Bill: Manifesto lacks sense of wide-spread support. How do we get agreement on part of social scientists broadly defined.

George: DDI wants to show that the science can be done better with a metadata-based workflow. Outside the DDI Alliance, there aren’t many social sceintists who don’t know what that means. We need to state the vision about what we mean and what the benefits are.

Chuck: Where do we fit into the research infrastructure? I listed a SWOT analysis of the DDI specification and the Alliance. To my surprise I had a lot more strengths than weaknesses or threats. Your principles should reflect your strengths.

Discussed SWOT analysis:

Strengths

* Community-based organization with standards setting processes well developed. Reflects being inclusive and representative of the community.
* 22 years of experience with multiple lines of specifications and controlled vocabularies
* Knowledgeable and skilled core technical development team + marketing and training groups
* Software implementations of DDI specifications and vocabularies
* Recognition by NSI’s (national statistical institutions) and data producers like MIDUS
* Active empirical SBE user communities (surveys, archives, etc)
* Adaptation to changes in information technologies (XML, RDF, Schema, …)
* International partnerships
* Skilled trainers
* A lot of data in DDI in the social sciences

Weaknesses

* Sustainable business model
* Domain isolation (SBE vs observational research)
* Lack of an organizational succession plan
* Primary foothold is at the latter stages of the lifecycle; little buy-in by data creators or researchers
* DDI is not built into the tools that researchers and data creators use
* Uncertainty about path from DDI 2 to 3 to 4.

Opportunities

* Digital Research Infrastructure provider in the cloud: registry services operated by the DDI community
* Alignments with the international RDM community (look at FAIR partnerships or the new OCLC + euroCRIS partnership)
* Connect to new data discovery ecosystem, such as schema.org and DataCite
* Enable discovery, interoperability, and integration of distributed data
* Build DDI into Common Data Element registries
* Integrate the CCSDS DAI (data archive interoperability) WG’s Architecture Concept
* Integrate DDI across stakeholders activities across the lifecycle - greater participation from groups across the lifecycle

Threats

* Loss of key experts in the community
* Loss of the DDI Host
* Failure of adoption by software producers
* Need to renew active membership in the community as leadership turns over
* Too many specifications and incompatible dialects

Steve: I stay away from the term metadata. In quantitative social sciences we have four tools: SAS, SPSS, Stata, R. If teaching, don’t talk about metadata. Do talk about value labels, questions, etc.

Data producers are important group. So are Data preservers.

The risk with NSI’s is that they change their mind frequently. They

WIth sustainable business model, we were really successful with attracting new users. But contribution is mainly content based, but what does that mean for maintenance. We are really dependent on volunteers. If we do not focus on our direction, the workforce can go their own way.

Community-based organization. It’s essential, but are we organized in a way to make most use of it. How do we have the resources to pay for an Executive Director, project manager, contractor, etc.? So much has been accomplished by in-kind contributions. Have we run out of the good will? To move into new opportunities, we need a project manager, standard maintenance.

George: we have to look for things to make researchers’ lives easier if we want them to use it.

Steve:

Examples:

1) <https://github.com/IQSS/dataverse-client-r>

Makes it really easy to deposit content into an archive. SWORD-based data archiving

2) Stata creating dynamic markdown to make things easier to do things

<https://www.stata.com/new-in-stata/markdown/>

3) qualtrics r package

https://cran.r-project.org/web/packages/qualtRics/index.html

Gets content out of Qualtrics, RedCap. These tools don’t use DDI now, but you need something like DDI .

Chuck: We need to be part of the greater research infrastructure environment. We should think of how to deliver a metadata service to techies and others.

Achim: Markdown is very successful. It could be a model for DDI. DDI is only partially successful in terms of software. If one tools exports it, if the next tool tries to import, often it breaks. That means DDI is not standardized enough. What did Markdown do? There were a lot of different branches. This community was large. They just developed tools for all branches. We can’t do that. We have to agree on one way to do DDI. Last week, we talked about doing test cases that agree on one dialect on DDI. Regarding softare, I’m not sure how far the Alliance should do. We should focus more on import/export and portability of DDI. Therefore, this is threat that we have too diverse of dialects.

