

Scenarios for the DDI-RDF Discovery Vocabulary

Johanna Vompras¹, Arofan Gregory², Thomas Bosch³ and Joachim Wackerow⁴

Abstract

Statistical domain experts - core members of the DDI Alliance Technical Committee, representatives of national statistical institutes and national data archives - and Linked Data community members have spent four years to develop the DDI-RDF Discovery Vocabulary (Disco). In total, 26 persons from 23 organizations and 12 countries have contributed to its development. Disco has emerged as a massive simplification of the DDI standard, optimized for querying using Semantic Web technologies such as SPARQL. Disco is intended to provide means to describe data by essential metadata to support the discovery of unit-record data sets and related metadata using RDF technologies in the Web of Linked Data.

In this paper, we introduce representative real world scenarios that show (1) the usage and importance of the DDI-RDF Discovery Vocabulary for the social, behavioural, and economic sciences and the Linked Data community and (2) the interaction between Disco and other vocabularies. These scenarios have in common that they represent real information needs of researchers. The Disco specification itself contains example data which can be consulted to get details of how to construct Disco instance data and to get a feeling of the full potential of Disco to represent metadata on statistical data.

Introduction

The DDI-RDF Discovery Vocabulary presents a good model for vocabulary development: it was the joint product of collaboration between members of the social, behavioural, and economic sciences community, DDI experts and implementers, and members of the Linked Data Community. It re-uses other popular vocabularies wherever possible, and can be applied to the research data from many different domains, rather than being specific to a single set of domain data (e.g., census). And it is based on a proven and widely implemented metadata model, sufficient for the demanding requirements of discovering and describing unit-record research data.

This document describes the scenarios which the DDI-RDF Discovery Vocabulary was designed to support. These are not formal UML use cases - instead, they are scenarios for the possible use of the vocabulary, based on an analysis of existing search interfaces and known behaviors for those looking for research data.

The process around these scenarios is to define them, posit the thinking of the researcher/user seeking to find data, to identify needed classes and properties in the vocabulary, and then to render the search as it might be implemented. An examination of the implementation of the searches/queries (in SPARQL) will show if the vocabulary is optimized for the most common scenarios.

DDI-RDF Discovery Vocabulary

Figure 1 gives an overview of the conceptual model containing a small subset of the DDI-XML specification¹. More detailed descriptions of the DDI-RDF Discovery Vocabulary are given in the Disco specification² and two conference papers (Bosch et al. 2012 and Bosch et al. 2013). To understand the DDI Discovery Vocabulary, there are a few central classes, which can serve as entry points. The first of these is Study. A **Study** represents the process by which a data set was generated or collected. Literal properties include information about the funding, organizational affiliation, abstract, title, version, and other such high-level information. In some cases, where data collection is cyclic or on-going, data sets may be released as a **StudyGroup**, where each cycle or "wave" of the data collection activity produces one or more data sets. This is typical for longitudinal studies, panel studies, and other types of "series". In this case, a number of Study objects would be collected into a single StudyGroup.

