Summary of Economic Sustainability Models Breakout Session

Chris Rusbridge and Fran Berman,
Co-leads

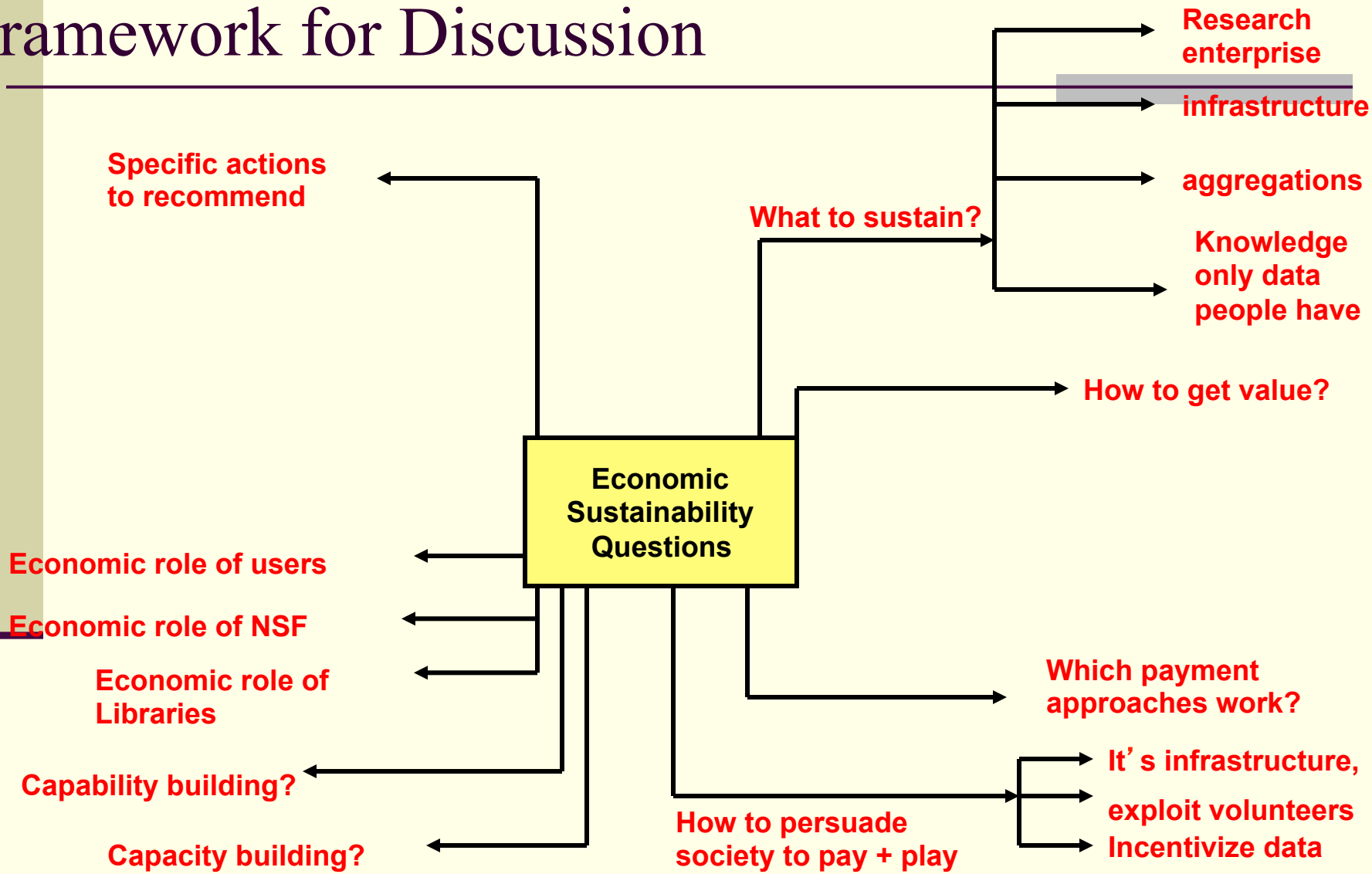
Julia Blixrud and Heather Joseph,
Scribes

ARL/NSF Digital Data Workshop, September 2006

Focus: Economic Sustainability

What models are required to sustain data management and preservation efforts over the long term?

Chris' Mindmap Provided a Framework for Discussion



What economic models are relevant?

- The group began the discussion by describing the economic models which support their current activities.
- During the discussion, we discussed a spectrum of traditional and non-traditional related economic models including
  - **ICPSR** (subscription, user fees, federal, private funding)
  - **The Mormon Church** (tithing, user fees, sales)
  - **PBS** (donations, federal, state?, volunteers [donated time, expertise], sales)
  - Volunteer activity (archiving @ home)
  - **Markets** (DRI, data "futures", shares, etc.)
  - **Hybrid** (federal+state, public+private, etc.)

