AN EXAMINATION OF MACHINE-LEVEL POLICY IMPLEMENTATION WITHIN IRODS DATA GRIDS

Jewel H. Ward

Doctoral Candidate

University of North Carolina at Chapel Hill

iRODS User Meeting

Harvard University

18 June 2014

I. Problems and Solutions

II. My Dissertation Research

III. The Questions

IV. Results Q1

V. Remaining Results

VI. Questions/Comments

The Truth

"*What people say, what people do, and what they say they do are entirely different things*." – Margaret Mead

IOW, what are people actually **DOING**?

The Problem

Q: How can archivists, researchers, and other digital preservationists "prove" that digital material is being preserved for some required or indefinite period of time?

The Solution

A: Develop international standards for repositories

- The Open Archival Information System (OAIS) Reference Model (ISO 14721:2012)
  - "Recommendation" for how a preservation system should be implemented by the preservation organization
  - Composite of Functional Models includes: Ingest, Data Management, Access, Archival Storage, Administration, and Preservation Planning
- Trusted Digital Repositories (ISO 16363:2012)
  - "Recommendation" for policies that should be implemented in or around the system by the preservation organization
  - Covers: Organizational Infrastructure, Digital Object Management, and, Infrastructure and Security Risk Management

New Problem

Q: How does an organization "prove" it is following these standards?

A: Receive an Audit and Certification "seal"

- Data Seal of Approval
- DRAMBORA self-audit
- NESTOR Seal for Trustworthy Digital Archives
- etc.

Solution, Cont.

Added by Stephen Marks, last edited by Stephen Marks on Jul 31, 2012 (view change) show comment

Go to Audit Progress Overview

Documentation Item	Started	Draft complete	Committee Approval	Final	Comments
Mission/Vision Statement	✓	✓	✓	✓	
Designated Community Definition	✓	✓	✓	✓	
Preservation Strategic Plan	✓	✓	✓	✓	
Collection Development Policy	✓	✓	✓	✓	
Access Policy	✓	✓	✓	✓	
Rights Policy	✓	✓	✓	✓	
Review Cycle for Documentation Policy	✓	✓	✓	✓	
Risk Analysis and Management Strategies	✓	✓	✓	✓	
Security Plan	✓	✓	✓	✓	
Disaster Recovery Plan	✓	✓	✓	✓	Available on request.
URI & File Naming Plan	✓	✓	✓	✓	
Backup Plan	✓	✓	✓	✓	
Succession Plan	✓	✓	✓	✓	

New Problem & Another Solution

Q: How do you really "prove" people are implementing their policies and standards?

I.e., "Trust, but verify."

A: Implement policies at the machine-level, so that whether or not the policies are actually implemented is verifiable.

OK, But Before You Do THAT…

…shouldn't you look at what people are currently doing?

My Dissertation Research

"Elevator Speech"

A gap analysis whereby I examine what people say they are doing (interviews), examine what they are really doing (core.re files, etc.), and analyze what they are supposed to be doing (i.e., implementing the appropriate sections of the OAIS and ISO 16363 recommendations).

The Questions

- ➜Q1: OAIS & ISO 16363 vs. iRODS msi and rules

  What policies should and can data managers and data scientists implement at the machine-level?

- Q2: Core.re files

  What policies are data managers and data scientists actually implementing at the machine-level?

- Q3: Interviews

  How are data managers and data scientists addressing digital stewardship with regards to the data repositories they manage and use? Whom do data managers and data scientists perceive as being responsible for data preservation?

- Q4: All

  What are the implications for the long-term stewardship of the data when the perception of personnel and an institution's data repository policies are considered against the OAIS and ISO 16363?

Method

1) ➔Compared iRODS microservices and rules against the OAIS and ISO 16363 recommendations
2) Examined core.re files (and other policies)
3) Interviewed iRODS users**

(Finished recently.)

