

Preserving Things that Count: exploring partnerships among domain specific repositories, institutional repositories, and social science researchers

iPRES 2006

Ann Green and Myron Gutmann

ICPSR

Inter-university Consortium for Political and Social Research

Overview

- Characterize the parallel life cycles of social science research and social science data
- Describe key elements of two different kinds of repositories
- Recommend how researchers and the two different kinds of repositories can forge partnerships to support long term access to social science data

What's the trouble in the story?

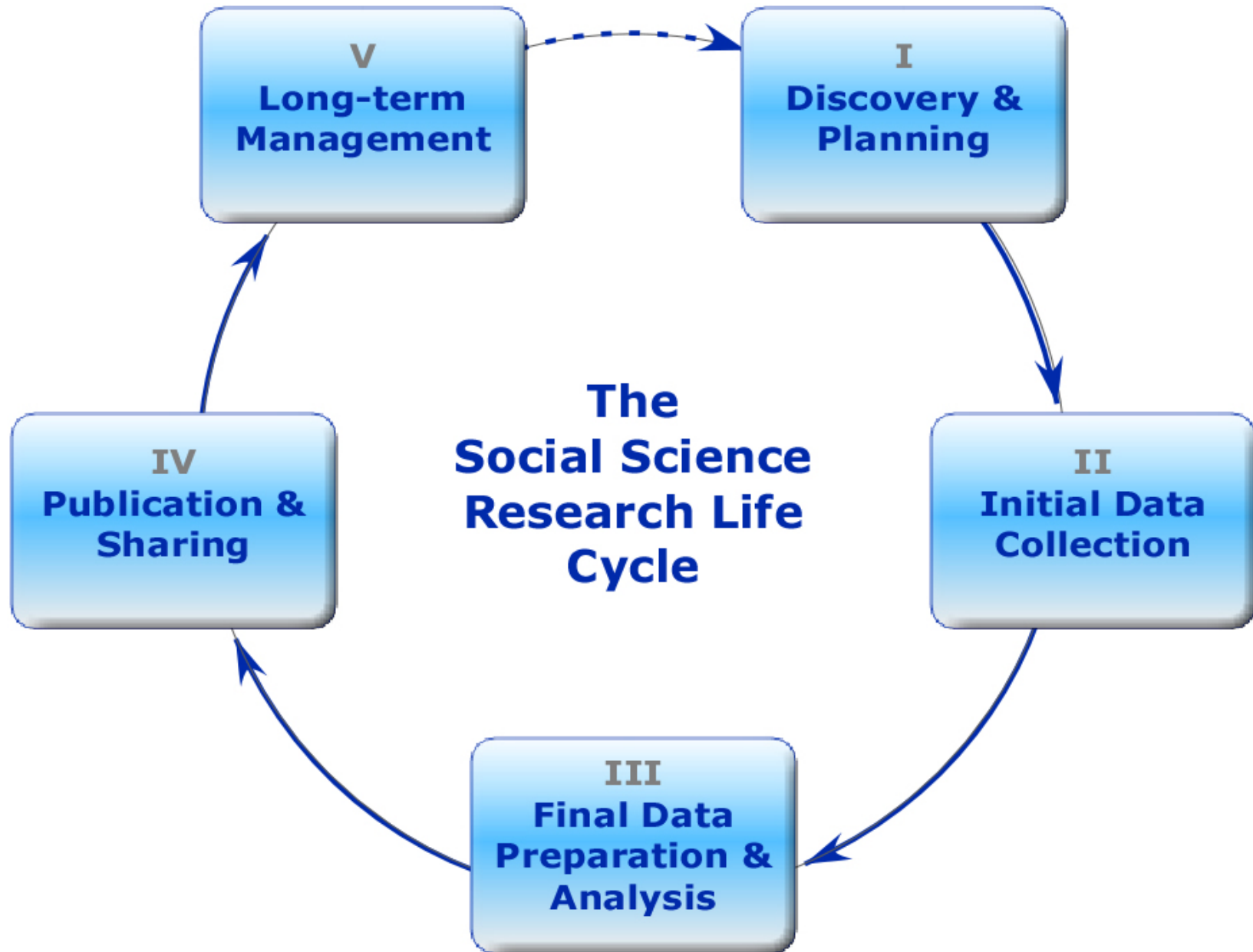
part one

- Data are being lost – from personal websites, hard drives, thumb drives, learning management systems, survey production systems, email attachments.
- Data archiving requirements are not being met; support and guidelines are not in the hands of researchers.
- Data documentation is not being produced at critical moments; undocumented or poorly documented datasets are useless.

What's the trouble in the story?

part two

- “Self-archiving” is at the heart of most e-print Institutional Repositories, yet self-archiving of data is problematic:
 - Incomplete and non-standard documentation
 - Proprietary file formats (e.g. SPSS, Stata, SAS)
 - Issues of confidentiality and disclosure avoidance may block data sharing, even though public datasets could be created
 - Most preservation oriented lifecycle models begin at the Acquisition stage. We propose a lifecycle model that starts at the beginning of the research life cycle.
- Repositories are not synchronized; decades of domain expertise are not being tapped
- Narrow institution focus and branding of digital assets can run counter to research activities that cross institutional boundaries



