Distilling Design Patterns From Agile Curation Case Studies (IN41A-1656)

Karl Benedict[1] W. Christopher Lenhardt[2] Joshua Young[3]

[1]University of New Mexico [2]Renaissance Computing Institute [3]University Corporation for Atmospheric Research

Abstract

In previous work the authors have argued that there is a need to take a new look at the data management lifecycle. Our core argument is that the data management lifecycle needs to be in essence deconstructed and rebuilt. As part of this process we also argue that much can be gained from applying ideas, concepts, and principles from agile software development methods. To be sure we are not arguing for a rote application of these agile software approaches, however, given various trends related to data and technology, it is imperative to update our thinking about how to approach the data management lifecycle, recognize differing project scales, corresponding variations in structure, and alternative models for solving the problems of scientific data curation. In this paper we will describe what we term agile curation design patterns, borrowing the concept of design patterns from the software world and we will present some initial thoughts on agile curation design patterns as informed by a sample of data curation case studies solicited from participants in agile data curation meeting sessions conducted in 2015-16.

Introduction

The challenges that must be addressed by current research data management and curation processes and strategies consist of a combination of established practices that are not compatible with increasing complexity in the data management landscape at the project level; increasing expectations by sponsors, publishers, and institutions relating to data management and curation; and rapid growth in the volume, variety and velocity (three dimensions commonly used to define "big data") of data generated by and used in research. In combination these challenges translate into an increasing need to develop effective data management and curation strategies that align with a set of *shared values and principles* that inform management and curation objectives, and implement processes that are *well documented and portable* across specific data management projects. It is this latter requirement that is addressed in this poster - the development of a framework for capturing elements of successful data curation activities and generalizing those elements into linkages with existing design patterns, or defining new design patterns when they don't exist.

Work to Date

Thus far the focus of the project's work has been on developing a framework within which the team can discuss the concept of *agile data curation* with the community, and iteratively evolving that framework through a series of meeting sessions, workshops and presentations that have been given at multiple venues including AGU (2014, 2015), ESIP Federation Meeting (2016), Research Data Alliance (2014, 2015, 2016), and SciDataCon (2016). In these various activities the team has worked on communicating the conceptual framework for our vision of agile data curation, presented a variety of initial values and principles derived from those defined in the *Manifesto for Agile Software Development* (*1*), and solicited the presentation of data management projects that exemplify (either intentionally or unintentionally) these principles.

Conceptual Model for Agile Data Curation Design Patterns

While this outreach and community engagement work described above is continuing, the work presented here is the starting point for our third goal of adapting the concept of design patterns that had been developed for object oriented software development (*2*), and extended into related domains (*3–7*), for use in documenting *named* data curation *problems*, *solutions*, and *consequences* that provide *descriptions of generalized data components that are customized to solve a general design problem in a particular context* (adapted from (*2*)).
The conceptual model that the research team has developed for mapping research data curation functional requirements into design patterns represents a combination of specific research activities that have data-related components (as exemplified in Figure 1) and linkages between those components as envisioned by a model such as the *Open Archival Information System* (OAIS - (*8–10*) - Figure 2). In particular, the research team is currently developing a model for collecting and synthesizing data curation case studies that can be used as exemplars for identifying existing design patterns or developing new ones that are relevant in data curation.

Illustration of the Design Pattern Conceptual Model to a Developed Data Management, Discovery and Access Platform - GSToRE

The Geographic Storage, Transformation and Retrieval Engine (GSToRE (*11*)) is a data management, discovery and access platform that was developed by the Earth Data Analysis Center as a tiered services oriented architecture (SOA - Figure 3) that was developed to meet the needs of multiple research data discovery, access and interaction models. While the GSToRE platform was not explicitly developed with any specific design patterns, there are a number of design patterns that are roughly represented within its design. The capabilities developed in GSToRE and related research and data managements steps are linked to each other in the diagram provided in Figure 3.

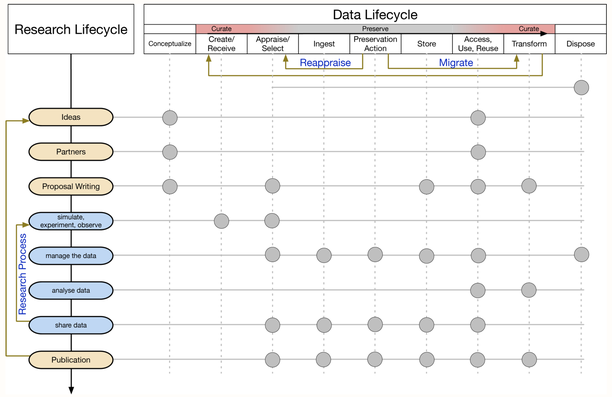

Figure 1: Intersection of Research Lifecycle (*12*) and Data Curation Lifecycle Actions (*13*) illustrating high-level research activities that involve data-related functions.

