



PoDRI

Policy-Driven Repository Interoperability: an Applied Research Project

Richard Marciano – Director, SALT Group
David Pcolar – Lead, CDR

PoDRI Background

- The **PoDRI** project investigates requirements for policy-aware interoperability mechanisms between repositories and demonstrates key features needed for their implementation.
- A project funded by IMLS





PoDRI Leadership

- Project Leads
 - Richard Marciano, Director, SALT Group
 - Dave Pcolar, Lead, CDR
- Partnership:
 - DuraSpace/Fedora Commons, UNC/DICE/iRODS, UNC/RENCI



PoDRI Policies

- In the Pledge project [PledgePolicyPrototype – DSpace Wiki, 2007], policies are described as:
 - *A policy is typically a rule describing (or prescribing) the interactions of actions that take place within the archive, or a constraint determining when and by whom an action may be taken. For example, a policy could demand that every Item being submitted include an approved deposit license. Another policy might demand that every Bitstream in the asset store be checked for content integrity (i.e. checksum recomputed and compared with the checksum on record) at least once in every six months.*
- Examples include:
 - auditing, replication of content,
 - automatic extraction and association of metadata,
 - validation of checksums, format migration, and trustworthiness.



PoDRI

Evolving Preservation

- Evolution of digital preservation policies
 - from managing digital content and electronic services over the last decade,
 - to developing e-strategies and infrastructure in recent years,
 - to the latest emphasis on *policy-based digital preservation strategies*.



PoDRI

Diversity of Tools/Repositories

- iBiblio, ContextMiner, Archivist Toolkit,
- Fedora, DSpace, ArchiveIT, ContentDM, LOCKSS,
- aDORe, ExLibris DigiTool, Archon, DocuShare,
- EPrints, Blackboard, SharePoint, Documentun,
- L-Store, SRB, iRODS...



PoDRI

Emerging Integrative Frameworks

- Big Digital Machine
- Cooperative Strategies for Distributed Digital Preservation
- DuraCloud: Preservation Infrastructure in the Cloud
- TIPR: Towards Interoperable Preservation Repositories (DSpace, aDORe, and DAITSS)



