Living the Future:
Digital Repositories & Large Data Sets

Christopher Cole
Associate Director
National Agricultural Library
24 July 2009

Our past

Predicted Future

The Reality

Today

- Mixed model
  - Part digital
  - Part paper
- Why
  - Availability
  - Cost
  - Expertise

Amidst all this Change

- Our role and value are being questioned
  - Why do we need the library when we have the Internet?
- But our key role as libraries remains to
  - Acquire
  - Organize
  - Serve it to the user

What changes

- Paper versions are no longer produced
- We are learning learn to store, curate, and serve digital versions
  - i.e.: develop digital repositories
- Skill sets are different
  - e.g.: Licensing, metadata skills beside MARC
- Some tasks are disappearing
  - e.g.: serials check-in, exchange

AgSpace

- NAL's digital repository
- Umbrella Term
  - DSpace
  - ZyLab
  - Image Storage
- Every Item has metadata in AGRICOLA
- Persistent hyperlinks
- Goal:
  - applications will change
  - Objects, Metadata and Hyperlinks survive & transfer

Staff Reorganization

- Converted gift & exchange staff to USDA repository activities
- Repurposed indexers and catalogers
- Developed IT skill sets to operate a repository

Changing programs

- Re-focused our AGRICOLA index to USDA and full-text items
- Devoted 15% of the indexing capacity to AgSpace items
- Exchange program has ended
- Gifts discouraged
- Serials check-in only crucial for indexed titles

Future Focus

- Completing the retro-conversion of all our metadata catalogs (card & print)
- Identifying which USDA resources to digitize
  - Uniqueness
  - Non-duplication
- Identify potential partners to increase pace & scope of digitization

Large Data Sets

- Ticking time bomb
  - Annual production of data sets will soon outstrip world-wide storage capacity
  - Information being created at all levels of USDA and other science organizations
  - Access to this data is crucial to ongoing research
  - Data Sets will begin to be more important than the printed article

The Library's Role in Large Data Sets

- Can't:
  - invent the data storage applications
  - Curate the live data sets
- Can:
  - Develop metadata schemas
  - Develop file conventions
  - Enable machine to machine searching
- The solutions will be machine intensive not staff intensive

Questions?

Christopher.Cole@ars.usda.gov