



Project N°: 262608



Acronym: **Data without Boundaries**

DELIVERABLE D7.5

DDI and SDMX.

A Report on Collaboration with Groups Developing DDI and SDMX

WORK PACKAGE 7

Standards Development

REPORTING PERIOD:	From: Month 18	To: Month 48
PROJECT START DATE:	1 st May 2011	DURATION: 48 Months
DATE OF ISSUE OF DELIVERABLE:	April 2015	
DOCUMENT PREPARED BY:	Partners 2, 5	UTA-FSD, GESIS

Combination of CP & CSA project funded by the European Community
Under the programme "FP7 - SP4 Capacities"

Priority 1.1.3: European Social Science Data Archives and remote access to Official Statistics

ACKNOWLEDGEMENTS

This document was prepared by Mari Kleemola, task leader (UTA-FSD), and Uwe Jensen (GESIS).

The authors gratefully acknowledge contributions and feedback from WP7 team partners Stefan Ekman, (UGOT-SND), Claus-Göran Hjelm, and Hans Irebäck (SCB).

Table of Contents

1	Introduction and Background	7
1.1	Description of Work	7
1.2	Relationship with other WPs and Deliverables	8
1.3	Outline	9
2	The Metadata Standards: DDI and SDMX	10
2.1	Data Documentation Initiative (DDI).....	10
2.2	Statistical Data and Metadata eXchange (SDMX)	10
3	Collaboration and Contribution by WP7	11
3.1	The Landscape	11
3.2	DwB Workshop on Metadata Standards December 2011, Gothenburg.....	12
3.3	DDI SDMX Dialogue Group 2010-2013.....	12
3.4	GSIM/DDI/SDMX Mapping 2013	13
3.5	Workshop on Statistical Metadata (METIS): Implementing	
	the GSBPM and Combining Metadata Standards, October 2011	14
3.6	Workshop on Strategic Developments	
	in Business Architecture in Statistics, November 2012.....	15
4	Discussion and Future Directions	16
5	Conclusions	19
	References	20
	Glossary of Abbreviations	25

1 Introduction and Background

The Data without Boundaries (DwB) project is an EU project within the 7th Framework Programme (FP7), aiming at making official statistics microdata from European countries available for researchers within the European Union.

With 29 partners belonging to the European Statistical System (10 National Statistical Institutes or statistical departments), to the CESSDA (11 Data Archives) and to the Research Community (7 universities and 1 SME involved in methodological research), the project aims at discussing and promoting common improvements, solutions and frameworks to be proposed to these communities.¹

When the goal is cooperation and interoperability, among the most important issues are common standards and protocols. Especially metadata standards are crucial since they form the basis for common understanding and for correct use and interpretation of microdata. This report describes the work done by WP7 in promoting collaboration between metadata standard bodies, and in advancing discussions about metadata in organisational and individual level.

The audience for this report are statistical institutes, data archives and all metadata experts that are interested in standards and their development.

1.1 Description of Work

The central purpose of Work Package 7 was to create a common platform for a lasting cooperation between national statistical institutes and data archives. This Deliverable is based on Task 7.7 that is conveyed following in the Description of Work (DoW):

- “This task will build and maintain effective collaborations with the DDI Technical Implementation Committee and the SDMX (Statistical Data and Metadata eXchange).”

The Deliverable D7.5 is described as a

- “report on Collaboration with Groups Developing DDI and SDMX”.

The Description of Work for the project and for Work Package 7 was written mainly in 2010, and the DwB project started in May 2011.

The DoW for WP7 emerged from extensive discussions with several stakeholders and reflected the situation and the needs of the various communities and groups in the metadata field in 2010. In the discussions, swift collaboration between metadata standards like DDI (Data Documentation Initiative) and SDMX (Statistical Data and Metadata eXchange) was seen unanimously as beneficial.

¹ Data without Boundaries - DwB website: <http://www.dwbproject.org/about/> [Accessed 14.1.2015]

Therefore it is not surprising that things moved forward quickly and collaboration had already begun before the DwB project kick-off. For example, the SDMX/DDI Dialogue Process was initiated in December 2010², shortly before this project started. However, WP7 participants were involved in the Dialogue Process from the beginning.

Because of these developments, this Work Package was able to “jump onto a moving train” and did not need to start with building nor maintaining the collaborations as such; instead, the focus was shifted to networking, creating opportunities for, and participating in the collaboration³.

Consequently, the main goals in Task 7.7 were set as:

- to identify best opportunities for networking and collaboration;
- to actively create opportunities for collaboration;
- to closely follow the progress in this fast-moving field.