A thought experiment: Abstracting ICPSR

- **What has made ICPSR successful as a model?**
  - Robust environment with low barrier to access
  - Content which is of great value
  - Business model and structure which reflect the culture of the domain and constituent users
  - Useful tools associated with data
  - Trusted repository

Key to start from state-of-the-art rather than to reinvent the wheel

- **Economic sustainability models should utilize existing theory and practice as a foundation** – critical to have economists and sustainable infrastructure expertise in the discussion.
  - This is **symptomatic of a more general problem** – we shouldn't reinvent the wheel in economics, business, archiving, etc. Rather we should use the existing knowledge and experience base as a starting off point
  - This will mean the need for venues for more in-depth cross-cultural discussion and projects to help educate communities
- **Preservation will require both research into new viable models, and experimentation with new ideas**
  - Five years is short for an experiment, 5+5 is better
  - Risk taking: failure is an option!

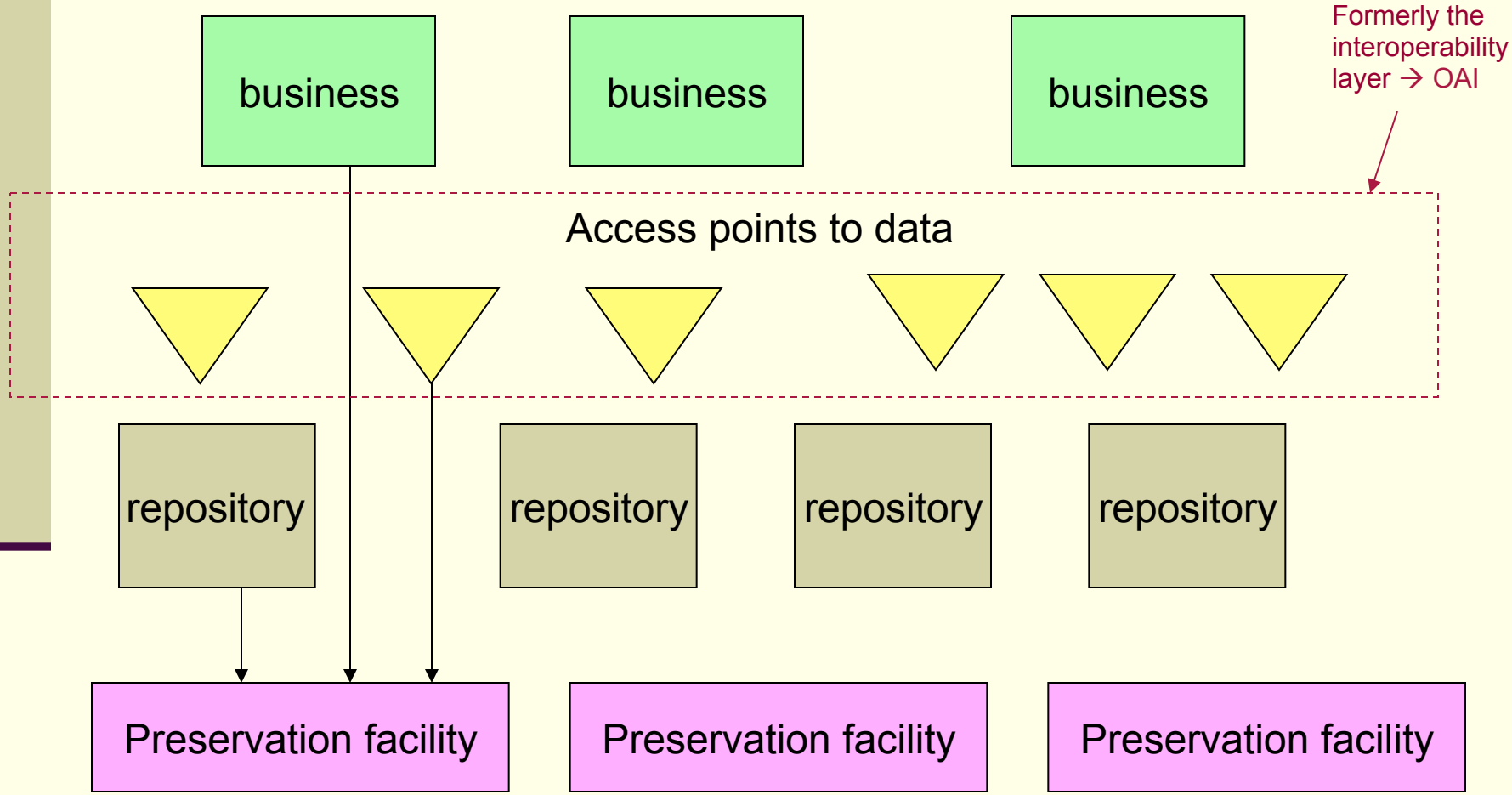
Many "building block" issues

- How do researchers and librarians sort through the **legal and policy issues** regarding ownership, use, confidentiality, privacy, liability etc.?
- What is the **minimal level of service** that makes data preservation worthwhile?
- What is the **cost of not keeping data**? When is it productive to re-compute, replicate experiments, re-do?
- What is the **data version of the "Earth Simulator"**? (i.e. what is the newsworthy item that will get U.S. competitive juices flowing and help generate new funding for data management and preservation)

