2021-01-12 Statusmeeting

- Update on NAS latest tests and developments (if any)
- Status of the production sites
- Next meetings
- Any other business?

Agenda for the joint NetarchiveSuite tele-conference 2021-01-12, 13:00-14:00.

Participants

- BNF: Sara, Aurianne
- ONB: Andreas
- KB/DK - Copenhagen: Tue, Stephen, Anders
- KB/DK - Aarhus: Kristian, Colin
- BNE: Alicia, Juan Carlos?
- KB/Sweden: Pär, Peter

Update on NAS latest tests and developments (if any)

Status of the production sites

Netarkivet

- Broad crawl.
  - Looking into slow harvest.
  - Workshop on test system to troubleshoot better
  - We suspect throttling issues/slow response times from hosting companies
  - Would like to end it soon
- Facebook harvesting
  - mbasic.facebook.com looks interesting.
    - We are investigating if we can harvest this with Heritrix.
    - In ArchiveIT we can get some comments - we are investigating if we can get all comments.
  - Just started ArchiveIT harvest of 1700+ selected Facebook profiles.
- SolrWayback
  - Meeting with legal department today. Will probably be ok to put in production
  - 2 IIPC blogposts (UX and technical) coming up
  - 3 webinars coming up (1 technical webinar for IIPC members only and 2 for all interested)
- Netarkivet.dk to transfer to kb.dk/....
- .....
Corona event harvest

We are still working in this event crawl.

We have more than 5,000 seeds selected and its size is 62 TB which is more than our broad crawl.

Broad Crawl

We made a broad crawl of the .gal domain, a regional domain.

It lasted 24 hours and harvested more than 5,000 web pages. 214 GB.

Twitter harvesting

We are having problems to harvest Twitter since mid-December. We are receiving a 400 error but we don’t know how to solve it.

BCWeb version 6.1

We are finally working on the 6.1 BCWeb version

KB-Sweden

Next meetings

- February 9th
- March 9th
- April 6th
- May 4th
- June 8th
- July 6th
- September 7th
- October 5th
- November 2nd
- December 14th
- January 11th, 2022

Any other business?