

Maintaining Standards

Managing metadata consistency across collections and tools

Michael Durbin



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Outline

- About Policies and Standards
- Policy Creation
- Policy Enforcement
- Conclusions, Questions, Discussion



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

ABOUT METADATA POLICIES



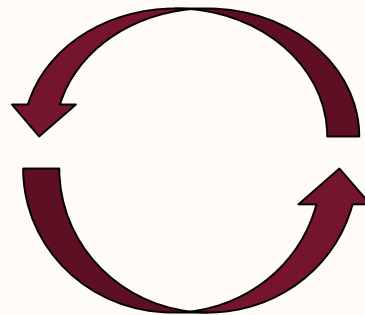
DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Metadata Standards vs. Policies

- What do we mean by “standards”?
 - A level of conformance to norms/expectations.
- What do we mean by “policies”?
 - Our agreement to meet certain standards.

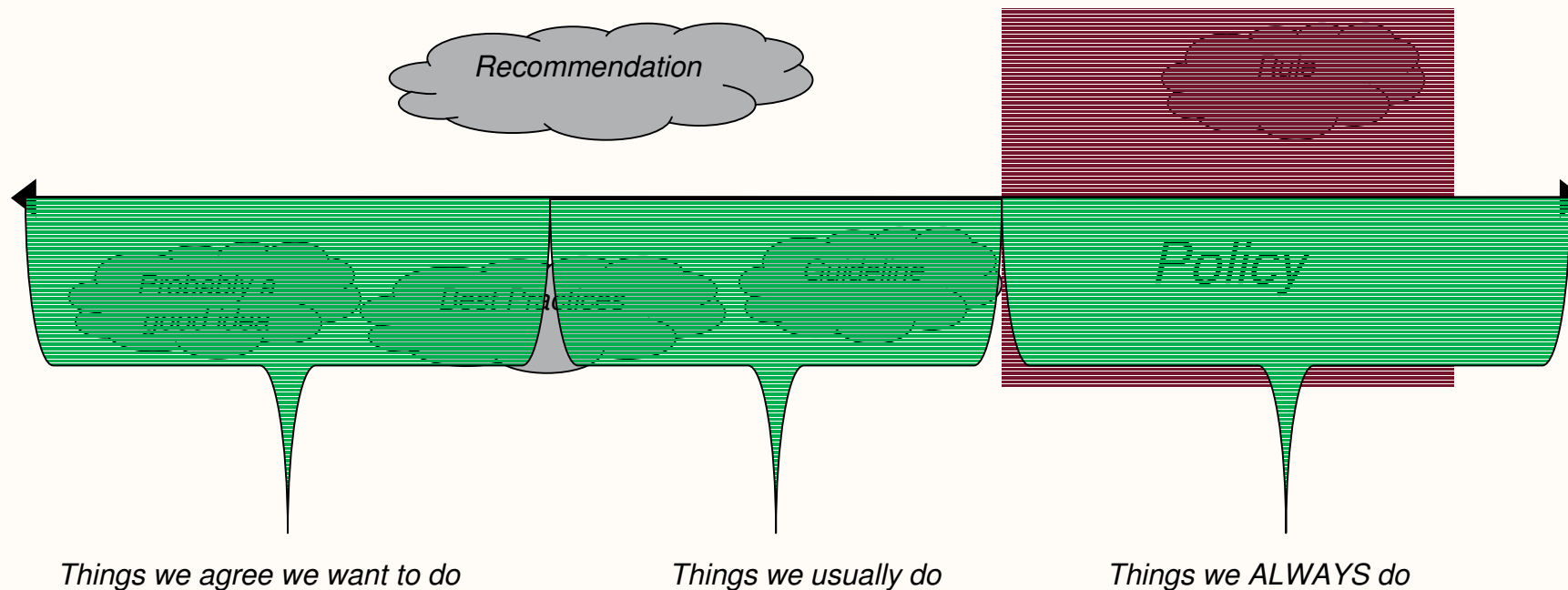


DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Policy Continuum



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Policy Importance - Observations

- Using the terminology from the previous slide, whether you define it up front or not, some “guidelines” will evolve into “policies” and some “policies” will have insufficient compliance and become mere “recommendations”.
- Once production systems depend on, or assume anything about metadata or baseline functionality, it must become policy to conform to that assumption or standard.



Types of Policies

- Metadata Policies
 - Metadata datastream presence
 - Metadata format
 - Minimum Metadata
- Other policies
 - Index requirements
 - That the resource index is enabled
 - That every item is present in another index
 - Data requirements
 - Datastream presence
 - Datastream properties



Why create policies?