Interesting Issues

- Large projects doing a reasonable job of putting data on the radar. **Small projects are the most at risk**.
- **Most libraries do not currently host substantive research data** – both library and research community need more experience with one another's cultures. Is there a way NSF can help foster greater engagement?
- **Good infrastructure must have a plan for "the end"** – how do we reappraise if necessary, how do we hand-off, how do we become self-sufficient?

Eric's Updated Version of the Cliff Lynch model



"Actionable" Recommendations 1

"We don't get anywhere if we don't start somewhere."

1. **Involve economics and social science experts** in developing economic models for sustainable data preservation – research should ultimately generate models which could be tested in practice.
2. **Set up multiple repositories and treat them as experiments**
   - Require that repository experiments develop plans to address key issues such as transition between media/ formats/institutions, self-sustainability, exit strategy, etc.
3. **Develop usable and useful tools for automated services and standards** which make it easier to understand and manipulate data. Develop incentives to encourage community use. Invisible metadata creation!

"Actionable" Recommendations 2

4. **Require data sharing plan in proposals** that has practical value (and appropriate support). Plans for resource and reference data should contribute to community data stewardship

5. **Create and enforce data sharing policies among NSF awardees** (e.g. final report not accepted unless awardee is compliant with stated data management plan)

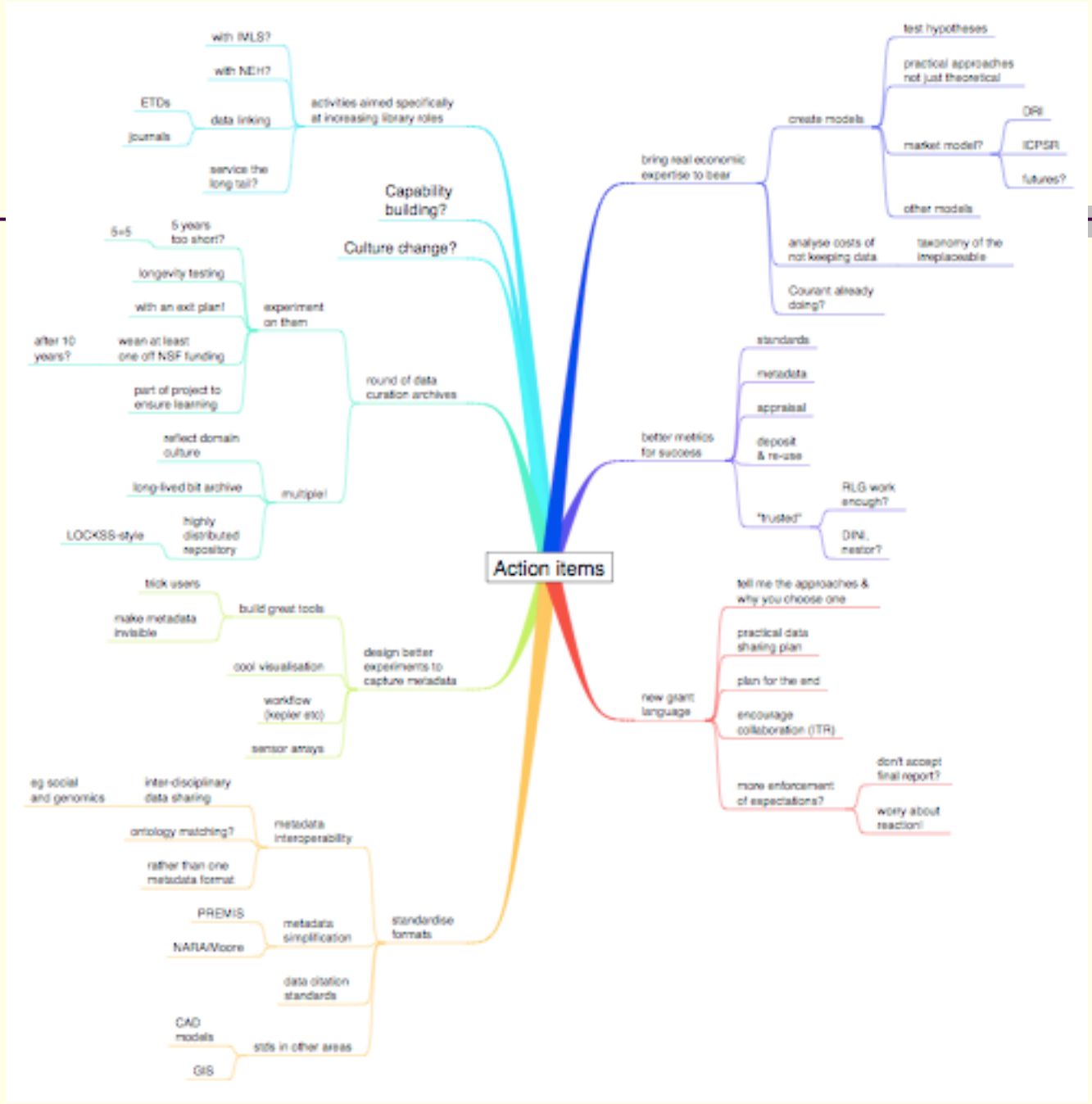
6. **Use NSF program process to help the library community take more responsibility** for the stewardship of research data (with other funders?)