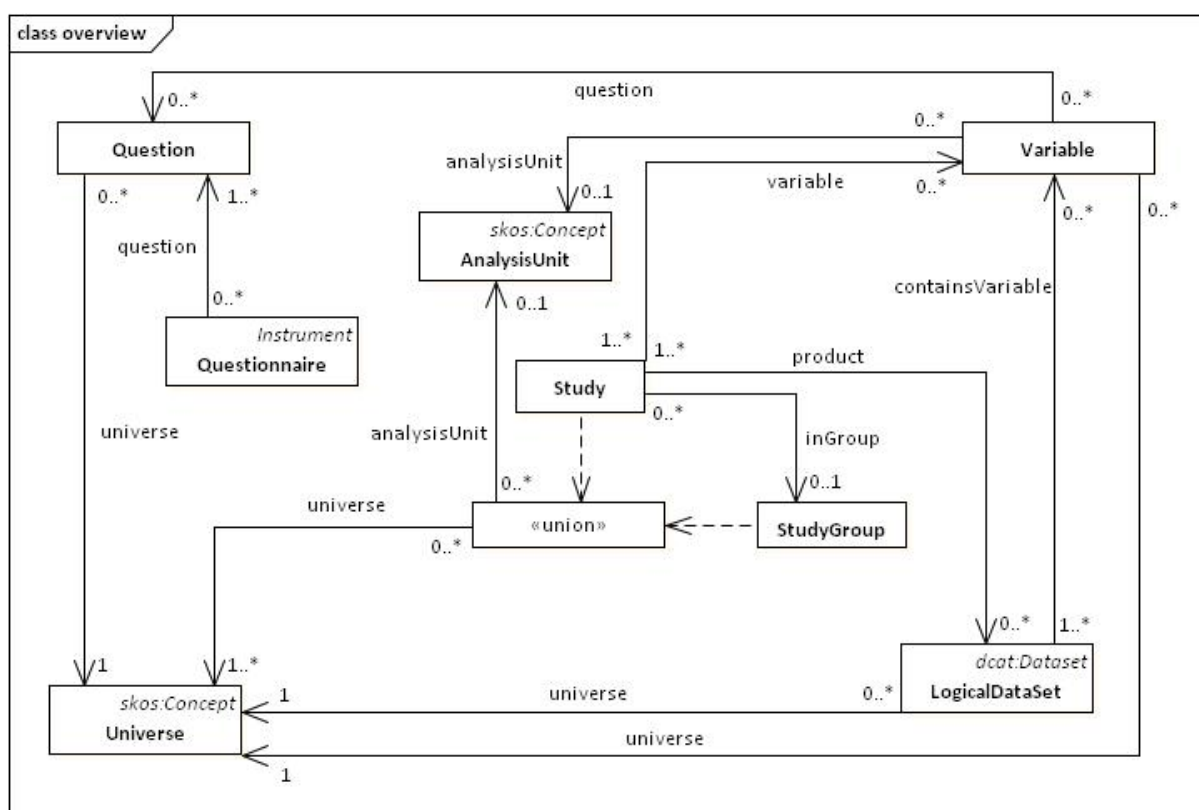


Figure 1. Conceptual Model of the DDI-RDF Discovery Vocabulary

Data sets have two representations: a logical representation, which describes the contents of the data set, and a physical representation, which is a distributed file holding that data. It is possible to format data files in many different ways, even if the logical content is the same. **LogicalDataSet** represents the content of the file (its organization into a set of Variables). The LogicalDataSet is an extension of the dcat:DataSet. Physical, distributed files are represented by the **DataFile**, which is itself an extension of dcat:Distribution.

¹ <http://www.ddialliance.org/Specification/>

² <http://rdf-vocabulary.ddialliance.org/discovery>

When it comes to understanding the contents of the data set, this is done using the **Variable** class. Variables provide a definition of the column in a rectangular data file, and can associate it with a **Concept**, and a **Question** (the Question in the **Questionnaire** which was used to collect the data). Variables are related to a **Representation** of some form, which may be a set of codes and categories (a “codelist”) or may be one of other normal data types (dateTime, numeric, textual, etc.). Codes and Categories are represented using SKOS concepts and concept schemes.

Data are collected about a specific phenomenon, typically involving some target population, and focusing on the analysis of a particular type of subject. These are respectively represented by the classes **Universe** and **AnalysisUnit**. If, for example, the adult population of Finland is being studied, the AnalysisUnit would be individuals or persons.

Unique identifiers for specific DDI versions are used for easing the linkage between Disco metadata and the original DDI-XML files. Every element can be related to any foaf:Document (DDI-XML files) using dcterms:relation. Any entity can have version information (owl:versionInfo). However, the most typical cases are the versioning of the metadata (the DDI-XML or the RDF file), the versioning of the study (as a study goes through the life cycle from conception through data collection) and the versioning of the data files. Every LogicalDataSet may have access rights statements (dcterms:accessRights) and licensing information (dcterms:license) attached to it. Studies, logical datasets, and data files may have a spatial (dcterms:spatial), temporal (dcterms:temporal), and topical (dcterms:subject) coverage.

