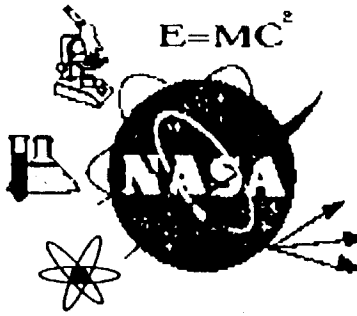
TP/IN/82

577823



ScienceDesk

Information Sharing for Scientific Project Teams

ScienceDesk Team

Rich Keller (PI)
Shawn Wolfe
Jim Chen
Kim Hubbard
Deepak Kulkarni
Steve Rich
Keith Swanson
Robert Carvalho

Computational Sciences Division

Early Microbial Ecosystems Research Group

Dave Des Marais
Brad Bebout
Lee Prufert-Bebout
Tori Hoehler
Scott Miller
Linda Jankhe
Mary Hogan

NASA Astrobiology Institute

Electron Microscope Lab

Dave Blake
Kathy Kato
Sherry Cady
Lori Crumbliss

Exobiology Branch

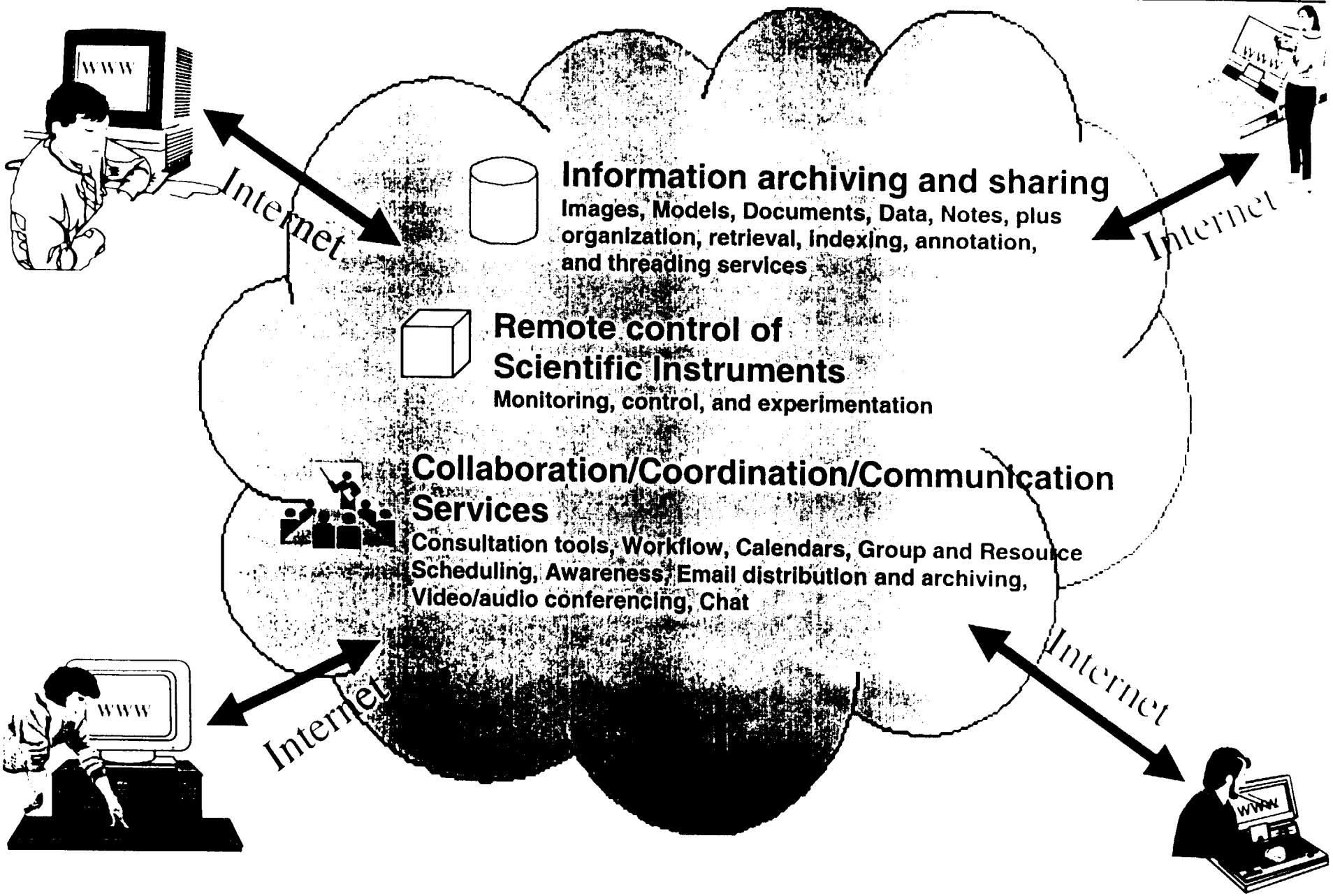
NASA Ames Research Center

Co-funded by the NASA Cross-Enterprise Technology Development Program, the NASA Astrobiology Institute, and the NASA Intelligent Systems Program

Ames Research Center

ScienceDesk Focus: Infrastructure support for distributed scientific teams

Ames Research Center



Research Areas

Ames Research Center

- **Scientific Knowledge Management:** capture, preservation, traceability of scientific knowledge
- **Intelligent Information Access:** intelligent indexing, visualization, and navigation
- **Collaboratories:** asynchronous and synchronous collaborative scientific teamwork
- **Agent-assisted Remote Experimentation:** intelligent monitoring and control

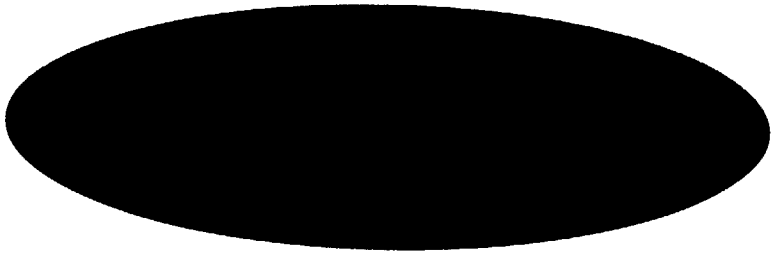
http://sciencedesk.arc.nasa.gov

Human-Centered Methodology

Ames Research Center

Study science team in operation

- interview
- observe (lab & field)
- videotape



Analyze and model

- workflow
- procedures
- communication flow

Develop new IT concepts & tools

- ScienceOrganizer
- Shared Image Annotator
- O_2 Microsensor Controller

Develop prototype

Evaluate prototype

Critique & Review

Critique & Review

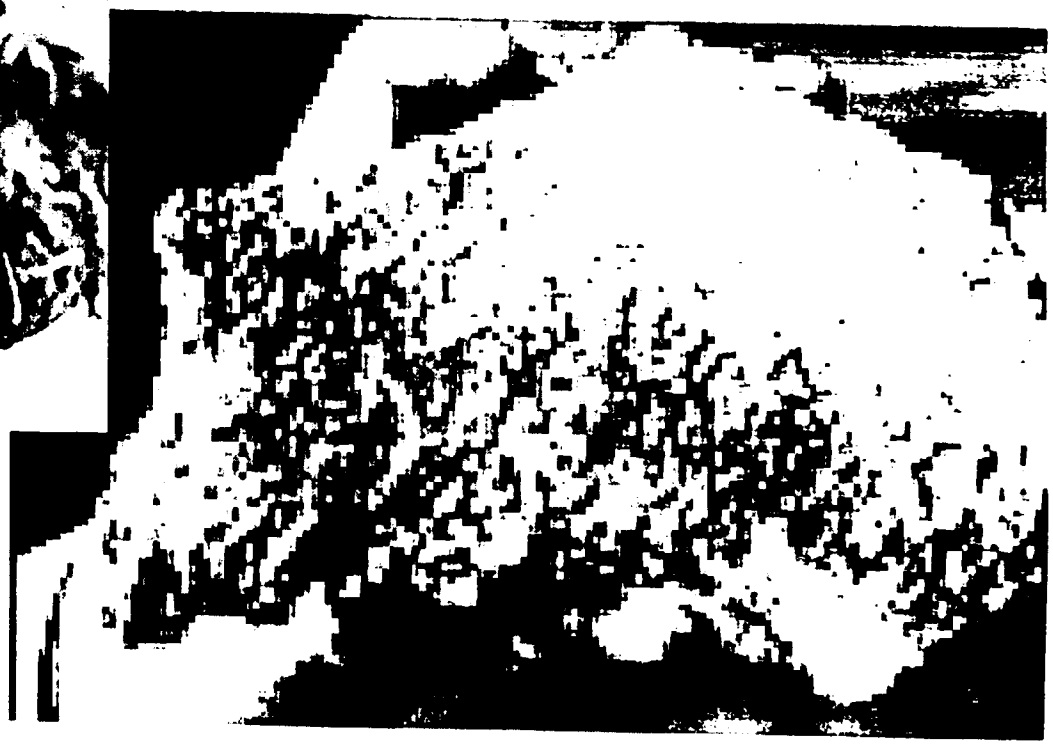
Critique & Review

ScienceDesk Testbed:
Early Microbial Ecosystems Research

Ames Research Center

Study of "Microbial Mats"

- Photosynthetic microbial mats are the oldest known ecosystem on Earth
- Can be found in extreme environments (thermal, saline, high carbonate)
- Serve as an early-Earth analog for studying development of life



Early Microbial Ecosystems Group: Distributed Field and Lab Work

Ames Research Center

Field Research Sites

Home Institutions

Portland State

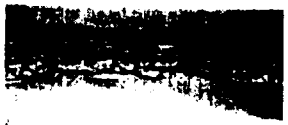
U. of Oregon

U. Conn

U. Miami

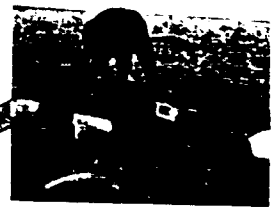
Arizona State

Field Data Collection & Preliminary Data Analysis



Highborne Cay, Bahamas

- field notes
- images
- measurements
- sensor data



Baja California, Mexico

Field Samples

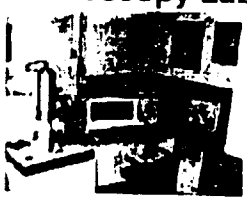
Geological Samples

Yellowstone National Park

Biological Samples

- lab notes and experiment data
- electron microscope images
- analysis results
- publications

Microscopy Lab



Greenhouse

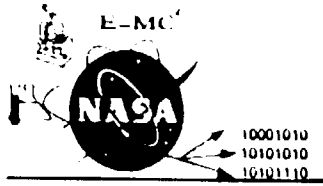
Microbiology Lab

Microbial Culture Facility

Other culture collections

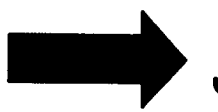
Ames Research Center

Microbial Cultures



Scientists' requirements

- Need to share information and data among distributed team
- Wide variety of data sources (sensors, instruments, data analysis software)
- Heterogeneous information formats (image, data)
- "Pedigree" of data and information products must be tracked in detail
- Information products must be archived to support historical preservation & replication of results



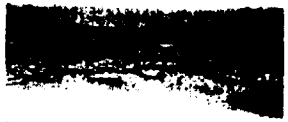
ScienceOrganizer

ScienceDesk Group Thrust Areas

Ames Research Center

Portland State
U. of Oregon
U. Conn
U. Miami
Arizona State

Field Research Sites
Home Institutions



Highborne Cay, Bahamas

1. Project Info. Repository:
ScienceOrganizer
- Images
- Cultures
- Datasets
- Procedures
- Documents
- Scientific Records

Baja California, Mexico

2. Remote Electron Microscopy Facility

Microscopy Lab

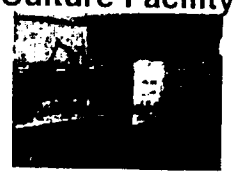
Yellowstone National Park

Greenhouse

3. Agent-assisted Greenhouse Collaboratory

Microbiology Lab

Microbial Culture Facility



Ames Research Center

What is ScienceOrganizer?

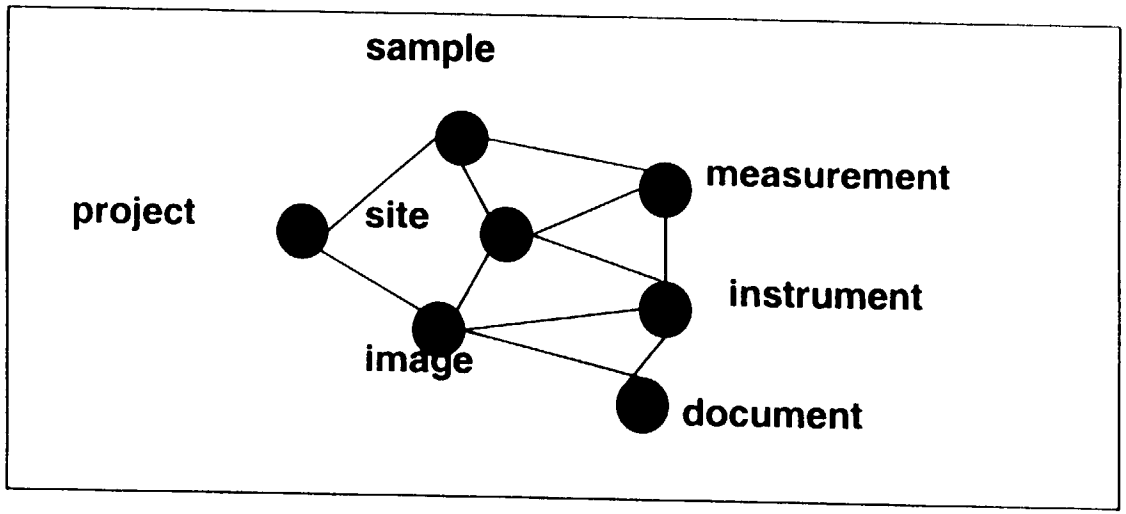
Ames Research Center

- An information repository / digital library for distributed scientific project teams
- A hybrid tool combining the functionality of:
  - a database
  - a document-sharing system
  - a web-like hypermedia information space
  - a semantic net
- Enables storage and retrieval of heterogeneous project information: *images, datasets, documents, and various types of scientific records* *(describing samples, field sites, measurements, instruments, personnel, etc.)*
- Supports cross-linkage among stored items to enable rapid access to interrelated information

The ScienceOrganizer "Project Information Web"

Ames Research Center

ScienceOrganizer maintains project information in an interconnected network or "information web"



- Nodes: *information resources*
- Links: *relationships among resources*

Information Resources ("nodes")

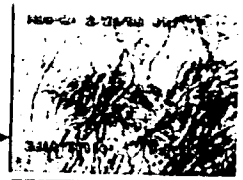
- Describe various types of project-specific information
  (e.g., field sites, samples, images, documents, personnel, instruments, measurements, etc.)
- Contain metadata (categorical, text, or numeric)
- Have (optional) attached file (e.g., image, document, dataset)

Information resource instances:

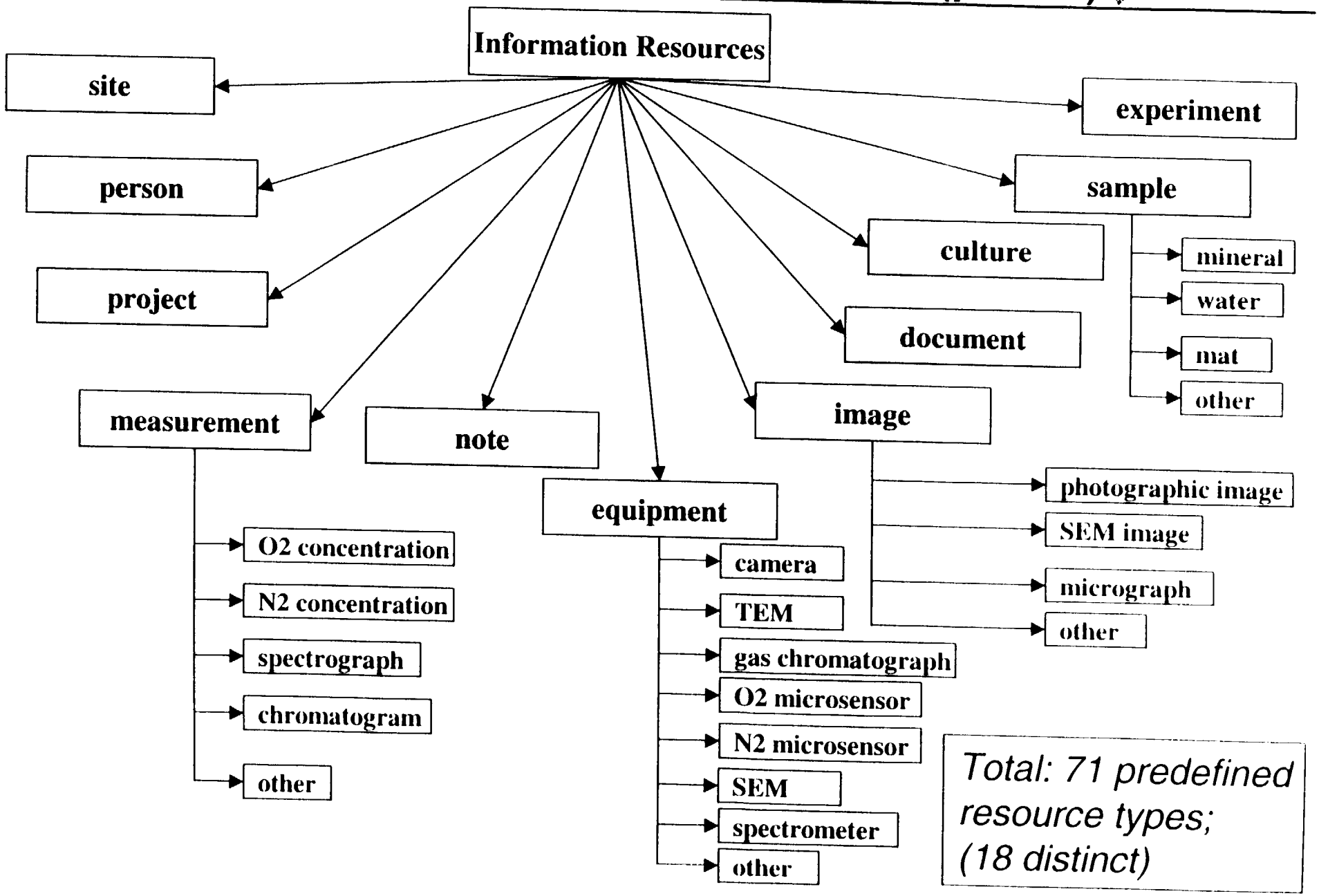
Microbial-Culture-123
Cultivated-by: R. Smith
Genus: phormidium
Growth medium: ASN
Date isolated: 03-04-00
field5: ...

Microbial-Mat-Sample -654
Collected-by: S. Jones
Collection date: 1/24/00
Collection site: Pond 6:
field4: …

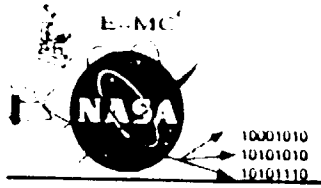
SEM-Image -654
Taken-by: R.Smith
Image date: 1/24/00
Equipment:
Image File:



Hierarchy of ScienceOrganizer Information Resource Types (partial)



Total: 71 predefined resource types; (18 distinct)



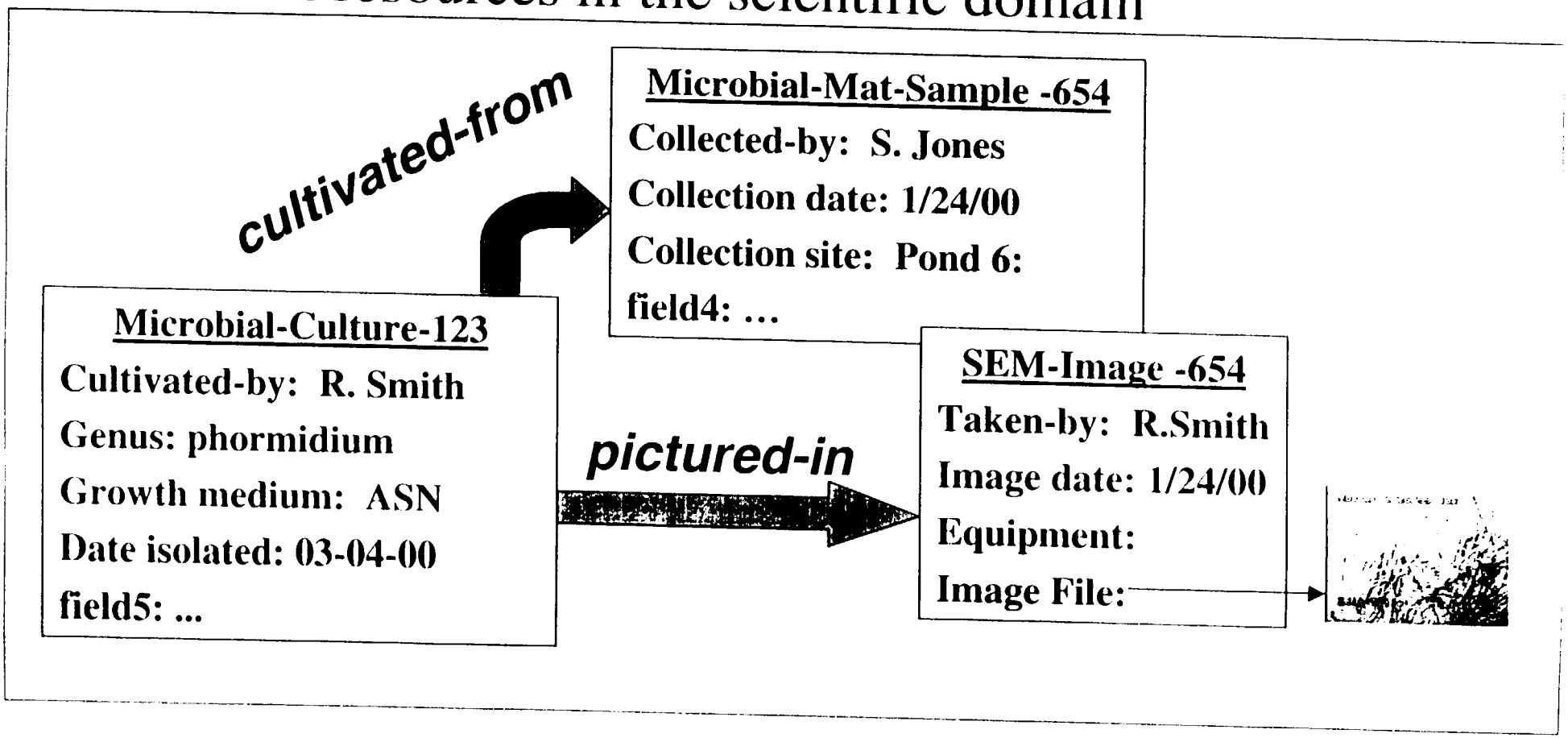
How do Information Resource instances get introduced into ScienceOrganizer?

- Manual input by users
  - information entered via Web forms
  - attached files are uploaded from local computers to the ScienceOrganizer server
- Automatic creation by programs
  - computers interface with ScienceOrganizer server to create information resources

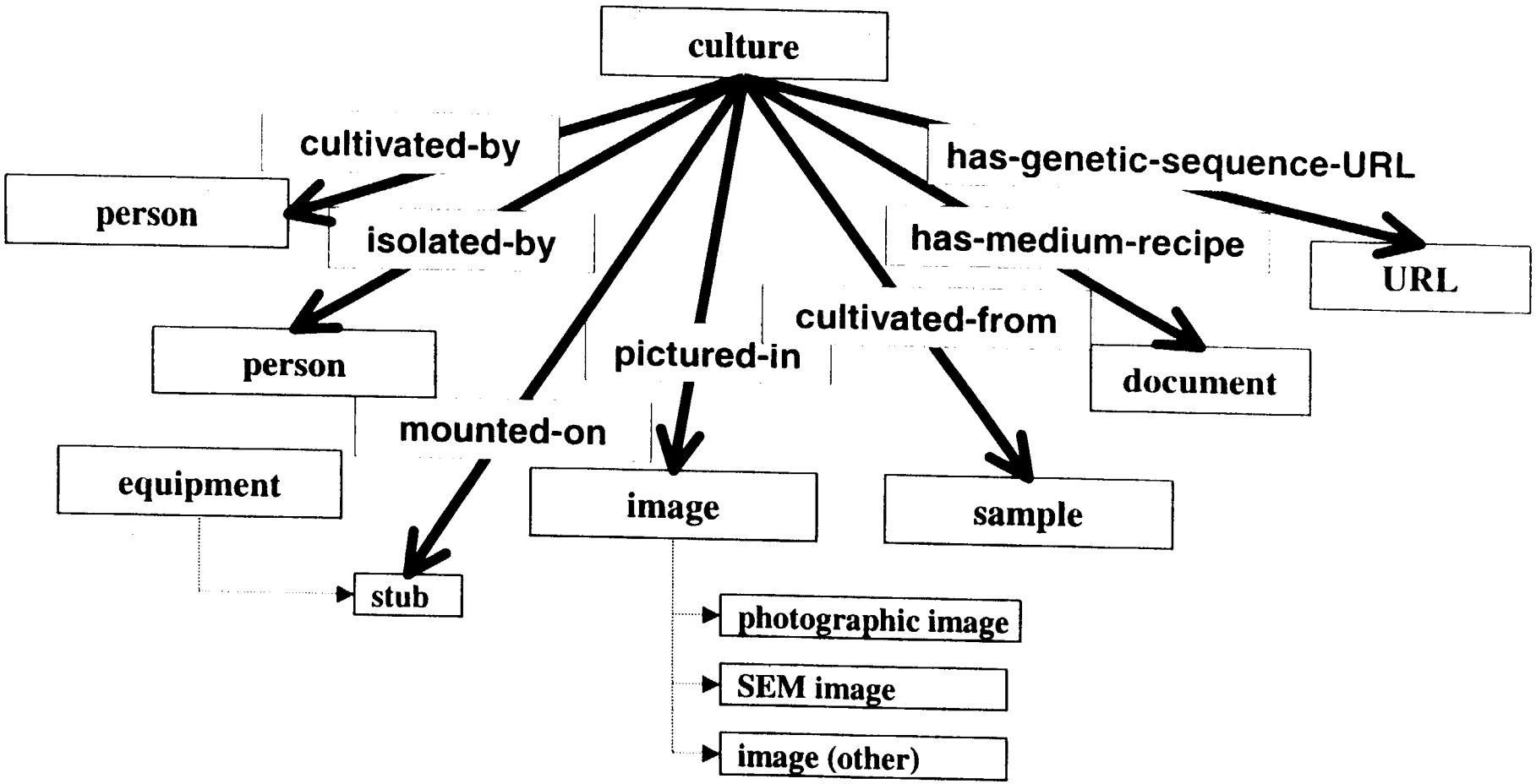
Relationships among Resources ("links")

Ames Research Center

- Information Resources are interrelated by means of named semantic links that characterize the nature of the relationship
- Relationships are defined based on an analysis of the information resources in the scientific domain



Sanctioned Links between "culture" and other types of resources



... a portion of the science information domain theory

Link Types (partial)

relation	domain	range	inverse	id
affiliated_with	participant	institution	has_staff	112
associated_project	sample	project	has_sample	169
authored_by	document	participant	authored	69
calculated_from	measurement	measurement	used_in_calculating	57
camera_used	micrograph	camera	produced_image	36
collected_at	sample	site	collection_site_for	34
collected_by	sample	participant	collected	21
component_instrument	x_ray_detector	TEM	has_component	164
conducted_at	experiment	site	experiment_site_for	49
contained_in	org_item	container	contains	26
cultivated_by	culture	participant	cultivated	24
cultivated_from	culture	sample	source_of	33
custodian_for	participant	culture	has_custodian	43
data_reduced_by	measurement	participant	performed_data_reduction	55
describes_logistics_for	document	trip	logistics_described_in	52
describes_objectives_for	document	trip	objectives_described_in	180
destination_for	study_area	trip	destination	103
employed_in	sample	experiment	employed_sample	183
experimental_measurement_for	measurement	experiment	has_experimental_measurement	51
has_experiment	trip	experiment	conducted_during	66
has_genetic_sequence_info	culture	document	describes_genetic_sequence for	63
has_genetic_sequence_URL	culture	link	describes_genetic sequence for	28
has_lab_user	laboratory	participant	uses_lab	110
has_mail_archive	project	link	mail_archive for	182
has_manager	laboratory	participant	manages	109
has_participant	trip	participant	participated in	67
has_preceding_measurement	acetylene_reduction	acetylene_reduction	has_subsequent_measurement	137
has_profile	participant	profile	profile_for	166
has_sample	trip	sample	collected_during	65
has_site	project	site	site_for	44
has_study_area	project	study_area	study_area_for	101
has_subassembly	flow_box	tray	subassembly of	171
imaged_by	image	participant	took_image	73
imaged_using	SEM_image	SEM	produced_image	71

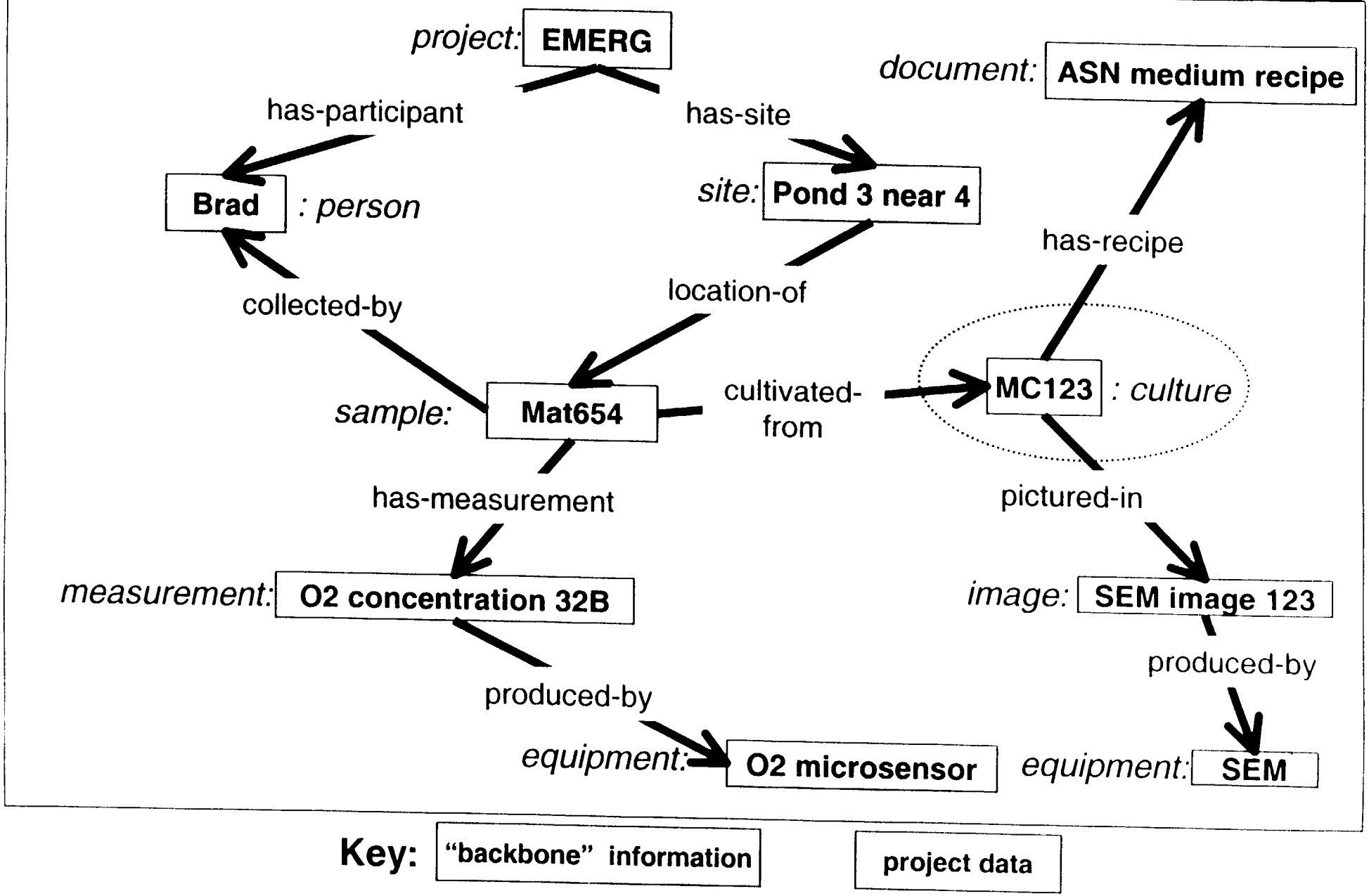
Total: 66 predefined link types

How do resources get linked within ScienceOrganizer?

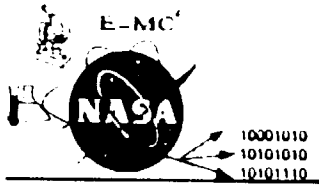
- Manual linking
  - user can explicitly link two existing resources
- Automatic linking
  - when creating a new resource, links are automatically established based on user selections
  - rules can establish links by exploiting knowledge-based inference

("if a person participated in an experiment,
& the experiment was conducted during a specific field trip,
then the person was a participant in that field trip")

Rich "Project Information Network" develops over time



Key: "backbone" information | project data

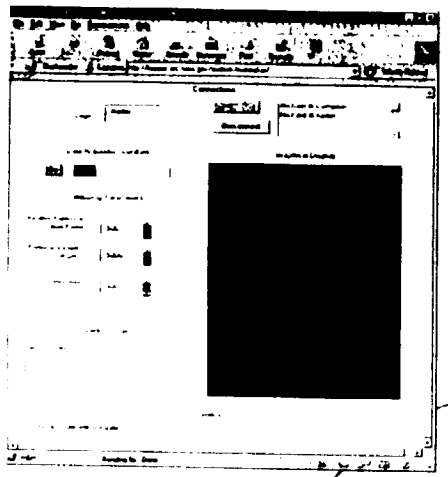


Why a Semantic Hypermedia Model?

- Hierarchical information access models (e.g., directory structures) don't adequately support information cross-correlation
- Semantics impose structure, discipline, and shared understanding across distributed users
- Semantics enable inference, which supports automated linking and field-filling

Auxiliary tools interfacing with ScienceOrganizer

Ames Research Center



real-time measurements

#2 Oxygen Microsensor Remote Controller

sensor

ScienceOrganizer

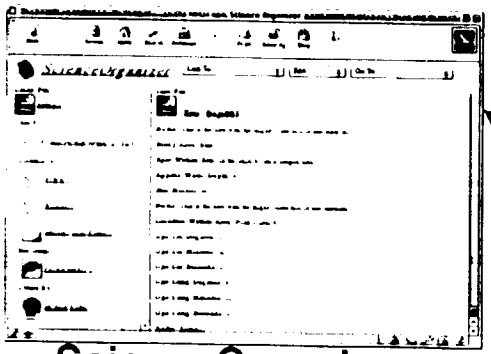
image

#1 Shared Image Annotator

annotated image

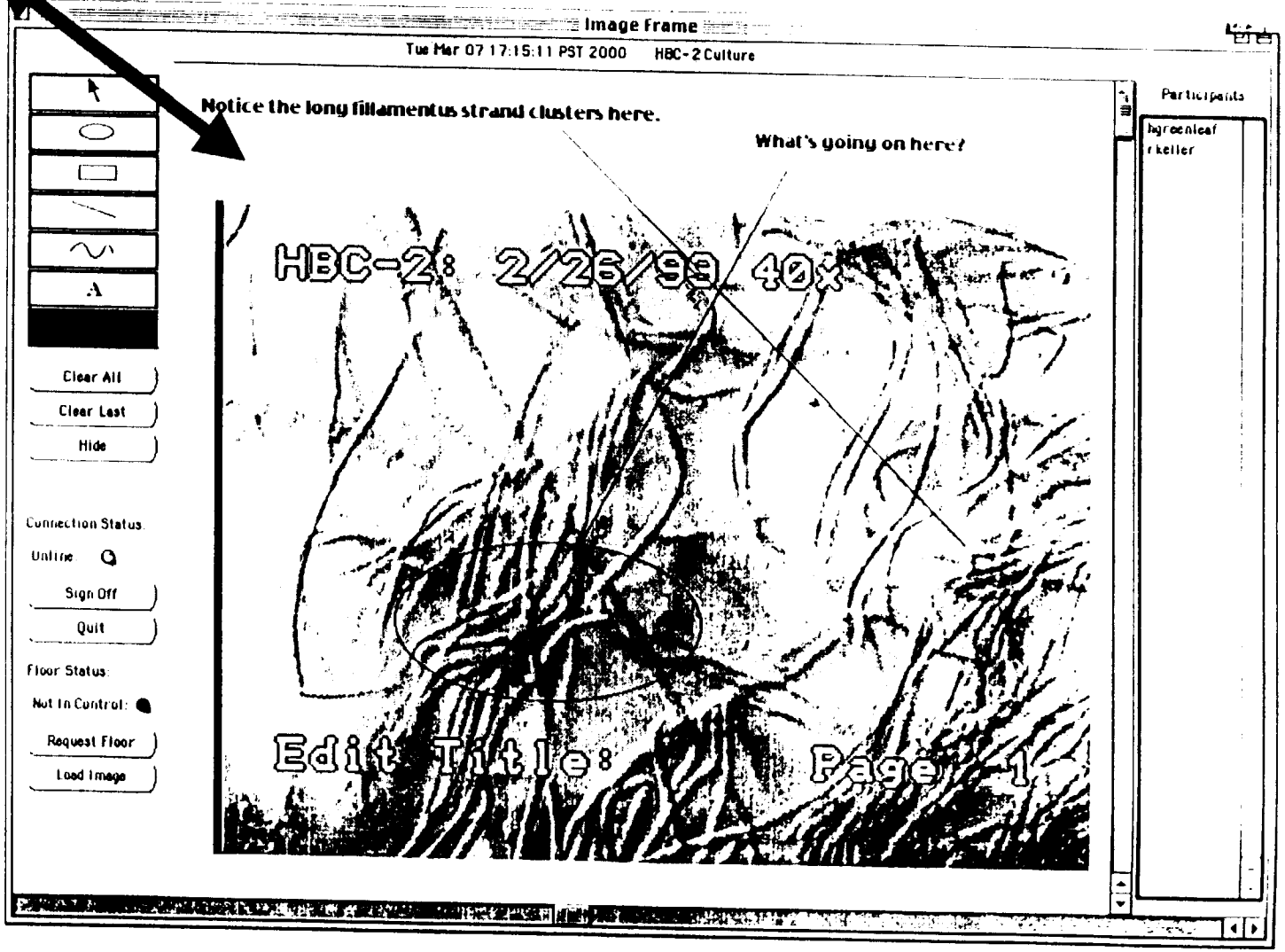
Auxiliary tool #1: Shared Image Annotator

Ames Research Center



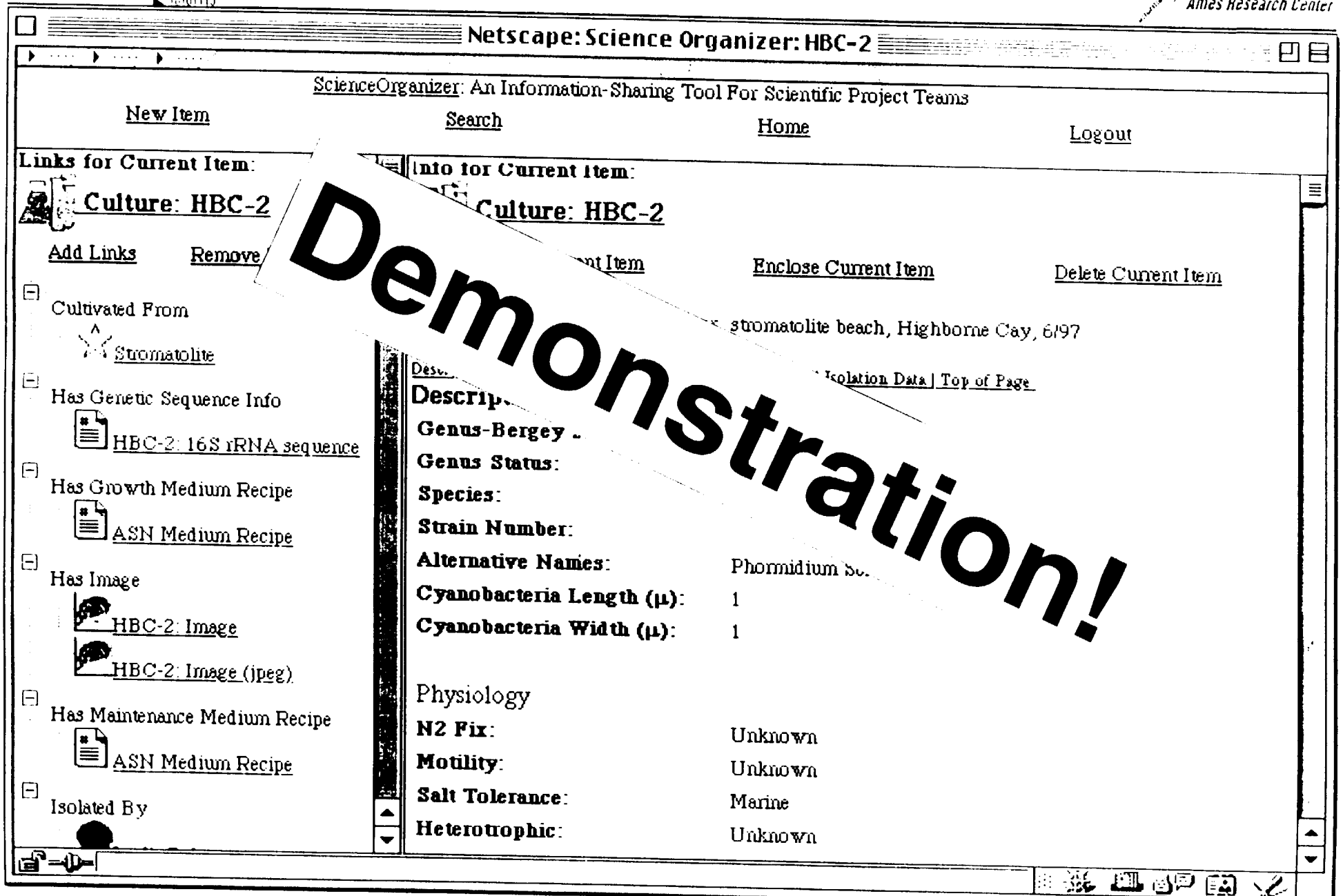
ScienceOrganizer

Collaborative Whiteboarding & Annotation Tool



Uses Sun's JSDT (Java Shared Development Toolkit) for underlying collaborative infrasturcture

ScienceOrganizer Interface



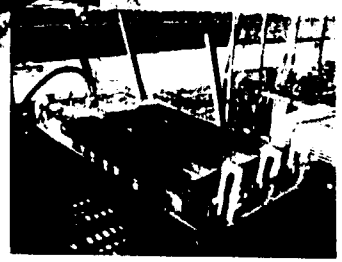
Demonstration!

Astrobiology Greenhouse Facility

Ames Research Center

EMERG

Early Microbial Ecosystems Research Group



Greenhouses Offer:

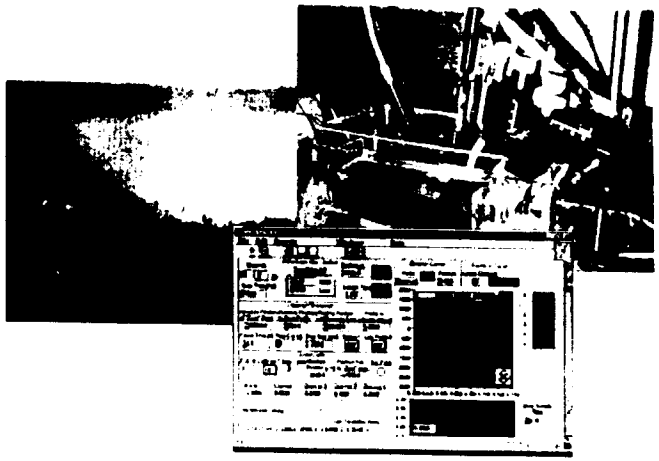
- Natural Light
- Realistic Flow Fields
- Natural Solar UV

Experimental Manipulations

Trace Gas Production and Consumption Under "Early Earth" Conditions

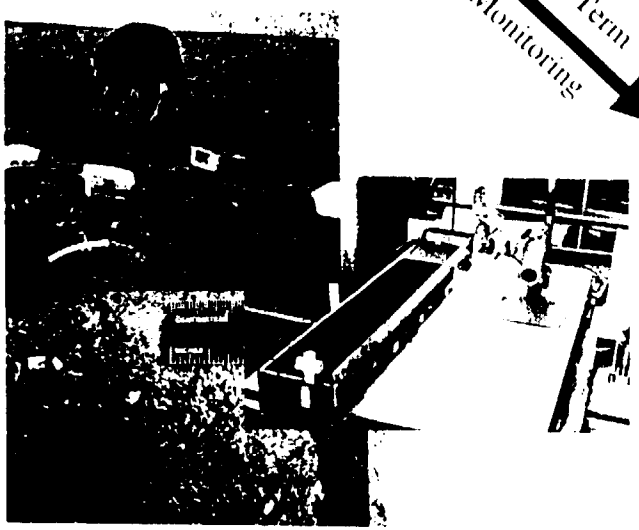
Technology Development

Microsensors, Remote Instrumentation Collaboratory



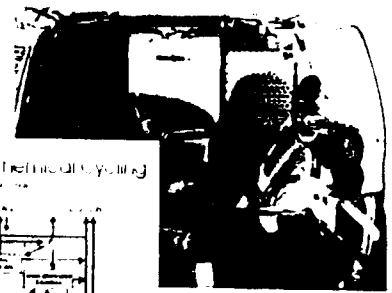
Used for

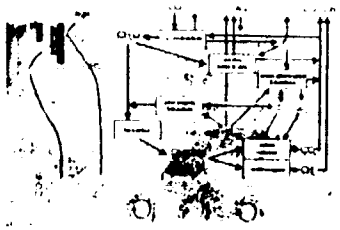
Maintenance of Field Collected Mats and Growth of Defined Communities



Long Term Monitoring

Detailed Studies of Mat Biogeochemistry with Improved Access to Analytical Instrumentation

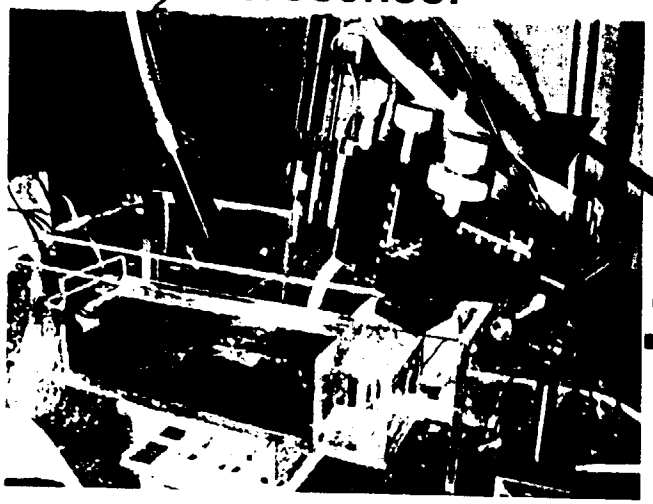




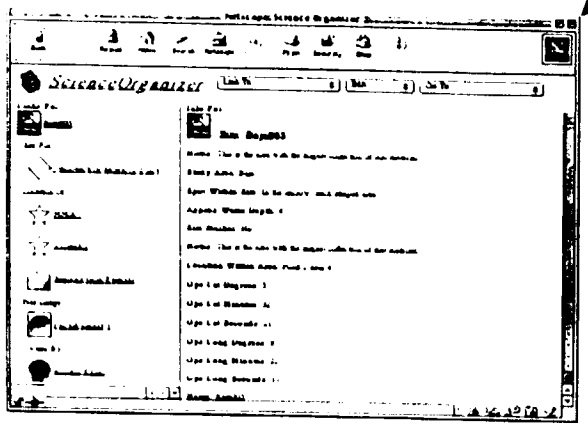
Remote Experimentation:
Oxygen Microsensor Remote Controller

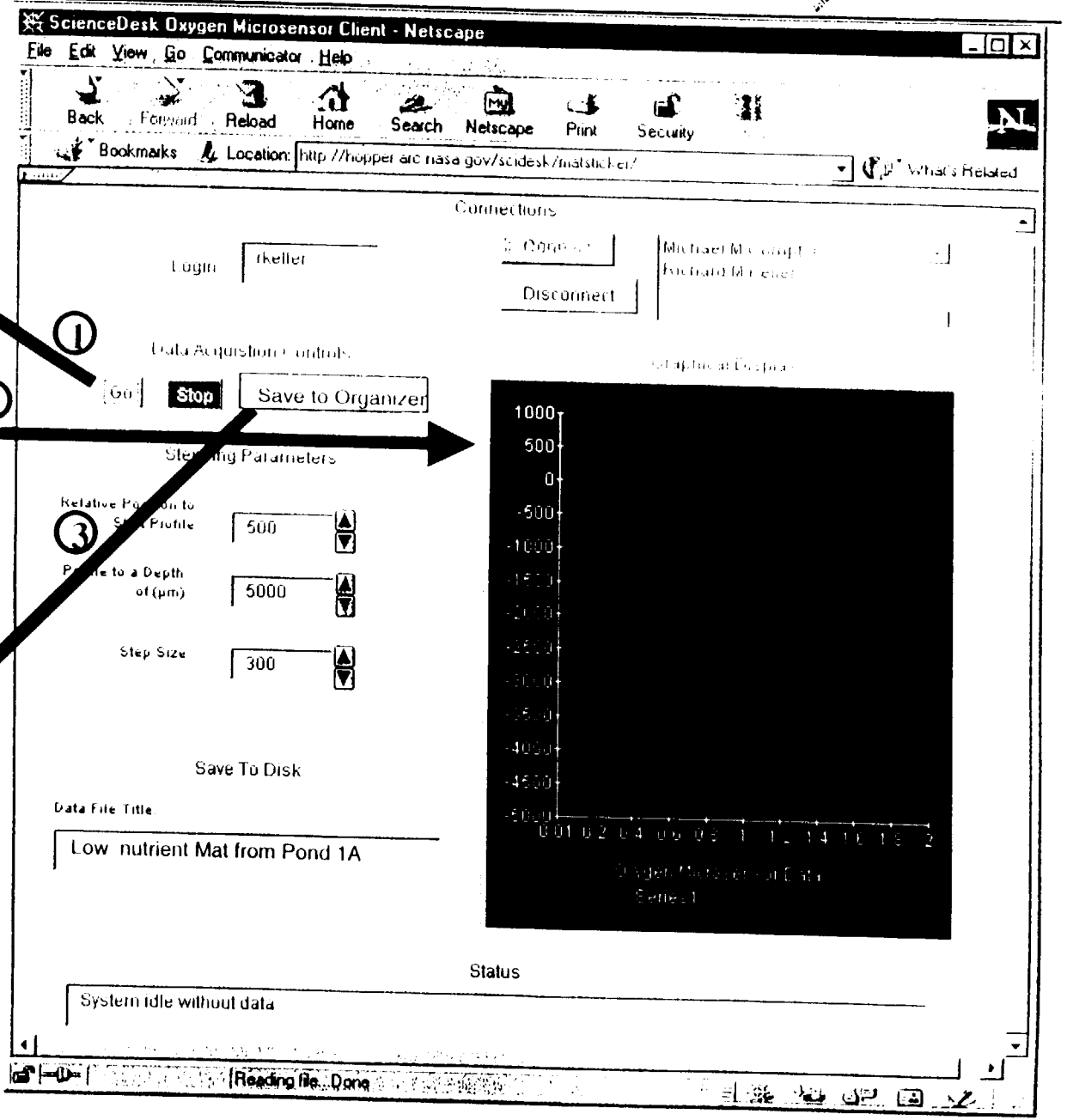
Ames Research Center

O_2 Microsensor



ScienceOrganizer

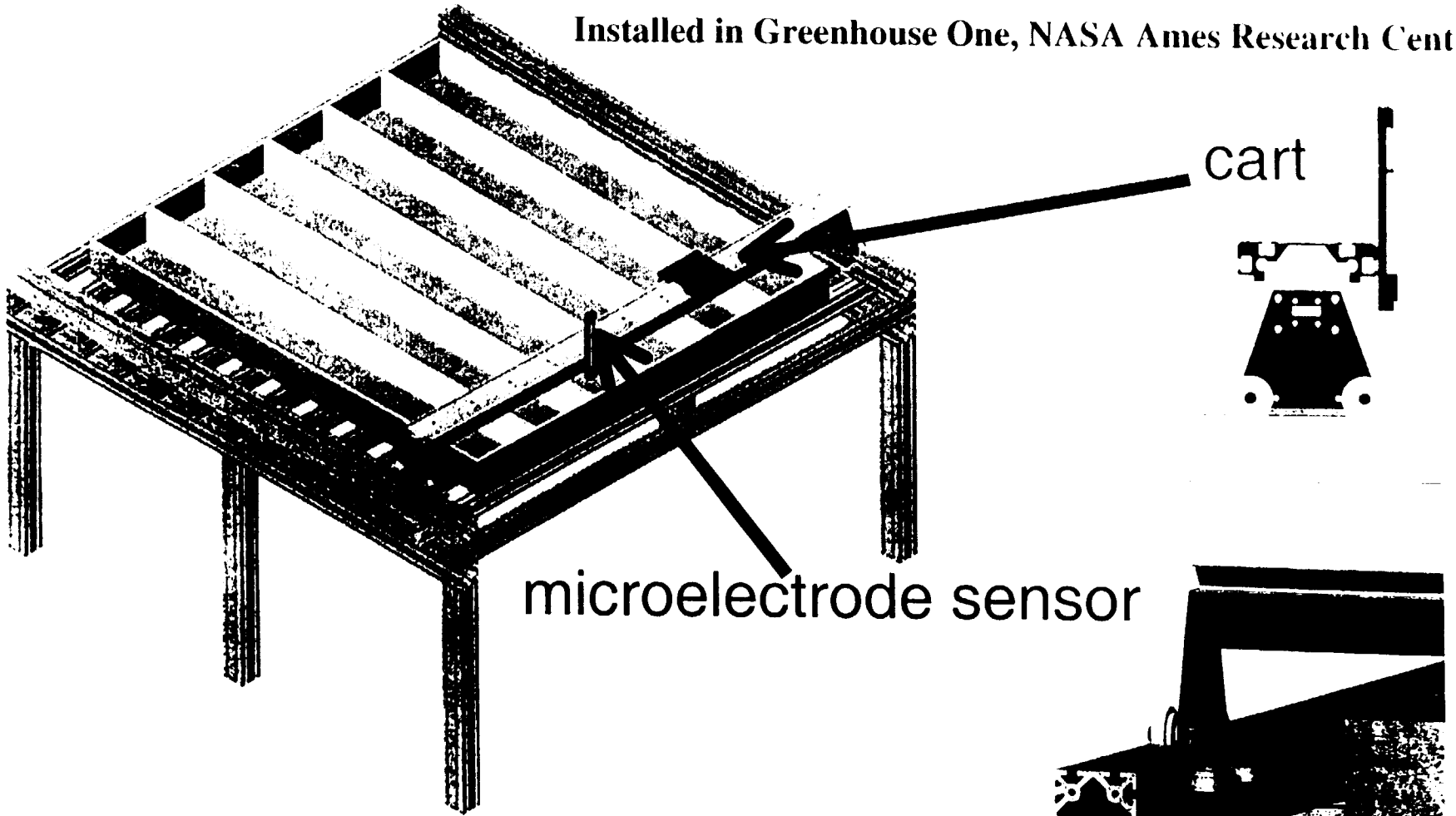




Remotely Operable X-Y-Z stage for Data Acquisition and Imaging

Ames Research Center

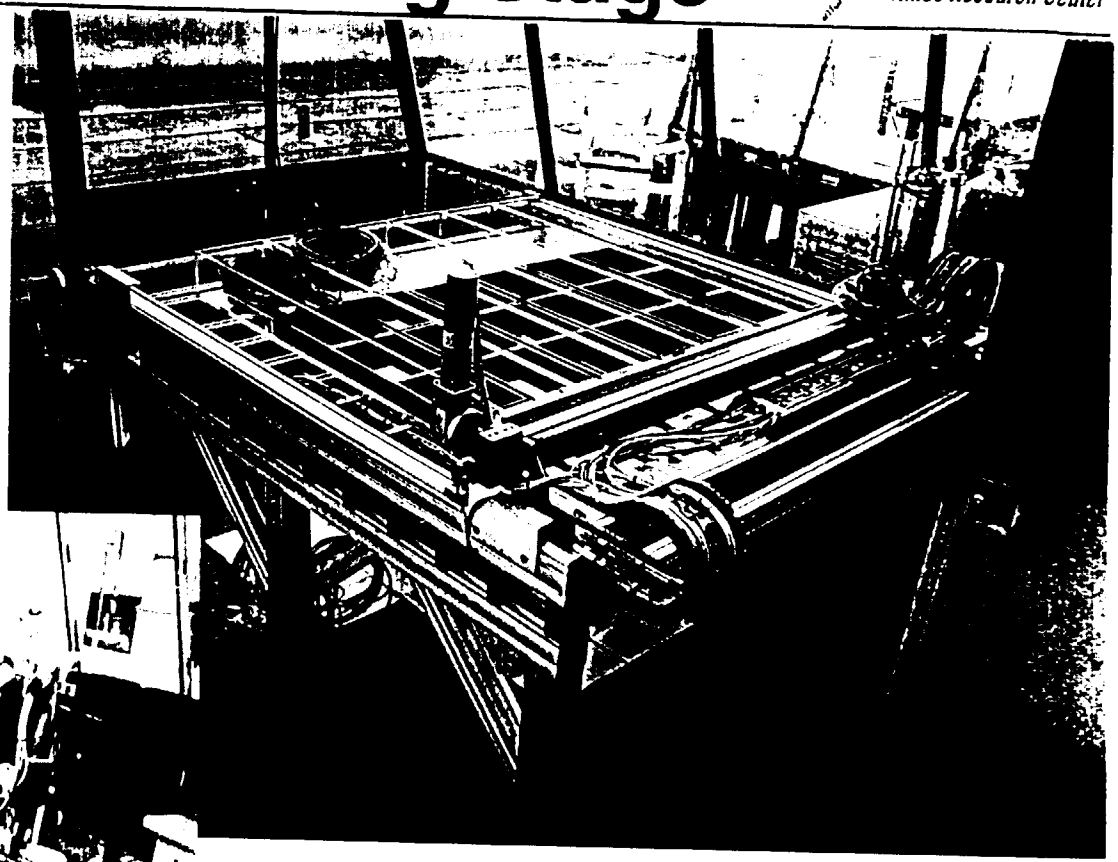
Installed in Greenhouse One, NASA Ames Research Center

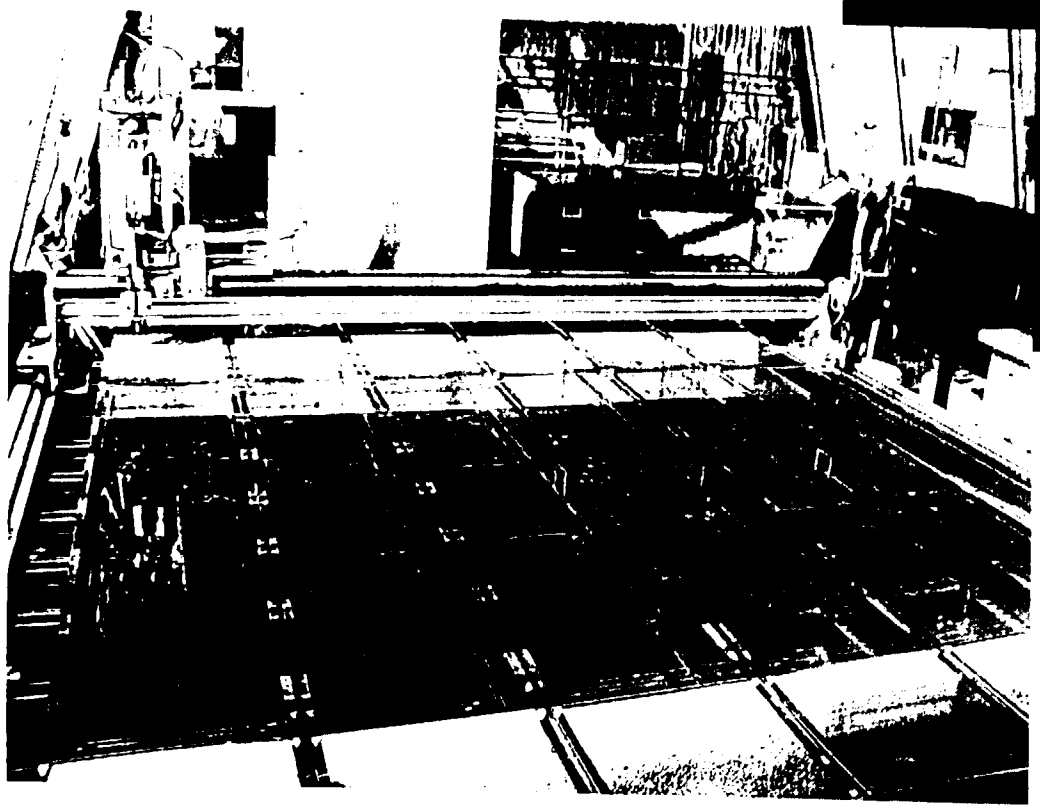


Design/Fabrication: Code FE, NASA Ames, Dan Andrews, Brian Koss

Finished Product:Remotely Operable Positioning Stage

Ames Research Center

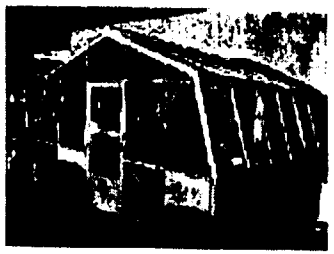




Agent-assisted Collaboratory:
A remote control experimentation facility

Ames Research Center

Greenhouse



Science team

Experiment management tools

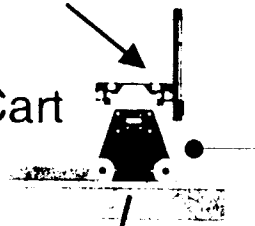
- experiment design
- experiment scheduling
- automated instrument control
- automated monitoring
- automated data acquisition

