

What Do Geologists Need to Know about Metadata?

Abstract: Over the life of an average interpretation project, gigabytes of information are stored. A geologist may generate an average of 2 to 3 important documents each week resulting in hundreds of files created yearly. The problem is how find the critically important files needed for the next project. What is the solution? Metadata is searchable information about a data resource. For an example, look at how the government manages data, they had to get organized and created geospatial metadata standards.

What are the metadata elements and standards needed by geologists? The goal is populate the metadata, make the metadata searchable and to maintain metadata elements for ownership and data retention time. Properties of the datasets such as title, author and creation date are harvested and recorded in metadata documents automatically, but the most useful search items will be manually entered through a metadata editor program. Theme codes and keywords from the metadata standards can be added in a special metadata profile that allows entry of metadata elements useful for hydrocarbon exploration. Custom metadata editors, metadata templates and xml stylesheets are especially useful for populating multiple datasets within a project.

Metadata standards and formats of image metadata, geospatial metadata for ArcGIS and the "document properties" of Microsoft Office files and Adobe Acrobat pdfs differ; however, the basic principles of data ownership and responsibility apply. By practicing adherence to standards, metadata is the key to sharing the data needed for geologic interpretation projects.

Metadata is... searchable information about a data resource.

or..."Metadata is hidden information in a computer file that may contain potentially dangerous or embarrassing information or lead to an accidental disclosure."

http://blogs.adobe.com/acrolaw/2005/10/metadata_and_pd.html

Introduction to purpose of poster

I wanted to update metadata for some feature classes that I created to analyze the precision of IHS well locations in Iran. I was sending the files back to IHS and I thought I should think about entering some metadata to show them that I realized just whose data I was using. After investigating how to use metadata within ArcCatalog, I started to develop workflows for my projects. I then started to look at current implementations of metadata workflows and became frustrated with the lack of tools and usage of metadata in the earth science community.

Metadata is used to...

Catalog/Search/Determine Usability/Document

- Seismic lines (2D, 3D data, navigation, balance)
- Well logs and cross section lines
- Maps (structure, isopach, facies)
- Stratigraphic columns
- Tabular data (picks, core analysis, geochem)

For spatial datasets, provide enough information for users to work with those datasets in a GIS

- ▶ Who is the owner/author/editor?
- ▶ Where is the data located?
- ▶ What is the coordinate system?
- ▶ When was the data created, is this the current version?
- ▶ Why is this data important?

The *purpose* of a map is only thing that we can't measure or determine by some means. Only the geologist knows why he made the map, so no matter what metadata workflow we use, the geologist must be involved at some level to provide the project-purpose metadata paragraph.

Metadata philosophy: Taggers vs. Searchers

- The **taggers** believe in adding complete metadata so we can search
- The **searchers** believe in searching everything using powerful search tools

Neither viewpoint gets the job done...

We are getting overwhelmed by data! Many datasets have legal restrictions!

Topics for discussion

- How to edit metadata in ArcCatalog
- What are the key metadata elements?
- Auto-updating: When does it happen?
- How to auto-generate keywords
- Using .xml templates and enclosures

Remember...

- **Metadata can get lost when re-saving files**
- **Don't put all the metadata into a single document**
- **Don't put metadata into all of your documents**
- **Do use metadata .xml templates**

Metadata Completeness

- What is the type of data being documented? Is it a geodatabase, shapefile, coverage, folder?
- What is the purpose of the data? Is it for internal use only, will it be made public, will it be shared internationally?
- Who is the audience for this dataset? Specialists or generalists? How much do they need to know?
- What is your organization's policy regarding metadata completeness?

Metadata Workplan

A metadata workplan is essential as we need to work together to create a successful implementation

- **Create metadata**
- **Maintenance**
- **QC and Validation**
- **Sharing permissions**

Problem statement

What is the problem?

- Complicated metadata editors
- Generic style sheets
- Numerous metadata fields
- Poor synchronization
- Poor metadata persistence
- *Very few geologists are actively entering metadata!*

What is the desired state?

- Author and owner recognized
- Administration of data enhanced
- Legal restrictions complied with
- *Metadata entered for all of our important data files!*

What is my solution?