DDI Social Science Data Life Cycle Model

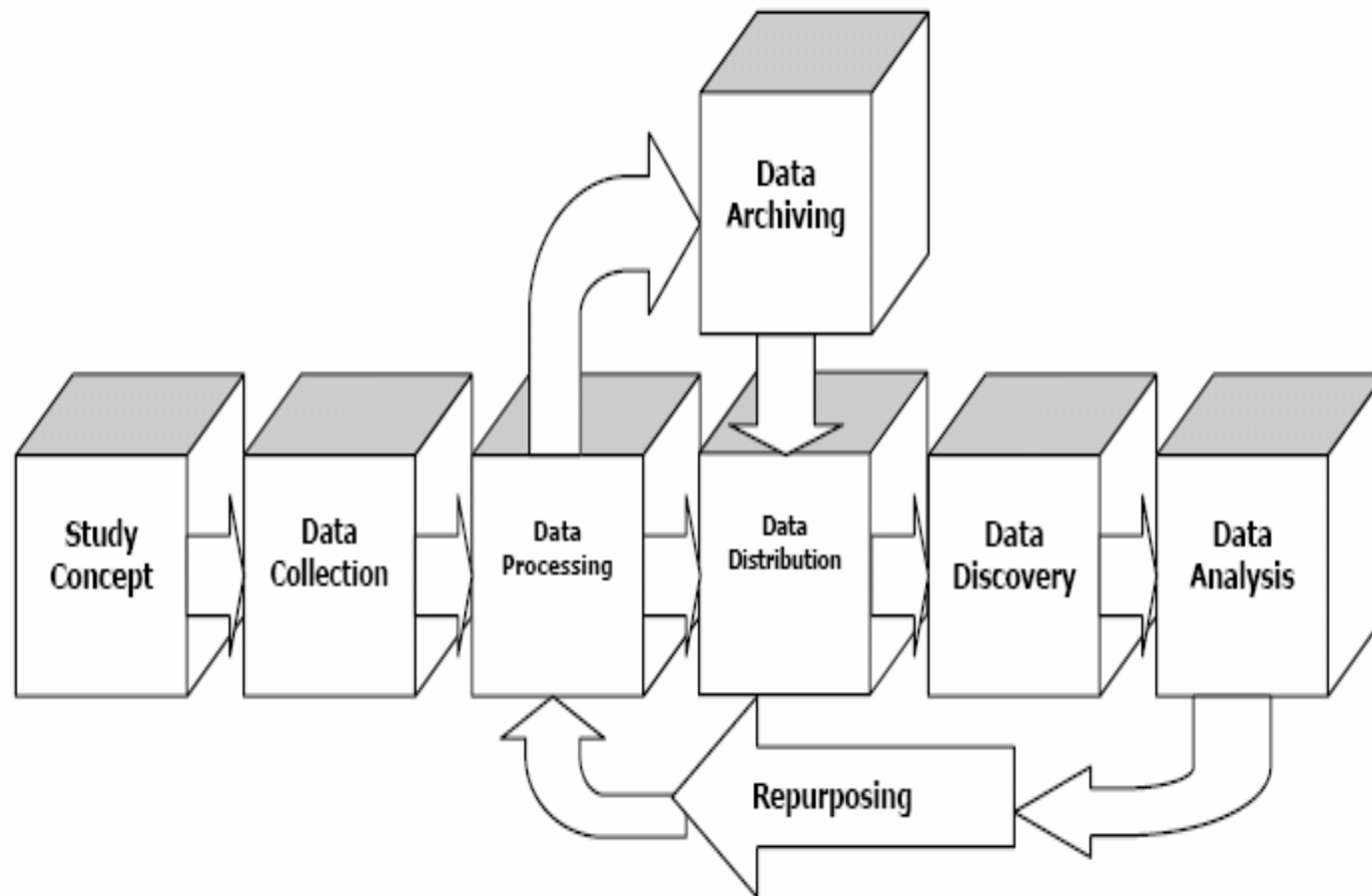
