Conference paper

Agile Data Curation as a Diversity of Practices Grounded in Shared Values and Principles

# Summary

Current research data practices typically fall along a continuum between highly engineered systems and ad-hoc practices. In recognition of the increasing investment in and importance of research data as an asset for doing research, evaluating current research results, and as a resource for new research, funding agencies, publishers and some research teams have instituted research data management practices aligned with a wide variety of *data life cycle* models that embody a circular process of steps that include such activities as *creation, assessment, documentation, use, preservation, discovery* and *reuse*. While these data lifecycle approaches are well aligned with the documentation and preservation of research data - particularly as they have been primarily developed by organizations with a mandate to provide for the preservation of data, this linear (or more appropriately cyclical) model does not necessarily focus on the level of effort required throughout the processes embodied in the lifecycle. The *agile data curation* conceptual model outlined in this presentation provides a draft set of *values* and *principles* that are inspired by the *Manifesto for Agile Software Development[[1]](#footnote-1)* and the associated set of principles[[2]](#footnote-2) that were developed in alignment with the values defined therein. The research team is seeking feedback from the community on the draft values and principles that will be presented, case studies that exemplify them, and the design pattern strategy that has been developed accelerate the adoption of practices that are aligned with the values and principles defined.

# Overview

The concept of *agile data curation* has the potential to provide an organizing framework for research teams to develop and adopt a set of research data management, documentation, sharing and preservation practices that meet the combined needs of efficient data management, analysis and sharing within a project and enabling effective reuse of developed research data beyond the life of a specific project. Ideally these practices are informed by and based up a core set of principles that have been adopted by the research data management community that focus on the discovery and delivery of useful data to users both within and outside of the research project that created them, enable unanticipated uses of research data by new communities of users, embody design principles that encourage technical excellence, enable machine-to-machine interaction with data assets. A draft set of principles have been developed for consideration by the community as a collection of foundational principles upon which an *agile data curation* community of practice may be developed. These principles are the following (structured after the Agile Software Development principles referenced in the Summary above):

* Maximize the impact of research data through accelerated capacity for discovery, access and use of valuable data
* Expect unanticipated needs for and uses of research data (and documentation) and develop flexible systems to support new uses and users without significant modifications
* Facilitate automated interaction with data and metadata assets through well documented public web services that enable disintermediated use and reuse of research data
* Data creators and data curators should work closely throughout planning, research and preservation activities to ensure the most efficient and streamlined process
* Identify key individuals in a data curation project that have the requisite knowledge and motivation to do the job and get out of their way
* Identify the most effective method(s) for maintaining close communication and *use* them
* Delivery, access, use and citation of research data are the primary measures of success
* Design principles that enable steady delivery of incremental improvements to research data discovery, access and use should be consistent with a sustainable level of effort and funding from sponsors, data creators and curators, and users
* Continuous attention to technical excellence and good design enhances agility
* Start with the basics and only make systems more complex as needed, while maintaining a low bar to entry
* Continuously work to develop and evolve a community of data providers, curators and users that all participate in the ongoing evolution of the research data systems that they interact with

This paper will outline these principles and associated concepts in an effort to elicit feedback from the session participants relating to these principles (and others that may be introduced during the discussion), provide a conceptual foundation for identifying and discussing exemplars of these principles through existing research data management programs, outline a strategy for accelerating adoption of these principles through the development of agile data management design patterns that may be used by practitioners in the planning and execution of their research projects.

1. http://www.agilemanifesto.org/ [↑](#footnote-ref-1)
2. http://www.agilemanifesto.org/principles.html [↑](#footnote-ref-2)