The DoW specifically mentions the collaboration with and between the DDI Technical Implementation Committee⁴ and the SDMX. Both are important standards developing groups, and actually the discussions about the possibilities of collaboration were initiated by experts familiar with both standards. The collaboration, however, has not been restricted to these two bodies and to the DwB project - instead, and for the benefit of all, it has been much wider and more extensive than expected in the time of writing the DoW. The broad collaboration and growing interest in metadata issues reflect the importance of metadata in the digital world as well as the realisation that creating long-lasting and robust yet flexible and interoperable solutions and best practices requires groups to align and coordinate their efforts.

1.2 Relationship with other WPs and Deliverables

In DwB, several work packages have been working with metadata issues. WP5 has provided structured metadata models for creating a study-level catalogue of official statistics microdata and for working with the actual data files (DwB D5.2⁵). WP8 has focused on the metadata needed for data discovery (DwB D8.2⁶). WP7 has produced a state-of-the art

² SDMX DDI Dialogue - Overview Page.
<http://www1.unece.org/stat/platform/display/metis/SDMX+DDI+Dialogue+-+Overview+Page>
[Accessed 14.1.2015]

³ It was also decided to move person months from Task 7.7 to other tasks, mainly to contribute to the DwB deliverables D7.1 and D7.2/D7.3.

⁴ Note: the organisational structure of the DDI Alliance changed in 2013 and the DDI TIC ceased to exist. The new DDI organisation supports the effective development of the DDI specifications. There is an Executive Board elected by the member representatives, a Scientific Board that oversees the substantive development of the DDI specifications, and a Technical Committee that creates and stewards the specifications and ensures their usability (Vardigan, M. 2013).

⁵ DwB D5.2 Report and databank documenting national OS Microdata, 2013.
(<http://www.dwbproject.org/about/deliverables.html>) [Accessed 18.3.2015]

⁶ DwB D8.2 Metadata Model, 2012. (Ibid).

review of metadata standards usage (DwB D7.1⁷) and charted standards with future relevance as well as the needs, key areas, rules and best practices in metadata standard selection and usage (DwB D7.2/D7.3⁸). In addition to this short Deliverable, the work done within Task 7.7 has provided input to the other WP7 Deliverables.

1.3 Outline

In this compact report we first introduce the DDI and SDMX standards very shortly. We then proceed to describe the metadata landscape from the DAs and NSIs point of view, as well as the input from Task 7.7 to the DDI and SDMX collaboration. In the process, we endeavour to give a picture of the context and development of the DDI and SDMX collaboration. Finally, we discuss our contribution to the creation of a lasting cooperation between national statistical institutes (NSIs) and data archives (DAs), the significance of collaboration between stakeholders, and future directions of metadata collaboration.

For the reader's convenience, we have also added a Glossary of Abbreviations at the end of this report.

⁷ DwB D7.1 Metadata standards - usage and needs in NSIs and data archives, 2012. (Ibid).

⁸ D7.2/D7.3 [Deliverables merged together] Standards with future relevance for European Social Science data infrastructure - Needs, key areas, rules & best practices in Metadata Standard selection & usage, 2014. (Ibid).

2 The Metadata Standards: DDI and SDMX

There are two metadata standards that are followed by a majority of the European NSIs and DAs: the SDMX metadata standard and the DDI metadata schema specification. There are various reasons to why those two metadata standards are widely accepted (DwB D7.2/D7.3, p. 11). Since both DDI and SDMX have been introduced in the previous DwB deliverables D7.1 and D7.2/D7.3, we will give just a short overview here.

2.1 Data Documentation Initiative (DDI)

The DDI is an international, structured metadata standard related to the observation and measurement of human activity. It started out in the mid-1990s as a replacement for traditional archival codebooks documenting research data, and then branched off to cover the research data lifecycle. Over time, the DDI XML specifications have evolved to add new coverage and functionality to respond to new user requirements.

While the origins of DDI are in the quantitative social sciences and in data archiving, it is increasingly being used by researchers and practitioners in other disciplines and in various organisations. It is also being used to document data types such as social media, biomarkers, administrative data, and transaction data. The specification itself is modular and can document and manage different stages of the data lifecycle, such as conceptualization, collection, processing, analysis, distribution, discovery, repurposing, and archiving (Vardigan, M. 2013). Currently, the “DDI Moving Forward” project is developing the next version of the standard. It will be an updated version that models the same information, and more, as DDI Lifecycle today, but in a more formal manner, and with increased ability to extend the content model in the future (DDI Moving Forward. 2014)⁹.

