# Virtual Dave Lankes Presentation

http://www.DavidLankes.org

**TITLE:** Archiving Human Intermediation: The Digital Reference Electronic Warehouse (DREW) Project

**AUTHOR(s): Scott Nicholson,** R. David Lankes

**PUBLICATION TYPE:** Presentation

**DATE:** 2004

**VENUE:** Virtual Reference Desk Annual Conference, Cincinnati, OH.

**ABSTRACT:** A brief presentation to discuss trends and thoughts on digital reference as it "grows up."

**KEYWORDS:** virtual reference, data warehousing, research, digital reference

# Archiving Human Intermediation:

# The Digital Reference Electronic Warehouse (DREW) Project

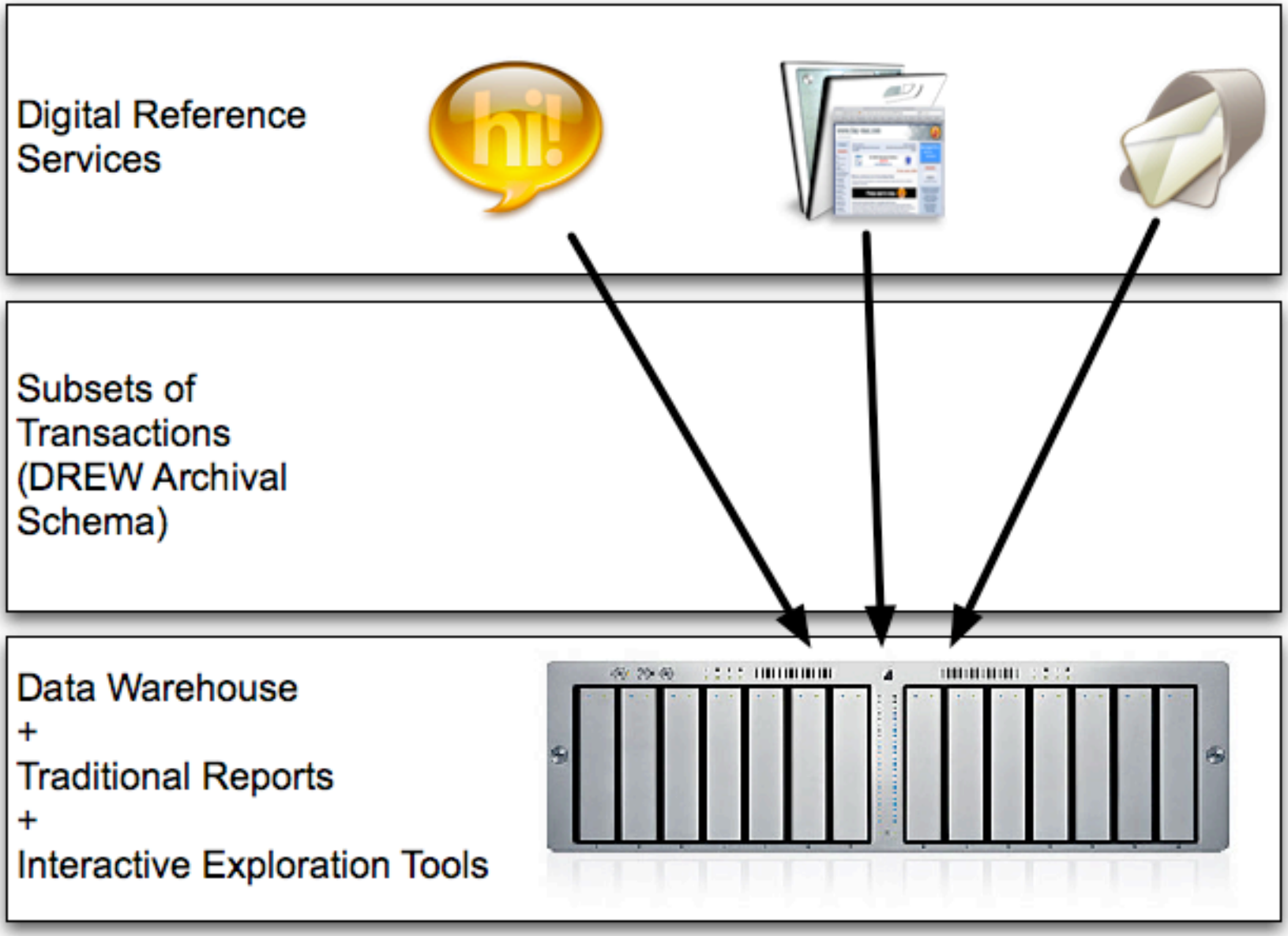
Scott Nicholson and R. David Lankes

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


Digital Reference Services
Subsets of Transactions (DREW Archival Schema)
Data Warehouse
+
Traditional Reports
+
Interactive Exploration Tools

