



Karl Benedict¹ W. Christopher Lenhardt² Joshua Young³
¹University of New Mexico ²Renaissance Computing Institute ³University 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.

Process: Values -> Practice -> 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)).

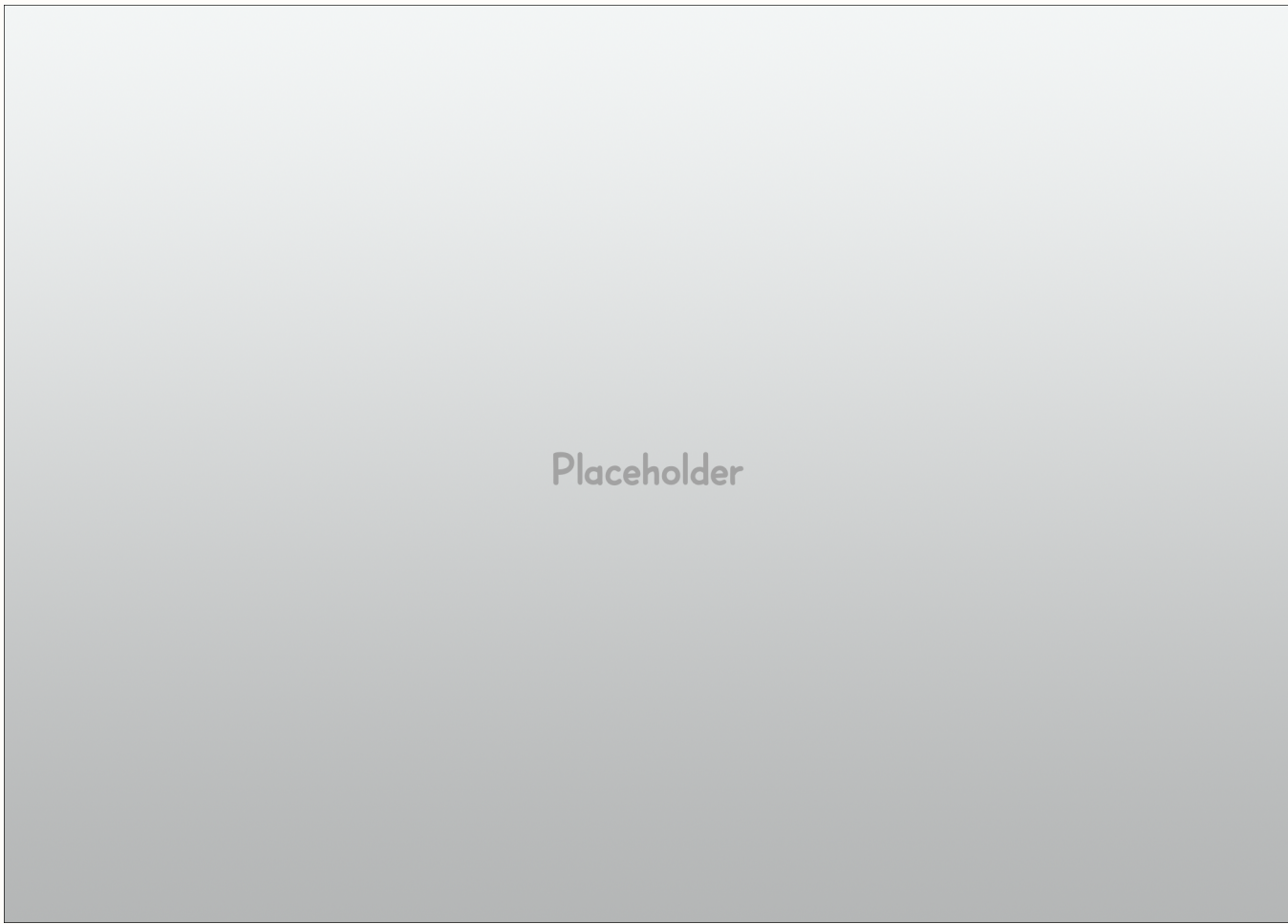


Figure 1: Information flow into developed design patterns.

