Alliance for the Data Documentation Initiative (DDI)

Strategic Plan 2007 and Beyond

Introduction

Transitioning to this new plan from the original DDI Alliance Strategic Plan written in June 2003 provides us with an opportunity to reflect on previous goals and to look ahead to what we hope to achieve over the next several years. For many reasons, this is a good time to assess our progress as a self-supporting membership Alliance and to set out an agenda for the future that takes the DDI project to a new level.

The four intervening years since the establishment of the Alliance have seen dramatic advances for the DDI project and the Alliance. The creation of DDI 3.0, long envisioned to deliver requested features and functionality, is a signal accomplishment. DDI 3.0 brings the DDI effort fully into the 21st century in terms of cutting-edge use of XML Schemas to provide machine-actionability to drive systems. It incorporates all of the key substantive areas recommended in 2003, with much more robust spatial coverage. Other important innovations in DDI 3.0 include:

- An underlying data model that permits the expression of the model in alternative technologies
- Coverage of more of the data life cycle, with an emphasis on data collection
- Modular design
- Enhanced support for multiple languages
- Support for variable comparison and harmonization
- Structured mechanisms for identification and versioning that enable the creation of registries like question banks
- Core HTML for formatting of unstructured text
- Elimination of redundancies through a new grouping model and an extensive set of reusable elements
- Capability to create "DDI profiles" for specific uses
- Mechanism to carry data inline
- Alignment with other metadata standards, including MARC (library catalog bibliographic format), Dublin Core (cross-domain information resource description), SDMX (time-series data), ISO 11179 (metadata registry), and FGDC and ISO 19115 (geographic standards)
- Extensibility

At the point of this writing, DDI 3.0 has been approved as a Candidate Release and is scheduled to be published by early 2008, with a user manual to accompany it. This will be followed closely by publication of DDI 3.1, which is likely to incorporate more of the research data life cycle – i.e., survey design and implementation elements – and other enhancements not fully realized at the time of publication.

Strategic Questions

With the delivery of DDI 3.0, the Alliance is now positioned to make good progress on strategic goals over the next several years. To set the stage for this next phase of DDI development, it is helpful to pose some big-picture questions about the future of DDI and our vision for success. Three such overarching questions are considered here, followed by a discussion of tactics that will be key to realizing our vision.

What should the DDI standard be in five years? In an ideal future, we envision an evolving standard used universally in the social science data community and serving as a model for other XML-based metadata standards for data. This DDI of the future should be a registered ISO standard with a committed advisory group that maintains and enhances it on a periodic basis. Systems used across the social science research enterprise should be built upon and import/export DDI so that the DDI is part of the mainstream of social science.

Who should be members of the DDI Alliance in five years? Again, thinking in broad terms, a significant number of academic and non-academic data producers and data archives should be members, as well as many universities and software publishers whose products lead to the collection, management, archiving, and analysis of data. These DDI members should support the evolution of DDI in various ways – through membership fees, by supporting specific projects financially, and through contributions of staff resources. The research community should understand the value of having structured and standardized metadata across the life cycle.

What should the DDI Alliance accomplish in the next five years? We need to develop the standard so that it keeps pace with changing needs; promote the development of tools to make DDI easier to use; train users and developers in working with DDI; and raise sufficient funds to continue our activities. We need to promote the DDI concept and develop materials and language that articulates the advantages of the DDI approach. These objectives are not new, but we need to envision innovative strategies to accomplish them and to move the DDI effort forward to new levels of achievement.

Tactics to Realize DDI Goals

To realize these goals for the future, it is important to have agreement on a clear set of priorities and projects. Below, we lay out some important areas for DDI and some tactics that will help us make progress toward our ideal future.

