2023-10-03 Statusmeeting

- Lastest 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 2023-10-03, 13:00-14:00.

Participants

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

Lastest NAS evolutions

Lastest release NetarchiveSuite 7.5: https://sbforge.org/display/NAS/NetarchiveSuite+7.x+Release+Notes

Discussion about topics for a future NAS workshop

Brainstorming: https://docs.google.com/document/d/1eO2cgcfwQ-7BOxjr2SApjpjTEgXc-VICPEgmqyVOriic/edit?usp=sharing

2019 NAS workshop

Status of the production sites

Netarkivet

- 3rd. Broadcrawl 2023- step 2 - Done - 592 Jobs
- 4th Broadcrawl 2023- step 1 will be launched this week
- Upgrading to NAS 7.5
- Testing on site installation of Browsertrix Cloud - will intensify the coming months,
- Still focusing on Paywall content.
- IA 1996-1999 flies of .dk-crawls ingested
- Proposals for IIPC WAC in Paris 2024 sent
  - Unlocking Access: Navigating Paywalls and Ensuring Quality in Web Crawling (Behind Paywall Websites - Crawl, QA & More)
  - Run Your Own Full Stack SolrWayback (2024
  - Towards A Formal Registry Of Web Archives For Persistent And Sustainable Identification
  - Browser-Based Crawling For All: Introduction to Quality Assurance with Browsertrix Cloud
  - Youtube-extractor in Heritrix who’s using this..

BnF

Firstly, our 2023 broad crawl will be launched within 3 weeks. We were obliged to modify our schedule slightly due to the Skyblog harvest. Our test broad crawl is about to end and the next step will be the estimation of the budget per domain.

The Skyblogs harvest continues and should end in a little less than 2 weeks. We have around 200 jobs left to crawl and the size should be 38 TB.

A virtual guided tour on webcomics is currently in preparation. Different themes are covered such as the history of webcomics, the comics economic players and the tools used to read comics online. It will be published in January 2024.

ONB

BNE

KB-Sweden
Next meetings

• November 7th
• December 5th
• January 9th 2024

Any other business?