alone transformations have been developed and are soon to be finalized, reading SPSS files and creating DDI 3.0, and then transforming this back into SAS, Stata or SPSS.

QuDEX: This is an initial draft of an XML standard for capturing the metadata output of CAQDAS packages for documenting qualitative data in text and other media. It is part of the overall DEX project at UKDA. ODaF acted in an advisory capacity helping to advise on the XML schemas. It is anticipated that this will become part of the DDI in future versions. (See Louise Corti's presentation for more complete information.)

¹ http://opendatafoundation.org/events/odaf_europe_2008.php

² <http://tools.ddialliance.org/?lv11=product&lv12=dext>



DDI Foundation Tools Program³: This is a project involving several organizations for working collaboratively to develop a set of utilities to make the creation of DDI 3.0 applications easier. All products will be open-source. Participants include ODaF (infrastructure and technical coordination), the DDI Alliance (administering funds), the Danish Data Archive (DDA), GESIS-ZUMA, NORC, the Canadian Research Data Centre Network, and the UK Data Archive. Contributions are monetary or in-kind.

NORC's Virtual Data Enclave⁴: This is a project whereby sensitive datasets from US agencies (such as NIST, the USDA or the Kauffman Foundation) are being made available to researchers in a secure virtual on-line environment, to overcome the challenges of having physical RDCs and improve on the researcher collaborative efforts. ODaF is being involved in an advisory capacity to support the metadata management framework and the implementation of tools for producers and researchers.

Canadian RDC Network: This is an on-going project to connect the many Research Data centers in Canada - ODaF has helped the Canadian RDC Network to write a specification for the development of DDI 3.0-based framework which will be used to manage metadata among the RDCs. (See Chuck Humphrey's presentation for more details).

EU FP7 Virtual RDC Proposal: This is a proposed project involving several European countries to create a network of virtual RDCs. Metadata Technology has been involved in helping draft the metadata management component of the proposal. (See Stefan Bender's presentation for more details.)

SDMX Browser: This is a soon-to-be-complete open-source development project to create a generic desktop client for storing and visualizing aggregate data formatted as SDMX-ML. It includes features for viewing the structural metadata (concepts, code lists, etc.) as well as data sets. Currently in beta.

Standards Mapping Project: This is an informal project with participants from several organizations, working informally to define mappings between SDMX, ISO/IEC 11179, DDI, XBRL, Dublin Core, and ISO 19115. Participants include members from METIS (the bi-annual UN/ECE metadata conference), national statistical organizations, ODaF, the DDI Alliance, the SDMX sponsors, and XBRL international. Hopefully some use cases and detailed mappings between these standards will be made available by the end of 2008.

One point raised by Chuck Humphrey was the need for a coordinated architecture and strategy for tools development around the use and support of DDI 3.0. This point was generally well-taken, and will probably be reflected in

³ <http://tools.ddialliance.org/?lv11=ftp>

⁴ <http://www.norc.org/DataEnclave>



some of the proposed DDI-related projects in which ODaF will be taking an active role as a result of this meeting.

The afternoon consisted of presentations by participants (see presentations on ODaF website):

- The UKDA Economic and Social Data Service ((ESDS) [Louise Corti, UK]
- The Institute for Study of Labor (IZA) [Nikos Askitas, Germany]
- The Institute for Employment Research (IAB) [Stefan Bender, Germany]
- Forschungsdatenzentrum der Rentenversicherung (FDZ-RV) [Conny Dippold, Germany]
- ICPSR/SRO Shared Data Model Project [Mary Vardigan, USA]
- Canada RDC Project [Chuck Humphrey / Raymond Currie, Canada]
- DANS Digital Preservation Project [Rene van Horik / Laurent Sesink, Netherlands]
- SSB Archiving / Dissemination project [Rune Gloersen, Norway]

Also presented was GESIS-ZUMA's contribution to the DDI Foundation Tools Program, which includes plug-ins for SAS and SPSS which provide for the generation of DDI metadata files.