2.2 Statistical Data and Metadata eXchange (SDMX)

The Statistical Data and Metadata eXchange (SDMX)¹⁰ initiative is committed to developing and promoting technical standards that are suitable for the electronic exchange of statistical information, with an emphasis on aggregate data. SDMX is primarily used by the official statistics community for the exchange of time series data as well as for the dissemination of data and metadata.

The SDMX standards offer a common model, a choice of syntax and, for XML, a choice of data formats which support the exchange of any type of statistical data. The SDMX Information Model, presented using UML, provides a way of modelling statistical data, and defines the set of metadata constructs used for this purpose (SDMX Standards: Section 1. Framework for SDMX Technical Standards Version 2.1. 2011).

⁹ Detailed information about DDI Moving Forward is available at the DDI Alliance Collaboration wiki: <https://dditools.atlassian.net/wiki/display/DDI4/DDI+Home> [Accessed 12.2.2015]

¹⁰ SDMX website: <http://sdmx.org/> [Accessed 12.2.2015]

3 Collaboration and Contribution by WP7

3.1 The Landscape

The field of metadata standards is overwhelming and complex. One of the main reasons for the intricacy has been a practical one: for an organisation, there is usually a pressing need to get a system running up to support a specific task. As a result, various discipline, organisation or country-specific metadata efforts have led to duplicate or overlapping standards (DwB D7.2/D7.3, p. 42).

Task 7.7 has tried to recognise and provide opportunities for communications between the metadata communities. In the context of NSIs, SDMX is the main metadata standard, and data archives use DDI. Our specific aim was to help bring these two communities together. However, the idea that SDMX and DDI are complementary standards and could be used together in powerful ways is not new. As a matter of fact, SDMX and DDI were intentionally designed to align themselves with each other as well as with other metadata standards (Gregory, A. and Heus, P. 2007).

During the project, there has been significant progress in the collaboration of the DDI and SDMX metadata communities. This progress started already before DwB, and can to a great extent be credited to the active role of the UNECE and the Conference of European Statisticians Steering Group on Statistical Metadata (METIS) in this process.

Traditionally, the main role of metadata in a statistical organisation was to support the production of official statistics. In 2004, the participants of the Joint UNECE-Eurostat-OECD Work Session on Statistical Metadata (METIS) meeting noted that statistical organisations were missing a common framework for their metadata systems. They published the Generic Statistical Business Process Model (GSBPM) in 2008, and the model was continuously adopted by NSIs in the following years.

Certain similarities between the GSBPM and the DDI Lifecycle Model lead to growing interest in NSIs to use DDI Lifecycle to document the earlier phases of the statistical production process (particularly microdata), while SDMX standards were seen as more appropriate for macro data (Vale, S. 2010). In April 2011, the METIS Steering Group approved a Strategy Paper for Promoting the Common Metadata Framework (Promoting the Common Metadata Framework. 2011).

In the Introduction to the Common Metadata Framework website¹¹ it is observed that:

- a statistical metadata system should be a tool to facilitate the efficient functioning and further development of the whole statistical information system
- statistical metadata is a developing field and several projects dealing with different aspects of statistical metadata management have been conducted; and that
- the metadata work continues to progress through the regular meetings of the joint UNECE/Eurostat/OECD work sessions on statistical metadata.

¹¹ <http://www1.unece.org/stat/platform/display/metis/1.++Introduction> [Accessed 14.1.2015]

In the last few years, metadata discussions have been actively going on in several places, different levels, and in various contexts.

In Task 7.7 we decided to foster the discussions and collaboration by organising a metadata seminar, by participating in the groups initiated by METIS, and by attending DDI- and SDMX-related events. Several WP7 team members have participated in the discussions and various events but as Task Leader, Mari Kleemola (UTA-FSD) has been primarily liable for the work. During the project period, she was also the DDI Alliance Expert Committee Vice Chair (2010-2013) and a member of the DDI Alliance Executive Board (2013-2015).

3.2 DwB Workshop on Metadata Standards

December 2011, Gothenburg

DwB organised a joint workshop on metadata standards in collocation with the European DDI User Conference EDDI 2011. The workshop was held on the premises of the University of Gothenburg, Sweden on 7th December 2011 and it brought together more than twenty participants from NSIs, DAs and standard developing organisations. For many, this was the first “cross-disciplinary” metadata event.

The workshop provided an opportunity to establish a common framework where the DwB stakeholders (DAs, NSIs, researchers) could situate their own activities and identify the metadata standards relevant to them. The presentations and discussions focussed on SDMX, GSBPM and DDI as well as on the needs of end users. Interestingly, not more than a year earlier, the most common heard question was “can we use DDI and SDMX together?” in Gothenburg, the question had changed to “how can we use DDI and SDMX together?”.