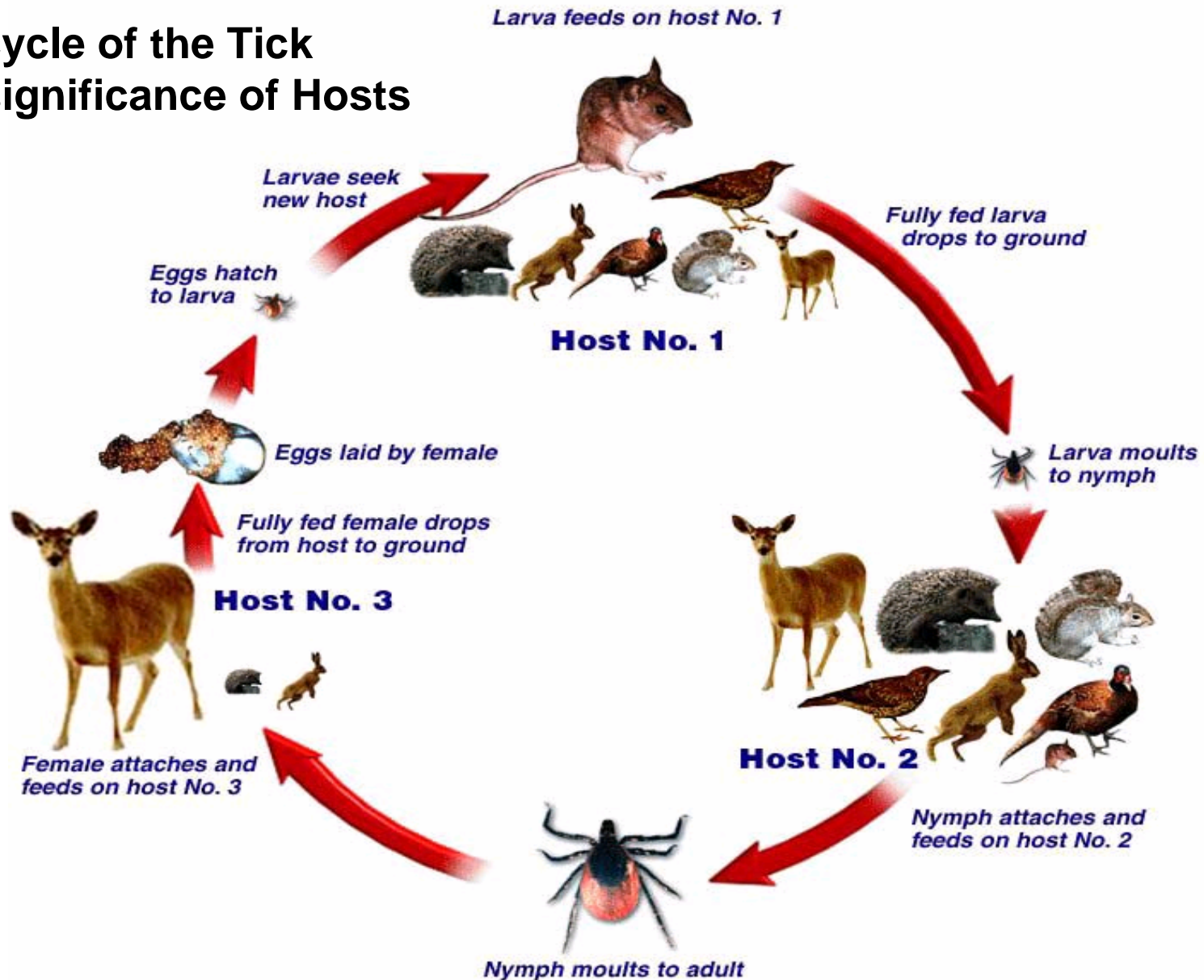