April 15th (Day 2)

The second day was organized around a series of topics and ideas which had emerged from the discussions and presentations of the preceding day. The intent is to come up with new project proposals or activities which ODaF could help coordinate, or projects in which ODaF could be a meaningful participant.

The list was as follows:

- (1) DDI Versions: Should a revision of DDI 2.* be created to better align existing implementations with DDI 3.0?
- (2) What should ODaF do to promote the outcomes of the standards mappings project?
- (3) Driving adoption of DDI within the official statistics community
- (4) DDI "Roadmap" describing all of the tools and work going on around DDI 3.0.
- (5) European DDI User's Group
- (6) Creation of a DDI 3.0 Editing tool
- (7) DDI Metadata Visualization tools
- (8) Distribution and extension of the UKDA QuDEX qualitative metadata browser
- (9) Meeting focused on RDCs
- (10) Tools for transforming from Beyond 20/20 format to SDMX
- (11) Adding archival standards (METS, PREMIS) to standards mapping
- (12) Create a pre-configuration of Fedora which supports DDI
- (13) Write a best practices paper for the use of DDI 3.0



The outcomes of discussions are noted here:

(1) DDI versions

It was decided that there was no current ODaF action around this idea. It is something which will be discussed internally in the DDI Expert Committee, and if any decision is made there then perhaps the topic can be revisited by ODaF.

(2) Standards Mapping Project:

It was felt that a paper describing the "standards landscape" would be useful. The detailed outcomes of the project should be published on/linked to from the ODaF website, and all detailed mappings should be submitted to the METIS Metadata Framework for publication. The overall standards landscape would also make a good poster-type of presentation, to help explain the idea. It was also suggested that an ODaF speaker be found to present at the next ISI meeting in Durban, South Africa in 2009.

(3) DDI Adoption in Official Statistics

ODaF should coordinate with Rune Gloersen to help increase awareness of DDI 3.0 within the NSOs, at METIS and in other fora.

(4) DDI "Roadmap"

It would be useful to have a document describing the overall lifecycle model, with an explanation of how DDI and various tools fit into the picture - this could be couched in terms of use cases or tasks. It would be very good if an end-to-end demo could be created. ODaF could contribute to the creation of such a roadmap, and could act also as a point of publication for it.

(5) European DDI User Group

The existence of such a user group was seen as desirable and useful, so long as it did not present issues of coordination with the DDI Alliance itself. It was noted that the September CESSDA meeting will focus on DDI. It was felt that a user group could cover many areas:

- Contribute to definition of best practices
- Provide descriptions of user experiences
- Help gather requirements for input into the standard as it evolves
- Provide information regarding updates in the DDI community
- Create tutorials for users



This is similar to some other user fora. ODaF could help to coordinate such a group, which might be organized into a business track and a technical track. The idea is to issue a call for papers in December 2008, with a targeted date for a conference in May of 2009.

(6) Creation of an Open-Source DDI 3.0 Editor

The idea is to launch a project similar to the DDI Foundation Tools Program in which several organizations contribute to the coordinated development of a DDI 3.0 editing tool, which would be made available as an open-source application to help adoption and use of the standard. Interested parties included DANS, DDA, the DDI Alliance, ODaF, UKDA (as a tester), IZA, IAB RDC, GESIS-ZUMA. Other organizations which might be interested include: the World Bank/IHSN, SRO, NORC, and NSD. All DDI Alliance members should be approached.

(7) DDI Metadata Visualization Tools

There was some discussion of creating also tools for metadata visualization (category statistics, use of variables, grouping and comparison of DDI structures, etc.) The idea is to have an open source project hosted by ODaF, rather than a broad collaborative effort like the DDI Foundation Tools Program. Funding sources might include NSF, IZA. The proposal was made to discuss this idea further, so that coherent proposal could be outlined. This activity is seen as taking place further down the road, given that the initial focus should be on the implementation of DDI 3.0 tools as well as the compilation of metadata.

