



Archiving Human Intermediation:

The Digital Reference Electronic
Warehouse (DREW) Project

Scott Nicholson and R. David Lankes

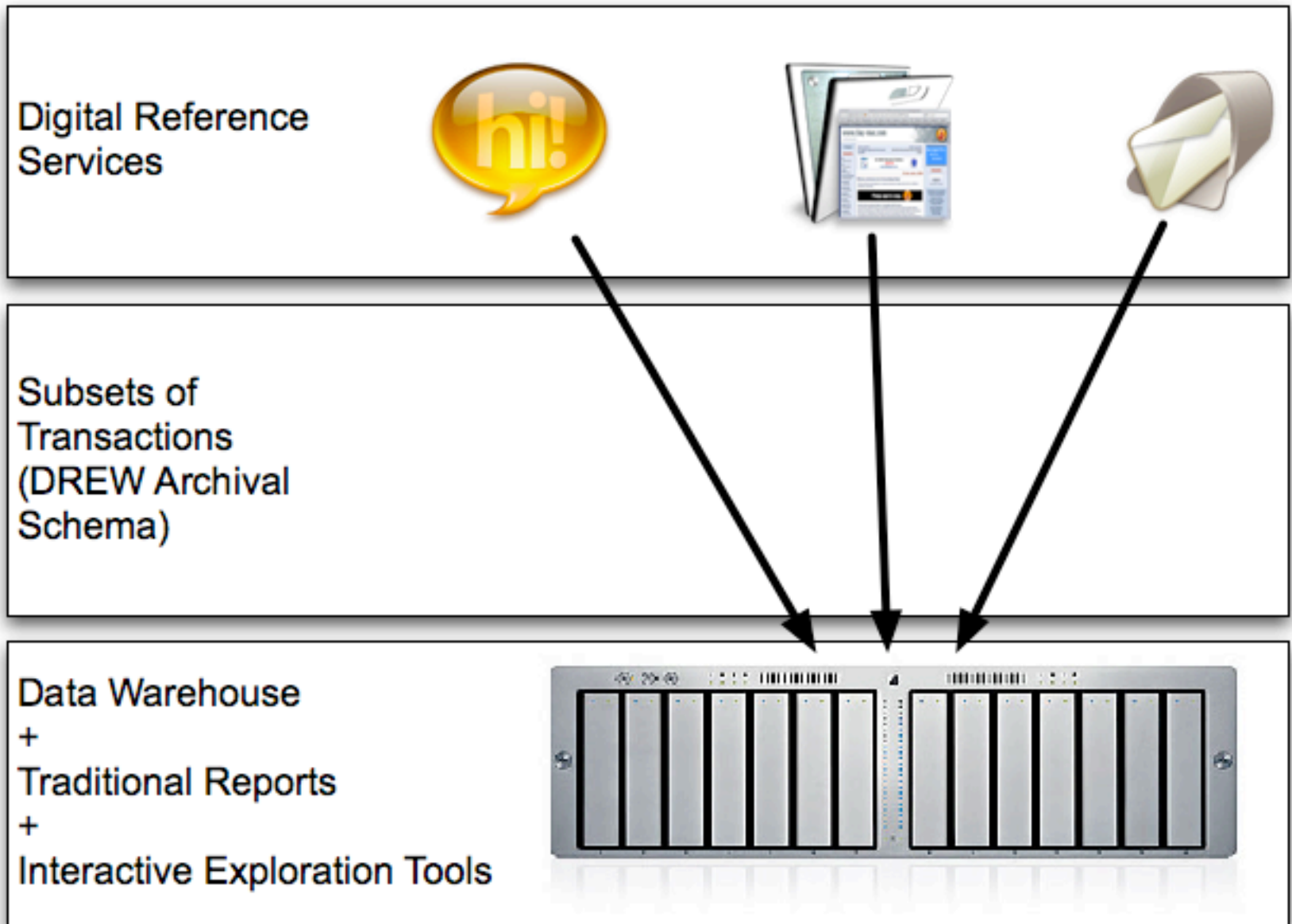
Syracuse University School of
Information Studies



Overview

- Overview of DREW
- Survey of Digital Reference Services
- Development of DREW Schema
- Privacy Issues
- Applications of DREW
- DREW as a Complex Adaptive System
- DREW Research Agenda