Searching for Data Sets by Accessibility

1. Description of Scenario

This scenario describes how to retrieve data sets which fulfill particular access conditions. Many research data sets are not freely available, and access conditions may restrict some users from accessing some data sets. It is common to want to search only for those data sets which are either publicly available, or which have specific types of licensing/access conditions. Access conditions vary by country and institution. Users may be familiar with the specific licenses which apply in their own context. It is expected that the researcher looking for data might wish to see the data sets which meet specific access conditions or license terms. Here, a researcher is using a tool which will generate a SPARQL query which returns the titles of data sets which are publicly available under the Canadian Data Liberation Initiative Community policy.

2. Typical Queries

One typical query would be to find titles of data sets which are publicly available under the Canadian Data Liberation Initiative Community policy. Optionally, this SPARQL query gives links to the rights statement and the license. Execute the following SPARQL query to answer this question.

3. Needed Properties

- dcterms:title
- disco:isPublic
- dcterms:accessRights
- skos:prefLabel

- rdfs:seeAlso
- dcterms:license

4. Technical Implementation

```
SELECT ?logicalDataSetTitle ?rightsStatementURL ?licenseDocument
WHERE {
  ?logicalDataSet a disco:LogicalDataSet ;
    dcterms:title ?logicalDataSetTitle ;
    disco:isPublic ?isPublic ;
    dcterms:accessRights ?rightsStatement .

  ?rightsStatement skos:prefLabel ?rightsStatementLabel .

  FILTER (
    ?isPublic = "true" &&
    ?rightsStatementLabel = "Data Liberation Initiative Community"
  )

  OPTIONAL {
    ?rightsStatement rdfs:seeAlso ?rightsStatementURL .
  }
  OPTIONAL {
    ?logicalDataSet dcterms:license ?licenseDocument .
  }
}
```

Searching for Studies by Publishing Agency

1. Description of Scenario

Researchers are often aware of the organizations which disseminate the kind of data they want to use. This scenario shows how a researcher might wish to see the studies which are disseminated by a particular organization, so that the data sets which comprise them can be further explored and accessed.

2. Typical Queries

“Show me all the studies for the period 2000 to 2010 which are disseminated by the ESDS service of the UK Data Archive.”

3. Needed Properties

- dcterms:title
- dcterms:temporal
- disco:startDate
- disco:endDate
- dcterms:publisher

4. Technical Implementation

```
SELECT ?studyTitle
WHERE {
  ?study a disco:Study ;
    dcterms:title ?studyTitle ;
    dcterms:temporal [
```

```

    a dcterms:PeriodOfTime ;
    disco:startDate "2000"^^xsd:date;
    disco:endDate "2010"^^xsd:date;
  ] ;
  dcterms:publisher ?publisher .

  FILTER (regex(?publisher, "ESDS service of the UK Data Archive", "i"))
}

```

Searching for Studies by Producer or Contributor

1. Description of Scenario

Researchers are often searching for data attributed to a specific researcher or researcher group in a particular agency. This scenario shows how a researcher might wish to see the studies which are disseminated by a particular person, so that the data sets which comprise them can be further explored and accessed.

2. Typical Queries

“Show me all the studies people are working on from University of Example.”

3. Needed Properties

- dcterms:title
- dcterms:contributor
- org:memberOf
- rdfs:label

4. Technical Implementation

```

SELECT ?studyTitle ?contributor
WHERE {
  ?study a disco:Study ;
    dcterms:title ?studyTitle ;
    dcterms:contributor ?contributor ;

  ?contributor a foaf:Person ;
    org:memberOf [
      rdfs:label "University of Example"@en;
    ] .
}

```

Finding Relevant Studies by Free Text Search/Keyword

1. Description of Scenario

The most natural way of searching for data is to formulate the information need by using free text terms and to match them against the most common metadata, like title, description, abstract, or unit of analysis. This scenario shows how a researcher might search for relevant studies which have a particular title or keywords assigned in order to further explore the data sets attached to them. The definition of an analysis unit might help to directly determine which data sets the researcher wants to download afterwards.

2. Typical Queries

Find all studies (titles, abstracts) with questions about commuting to work.

