Creating a Corporate Taxonomy

Internet Librarian 2001

7 November 2001

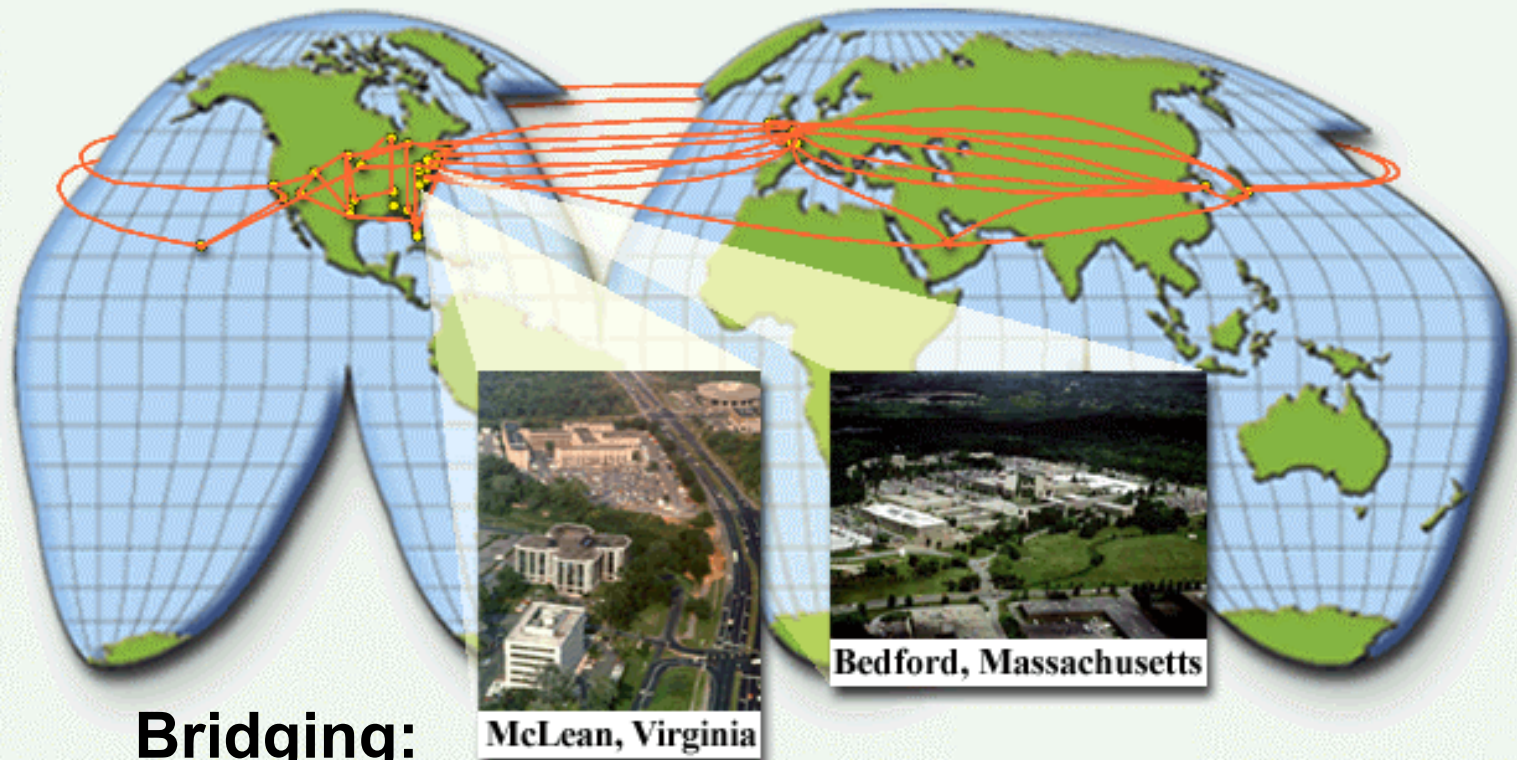**Betsy Farr Cogliano**

MITRE

Background

- **MITRE is a not-for-profit corporation operating three Federally Funded Research and Development Centers (FFRDCs):**