Conceptual Model for Agile Data Curation Design Patterns

Vivamus efficitur eros et luctus porttitor. Aenean non urna semper, sollicitudin odio sed, ullamcorper augue. Ut malesuada lorem tortor, a posuere urna tempus at. Suspendisse a lorem odio. Integer non metus eu lacus maximus malesuada vel sit amet purus. Maecenas lacinia nisl in justo pretium cursus sed sed mauris. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. (2)



Figure 2: Proposed *Agile Data Curation* design pattern elements.

Ut blandit nisl est, sit amet tempus odio pellentesque vitae. Sed auctor ornare diam, sit amet vehicula massa tristique in. Morbi mollis elit risus, id pretium dui gravida varius. Curabitur quis tristique odio. Vivamus nec erat non turpis faucibus luctus. Nunc a ante vitae massa commodo semper sed vel dui. Nulla accumsan odio et diam vestibulum viverra. Curabitur sed lorem eget velit feugiat cursus at cursus mauris. Nam dapibus nisl non quam maximus gravida. Integer semper cursus urna id ultricies. Cras imperdiet enim quis augue maximus, eu condimentum diam facilisis. Etiam et massa sodales elit sollicitudin auctor quis vitae purus. Maecenas vitae lacus tortor. Vestibulum rhoncus congue ullamcorper. Donec in aliquam nibh, ac suscipit elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

