



2007 Annual Meeting Report

Pascal Heus / Arofan Gregory
January 2008

Table of Contents

Overview	2
Agenda.....	2
Discussions Items	3
Business report.....	3
Ongoing Projects	4
Status of 2006 Resolutions.....	6
Marketing Activities.....	6
New projects / Ideas	7
Other Standards	8
Face-to-face meetings.....	9
Other Remarks / Ideas	10
Resolutions	12
Annex 1: Finances	13

Overview

The 2007 Open Data Foundation Annual meeting was held in Washington, DC, USA on Wednesday December 19th and attended in person or via conference call by the following participants:

Directors: Julia Lane, Ernie Boyko

Advisors: San Cannon, Tim Dunstan, Dan Gillman, Arne Ulhendorff (representing IZA), Wendy Watkins

Managers: Arofan Gregory, Pascal Heus

Observers: Tim Mulcahy (NORC), Bryan Fitzpatrick

Agenda

The agenda was finalized at the meeting and the order was reshuffled as needed to meet participant's availability.

- I. Business Report
 - a. Personnel Changes
 - b. Financial Report
 - c. IT Infrastructure
 - d. Insurance / Tax exemption Status
- II. Report on Ongoing Projects
 - a. NORC Data Enclave
 - b. UKDA DExT
 - c. DDI Tools
 - d. UKDA QuDEX
- III. Status of Resolutions from 2006
- IV. Marketing Activities
 - a. Website, Newsletters and Blog
 - b. Position Papers and Architecture Guidelines
 - c. Conferences
 - i. IASSIST 2007 (Montreal, Canada, May 07)
 - ii. Open Forum on Metadata Registries (New York, NY, Jul 07)
 - iii. Digital Curation Center (Washington, DC, Dec 07)
 - iv. XML 2007 (Boston, MA, Dec 07)
 - d. New Ideas for Further Marketing
- V. Face-to-Face Meetings
 - a. Report on St. Helena Meeting
 - b. Discussion of Next Meeting
 - i. European meeting in Spring
 - ii. US Meeting in NY (Sun?)
 - iii. Annual meeting
- VI. New Projects and Ideas
 - a. Prospects: EU Protected Area, ISO 11179 registry (w/Dan McCreary), Human Genome (www.biobricks.org), Canada RDC
 - b. Propose "Generic" projects: technical support to data enclave / RDC; DDI 3.0; SDMX
 - c. Other important standards? (Semantic Web; XMDR)
- VII. Resolutions

Discussions Items

Business report

Personnel Changes

- Mary Vardigan from ICPSR stepped down as an advisor to reduce risks of conflict of interest given her involvement in other related projects involving ODaF.
- Two new advisors recently joined the Foundation:
 - o Dr. Andy Nelson, a geography / GIS expert specializing in methods for modeling environmental and socio-economic data for natural resource management from global to local scales. Andy currently works at the Global Environment Monitoring Unit (GEM), Institute for Environment and Sustainability (IES) at the Joint Research Centre of the European Commission
 - o Dr. Nikos Askitas, an IT expert from the Institute for the Study of Labor (IZA) in Bonn Germany. Nikos brings on board a very strong IT expertise and a connection with the Research Data Centers network in Germany.
- As Pascal Heus recently joined Metadata Technology (a UK based private company), and given that Arofan Gregory and Chris Nelson are directors of that company, we feel that this situation over-represents Metadata Technology on the management team of ODaF. Chris Nelson has agreed to step down as a manager and we propose his replacement to be Brian Fitzpatrick. Brian is currently engaged in data-related consulting services, in particular in Vietnam and is retired from the Australia Bureau of Statistics (ABS) where he was Chief Technology Officer and leader of the ABS Enterprise Architecture planning. His full resume will be circulated amongst the director. Brian was involved in helping with the development of SDMX, and has been a proponent for technology standards and for good metadata management practices within the official statistics world.

