Federating Knowledge Searches: Integrating the Enterprise

Jeanne Holm
Chief Knowledge Architect
NASA/Jet Propulsion Laboratory

January 17, 2006

Including presentations from Jayne Dutra, Andrew Schain,
Rob Raskin, Brian Wilson, Rich Keller, and Walt Truszkowski

Objectives

Collaborate

Communicate

Innovate

Motivate

- Address human factors for delivering the right information to the right people to avoid duplicative research and efforts
- Uncover the ability of semantic search and ontologies to bring the most relevant information to the top
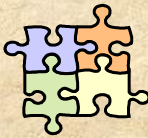- Re-utilize your existing infrastructure to reach deeper and broader into your organization's resources to decrease costs in IT
- Knowledge management is getting the right information to the right people at the right time, and helping people create knowledge and share and act upon information in ways that will measurably improve the performance of an organization and its partners

January 17, 2006   NASA KM   2

Defining the Jargon

- Taxonomy
  - Structured view of an organization's (NASA, Boeing, PRW) information
- Ontology
  - Categorical view of a knowledge domain that may transcend organizational definitions (Earth science, physics, astronomy, mechanical engineering)
- Semantic Web
  - Today's Web is made for people to read and understand
  - The Semantic Web is being structured for computers to read and understand
    - Systems can perform transactions across applications without human help
    - Leverages the vast amount of data accessible on the Web for machine processing

Collaborate

Communicate

Innovate

Motivate

January 17, 2006   NASA KM   3

Now and Later Approaches

- Today we provide an integrated infrastructure that allows sunk costs and best-of-breed approaches
- Tomorrow, we are pursuing two additional solutions
  - Provide dynamic modeling of categorical data to place content with context (with assumptions)
  - Manage data problem through an Integrated Services Architecture that emphasizes structuring the data and solving deep-seeded problems of data integrity and integration
  - In reality, each approaches the problem from a different and complementary point of view
    - Integrate with business processes and policies
    - Keep independent of any specific repository or technology
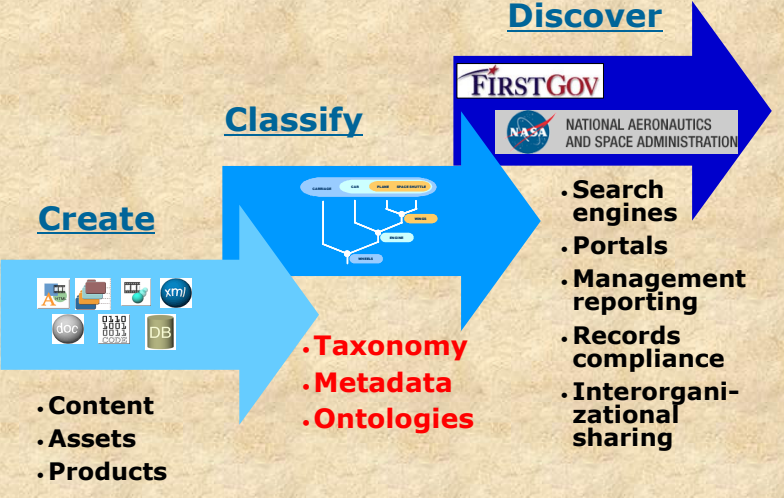   - Deal with harmonization of information according to a larger point of view

Collaborate

Communicate

Innovate

Motivate

January 17, 2006   NASA KM   4

Content Publishing Life Cycle

Discover

Classify

Create

FIRSTGOV

NASA NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

- **Search engines**
- **Portals**
- **Management reporting**
- **Records compliance**
- **Interorgani-zational sharing**

