Foundations of Digital Curation

Tyler Walters
Dean, University Libraries
Virginia Tech

Digital Directions conference
Boston, MA, June 13, 2012

Digital Curation: We will discuss...

- Body of Knowledge

- Definitions

- Why Curate?

- DCC Lifecycle Model

- Roles & Responsibilities

- Collaborative Strategies

Much of the material for this presentation is taken from New Roles for New Times: Digital Curation for Preservation. By Tyler Walters and Katherine Skinner (ARL, 2011)

Digital Curation…

Incorporates aspects of…

- Library and information science

- Archival science

- Computer science

- Data management

- Information systems design and management

- Informatics fields (bio, geo, eco/env, health, hydro, neuro etc.)

- Domain-based information management practices

- Several others…

VirginiaTech
Invent the Future

Definition – Digital Preservation

- **Digital preservation** focuses on the "series of managed activities necessary to ensure continued access to digital materials for as long as necessary."

http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts

VirginiaTech
Invent the Future

Definition – Digital Curation

- **"Digital curation** refers to the actions people take to maintain and add value to digital information over its lifecycle, including the processes used when creating digital content**"**

- *DP and DC are highly complementary concepts; DC must facilitate preservation of digital objects. Curatorial actions must serve needs of current and future users*

- http://www.dcc.ac.uk/digital-curation/what-digital-curation

VirginiaTech
Invent the Future

Why Curate?

Somebody has to do it!

- <u>Most content creators</u>:

- Are not taught DC / content management
- Don't know what metadata are
- Can't name services, data centers, or repositories
- Don't always share content publicly or store it in some manner of an archives
- Aren't convinced they should always share their content

VirginiaTech
Invent the Future

Why Curate?

- Question of <u>authenticity</u> (changes)

- Question of <u>provenance</u> (custody)

- Uncertainty of <u>rights</u> (usability)

- Need for <u>documentation</u> (metadata)

- Need for <u>info. organization</u>, findability (share)

- <u>Relationship management</u> between different versions of static and dynamic content (versions)

- Preserve <u>cultural memory</u> for future (learning about human societies, activities)

- <u>Research</u>: verify/reproduce results, reuse for future research (advance knowledge)

DCC Lifecycle Model

Provides a graphical high level overview of the stages required for successful curation and preservation of data.

It can be used to plan activities within an organisation or consortium to ensure all necessary stages are undertaken, each in the correct sequence.

1. Full Lifecycle Actions
2. Sequential Actions
3. Occasional Actions

http://www.dcc.ac.uk/lifecycle-model/


WⅢ VirginiaTech
Invent the Future
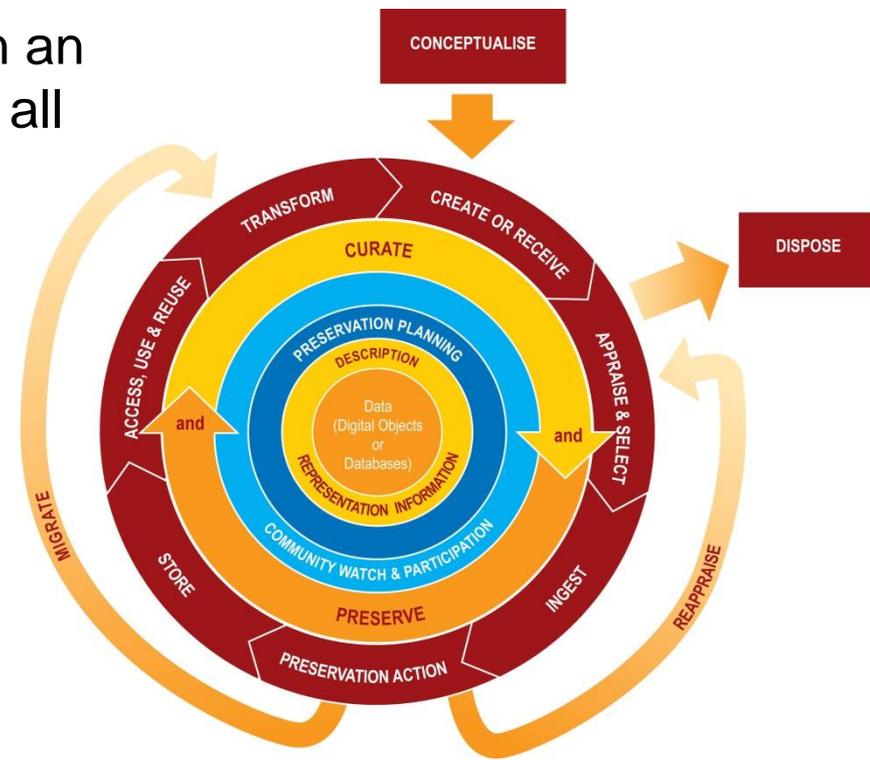
DCC Lifecycle Model

