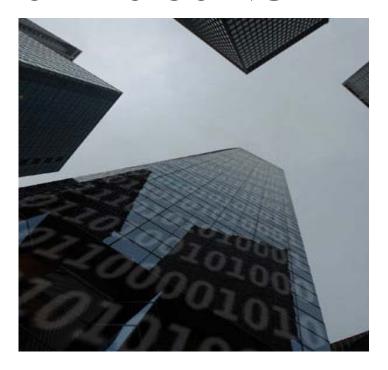# GROUPING OF SURVEY SERIES USING DDI 3

By Uwe Jensen with Sanda Ionescu, Mari Kleemola, Agostina Martinez, Wendy Thomas, Mary Vardigan, and Wolfgang Zenk-Möltgen

04/30/2010	DDI Working Paper Series – Use Cases, No. 6

Grouping of Survey Series Using DDI 3

BY UWE JENSEN WITH SANDA IONESCU, MARI KLEEMOLA, AGOSTINA MARTINEZ, WENDY THOMAS, MARY VARDIGAN, AND WOLFGANG ZENK-MÖLTGEN

ABSTRACT

Ongoing data series tend to be complex in terms of management, processing, and analysis. A team in Europe made an exploration and evaluation of DDI 3 to determine its suitability for marking up and managing cross-national comparative surveys like the International Social Survey Programme data. Their findings may help others interested in organizing survey series using the DDI's Group and Resource Package mechanisms.

BACKGROUND

This use case had its origins in the Council of European Social Science Data Archives (CESSDA) Preparatory Phase Project, Work Package 8, which mandated an evaluation of DDI 3 and its major features, to be completed during 2009. As part of this evaluation, work package staff began to explore the use of DDI 3 in archives and in particular, how they might leverage DDI's organizing structures to facilitate processing and management of as well as access to complex survey series. Of particular interest to data archives are the new features that support reuse of metadata within the workflows, sharing common metadata within the archive and research communities, and managing collections of complex survey data.

There are several ongoing data series in Europe, including the International Social Survey Programme (ISSP), the Eurobarometer series, the European Values Survey (EVS), and the European Social Survey (ESS). For the purposes of this use case, we selected the ISSP as a good example of the layers of complexity to be addressed with such a series. However, the overarching goal is to provide guidance on the DDI grouping structure in a broad way to encourage the use of DDI for other series as well.

USE CASE / REQUIREMENTS

About ISSP

The ISSP is a continuing annual programme of cross-national and cross-cultural surveys covering topics important for social science research. The programme started in 1984 with four founding members - Australia, Germany, Great Britain, and the United States — and has grown to 45 member countries from all over the world in 2008. ISSP researchers concentrate on developing questions that are meaningful and relevant to all countries, and can be expressed in an equivalent manner in all participants' languages.

The ISSP is an example of a data series featuring comparison by design. The surveys are fielded as substantively themed modules — for example, the Role of Government -- which may be repeated at various intervals to gauge trends and patterns across the participating countries. ISSP questions are fielded in each country in the native language(s). Countries may also gather country-specific data during survey

administration. All countries' common ISSP data from the fielded module are then merged into a cumulative file by the data archive responsible for the integration process. All information exchanged between the integrating archive and the countries is documented in English.

At the time this use case was written, the ISSP consisted in total of 10 topical modules, surveyed at 23 temporal instances until 2007, with 522 country samples from 44 different countries.

The ISSP data support several approaches for analysis: changes within a range of topics in a single country over time, differences and similarities between countries on various topics, and most commonly, examination of specific topics over time, and between countries.

The integration process does not deal with the full data collected in each national survey used to field the ISSP modules The ISSP modules are generally fielded as part of a national survey like the ALLBUS in Germany. However, these national surveys are often made available with full national language(s) like the Finnish versions of the ISSP documented by the Finnish Social Science Data Archive (FSD). National archives or agencies may choose an alternative packaging or grouping approach to meet their detailed needs and requirements. A related use case that focuses on the 2006 Finnish ISSP Role of Government survey is described in "DDI 3: An Archive's Perspective on DDI3" as published in the DDI Working Paper Series.

About DDI Group, Resource Package, and Comparison

DDI 3 offers a number of mechanisms to support comparison and reuse of metadata within series. These include Group, Resource Package, and Comparison. Each mechanism facilitates organization of metadata around slightly different structures and can be used alone or in combination depending on the metadata themselves and the applications that the metadata need to support. For this use case it was necessary to decide which framework was most appropriate for each type of information in the series and to understand the impact of the selected framework on different points in the collection and integration process.

Group

A Group can be comprised of Study Units and Sub-Groups. Studies may be grouped for a number of reasons but generally fall into the category of grouping by design or ad hoc groups. Grouping by design takes place when studies are either intended to be a series or when a repetition of the study takes place. The key factor is that the second study in the series is intended to inherit features of the first study (questions, variables, study design, universe, etc.) for the purpose of comparability. Grouping allows one to define which parts of the major components are shared, where overrides take place, and how to relate or link data in one study to data in a subsequent survey.

When using inheritance within groups to show comparability – or even just to reuse metadata – it is important to understand how local overrides work, as this can impact the way the metadata are grouped. Within each group, all metadata are inherited down the grouping structure. At any level, it is possible to override any inherited metadata using the Add, Update, or Delete attributes. Essentially one should organize the metadata so that the most stable sections are near the top of the tree, with sections of greater change and variability in the lower Sub-Groups.

The payoffs of this type of grouping structure include the following:

- Reuse of metadata – Information needs to be entered only once and is then inherited down the tree structure

- Elimination of error-prone and expensive data entry – Because metadata are reused, only differences and deviations need to be documented, which saves time and money

- Consistency – Standardization provides consistency in structure

- Implicit comparability – With the inheritance approach, comparability across years and other dimensions can be assumed because they use the same metadata

- Flexible approaches to data based on alternative grouping strategies – Users can organize data in the hierarchy that supports their needs. Once one grouping structure is established, it is possible to create new kinds of groups through referencing of the original information.

