A faceted approach to metadata management

Bradley P. Allen

10 October 2004

navigation
for the
digital
universe

Agenda

- Market research findings
  - Types of metadata usage
  - Approaches to metadata creation and maintenance
  - Perception of available tools
- Efforts at Siderean
  - Faceted navigation, RDF and Web services applied to metadata management

Market research methodology

- 22 interviews with U.S. organizations from August to September 2004.

- Market Segments covered: Life Sciences (Health Care, Pharmaceutical, Manufacturing (Aerospace, Automotive, Chemical, High Tech), Energy, Financial Services (Banking, Insurance), State Government

- Recruited individuals in each of the segments for a discussion of their use of metadata and their pain in finding and using information across repositories

- Discussed their business objectives, approach, technologies, challenges in their implementations, ideal solution

Types of metadata usage

- Metadata usage is not yet widely adopted but is emerging
  - The use and understanding of metadata varied widely across the enterprises interviewed
- Three types of usage were identified:
  - General employee information finding across repositories
    - Dependent on metadata about people and organizations
  - Linking people to distributed business processes
    - Provide the information to complete the process
    - Notion of document type not easily applied
  - Content management/publishing to intranet or portal
- Better information access, not compliance is the driver

Approaches to creating and maintaining metadata

- Top-down: 5 respondents
- Bottom-up: 5 respondents
- Portal-driven: 3 respondents
- None
  - Tried and failed: 3 respondents
  - Haven't tried yet

Top down approach

- CEO says "We must be an information driven company" which drives an effort
- "We'll develop a corporate controlled vocabulary that all divisions will use"
  - Most mentioned use of Dublin Core
  - Most creating their own controlled vocabularies for subject tagging
- The effort is multi-year, ROI hard to track, and may not be implemented or adopted widely

Bottom up approach

- Groups determine their vocabulary describing their process
  - Often in a collaboration environment
- Light tagging of content when it is created or when the content is published to a portal
  - Again, most mentioned use of Dublin Core and creating their own controlled vocabularies

Portal-driven approach

- "Using a portal we have an integrated information source for our internal process users or external customers"
- "Now we have to architect the information related to business processes differently to search across multiple repositories"
- But they lack tools and applications that support this

Tried and failed

- Started from content popping up everywhere and think that it is too difficult to create metadata from existing content
- "We can't ever hope to organize this morass of content, so let's put in a search appliance like Google"
- "Our internal needs are like the public internet and users are familiar with Google searches"
- But still feel that metadata would improve matters, particularly within business units

Use of metadata management tools

- Perceived lack of tools to apply metadata to content
- The process and tools are very disconnected.
  - Point solutions not well integrated
- End to end solution not really available, much custom development done
  - Not easy to tag or incorporate content into the desired structures
  - No easy way for groups creating the vocabularies to deliver them to production environments

Desired solution

- Capture existing metadata cheaply and make it easy to enhance its value, but get started with what's easily available

- Plug "governance" gap between portals, search engines and repositories

- Provide metadata-enhanced information access components that can be simply incorporated into web applications and portals

- Incorporate usage feedback loop to further enhance value of metadata

Efforts at Siderean

- Developing an end-to-end platform for faceted navigation of corporate information

- Finding that faceted navigation has an enabling role to play in metadata creation and maintenance as well as delivery