3. Needed Properties

- disco:instrument
- dcterms:title
- dcterms:abstract
- disco:questionnaire
- disco:question
- disco:questionText

4. Technical Implementation

```
SELECT ?studyTitle ?studyAbstract
WHERE {
  ?study a disco:Study ;
    disco:instrument ?instrument ;
    dcterms:title ?studyTitle ;
    dcterms:abstract ?studyAbstract .

  ?instrument disco:questionnaire ?questionnaire .
  ?questionnaire disco:question ?question .
  ?question disco:questionText ?questionText .

  FILTER (regex(?questionText, "commut.*work"))
}
```

Searching for Studies by Coverage/Universe

1. Description of Scenario

Researchers are often interested in finding studies or issues concerning a particular part of a population or/and geographic location. In addition, this scenario shows how this kind of information can be accessed and gives examples of using information from external data sets to characterize coverage and universe.

2. Typical Queries

“Search for studies about social and historical problems caused by the late German migration and resettlement from Poland after 1945.”

To make it more complicated, we could also consider the temporal change of “spatial coverage” (e.g, the fact that after 1945, the area of Silesia or Hither Pomerania was attached to Poland) and expand the query “resettlement from Poland” by “resettlement from Silesia”.

3. Needed Properties

- dcterms:title
- disco:coverage
- dcterms:subject

- dcterms:spatial
- dcterms:temporal
- skos:definition
- disco:startDate

4. Technical Implementation

```

SELECT ?studyTitle
WHERE
{
  ?study rdf:type Study;
    dcterms:title ?studyTitle;
    dcterms:subject ?subject;
    dcterms:spatial ?location;
    dcterms:temporal ?periodOfTime.
  ?subject rdf:type skos:Concept;
    skos:definition ?subjectDefinition.
  ?location rdf:type dcterms:Location;
    skos:definition ?locationDefinition.
  ?periodOfTime rdf:type dcterms:PeriodOfTime;
    disco:startDate ?startDate.
  Filter(
    ?subjectDefinition = 'migration'.
    ?locationDefinition = 'Germany'.
    ?startDate = 1945.)
}

```

Searching for Studies by Data Source

1. Description of Scenario

It is often the case that researchers use more than one data collection sources (or instruments) as key indicators. This can help them to validate the data and to gain a richer understanding of the topics being explored. However, in order to address similar research questions or for the replication of the experiments a researcher might be only interested - - due to his/her resources, available equipment, or expertise -- in data collected by only a specific instrument (e.g., by an online survey).

2. Typical Queries

“Find all studies about *‘social preference’* with data collections performed online or computer-based, and additionally used eye tracking methods on subject persons to analyse their motives for the decisions they made.”

3. Needed Properties

- dcterms:title
- dcterms:subject
- disco:instrument
- skos:definition
- dcterms:description

4. Technical Implementation

```
SELECT ?studyTitle
WHERE
{
  ?study rdf:type Study;
    dcterms:title ?studyTitle;
    dcterms:subject ?subject;
    disco:instrument ?instrument.
  ?subject rdf:type skos:Concept;
    skos:definition ?subjectDefinition.
  ?instrument rdf:type disco:Instrument;
    dcterms:description ?instrumentDescription.
  Filter(
    ?subjectDefinition = 'social preference'.
    ?instrumentDescription = 'performed online or computer-based, and
    additionally used eye tracking methods on subject persons to analyse their
    motives for the decisions they made')
}
```

Searching for Data Sets Using Variables

1. Description of Scenario

In the easiest case, a search for data sets is performed by the variables. A researcher might be interested in finding data sets by searching the properties of variables, e.g., their associated concepts and represented variables, representation, description, and analysis unit.

2. Typical Queries

Find study groups (titles, abstracts) where the study uses the species variable and has a variable defined as *Bufo alvarius*.

3. Needed Properties

- disco:inGroup
- disco:variable
- dcterms:title
- dcterms:abstract
- disco:concept
- disco:basedOn

4. Technical Implementation

```
SELECT ?studyGroupTitle ?studyGroupAbstract
WHERE {
  ?study a disco:Study ;
    disco:inGroup ?studyGroup ;
    disco:variable ?variable .

  ?studyGroup dcterms:title ?studyGroupTitle .
  ?studyGroup dcterms:abstract ?studyGroupAbstract .

  ?variable disco:concept ?variableConcept .
}
```



```

?variableConcept rdf:type skos:Concept .
FILTER (regex(?variableConcept, "species", "i"))

?variable disco:basedOn ?representedVariable .

?representedVariable rdf:type disco:RepresentedVariable .
?representedVariable disco:concept ?variableDefinitionConcept .
FILTER (regex(?variableDefinitionConcept, "Bufo alvarius", "i"))
}

```

Browsing Data Collections Using Concepts and Subject Keywords

1. Description of Scenario

It is a common strategy for browsing data collections to use concepts or predefined categorizations instead of the exact textual search on the data. The mostly used approach is to have a (high-quality) domain-specific classification of the underlying data collection and additionally link the concepts of this classification to external thesauri or controlled vocabularies which describe the world as Linked Data.