    - **Department of Defense**

    - **Federal Aviation Administration**

    - **Treasury Department**

- **Mission**

As a public interest company in partnership with the government, MITRE addresses issues of critical national importance, combining systems engineering and information technology to develop solutions that make a difference.
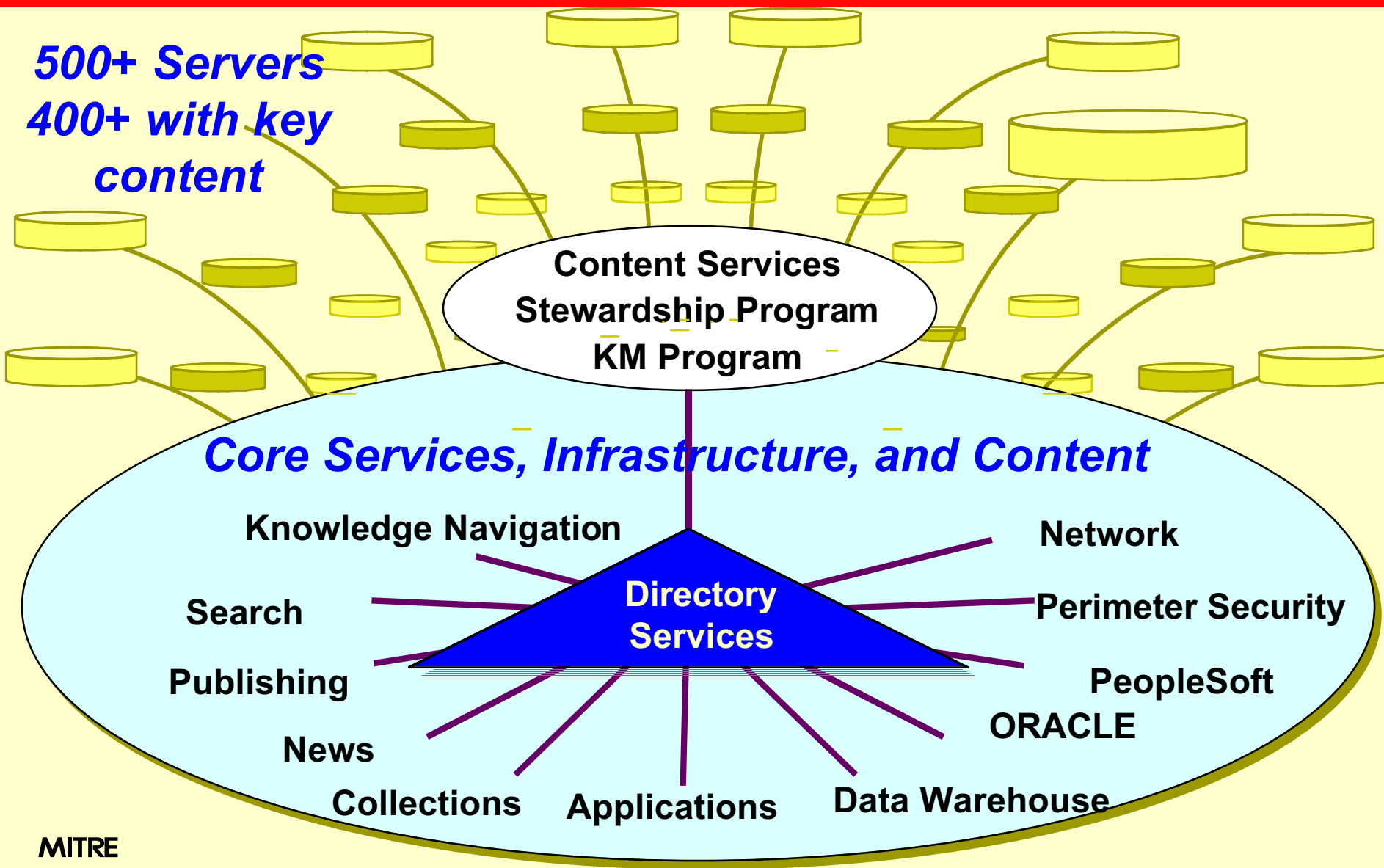
MITRE

MITRE Collaborative Environment

McLean, Virginia

Bedford, Massachusetts

Bridging:
- Technology
- Geography
- Organization
- Time

M

MITRE's Federated Intranet--the MII

**500+ Servers
400+ with key
content**

Content Services
Stewardship Program
KM Program

Core Services, Infrastructure, and Content

Knowledge Navigation

Network

Search

Perimeter Security

Publishing

Directory
Services

PeopleSoft

ORACLE

News

Collections   Applications   Data Warehouse

MITRE

One Definition of Taxonomy

A Taxonomy is a classification of information objects that facilitates the discovery of information.