Refinement of the Specification

While we have a solid foundation on which to build, we need to continue to develop and refine DDI 3.0 so that it meets the real needs of users. We have just undertaken a radical change in moving to DDI 3.0, and it is now time to evaluate and fine-tune what we have created so that it works optimally for its intended purposes. We need to pay particular attention to the areas of complex files and longitudinal data, instrument documentation,

aggregate data, and ex-post comparison as we move forward. This will involve continuing to apply DDI 3.0 to real world examples and creating a suite of use cases that provide proof of concept and convincing evidence that DDI can handle the variety of social science data scenarios.

Tools

Tool development is critical for widespread use and adoption of DDI 3.0, and thus we designate this activity as a top priority moving forward. The complexity of the DDI 3.0 data model and the XML Schemas environment means that the standard and tools to use it must be developed hand in hand. We have made good progress in this area with the creation of comprehensive DDI Help documentation, a DDI 2 to 3 converter, a statistical package converter, stylesheets, forums, and templates. In addition, at the time of this writing, a coalition of stakeholders including the Alliance is coming together to create an open source DDI startup toolkit with financial and inkind support from several sources. This toolkit will provide potential DDI users with a set of tools that will meet their needs in terms of getting information into DDI.

The goal for the future is build upon the existing tools to create a suite of utilities that permits users to take full advantage of the capabilities of DDI 3.0 in an efficient way. For those interested in using DDI 3.0 we need to provide a clear path to adopting the standard with tools that remove the barrier of complexity so that markup, management, and display tasks are relatively easy and straightforward.

The structure of DDI 3.0 was designed to enable the development of metadata registries, and we need to capitalize on this new functionality. By sharing and reusing metadata in this way, the social science research community can ultimately realize some savings and also forge new relationships.

Further, we need to establish an ongoing mechanism for tools creation and updating and an active community of developers. Open source appears to be the best technological environment of choice to make this happen. To that end, we will establish a DDI Tools Working Group that will recruit participants who have expertise this domain and can participate in a tools community. The Alliance will also continue to put up as much financial support as possible to facilitate the creation of the needed suite of tools.

Outreach

Outreach for DDI needs to go beyond simply promoting the DDI approach. Looking ahead, we need to obtain feedback from the broader user community on how the current specification meets their needs, and we must solicit new ideas on features and functionality not yet implemented. This two-way communication with users and with experts across the data life cycle is essential to making the standard the best it can be.

In terms of a communication strategy, we need to develop a library of presentations, targeted to different audiences, that our members and others can use for specific venues

and forums. We might also publish a newsletter or a blog to highlight uses of DDI and how it is being used in different settings. FAQs or an Ask the Experts column would be useful as would DDI "profiles" listing the element sets used in different organizations and for specific purposes.

These communication pieces must be professional in quality and part of an overall communications strategy. In taking the DDI to the next level, it will be necessary to set up strategic meetings with key individuals with the goal of urging adoption of the DDI or recruiting new members. These targeted meetings must be part of the communications plan.

We suggest the following constituencies as potential outreach targets:

Vendors. With publication of DDI 3.0, we are now in a position to begin serious outreach efforts to new communities. While we have made inroads to some data producers and are working with Blaise and CASES computer-assisted interview software to produce DDI XML, we need to expand these efforts to reach other CAI firms. Having tools and examples that showcase the potential of DDI markup is important to this endeavor.

Other types of software vendors, including the statistical package vendors and companies like StatTransfer, are another important outreach target. Having statistical packages support DDI markup as an output format is highly desirable.

Potential members and users. Outreach to potential DDI Alliance members is another objective in this area. A review of the DDI Papers, Presentations, and Reports on the DDI Web site shows over 60 publications, mainly presentations, about the DDI since 2000. IASSIST conferences have had a significant focus on DDI in recent years, and the support from IASSIST for the standard has been excellent. We have begun to present DDI to new audiences (e.g., FedCASIC and the Comparative Survey Design and Implementation group during 2007), and now need to expand those efforts by encouraging Alliance members to be effective advocates for the effort. Spreading the word in this way has the potential to bring in new members of the Alliance and to speed the adoption of DDI.