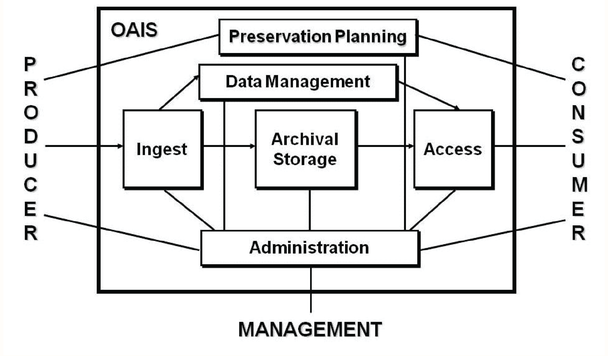

Figure 2: OAIS Reference Model (*8–10*) as a high level interaction model between functional components of a preservation system

Conclusions

While the research team is in the early stages of the identification of data curation design patterns as part of a larger program of defining an *agile data curation* approach to research data management and preservation, the work presented here represents our first attempt to map an existing data management platform into a set of design patterns that were developed in support of object oriented software development. This mapping provides a frame of reference for engaging with the research data management community in developing a community-based process for documenting design patterns that have proven effective in multiple contexts and implementations.

Acknowledgements

This work has been partially supported through funding from the National Science Foundation (Award no. IIA-1301346)

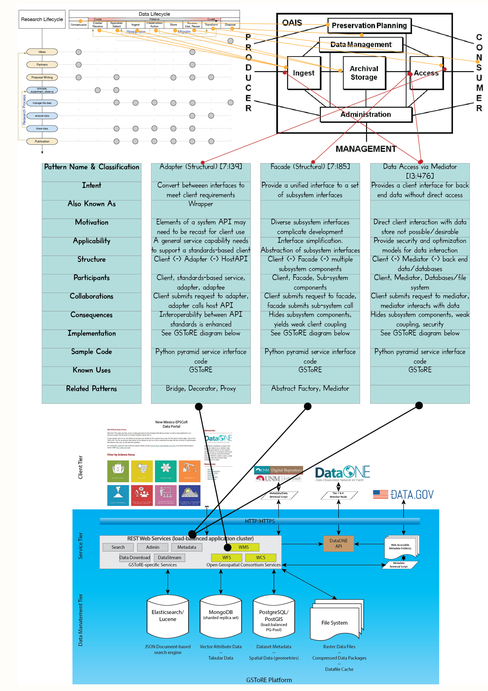

Figure 3: Mapping of the GSToRE Platform's Capabilities into a Set of Design Patterns and corresponding linkages between the OAIS Framework and the Research Lifecycle

Bibliography

1. K. Beck *et al.*, Manifesto for Agile Software Development (2001).
2. E. Gamma, R. Helm, R. Johnson, J. Vlissides, *Design patterns: Elements of reusable object-oriented software* (Addison-Wesley, Reading, Mass., 1995), *Addison-wesley professional computing series; addison-wesley professional computing series.*
3. R. Daigneau, *Service Design Patterns: Fundamental Design Solutions for SOAP/WSDL and RESTful Web Services* (Addison-Wesley Professional, 2011).
4. C. G. Lasater, *Design Patterns* (Jones & Bartlett Learning, 2010).
5. L. Ackerman, C. Gonzalez, *Patterns-Based Engineering: Successfully Delivering Solutions via Patterns* (Addison-Wesley Professional, 2010).
6. A. Schwinn, J. Schelp, Design patterns for data integration. *Journal of Enterprise Information Management*. **18**, 471–482 (2005).
7. G. Hohpe, B. Woolf, *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions* (Addison-Wesley Professional, 2003).
8. Consultative Committee for Space Data Systems (CCSDS), "Reference Model for an Open Archival Information System (OAIS)" (CCSDS 650.0-M-2, Consultative Committee for Space Data Systems (CCSDS), 2012).
9. International Organization for Standardization (ISO), ISO 14721:2012 - Space data and information transfer systems – Open archival information system (OAIS) – Reference model. *ISO* (2012).
10. B. Lavoie, "The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition)" (Digital Preservation Coalition, 2014).
11. Earth Data Analysis Center (EDAC), GStore V3 API (2016).
12. JISC, How Jisc is helping researchers : Jisc (2014).
13. Digital Curation Centre (DCC), DCC Curation Lifecycle Model | Digital Curation Centre (nd).