

ISSP DataWizard – Computer Assisted Merging and Archiving of Distributed International Comparative Data

Since 1985 the International Social Survey Programme (ISSP) has conducted annual social surveys in the participating countries covering relevant topics from the Social Sciences. The ISSP was founded in the early eighties by four Social Science Research Institutes for the purpose of adding an international comparative aspect to the existing national social surveys. The founding members were:

by Robert Strötgen & Rolf Uher*

background variables, and
4. make the data available to the social science community as soon as possible.

As of 2001 the number of participants has grown from the original four countries to 38 countries worldwide. The map shows the geographical distribution of the current ISSP members.

Country	Institute	Survey
USA	NORC National Opinion Research Center, University of Chicago	GSS General Social Survey
Germany	ZUMA Zentrum für Umfragen, Methoden und Analysen	ALLBUS Allgemeine Bevölkerungsumfrage \Sozialwissenschaften
Great Britain	The National Center for Social Science (former Social and Community Planning Research – SCPR), London	BSA British Social Attitudes
Australia	RSSS Research School of Social Sciences, Australian National University, Canberra	NSSS National Social Science Survey

The ISSP agreed to four general principles:

1. jointly develop topical modules dealing with important areas of social science
2. field the modules as a fifteen-minute supplement to the regular national surveys (or a special survey if necessary)
3. include an extensive common core of



The member states are:

Australia	Japan
Austria	Latvia
Bangladesh	Mexico
Brazil	Netherlands
Bulgaria	New Zealand
Canada	Norway
Chile	Philippines
Cyprus	Poland
Czech Republic	Portugal
Denmark	Russia
Finland	Slovakian Republic
Flanders	Slovenia
France	Spain
Germany	South Africa
Great Britain	Sweden
Hungary	Switzerland
Ireland	Taiwan
Israel	USA
Italy	Venezuela

The first ISSP Survey in 1985 was conducted in 6 countries, the original 4 plus Italy and Austria and had the topic ‘Role of Government’. The following topics have been fielded since then or are being planned for the near future:

Topics 1985 - 2004	
* Role of Government	> 1985, 1990, 1996
* Social Networks	> 1986, 2001
* Social Inequality	> 1987, 1992, 1999
* Family	> 1988, 1994, 2002
* Work Orientations	> 1989, 1997
* Religion	> 1991, 1998
* Environment	> 1993, 2002
* National Identity	> 1995, 2003
* Citizenship	> 2004

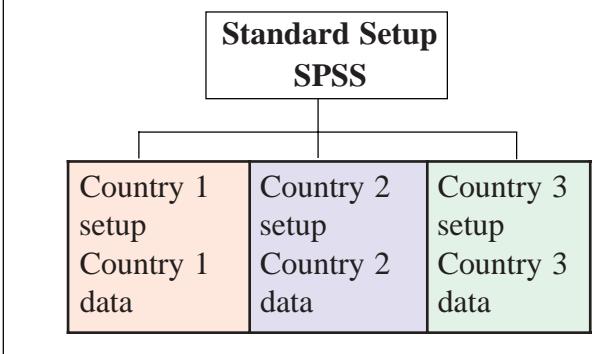
Replicating topics over time has enriched the international comparative aspect by adding a time-series component.

At the 1986 ISSP meeting in Mannheim the Zentralarchiv¹ was chosen as the ‘Archive of the ISSP’. The tasks of the Zentralarchiv are to archive, check and maintain data and documentation of the country-specific studies and distribute it to the scientific community. Following the structure of the questionnaire and the set of standard background variables, an international comparative data set is prepared accompanied by extensive and detailed comparative documentation. This documentation, also known as a codebook, includes all information that is necessary to analyze and interpret the data set. In addition to the methodological and technical description of the survey the codebook includes the complete questions and answer categories, the frequency-distributions broken down by countries, and also country-specific details like deviations from the agreed standard, problems with translations of indicators and the like.

The first step in creating the international file is the production of a ‘Standard Setup’. This includes the desired structure of the integrated file, the variable-names, variable-labels, codes, value-labels and the definition of the missing values. The starting point for the production of the ‘Standard Setup’ is the basic questionnaire of the respective ISSP module and the set of standard background variables defined for the ISSP.