Finances

- The total income for the ODaF in 2007 was around \$106K, primarily coming from the NORC Data Enclave Project (45%) and the UKDA DExT Project (40%) with the remaining balance coming out of personal contributions from the management team.
- The net income was around \$26K.
- Operating expenses totaled \$14K and consisted in supporting the IT and administrative infrastructures, membership in the DDI Alliance, sponsoring and participating in the IASSIST 2008 conference, and organization of the Napa Meeting.
- Forward looking cash balance for 2008 is \$12K
- Anticipated expenses for next year are expected to increase to about \$30K to include liability insurance and provide further support for participation in conferences and organization of meetings.
- 2008 income will come from ongoing projects such as the NORC Data Enclave and prospective collaboration work with the Canada RDC, Chaplin Hall, UKDA, EC and others. We also propose to more actively seek donations, possibly through sponsorship of ODaF services such as the G-Forge repository, web site hosting, and other.
- Regarding accepting donations, the following points were made:
 - o We can seek donations from private companies but need to make sure we have more than one sponsor to avoid being tied to a single vendor.

- Donations should have 'no strings attached'
- Potential vendors include Sun, IBM, MS, Oracle, SAP, Adobe, Red Hat, Yahoo, Google, Swivel
- Care should be taken if sponsorship by Microsoft is sought, given the company's negative image in the open source and standards communities and with other organizations
- We first need liability insurance and tax exemption status
- We should set a maximum amount – the exact amount was not decided.

IT Infrastructure

- ODaF currently operates two dedicated servers hosted by the aplus.net company. One server is of web hosting and email and the other is dedicated to the G-Forge based repository to support development of open source projects.
- It was suggested by meeting participants to setup a web accessible secure area to share documents / presentations amongst ODaF with various access level (management/directors, advisors, members).

Insurance / Tax Exemption status

- Filing for 501(c)(3) status has not been completed at this time. One reason for the delay is that the company needs to show some legitimate track record before applying for tax exemption. The filing is expected to proceed shortly.
- We are also seeking to purchase a liability insurance to protect the company and its members. Given that we operate globally, an international coverage is required which considerably increases the insurance premium. We are negotiating with insurance agents and are close to a decision.
- Current coverage is provided only through stipulation in contracts with providers that they assume professional liability for the tasks they perform. While this is a reasonable clause to have in our contracts, it does not provide some types of coverage which are desirable.

Ongoing Projects

NORC Data Enclave

- The project has been progressing very well and the data enclave went live last September. Detailed information is available at <http://dataenclave.norc.org>.
- Current data providers include the NIST Technology Innovation Program (formerly known as Advanced Technology Program), the UN Department of Agriculture and the Kauffman Foundation.
- ODaF's role is primarily technical assistance for the adoption of metadata standards and best practices, and support for the management of survey data and documentation.
- This first year focused on establishing the data enclave and the research collaboratory.
- 2008 will continue to support and promote collaborative research through the enhancement of metadata and development of new tools such as a citation management utility or a statistical source code metadata parser. New data and data providers are also expected to join the enclave.

UKDA DExT

- This is a collaborative project with the UK Data Archives for the development of data archiving and conversion tools.
- This initial phase is a proof of concepts and aims at the development of a utility that (1) reads SPSS files, (2) exports the data to ASCII and generate DDI 3 metadata for storage and preservation, and (3) generates setup script to import the ASCII data in statistical packages such as SAS, Stata and SPSS.
- The project initially scheduled to conclude at the end of November is slightly behind schedule and we're looking at a January 08 delivery.
- Beside the output product that will perform the above functionalities, the project has produced an interesting stand-alone package, as the SPSS reader can be reused by other applications to read SPSS file and generate DDI 2 or DDI 3 compliant metadata. This would be one of the required components for a new DDI Editor for the data archivist. It could also be used to deliver web-based data/metadata conversion services.
- The DExT initiative is ending and we have no guarantee at this point the project will be renewed under another UK funded initiative. We however expect to continue to maintain and enhance this product as it will be donated to the DDI Foundation Tools Program as a UKDA contribution.

