

April 24, 2001

Addendum to Final Report
“Electronic Preservation of Data Documentation: Complementary
SGML and Image Capture”
SBR-9617813

Results of the Evaluation of the
Data Documentation Initiative (DDI)

Background

At the request of NSF, ICPSR conducted an external evaluation of the Data Documentation Initiative (DDI), a project funded by NSF over the period 1997-2000. Four independent evaluators were selected to review the DDI project. These individuals were:

- Daniel Greenstein, Director, Digital Libraries Federation
- James Jacobs, Data Services Librarian, University of California, San Diego
- Tom W. Smith, Director, General Social Survey
- Stuart Weibel, Director, Dublin Core Metadata Initiative

The evaluators met in Ann Arbor on February 23, 2001, along with the DDI Technical Working Group and with other members of the DDI Committee. The purpose of the meeting was to familiarize the evaluators with the DDI and to ensure that they understood the charge to review the initiative. The meeting format was open and informal with time set aside for the evaluators to meet on their own as a group. Evaluators' reports were due three weeks after the meeting and are attached to this report.

Summary of Evaluators' Findings

The four evaluators were in agreement that the DDI is a worthwhile scientific effort and that it fills an urgent need for standardization of social science technical documentation and interoperability. One termed it “a strategic component of the infrastructure necessary to support the exchange of structured social research survey data.”

The evaluators all felt there would be a significant payoff to their respective communities. They were impressed by the significant buy-in and international participation that the DDI has garnered and by the in-kind contributions made by so many organizations.

In terms of the substance of the emerging standard, the evaluators concurred that the DDI does an “excellent job for survey research,” but each noted that more work is necessary to extend the standard to more complex types of data – most notably, time series and aggregate/tabular data. It was further suggested that the DDI incorporate elements from

the American Association for Public Opinion Research (AAPOR) standards and best practices.

With respect to the technical format of the DDI – it is written in XML (eXtensible Markup Language) in the form of a Document Type Definition, or DTD – all agreed that this was the “best available choice today” but that the DTD approach has limitations in terms of extensibility. Most believed that moving to another XML format such as Schemas or RDF will be necessary to achieve the modularity necessary for future development and extension of the standard.

The evaluators offered several ideas and approaches to consider as the DDI moves forward and evolves:

- Create “best practices” materials to guide markup authors in applying the DDI to their documentation
- Develop tools for markup of both new documentation and legacy materials and make existing tools widely available
- Define the DDI’s user communities more precisely and tailor it for them
- Provide promotional materials
- Develop controlled vocabularies to govern data entry
- Create a more “light-weight” version of the standard
- Craft a mechanism and procedures for revising the standard
- Work on persistence of names and locations of datasets as well as naming registries
- Be mindful of security and access control issues
- Harmonize the DDI with other existing metadata standards
- Continue to push for adoption of the standard by major communities, including the federal government, private data collectors, CATI/CAPI implementers, digital libraries, digital archiving projects, funding agencies, and software developers

Summary of Beta Testers’ Findings

To assess the experiences of actual DDI users, we recontacted the 13 organizations that participated in the beta test in 1999. Overall, this group is unanimous in viewing the DDI as an important initiative – indeed, one tester termed it “the most significant development in the social science data community in decades.” Another described it as “very commendable” and “very necessary.”

However, the testers are also aware of some of the difficulties involved in using the DTD and the issues that remain to be resolved. The testers highlighted many of the suggestions mentioned as future directions by the evaluators. They were most interested in the development of an aggregate data model, tools for input and translation, and materials on best practices, including a library of examples of DDI-compliant codebooks marked up properly. Other suggestions included adding descriptions of spatial data to the standard and offering training in use of the DDI.

Conclusion

As is evident from the attached reports and the summaries of different perspectives presented above, the DDI effort has been well received by the social science data producer and archiving communities. The evaluators were struck in particular by how quickly this activity has moved forward in comparison with other standards-setting projects, which are notoriously slow to come to fruition. Given the enthusiastic buy-in that has taken place and the investments of the many stakeholders in the DDI, continuing this enterprise seems a worthwhile goal that will ultimately improve the conduct of social science research and move us in the direction of increased interoperability.

DDI Evaluation Report

March 19, 2001

Daniel Greenstein
Director, Digital Libraries Federation

The following points arise from my reading of the extensive materials that have been made available by the DDI via its website and from the review panel's discussion at the end of February.