Taxonomies range from a list of terms on a Web site to a hierarchical arrangement of terms within a Web collection to a controlled vocabulary to a thesaurus.

Taxonomies can be topical, organizational, role-based, etc.

MITRE

Evolution of Taxonomies at MITRE

Primarily Print

Corporate Library System (CLS)

- BASIS Thesaurus Module

- 3 Thesauri in 1
  - LC
  - DTIC
  - MITRE-Unique Terms (Acronyms)

Intranet & Internet

Information by Technical Topic

- ~ 30 Topical Headings
- CLS Terms
- Internal Sources

Knowledge Navigator

- Same Headings
- External Sources

Knowledge Zones

Integration of Internal and External Links
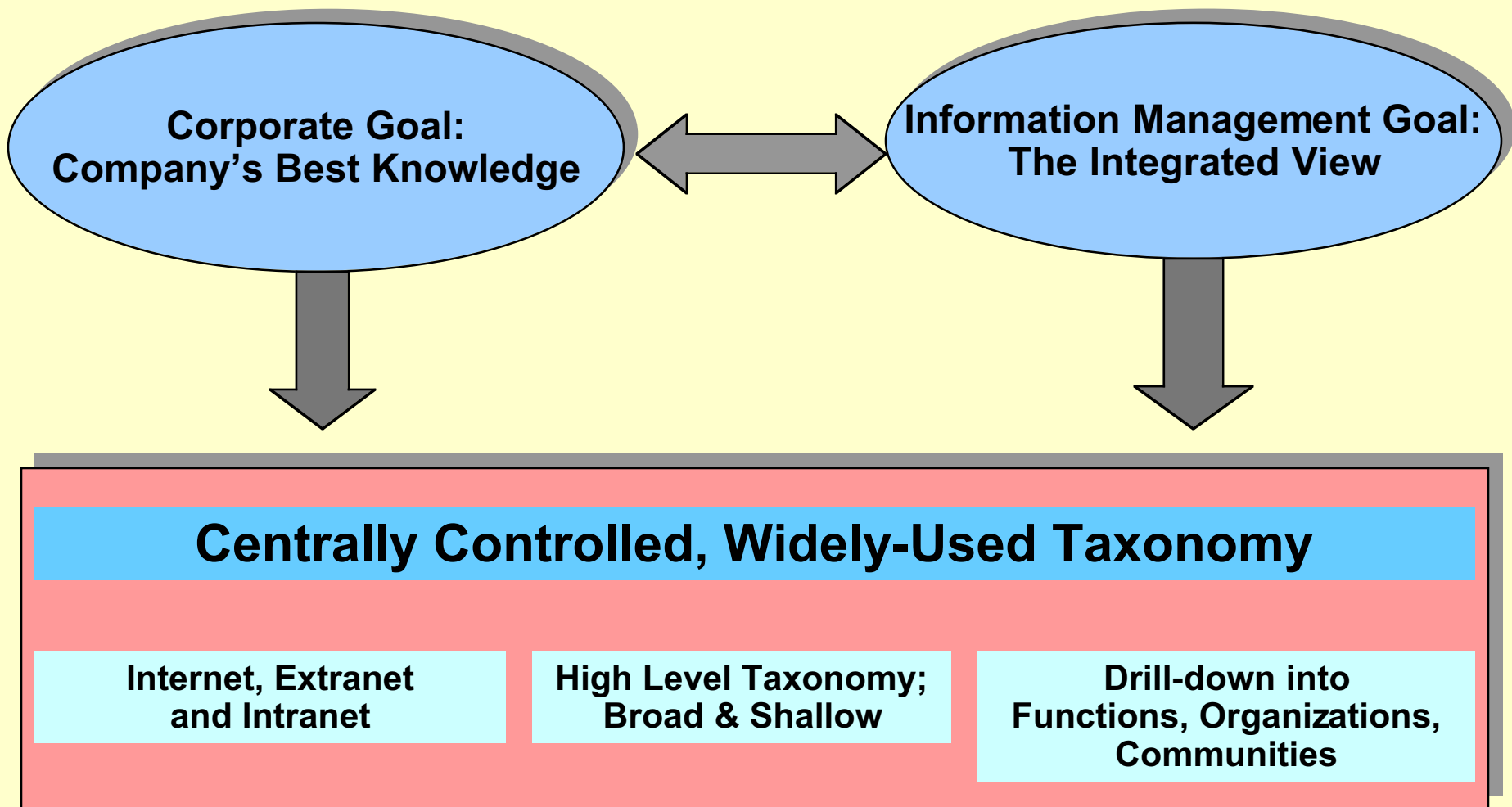
- 18 Broad Categories
- 5 to 15 Sub-Categories
- Based on Usage
- Zone Stewardship
  - Content Stewards
  - Subject Matter Advisors