Overview



Overview

DREW



LIS Researchers

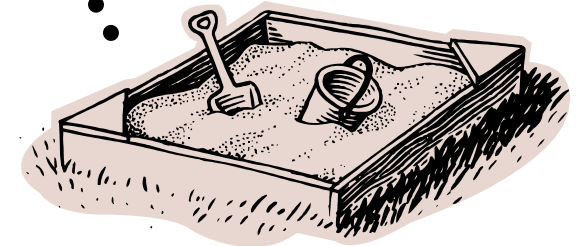
Archive of Transactions
Standard and Custom Reports
Exploratory Tools



Individual
Services

Support Management
and Decision-Making

Multi-Service
Data Needed for
Research Projects



Sandbox of Exploration to
Improve Knowledge and
Inform Practice



Introduction

- Goal: Create a
 - shared archive of
 - digital reference transactions
 - from multiple services and
 - different disciplines
 - for research purposes

Result: A multi-disciplinary knowledge base capturing human expertise

Related Projects



Other archives

- Google Answers
- QuestionPoint's Knowledge Base
- MadScientist & other single service archives

NetRef – NISO standard for exchanging questions between service

- Focused on process during question answering
 - DREW will complement NetRef
-
- NetRef – Standard for in-process
 - DREW – Standard for archival purposes



Survey of Digital Reference Services

- Starting point - Janes (2003)* and user view (patron and question)
- Expand to include information on the
 - Patron
 - Question
 - Responder
 - Response
- Distributed at VRD2003 and online

* Janes, J. (2003). Question Negotiation in an Electronic Age. In R. D. Lankes, S. Nicholson & A. Goodrum (Eds.), *The Digital Reference Research Agenda* (pp. 48-60). Chicago, IL: Association of College and Research Libraries.

Survey Format



- For each field, services were asked if they:
 - Currently collect that field
 - Do not currently collect, but are willing to collect that field
 - Are not willing to collect each field
- Write-ins and comment areas for each topic

Survey Fields

Patron Information	Expert/Responder Information
Name	Name
E-Mail	E-mail
Telephone	Telephone
City	City
State	State
Country	Country
Grade/Education Level	Title
Professional Role	Institution
Member of organization (library, school, etc.)	Qualifications

Question Information	Response Information
Subject (From a List)	Response Text
Subject (Free text supplied by User)	Resources consulted
Text of Question	Date of response
Purpose	Time of response
Desired form of answer	
Previously consulted sources	
Requested deadline for response	
Date of question	
Time of question	
Routing information (i.e. question referrals)	

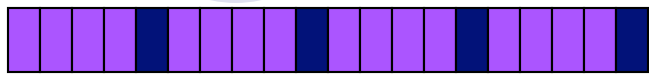
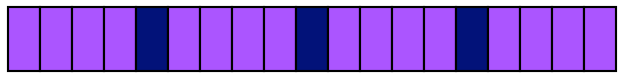
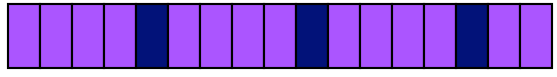



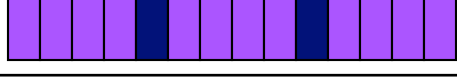
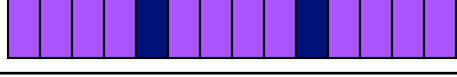

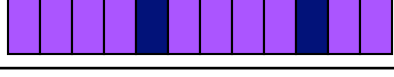

Percentage of Services Collecting Different Fields

25% 50% 75% 100%

Question Text, Date, Time	
Response Text, Date, Time	
Patron & Responder Identifiers	
Resources Consulted for Response	
Routing & Referral Information	
Responder Institution	
Question Subject (free-text)	
Patron Location	
Patron Level	

