The Biodiversity Heritage Library

Preserving a Knowledge Ecology

Thomas Garnett, BHL Program Director
27 September, 2010
The cited half-life of publications in taxonomy is longer than in any other scientific discipline

-Macro-economic case for open access, Tom Moritz

-Current taxonomic literature often relies on texts and specimens > 100 years old.
Digitize the core literature on biodiversity. **Open Access**: all content can be repurposed, reused, reformatted. **Congruent**: must fit in to a dynamic knowledge ecology.

*Reptilia and Batrachia. (1885-1902) by Albert C.L.G. Günther*
BHL Content

- The Biodiversity Heritage Library has the largest collection of full-text digitized core biodiversity literature in the world.
- No other repository is even close for this class of texts.
- 31,000,000 pages – 83,000 volumes – growing. 80 terabytes today – double in two years.
Preserving a Knowledge Ecology

Biodiversity information is created, used, and maintained in a complex knowledge ecology of taxonomists, botanists, zoologists, conservationists, field researchers, professional societies, publishers, librarians and others. Making it available for future generations requires more than preserving bit streams.

Self-organization, resilience, heterogeneity, & fractal dynamics
Open Access is a form of preservation

• distribute responsibility to the wider community and thereby pool resources.

• digitize/make available the greatest amount of biodiversity texts in as open a manner as possible. This includes open to human reading but also open to algorithmic investigation by projects that will push the frontiers of biodiversity research.

• To ensure and demonstrate that the legacy of biodiversity research belongs to all of humanity and is the exclusive provenance of no one country
BHL Digital Preservation Principles

• Ensure long-term continued access to the content of the legacy biodiversity published record by eliminating any single point of failure.

• Distribute responsibility to the wider biodiversity community and thereby pool resources.

• Create responsible ownership of the content by partner projects.
BHL Digital Preservation Principles 2

- Replicate the content in different locations
- With different system architectures.
- Under different managements and different funding streams.
- Each partner nodes is self-sustaining.
- Transparent processes.
BHL Vision: Global Infrastructure

- Preservation System – multiple redundant copies of all digitized content.
- Access System – files, metadata & services needed to deliver content.

Replicate

Sync

Data Ingest

Data Ingest

Data Ingest

Data Ingest

Asia

Indian Ocean

Africa

Europe

Atlantic Ocean

North America

South America
• The Long Now Strategy
  – Institutions that are creating the BHL exist to persist through time. That’s an important part of their business. Use them.
  – Do not depend on any one institution.
  – Preservation is both a *technical* and a *social* process.
  – The future is uncertain, the technology landscape changes, people pass on. So create consortial structures that are low-overhead, flexible, and can respond quickly.