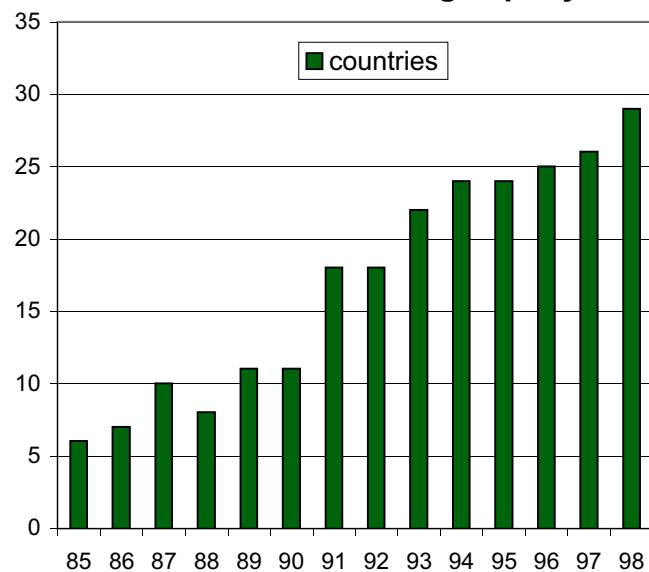
Even though this ‘Standard Setup’ is being distributed to the participating countries in advance, a number of cases have remarkable deviations from the desired standard, which must be considered in the process of merging the country-data to the integrated file prior to archiving the data.

Merging Process



Starting with 6 country data sets in 1985, by 1998 the ISSP had grown to include 28 different data sets, dramatically increasing the workload for preparing the merged file.

of data-sets to be merged per year



The amount of work that is necessary to harmonize one country-data-set to the standard can be measured by the amount of different documents that have to be viewed and used in order to understand all the details of the country-specific conditions.

Documents needed for processing:

- * ISSP Basic questionnaire
- * Standard setup
- * Standard background variable
- * Codebooks of earlier ISSP Modules
- * Frequencies of original data
- * Original questionnaire

- * Dictionaries
- * English documentation
- * Re-coding documentation
- * Frequencies of re-coded data
- * Other sources: internet, ISCO, statistical yearbooks, etc.

The predefined standard is being considered in different degrees of quality. The number of recode-statements to harmonize the country data sets to the final standard varies from 50 to 300, in one case over 1000 recode-statements were needed to merge the file. In all of these cases each recode-statement must be checked and proofed to determine whether the desired results have been attained.

The time and resource consuming nature of this process lead to ideas for developing a tool that would efficiently support the harmonizing process. The instrument needed to compare the pre-defined standard with the actual country specific setup with the following processing criteria. The first comparison should be done automatically resulting in a list of all deviations. The deviations in the setup and in the dataset itself are then corrected through individual recoding. The interface for this process must be user-friendly. The system should create a detailed report of all steps in the process.

ISSP Data Wizard:

- * Maps original with standard setup
- * Provides a comparative view
- * Assigns variables and values to standard
- * Recodes data to standard
- * Allows for individual data-processing
- * Reports processing steps

The concept of the ISSP DataWizard is being developed cooperatively by the GESIS (German Social Science Infrastructure Services) institutes of the Central Archive for Empirical Social Research (ZA) in Cologne and the German Social Science Information Center (IZ) in Bonn.

The first 'real' application after the explicit test-phase will be done with the 2000 ISSP module on 'Environment II'. In a further stage of development the ISSP DataWizard will be prepared for distributed use so that the ISSP participants can prepare their data-sets at their home institutions. Thus the expectation is high that the quality of the data delivered to the archive will be significantly higher according to the pre-defined standard, thereby facilitating the final steps of merging the files to a common international data-set.

Implementation of the ISSP DataWizard

The ISSP DataWizard is implemented at the German Social Science Information Center (IZ) as a Java/Swing

Application.² The use of a platform independent application is an advantage, particularly when distributing the tool to ISSP project partners in different countries. Although the data are stored in an Oracle database, the tool can be used just as readily on any relational JDBC-capable database server.

Great importance is attached to providing an ergonomic user interface in order to provide users with a sound level of support, focus their attention on the main aspects of their work and not burden them with the unimportant areas or information. The WOB model («tool metaphor based strictly object orientated graphic direct manipulative user interface») is a set of software ergonomic proposals which, in their entirety, were designed to create efficient and «natural» user interfaces (Krause 1995). The development of the ISSP DataWizard was orientated on this model.