UKDA QuDExT

- As part of the overall DExT project, UKDA has designed a standard XML format for capturing metadata added by analysts working with multi-media interviews and other non-numeric sources. There are about 15 CAQDAS vendors, 9 of whom have been involved in this project. Funding is soon to end, and ODaF has been assisting UKDA in finding a standards organization to adopt and maintain this work, and also providing technical advice. It is possible this work will become part of DDI.
- It was suggested (JL) to bring the project outputs to the attention of other users and archives. This should be advocated in conferences and we should also discuss with Louise to see how they can support advocacy of the tools and ODaF (Arofan to follow up).

DDI Foundation Tools Program

- The DDI-FTP is a collaborative effort from several partner organizations to support the development of core libraries and tools for the management of DDI metadata and statistical data. The initiative was launched last September is expected to last for at least 2 years.
- ODaF is providing the web infrastructure to host a new DDI Tools web site and manage the open source component development. See <http://tools.ddialliance.org> for a web site preview, expected to launch around mid-January. We are also contributing to the development efforts
- Short term anticipated delivery are a "road map" document for the development of tools, core Java classes for the management and validation of DDI documents (1/2.0 and 3.0) and the new web site. A SPSS export tool developed by Achim Wackerow (GESIS/ZUMA) will also be made available (requires SPSS 14).

Standards Mapping Paper

- There has been a project proposal draft for mapping SDMX, DDI, ISO/IEC 11179, and various other related standards to each other for some time. This project has not

been approved or launched. However, significant work in this area has been performed by various involved individuals, and it was suggested that the proposal be completed and proposed. This would take place in the first quarter of 2008.

- Suggested title / keywords: "Metadata Mosaic", "Rosetta Stone 2.0"

Status of 2006 Resolutions

The resolutions from the 2006 meeting were not reviewed in detail due to time constraints. Progress was summarized as being good on the project front, but as not being as well advanced in terms of processes, job descriptions, and other day-to-day tasks. The decision was to focus on getting these resolutions completed, rather than trying to further expand the scope of the organization at this time.

Marketing Activities

Conferences/Workshops

- We made ODaF presentations at the following conferences
 - o IASSIST 2007 (Montreal, Canada, May 07)
 - o 10th Open Forum on Metadata Registries (New York, NY, Jul 07)
 - o 3rd Digital Curation Center (Washington, DC, Dec 07)
 - o XML 2007 (Boston, MA, Dec 07)
- We also held a two day workshop on Metadata Standards and Best Practices at the Institute for the Study of Labor in Bonn, Germany (Nov 07)
- The above presentations are available on ODaF's web site in the "Publications" section
- We participated in the DDI 3.0 training in Germany, at the Dagstuhl Computer Science Conference Center. This was arranged by Joachim Wackerow of GESIS/ZUMA, and involved more than twenty students from research institutes and archives around the world. Another similar training is being arranged for next year.
- For 2008, we plan to continue show casing ODaF, metadata standards, best practices and related tools at various events.
- We were sponsors for the IASSIST 2007 conference and expect to do likewise for the 2008 event at Stamford in Palo Alto, California.

White papers / publications

- A "DDI and SDMX: Complementary, Not Competing, Standards" paper was published on the ODaF web site to clarify how these two standards can work together.

Ideas for further marketing

- It was emphasized during the meeting that we need to produce more papers and short publications to further advocate our existence and work.
 - o Potential publication places: IASSIST Quarterly, Chance (JL)
- We went through a full review of the XML 2007 presentation which we feel was appropriate for an audience of technologists. A paper will be prepared along this presentation by Arofan.
- This led us to a long discussion emphasizing the need for a more "general public" oriented paper that describes the vision of ODaF (without technological terms). Many

people do not understand what we do and we need to reach out of our usual network. This idea will be followed up upon by Pascal who will prepare a first draft and circulate for review and contributions.

- We will also down the road need ODaF overview papers targeted for our core audiences (producer, archivist, researchers, etc.)

