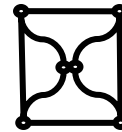

SUBSEQUENT HARVEST PLAN

MetaArchive of Southern Digital Culture Project



2005-08-15

Summary

This document describes subsequent plans for harvesting collections into the MetaArchive of Southern Digital Culture Preservation Network (hereafter referred to simply as the preservation network). Harvesting activities include both 1) preparatory creation of required LOCKSS manifests and plug-ins, 2) the activation of the harvesting process in the individual site installations. This document makes reference to the project plan and other project documents which provide more details about other aspects of the project.

Where

All project preservation sites (Emory, FSU, GA Tech, VA Tech, Louisville, and Auburn) will archive additional collections into the preservation network via harvesting activities. Such activities may also be conducted at the Library of Congress and other potential sites that may join the cooperative over time.

What

Subsequent collection archiving/harvesting activities are slated to occur from 8/29/05 through 8/25/06. This is the "subsequent harvesting phase." There are currently more than 30 collection entries in the MetaArchive conspectus database. There will likely be additional conspectus entries added in coming months, but this is the general number of collections we will likely be dealing with. This suggests a target of 5-6 collections per institution to be archived in the preservation network.

Rather than attempt to target particular collections at particular dates, each institution will endeavor to prepare one collection for harvesting each 1-2 months during the subsequent harvesting phase. Our goal will be to complete all harvesting activities (at least as regards the purposes of this project) by the end of August 2006.

In addition, FSU has volunteered to prepare a supplementary document that examines the ways that the LOCKSS manifest/plugin approach has been implemented by the various partners and identify best practices for creation of such tools. This document will identify exemplars of the various relevant approaches.

Who

Following the pattern of the first harvest plan, the respective members of the Steering Committee will be responsible for ensuring that the required activities for the subsequent content harvests take place at the six primary preservation sites. Technical staff working on the project may be the ones who actually do much of the work in practice, but the Steering Committee members are the individuals responsible for ensuring that the work gets done in time to meet the project timeline, as well as reporting out to the rest of the project on collection harvesting accomplishments. Specifically, the

Steering Committee members will: 1) provide updates on both conference calls and the listserv about individual collections as they become available for harvesting and as respective institutions in fact harvest them, and 2) update entries in the conspectus reflecting their changing status as they become harvestable.

Harvest Prioritization Decisions and Quotas

Decisions by an individual institution to add new collections to the MetaArchive should take into account the previously established ranking algorithm and the amount of storage space that each new collection represents. The question of quota arrangements will be discussed in subsequent project conference calls immediately, and project meetings (starting with the February 2006 meeting), to see if they should be renegotiated using a more complex system reflecting both collection value and other factors. The group has recognized that the question of institutional quotas is complicated, and should be carefully considered in the overall context of cooperative agreement analysis.

For the next few months, each collection decision should be vetted to the group, and unless there are major concerns about the size of a new collection or particular impacts of adding the new collection to the MetaArchive, we should try to be inclusive and allow new collections into the preservation network. The Steering Committee will carefully monitor the status of the network in the intervening months to ensure that it grows at a reasonable rate, neither too fast or too slowly.

Next Steps Targets

Each preservation site should endeavor to prepare 2 additional collections (beyond the initial collection identified in the original harvest plan) for harvesting by the end of calendar year 2005. We are slightly behind on the harvesting activities and need to try to make up some time on the schedule.

Each institution will continue to prepare a new collection for archiving/harvesting every 1-2 months during 2006, up until the point when the institution no longer has further collections under consideration.

Notes on this document

This document was completed at the third MetaArchive All-Project Meeting held at Emory University on Monday 2005-08-15, by the Steering Committee and additional attending project participants. The document was last updated by Martin Halbert on 2005-08-15.