The workshop contributed to the DwB’s goal of building more co-operation at the European level between the NSIs and the DAs. Cooperation builds on trust, and one way to help develop trust is creating networks and personal connections between individuals. The workshop also helped to chart the interests of NSIs and DAs, and to recognise specific areas where there was scope for cooperation.

The agenda and presentations are available at the DwB website¹².

3.3 DDI SDMX Dialogue Group 2010-2013

An informal dialogue process between DDI and SDMX standards began in the 2010 EDDI meeting in Utrecht and was active during 2010-2013. The dialogue engaged representatives of the SDMX Secretariat and the DDI Alliance as well as other stakeholders who had an

¹² http://www.dwbproject.org/events/workshop_eddi.html [Accessed 15.1.2015]

interest in the two standards and how they can work together. The dialogue was endorsed by both the DDI Alliance and the SDMX¹³.

The goal of the dialogue process was to explore how the two standards can best be used together, and to see if the standards groups can better support their users (Gregory. 2011). DwB WP7 members attended the Dialogue meetings and teleconferences actively. Iris Alfredsson (UGOT-SND) attended the SDMX/DDI Dialogue session on 30th. March 2011 in Lisbon, and Mari Kleemola attended the SDMX/DDI Dialogue teleconference on 4th. August 2011 and the Dialogue Session on 5th. October 2011 in Geneva. Since February 2012, Mari Kleemola represented DwB in the Microdata Access Task Team of the Dialogue, which held several teleconferences during 2012.

By attending the Dialogue, DwB WP7 members were able to identify further opportunities for networking and collaboration, and to follow the progress in the metadata field. In the meetings, they were able to advocate and bring forward both researchers' needs and the DwB's objective of building a single point of access for microdata in Europe. Specifically, DwB WP7 contributed to the use case of remote access of microdata.

Parallel to the SDMX/DDI Dialogue, work on the Generic Statistical Information Model (GSIM) was taking place, and the version 1.0 of GSIM was released in December 2012.

The GSIM is a key element of the strategic vision prepared by the High-Level Group for the Modernization of Statistical Production and Services (HLG), and endorsed by the Conference of European Statisticians. The GSIM is a reference framework of conceptual information objects and as a such, it has complementary relationship with various standards. There are a number of synergies between use of GSIM as a reference framework and the application of representation standards like SDMX and DDI (Generic Statistical Information Model (GSIM): Communication Paper. 2013).

Hence, the SDMX/DDI Dialogue was followed by the GSIM to DDI to SDMX mapping work.

3.4 GSIM/DDI/SDMX Mapping 2013

The GSIM to DDI to SDMX mapping was a formal project initiated by "The High-level Group for the Modernisation of Statistical Production and Services" (HLG). The work took place in 2013, and it included detailed mappings between the information objects in GSIM and those in the information models of DDI and SDMX. The aim was to identify any issues affecting the coherence of these standards, and propose solutions where possible.

The main outputs were mappings between the standards¹⁴ and feedback to the groups developing DDI and SDMX.

¹³ SDMX/DDI Dialogue - Overview Page
<http://www1.unece.org/stat/platform/display/metis/SDMX+DDI+Dialogue+--+Overview+Page> [Accessed 15.1.2015]

¹⁴ The mappings are available at <http://www1.unece.org/stat/platform/display/gsim/GSIM+and+standards> [Accessed 12.2.2015]

All in all, the SDMX and DDI standard developing bodies collaborated fruitfully in this process. One outcome of the mapping work was that some changes to the DDI Lifecycle 3.2 were considered in order to make variable descriptions of microdata better. The DwB WP7 members participated in the mapping work and were thus able to share the lessons learned especially with the DDI Alliance, and contribute to the DDI Alliance's endorsement of the work.

Since the mapping work proved successful, several statistical organisations expressed an interest in further developing this work to create a set of DDI profiles (High-level Group for the Modernisation of Statistical Production and Services. 2014). These profiles identify the specific parts of the DDI standard to use in particular scenarios related to the production of official statistics¹⁵.

Mari Kleemola attended the DDI/GSIM mapping work and the DDI Profiles team, both of which worked via teleconferences. Since January 2014, the work has continued under the new Modernisation Committee on Standards, which reports to HLG. The DwB Work Package 7 has not participated in this work.