2. Typical Queries

“Find questions about the education system”

3. Needed Properties

- disco:concept
- skos:topConceptOf
- dcterms:title

4. Technical Implementation

```

SELECT DISTINCT ?questionTitle
WHERE {
  ?question a disco:Question ;
  disco:concept [
    skos:topConceptOf conceptSchemeEducationSystem
    skos:definition 'education system'@en ] ;
  dcterms:title ?questionTitle .
}

```

Searching for Data Sets Using Related Questions

1. Description of Scenario

Variable search by one or more questions across studies.

2. Typical Queries

“Find all variables which are populated by specific questions (description).”

3. Needed Properties

- disco:question
- disco:questionText

4. Technical Implementation

```
SELECT ?variable
WHERE
{
  ?variable
    rdf:type disco:Variable,
    disco:question ?question.
  ?question
    rdf:type disco:Question,
    disco:questionText "How old are you?".
}
```

Searching for Data Sets Using Classifications

1. Description of Scenario

Variable searched by Representation with skos:ConceptSchemes.

2. Typical Queries

"Find all variables which have a union of datatypes and an unordered code (skos:ConceptScheme) list as representation."

3. Needed Properties

- disco:representation

4. Technical Implementation

```
SELECT ?variable
WHERE
{
  ?variable
    rdf:type disco:Variable.
    disco:representation ?unionrdfsDatatypeskosConceptScheme.
  ?unionrdfsDatatypeskosConceptScheme
    rdf:type union ( rdfs:Datatype, skos:ConceptScheme ).
}
```

Searching for Reusable Questions Using Related Concepts

1. Description of Scenario

Researchers often use existing question banks or existing study catalogues as the basis for their own questionnaire construction which is a very crucial and important part of the research. A poorly designed questionnaire can be a reason for a low response rate achieved, the quality of responses obtained, and consequently the conclusions drawn from the study results.

The next scenarios (K-M) provide examples how systems for questionnaire design or question suggestion could benefit from the usage of a DDI-RDF Discovery Vocabulary.

2. Typical Queries

“Find all questions which comprise ‘age’ in the context of Education -> School (subconcept)”.

3. Needed Properties

- disco:concept
- skos:definition

4. Technical Implementation

```
SELECT ?question
WHERE
{
  ?question
    rdf:type disco:Question;
    disco:concept ?skosConcept.
  ?skosConcept
    rdf:type skos:Concept;
    skos:definition "age".
}
```

Searching for Reusable Questions Using Related Variables

1. Description of Scenario

A very common scenario is to search cross-survey for questions by variable-level information (like name, label, representation) to retrieve question texts which might be relevant for a researcher to design his own questionnaire.

2. Typical Queries

“Find questions and their corresponding response categories which are using the variable ‘age’”

3. Needed Properties

- skos:notation
- disco:representation
- disco:question
- skos:inScheme

4. Technical Implementation

```
SELECT ?question ?category
WHERE
{
  ?variable
    rdf:type disco:Variable;
    skos:notation "age";
    disco:representation ?unionrdfsDatatypeskosConceptScheme;
    disco:question ?question.
  ?question
    rdf:type disco:Question.
  ?unionrdfsDatatypeskosConceptScheme
    rdf:type union ( rdfs:Datatype, skos:ConceptScheme ).
}
```

```
?category
  rdf:type skos:Concept;
  skos:inScheme ?unionrdfsDatatypeskosConceptScheme.
}
```

Searching for Reusable Questions by the Question Text

1. Description of Scenario

The pure textual search on question text is the easiest way to access relevant questions. The DDI-RDF Discovery Vocabulary could be used here to incorporate “similar” terms to extend the query parameters and thus optimize the query results.