Resource Package

A Resource Package is a structure that allows for publication of metadata that are intended to be used by multiple studies whether they are related in a series or not. Resource Package can publish any maintainable module or scheme except for Group, Study Unit, or Local Holding Package. This mechanism is commonly used to publish schemes for reuse within organizations or research communities, such as industrial classification codes, geographic locations, commonly agreed upon concept descriptions, and question or variable structures.

A Resource Package that is published for reuse either within an organization or by an larger audience will have the attribute "isPublished" set to "true" in the published maintainable. This ensures the user that any future changes to the Resource Package will result in a version change and that the current version will not disappear, thereby breaking a link for future users of the metadata.

Comparison Module

While inheritance of metadata within a group or the reuse of metadata published in a Resource Package can express explicit comparison (equality), the expression of equality or similarity in all other situations requires the use of the Comparison Module (see the use case "Using DDI 3 for Comparison," which is part of the DDI Working Paper series). Currently comparison addresses specific sections of well defined metadata: Universe, Concept, Question, Variable, Category, and Code Scheme. Similarities and dissimilarities can be described for each of these element types and, in the case of Code Schemes, the process of translating one code scheme into another can be captured explicitly. In the case of comparison by design there is no need to express equalities that are described through the inheritance structure. However in the case of a change in question wording or content of the response domain, comparison between the old and new version could be captured in the Comparison Module.

Applying DDI 3 Data Management Structures to ISSP

The first step in this use case was to look at the types of information in the ISSP, which is basically a matrix of country, time, and topical coverage. In addition to a set of background variables asked in each survey, different sets of topical questions are grouped in topical modules, which are fielded approximately every five years. Table 1 shows the relationship of the number of countries participating in each year with respect to each topical module. This highlights the complexity of the ISSP survey series organization and helps to determine the implication of various nesting patterns for Groups and Sub-Groups within the DDI framework.

	ISSP Topical modules					Total. no. of participating countries	Total no. of different countries
						522	**44**
M1	**Citizenship**	**2004**					
	Countries	**38**				**38**	**38**
M2	**+ Environment**	**1993**	**2000**				
	Countries	21	26			**47**	**30**
M3	**+ Family & Changing Gender Roles**	**1988**	**1994**	**2002**			
	Countries	8	23	34		**65**	**36**
M4	**+ Leisure Time & Sports**	**2007**					
	Countries	**34**				**34**	**34**
M5	**+ National Identity**	**1995**	**2003**				
	Countries	23	34			**57**	**35**
M6	**+ Religion**	**1991**	**1998**	**2008**			
	Countries	17	31	x		**48**	**31**
M7	**+ Role of Government**	**1985**	**1990**	**1996**	**2006**		
	Countries	6	10	27	35	**78**	**38**
M8	**+ Social Networks**	**1986**	**2001**				
	Countries	7	28			**35**	**28**
M9	**+ Social Inequality**	**1987**	**1992**	**1999**	**2009**		
	Countries	10	17	26	x	**53**	**29**
M10	**+ Work Orientations**	**1989**	**1997**	**2005**			
	Countries	11	25	31		**67**	**38**

Table 1: Number of Countries Participating in Fielding Topical Modules by Year (x: Fielding not yet finished)

Note that not only does the count of countries fielding a topical module within a given year change, but the mix of countries may also change. For example, in M3 Family and Changing Gender Roles a total of 65 surveys were fielded over the three years by different countries. However, only 36 different countries were involved. Countries may join or drop out at any point in time[1]. This is important in that it helps to identify the most changeable part of the country/time/topic matrix, the countries involved in a specific time/topic combination.

Once the matrix showing the relationship was created, the next step was to think about the ISSP and where intended commonalities and differences are found. From the perspective of the integrating archive the primary commonality lies in the ISSP content modules because each question in the module is designed to be fielded across all participating countries.

---

[1] Reference: ISSP Modules - Study Overview

Additional areas of commonality are the set of standard demographic questions, and the common coding schemes for geography, occupation, etc. Occasional deviations and changes appear over time and it would be helpful to develop a means of capturing and comparing these differences as part of a larger grouping tool.

POSSIBLE SOLUTIONS and DESIGN CHOICES

The structure of DDI 3 allows for considerable flexibility in organizing the metadata to support different activities at different points in the data and metadata life cycle. The approach for capturing the design and evolution of a research project may differ from that of a data extraction and analysis system that is trying to facilitate discovery and analysis by topic, year, language, or country. While this level of flexibility allows users to organize metadata to meet their needs, it means that a number of decisions need to be made regarding which modules to use and what the optimal metadata structure would be to achieve current and future goals. For example, use of Group and the inheritance structures may be optimal for a closed system dealing with a study series in isolation, but may not be optimal if much of the content is heavily reused outside of the specific study series.

The ISSP integrating archive is concerned with the series as a whole and how the data contained in that series are commonly analyzed. Therefore the first step is looking at the series as a whole and determining primary commonalities and deviations within the ISSP series. If approached from the perspective of data collection, the initial organizing structure would probably be country and then time. However, in viewing it from a series perspective, the focus shifts to looking at topics over time and across countries. In other words, the interest of the user is focused on the ten topical modules and the related questions and variables. It is at first a static view on the present status of a completed and published survey series. This provides an initial organizing structure for the series.

While this addresses the organizing structure of the Group module, there is the secondary consideration of what metadata should be placed within the Group inline and what should be placed in a Resource Package and referenced by the Group structure. This decision is made by analyzing which metadata are specific to the ISSP series and which would be considered ISSP-independent metadata and placed more effectively in Resource Packages for external reuse.

From a broader life-cycle perspective design decisions should consider both the organizational and work principles of the ISSP. The approach selected by the use case was based upon:

- The ex-ante defined standards (survey design, basic questionnaire; question, and variable structure)

- The integrated versions of the single national ISSP survey instances

- The documentation and communication language, which is English

Building the Initial Group Structure

Exploring the commonalities within the ISSP as a complete series is an important step in the process of determining grouping structures. The ISSP is fielded in multiple countries on a periodic basis and the content is sub-ordered by the ten topical modules within the individual Study Units. While researchers may approach the ISSP from any of these perspectives, we have found that most researchers approach the data from a topical perspective. Because our purpose is to facilitate research using the integrated set of ISSP data, the extensive use of a topical approach drove a number of grouping decisions.

