SciDataCon - Denver, CO - September 13, 2016

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

- **Agile Data Curation as a Diversity of Practices Grounded in Shared Values and Principles** - Karl Benedict, Joshua Young, W. Christopher Lenhardt
- **Agile based data curation to handle manuscript related data – ORNL DAAC case study** - Suresh Kumar Santhana Vannan, Tammy Beaty, Daine Wright, Yaxing Wei, Alison Boyer
- **The Data Management Resource Center: An Agile Data Curation approach to expanding data availability and paving the way for others.** - Joshua Young, Larissa Gordon, Doug Dirks, Jeff Weber
- **Dynamic creation of Data Management Policy Rules** - Heike Görzig, Benjamin Gernhardt, Felix Engel, Matthias L. Hemmje
- **Agile Data Management in Practice at a State Geological Survey** - Denise J Hills

Karl Benedict, Joshua Young, W. Christopher Lenhardt

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

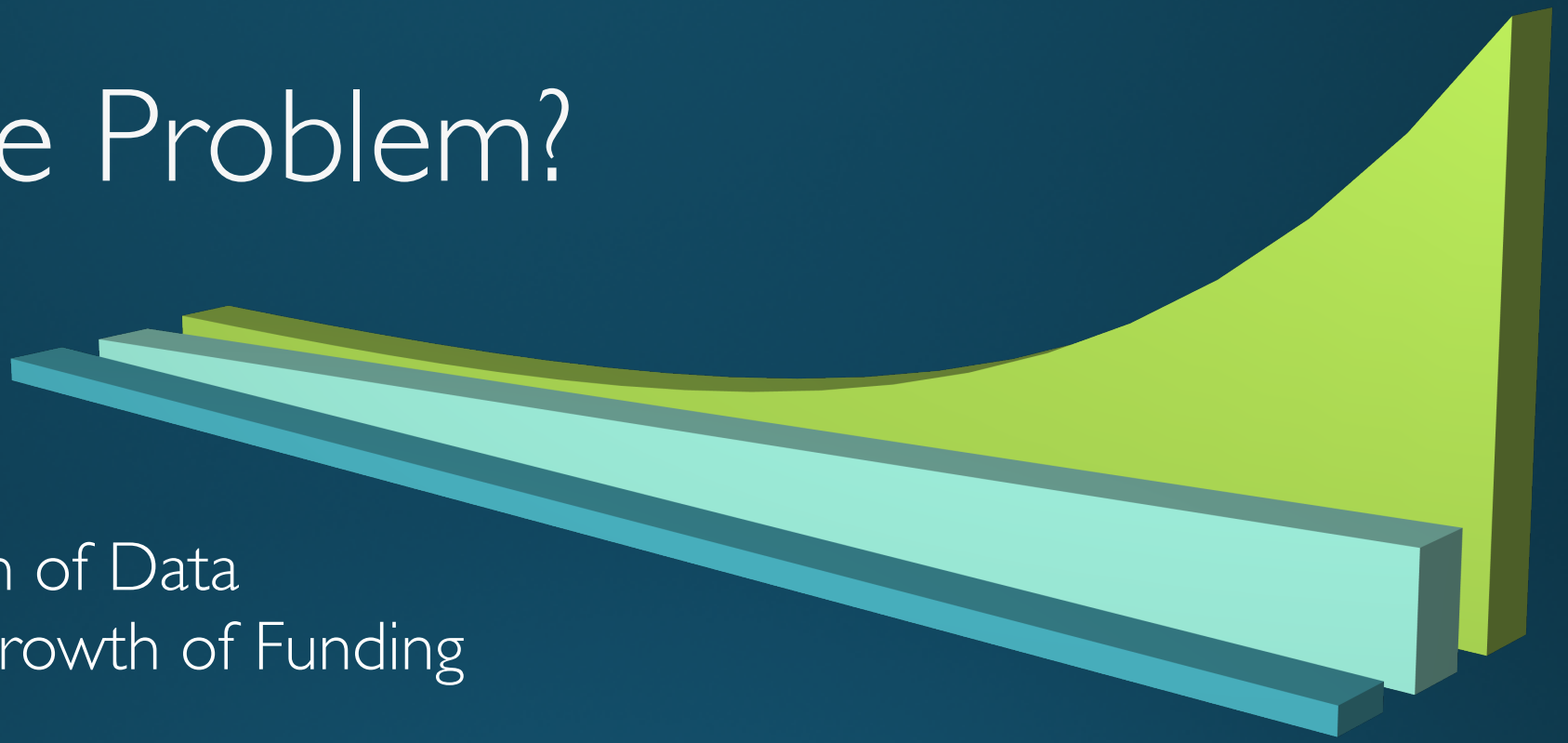
Overview

- What's the problem?
- Parallels with Software Development
- Conceptual Mapping
  - Values
  - Principles
  - Technical Debt
- Next Steps
- Contact Information

What's the Problem?