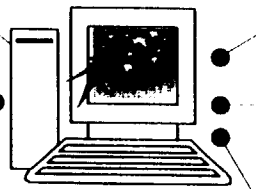
Instrument package

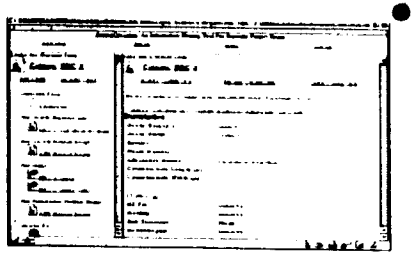
Greenhouse server w/ intelligent agent

Microbial mats

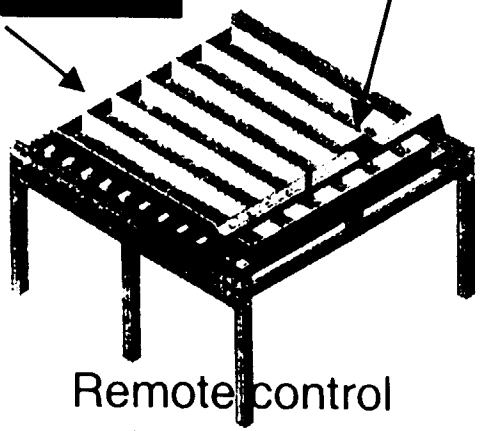
Cart



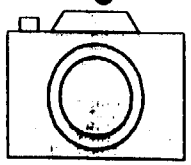




ScienceOrganizer Information Repository



Remote control positioning stage

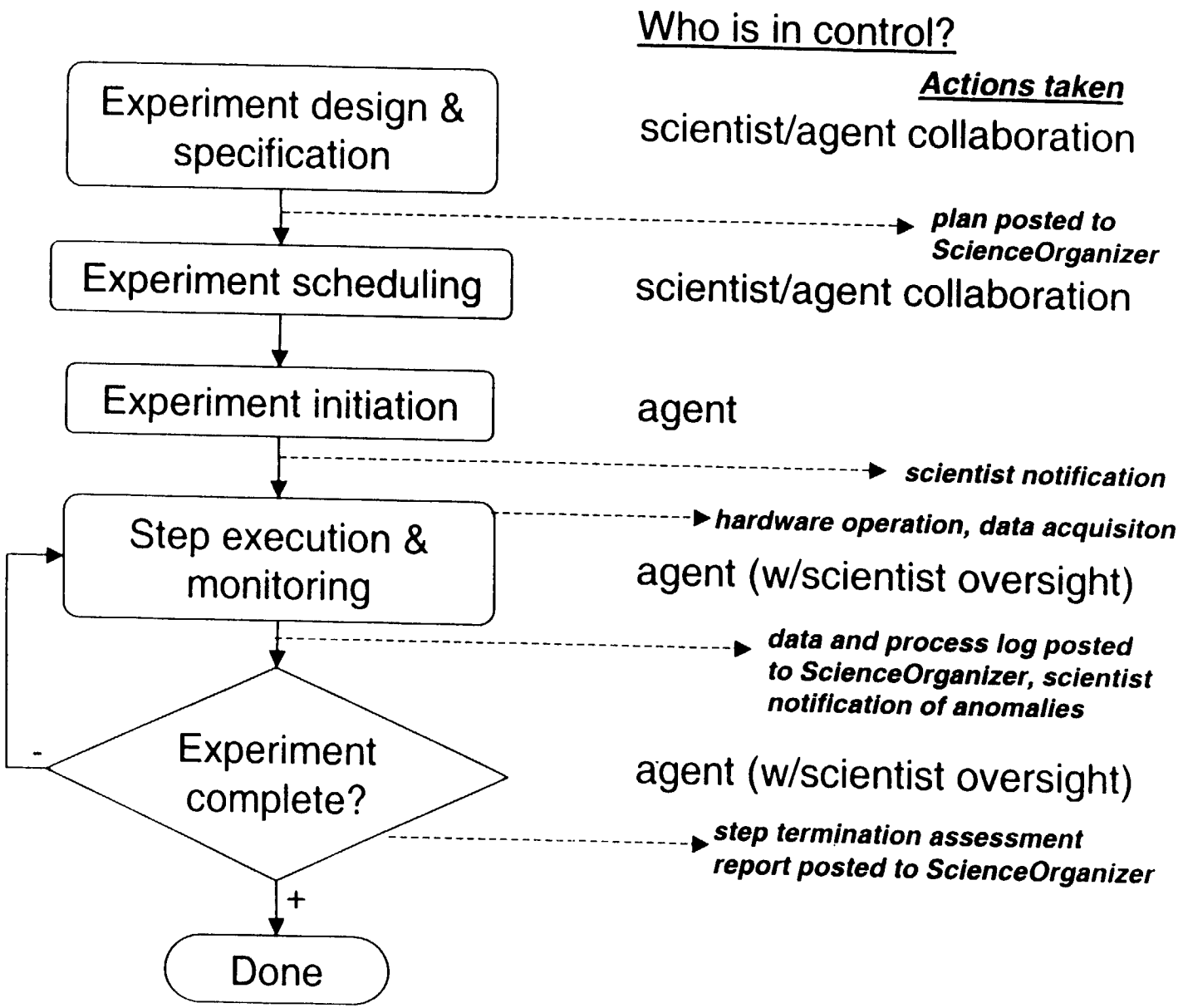


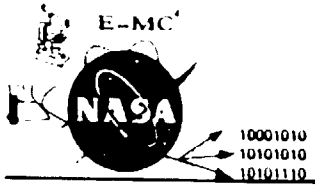
Cameras:

- environmental pan/zoom Webcam
- zooming closeup sensor videocam
- wide-field mat still camera

Remote access & collaboration tools

Experimentation Process





Technical Challenges & Open Issues

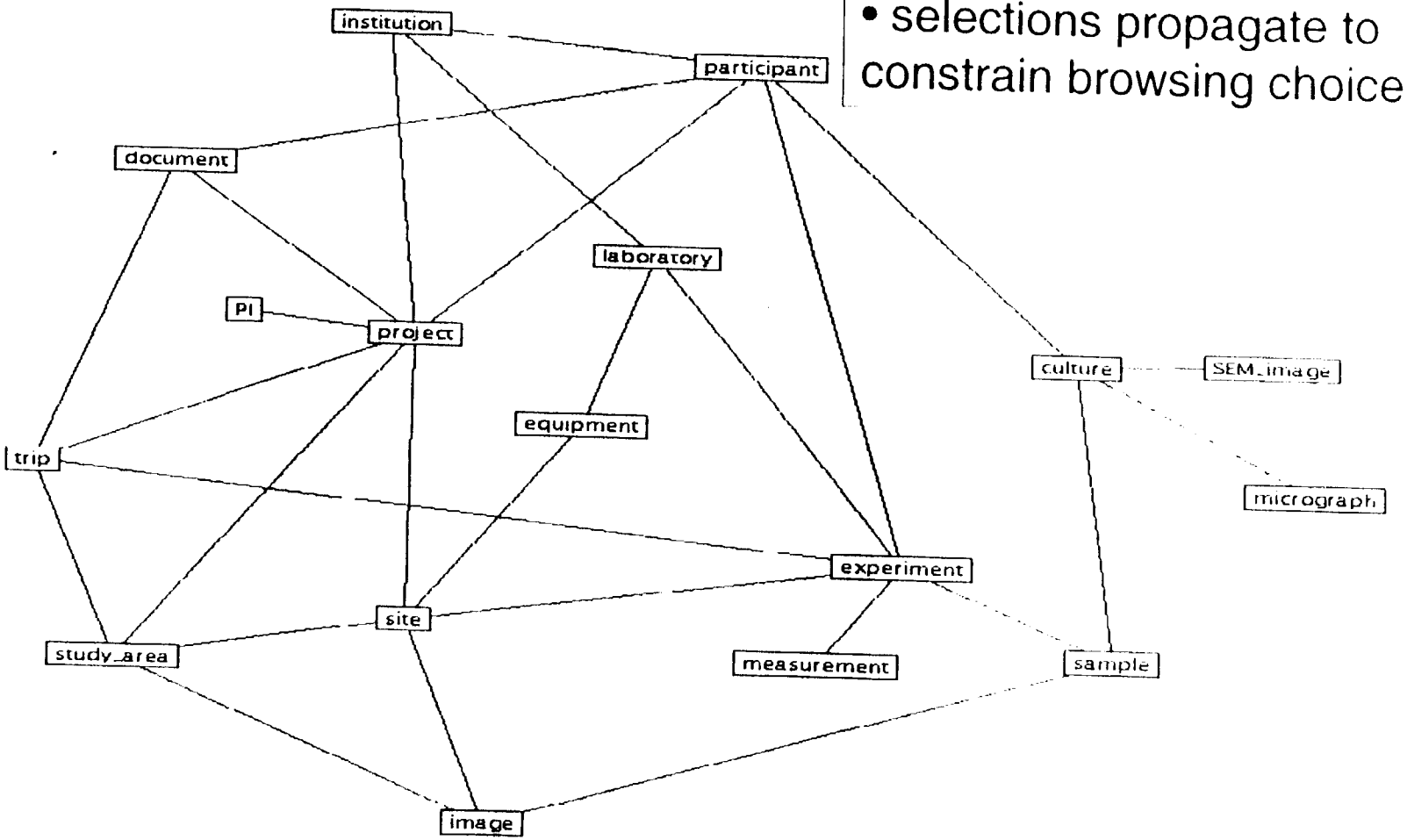
- Hyperspace navigation & visualization
- Customization facilities: interface & domain
- Access control and information-shielding
- Grouping of information resources (folders)
- Scale-up: as of 2/01,~550 nodes/~1700 links/30 users
- Automated knowledge capture: ease user burden
- Novel input modes (pen/tablet, speech)
- Mobil/wireless setting/handhelds

Hyperspace Navigation Aids: Constraint-based browser

Ames Research Center

Applet

- abstract graph
- selections propagate to constrain browsing choices



Scramble | Shake | Stress | Random | Sleep | Wakeup | Stop | Start | Stay | Move

Applet started.

Hyperspace Navigation Aids: Dynamic "hotspots"

- Determine useful navigation hubs ("hotspots") based on a heuristic scoring function that takes into account:
  - number of nodes reachable in one step
  - frequency that node is visited
  - count of outgoing links

Related Work

Repositories:

- Document-sharing systems: BSCW, DocuShare, Postdoc
- Electronic lab notebooks
- Laboratory information mgmt. systems (LIMS)
- Personal information mgmt. systems (PIMS): Outlook, Palm

Hypermedia:

- Semantic hypertext systems / semantic web
- Knowledge-based hypermedia
- Concept maps
- Topic maps

Knowledge representation:

- Semantic nets
- Ontologies