PoDRI

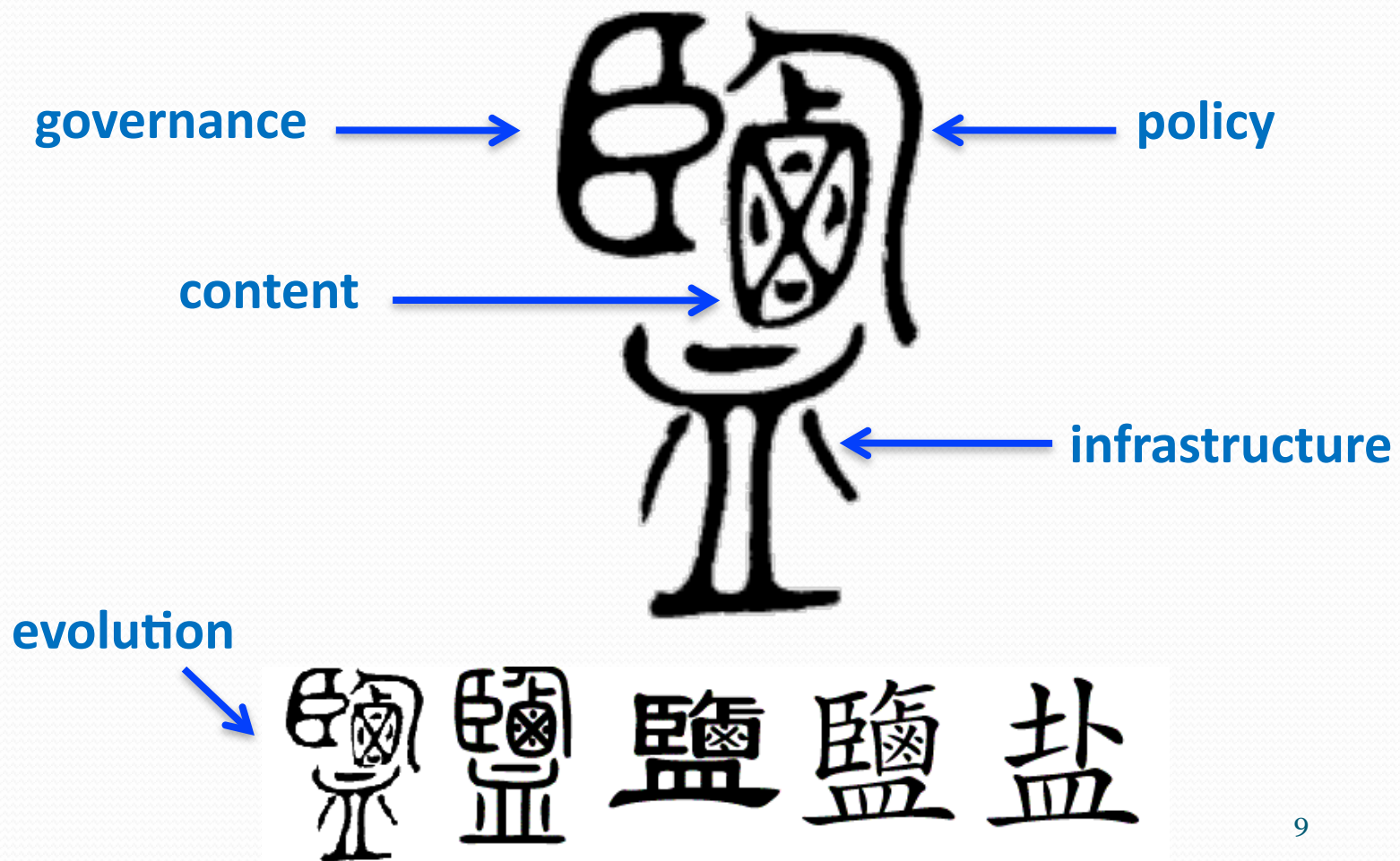
Common Themes

- Scalability, service-oriented frameworks,
- Distributed management, interoperability,
- Business models, digital preservation

SustainAbiLiTy & Evolution

SALT/"yan": a metaphor for sustainability

Richard Marciano





PoDRI

Research Questions

- What is the feasibility of repository interoperability at the policy level?
- Can a preservation environment be assembled from two existing repositories?
- Can the policies of the federation be enforced across the repositories?
- What fundamental mechanisms are needed within a repository to implement new policies?



PoDRI

Requirements for Architecture

- Persistent
- Interoperable
- Scalable
- Distributed
- Flexible
- Semantic

PoDRI

Component Software

- iRODS 
 - open source
 - federated storage
 - rule engine and microservices
- Fedora  **FedoraCommons™** 
 - open source
 - file-centric
 - front-end APIs
 - richer metadata



PoDRI

Building Technology

- Lab for the digital future
- Sustainability of components
- Diverse ecosystem of resources and partners
- Expanding the universe of interaction
- Evolution of management policies
- Lowers the cost of entry for repository implementation
- Policy frameworks



PoDRI

Management Architecture

- Integrate views of content, original arrangement (hierarchy) and metadata
- Create an audit trail of policy execution events and related provenance information
- Demonstrate policy management through Fedora
- Demonstrate policy management from iRODS



