

# 2020-05-05 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 2020-05-05, 13:00-14:00.

## Participants

- BNF: Clara, Sara, Géraldine
- ONB: Andreas
- KB/DK - Copenhagen: Tue, Stephen, Anders
- KB/DK - Aarhus: Sabine, Kristian, Colin
- BNE: Alicia, Nuria
- 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](mailto: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

Feedback on latest developments and tests.

BnF PR on H3 have been accepted:

- Add support for FTPS: <https://github.com/internetarchive/heritrix3/pull/320>
- Extend the HTML extractor to extract data- prefixed attribut: <https://github.com/internetarchive/heritrix3/pull/323>
- Fix the crawl status in the CrawlSummary report on H3 console: <https://github.com/internetarchive/heritrix3/pull/326>

## Status of the production sites

## Netarkivet

- Preparation for implementation of SolR wayback for external users are nearly finished. We had to clear security issues and are now waiting for the final go from our legal experts
- We have send a recommendation for decision to our directors: We want to implement Warc-files created by crawling with [webrecorder.io](https://webrecorder.io) into our preservation system
- Our broad crawl is ongoing, some of us are busy with the job follow up
- We are negotiating with our IT department: we want them to allot time to implement the new Heritrix release, which hopefully – among others – will solve our problems with harvesting “lazy load”
- The event crawl "Coronavirus in Denmark" is ongoing, of course. We get help from people outside the Netarchive Team, among others, colleagues who are not able to do their ordinary work

## BnF

We have finished our tests on the new Heritrix IIPC version and plan to put it into production before the end of June. This version will integrate also the migration to Postgresql 11.

After this deployment, we will be ready to launch a crawl of YouTube channels about the coronavirus. To enrich this collection, we will also launch an Instagram crawl : we are targeting a selection of 150 instagram profiles. Images and text will be crawled from picuki.com.

## ONB

## BNE

- We are focused on our coronavirus collection. We are collecting proposals from the regional web curators because they don't have access to the tools due to the situation. We are also accepting public nominations using a web form. We have more than 5 Tb of information and almost 2,000 seeds.
- We had planned to launch our annual broad crawl in April but we postpone it until the situation is normalized.
- We have already installed the version 6.1 of BCWeb in a test environment. He will test it these days before uploading it to the production environment

## KB-Sweden

## Next meetings

- June 9, 2020
- July 7, 2020
- September 8, 2020
- October 6, 2020
- November 3, 2020
- December 8, 2020
- January 5, 2021

## Any other business?

€€€€€€€€€€