OAIS Composite of Functional Entities
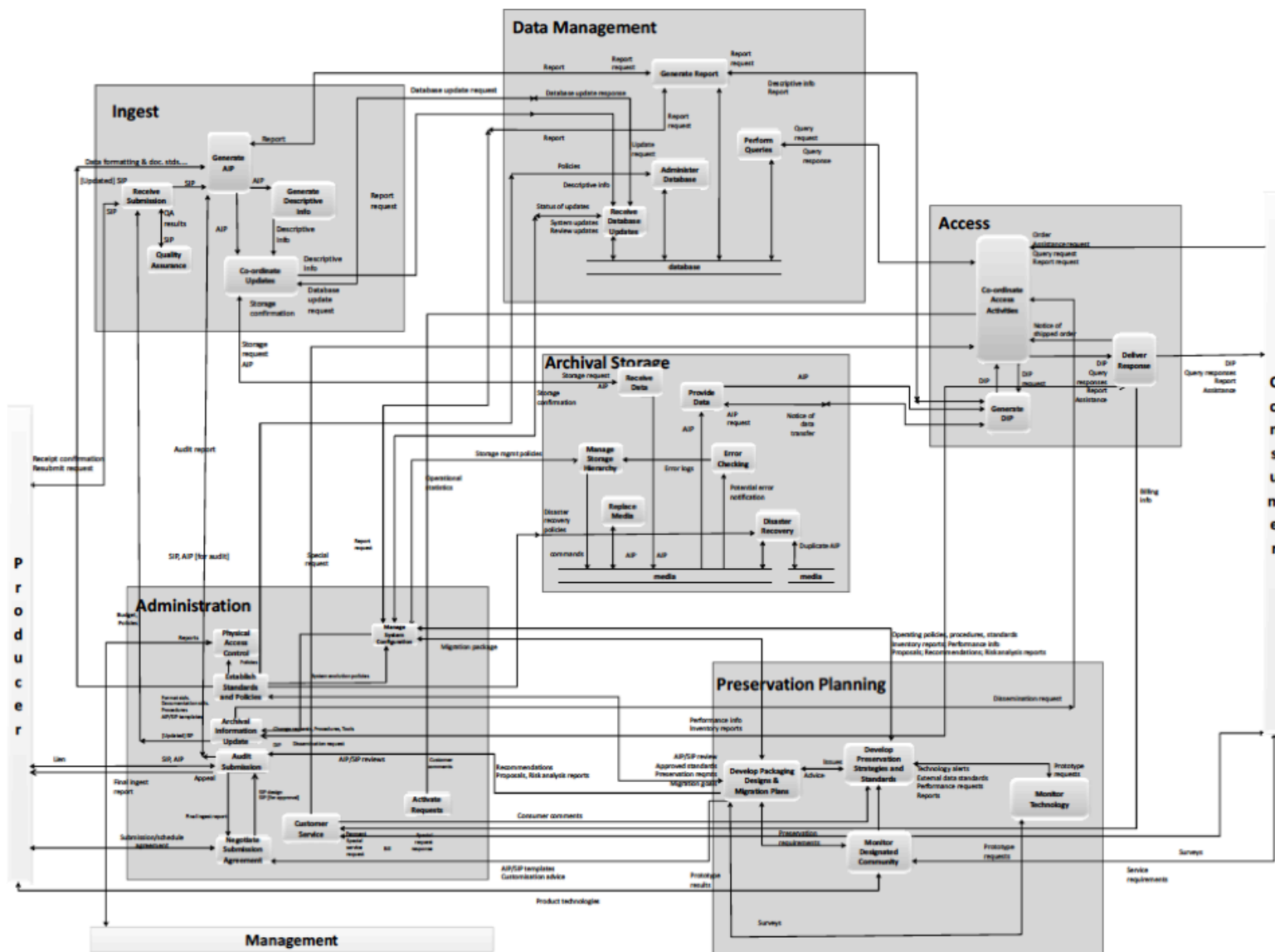
Results (Q1)

Q1: What policies should data managers and data scientists implement at the machine-level? (And are those policies represented in a preservation system/iRODS?)

- **255 iRODS 3.x microservices vs. the OAIS Composite of Functional Entities (CFE)**
  - Msi represented 308 times (1➔ many).
  - Eight msi --- incl. msiReplColl() and msiExtractTemplateMDFromBuf() –– did not map to any OAIS CFE. Ohters incl. msiGetQuote(), msiHello(), etc.
  - Of the OAIS CFE, only "Preservation Planning" is not represented in iRODS

    From the coding notebook: "Administration and Preservation Planning are more the people or human side, where the policies are being developed."