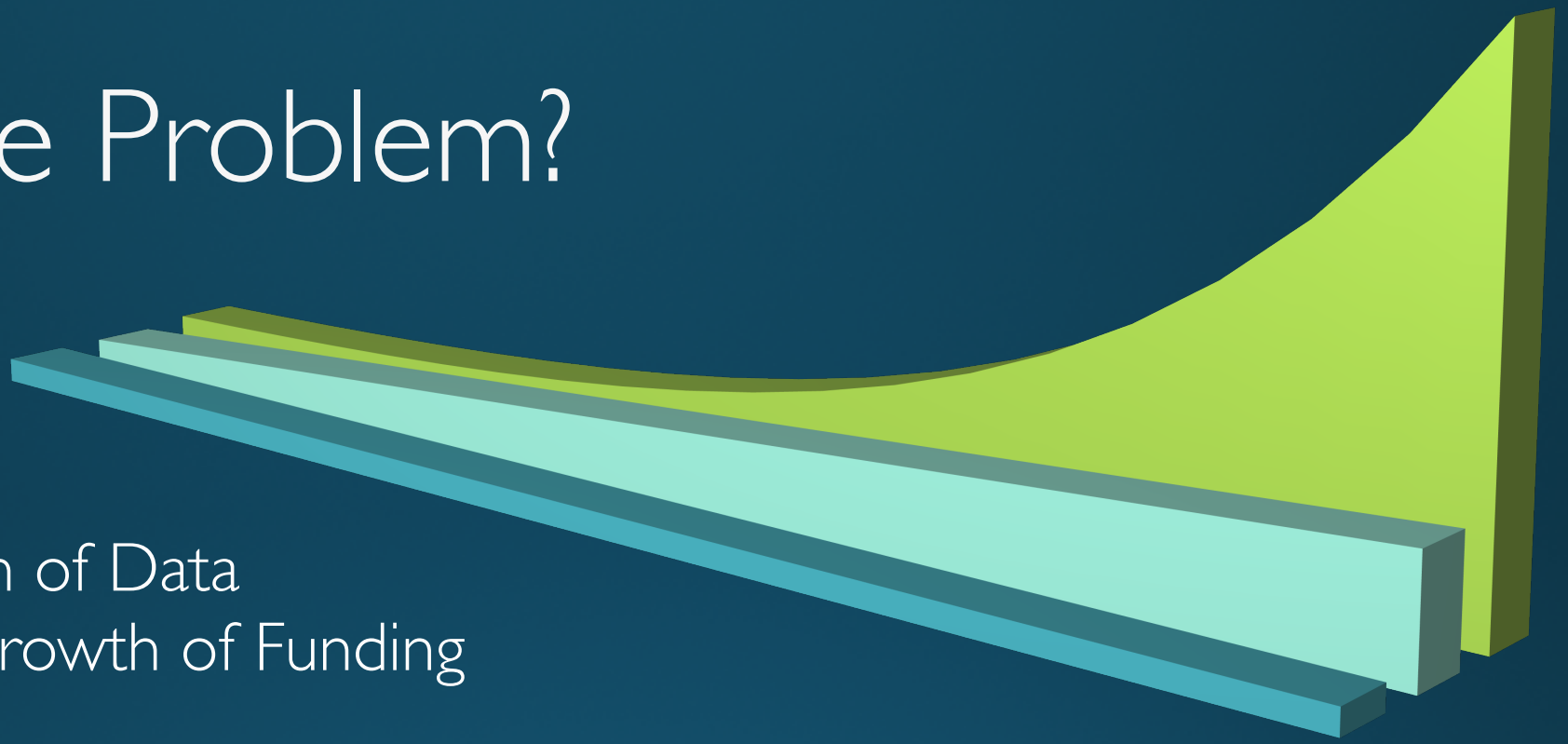
What's the Problem?

What's the Problem?



Exponential Growth of Data
Non-exponential Growth of Funding

What's the Problem?