Steve: This is why I mentioned data description in R package. The question for us is about strategic points to think about it. One is best practices. What helped us standardize DDI? Nesstar helped define a lot of things. Have reservations on going too far on software development, but there are some advantages.

George: Regarding dialect issues, there could be a dialect validator. The Alliance could create this. W3C did that years ago with HTML.

Agreement that validation would be a good thing.

Steve: a lot of tools being used look a lot like DDI.

Chuck: Everyone hears about FAIR. That’s maybe helpful for repositories, but it doesn’t describe whole context of DDI. It’s more like PLAY FAIR. Preservation, etc. There is now in the RDM community a movement to repositories. We should support that. Find doesn’t do it. It’s more about discovery. You can find 1,000 datasets, but you should be able to discover the relationships among those data. What’s missing is bringing together 10 datasets and figure out how to integrate. That puts us into alignment into the international RDM community.

Chuck: Right now, lots of discussion about supporting research data is about the Cloud. Don’t know if we’ve thought about how we built our standard in this context. Standard for the codebook was built with desktop context.

Steve: We’ve got there without thinking about it. A lot of reusable components could be supported by registries.

Chuck: We’ve taken it for granted, but we haven’t made it as explicit. Our opportunity with DDI 4 is to do machine-actionable metadata.

Chuck: FAIR has become a movement. We want to bring a repository perspective with PLAY FAIR.

George: You need a registry rather than a question bank so it can be machine-actionable.

Chuck: Don’t make something a portal, but have a machine capture it.

Achim: Regarding registries, DDI was only thinking on the specification and not on software or a framework around DDI. SDMX did much better in this regard. They produced a specification and an entire framework. They have registry, REST protocol, glossary, etc. It’s pretty complete. That could be a role model. We already have a globally unique identifiers, but we don’t have exchange protocol, etc.

Achim: Chuck mentioned domain isolation. DDI did technology isolation with DDI 3. We just used XML. With DDI 4, we have new opportunities. If we take modeling seriously and use model and related RDF representation, new worlds are available. It’s a big step forward for interoperability and discovery. People currently only partially realize about DDI 4 opportunities.

Chuck: Poeple think about what it will add to things now, but it gives us lots more opportunities 10 KM ahead.

Achim: We’re solving the technical interoperability issues, but we’re still not dealing with domain barriers/isolation.

Chuck: DDI’s strength is in observational research. If you look at domains looking at observation, DDI can support those elements. We’ve never really focused on that. A lot of it is because a lot of our elements have social science elements associated with it and they don’t call it that. Now’s our opportunities with registries. We’re not training them to code DDI from scratch -- we’re simply saying they have observation and our system says do this. One way to bridge the domain thing is to think bigger in terms of observational research and how we can deliver support.

George: I’m skeptical of that due to contacts in biomedical world, which never gets rid of old ontologies (E.g. FIHR on top of OMAP, etc.). What’s important is how to make things interoperable (doing mappings in between). bioCADDIE created its own metadata standard and showed how to map between standards. DDI does well with this. I think you’re right, but you need to approach it differently. We shou

Chuck: We have shared common elements that all observational researchers have. It would help them since they would have core elements to draw on.

Achim: Codata is organizing a workshop on data integration across domains….

Chuck: Interdisciplinary research is being promoted as solving science problems. How do you do that? Discovery. That’s where DDI has natural advantage.

Hossein: Do we have a methodology to create this strategy? We should know about the current situation of inside and outside? Have we done it? Some service providers don’t use DDI. Why don’t they use DDI or DDI 3. If we know why they don’t do it, we’ll understand weaknesses.

Chuck: We’re using this as a means of opening the discussion. It’s not the end. There should be a wider process. We’ll eventually take this back to the community. Methodology of DDI: nothing gets proclaimed without review and approval.

Hossein: What is the DDI business model?

