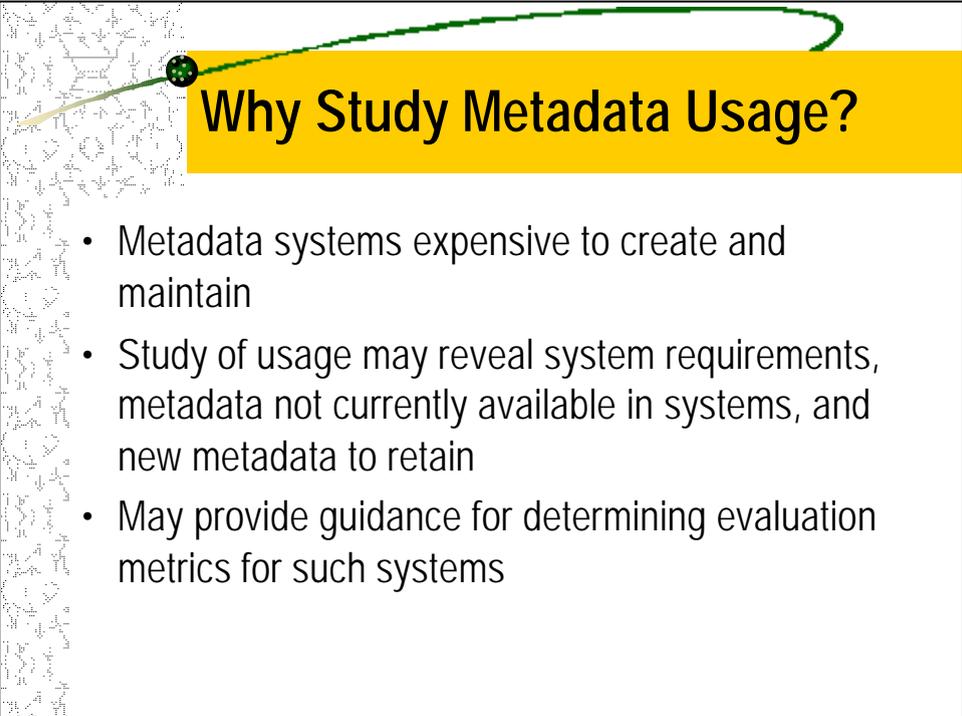


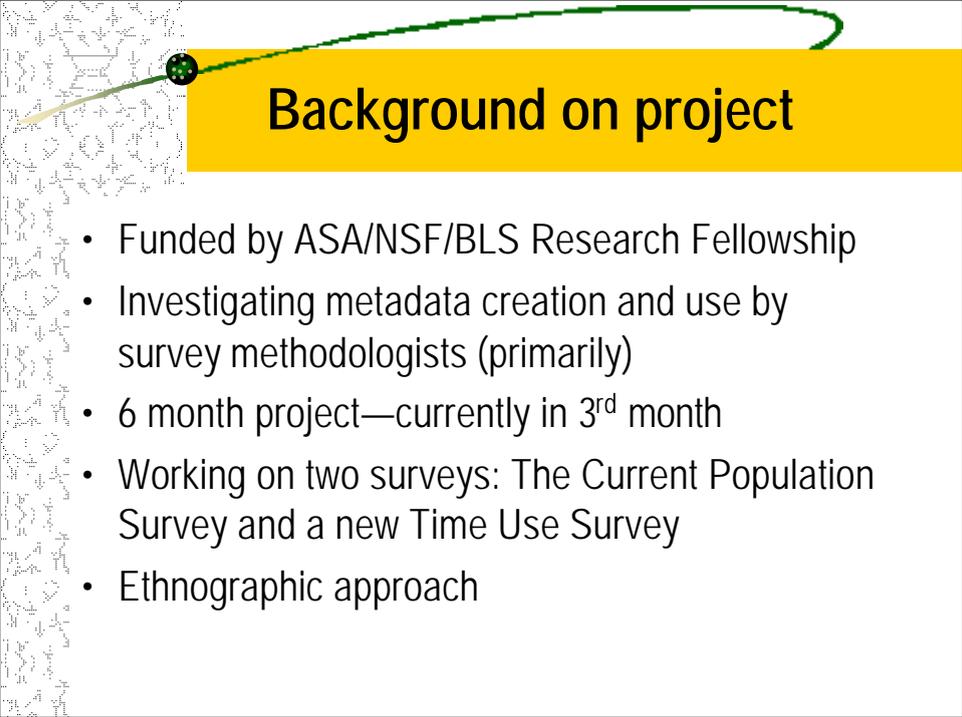
## **Understanding and Modeling Metadata Creation and Use**

