

2020-10-06 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 2020-10-06, 13:00-14:00.

Participants

- BNF: Clara, Sara
- ONB: Andreas
- KB/DK - Copenhagen: Tue, Stephen, Anders
- KB/DK - Aarhus: Kristian, Colin
- BNE: Alicia, José Carlos, María
- KB/Sweden: Pär, Peter

Join from PC, Mac, Linux, iOS or Android:

<https://kjdk.zoom.us/j/104443571>

Or an H.323/SIP room system:

H.323: 109.105.112.236
Meeting ID: 104 443 571

SIP: 104443571@109.105.112.236

Or Skype for Business (Lync):

<https://kjdk.zoom.us/skype/104443571>

Or Telephone:

Denmark: +45 89 88 37 88 or +45 32 71 31 57
United Kingdom: +44 203 051 2874 or +44 203 481 5237 or +44 203 966 3809 or +44 131 460 1196
Finland: +358 9 4245 1488 or +358 3 4109 2129
Sweden: +46 850 539 728 or +46 8 4468 2488
Norway: +47 7349 4877 or +47 2396 0588
US: +1 669 900 6833 or +1 646 558 8656
Meeting ID: 104 443 571

International numbers available: <https://zoom.us/u/acRu0MV3xJ>

You can join a meeting by using apps from a pc, a tablet or a smartphone, but you can also use the browser based version (it works with newer versions of Chrome or Firefox)

Please click OK if you see the system dialog.

Launching...

If nothing prompts from browser, [download & run Zoom](#).

If you cannot download or run the application, [click here](#) to help your browser.

Update on NAS latest tests and developments

Any feedback on NAS 6.0 ?

Status of the production sites

Netarkivet

Broad crawl

Step 2 is proceeding in a great fashion.

Event crawl

We decided to continue with the event crawl on Corona in Denmark but with lower frequency and. 0-hop sites reduced greatly, and with minimal curatorial activity.

Alexandre, trainee

Arrived and is up and running, working remotely from Copenhagen with the rest of the team. We are almost done with the intro-program and are looking into what will give most value to Alexandre, Netarkivet and also BnF.

IT-University in Copenhagen:

The collaboration with the IT-University in Copenhagen is moving forward.

Youtube

We have experimented with getting embedded video-content and so far the results are great (except WARC-validation is not valid with re-visits)

WARC-file-validation

We are working on finalizing a workflow from Webrecorder/[Conifer.org](https://conifer.org) to Netarkivet. To be able to validate WARC-files correctly is a big part of getting the right level of preservation (we use JWAT for this). But it's a bit complicated – see for instance this OPF blog by Remco van Veenendaal from Holland: <https://openpreservation.org/blogs/warc-validation-tool-experiences/>

(How) are you validating WARC-files? And what is the future on this?

BnF

Our annual broad crawl will be lunched this tuesday, 6th of October. This will be our first broad crawl with the new NAS version including the official Heritrix 3 IIPC and we expect a better efficiency of the crawling. We have reduced to 2000 the maximal number of URLs per domain (instead of 2500 last year) and we expect to harvest between 110 and 115 TB.

Next week, we'll put in production a new version of our public GUI "Archives de l'internet" making available for the readers the video channels harvested in July. These channels cover the topics of covid19 outbreak and French local elections. The new version includes also a mechanism allowing to display Instagram account pages in our web archive and to browse the posts using Picuki directly from the Instagram page. Finally, this new version gives access to 3 titles of paid online newspapers, harvested with authentication.

ONB

BNE

Even Crawl

We are still working in Coronavirus collection. Now we are focused on audio and video content.

Serials Broad Crawl

At the beginning of the year we launched a broad crawl of open-access electronic serials and now we are focused in the quality assurance of this collection

BCWeb

We are having some problems with the old version of BCWeb, so we hope this will speed up the installation of the new version that we have available, 6.1

IT Department

The new head of the IT department has just arrived.

KB-Sweden

Next meetings

- November 3, 2020
- December 8, 2020
- January 5, 2021

Any other business?

€€€€€€€€