2. Typical Queries

“Find questions in which subject persons were asked: *‘How do you assess your physical fitness?’*”

3. Needed Properties

- disco:questionText

4. Technical Implementation

```
SELECT ?question
WHERE
{
  ?question
    rdf:type disco:Question;
    disco:questionText "How do you assess your physical fitness?".
}
```

Searching for Reusable Questions Using Related Universe/Coverage

1. Description of Scenario

In the questionnaire design phase, a possible scenario could be that a researcher wants to extend a given set of general questions about a topic by questions which have been posed to a particular part of population (or within a different geographic location). To do that, he/she has to find questions by searching on the related coverage in the universe links of studies.

2. Typical Queries

A set of questions about questions which have been posed to collect data about the *quality of life among people aged 75 and over* in the UK is already given. Now the researcher would like to extend his set by questions which discover the correlation between the *quality of life and professional status of employed people*.

3. Needed Properties

- disco:universe
- skos:definition

4. Technical Implementation

```
SELECT ?question
WHERE
{
  ?question
    rdf:type disco:Question.
    disco:universe ?universe
  ?universe
    rdf:type disco:Universe;
    skos:definition "quality of life and professional status of employed people".
}
```

Searching in RDF Metadata of a DDI Instance

1. Description of Scenario

There are many advantages for querying and browsing DDI instances enriched with RDF. For instance, this could provide more extended exploration possibilities of the study space than with keywords or give researchers the possibility to discover additional knowledge which is implicitly stored in linked data sources or facilitate extended aggregation. In contrast to the inflexible and pure metadata search, the use of RDF opens new perspectives of developing new (fuzzy) similarity measures which can be applied to data for comparison purposes or to rank query results.

2. Typical Queries

“Search for studies which have been conducted in cooperation by at least three large international funding organizations.”

3. Needed Properties

- dcterms:title
- disco:fundedBy

4. Technical Implementation

```
SELECT ?study ( COUNT( ?organization ) AS ?count )

WHERE
{
  ?study
    rdf:type disco:Study;
    dcterms:title ?studyTitle;
    disco:fundedBy ?organization.
  FILTER ( ?count >= 3 )
  ?organization
    rdf:type org:Organization
}
GROUP BY ?study
ORDER BY desc ( ?count
)
```

Conclusion

Many archives and other organizations have large amounts of data, sometimes publicly available, but often confidential in nature, requiring applications for access. Such organizations use the proven and highly detailed DDI-XML standard for describing data sets of this type. We make use of the DDI specification to create a simplified version of this model by choosing the DDI elements which are needed for the discovery purpose. With the background of the broadness and complexity of DDI, DDI-RDF focuses on a subset of DDI. The selection relies on scenarios which are oriented on the discovery of data in the Linked Data context and possible usage within the web of data. The conceptual model is based on XML Schemas describing the DDI domain data model with extensions that partly borrow from existing vocabularies and partly reside in a new DDI vocabulary.

When we consider how such a standard could be used as the basis for an RDF vocabulary, we realize that the requirements are very different. The most obvious use case is that of discovery, given that much of the data is highly confidential, and that access to the data must be applied for in most cases. Further, the challenges of searching the Web of Linked Data are enormous – the sheer range of information is incredibly broad. Thus, the almost twelve hundred metadata fields within DDI is itself a problem. The DDI model must be significantly reduced in complexity to be meaningful to cover these requirements. The fact that DDI is not specific to any particular research domain or type of research data is a positive feature, however, as the range of data to be exposed into the Web of Linked Data is also very broad.

In this paper, we introduce representative real world scenarios that show (1) the usage and importance of the DDI-RDF Discovery Vocabulary for the social, behavioural, and economic sciences and the Linked Data community and (2) the interaction between Disco and other vocabularies. These scenarios have in common that they represent real information needs of researchers. We describe the scenarios which the DDI-RDF Discovery Vocabulary was designed to support. These are not formal UML use cases - instead, they are scenarios for the possible use of the vocabulary, based on an analysis of existing search interfaces and known behaviors for those looking for research data. The process around these scenarios is to define them, posit the thinking of the researcher/user seeking to find data, to identify needed classes and properties in the vocabulary, and then to render the search as it might be implemented. An examination of the implementation of the searches/queries in SPARQL shows if the vocabulary is optimized for the most common scenarios.

