

Agile Data Curation at a State Geological Survey: A Use Case

Denise J. Hills, Geological Survey of Alabama; dhills@gsa.state.al.us

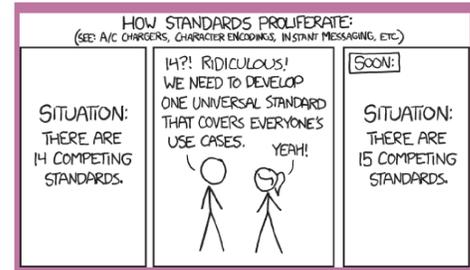


image: "Standards" <http://xkcd.com/927/>
Standards fail for a variety of reasons. Therefore, it's a good idea to use existing standards that have already been tested (possibly with modification) to ensure long-term usability.

Data Curation at GSA: Unintentionally Agile
*Four basic principles from <http://agilemanifesto.org>

Semiautomated workflows can help make it easier to capture information from current researchers with minimal effort. By simplifying the process, personnel are more likely to adopt and adhere to standards. Template workflows also streamline training of additional personnel to assist in data preservation, rescue, and registration.

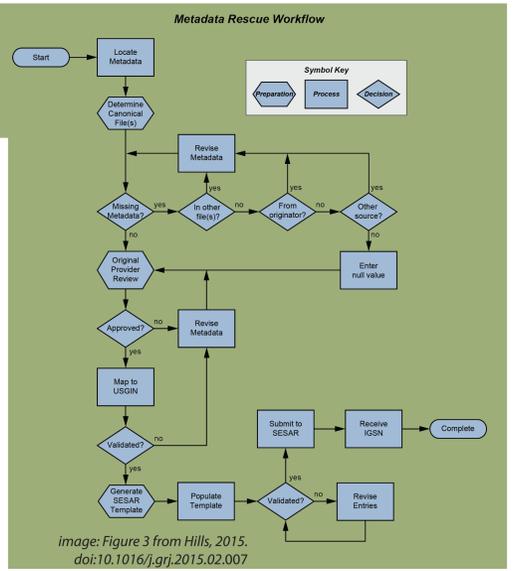


image: Figure 3 from Hills, 2015. doi:10.1016/j.grj.2015.02.007

*Individuals and interactions over processes and tools

*Working software over comprehensive documentation

*Customer collaboration over contract negotiation

*Responding to change over following a plan

Much data and metadata can be lost when people leave or retire, or even when someone simply forgets their own organizational system. Periodic data interviews as well as exit surveys can help capture some of that potentially dark data.



Geophysical well logs housed at the GSA and OGB has been scanned and are thus "digital" in one sense. However, limited metadata was captured at the time of the scan, and so the scans are not readily discoverable. Through other projects, some of the metadata has been captured on an as-needed basis.

Name	Date Modified	Size	Kind
120026.tif	Jul 20, 2005, 3:00 PM	2.7 MB	TIFF image
120027.tif	Jul 20, 2005, 3:04 PM	7.7 MB	TIFF image
120028.tif	Jul 20, 2005, 3:08 PM	4.1 MB	TIFF image
120029.tif	Jul 20, 2005, 3:17 PM	4.4 MB	TIFF image
120030.tif	Jul 20, 2005, 3:21 PM	2.6 MB	TIFF image
120031.tif	Jul 20, 2005, 3:24 PM	1.4 MB	TIFF image
120032.tif	Jul 20, 2005, 3:26 PM	1.4 MB	TIFF image
120033.tif	Jul 20, 2005, 3:27 PM	552 KB	TIFF image
120034.tif	Jul 20, 2005, 3:28 PM	702 KB	TIFF image
120035.tif	Apr 20, 2006, 4:00 PM	699 KB	TIFF image
120036.tif	Jul 20, 2005, 3:30 PM	698 KB	TIFF image
120037.tif	Jul 20, 2005, 3:31 PM	710 KB	TIFF image
120038.tif	Jul 20, 2005, 3:33 PM	1.1 MB	TIFF image
120039.tif	Jul 20, 2005, 3:18 PM	180 KB	TIFF image
120040.tif	Jul 21, 2005, 8:01 AM	17.5 MB	TIFF image
120041.tif	Jul 21, 2005, 8:05 AM	4.5 MB	TIFF image
120042.tif	Apr 21, 2006, 8:33 AM	3 MB	TIFF image
120043.tif	Jul 21, 2005, 8:12 AM	4 MB	TIFF image
120044.tif	Jul 21, 2005, 8:14 AM	1.9 MB	TIFF image

The GSA aspires to a fully comprehensive data management and curation scheme. However, to make any progress we have had to accept that "good enough" is better than nothing at all. This in turn makes access to our data easier and more productive, and allows us to make incremental progress.

PLAN	REALITY
<input checked="" type="checkbox"/> Assess collection status	<input checked="" type="checkbox"/> Assess collection status
<input checked="" type="checkbox"/> Evaluate needs	<input checked="" type="checkbox"/> Evaluate needs
<input checked="" type="checkbox"/> Develop management plan	<input type="checkbox"/> Develop management plan
<input type="checkbox"/> Perform maintenance as needed	<input checked="" type="checkbox"/> Perform maintenance as needed
<input type="checkbox"/> Update management plan as needed	<input type="checkbox"/> Update management plan as needed

Workflows

"Digital" Well Logs

Plan vs. Reality

Data Interviews