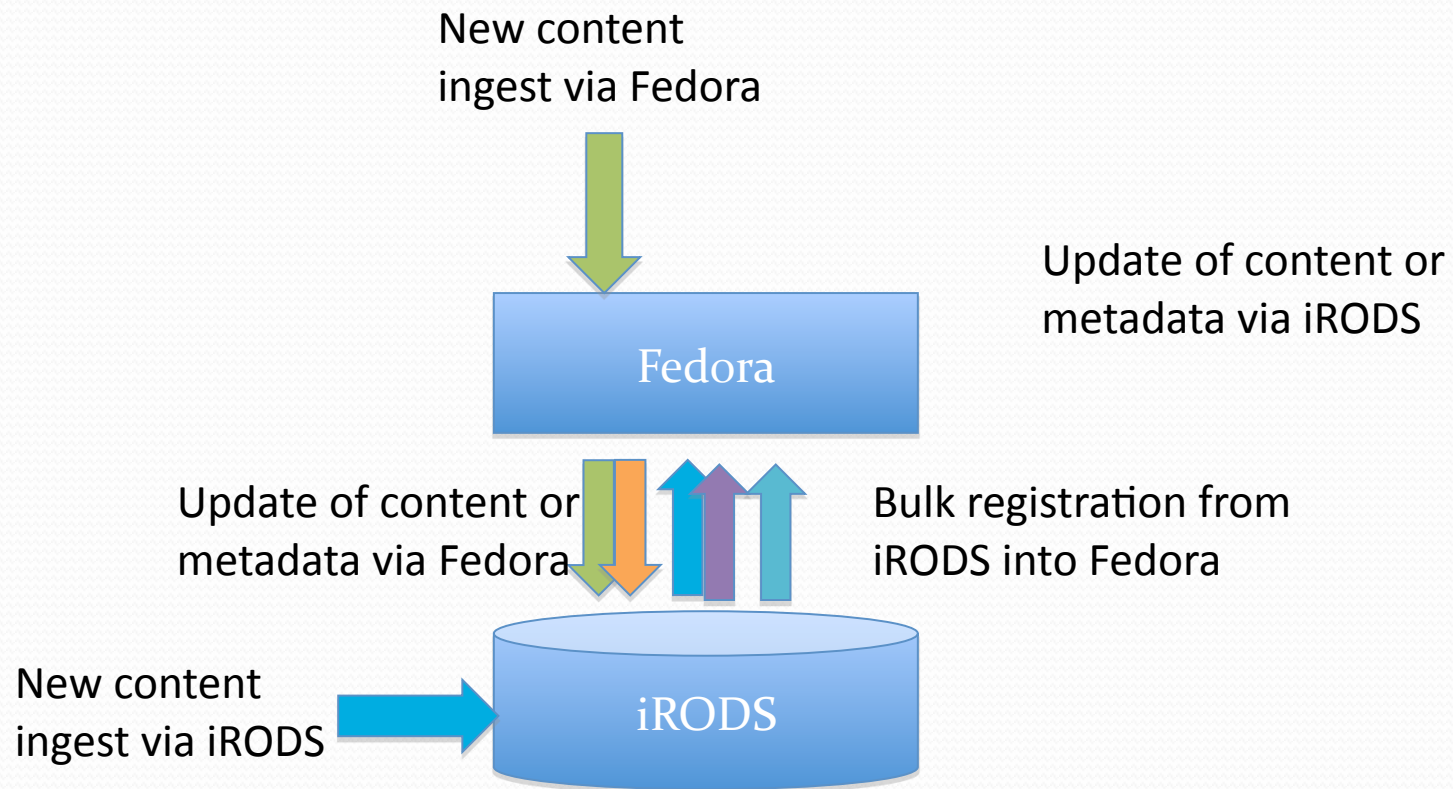
PoDRI

NATIONAL IMPACT AND INTENDED RESULTS

- Advance the state of open repositories by contributing to repository integration and interoperability
- Research interoperability approaches that address the repository policies themselves
- Support evolution through the virtualization of the underlying technology
- Support changing business requirements through a high-level customization of repository behaviors
- Support openness through a data type agnostic treatment of content

PoDRI

Use Cases





PoDRI

Lessons Learned

- **Finding 1:** The transfer of information packages between repositories is not sufficient to guarantee the integrity of the content. In addition, management policies from one repository need to be enforced in the repository where the content is replicated.
- **Finding 2:** Enforcing behaviors and relationships as machine-actionable policies at a remote repository needs to be further researched, as well as other policy-based mechanisms for validating assessment criteria for successful repository integration.
- **Finding 3:** Additional research is needed to explore the feasibility of repository-independent policy representations that lend themselves to interoperability, based upon tests of policy migration between archives



