OCLC Meeting Notes

Taylor Surface/CDM Staff (Seattle)

I had the opportunity to spend a few hours talking a bit more to Taylor about the OCLC Gateway and OCLC’s interest in potentially pursuing a partnership with the DPLAOhio effort. I’m not sure that I learned anything new during the meeting, but I do have a better idea of how we could move forward if we were interesting in exploring this as an option, as well as a few other potential thoughts for discussion.

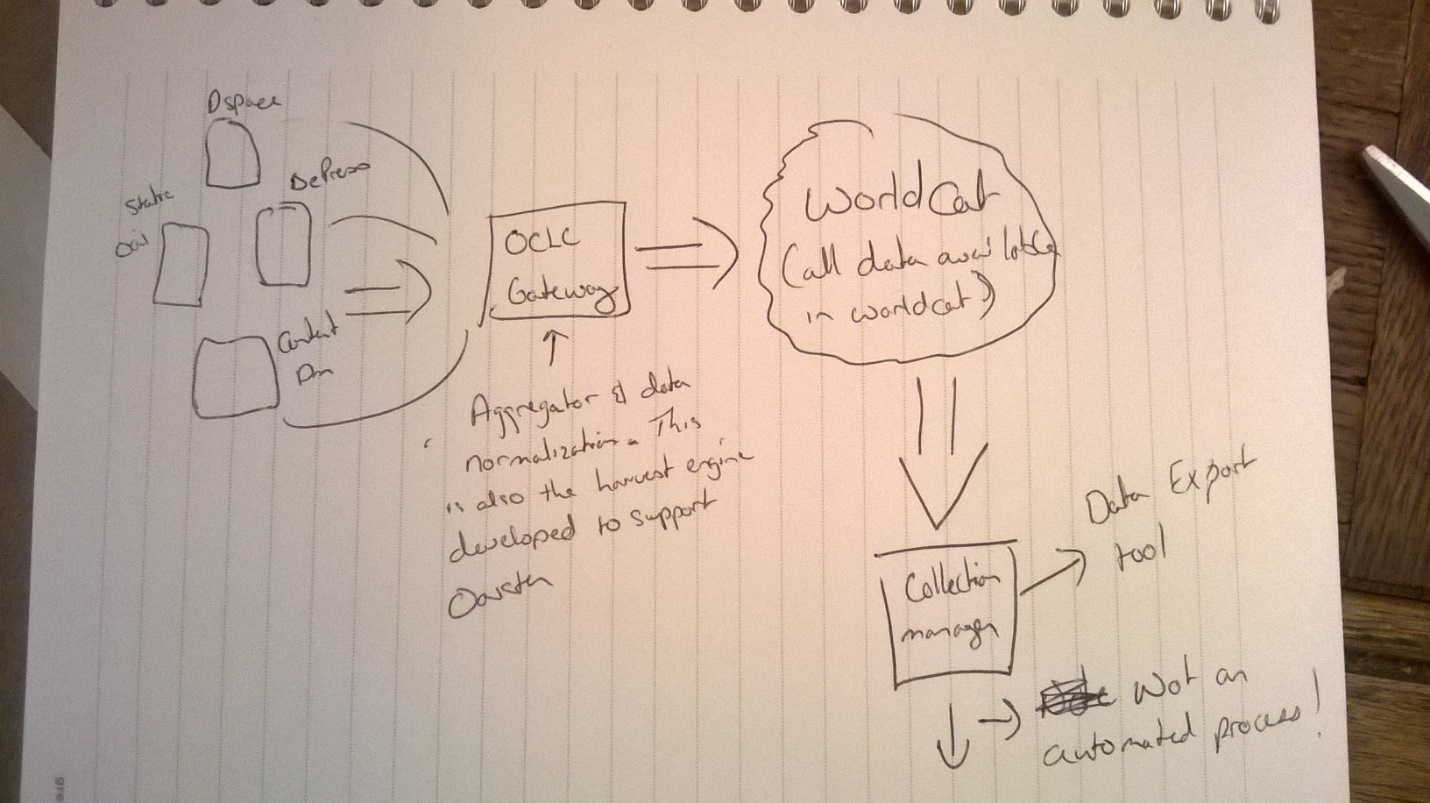
**Background**

A little background on why OCLC is interested in this. OCLC’s gateway export infrastructure was developed when they took over OAISTER from the University of Michigan. The gateway tool was developed as a way to simplify metadata harvest, provide some simplified metadata normalization tools, and encourage organizations to register their digital collections with OCLC. What OCLC got out of this project was more content to add to WorldCat. All data harvested through the gateway tool is loaded into WorldCat. By loading them into WorldCat, these records become available through OCLC’s other management tools, including metadata export using their collection manager tool.

So why is OCLC interested in this now? Well, this isn’t altruistic. OCLC is looking to sell a service and is looking to make sure that they are not left out of the digital collections space. The development of Hydra in a Box and the push by DPLA to have a presence in each state raises the very real possibility DPLA, and note WorldCat, would become the preferred location for exposing digital collections in aggregate. Given that, they are looking to develop a service offering that they could take to other states looking to join DPLA by providing the infrastructure necessary to handle the data aggregation, allowing the hubs to focus on education, outreach, recruitment, and sustainability.