# Overview

**DREW**

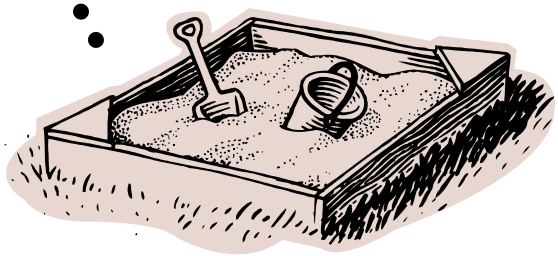
LIS Researchers

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) | |

# Survey Response Demographics

- 53 responses from 49 different services

- Academic Library – 53%
- Public Library – 15%
- Special/Other Libraries – 17%
- AskA Services (no library affiliation) - 14%

# Technology Definitions

- **Different platforms for digital reference:**
- Chat = Synchronous, free-form, may be a pre-chat form to fill out (Chat or IM)
- Webform = Asynchronous, controlled fields
- E-mail = Asynchronous, free-form

# Survey Response Demographics

| | Chat | Webform | Email |
|---|---|---|---|
| Overall | 47% | 38% | 15% |

# Survey Response Demographics

| | Chat | Webform | Email |
|---|---|---|---|
| Overall | 47% | 38% | 15% |
| Academic | 54% | 30% | 17% |
| Public | 29% | 71% | 0% |
| Special/ Other | 50% | 50% | 0% |
| AskA Services | 34% | 50% | 17% |

# Average transactions per Month

- Ranged from 10 to 30,000 (Tutor.com's Online Classroom)

| | Mean (StD) | Median |
|---|---|---|
| Chat | 1906 (6410) | 120 |
| Web form | 164 (192) | 80 |
| E-mail | 30 (31) | 18 |

# Reference Platform Used

| | |
|---|---|
| Question Point | 23% |
| Tutor.com | 21% |
| 24/7 | 8% |
| Altarama RefTracker | 4% |
| QABuilder 2.0 | 4% |
| Docutek VRL Plus | 2% |
| eAssist NetAgent | 2% |
| ExpertCity's Desktopstreaming | 2% |
| LivePerson (HumanClick) | 2% |
| Open Ask A Question | 2% |
| PHP Live Support | 2% |

# Including In-House Solutions

| | |
|---|---|
| *E-mail, Web form, or In-House tool* | 27% |
| Question Point | 23% |
| Tutor.com | 21% |
| *E-mail* | 13% |
| 24/7 | 8% |
| *Web form* | 8% |
| *In-house tool* | 6% |
| Altarama RefTracker | 4% |
| QABuilder 2.0 | 4% |
| Six other services | 2% each |

# Fields Collected by Services

| Patron Information | Overall | Web form | Chat |
|---|---|---|---|
| E-mail (IP address ) | 77% | 90% | 68% |
| Name | 72% | 80% | 68% |
| Country | 36% | 65% | 20% |
| State | 34% | 55% | 24% |
| Member of Organization | 34% | 35% | 32% |
| City | 32% | 55% | 20% |
| Educational level | 30% | 40% | 28% |
| Phone number | 23% | 25% | 16% |
| Professional Role | 23% | 30% | 16% |

# Fields Collected by Services

| Question Information | Overall | Web form | Chat |
|---|---|---|---|
| Text of question | 93% | 100% | 88% |
| Date | 91% | 95% | 92% |
| Time | 85% | 85% | 92% |
| Routing/Referral | 45% | 30% | 60% |
| Subject (free-text) | 43% | 35% | 44% |
| Deadline for answer | 17% | 30% | 4% |
| Desired form of Answer | 11% | 10% | 8% |
| Purpose | 9% | 20% | 4% |
| Prev.consulted resources | 9% | 10% | 8% |
| Subject (from a list) | 8% | 10% | 8% |