Faceted navigation and metadata management

- Easy query construction in faceted UIs unify and simplify many maintenance tasks when coupled with basic operations
  - Tagging rule specification
  - Train/test set construction
  - Cleansing
- Faceted approach eases taxonomy development
  - Shifts focus from monolithic taxonomies to smaller vocabularies, schemas and application profiles
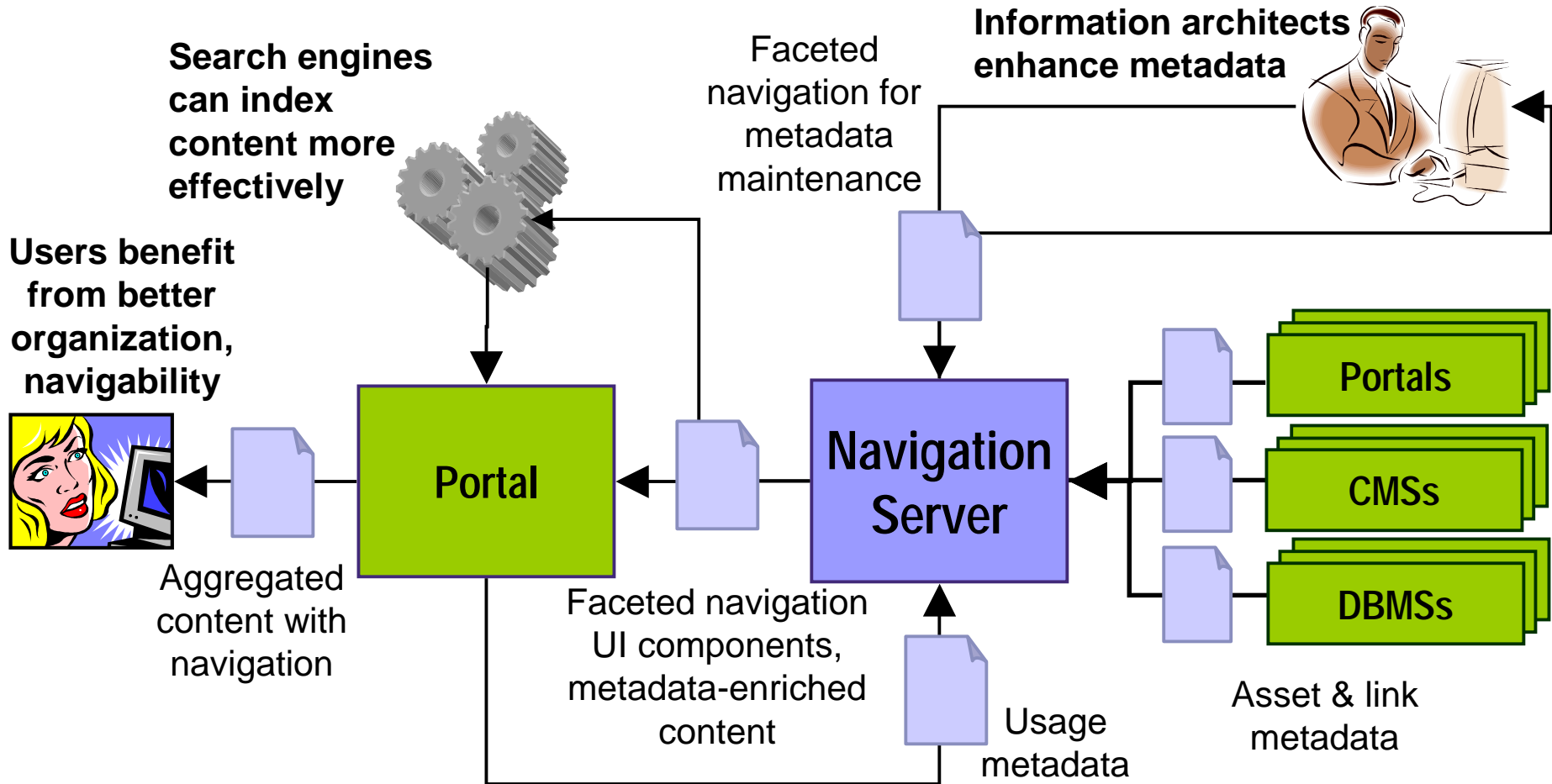
RDF and metadata management

- RDF unifies the entities used in creation and maintenance so that a standard set of tools can be applied across them
  - Ontologies
  - Controlled vocabularies
  - Documents
- RDF advances the notions of reusability and open standards to the benefit of maintainability

A virtuous metadata enhancement cycle

Search engines can index content more effectively

Faceted navigation for metadata maintenance

Information architects enhance metadata

Users benefit from better organization, navigability

Portal

Navigation Server

Portals

CMSs

DBMSs

Aggregated content with navigation

Faceted navigation UI components, metadata-enriched content

Usage metadata

Asset & link metadata

Siderean™
SOFTWARE

navigation for the digital universe