May 2014 Documentation Telecon **[1]**

Submitted by Krbm on Thu, 2014-04-17 16:45   Thursday, May 15, 2014 - 14:00 to 15:00
Event: Telecon Minutes [2]
Session Type: Other [3]
Collaboration Area: Documentation [4]
Abstract/Agenda: 1) Summer meeting track
  2) Hack-a-thon prep
  3) Data Quality (Ted)

Notes:
Summer Meeting Sessions:

Google Doc:
https://docs.google.com/spreadsheet/ccc?key=0ArDAFB2BsbfRdEI1NGtUa2tpNFdlODlWUDJGNHZYeHc&usp=drive_web#gid=1 [5]

5 sessions have been identified as Documentation (in yellow). (Data stewardship are in pink)

1)   Steve Richard – Information exchanges and interoperability architecture (Wed 4 – 5:30)

2)   Katie Baynes – Streamlining Metadta using the CMR (ECHO and ESIP) (Thurs 9-1030)

3)   Barry Weiss – Design and Implementation of ISO Metadta in Science Data Products

4)   Ge Peng – Identifying and Assessing Best Practices in Data Quality

5)   Anna Milan - <MD_HackAThon>

4) 2 talk on quality a 1Hr discussion (Peng) – possibly include talks by Ted, Ed, and Alek)

3) Berry and Jeff Lee – OCDD

Ted is not sure about what will be included in 1 and 2

MD_HackAThon - Anna

·   See is as 2 things – a way to address problems with MD and how to apply ISO to special data sets

·   Do we want to recommend a tool set? If so oXygen (has 1 month free trail)

·   Peng – it is metadata or other – Archieve, Data Quality, seach/distribution

·   Ed – maybe take model from satellite mission and see hhow best ot encapsulate in ISO a single GRIST granule

·   Anna – have people volunteer problems before the meeting

·   Use 19157 and 19115-2 (not 19115 or 19115-1)

·   Peng – Data quality could use more guidance

·   Anna – should we split into 2 groups – how to use tool and problems or just pick 1

· Ted have about 10 things (lineage, acquisition, resources, etc) have have people vote on what they want to see (on a wiki page)

o Kelly – might want to use survey monkey or survey gismo

· Anna – people's data are special and want to help them – also ask about learning tools in the survey

· Anna – will other sessions cover data.gov – yup – Tuesday US Geo

· Ed – what is the difference between Data Stewardships' and Peng's data quality sessions

o Peng's focuses more on best practices in DQ – intro maturity matrix

o Data Stewardships' has 2 invited talks and a discussion

ISO 19157 – A framework for Progress on Data Quality

· Data quality moved from ISO 19115 (black) to 19157 (red) (colors are for Ted's presentation)

· It is currently an approved concept model – will be edited in June and then become an international technical specification

· Elements are abstract (triangle) means can be implemented in different ways

· Data Quality has elements

o Three parts – measure, evaluation method, results

o Has a standalone quality report

· Scope – defined range or elements for a quality set or result

o Can have quality for subsets

1) Done with scopecode – text term (has list of valid elements)

2) Extent – geographic, temporal, spatiotemporal, or vertical

3) Scope description – what attribute/feature quality applies to

· Standalone quality reports are new – link a citation to a report (possibly a journal article)

· Modular quality information

· Measure to database – each group builds database

· There are many types of reports – did not change from 19115

o Data quality usability (191152) – usability implied elements of data quality

· Data Quality Evaluation Methods

o Direct internal = use data in data to evaluate

o   Direct external – between dataset

o   3 types of results (include time and scape for each result)

1)   Quantitative

2)   Conformance

3)   Converge

4)   Descriptive (NEW)

·   Metaquality – quality of quality information

o   Way each was applied – homogeneity, confidence, representatively

·   Peng – what happens to issues of data quality identified after the data is released

o   Reports include dat and time – can querry reports for specific times/data and elements

o   Ex. SEDAC (socioeconomic data and application center) – has a feedback system for their data

·   Peng – challenge is who will implement

·   Anna – what about content, quality, and lineage –these are often overlooked

Participants:
Ajay Krishnan, Ed Armstrong, Ge Peng, Kelly Monteleone, Ted Habermann, Erin Robinson, Anna Milan, J. Wei, Robert Casey, Suhung Shen, Feng Ding

Creative Common License: Creative Commons Attribution 3.0 License
Accepted:

  Source URL: https://commons.esipfed.org/node/2394

Links
[1] https://commons.esipfed.org/node/2394
[2] https://commons.esipfed.org/taxonomy/term/284
[3] https://commons.esipfed.org/taxonomy/term/321
[4] https://commons.esipfed.org/collaboration-area/documentation
[5] https://docs.google.com/spreadsheet/ccc?key=0ArDAFB2BsbfRdEI1NGtUa2tpNFdlODlWUDJGNHZY eHc&usp=drive_web#gid=1