# Fields Collected by Services

| Responder Information | Overall | Web form | Chat |
|---|---|---|---|
| Name | 53% | 50% | 60% |
| E-mail | 45% | 35% | 52% |
| Institution | 45% | 45% | 52% |
| State | 34% | 40% | 32% |
| Country | 32% | 40% | 28% |
| City | 28% | 35% | 28% |
| Title | 25% | 30% | 24% |
| Telephone | 17% | 20% | 16% |
| Qualifications | 17% | 20% | 16% |

# Fields Collected by Services

| Response information | Overall | Web form | Chat |
|---|---|---|---|
| Date | 93% | 90% | 96% |
| Text of response | 89% | 95% | 88% |
| Time | 87% | 80% | 96% |
| Resources consulted | 51% | 65% | 40% |

# Services Collecting Different Fields

| | |
|---|---|
| Question Text, Date, Time | 90% |
| Response Text, Date, Time | |
| Patron & Responder Identifiers | 75% |
| Resources Consulted for Response | |
| Routing & Referral Information | 50% |
| Responder Institution | |
| Question Subject | |
| Patron Location | 40% |
| Patron Level | |

# Fields services are Willing to Collect

| | |
|---|---|
| Question Text, Date, Time | 100% |
| Response Text, Date, Time | |
| Patron & Responder Identifiers | |
| Routing & Referral Information | 80% |
| Responder Institution | |
| Resources Consulted for Response | |
| Question Subject | |
| Patron Location | |
| Pre-Session Resources Consulted | 70% |
| Patron Level | |
| Responder Role | 50% |