Researchers. One factor that has slowed success of the DDI endeavor is the lack of support from social science researchers themselves. Many do not understand the advantages of DDI for their work and question the investment in metadata of this type. We need to counter this attitude by building a cogent business case that persuades the research community of the usefulness of this approach to documentation. Until we have a good argument to present to this community, we will not be able to fulfill the promise of DDI. We also need to continue to present information about the DDI to this group to reinforce the value of the standard.

Other standards developers. The social science research environment has a multiplicity of relevant metadata standards at its disposal. How they all relate is not clear and there is confusion about best practice in a world of so many standards.

DDI can make a contribution by presenting clear mappings between DDI and other standards and by explaining how they fit together. We should also be talking to developers of the other standards to determine how we might work together to help the community. Otherwise, we continue to work in isolation and run the risk of not understanding the environment in which we are working.

Membership

This is an area in which the Alliance can improve over the next few years. At this writing, there are 29 official members of the Alliance with another promised for the next fiscal year, an increase of five from the inception of the Alliance in 2003 with the original 25 members. The membership has largely been stable with little churn: only a handful of members (five) have left the Alliance since it began, in almost all cases because the representative moved elsewhere. However, we need to do better at recruiting new members.

Having said that, it is also the case that there are dedicated DDI users all over the world who support the effort enthusiastically but are not official members. Some are members of DDI working groups. There is an especially large group of DDI users who use the International Household Survey Network Microdata Toolkit to produce documentation in DDI format. We need to think about whether we could involve these individuals in the Alliance, perhaps at a reduced or subsidized membership level.

We might also look again at government statistical agencies as potential members, along with the other groups that we are targeting for outreach. This will require customizing our message to articulate the benefits of joining the Alliance in ways that are meaningful to the various groups.

Training

As was the case four years ago, we need to broaden our training efforts with additional workshops and develop a wider array of useful training materials. Formal training courses need to be created and widely offered, perhaps through the vehicles of the ICPSR and Essex Summer Programs, as well as elsewhere in Europe and the Americas. Many have pointed out that it is difficult to get started in using the DDI and that there is a steep learning curve.

The Alliance Web site can play an important role in providing access to tools and other materials, such as step by-step instructions in getting started with the DDI.
Related to this, we need to "train the trainers" so we have a solid group of trained presenters to represent the Alliance in various venues. Ideally, training should be offered on a regular schedule at a set location and also in a flexible, customized manner so that DDI trainers can travel to specific venues when this is desired.

Funding – External and Internal

The previous Annual Report suggested that the Alliance would not be able to fulfill its goals without significant external support. While this has not happened, it is still the case that external funding is a worthwhile aim. We need to be thinking about how to partner to find external funding sources and to have a mechanism whereby the Alliance can accept contributions from member organizations that might be targeted at specific tasks.

As of this writing, a new funding model for projects has been created to permit multiple partners to contribute to a project and to enable matching of DDI Alliance funds. This seems an excellent model to use going forward, one that we should attempt to build on in coming years to leverage resources.

Related to this, now that we have a standard in place, we need to be more aggressive about using the Alliance monies to further tools development and the adoption of the standard. We need to earmark funds for specific purposes and outsource activities that we have previously relied on committee members to do, such as documentation writing, use case/example development, evaluation, etc. Using our internal funds more proactively to accomplish specific tasks and goals, we can move away from our previous model of in-kind contributions, which has been difficult to sustain. This new approach will enable the Alliance to function more professionally and efficiently going forward.

ISO Standard Status

Now that the DDI has a data model underlying the specification, the DDI is positioned to become a recognized standard sanctioned by the International Standards Organization, ISO. Having the status of an ISO standard would serve to legitimate the DDI specification more effectively than almost any other activity we could pursue. In our next phase of activity, we need to craft a timetable to set us on the path to becoming an official standard for the social sciences.