Documentation Cluster – February Telecon February 20, 2014

Attendees: Ted Habermann, Erin Robinson, Aleksandar Jelenak, Anna Milan, Kelly Monteleone, Ed Armstrong, Stephen Richard, John Graybal, Rich Signell

1. Strategic plan (<https://docs.google.com/document/d/10lKi88rReL0rF4WhZ8yEMabBoQTuALX-vu2VXBVokCo/edit?usp=sharing>)
	* Looks finalized
	* (Ed) should we move forward with the hack-a-thon
	* ACTION – Anna will submit a session for the summer meeting for the hack-a-thon
		+ Need a conceptual framework for the hack-a-thon (structure)
		+ ½ day (3 hr)?
		+ People would bring their own hardware & metadata
		+ Tie it into a particular spiral or goal for metadata
			- ISO 19157 – data quality group in ESIP… new and has interesting capabilities for data quality
	* Provide an example to work through – have people volunteer their data sets and then in teams work to document the quality and then reviewing together
	* Erin – this sounds more like “training”
	* Anna – might have more interest if more open than training
	* Ted – 3 or 4 data quality use cases (such as Ed’s example of a satellite granule, data and quality information for swath)
	* In 19157 includes a stand-alone quality report
	* Introduction to iso quality metadata and then bring your data quality examples for the second session (2 sessions)
		+ - GRIST granules has lots of good examples (Ed)
		+ Building off resources for the community and their data sets
	* ACTION – Kelly will put strategic plan on the Commons
		+ Question about the ESIP values in the strategic plan
		+ Erin – context of how fit into ESIP at large
2. "Drop It! Good Boy! (and Other Metadata Strategies)" by John Graybeal
	* This is about motivations and motivating good metadata into systems
		+ People’s motivations can be different – need to take steps to help people reach their goals to meet our goals
	* Start with reality of cost and benefits
		+ Cost - Time, money, privacy/secrecy, simple life, free will, independence
		+ Benefits – publicity, functionality, reusability, interoperability, understanding, social grace
		+ When you first enter metadata – costs are immediate (time, money)
			- The benefits are delayed. Maybe immediate publicity
		+ Costs are very concrete (measureable). The benefits are less concrete (they are abstractions)
		+ Balance is lop sided because of the costs
	* Change ahead – more integrated systems are coming online – trying to change the cost-balance equations
	* Mandate – ex like commanding your dog to drop something
		+ When mandate is strong and well enough enforced they people do it (like taxes)
		+ 2 examples
			- fgdc – they were bringing in csdgm and trying to get it adopted by all providers of environmental data sets – had mixed uptake
			- inspire – has had astonishing amount of uptake of metadata requirements
		+ not just about providing a mandate
	* Other motivations – social persuasions
		+ on marine explore (marinexplore.org) – uses points and achievements
		+ tangible rewards – ex. Being able to use the system
			- ex. LinkedIn – need to provide information to get at more features
			- in marineos (the private installation of marine explore) – you can see what you have provided
	* Recommendations
		+ Why do you care about making the metadata accessible
			- People will provide information because they have to or because it provides value to them
		+ How can I had more value into my system
			- Need…. more data, more data functions, more data relations
			- Thus need more better metadata
		+ Minimize work
			- Autoimport (& API), Only ask once, Autosuggest/complete (whenever possible… control vocabulary), Autoverify, Minimize keys/clicks, Visual guidance (make things clean and easy to follow), help and examples, prioritize fields (not always able to fill all of it), normalize/autolookup, templatize/autolookup
		+ Maximize reward
			- Show progress (feedback at every step, increase social rewards, overall competition), show reuse of metadata, show benefits on the fly (relations/groups of data, possible overlaps/duplicates, similar data sets/providers)
		+ Make the process fun!
	* Ed – any comment on prioritizing the mimize work and maximize rewards column
		+ 1) auto import 2) only fill in once
	* Anna – liked reward – help you understand your data better
		+ Find the researchers (for example) don’t completely understand the units that their data was collected in
	* Ted – talk about metadata in relation to “trust”
		+ Can talk about scales of trust
3. Metadata of the Month – Ted – Identifying Metadata records
	* Significant change – identifier for the metadata record – it was a character string (unstructured)
		+ At NGDC put namespaces into character strings
		+ It didn’t use MD\_Identifier class
		+ NOW – it is an Identifier … it is a code, an authority for the code, a code space, and a description
			- More robust
	* Concept of parent and child … in old it was a character string and is now a citation
		+ This can include a url for the metadata
	* Stephen – usually use GUI (globally unique identifier) or UUIDs – but they identify the record themselves and some use the same for metadata and data…
		+ John – if it is a NetCDF file with the data in the same file then it works
		+ Ted – think there are 2 pieces – 1 technical discussion of UUI, DOI, … and then the social benefits