Researchers and content creators tend to focus on:

- conceptualise
- create or receive
- ingest
- store
- access, use and reuse
- data
- description
- community watch and participation
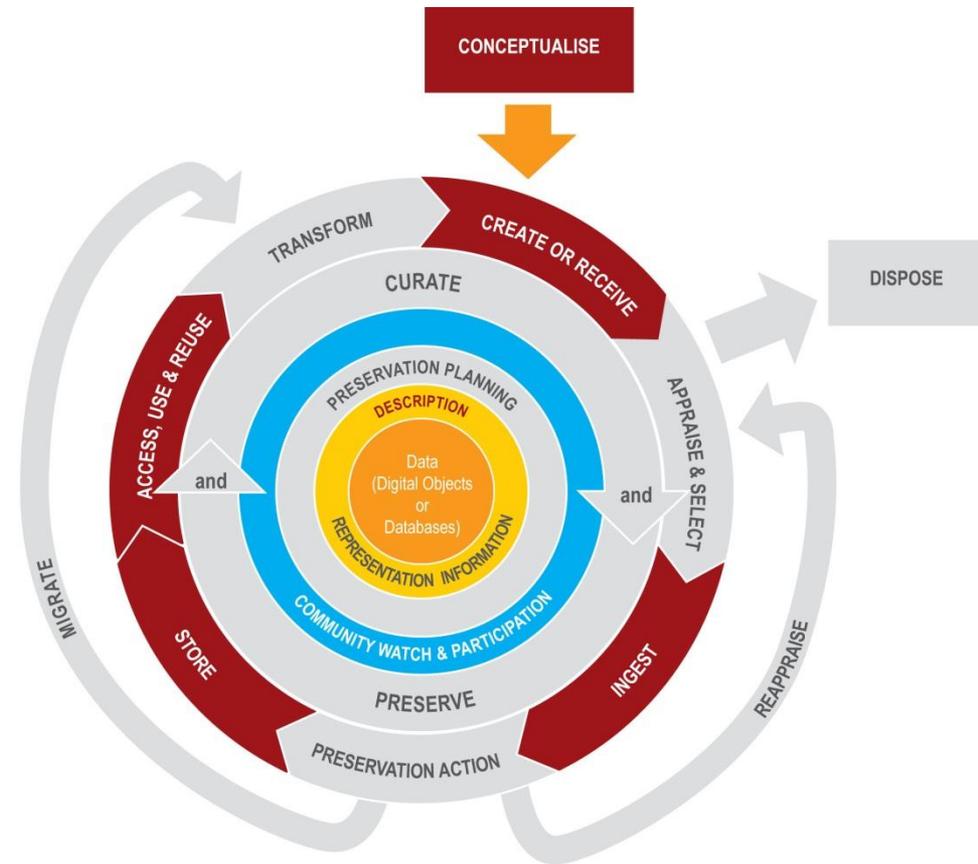
DCC Lifecycle Model

Researchers and content creators
tend to focus <u>less</u> on:

- appraise and select

- dispose

- preservation action

- transform

- representation information

- preservation planning

- curate and preserve

- migrate

- reappraise

VirginiaTech
Invent the Future
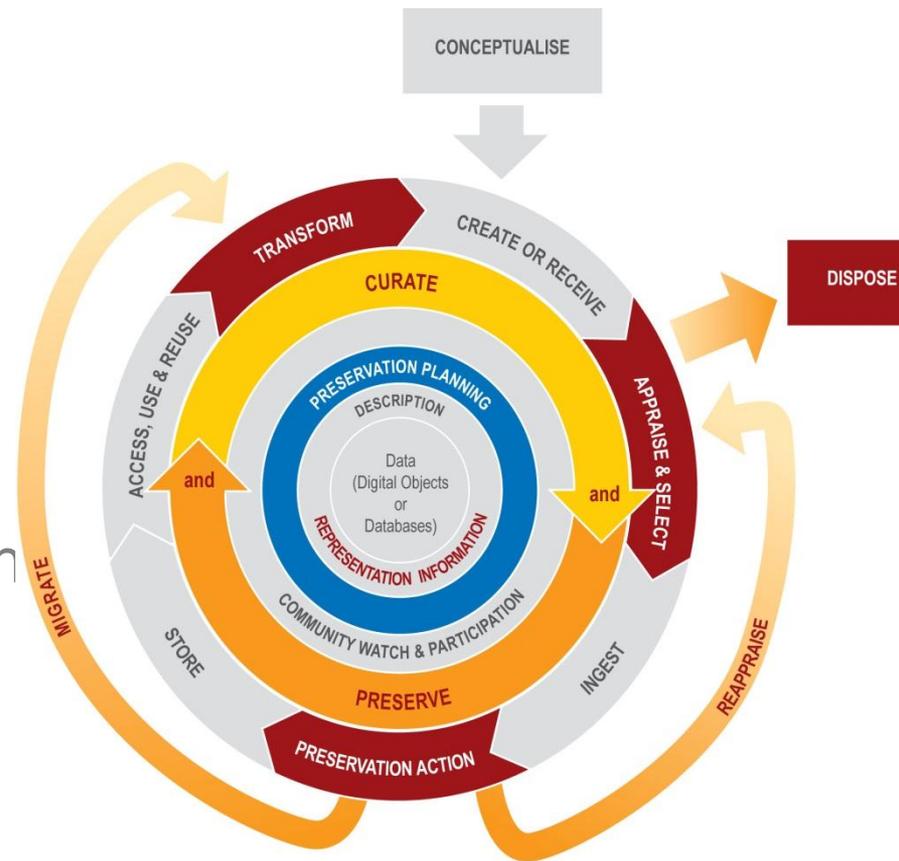
CORE SKILLS FOR DATA MANAGEMENT

A follow-up from the second DCC Research Data Management Forum (November 2008)

DATA MANAGER
- Risk / Disaster Management, Contingency
- Data Legislation (rights, IPR, DP, etc)
- Data Security, Access, Authentication
- Conditions of Use

DATA CREATOR
- Monitoring Processes
- Metadata
- Documentation (research, environmental, temporal) Context

- Value of Data, Economic Issues
- Data Preservation
- Facilitation / Communication
- Data Modelling

DATA LIBRARIAN
- Negotiation Skills
- Complaints and Expectation Management
- Coordination of Practice across Institution
- Data Appraisal and Retention
- Advocacy, Promotion, Marketing, Raising Awareness
- Standards Development

DATA SCIENTIST
- From Information Management to Knowledge Management
- Data Analysis and Manipulation
- Extracting Information from Data Models (and People)
- Merging, Mash-ups, Integration

Digital Curation Centre: Roles. http://www.dcc.ac.uk/resources/roles

Virginia Tech — *Invent the Future*

# Not 'Public' and 'Technical' but rather…

A New Triad:

Cyberinfrastructure:
- Includes policy

Content
- Licensed
- Purchased
- Donated
- Locally generated

Services
- Information production
- Access and dissemination
- Long-term curation / preservation
- Rights

Infrastructure
(facilities, technologies, expertise)

Content
(external & internal resources)