New projects / Ideas

Generic Activities

- On several occasions, we had in 2007 the opportunity to engage ODaF in activities aligned of the mission of the organization but we could not do so as no official ODaF “project” existed to execute it. These activities were often short-term work items such as training, advocacy, workshops or provision of minor technical assistance that would actually not fall into the scope of a full-fledged project.
- To address this in the future, we propose authorizing ODaF managers to be allowed to contractually engage the organization in such activity without prior authorization from the Directors and the need to file for a project proposal. It is understood that such work would need to be short-term or of limited scope, and not fall in the category of a Project and that Directors will be kept informed of such activities.

EU Protected Area

- The Global Environment Monitoring Unit (GEM) at the Institute for Environment and Sustainability (IES) of the Joint Research Centre of the European Commission its is currently supporting a project aiming at a large scale assessment of protected areas (see <http://www-tem.jrc.it/PA/index.html>)
- Every 10-days the project generates a large number of indicators compiled from many different sources and the project team lead by Dr. Andy Nelson (ODaF new advisor) is looking for a solution to disseminate these results on the web.
- Through informal discussions, we suggested to look into SDMX as an approach to deploy a global registry.
- This is currently being evaluated and there is a good possibility that this approach will be adopted. If this is the case, the Open Data Foundation would be in a position to provide technical assistance for the implementation of the solution.
- This would foster the adoption of the SDMX standard and related technologies and will demonstrate its use outside the world of official statistics. It would also likely make the standard visible to other related initiatives that would then be likely to adopt similar solutions.

Open Source ISO 11179 Registry

- Following a presentation about ODaF at the XML 2007 conference in Boston, a volunteer developer asked if it might be a good idea to create an open-source ISO 11179 registry. This would depend on the status of the standard ISO bindings to such a registry, which is a separate ISO standard currently being developed.
- Both Tim Dunstan and Dan Gillman found felt this project would be very useful and benefit the ISO 11179 community. Making such product available would also bring ODaF to the attention of the ISO 11179 community.

- The decision was taken to look further into this idea, and to propose a project if it was felt that sufficient resources could be found to conduct the development.

Human Genome (“biobricks”)

- After the presentation at XML 2007, Arofan was approached by someone who had been involved in a project creating a registry of information about the human genome, with a not-for-profit known as “bio bricks”. While the idea behind this not-for-profit is similar to that of ODaF, the subject is well outside our current area of focus (social sciences and official statistics).
- This project was ruled out of scope of the ODaF current work. We can keep in touch with the community but should not invest extra time/resources.

SDMX Community Support

- During the meeting, the need for a SDMX User Community space was identified as a major need for the standard.
- We therefore suggested to further explore who is involved in SDMX and how wide the community really is (new SDMX related projects seem to emerge from many spaces) and propose to establish a web based facility (primarily a web page and a mailing list) to facilitate the dialog between all SDMX users, provide a space to compile SDMX projects / resources, and give a chance to users to share their stories.
- If this takes off, we should also probe the groups for a potential meeting or technical conference focusing on tools and implementation, strategies, controlled vocabularies.
- This project was strongly supported by San Cannon and other participants.

DDI 3.0 and Questionnaire Design Tools

- During recent email exchange, a large number of open source and commercial survey questionnaire design tools were identified. We feel that this list could be useful to advocate the new DDI 3.0 standard amongst the questionnaire design tools vendor / developer community
- This activity could be conducted by ODaF (possibly on behalf of the DDI Alliance) to make these companies aware of the existence of DDI 3.0 and its ability to manage question banks or capture questionnaire instruments in a standard XML based fashion. This could foster the adoption of the specification, support the import/export of information between tools and promote the development of new DDI-compliant tools.
- If approved by the directors, we would first discuss the idea with the DDI Alliance and collaborate with them to formulate the overall approach strategy (timing, content, etc)

Other Standards

XMDR

- XMDR is a project to produce input to the revision of ISO/IEC 11179 and produce a reference implementation of the proposed revision. It is designed to bridge the gap between ISO 11179 and the emerging semantic web standards with their focus on

ontologies. Because ODaF is working in some related areas (standard classifications and ontologies, and the promotion of ISO 11179) mention of XMDR at conferences and at other events has been frequent. A representative of the group was invited to the St. Helena meeting, but did not respond to the invitation.