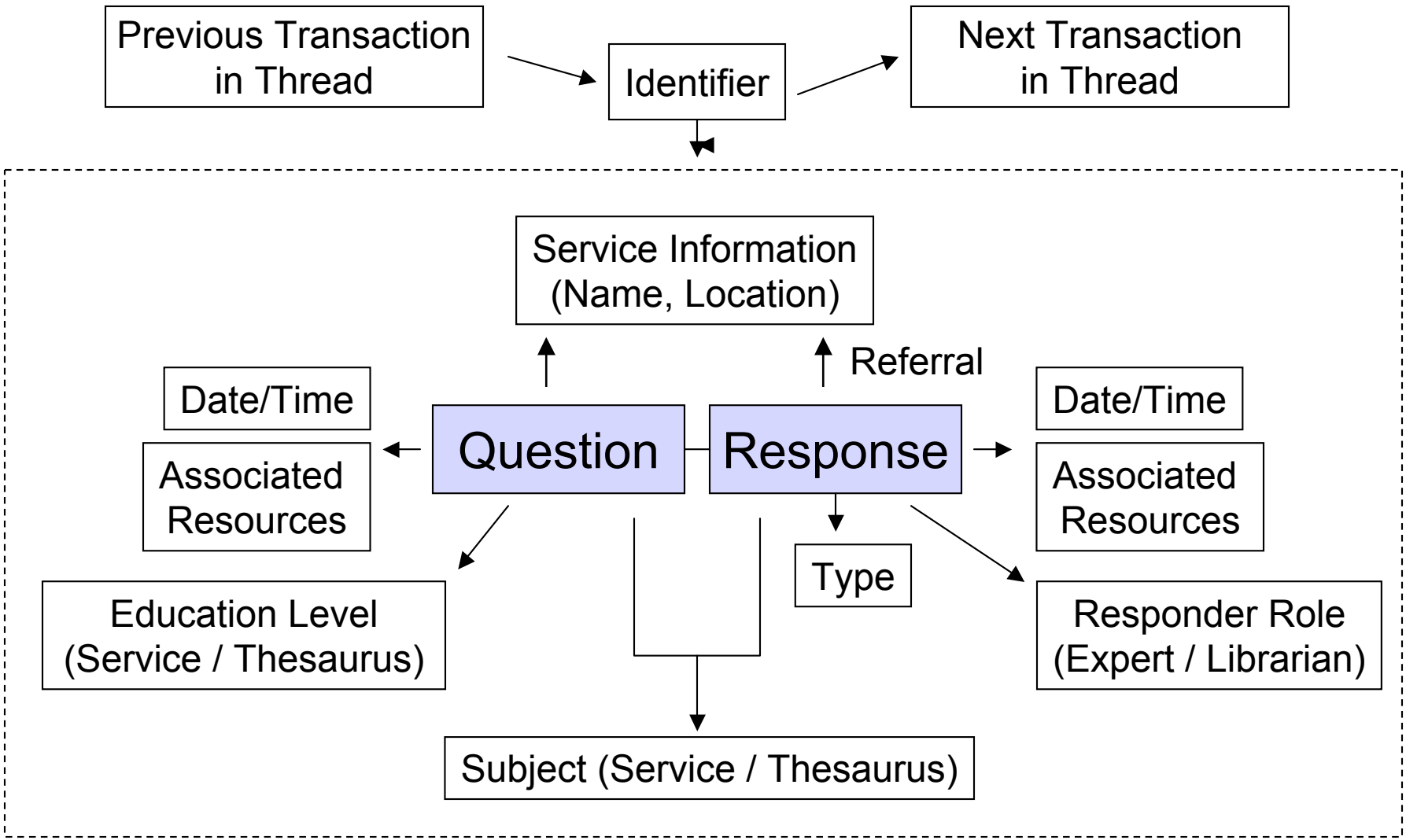
# Observations

- Webform services currently collect more information than chat
- Chat services are *willing* to collect more information than webform

# Constructing the DREW Schema

| |
|---|
| Question Text, Date, Time |
| Response Text, Date, Time |
| ~~Patron & Responder Identifiers~~ |
| Routing & Referral Information |