Before launching into the detail, let me reiterate what I said when we met in February about the very substantial progress that has been made with the DDI. I am familiar enough with standardization efforts to appreciate the significant intellectual, political and organizational work that they entail. I am as familiar with the very long road that the data community has traveled in order to converge the many different approaches to data documentation. The DDI has made outstanding progress not only in developing a viable DTD but also in bringing the data community so much closer to the consensus that it has sought for so long. I am convinced that in as little as 5 years, the very substantial debt that is owed the DDI by information professionals in the data community and by scholars who rely so heavily in their research on data, will be well and widely recognized. Kudos to all who have been involved in this truly impressive effort.

1) The materials available from the DDI website are comprehensive and well presented. They would benefit from some more sample applications, perhaps some explicit examples of both good and bad practice (that is, good and bad practices with data generally as well as good and bad practice in application of the DDI). There should also be a clear statement somewhere about who/what the DDI is for (see comment 6 below).

2) The DDI will inevitably follow a developmental trajectory that is now familiar with similar standardization efforts as they turn their attention from the standard's initial development to equally important task of promoting its effective application.

Here the DDI has a great deal to learn from complementary initiatives (none of them so terribly successful) such as the Dublin Core, the Encoded Archival Description, the Text Encoding Initiatives, and the Visual Resource Association's Core Categories for the works of art. These initiatives succeed best where they focus on implementation guidelines, tools, and incentives for application and do not at least initially overextend themselves by extending the DTD so that it can cater (at least in theory) to more generic types of information. The difficulty of course is that the funding agencies upon which standardization initiatives rely emphasize innovation over consolidation. Here, the DDI with its preponderant organizational and financial reliance on the data community may be significantly advantaged with respect to many other standards initiatives.

3) In supplying incentives to practitioners to actually adopt and deploy the DDI, the DDI may need to consider the phased (prioritized) development of the following.

- Application guidelines that provide examples of practice as fit for particular kinds of data or for particular uses;
- Tools for generating DDIs for those new datasets that have not yet been documented;
- Tools for generating DDIs for legacy data for which codebooks and documentation has already been prepared;
- Promotional materials that explain to ancillary but vital scholarly and library communities what the DDI is and why its application will enhance their work;
- Controlled vocabularies that guide (if not actually govern) the use of key type and subject attributes that are permitted throughout the DDI; and
- Some more light-weight version of the DDI; one that covers application of high-use elements (and probably 85% of the datasets for which the DDI actually applies).

4) The DDI will need to establish a process that will enable it to manage the change requests and other refinements that will inevitably result from its more widespread practical application. Some authoritative voice needs to be established in that process, though needn't necessarily be centrally administered.

5) The DDI should limit the effort spent on extension until the costs and benefits of implementation are better understood.

6) To achieve many of the goals outlined above, the DDI may wish to develop a more refined and prioritized sense of who its users are and of why the DDI may be important to them. At present the form and content of the documentation suggests a specific emphasis on organizations that take responsibility for the long-term management of data. Discussion about the DDI's importance and potential application tends to infer a much wider user community. My sense is that the DDI is likely to achieve more (at least initially) by focusing on the needs and interest of the smaller rather than the broader and more diverse user base.

One senses generally that the days of the "omnibus" DTD (or even standard, if you follow the recent declining enthusiasm for Z39.50, for example). Where DTDs are concerned, the TEI's experience should be instructive. As a result of political and fund-raising pressures, the TEI fashioned itself as a mark-up scheme appropriate for virtually any encoded text and encoded text application. Some would even argue that the TEI provides a wrapper for any digital object. The result is a DTD that is at once impenetrable and infinitely extensible (I suspect that without much effort it would be possible to demonstrate that the DDI is in fact TEI conformant). One consequence is that any random collection of conformant TEI-encoded texts cannot be processed meaningfully as a collection without very significant editorial and other costly interventions in the structure and markup of specific items in the collection. In extending its scope, the TEI ultimately defeated the aims of interchange and interoperability that it set out initially to achieve. In sum, it may be better for the DDI to supply a vital tool for data archives and other data managing organizations than to supply frustratingly partial (and ultimately non-interoperable) solutions to a broader community.