The decision to use the topical module as the top level in the Group structure was based on a pragmatic top-down view according to the present ISSP Internet presentation. It provides a view that is consistent and stable

over time. Although the metadata will be organized by topic, they could be organized by country or year as well. This initial organization does not preclude a later development of a more dynamic presentation, offering data users additional options for browsing study information and accessing data from other perspectives.

It is important that the structure be open and flexible enough to permit analysis by country or year and not just through an integrated topical perspective. Single country-year Study Units should be available in multiple languages (English and native languages) and country-dependent information (variables unique to a single country) could be derived from these country instances.

Building the initial group structure was based on the following selection and decision principles:

1. The top-level Group "ISSP" contains the information that is common to the whole series. This includes information such as Series Name and Description, the Concept Scheme, and common Universe structure, as well as Data Collection and Logical Product information on the core questions and variables found in all the country/year study instances (referred to hereafter as the individual Study Unit). The group "ISSP" also contains the primary set of ten topical Sub-Groups.

2. The ten Sub-Groups contain the metadata specific to the topical module (concepts, questions, and variables as well as specific purpose statements). Sub-Groups within each topic module reflect changes over time like versioning of a question and comprise the related country instances.

    A simplified example is presented by the Module "Environment" as one of ten ISSP Sub-Groups:

    - The Sub-Group provides the name, description, and purpose of the module. The module was fielded two times (1993, 2000). Metadata specific to this module and not inherited from the Group, such as additional universe or concept descriptions, would be placed here in the Sub-Group. The metadata for the module-related common set of questions (Data Collection; Q. 1 to Q.20) and the related schemes for categories and codes (Logical Product V1 to V40) are provided at this level.

    - The module Environment is further divided into two Sub-Groups for 1993 and 2007. Both Sub-Groups inherit the ISSP common core of variables and related metadata and the module Environment-specific content noted above. Each Sub-Group contains specific information on the timing of data collection and other information specific to each fielding of the module. In addition, 2007 would contain any information that has changed between the 1993 and 2007 fielding of the questionnaire such as the deletion, addition, or versioning of a question or variable.

3. Within each of the year Sub-Groups, the individual Study Units for each country taking part in the particular module are found. The Study Unit contains study-dependent metadata, such as the fieldwork-specific sample design (Data Collection), the native language(s) of the question wording, the country-specific variables, deviations on the question and variable level, and other materials like the national questionnaire(s), etc.

Specific Issues for Higher and Lower Group Levels

Top Level -- Group

Everything at the top level of the Group structure is inherited downward. Changes in these metadata will result in further Sub-Grouping over time so care should be taken to include information that is not only common to all the Sub-Groups, but that is relatively static over time. Changes in some metadata do not

require Sub-Grouping because specific information is referenced directly and the appropriate version will be used. For example, an Organization Scheme may change as individuals or organizations are created or versioned. However, an element in a specific Study Unit will reference a specific version of an organization or individual. The fact that the complete Organization Scheme is relevant to the complete series does not change. In the same way, a Concept or Universe Scheme can be modified without changing the specific content of the question or variable that references a specific concept or universe

Sub-Groups: Topical Modules

The majority of metadata in the ISSP series are specific to a topical module or an individual Study Unit and therefore the majority of metadata will be held in these levels. The Sub-Group at the topical level will contain the purpose of fielding of this particular topical questionnaire, the Data Collection elements describing the questions, and Logical Product information related to variables. Concept and Universe additions are generally made at this level.

Sub-Groups: Topical Modules/Year

Individual year Sub-Groups provide information on information that has changed over time such as the addition, removal, or versioning of a question or variable. In particular, Data Collection information on the overall data collection process that is common to all countries fielding this questionnaire is included here. For example, a new review process may be put in place or the questionnaire needs to be fielded during a specific month.

Study Units: Individual Country

At this point the Study Unit has inherited overall series information and core questions/variables from the Group, specific questions and variables, module description, and purpose from the Sub-Group: Topical Module, and time-related changes as well as collection information for the specific collection period from Sub-Group: Year. What is contained in the individual country's Study Unit is only that information specific to the country. This will include references to the organizations and/or individuals responsible for the collection, additional questions or variables specific to the country, and any country-level versioning of the standard question or variable set (i.e., an expanded response domain for a standard question).

Question and Variable Documentation from Sub-Groups to Study Units

The standard documentation of these core metadata elements is housed in the appropriate schemes in the Data Collection and Logical Product of the topical Sub-Group (i.e., Environment).

Local (country/year combination) changes in questions or variables are recorded using the attribute "action." Action allows one to specify changes that occur only in a specific instance (changes that *cannot* be inherited). The following are common examples of these local changes:

- Add: A specific question or set of questions may be asked in an individual country in order to collect related locally significant information.

- Delete: Some questions or variables do not occur in all instances of questionnaires for Environment. In this case the metadata would be deleted by entering the full object (question in this case) with the same id as the inherited object and setting the flag to "Delete".[2]
  Note that if this occurs at a Sub-Group level, (i.e., questions Q1 and Q 16 were not asked in 1993),

---

[2] This reflects a change in the DDI 3 documentation clarifying the behavior of the three values of the action attribute, published in a corrigendum on 2010-10-01.

this would either need to be handled as an additional Sub-Group splitting within the topical module or reiterated as a "Delete" for each country as the information related to these action codes cannot be inherited. Note also that the original question remains accessible.

- Update: A common example is a change in a small feature of an element that occurs only once. Variable names sometimes change from wave to wave, for example, but nothing else about the variable may change. The inheritance structure allows one to indicate that variable name is the only variable component that has changed and that it needs to be replaced as a local override. This is done by listing all values of VariableName that are valid at the local level even if previously listed at the Group level. All other properties of the Variable are inherited from the Group.

Native Language Issues from Group to Study Units

All language versions can be captured in the upper level Group or in Sub-Groups where language first occurs. This is done by replicating the element with different xml:lang codes. In this way users can easily understand the implicit comparison of the different language equivalents. At the local level of the Study Unit the language used can be set using the xml:lang value as an attribute of the Study Unit element tag. Countries that field multi-language questionnaires can include a Data Collection element within the Study Unit by reference indicating a single language on each.