Percentage of Services Willing to Collect Different Fields

25% 50% 75% 100%

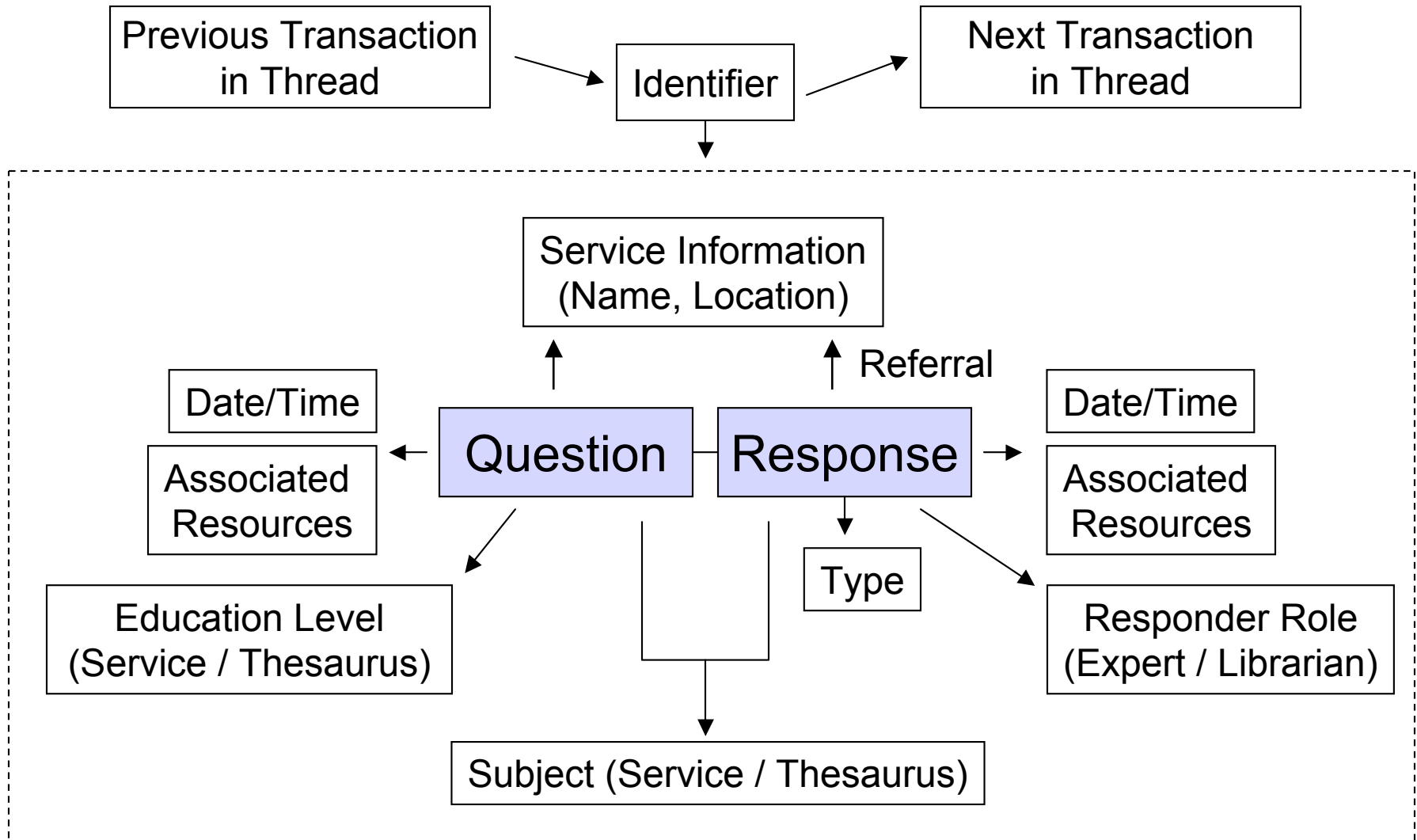
Question Text, Date, Time	
Response Text, Date, Time	
Patron & Responder Identifiers	
Routing & Referral Information	
Responder Institution	
Resources Consulted for Response	
Question Subject (free-text)	
Patron Location	
Pre-Query Resources Consulted	
Patron Level	
Responder Role	



Observations

- Asynchronous Webform-based services currently collect more information than chat
- Synchronous Chat/IM services are *willing* to collect more information than webform

Constructing the DREW Schema



Privacy



- Current goal = For research only
 - Contributing Services – Own archives
 - LIS Researchers – Subset needed for Research
- Privacy of digital reference is a challenge
 - Easy to remove fields of personal information
 - Difficult to strip out personal information from full-text
- Research agenda on Privacy
 - Similar research on de-identification of medical records

The Next Frontier: Knowledge Bases

- Possible Utility of Knowledge Bases
 - Alternative Source of Answers
 - Help Desk Model, Saturation
 - Resource for Expert
 - “Brain Box”
 - “First Order” Resource
 - Disconnected from Reference Process

