



Demystifying Born-Digital: Concepts, Trends, Challenges

Anne R. Kenney

Educational Forum on Electronically Generated
Archival Materials

November 19, 2013



Cornell University


MITLibraries

Ten Myths for *not* doing digital preservation:

1. Someone else is already doing it
2. We rely on paper copy not the digital file as our preservation master
3. We're just licensing content
4. There are lots of copies
5. It costs too much
6. Vendors support backwards compatibility
7. We don't know how to do it
8. Storage is cheap, so we just keep everything
9. There are too many other priorities
10. If we wait long enough, someone else will figure it out

10 Reasons to do digital preservation:

1. An increasing % content is at risk
2. We need to protect institutional investments and resources
3. Things can be done now, even if the complete solution isn't clear
4. Libraries, archives, and museums are the only institutions with a preservation mandate based on value to scholars and society
5. There is no evidence that other stakeholders will commit to long-term preservation when it's no longer in their economic interest to do so
6. It's necessary to ensure access
7. Legal and financial requirements may mandate it
8. It's more cost effective to develop preventive measures than recover from data loss/disaster
9. Our covenant with users and donors is based on a confidence that we can preserve the material – if we can't, both parties will lose faith and we'll soon be out of business
10. Decisions on digital preservation should not be driven solely by technology



Community Traces

- **1960s:** National archives, data archives
 - **1970s:** Increasing interest and concern
 - **1980s:** Digitization developments
 - **1990s:** Library, museum, Web collections
 - **2000s:** Digital art, geospatial, e-science...
 - **2010s:** Research data, analog archives...
- variations by nation, domain, size, complexity...



1996 PDI Report

Challenges:

- obsolescence
- migration
- legal issues
- infrastructure
- lack framework

Preserving Digital Information

Report of the Task Force on Archiving of Digital Information

**commissioned
by
The Commission on Preservation and Access
and
The Research Libraries Group**

May 1, 1996



Two Foundation Documents

Trusted Digital Repositories: Attributes and Responsibilities (RLG/OCLC)

<http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>

OAIS Reference Model (CCSDS) [ISO 14721]

<http://public.ccsds.org/publications/archive/650x0m2.pdf>



Attributes of TDR

- OAIS Compliance
- Administrative Responsibility
- Organizational Viability
- Financial Sustainability
- Technological and Procedural Suitability
- System Security
- Procedural Accountability



DPM Model Development

- Mapped requirements to TDR attributes
- Identified interdependencies
- Attributes: not hierarchical or equal
- TDR Model underscores work to be done by an organization – beyond technology

