

2019-04-09 Statusmeeting

- [Update on NAS latest tests and developments](#)
- [Feedback from the 2019 NAS workshop](#)
- [Status of the production sites](#)
- [Next meetings](#)
- [Any other business?](#)

Agenda for the joint NetarchiveSuite tele-conference 2019-04-09, 13:00-14:00.

Participants

- BNF: Sara, Clara
- ONB: Andreas, Michaela
- KB/DK - Copenhagen: -
- KB/DK - Aarhus: Colin, Sabine, Kristian, Knud Åge
- BNE: Alicia
- KB/Sweden: Par, Thomas, Peter

Update on NAS latest tests and developments

- Umbra - se also: <https://github.com/netarchivesuite/netarchivesuite-umbra-docker>

Feedback from the 2019 NAS workshop

Internal discussions on priorities, international discussions.

Status of the production sites

Netarkivet

Broad crawl

We have almost finished (only 3 slow jobs back) the first full broad crawl for 2019 without any software issues and we have never **harvested** so much as this time (more than 60 TB in 2 ½ month) withdrawn the -5003 returncodes.

It seems that our harvesters have seen a lot more than the quota is set to before they stop, so we still have a lot of -5003 in the crawllogs.

Selective crawls

We still focus on IP-validated access to content behind paywalls. We ran into problems with I-frames but they seem to be solved with Umbra. Another issue is to keep the site owners contact informations according to GDPR.

Event crawls

We are preparing for the elections for the EU Parliament and for Danish parliamentary elections. The latter has to take place at the latest in June 2019

Together with a colleague – a researcher we had a mini event crawl on April Fools using BCWeb for the nominations of URL's. As you never know, where April Fools pop up, the researcher wanted us to crawl with 4 hops. Thus, the crawls are still ongoing. Part of the evaluation would be to crawl with less hops next time.

After the war in the 1860th Denmark lost a part of Southern Jutland to Prussia/Germany. After WW1 Southern Jutland became Danish again. Next year we will celebrate the centennial of this reunion – preparations are popping up on the internet. Thus, we are preparing an event crawl: we already have collected about 40 URL's

Most urgent technical issue

Our citrix wayback access platform is performing very badly – among others it may take over 5 minutes to load a page and many images are not displayed

BnF

ONB

BNE

Broad crawl

We usually launch our annual broad crawl in April, but there is a new manager in our IT Team and they are working hard in restructuring all the IT Department.

Selective Crawls

We continue with our activity in selective crawls. We want to launch a new selective crawl about popular heritage. We focus on websites with important and unique information about small countries and villages. This websites are maintained by individuals or local associations and most of them disappear rapidly. Furthermore, many of them have content difficult to harvest as sounds records, video or interactive maps. We are thinking about requesting the deposit of those files when we cannot collect them.

Event crawls

We are working in three different elections: European Parliament elections, local elections and Spanish Government elections.

We have a daily crawl for twitter and facebook accounts.

KB-Sweden

Next meetings

- April 9
- May 7
- June 4
- July 2
- September 10
- October 8
- November 5
- December 3
- January 7, 2020

Any other business?

€€€€€€€€