

Google Surrogates

Problem Statement

Google provides partner institutions with digitized copies of works from various partner libraries – these digitized copies are referred to as surrogates. When Google receives a book from a library partner and recognizes it was already scanned from a different library partner, Google rejects the book. For example, Google may reject a volume from Wisconsin because it previously scanned the same volume from Indiana. Google will then make available to Wisconsin the digitized copy, or surrogate, from Indiana. HathiTrust has not yet begun ingesting surrogates into the repository because of the many complex issues associated with these materials. This issue will get more complicated as we get into settlement works that (a) include illustrative content and (b) where the publisher is or may not be the rights holder for that illustrative content. In those cases, Google will make available an un-redacted copy to the source institution and a copy with images removed to other institutions. Currently (prior to the settlement) Google is making public domain surrogates available to partner libraries now.

Background Information

Paul Soderdahl, Director, Library Information Technology, University of Iowa, and HathiTrust Strategic Advisory Board member, prepared an overview outlining the various issues surrounding surrogates; *Surrogate Problem Overview*. The overview concisely describes the storage, display and repository management issues that result from Google designated duplicates.

Teams from University of Michigan (UM) and University of California (UC) have investigated implications for HathiTrust when receiving surrogates and have prepared two white papers that provide additional information on issues associated with surrogates.

New Google Duplicate Detection and Return Procedure: Impacts on HathiTrust Bibliographic and Item Level Metadata, Jon Rothman, University of Michigan, October 27, 2009.

Google Designated Duplicates: Implications for HathiTrust End User Display, Heather Christenson, California Digital Library, February 11, 2010.

The Google surrogates is a complicated issue and many issues that need further investigation including legal issues, ownership, storage, branding, display, preservation and repository management issues. For example, do we store the surrogates in the source institution's namespace? How do we inform the end user that the original comes from Stanford and is provided as a substitute to California, Wisconsin and Michigan? What metadata do we need to store and preserve?

Working Group Charge

The HathiTrust Strategic Advisory Board charges a Google Surrogates Working Group to:

- Review this charge, recommend modifications as needed.
- Study the information previously prepared in order to further understand the issues.
- Interview stakeholders to gain additional understanding of the issues.
- Succinctly identify the issues for the SAB, the Executive Committee and the library directors.
- Specify high-level principles reflecting HathiTrust's needs and an approach in this area.
- Estimate the benefits and full cost (storage, preservation, programming time, etc.) of ingesting surrogates, as well as the consequences and opportunity costs of not ingesting them. Provide a recommendation on whether the benefits of ingesting the surrogates is worth the cost (or not; or something in between).
- Given the SAB wants to keep both the Source of the Print Volume (SOPG) and the Contributor of the Digital Item to HathiTrust (CDIH) in the metadata, make specific recommendation on where the SOPG and CDIH should display, including related branding considerations and the cost of making these changes.
- Provide recommendations for the prioritization of work.

Working Group Membership:

Heather Christenson, California Digital Library

Bernie Hurley (liaison to the SAB), University of California-Berkeley

Jon Rothman, University of Michigan

Paul Soderdahl (liaison to the SAB and consultant to the working group)

Timeline:

Members are charged through the end of June 2011.

- Charge recommended for approval by the HathiTrust Strategic Advisory Board on May 20, 2010.
- Sent to the HathiTrust Executive Committee for approval on June 8, 2010.