- Remember that AAPG Ethics requires citation and protection of others' datasets and intellectual property
- Auto-fill higher-level information by cascading directory path information from enclosing folders and project work orders
- Auto-fill lower-level metadata fields using data analysis
- *Improve synchronization between bibliographic and spatial metadata*

AAPG Code of Ethics

- Members shall not use or divulge any employer's or client's confidential information without their permission and shall avoid conflicts of interest that may arise from information gained during geological investigations.
- Members shall freely recognize the work done by others, avoid plagiarism, and avoid the acceptance of credit due others
- Members shall endeavor to cooperate with others in the profession and shall encourage the ethical dissemination of geological knowledge.

<http://www.aapg.org/business/codethic.cfm>

**Maintain the spirit of the standard
bibliographic citation for
*digital metadata!***

Citation examples

California Division of Mines and Geology, 1992c, Geologic Atlas of California: San Luis Obispo, compiled by Jennings, C.W., California Department of Conservation, Sacramento, CA, GAM015, 1:250,000.

Crowell, J.C., 1974, Origin of late Cenozoic basins in southern California, in Tectonics and sedimentation, Dickinson, W.R., ed., Society of Economic Paleontologists and Mineralogists Special Publication 22: Tulsa, Oklahoma, SEPM, p. 190-204.

Citation examples

Graham, S.A., Ingersoll, R.V. and Dickinson, W.R., 1976, Common provenance for lithic grains in Carboniferous sandstone from Ouachita Mountains and Black Warrior Basin: Journal of Sedimentary Petrology v. 46(3), p. 620-632.

Hall, C.A., Jr., 1973b, Geology of the Arroyo Grande 15' quadrangle, San Luis Obispo County, California, California Division of Mines and Geology Map Sheet 24: Sacramento, CA, California Division of Mines and Geology, 8 p.

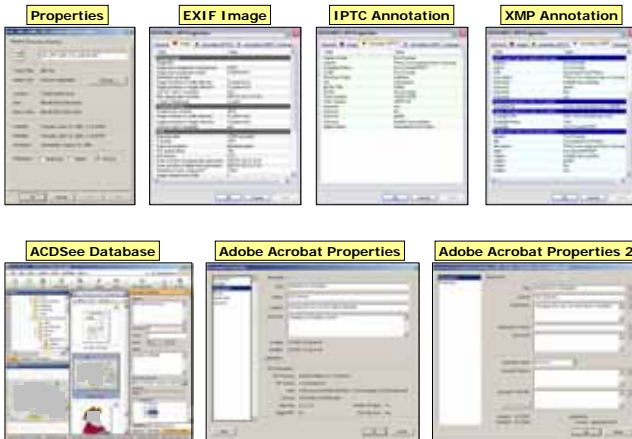
Metadata synchronization

Metadata creation during...

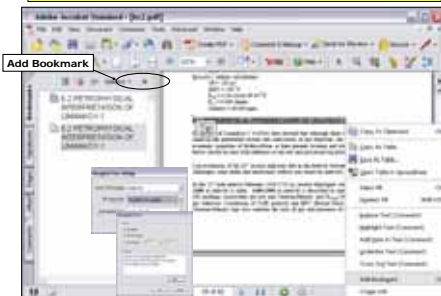
- Project approval
- Initial image generation
- Image processing
- Georeferencing
- Vector feature editing
- Project results report

These programs all have their own unique metadata!:

- Exchangeable Image File (EXIF)
- Intl. Press Telecomm Council (IPTC)
- Adobe Photoshop Document Properties
- ACDSee database
- ArcGIS (ArcCatalog) .xml
- Microsoft Windows NTSF File Properties
- Microsoft Office File Properties
- Adobe Acrobat Document Properties

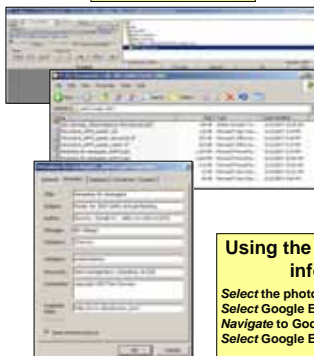


Adobe Acrobat Text Capture to Bookmarks and Keywords



- Create table of contents page text (type or OCR)
- highlight each line of text to be bookmarked
- right-click "Add Bookmark"
- go to each figure & click on the bookmark name & right-click "Re-link bookmark" to create a bookmark link useful for indexing large reports