3.5 Workshop on Statistical Metadata (METIS): Implementing the GSBPM and Combining Metadata Standards, October 2011

The main substantive topics for the METIS Workshop in October 2011, hosted in Geneva by UNECE, were DDI and GSBPM (Generic Statistical Business Process Model). The presentations discussed issues such as DDI in statistical organisations, national implementations of the GSBPM, process quality management and metadata flows within the GSBPM. The presentations gave an overview of the business and metadata practices in various statistical institutions. The practices vary to a great extent, but many NSIs have adopted or were planning to adopt the GSBPM, which will probably mean more harmonised processes, metadata and statistics in the future. Arofan Gregory's presentation¹⁶ on DDI and in particular the DDI Lifecycle concept inspired lively discussion and there was increasing interest in DDI from the NSIs. The discussions also revealed the need for "SDMX Lite" and

¹⁵ The profiles are:

- Basic Technical Objects (BTO) Profile
- Variable Profile
- Represented Variable Profile
- Questionnaire Profile
- Codelist Profile

They are available at <http://www1.unece.org/stat/platform/display/gsim/DDI+Profiles> [Accessed 12.2.2015]

¹⁶ Arofan Gregory: Overview of DDI, available at: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.33/2011/mtg2/WP_1_Arofan.ppt [Accessed 12.2.2015]

“DDI Lite”. It was pointed out that there is a mechanism for solving this problem: the statistical organisations need to agree on a common profile of most important objects and features in each standard. Several participants also expressed interest in the DwB Metadata Workshop on Metadata Standards in Gothenburg (see p. 10).

Mari Kleemola represented DwB at this METIS Workshop and also attended the collocated DDI/SDMX Dialogue meeting that took place on 5th October. The Workshop Agenda, Summary and presentations are available at UNECE’s website¹⁷.

3.6 Workshop on Strategic Developments in Business Architecture in Statistics, November 2012

This Workshop was hosted by UNECE in Geneva, and it was attended by by representatives of the High-Level Group for the Modernisation of Statistical Production and Services (HLG), the Bureau of the Conference of European Statisticians, and twenty-one expert groups whose work is related to the modernisation of statistics. The Workshop focussed on the Generic Statistical Information Model (GSIM) project. It was noted, for example, that GSIM will facilitate implementation of standards like DDI and SDMX. However, there is a need to agree on how the standards will be used to represent information in a standard way. Brian Pink (Head of the Australian Bureau of Statistics, ABS) observed that thinking about managing information as corporate resources will mean that it is necessary to accept solutions that are optimal for the organisation but may not always be optimal for all individual business areas. This requires cultural changes in the NSIs.

DwB was represented at the Workshop by Roxane Silberman (Scientific Coordinator DwB) and Mike Priddy (DANS-KNAW). Mari Kleemola attended the Workshop representing the DDI Alliance. The Workshop Agenda, Summary and presentations are available at UNECE’s website¹⁸.

¹⁷ UNECE/Meetings: Workshop on Statistical Metadata: Implementing the GSBPM and Combining Metadata Standards, 5 - 7 October 2011, <http://www.unece.org/stats/documents/2011.10.metis.html#/> [Accessed 13.2.2015]

¹⁸ UNECE/Meetings: Workshop on Strategic Developments in Business Architecture in Statistics, 7 - 8 November 2012, <http://www.unece.org/stats/documents/2012.11.hlgbas.html#/> [Accessed 13.2.2015]

4 Discussion and Future Directions

All in all, Task 7.7 has been a continuous networking effort that has been running the whole duration of the project. However, the main bulk of work and effort took place during the first half of the project in 2011-2013 when data archives and statistical institutes were still taking their first steps in metadata collaboration. A milestone in this collaboration was the DwB Workshop on Metadata Standards in 2011 in Gothenburg that brought together metadata experts from NSIs, DAs and standard developing organisations like DDI Alliance and SDMX. The meeting turned out to be a successful opportunity for sharing practises and for networking. In Gothenburg we also saw how the concern - “can we use DDI and SDMX together?” - had shifted from a possibility to advanced uses cases and practices “using DDI and SDMX together”.

From the beginning of the project it was clear that one of the best avenues for enhancing the DDI/SDMX collaboration were the METIS groups; first of these was the informal “SDMX/DDI Dialogue” group, which was followed by the “GSIM to DDI to SDMX Mapping”, a formal project initiated by The High-level Group for the Modernisation of Statistical Production and Services. Endorsing these Working Groups, and participating in their work, gave us a vantage point to this swiftly moving field.

Beyond the participation in working groups and seminars as part of our Task 7.7 efforts, WP7 members have actively taken part in discussions about metadata standards and practices in various other forums and events. These include, for example, the IASSIST and EDDI conferences, the DDI Alliance meetings and committees and CESSDA seminars. All the discussions outside the actual DwB project work have allowed us to influence and shape the changing metadata landscape. Similarly, since the WP7 members represent both data archives and national statistical institutes, the metadata discussions within the WP7 group as well as the resulting professional networks have been and continue to be highly beneficial.