The MITRE Catalog

Web-based Catalog of Collections and Items

- Applied Metadata Standard
- 30 Top Terms
  - Based on LC
  - All terms relate to at least 1 Top Term
  - Standard for External Crawlers

MITRE

The Need for Taxonomy

**Corporate Goal:
Company's Best Knowledge**

**Information Management Goal:
The Integrated View**

Centrally Controlled, Widely-Used Taxonomy

**Internet, Extranet
and Intranet**

**High Level Taxonomy;
Broad & Shallow**

**Drill-down into
Functions, Organizations,
Communities**

MITRE

A Taxonomy Pilot - Knowledge Mapping

- **Pilot Goals**
  - **Evaluate taxonomy development tools and methods**
  - **Recommend an approach for taxonomy development at MITRE**
- **Methodology**
  - **Phase I**
    - **Identified requirements**
      - **Taxonomy generation**
      - **Categorization**
      - **Visualization**
    - **Evaluated tools**
      - **Sent Request for Information to 8 vendors**

        » No one vendor met all requirements
      - **Chose Semio and The Brain - existing partnership**

MITRE

A Taxonomy Pilot, cont'd.

- **Phase II**
    - **Identified scope of content for pilot taxonomy**
        - **Three taxonomies: organizational, topical, process**
            - » HR Web site (organization) (161 documents; 738K data)
            - » Documents in published spaces (technical topics) (1952 documents; 10 MB data)
            - » Project Management (process) (973 documents; 4MB data)
    - **Identified current navigation tools for comparing to pilot taxonomies**
        - **MII Search (Verity)**
        - **Browsing tools**
            - » Table of Contents
            - » Alphabetical index
            - » FastJump keywords

MITRE

A Taxonomy Pilot, cont'd.

- **Created taxonomies and visual maps**
  - **Manually identified sites or document sets to be crawled**
    - » Ran crawler against these sites to create categories - not successful
  - **Manually defined high level categories using existing sources**
    - » HR Website structure
    - » Technical Area Teams - names of subject specialties
    - » Manager's View of MII - project management area
  - **Reviewed content from crawl and identified concepts within categories**
  - **Crawled document sets and extracted additional relevant concepts; added 1-2 levels within categories**
  - **Ran crawler multiple times to enrich categories and associated concepts**
  - **Output XML tags for displaying maps via The Brain**

MITRE

A Taxonomy Pilot, cont'd.

- **Tested maps with users**
  - **Twenty-four testers in user lab setting**
  - **Performed a set of nine tasks, using**
    - » Maps
    - » MII Search
    - » MII navigation

  - **Answered a series of questions about the maps and their user experience**
  - **Monitors timed testers and noted significant problems and concerns**
- **Collected lessons learned on the taxonomy development process**

MITRE

A Taxonomy Pilot, cont'd.

- **Methodology**
  - **Phase III**
    - **Assessed results - Maps**
      - **Maps performed best in 5 out of 9 tests**
        - » Maps were most effective when looking for topics; less effective when looking for known items are trying to solve a problem
      - **Users found the maps useful in quickly finding information, but the maps were not intuitive to use**
      - **There were some performance problems with some platforms**

MITRE

A Taxonomy Pilot, cont'd.

