

2017 NAS workshop

The 2017 NAS workshop will take place on April 26-28 and will be hosted by National Library of Austria, Vienna

Location:

Address: Austrian National Library / Training Department, Augasse 2-6, 1090 Vienna, **Attention** - not on historic library premises in the city center!

Map: <https://www.onb.ac.at/bibliothek/ausbildung/neuer-standort/>

Public transport: metro lines U4 and U6, tram D

Route planner: <http://www.wienerlinien.at/eportal3/ep/channelView.do/pageTypeld/66533/channelId/-48703>

Hotels nearby:

Arthotel ANA Katharina, <http://ana-hotels.de/katharina> (2 min walk)

ibis Styles Wien City, <http://www.accorhotels.com/de/hotel-9034-ibis-styles-wien-city/index.shtml> (5 min walk)

Hotel Boltzmann, <http://www.hotelboltzmann.at/index.php> (15-20 min walk)

Participants:

Organization	Technical	Curator
Netarkivet	Colin Rosenthal Tue Larsen Søren Vejrup Carlsen	Sabine Schostag Stephen Hunt
ONB	Andreas P.	Michaela Mayr
BnF	Sara Aubry Thomas F.	Géraldine Camile
BNE	Juan Carlos García Arratia Fernando Monzón	Mar Pérez Morillo
NL of Sweden	Eva Meszaro	Pär Nilsson Daniel Jansson

- Topics to be discussed:
- Agenda
 - Schedule for 26.04.2017 (12:30-17:30)
 - Schedule for 27.04.2017 (9:00-17:00)
 - Schedule for 28.04.2017 (9:00-14:30)

Topics to be discussed:

NAS5 / Heritrix 3 - technical	Heritrix 3 - curatorial
-------------------------------	-------------------------

- State of the art of current developments
- Upcoming developments
- Introduce a multiple crawlers approach into NAS
- Videos/social media harvesting
- What CDX format are you using today and plan to support within next year?
- Which version of (Open)Wayback are you using today and what do think about the future development of OpenWayback?
- Performance of Wayback-Index. How to speed it up? Any experience with splitting up the index in several chunks or serving the index from multiple hosts?
- Which social media can you archive today?
- How to consolidate crawl.log and frontier search features in NetarchiveSuite?
- BnF's freetext search (better than KB DK's) - anything to share with the community?
- Automatic quality assurance. Any Ideas? Proof of concepts?
- Others ?

- Feedback on using NAS 5 and Heritrix 3
- Missing features
- Priorities for future development
- Is it possible to connect other tools than Heritrix to NAS (tools that can produce WARC files and capture content, which Heritrix is not able to catch) If so, which tools to we want to use?
- Revival and update of the curator roadmap
- Harvest the electoral web: selection, harvest parameters
- Experiences with harvesting pages with login content (pay walls)
- Experiences with harvesting images embedded in javascript (and replay them in the archive)
- Exchange of experiences with documentation of the crawls (in and outside NAS)
- Others ?

Agenda

Schedule for 26.04.2017 (12:30-17:30)

12:30 - 14:00 Welcome, sandwiches and coffee

14:00 - 14:30 [Workshop introduction](#) (Michaela, Sara)

14:30 - 16:00 Institution updates and plans for 2017

- [Update from ONB](#) (Michaela)
- [Update from BNE](#) (Mar, Juan Carlos)
- [Update from BnF](#) (Géraldine)
- [Update from KB Danemark](#) (Sabine, Tue)
- [Update from KB Sweden](#) (Daniel, Pär)

16:00 - 16:15 Coffee break

16:15 - 17:30 [NetarchiveSuite 5.3: demo and discussion of latest features and installation challenges](#) (Colin, Sara)

19:30 Dinner (at own expense)

Schedule for 27.04.2017 (9:00-17:00)

09:00 - 11:30

Technical track:

- Share experiences in using NAS 5 and Heritrix 3 from a technical perspective
- Discuss issues and future developements regarding H3 integration
- Establish a list of NAS bugs and missing features [DK list](#) - [BnF list](#)
- Sum up, define priorities

Curator track:

- Share experiences with crawl documentation in and outside NAS (start with input from [Netarchive](#), [ONB](#))
- Share experiences in using NAS 5 and Heritrix 3 from a curatorial perspective ([BnF](#))
- Establish a list of NAS bugs and missing features
- Sum up, define priorities

10:30 - 10:45 Coffee break

11:30 - 12:30 [Common track](#): presentation of curators and technical priorities, update of NAS curator road map: <https://sbforge.org/jira/projects/NASC/summary/statistics>

12:30 - 14:00 Lunch

14:00 - 15:30

Common track: Harvesting videos and social medias:

- State of the art and considerations in the different institutions (start with input from [BnF](#), [Netarchive](#))
- Discuss possible external tool integration

15:30 - 15:45 Coffee break

15:45 - 17:00

Technical track:

- Further technical considerations
- Discuss or make a proof of concept

Curator track:

- Share experiences with harvesting content beyond paywalls (start with input from [ONB](#) - see also [Documentation](#) and [presseCredential.txt](#) -, [BnF](#), [Netarchive](#))

17:00 - 19:00 Guided tour at [Austrian National Library State Hall](#) (optional)

Schedule for 28.04.2017 (9:00-14:30)

09:00 - 09:45

Common track: Demo of BCweb functionalities (Sara)

9:45 - 11:45

Technical track:

- Discuss BCweb and NAS articulation
- Discuss how to cooperate on BCweb and push new features
- Discuss NAS development cycles and how to best contribute to the code
- Discuss on access tools (OpenWayback, full text search) and cooperation opportunities

Curator track:

- Collection strategies (start with input from [Netarchive](#). In Denmark web content is part of legal deposit. See the [unofficial English translation](#))
- BCweb: use at [BNE](#) and [BnF](#), further cooperation at curator level

10:30 - 10:45 Coffee break

11:45 - 13:00 Tracks sum-up, community next steps

13:00 - 14:30 Lunch and goodbye