Looking ahead, possibly the most exciting developments in metadata standards from a DDI/SDMX perspective are the ongoing DDI Moving Forward Project and the work done under the Modernisation Committee on Standards. The connection between the metadata communities is good and there is much potential for partnerships and collaboration that can move all stakeholders forward.

DwB has been one of the initiatives that has paved the way for this new level of collaboration. The contribution of WP7 in this process has not been in the form of papers or presentations; instead we have organised a successful workshop and participated in ongoing discussions and actual standards development work, bringing forward researchers’ metadata needs. The work has created networks and opportunities for collaboration and shaped the way metadata issues are approached in the NSIs. We have helped gain the critical mass needed to affect the current metadata practises and to advance the use of SDMX and DDI standards in collaboration. This kind of common understanding of metadata standards and their use is crucial when the aim is to build a single point of access for micro data in Europe.

It will be interesting to see how NSIs and DAs will use the standards in the future. Will they, for example, use standards to run data and metadata systems in fully machine-actionable

ways? And will standards and their use evolve to cover new kinds of data, for example big data, administrative data, linked data, social media data, experimental data and health research data?¹⁹

In recent years, one of the big movements in science has been Open Access - the notion that results (both publications and data) from publicly financed research should be freely available to researchers and also to wider audiences. The European Union has committed itself to the cause and states that the information should benefit European companies and citizens to the full (European Commission. 2013a). Under the research and innovation programme Horizon 2020, the Commission has introduced a new element: the use of Data Management Plans that outline how research data will be handled during a research project, and after it is completed. The beneficiaries of EU research funding must make it possible for third parties to access the data, including associated metadata (European Commission. 2013b).

A new initiative, the Research Data Alliance (RDA), began in August 2012 and is supported by the European Commission, the National Science Foundation and other U.S. agencies, and the Australian Government²⁰. The RDA's working groups are focusing on "the key challenges which must be overcome in order to create a thriving datasharing landscape" and one of their observations is that "metadata practice needs to be improved in order to help discovery and reuse" (RDA Europe Magazine. 2014²¹). Currently there are several metadata-related Working Groups, for example Metadata Standards Directory Working Group, Data Citation Working Group, and Standardisation of Data Categories and Codes Working Group²².

Recent research on sharing research data suggests that data infrastructure and standards are needed at various levels and stakeholders, and guided by dedicated data policies of research funders, research disciplines, research institutions, publishers, data centres and repositories. Guiding policies can be established, e.g. when a research community agree on standards for data formats and documentation, while publishers can set standards and requirements for data files. Finally data infrastructure is needed in the form of repositories that care for sustainable and long termed availability of these data and the required context information (Van den Eynden, V. and Bishop, L. 2014). In this respect the report "Incentives and motivations for sharing research data: researcher's perspectives" (ibid) recommends:

"Continue to invest in data infrastructure that also provides rich context, detailed metadata and even a narrative account of the data creation. The kind of infrastructure researchers find most useful is where research data, papers and other outputs or resources are jointly available within a single data resource." (Ibid p.35)

¹⁹ These issues are discussed in DwB D7.2/7.3 (2014) and D7.6 (2015), and they are also in the agenda on the New Techniques and Technologies for Statistics (NTTS) 2015 conference. See <http://www.cros-portal.eu/content/ntts-2015-programme> [Accessed 16.2.2015]

²⁰ The Research Data Alliance website on Organisation see: <https://rd-alliance.org/organisation.html> [Accessed 16.2.2015]

²¹ RDA Europe Magazine - 2nd issue. 21, Nov 2014. Publications & Reports. <http://europe.rd-alliance.org/documents/publications-reports/rda-europe-magazine-2nd-issue> [Accessed 16.2.2015]

²² RDA website Working and Interest Groups, <https://rd-alliance.org/groups> [Accessed 16.2.2015]

It is worth noting that the idea of documenting data and creating metadata for the purposes of storing and reusing is not new. Social scientists started to store digital data and recognised the potential of sharing data via data archives, i.e. research infrastructures, already in the early 1950s. The various social science data archives were actually the very first institutes to handle and preserve digital material (Doorn et al. 2007).

Now this story is about to repeat, however under changing conditions of an expanding data landscapes: the larger scientific community is opening its data for re-use. The social science data archive community will be able to share their long experience with other scientific disciplines in this respect. However new metadata challenges are arising due to new data sources and advanced approaches in the interdisciplinary use of domain specific produced data, analytical results and / or research methods.