See the grouping markup in Appendix C for examples of:

- Variables in multiple languages

- Sub-Groups on topic and time

- Study Units by country

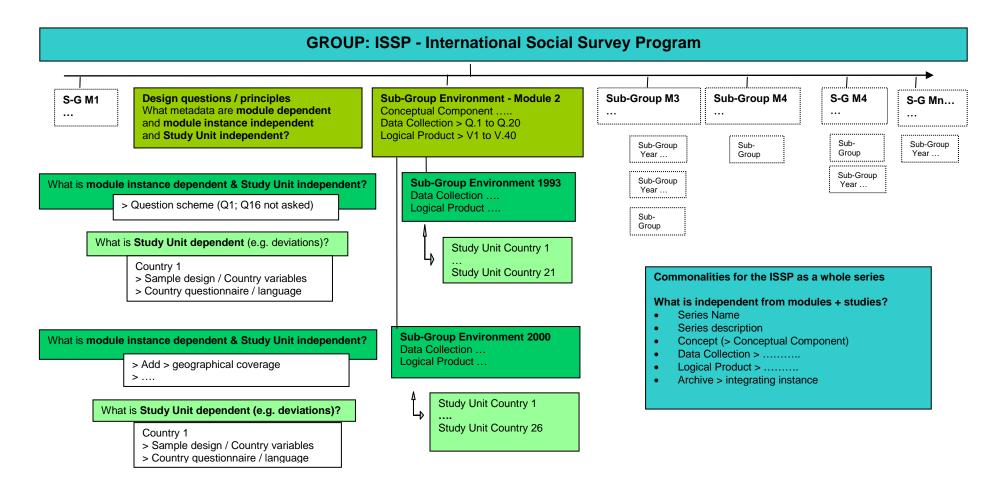- Use of the "Add," "Delete," and "Update" actions

GROUP: ISSP - International Social Survey Program

| S-G M1 … | **Design questions / principles**
What metadata are **module dependent**
and **module instance independent**
and **Study Unit independent?** | **Sub-Group Environment - Module 2**
Conceptual Component …..
Data Collection > Q.1 to Q.20
Logical Product > V1 to V.40 | **Sub-Group M3** … | **Sub-Group M4** … | **S-G M4** … | **S-G Mn…** … |

| Sub-Group Year … | Sub-Group | Sub-Group | Sub-Group Year … |

| Sub-Group Year … | | Sub-Group Year … | |

| Sub-Group | | | |

What is **module instance dependent & Study Unit independent?**

> Question scheme (Q1; Q16 not asked)

Sub-Group Environment 1993
Data Collection ….
Logical Product ….

What is **Study Unit dependent** (e.g. deviations)?

Country 1
> Sample design / Country variables
> Country questionnaire / language

Study Unit Country 1
…
Study Unit Country 21

Commonalities for the ISSP as a whole series

What is independent from modules + studies?
- Series Name
- Series description
- Concept (> Conceptual Component)
- Data Collection > ………..
- Logical Product > ……….
- Archive > integrating instance

What is **module instance dependent & Study Unit independent?**

> Add > geographical coverage
> ….

Sub-Group Environment 2000
Data Collection …
Logical Product …

What is **Study Unit dependent (e.g. deviations)?**

Country 1
> Sample design / Country variables
> Country questionnaire / language

Study Unit Country 1
….
Study Unit Country 26

Figure 1: Schematic Group structure of the ISSP with 10 topical modules

Modifications to the Initial Group Structure

Once the initial structure and basic hierarchy was determined for the first ISSP structure (see Figure 1) we saw the potential to eliminate some redundancies within the metadata by using another DDI grouping structure, Resource Package, in conjunction with Group (see Figure 2). Modifications were made to allow broader reuse of metadata found in the ISSP metadata currently packaged in the Group structure. The modified design is a combination of Group and Resource Package that facilitates reuse of metadata in two different ways:

- Makes standard metadata used in the ISSP series available for use in studies outside of the series

- Makes it easier to manage ISSP metadata within the series through Resource Packages

Study-Independent Metadata Placed in Resource Packages

Study-independent metadata refers to standard sets of categories and variables that might be used by multiple studies and audiences. This allows management of this metadata over time (such as version updates) without affecting the versioning of the ISSP Group. Thus, Resource Packages are recommended for:

- Countries (ISO)

- Regions (NUTS, etc.)

- Employment (ISCO, etc.)

- Standard Demographic variables

Study-Dependent Metadata Placed in Resource Package

Reorganizing study-dependent metadata in a combined Group / Resource Package approach allows optimized reuse of metadata within the survey series. This metadata design can minimize the amount of metadata for items such as standard answer scales or category schemes used across all topically grouped question wordings or variables. Technically, the Group inheritance mechanism is not able to share metadata across the borders of the designed Sub-Groups, but the RP mechanism facilitates this by reference. These category and code schemes could be held in the Group level metadata and then used by reference by various questions and variables in different Sub-Groups, but it is more easily managed within a Resource Package. This is particularly true when the addition of a country may require the addition of a new language label or text to a category or similar description.

The contents of the Resource Packages should be constructed from study-level information that may appear in multiple Study Units, not necessarily in an inheritance hierarchy. The value of moving these metadata into Resource Packages is that it makes it easier to version and provides implicit comparability between the use of the element in two studies that are not in an inheritance hierarchy. It also helps the user understand the interrelationships between the various studies whether they inherit from each other or not.

We distinguish between two types of possible Resource Packages in this context:

Type 1: Study Unit- and module-independent Resource Package (e.g., Category scheme, Code scheme / question schemes if they are used with standard wordings) for

- Standard Archive and ID variables, e.g., Study Number; Edition of the dataset, Respondent ID Number, Country/Sample, ISO code of country

- Standard answer scales in the topical sections, such as "Agree / Disagree …"

- Standard schemes for variables in the demographic section such as sex, age groups, marital status, religion, etc.