7) There may be a tension within the DDI about the processing aims it is intended to support. Resource discovery and location, and resource management are clearly of paramount importance. Given the archive-centric feel of the initiative, this seems right and proper. There is a budding sense, however, that the DDI may provide a toolkit capable of supporting online data exploration. This hypothesis should probably be tested before any further development work is undertaken in this area. As part of that empirical test, the initiative should evaluate the various business models that apply in the provision of online services that support data exploration. One senses that they apply differently to different institutions in a way that is likely to shape (if not actually undermine) the extent to which it will be possible cost effectively to develop and implement tools that permit interoperability for more than the most basic data exploration tools.

Evaluation of the Data Documentation Initiative

Jim Jacobs

Data Services Librarian

University of California, San Diego

jajacobs@ucsd.edu

March 16, 2001

How do you view the basic concept of the DDI enterprise and the principles underlying it?

This is an excellent concept that is greatly needed and will, I think, be both appreciated and used by data collectors, libraries and archives, distributors, and researchers.

Do you believe that the DDI is a worthwhile scientific effort? Do you agree with the substance of the DDI in terms of elements and attributes included or are there significant omissions and/or redundancies? If so, what are these omissions and/or redundancies?

Yes, this is a worthwhile scientific effort. This is just the kind of scientific research that needs to be done to further social science research and data exchange, and digital library projects. It will provide the foundation for both.

The DDI as it stands does an excellent job for survey research. It provides the opportunity to encode detail down to the level of each individual variable, thus even making it possible to accurately mark up documentation for data that come from multiple sources. I can find no elements that are necessary that have been omitted.

The DDI is less complete for other kinds of social science data. Specifically, time series data and aggregate data (such as census tables) are treated much less thoroughly than survey data. The DDI committee is well aware of this. In fact, they have explicitly stated that they decided early on to concentrate on survey data only and delay doing aggregate data and more complex file types. This was a wise choice, given the amount of work and detail that doing a relatively straightforward format (survey data) took.

The Committee should be commended, in fact for then going beyond that goal and modifying the DDI and doing some initial markup and of aggregate and time series data. These initial efforts are instructive and much can be learned from them. More work is needed in this area.

One issue that future work will have to address is, "Can a single DTD efficiently describe all data types?". It might be more efficient, for instance, to use a modular approach for future DDI development allowing the user of the DDI to

include the module or modules appropriate for the particular type of data being marked up at the moment. This could also help simplify use of the DDI for those who do not want to mark up documentation in complete detail; a simple module with top level elements could make this easy to do without limiting the availability of a complete module for those who wanted to do markup down to the level of each variable.

Such a modular approach could also ensure that future development of the DDI would not either become overly complex or incorporate redundant information. The present design does not have redundant or unnecessary information, but keeping a single DTD for all data formats non-redundant will, I think, be more difficult than creating a modular DTD.

What do you see as the payoff for the communities you are involved with?

The payoffs are already beginning to be apparent. The Nesstar project (<http://www.nesstar.org/>), the work at University of California Berkeley (<http://socrates.berkeley.edu:7502/archive.htm>) and the California Digital Library (<http://countingcalifornia.cdlib.org/>), the Harvard/MIT Virtual Data Center (<http://thedata.org/>), and GRETA project at the University of Minnesota (<http://www.socsci.umn.edu/pdas/greta.html>), the work on DDI stylesheets at the University of Virginia (<http://quincy.lib.virginia.edu/dl/demo.html>), and others are all making use or plan to make use of the DDI. Some of these projects could not have been conceived without a DDI.

Each of these projects provides payoff in three ways for the social science community. First, data collection, once the DDI is adopted widely, will actually improve as researchers who gather data go through the process of answering essential questions about the data collection even before the data are collected. The DDI will help researchers on the long term benefits of their data at the time of collection. Second, the DDI will make it easier to develop interfaces for users to data and to include the holdings of geographically distributed data libraries and archives. Users will then be able to find appropriate data for secondary analysis without having to know which archive has the data. Users will also be able search at the variable level across datasets and retrieve contextual information about the variables. This will literally create a new way to do secondary research as researchers will much more easily discover the wealth of the contents of datasets. Third, data libraries and archives and data distributors will be able to more easily exchange meta-data thus making it easier to create local and specialized catalogs that include information about diversely held data. This too will make it easier for researchers to locate data.

These payoffs will result in better, more informed research and society will benefit from this.

Do you think the technical format (XML markup according to a Document Type Definition, or DTD) is appropriate and optimal? Would another format be preferable? If so, please describe. If you have experience with XML, please comment also on the availability and adequacy of XML and XSLT applications now and in the future.