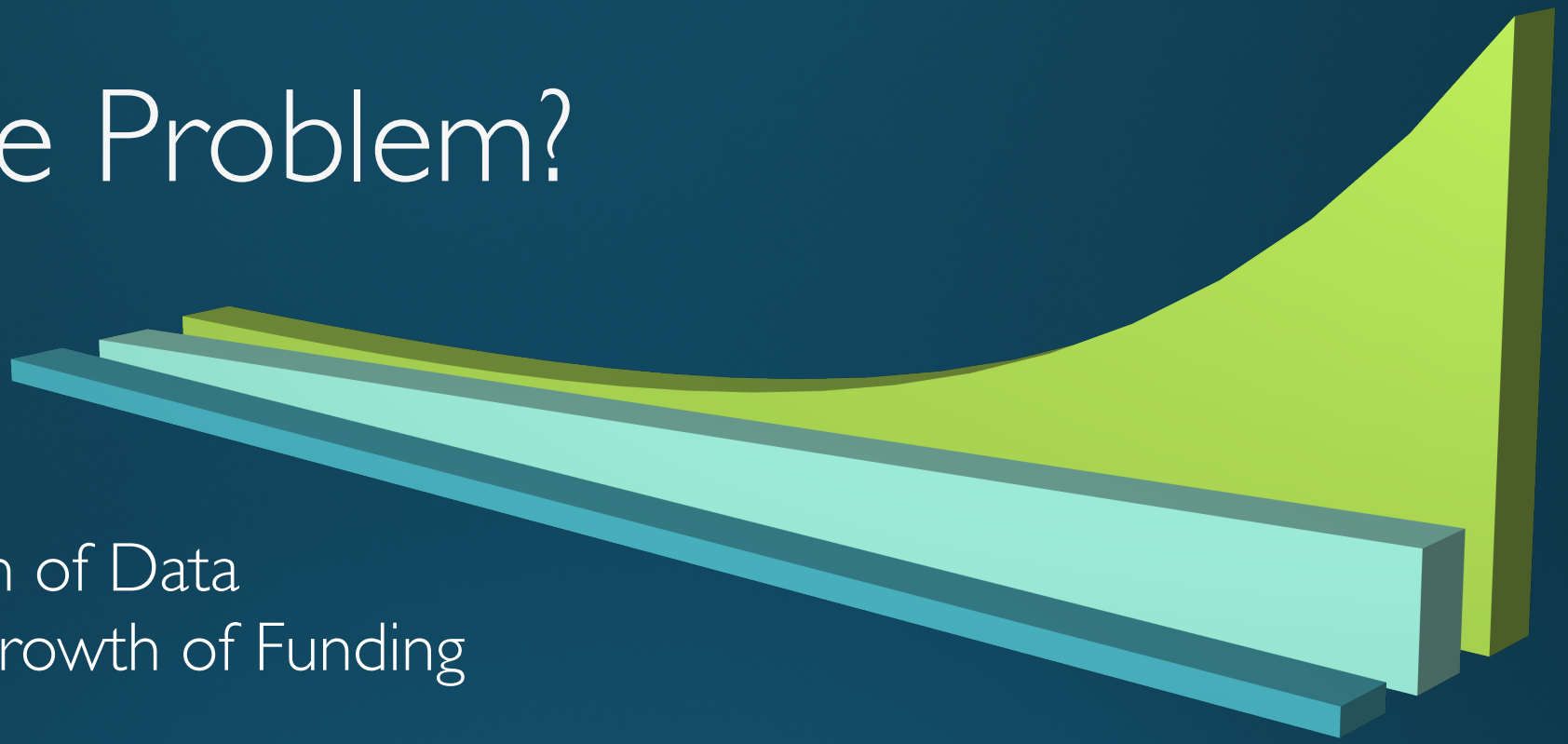


Exponential Growth of Data
Non-exponential Growth of Funding

Increased Demand For:

- Documentation
- Sharing
- Preservation

What's the Problem?



Exponential Growth of Data
Non-exponential Growth of Funding

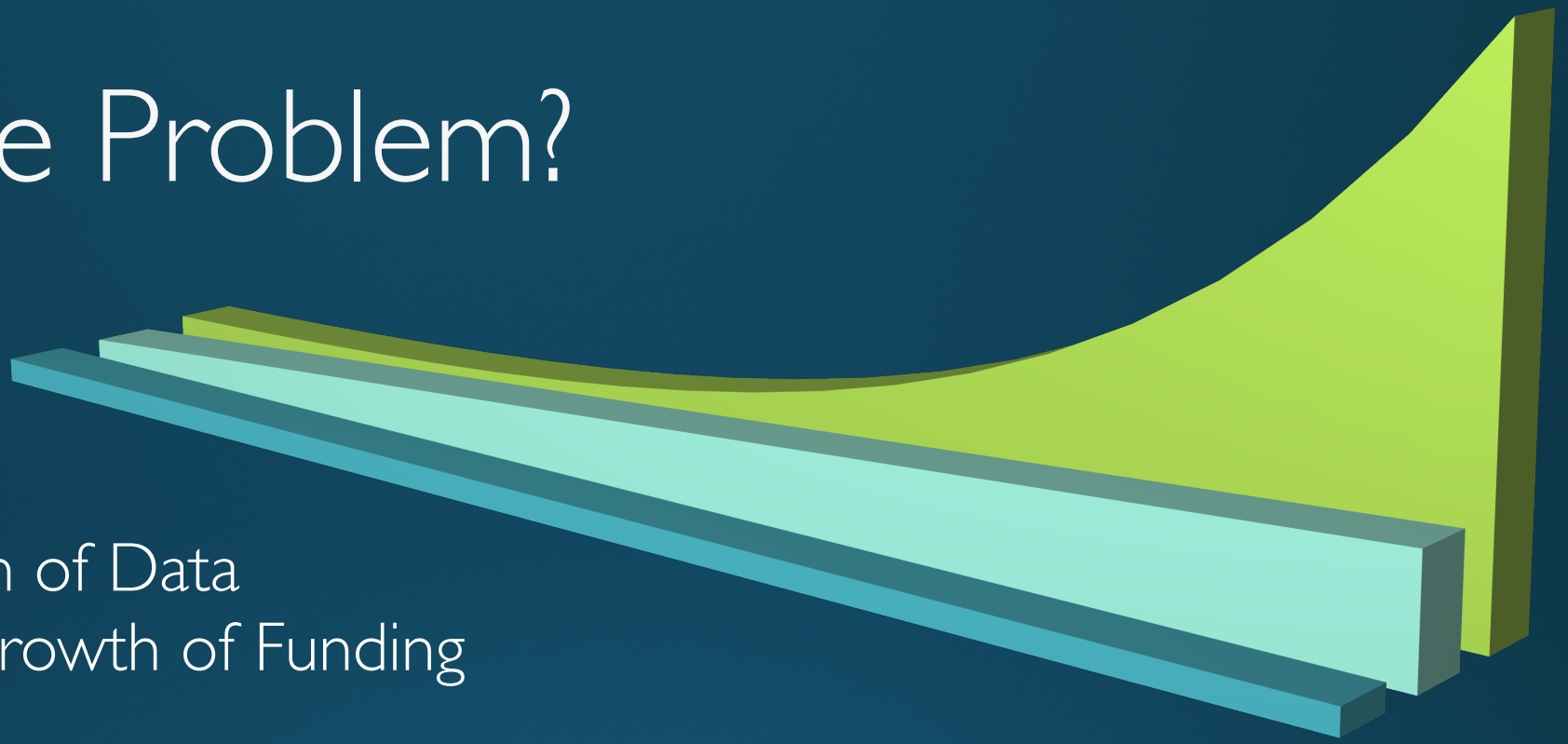
Increased Demand For:

- Documentation
- Sharing
- Preservation

Management

- Organize
- Structure
- Format
- Analyze
- Publish

What's the Problem?



Exponential Growth of Data
Non-exponential Growth of Funding

Increased Demand For:
- Documentation
- Sharing
- Preservation

Management
- Organize
- Structure
- Format
- Analyze
- Publish

Curation
- Selection
- Preservation
- Maintenance
- Collection
- Archiving

Parallels with Software Development

Parallels with Software Development

Parallels with Software Development

Parallels with Software Development

Long development & release cycles
Tightly coordinated development teams
Detailed written specifications
Dedicated documentation development
Targeted at clearly defined target users
High code reuse

Software
Development

Continuous Development
Small-scale shared development
Continuous change in functional objective
Irregular/unstructured documentation
Developers are the users
Minimally structured and low reuse

Long development & release cycles
Division of labor around data lifecycle
Detailed written specifications
Dedicated documentation development
Targeted at clearly defined target users
High degree of discovery & reuse

Data
Curation

Focus on in-project continuous management
Individual-scale data management structure
Objective aligned with immediate analysis needs
Unstructured documentation
Managers are users
Local (research team) reuse

Parallels with Software Development

Long development & release cycles
Tightly coordinated development teams
Detailed written specifications
Dedicated documentation development
Targeted at clearly defined target users
High code reuse

Continuous Development
Small-scale shared development
Continuous change in functional objective
Irregular/unstructured documentation
Developers are the users
Minimally structured and low reuse

Long development & release cycles
Division of labor around data lifecycle
Detailed written specifications
Dedicated documentation development
Targeted at clearly defined target users
High degree of discovery & reuse

Focus on in-project continuous management
Individual-scale data management structure
Objective aligned with immediate analysis needs
Unstructured documentation
Managers are users
Local (research team) reuse