Type 2: Study Unit-dependent but module-independent Resource Package (i.e., Category scheme, Code scheme) for

- Country-related schemes to reuse variable structures on education, regions, parties, income, etc., within each topical module the country participated in

For both types it may be useful to analyze the rate of change (versioning) in ex-ante and ex-post defined metadata. A more in-depth understanding of the background for particular variables may help optimize the construction and use of Resource Packages.

The use of Resource Packages is an option for managing metadata that are reused within the series or more broadly. In any case the use of Resource Packages must be balanced with the needs of the responsible organizations involved in the collection and management of the metadata. Ideally those organizations who collect and manage metadata within a country or topical area can find a balance that will allow them to effectively share metadata and reorganize it to optimize different use cases, such as the perspective of the collecting country and the perspective of the organization supporting an integrated dataset for topical research.

Further Options to Reuse Metadata

The following options related to metadata reuse were not discussed for the ISSP use case but deserve further review:

- DDI profile to provide specifications for submitted DDI instances. The profile can be used for automated controls to verify that a submitted instance meets the expected or required coverage.

- Controlled vocabularies throughout an archive and or / series specific vocabularies
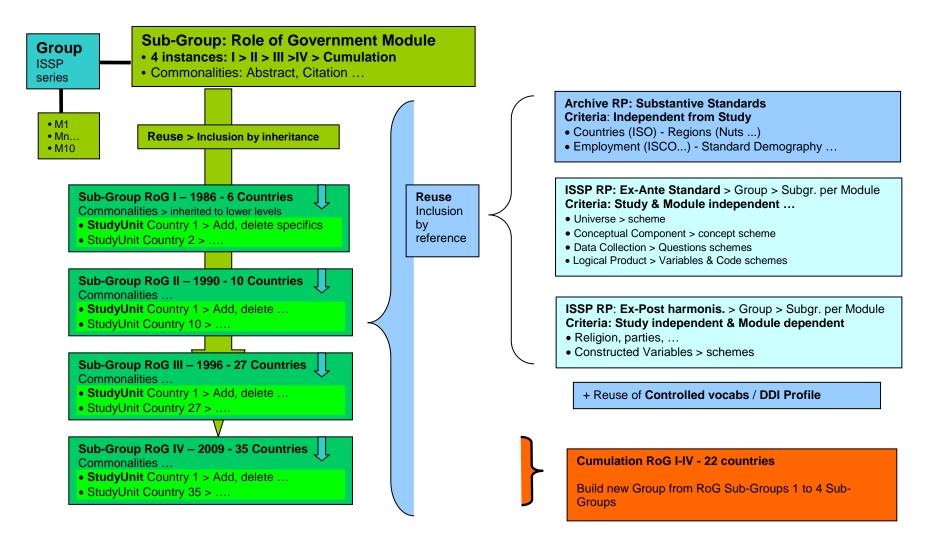
Group
ISSP
series

- M1
- Mn…
- M10

Sub-Group: Role of Government Module
- **4 instances: I > II > III >IV > Cumulation**
- Commonalities: Abstract, Citation …

Reuse > Inclusion by inheritance

Sub-Group RoG I – 1986 - 6 Countries
Commonalities > inherited to lower levels
- **StudyUnit** Country 1 > Add, delete specifics
- StudyUnit Country 2 > ….

Sub-Group RoG II – 1990 - 10 Countries
Commonalities …
- **StudyUnit** Country 1 > Add, delete …
- StudyUnit Country 10 > ….

Sub-Group RoG III – 1996 - 27 Countries
Commonalities …
- **StudyUnit** Country 1 > Add, delete …
- StudyUnit Country 27 > ….

Sub-Group RoG IV – 2009 - 35 Countries
Commonalities …
- **StudyUnit** Country 1 > Add, delete …
- StudyUnit Country 35 > ….

Reuse
Inclusion
by
reference

Archive RP: Substantive Standards
Criteria: **Independent from Study**
- Countries (ISO) - Regions (Nuts ...)
- Employment (ISCO...) - Standard Demography …

ISSP RP: Ex-Ante Standard > Group > Subgr. per Module
Criteria: Study & Module independent …
- Universe > scheme
- Conceptual Component > concept scheme
- Data Collection > Questions schemes
- Logical Product > Variables & Code schemes

ISSP RP: **Ex-Post harmonis.** > Group > Subgr. per Module
Criteria: Study independent & Module dependent
- Religion, parties, …
- Constructed Variables > schemes

+ Reuse of **Controlled vocabs** / **DDI Profile**

Cumulation RoG I-IV - 22 countries

Build new Group from RoG Sub-Groups 1 to 4 Sub-Groups

Figure 2: Advanced Reuse Combining Group and Resource Package (RP)

ISSUES / RESTRICTIONS

The construction of the metadata for the data cumulation of Role of Government instances I to IV has not been included in the use case as presented here. The idea to design a metadata structure for an existing data accumulation was meant to focus the overall scope of the available metadata on the Sub-Group "Role of Government" and down through the hierarchy of Sub-Groups and Study Units. A design for such a metadata structure is still on the agenda and might be added to an updated version of the working paper.

An issue open to further discussion concerns the general possibility to change the future workflow of the ISSP or EVS to allow for the primarily virtual integration of data (in addition to the physical integration) for all Study Units included in a fielded topical module or a new wave. In particular the creation of study level metadata (Study description) would benefit from the availability of various technical options to merge metadata from standardized country study descriptions with additional information obtained from the integrating archive. The desired consequence would be richer metadata in both versions of the study as presented to the community. If achieved, the metadata from the country fieldwork and the integrated (comparable) version would both be compliant with a defined project standard.

This use case has not explored the use of the Comparison Module in combination with the Group and Resource Package model to see how we could provide explicit documentation of comparable metadata not covered by inheritance and comparison through reuse. It would be helpful to explore the impact on advances in management, publication, and reuse of comparative and comparable data at large in the scientific community.

Some language issues remain to be resolved.

Ideally, this Group structure should be supported by others further upstream in the data life cycle at the point of data collection. This would ensure standardization across agencies and a consistent approach to data management. For example, the team fielding a questionnaire should make the determination that an item deviates from the equivalent item in the previous round and then pass that information down the system. In reality, this is usually not the case, so it may turn out that we need to try to influence others on these decisions.

