DCAT Meeting September 2017

Date & Time

- September 12th **15:00** UTC/GMT - **11:00** ET

Dial-in

We will use the international conference call dial-in. Please follow directions below.

- U.S.A/Canada toll free: 866-740-1260, participant code: 2257295
- International toll free: http://www.readytalk.com/intl
    - Use the above link and input 2257295 and the country you are calling from to get your country's toll-free dial in #
    - Once on the call, enter participant code 2257295

Agenda

The September 12th meeting will focus on a discussion of DSpace Statistics:

- Should DSpace maintain its own statistics repository or should it only focus on communicating with existing analytics vendors.
- If a statistics store is not maintained within DSpace, how many analytics vendors would need to be supported?

Preparing for the call

If you can join the call, or are willing to comment on the topics submitted via the meeting page, please add your name, institution, and repository URL to the Call Attendees section below

Meeting notes

Discussion about DSpace statistics, and how bots are or can be filtered out. The DSpace community needs to be on the same page in terms of using third party services and who maintains bot lists that are reliable. It is difficult to maintain up to date lists, and no one institution can be responsible for updating a list on a regular basis. Could we somehow automate the inclusion of an up to date bot list during software upgrades?

Jose mentioned that there is a group working to create a list of robots and spiders to be blocked for DSpace, and that the community could look at the COUNTER 5 platform in the way it counts downloads and views. He mentioned that everyone should be wary of services like Google Analytics since it is not open source, and it can be difficult to understand how they derive their stats numbers. Jose explained his preference for using internal DSpace stats numbers versus those of 3rd party services like Google Analytics and Piwick. Instead of using these vendors to provide numbers, his group has built a plugin that exports statistics from DSpace into the OpenAIRE portal to manage the DSpace logs and data. (Link to a webinar with more information has been shared in the meeting page.)

Discussion about other vendors that repository managers have used: any others besides Google Analytics and Piwick? Some institutions built custom solutions with AW Stats and Elastic Search, but those are no longer supported. Some of these services cost money, while many institutions in the DSpace community will need to rely on free and open source solutions.

Discussion about how long institutions keep their web logs, since while they can get quite large, they provide monthly usage statistics that administrators rely on. Some institutions use these logs, while others use SOLR to report the same numbers.

The University of Kansas' ScholarWorks repository has implemented a custom solution to address the lack of aggregated statistics available in DSpace and to provide faculty members with more granular data about their item usage. The code provides aggregated stats that display the most popular items within a community, for a specific author, within a date range, etc. The code has not yet been released open source, but it will be in the future.

Ultimately the discussion about whether we should embrace vendor or 3rd party solutions or work to improve DSpace internal statistics is one that should happen on a larger scale within the Open Repositories community as a whole. There is no easy solution, and we will need to look carefully at the tools we use and evaluate them for sustainability moving forward. We also must be cognizant of our international community and of everyone's different statistics reporting needs and responsibilities.

Maureen asked call attendees to share links to resources and documentation that was discussed. It would be helpful to hear what different institutions' needs are in terms of statistics so we could create a comprehensive list of needs for the community and develop some aspirational goals around stats.

Next month's call will be coordinated by Marianne, and will be about strategies for determining the levels of support in the DSpace file format registry and possible implications for preservation.

Call Attendees

- Maureen Walsh (Ohio State University)
- Bram Luyten (Atmire)
- Ignace Deroost (Atmire)
- Claudio Cortese (4Science)
- Jose Carvalho (University of Minho)
- Marianne Reed (University of Kansas)
- Valerie Collins (University of Minnesota)
- Anne Lawrence (Virginia Tech)

- Suzanne Chase (Georgetown University)
- Iryna Kuchma (EIFL)
- Terrence W Brady (Georgetown University)
- Felicity Dykas (University of Missouri)
- Pauline Ward (University of Edinburgh)