Poster #14

Qualitative study of data preservation practices at State Geological Surveys

by Sarah Ramdeen, School of Information & Library Science, University of North Carolina at Chapel Hill

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Abstract

In 2007 the USGS started the National Geological and Geophysical Data Preservation Program (NGGDPP). In the second phase of this program they awarded grants to state geological surveys to develop long range plans for data preservation and create metadata for their holdings to be uploaded in to the newly created national catalog for geological data sets. The USGS identified a need for the preservation of geological data and encouraged states to develop plans to sustain their data into the future. However the people tasked with creating these plans are not trained archivists. It is common for the role of archivist or curator of a geological collection to fall upon a professional geologist who has not been trained in the methods of access and preservation.

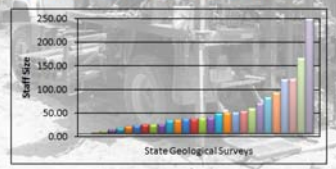
This research will address the questions of how these non archivists developed long range plans, what they did and why they chose these methods. I want to look at how they faced these issues, if they are aware of the existing archival knowledge and how they approached this task on their own. I will also look at their perception of success and perceptions of importance within their organizations - how they determined what materials to preserve and how they measure the success of their efforts. I am also interested in what motivates them to save the data they collect and how they share it outside of their organization.

Demographics

Base map © The Geological Society of America, Inc. Retrieved from http://www.geosociety.org/graphics/sections/005sections-map.pdf

Participants:

- 6 interviews were conducted with participants selected from the 29 State Geological Surveys taking part in the USGS funded NGGDPP program. It began as stratified sampling and further participants were gathered through the snowball sampling
- The State Surveys interviewed range in size from roughly 16-60 staff members. This represents the bottom 4 quintiles for those states taking part in the NGGDPP. (The 5th quintile are currently not represented by this study.)
- Participants represent 5 of the 6 geographical regions defined by the Geological Society of America (The Southeasten section is not represented by this study at this time.)
- 4 of the 6 participating States are attached to research institutions, the remaining are part of their State's government system



Typical collections at a Survey

- Rock core and cuttings
- Oil and Gas logs
- Sample strip logs
- Crude Oil samples
- Graphic records
- Geophysical logs
- Hand samples of rocks
- Survey field maps and notes
- Environmental impact statements for nuclear sites, dams, locks
- Reprint collections – both paper and physical records
- Geochemical data from limestones, coal beds, glacial tills, sands and gravels
- Side wall cores and core plugs
- Samples and results from sand and gravel investigations
- Photographs
- Records for cores and soft sediments
- Other paper records

Quotes from Participants

Significance of the collections

"We're not doing "Research for research – if you can't turn it in to a product for customers, if it's not utilized for mining, earthquake planning, environmental planning, why do the work? … You need public outreach and access to justify doing science otherwise it's just for the file cabinet."

"Many people come here and do not understand why we have a warehouse of rocks. But when you explain relevant issues, climate change, gas prices they get a better understanding"

Example of preserving for unintended nature – "CO2 sequestration, climate change. 30 years ago we never would have dreamt of it."

"We've been legislatively mandated by the state to conduct this kind of data gathering and data preservation."

What does preservation mean for you?

"When we started, preservation was not at the front of our mind; staff time and access was the priority."

"Basically two steps for preservation, preserve it and make it available to the public."

"My position is considered highly necessary… He (the State Geologist) regards this as a very important part of the mission of this survey."

Most important task – "To ensure a balance between responding to immediate issues and ensuring that systematic information is available for unanticipated requirements in the future."

Important issues with preservation

"We have put a lot of emphasis on inventorying the collections because people who knew what was in the collection have been retired for about 5-10 years…no one on the staff really had an idea what we had up here. So the first step is to get a really detailed inventory of as many collections as we can."

"I can tell you how many times I've received a collection where a lot of the data is on the old floppy discs, not just the little discs but the big discs or the Bernoulli ones. You don't have the equipment to actually pull the information off."

"About 15 major collections, if I counted again I would probably get more."

"Once you have a digital collection you must continue to upgrade as computer technology and operating systems come along"

What we need is to "bring it all together so that when you search for something … and find an outcrop, it will then also search all the other ancillary databases and find the information from all of the other databases that might have something to do with that one location."

Problems faced with collections

"Budgets are a constant battle, a lot of it is simply people do not realize how valuable the survey is and its products."

