

We can build amazing private islands



Repository Island

But should we?

Data Cyberinfrastructure

... visualize it



Understanding Infrastructure: Dynamics, Tensions, and Design



Report of a Workshop on "History & Theory of Infrastructure:
Lessons for New Scientific Cyberinfrastructures"

Paul N. Edwards
Steven J. Jackson
Geoffrey C. Bowker
Cory P. Knobel

January 2007



NSF Grant 0630263 • Human and Social Dynamics • Computer and Information Science and Engineering • Office of Cyberinfrastructure

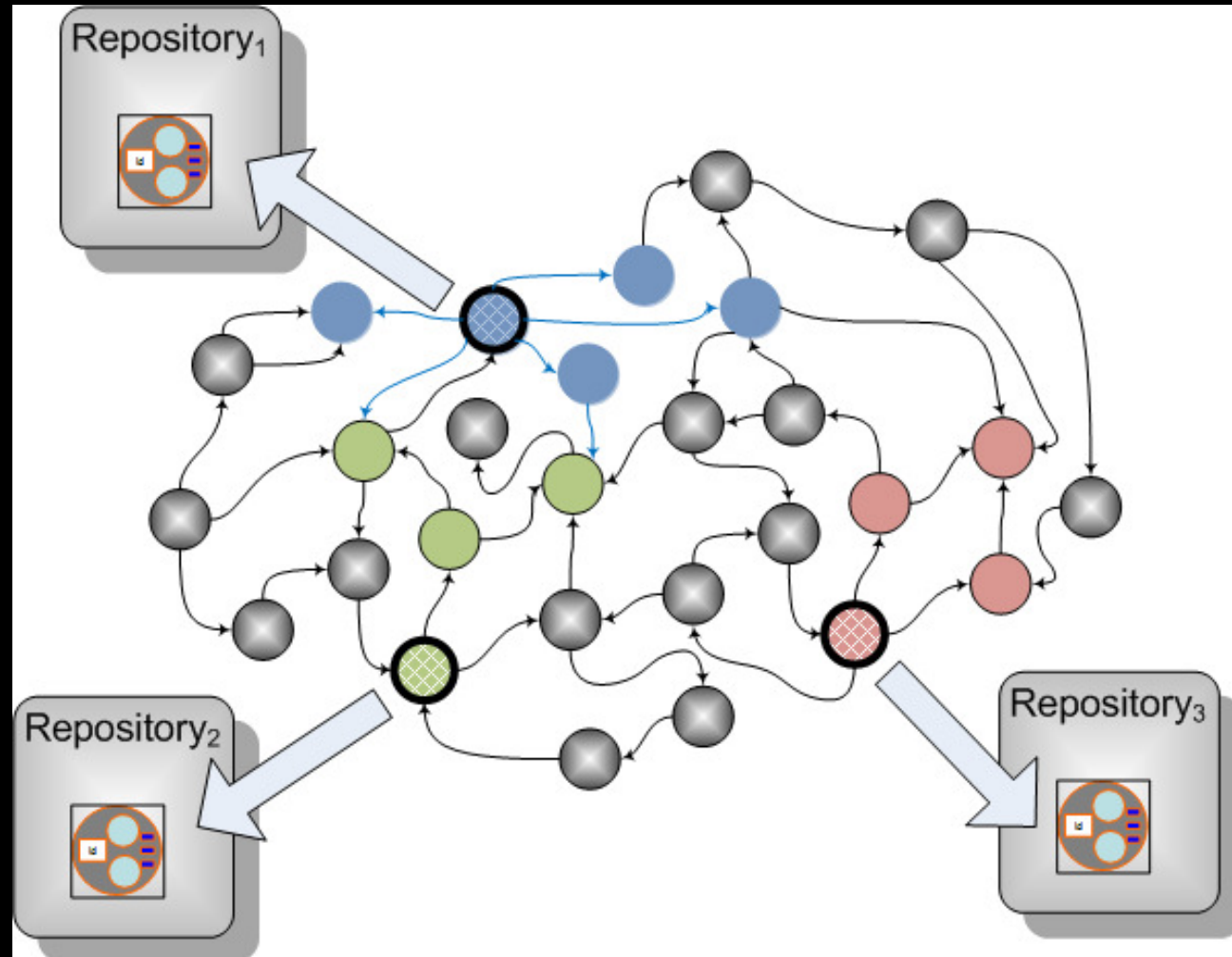
Not a rigid road map, but principles of navigation... no one way to design.

Emergence starts when formerly stand-alone or incompatible systems are interconnected via adapters and gateways

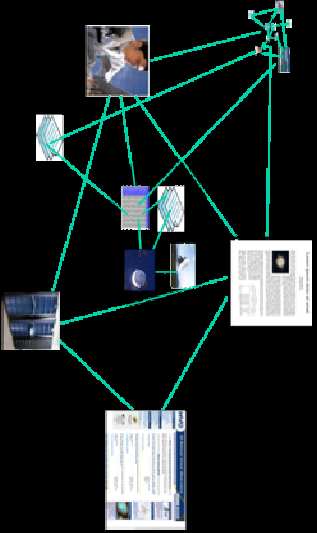
Maturing involves the transition away from homogeneous, centrally controlled, local systems towards heterogeneous, widely distributed, coordinated networks.