| ~~Responder Institution~~ |
| Resources Consulted for Response |
| Question Subject |
| Patron Location |
| Pre-Session Resources Consulted |
| Patron Level |
| Responder Role |

# Constructing the DREW Schema



# 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

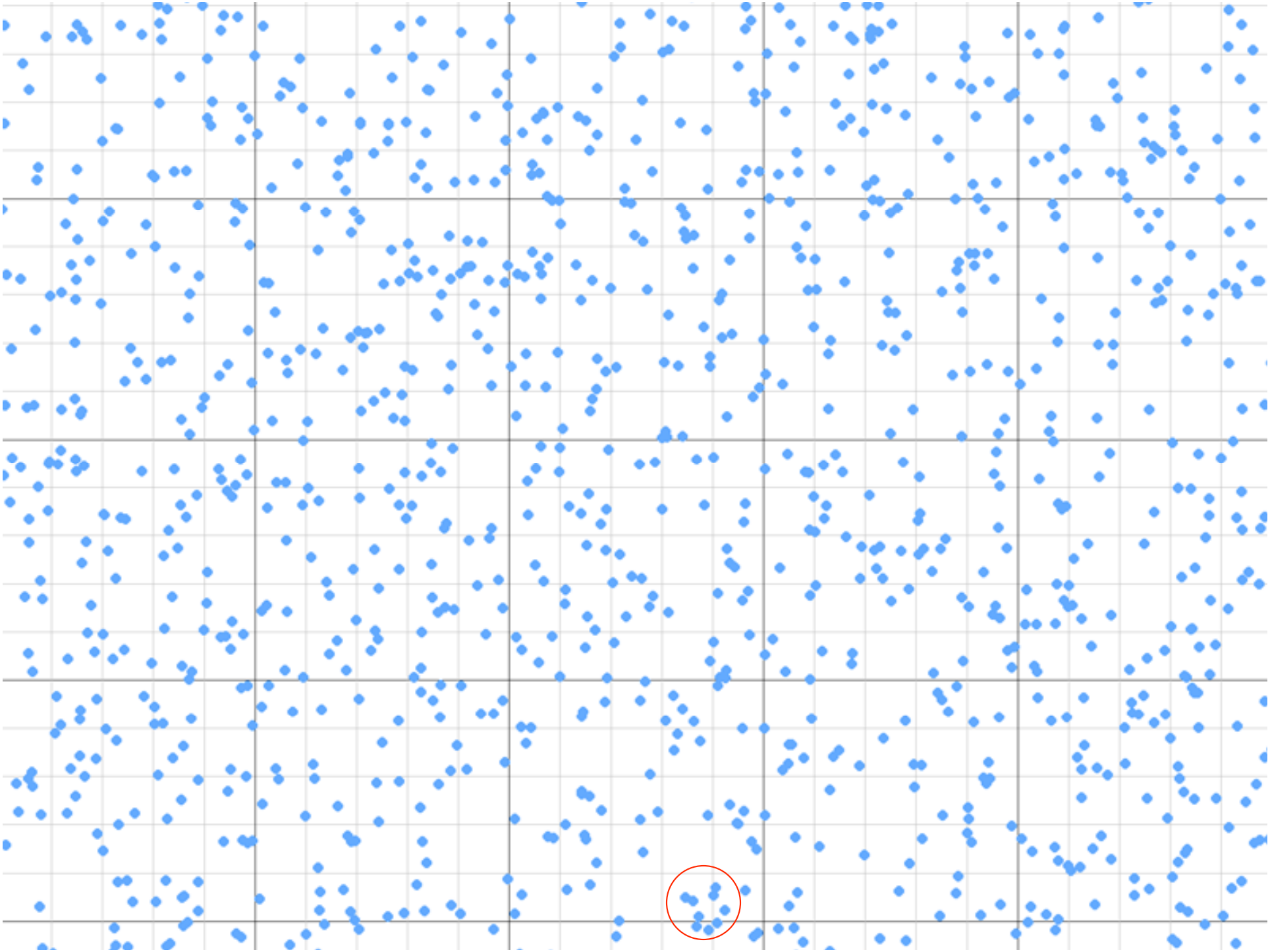
Expert
Patron
Question
Age
Topicality
Ratings
Annotations