Conceptual Mapping

Agile Software Values

Agile Software Values

Individuals and interactions over processes and tools

Working software over comprehensive documentation

Customer collaboration over contract negotiation

Responding to change over following a plan

Agile Software Values

Individuals and interactions over processes and tools

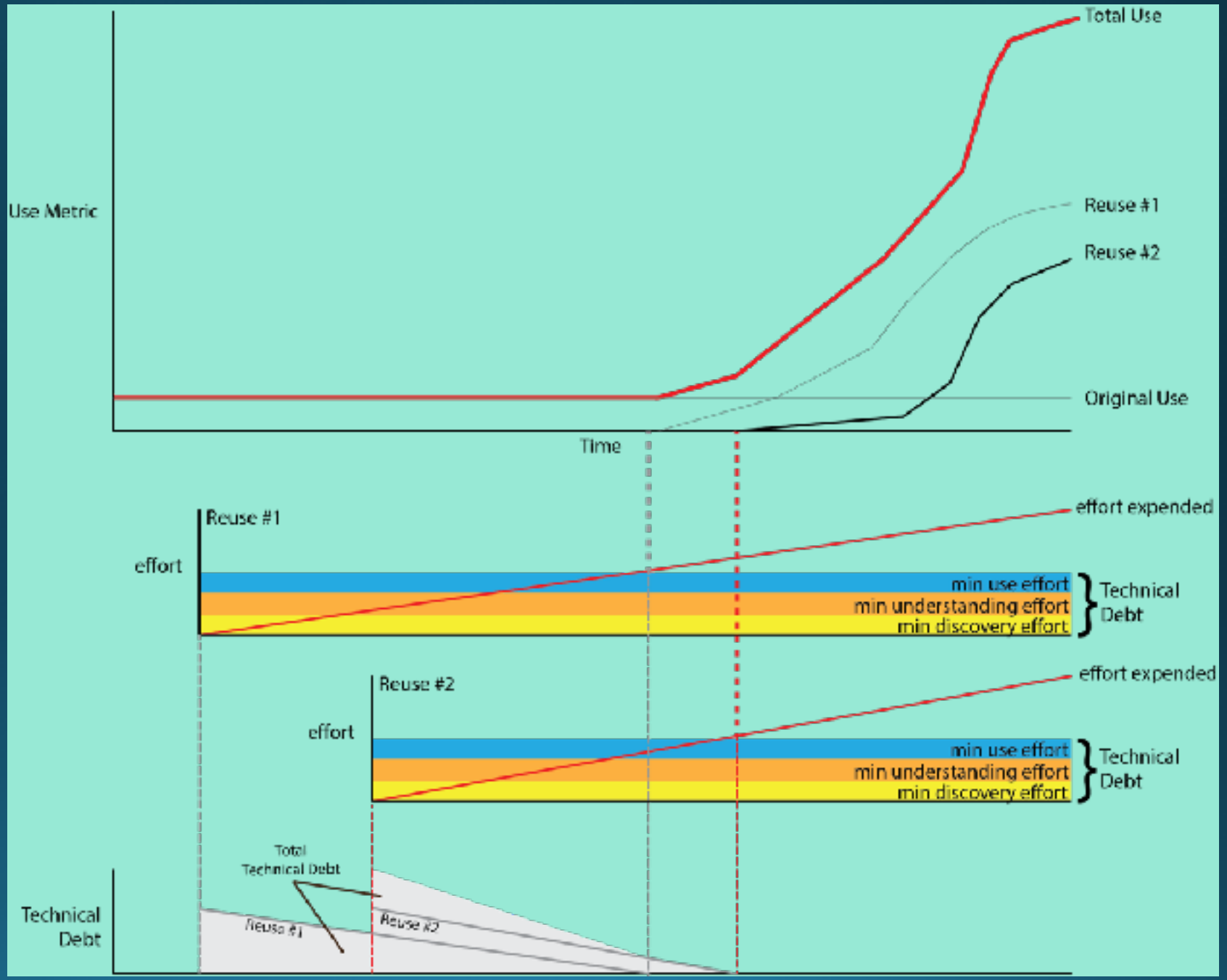
Working software over comprehensive documentation

Customer collaboration over contract negotiation

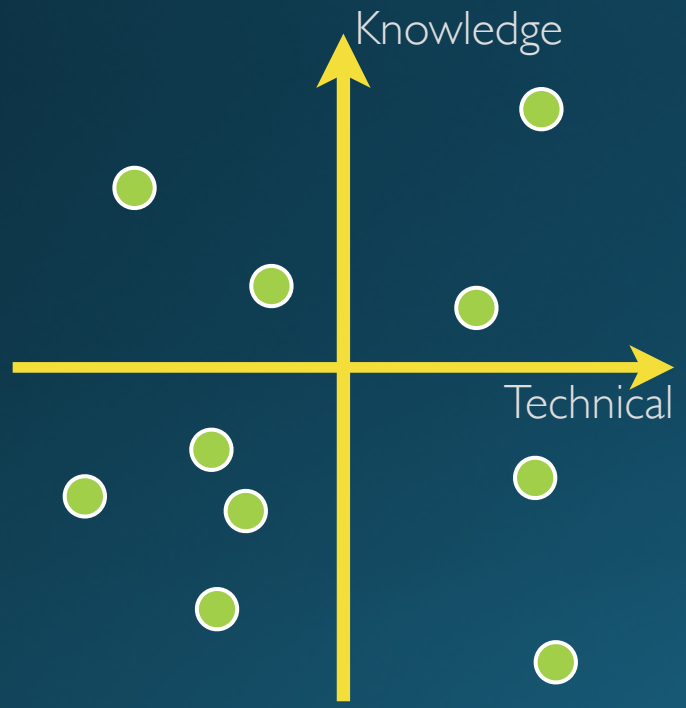
Responding to change over following a plan