Results (Q1)

- [msi vs OAIS CFE]
  - Common Services (101)
  - Administration (62)
  - Data Management (45)
  - Ingest (37)
  - Access (28)
  - Archival Storage (24)
  - None (8)
  - Preservation Planning (0)

Results (Q1)*

Common Services**

open
close
read
write
lseek
rename
copy
create
delete add
chksum
phymv
unlink
rename

*coding notes
**primitives

Not Common

get
put
repl
trim
rsync/get

Results Q1

Core.re (standard install) vs. OAIS CFE

- The 68 iRODS rules were represented 95 times vs OAIS CFE sections. Again an 1➔ many relationship.
- Again, "Preservation Planning" is not represented in iRODS rules.

Coding Notes (some highlights)

- Had to decide if we are classifying rules or a Policy Enforcement Point. We decided we will classify the PEPs. That means, I deleted any rules that are not PEPs.
- If a rule is called by the C Code, then it is a Policy Enforcement Point (PEP). If a rule calls another rule, then the called rule is not a PEP.
- Look at the rule name, and if it is a PEP, then it can be empty or non-empty.
- All rules are turned on, because they are Policy Enforcement Points.
- When a rule is customized, it is not empty anymore. Only look at the empty ones because they are PEPs.

Results (Q1)

- [core.re vs OAIS CFE]
  - Common Services (28)
  - Data Management (26)
  - Administration (20)
  - Archival Storage (8)
  - Ingest (8)
  - Access (5)
  - None (0)
  - Preservation Planning (0)

ISO 16363 Sections

Results (Q1)

- msi vs ISO 16363 sections
  - 69 ISO 16363 policies represented in 251 msi
    (msi number is different, removed deprecated msi this time).

    - Organizational Infrastructure (34)
    - Digital Object Management (35)
    - Infrastructure and Security Risk Management (0)

- Note: very narrow mapping, could have had broader definition

Results (Q1)

- Core.re (standard install) vs. ISO 16363
  - Problem: we could not do a direct mapping, so we had to do a two-step process
    - Go through ISO 16363 step-by-step, and ID the section, the Event, and the Artifact.
    - Then map the rules to the following definitions defined in Step 1
      - **What we agreed to**
      - Doc1 = human-created documents, both executable and non-executable, includes some metadata (bibliographic, descriptive, etc.).
      - Doc2 = machine-generated, constantly updated, (incl. preservation metadata, etc.), current state + history, logs, etc.
      - Doc3 = product (reports, "doc2 + procedures that are applying assessment criteria"). Ex.:: stepping through Tufts' B2 Ingest. ReMo: have docs/digital objects...I have time-dependent entries, I have DOs that are the products of the procedure

Results (Q1)

	A	B	C
	ISO 16363 Section	Artifact	Event
1	ISO 16363 Section	Artifact	Event
2	3.1.1	shall have a mission statement	
3	3.1.2	shall have a Preservation Strategic Plan	
4	3.1.2.1	shall have an...succession plan,	repository ceases to exist or the
5	3.1.2.2	shall monitor its organizational environment	
6	3.1.3	shall have a Collection Policy or other	
7	3.2.1	duties...appointed staff	
8	3.2.1.1	the duties	
9	3.2.1.2	staff	
10	3.2.1.3	active professional development program	

Results (Q2)

iRODS	State Transitions		Note: Doc1 = Doc2			
Core.re code	**Based on ISO 16363**		Note: Doc3 producing/consuming			
Standard core.re "turned on" (default install)	**Documentation1**	**Documentation2**	**Documentation3**	**Procedure**	**Planning**	*TOTAL*
acCreateUser {						
acPreProcForCreateUser;						
acCreateUserF1;						
acPostProcForCreateUser; }	0	0	0	0	0	0

Results

- Documentation1 (21)
- Documentation2 (21)
- Documentation3 (0)
- Procedure (0)
- Planning (0)

Results (Q1)

- In summary, for the last part of Q1, ISO 16363 vs. core.re

  - We mapped policies to code via state transitions

  - ISO 16363 policies ➔ Doc1, Doc2, Doc3 ➔ iRODS rules

So What?

- msi and core.re mapping to the OAIS CFE
  - The results show the "state of the state" of a current preservation system vs. established recommendations.

  Do some areas of iRODS need to be brought up to speed, or is the system fine like it is?

- msi and core.re mapping to ISO 16363
  - This shows the preservation capabilities of the system, and closes the gap between specification and implementation

Remaining Results

- Answers to Q2-4 coming to a dissertation near you soon….

Questions? Comments?

jewel_ward@unc.edu