2018-11-06 Statusmeeting

- Update on NAS latest tests and developments
- Preparation of 2019 NAS workshop
- Status of the production sites
- Next meetings
- Any other business?

Agenda for the joint NetarchiveSuite tele-conference 2018-11-06, 13:00-14:00.

Participants

- BNF: Sara, Géraldine
- ONB: Andreas, Michaela
- KB/DK - Copenhagen: Tue, Stephen, Anders
- KB/DK - Aarhus: Colin, Sabine
- BNE: Mar
- KB/Sweden: Bengt

Update on NAS latest tests and developments

The Release Test for NetarchiveSuite 5.5 is under way and expected to be finished this week, followed by a formal release as soon as possible after that. See NetarchiveSuite 5.5 Release Notes for more details about the content of the release.

We are beginning a new development phase concentrating on updating our backend storage and processing architecture. We are moving to bitrepository.org software and also investigating a more modern mass-processing platform - for example index-generation via hadoop instead of our own NAS batch framework. For this reason we will be surveying our NAS partners to find out which parts of the NAS package are actually in use in different institutions.

Preparation of 2019 NAS workshop

2019 NAS workshop

Status of the production sites

Netarkivet
We launched step 2 of our 3rd broad crawl this year (with a limit of 14GB per domain) on 2018/10/23.

We looked at all our open issues and grouped them thematically:

- Harvesting problems
- Replay problems
- Improving existing functionalities
- New functionalities
- Automation of operations, which are solved manually at the moment
- Will be solved by existing projects

The aim was to find the most urgent problems, which we cannot solve without developers help.

We are working on the implementation of SOLR wayback to search in Netarchive. By now SOLR Wayback still is a prototype. Amongst others we need to clarify UX, security questions, how to do the logging and to chose a platform for the user access.

But the display of the results is much better than Blacklight.

We are working on a procedure for a new type of usage from the archive: data extraction for research project from the archive. The data to be extracted are determined by a search string – hopefully this would be rather easy with SOLR-wayback.

We are going to prepare a mini-event harvest “Week 46”. In week 46 the Royal Library collects local broadcast stations’ (both radio and television) productions. For a couple of years ago Netarchive started collecting there home pages.

BnF
ONB

- We are still running our domain crawl for this year and we are in the last quarter of the crawl. We are expecting the end in a few weeks.

BNE

KB-Sweden

Next meetings

- December 4th
- January 8th 2019

Any other business?