

Recent Data Documentation Initiative (DDI) Related Events

Personal Report: Scott Crosier – CSISS / ESRI

Describing Social Science Data – A Brief history of the DDI

(An excerpt from: Crosier, S. J. [2002] "Defining Space - A Metadata Approach" [in publication] Of Significance...)

Metadata for the social science community is not new. Throughout the development of social data, the use of codebooks has been prevalent. These codebooks acted to help the data user understand the results of the research. They included such elements as details about the group that was surveyed or studied, the questions and methods of data collection, and the final evaluation of the data. The basic principles behind these codebooks have been modified into a standardized, digital format. The Inter-University Consortium for Political and Social Research (ICPSR) has been the leading force in standardizing the documentation of social science datasets. In this effort, they have developed a metadata standard entitled the Data Documentation Initiative (DDI) (ICPSR, 2001.) Although this standard has many of the traditional properties of social science data it is less specific in addressing the spatial characteristics of the data.

For example, the finest resolution that is specifically defined is the name of the country or countries to which the data corresponds. Most other spatial aspects are addressed through a free textual entry entitled iUniversei in which the data documenter can specify other spatial properties of the study. With less emphasis on the spatial properties of the data, it is difficult, if not impossible to search for data based on any other spatial extent smaller than the country. To overcome this limitation, a proposed modification to the spatial definitions of the DDI is underway.

DDI Geography Workshop

Upham Hotel, Santa Barbara, CA
26 August 2002

General Description:

After expressing interest in the development of the DDI, representatives from the Center for Spatially Integrated Social Science (CSISS), the Alexandria Digital Library (ADL), and the Map and Imagery Library (MIL) from the University of California, Santa Barbara (UCSB) were asked to participate in the workshop discussing the development of additional elements of the DDI relating to geospatial properties. The meeting brought together a number of those influential parties involved with the Inter-Consortium for Political and Social Research (ICPSR), those interested in the development of the DDI to describe geospatial elements, as well as those experienced in developing metadata schemas and describing and cataloging geospatial properties.

Participants:

Ronald Wilson – ICPSR, Univ. of Maryland; Head of the DDI Geography Working Group
Peter Granda –ICPSR, Univ. of Michigan
Wendy Thomas – Minnesota Population Center, Minneapolis, MN
Michael Goodchild –CSISS, Geography, UCSB
Scott Crosier – CSISS, UCSB / ESRI, Redlands, CA
Eric White – CSISS, UCSB
Mary larsguarrd – Map and Imagery Library, UCSB
Linda Hill – Alexandria Digital Library, UCSB

Agenda Topics and Resolution:

Introduction to the Center for Spatially Integrated Social Sciences in regards to mission and objectives, key personnel and their role in aiding the Geography Working Group (GWG).

- Dr Goodchild made a presentation discussing the basic need to describe a location in the same manner. He also expounded on the six principles of CSISS, mainly: linking data through a common location, spatial analysis, spatially explicit theory, place based analysis, knowledge and policy, and place based search.
- Each participant introduced him or herself and briefly described their interest in the development of spatial elements in the DDI. Each also discussed their exposure to the DDI as well as their experiences in developing and examining metadata schemas.

A follow-up on previous meeting item on the relationship of International Standards Organization (ISO) vs. FGDC standards what it meant to be 'FGDC compliant'.

- Ron Wilson briefly discusses the results of the previous meeting with the DDI working group from the ICPSR over the development of the DDI, including his assignment from them to develop the spatial or geographic elements for the DDI.
- The topic of relating the DDI to the ISO standard on geographic elements or the FGDC's standard was begun.
- It was discussed that the elements that the ISO and the FGDC use are similar, and that as long as the DDI described spatial elements in a method that defines the spatial elements, tools are available to translate one standard to another. Also, the incorporation of ISO and the FGDC into single standard were discussed. In essence, if you map to one, you will map to both.
- Ron Wilson was interested to get the perspective of 'FGDC compliance' from those involved in the ADL as well as the MIL. The general consensus was that to be 'FGDC compliant' had little bearing, as the spatial elements involved were the only part that was needed, not to comply with every aspect of the FGDC. Many elements in the FGDC do not relate to social science data.

