A Semantically Structured Information Repository for Distributed NASA Project Teams

Richard M. Keller, Ph.D.

Principal Investigator

Semantic Technologies Applied Research Team
Information Sharing and Integration Group
Intelligent Systems Division

NASA Ames Research Center

Supported by NASA Programmatic R&D Funds

http://sciencedesk.arc.nasa.gov *rkeller@arc.nasa.gov*

Contributors

Core Team

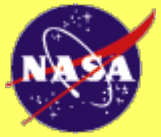
- Dan Berrios
- Robert Carvalho
- David Hall
- Rich Keller
- Ian Sturken
- Shawn Wolfe

Former Teammates

- Jim Chen
- Mike Compton
- Jon Guice
- Dennis Heher
- Kim Hubbard
- Larry Kiser
- Deepak Kulkarni
- David Nishikawa
- Steve Rich
- Keith Swanson
- Sergey Yentus

Key Engineering and Science Collaborators

- Brad Bebout
- Dave Blake
- Dave Des Marais
- David McKay
- Tina Panontin
- Maarten Sierhuis
- James Williams



Outline

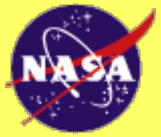
- Introduction & Motivation
- SemanticOrganizer system
- Applications
- Lessons learned

Motivation:
Information Sharing across Collaborating Teams

- **Large distributed NASA project teams must work together to address complex science and engineering questions**
- **Examples of distributed "virtual" teams:**
  - **NASA Astrobiology Institute**
  - **NASA Accident Investigation Teams**
  - **NASA Mission Design Teams**
- **Paramount need: *<u>to share complex science & engineering data/information across teams</u>***

What solutions are employed?

(e.g., use FTP Server, WebDAV, Document Server, Content Management System)



Our Contention: Conventional file-sharing solutions are insufficient!

- Little discipline or organization to files; no “librarian”
- Lack of metadata
- Lack of shared terminology/vocabulary for naming files, folders, or metadata
- Many files not inherently indexable/searchable via text-based approaches (lots of data & images – not much text)
- Complete reliance on hierarchical file structure for indexing – no multidimensional cross-indexing

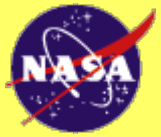
"Islands of Information"

Co-locating information is not enough!

- Disconnected "islands" of information
- File-shared information is:
  - not integrated
  - not interconnected
  - not contextualized

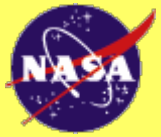
  } *difficult to 'connect the dots' and see the big picture*

- Difficult to see the connections between data/information gathered by different people in different places



Desiderata for an Improved Knowledge Sharing System

- Integrate/interrelate information across the “islands”
- Permit rapid, intuitive access and search of the information space
- Provide high degree of customizability to accommodate each project’s unique work products, processes, and goals
- Enhance consistency and completeness of shared information
- Capture contextual information associated with shared information & data products



Model-Based Approach to Information-Sharing

1. Centralize: Establish centralized access to all project information within a common repository
2. Standardize: For a given distributed project team, *engineer* a common information model and vocabulary for describing and interrelating all project information
3. Contextualize: Annotate the data with ample semantic context to provide a basis for improved access and search

Develop a Semantically-structured Information Sharing Repository

What is a Semantic Information Repository?

An information management system

(i.e., database / web server / file server / document mgmt. system)

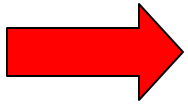
+

an ontology!

Repository users (people *and* intelligent agents)**:**

- ❑ **Upload / enter heterogeneous content:**
  1. datafiles/documents/images
  2. *records* documenting people, places, things, activities that provide *context* to data
- ❑ **Interlink / interrelate content** within a ***semantic information network***
- ❑ **Search, access, and download content**

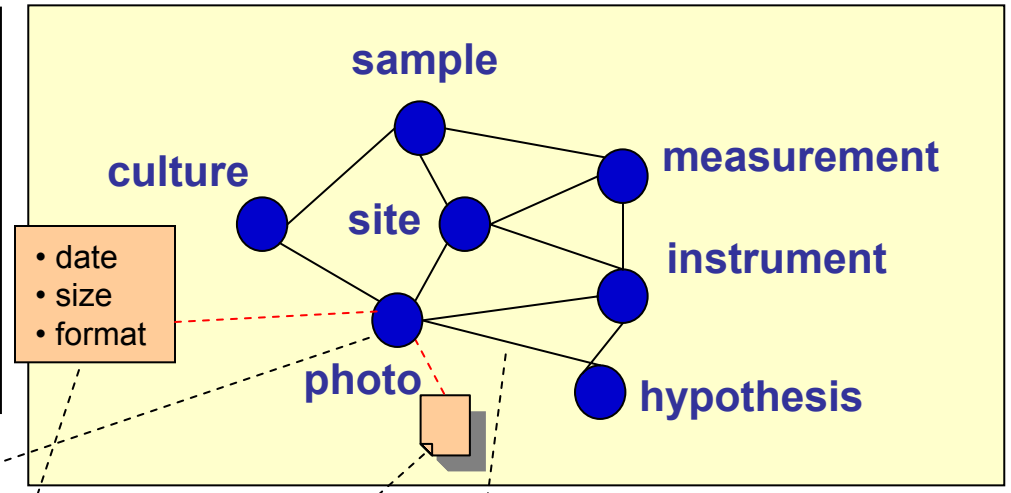
Outline



- Introduction & Motivation
- SemanticOrganizer system
- Applications
- Lessons learned

SemanticOrganizer Core Representation: Semantic Network Structure

Ontology: Specifies the types of nodes, attributes and links defined for project team



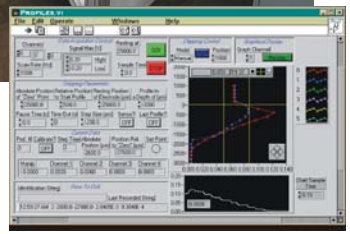
Rules: Add/modify nodes, links & attributes in the network

- **Nodes**: key project resources (e.g., people, places, artifacts, activities, processes)
- **Attributes**: properties of resources (metadata)
- **Links**: relationships among resources
- **Attached files**: electronic products associated with resources (in almost any type of format)

SemanticOrganizer Sample Application: Astrobiology Research Investigations

Microbial mat (algae)

Collection of microbial mats in the field



Detailed studies of mat biogeochemistry

- monitoring
- analysis
- experimentation

Greenhouse Incubator

geographically-disbursed team of collaborators

Semantic Network Fragment from Astrobiology Application

Spring Beach

Brad Bebout

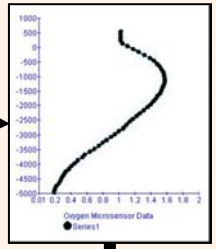
Greenhouse

collected-by

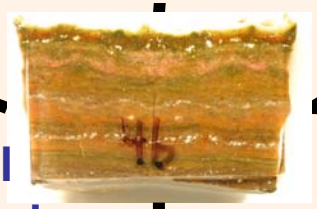
collected-at

stored-in

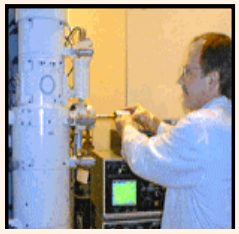
O_2 Concentration



has-measurement



Microbial Mat Sample

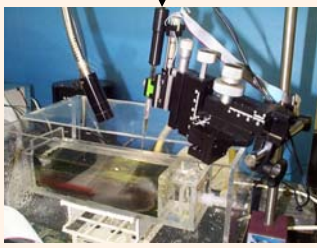


Electron Microscope

has-culture

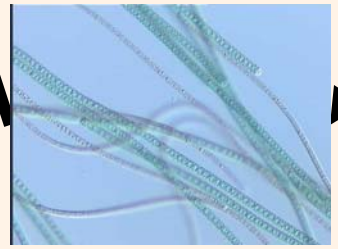
HBC-2 Microbial culture

measured-with



O_2 Microsensor

imaged-with

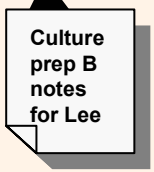


has-image

cultivated-by

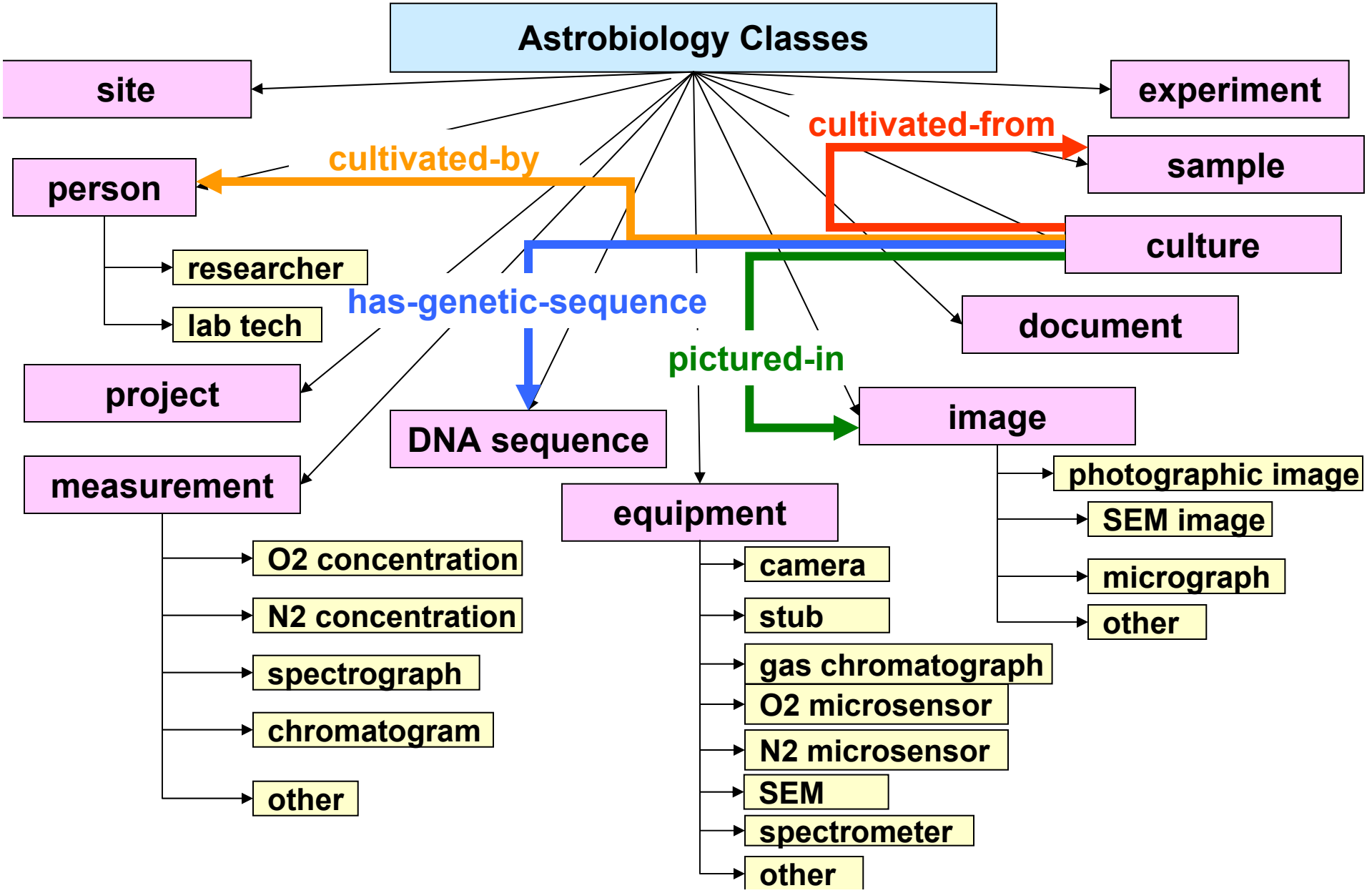
Mary Hogan

has-recipe

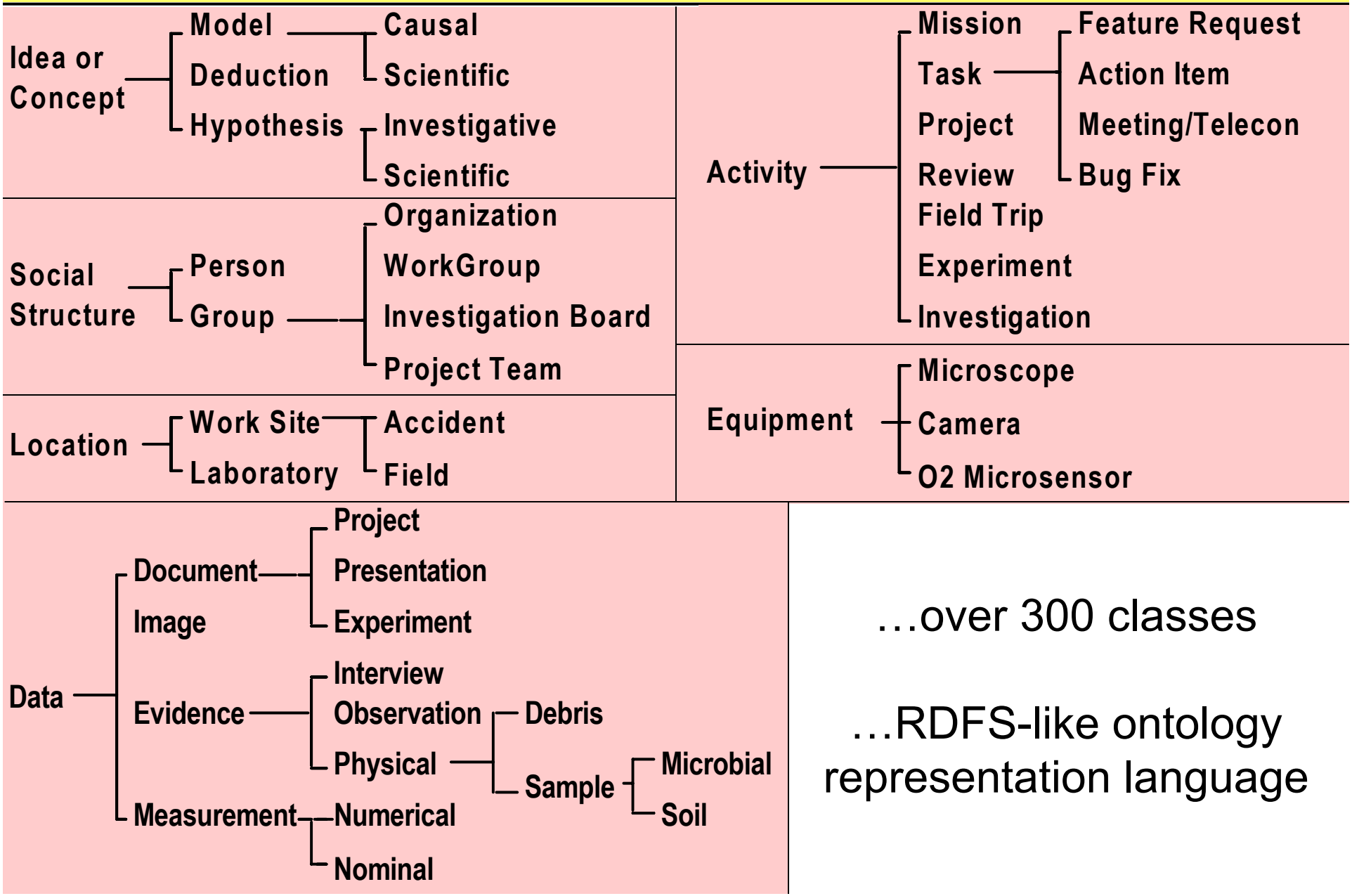


Culture recipe

Astrobiology Ontology (partial)



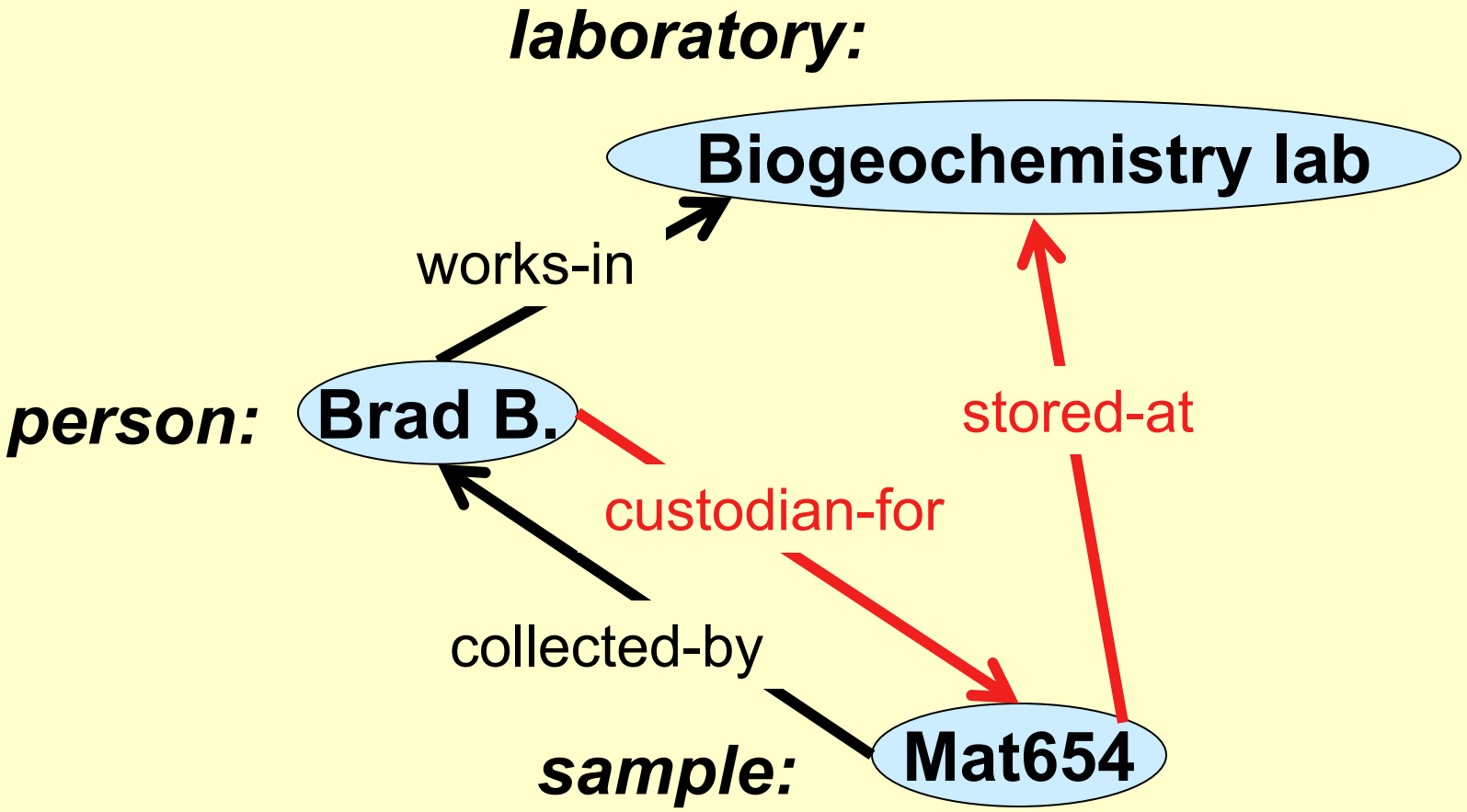
Unified Master Ontology



…over 300 classes

…RDFS-like ontology representation language

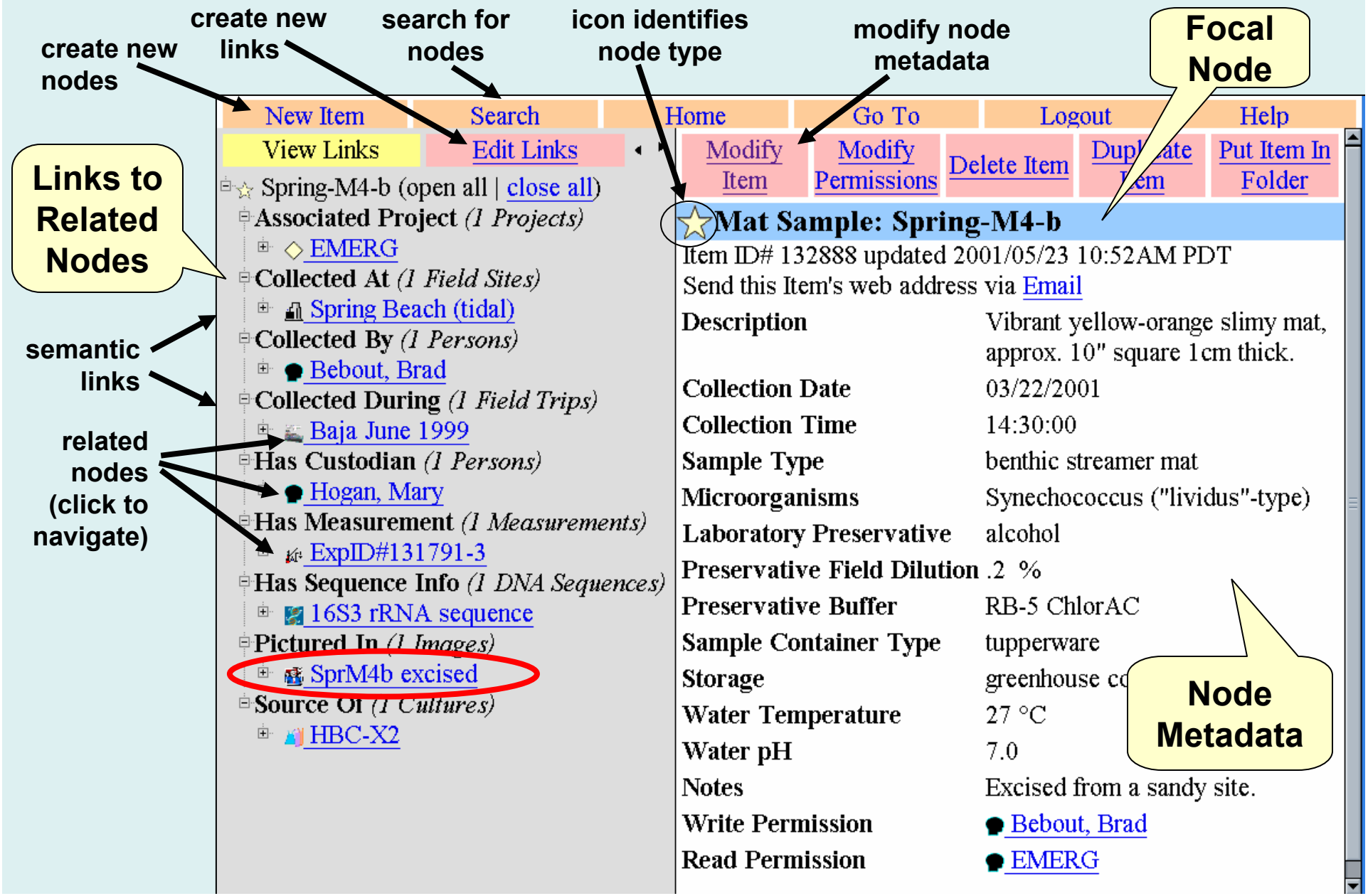
Simple Example of Rule-based Linking



If person P is *custodian-for* sample S, and P *works-in* laboratory L, Then S is *stored-at* L

If sample S is *collected-by* person P, and S has no custodian, then P is *custodian-for* S

SemanticOrganizer's Browser-based Interface (*Algae Mat Sample Node: Spring-M4-b*)



Sample is *pictured-in* Photo

New Item | Search | Home | Go To | Logout | Help

View Links | Edit Links

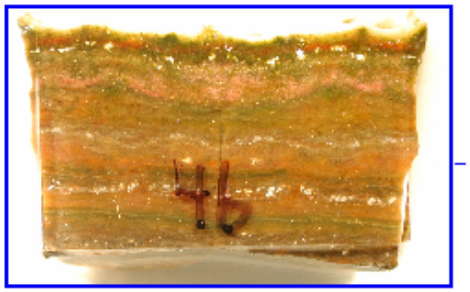
- SprM4b excised (open all | close all)
  - **Camera Used** *(1 Cameras)*
    - Brad's Digital Olympus
  - **Image Of** *(1 Samples)*
    - Spring-M4-b
  - **Imaged By** *(1 Persons)*
    - Bebout, Brad

Modify Item | Modify Permissions | Delete Item | Duplicate Item | Put Item In Folder

Photo: SprM4b excised

Item ID# 132919 updated 2003/06/14 04:28PM PDT

Send this Item's web address via Email



Click image to view at ACTUAL size.

Right Click (or click and hold on Macintosh) to download image.

Click here to download associated jpg file

Help with downloading

Annotate Image

Caption	Sample exhibits classic layering associated with mats from the Baja region.
Image Date	03/22/2001
Notes	Original photo stored in N239 room 433.
Write Permission	Bebout, Brad
Read Permission	EMERG

Edit Links

ScienceOrganizer: An Information-Sharing Tool For Scientific Project Teams — NASA Ames

New Item | Search | Home | Go To | Logout | Help

View Links | Edit Links

SprM4b excised (open all | close all)

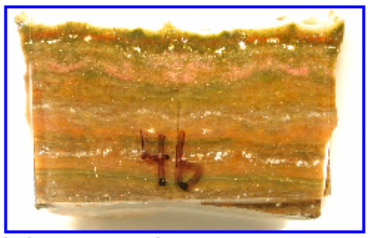
- **Camera Used** *(1 Cameras)*
  - *Link to existing / new Cameras*
  - Brad's Digital Olympus
- **Image Of** *(1 Mat Samples/0 Sediment Samples/0 Study*
  - *Link to existing / new Sediment Samples*
  - *Link to existing / new Study Areas*
  - *Link to existing / new Mat Samples*
  - *Link to existing / new Stromatolite Samples*
  - *Link to existing / new Field Sites*
  - *Link to existing / new Water Samples*
  - *Link to existing / new Gas Samples*
  - Spring-M4-b
- **Imaged By** *(1 People)*
  - *Link to existing / new People*
  - Bebout, Brad
- Other Permissible Links...
  - **Contained By** *(0 Photo Folders/0 Image (other) F*
    - *Link to existing / new Photo Folders*
    - *Link to existing / new Image (other) Folders*
    - *Link to existing / new Item (other) Folders*
  - **Figure For Pap** *(0 Publications)*
    - *Link to existing / new Publications*
  - **Illustration For** *(0 Models)*
    - *Link to existing / new Models*
  - **Is Discussed In** *(0 Email Messages)*
    - *Link to existing / new Email Messages*
  - **Map For** *(0 Study Areas/0 Field Sites)*
    - *Link to existing / new Study Areas*

Modify | Permissions | Delete | Duplicate | Put in a Folder

Photo: SprM4b excised

Item ID# 132919 updated 2003/06/14 04:28PM PDT by Demo, EMERG

Send this Item's web address via Email



Click image to view at ACTUAL size.

Right Click (or click and hold on Macintosh) to download image.

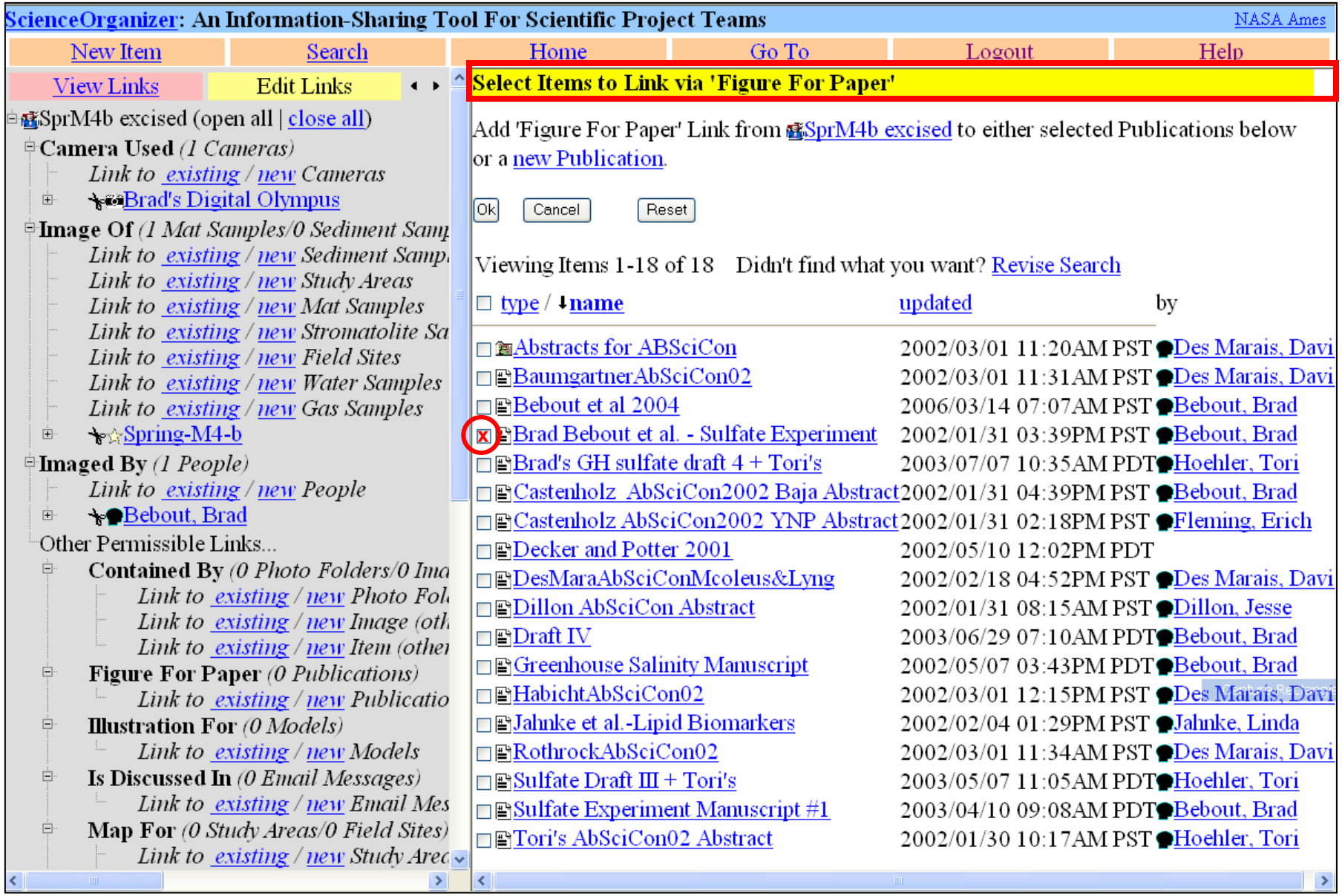
Click here to download associated jpg file (389214 bytes)

Help with downloading

Annotate Image

Camera Used	Brad's Digital Olympus
Caption	Sample exhibits classic layering associated with mats from the Baja region.
Image Date	03/22/2001
Imaged By	Bebout, Brad
Sample	Spring-M4-b
Notes	Original photo stored in N239 room 433.
Write Permission	Demo, EMERG
Read Permission	Demo, EMERG

Linking to an Existing Item



Edit Links

ScienceOrganizer: An Information-Sharing Tool For Scientific Project Teams — NASA Ames

New Item | Search | Home | Go To | Logout | Help

View Links | Edit Links

SprM4b excised (open all | close all)

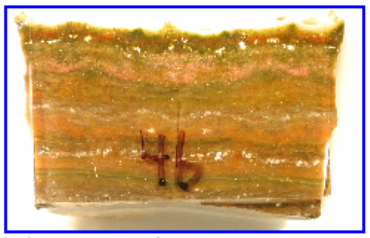
- **Camera Used** *(1 Cameras)*
  - *Link to existing / new Cameras*
  - Brad's Digital Olympus
- **Image Of** *(1 Mat Samples/0 Sediment Samples/0 Study*
  - *Link to existing / new Sediment Samples*
  - *Link to existing / new Study Areas*
  - *Link to existing / new Mat Samples*
  - *Link to existing / new Stromatolite Samples*
  - *Link to existing / new Field Sites*
  - *Link to existing / new Water Samples*
  - *Link to existing / new Gas Samples*
  - Spring-M4-b
- **Imaged By** *(1 People)*
  - *Link to existing / new People*
  - Bebout, Brad
- Other Permissible Links...
  - **Contained By** *(0 Photo Folders/0 Image (other) F*
    - *Link to existing / new Photo Folders*
    - *Link to existing / new Image (other) Folders*
    - *Link to existing / new Item (other) Folders*
  - **Figure For Pap** *(0 Publications)*
    - *Link to existing / new Publications*
  - **Illustration For** *(0 Models)*
    - *Link to existing / new Models*
  - **Is Discussed In** *(0 Email Messages)*
    - *Link to existing / new Email Messages*
  - **Map For** *(0 Study Areas/0 Field Sites)*
    - *Link to existing / new Study Areas*

Modify | Permissions | Delete | Duplicate | Put in a Folder

Photo: SprM4b excised

Item ID# 132919 updated 2003/06/14 04:28PM PDT by Demo, EMERG

Send this Item's web address via Email



Click image to view at ACTUAL size.

Right Click (or click and hold on Macintosh) to download image.

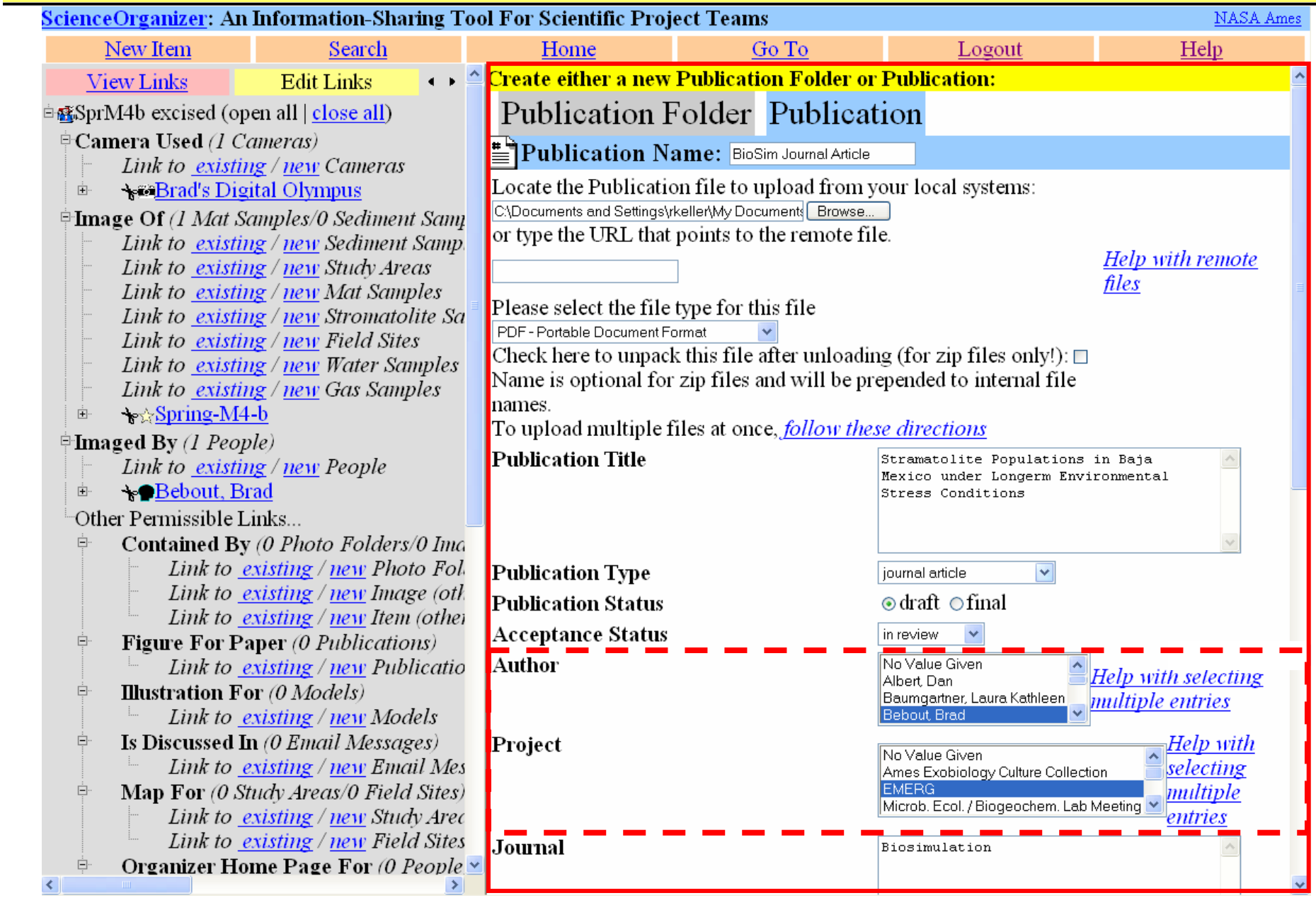
Click here to download associated jpg file (389214 bytes)

Help with downloading

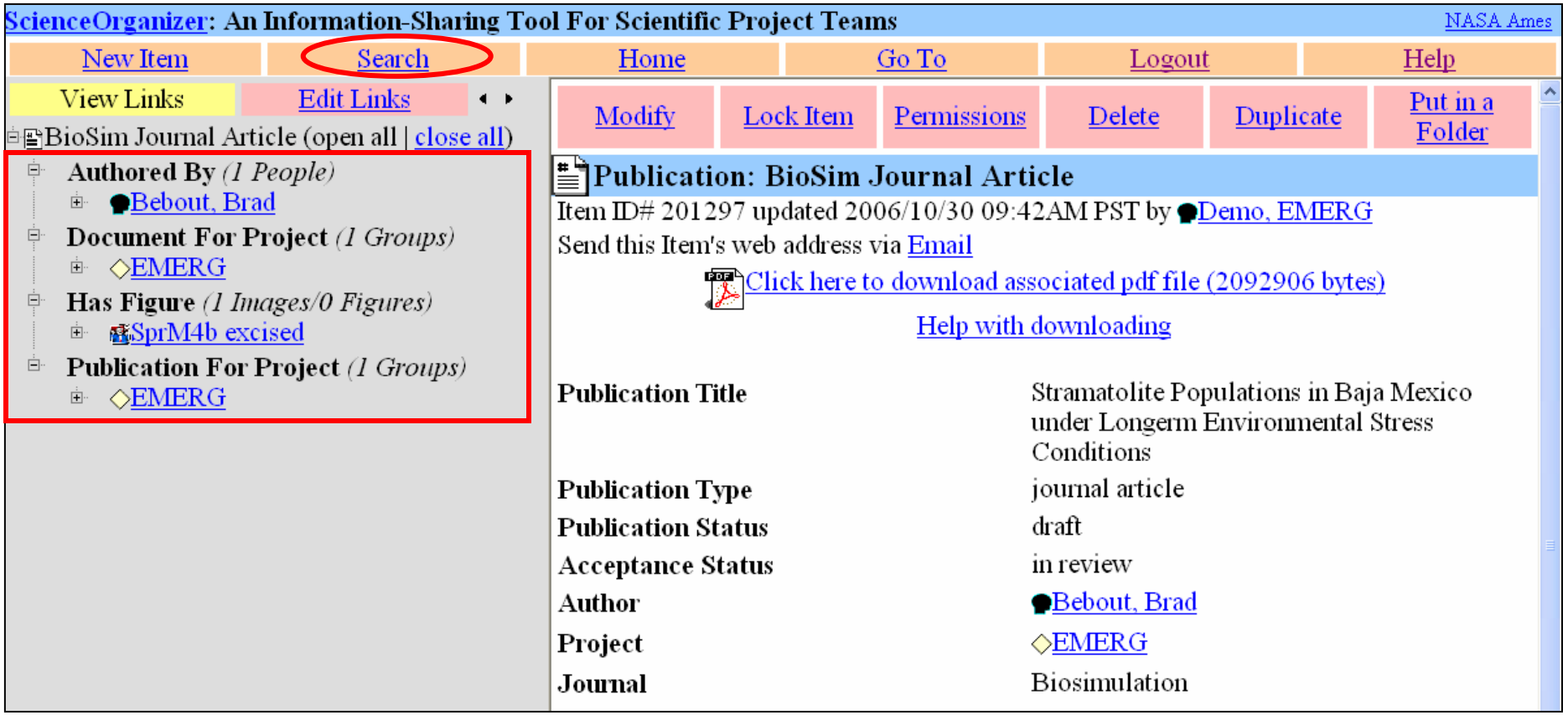
Annotate Image

Camera Used	Brad's Digital Olympus
Caption	Sample exhibits classic layering associated with mats from the Baja region.
Image Date	03/22/2001
Imaged By	Bebout, Brad
Sample	Spring-M4-b
Notes	Original photo stored in N239 room 433.
Write Permission	Demo, EMERG
Read Permission	Demo, EMERG

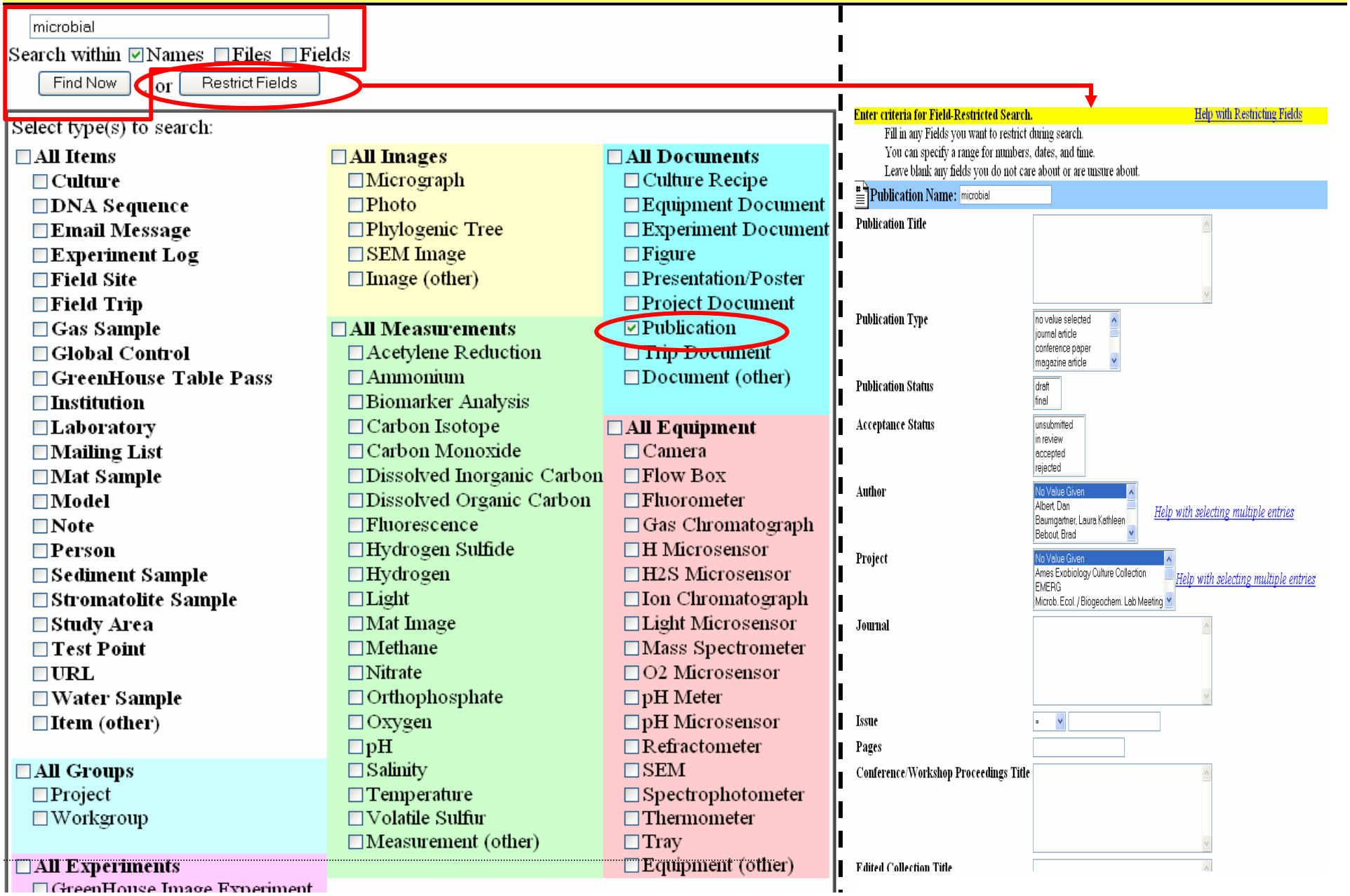
Linking to New Item

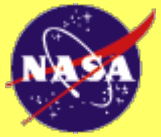


Resulting Publication with Links Established



Search


microbial
Search within Names Files Fields
Find Now
or
Restrict Fields
Select type(s) to search:
All Items
Culture
DNA Sequence
Email Message
Experiment Log
Field Site
Field Trip
Gas Sample
Global Control
GreenHouse Table Pass
Institution
Laboratory
Mailing List
Mat Sample
Model
Note
Person
Sediment Sample
Stromatolite Sample
Study Area
Test Point
URL
Water Sample
Item (other)
All Groups
Project
Workgroup
All Experiments
GreenHouse Image Experiment
All Images
Micrograph
Photo
Phylogenic Tree
SEM Image
Image (other)
All Measurements
Acetylene Reduction
Ammonium
Biomarker Analysis
Carbon Isotope
Carbon Monoxide
Dissolved Inorganic Carbon
Dissolved Organic Carbon
Fluorescence
Hydrogen Sulfide
Hydrogen
Light
Mat Image
Methane
Nitrate
Orthophosphate
Oxygen
pH
Salinity
Temperature
Volatile Sulfur
Measurement (other)
All Documents
Culture Recipe
Equipment Document
Experiment Document
Figure
Presentation/Poster
Project Document
Publication
Trip Document
Document (other)
All Equipment
Camera
Flow Box
Fluorometer
Gas Chromatograph
H Microsensor
H2S Microsensor
Ion Chromatograph
Light Microsensor
Mass Spectrometer
O2 Microsensor
pH Meter
pH Microsensor
Refractometer
SEM
Spectrophotometer
Thermometer
Tray
Equipment (other)
Enter criteria for Field-Restricted Search.
Help with Restricting Fields
Fill in any Fields you want to restrict during search.
You can specify a range for numbers, dates, and time.
Leave blank any fields you do not care about or are unsure about.
Publication Name: microbial
Publication Title
Publication Type
no value selected
journal article
conference paper
magazine article
Publication Status
draft
final
Acceptance Status
unsubmitted
in review
accepted
rejected
Author
No Value Given
Albert, Dan
Baumgartner, Laura Kathleen
Bebout, Brad
Help with selecting multiple entries
Project
No Value Given
Ames Exobiology Culture Collection
EMERG
Microb. Ecol. / Biogeochem. Lab Meeting
Help with selecting multiple entries
Journal
Issue
Pages
Conference/Workshop Proceedings Title
Edited Collection Title



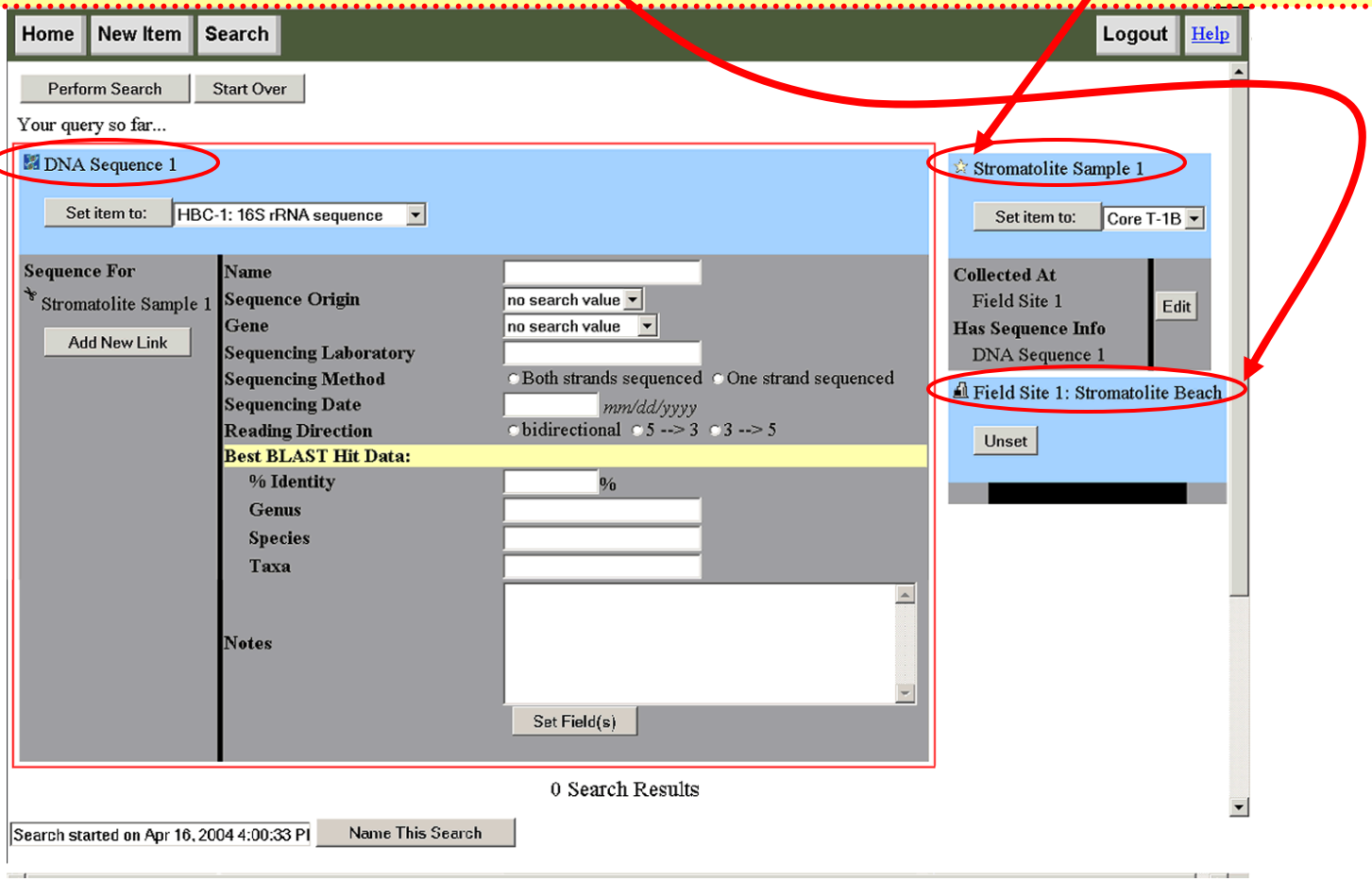
Additional SemanticOrganizer Features

- Semantic Search
- Email discussion list integration
- Automated semantic hyperthreading
- Microsoft office integration
- Collaborative image annotation
- Interoperation w/external agents

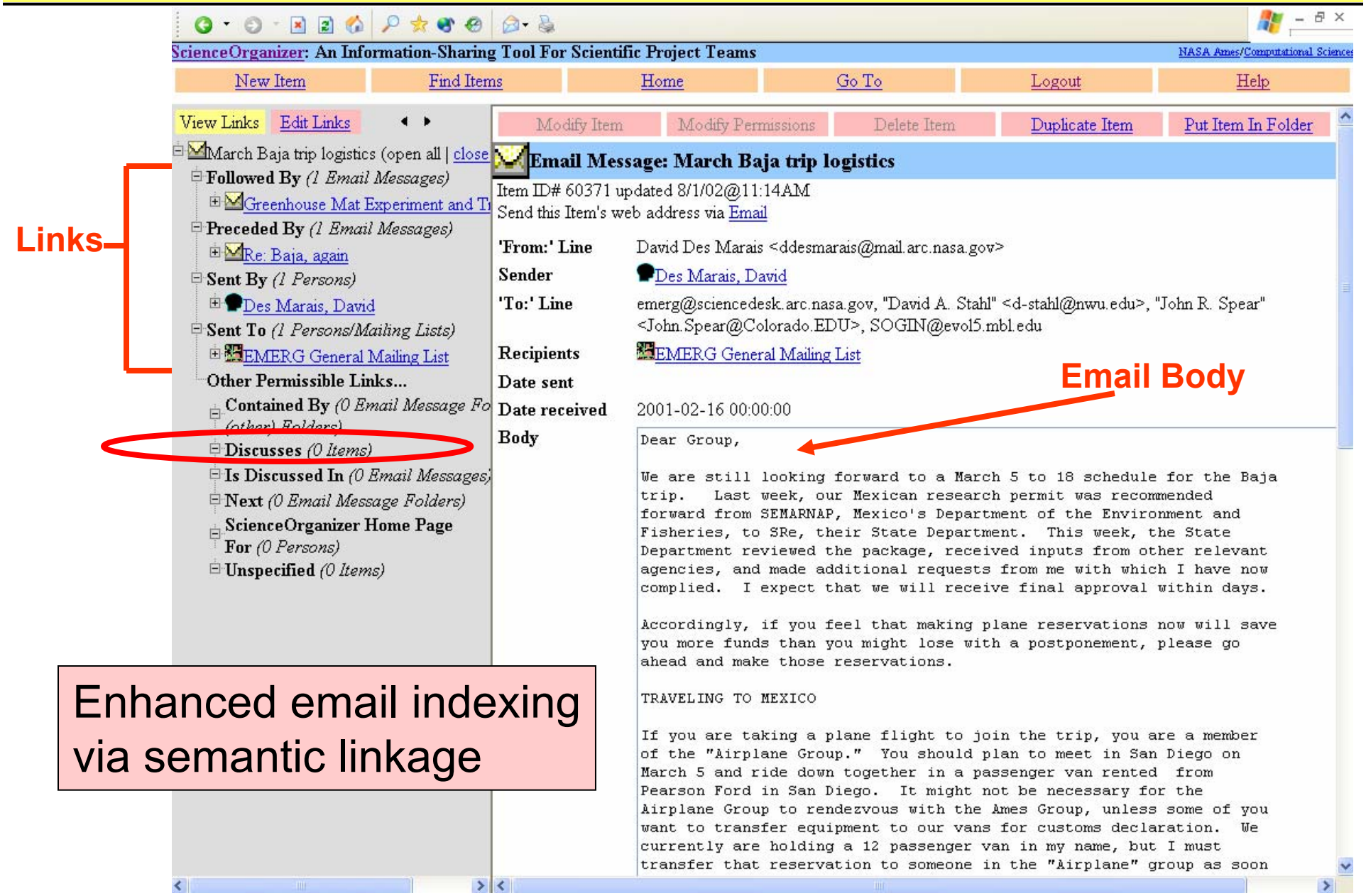
Semantic Search

- Search by semantic pattern (patterns involving extended chains of nodes, links, and values)
- Executed by matching pattern against the instance nodes and links in the semantic network

"Find all DNA Sequences associated with Stromatolite Samples collected at the Field Site named "Stromatolite Beach""



Email Discussion List Integration



Links

Email Body

Enhanced email indexing via semantic linkage

Automated Semantic Hyperthreading [Concept Extraction from Email] (experimental)

Semantic Organizer

Email body (extract)	Relevancy rank	Node	Type
"FIELD OBJECTIVES This field trip will address a number of objectives that include those of the Ecogenomics project as well as those associated with previously-funded grants. Accordingly, we will be performing biogeochemical measurements and collect samples for various analyses back at our respective laboratories…	1.063660	Spear, John	(participan)
	1.063660	Stahl, David	(participan)
	1.063660	Bebout, Brad	(participan)
	1.036137	Ames	area)
	1.036137	Baja	a)
	1.031293	Ecogenomics Focus	(proje
	1.020290	Pond 5 near 6	(aque i)
	1.020290	Pond 4 near 5	(aque i)
Straw sampling Protocols In situ measurements of fluxes: … number of diel periods : 2 (once in Pond 4 near 5 , once in Pond 5 near 6) … number of chambers: 6 (3 mats, three blanks) … sampling times: 0600h, 0900h, 1200h, 1500h, 1800h, 2100h, 0000h, 0300h … Water volume taken per sampling ti … Gas volume taken per sampling inter	1.009677	Visscher, Pieter	(participan)
	1.009677	Hoehler, Tori	(participan)
	1.009677	Hogan, Mary	(participan)
	1.009677	Garcia-Pichel, Fer	(participan)
	1.009677	Dillon, Jesse	(participan)
	1.009677	Turk, Kendra	(participan)
	1.009677	Miller, Scott	(participan)
…SAMPLING We recognize that the acquisition ar preserv later use will be a major objective of s trip. discussion of sampling preferences du Ames.			

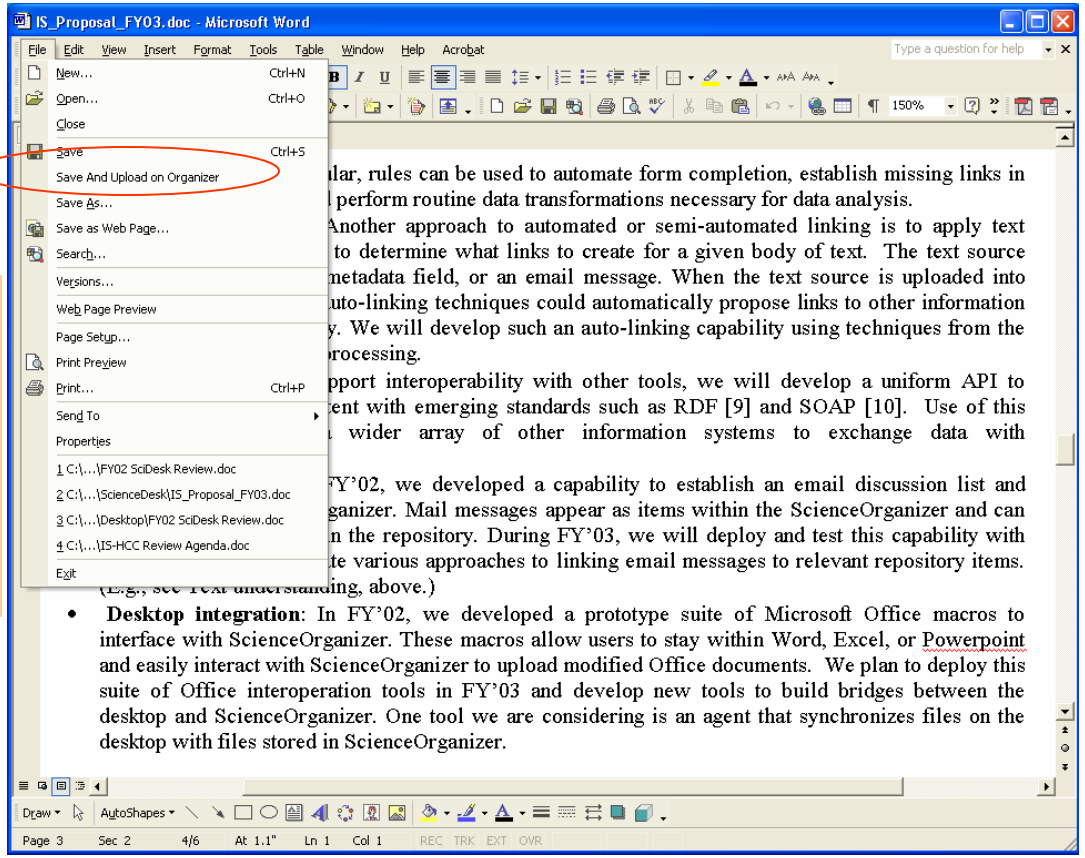
2. Relevancy ranking score computed:
- based on heuristic combination of factors

1. Text marked up to identify:
- nouns matching node names/types
- verbs corresponding to links
- references to attribute values
- WordNet synonyms/hyponyms used to enable inexact semantic matching

Microsoft Office Interoperation

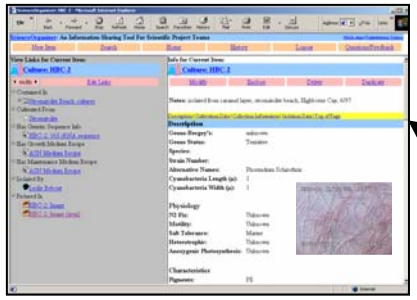
- **Developed suite of Microsoft Office macros to enable direct upload and subsequent modification of documents without leaving Office application**

"Save and upload to Organizer"



- Macro communicates w/server
- User fills out metadata using standard Organizer form on creation
- Subsequent saves are transparent

Shared Image Annotator



SemanticOrganizer

Collaborative Whiteboarding & Annotation Tool

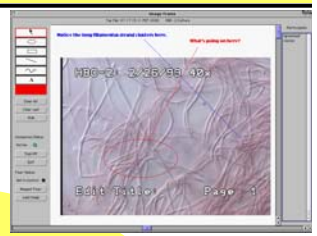
Interoperation w/ External Agents & Systems via API

Agent-assisted Scientific Experimentation System

real-time measurements, images

Agent-based Mars exploration testbed

images, measurements



image

SemanticOrganizer

Shared Image Annotator

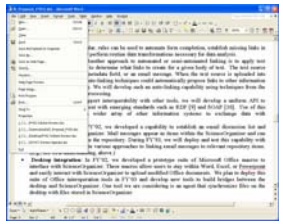
annotated image

Microsensor Remote Controller

sensor

documents

Microsoft Office interface

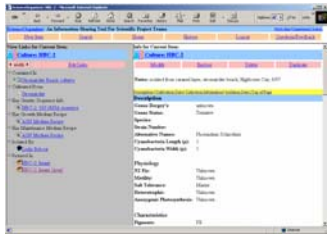


Architecture

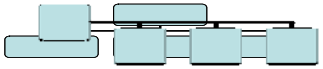
Semantic Organizer

Interface Layer

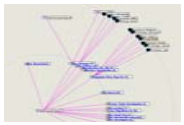
Node Browser Interface



Structure Editors/ Viewers

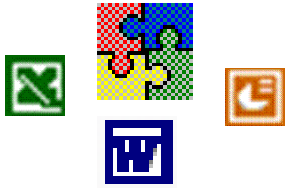


Radial Network Browser



Programmatic/Agent Interface

Microsoft Office Macro Integration



Representation & Reasoning Layer

Email Ingestor
(email & attachments)

Semantic Repository
(items, links, attribute values)

Ontology
(item types, link types, attributes, rules)

RDFS-equivalent

Semantic Annotation

WordNet thesaurus

Inference Engine

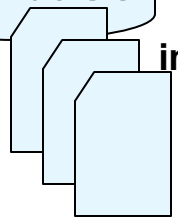
Implementation Layer

Java servlets

Apache-Tomcat

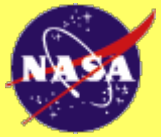
MySQL Database

Jess



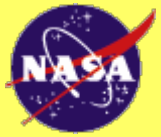
images
documents
datafiles

file system



Outline

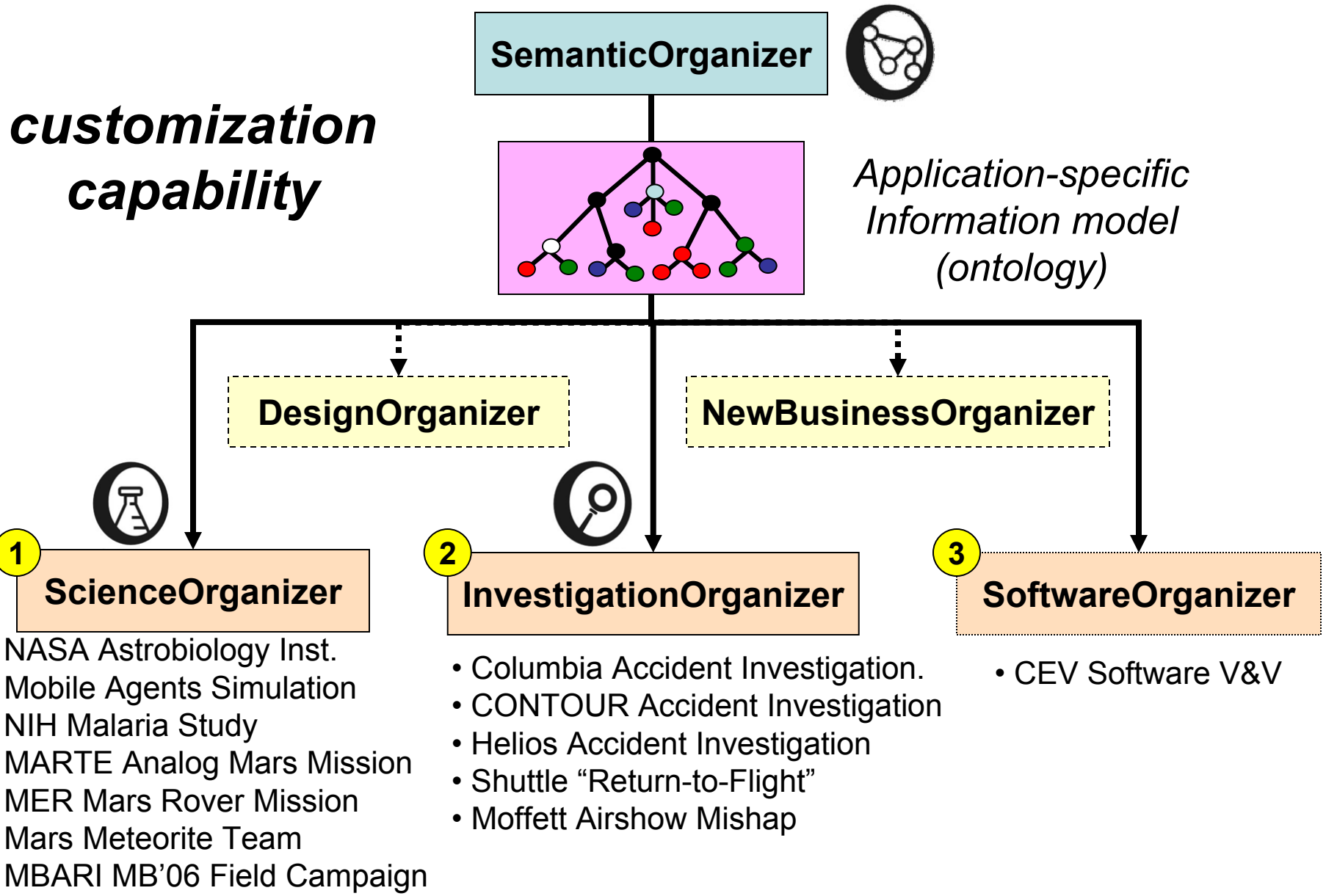
- Introduction & Motivation
- SemanticOrganizer system
- Applications
- Lessons learned



Operational Status

- System deployed initially in 2001
- Over 400 registered individual users from over 50 organizations within NASA at peak
- Over 30 projects hosted
- Over 45,000 nodes & 160,000 links in repository
  - Over 10,000 electronic files stored (documents, image, datasets)
  - Over 10,000 archived email messages

SemanticOrganizer Applications

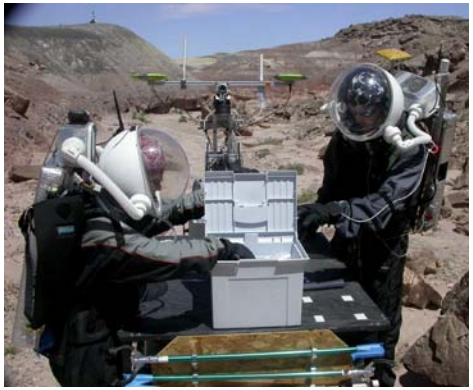


1. ScienceOrganizer Deployments

Astrobiology Field
and Lab Science

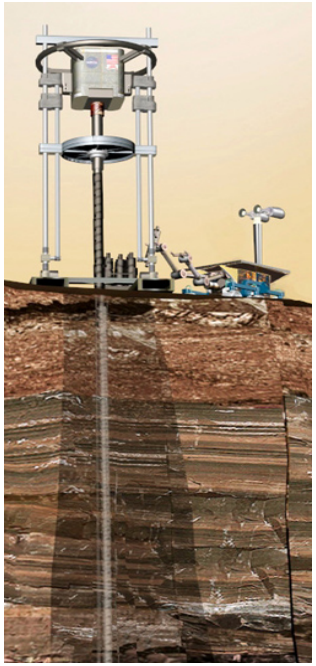
NIH-NASA
African Malaria Study

Electron Microscopy
Image Archive
(Martian Meteorites)



Simulated Mars Surface
Exploration

Monterey Bay
Oceanographic
Campaign



MARTE Mars Analog
Drilling Mission

Mars Exploration Rovers
Hypothesis Management
(prototype)

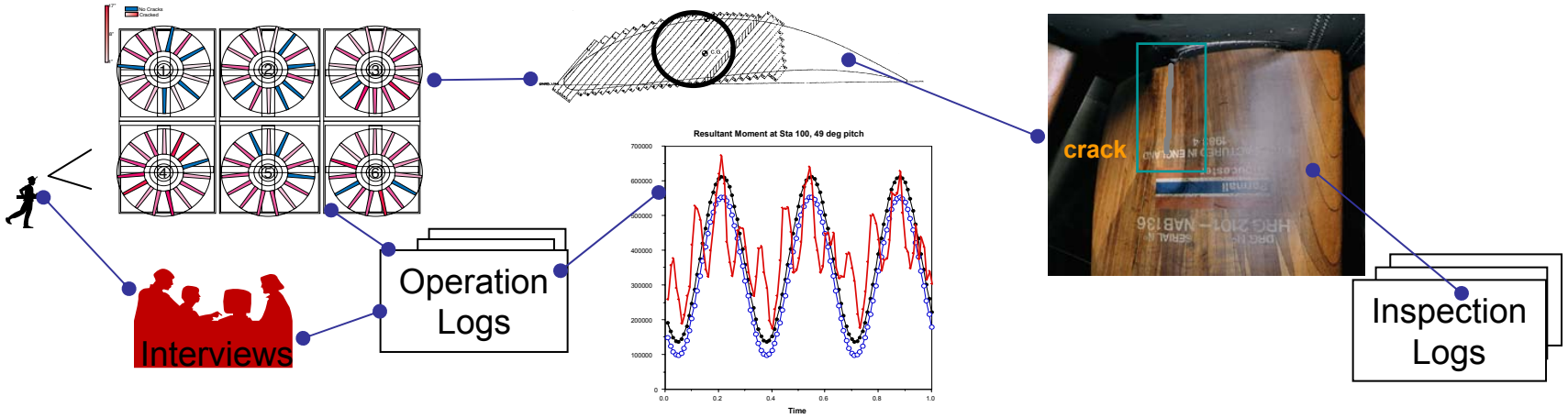
Activities Supported by ScienceOrganizer

- Collecting scientific field data
- Initiating/monitoring laboratory experiments
- Acquiring data from remote instruments
- Archiving samples and images
- Tracking scientific hypotheses
- Writing scientific papers & proposals
- Conducting education and outreach

2. InvestigationOrganizer:
Support for Accident Investigation Teams

Large amount of data/evidence must be organized/shared rapidly:

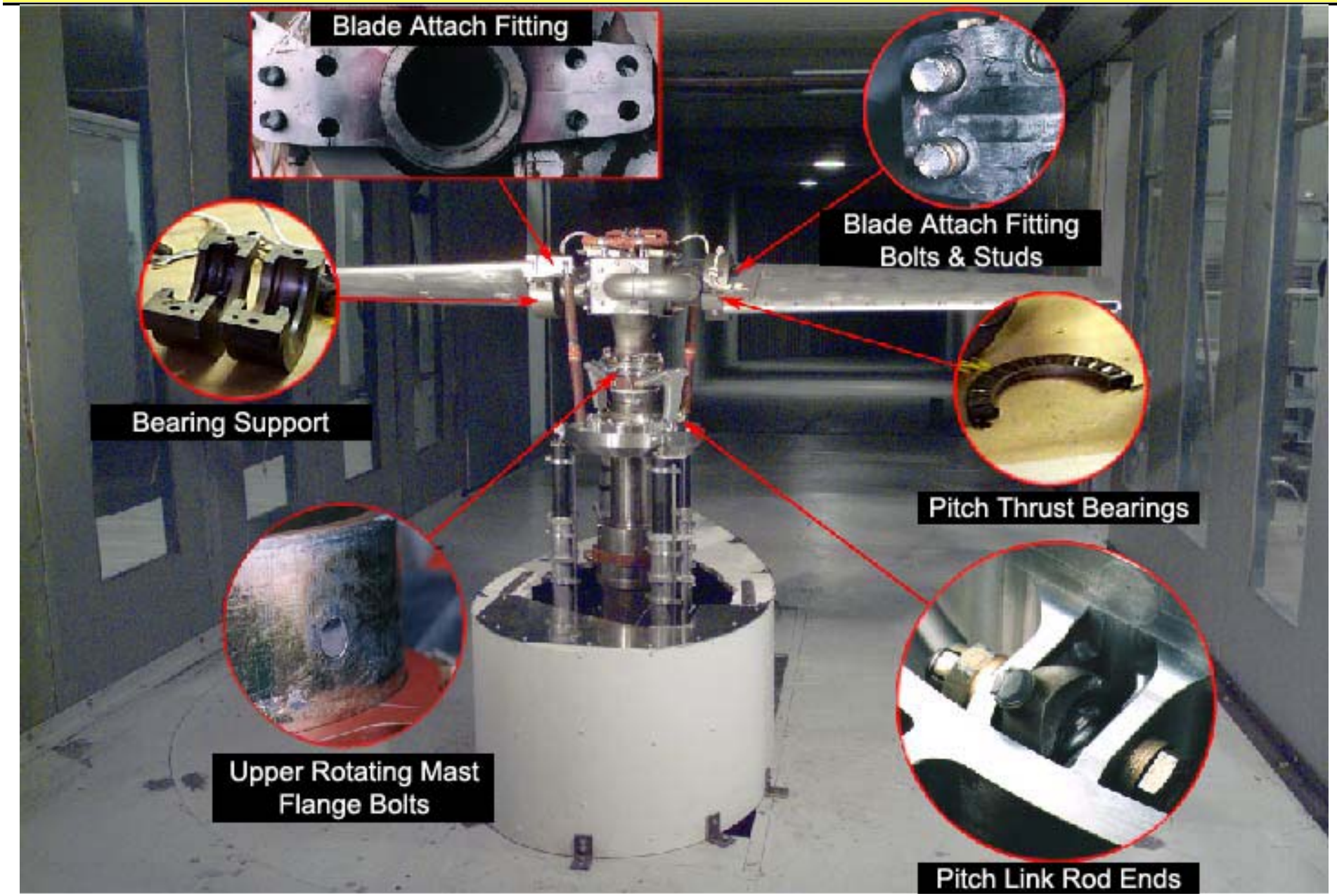
- **design specifications**
- **materials analyses**
- **procedure manuals**
- **operations logs**
- **inspection logs**
- **defect images**
- **witness transcripts**
- **investigation notes**
- **site maps, images**
- **lab reports**
- **engineering drawings**
- **risk analyses**
- **test data**
- **training records**



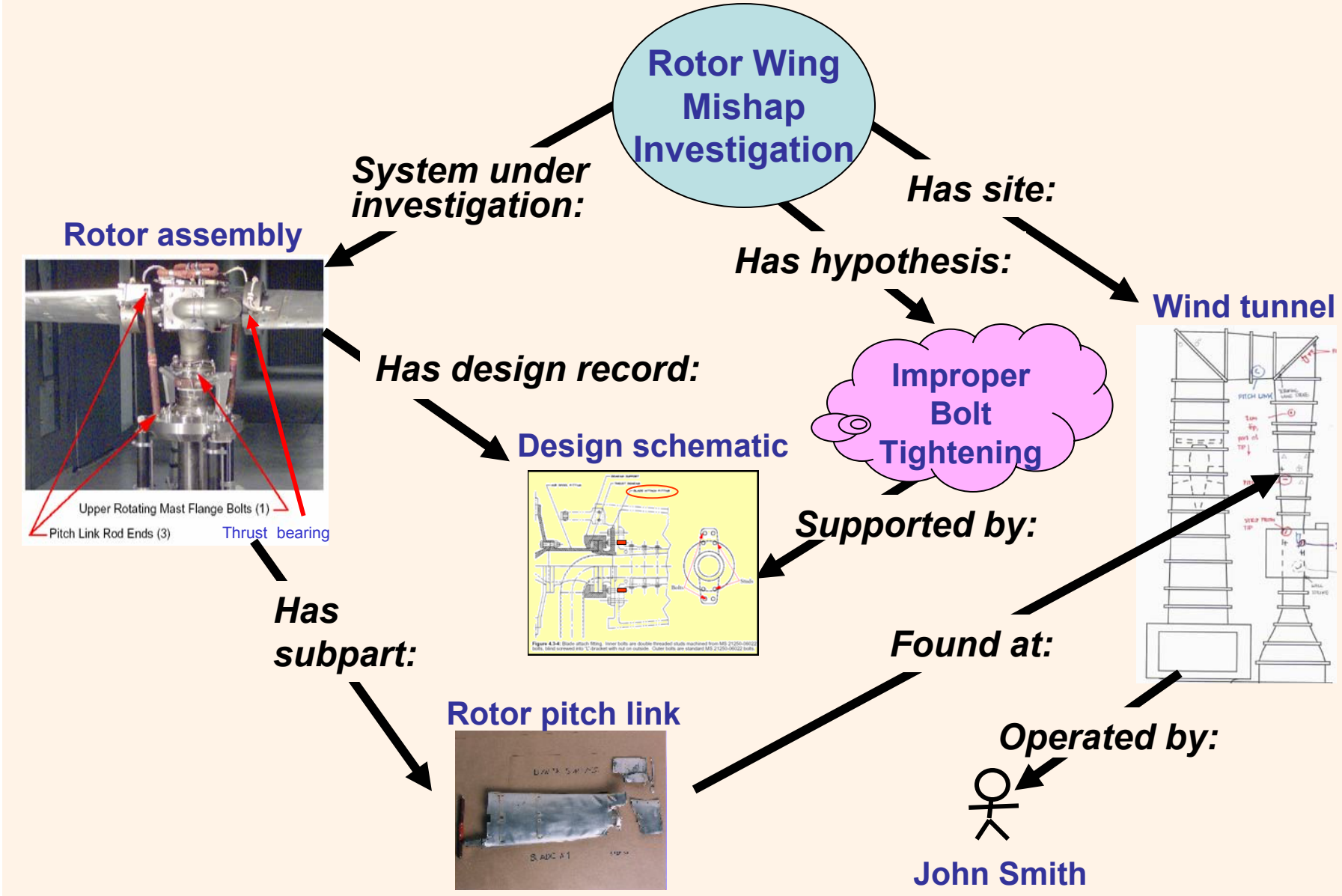
- Much of the information is interrelated

- Interrelationships & information context are important!

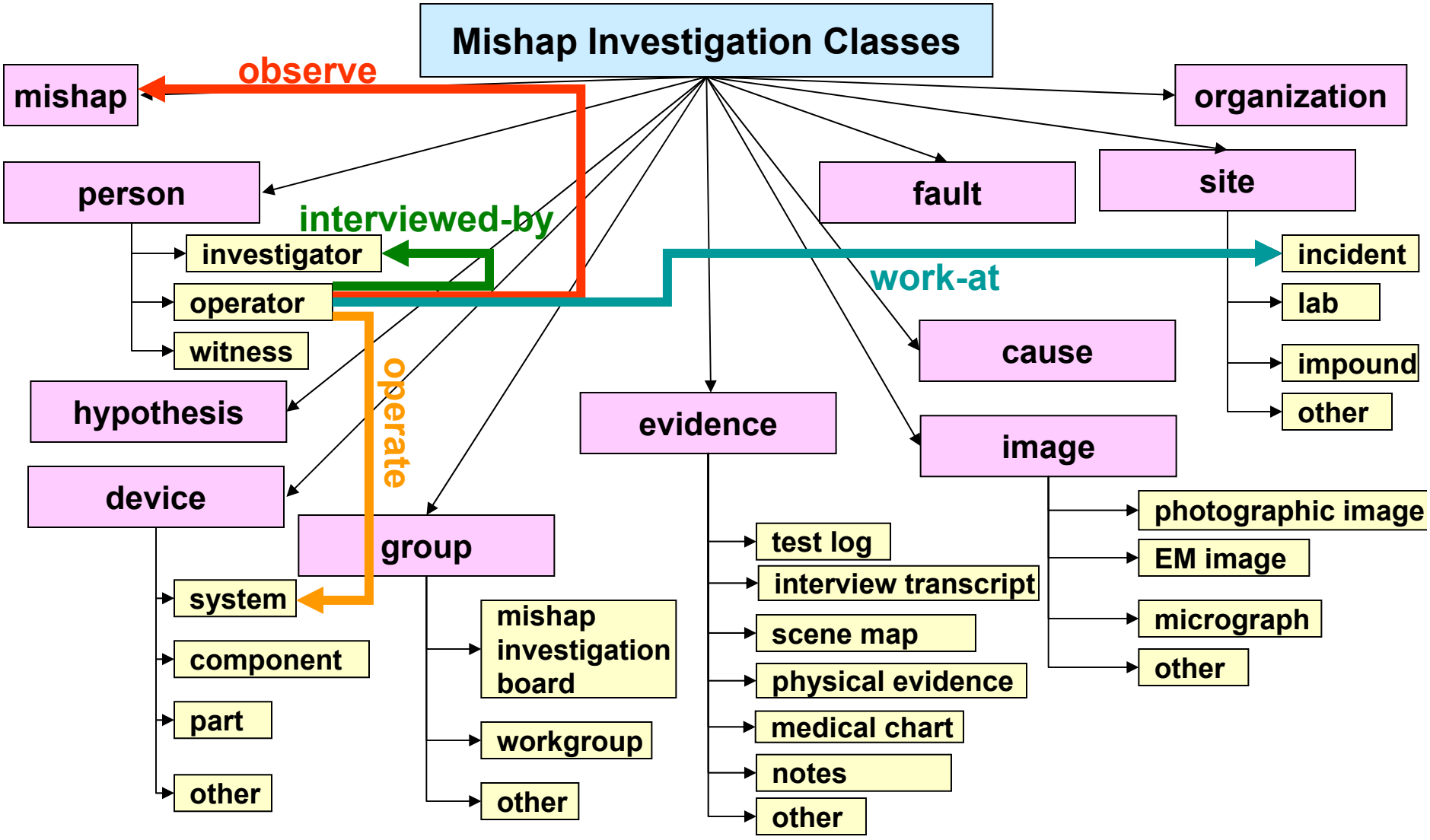
InvestigationOrganizer Example: Helicopter Rotor Wing Mishap Investigation



Semantic Network Fragment from Helicopter Rotor Mishap Investigation



Mishap Investigation Ontology *(partial)*



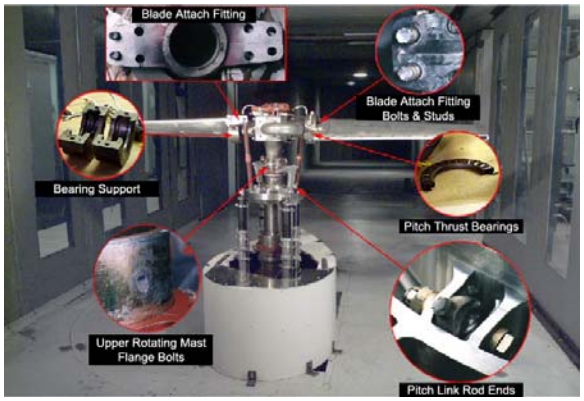
InvestigationOrganizer Deployments

CONTOUR Comet-Follower

Shuttle Columbia

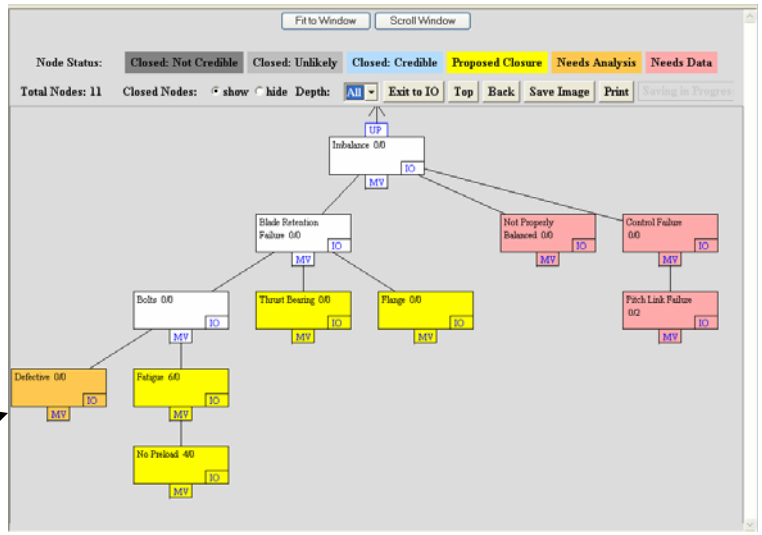
Helios Solar-powered UAV

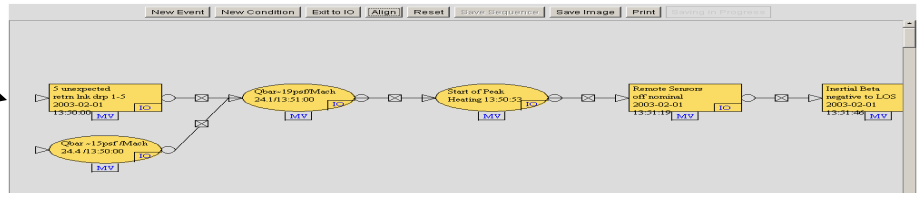
Helicopter Rotor Wing

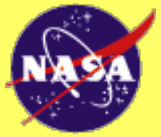


Activities Supported by InvestigationOrganizer

- Collecting evidence
- Tracking hypotheses
- Modeling accidents
  - fault tree development
  - event timelines
- Analyzing data
- Supporting investigative action-tracking/workflow
- Writing reports



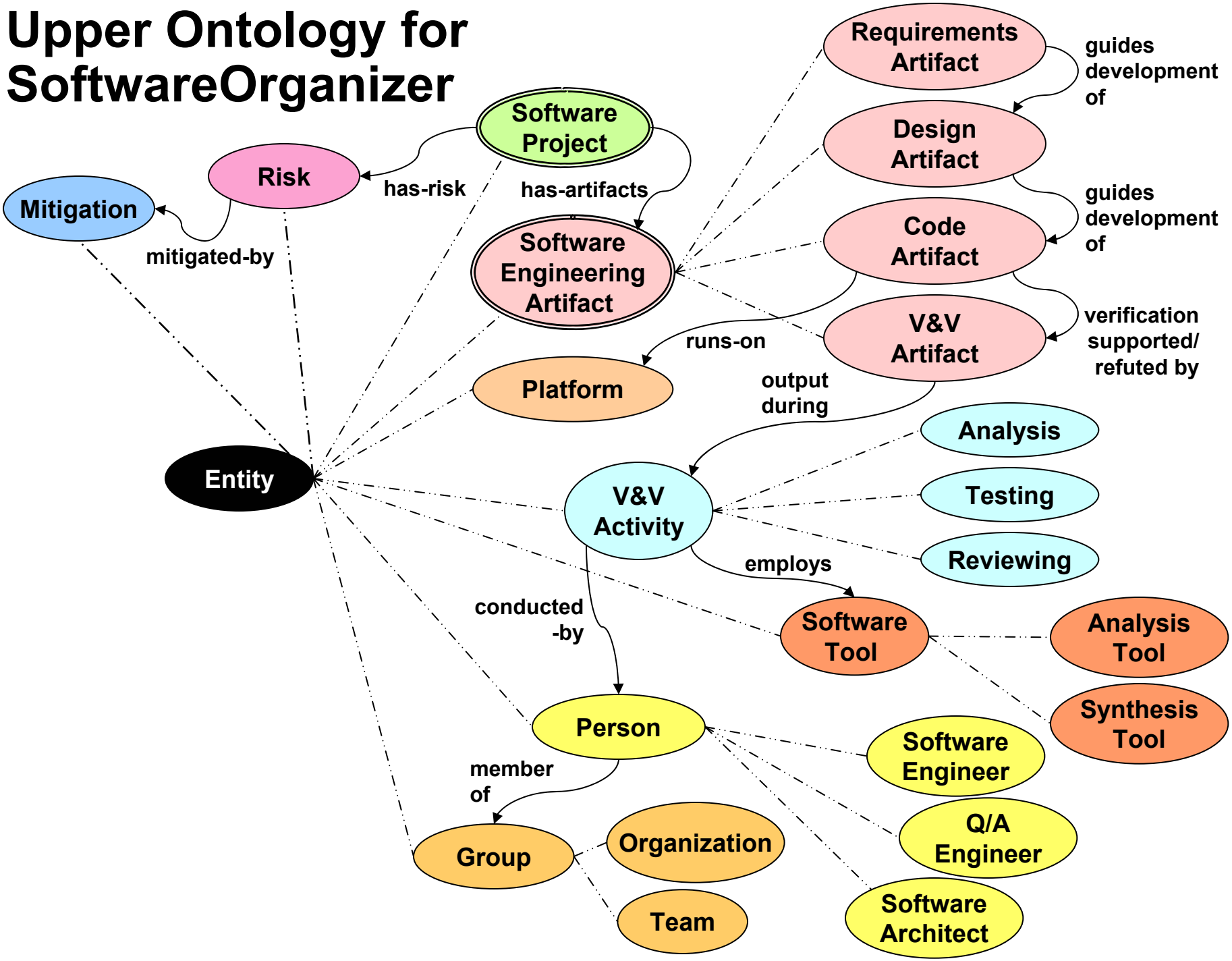


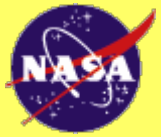


3. SoftwareOrganizer Application (under development)

- **Target Users**: NASA software acceptance, V&V, QA personnel
- **Goals**:
  - Store and interrelate artifacts from throughout the software lifecycle: requirements, design, coding, V&V
  - Support end-to-end requirements traceability
  - Provide an historical record of activities conducted and artifacts produced as part of a NASA software delivery
- **Motivation**:
  - Due to large number of contractors, no single place to examine all software engineering artifacts for a project
  - Existing vendor tool suites capture only a subset of the desired artifacts and are not customizable for NASA needs
  - Closed, vendor-specific solutions not suitable for diverse NASA contracting environment

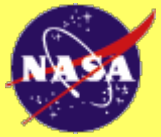
Upper Ontology for SoftwareOrganizer





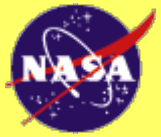
Outline

- Introduction & Motivation
- SemanticOrganizer system
- Applications
- Lessons learned



Lessons Learned

- **Lesson #1:** Network-structured Storage Models Confound Users
- **Lesson #2:** Need to “Loosen” Semantics to Accommodate Different Organizing Styles
- **Lesson #3:** Navigating the Information Space can be Difficult
- **Lesson #4:** Automated Knowledge Extraction is Key



Lesson #1: Network-structured Storage Models Confound Users

- Despite ubiquity of the Web, users' comfort level with networked storage schemes is low
- Networks can be disorienting
- Hierarchical storage schemes rule! (e.g., folders, directory structures)

- Mitigation strategy: Introduction of typed folders and subfolders into network
  - Largely unsuccessful: semantics and type restrictions are unintuitive and counterproductive

Lesson #2: Need to "Loosen" Semantics to Accommodate Different Organizing Styles

- **Personal Organizational style depends on:**
  - **Individual differences and preferences**
  - **Task demands**
- **SemanticOrganizer frustrated:**
  - **Casual (and chaotic) organizers**
  - **Users whose task demands or time available didn't justify the overhead required to provide semantic linkages: cost-benefit proposition wasn't clear**
- **Compromises that serve all users are difficult to achieve**

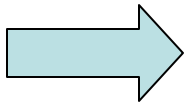
Lesson #3: Navigating the Information Space can be Difficult

- Information space is very large
  - Over 45,000 nodes
  - Over 150,000 links
- Can't display entire information space
- Difficult to comprehend the scope and extent of the information space
- Difficult to overview and navigate the space

→ *Semantic web visualization research*

Lesson #4: Automated Knowledge Extraction is Key

- **Relying on purely manual entry and linkage of information is not tenable**
- **Automated methods of creating and linking information are essential:**
  - **Inference rules**
  - **Object and link extraction from text**
  - **API for communication with agents/processes**
  - **Integration with document management software (MS Office)**
  - **Interoperation with stand-alone sources (databases, Web sites)**



Semantic annotation research

Key Idea:
Model-based Information Integration

- Use a common information model to “glue” together disparate data and information into a coherent, integrated whole
- Use model semantics to check validity of information and improve consistency and completeness of information
- Use semantic context to facilitate information access, retrieval, and search

Questions?

Web Site http://sciencedesk.arc.nasa.gov

Paper Keller et al.: “SemanticOrganizer: A Customizable Semantic Repository for Distributed NASA Project Teams”, International Semantic Web Conference, Hiroshima, Japan, November 2004.

Honors

ISWC Winner, Best paper award, 2004 International Semantic Web Conference

Finalist in the 2003 NASA Software of the Year Competition and recipient of a Space Act Award for significant technical contribution to NASA.

Winner of a 2004 NASA Turning Goals Into Reality Award for outstanding accomplishment in supporting the Columbia Accident Investigation Board

Finalist in the 2004 Semantic Web Challenge Competition