Biodiversity Heritage Library Datasets Now Openly Accessible on the Amazon Web Services Cloud

Sixty-two million pages of scientific text, images, and metadata, representing 500 years of biodiversity data will be openly accessible via the Registry of Open Data on AWS

Washington, DC, November 27, 2024 — The BHL Technical Team is thrilled to announce that Biodiversity Heritage Library (BHL) datasets will be openly accessible on the Amazon Web Services (AWS) cloud, thanks to the AWS Open Data Sponsorship Program. Moving BHL data to the cloud allows researchers globally to explore and analyze over 500 years of biodiversity data, enhancing their ability to derive scientific insights from our shared past to inform future global environmental policy.

BHL data is now hosted on AWS, and comprises over 62 million pages of scientific text from the 15th to the 21st centuries. BHL's vast collection represents an unparalleled biodiversity resource with enormous potential to be used for longitudinal studies and conservation efforts.

Key Use Cases for BHL's Data:

- **Historical Baseline Data:** Spanning centuries, BHL provides a unique record of biodiversity literature, allowing researchers to establish long-term species baselines and track changes in species distributions and abundances over time.
- **Longitudinal Studies:** The dataset supports analysis of biodiversity trends and ecosystem responses to environmental changes, offering valuable insights into historical and contemporary biodiversity dynamics.
- **Rare and Endangered Species:** BHL's historical records include species that may no longer exist or have become rare, aiding in the understanding of past biodiversity and informing current conservation efforts.
- **Taxonomic Stability:** The collection includes taxonomic descriptions and classifications from various time periods, essential for understanding species relationships.
- **Cultural and Scientific Heritage:** Beyond scientific data, BHL preserves historical texts, scientific illustrations, and annotations, enriching our understanding of past scientific practices and societal attitudes toward nature.

Transitioning to AWS services represents a transformative opportunity for BHL. As BHL currently operates outside a cloud environment, this move will optimize scalability,

accessibility, performance, and cost-efficiency. AWS's advanced computing capabilities will enable faster data processing and analysis, while AWS's cost management tools will in the long-term help reduce infrastructure costs associated with on-premise storage and upgrades.

As a vital component of the global biodata infrastructure, the demand for BHL's content and services continues to rise. By leveraging AWS services, we can address several critical needs, including:

- **Complimentary storage and access** to a comprehensive suite of AWS cloud computing tools.
- **Ensured page persistence and unique page identifiers** essential for biodiversity informatics applications.
- **Enhanced page-serving speeds in low-bandwidth regions** where biodiversity research is critical.

BHL's Data Manager, JJ Dearborn remarked, "Collaborating with AWS unlocks the full potential of BHL data. Cloud hosting not only provides an additional data safe harbor but also makes it accessible to researchers worldwide. We're excited to democratize access to BHL data and see it used to spark innovative cross-disciplinary research in biodiversity informatics."

The AWS Open Data Sponsorship Program is committed to making high-value datasets freely available to encourage innovation and collaboration. By supporting the free access to a diverse array of big datasets, AWS helps advance research and development in numerous fields.

To explore the BHL dataset and learn more about its content, visit the [AWS Data Exchange](#).

About the Biodiversity Heritage Library

The Biodiversity Heritage Library (BHL) is the world's largest open access digital library for biodiversity literature and archives. Headquartered at the Smithsonian Libraries and Archives in Washington, D.C., BHL is a global consortium that digitizes and freely shares natural history literature and archives. By providing open access to critical biodiversity knowledge, BHL addresses major research challenges and supports the global scientific community in understanding and conserving Earth's species amid climate change and extinction crises. BHL also collaborates with rights holders, volunteers, and international experts to enhance content accessibility and interoperability. Since its inception in 2006, BHL has been working towards a shared future where biodiversity knowledge is globally accessible and supports universal bioliteracy.

For more information, visit www.biodiversitylibrary.org.

###