To be further addressed from the previous meeting is the issue of the importance of descriptive issues to the Working Group and recommended that they do not incorporate all geographic concepts into the DDI but only those that applied to social science data.

- Concerns were raised by those from the different affiliations of UCSB at the idea of liberating the user/documenter from dictating the presence or absence of geographic concepts in the resource and its related metadata. This would permit the user/documenter to omit potentially useful spatial material from the documentation, considering it miniscule or unnecessary.
- Those more familiar with social science material suggested that the spatial elements involved in some studies were at times more vague than many of the resources that geographers or earth scientists might use.
- The issue was raised for dealing with material that had already been cataloged or for which metadata had already been produced termed as 'legacy material'. Wendy Thomas particularly raised this point, whose organization has already invested a large amount of time in documenting historic census records. The argument was that for those materials that had already had their spatial elements described in a particular, perhaps obscure format, a method of describing the translation technique should be made available with the prior method of spatial description remaining. This would produce a type of indirect reference to the spatial properties of the material. In order to determine the spatial footprint of a

data source, a user would need to find the datasets, determine the method of translation into traditional, XY coordinate values, make the translation, and then find the data according to the obscure spatial reference technique. While this technique will still permit a spatial digital search to be made, it will increase the steps in the process. The argument still stands between those interested in getting the DDI to accept their legacy methods previously used to catalog geographic properties and those interested in making the DDI a key part of digital search tools.

What elements should be selected from the FGDC model for inclusion into the DDI DTD? And, should they follow the same format, terminology and attribute? Would it be wise to incorporate the minimum standards of the model? If so, then should they be enhanced or left as is? If no, then what should be used?

- The FGDC is organized into a format that has a section dedicated to the spatial location of the data. It was recommended that this section be used as a guide in developing the geographic elements for the DDI.

Discuss an example of modified elements for the DTD that includes geographic and spatial data. Specific questions need to be addressed as to why the lowest level of geographic aggregation covered by data be repeatable and whether or not a warning description be used to state the potential dangers of doing analysis based on the lowest level of geographic aggregation

- Several methods were discussed however, at this point in the workshop, sufficient ideas had been shared in what geographic elements should be addressed in the DDI to enable Ron Wilson to develop a sensible relationship to basic geographic elements.
- Universal consensus was made in the fact that as we make a resource available to others, a potential user might misuse the results of a particular study and base their research or project on their own misuse of the original study or material.
- It was further pointed out that all elements of metadata should assist the potential user in determining the proper use of the resource. Other elements could also be added to directly address that issue.

What other options are available for incorporating geographic and spatial data elements into the DDI DTD?

- This point was briefly discussed. It was decided that the best option would be to emulate either the elements in the FGDC or ISO standard.

Discuss the inclusion of a Tag in the DTD that includes multiple examples of how to describe and handle geographic and spatial data, as well as aggregate and hierarchical data for the purposes of educating users on how to employ these elements correctly in their markup

- This returned to Wendy's point in regards to legacy material. Her point also arose that the spatial footprint of some social data is vague, covers only particular, disconnected points, or has a global footprint. The need to address both of these issues was discussed.
- The plan that was formulated during this workshop will result in a weaker emphasis on the spatial elements of the DDI. A documenter will be posed with the question if the data contains spatial elements. If so, the question will be addressed if the geography is described directly or indirectly. A direct description will include elements similar to those found in the FGDC metadata standard. An indirect description would include one where the standard geographic descriptors are provided through a translation tool. This approach will, in theory satisfy

spatial searching as well as accommodating those with legacy material without the resources to update their metadata.

Discuss the inclusion of lines as a geographic feature in the DTD. Much has been discussed in regards to locations [points] and boundaries [polygons] with little emphasis, or even consideration of networks [lines], which there seems to be little interest in doing so.

