Closing the Loop
Digital Library Services in the Age of Data and Open Access

Tom Cramer
Chief Technology Strategist
Stanford University Libraries
June 27, 2013
What Does A Library Do?

• **Select** resources to add to its collections
• **Acquire** those resources
• **Catalog** the resources to enable discovery
• **Hold** the items until such time as they are needed
• **Circulate** the items to patrons on request
• **Preserve** the items so they are available to future scholars
The Research & Information Lifecycle

Conceive → Plan & Propose → Collect Data → Refine & Analyze → Publish
The Research & Information Lifecycle: Plus Data

Conceive → Plan & Propose → Collect Data → Refine & Analyze → Publish

Deposit Data → Manage Data → Preserve Data → Disseminate Data

Cite
The Research & Information Lifecycle: Closing the Loop

1. Conceive
2. Plan & Propose
3. Collect Data
4. Refine & Analyze
5. Publish
6. Deposit Data
7. Manage Data
8. Preserve Data
9. Disseminate Data
10. Cite
The Stanford Digital Repository (SDR) is...

- a core Library service
- Stanford’s digital preservation system
- a platform for asset management & delivery
- ~125 TB, 250k objects, 100M files
  - books, mss, articles, theses, images, media, data
Access, Embargo, Licensing & Attribution

• **Access**
  – Discover, View, Download
  – World, Stanford, Group, Private settings

• **Licensing**
  – Creative Commons, Data Licenses, Software Licenses, All rights reserved
  – Depositors maintain IP; SDR licensed to provide a stewardship environment

• **Embargo**
  – Support for restricted access for pre-determined timeframes (e.g., Stanford-only for 6 months...)

• **Attribution & Use Statements**
  – Depositors enabled to specify access & use statements
PURLs (Persistent URLs)
Openness in Science

Need infrastructure to facilitate...

- Deposit/curation of data and code
- Links to published articles
- Permanence of links

Digital Preservation
What Is DPN?

57 member organizations cooperatively investing in long-term, scalable, digital preservation
What Is DPN?

technical staff and systems from 5 large scale preservation repositories
What Is DPN?

...working groups of experts in succession rights, business services, communications and research data...
What is DPN?

All building a digital preservation backbone for the academy
Scenario 1: Ingest & Replication
Scenario 2: Restoration of Content
Scenario 3: First Node Cessation
Scenario 4: Successioning
What Does DPN Do?

1. Establishes a network of heterogeneous, interoperable, trustworthy, preservation repositories (Nodes)
2. Replicates content across the network, to multiple nodes
3. Enables restoration of preserved content to any node in the event of data loss, corruption or disaster
4. Ensures the ongoing preservation of digital information from depositors in the event of dissolution or divestment of depositors or an individual repository
5. Provides the option to (technically and legally) "brighten content" preserved in the network over time
DPN Benefits

1. Resilience
2. Succession
3. Economies of scale
4. Efficiency
5. Extensibility
6. Security
Big Challenges/Questions from Library Perspective

• Technical integration
  – Publishers, domain repositories, research systems, Web-based sites

• Distribution of curation responsibilities
  – Who selects? deposits? describes?

• Funding mechanisms

• Integration into the existing workflow(s)