Carol A. Hert  
U.S. Bureau of Labor Statistics and  
Syracuse University



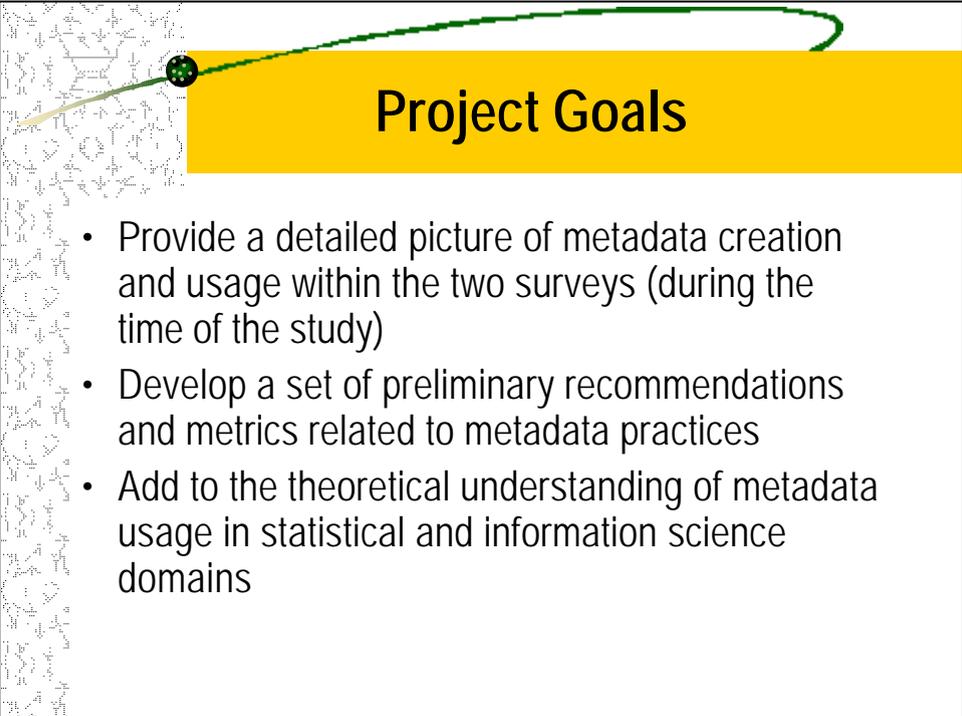
## **Why Study Metadata Usage?**

- Metadata systems expensive to create and maintain
- Study of usage may reveal system requirements, metadata not currently available in systems, and new metadata to retain
- May provide guidance for determining evaluation metrics for such systems



