

# 2021-12-14 Statusmeeting

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

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

## Participants

- BNF: Clara, Sara, Auriane
- ONB: Andreas
- KB/DK - Copenhagen: Anders, Thomas, Stephen, Tue
- KB/DK - Aarhus: Colin
- BNE: José, Alicia, Miguel
- KB/Sweden: Peter, Jonas

## Update on NAS latest tests and developments

### Status of the production sites

#### Netarkivet

- NAS and Bitmagasinet. We are having quite some difficulties getting everything perfect, but we are working hard to iron out any issues. Seems like it's mainly Hadoop-issues
- In the future how do we test NAS-releases as a community?
- We are planning to test RHEL 8 - but we need a stable test-system first, so we can see implications from upgrading. Any feedback from others who have updated would be appreciated
- We have a contact at YouTube/Google and they have asked us for IP- ranges, useragent and motivations - so there is hope to harvest YouTube in the future like BnF
- We are awaiting that the IIPC steering committee decides to proceed (or not) with [User-Friendly High Fidelity Browser-Based Crawling for All - IIPC Tools Portfolio Funding Proposal 2022-2023](#)
- Broad harvest is postponed until January/we have a working system
- We still do selective and special harvests.

#### BnF

First of all, we are pleased to welcome, in our team, Kevin Locoh-Donou, data engineer for the LIFRANUM research project, for a period of 9 months.

On the occasion of "Fantastic Futures" the 3rd international conference on Artificial Intelligence (AI) for libraries, archives and museums which took place at the BnF on December 10th 2021, the digital legal deposit service highlights its AI websites collection:

- A slide show presents about 30 captures from our collections for which we have obtained reproduction rights: [https://www.bnf.fr/sites/default/files/2021-12/L%27intelligence\\_artificielle\\_sur\\_le\\_web.pdf](https://www.bnf.fr/sites/default/files/2021-12/L%27intelligence_artificielle_sur_le_web.pdf)
- For the first time, the BnF invites the public to take part in the selection of French websites, concerning artificial intelligence in several areas: scientific and technical applications of AI, art, literature, economy, society, law, ethical issues. It is possible to make suggestions at this address: <https://www.bnf.fr/fr/collecte-du-web-consacree-lintelligence-artificielle>
- We also participated in the Fantastic Futures poster session with a poster presenting our Artificial Intelligence collection and referring to the harvest participation form via a QR code.

Finally, our 2021 broad crawl ended on November 14th and lasted a little less than 5 weeks. 2.5 billion URLs were crawled for a total of 114TB.

#### ONB

- Working on strategy working groups for our vision 2035 in which webarchiving should play a bigger role.
- For reasons we started again the Corona-Crawl
- Updated Frontend <https://webarchiv.onb.ac.at> from Bootstrap 3 (jquery 1.x) to Bootstrap 4 (jquery 3.6)
- We ingested the first 20 TB (20%) into the long time Preservation system

## BNE

- Ending the broad crawl of the .gal domain (regional domain from Galicia) 2.500 domains and 315 GB of information
- We are studying the creation a new comic collection to harvest the webcomics and comics on the internet, an all this short-lived production in free access.
- Updating of Nas 7.2 is ready in Preproduction. We are waiting for a new powerful hardware on Janaury to carry it out

## KB-Sweden

## Next meetings

- January 11th, 2022

## Any other business?

€€€€€€€€