Repositories as components of networked information infrastructure

Repositories and Cyberinfrastructure



Implications for our future work



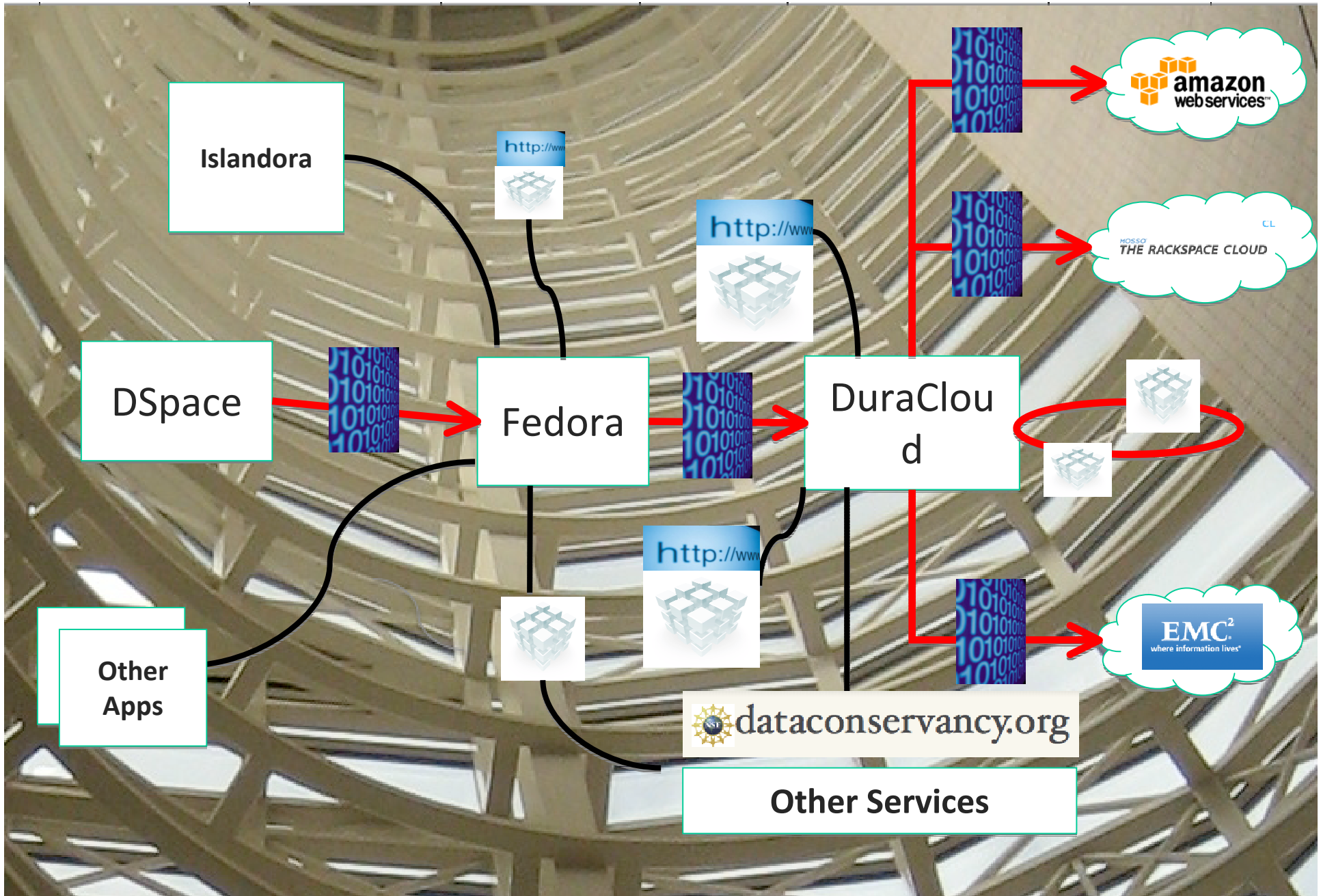
more distributed

more collaborative

more web-oriented


more open

more interoperable



Integration is Key: "DuraInfrastructure" Pattern

JOHNS HOPKINS
UNIVERSITY

 dataconservancy.org



PI, Sayeed Choudhury, Johns Hopkins, Sheridan Libraries



Co-PIs and Partners



Carl Lagoze

Cornell University



Mary Marlino

National Center for Atmospheric Research

Carole Palmer

CIRSS, GSLIS, University of Illinois at U-C

Paddy Patterson

Marine Biological Laboratory

Sandy Payette

DuraSpace

Chris Borgman

University of California Los Angeles

Ruth Duerr

National Snow and Ice Data Center

Eileen Fenton

Portico

Mark Evans

Tessella, Inc.