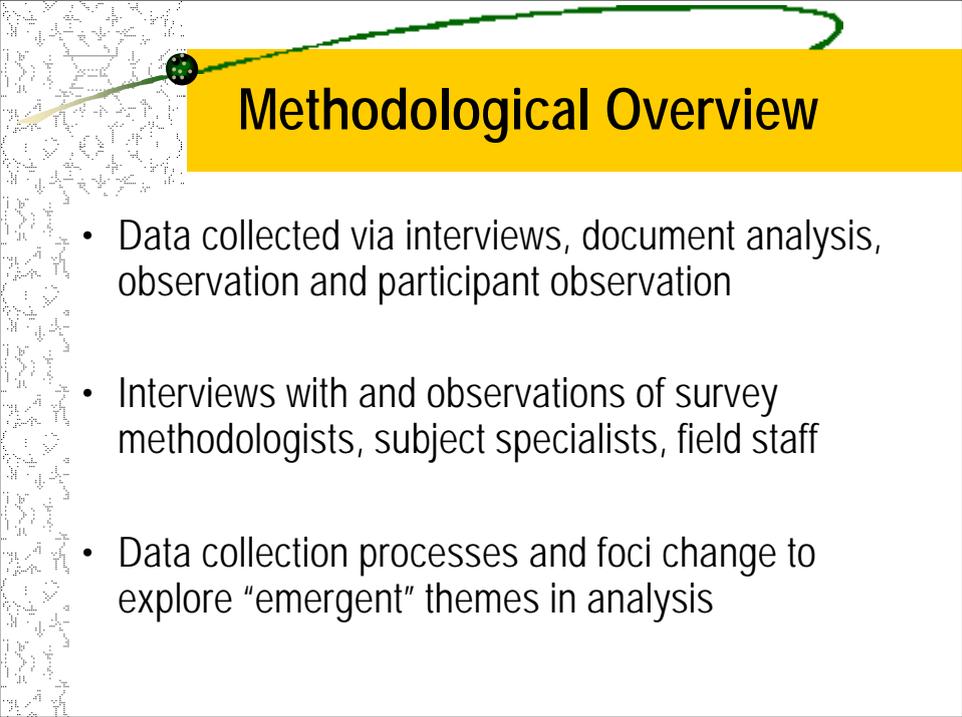
## Background on project

- Funded by ASA/NSF/BLS Research Fellowship
- Investigating metadata creation and use by survey methodologists (primarily)
- 6 month project—currently in 3<sup>rd</sup> month
- Working on two surveys: The Current Population Survey and a new Time Use Survey
- Ethnographic approach



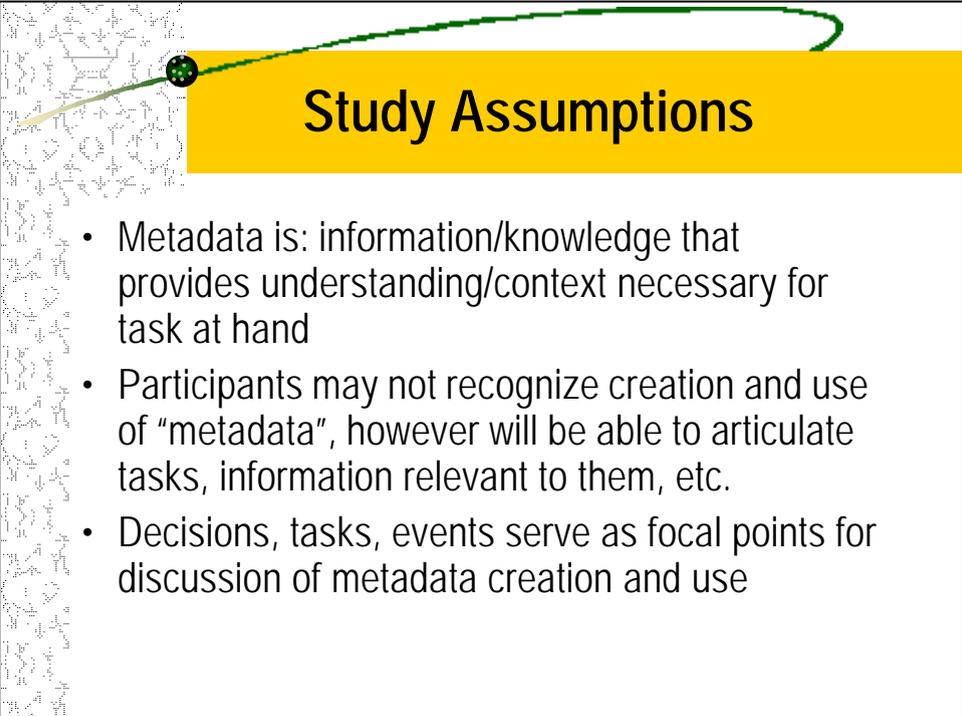
## Project Goals

- Provide a detailed picture of metadata creation and usage within the two surveys (during the time of the study)
- Develop a set of preliminary recommendations and metrics related to metadata practices
- Add to the theoretical understanding of metadata usage in statistical and information science domains