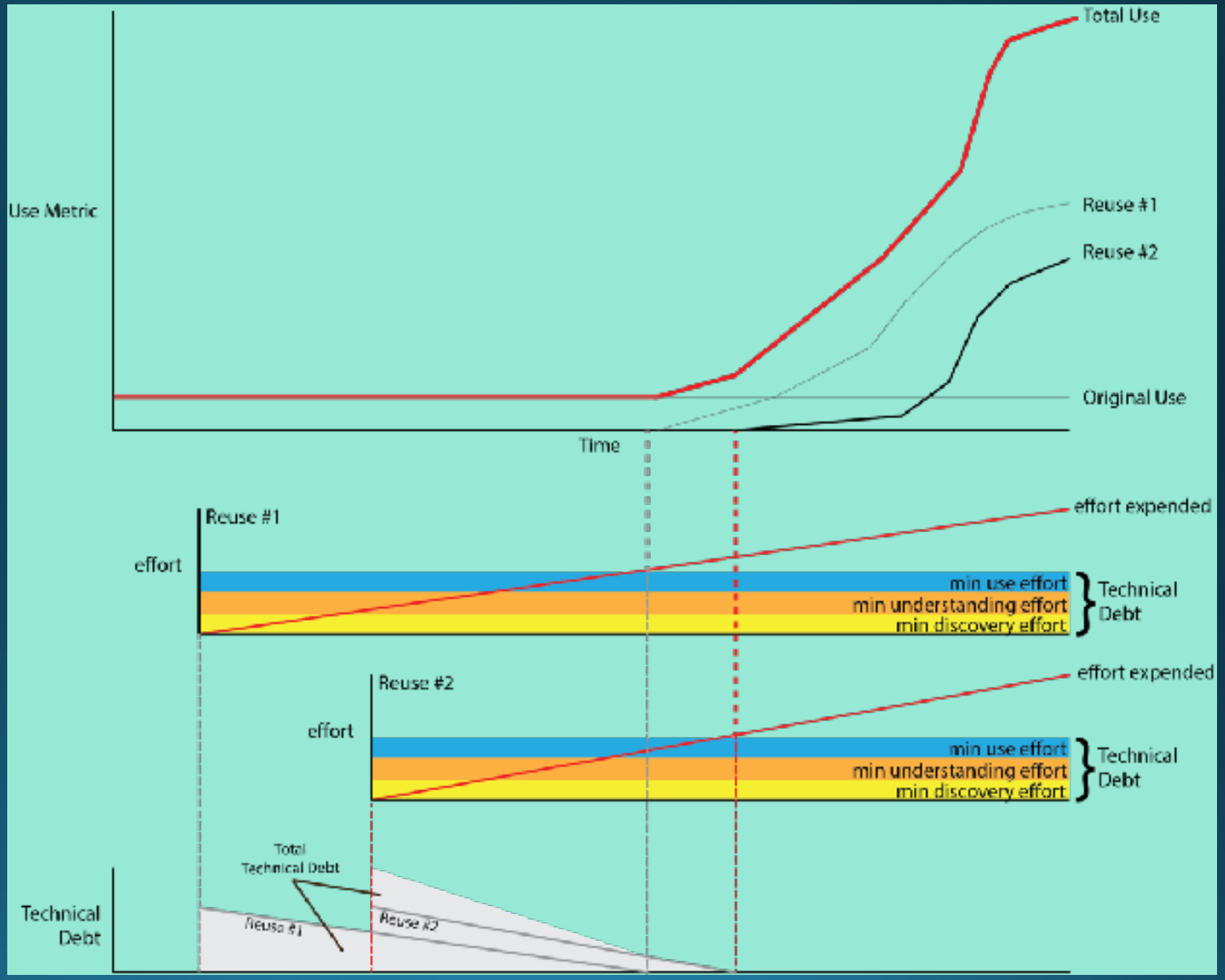
incremental delivery of value to users

Technical Debt

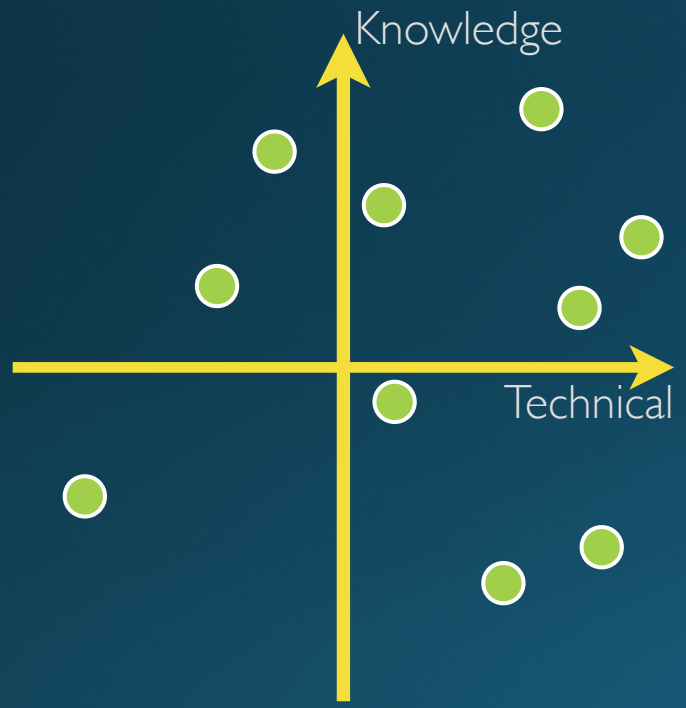


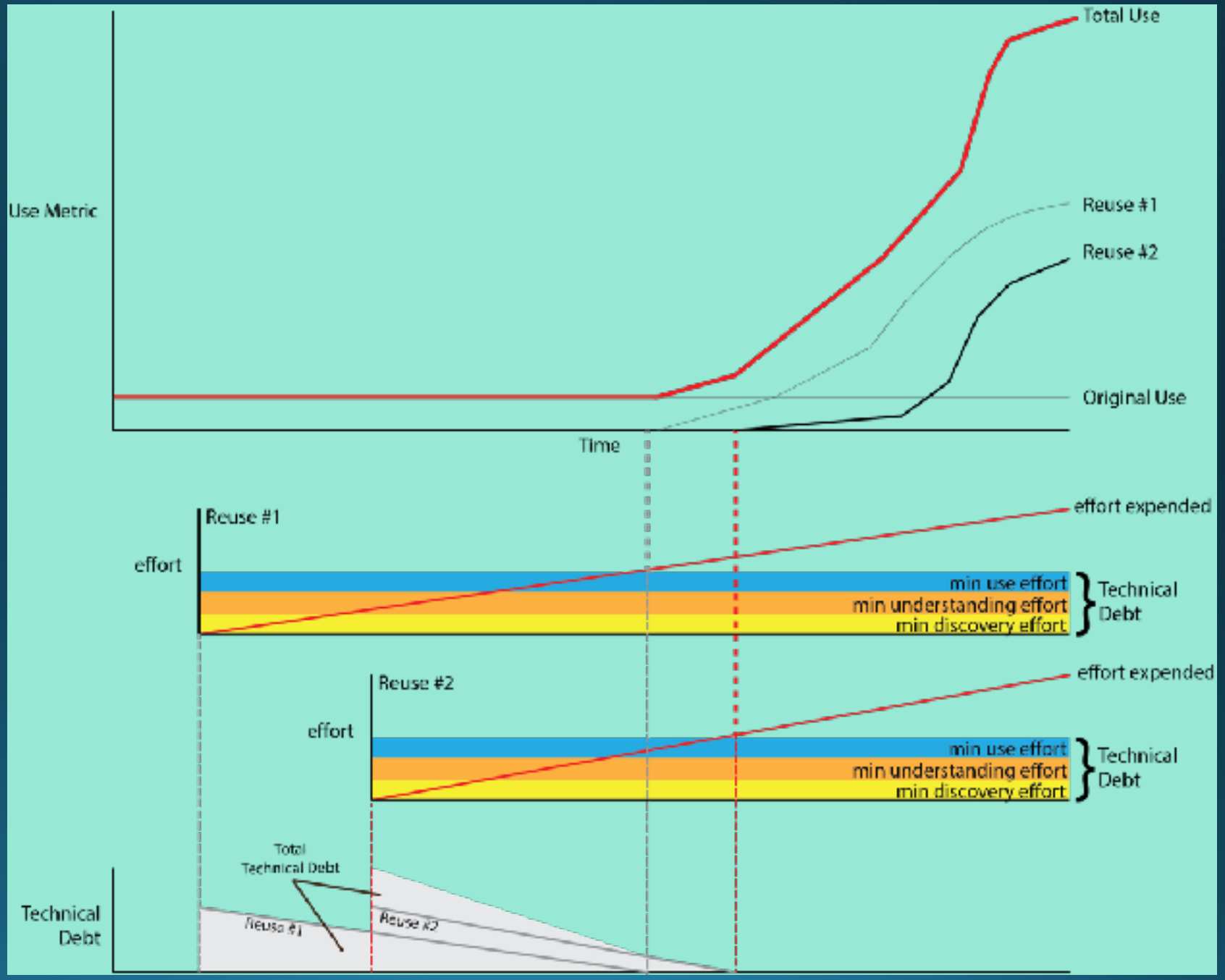
Technical Debt



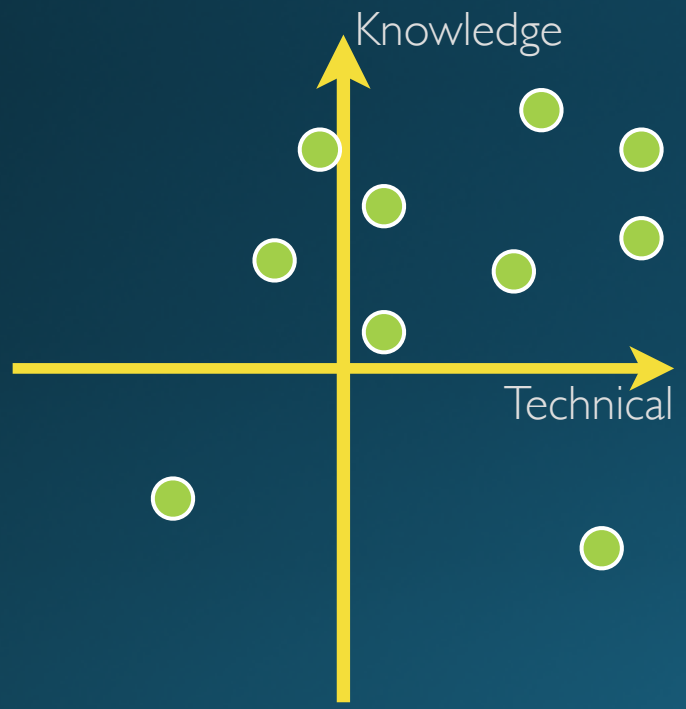


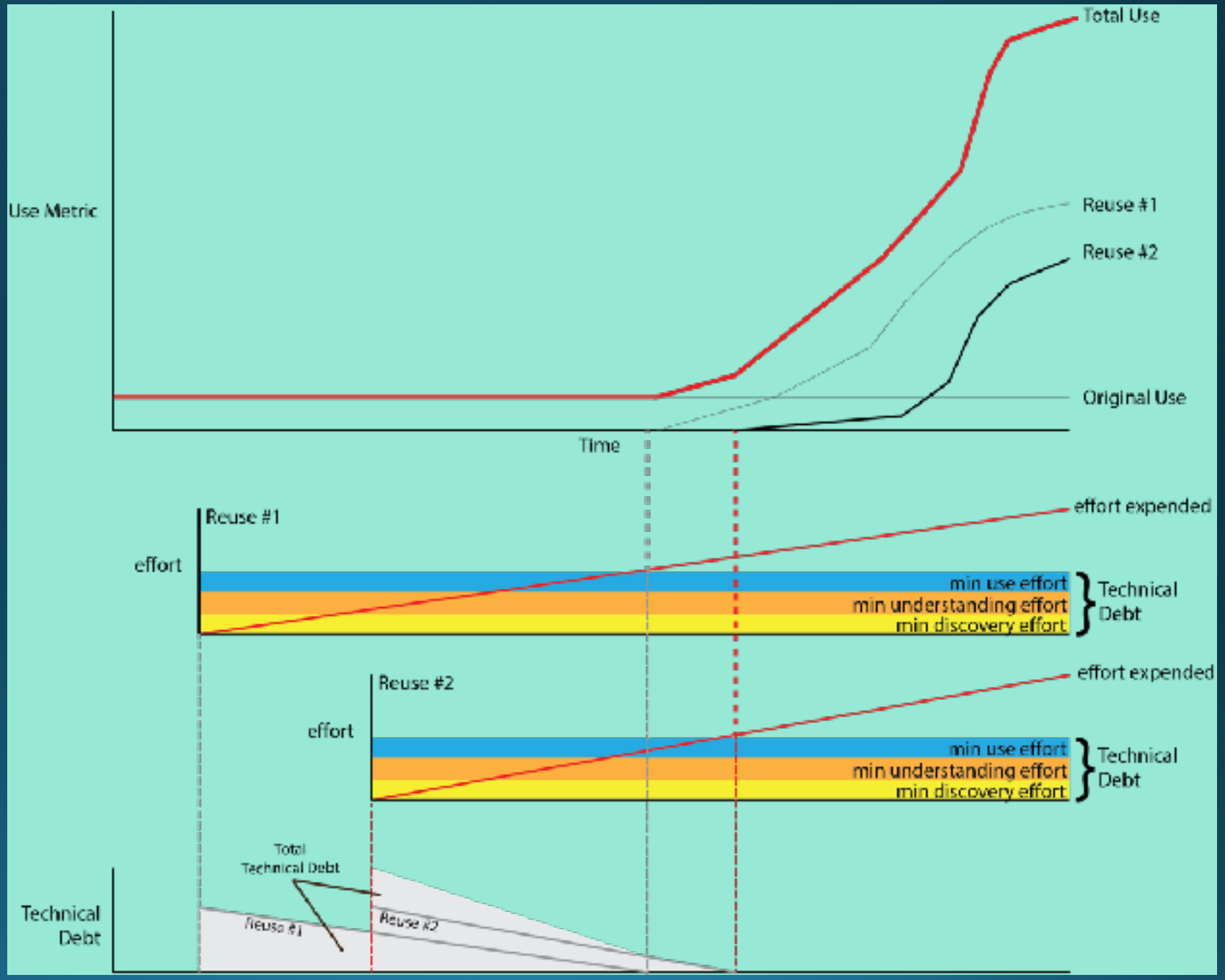
Technical Debt





Technical Debt





Next Steps

- Begin community conversation and development of *shared* agile data curation *values and principles*
- Derive a set of explicit and implicit values and principles from the literature (e.g. FORCE11 FAIR Data Principles)
- Develop capacity to capture structured information about *exemplars of agile data curation* value and principles to contribute to …
- The development of *data curation design patterns* informed by practice and generalized for reuse in developing new data management and curation processes, and assessing existing ones

Contact Us …

- Karl Benedict – kbene@unm.edu
- Chris Lenhardt - clenhardt@renci.org
- Joshua Young - jwyoung@ucar.edu

- Open Science Framework Coordination Site: https://osf.io/d2bac/

https://docs.google.com/forms/d/17IWFYcp63Vr74WaFL5dO3V3LgmQdPJ-qamQi3rGGxZQ/edit

Draft Case Study Capture Form

https://www.surveymonkey.com/r/agile-values-principles

Draft Values & Principles Survey

Principles

- 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

Principles

- *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*

Principles

- 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