Services
(information services – info. production & use)
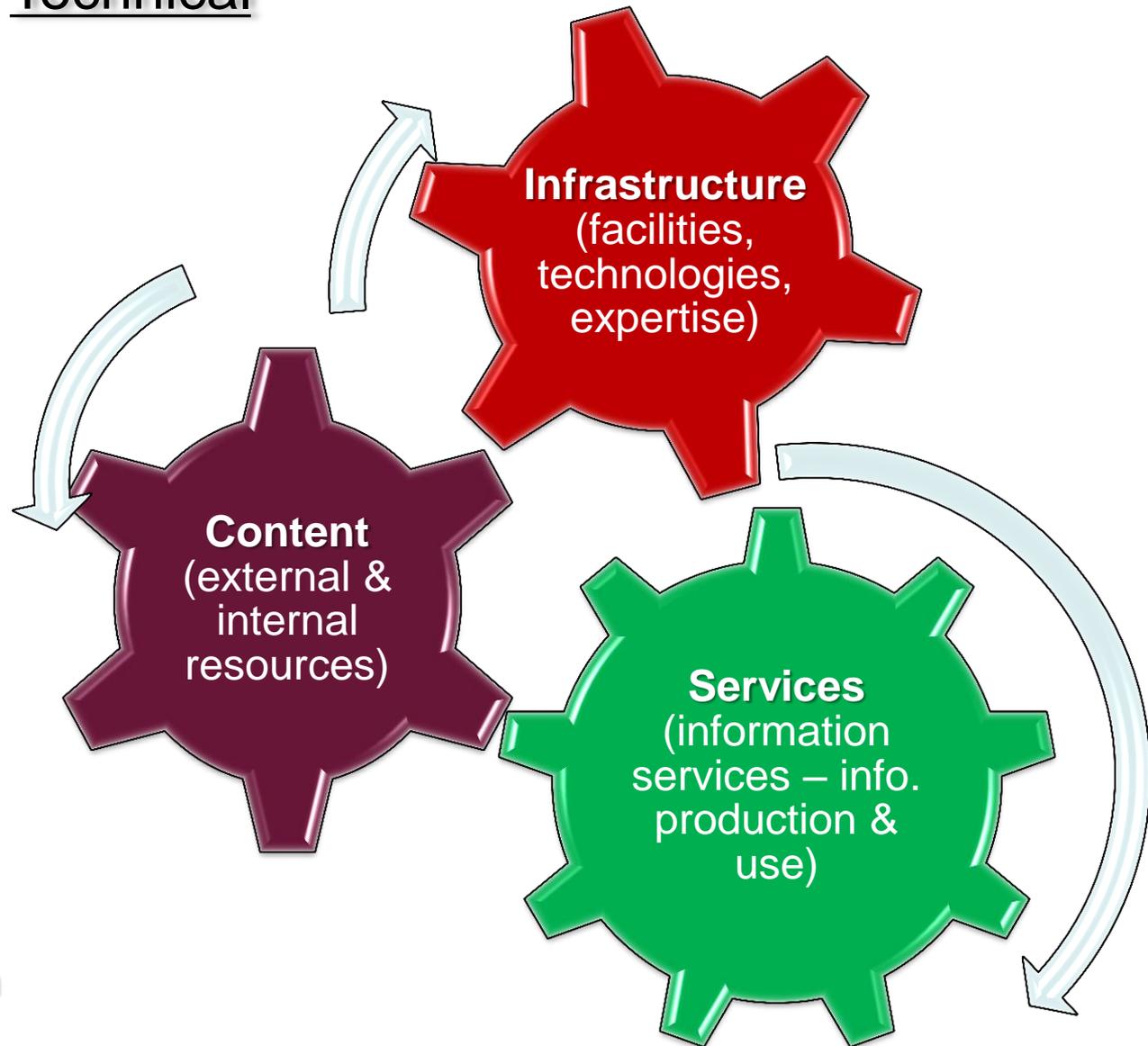
New Roles: Cyberinfrastructure & Collaboration

CI: Technologies, policies, people involved in supporting digital and internet-based research and learning

- Enables researchers to communicate efficiently

- Ability to create and disseminate their work without inventing new tools every time they undertake a new project

- Hard to build stable CI unless we collaborate and invest in community-generated solutions, rather than insisting on building individualized workflows and systems (wont interoperate)