## Methodological Overview

- Data collected via interviews, document analysis, observation and participant observation
- Interviews with and observations of survey methodologists, subject specialists, field staff
- Data collection processes and foci change to explore “emergent” themes in analysis



## Study Assumptions

- Metadata is: information/knowledge that provides understanding/context necessary for task at hand
- Participants may not recognize creation and use of “metadata”, however will be able to articulate tasks, information relevant to them, etc.
- Decisions, tasks, events serve as focal points for discussion of metadata creation and use

## Preliminary Insights (Not Findings!)

- Analytic frames used in data analysis
- Analytic themes currently being explored
- Possible metrics and issues relevant to metadata system recommendations

## Analytic Frames: Metadata Types

**Statistical metadata**

**Methodology metadata**

survey methodology

Non-survey research methodology

Agency-specific survey methodology

**Administrative**

Staffing/personnel

Research project management

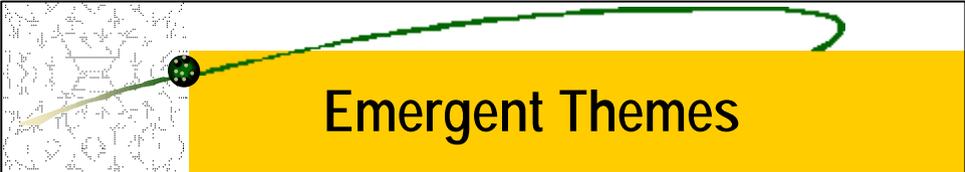
**Referral/person information**

**Rationales for decisions**



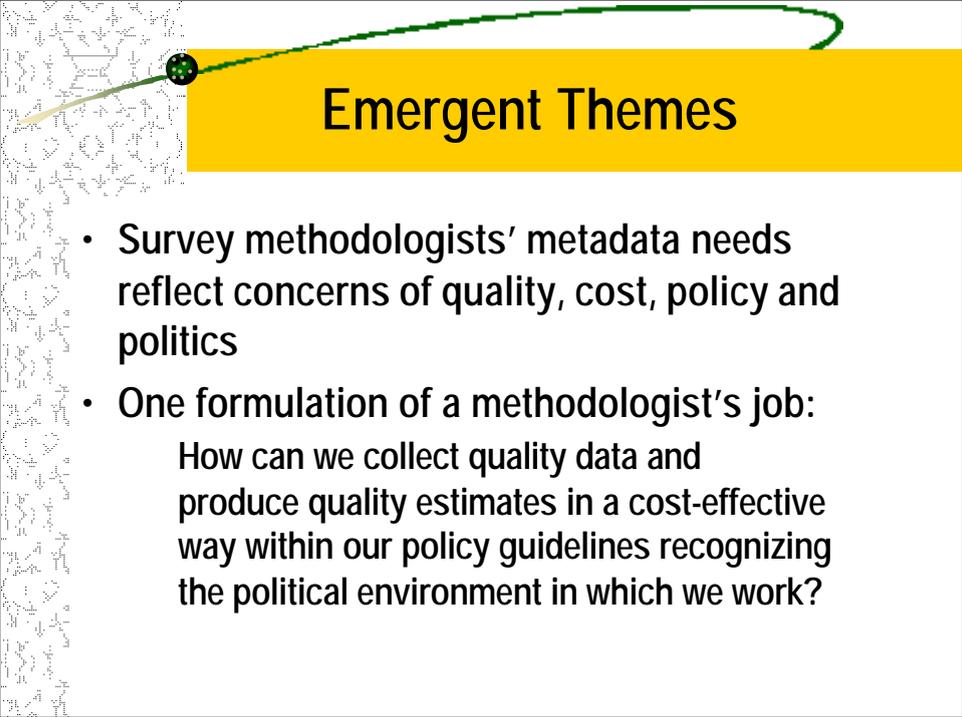
## Analytic Frames

- Aspects of metadata in context
  - How/when created
  - How organized
  - What format
  - How stored
  - How retrieved
  - How displayed/presented
  - How used (during a task)