The XML DTD was the appropriate choice at the time the decision was made to use it. It is still the best available choice today. Other, perhaps better, choices may become available soon. These include XML Schema (<http://www.w3.org/TR/xmlschema-0/>) and Resource Description Framework (“RDF” see <http://www.w3.org/TR/rdf-schema/>). Because these are still in draft form and because their intents overlap, it is not yet time to adopt either.

One of the reasons that XML in general is the right standard to use is that whatever work is done on the DTD and in marking up documentation using the DDI will not be wasted when new standards emerge. Transferring to new technologies will be (relatively) easy and the markup itself should be compatible with future standards.

My personal experience with XML applications is very limited, but I do understand from talking with those who have used the DDI extensively that DDI-specific application may well be necessary. The initial work on XSLT looks very promising. I understand that the project has already created some conversion programs (converting SPSS and SAS data definition files to DDI-XML), and these, with some slight modifications, could well be used by data collectors, libraries, and archives to help through the transition from older technologies to XML.

Can you envision other ways to accomplish the same purpose, and if so, what are the advantages of these other approaches?

As mentioned above, XML Schemas and RDF both hold promise in this area but are not ready to be used yet. The main advantage to using XML over other “standards” (OSIRIS, SPSS, SAS, S, etc), is that the XML standard is being adopted widely outside the social sciences and statistics communities thus making it more robust and viable for the long term.

Assuming you find the DDI endeavor worth continuing, how should the project best move forward in terms of substance and format?

There are seven areas that warrant exploration and further work:

1. Development of DDI for time series, aggregate, and other more complex data formats. Explore the possibility of making the DDI modular to help accomplish this.
2. Also explore modularity for the purpose of making the DDI more easily used in different environments for different purposes. Encouraging the

use of DDI for minimal-level markup of documentation will enhance its visibility and viability in general and make the more detailed markup that promises so much more interchangeability easier to accomplish in the long run.

3. Continue to explore the emerging standards of XML Schemas and RDF and related standards that may offer more flexibility than the DDI.
4. Modify the tools used by the project to convert other standard formats (e.g., SAS, SPSS, OSIRIS) to XML so that those tools may be distributed widely for use in the social science community by researchers wanting to deposit data, by libraries and archives wishing to convert existing holdings, etc.
5. Ensure that the funding agencies and grant seekers are aware of the DDI for development of other social science data-access projects. There are many projects already in development or on their way to granting agencies that could benefit from the use of DDI.
6. Encourage use of DDI and XML by digital library projects. Although data files represent an early and major part of digital collections in libraries, their use and the tools to access them have not always been well integrated into digital libraries. Increasingly, digital libraries will have many different kinds of objects, and data should be viewed as a peer of other kinds of objects. The DDI will facilitate that.
7. Encourage the use of DDI by digital archiving projects such as those done by the National Historical Publications and Records Commission.

DDI Evaluation Report

February 26, 2001

Tom W. Smith

National Opinion Research Center (NORC)

International standards for the archiving and documenting of social science data (especially, but not exclusively, survey data) for on-line searching and access are a vital and urgent need. The DDI is both approaching this task through the best organizational mechanism, a cooperative collaboration of the major social science archives, leading data producers, and other important stakeholders, and proposing a highly suitable technical application involving XML. The goal, organizational, and technical principles underpinning DDI are very appropriate.

DDI's list of attributes and elements, what I'd call fields, covers most important information in a clear and concise manner. I think that a few extensions would be useful. I understand that DDI will incorporate the information required in the standards of minimal disclosure of the American Association for Public Opinion Research and those recommended in its Best Practices booklet, but does not cover all of these. I strongly recommend their complete inclusion and further suggest that AAPOR's Standard Definition of Outcome Rates and Final Disposition Codes be used as the basis for the response rate field.

DDI's payoff will be enormous. With data archived according to DDI standards, tools like NESSTAR will allow the interrogation of data holdings and archives across institutions and countries. This will greatly assist social science research and increase productivity notably.

