Chronicles Committee Meeting
Minutes
04/25/2012

Attendees

1. Kenning Arlitsch (Univ. of Utah)
2. Bill Donovan (Boston College)
3. Mike Furlough (Penn State)
4. Cathy Hartman (UNT)
5. David Minor (Chronopolis)
6. Mary Molinaro (Univ. of Kentucky)
7. Mark Phillips (UNT)
8. Matt Schultz (Educopia)
9. Katherine Skinner (Educopia)
10. Ryan Speer (VA Tech)

Agenda & Minutes

Preview of draft outline for the Guidelines

1. Matt provided a brief overview of the draft outline for the Guidelines to Digital Newspaper Preservation Readiness
   a. This document has been shared with the Advisory Group, and is awaiting further feedback for improvements on their next scheduled call – Friday, May 18th (1pm-2pm ET)
   b. Chronicles Committee members were advised to read through the outline thoroughly and prepare feedback and/or additional questions for discussing on the next Chronicles Committee call scheduled for Wednesday, May 30th (2pm-3pm ET)
   c. The Guidelines will help form the basis for more individualized preservation readiness plans to assist each of the content curators (Chronicles Committee sites) with preparing their digital newspaper collections for ingest into the three DDP sites in late 2012/early 2013

Research Update

2. Matt provided a brief update on progress with project research
   a. Matt and Ryan Speer (VA Tech) conducted a follow-up interview with Boston College on their use of Veridian’s hosted service for their digital newspaper collections
      i. Notes from this interview are forthcoming – Matt is following up w/Frederick Zarndt (DL Consulting) on additional Veridian features of interest to the project – stay tuned!
   b. Matt and Ryan are completing work over the next couple of weeks on the Standards Gap Analysis – see here: http://tinyurl.com/7jkwtjz
i. This resource will help to make determinations on the essential to optimal standards, guidelines, specifications, tools, etc as they rightly apply

Versioning Discussion

1. Matt opened discussion to the topic of “versioning” – both in regard to changes with digital newspaper content & how these can be responsibly tracked and managed by the three DDP sites (MetaArchive, UNT & Chronopolis)
   a. Matt started by referring to the versioning document shared just prior to the call – Digital Object Storage and Versioning in the Stanford Digital Repository (http://tinyurl.com/18r)
      i. This is a great overview and case study of versioning approaches and utilities
   b. Revisited the Chronicles Survey – no active change factors were reported at that time by content curators with respect either content, OCR, or metadata
   c. Matt requested some new or additional details on known content changes
      i. Kenning Arlitsch (Univ. of Utah) brought forward their need to update 1/3 of their newspaper titles to make them conformant to METS-ALTO
         a. Kenning reported that previous versions would likely be retained for some specified period of time in a dark archive fashion
      ii. Bill Donovan (Boston College) offered the case study of user-corrected OCR – question of how and when to capture such changes and update METS-ALTO records
         a. Matt mentioned previous conversations with Frederick Zarndt (DL Consulting) on the Advisory Board – he mentioned an upcoming METS-ALTO Editorial Board Meeting on just this topic – stay tuned for more details
      iii. Mark Phillips (UNT) mentioned the occasional need to update metadata description or subject headings in the UNT system
         a. Each update is registered as new object
      iv. Mark also mentioned the occasional need to regenerate derivatives from master objects – just an example of a change factor – no real version tracking applied
      v. David Minor (Chronopolis) referred back to the Guidelines outline and urged that we make some distinctions between versioning concerns at an access/discovery level and versioning concerns within a repository environment itself
      vi. Bill Donovan inquired how format migrations may need to be handled by versioning protocols – BC is migrating away from
TIFFs as master objects –using JPEG2000 instead—to save space

a. Mark Phillips shared experiences from NDNP meetings that were wrestling with similar resource-driven decisions—some voices advocating strong preservation support for negative masters—this may not be realistic or achievable

b. This led to discussion about economical approaches to regenerating derivatives

   i. Mary Molinaro (UKY) shared their decision to generate TIFFs from their born-digital publisher PDFs

   ii. Mark Phillips echoed this approach—UNT is generating high-res JPEGs from publisher PDFs

d. Matt steered the conversation toward the three DDP sites (MetaArchive, UNT & Chronopolis)—what are the current versioning support or approaches?

   i. Mark Phillips (UNT) explained that any changes to a digital object in CODA result in a newly stored object that is held together by identifiers—right now this is not a resource issue because there are not a lot of frequent changes

   ii. David Minor (Chronopolis) echoed Mark’s approach of storing all objects—with the exception of data sets Chronopolis does not see a lot of active changes—still manageable from a resource perspective

   iii. Matt Schultz (MetaArchive)—not a lot of content change occurring in MetaArchive either, w/the exception of manifest pages (content inventories)—have moved to a policy of change notations on these manifest pages when content or metadata is being updated by a contributor

2. The versioning conversation closed out by advocating that the Guidelines aim to document the various change factors related to digital newspapers, briefly review versioning strategies, but avoid recommending untested or unproven approaches