Microsoft NTSF/Office



Picasa2 => Tools

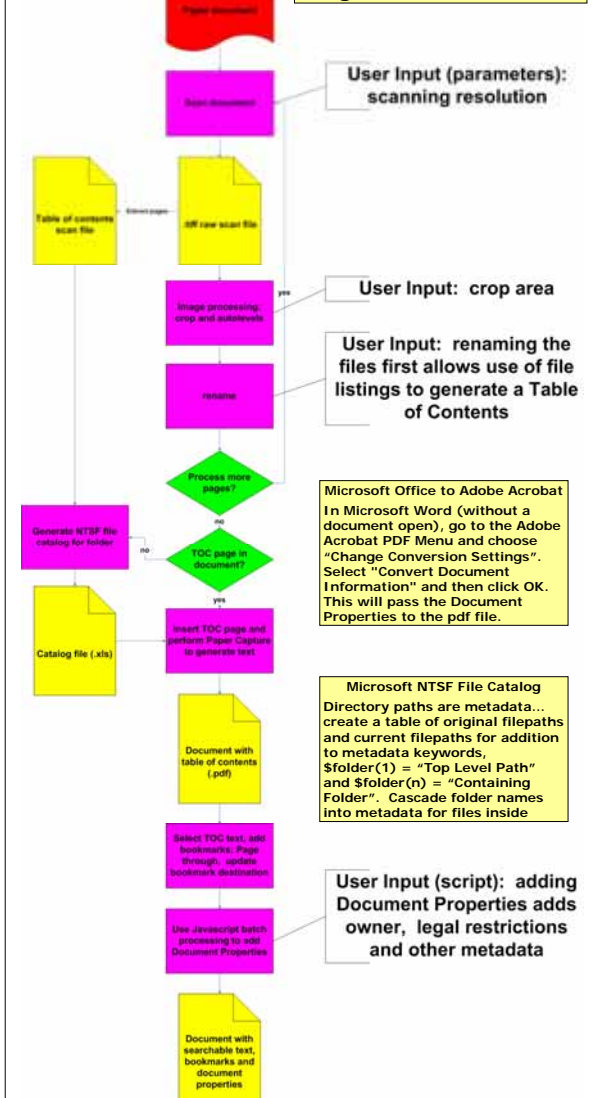


Using the "Geotag" feature of Picasa2 to add location coordinate information to EXIF GPS metadata for graphic files

Select the photo in Picasa2.
Select Google Earth Tools => Geotag => Geotag
Navigate to Google Earth photo location
Select Google Earth Geotag/Geotag All button, and "Image Properties" in Picasa2 to view metadata

<http://picasa.google.com/support/bin/answer.py?answer=43896>

Image metadata workflow



Metadata is fragile, hard to transport and is input by individuals with a variety of software tools, languages, and formats.

Metadata standards and formats of image metadata, geospatial metadata for ArcGIS and the "document properties" of Microsoft Office files and Adobe Acrobat pdfs differ; however, the basic principles of data ownership and responsibility apply. By practicing adherence to standards, metadata is the key to sharing the data needed for geologic interpretation projects.

Metadata for geoprocessing

Models are geoprocessing workflows

Model metadata for **casual users**

- Primary purpose is to search for data
- Does not need details of processing
- Create metadata *after* processing steps are done

Model metadata for **specialists**

- Primary purpose is for QC analysis
- Details of processing and results of analysis
- Need to create and edit *while* doing processing

Possible enclosures:
project overview
graphic index map
model history files

http://support.esri.com/KnowledgeBase/documentation/FAQs/sde/_webhelp802/ArcCatalog/Metadata_Support.htm

**Do geologists
need tools to
document the
geoprocessing
steps used in
model building?**

ArcCatalog Enclosure dialogs

Metadata properties dialog

Add New Enclosure dialog

Xml Tag:

Path: Browse...

Description:

☐ Image

OK Cancel

Enclosures allow you to embed almost any file within your metadata

Metadata Properties

Enclosures | Options

Add... Delete Save As...