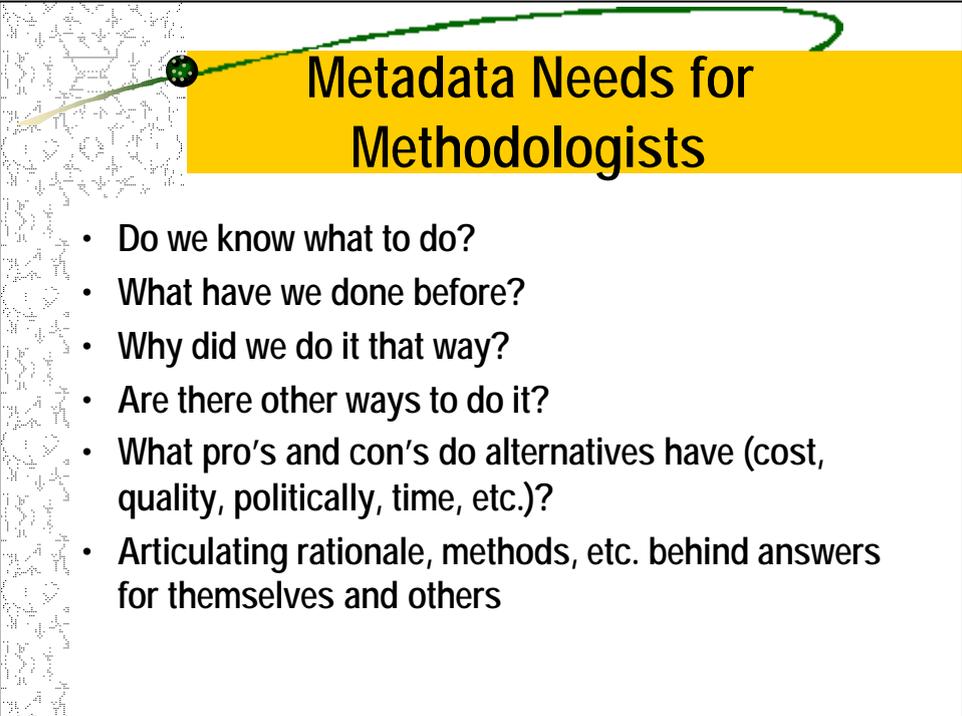
## Emergent Themes

- Survey methodologists' metadata needs reflect concerns of quality, cost, policy and politics
- The physicality of metadata
- Decisions as fulcrums for metadata creation and use
- Translation of metadata across boundaries



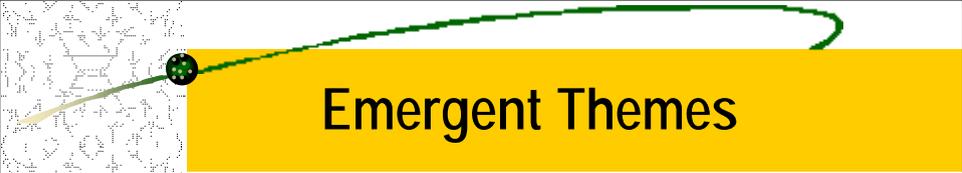
## Emergent Themes

- Survey methodologists' metadata needs reflect concerns of quality, cost, policy and politics
- One formulation of a methodologist's job:  
How can we collect quality data and produce quality estimates in a cost-effective way within our policy guidelines recognizing the political environment in which we work?



## Metadata Needs for Methodologists

- Do we know what to do?
- What have we done before?
- Why did we do it that way?
- Are there other ways to do it?
- What pro's and con's do alternatives have (cost, quality, politically, time, etc.)?
- Articulating rationale, methods, etc. behind answers for themselves and others