- Working with other insts. can deliver stronger CI and collections that can work interoperably, and do not depend unduly on a single institution

Content Acquisition & Hosting

- Offering reliable digital curation/preservation services in support of the content production activities taking place

- Achieve goal through hosting broad range of content:

  - Digitized collections (manuscript / archival)
  - Licensed content (journals, databases)
  - Acquired web-based materials (web archiving)
  - Research data, primary source material generated by parent
  - Published intellectual output (papers, books, ETDs),
  - Instructional / multimedia materials, digitally captured lectures symposia, and other parent institution events

Digital Publishing, Curation, and Preservation Services

- New roles in content acquisition / hosting are complemented by new roles in content production, curatorial, preservation activities

- Significance of creating strong partnerships between curating inst. and community it serves cannot be overstated

- Libraries well positioned to provide e-publishing options for parent institution

- Experiments with e-publishing successful, particularly with open access journals, repository-based publishing, and digital scholarly resources in the humanities and social sciences

Roles & Responsibilities:
Content Producers & Disseminators
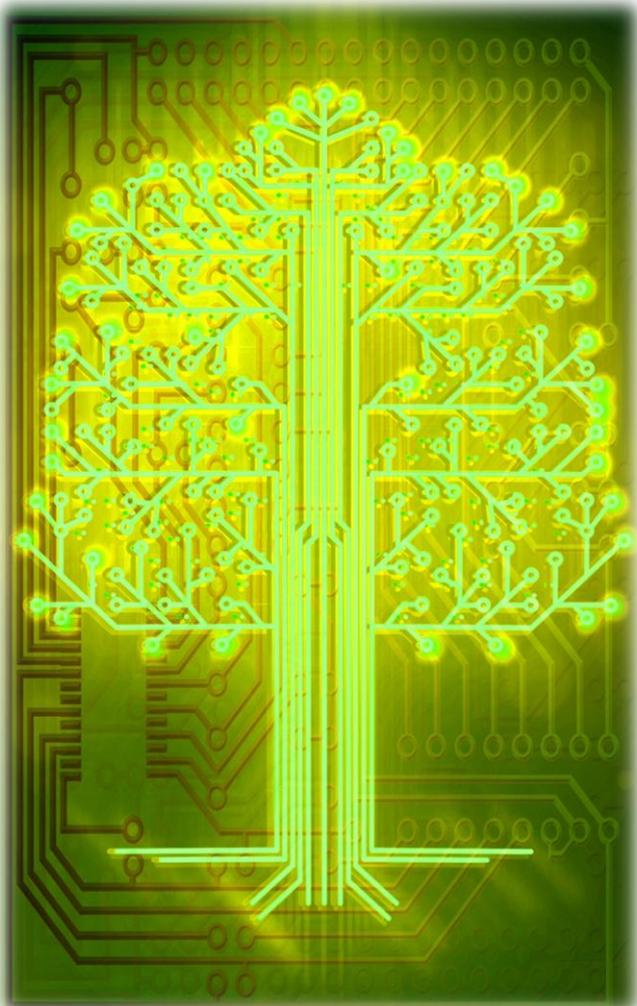
New role of Digital Curation is emerging — or, attention to the lifecycle management of digital objects and collections

o   Historical focus: access and preservation

o   Access-oriented realm: libraries are managing more information than ever before, ultimately assist users in accessing & disseminating content

o   Preservation-oriented realm: long-term stewardship of digital content. Services evolving in response to growing concerns about the viability, authenticity, and sustainability of digital content

o   Production-oriented realm: curating insts. provide services, training, consulting for content producers on production processes, issues

Content Selectors & Rights Advisors

Content Review

Quality, relevance, documentation to justify curation.

Issues of stability, accessibility, authenticity

Acquisition Policy

With legal counsel, review policies to ensure sufficient rights are transferred to effectively curate/ preserve content

Rights Management

Employing rights mgmt specialists rather than relying on the parent's general counsel.