- **Assessed results - Taxonomy Development Process**
  - **The tool quickly identified important noun phrases and categorized documents into appropriate categories**
  - **Identified documents not grouped and categories not used**
  - **Produced metrics on balancing of categories**
  - **Needed significant up front work in creating categories**
- **Conclusions**
  - **Focus on a topical taxonomy; will provide greatest benefit**
  - **Focus on taxonomy development rather than visualization techniques**
  - **Need further test and evaluation of taxonomy development processes and methods**

MITRE

A Second Pilot

- Goals
  - Further test and evaluate processes and methods for creating a topical taxonomy
  - Recommend a methodology for creating a Corporate Subject Taxonomy
- Methodology
  - Developed subject taxonomy proof-of-concept
    - Determined subject area coverage and scope
      - 5 broad subject areas identified in needs assessment
    - Identified internal and external taxonomies and thesauri to use in concept analysis
    - Selected sample documents for term extraction and clustering
      - Located ~ 500 documents using internal search engines and browsing tools

MITRE

A Second Pilot, cont'd.

- Methodology, cont'd.
    - Developed proof-of-concept, cont'd.
        - Ran sample documents through Semio categorization tool for term extraction and clustering
        - Performed human concept analysis on a sample of the document set
        - Further refined the taxonomy based on this analysis
- Identified Issues
    - Current document storage structures impact the 'availability' of documents for indexing
    - There is no one taxonomy/thesaurus that meets MITRE's needs
    - A taxonomy management tool is needed to collect and control changes to the taxonomy

MITRE

Current Taxonomy Development

- Goal
  - As part of a larger Information Architecture Initiative, deploy a two level subject taxonomy for use in browsing, searching, publishing, and delivering MII content
- Methodology
  - Collect terms
    - Collect applicable terms from a variety of sources, both internal and external
    - Using user focus groups, organize terms into a two level taxonomy
    - Vet results with additional users
  - Re-evaluate automatic categorization tool selection
  - Select a taxonomy management tool
- Coordinate Taxonomy Deployment to Search, Catalog, Subscribe, etc.

MITRE

Future Taxonomy Development

- ● **Extend the Taxonomy into Communities**
  - − **Develop hierarchical vocabularies based on terms used within communities**
  - − **Link vocabularies to taxonomy categories**
  - − **Broaden the taxonomy to include organizations and functions**
    - ● **Human Resources**
    - ● **Administration**
    - ● **Finance**
    - ● **Project Management**
- ● **Deploy the Taxonomy to New Services**
  - − **Publish**
  - − **Profiles**

MITRE

MITRE's Information Architecture

User Interface

MII Portals   Dynamic Views   Activity Views

Current
FY02
Future

Core Capabilities

Full Text Search   Navigation

Security Services   Quality Management

Dynamic Browse   Metadata Search

Document Life-Cycle Mgmt   Workflow

Profile Tools   Community Extraction

Core Processes

Web Content Publishing   Retrieval & Reuse

Project Document Mgmt

Publish & Subscribe   Notification

Automatic Categorization

Collaboration   Online Training

Audience-Targeted Sharing

The Catalog

Registration

Document Attribute Database

APIs to Services

The Profiles

Registration

People Attribute Database

APIs to Services

Information Object Model

Classes   Metadata   Attributes   Taxonomies

Distributed Information Space

Project Share   Transfer Folders   Web Collections   Emails   Electronic Records   Other

MITRE

An Example - Before Information Architecture

Question: What FY01 MITRE projects have a significant KM component?

Process:

MII Search: KM & MITRE in last year

E-mail to all ACs

Browse some possible projects

Ask colleagues

Results:

What do you mean by KM?

Spent 2 hours browsing; no luck!

928 hits; 368 duplicates; 206 document pieces; 292 warranting further investigation; 3 valuable items found

What about experts?

I think Joe's project is working on KM.

Are you including collaboration in KM?

And CoPs?

Have you tried Gxxx?

