2019 NAS workshop

The 2019 NAS workshop will take place on February 20-22 and will be hosted by National Library of Spain in Madrid.

Location: National Library of Spain (entry at the ground floor)

Address: Paseo de Recoletos, 20-22 - 28071-Madrid

Map: [https://www.google.com/maps/place/Biblioteca+Nacional+de+Espana/@40.4235049,-3.6916211,17z/data=!3m1!4b1!4m5!3m4!1s0xd4228907e039627:0xd5b5764d8a0a53f78m213d40.4235049/4d-3.6894324](https://www.google.com/maps/place/Biblioteca+Nacional+de+Espana/@40.4235049,-3.6916211,17z/data=!3m1!4b1!4m5!3m4!1s0xd4228907e039627:0xd5b5764d8a0a53f78m213d40.4235049/4d-3.6894324)

Tourist information: [recommended things to see in Madrid](#)

Hotels: see below

Participants:

<table>
<thead>
<tr>
<th>Organization</th>
<th>Technical</th>
<th>Curator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Netarkivet</td>
<td>Colin Samuel Rosenthal</td>
<td>Anders Klindt Myrvoll</td>
</tr>
<tr>
<td></td>
<td>Knud Aage Hansen</td>
<td>Sabine Schostag</td>
</tr>
<tr>
<td></td>
<td>Tue Hejlskov Larsen</td>
<td>Stephen Hunt</td>
</tr>
<tr>
<td></td>
<td>Kristian Bak</td>
<td></td>
</tr>
<tr>
<td>ONB (via Skype)</td>
<td>Andreas P</td>
<td>Michaela</td>
</tr>
<tr>
<td>BnF</td>
<td>Sara Aubry</td>
<td>Géraldine Camile</td>
</tr>
<tr>
<td></td>
<td>Clara Wiatrowski</td>
<td></td>
</tr>
<tr>
<td>BNE</td>
<td>Juan Carlos García</td>
<td>Alicia Pastrana</td>
</tr>
<tr>
<td></td>
<td>José María Martín</td>
<td>Maria Bueno</td>
</tr>
<tr>
<td></td>
<td>Fernando Monzón</td>
<td>María Ezquerra</td>
</tr>
<tr>
<td></td>
<td>Luis Sánchez</td>
<td>Yasmín Rommaneh</td>
</tr>
<tr>
<td></td>
<td>Nuria Serrano</td>
<td>Mar Pérez</td>
</tr>
<tr>
<td>NL of Sweden</td>
<td>Thomas Roos</td>
<td>Pär Nilsson</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Peter Svanberg</td>
</tr>
</tbody>
</table>

Topics to be discussed:

- Technical
- Curatorial

Topics to be discussed:

- Agenda
  - Schedule for 20.02.2019 (12:30-17:30)
  - Schedule for 21.02.2019 (9:00-17:00)
  - Schedule for 22.02.2019 (9:00-14:30)
• State of the art of current bugs and possible fixes
• State of the art of current developments and upcoming developments in NetarchiveSuite
• Integration of latest H3 stable release
• Videos, social media, Umbra (installation, configuration tests and usage)
• Introducing WARC 1.1
• Brainstorming on priorities for future developments
• Brief state of the art on access tools: use and perspectives in the different institutions
• SolrWayback demo
• How does your NAS Deployment/Configuration look like (Settings, Hardware)?
• Oracle java is subject to a fee. ONB's IT department would like to switch completely to OpenJava if possible. Does NAS work as usual on OpenJava? What do you think? What will you do?

• NAS missing features
  ◦ The curator roadmap NASC and comments from Vienna meeting
  ◦ Brainstorming on priorities for future developments
  ◦ Scheduling harvests at a precise date or period
  ◦ Presentation of BCweb latest release and futures evolutions
  ◦ Presentation of how do we collect and crawl youtube and give access
  ◦ Coordination of external selections
  ◦ What documentation shall we provide for researchers?
  ◦ How do we make workspaces for researchers - tools, limits?
  ◦ Capturing social media
  ◦ Webarchives and digital preservation
  ◦ Which browser and version do we support today and in the near future in harvester requests, Umbra and in archive access? Umbra usage and experiences
  ◦ OpenWayback and CDX creation issues and development - and experiences with other tools e.g. pywb, SOLRWayback?
• Broad crawls
  ◦ How do we make job monitoring during broad or big "deep" crawls?
  ◦ How do we manage huge webhotels?
  ◦ How do we manage byte/objects limits for different groups of domains?