Thus development of standards and metadata systems needs to reflect various requirements and fields of the data they shall manage and inform about. While content related needs are slightly going to change over time and related to scientific substance and communication, the challenges to develop metadata-driven systems and infrastructures are to consider.

Technical issues in shaping a future oriented metadata roadmap will be covered and discussed in the DwB deliverable D7.4 on “Software development and metadata standards”. Content driven metadata topics in adjacent areas of social science research will be described in DwB report D7.6 “Metadata standards and practices in related disciplines and standards for linking different sources”

5 Conclusions

It is obvious that there is not and probably never will be a single metadata standard that would be simple, cover everything and satisfy all the metadata needs. The solution is to apply several complementary standards and relevant conceptual models. The most important ones for the NSIs and DAs are currently DDI and SDMX as well as GSIM and GSBPM.

Continuous work, discussions and collaboration is needed on both developing the existing and new standards and models, and also on finding out how the standards can be used together in an effective way. In many cases, cooperation starts at individual level. However, in order to gain sustainability, organisations need to commit themselves to developing, maintaining and using the standards. Political decisions and funding are needed to create research environments where sharing data is routine and data infrastructures that enable data opening and sharing.

At the moment, the connection between the metadata communities is good and there is great potential for partnerships and collaboration that can move all stakeholders forward. Dwb has been one of the initiatives that has paved the way for new levels of collaboration.

It is easy to picture a future where the stakeholders have together achieved the goal of more efficient and compatible production of good-quality data and metadata, which in turn will result in easier discoverability and reusability of metadata and data as well as easier access to microdata for researchers.

6 References

Note: [last visit] indicate last access to a linked resource in the reference list.

Data without Boundaries.

DwB website:

<http://www.dwbproject.org/about/>

[14.1.2015]

DDI Moving Forward Process Summary. 2014.

DDI Alliance website:

<http://www.ddialliance.org/ddi-moving-forward-process-summary>

[15.1.2015]

Doorn, Peter and Tjalsma, Heiko. 2007.

Introduction: archiving research data.

Archival Science 7(1), 1-20.

DOI:10.1007/s10502-007-9054-6

DwB D5.2 Report and databank documenting national OS Microdata. 2013.

Available at:

http://www.dwbproject.org/export/sites/default/about/public_deliverables/dwb_d5-2_databank-national-survey_report_final2.pdf

[14.1.2015]

DwB D7.1 Metadata standards - usage and needs in NSIs and data archives. 2012.

Available at:

http://www.dwbproject.org/export/sites/default/about/public_deliverables/dwb_d7-1_metadata-standards-usage_report.pdf

[14.1.2015]

DwB D7.2/D7.3 Standards with future relevance for European Social Science data infrastructure - Needs, key areas, rules & best practices in Metadata Standard selection & usage. 2014.

Available at:

http://www.dwbproject.org/export/sites/default/about/public_deliverables/dwb_d7-2_7-3_future-metadata-standards-usage-selection_integrated-report.pdf

[14.1.2015]

DwB D8.2 Metadata Model. 2012.

Available at:

http://www.dwbproject.org/export/sites/default/about/public_deliverables/d8_2_metadata_model_report.pdf

[14.1.2015]

European Commission. 2013a.
Scientific Data: open access to research results will boost Europe's innovation capacity.
IP/12/790: 17/07/2012.

Available at:

http://europa.eu/rapid/press-release_IP-12-790_en.pdf

[13.2.2015]

European Commission. 2013b.
Guidelines on Data Management in Horizon 2020. Version 1.0, 11 December 2013.

Available at:

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[13.2.2015]

Generic Statistical Information Model - GSIM

Overview:

<http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model>

[19.3.2015]

Communication Paper for a General Statistical Audience (Version 1.1, December 2013).

Available at:

<http://www1.unece.org/stat/platform/display/gsim/GSIM+Communication+Paper>

[11.2.2015]

GSIM and other standards

Available at:

<http://www1.unece.org/stat/platform/display/gsim/GSIM+and+standards>

[19.3.2015]

GSIM and other standards: DDI Profiles

Available at:

<http://www1.unece.org/stat/platform/display/gsim/DDI+Profiles>

[19.3.2015]

Gregory, Arofan and Heus, Pascal. 2007.

DDI and SDMX: Complementary, not Competing, Standards.

Open Data Foundation.

Available at:

www.opendatafoundation.org/papers/DDI_and_SDMX.pdf

[15.1.2015]

Gregory, Arofan. 2011.

The Standards Working Together.

A presentation at the Workshop on Metadata Standards, Gothenburg, 7th. December, 2011.