Current Approaches

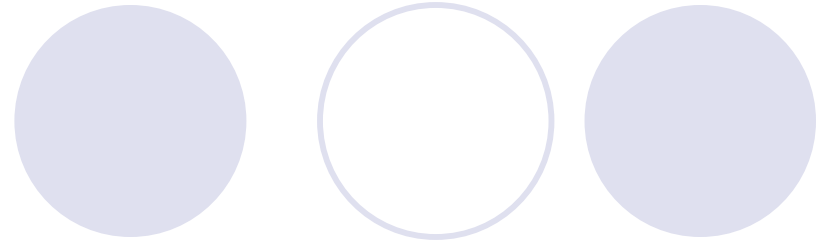
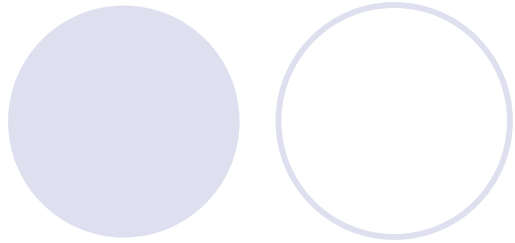


- All or Edit
 - Either all transactions are searchable or services use an extensive deductive editing process
- Primarily Deductive
 - Context Dependencies
 - Metadata Creation
 - Chunking
 - Fact Shifting and Temporal Dependencies
- Seed and Weed
 - Edit them in, then have to weed the archive



A New Approach: Induction

- Treat the Output of Reference Transactions as Semi-Structured Digital Object
- Semi-Structured Objects have Static and Dynamic Attributes
 - Static: User ID, Expert ID, Content
 - Dynamic: Age, Topicality, Annotations
- Create a “Space” for These Objects/Agents to Interact
- Create Performance Systems for Agents

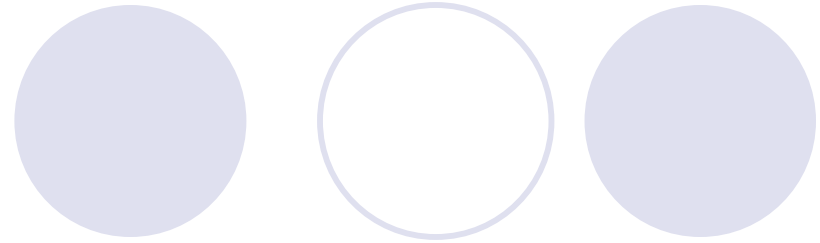
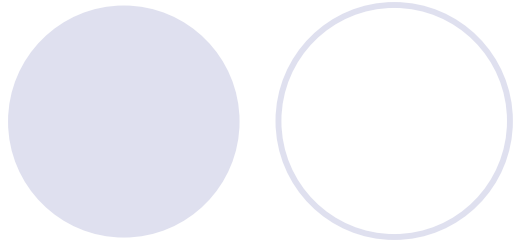


Expert
Patron
Question
Age
Topicality
Ratings
Annotations



Expert
Patron
Question
Age
Topicality
Ratings
Annotations

```
IF expert (STRING)=expert' THEN MOVE(+1)  
IF ABS(age(NUM)-age(NUM))>365 THEN MOVE (-3)
```



Dave
Anna
Why?
1
300
√√√
100 Links



Dave
Anna
Where?
440
56
√
0 Links

IF expert (STRING)=expert' THEN MOVE(+1)
IF ABS(age(NUM)-age(NUM))>365 THEN MOVE (-3)



A Complex Knowledge Space

- Performance System

- Tags (Object Attributes)

- Internal Model

- Weighting Scheme for Object Comparison

- Rule Development Schemes (Inductive Rule Development)

- Geometry

- Means of Determining Proximity for Clustering

- Attributes of the Environment

- Synchronization Scheme

- Means of Comparison (Pairwise? Neighboring Nodes?)



DREW Research Agenda through IIS

- Map out standard for digital reference
- Create tools to extract fields from current systems (both synch. and asynch.)
- Explore thesaurus for mapping subjects
- Resolve privacy concerns
- Create evaluation and visualization tools
- Understand life of a reference transaction

A decorative graphic at the top of the slide consists of two groups of three circles. The left group has a solid light blue circle on the left, a white circle with a light blue outline in the middle, and a solid light blue circle on the right. The right group has a solid light blue circle on the left, a white circle with a light blue outline in the middle, and a solid light blue circle on the right.

Conclusion

- Overall Goals:
- Create a place for researchers to access transactions from multiple digital reference services
- Make it easier for library services to explore their own archives using the tools created by researchers