The DDI should proceed on several fronts. First, it should disseminate information about itself to those not already in its considerable network of collaborators. In particular, data providers should be brought on board. This would include both adoption by government agencies such as the US Census and private data collectors. More contacts with organizations like AAPOR, WAPOR, CASRO, and ESOMAR would facilitate the latter. Second, data-transformation interfaces with the main analysis software developers (SAS, SPSS, STATA) and the main CATI and CAPI data collection programs should be developed. Hopefully the main burden of this can be taken up by the software companies with consultation from DDI. High priority should be given to these and other easy to use interfaces so that data providers will start to follow DDI protocols and the enormous job of data archives is reduced. Third, DDI should press forward with improvements in the classification structure for typical survey data and for the extension of DDI to handle other important social science data types such as time series and aggregate data. I'd put less immediate emphasis on exotic mediums such as audio-visual material that may appear in web surveys, but would keep such in mind for the future. The idea of a keyword field on the variable level is a good one.

Evaluator Background

Stuart Weibel is Director of the Dublin Core Metadata Initiative, hosted by the OCLC Office of Research. He has led the initiative since 1994 and consults widely with potential adopters.

Dr. Weibel has not consulted with or for ICPSR or DDI prior to this evaluation, nor does he have close professional or personal relationships with any of the members of these teams.

I. Questions for Evaluators

The following general questions were asked of the external evaluators:

1. Basic concept of DDI and underlying principles

Perhaps the most useful metaphor for the Web is that it reduces friction for the exchange of data and information. The DDI is a strategic component of the infrastructure necessary to support the exchange of structured social research survey data. This important class of research data has long been the focus of international collaborative attention, providing as it does a wealth of long-term information about issues in policy and social research. Effective use of such data has been hampered by lack of standardization of protocols, methodology, and interchange formats. Providing standard means for capturing and structuring such information influences each of these limitations, reducing the friction for discovery and reuse of such data.

Promoting machine-readable, standardized, structured encodings for the data and metadata of such surveys will foster the emergence of tools to assist in the creation and management of such data sets according to commonly accepted principles. The resulting data will be easier to use and compare to other data sets.

Perhaps the most important impact of standardization such as is represented by the DDI effort is that the data will be far easier to process, analyze and compare, greatly increasing the impact of the funds spent on creating the data initially, and creating new opportunities to use the data productively. Without efforts such as this, such use is limited by the costly effort of hand-crafted analysis.

2. Is DDI a worthwhile scientific effort?

The DDI is not just worthwhile, it is essential if the field of political and social sciences research is to realize its potential. The Web has made possible a level of collaboration unthinkable a decade ago. Capitalizing on this greater capacity for information flow requires development of standards that will support the necessary interoperability and interchange among systems and data sets.

Achieving this degree of interoperability is far more than the elaboration of a set of elements... it represents substantial intellectual investment in conceptual schema formation as well as the formulation of technological strategies in a world where the

technology is evolving rapidly. All of this must be done against a backdrop of international and inter-institutional collaboration that is challenging to say the least.

3. Substance of DDI in terms of elements and attributes

I am not sufficiently familiar with the field of social and political science research to comment on the element level organization of the DDI; my perusal of the DTD suggests that it is quite reasonable. The higher level organization of the metadata structure seems sensible and useful:

- Document description – bibliographic description of the rendition
- Study Description - converging with AAPOR (American Association Public Opinion Research) best practice about study description, methodology,
- File description – access to file level characteristics of the data
- Data Description – structure of the data
- Other material

Compartmentalizing the metadata in this way will support the modular development of system functionality and ease the task of improving and maintaining the metadata and associated data sets.

4. Payoff for the communities you are involved with

Every community concerned with interchange of data must come to grips with identifying the characteristics of their data that will promote discovery, access, reuse, and appropriate management characteristics (such as intellectual property rights and preservation characteristics and responsibilities). The DDI effort is an excellent model for communities with formal data structuring challenges.

5. Appropriateness of technical Format – XML and DTDs, and availability of tools and applications

One of the great challenges of the DDI effort is the rapid evolution of the enabling infrastructure that underlies their efforts – initially SGML, more recently XML and its many sub-parts. The state of the current effort is about as sensible as it can be in light of this ongoing technological shift. The use of XML is the only sensible thing one could do... any other approach would be seriously flawed. The use of DTDs is less clear. The ICPSR team is well aware of the existing options – DTDs, XML Schemas, RDF Schemas, and continues to explore them. DTDs are stable and a known approach, but have a significant liability (they are difficult to extend or modularize). Plans for Version 2 of the DDI standard are exploring the options for overcoming this liability.

6. Are there other ways to accomplish this work?

As indicated in the previous section, there are alternative strategies that the design team is exploring.