## Emergent Themes

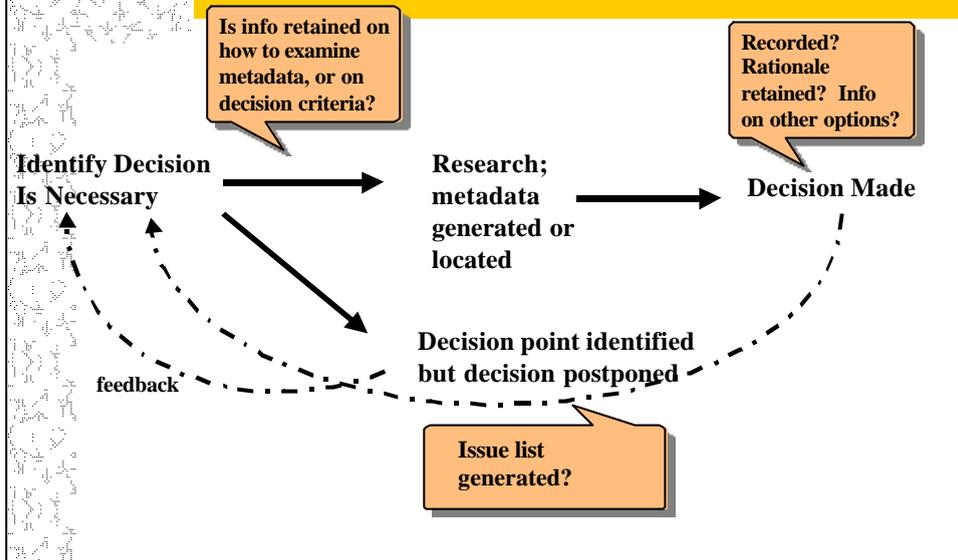
- Metadata “aspects” are shaped by physical world
  - Use of color to manage data collection materials, different notebooks for different projects
  - Physical space may be used to present metadata needed often or to facilitate recall and retrieval
  - Office space organization influences what metadata are retained and or shared



## Emergent Themes

- Decisions are fulcrums for metadata creation and use
  - However these decision fulcrums are perceived and acted on differently by different stakeholders

## A Preliminary Decision-oriented Model of Metadata Usage and Creation



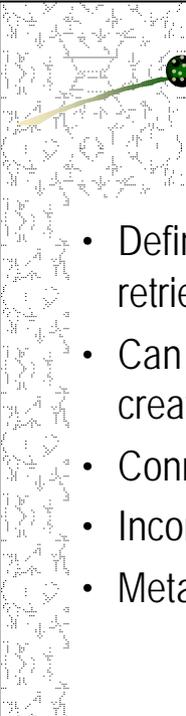
## Emergent Themes

- There is a “metadata” interface/translation point where survey methodological concerns meet field concerns
- Shared purpose but not necessarily shared aspects of decision
  - The SCHIP sample
  - How much of the Time Use sample should be assigned to the field (versus CATI)
  - CPS and data collection during Thanksgiving week



