

# 2018-06-12 Statusmeeting

- [Update on NAS 5.4.1](#)
- [Harvesting Youtube with NAS and H3: feedback from BnF](#)
- [Status of the production sites](#)
- [Next meetings](#)
- [Any other business?](#)

Agenda for the joint KB, BNF, ONB and BNE NetarchiveSuite tele-conference 2018-06-12, 13:00-14:00.

## Participants

- BNF: Sara, Géraldine
- ONB: Andreas, ~~Michaela~~
- KB/DK - Copenhagen: Tue, Nicholas, Søren, Anders
- KB/DK - Aarhus: ~~Celine~~, Sabine
- BNE: Mar
- KB/Sweden: Bengt

## Update on NAS 5.4.1

NAS 5.4 is available for download [here](#) but we are awaiting completion of the acceptance test before making a formal announcement.

We have actually found a bug (memory leak) in NAS 5.4 [NAS-2751 - Getting issue details...](#) **STATUS** which affects the new functionality to manage the number of Jobs on-queue. The feature is, in fact, disabled by default, but we are working on a quick patch-release so there will be a 5.4.1 within days.

## Harvesting Youtube with NAS and H3: feedback from BnF

In March and May, the team has worked on defining an integrated workflow to harvest Youtube channels and videos. We will present the results of our work.

## Status of the production sites

### Netarkivet

#### Broad crawl:

- Our second broad crawl 2018 is ongoing. We started with running step 1 with a limit of 10 MB/domain from May 20 to May 28 (552 jobs). We relaunched a run with a limit of 10 MB/domain. We have a problem with too many -50 return codes.

#### Event crawls:

- The collection on the collective negotiations on pay is almost finished, as all unions have accepted the results.
- We just launched a new event crawl on a political and cultural meeting on the island of Bornholm – called “Folkemødet”, which means “peoples meeting”. This a pilot project for the use of BCWeb.

All technical and practical obstacles for the use of BCWeb are “surmounted” – we have now two registered external users.

We have installed OpenWayback in our test environment, among others are we focusing on the replay of https-pages. It works for quite some sites, but not for all sites.

### BnF

### ONB

- Our Crawl of [twoday.net](#) is almost finished, but still running. The austrian blogging platform wanted to shutdown by end of May, but they postponed the shutdown to end of June. This gives us extra time to finish our jobs.
- As soon these jobs are finished we are upgrading to Nas 5.4 or 5.4.1 and preparing our domaincrawl
- We have a request for crawling a website regularly. It's the website of Vienna [wien.at](#). They want to support us with resources. Our Management has also interest in offering such a service.

### BNE

Our 2018 broad crawl finished a couple of weeks ago. Comparing with the one launched in 2016, that lasted 3 months, this one has been considerable shorter: only 42 days.

The number of .es domains is more than 1.900.000. The limit per domain was 150 MB. And around 50 TB were archived.

The event crawl on the Catalan elections has been closed. It lasted around 7 months and contains 1.800 seeds.

Recently we've been very busy with the National Politics collection, due to the many changes have been taking place in relation to the Government change.

We have plans to upgrade to NAS version 5.4 soon.

We have also been designing a web archive interface for the users, that includes search for subject, collection and titles along with the default url search. The design is more or less ready and now we are in the development phase.

A couple of months ago we heard about the closing of Wikispaces by the end of July. Wikispaces is a free hosting service, that hosts mainly academic and learning content. As there is no way to discriminate by language or country, it was necessary to count with some help from outside our team. We launched a social media campaign (a press release on the Library website and a call on Twitter) calling for nominations from the academic and research community along from individuals who know some Spanish wikispaces. We received many nominations. We consider this collection "at-risk" and we already have crawled more than 300 Spanish wikispaces.

#### **KB-Sweden**

#### **Next meetings**

- July 17th
- September 11th
- October 9th
- November 6th
- December 4th
- January 8th 2019

#### **Any other business?**