Increasing IP issues require info. mgmt. expertise as well as legal expertise

Instructional Partners & Consultants

This role includes physical & virtual environments. Both have been directly impacted by growing information mgmt. needs of digital scientific, social science, record, and cultural heritage materials

- Librarians and other instructional specialists are increasingly involved in parent inst. engagement:
  - New instructional design and class partnerships
  - Various help services in virtual and physical environments
  - Continuing education-oriented outreach to the local community

- With research data, curating insts. are developing Data Literacy programs, teaching producers about curation techniques

Ethnographers of Information Producers and Users

Librarians are spending time "living among the natives" in their parent insts. and studying their information production and consumption habits. They are "participant observers"
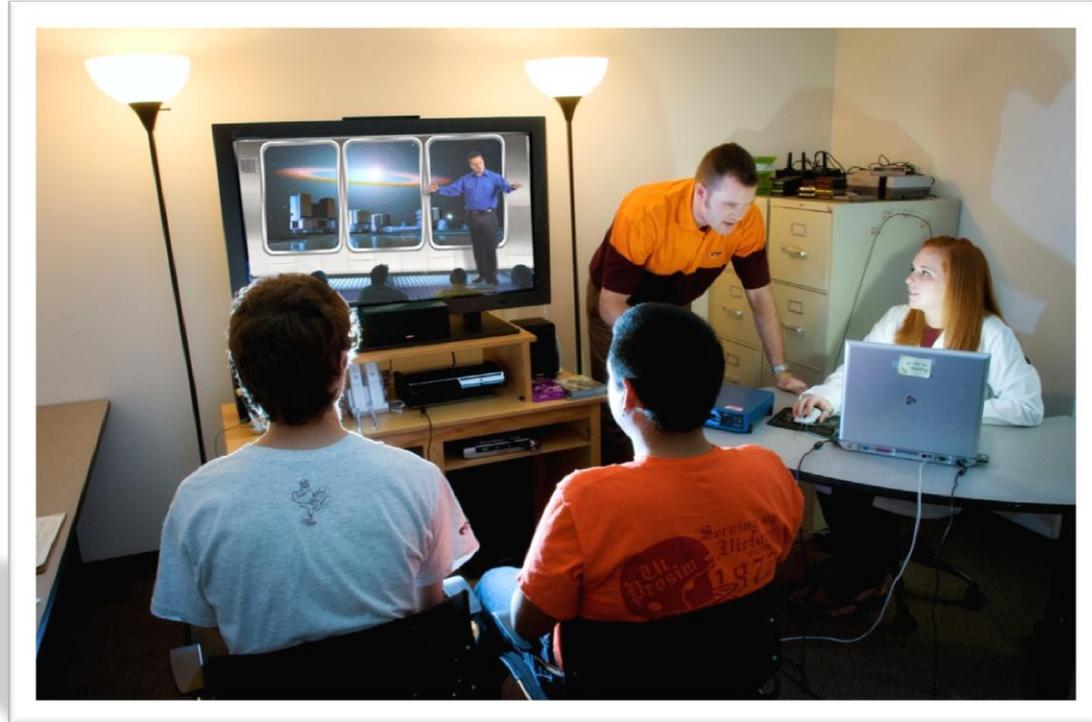
- If curating insts. hope to meet the future needs of their communities, they must know their constituents—from administrators and staff to professors and students
- As technology continues to develop at a rapid pace, the practices and needs of parent institutions will likewise quickly shift

Systems Builders

This role is about designing and architecting environments:

o Physical spaces such as information / learning commons

o Virtual spaces, such as websites, digital repositories, and other information systems

o Discrete system-building activities such as licensing, metadata creation, and copyright research

o It also includes providing content producers with new tools and processes for publishing their works such as e-books, e-journals, database-driven web sites, as well as more specialized digital resources