Based on the analyses of current work processes carried out cooperatively by ZA and IZ, an attempt has been made to optimize support for the steps involved in processing the ISSP modules. In this context, the Wizard manages modules covering the associated default and national setups, country-specific study descriptions and survey records. (*See Fig 1 : Managing study descriptions with the ISSP DataWizard*)

An import and export interface provides the capability of reading and writing SPSS setup and data files. The open XML standard of the Data Documentation Initiative³ is also supported. This permits the use of a flexible and maker-independent data format and also enables a straightforward exchange of data with other tools supporting this standard.

The setup that defines variables and values with codes, labels and comments - thus matching the questionnaire - can be viewed, edited or even re-created using the DataWizard. It is also possible to alter the sequence of variables and values. This can be done both for the default questionnaire of a module as well as for country-specific questionnaires that implement this standard. In addition to making entries in the editor, users can also adopt variables and values from other modules, e.g. in the case of permanent demographic variables, where questions are asked that pertain to a module with a subject area covered in a preceding module, or in conjunction with values on scales that are used for several variables. It is also conceivable for partners to copy the entire default setup and make the requisite country-specific changes to the copied version so as to reduce the number of transfer errors, such as scale reversals, etc. (*See Fig 2 : Managing setups with the ISSP DataWizard*)

The central function of the ISSP DataWizard is to compare country-specific setups with the default setup. To this end, the work carried out 'intellectually' at the ZA has been

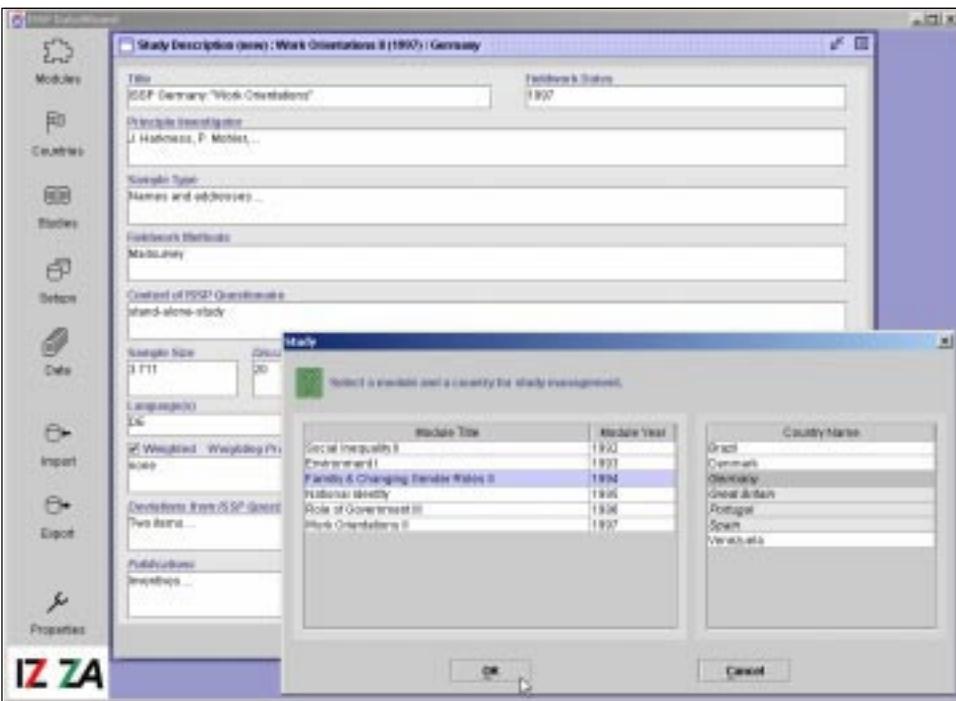


Fig 1.: Managing study descriptions with the ISSP DataWizard

analyzed and – in the simple cases – translated into rules wherever possible. An automatic mapping process, for instance, identifies reversed scales, incorrect variable codes

values into the default setup that have been «overlooked» for a particular country without having to switch to the part for processing the default setup. (See Fig 3 : Post-editing setup assignments using the ISSP DataWizard)

or similar problems and marks them as errors. Variables and values capable of being assigned beyond doubt are marked «OK», and dubious assignments highlighted for intellectual investigation. In the process of intellectual post-processing, which cannot be avoided, these problem cases are highlighted (e.g. using colour markings and an error browser) whereas unequivocal assignments require no further attention. The editor for post-processing compares a country-specific setup with its allocated default setup and synchronizes the display in order to quickly provide a clear view of the information relevant to a variable. Assignments can be made from context-related selection lists, thus reducing the likelihood of making incorrect entries and easing the demands on the user's attentiveness. This way, it is also possible to adopt