“Actionable” Recommendations 3

7. **Use NSF program process to change culture** in research community
8. Undertake capacity & capability building activities

A Bolder Vision? Remember Dli!

- US Digital Curation Initiative
  - A major, inter-disciplinary, cross-directorate, inter-agency program, with options built in for international collaboration (UK, EU, Australia at least), that will both experiment on models and build sustainable curation services!


Action items
Capability building?
activities aimed specifically at increasing library roles
with IMLS?
with NEH?
data linking
ETDs
journals
service the long tail?
Culture change?
round of data curation archives
experiment on them
5=5
5 years too short?
longevity testing
with an exit plan!
after 10 years?
wean at least one off NSF funding
part of project to ensure learning
multiple!
reflect domain culture
long-lived bit archive
highly distributed repository
LOCKSS-style
design better experiments to capture metadata
build great tools
tick users
make metadata invisible
cool visualisation
workflow (kepler etc)
sensor arrays
standardise formats
metadata interoperability
eg social and genomics
inter-disciplinary data sharing
ontology matching?
rather than one metadata format
metadata simplification
PREMIS
NARA/Moore
data citation standards
stds in other areas
CAD models
GIS
bring real economic expertise to bear
create models
test hypotheses
practical approaches not just theoretical
market model?
DRI
ICPSR
futures?
other models
analyse costs of not keeping data
taxonomy of the irreplaceable
Courant already doing?
better metrics for success
standards
metadata
appraisal
deposit & re-use
"trusted"
RLG work enough?
DINI, nestor?
new grant language
tell me the approaches & why you choose one
practical data sharing plan
plan for the end
encourage collaboration (ITR)
more enforcement of expectations?
don't accept final report?
worry about reaction!

Bring real economic expertise to bear

- Create models
  - test hypotheses
  - practical approaches not just theoretical
  - market model?
    - DRI, ICPSR, futures?
  - other models
- Analyse costs of not keeping data
  - taxonomy of the irreplaceable
- Courant already doing?

Design better metrics for success

- Standards
- Metadata
- Appraisal
- Deposit & re-use
- "trusted"
  - RLG work enough?
  - DINI, nestor?

New grant language

- Practical data sharing plan
  - tell me the approaches & why you choose one
- Plan for the end
- Encourage collaboration (cf ITR)
- More enforcement of expectations?
  - don't accept final report?
  - worry about reaction!

Work to standardise formats

- Metadata interoperability
  - inter-disciplinary data sharing
    - eg social and genomics
- Preservation metadata simplification
  - PREMIS
  - NARA/Moore
- data citation standards & promotion
  - To change culture
- Support standards in other areas
  - CAD models, GIS, etc

Design better tools & experiments

- Sensor arrays & other experimental engineering to capture metadata
- Build great tools
  - Robust, reliable, useful, usable
  - “trick users”, make metadata generation invisible
    - Eg cool visualisation, workflow (Kepler etc)

New round of data curation archives

- Multiple!
  - reflect domain cultures
- Experiment on them (take risks)
  - 5 years too short- 5+5?
  - with an exit plan!
  - longevity testing
  - part of project to ensure learning
  - wean at least one off NSF funding
    - after 10 years?

New round of data curation archives

- Specific possibilities
  - long-lived bit archive
  - Build on IR work
  - highly distributed repositories
    - LOCKSS-style
    - Archiving @ home

Activities aimed specifically at increasing library roles

- Libraries have significant opportunities to extend their roles in info discovery, archiving etc in research data
- Partnerships with domain researchers
- Forum for outreach & scientific communication
- Data linking from
  - ETDs
  - journals
- Capacity building opportunities in library educaiton etc with NEH and/or IMLS?

Other things

- Capability building
  - Education: librarians, data scientists, researchers
- Capacity building
- Culture change?