Deconstructing the Data Lifecycle BoF

Karl Benedict - University Libraries, UNM
Chris Lenhardt - RENCI
Mark Parsons - RDA
Joshua Young - UCAR/Unidata

March 10, 2015

Contents

Agenda - Context

Introduction

- Background regarding the genesis of this effort, work to date.
- Why Agile?

Presentations on Agile Curation Case Studies and Examples of the Technical Debt Model

- Provide a template for both documenting an A.C. case study and a T.D.M. example (the T.D.M example may contribute to an Agile Curation case study but not all T.D.M. examples will constitute a case study) - http://goo.gl/P7wuFY

Examples

- GSToRE and RDS at UNM
- Sea Ice as an Agile Curation example
- Others???

Agenda - Where from Here?

Discussion / Next Steps

- Identify more case studies
- Examples of agile curation
- Data reuse examples with an explicit consideration of investments required to facilitate data reuse
- How far can the agile model go when applied to data curation?

BoF Results

- Discuss the potential for a white paper or publication
- Outline paper / writing assignments / timetable
- Does the group want to continue as an RDA IG.

Background

- Conversations started at GeoData 2014 and continued at the ESIP Federation Summer 2014 meeting

- Poster presented at the 2014 AGU Fall meeting to broaden the dialog - http://www.slideshare.net/JoshYoung8/ag-cu-11

Taking Another Look at the Data Management Life Cycle: Deconstruction, Agile, and Community

Why Agile?

Key Ideas Presented in the AGU Poster

Research data management and curation is

- resource constrained
- dynamic
- may require expertise/knowledge beyond the research team that is creating the data

While agile development includes methods and practices, it also embodies a fundamentally *different philosophical approach* as embodied in the *Manifesto for Agile Software Development*

Manifesto for Agile Software Development

> We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:

Individuals and interactions over processes and tools

Working software over comprehensive documentation

Customer collaboration over contract negotiation

Responding to change over following a plan

That is, while there is value in the items on the right, we value the items on the left more. http://agilemanifesto.org/.

Assumptions for Agile Curation

1. Access to data is the first goal
2. Generative value is supported (Zittrain, 2006)
3. Researcher involvement through a participatory framework that aligns data management with scientific research processes (Yarmey and Baker, 2013)
4. Projects will utilize free open-source resources to the greatest extent practical
5. Community participation increases project capacity
6. Data management requirements and practices evolve as the research project proceeds
7. Bright and dedicated individuals can learn appropriate skills and respond to the demands of their particular project, as they proceed
8. Approaches apply across scales
9. Consider technical debt when making data management and curation decisions
10. Data evaluation can be conducted through use and feedback

- based on the Agile Underlying Assumptions found in Turke, et al (2002)

Technical Debt

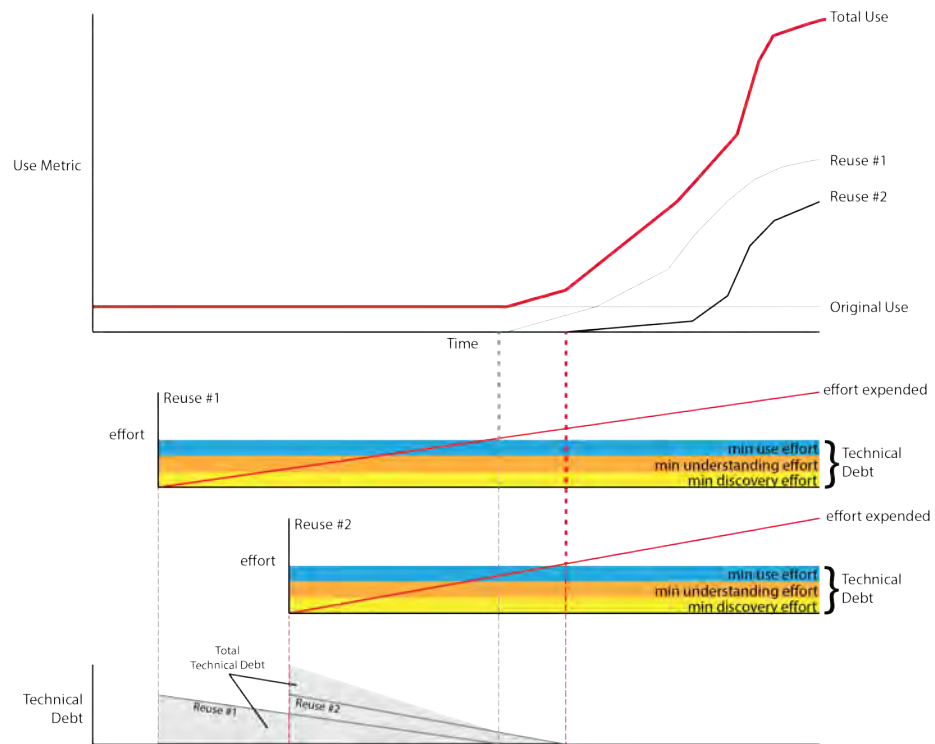The evolving technical debt conceptual model can be found at: https://github.com/karlbenedict/agilecuration

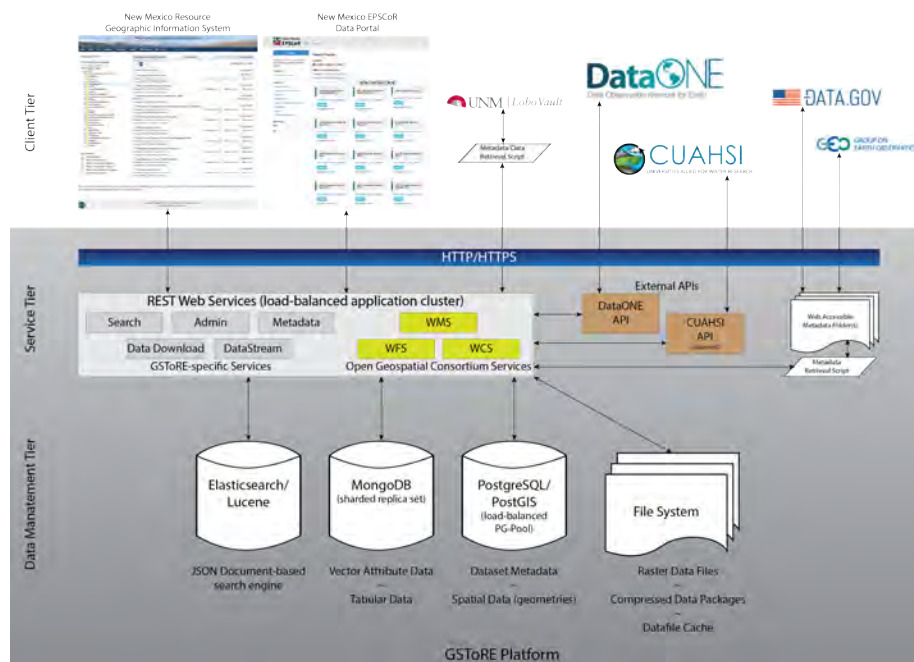Figure 1: Technical Debt Illustration

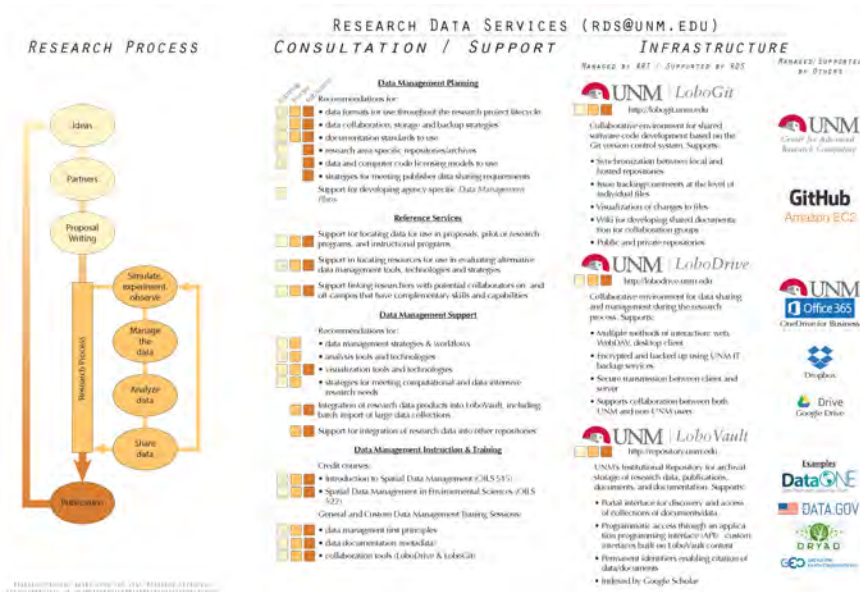Figure 2: GSToRE Component Diagram - http://gstore.unm.edu

Figure 3: RDS Service Catalog - rds@unm.edu

Examples

Geographic Storage, Transformation and Retrieval Engine (GSToRE)

Research Data Services at UNM

NSIDC - Ruth Duerr

Where from Here?

Discussion / Next Steps

- Identify more case studies
- Examples of agile curation
- Data reuse examples with an explicit consideration of investments required to facilitate data reuse
- How far can the agile model go when applied to data curation?

BoF Results

- Discuss the potential for a white paper or publication
- Outline paper / writing assignments / timetable
- Does the group want to continue as an RDA IG.

Key Links

- Case Study Submission Form - http://goo.gl/P7wuFY
- Git Repository for evolving collection of agile curation documents and other materials - https://github.com/karlbenedict/agilecuration
- Deconstructing the Data Life Cycle – Agile Curation BOF - https://www.rd-alliance.org/node/47138

References Cited

Anderson, D. J., (2003) Agile management for software engineering: Applying the theory of constraints for business results. Prentice Hall Professional

CEOS.WGISS.DISG. "Data Life Cycle Models and Concepts – Version 1". TNO1, (2011), Issue 1. http://wgiss.ceos.org/dsig/whitepapers/Data%20Lifecycle%20Models%20and%20Concepts%20v8.docx

Heidorn, P. B., (2008), Shedding light on the Dark Data in the Long Tail of Science, Library Trends, 57, 2, 280- 299

Paulo Sérgio Medeiros dos Santos, Amanda Varella, Cristine Ribeiro Dantas, and Daniel Beltrão Borges. "Visualizing and Managing Technical Debt in Agile Development: An Experience Report". H. Baumeister and B. Weber (Eds.): XP 2013, LNBIP 149, pp. 121–134

Rusbridge, C., Burnhill, P., Ross, S., Buneman, P,. Giaretta, D., and Atkinson, M. (2005) The Digital Curation Center: A vision for digital curation. In Proceedings to Global Data Interoperability-Challenges and Technologies, 2005. Mass Storage and Systems Technology Committee of the IEEE Computer Society, June 20- 24, 2005, Sardinia, Italy, Retrieved November 13, 2014 from http://eprints.erpanet.org/82/

Turke, D., France, R., and Rumpe,B. (2002), Limitations of agile software processes., Third International Conference on eXtreme Programming and Agile Processes in Software Engineering, Cambridge University Press

Yarmey, L. and Baker, K.S. (2013) Towards Standardization: A Participatory Framework for Scientific Standard- Making, International Journal of Digital Curation, 8,1, 157-172 Zittrain, J., (2006) The Generative Internet, 119 Harvard Law Review 1974 Published Version doi:10.1145/1435417.1435426 Accessed December 3, 2014 1:47:07 PM EST Citable Link http://nrs.harvard.edu/urn-3:HUL.InstRepos:9385626

Acknowledgements