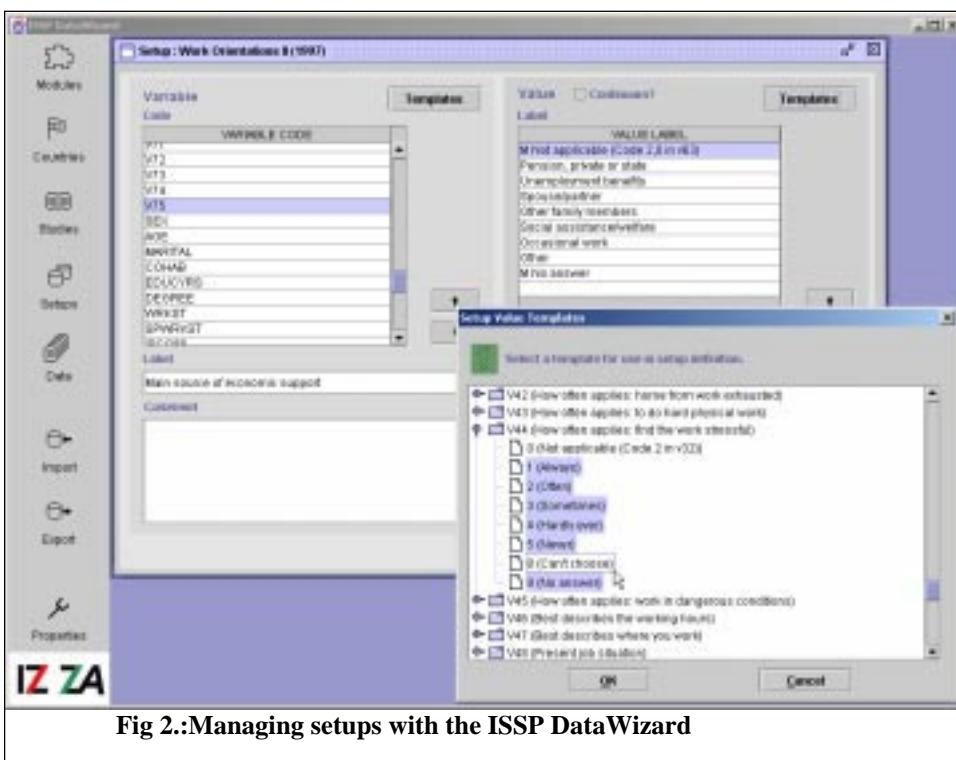


Fig 2.:Managing setups with the ISSP DataWizard

The database archives both the reference setups of ISSP partners as well as the edited versions. This means it is possible at any time to access the data submitted and check the adaptations made. In addition, it is possible to document all checks and assignments in a text file that additionally contains a complete concordance between a country-specific setup and the relevant default setup. This creates transparency in terms of merging and mapping; any suspected error can be reliably investigated.

Adaptations to the setups (to the structure of the data collected) must also be applied to the data itself. Divergent value codes, reversed scales, etc. must – as adapted in the setup – also be corrected in the data. The

necessary re-coding of data takes place automatically on the basis of the mapped setup. Here too, the database archives the data originally submitted as well as edited version.

It is also possible to employ customized rules in order, for example, to examine the consistency of filter queries or to re-code country-specific indices to a default index, provided this is possible without intellectual scrutiny. Users are provided with complex tools for creating and applying these rules. (See Fig 4 : Simple data visualization with the ISSP DataWizard)

Finally, users have the capability of viewing data in straightforward counts that allow them to carry out simple checks, e.g. what age distributions to expect, etc. Here, it is also

possible to compare the reference data version with the edited version and thus identify the result of the mapping and re-coding process. However, these capabilities only serve to monitor the success of merging; the actual data analysis is still to be carried out using the normal statistics program packages.

