2023-06-27 Statusmeeting

- Presentation about ONB integration of Browsertrix
- Proposal for NAS evolutions
- Discussion about topics for a future NAS workshop
- Status of the production sites
- Next meetings
- Any other business?

Agenda for the joint NetarchiveSuite tele-conference 2023-06-27, 13:00-14:00.

Participants
- BNF: Sara, Clara, Auriane, Nola
- ONB: Andreas, Antares
- KB/DK - Copenhagen: Anders, Thomas, Stephen, Tue
- KB/DK - Aarhus: Colin
- BNE: José, Miguel, Eva
- KB/Sweden: Peter, Pär

Presentation about ONB integration of Browsertrix

20230627_browsertrix.pptx

Proposal for NAS evolutions

[NAS-2886 - Getting issue details...]

[STATUS]

[NAS-2887 - Getting issue details...]

[STATUS]

Discussion about topics for a future NAS workshop

Brainstorming: https://docs.google.com/document/d/1eO2cgcfwQ-7BOxrjzSApjvTEgXc-VICPEgmqyVOric/edit?usp=sharing

2019 Workshop: 2019 NAS workshop

Status of the production sites

Netarkivet

- 2. Broadcrawl just finished.
- Pyweb - experiments: making new CDX-index from WARC-files and more. Awaiting approval to do it on larger scale
- 2 data delivery projects: LLM (40/15TB) and soccer!
- JWAT - final testing being done
- Focusing on Paywall content. A dashboard with visual feedback of access/no access for all our IP’s have been made using some clever text-comparison.

BnF
Last month we have been working in two election events simultaneously. The local and regional elections that were succeed on May 28th, and national elections will take place on July 23th. Both of them are supporting by more than 40 half-time web curators from different regional libraries, they are coordinated by National Library of Spain. So far, we have collected more than 3,500 webpages for local and regional elections and more than 500 in national elections.

Currently, collecting Twitter's accounts is not a problem, we are collecting more than 2,000 daily. The 429 error has disappeared and quality of harvesting is pretty good.

This week we carry out a special collecting for the “LGBTI proud” as we do every year.

The Spanish Web Archive has launched a new search engine allows users to find the webpages has been collected and easier access to Openwayback: https://www.bne.es/es/coleccion/archivo-web-espanola/buscar

Our IT has working in updated NAS from 5.4.2 to 7.4. The forecasting is finish in September.

Next meetings

- September 5th
- October 3rd
- November 7th
- December 5th
- January 9th 2024

Any other business?