Figure 3: Combined Life Cycle Model

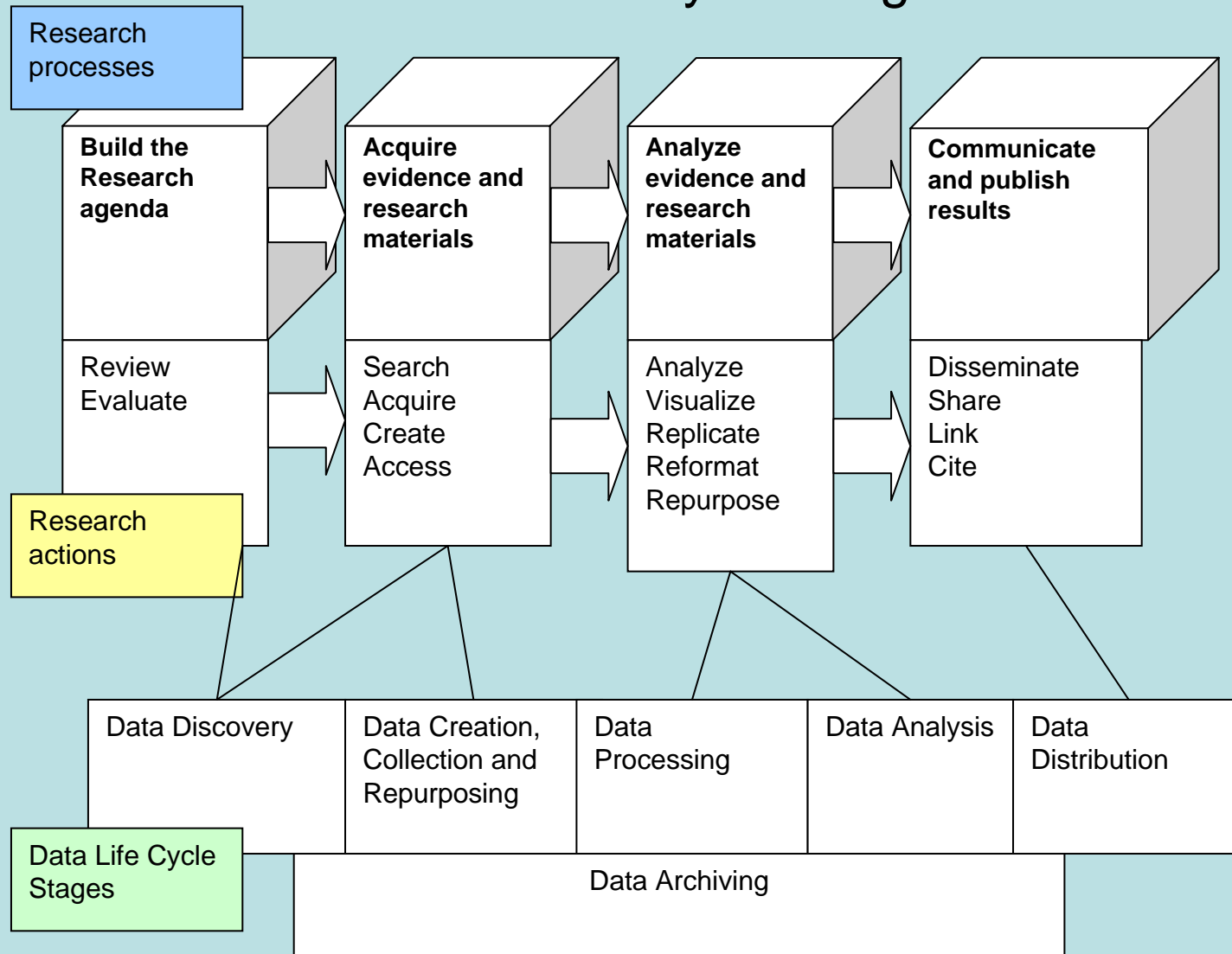
Life cycle of the Tick and significance of Hosts



