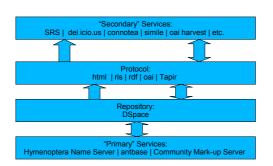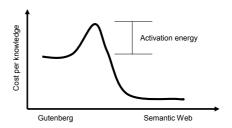# Plazi.org: Using DSpace as a Repository of Species Descriptions

PLAZI

PLAZI.ORG

- Repository for stable, persistent access to digitized taxonomic literature
- Integrated with tools for creation and enhancement
- Facilitate external services to draw data from repository
- Catalyst for the shift to digital enhanced scientific communication
- Provide tools to overcome the inherent transitional costs of including legacy publications into rapidly growing semantic Web environment

Cost per knowledge | Activation energy | Gutenberg | Semantic Web

"Secondary" Services:
SRS | del.icio.us | connotea | simile | oai harvest | etc.

Protocol:
html | ris | rdf | oai | Tapir

Repository:
DSpace

"Primary" Services:
Hymenoptera Name Server | antbase | Community Mark-up Server

Rationale

- Use DSpace to serve as "third party" repository/depository
- Provide stable, persistent locations for species descriptions through identifiers for publications and treatments
- Enhance publications by refining the granularity to improve access to content
  - species names
  - specimens (geographic names and collecting events)
  - bibliographic references
  - images
  - gene sequences
- Expose data to external services via variety of channels

Taxonomic Literature

- Formal descriptions of species
- Quasi legal documents
- Species naming governed by international codes requiring both publication and citation of prior descriptions
- Rich source of data potentially beneficial to many domains (e.g., Conservation)
- Demands Open Access
- Scope: description of 1.8M species, >100M published pages, 22,000 new descriptions/year; current scope of plazi: >12,000 species, 3,400 publications, 150 enhanced xml versions

Workflow

- A workflow has been established to involve the community to create enhanced publications
- Digital documents submitted to DSpace at Plazi.org
- Community Mark-up Server creates XML TaxonX-XML document using GoldenGate as advanced editor, and added to the respective DSpace items.
  - Identifiers are found for scientific names, bibliographic references
  - New names are entered into dedicated Name Servers (Hymenoptera Name Server and Zoobank)
  - Georeferencing of locations
- Idaho Search and Retrieval Server imports the TaxonX-XML documents into Postgres-DB, builds indices and provides search capabilities
- Data is exposed via OAI and TAPIR protocols and pushed to services such as Antweb, Ispecies, connotea, del.icio.us, etc.

Terry Catapano
Columbia University/Plazi.org
thc4@columbia.edu

Donat Agosti
American Museum of Natural History/Plazi.org
agosti@amnh.org

Guido Sautter
University of Karlsruhe
sautter@ira.uka.de