- **Taxonomy**
- **Metadata**
- **Ontologies**

- **Content**
- **Assets**
- **Products**


Creating a Learning Organization

- ◆ Integrated approach to ensuring best practices and key lessons learned are applied on missions
  - – NASA Engineering Network
    - • Capitalizes on best ways engineers currently work, while solving cultural and process areas that NASA for which has been criticized
    - • Builds on shared infrastructure and seamlessly integrates with NASA initiatives, distributed systems, and KM infrastructure
    - • Distinguished by integrating lessons and learnings that come out of engineering discussions and repositories into day-to-day engineering processes, policies, and training curriculum
    - • Integrates information broadly from academia, industry, contractors, government, and NASA personnel
      - – Portals to organize community and individual access to information
      - – Collaborative tools expanded for secure access with our partners
      - – Expertise and expert directories organized around sharing knowledge person-to-person over virtual social networks
      - – Metasearch across distributed repositories

Learning Process Occurs Behind All Components: Embed learnings into tools and communities
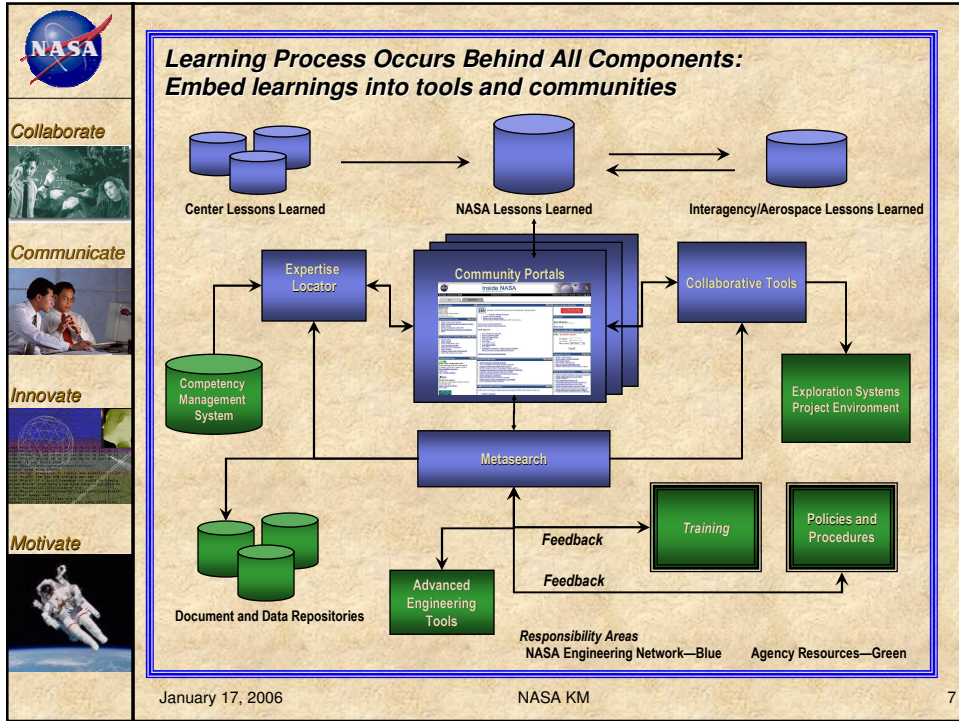
Accessing and Gathering Lessons Learned

Slide 9

Communities for Collaboration

Collaborate
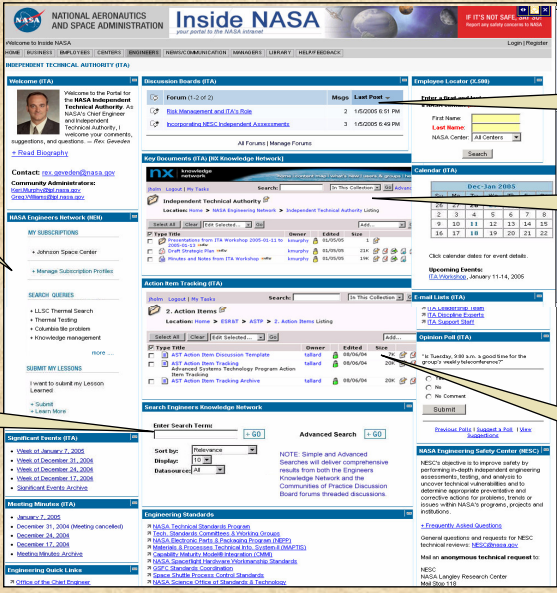Communicate
Innovate
Motivate

Saved searches and subscriptions

Find information

Discussions and Q&A

Integration to document management

Action item tracking

January 17, 2006 — NASA KM — 9

Slide 10

The NASA Public Portal

Collaborate
Communicate
Innovate
Motivate

- Was designed and intended to be a dramatic, interactive interface to NASA by the public, kids, media, educators, and students
  - Create "One NASA" on the web to find content faster and easier
  - Exemplar of the President's Management Agenda
  - Tie together NASA's public-facing web resources
- Streaming live coverage, dynamic and distributed publishing, and automatic image upload brought fresh images within minutes of the spacecraft sending them to Earth
- Rich, interactive media at the home page helps people see NASA's message and understand our discoveries
- Great partners such as eTouch Systems Corporation
- Largest government site in 2005
  - 17.5 billion hits
  - 140 million unique visitors
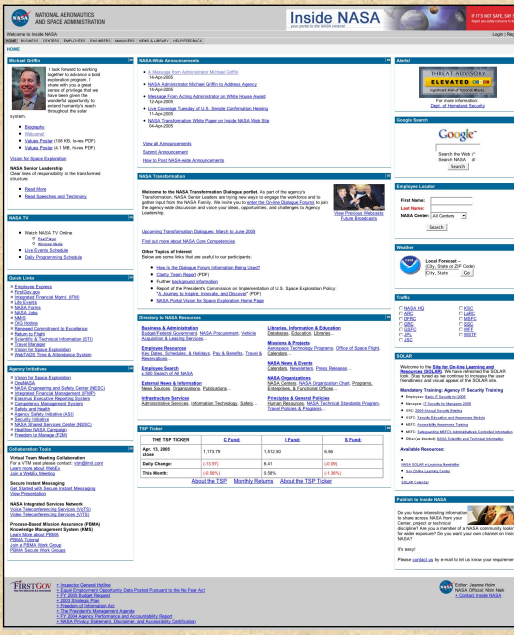 - 300 TB of data

January 17, 2006 — NASA KM — 10

Inside NASA

- Intended for employees and partners
- Customizable
- Access to e-mail
- Instant messaging
- Collaborative tools

What Defines Reuse?
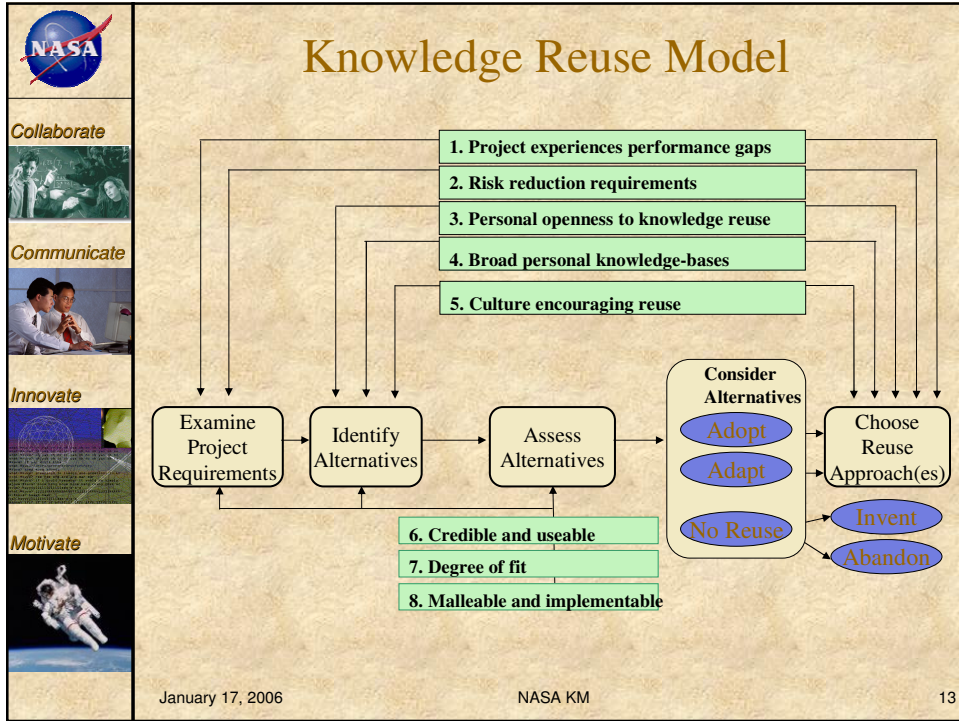
- Content (images, records, sites) that is accessed, understood, or modified to create new content or learning
- Some is measured (visits, downloads, feedback), some anecdotal, and some difficult to measure and track
- Confounded by human nature--sometimes people do not remember what led them to certain conclusions, other times downloaded information is not used or useful
- Study by Ann Majchrzak (USC), Olivia Neece, and Lynne Cooper (next slides are derived from their study)
- Two Mars projects with 6 targeted technologies

Adopt   ← →   **Adapt**

Knowledge Reuse Model

1. Project experiences performance gaps

2. Risk reduction requirements

3. Personal openness to knowledge reuse

4. Broad personal knowledge-bases

5. Culture encouraging reuse

Examine Project Requirements → Identify Alternatives → Assess Alternatives → Consider Alternatives (Adopt, Adapt, No Reuse) → Choose Reuse Approach(es) → Invent, Abandon

6. Credible and useable

7. Degree of fit

8. Malleable and implementable

January 17, 2006 — NASA KM — 13

Reuse: Community Management

- Expert or manager is leading an online community and reaches to and reuses the knowledge of the group
  - Promotion and adoption of best practices
  - Disseminate knowledge
  - Facilitate sharing across Centers and disciplines
  - Process for waivers and deviations
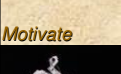 - Review documents and standards
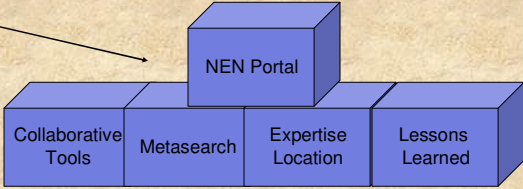 - Encourage open discussion

January 17, 2006 — NASA KM — 14

7

Reuse: Lessons Learned

- ◆ Engineer capturing and sharing lessons learned
  - – Publish and find lessons learned and relevant information
  - – Access content that represents previously tacit and informal knowledge
  - – View and promotion of discussions, exemplars, and documents to official lessons learned
  - – Individual learning new skills in a community
    - • People and tools to enable informal, Socratic method of learning new skills (engineering discipline portals and collaborative tools)

NEN Portal

| Collaborative Tools | Metasearch | Expertise Location | Lessons Learned |

January 17, 2006   NASA KM   15

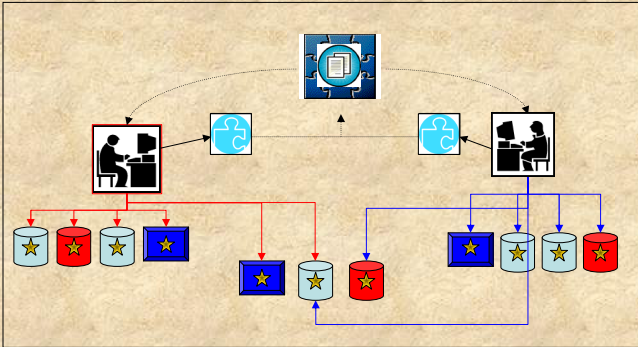
Reuse: Joint Analysis

- ◆ Understanding what others are searching for and connecting similar searchers together to gain greater understanding and better analysis

January 17, 2006   NASA KM   16

8

Taxonomy: A Solid Foundation

- ◆ RDF files apply metadata and controlled vocabularies to content
  - – http://spongebob.siderean.com:7880/test/test2query3.jsp
- ◆ Collection from different sources
  - – Faceted search using the Seamark (Siderean Software)
  - – For more information, contact Jayne.Dutra@jpl.nasa.gov

Facets	Strategic Value
Access Requirements	Sensitivity and access control
Audiences	Who the content is intended for
Business Purpose	Why the content was created
Competencies	Relevant field or discipline
Content Types	The genre of the content
Industries	External partners and businesses
Instruments	Flight payloads that yield science
Locations	Sites where work occurs – on and off Earth
Missions/Projects	NASA's lines of business
Organizations	NASA organizations
Subject Categories	The topic of the content

January 17, 2006   NASA KM   17

NASA Taxonomy Site and Resource

Link to Metadata Specification

Link to XML DTDs and RDFs

Background and training materials

Links to Controlled Vocabularies

http://nasataxonomy.jpl.nasa.gov

January 17, 2006   NASA KM   18

Search Across Two Partners

Agency Index

Common Metadata Catalogue
DMZ/www

Contractor Index

| Schema Translation Models | Metadata Rationalization Rules | Enterprise Ontology Models | Common Vocabularies |

Semantic Hub

Agency User

Contractor User

Collaborate

Communicate

Innovate

Motivate

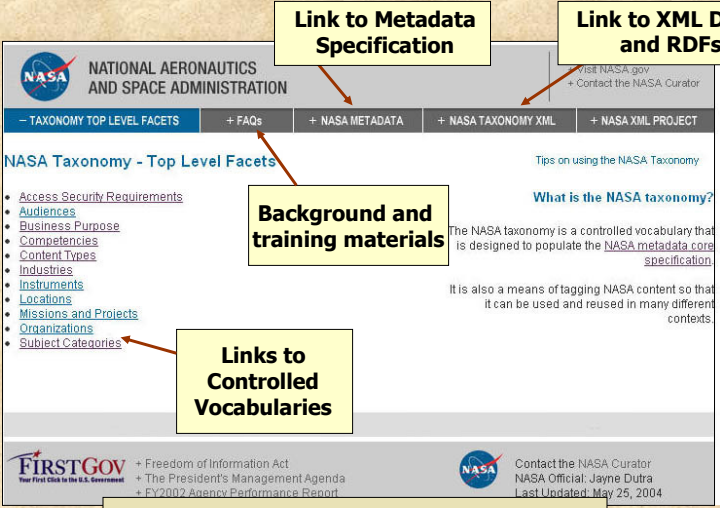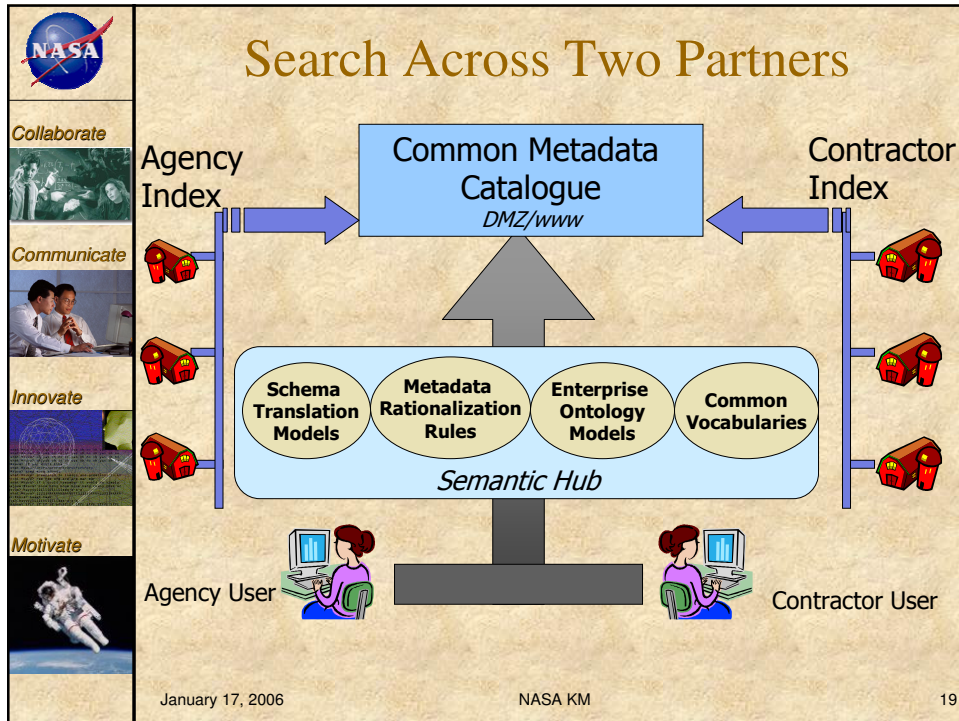---