The relative size of the animals approximates their significance as hosts for the different tick life cycle stages in a typical woodland habitat.

Courtesy of Dr Jeremy Gray and Bernard Kaye

Social Science Research Processes, Research Actions, and Data Life Cycle Stages



Domain Support

throughout the research life cycle

- Services provided by domain experts:
 - Knowledge of how particular datasets fit into the overall domain knowledge base
 - Data documentation: standards and production
 - Methods of disclosure avoidance, privacy and security
 - Support for discovery and reuse of data
- In the social science domain, this network is connected internationally through IASSIST and IFDO, in Canada through CAPDU, and in Europe through CESSDA

Institution-based digital repositories

Collect, manage, and maintain the intellectual output of an institution.
Focus primarily on text and images; eprints model.
Acquire content through 'self archiving' mechanisms.

Strengths:

- Local contacts and support services
- Open access goals
- Recent attention brings funding and staff resources – great progress made
- Preservation commitment may be strong, but varies across institutions

Domain-specific digital repositories

Types:

- Data libraries in single institutions
 - Acquire and manage data to support the research and learning activities of the institution
 - Extensive support services in locating, evaluating, and using data
- Topical data archives and government based data archives
 - CPANDA, American Religion Archive, NARA, etc.
- Social Science data archives
 - ICPSR, Roper Center, Odum Institute, Murray Center, etc.

Strengths:

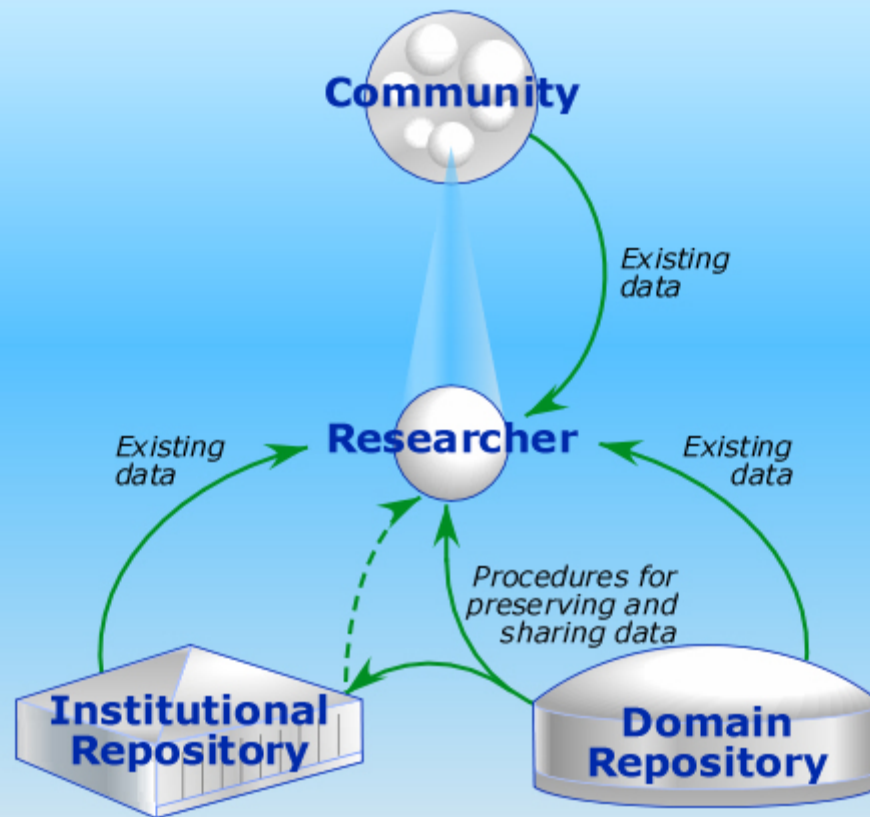
- Knowledge of content, file types, migration and emulation strategies, research requirements, analytical techniques
- Preservation commitment is as important as access
- Support services well established; network of professionals
- Commitment to supporting the quality of social science research; active at national and international levels of funding, infrastructure planning, and dataset production
- Training and community building across institutions, faculty, graduate students, other archives
- Not controlled by local agendas

Seeking partnerships in the repository landscape

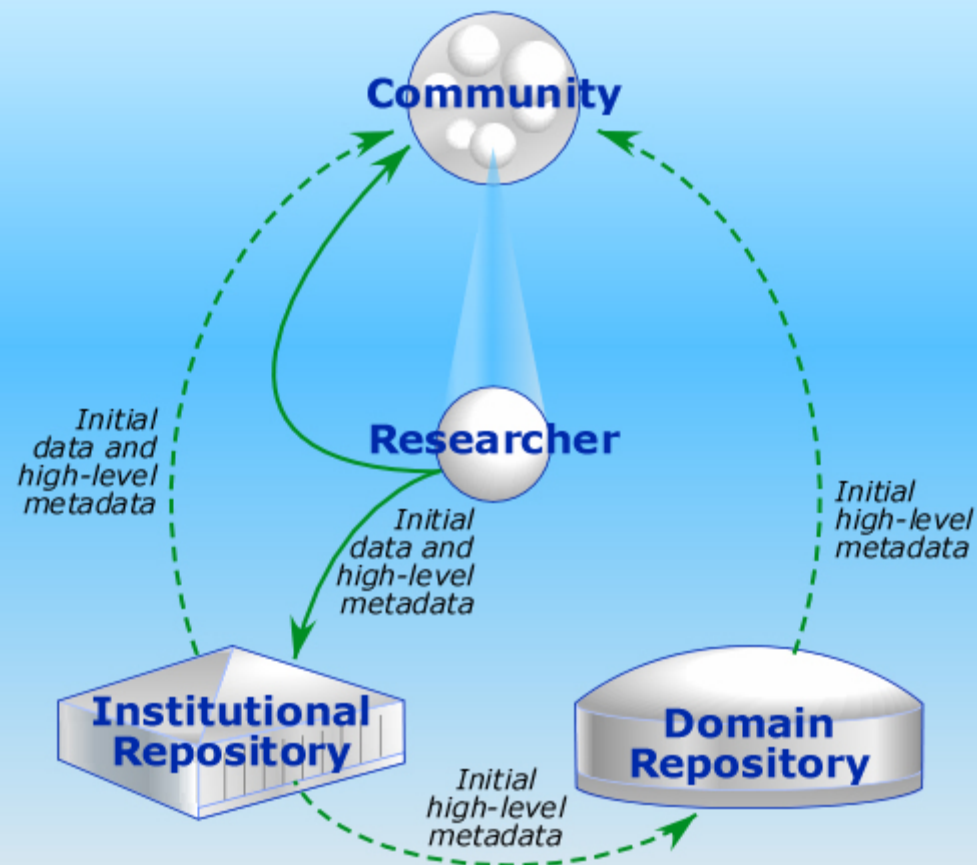
This approach envisions both cooperation and specialization.