Whereas the current version of the ISSP DataWizard was primarily developed to support the ZA in merging international data, future versions will have to be developed towards the work distribution between ZA and ISSP project partners. It will be necessary to resolve questions of data exchange, access rights etc.

Conclusion

The ISSP DataWizard represents a tool that reduces the effort involved in merging ISSP records as well as the possibility of errors by automating simple activities and offering support to users in the necessary process of intellectual post-editing. Once the tool has proven its worth in practice, it must be further improved and optimized towards the requirements of the

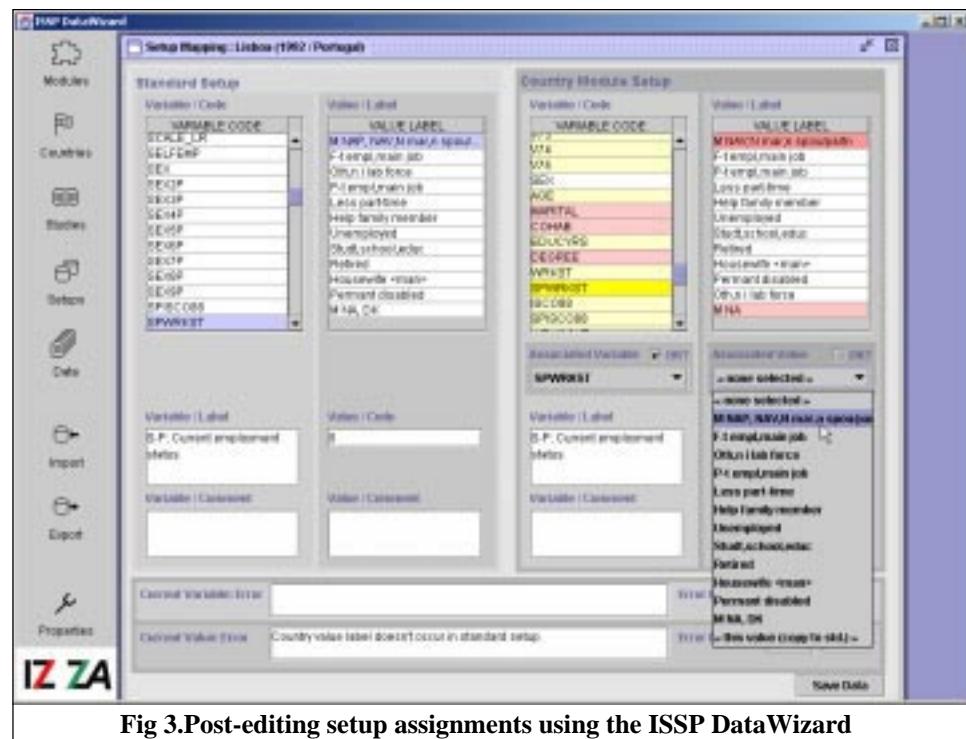


Fig 3.Post-editing setup assignments using the ISSP DataWizard

applications for which it is used. A general extension of the methods used beyond the ISSP context for similar application areas cannot be ruled out.