- Collect/Retain Preservation metadata for the historical record
- Guarantee baseline user experience
 - Standardization and functional assumptions allow for richer application development
- Allow for current and future management and reporting
- Meet the needs or expectations of others
 - Granting agencies
 - Repository architect
 - Community members
 - Aggregators (OAI-PHM)
 - Potential Collaborators



POLICY CREATION



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Starting point: Fedora Policies

- Every object has a unique PID that serves as its identifier.
- Every object has DC datastream.
- ... and pretty much nothing else.

Low requirements, low baseline functionality



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Policy Creation Considerations

o Do

- support your workflow
- enable useful features
- consider the community and future collaboration (sometimes peer pressure is good)

o Don't

- hamper progress for policies without just cause
- sacrifice user experience for ideals



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

POLICY ENFORCEMENT



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

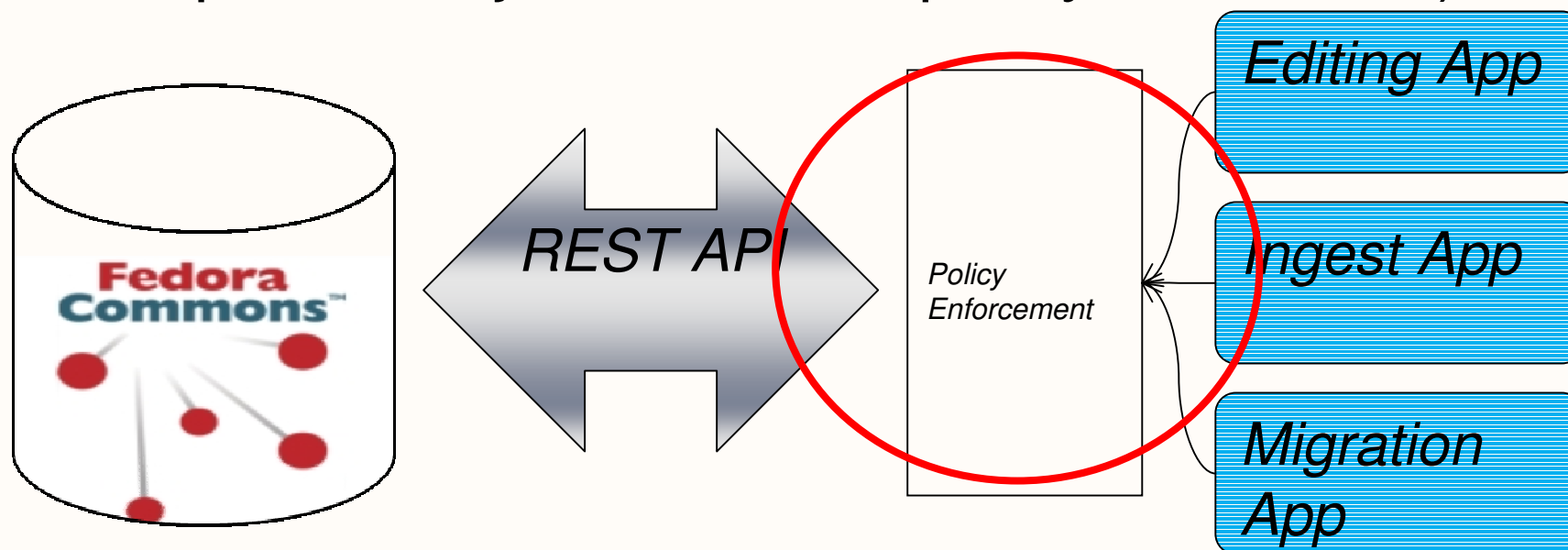
Policy Enforcement

- Conformance to policies may change at any time that
 - an object is born
 - an object changes
 - policies change



Policy Enforcement on Object Creation or Change

- Use a single codebase for Fedora updates that ensures policy adherence (or ensure that all code that updates objects ensures policy adherence)



Policy-enforcing Interface

o Advantages

- Relatively simple
- Relatively few places to introduce problems
- Allows you to reasonably assume that all items in the repository conform to your policies

o Disadvantages

- Prevents or complicates the use of new (potentially useful) applications that modify your repository



Policy Enforcement on Object Change: Other options

- Periodic Policy Validation
 - Requires tolerance and asynchronous correction of non-conformant objects
- Selection of policies that can be enforced by Fedora/disseminators or user-facing applications
 - Severely limits policy options, or requires extensive changes to Fedora