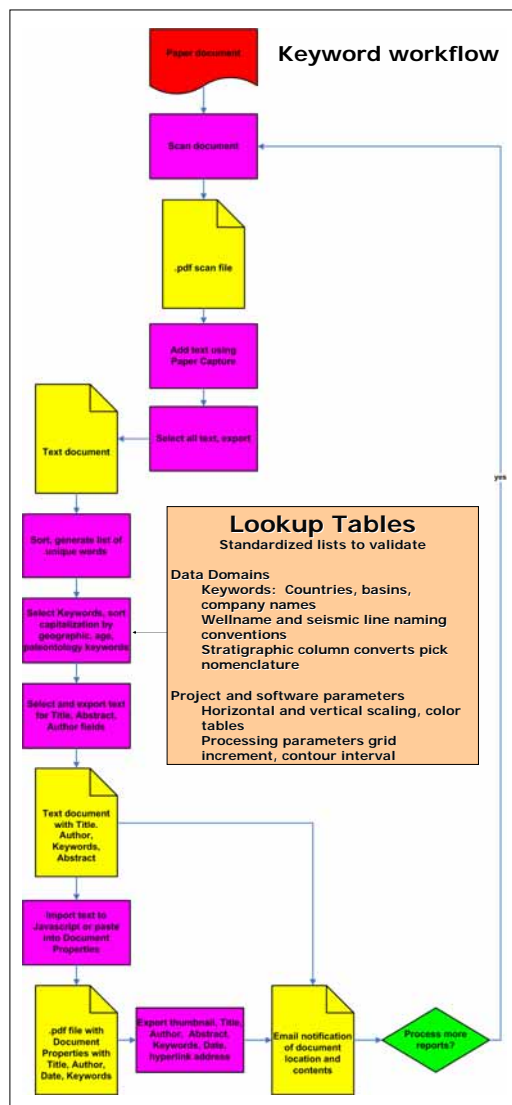
OK Cancel Apply

Auto-generating metadata

Digest existing metadata sources to create a database

- Capture text in:
 - project summary*
 - attribute tables
 - text in document
 - similar datasets
- Sort unique words
- Sort capitalization
 - placenames
 - stratigraphy
 - lithology
 - paleontology
- Select keywords
- Validation with look-up tables
- Intersection with spatial features

Metadata is data about data...
extract and generate it from the data!

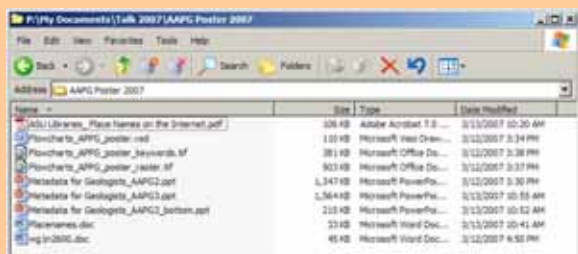


***Project Summary Workflow**

Project proposals (Work Orders) can serve as the basis for a metadata database containing the Project Name, Project Owner and Project Purpose fields. The Internal Charge Code field can serve a key field to interlink with project time-writing and with the data files generated for the project. Fill the project-level charge code information into the data files created during the project (auto-insert during creation or cascade from folder metadata) and then update Project Name, Project Owner and Project Purpose as needed.

File path workflow

Directory paths (folder names) *are* metadata. Create a table of original filepaths and current filepaths for addition to metadata keywords, i.e. \$filefolder(1) = "Top Level Path" and \$filefolder(n) = "Containing Folder". Then cascade the folder names into metadata keywords for each file inside each folder. All of these files should have "AAPG Poster 2007" in the metadata keywords.