However, many view the current system of data management and access as adequate. Thus, they may not want to make these types of changes without a clearer picture of the payoffs.

In addition, this combined Group and Resources Package structure requires an IT infrastructure for which there must be an organizational commitment. It will require a sustained investment in tool and system development as well as a development in the intellectual capacities of the organizations involved to plan and implement a common set of standardized metadata repositories. We need to explore the usability of such complex resources in daily documentation work. Technical options must balance the needs for series-specific processes with the general workflows of data and metadata management within an archive or research project.

A general modular DDI tool suite is required to capitalize on the advantages of DDI 3 for supporting seamless daily work routines. Tools to analyze comparability and to manage Group structure and their relations to Resource Packages will be needed to take full advantage of the possibilities presented in a move to DDI 3.

RESULTS

As outlined above, decisions were made on a grouping structure and what the groups will contain, as pictured in the diagram in Figure 1. A further step in integrating the use of Resource Packages and Groups was described along with a specific example for one of the ten topical ISSP modules was presented in the diagram in Figure 2.

We also provide some XML code in Appendix C to show the structure of Groups and Sub-Groups.

OUTLOOK / CONCLUSION

Beyond those issues, this grouping approach should be generic enough to apply to the major European data series. The CESSDA PPP ends at the close of 2009, and Work Package 8 will provide a working paper, which is a report of the exploration of DDI3.

The evaluation of the design options to reuse metadata using the DDI 3 modules Group and Resource Package supports the following conclusions:

1. Group supports the organization of dependent studies (comparative by design) to capture metadata one time within a Group or Sub-Group and allow inheritance by lower-level Study Units.

2. Resource Package advances the reuse of study-independent metadata (agency internal reuse) or publication of metadata for broader public reuse (e.g., for question databases, harmonisation, or shared standardized demographic classifications)

3. The integrated use of a Group, Resource Package, and the Comparison Module fosters the management and documentation of complex international comparative survey series by

- The economic management of metadata and the provision of advanced content

- Reuse of schemes for Questions, Variables, Codes, etc., for comparative studies

- Standardized documentation of changes over time

- Extended description of comparative and comparable metadata

In addition to a more archive-centric view the application of the Group approach along the life cycle would foster an increased and seamless reuse of metadata for archives, related project partners, and the scientific community at large through:

1. Provision of ex-ante defined standards (questions and variables) to prepare the fieldwork in light of the intended metadata exchange between the research team and fieldwork agency

2. Transfer of grouped metadata following the initial fieldwork, first local (national) study, and data documentation from the data producer to the data provider for efficient standardization and harmonisation in the processing of integrated study formats

3. Publication and presentation of surveys as preprocessed views on metadata structures that support

- Browsing in specially grouped structure (country views, integrated views, etc.) at public portals

- Search for comparative or comparable items based on metadata from the Comparative Module

APPENDIX A

The paper is one of several papers which are the outcome of a workshop held at Schloss Dagstuhl - Leibniz Center for Informatics in Wadern, Germany, November 2-6, 2009.

Workshop title:
Workshop on Implementation of DDI3 - Advanced Topics

Organizers:
Arofan Gregory (Open Data Foundation, Tucson, Arizona, USA)
Wendy Thomas (Minnesota Population Center, University of Minnesota, USA)
Mary Vardigan (Inter-university Consortium for Political and Social Research [ICPSR], University of Michigan, USA)
Joachim Wackerow (GESIS, Leibniz Institute for the Social Sciences, Germany)
Link: http://www.dagstuhl.de/09452

This series was edited by Michelle Edwards, Larry Hoyle and Mary Vardigan.

The authors of the paper would like to acknowledge others who participated in this workshop.

Alerk Amin, CentERdata, Tilburg University, the Netherlands
Michelle Edwards, University of Guelph, Canada
Bryan Fitzpatrick, Rapanea Consulting, United Kingdom
Oliver Hopt, GESIS, Leibniz Institute for the Social Sciences, Bonn, Germany
Larry Hoyle, Institute for Policy and Social Research, University of Kansas, USA
Sanda Ionescu, Inter-university Consortium for Political and Social Research (ICPSR), University of Michigan, USA
Jannik Jensen, Dansk Data Archive (DDA), Denmark
Uwe Jensen, GESIS, Leibniz Institute for the Social Sciences, Köln, Germany
Mari Kleemola, Finnish Social Science Data Archive (FSD), University of Tampere, Finland
Dan Kristiansen, Dansk Data Archive (DDA), Denmark
Agostina Martinez, University of Cambridge, United Kingdom
Martin Mechtel, Institute for Educational Progress, Humboldt-Universität zu Berlin, Germany
Olof Olsson, Swedish National Data Service (SND), Sweden
Ørnulf Risnes, Norwegian Social Science Data Services (NSD), Norway
Wolfgang Zenk-Möltgen, GESIS, Leibniz Institute for the Social Sciences, Köln, Germany