"When people retire and walk out the door they take a lot of information with them, not in the form of data just in the form of 'oh yeah there were these two guys...' They may be the only person who knows where that book or report is and why it was done."

"For the longest time I was the IT staff."

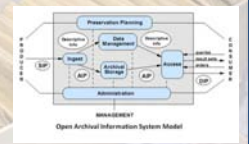
"We are currently at about 92% full which most states are in the same boats. 2/3 of the state repositories and maybe even federal repositories are 90% full or more."

Archival model for Digital Preservation

The tools created by Curators and Archivists may be of use to the State Geological Surveys while developing their long range plans.

In the OAIS model for digital preservation, an Archival Information Packages (AIP) is created for preservation and a Dissemination Information Package (DIP) is created for consumer access. The creation of these two different versions of a digital item is important in order to protect the information they contain.

The same process may be applied to physical samples. Percentages of cores or cuttings maybe retained as an archival record while the rest is available for the consumer to sample, study, test and even destroy in the process.



© CCSDS Retrieved from http://public.ccsds.org/publications/archive/650x0b1.pdf

Resources used for making decisions on preservation:

- Their own "on the job" knowledge and experiences
- Fellow staff members
- Counterparts at other Geological Surveys
- Course work–paleontology/museum related
- Websites – Smithsonian, NARA and others
- none

Discussion

A common theme among participants was the desire for the creation of a front end interface for their databases similar to Google Earth that would allow them to display various layers of the data in their collections that would also link between the map interface and full text versions of their publications. The idea being to link all of the data associated with a geographic location to a visual interface for retrieval based on a geographical range not just on keywords.

Future work

The next step for this project will be to continue to collect data through more interviews. I would like to use the knowledge gained here to develop tools and tutorials to help introduce archival methods that may be of use to this community. After preliminary results in talking to the participants they seem aware of the issues and in most cases the big picture solution to these problems. But they might not have the funding or the knowledge to physically complete the tasks.

For more information please contact Sarah Ramdeen at ramdeen@email.unc.edu.

Poster presented at the Archival Education and Research Institute July 9th 2009

Images courtesy of the Florida Geological Survey

Qualitative study of data preservation practices at State Geological Surveys

by Sarah Ramdeen, School of Information & Library Science, University of North Carolina at Chapel Hill

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Abstract

In 2007 the USGS started the National Geological and Geophysical Data Preservation Program (NGGDPP). In the second phase of this program they awarded grants to state geological surveys to develop long range plans for data preservation and create metadata for their holdings to be uploaded in to the newly created national catalog for geological data sets. The USGS identified a need for the preservation of geological data and encouraged states to develop plans to sustain their data into the future. However the people tasked with creating these plans are not trained archivists. It is common for the role of archivist or curator of a geological collection to fall upon a professional geologist who has not been trained in the methods of access and preservation.

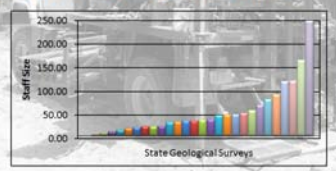
This research will address the questions of how these non archivists developed long range plans, what they did and why they chose these methods. I want to look at how they faced these issues, if they are aware of the existing archival knowledge and how they approached this task on their own. I will also look at their perception of success and perceptions of importance within their organizations - how they determined what materials to preserve and how they measure the success of their efforts. I am also interested in what motivates them to save the data they collect and how they share it outside of their organization.

Demographics

Base map © The Geological Society of America, Inc. Retrieved from http://www.geosociety.org/graphics/sections/605sections-map.pdf

Participants:

- 6 interviews were conducted with participants selected from the 29 State Geological Surveys taking part in the USGS funded NGGDPP program. It began as stratified sampling and further participants were gathered through the snowball sampling
- The State Surveys interviewed range in size from roughly 16-60 staff members. This represents the bottom 4 quintiles for those states taking part in the NGGDPP. (The 5th quintile are currently not represented by this study.)
- Participants represent 5 of the 6 geographical regions defined by the Geological Society of America (The Southeasten section is not represented by this study at this time.)
- 4 of the 6 participating States are attached to research institutions, the remaining are part of their State's government system