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

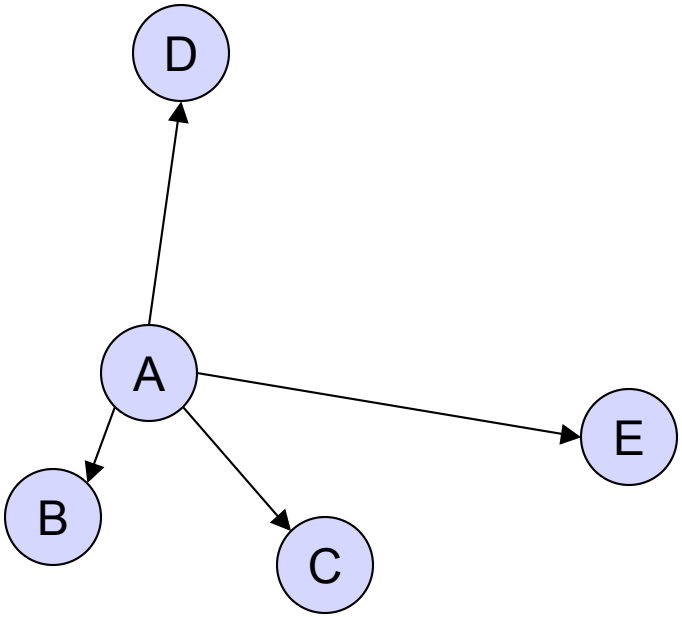
Dave
Anna
Why?
1
300
√√√
100 Links

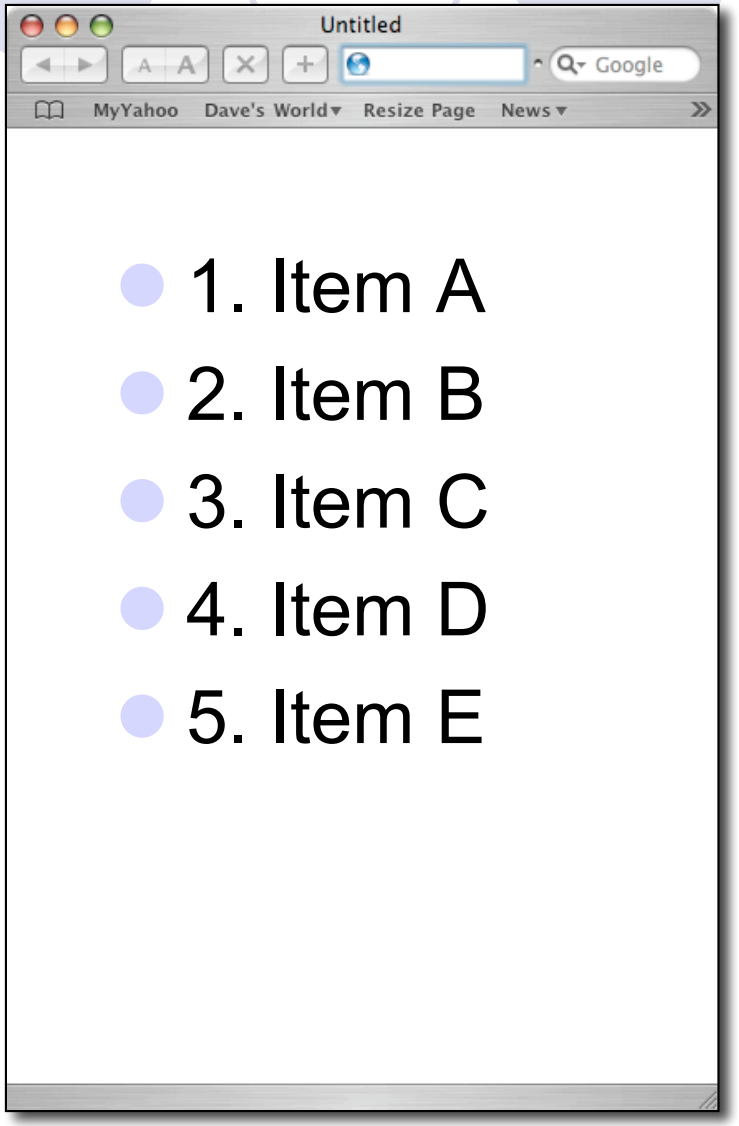
Dave
Anna
Where?
440
56
√
0 Links

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



# From Clusters to Screen



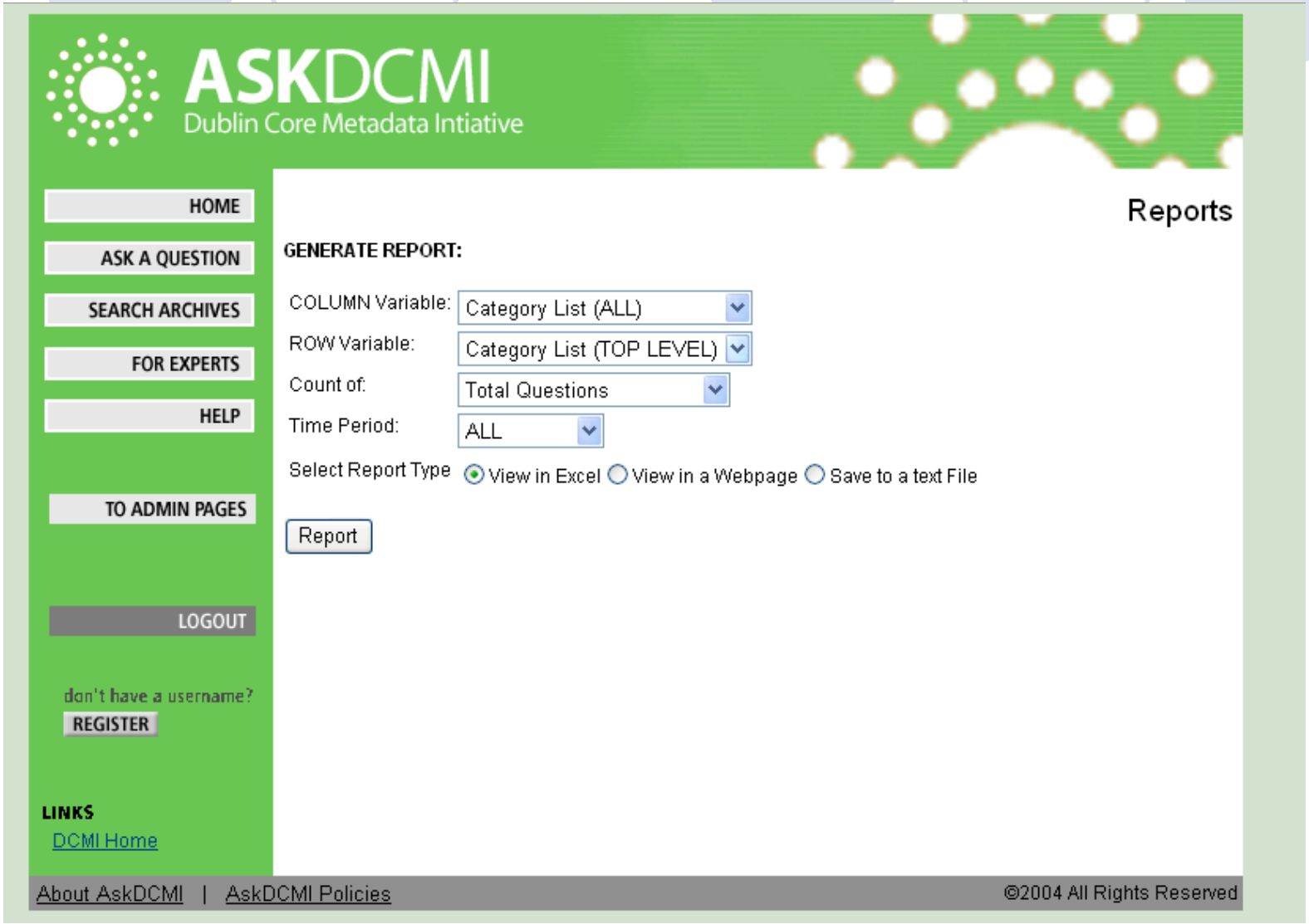
# Privacy

- Current goal = For research only
  - Managers and researchers
- 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