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

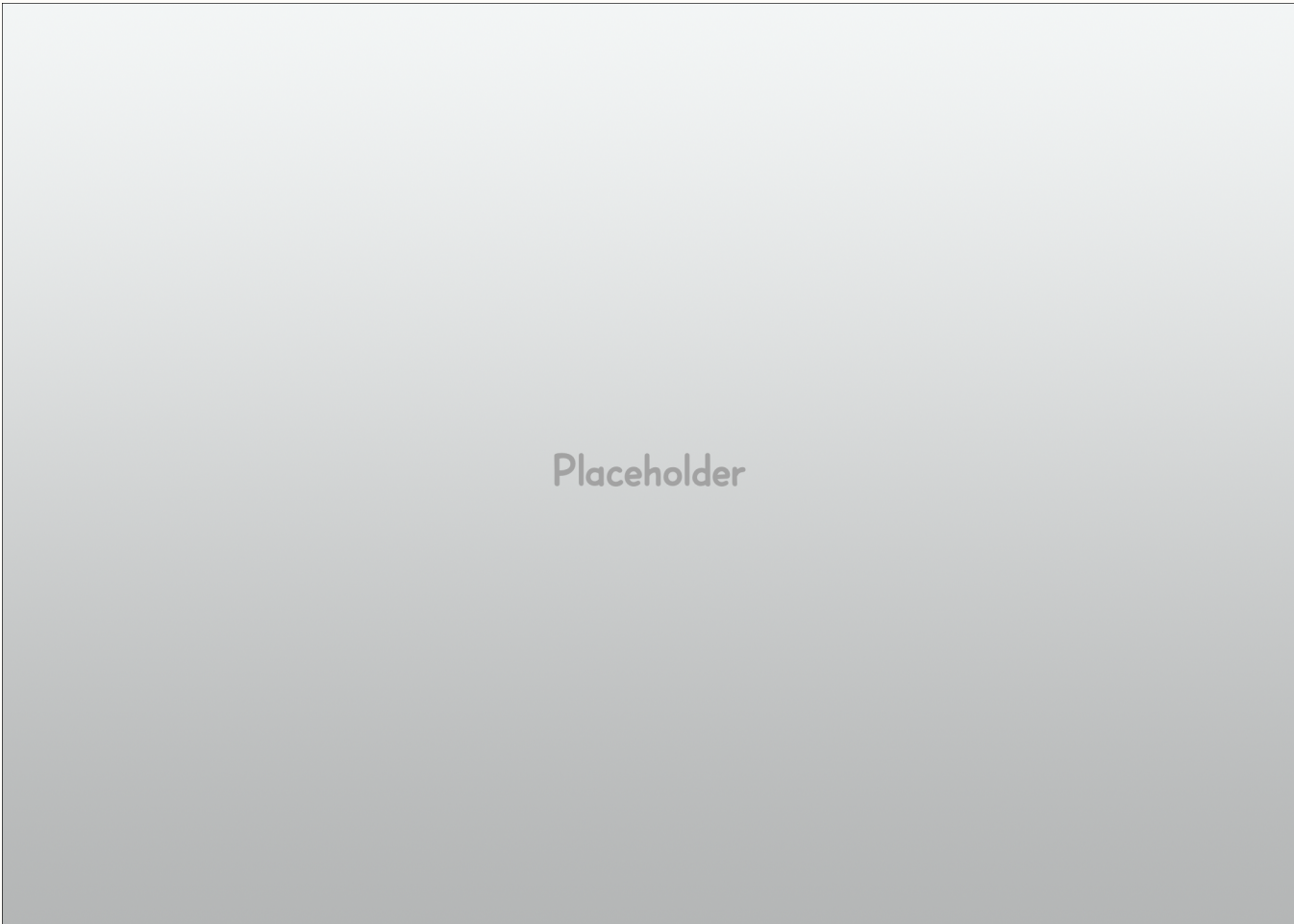


Figure 3: The Geographic Storage, Transformation and Retrieval Engine (GSToRE) Platform .