Bill: It’s evolving

Chuck: My immediate inclination is to go to the charter, which discusses ongoing relationships with the Alliance. The problem, though, is that it doesn’t resolve the operational needs that we’re discussing. We have not monetized DDI. It’s open and free. We’ve leveraged funding through grant applications and membership. The business model needs to continue to be discussed.

Chuck: Cannot see us moving to a cloud based registry implmeentation of DDI with either DDI 2 or DDI 3.

Hossein: Should we have IP strategy as part of the plan?

Mari: This is covered by the Producer element.

Mari: DDI is about re-use (the R in FAIR)

Next steps:

In addition to three documents:

* Mission and Guiding Principles
* Strategy Plan for the DDI Alliance
* Infrastructure Vision

Also, possible 4th document to cover:

* Provide examples that show new capabilities of DDI 4.
* Show relationship of DDI 4 to 3 and 2.

Chuck: once DDI 4 comes out, it would be good to show a chart providing comparison on what features each version offers. Ideally, DDI 4 features would cover everything done in DDI 2 and 3.

Tomorrow morning, we’ll discuss all documents and determine what should be changed. Then in the afternoon, we’ll start writing.

Chuck: We should promote and partner about DDI. DDI is part of the wider research community digital infrastructure. We need to strategically identify with other partners and collaborate. For example, OCLC announced they’re partnering with EuroCRIS, the organization around research information systems. They’re looking at the management of research information and working together about cataloging interests and creating large base for doing research. ARL has created SHARE. Looking at how to make connections between metadata. Who is DDI looking to partner with to improve discovery infrastructure for research data. Synergy and knowledge gained. DDI is an integral component of global infrastructure. We aren’t going to do this by ourselves.

Chuck: OAIS reference model people are doing an inter-operabilty specification that basically describes our infrastructure approach. Do we need to align ourselves with a partner like them? This could fit within mission, vision, and strategic plan. That is, we fit in as an essential component of global research infrastructure.

Chuck: Richness of DDI is that we identify attributes at the study level, as well as the detailed data level. WIth other standards, you don’t see that opportunity. That translates into the discovery layer, especially for cross-disciplinary discovery. We need to think of ways to capitalize on study- and variable-level discovery -- and how we can contribute this to the larger research community.

Chuck: We need to think of ourselves as part of the research infrastructure community. The mission statement should identify DDI as key component of the global research infrastructure, much like the astronomers talk about their infrastructure as part of the global community.

\*\*\*Wednesday\*\*\*

Chuck: I was involved with meeting of supercomputer community. It was clear that what they needed to know how to collaborate between libraries and supercomputer community. Key is trust. Have to understand history and respect where they’re coming from. Develop a common vocabulary, etc. Guiding principles for being a welcome collaborator in diverse community.

Jon: What is infrastructure for? I’ve just been sent UK plan for research infrastructure.

Achim: We were talking about queries against the common registry. Types of queries by user group.

User groups:

* People looking for data
* Integrating data across multiple sources and multiple distribution sites (e.g., if want to know if attitudes of crime are correlated with unemployment rates, you need to make sure same measures were used). Finding data, but finding data with a certain kind of specification.
* Designing an instrument who wants to know what instruments have been used in the past. What questions are there? How well did those questions perform?
* Data curators -- curation of standards and registries. We will need to redefine what curation means. We’ll need to make sure the metadata conforms to international standards.
* Data curator -- query the CDE to embed geographical representation in registry
* Administrator function -- make sure the CDE is up and running
* Governance process -- people propose CDEs, but is it well enough described?
* If you want data elements to be linked to concepts, that requires a lot more effort and we don’t have good ways of doing that.

Jon: how to get buy-in of other organizations to allow use of standards in another place. Linked data would be a good way to use it rather than archives pulling it.

George: Getting things to work is a chicken and egg thing. Some repositories will put it in DDI compliant format. Once it’s in there and people using it, ILO might say I have control and others can’t use it.

Arofan: This exact issue of ownership and agency became huge debate in SDMX. In the identifiers of SDMX you have an agency (owner). Can represent one or more organization. It’s a really important policy. You have to get agreement of other organizations.