(8) QuDEX Browser Tools

The existing QuDEX browser developed by UKDA could easily be used as part of a Flex application which could also support multi-media (audio, visual, video, etc.) The idea is to take the existing code, extend it to support some additional media, and release this as open-source through ODaF. This was seen as a viable option, since it is not resource-intensive and will be help to drive a standard approach to the metadata related to qualitative data. Several attendees expressed an interest in this type of application: UKDA, ICPSR, and the Canadian RDC Network.

(9) RDC-Focused Meeting

The question was raised whether it would be useful to have an ODaF meeting focused on the issue of metadata management for RDCs. This idea was not endorsed, as meeting is planned for September of 2008 (in Winnipeg, Canada). ODaF would be invited to such a meeting to contribute, but it is not needed to help organize.



(10) Tools for Beyond 20/20 Transformations

Because many archives have aggregate data in Beyond 20/20's format, it was seen as useful to have tools which would transform this into SDMX, so that other tools could also be used for this data. People to address this issue should include Michel Seguin at Statistics Canada (michel.seguin@statcan.ca) Additionally, it was agreed that the developers of the Australian Superstar/Supercross applications might be approached to provide support for SDMX. Warren Richter is the person to contact, and it was agreed to do this. Note that SDMX support is already planned for PCAxis, which is another of the tools widely used by NSIs. This was seen as a good project to propose for ODaF if the right expertise could be identified to do the mappings.

(11) Adding Archival Standards to the Standards Mappings

There are now some standards emerging out of the archival world which could usefully be added to the work on standard mapping, notably METS - a packaging standard for digital preservation - and PREMIS, an XML based on the OAIS model. It was agreed that these should be addressed as part of the the work on standards mapping.

(12) Creating a Fedora Configuration for DDI/METS/QuDEX

It was noted that many people are now using Fedora⁵ and related open-source software for archiving (eg, "Elated"). Much of the work in using Fedora is to build a "content model" which is essentially a configuration of Fedora. It was agreed that a project providing a pre-configuration of Fedora to support DDI, METS, QuDEX and other relevant standards would be useful to have, and that organizing and investigating such a project should be pursued by ODaF.

(13) DDI 3.0 Best Practices Paper

It was agreed that it would be useful to put together a paper on DDI 3.0 best practices. This could be a joint effort between several organizations, and it was proposed that further discussion be held between interested parties. These included ODaF, the Canadian RDC network (Chuck Humphreys), and UKDA (Ken Miller). It was noted that CESSDA already has some material in this area, and that it should be used as input to the discussion. The International Household Survey Network has also already produced a set of guidelines for data archivists⁶ that can be used as a reference.

⁵ <http://www.fedora-commons.org/>

⁶ http://surveynetwork.org/home/docs/DDI_IHSN_Checklist_OD_06152007.pdf



(14) ODaF Meetings

It was agreed that an annual ODaF Europe meeting was useful, and that another would be organized for 2009, to be hosted by IZA in Bonn. Also, the next working meeting in the US will be hosted by Sun Microsystem's facility in New York City, during the fall of 2008 (currently planned for Oct 2-3).

Other

Note that during the discussion the subject of the HASSET classification came up, and that this was something which could possibly be expressed in some standard formats. This is a point for investigation.

Conclusions

We would like to thank all the participants for coming to the event and for sharing their thoughts, ideas, suggestions and for making this first regional event a success. The above issues and actions will be followed-up upon in the coming weeks by ODaF and other agencies / individuals. Results and outcomes will be reported through the ODaF web sites and the members' mailing list. Many thanks to the UK Data Archive staff – especially Kevin Schurer and Louise Corti - for their hospitality. We are looking forward to next year's meeting at IZA in Bonn.

For feedback / comments on this report, please contact
agregory@opendatafoundation.org or pheus@opendatafoundation.org