

Open Standards for the Research Industry

Nancy Tienhaara DASHcati Software Ltd

Moving questionnaire components across software package boundaries is much like moving material goods between countries; at the borders there are delays and hassles as well as blockages that arise from incompatibilities in export/import rules in the different jurisdictions.

Open standards promise a better alternative. Consider just two of the changes that would ensue if research software providers adhered to a common set of rules independent of any specific vendor. For research suppliers, all parts of a survey would pass more flexibly and quickly from one research task to another without tedious re-programming. Research buyers could choose suppliers worldwide for a single project on the basis of their individual merit rather than confining their choices to a common software package.

How are surveys transferred now?

Most data collection and tabulation software packages provide custom interfaces to the software their clients most commonly use, so that moving information to and from word processors, spreadsheets, PC statistical packages, databases, etc. are routine.

However, these current options are inadequate. Information that is available in one package is often lost in the transmission to another package; for example, data may be passed easily but question text and option categories may not be included. Also, the number of import/export software targets is usually limited. At the same time, the Canadian research industry is broadening its scope into new areas and also expanding its umbrella to include more and different practitioners; this suggests that the number of interfaces required will expand as well to include their software.

How would standards work?

Think of standards as a common interface. It would be the responsibility of software vendors to provide the translation of their package to and from this common interface. When each program can translate both into and from the shared format, then the survey components can move from one program to any other automatically via this standard format.

Why should software providers make it easier for their clients to move surveys to competing software packages?

There are three answers.

First of all, it is time-consuming and expensive for software companies to develop, support and update interfaces to other software which, like their own, constantly changes.

Secondly, with survey research expanding into new areas, the overall market for research software is growing, and there is every indication that higher market share will go to the software which adheres to a standard.

Third, and most important, once current clients of each software package recognize the value of standards, they will insist on them.

What are "open" standards?

The major defining characteristic of open standards is that the specifications on which they are based are owned by a vendor-neutral organization rather than by software developers. For example, the World Wide Web Consortium (or the W3C), the leading standards organization for web-related standards, has 500 corporate and academic members, including legendary rivals Sun Microsystems

and Microsoft. Its goal is to provide "a vendor-neutral forum for its members to address web-related issues ... to produce free, interoperable specifications and sample code.

A paper for the Standards Engineering Society outlined 10 concepts that constitute the principles of open standards. They include these three provisions: that all stakeholders may participate in the development process and have full access to meetings and documents; that agreement is reached through broad consensus and due process rather than domination; and that the same standards hold for the same function world-wide.

The universally-accepted definition of open standards means that no individual software company can legitimately claim to provide "open standards"; rather they can only support them. This is an important point because the word "open" is frequently used in software marketing for a variety of agendas, both honest attempts to provide commonality and deceptive efforts to control and monopolize. Open standards cannot, by definition, be owned by any one vendor.

How are standards implemented?

Whatever the industry, standards are being implemented using the same tool, XML. The survey research standards initiatives are all following this same route, and the common interfaces described above are being prepared in XML.

XML, which stands for eXtensible Markup Language, is a way to structure documents. Consider a questionnaire. XML uses tags to identify questions, option lists, interviewer instructions, and all the other features specific to surveys. There is no single set of XML tags that every industry uses, but instead each industry can create what it needs for its own application. This is the basis for the word "extensible" in XML.

Extensibility can deliver power, but with it comes extra complexity and planning. Teamwork becomes important. For an industry-wide project, you need to work with your counterparts to share common name tags. This is where industry organizations enter the picture.

Who is discussing open standards for surveys?

Discussions about open standards for survey research are occurring in several locations around the globe. For market research, the major impetus is coming from England, where several initiatives are under way. The first is the most developed, the second is at the feedback stage, and the third is the most ambitious.

Triple-S XML (www.triple-s.org)

The Triple-S standard defines "a means of transferring the key elements of entire surveys between different survey software packages across various hardware and software platforms". Although initially presented as a method to exchange variables as well as data between different tab packages, Triple-S is now being used to transfer question text and response lists from CAI packages as well. The meaning of three S's? Apparently the name originated from the fact that all of the relevant topics started with s: standardization, surveys, software, systems, simplicity.

The Triple-S group was initially comprised of key developers from 3 commercial survey software houses; Keith Hughes (Merlinco), Stephen Jenkins (Mercator) and Geoff Wright (Pulse Train). More recently, Laurance Garrard (Maritz-TRBI) and Ed Ross (Open Survey) have joined the group. The first standard was published in 1994 with later updates including a transition to XML in Dec. 1999. In March 2001, Triple-S was recognized as a survey interchange standard by OASIS (Organization for the Advancement of Structure Information Standards), which promotes open XML standards for all types of software applications.

TabXML (www.opensurvey.org)

The impetus behind this project is the recognition that there are many different pieces of software used to work on tables, and exchanging information between them is often difficult. TabXML is an

XML representation of survey research cross tabulations so that the information in tables can be absorbed easily into a variety of mainstream and specialized software packages.

TabsML is a cooperative venture that includes E-Tabs, DataTree and Open Survey. The TabsML v1.1 specification is posted on the OpenSurvey web site with a request for comments.

AskML (www.opensurvey.org)

AskML is an ambitious attempt to make questionnaires independent of the CAI systems on which they run. It goes beyond the data descriptions of Triple-S and TabsML to the actual execution of a CAI (Computer Assisted Interviewing) survey. It would mean that a single survey instrument could be designed once, perhaps in an authoring language separate from all commercial packages, and that instrument could be implemented consistently in different commercial packages, software versions, languages, and modes (e.g. web, handheld, CATI).

AskML is a major project of Open Survey, a non-profit organization founded in 2000 by Ed Ross (founder of Quantime) and Andrew Jeavons (former president of Surveycraft Systems Inc.) with Harris Interactive involved as a founding sponsor. OpenSurvey is a non-profit organization dedicated to the development, discussion and distribution of open source software and open standards for survey research.

This is not an exhaustive list of the projects that are underway. The Object Management Group (www.omg.org) is developing a Common Warehouse Metamodel for defining the storage and handling of data in data warehouses for adoption across the IT industry; the involvement of Triple-S and Open Survey will hopefully mean that the CWM will be able to accommodate questionnaire data, unlike most of the existing data warehouse models. The University of Michigan's Data Documentation Initiative DDI (www.icpsr.umich.edu/DDI) is developing tools to make past surveys more readily available for re-analysis. TADEQ, a Tool for the Analysis and Documentation of Electronic Questionnaires is a project to make a human-readable presentation of the CAI questionnaire; it is mainly an initiative of the statistical agencies of various countries. And there are still others!

What next?

The development of standards is a huge effort, undertaken by a group of people who are volunteers. In some cases their companies help underwrite the time they spend on standards development, in others it is work done on their own time. Those involved in the standards process share a passion for their industry to work and work well. But it takes time, and more assistance is always needed.

For us here in Canada, awareness of the standards process is the first step, active involvement is next. Although the major initiatives are an ocean and several time zones away, the web provides many ways to get involved in the discussions. Visit the web sites listed above and learn more about each of the initiatives. Comment on the current initiatives posted for public review. Volunteer assistance in devising standards, programming the standard interfaces, funding projects, or educating the research community. Plan your next trip to England to coincide with the conference "Open Standards: Breaking Down the Barriers" being presented by ASC, the Association for Survey Computing (www.asc.org.uk) at Imperial College in London on Sept. 19, 2002. And finally, most important of all, talk to your software providers about where they stand on open standards.

Nancy Tienhaara is a partner in [DASHcati Software Ltd.](http://www.dashcati.com) in Victoria BC.