Chuck: Curator’s going to ensure agency has been validated. This is going to need curation.

Chuck: Another curator’s challenge is versioning.

Chuck: Those queries are more of management of a CDE than its use. We did raise a whole set of issues of operational management of a registry.

Achim: Two purposes. The vision should outline on a high level some of the techniques. Techniques are some measures to achieve something. The queries can show that. The other purpose are these requirements expressed by the description of the complex queries. Are they good enough to discuss technical solution which is more a part of the strategic plan. We should express this at a high level good enough so the lower level work can be done.

George: There’s another use case: one of the problems that researchers commissioning a dataset in the survey world. They have an idea of a concept of a question, but they have to communicate that to a data collection agency. National Election study, this is a big problem since they change questions over time. They have to check that questino is implemented properly in computer system. CDE registry solves this problem. The client says they want this CDE with this number. THe data collection firm can go to the CDE registry witth the number and download all the attributes with the number. Anything you can put into DDI to describe a variable could reside in a CDE registry. The communication between the client and the data collection agency becomes much simpler.

Chuck: If there are search engines that could take advantage of a CDE to improve searches.

George: bioCADDIE search engine is looking to do this. What if you knew that this term is part of this ontology. Search returns not just all synonyms, but it can also say this item is part of this ontology. Do you want to see other things that are related to this ontology. In linking to CDE registries in biomedical world, some of those registries have ontology to improve searches.

George: One of ways to tying into existing things: you have big long-running projects (ISS, national election studies), internal to projects, they know which questions are the same from year to year. If they could put that knowledge into existing registry.

Jon: Why don’t they put it into the archives?

George: It is in the archives, but going from social survey to election survey isn’t easy.

Chuck: Another user example: publisher. What might a publisher use the CDE for?

George: What Elsevier wants to do is put on their publication a box where you can play with the data analysis. They have a prototype linked to the protein database where you can play around with the protein directly within the browser. The DDI-Model enables that.

Chuck: ANother aspect: Funding councils. They spend a lot of time where new investments should be made in research. They spend a lot of time monitoring. The tools we are building could help them understand trends and changes. New stuff going into CDEs could help funding agencies understand new directions, especially interdisciplinary projects.

George: A couple other examples. The research council wants science to be cumulative. Every study now tends to be a one-off, especially small studies. If those small studies used data elements linked to other things, you get better, more cumulative science.

George: Agencies also interested in places where you can make connections. If you see two disciplines using the same measures, even the same data but not talking together, the funding agency might create new funding program to encourage the programs to talk to each other.

Chuck: This gives you an idea of wider range of uses in wide research infrastructure community.

Jon: There’s some confusion of research registry. There are views partly directed at different communities.

Achim: Purposes of this discussion: 1) Do we have enough meaningful complex queries as examples of showing the purpose and advantages of registry? 2) Are these queries detailed enough to describe requirements of the technical solution? Maintenance, user perspective, and techincal levels.

Jon: What sort of requirements would be required in a specific amount of time to create a registry?

Chuck: The compromise is to show diagrams to show user requirements, but don’t have much description to help technical description.

Achim: Reviewing the diagrma: We need to define the interface to the repository and registry. We need communication between repoistory and registry for legacy stuff and adminstrator/curator and registry. FOr complex queries, we need something. We didn’t talk about persistent identifiers, we need mechanism to define the PIDs on the CDE level (not the registry level, since repositories already assign PID). Then we would like to expose both metadata from the repoistory but also CDEs for the GOogle search. Currently it would be possible by schema.org. Then we have some interaction with DataCite. We have a reference to DDI instance in a repository. We should make a best practice for repositories when they put things in DataCite, they’re following best practice.

George; We should put in strategic plan a group to advise Alliance and metadata repositories about licensing agreements. The Alliance should have guidelines on terms of use and licensing arrangement.

Achim: Regarding licensing, there should be written best practice that could be given to repositories.

Arofan: Also should address service level agreements.