A conceivable area for future development is the addition

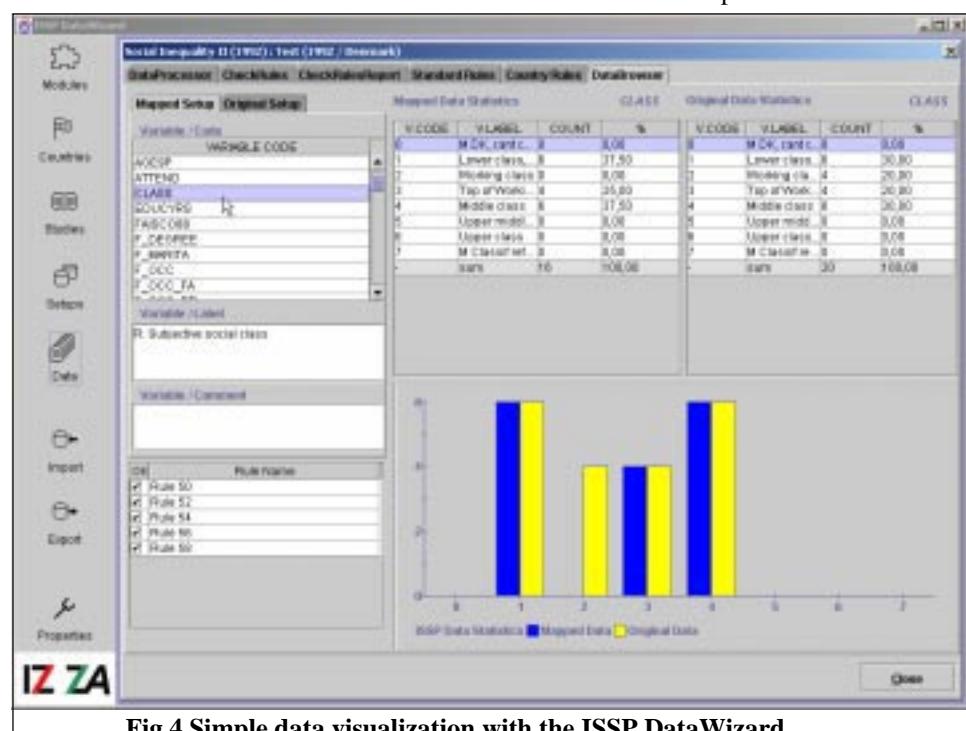


Fig 4.Simple data visualization with the ISSP DataWizard

of further elements of artificial intelligence to the ISSP DataWizard that do not focus solely on the data structure of default setup. For instance, plausibility rules could be introduced that contain the anticipated spread of characteristics and provide warnings where country-specific data stray beyond the given confidence intervals. This way, the user would be drawn to critical points and could consider taking a closer look at the data and documents. Developments of this type will need to be the subject of further concept formulation.

Literature:

J.W. Becker, James A. Davis, Peter Ester, and Peter P. Mohler, eds. (1990), Attitudes to Inequality and the Role of Government. Rijswijk, The Netherlands: Sociaal en Cultureel Planbureau

Petra Beckmann, Peter Ph. Mohler, Rolf Uher (1991), ISSP. International Social Survey Programme – Basic Information on the ISSP Data Collection - 1985-1994, ZUMA-Arbeitsbericht, Nr. 15

Alan Frizzell and Jon H. Pammett, eds. (1996), Social Inequality in Canada. Ottawa: Carleton University Press

Alan Frizzell and Jon H. Pammett, eds. (1997), Shades of Green. Ottawa: Carleton University Press

Roger Jowell, Sharon Witherspoon, and Lindsay Brook, eds. (1989), British Social Attitudes: Special International Report. Aldershot: Gower

Roger Jowell, Lindsay Brook, and Lizanne Dowds, eds. (1993), International Social Attitudes: The 10th BSA Report. Aldershot: Dartmouth Publishing

Jürgen Krause (1995), Das WOB-Modell, (IZ-Arbeitsbericht 1) Bonn: Informationszentrum Sozialwissenschaften

N. To_, P.Ph. Mohler and Brina Malnar, eds. (1999), Modern Society and Values, A Comparative Analysis Based on ISSP Project. FSS, University of Ljubljana und ZUMA Mannheim

Rolf Uher, Irene Müller (1988), The International Social Survey Programme – ISSP, in: IASSIST Quarterly, Vol. 12, No. 4, S. 3 ff

Rolf Uher (2000), The International Social Survey Programme (ISSP), in: Gert G. Wagner et. al. (eds.) Schmollers Jahrbuch, Zeitschrift für Wirtschafts- und Sozialwissenschaften, 120. Jahrgang, Heft 4, Berlin 2000, S. 663-672

Footnotes

¹ Since 1997 the Zentralarchiv is supported in its work by

the Spanish ISSP partner¹ In addition to the authors, Siegfried Schomisch, Udo Riege and Max Stempfhuber are involved in developing this tool.

2 <http://www.icpsr.umich.edu/DDI/>

* Paper presented at the IASSIST/IFDO Conference, Amsterdam 2001. Robert Strötgen, German Social Science Information Centre (IZ) Bonn, and Rolf Uher, Central Archive for Empirical Social Science Research (ZA) at the University of Cologne.