Trusted Digital Repository Model

6. Procedural Accountability (Certification)

1. Administrative Responsibility

2. Organizational Viability

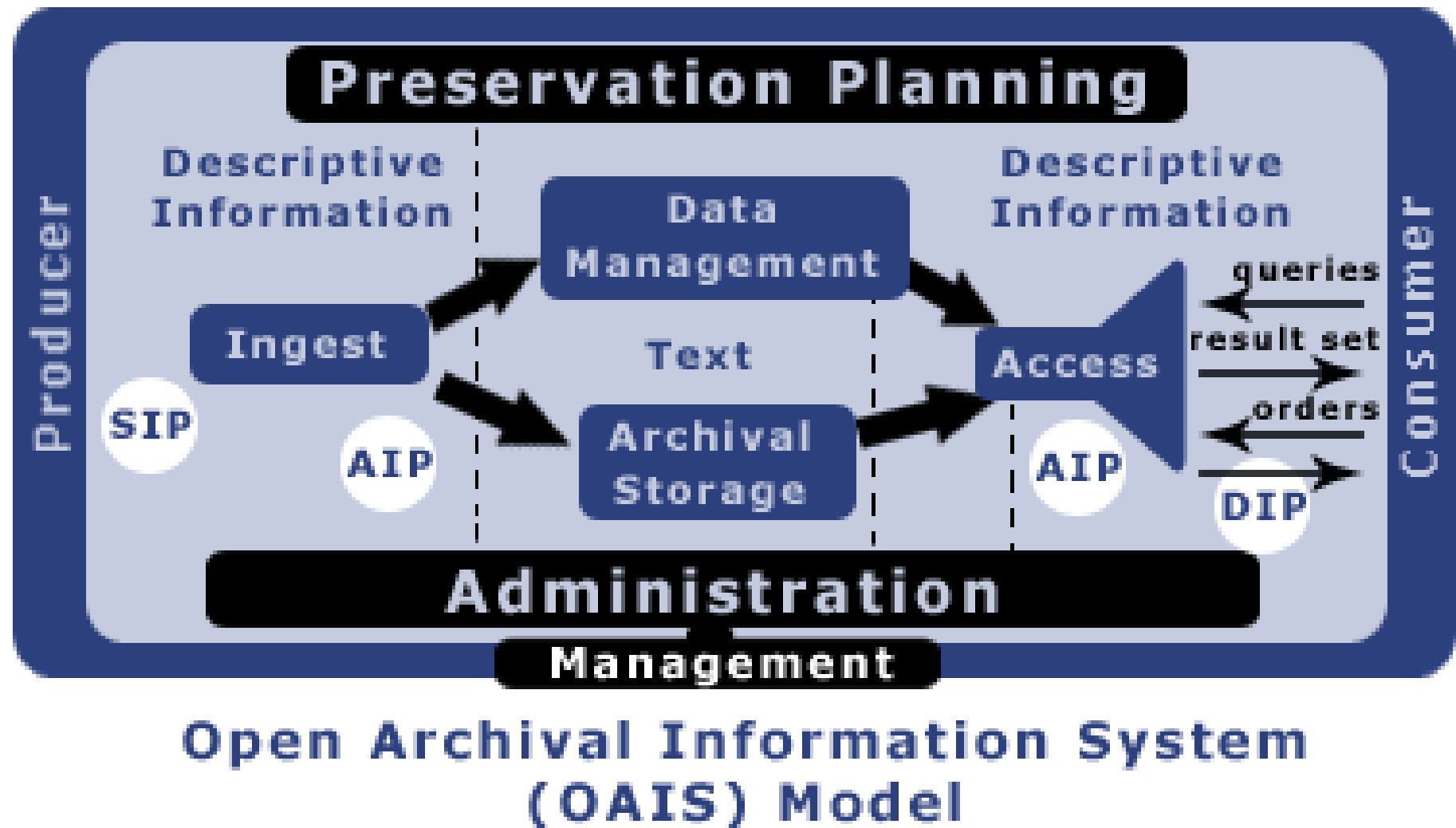
Digital Archives Border

3. Financial Sustainability

4. Technological and Procedural Suitability

5. System Security

OAIS High Level Reference Model





OAIS Functionality

Generic Features

- Ingest
- Data Management
- Access
- Common Services

Preservation-Specific

- Administration
(includes policies)
- Archival Storage
- Preservation Planning



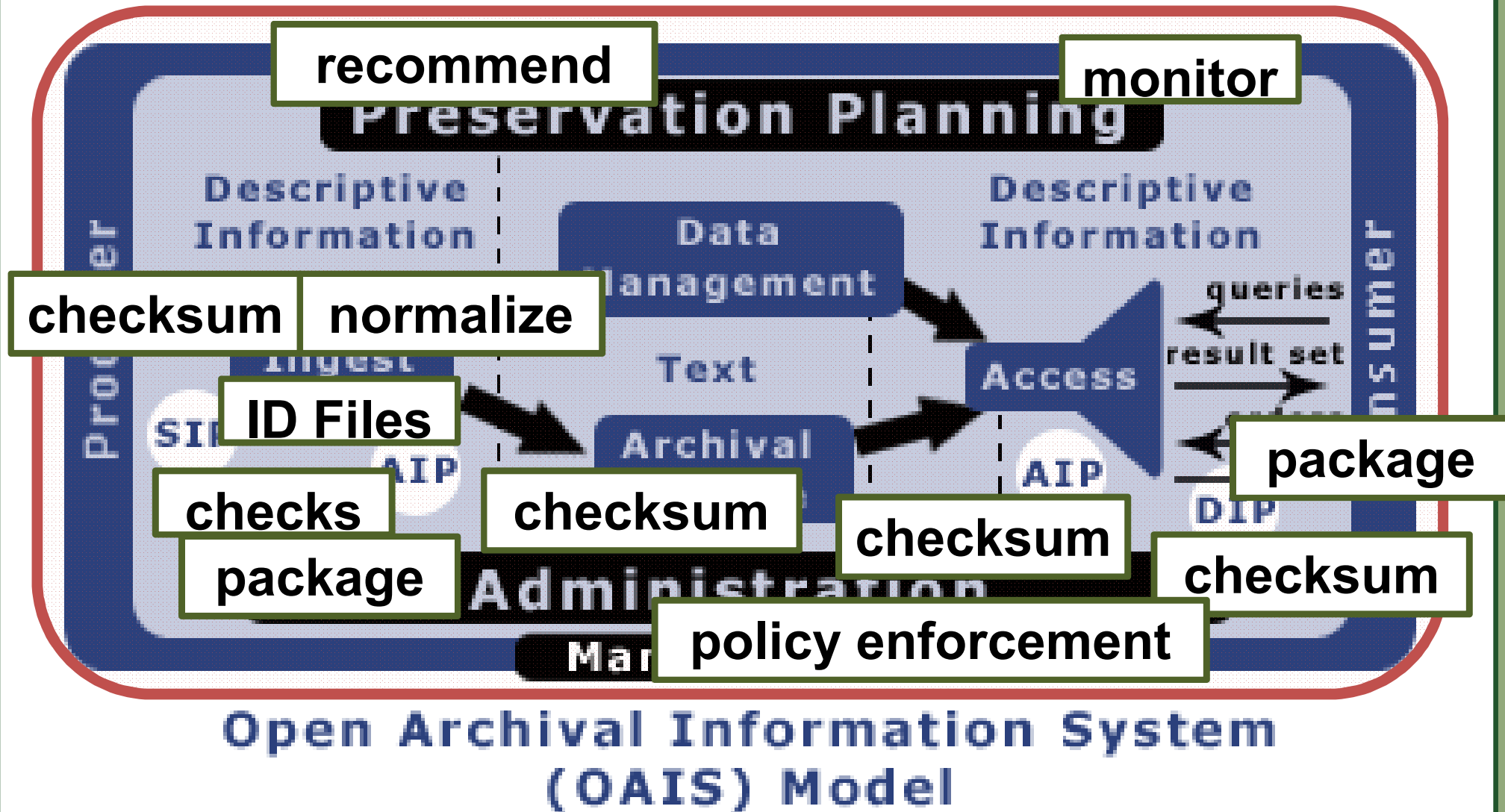
Foundation Documents

- TDR itself lacks an implementation model
- OAIS itself lacks organizational context
- Together, they leverage community-based efforts and enable collaborative initiatives

Merged Model



Tools – solving problems...





Other Standards & Practice

- Producer Archive Interface Method Abstract Standard (PAIMAS), 2005 and update pending
- NISO Building Good Digital Collections, 3.0 2007
- Preservation Metadata Implementation Strategies, v. 1.0 in 2005 and v. 2.1 in 2011
- Blue Ribbon Task Force on Sustainable Preservation and Access, 2010
- Audit and Certification of Trustworthy Digital Repositories (ISO 16363:2012)