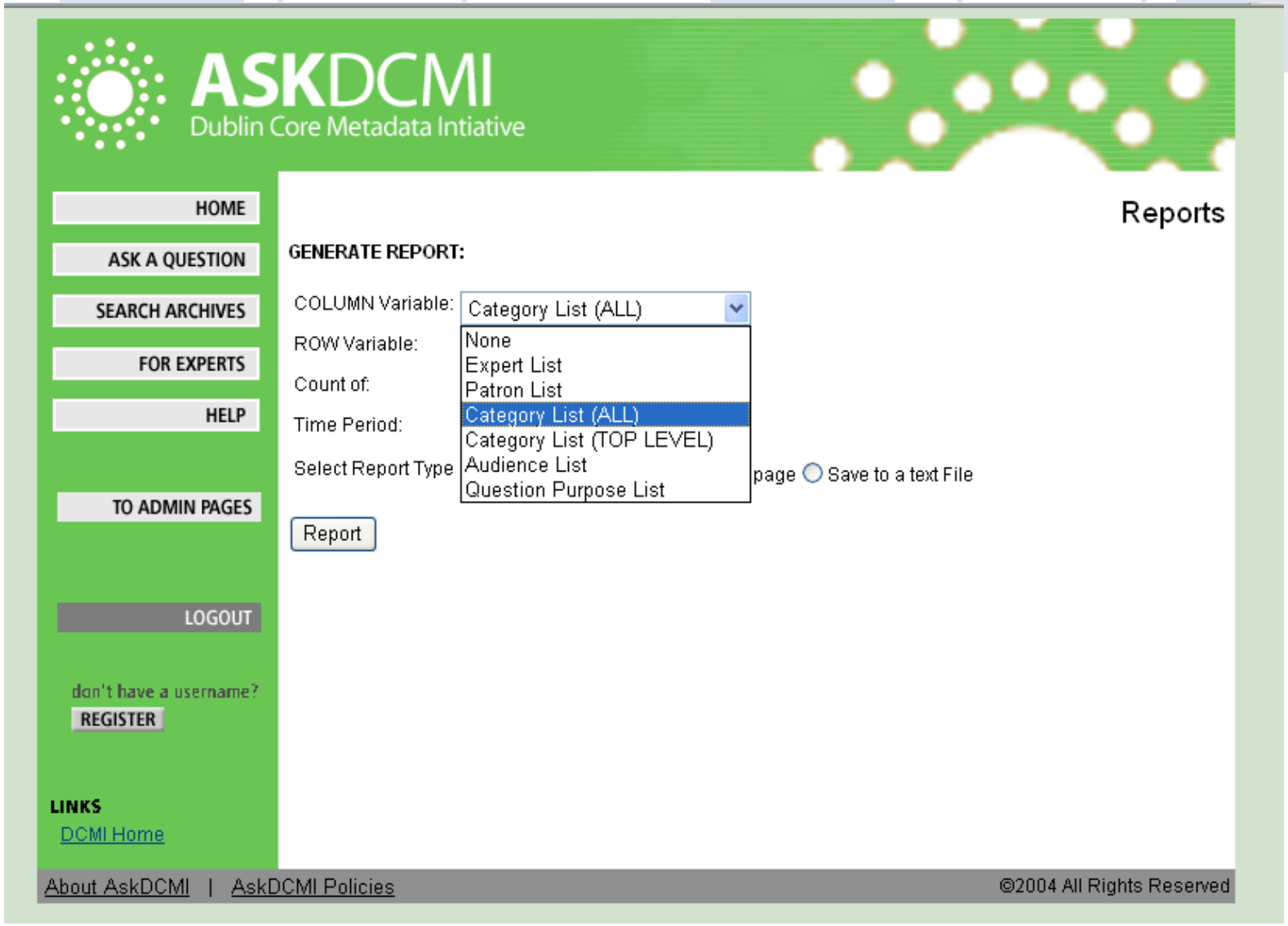
# Usefulness of DREW

- Support of Teaching
  - Work closely with DREI project
- Service Management & Decision-Making
  - Standardized format allows standardized statistical reports and bibliomining tools
  - Individual service and consortial level
- Support of Research
  - Information seeking
  - Human intermediation
  - Connecting resources to questions & topics

