RWX: Foundations of DDI in Scientific Literature

Read-Write-Execute (RWX) wants to draw attention towards DDI related publications. In this first column I want to present some foundations of DDI in scientific literature. (I thank Achim Wackerow for his suggestions.) At the same time a bibliography of DDI articles, working papers and presentations starts to be available at Bibsonomy.org (<http://www.bibsonomy.org/user/ddialliance?lang=en>) with easily reusable bibliographic metadata. Suggestions for RWX or the bibliography are greatly appreciated. [Knut Wenzig](http://www.diw.de/cv/en/kwenzig), [kwenzig@diw.de](mailto:kwenzig@diw.de).

In her paper “The DDI matures: 1997 to the Present” Mary Vardigan (2013), the former Director of the DDI Alliance, presents a timeline of the conceptual and organizational development of the world of DDI. While there was a first meeting already in 1995, the year 1997 (to which the title refers) is the year of the “instantiation in XML” (Vardigan 2013: 45). DDI startet (in versions 1 and 2) to describe data sets by the still supported codebook approach. From version 3 on the scope has been broadened to the whole lifecycle of data, with the concept of data collection as a starting point and finally enabling repurposing and reuse of elements.[[1]](#footnote-0) Referring to “Developing a Model-Driven DDI Specification” (Participants in 2012 Dagstuhl Seminar on DDI Moving Forward, 2012) the paper ends with a list of high level design goals, which the upcoming DDI version tries to achieve.

The earliest entry in Mary Vardigans references is the paper “Providing Global Access to Distributed Data Through Metadata Standardisation - the Parallel Stories of NESSTAR and the DDI” submitted by the Norwegian Social Science Data Services, prepared by Jostein Ryssevik (1999). The “relative distance between the end-users of a statistical material and the production process” (p. 2) is identified as the fundamental problem to be solved. Especially by addressing this problem by providing discovery systems the need for metadata standards like DDI emerged. The authors recall, that DDI has been developed using the (then) new XML language and the defined XML code could contain the description of the document itself, of the study, the file, the variables, and other study-related materials. Already in this early paper RDF (Resource Description Framework[[2]](#footnote-1)) is described as an application “that provides the foundation for metadata interoperability across different resource description communities.” (p. 5) Using DDI as a language, the medium NESSTAR should deliver a great range of interconnected services and platforms. Even if the last release of the NESSTAR is more than [one year old](http://www.nesstar.com/), the ideas in the article - whether or not realized by the software - deserve to be revisited. Using the metaphoric antonym of Bazaars vs. Cathedrals the same authors (Ryssevik 2000) conceptualize their vision of - even then! - metadata systems which cover the complete life-cycle.

The article “The Data Documentation Initiative” by Grant Blank and Karsten Boye Rasmussen (2004) was published in Social Science Computer Review, one of the top ranked academic journals in the category “Information Science & Library Science”. The authors describe the requirements of data documentation in the social sciences, how DDI as an XML based standard can be used to store information presented in codebooks, and how “standardization creates new opportunities for software development to aid users.” (p. 314)

Today, after up to 20 years, we only can read and reevaluate those ideas, because people took the time to write them down. In this sense to contribute to the scientific inventories of knowledge should be understood as a best practise and an integral part of software-development for the acedemic community.

### **References** (also available at <http://www.bibsonomy.org/user/knutwenzig/RWX072016>)

Blank, G. & Rasmussen, K. B. (2004), 'The Data Documentation Initiative: The Value and Significance of a Worldwide Standard', Social Science Computer Review 22 (3), 307-318, [doi:10.1177/0894439304263144](http://doi.org/10.1177/0894439304263144).

Participants in 2012 Dagstuhl Seminar on DDI Moving Forward (2012), Developing a Model-Driven DDI Specification, DDI Working Paper Series (Other Topics) DDI Alliance, [doi:10.3886/DDIWorkingPaper04](http://doi.org/10.3886/DDIWorkingPaper04).

Ryssevik, J. & The Norwegian Social Science Data Services (1999), Providing Global Access to Distributed Data through Metadata Standardisation--the Parallel Stories of Nesstar and the DDI, Conference of European Statistics, UN/ECE Work Session on Statistical Metadata (Geneva, Switzerland, 22-24 September 1999), Working Paper 10, <http://www.unece.org/stats/documents/1999/09/metis/10.e.pdf>.

Ryssevik, J. & The Norwegian Social Science Data Services (2000), Bazaar Style Metadata in the Age of the Web--An 'Open Source' Approach to Metadata Development, Conference of European Statistics, UN/ECE Work Session on Statistical Metadata (Washington D.C., United States, 28-30 November 2000), Working Paper 4, <http://www.unece.org/fileadmin/DAM/stats/documents/2000/11/metis/4.e.pdf>.

Vardigan, M.; Heus, P. & Thomas, W. (2008), 'Data Documentation Initiative: Toward a Standard for the Social Sciences', International Journal of Digital Curation 3 (1), 107-113, [doi:10.2218/ijdc.v3i1.45](http://doi.org/10.2218/ijdc.v3i1.45).

Vardigan, M. (2013), 'The DDI Matures: 1997 to the Present', IASSIST Quarterly 37 (1--4), 45-50, http://www.iassistdata.org/sites/default/files/iqvol371\_4\_vardigan.pdf.

1. Version 3 is described in Vardigan, Heus, Thomas 2008. [↑](#footnote-ref-0)
2. Vardigan (2013: 48) expects, that RDF will be used in the upcoming version of DDI as a connection to the semantic web. [↑](#footnote-ref-1)