Three Legs for Balance



Organizational Infrastructure



Technological Infrastructure



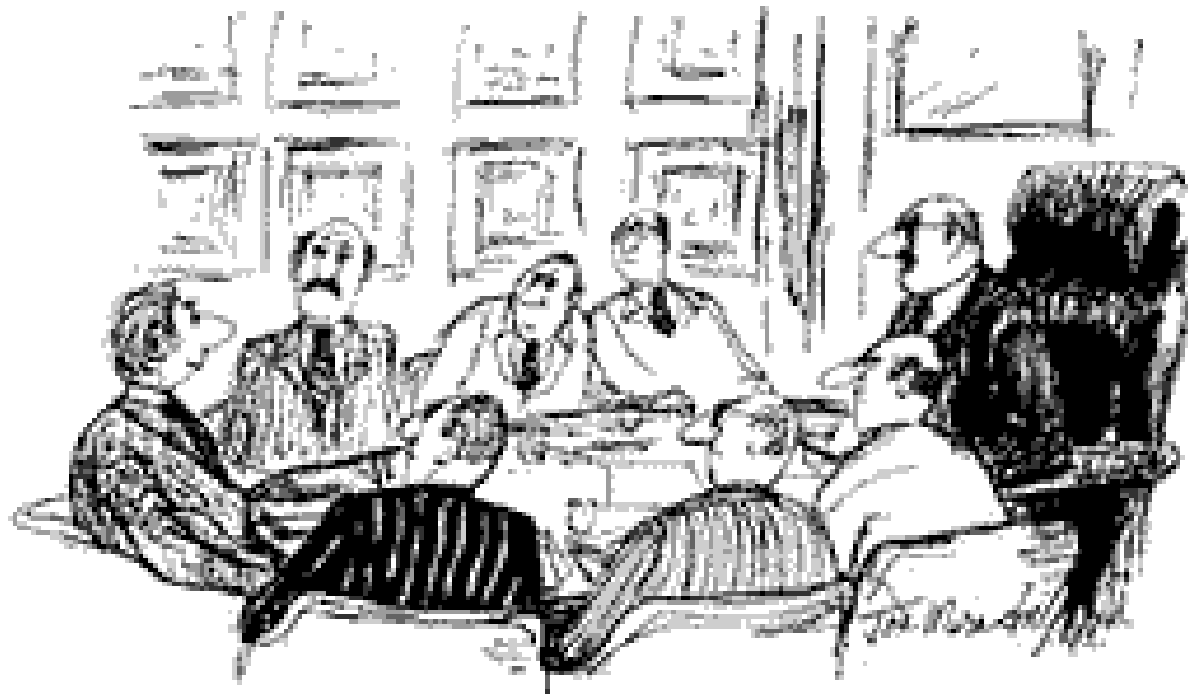
Resources Framework



Organizational Infrastructure



- Best expressed by TDR
- Best reflected in:
 - acknowledgement
 - policy development
 - implementation
 - preservation planning
 - institutional commitment



“Well, that makes it four ‘aye’s, two ‘nay’s, and one ‘hey, no problem.’ ”



Technological Infrastructure

0101

- Best expressed by OAIS
- Combination of:
 - hardware and software
 - formats and storage
 - network and security
 - functions and workflow
 - procedures, protocols, documentation
 - technical and archival skills



***“We have lots of information technology.
We just don’t have any information.”***



Resources Framework

\$\$\$\$

- No community-based articulation comparable to TDR or OAIS – yet...
 - Life³ Predictive Costing Tool (2010)
 - Blue Ribbon Task Force on Sustainable Preservation and Access (2010)
 - Upcoming JISC work on clarifying cost of curation



MANKOFF

*"No, I'm not backing up our files—I'm just assuming
that the F.B.I. is making copies."*



Five Stages for Growth

- 1. *Acknowledge:*** understanding that digital preservation is a local concern
- 2. *Act:*** initiating digital preservation projects
- 3. *Consolidate:*** segueing from projects to programs
- 4. *Institutionalize:*** incorporating the larger environment and rationalizing programs
- 5. *Externalize:*** embracing inter-institutional collaboration and dependency



Stage 1: Acknowledge



Organizational infrastructure:

often non-existent; implicit policy, or very high level



Technological infrastructure:

non-existent or heterogeneous and decentralized; disparate elements



Resources:

generally low, finite, ad hoc financial commitment



“Good news, chief, a computer virus destroyed all our documents.”



Stage 2: Act



Organizational infrastructure:

implicit policy or expressed in general terms,
increased evidence of commitment



Technological infrastructure:

project-specific and reactive; ad hoc location



Resources:

often project-based funding



“Be patient, madam. At this very moment, high-speed computers are working to eliminate or aggravate your problem.”



Stage 3: Consolidate



Organizational infrastructure:

development of basic and essential policies



Technological infrastructure:

assessment of technology investment and requisite infrastructure, shift to proactive mode



Resources:

some funding and support beyond projects, but limited



“All I’m saying is now is the time to develop the technology to deflect an asteroid.”



Stage 4: Institutionalize



Organizational infrastructure:

consistent, systematic management;
comprehensive policy framework



Technological infrastructure:

technology planning anticipates needs;
infrastructure investments planned/implemented



Resources:

sustainable funding identified for core program
areas and enhancement



“Wentworth, could I take another look at that reorganization plan?”



Stage 5: Externalize



Organizational infrastructure:

virtual organizations complement institutional ones;
collaboration inherent feature in resource planning



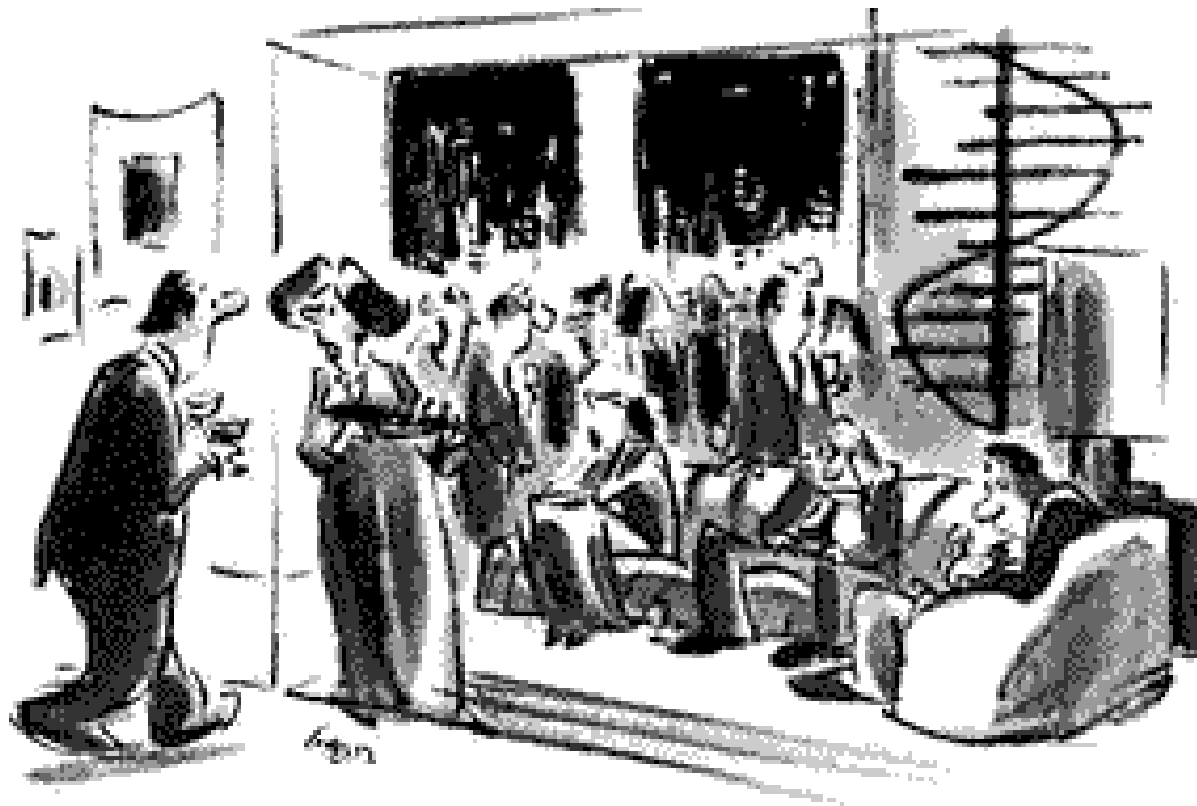
Technological infrastructure:

distributed and highly integrated; extra-organizational
features and services



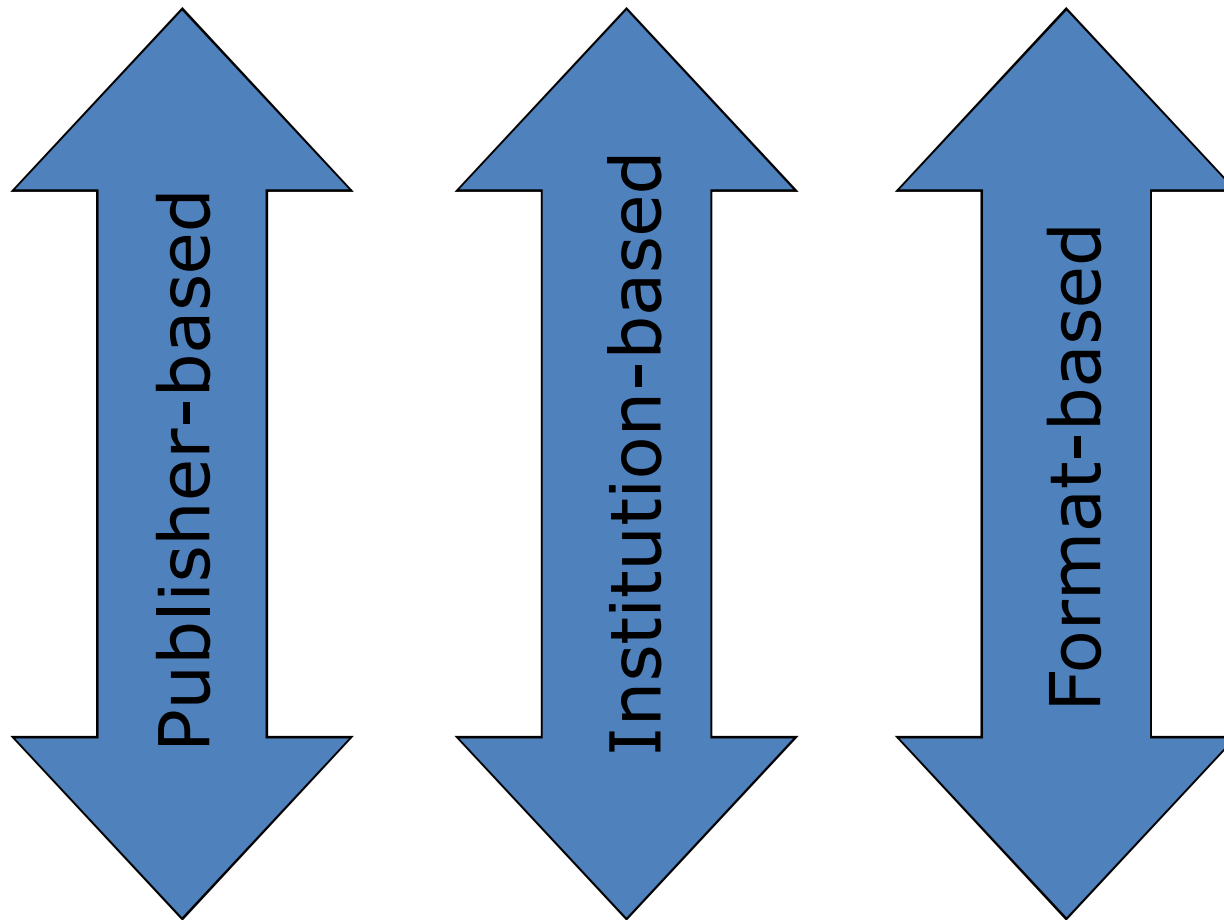
Resources:

varying levels of investment, but sustainable funding;
possibly distributed financial management

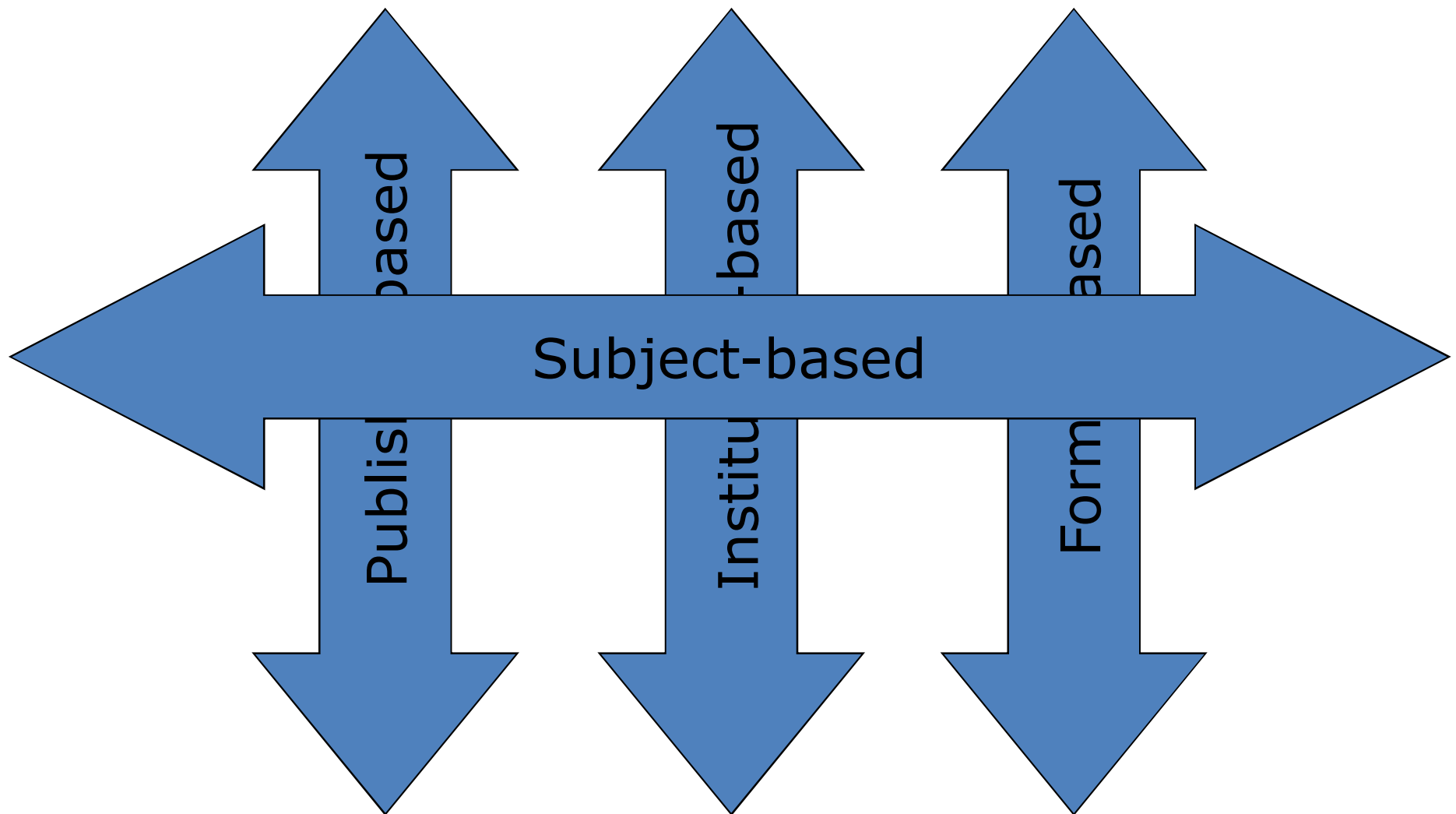


***“Oh, Lord! We forgot to invite any
content providers.”***

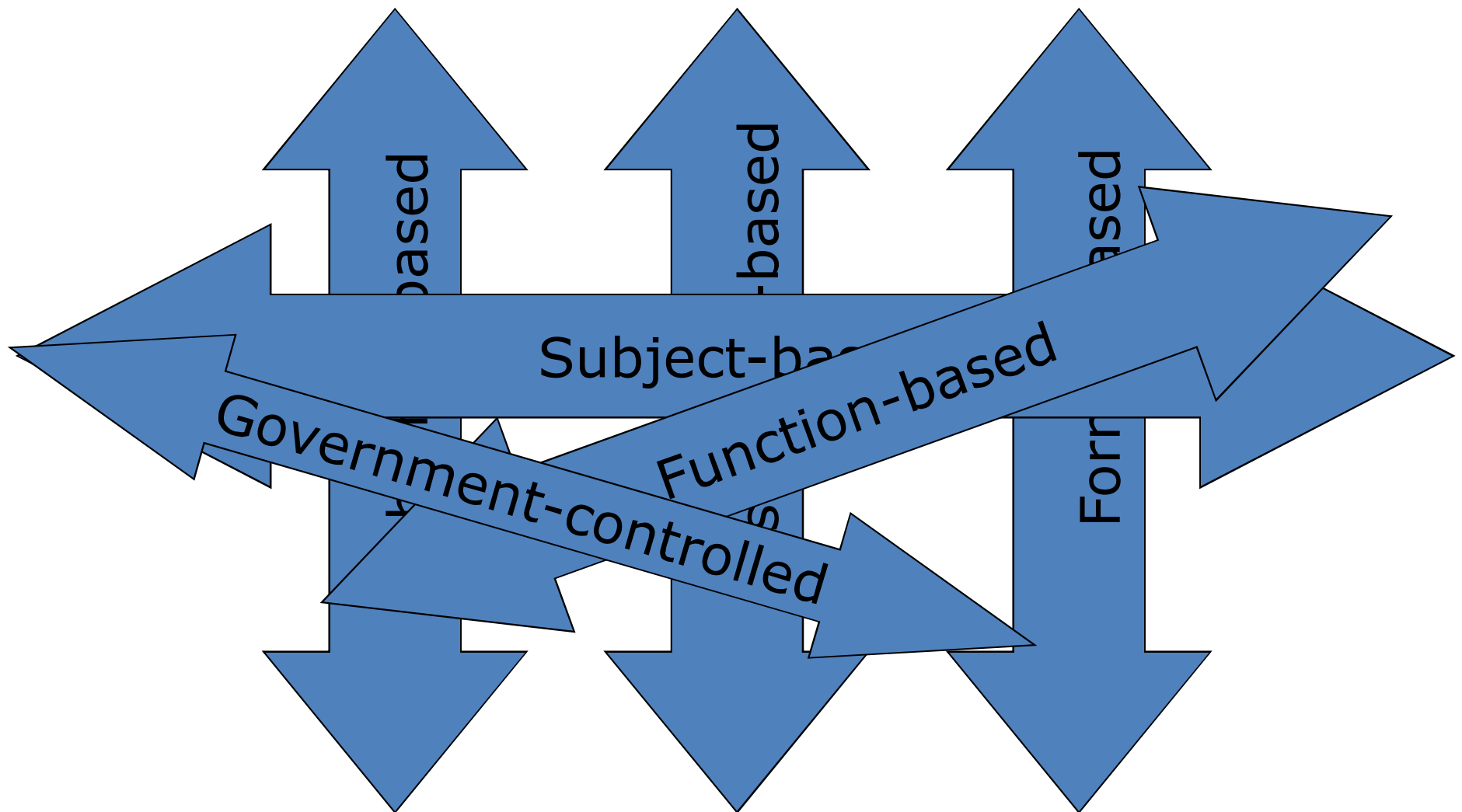
Vertical Archive Silos



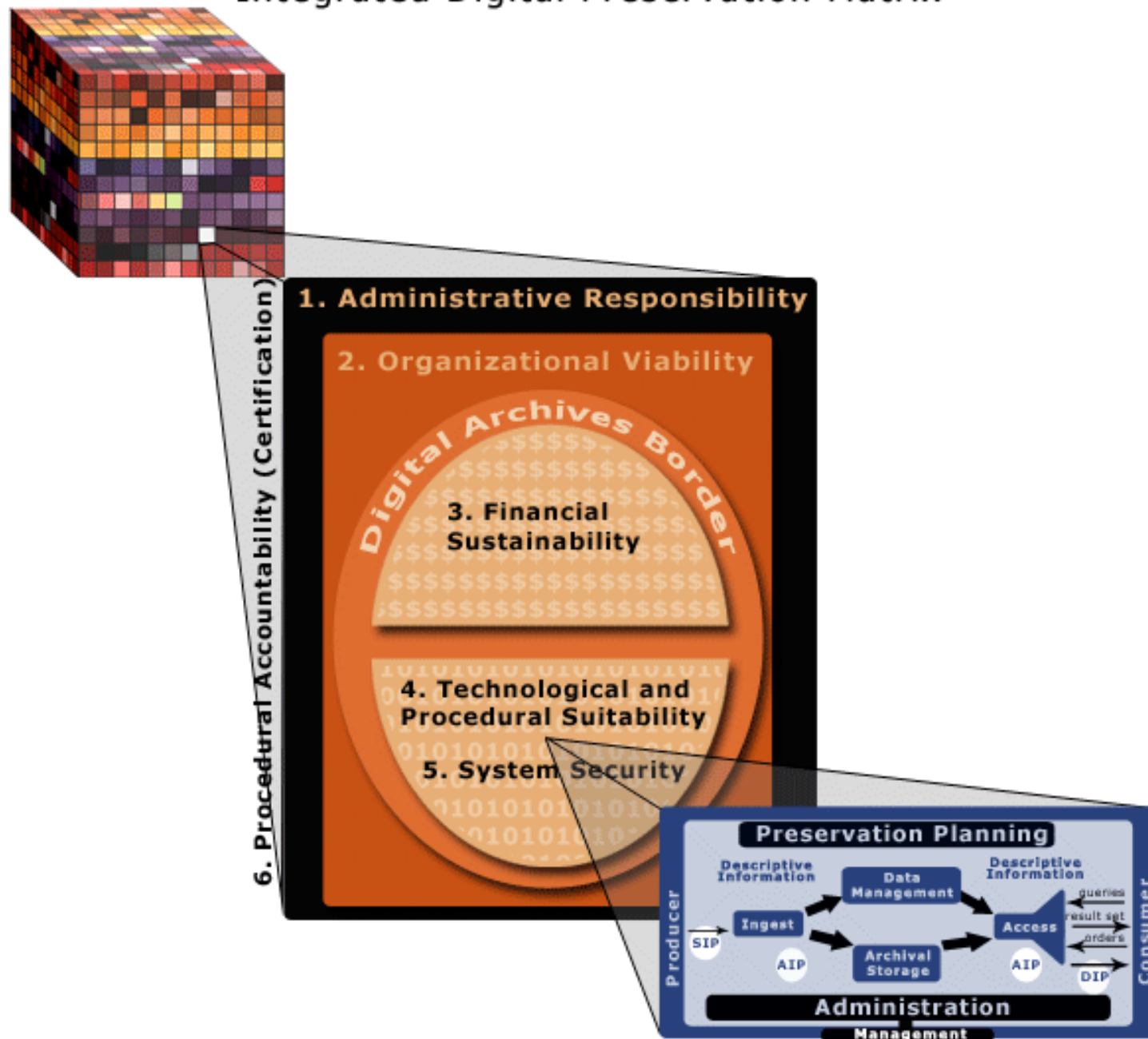
Cutting Across



From LOCKSS to Gridlocks



Integrated Digital Preservation Matrix





Extending Responsibility Beyond the Institution

- Building consortial digital archives
- Understanding interconnectedness
- Developing national programs
- Developing methods and tools for the matrix (prevent, predict, detect, respond, repair)
- Ensuring a quality control/audit process



Using the Stages

The 5 Stages:

- identify steps for **developing** an organization's digital capacity (defines a maturity model)
- provide a way of **communicating** about digital preservation development
- enable **measuring** progress towards programmatic digital preservation goals



"Faced with the choice between changing and proving there is no need to do so, most people get busy on the proof."

John Kenneth Galbraith