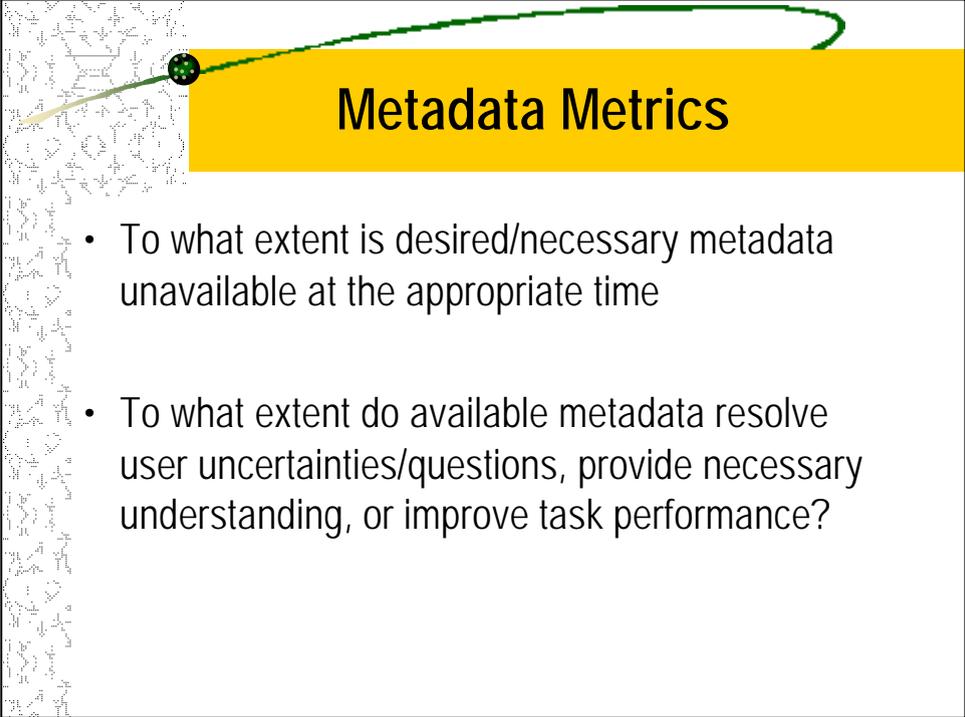
## Summary of Study Status

- First round of data collection completed
  - Low level categorizations of metadata *in situ*
  - Set of “tantalizing” insights to investigate
- Further analysis of collected data followed by feedback from stakeholders
- Further data analysis to confirm/disconfirm generalizations
- Recommendations arising from empirical work



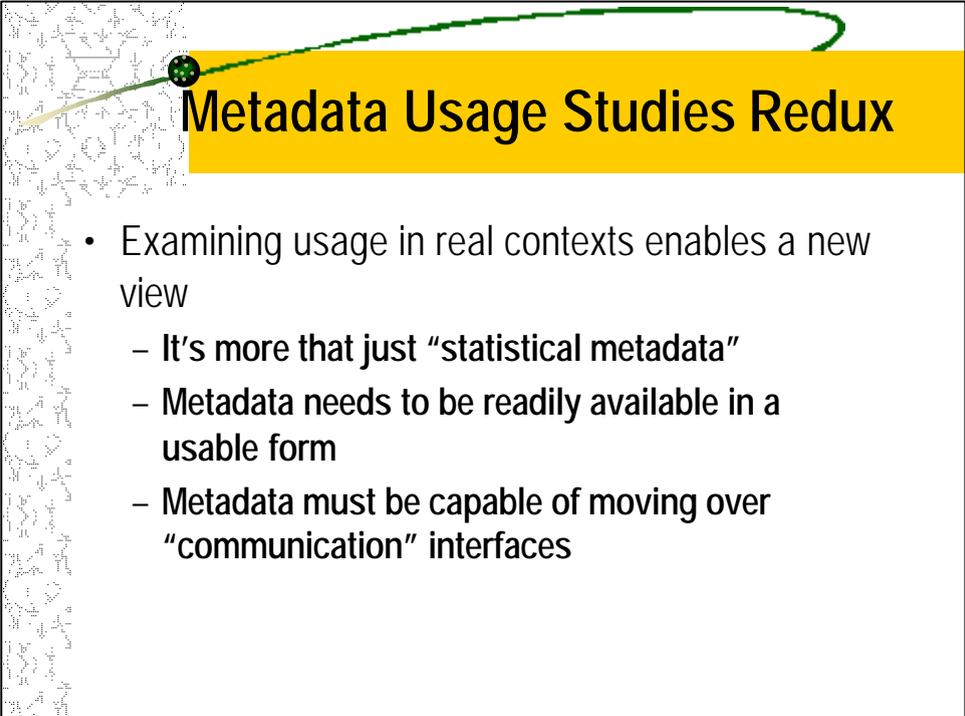
## Improving Metadata Systems

- Defining metadata in ways to enable capture and retrieval
- Can we recognize “metadata” during its creation?
- Connecting related sets of metadata
- Incorporating non-textual metadata
- Metadata to bridge “interfaces”



## Metadata Metrics

- To what extent is desired/necessary metadata unavailable at the appropriate time
- To what extent do available metadata resolve user uncertainties/questions, provide necessary understanding, or improve task performance?



## Metadata Usage Studies Redux

- Examining usage in real contexts enables a new view
  - It's more than just "statistical metadata"
  - Metadata needs to be readily available in a usable form
  - Metadata must be capable of moving over "communication" interfaces



## My contact information

- Email: [cahert@syr.edu](mailto:cahert@syr.edu)
- Phone:
  - 202-691-7399
  - 315-443-5441 (after January 3, 2001)
- Websites with my other work related to statistical information
  - <http://istweb.syr.edu/~hert> (technical report section)
  - <http://istweb.syr.edu/~tables>