Issues

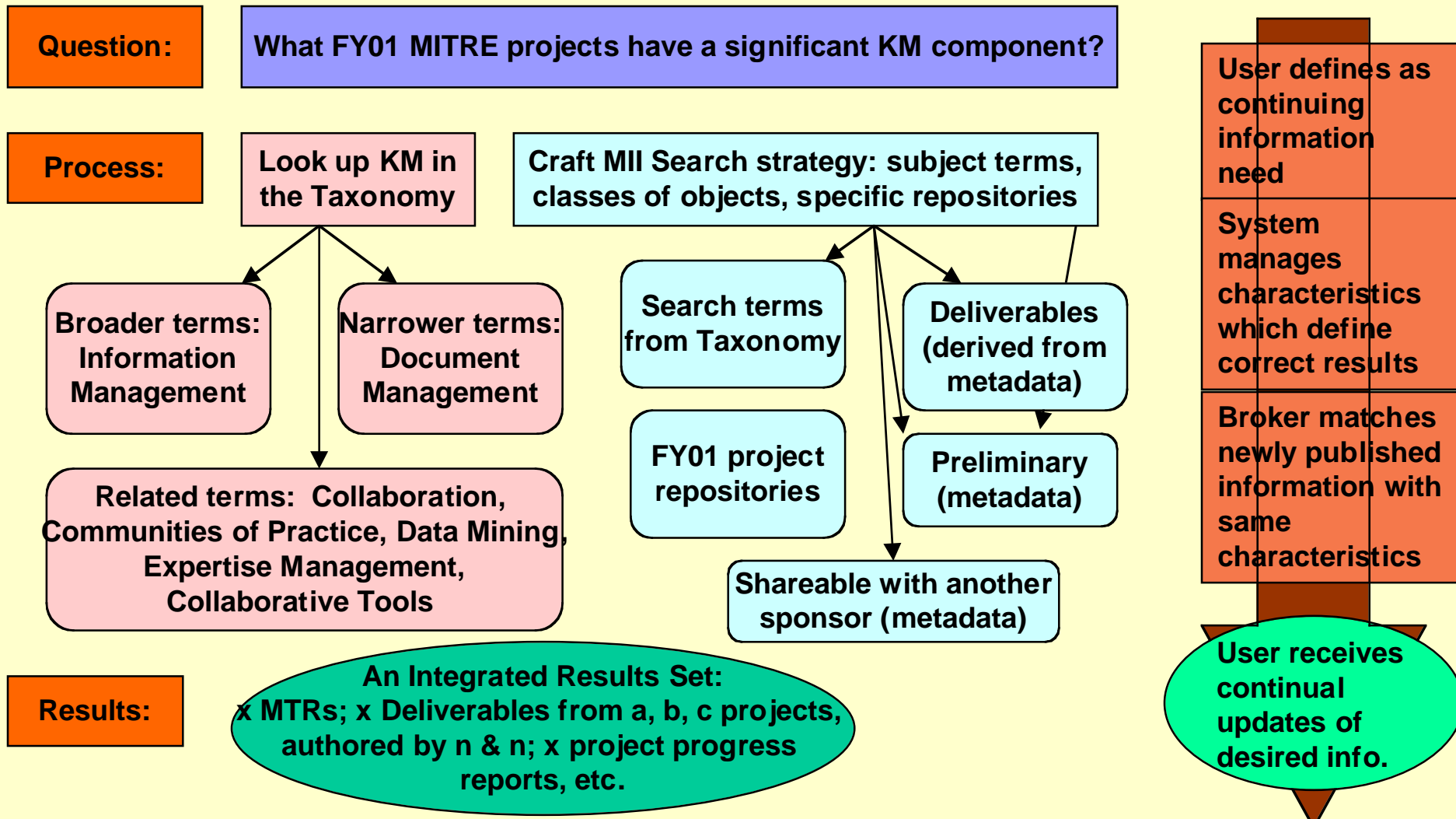| Definition	Definition	Completeness
Repository	Completeness	Productivity

MITRE

With Information Architecture

Question: What FY01 MITRE projects have a significant KM component?

Process:

Look up KM in the Taxonomy

Craft MII Search strategy: subject terms, classes of objects, specific repositories

Broader terms: Information Management

Narrower terms: Document Management

Search terms from Taxonomy

Deliverables (derived from metadata)

Related terms: Collaboration, Communities of Practice, Data Mining, Expertise Management, Collaborative Tools

FY01 project repositories

Preliminary (metadata)

Shareable with another sponsor (metadata)

Results:

An Integrated Results Set: x MTRs; x Deliverables from a, b, c projects, authored by n & n; x project progress reports, etc.

User defines as continuing information need

System manages characteristics which define correct results

Broker matches newly published information with same characteristics

User receives continual updates of desired info.

MITRE

For More Information

- **About MITRE**
  - Visit our Web site at:
    - **www.mitre.org**

- **About the Briefing**
  - Contact Betsy Cogliano
    - **bfc@mitre.org**
    - **781-271-7834**

MITRE