Policy Enforcement on Policy Change

- When policies change, your guarantees or expectations for compliance are immediately voided.
- Because policies are a construct or concept outside of Fedora, there is no internal mechanism for identifying non-compliant objects or correcting them



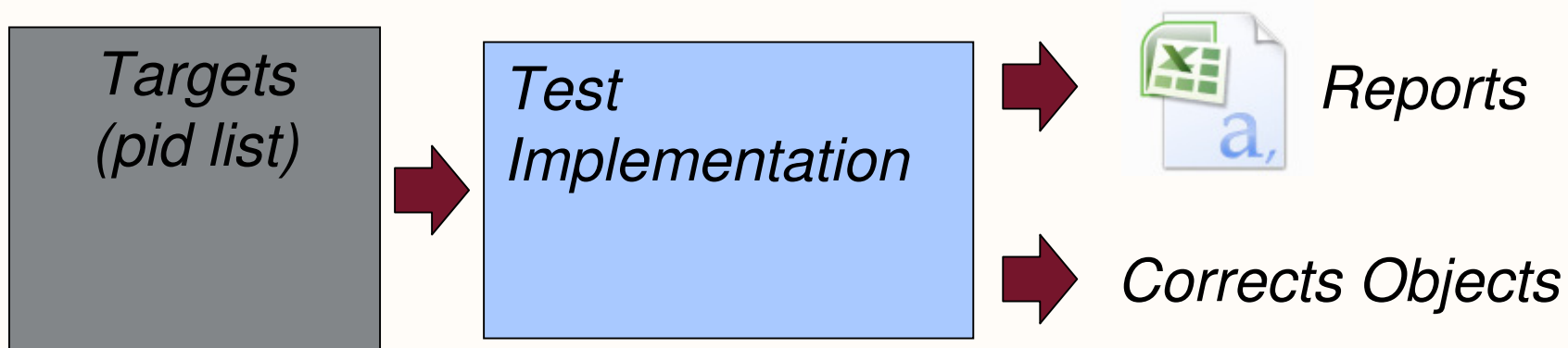
Policy Change: Tools and Procedures

- Local Tool: Repository Audit
 - Tests the relevant subset of the repository for compliance with a given policy
 - Reports which objects pass or fail
 - In some cases, corrects those items that fail



Repository Audit Tool: Technical Details

- Uses Resource Index to generate list of items to test
- Uses REST API to access objects for testing
- Outputs test results in a CSV file, and corrects problems if an automatic solution is available



Repository Audit Tool: Example Cases

- o Ensuring there was a “correctly” placed link in MODS for the persistent URL for full record view of every digitized item in a collection.

```

http://purl.dlib.indiana.edu/iudl/lilly/hohenberger/Hoh006.001.0045
</mods:url>
- <mods:url access="object in context" usage="primary display">
  http://purl.dlib.indiana.edu/iudl/lilly/hohenberger/Hoh006.001.0045
</mods:url>
</mods:location>
<mods:accessCondition type="restriction on access">There are no restrictions on access.</mods:accessCondition>
- <mods:accessCondition type="use and reproduction">
  Copyright and reproduction rights for all Frank Hohenberger photographs are held and administered by the Lilly Library. It
  required prior to any reproduction of images of works by artists and photographers other than Frank M. Hohenberger that
</mods:accessCondition>
- <mods:recordInfo>
  <mods:recordContentSource>InU-Li</mods:recordContentSource>
  <mods:recordIdentifier>lilly/hohenberger/Hoh006.001.0045</mods:recordIdentifier>
  <mods:recordCreationDate encoding="w3cdtf">2006-11-15</mods:recordCreationDate>
  <mods:recordChangeDate encoding="w3cdtf">2010-05-23</mods:recordChangeDate>
- <mods:recordOrigin>
  MODS record generated from EAD finding aid for the collection.
  </mods:recordOrigin>
- <mods:recordSourceOfCataloging>

```



Other Experiences

- METS Navigator (Page-Turner Application) upgrade
 - Required slight change in METS profile
- MODS 3.3 upgrade
 - Some older collections used older version
- Image blocking based on Copyright Date



CONCLUSIONS



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Conclusions

- The following three step strategy to policy management is effective and well suited to evolving repositories of moderate size:
 1. Identifying policies
 2. Enforcing policies on object creation/update
 3. Exhaustively updating non-conformant objects on the creation or modification of a policy



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Lingering Concerns / Room for Improvement

- All of these approaches lack encapsulation of policies within Fedora
- The cost of policy modification increases linearly with repository size
 - stifles flexibility or encourages planning ahead



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Questions?

Michael Durbin – midurbin@indiana.edu



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services