Organizational Designers

- **Provide for the agility and experimental atmosphere necessary to transform an institutional structure to fit its new terrain**
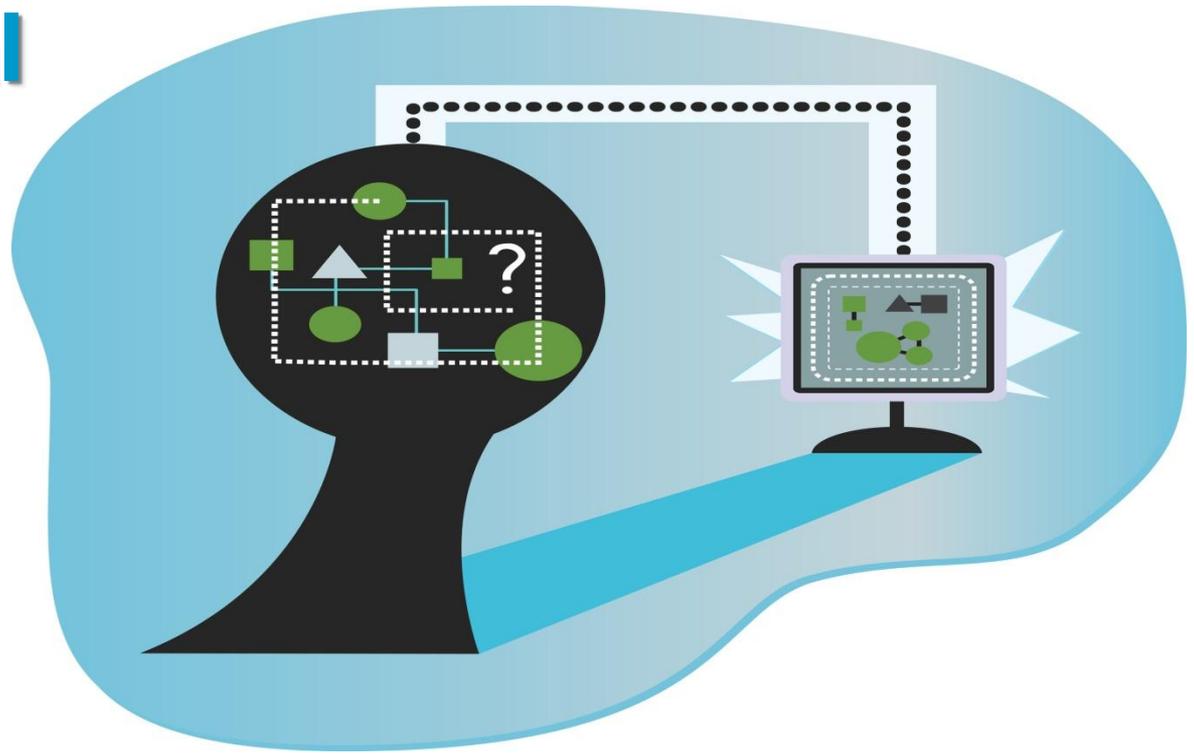
- Experiment with new organizational forms, distributed and decentralized inter-institutional models of cooperation. Imperative for digital curation.

- Designers must 'design in' paths to help existing employees find ways to adapt and learn new skills

- Redefining how research libraries organize their work in ways that better fit the digital and physical landscapes they balance

Collaborative Network Creators and Participants
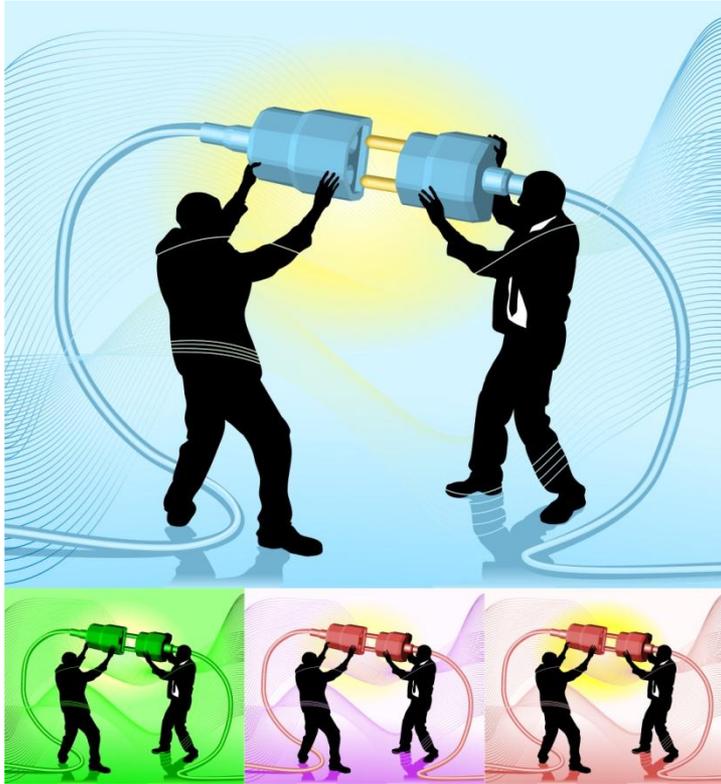
- Curating insts. are forming collaborative entities that reach well beyond grant-project timeframes