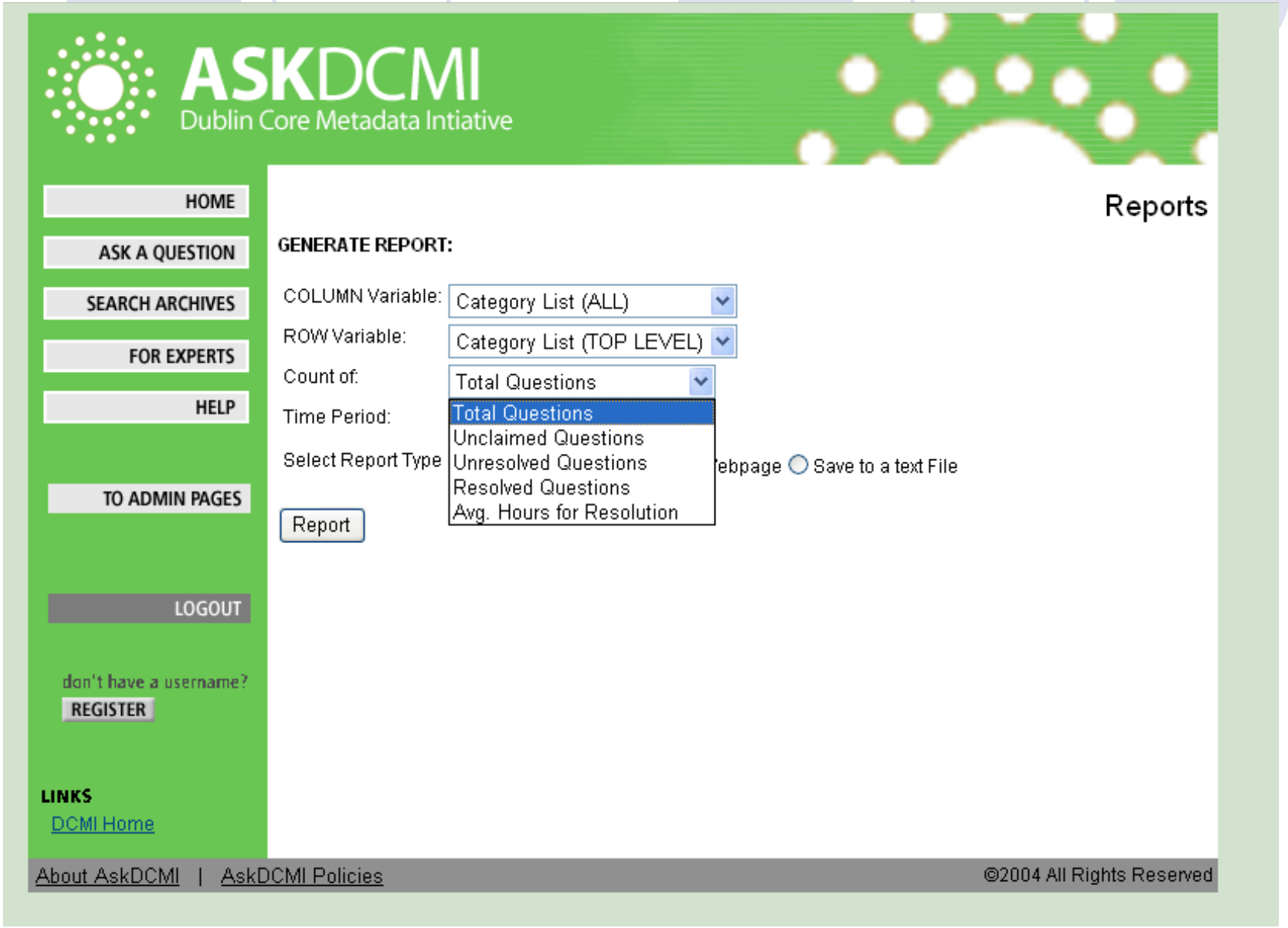
# QABuilder Data Exploration Tool



# QABuilder Data Exploration Tool



# QABuilder Data Exploration Tool



# Complexity Theory and DREW

- DREW as Complex Adaptive System
- Inductive clustering of transactions
- Self-organizing knowledge bases
- Transaction -> Agent
  - Static and dynamic attributes
- Visualizations of knowledge space

# 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