DURASPACE™

UCLA



National Snow and Ice Data Center



PORTICO



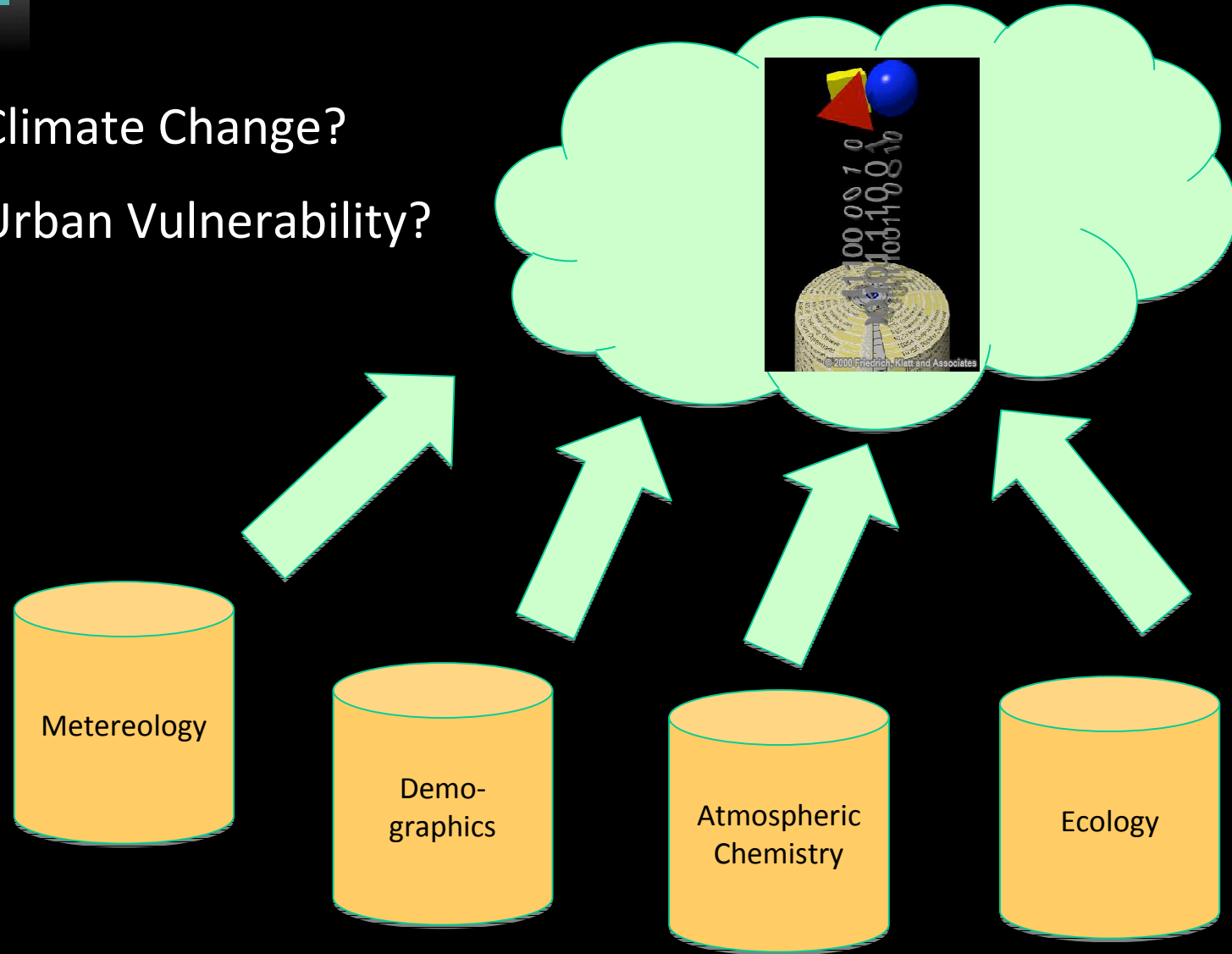
Data Curation

The Data Conservancy embraces a shared vision:

Data curation is a means to collect, organize, validate and preserve data so that scientists can find new ways to address the grand research challenges that face society.

Multi-disciplinary Data Integration

Climate Change?
Urban Vulnerability?



Repositories and Data Curation

- **Data curation**
 - What levels (e.g., raw; cooked)?
- **Management of very large data sets**
 - In situ strategies (e.g., DC with Sky Survey)
 - Registry repositories? Surrogate repositories?
- **Facilitating computational services**
 - Data proximity to computational services
- **Content models for data**
 - Many models; how synthesize?
 - Ontologies to enable cross-domain analysis?

Emergence of Infrastructure

Systems



- Heterogeneous components
- Central control
- Closed, stable
- Dedicated/improvised gateways



Networks



- Heterogeneous systems
- Distributed control
- Coordination
- Generic gateways
- Open, reconfigurable