- Fostering long-term, collaborative, sustainable practices to support common needs, including technology development, repository management, and preservation

- Coordinating intra- and inter-institutional alliances, maintain clear documentation about roles and responsibilities for each institution in these multipronged efforts

They are learning how to forge healthy collaborative relationships through credibility, reliability, trust development, and self-awareness

Collaborative Strategies for Digital Curation / Preservation

- Wherever possible, collaborative or community-based approaches to digital curation are likely to be more effective and sustainable for a variety of reasons
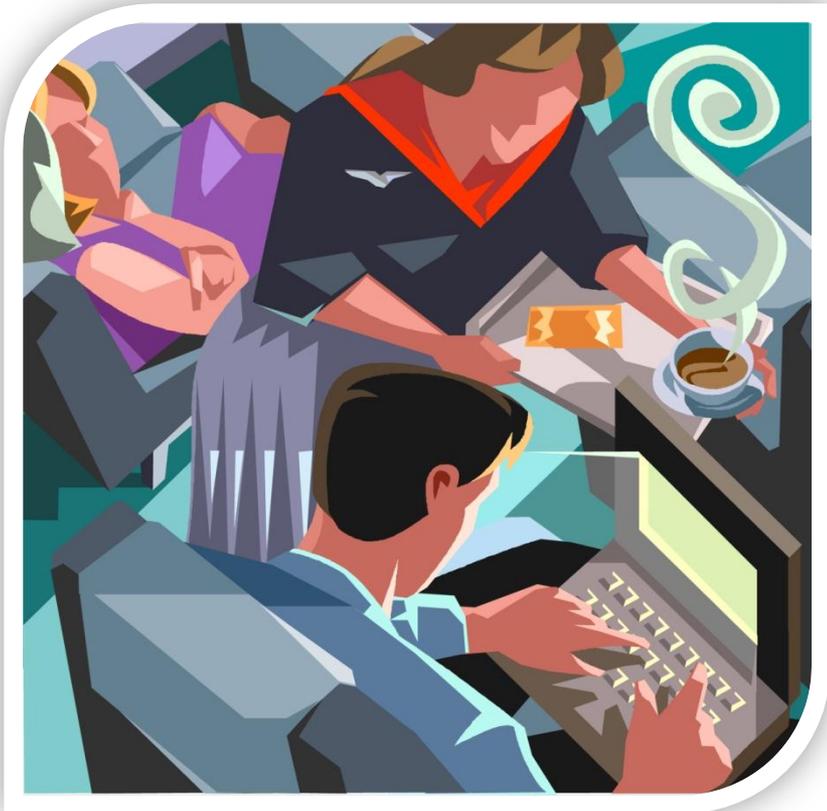
However, effective use of collaborative strategies and community-managed resources requires local investment and capacity building

Collaborative Strategies

Recommendations （6）:

- **Stop waiting, start local proactive relationships**
  - Talk with content producers, policymakers
- **Stake a claim in the production cycle**
  - Consider content production, publishing, hosting
- **Start retraining and repurposing staff**
- **Be a doer, not a broker, wherever possible**
  - Are we outsourcing a core purpose?
- **Consider digital curation collaborations**
  - Who else is facing similar challenges, needs?
- **Actualize collaborative engagement**

VirginiaTech
Invent the Future

Thank you

tyler.walters@vt.edu

540-231-5595

tywalters1 Skype / Twitter / FB