- Dan Gillman expressed some reservations regarding this project. It is often touted as representing the US position on the revision of ISO/IEC 11179, and this is not the case. XMDR is a research project centered at Lawrence - Berkeley National Lab providing input to the US technical standards committee responsible for ISO/IEC 11179, among other things.

Face-to-face meetings

2007 Napa Meeting

- An Open Data Foundation Meeting was held St. Helena, California, September 13 – 15, 2007. The intent of the meeting was to discuss current work and brain-storm ideas for future projects, rather than to formally conduct the business of the organization. Due to limited space, only Directors, Advisors, and Managers were invited to the meeting, along with a few presenters from outside the organization.
- Outside ODaF directors, advisors and managers, participants included members of the following organizations: UK Data Archive, Dutch Data Archive (DANS), Canada Research Data Centres, US National Cancer Institute, Adobe Ltd. and Swivel.com
- A full meeting report was circulated earlier.
- We continue to build upon the meeting outcomes.

2008 Planned Meetings

- Based on increasing activities and on request from our members, we plan to organize a European ODaF meeting in the spring of 2008. Several places have been proposed to host such meeting (UK, Netherlands, and Germany). Given the fast emerging interest in DDI in Germany and taking into other aspects such as traveling and institutional support, we agreed that Germany would strategically be the most appropriate place at this time. We will therefore pursue the Institute for the Study of Labor (IZA) invitation to host a meeting in March or April at their facility in Bonn (PH).
- Following the success of the Napa meeting, we also feel a North America get together in 2008 would be useful and we propose to host such event in the Fall. During the XML 2007 conference, Jon Bosak from Sun Microsystems suggested that we could use their New York headquarters facility for such event. We will follow-up with him to further pursue this idea.
- Overall, we feel that hosting technical / brain storming meetings with external organizations invited to present their perspectives and idea leads to very positive developments. To make the event attractive, we should make sure the facility is of high standards and provides a rewarding experience for participants.
- The 2008 Annual Meeting will be planned for December in the US, likely at NORC's office in Washington DC.

Other Remarks / Ideas

Geography / GIS

- This topic is often mentioned in our presentation but we do not have much expertise or activities in the area.
- Geography is a very large field and the space is already well filled by several communities.
- We should in general be more aware and active in this area and need to:
 - o Add geography / GIS related expertise to the board of advisors
 - o Network with geography / GIS communities to ensure bi-lateral awareness of our standards and activities
- issues that were highlighted:
 - o Dynamic changes of boundaries across time
 - o Standard definitions (e.g. housing)
- Wendy Watkins can connect us with geography group at Carleton
- Additional contact can be provided by Julia Lane
- Dan Gillman also suggested the US Census Bureau

Statistical Data Disclosure

- Question was raised whether we should get involved in the area of data anonymization / statistical data disclosure
- It was agreed that this is a fairly closed community where it takes a very long time to gain credibility. The decision was taken to not attempt to involve this community for the time being..

Data Shift / New Directions

- Less and less data will be released, we are shifting more and more towards Research Data Centers / Data Enclaves.
- This opens the door to new opportunities:
 - How do research communicate with each other (collaborative / metadata)?
 - Distributive research teams.
 - o AG: need processes
 - o JL: research requires trusted relationships
- Look into virtual collaborative:
 - How can we learn from myspace, facebook, second life?
 - NSF workshop on virtual organizations
- This however doesn't work in developing world where we basically need more open source tools, how do we reconcile?

Regarding Outreach to Other Communities

- This is OK for networking purposes but we must minimize investment of resources.
- We first need to establish ODaF as a strong player in SBE.
- Target for other sectors should be 2-3 years down the road.
- We should not try to solve other people's problems.

DDI / SDMX Training

- Metadata Technology just started offering 3-day courses in SDMX and DDI 3.0
Given that this is not an ODaF product and MT is a commercial company, we do not

want to directly advertise this directly but want to make sure the interested communities are made aware of this.

- Do we have permission from the Directors to advertise this amongst ODaF members (through the mailing list or web site)?
- Other suggested advocacy channels include IASSIST and DDI mailing lists. The Joint Program in Survey Methodology was also mentioned as a potential delivery space (<http://www.jpsm.umd.edu/jpsm/>)

Fedstat

- Dan Gillman pointed out that the FedStat project, a “portal” to statistical resources in the US may soon be revised. If the project reconstitutes, it could present an opportunity for ODaF to participate (<http://www.fedstats.gov/>)
- Related to this topic, a DRM/SDMX mapping paper would be most useful. Arofan will look into this with San and Dan as reviewers.

Priorities / Strategies

- Top priorities are to:
 - o Take care of administrative/legal issues (tax exemption, liability insurance)
 - o Get the mission statement and vision papers out
- Focus and build on success stories, strengthen position within network. Go for the low hanging fruit (RDCs)
- Deepen existing portfolio (Hardening and deepening)
- In general, we need for any project / product to clearly identify the users and link to ODaF’s mission statement.
- Secure funding

Resolutions

1. Brian Fitzpatrick should become an executive manager for ODaF to replace Chris Nelson.
2. A secure area on the ODaF web servers should be established to share internal documents and presentations amongst directors, management, advisors and members.
3. ODaF managers should be authorized to contractually engage the organization in any activity related to its mission as long as such activity is of limited scope and duration and does not fall in the category of a Project. This is primarily to cover advocacy, training or short-term technical advisory work.
4. A proposal should be submitted to the directors for a collaborative project with the Joint Research Centre of the European Commission for the use of SDMX for the dissemination of data on protected areas.
5. A project for the development of an open source ISO 11179 registry implementation if sufficient resources can be identified, and assuming that the ISO standard XML bindings are in a sufficiently mature state.
6. A proposal for the establishment of a virtual community to support SDMX Users should be submitted.
7. A proposal for a project to promote DDI 3.0 amongst communities of questionnaire design and question management tools developers should be submitted.
8. ODaF should be required to produce and publish at least 5 white papers in 2008.
9. A regional European ODaF meeting should be organized in Germany in the Spring of 2008, at the IZA offices if possible.
10. A regional ODaF meeting in the US should be organized in the Fall of 2008.

Annex 1: Finances

Accounts

- Current Balance : \$21,504.72
- Account receivable: +\$10,400.00
- Accounts payable : -\$22,225.13
- *Balance* : \$9,679.59
- To bill later
 - NORC \$2.6K : \$390.00 (15%)
 - UKDA £8K/\$16K : 2,400.00 (15%)
- Projected balance : \$12,079.96K

Income

- NORC Data Enclave Project
 - \$45K → 15% → \$6.75K
 - Consultants: PH / Metadata Tech
- UKDA DExT
 - £20K / \$40K → 15% → \$6.00K
 - Consultants: PH/JG (75%), AW (10%)
- Canada RDC
 - \$7K → \$3.6K (donation)
 - Consultants: JR/AG/PH (travel only)
- Donations
 - PH: \$10K (NORC work not invoiced)
- Total income: ± \$102K
- Net income: ± \$26K

Expenses

- APLUS Servers : \$2.8K
- DDI Alliance : \$2.5K
- Napa meeting : \$3.5K
- IASSIST 2007 : \$3K
- Misc. : \$2K
 - Bank, USPS, quicken, printing

2008 Estimates

- Expenses
 - APLUS : \$3K
 - GForge : \$2K
 - DDI Alliance : \$2.5K
 - IASSIST 2008 : \$3K
 - Meetings/conf : \$10K
 - Insurance : \$5K
 - Miscellaneous : \$2.5K
 - TOTAL : ±\$30K
- Income
 - NORC : \$10K
 - Projects? : \$10-\$20K (UKDA, RDC, EU, etc.)
 - Donations? : \$10K (sponsors)