

UNCLASSIFIED

Facilitating Semantic Data Integration and Interoperability between Collaboration Tools and Digital Object Repositories

Linn Marks Collins

Mark Martinez, James Powell, Alex Baker, Tiago Simas, and Susan Heckethorn
eScience and Human-Computer Interaction Team
Los Alamos National Laboratory Research Library



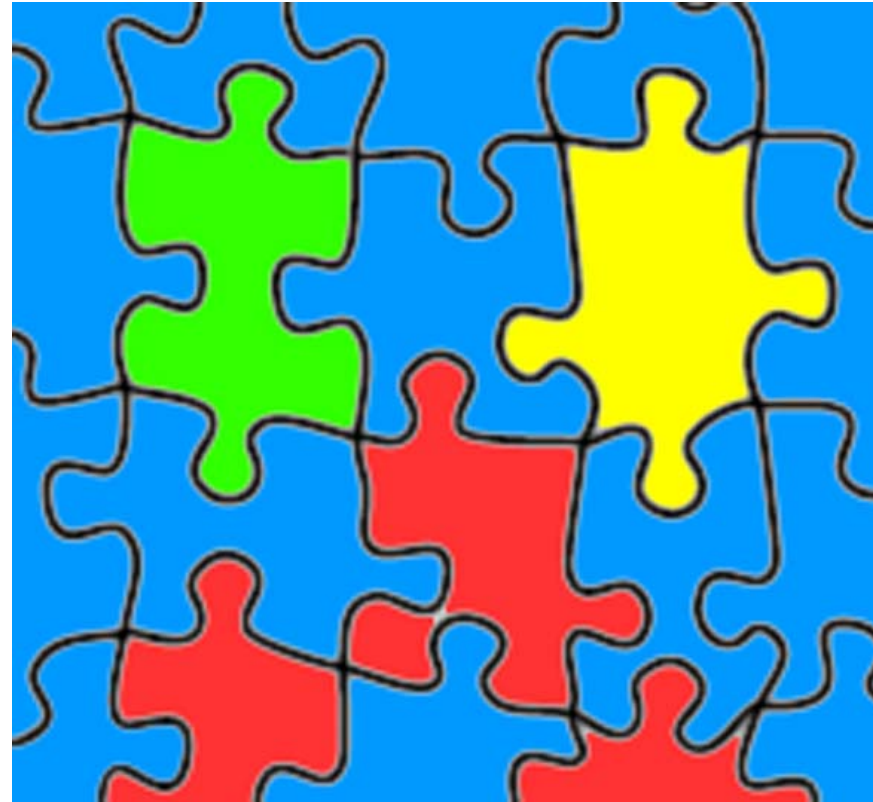
UNCLASSIFIED

Operated by the Los Alamos National Security, LLC for the DOE/NNSA



Topics

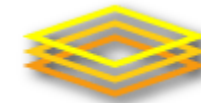
- Context
 - Collaborative eScience
 - Distributed, interoperable resources and repositories
- Social-technical model
- Pilot tests
 - Creating metadata for data sets
 - Discussing collaborative research
 - Writing research proposals
 - Aligning research with strategy
 - Building a community of interest
- Analysis
- Future work



The social-technical puzzle

UNCLASSIFIED

The Open Science Grid



Supported by the National Science Foundation and the US DOE Office of Science



UNCLASSIFIED

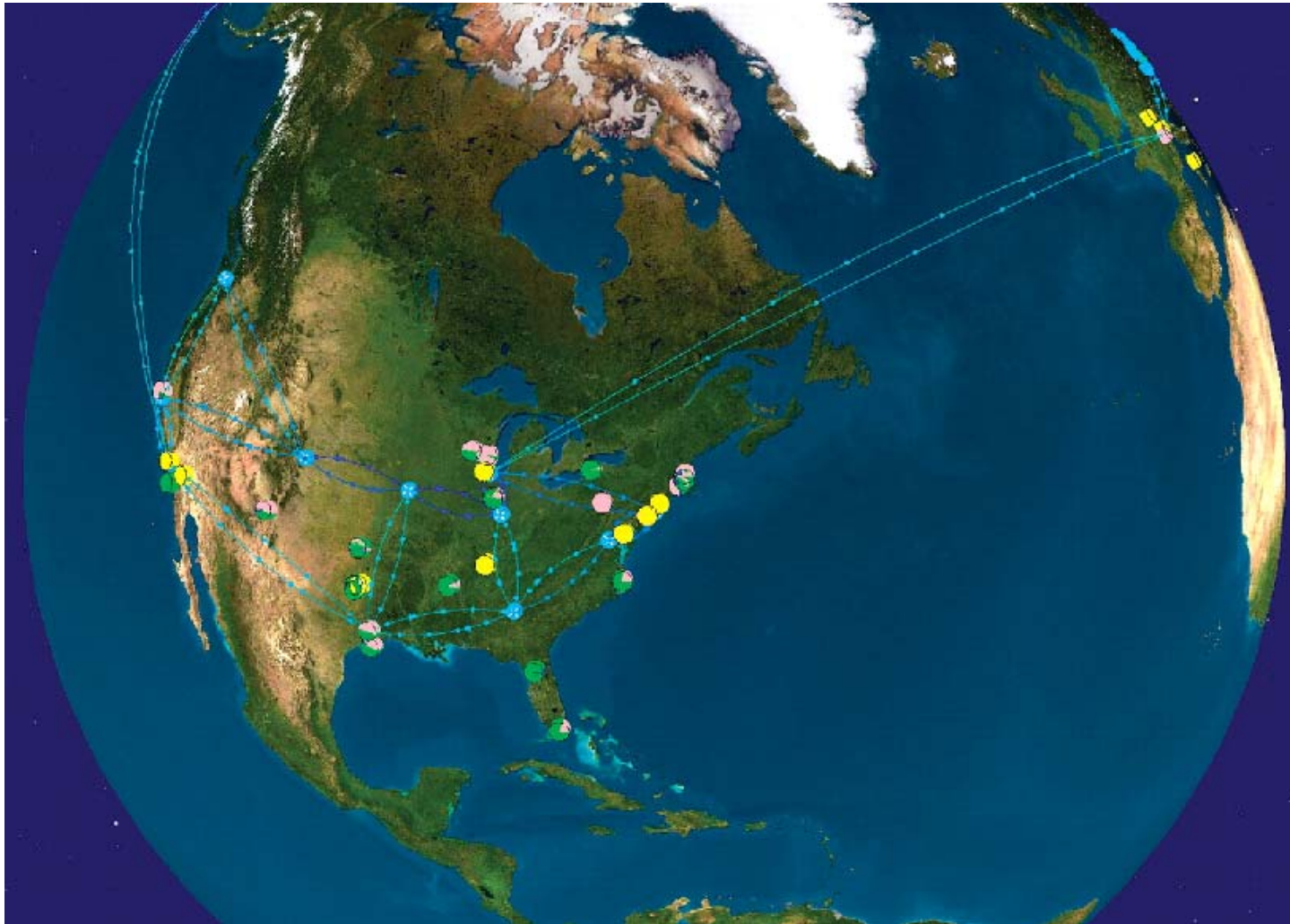
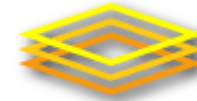
Slide 2

Operated by the Los Alamos National Security, LLC for the DOE/NNSA

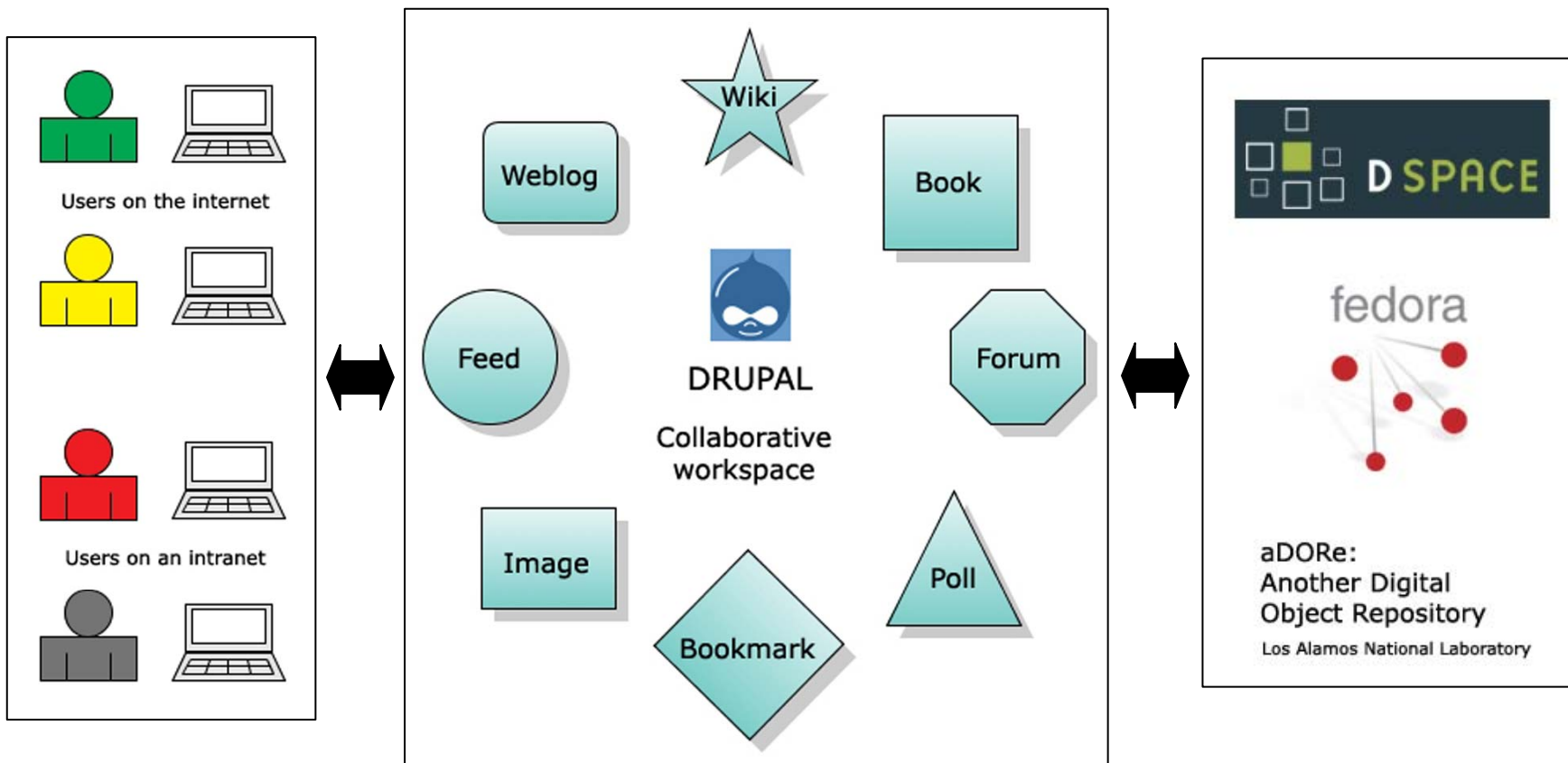


UNCLASSIFIED

The Open Science Grid



Social-technical model



Components of the model

- Users on the Internet or on an intranet
- Open-source collaboration tools for creating content (digital objects)
- Content (digital objects) that can be exported and imported in various XML formats
- Distributed, interoperable digital object repositories
- Unique identifiers for integrating digital objects
 - Metadata with digital objects
 - Digital objects with each other
- Standards for ensuring interoperability
 - Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
 - Digital Item Declaration Language (DIDL)
 - Pathways Core

Collaboration tools

- Drupal
 - Content management system
 - Many types of content nodes
 - Gallery module
 - RSS aggregator
 - <http://drupal.org>
- Gallery
 - Image management system
 - <http://gallery.menalto.com>
- Connotea
 - Social bookmarking tool
 - RSS feeds
 - <http://www.connotea.org>



Drupal

- Includes several kinds of content nodes and tools
- Left sidebar: menu and information about users
- Middle of the page: weblog posts
- Right sidebar: search tool and information feeds

The screenshot shows the Drupal 4.7.3 demo website. The header includes the 'inside Los Alamos' logo, navigation links (ABOUT LANL, NEWS, LIBRARY, JOBS), and a search bar. The main content area features a 'DRUPAL.LANL' banner and a 'Drupal 4.7.3 demo' article. The left sidebar contains a 'Site Menu', 'Linn Collins' profile with links to 'my blog', 'create content', 'forums', 'site map', 'tags', 'my account', 'news aggregator', and 'administer', and a 'Who's online' section showing 1 user and 0 guests. The right sidebar includes a search bar, 'LANL in the news' feed with various news items, and a 'How to create content in Drupal' article with instructions for creating blog entries and book pages.

Access control in Drupal

- Two default user roles
 - Anonymous
 - Authenticated
- Other roles can be added
 - Admin
 - Moderator
- Permissions can be configured by role
- Finer-grained access controls are available
- Some will be integrated in the Drupal core

you are here: [Home](#) » [administer](#) » [Access Control](#)

access control

Help:

Permissions let you control what users can do on your site. Each user role (defined on the [user roles page](#)) has its own set of permissions. For example, you could give users classified as "Administrators" permission to "administer nodes" but deny this power to ordinary, "authenticated" users. You can use permissions to reveal new features to privileged users (those with subscriptions, for example). Permissions also allow trusted users to share the administrative burden of running a busy site.

[permissions](#) | [roles](#) | [access rules](#)

Permission	admin user	anonymous user	authenticated user	moderator user
aggregator module				
access news feeds	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
administer news feeds	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
block module				
administer blocks	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
use PHP for block visibility	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
blog module				
edit own blog	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
book module				
create book pages	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
create new books	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
edit book pages	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
edit own book pages	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
outline posts in books	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
see printer-friendly version	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
comment module				
access comments	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Tags and tag clouds in Drupal

- Tags are like keywords
- Drupal allows both structured and unstructured (free) tagging
- Free tags can be considered user-created, impromptu metadata
- Size of font indicates relative number of entries

inside Los Alamos

ABOUT LANL · NEWS · LIBRARY · JOBS

Search

Los Alamos >> eScience and HCI Team

What is eScience? Bibliography

you are here: Home » Tags

tags

AccessGrid Announcement Configuration Drupal
eScience2006 ESIS IJDJL Karst
Milestones Modules PathForward Reading
RecommendationSystem RIS ScienceSifter Semantic
interoperability SocialNetworkAnalysis Tagging Update

Site Menu

- ▣ Path Forward
- ▣ Karst Project
- ▣ RIS Paper

Search

Search

linn

- ▣ my blog
- ▶ create content
- ▣ forums
- ▣ site map
- ▼ tags
 - ▣ Administrivia
 - ▣ eScience
- ▣ my account
- ▣ recent posts
- ▶ news aggregator
- ▶ administer

Who's online

There are currently 1 user and 0 guests online.

Online users:

linn

Active forum topics

Questions about TR Drupal

eScience Project Taxonomy: What tags/terms should we use?

Should forum topics be promoted to the front page?

Changes in username and lanlauth.module

Search module observations

more

Connotea Bookmarks

Web journals threaten peer-review system

K-Logging: Supporting KM with Web Logs

Instruments Sensors and the Semantic Grid

cWebsite: ancer

Biomedical Informatics Grid or caBIG™

Decentralized Information Group


OpenAcademic :: bringing education to all

Using Drupal: A Web Experience

Using open source software to design, develop, and deploying a collaborative

Gallery

- Image management system
- Can be integrated with Drupal
- Includes photo albums and layouts for organizing images
- Supports over 30 languages

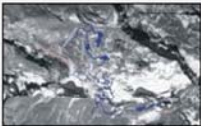
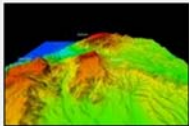


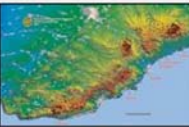




Cave Areas Map/Image Gallery - Speleoukraine.net
 Gallery/Archive of maps and images of cave areas and caves

Home :: Login
 Album list :: Last uploads :: Last comments :: Most viewed :: Top rated :: My Favorites :: Search

Choose your language [v]

Top rated


 <p style="font-size: x-small;">Άεροφωτό Ιδρυματισμού της Βαλκικής Ημιορεινής Περιοχής, Αδριατική (Ortobalagan Valley aerophoto, with main caves outline)</p> <p style="font-size: x-small;">78 views</p> <p style="font-size: x-small;">Άεροφωτό βασισμένο σε αεροφωτό Ιδρυματισμού της Βαλκικής Ημιορεινής Περιοχής, Αδριατική με χάρτες των κύριων σπηλιών, Αδριατική.</p> <p style="font-size: x-small;">Aerophoto of the Ortobalagan Valley, with maps of the main cave outlined, Arabika Massif.</p> <p style="font-size: x-small;">☆☆☆☆☆ (2 votes)</p>	 <p style="font-size: x-small;">Ορεινά 3D της Αδριατικής Ημιορεινής (3D view of the Chatyrdag massif from north-east)</p> <p style="font-size: x-small;">57 views</p> <p style="font-size: x-small;">Εικόνα 3D της Αδριατικής Ημιορεινής. Η εικόνα δημιουργήθηκε από το DEM της Ουκρανίας, με βάση τα δεδομένα του SRTM.</p> <p style="font-size: x-small;">The image was generated from DEM of Ukraine based on the radar topography data (SRTM).</p> <p style="font-size: x-small;">☆☆☆☆☆ (1 votes)</p>	 <p style="font-size: x-small;">Διαμορφωμένο χάρτη της Αδριατικής Ημιορεινής, περιγράμμιση 50m (Arabika-Bzybsky relief map, contours 50m)</p> <p style="font-size: x-small;">70 views</p> <p style="font-size: x-small;">Διαμορφωμένο χάρτη της Αδριατικής Ημιορεινής, περιγράμμιση 50m.</p> <p style="font-size: x-small;">A relief map of the Arabika Massif. Generated from a 3D-model based on radar topography data. Hypsometric contours are drawn at 50m.</p> <p style="font-size: x-small;">☆☆☆☆☆ (1 votes)</p>
 <p style="font-size: x-small;">Διαμορφωμένο χάρτη της Ποδολίας και Βουκοβίνης, με σπηλιές (Shaded map of the Podolje and Bukovina region, with caves)</p> <p style="font-size: x-small;">98 views</p> <p style="font-size: x-small;">Διαμορφωμένο χάρτη της Ποδολίας και Βουκοβίνης, με σπηλιές. Δημιουργήθηκε από το GIS "Karst and Caves of Ukraine", με βάση τα δεδομένα του SRTM και τον Ουκρανικό Κατάστιχο Σπηλιών της Ουκρανίας. Σπηλιές - κόκκινα σημεία (μεγαλύτερα σπηλιές - πάνω από 1 χμ μήκος).</p> <p style="font-size: x-small;">☆☆☆☆☆ (4 votes)</p>	 <p style="font-size: x-small;">Διαμορφωμένο χάρτη των βουνών της Κριμαίας, με σπηλιές και πηγές (Shaded map of the Crimean mountains, with caves and springs)</p> <p style="font-size: x-small;">98 views</p> <p style="font-size: x-small;">Διαμορφωμένο χάρτη των βουνών της Κριμαίας, με σπηλιές και πηγές. Δημιουργήθηκε από το GIS "Karst and Caves of Ukraine", με βάση τα δεδομένα της Κριμαίας (SRTM) και τον Ουκρανικό Κατάστιχο Σπηλιών της Ουκρανίας. Σπηλιές - κόκκινα σημεία, Πηγές - μπλε σημεία.</p> <p style="font-size: x-small;">☆☆☆☆☆ (1 votes)</p>	 <p style="font-size: x-small;">Διαμορφωμένο χάρτη της Αδριατικής Ημιορεινής, περιγράμμιση 50m (Arabika massif shaded map, contours at 50m)</p> <p style="font-size: x-small;">75 views</p> <p style="font-size: x-small;">Διαμορφωμένο χάρτη της Αδριατικής Ημιορεινής, περιγράμμιση 50m.</p> <p style="font-size: x-small;">A shaded map of the Arabika massif. Generated from a 3D-model based on radar topography data. Hypsometric contours are drawn at 50m.</p> <p style="font-size: x-small;">☆☆☆☆☆ (1 votes)</p>

 **Los Alamos**
 NATIONAL LABORATORY
 EST. 1943

Operated by the Los Alamos National Security, LLC for the DOE/NNSA

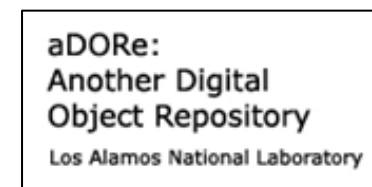
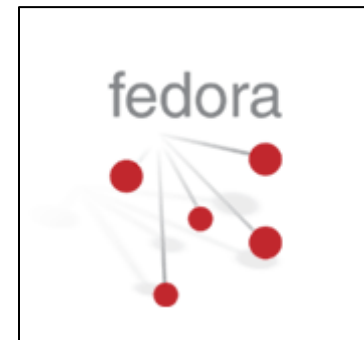
UNCLASSIFIED

Slide 10



Digital object repositories

- DSpace
 - MIT
 - Hewlett-Packard
 - Institutional repository
 - China Museum Project
 - Includes a wiki (Mediawiki)
- Fedora
 - Cornell
 - University of Virginia
 - Includes a wiki (Mediawiki)
- aDORe
 - LANL Research Library
 - Herbert Van de Sompel et al.



Open Archives Initiatives Protocol for Metadata Harvesting (OAI-PMH)

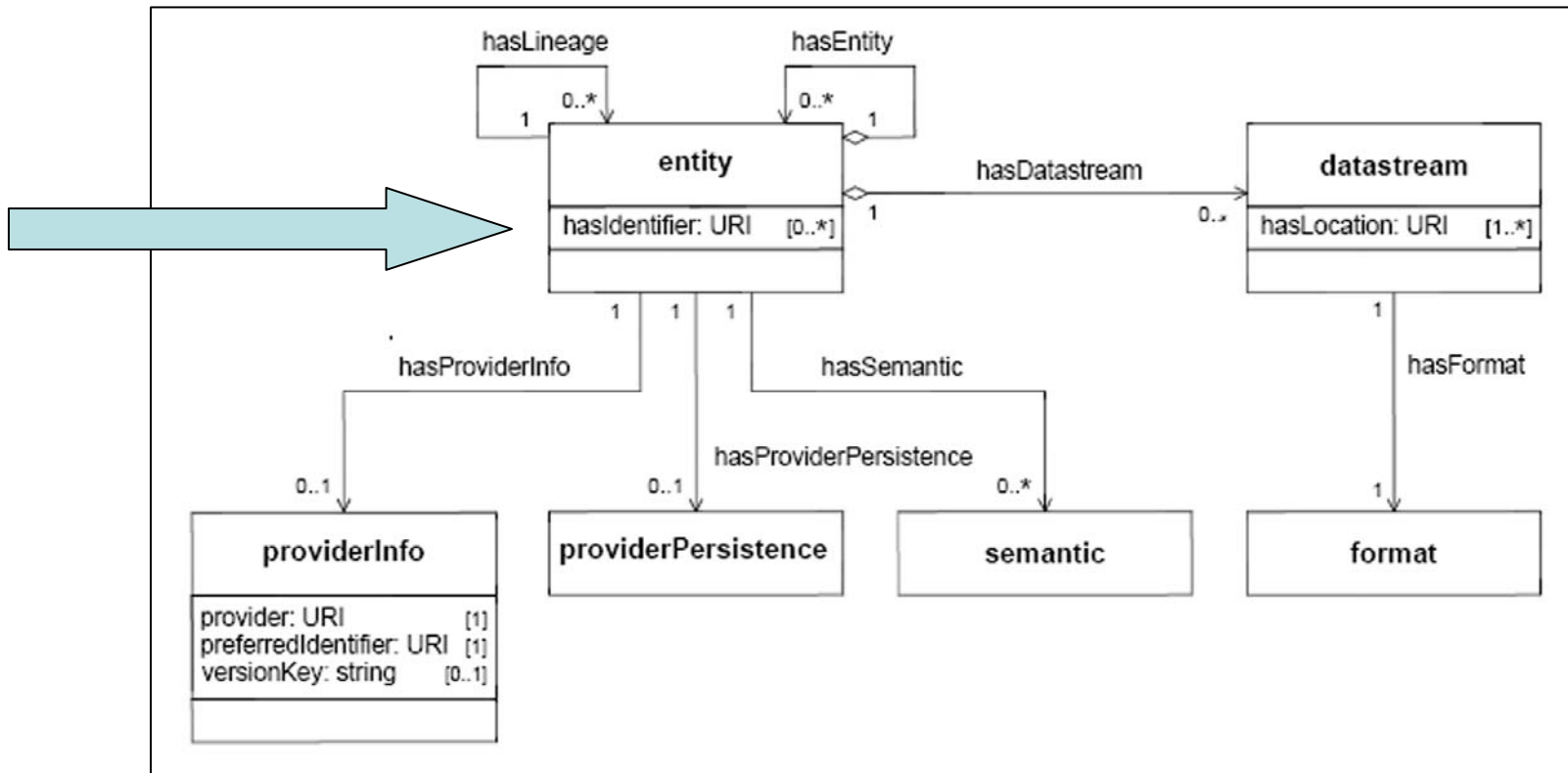
- Facilitates interoperability of digital object repositories through metadata harvesting
- Definitions
 - Repository: Network-accessible server that can process OAI-PMH requests
 - Harvester: Client application that issues OAI-PMH requests
 - Item: Constituent of a repository from which metadata about a resource can be disseminated
 - Unique identifier: Information that unambiguously identifies an item within a repository and is used in OAI-PMH requests for extracting metadata from the item



Further information:

<http://www.openarchives.org/OAI/openarchivesprotocol.html#DefinitionsConcepts>

Unique identifiers



Pathways Core: A Data Model for Cross-Repository Services.
 Bekaert, Liu, Van de Sompel, Lagoze, Payette, and Warner. Poster for JCDL 2006.

Pilot test 1: Creating metadata for data sets

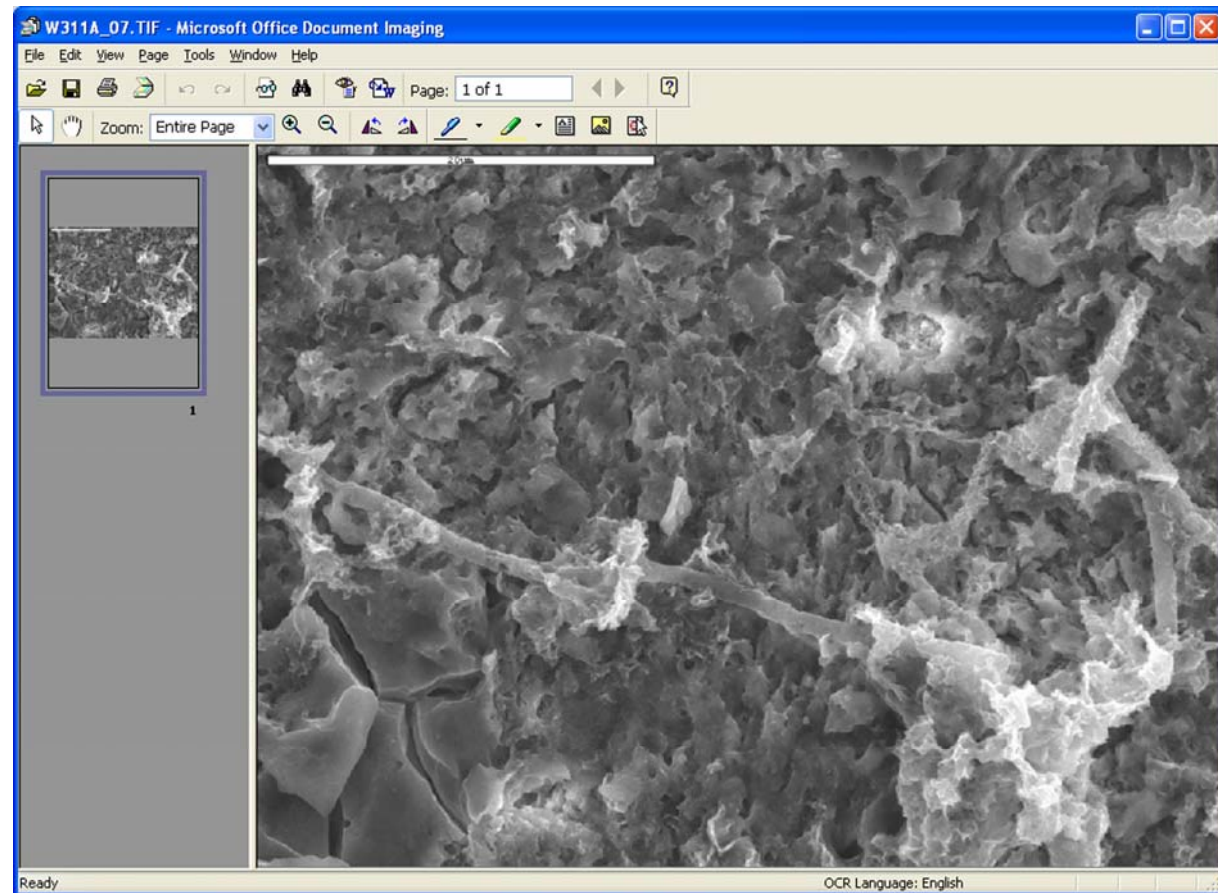
- Many data sets have little or no metadata
 - Stored in scientists' file cabinets or on their hard drives
- Many data sets have inadequate metadata
 - Time-sensitive, so needs to be updated
 - Domain-specific, so requires multidisciplinary perspectives
- Data sets need to be curated and archived
- National Science Foundation Cyberinfrastructure vision
 - Research vs. resource vs. reference data sets
 - Publishing data sets for re-use
 - Archiving reference data sets

Social and technical obstacles

- Creating metadata needs to be automated to the degree possible
 - Self-archiving data sets
- Expert curation requires human expertise, however
- System should be designed to:
 - Support collaborative annotation of data sets by experts from various institutions and locations
 - Include a workflow that is data-centric
 - Associate data with metadata by way of unique identifiers
 - Archive data sets with their associated metadata
 - Support semantic integration and interoperability in and across domains

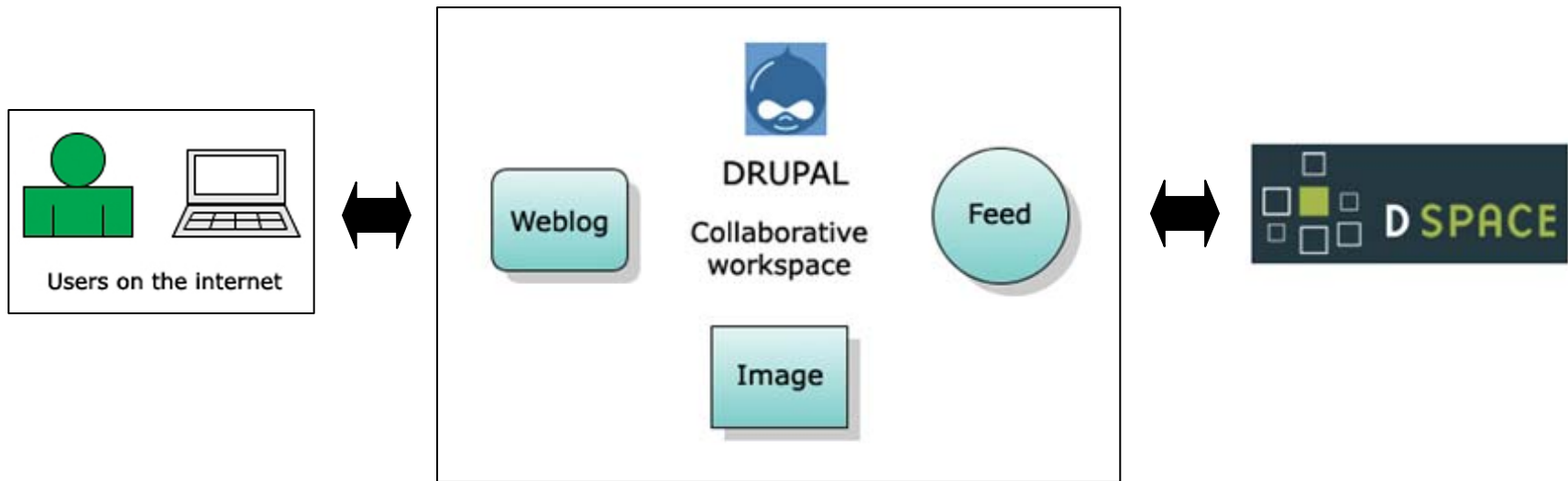
Image data sets

- Karst science
- Scanning electron microscope (SEM)
- SEM images of microbial life forms



DSPACE - University of New Mexico - Karst and Cave Studies

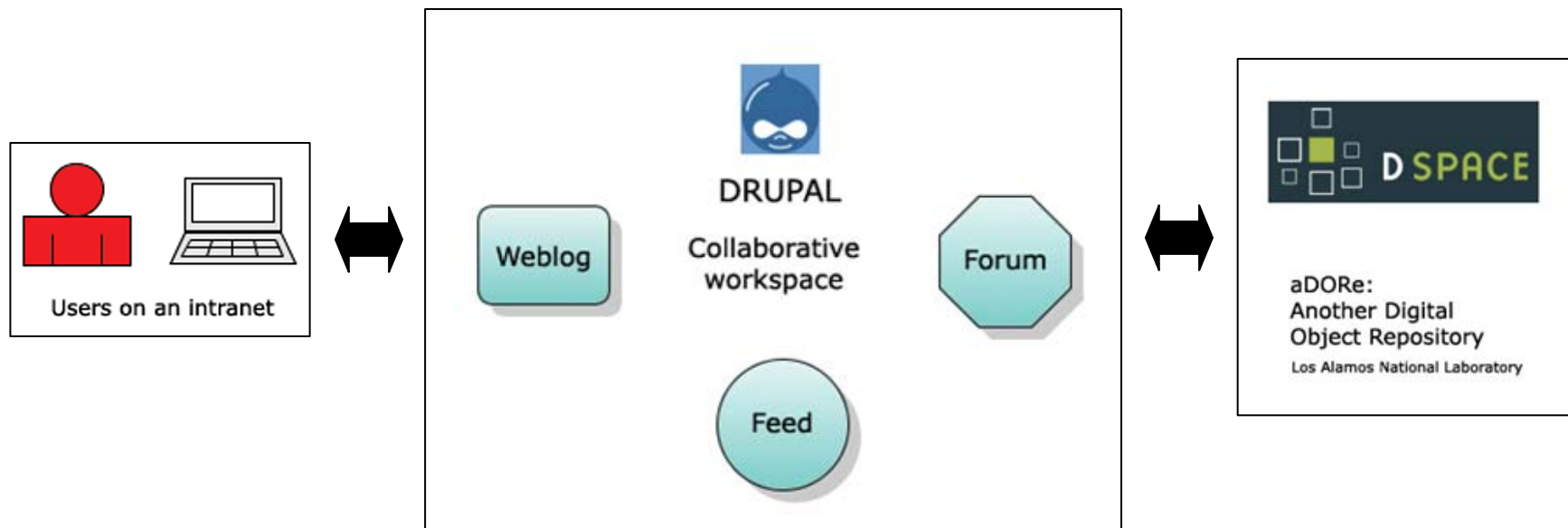
Social-technical model for pilot test 1



Pilot test 2: Discussing collaborative research

- Professional scientists at LANL
- Working in multiple organizations, in various locations, and on similar problems
- System should be designed to:
 - Restrict access to groups of users on an intranet
 - Provide access to shared information
 - Facilitate open-ended discussion of information
 - Enable the creation of shared knowledge
 - Support related research activities
 - Preserve institutional memory

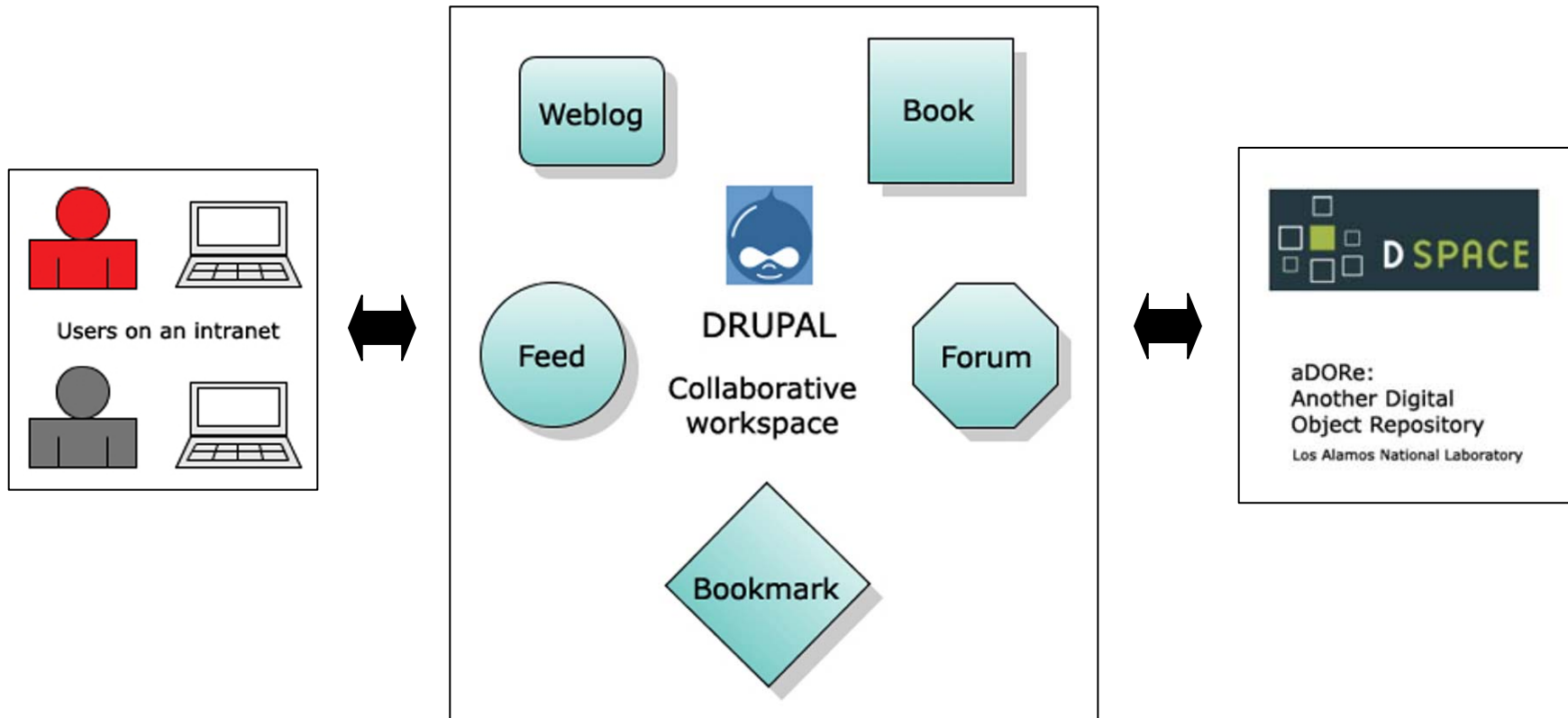
Social-technical model for pilot test 2



Pilot test 3: Writing research proposals

- Scientists who write research proposals
- Important work for the scientists and for the institution
- Currently write proposals collaboratively in email and in text processing tools
- Often store proposals on their hard drives or in their file cabinets
- System should be designed to:
 - Restrict access to specific groups of users on an intranet
 - Facilitate collaborative discussion and authoring
 - Track individual contributions
 - Preserve institutional memory

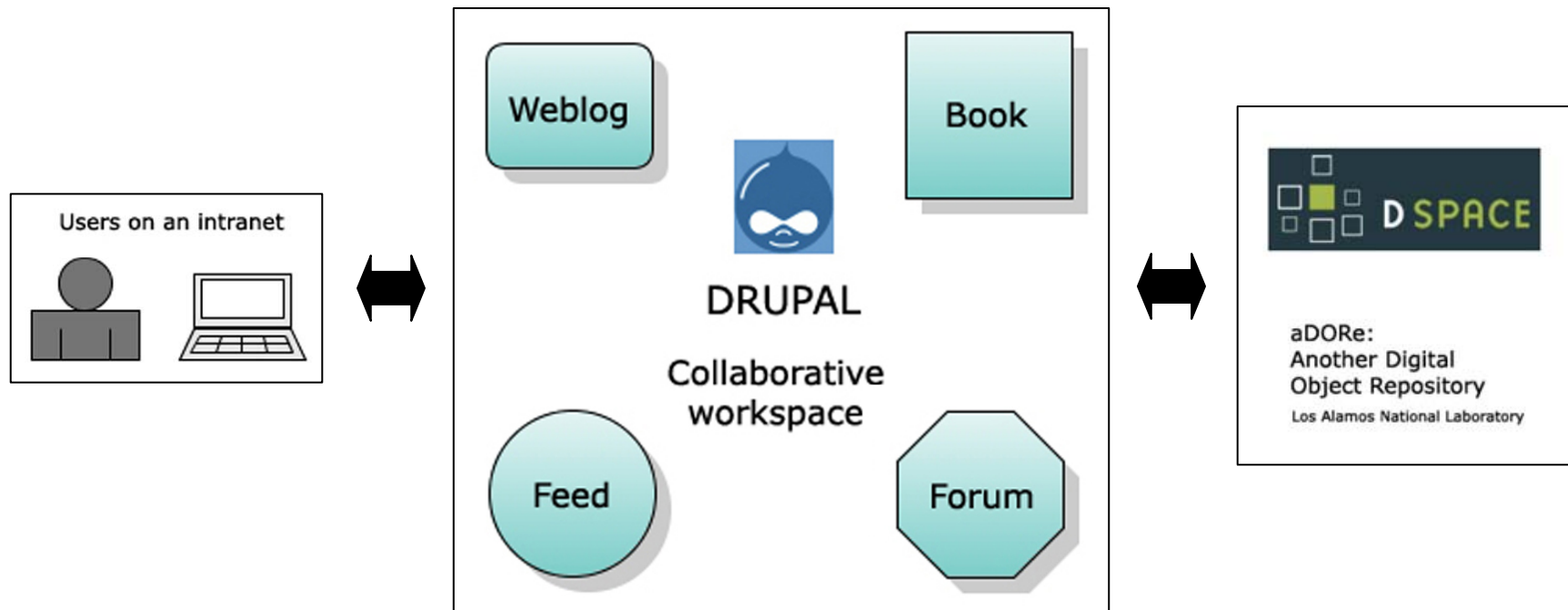
Social-technical model for pilot test 3



Pilot test 4: Aligning research with strategy

- Interdisciplinary teams of managers who participate in strategic planning
- High-level conceptual work
- Extremely important for the managers, their organizations, and the institution
- System should be designed to:
 - Restrict access to specific users on an intranet
 - Facilitate collaborative discussion and authoring
 - Track individual contributions
 - Preserve institutional memory

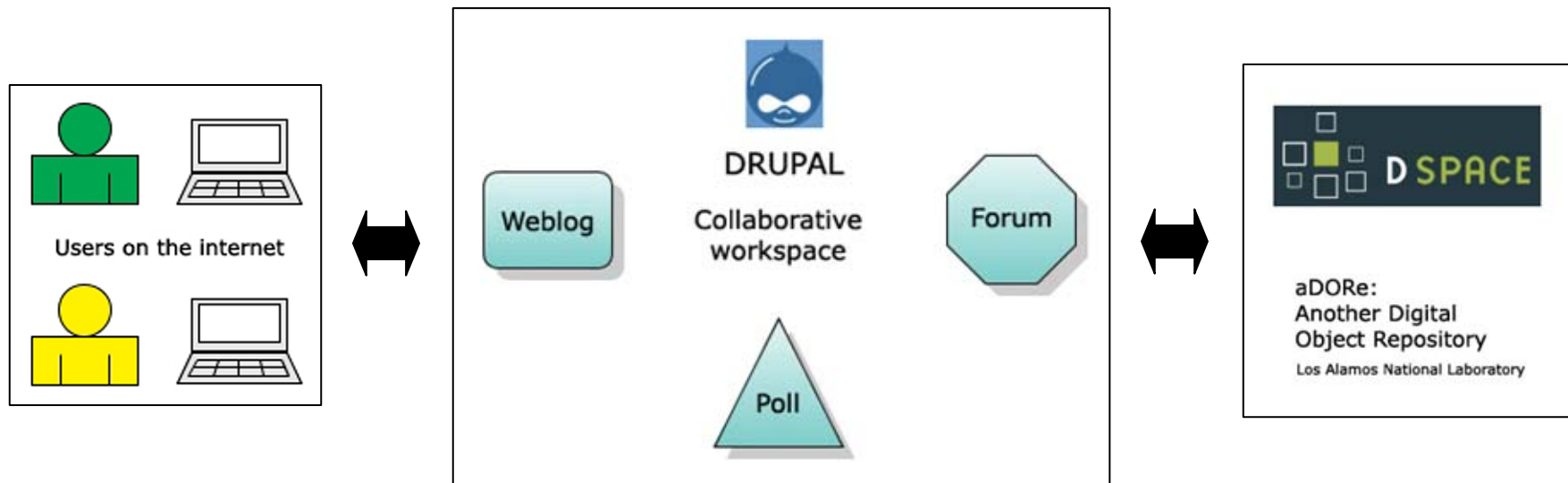
Social-technical model for pilot test 4



Pilot test 5: Building a community of interest

- Group of scientists with common concerns
- Currently communicate with each other using email and a listserv
- System should be designed to:
 - Restrict access to groups of users on the Internet
 - Support one-to-many communication of policies
 - Facilitate open-ended collaborative discussion within the group
 - Archive discussions as a contribution to institutional and social history

Social-technical model for pilot test 5



Analysis

- Flexible
 - Each pilot test involves a different configuration of components
- Inexpensive
 - Open-source software and standard hardware
- Easy to use
 - Requires small amount of time and effort to install and configure
- Results in:
 - Collaboratively-created metadata → Harmonizing metadata for a given domain
 - Collaboratively-created tags → Creating and/or revising scientific and institutional ontologies
 - Rich record of discussion → Studying content of scientific, institutional, historical, and social importance

Future work

- Improve metadata for the domain of karst science
 - University of New Mexico
- Analyze tags in order to create and/or revise scientific and institutional ontologies
 - Los Alamos National Laboratory
- Build a recommendation system for Drupal
- Integrate a visualization tool into Drupal

Acknowledgements

- The Alliance for Innovation in Science and Technology Information (AISTI)
- Scientists and librarians at the University of New Mexico
- Managers and scientists at the Los Alamos National Laboratory
- Colleagues at the Los Alamos National Laboratory Research Library