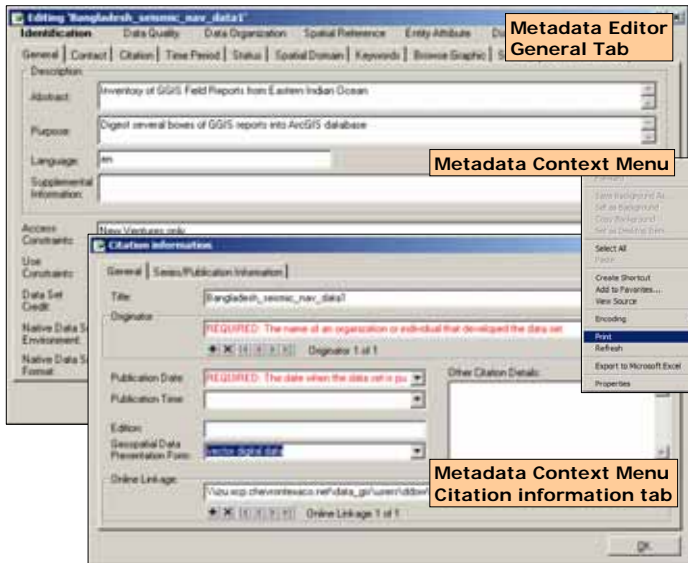
Template import and export

.xml templates are very useful for updating metadata. A better solution is to use a database to synchronize metadata between files.



ArcCatalog metadata editor

ArcCatalog provides an excellent GIS metadata editor, geared towards the GIS Analyst. Specialized (simplified) style sheets allow geologists to view only the needed and customized fields. Theme codes and keywords from the metadata standards can be added in a special metadata profile that allows entry of standardized metadata elements useful for hydrocarbon exploration.



ISO 19115 Core Metadata Elements

Mandatory Elements:

Dataset title
Dataset reference date
Dataset language
Dataset topic category
Abstract
Metadata point of contact
Metadata date stamp

Conditional Elements:

Dataset responsible party
Geographic location by coordinates
Dataset character set
Spatial resolution
Distribution format
Spatial representation type
Reference system
Lineage statement
On-line Resource
Metadata file identifier
Metadata standard name
Metadata standard version
Metadata language
Metadata character set

ESRI ArcGIS ArcCatalog Auto-update example from on-line training course

Why be concerned with the update setting for metadata properties? Because even though you can control whether to update properties when you create metadata, you cannot control what happens when someone else imports one of your templates. The person who imports your metadata may want to view the content of the template file before updating the metadata for their dataset.

Although some metadata properties, such as the title or the name of the metadata standard, are innocuous, other information may be sensitive. For example, the online linkage property can include the name of your computer or network server (the location where the data is stored). From a security standpoint, do you want those viewing the metadata to see this information?

The three tables below outline what happens to metadata properties when you export and import templates.

The first table shows the status of metadata properties when you export from two datasets. One is named With_Props and its metadata properties have been updated. The properties of the other dataset, No_Props, have not been updated.

Dataset Being Updated	Automatic Updates ON?	Click Metadata tab	Properties Updated?	Export to Template File	Template Contains the Dataset's Properties?
With_Props	Yes	Yes	Yes	Yes	Yes
No_Props	No	Yes	No	Yes	No; there are no properties.

In this table, you see how the update setting during import affects metadata properties when automatic updates are turned on.

Dataset Being Updated	Automatic Updates ON?	Import Template	Import Updates ON?	Properties Displayed When Viewing Metadata
Parcels	Yes	With_Props	Yes	Updated Parcels
Parcels	Yes	With_Props	No	With_Props
Parcels	Yes	No_Props	Yes	Updated Parcels
Parcels	Yes	No_Props	No	Updated Parcels

The table below illustrates how the update setting in the Import Metadata dialog box affects metadata properties when automatic updates are turned off.

Dataset Being Updated	Automatic Updates ON?	Import Template	Import Updates ON?	View Parcels Metadata—Which Properties?
Parcels	No	With_Props	Yes	Updated Parcels
Parcels	No	With_Props	No	With_Props
Parcels	No	No_Props	Yes	Updated Parcels
Parcels	No	No_Props	No	No properties

Regardless of the status of automatic updates when you export and import metadata, you can update metadata properties at any time—either manually (by clicking a button) or with automatic updates. Metadata documentation always updates.