APPENDIX B

APPENDIX C: ISSP SURVEY SERIES MARKUP EXAMPLE

```
<?xml version="1.0"?>
<DDIInstance xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="ddi:instance:3_1
http://www.ddialliance.org/schema/ddi3.1/instance.xsd" xmlns="ddi:instance:3_1" xmlns:r="ddi:reusable:3_1"
xmlns:xhtml="http://www.w3.org/1999/xhtml" xmlns:dce="ddi:dcelements:3_1"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:a="ddi:archive:3_1" xmlns:g="ddi:group:3_1"
xmlns:cm="ddi:comparative:3_1" xmlns:c="ddi:conceptualcomponent:3_1" xmlns:d="ddi:datacollection:3_1"
xmlns:l="ddi:logicalproduct:3_1" xmlns:pd="ddi:physicaldataproduct:3_1" xmlns:ds="ddi:dataset:3_1"
xmlns:pi="ddi:physicalinstance:3_1" xmlns:m1="ddi:physicaldataproduct/ncube/normal:3_1"
xmlns:m2="ddi:physicaldataproduct/ncube/tabular:3_1" xmlns:m3="ddi:physicaldataproduct/ncube/inline:3_1"
xmlns:s="ddi:studyunit:3_1" xmlns:pr="ddi:profile:3_1" isMaintainable="true" id="Example" version="1.0.0"
agency="de.gesis">
  <g:Group isMaintainable="true" id="ISSP_Integrated" agency="de.gesis" time="T0" instrument="I1"
    panel="P4" geography="G0" dataSet="D3" language="L3">
    <r:Citation>
      <r:Title>Integrated ISSP</r:Title>
      <!-- Citation for the Integrated ISSP -->
    </r:Citation>
    <g:Purpose id="GPurpose">
      <r:Content>Provide integrated access to ISSP over time and topic.</r:Content>
    </g:Purpose>
    <r:SeriesStatement><r:SeriesName>ISSP</r:SeriesName></r:SeriesStatement>
    <g:Concepts>
      <c:ConceptualComponent id="G_CC">
        <c:UniverseScheme id="Univ">
          <c:Universe id="GU">
            <c:HumanReadable>Population of Germany, Finland, and the United Kingdom</c:HumanReadable>
            <c:SubUniverse id="GU_DE">
              <c:HumanReadable>Germany</c:HumanReadable>
            </c:SubUniverse>
            <c:SubUniverse id="GU_FN">
              <c:HumanReadable>Finland</c:HumanReadable>
            </c:SubUniverse>
            <c:SubUniverse id="GU_UK">
              <c:HumanReadable>United Kingdom</c:HumanReadable>
            </c:SubUniverse>
          </c:Universe>
        </c:UniverseScheme>
      </c:ConceptualComponent>
```

```
41        </g:Concepts>
42        <g:DataCollection>
43          <d:DataCollection id="GDC">
44            <d:DataCollectionModuleName>Common Data Collection Information</d:DataCollectionModuleName>
45            <!-- Include any Methodology information that pertains to the whole series, the overall sample
46            description, and basic question set that is used by all surveys (be sure to include each language
47            variant) -->
48          </d:DataCollection>
49        </g:DataCollection>
50        <g:LogicalProduct>
51          <l:LogicalProduct id="GLP">
52            <!-- Include category schemes, code schemes, and variable schemes common to all studies. Be sure to
53            include all language variants. -->
54            <l:VariableScheme isMaintainable="true" id="VS_group">
55              <l:Variable isVersionable="true" id="V1" version="1.0.0">
56                <l:VariableName xml:lang="en">Age</l:VariableName>
57                <l:VariableName xml:lang="de">Alter</l:VariableName>
58                <l:VariableName xml:lang="fn">Ikä</l:VariableName>
59                <l:VariableName xml:lang="sw">Âlder</l:VariableName>
60              </l:Variable>
61              <l:Variable isVersionable="true" id="V2" version="1.0.0">
62                <l:VariableName xml:lang="en">Sex</l:VariableName>
63                <l:VariableName xml:lang="de">Sex</l:VariableName>
64                <l:VariableName xml:lang="fn">Sukupuoli</l:VariableName>
65                <l:VariableName xml:lang="sw">Kön</l:VariableName>
66              </l:Variable>
67            </l:VariableScheme>
68          </l:LogicalProduct>
69        </g:LogicalProduct>
70        <g:SubGroup id="M1" time="T4" instrument="I1" panel="P4" geography="G0" dataSet="D3" language="L3">
71          <g:Purpose id="M1_Purpose"><r:Content>To study Citizenship</r:Content></g:Purpose>
72        </g:SubGroup>
73        <g:SubGroup id="M2" time="T4" instrument="I1" panel="P4" geography="G0" dataSet="D3" language="L3">
74          <g:Purpose id="M2_Purpose"><r:Content>To study Environment</r:Content></g:Purpose>
75          <g:DataCollection>
76            <d:DataCollection id="M2_DC">
77            <!-- Include question scheme of all questions regarding Citizenship -->
78            </d:DataCollection>
79          </g:DataCollection>
80          <g:LogicalProduct>
81            <l:LogicalProduct id="M2_LP">
82              <!-- Include category schemes, code schemes, and variable schemes in this topical series -->
```

```
 83              </l:LogicalProduct>
 84           </g:LogicalProduct>
 85           <g:SubGroup id="M2_1993" time="T1" instrument="I1" panel="P4" geography="G4" dataSet="D0"
 86           language="L3">
 87             <g:Purpose id="M2_1993_Purpose"><r:Content>To study Environment in 1993</r:Content></g:Purpose>
 88             <g:StudyUnit>
 89               <s:StudyUnit id="M2_1993_DE" xml:lang="de">
 90                 <r:Citation>
 91                   <r:Title>1993 Germany ISSP Module 2</r:Title>
 92                 </r:Citation>
 93                 <s:Abstract id="A2de"><r:Content>ISSP Module</r:Content></s:Abstract>
 94                 <r:UniverseReference><r:ID>GU_DE</r:ID></r:UniverseReference>
 95                 <s:Purpose id="P2de"><r:Content>To study Environment in Germany in 1993</r:Content></s:Purpose>
 96               </s:StudyUnit>
 97             </g:StudyUnit>
 98             <g:StudyUnit>
 99               <s:StudyUnit id="M2_1993_UK" xml:lang="en">
100                 <r:Citation>
101                   <r:Title>1993 United Kingdom ISSP Module 2</r:Title>
102                 </r:Citation>
103                 <s:Abstract id="A2uk"><r:Content>ISSP Module</r:Content></s:Abstract>
104                 <r:UniverseReference><r:ID>GU_UK</r:ID></r:UniverseReference>
105                 <s:Purpose id="P2uk">
106                   <r:Content>To study Environment in the United Kingdom in 1993</r:Content>
107                 </s:Purpose>
