

A Data Architecture Fram

Greg Colati, University of Connecticut

Patrick L. Carr, University of Connecticut

CNI Spring Meeting, 2018

The Meaning of Yes

Our original research was done by a group of librarians, archivists, and technologists trying to create a more systematic and strategic approach to managing digital collections.

We believe that the results are generalizable to others facing similar issues.

The Working Group: Steve Batt, Roger Brisson, Patrick L. Carr, Greg Colati, and Betsy Pittman

How do we even talk about this?

Understanding how to steward analog collections is well understood.

However, there is not even a common vocabulary for talking about digital collections.

Data Architecture

A way to structure information for inventory control

”A set of rules, policies, standards and models that govern and **define** the type of **data** collected and how it is used, stored, managed and integrated within an organization and its **database** systems”*

[*https://www.techopedia.com/definition/6730/data-architecture](https://www.techopedia.com/definition/6730/data-architecture)

Information Architecture

”The practice of deciding how to arrange the parts of something to be understandable”*

A way to structure information for use by people

*<https://www.iainstitute.org/what-is-ia>

Information Architecture: A
framework of inventory
presentation for humans

Data Architecture: A framework of
inventory **management for**
systems

The Big Idea

Typically we think only about information architecture and let data architecture “happen.”

We must think strategically about **both** data architecture and information architecture.

The Current State of Confusion

Because of siloed systems that don't inter-operate well, there is a complex series of relationships that make it difficult to discover, present, and re-use collections.

The Current State of Confusion

We have created a proliferation of systems and solutions that are inefficient, redundant, expensive, and do not serve the needs of the data or the users.



Our Solution

Content from data producing programs flow through data architecture policies and are sorted into appropriate inventory systems based on the characteristics of the data.

Architecture to the Rescue

Digital objects are pulled from inventory systems, routed through presentation policies (IP rights, access and use limits, licenses, etc.) and distributed to presentation systems.

Data Classification

Data types are matched against stewardship requirements in three categories:

- Reliability or Digital Integrity
- Sustainability or Digital Curation
- Accessibility or Rights and access management

We are proposing a way of thinking, not a universal criteria.

System Classification

Characteristics of management systems in both custodial and non-custodial circumstances are considered when assigning data to inventory systems.

The systems themselves must be classified as to their ability to deliver required services.

Six Principles of Data Arch

1. All data stewarded by the library will meet its stewardship requirements.
2. For each digital object, only one record will have stewardship.
3. All item types will be evaluated and classified.
4. A list of item types and their classification will be updated as item types are defined or added.
5. All library data management systems will be evaluated at stewardship level.
6. A list of management systems and their classification will be updated as these systems are implemented and old systems are decommissioned.

Data Architecture in Action

The G Protein-Coupled Receptor,
GPR3 Promotes the Acquisition of
Oocyte Meiotic Competence: an
electronic dissertation

Data Architecture in Action

Characteristics/Activities	Levels of Stewardship		
	Level 1: Minimal	Level 2: Enhanced	Level 3: Full
Integrity			
Fixed	M	Y	Y
Permalink	M	Y	Y
Global identifier	M	M	Y
Provenance	Y	Y	Y
Context	Y	Y	Y
Curation			
Bitstream maint.	Y	Y	Y
Review	M	Y	Y
Backup	M	Y	Y
Format normalization	N	M	Y
Format migration	N	N	Y
Redundancy	N	N	Y
Audit Trail	N	N	Y
Error checking	N	N	Y
Access			
Reusable	M	M	Y
Interoperable	M	M	Y
Open	M	M	M

Data Architecture in Action

Women and the Reformation: an e-
book

Data Architecture in Action

Characteristics/Activities	Levels of Stewardship		
	Level 1: Minimal	Level 2: Enhanced	Level 3: Full
Integrity			
Fixed	M	Y	Y
Permalink	M	Y	Y
Global identifier	M	M	Y
Provenance	Y	Y	Y
Context	Y	Y	Y
Curation			
Bitstream maint.	Y	Y	Y
Review	M	Y	Y
Backup	M	Y	Y
Format normalization	N	M	Y
Format migration	N	N	Y
Redundancy	N	N	Y
Audit Trail	N	N	Y
Error checking	N	N	Y
Access			
Reusable	M	M	Y
Interoperable	M	M	Y
Open	M	M	M

DATA PRODUCING
PROGRAMS

PUBLISHING
PURCHASED
RESOURCES

POLICIES

DATA
CURATORIAL

TECHNOLOGIES LTPR

LONG TERM
PRESERVATION
REPOSITORY

Thank you!

- Greg Colati greg@uconn.edu
- Patrick Carr patrick.l.carr@uconn.edu