ISO metadata standards

- ISO metadata has more "optional" elements that address deficiencies in the FGDC standard
- Developed by ISO Technical Committee with FGDC, European, Australian Data Agencies
- Extensible and allows metadata profiles customized for a particular user community
- Includes Data Dictionary which characterizes the dataset, including its intended use and limitations.
- Models - Unified Modeling Language (.uml)
- Exports - eXtensible Markup Language (.xml)

FGDC Metadata Standards

The Federal Geographic Data Committee (FGDC) is an interagency committee that promotes the coordinated development, use, sharing, and dissemination of geospatial data on a national basis. This nationwide data publishing effort is known as the National Spatial Data Infrastructure (NSDI). The NSDI is a physical, organizational, and virtual network designed to enable the development and sharing of this nation's digital geographic information resources. FGDC activities are administered through the FGDC Secretariat, hosted by the National Geospatial Programs Office (NGPO) of the U.S. Geological Survey. <http://www.fgdc.gov/>

Definitions from the ESRI Dictionary

XML

Extensible Markup Language. Developed by the World Wide Web Consortium (W3C), XML is a standard for designing text formats that facilitates the interchange of data between computer applications. XML is a set of rules for creating standard information formats using customized tags and sharing both the format and the data across applications.

style sheet

A file or form that provides style and layout information, such as margins, fonts, and alignment, for tagged content within an XML or HTML document. Style sheets are frequently used to simplify XML and HTML document design, since one style sheet may be applied to several documents. Transformational style sheets may also contain code to transform the structure of an XML document and write its content into another document.

Conclusions

What can we do?

- Utilize basic bibliographic metadata standards
(author, title, owner, legal disclaimers)
- Create simple editors and style sheets
- Improve synchronization and persistence
(templates, autofill, keywords)
- Create additional standards to cover specialities

*Provide a basic solution
(ownership, legal, technical)
which interfaces with the
types of geologic data and
software used in our offices!*

Workflow suggestion:

The plan must be **easier**
than the present metadata
editing workflow!

Metadata Workflow Strategies

Raw data versus interpreted data

We can also think about raw data versus interpretations. We may think that the most important data is interpreted data, but for new ventures hydrocarbon exploration, the most important data is the original raw data. Even excellent previous interpretations become data for the next round of interpretations because most new big discoveries come from interpretation of NEW PLAY CONCEPTS.

Geologists are not geographers

In first job with Gulf Oil before the Chevron merger, I indexed or created metadata. They would wheel in map racks full of maps and we counted how many maps we indexed. We were using metadata as an administrative tool. During the latest downturn, the well files personnel moved to IT positions and the earth scientists were to enter the metadata. Those of us who maintain metadata for geologic projects know that it hasn't worked out so well.

Geologist are creating data seismic lines and coring wells, but it difficult to enter all the metadata required in these projects when you work for a business. I have contacted other GIS analysts, even some who have presented metadata management solutions at the ESRI Users Forum and I found that metadata solutions that are not managed and that depend too much on casual user input are not successful. I believe we need 80-90% compliance to metadata standards across the industry to be successful in long-term management of our geologic data.

Earth scientist are not GIS experts, they may like to look at maps, they may author a lot of maps, but they are not GIS experts and the vast majority of geologists do not know how to use ArcGIS beyond displaying data. They ARE very intelligent, they do get excited when they see a pretty map, but they don't REALLY want to spend their time doing ArcGIS or they would have majored in GEOGRAPHY not GEOLOGY. ArcGIS is a tool that they may use to investigate a problem, but it is not the center of their work and side issues like data management and metadata editing are going to get short shrift. What we need is digital data librarians or GIS Analysts, they know the specialized nomenclature of the petroleum industry and they like to index and file things! Asking a geologist to be your GIS expert is cheaper in the short-term, but long-term it is inherently inefficient.

***The easier the metadata workflow, the more it will adopted.
I believe we will need a heavily-automated solution,
no matter how complex the programming is.***

**ESRI Petroleum User's Group
Metadata Working Group Wiki
<http://www.pug-steering.com>**

Contact Info

Donald W. Downey

CA RG #7487 AAPG CPG #5838 GISP #51955
Earth Science Technology Department (ESD)
Exploration and New Ventures Team
6001 Bollinger Canyon Road, Room C-1188
San Ramon, CA 94583-0946
Tel 925 842 3448 Fax 925 842 3030
<mailto:dondowney@chevron.com>