- The researcher produces the scientific product, both data and publications;
- the institutional repository has specialized knowledge of campus conditions and the opportunity to interact frequently with the researcher;
- the domain-specific repository has specialized knowledge of approaches to data in a specific scientific field, for example domain-specific metadata standards, as well as the ability to give high-impact exposure to research products.

I Discovery & Planning

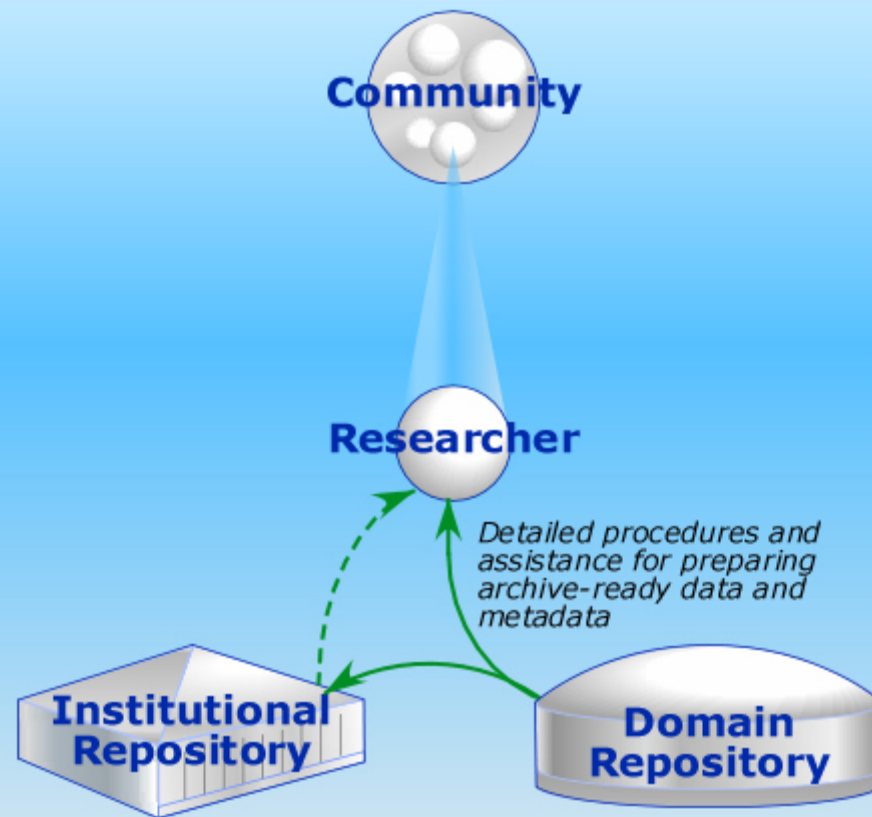