**Infrastructure model**

OCLC’s aggregation infrastructure currently looks like the following:



This is the infrastructure as it exists today. OCLC manages something like 2100 collections and approximately 40 million items through its gateway. The gateway service provides user the ability to select and profile their records – offers some limited normalization functionality, and passes information into WorldCat. As envisioned now, the hub would then gather all the metadata together on a particular schedule, and submit that data for ingest into DPLA.

So what would need to be done? The Gateway export tool is pretty limited in what it does – but there is the potential to create a DPLA profile that would enable OCLC and a partner to define a set of metadata normalizations so that data could be enhanced, with enhancements potentially being passed back to the local user. This enhancement work would need to be identified and done. Likewise, the collection manager – DPLA envisions a service that isn’t a data dump but more of an interactive harvest. This would need to be enhanced to support that type of use case.

Of course, I learned some interesting things – obviously, in order to use the harvester and infrastructure, everyone would need to be an OCLC member and have an OCLC symbol. Is this a barrier? I’m not sure. One area of discussion that we had was a configuration that looked like the Orbis-Cascade catalog in Oregon. There, the consortia had its own symbol, and consortia records would have that symbol attached. If individual membership was a barrier, then something like that might be possible. Of course, I also learned that this would be a subscription service. If there was interest in this as a potential technical solution, I’d recommend the governance group having early conversations about what the subscription cost might look like, and what type of subscription model DPLAOhio would want to enter into. There are a couple different options, ranging from individual organizations subscribing to the service, to the hub contracting for all members. Would this be cost prohibitive? How does this compare to developing and running a local solution? All questions that would need to be sorted out.

**Pitfalls**

There are a few – but the biggest is the licensing issue. Probably the most important thing to come out of this meeting is that Taylor will take this question to OCLC’s legal council today. Unless OCLC was willing to allow data to be contributed to DPLA as CC0, it’s a non-starter. This has been relayed as part of the conversations with Taylor and DPLA folks, it’s what Emily told this group during the symposia…this is the price to play if you want to be in the DPLA. The answer to this question though is important regardless of the option that we might utilize. Unless OCLC makes this allowance, DPLAOhio couldn’t use as part of our aggregation any metadata that was obtained via OCLC’s gateway harvester or pulled out of WorldCat. So, if we had members using the gateway export themselves – we’d still have to require them to profile their data into the DPLA aggregator in order to contribute their data to DPLA. Without a clear statement from OCLC, the metadata couldn’t be used due to the existing license. However, as I say, this is something Taylor recognizes and is taking up with their legal department and will be hoping to have with DPLAOhio and DPLA in general.

**Where could we go from here**

So, after talking to Taylor, I asked him how he envisioned this moving forward. The reality is that DPLAOhio doesn’t have a hub at this point, so there isn’t an organization that can take on a contract for this project, and, honestly, we can’t go forward discussing this in a proposal to DPLA until the licensing issues are sorted out.

So – here’s what we came up with. They are interested in planning out some work, the idea being that they would like to put developers on enhancing the gateway project by this summer. Also, knowing that DPLAOhio couldn’t move forward that quickly, we discussed a focus pilot. Working just with OCLC members with digital content, OCLC, DPLAOhio and DPLA would discuss and test a set of workflows that could facilitate this process for Ohio organizations. At the same time, governance or the steering committee could begin to engage OCLC around potential subscription scenarios to determine potential cost and sustainability options.

**Why might we want to do this**

It sounds like OCLC is motivated. I think that they are trying to find a way to stay relevant here – and this offers them an opportunity to potentially develop a new long-term subscription service offering. I’ve often found that when OCLC is motivated, good things happen. Obviously, there are a lot of unknowns, but a limited pilot would allow dplaohio to look at another potential option, and begin profiling metadata (which we need to do anyway).

However, the technology is likely the smallest reason to potentially participate in a pilot. As of right now, finding a technical home for this project is one of the hang-ups. We have some potential technical solutions available to us – but these will involve technical staff and infrastructure. This is going to have a high upfront cost, and will ask one institution to shoulder infrastructure for the state. I think we have a lot of folks excited about doing this – I don’t think we have anyone excited about hosting the infrastructure. Does it make sense from a resources and sustainability perspective to focus the hub on education and outreach, and partner on technology? Would that make it easier to find a central hub?

**Why wouldn’t we want to do this**

OCLC doesn’t always have a great record of moving things from pilot to production. While they seem motivated, we may be asking them to do a lot of unfunded work – and even after the pilot, potentially not go with that option. My guess is that any pilot likely would be just using the current infrastructure, with potential enhancements identified but not implemented unless this went beyond this phase, which may be a harder sell when making a bid to DPLA. Of course, there is also the license issue. Unless this can be resolved, DPLA won’t touch this information. By not pushing forward with a pilot and continuing to push for resolution on the licensing issue, we may be able to influence OCLC’s thinking on this issue.

I’m happy to answer questions tomorrow.