- The purpose of considering lines as a type of geography was discussed considering a number of aspects.
 - In terms of current spatial search tools, a network of lines would not differ in that, much like a multi-point feature, a bounding box surrounds the network. To search for a line feature, the bounding box coordinates or the encapsulating polygon of the network would need to be recorded. In this way, a search tool would determine if the region of inquiry coincides with the region of the line feature. Current search tools would not search for line features in any line-specific method.
 - Incidents on a line feature, such as a network, are, in essence, still point features. Social issues such as immigration or the spread of disease are often envisioned in terms of a network, however these networks consist of the joining of several point features.
 - The current techniques of describing spatial elements consider lines as point that are joined together. The properties of those points weigh stronger than the lines that join them.

Addressing used vocabulary towards geographic and spatial data is imperative. The terms *codeset* to describe polygons and *planar* to describe points is ambiguous and incorrect respectively. The use of this vocabulary will likely cause miscommunication if between geographers, social scientists and any others who use the DDI due to the conflicting context for which they are used among the aforementioned groups.

- It was discussed that the majority of data users who would be interested in the spatial aspects of the DDI would also be familiar with the tools used in spatial analysis. These tools have their roots in the more traditional spatially minded sciences. Therefore, it was decided to use many of the terms familiar in the GIS community, rather than attempting to use terms familiar in other social science fields.

Discuss the concept paper “A Reuse Approach to Metadata Documentation for Geographic & Spatial Data” in order to identify if there is one, any merit to this concept and two, can it be mapped to the FDGC and DDI Metadata Models?

- This is a paper that Ron Wilson was developing and originally wanted some feedback on it. At the time of the workshop, he dismissed this bullet point.

Results:

The ideas that were shared will be incorporated into the report and work being done by Ron Wilson, acting in behalf of the ICPSR, as the head of the DDI geography working group.

Metadata for Time and Place

Conference Sessions

Social Science Historical Association Annual Conference, St. Louis Missouri

24 - 27 October 2002

General Description:

Several sessions at the conference were developed to both accommodate the agendas of historic social scientists who were interested in the topic of metadata and the DDI and also to accommodate the theme of this year's conference "Thinking globally, researching locally, working cooperatively." Presentations were made ranging from general metadata material, the application of metadata in a number socio-spatial projects, and development with the DDI in terms of spatial elements.

Participants and Presentations:

Adding historic maps to the National Spatial Data Infrastructure, James W Wilson, James Madison Univ.

- James Wilson addressed some of the issues that they were facing and accomplishments they had made in collecting and distributing temporal data via the NSDI.

ECAI / TimeMap metadata: there's more to metadata than just finding things, Ruth Mostern, Electronic Cultural Atlas Initiative, Univ. of California, Berkeley

- The TimeMap project is researching methods for recording, indexing, analyzing and delivering Humanities data with spatial and temporal components and make it accessible on the web through indexed access to scattered resources, interactive maps, timelines and map animation.
- Using web or Windows-based software, one- can register and share cultural data through a central clearinghouse and combine-dispersed datasets to generate interactive maps and online publications.
- The TimeMap website can be found at <http://www.timemap.net/>

Federal Geographic Data Committee Standards, Sharon Shin, Metadata Coordinator, FGDC

- Sharon Shin, who has only recently become the Metadata Coordinator for the FGDC briefly discusses the

Standards for gazetteer data and gazetteer services, Linda Hill, Alexandria Digital Library, Univ. of California, Santa Barbara

- Linda Hill gave an introduction to gazetteers, including what they are and how they can help in a placed based search.

DDI and the National Historic Geographic Information System, Wendy Thomas, Minnesota Population Center, Minneapolis, MN

- Wendy Thomas presented information about the advancements and purpose of the NHGIS and then related some of the geographic elements and issues being addressed by the DDI.

Geographic elements and social science data: The archival perspective, Peter Granda, ICPSR, Ann Arbor, Mich.

- Peter Granda gave a very thorough presentation of the DDI including the role, development, current products and information available, goals, issue and future advancements.