7. How should the project best move forward in terms of substance and format

The DDI effort is an ongoing effort with a good track record. It represents excellent value for the funding agencies that have supported it thus far, which pay a small fraction of the overall committed funds. The fact that it is international in participation and scope

is a strong point that should encourage support from funding agencies in both Europe and North America, especially in light of the increasing importance of trans-national funding programs. This dimension of the work should be encouraged and even expanded to embrace a broader spectrum of international participation.

II. Issues arising from the evaluation sessions

The following is a summary of major specific issues (and this evaluator's commentary) that arose during the course of the evaluation session.

1. Data structure and encoding issues

a. Dynamic character of survey data

The nature of the data characterized by the DDI effort is, not surprising, constantly changing. The protocols and methods are diverse, and new means of web-based polling, for example, increase such diversity. Static means for describing such data are doomed to immediate obsolescence. DTDs will not work well in the long run, as the development team is aware.

b. Machine understandable, rather than just machine readable?

Syntactic standards for data exchange are essential, but not sufficient, for effective use and reuse of data. Syntactically correct gibberish is still gibberish. One of the goals of the DDI effort is to push the data encoding envelope towards structural and semantic coherence as well, which should improve the usefulness of DDI data sets. The goal is challenging, and should be looked at as a guiding principle rather than as 'the grail'. True machine understandability will continue to be a distant target, though one that can inform strategic directions.

c. Interoperability:

The long term history of ICPSR standardization of data formats is continuing to evolve. Interchange Formats began with flat tables characterized in paper code-books under the OSIRIS code book standard, evolved into the SGML world of structured markup, now into XML, SGML's simplified successor on the Web. XML holds the promise of well-structured and strongly typed data for Web interchange. The standards for syntactic interchange have thus improved markedly. Semantic Interoperability is much harder, and requires far more intellectual investment and consensus building.

The investigators are well aware of the importance of keeping content separate from syntax and structure, a mistake often made in the past in many data communities. Planning on the necessity for future conversions from a given syntactic encoding to another is an important feature of the digital age.

The near term challenges will involve issues such as:

- How can the DTD be extended?
- Should either XML or RDF schema technology replace the use of DTDs?

2. Supporting Infrastructure

The metadata requirements of the DDI effort touch many of the classic issues of developing metadata in other domains:

- How should controlled vocabularies be incorporated into codebook elements... for example, for subject, spatial and temporal coverage?
- How might thesauri be deployed in ways that advance the interests of the DDI community and also promote interoperability with other communities??
- How can persistence of names and locations of data sets be assured? For example, the data-bookmark facility in NESSTAR is a significant feature that will be brittle and unreliable in the absence of persistent identifiers and reliable service locations)
- Naming registries might be an appropriate service for ICPSR to provide for its constituents. URNs, PURLS, DOIs, and the OCLC Naming service should be explored.
- Security and access control issues are important components of ICPSR services. The marketplace will provide technical solutions for these problems, but it will fall to ICPSR to promote useful solutions within its community.

3. Process

Evolution of the DDI standard requires some level of process to support change while maintaining stability. It will be essential to maintain stability that stakeholders have invested in while assuring deliberate migration to new technologies as these technologies become important. It will be important for DDI to stay relevant without succumbing to the dangers of “creeping elegance” - the natural desire of technologists to do the newest, best thing, which may cause more problems for stakeholders than it solves.

4. Marketing Issues

Attention needs to be paid to the question of how to get the strategic adoption of the DDI standard. Clearly an important collection of international adopters has committed to DDI. Yet, there are important agencies missing that could bring the force of the standard to a stronger level. This problem will benefit from leadership within a group (ICPSR?) that has the authority of the community and the credibility of a major international consensus.

The emergence of tools to support the creation and management of such data is probably an important key to further adoption. The growth strategy should address the question of how to take advantage of the emergence of general metadata and XML tools and foster their specialization in ways that serve the DDI community.

5. Standardization

One way to assist the growth of deployment of DDI is to couple its use to other standards that are important to the stakeholders. The list of such standards is potentially long; it will be worthwhile for the leadership of DDI to give thought to which strategies might particularly enhance the value of DDI. Possible candidates (mostly well known to the DDI team) include:

- Dublin Core

- MARC standards
- Uniform Resource Names
- PURLS
- Digital Object Identifiers
- Open Archive Initiative
- W3C standards (Many, including of course the XML family and RDF)