108               </s:StudyUnit>
109             </g:StudyUnit>
110             <g:StudyUnit>
111               <s:StudyUnit id="M2_1993_FN">
112                 <r:Citation>
113                   <r:Title>1993 Finland ISSP Module 2</r:Title>
114                 </r:Citation>
115                 <s:Abstract id="A2fn"><r:Content>ISSP Module</r:Content></s:Abstract>
116                 <r:UniverseReference><r:ID>GU_FN</r:ID></r:UniverseReference>
117                 <s:Purpose id="P2fn"><r:Content>To study Environment in Finland in 1993</r:Content></s:Purpose>
118                 <d:DataCollection id="M2_1993_FN_DC1" xml:lang="fn">
119                 </d:DataCollection>
120                 <d:DataCollection id="M2_1993_FN_DC2"  xml:lang="sw">
121                 </d:DataCollection>
122               </s:StudyUnit>
123             </g:StudyUnit>
124           </g:SubGroup>
```

```
125       <g:SubGroup id="M2_2000" time="T1" instrument="I1" panel="P4" geography="G4" dataSet="D0"
126       language="L3">
127         <!-- Germany fielded the standard module. England fielded the module with a onetime change to V2,
128         removal of V89, and addition of V100. Finland did not field this module in 2000.-->
129         <g:Purpose id="M2_2000_Purpose"><r:Content>To study Environment in 2000</r:Content></g:Purpose>
130         <g:StudyUnit>
131           <s:StudyUnit id="M2_2000_DE" xml:lang="de">
132             <r:Citation>
133               <r:Title>2000 Germany ISSP Module 2</r:Title>
134             </r:Citation>
135             <s:Abstract id="A_2000_de"><r:Content>ISSP Module</r:Content></s:Abstract>
136             <r:UniverseReference><r:ID>GU_DE</r:ID></r:UniverseReference>
137             <s:Purpose id="M2_2000_DE_P">
138               <r:Content>To study Environment in 2000 in Germany</r:Content>
139             </s:Purpose>
140           </s:StudyUnit>
141         </g:StudyUnit>
142         <g:StudyUnit>
143           <s:StudyUnit id="M2_2000_UK" xml:lang="en">
144             <r:Citation>
145               <r:Title>2000 United Kingdom ISSP Module 2</r:Title>
146             </r:Citation>
147             <s:Abstract id="A_2000_uk"><r:Content>ISSP Module</r:Content></s:Abstract>
148             <r:UniverseReference><r:ID>GU_UK</r:ID></r:UniverseReference>
149             <s:Purpose id="P2uk">
150               <r:Content>To study Environment in the United Kingdom in 2000</r:Content>
151             </s:Purpose>
152             <l:LogicalProduct id="M2_2000_UK_LP">
153               <l:VariableScheme id="VS_M2_2000_en">
154                 <l:VariableSchemeReference><r:ID>VS_Group</r:ID></l:VariableSchemeReference>
155                 <l:Variable isVersionable="true" id="V2" version="1.0.0" action="Update">
156                 <l:VariableName xml:lang="en">Gender</l:VariableName><!--change made only in the English text-->
157                 <l:VariableName xml:lang="de">Sex</l:VariableName>
158                 <l:VariableName xml:lang="fn">Sukupuoli</l:VariableName>
159                 <l:VariableName xml:lang="sw">Kn</l:VariableName>
160                 </l:Variable>
161                 <l:Variable id="V89" action="Delete"> <!-- include complete content of "V89" -->
162                 </l:Variable>
163                 <l:Variable id="V100" action="Add">
164                   <l:VariableName>Special Variable</l:VariableName>
165                 </l:Variable>
166               </l:VariableScheme>
```

```
167            </l:LogicalProduct>
168          </s:StudyUnit>
169        </g:StudyUnit>
170      </g:SubGroup>
171    </g:SubGroup>
172    <g:SubGroup id="M3" time="T4" instrument="I1" panel="P4" geography="G0" dataSet="D3" language="L3">
173      <g:Purpose id="M3_Purpose">
174        <r:Content>To study Family and Changing Gender Roles</r:Content>
175      </g:Purpose>
176      <!-- similar substructure as M2 showing year and country divisions -->
177    </g:SubGroup>
178    <g:SubGroup id="M4" time="T4" instrument="I1" panel="P4" geography="G0" dataSet="D3" language="L3">
179      <g:Purpose id="M4_Purpose"><r:Content>To study Leisure Time and Sports</r:Content></g:Purpose>
180      <!-- similar substructure as M2 showing year and country divisions -->
181    </g:SubGroup>
182    <g:SubGroup id="M5" time="T4" instrument="I1" panel="P4" geography="G0" dataSet="D3" language="L3">
183      <g:Purpose id="M5_Purpose"><r:Content>To study National Identity</r:Content></g:Purpose>
184      <!-- similar substructure as M2 showing year and country divisions -->
185    </g:SubGroup>
186    <g:SubGroup id="M6" time="T4" instrument="I1" panel="P4" geography="G0" dataSet="D3" language="L3">
187      <g:Purpose id="M6_Purpose"><r:Content>To study Religion</r:Content></g:Purpose>
188      <!-- similar substructure as M2 showing year and country divisions -->
189    </g:SubGroup>
190    <g:SubGroup id="M7" time="T4" instrument="I1" panel="P4" geography="G0" dataSet="D3" language="L3">
191      <g:Purpose id="M7_Purpose"><r:Content>To study Role of Government</r:Content></g:Purpose>
192      <!-- similar substructure as M2 showing year and country divisions -->
193    </g:SubGroup>
194    <g:SubGroup id="M8" time="T4" instrument="I1" panel="P4" geography="G0" dataSet="D3" language="L3">
195      <g:Purpose id="M8_Purpose"><r:Content>To study Social Networks</r:Content></g:Purpose>
196      <!-- similar substructure as M2 showing year and country divisions -->
197    </g:SubGroup>
198    <g:SubGroup id="M9" time="T4" instrument="I1" panel="P4" geography="G0" dataSet="D3" language="L3">
199      <g:Purpose id="M9_Purpose"><r:Content>To study Social Inequality</r:Content></g:Purpose>
200      <!-- similar substructure as M2 showing year and country divisions -->
201    </g:SubGroup>
202    <g:SubGroup id="M10" time="T4" instrument="I1" panel="P4" geography="G0" dataSet="D3" language="L3">
203      <g:Purpose id="M10_Purpose"><r:Content>To study Work Orientations</r:Content></g:Purpose>
204      <!-- similar substructure as M2 showing year and country divisions -->
205    </g:SubGroup>
206  </g:Group>
207 </DDIInstance>
208
```