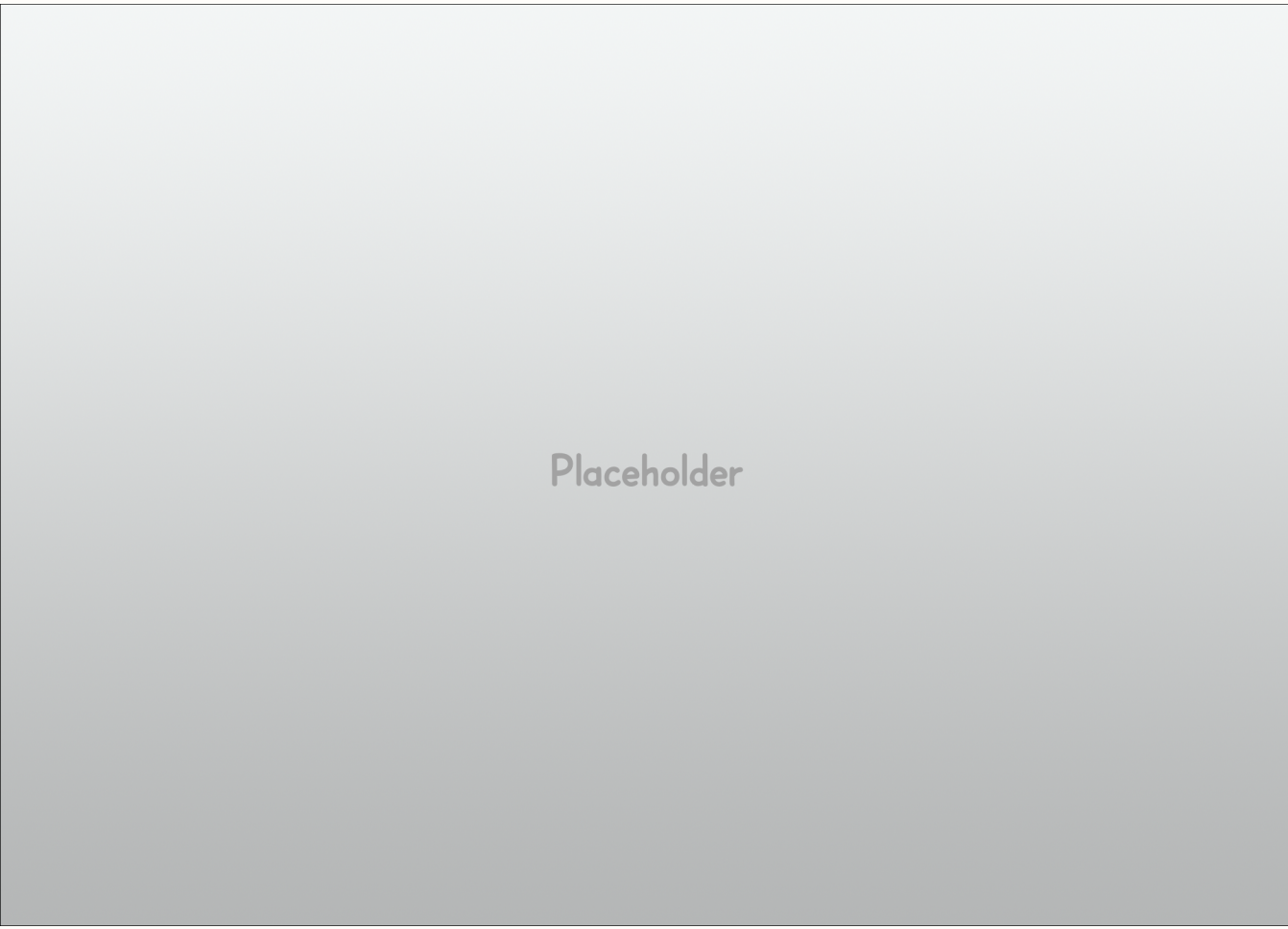


Figure 4: Mapping of the GSToRE Platform's Capabilities into a Set of Design Patterns.

Duis mauris urna, vulputate ac ex vitae, rhoncus lobortis nisi. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; In dignissim purus lacus, ut ultrices eros varius et. Mauris dictum cursus diam, vitae pharetra felis tincidunt pharetra. Suspendisse sem enim, lacinia eget consectetur ut, tempus vel erat. Proin vitae enim sit amet urna elementum rutrum. Nullam consequat eros sit amet est vestibulum rhoncus. Donec ullamcorper tempor finibus. Proin vestibulum nulla ut metus pretium ultricies. Vestibulum diam enim, laoreet eu luctus nec, dignissim sit amet nibh. Aenean et imperdiet turpis, suscipit posuere orci. Phasellus sed velit maximus, facilisis magna vitae, ornare elit. In hac habitasse platea dictumst. Ut non lacus eu tortor varius aliquet. Sed eu urna blandit, porta nibh in, mattis nibh.

Conclusions

Sed auctor nisl elementum leo eleifend, vel tincidunt eros vehicula. Duis consectetur augue erat, a bibendum nisi efficitur ac. Maecenas in justo vitae velit efficitur gravida ac in nisl. Pellentesque sit amet nunc magna. Aliquam mollis vulputate scelerisque. Aenean et scelerisque tellus. Donec et metus ante. Mauris dapibus lectus eget leo pretium, sit amet pharetra urna blandit. Suspendisse scelerisque ante mi, quis imperdiet quam suscipit quis. Mauris felis augue, rutrum a dignissim non, semper eu turpis. Donec tempus neque non dignissim dignissim. Duis semper ante risus. Vivamus id consectetur lectus. In enim odio, iaculis et libero ac, aliquam dapibus mauris. Cras eu velit ornare, tincidunt dolor quis, pellentesque nulla.

Bibliography

1. K. Beck *et al.*, Manifesto for Agile Software Development (2001).
2. E. Gamma, *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).

Acknowledgements

This work has been partially supported through funding from the National Science Foundation (#IIA-1301346)