Typical collections at a Survey

- Rock core and cuttings
- Oil and Gas logs
- Sample strip logs
- Crude Oil samples
- Graphic records
- Geophysical logs
- Hand samples of rocks
- Survey field maps and notes
- Environmental impact statements for nuclear sites, dams, locks
- Reprint collections – both paper and physical records
- Geochemical data from limestones, coal beds, glacial tills, sands and gravels
- Side wall cores and core plugs
- Samples and results from sand and gravel investigations
- Photographs
- Records for cores and soft sediments
- Other paper records

Quotes from Participants

Significance of the collections

"We're not doing "Research for research – if you can't turn it in to a product for customers, if it's not utilized for mining, earthquake planning, environmental planning, why do the work? … You need public outreach and access to justify doing science otherwise it's just for the file cabinet."

"Many people come here and do not understand why we have a warehouse of rocks. But when you explain relevant issues, climate change, gas prices they get a better understanding"

Example of preserving for unintended nature – "CO2 sequestration, climate change. 30 years ago we never would have dreamt of it."

"We've been legislatively mandated by the state to conduct this kind of data gathering and data preservation."

What does preservation mean for you?

"When we started, preservation was not at the front of our mind; staff time and access was the priority."

"Basically two steps for preservation, preserve it and make it available to the public."

"My position is considered highly necessary… He (the State Geologist) regards this as a very important part of the mission of this survey."

Most important task – "To ensure a balance between responding to immediate issues and ensuring that systematic information is available for unanticipated requirements in the future."

Important issues with preservation

"We have put a lot of emphasis on inventorying the collections because people who knew what was in the collection have been retired for about 5-10 years…no one on the staff really had an idea what we had up here. So the first step is to get a really detailed inventory of as many collections as we can."

"I can tell you how many times I've received a collection where a lot of the data is on the old floppy discs, not just the little discs but the big discs or the Bernoulli ones. You don't have the equipment to actually pull the information off."

"About 15 major collections, if I counted again I would probably get more."

"Once you have a digital collection you must continue to upgrade as computer technology and operating systems come along"

What we need is to "bring it all together so that when you search for something … and find an outcrop, it will then also search all the other ancillary databases and find the information from all of the other databases that might have something to do with that one location."

Problems faced with collections

"Budgets are a constant battle, a lot of it is simply people do not realize how valuable the survey is and its products."

"When people retire and walk out the door they take a lot of information with them, not in the form of data just in the form of 'oh yeah there were these two guys...' They may be the only person who knows where that book or report is and why it was done."

"For the longest time I was the IT staff."

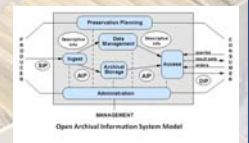
"We are currently at about 92% full which most states are in the same boats. 2/3 of the state repositories and maybe even federal repositories are 90% full or more."

Archival model for Digital Preservation

The tools created by Curators and Archivists may be of use to the State Geological Surveys while developing their long range plans.

In the OAIS model for digital preservation, an Archival Information Packages (AIP) is created for preservation and a Dissemination Information Package (DIP) is created for consumer access. The creation of these two different versions of a digital item is important in order to protect the information they contain.

The same process may be applied to physical samples. Percentages of cores or cuttings maybe retained as an archival record while the rest is available for the consumer to sample, study, test and even destroy in the process.



© CCSDS Retrieved from http://public.ccsds.org/publications/archive/650x0b1.pdf

Resources used for making decisions on preservation:

- Their own "on the job" knowledge and experiences
- Fellow staff members
- Counterparts at other Geological Surveys
- Course work–paleontology/museum related
- Websites – Smithsonian, NARA and others
- none

Discussion

A common theme among participants was the desire for the creation of a front end interface for their databases similar to Google Earth that would allow them to display various layers of the data in their collections that would also link between the map interface and full text versions of their publications. The idea being to link all of the data associated with a geographic location to a visual interface for retrieval based on a geographical range not just on keywords.

Future work

The next step for this project will be to continue to collect data through more interviews. I would like to use the knowledge gained here to develop tools and tutorials to help introduce archival methods that may be of use to this community. After preliminary results in talking to the participants they seem aware of the issues and in most cases the big picture solution to these problems. But they might not have the funding or the knowledge to physically complete the tasks.

For more information please contact Sarah Ramdeen at ramdeen@email.unc.edu.

Poster presented at the Geoscience Data Preservation Techniques Workshop July 2009

Images courtesy of the Florida Geological Survey