II Initial Data Collection



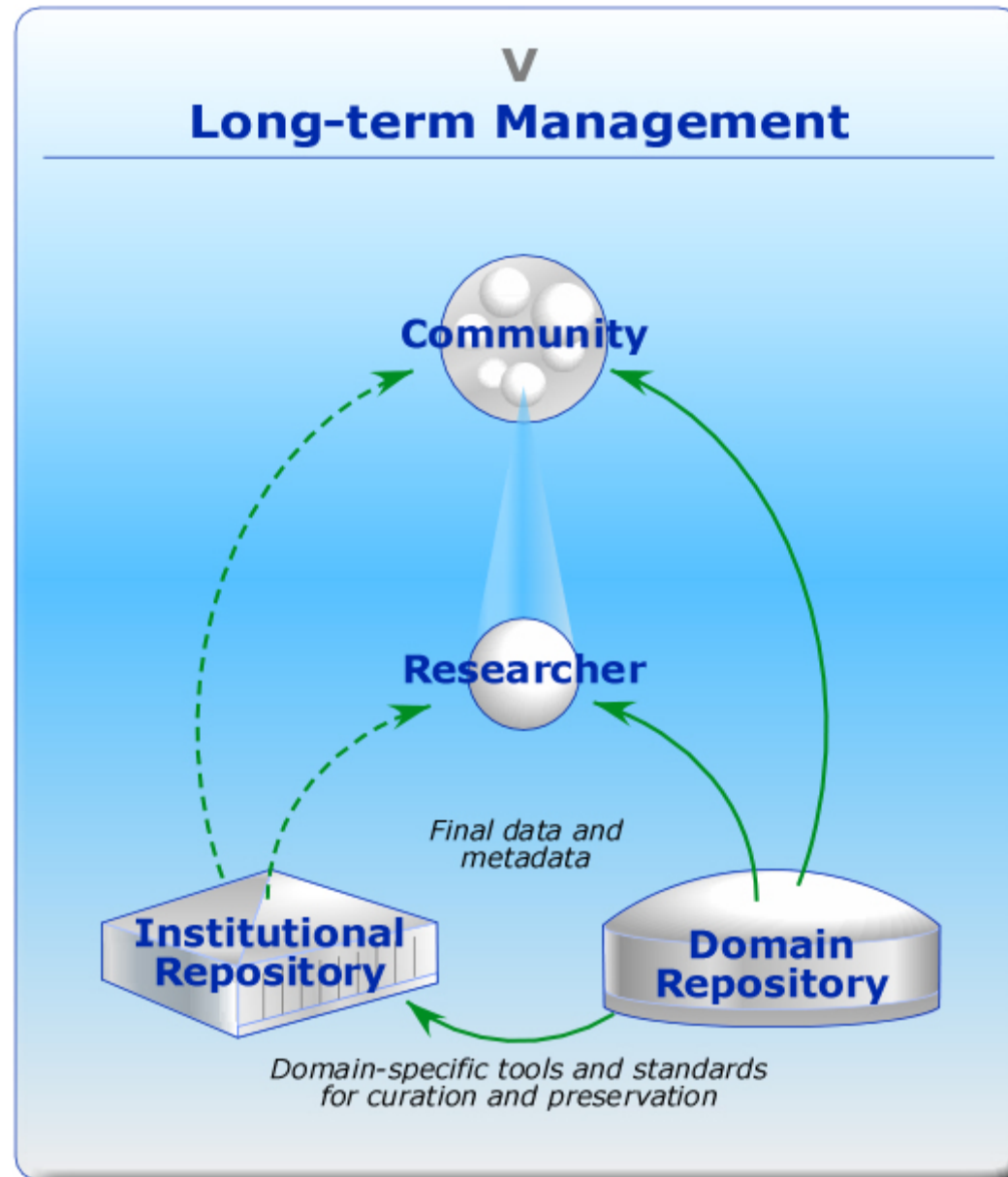
III

Final Data Preparation & Analysis



IV Publication & Sharing





iPRES 2006. Ann Green and Myron Gutmann. Graphic by Cole Whiteman, ICPSR

Further information:

UMich Deep Blue preprint:

“Building Partnerships Among Social Science Researchers, Institution-based Repositories and Domain Specific Data Archives”

<http://hdl.handle.net/2027.42/41214>

deepblue@umich.edu

IASSIST Conference, Montreal:

May 15-18, 2007

www.iassistdata.org