

The triple-s standard

Geoff Wright - Computable Functions

Presented at the Association for Survey Computing conference
"Open Standards: Breaking down the barriers"
at Imperial College on 19th September 2002.

Abstract

The triple-s standard has been successful in making it easier for users to transfer survey data between different data collection and analysis systems. But to continue with this objective the standard must develop and expand. Using the recently published triple-s XML version 1.2 we look at what is now available within the standard.

Introduction

The triple-s standard is an open standard for moving surveys between survey packages on various hardware and software platforms. The author has been involved with triple-s since its inception ten years ago and is a founder of Computable Functions, an IT consultancy based in Guildford that specialises in applying new technologies to survey research and analysis.

A standard is like a product

Producing a standard is much like the process of producing a software product. Initially you need someone who has an idea about a need that is currently unfilled – thank you Peter¹. You then clarify that initial idea and design a solution. A particular problem for an interchange standard to work is that you need both exporters and importers, so to keep the barriers low we tried to make the exporting process easy.

With a working prototype we can start to "sell" the standard – publishing, speaking at conferences, and getting some implementations. But adoption starts slowly – no-one uses version 1.0 of anything! However this does allow the early implementers to influence changes in the standard so that version 1.1 can be more useful. This process of incremental development has allowed the standard to react to increasing demands and newer technologies. In particular the early recognition that XML could replace all that home-grown syntax.

A successful standard has two types of "Users", the implementers and the users of products that include triple-s support. These need support and assistance, which is a role that the members of the triple-s Group have found to be taking an increasing importance. We now have a web-site for reporting news, downloading information, registering implementations and acting as a contact point. Also members of the group have been able to assist implementers and provide interpretation of the standard.

¹ **Data Use and Reuse**, P Wills, SGCSA, Bristol England, September 1992

The design of triple-s

Like many organisations in the 1990's we had a mission statement: -

The aim of the triple-s standard is to define a means of transferring the key elements of entire surveys between different survey software packages across various hardware and software platforms.

The main things to note are that this was to be an *interchange* standard for moving surveys, and that it would concentrate on the *key elements* of the surveys. To adapt a well-known acronym, we used a KISSS (Keep It Simple Standard Stupid) approach to the design of triple-s.

We decided on character format files, partly to make debugging and investigation easier, but also because ASCII character files were the most portable. Of course with XML we now have Unicode but the readability aspect is retained.

The main design decision was to adopt a two file approach. The metadata file essentially describes the questions, and the case data file the answers. The format of the description or metadata file should be irrelevant to users of triple-s, but be easy to use for implementers. We chose to adopt a block structured keyword and value syntax, which subsequently transferred well to the syntax of XML. For the case data we chose the simplest representation - fixed format ASCII with one record per respondent.

Within the description file the information is held in 4 layers: -

- The triple-s file itself (e.g. creation date, origin)
- The survey in general (e.g. descriptive title)
- The case data record (e.g. identifier)
- The variables (or questions) within the case data record (e.g. name, type)

The bulk of the information is about the variables themselves, such as the name, type, location and question text. To determine the variable types we looked at a number of widely used survey packages and extracted the most common. This set includes integer numbers and character fields, so any special question type can be exported although with the loss of specific type information.

For categorical variables we settled on positive numeric codes. As this was an interchange standard, rather than a means to describe an existing dataset, any other coding convention could always be mapped onto these code values.

triple-s XML version 1.2

Over time the triple-s Group have revised and extended the standard, adopting a policy of incremental development based primarily on the experience and expectations of users. The most significant was the move to XML as the language for the description file.

In July 2002 we published the *triple-s XML version 1.2* standard². This is fully compatible with the XML version 1.1 standard; in fact a valid XM 1.1 description is also valid XML 1.2. The enhancements provide useful additional information in the following areas: -

- Descriptive information such as the survey name and version. The latter is particularly relevant to tracking surveys and multi-site surveys.
- Variable use indicator which specifies the intent of the variable. The initial values of this attribute are "serial" (the case serial number) and "weight" (the case weight).
- Indicate the Base for a variable. The optional <filter> element can refer to a previous logical variable whose value determines if the current variable is available for that case.
- Multilingual texts can be specified for the survey title, variable label, and value text. These use the standard *xml:lang* attribute to specify alternate versions of the texts.
- Location of the case data file. An <href> attribute can be used to specify the explicit location of the case data file.

triple-s version 2

Whilst the release of triple-s XML version 1.2 has provided some of the features that users of triple-s have requested, we feel that we need to be looking further ahead. A key basis for any triple-s XML version 2 would have to be compatibility. We need to consider the existing triple-s standard, AskML³ and similar initiatives, and also the more general XML-based data interchange standards from both within and outside the survey research community.

Reverting back to our original mission statement the most important considerations will be what key elements of a survey are missing now, and which will be expected in the future: -

- More flexible format of case data records. At present triple-s uses fixed format ASCII records, but future options might include comma/tab-separated values, an XML representation of data values, and Unicode data values.
- Structured datasets. At present the standard only supports a simple one record per case for the data. This can cope with many existing surveys, but permitting more than one record type per survey would give much more flexibility. In particular when handling data that has been stored in relational databases.
- More variable types (e.g. date and time).
- More flexible codes and values. The existing scheme with numeric codes could be extended to permit alphanumeric codes for single/multiple codes.

² **triple-s XML**, triple-s Group, July 2002

³ **The AskML Project – An Effort To Develop A Standard XML Survey Representation**, OpenSurvey, <http://www.opensurvey.org>

- More options for labels, texts etc. (e.g. including images/sounds within the description).

Support for triple-s

The standard can not exist by itself, and we have found that increasing more of our efforts are going into support for the standard. This comes primarily in two ways: -

1. The triple-s website (<http://www.triple-s.org>) which provides:-
 - News about triple-s
 - Documentation on the latest version of the standard
 - An on-line copy of the DTD
 - Conversion software
 - Register of implementations
2. The triple-s Group
 - A contact point for information
 - Help and advice to implementers
 - Interpretation of the standard
 - Compliance to the spirit of the standard
 - Examples and test scripts

Conclusions

The triple-s standard was devised as an *interchange* standard based on Peter Wills initial ideas. It has developed over the past 10 years, and in particular was implemented in XML over 3 years ago. Our policy of continual development has resulted in XML version 1.2 being published earlier this year. Over 25 suppliers, including many of the leaders in survey research software now support the standard.

Geoff Wright, Computable Functions, Guildford, UK
geoff@computable-functions.com