Information Architecture Approach

Collaborate

Communicate

Innovate

Motivate

- In addition to pursuing taxonomy work, the underlying data problem can be resolved through an Information Services Architecture
- Linking, expression, extension, and relevance
  - Where information about an application, a service, a dataset is available we should tag it or add it incrementally
  - Tags require context, relationship and meaning which enables inference and pattern recognition
- "Annotate and share"
  - Provide a mechanism where subject matter experts add metadata or context and leave it for others to build on incrementally
  - Strategy enables data and information customers to drive incremental metadata organization based on their needs
  - Each valid construct can be left for others to reuse and repurpose

Orchestrated Tagging and Organizing

- Making it discoverable
  - Search, browse, and query
  - Machines connecting the dots, not just people
  - Organize metadata so that logic and relationships within applications or vital contextual instrument information is maintained
- Leverage what we have and what we know
  - Reuse and leverage available ontologies (e.g. Swoogle)
  - Import and translate existing schemas into RDF ontologies or OWL depending on requirements and opportunities

Collaborate

Communicate

Innovate

Motivate

A Little Bit of Infrastructure

- Knowledge representation (KR) libraries
  - Establish manageable, semantically rich official libraries
  - XML schemas, RDF, ontologies, thesauri, and taxonomies need to be valid, trustworthy, and available for reuse
- Service advertisement repositories
  - Enable computers to know when new service is online, understand what it does, employ its functions as part of generalized tasks
- Metadata collection and KR construction
  - Provide tools that harvest existing or assert new metadata
  - Evaluate natural language processors help determine metadata elements and simple mechanisms for customers to add semantic annotations
- Drawer of kitchen utensils
  - Have code that adds metadata awareness to infrastructure components
  - Collaboration awareness for Wikis and so on
  - Generate RDF from office-type applications and ontologies from database schemas

Collaborate

Communicate

Innovate

Motivate

Establish "Attractor Services"

- The *network effect* describes how a service becomes more valuable as more and more people adopt it. As more services and capabilities get incorporated, it motivates more individuals and more services to participate. The more services we tie together, the greater the utility. The greater the utility, the more services get incorporated.
  - Linking people, organizations, projects, and skills (POPS via the NASA Engineering Network)
  - Metadata search and inference in image inventories
  - Federal Enterprise Architecture and capital investments
  - Semantically enriched document management
  - Integrating science knowledge
  - Semantically enabled workflows

Information Architecture Benefits

- Overall order, consistency, predictability in how we handle common project development products and artifacts
  - Improved search and retrieval abilities
  - More accurate status reporting to management
  - Identification of design and engineering rationales
  - Better decision making – reuse of information
  - Improved problem solving abilities
  - Quicker turn around of trade studies and design development
- For more information contact Andrew.Schain@nasa.gov

Semantic Web Technologies Future

- The big question: is this real or can it be made real?
- What are *you* going to do about concerns about privacy and aware of how people react to the ability to find information in new ways

---

Thanks!

- Many thanks to my colleagues and our partners who contributed to these ideas and to the excellent work they are doing in implementing knowledge management solutions at NASA, particularly to Andy Schain and Jayne Dutra
- If you have any additional questions, please contact me
  Jeanne Holm, NASA/JPL
  4800 Oak Grove Drive, Mail Stop 202-204, Pasadena, CA 91109
  Jeanne.Holm@jpl.nasa.gov   (818) 354-8282
- More information can be found about
  – NASA's KM program: http://km.nasa.gov
  – NASA's portal: http://www.nasa.gov