CSISS and the DDI, Scott Crosier, CSISS, UCSB / ERSI

- I began my presentation by discussing spatial patterns, using the example of a map where the tem used to identify soda was mapped across the continental United States. In this we discussed the spatial arrangement of social behavior, spatial extent and resolution of the map and study.
- I then discussed the relationship of varying fields of research through a common location, citing several examples.
- I concluded by discussing the role of the DDI in providing a framework to search and analyze data based on location.

Discussant's Remarks, Humphrey Southall, Great Britain Historic Geographic Information System, University of Portsmouth

- Humphrey Southall concluded with an overview of the two sessions, raising several issues that were discussed during the sessions.
 - He reexamined the purpose of metadata, namely that it fulfilled three major roles: Data discovery, data quality evaluation and data integration and sharing.
 - The role of metadata was also mentioned. Termed as the “Killer App” Dr. Southall discusses some of the current applications and some of the other tools, including gazetteers used in data discovery and sharing.
 - He then examined some other well-known formats of metadata, the Dublin core, for example and discussed the advantages and disadvantages between the minimal approach versus a more substantial documentation.
 - The issues of time and space are easy to describe. Time and space have given, accepted methods of description. The challenge will lie in describing the spatial and temporal aspects of a variety of material, such as printed text.
 - Dr. Southall raised the question “Has GIS got in the way?” In this he discussed that many proprietary GIS software products limit the interaction of data between the different software companies, terming them as islands.
 - I have found that since my change of employment to ESRI, that by carrying the title of ERSI employee, regardless of my position in the company, I often become the light-hearted target of many comments and complaints about ESRI’s software or services as well as the GIS development community in general. Such was the case in both of these events.
 - Dr. Southall concluded his remarks by expressing his concept that historians can play a major role in the development and implementation of the DDI, data dissemination programs, and social science in general, expressing the need that historians have in communicating with government officials.

Personal thoughts and reflections – The development of the DDI:

The future of the DDI is, as any metadata standard, in the hands of the users. The success or failure of the proposed changes cannot be forced onto any party or individual. Only if those involved in developing the framework and the potential users of such a framework can come together to the necessary agreements, will the development and application of the DDI be successful.

Traditionally, metadata has been something produced after the research is over, and many of the studies involving social behaviors and ideas are not dependent of their spatial arrangement. These changes to the metadata schema will be asking the scientist questions that they have, perhaps, not asked themselves. Social scientists will need to be indoctrinated into the ideas involved with spatial patterns as well as the advantages of spatial tools and analysis. Further, they will need to broaden their understanding of the results of their particular study. If the results that they have come to; the answer to their problem, can be presented in a format that will permit others to relate the findings of a population to other fields of research, then we can begin to recognize new patterns, develop new ideas, and possibly solve problems facing the world today. One of the best methods of tying people, ideas and other fields of study together is through the spatial footprint of the wide range of locations.

This new frontier to social scientists is beginning to be explored and populated. Computational power has permitted a number of groups to begin to organize and document information that has historically been a tedious and an unreasonable practice. A large collection of historic social science material, such as the census, is being converted to a digital format. Many of the social scientists are seeing this development, and interest is rising in using this information to create other forms of socio-spatial information and develop research topics. Those currently involved in this research and those who will be incorporating spatial patterns into their research are in need of methods and technologies to collect, store and present socio-spatial information. Metadata, computational storage facilities and geographic information systems are the cure for these needs. The ICPSR is attempting to develop a metadata schema to describe such studies and results. Organizations such as CSISS are attempting to teach the social science community to think spatially; to become familiar with the tools and techniques involved in spatial research. ESRI, and presumably the other GIS software providers, are working to make spatial tools easier for those interested in spatial analysis to use GIS as a tool in their research. As we, in the spatially thinking community begin to reach out to those in the social science community we will begin to understand the methods and techniques shared by each other. We must make the efforts to reach out to one another and close this gap. This joining will broaden the understanding on both parts and will contribute to the development of further research and understanding in the interactions of people and space.