Acknowledgements

The work described in this paper has been started at the first workshop on “Semantic Statistics for Social, Behavioural, and Economic Sciences: Leveraging the DDI Model for the Linked Data Web”³ at Schloss Dagstuhl - Leibniz Center for Informatics, Germany in September 2011. This work has been continued at these three meetings: follow-up working meeting in the course of the 3rd Annual European DDI Users Group Meeting (EDDI11)⁴ in Gothenburg, Sweden in December 2011, second workshop on “Semantic Statistics for Social, Behavioural, and Economic Sciences: Leveraging the DDI Model for the Linked Data Web”⁵ at Schloss Dagstuhl - Leibniz Center for Informatics, Germany in

³ <http://www.dagstuhl.de/11372>

⁴ <http://www.iza.org/eddi11>

⁵ <http://www.dagstuhl.de/12422>

October 2012, and follow-up working meeting at GESIS - Leibniz Institute for the Social Sciences in Mannheim, Germany in February 2013. This work has been supported by contributions of the participants of the events mentioned above: Archana Bidargaddi (NSD - Norwegian Social Science Data Services), Thomas Bosch (GESIS - Leibniz Institute for the Social Sciences, Germany), Sarven Capadislis (Bern University of Applied Sciences, Switzerland), Franck Cotton (INSEE - Institut National de la Statistique et des Études Économiques, France), Richard Cyganiak (DERI, Digital Enterprise Research Institute, Ireland), Daniel Gilman (BLS - Bureau of Labor Statistics, USA), Arofan Gregory (ODaF - Open Data Foundation, USA and DDI Alliance Technical Implementation Committee), Rob Grim (Tilburg University, Netherlands), Marcel Hebing (SOEP - German Socio-Economic Panel Study), Larry Hoyle (University of Kansas, USA), Yves Jaques (FAO of the UN), Jannik Jensen (DDA - Danish Data Archive), Benedikt Kämpgen (Karlsruhe Institute of Technology, Germany), Stefan Kramer (CISER - Cornell Institute for Social and Economic Research, USA), Amber Leahey (Scholars Portal Project - University of Toronto, Canada), Olof Olsson (SND - Swedish National Data Service), Heiko Paulheim (University of Mannheim, Germany), Abdul Rahim (Metadata Technologies Inc., USA), John Shepherdson (UK Data Archive), Dan Smith (Algenta Technologies Inc., USA), Humphrey Southall (Department of Geography, UK Portsmouth University), Wendy Thomas (MPC - Minnesota Population Center, USA and DDI Alliance Technical Implementation Committee), Johanna Vompras (University Bielefeld Library, Germany), Joachim Wackerow (GESIS - Leibniz Institute for the Social Sciences, Germany and DDI Alliance Technical Implementation Committee), Benjamin Zapilko (GESIS - Leibniz Institute for the Social Sciences, Germany), Matthäus Zloch (GESIS - Leibniz Institute for the Social Sciences, Germany).

References

Bosch, T., Cyganiak, R., Wackerow, J., and Zapilko, B. Leveraging the DDI Model for Linked Statistical Data in the Social, Behavioural, and Economic Sciences. International Conference on Dublin Core and Metadata Applications, (2012), 46–55.

Bosch, T., Cyganiak, R., Gregory, A., Wackerow, J.: DDI-RDF Discovery Vocabulary: A Metadata Vocabulary for Documenting Research and Survey Data. In: Proceedings of the Linked Data on the Web Workshop, 22th World Wide Web Conference. CEUR Workshop Proceedings, Aachen (2013)

¹ Johanna Vompras | University Bielefeld Library, Bielefeld, Germany | E-Mail: johanna.vompras@uni-bielefeld.de

² Arofan Gregory | Open Data Foundation, Tucson, USA | E-Mail: agregory@opendatafoundation.org

³ Thomas Bosch | GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany | E-Mail: thomas.bosch@gesis.org

⁴ Joachim Wackerow | GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany | E-Mail: joachim.wackerow@gesis.org