PoDRI

Institutional Content Resources and Repositories

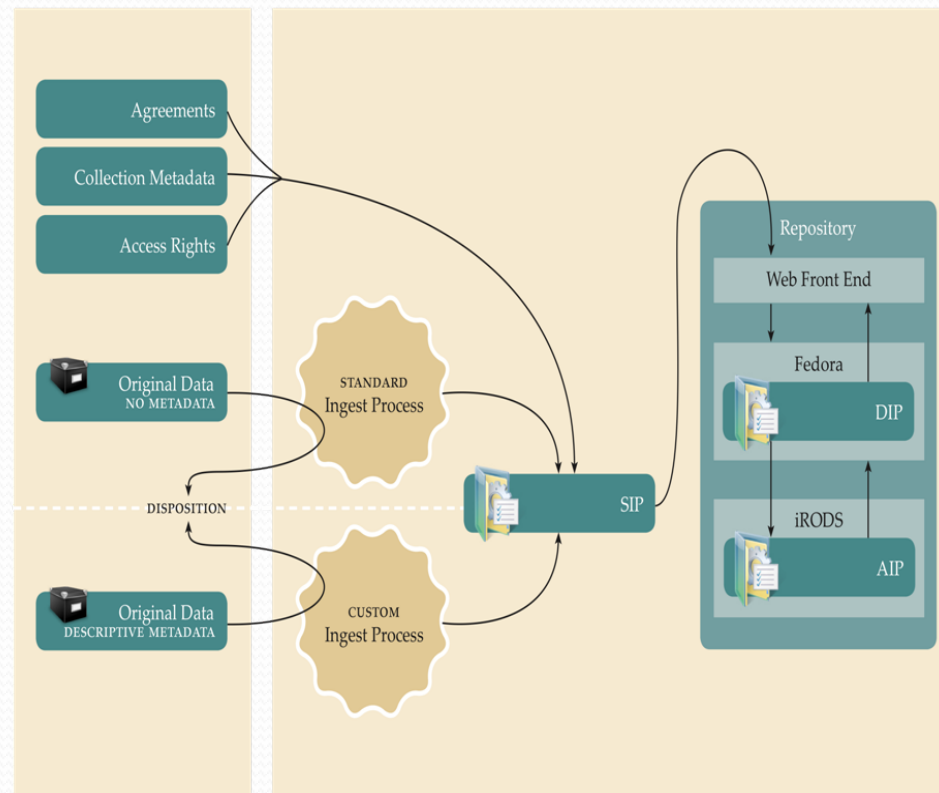
- Carolina Digital Repository (CDR)
 - CDR is designed as a repository for material in electronic formats produced by members of the University of North Carolina at Chapel Hill community. Its chief purpose is to provide for long-term preservation of digital assets. By preservation we mean the ability to ingest the material, index and search it, replicate it, and keep it safe from alteration.
 - Following standards developed in the reference model for an Open Archival Information Systems, the CDR employs Fedora for data content models and uses iRODS as a data store.
 - The University Library is a partner in the PoDRI, to build on the initial Fedora to iRODS connector and investigate interoperability issues in greater depth.

PoDRI and CDR

- Functionality focused on ‘Core Services’
 - Ingestion
 - Indexing & Discovery
 - Preservation
 - Dissemination
- Metadata Standards
 - PREMIS, METS & MODS
- Dublin Core based (FOXML)
- Minimal mandatory data elements
- Rights/Access/Use Metadata

PoDRI and CDR

CAROLINA DIGITAL REPOSITORY
Approved Collection Workflow





PoDRI

CDR's Approach to Data Life Cycle

- Custodial shift after objects are ingested into CDR
- Data management needs increase during SIP preparation
- OAIS model applied throughout architecture
- OAIS concepts (SIP, AIP, DIP) employed
- Implemented via software interaction



PoDRI and CDR

Challenges: IRs & the Life Cycle Model

- Life cycle assumes data is inactive and won't change once ingested into repository
- Develop approach to active, at-risk data
- Potential approach: Pre-custodial intervention



PoDRI and CDR Ongoing Work

- Expansion of use cases
- Bulk registration from iRODS into Fedora
- Applying preservation policies on iRODS:
 - integrity check
 - Replication



PoDRI

Open Repositories 2010, July – Madrid

- **Projects / Partners include:**

- **PoDRI,**
- **CDR,**
- **SALT,**
- **Sustaining Heritage Access through Multivalent ArchiviNg (SHAMAN),**
- **DreSNet,**
- **WissGrid**

PoDRI

Acknowledgements

- SALT:
 - Chien-Yi Hou
 - Alex Chassanoff

Libraries/CDR:
Sarah Michalak
Will Owen

- RENCI:
 - Leesa Brieger

SILS:
Helen Tibbo
Carolyn Hank

- iRODS: Bing Zhu

- Fedora: Dan Davis

- DICE: Reagan Moore

DURASPACE



PoDRI

Acknowledgements



TRIANGLE RESEARCH LIBRARIES NETWORK

DUKE UNIVERSITY

NORTH CAROLINA CENTRAL UNIVERSITY

NC STATE UNIVERSITY

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL