Gigabytes, Terabytes, Petabytes.....Now What?
Mike Thuman, Ex Libris Director, Digital Preservation

Abstract
Institutions have around the world are reporting continued concerns about the long-term preservation of digital content. Given the multiple layers of complexity that make up a digital archive, the simple addition of a preservation plan is not good enough. The risk that digital content will be lost due to technical, legal, and commercial factors is increasing. The problem of data loss at the federal government level is well documented. The National Library of Medicine’s HathiTrust Digital Library needs to develop and implement a preservation plan to prevent data loss. One practical way to get started is by focusing on the implementation of a comprehensive middleware solution that can support the long-term preservation of digital content.

Knowledge Base - Format Library
A library of different formats with metadata on each format
Based on the PRONOM global library

Overall Summary/Results
Preservation planning doesn’t start from the action itself. It’s a process which includes: format identification, metadata extraction, risk identification, evaluating alternatives, running the migration action and other steps.

Preservation planning is largely about formats and risks that should be shared among institutions around the globe.

Rosetta and the Global Library is an infrastructure that allows that to take place today.

Digital preservation is still a growing issue. More research, collaboration, sharing and standardization are necessary for success.