Available at

http://www.dwbproject.org/export/sites/default/events/doc/dwb_metadata_standards_session4_07122011.pdf

[15.1.2015]

The High-level Group for the Modernisation of Statistical Production and Services - HLG Overview

<http://www1.unece.org/stat/platform/display/hlgbas/High-Level+Group+for+the+Modernisation+of+Statistical+Production+and+Services>

[19.3.2015]

HLG Meeting, 5th. March 2014. 2014. Outcomes and Progress of HLG Projects.

Available at

<http://www1.unece.org/stat/platform/download/attachments/99484985/4%20HLG%20Projects%20Outcomes%20and%20Progress.docx?version=1&modificationDate=1395218009745&api=v2>

[15.1.2015]

METIS. Statistical Metadata; Information system (METIS-wiki) provided by UNECE

Available at

<http://www1.unece.org/stat/platform/display/metis/METIS-wiki>

[19.3.2015]

New Techniques and Technologies for Statistics – NTTS

Conference 2015, 10-12 March 2015 in Brussels

Available at

<http://www.cros-portal.eu/content/ntts-2015-programme>

[19.3.2015]

Promoting the Common Metadata Framework. 2011.

Strategy Paper Approved by the METIS Steering Group. April 2011.

Available at

<http://www1.unece.org/stat/platform/download/attachments/8683549/Promoting%20the%20Common%20Metadata%20Framework.doc?version=1&modificationDate=1305895861545&api=v2>

[14.1.2015]

RDA. Research Data Alliance

Overview

<http://europe.rd-alliance.org/>

[19.3.2015]

RDA Working and Interest Groups

Available at:

<https://rd-alliance.org/groups>

[13.2.2015]

RDA Europe Magazine 2nd issue 21 November, 2014.

Available at:

http://europe.rd-alliance.org/sites/default/files/report/RDA_Europe_Magazine_2014.pdf

[13.2.2015]

SDMX DDI Dialogue.

Overview Page.

<http://www1.unece.org/stat/platform/display/metis/SDMX+DDI+Dialogue+-+Overview+Page>

[14.1.2015]

SDMX Standards: Section 1 - Framework for SDMX Technical Standards

Version 2.1, April 2011.

Available at:

http://sdmx.org/wp-content/uploads/2011/04/SDMX_2-1_SECTION_1_Framework.pdf

[15.1.2015]

UNECE. United Nations Economic Commission for Europe

Overview Page.

<http://www.unece.org/>

[19.3.2015]

UNECE. Workshop on Statistical Metadata:

Implementing the GSBPM and Combining Metadata Standards 5 - 7 October 2011

<http://www.unece.org/stats/documents/2011.10.metis.html#/>

[19.3.2015]

UNECE. Workshop on Strategic Developments in Business Architecture in Statistics, 7 - 8 November 2012

<http://www.unece.org/stats/documents/2012.11.hlgbas.html#/>

[19.3.2015]

Vale, Steven. 2010.

Exploring the relationship between DDI, SDMX and the Generic Statistical Business Process Model.

Second Annual European DDI User Group Meeting (EDDI).

Utrecht, Netherlands 8-9 December 2010.

Available at

<http://www1.unece.org/stat/platform/download/attachments/57835554/EDDI+paper.pdf?version=1>

[13.2.2015]

Van den Eynden, Veerle and Bishop, Libby. 2014.
Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective.
A Knowledge Exchange Report.

Available at:

www.knowledge-exchange.info/Default.aspx?ID=733

[13.2.2015].

Vardigan, Mary. 2013.

Strategic Priorities of the Data Documentation Initiative (DDI) Alliance.

Working Paper in Work Session on Statistical Metadata, Geneva 6-8 May 2013.

Available at:

<http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2013/WP6.pdf>

[15.1.2015]

7 Glossary of Abbreviations

CESSDA	The Council of European Social Science Data Archives
DA	Data Archive
DDI	Data Documentation Initiative
DwB	Data without Boundaries
EDDI	Annual European DDI User Conference
GSIM	Generic Statistical Information Model
GSBPM	The Generic Statistical Business Process Model
Horizon 2020	EU Research and Innovation programme 2014-2020
HLG	High-Level Group for the Modernisation of Statistical Production and Services
IASSIST	International Association for Social Science Information Services & Technology
METIS	Statistical Metadata; Information system (METIS-wiki) provided by UNECE
NSI	National Statistical Institute
OECD	Organisation for Economic Cooperation and Development
RDA	Research Data Alliance
SDMX	Statistical Data and Metadata eXchange
UNECE	United Nations Economic Commission for Europe

