

Subject: DPLA Ingest Update

March 11, 2014

Hello all,

I'm writing to give you an update on our progress with the new ingestion system.

Our initial hope was that the system would be ready for new harvests in January, but we underestimated how complex this technology would need to be and the time we would need to implement it properly. We created a revised schedule in February that we hoped would be accurate. We emailed all of you at some point in the last month to say we would be getting started. But then some unforeseen issues arose. These are typical kinds of bugs that one has to deal with when developing a new system, but we did not plan adequately for them. However, we feel that the right thing to do is address them completely and responsibly now, despite the further delays in the schedule. The point of this redesign was to fix the earlier problems incurred by developing a system too quickly, with too many temporary fixes.

Our new system will be comprised of 4 main workflow steps: harvest of records from the provider, mapping of those records to DPLA MAP, "enriching" those records (any enhancement or correction of records, such as when we remove extra white spaces or ending punctuation, or when we enrich geographic terms with standard names and coordinates), and finally performing QA. Once those steps are completed, the record will be moved into the live dataset accessible through the API and portal.

Over the next two to three weeks we will be finishing up work on the QA environment, updating some indexing issues to allow for the new records to appear in our API and portal, and fine-tuning the enrichment process. The first few providers we are working with will be our guinea pigs in this system, so by the time we are ready to start working with your records we hope to have many of the kinks worked out.

We are again sincerely sorry for the delays. Please let me know if you have follow-up questions or would like to schedule a call to talk about any issues. Mark, Amy, and I are all available to work on this with you and provide as much information as you need.

Best,

--

Gretchen Gueguen
Data Services Coordinator
Digital Public Library of America
<http://dp.la>