Agenda

Schedule for 20.02.2019 (12:30-17:30)

12:30 - 14:00 Arrival, sandwiches and coffee
14:00 - 14:15 Welcome (Ana Santos Aramburo, director of BNE)
14:15 - 14:30 Workshop introduction (Mar, Sara)
14:30 - 16:00 Institution updates and plans for 2019 (15 min each)
  • Update from ONB (Michaela)
  • Update from BNE (Mar, Juan Carlos)
  • Update from KB Sweden (Pär)
  • Update from BnF (Géraldine)
  • Update from KB Danemark (Sabine, Tue, Anders)
16:00 - 16:15 Coffee break
16:15 - 17:30 NetarchiveSuite 5.5: demo and discussion of latest features including Umbra (Colin), Umbra usage and experiences, Feedback on tests with input from Clara (ppt), Tue (ppt)
20:00 Dinner (at own expense). Inclan Brutal Bar (Calle Álvarez Gato, 4)

Schedule for 21.02.2019 (9:00-17:00)

09:00 - 12:00

Technical track:

• Share NAS deployment and configuration in our institutions to identify used/unused components: See form.
• Discuss state of the art of current bugs and possible fixes
  ◦ Current bugs identified during ONB Daily Crawls & Domain Crawl 2018 (Andreas)
  ◦ Current bugs identified during BnF Domain Crawl 2018 (Clara)
• Review lists of NAS bugs and missing features and internal lists: NAS curator roadmap (NASC), BnF 2019 list
• JIRA issues labelled "Madrid":

<table>
<thead>
<tr>
<th>Key</th>
<th>Summary</th>
<th>Created</th>
<th>Reporter</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAS-2794</td>
<td>Strange Design Choice for Harvester Registration</td>
<td>Aug 20, 2018</td>
<td>Colin Rosenthal</td>
</tr>
<tr>
<td>NAS-2790</td>
<td>Creation of domains out of a domainlist with two identical domain names fails</td>
<td>Jul 17, 2018</td>
<td>Andreas P</td>
</tr>
</tbody>
</table>
Discuss possible integration of OpenJava, latest H3 stable release and WARC 1.1
Brainstorm on priorities and NAS codebase evolution for future developments
Discuss the possibility to submit an IIPC project

Curator track:
- Review and update the NAS curator roadmap (NASC)
- Brainstorm on priorities for future developments from a curatorial perspective
- Discuss practices and challenges in coordinating external selections (Géraldine, Sabine, Mar)

10:30 - 10:45 Coffee break
12:00 - 12:30 Sum up of curators and technical priorities
12:30 - 14:00 Lunch
14:00 - 17:00 Complex harvesting
Share experiences, practices and questions in the management of broad crawls:
- How do we make job monitoring during broad or big “deep” crawls?
- How do we manage huge webhotels (companies that host many websites)?
- How do we track web parkings?
- How do we manage byte/objects limits for different groups of domains?

Share experiences and practices in crawling and giving access to YouTube videos (Sara)
Share experiences in crawling social media (Facebook, Twitter, SlideShare, Flickr, Instagram)
Discuss possible further cooperation on these topics, common tools integration
15:30 - 15:45 Coffee break
17:00 - 19:00 Guided tour of the BNE

Schedule for 22.02.2019 (9:00-14:30)
09:00 - 10:30 Update on BCweb (Géraldine, Clara) - CSV sample
  - Demo of BCweb new functionalities
  - Update on BnF current and upcoming developments
  - Update on open source status
  - Discuss interest in upgrading and possible community developments
10:30 - 10:45 Coffee break
10:45 - 12:30 Access tools to webarchives
  - Demo of the SolrWayback search interface and playback engine for WARCs (Anders)
  - Discuss perspectives, projects and questions in the different institutions (input Sara, Mar, ?)
  - Browser, OpenWayback and CDX creation issues and development, experiences with other tools e.g. pywb, SOLRWayback
12:30 - 13:00 Community next steps
13:00 - 14:30 Lunch and goodbye
Hotels nearby: We suggest to look for some offers on www.booking.com for these hotels below, as the Library can't provide special offers.

- **Hotel Gran Versalles ****
  http://www.hotelgranversalles.es/
  How to get: https://www.google.com/maps/dir/Gran+Versalles,+Calle+Covarrubias,+Madrid/Biblioteca+Nacional+de+España,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+España/@40.4283965,3.6971013,17z/data=!3m1!4b1!4m14!5i0!7i11019!8i29169

- **Hotel Mediodía **
  http://www.mediodiahotel.com/
  How to get: Bus 27 https://www.google.com/maps/dir/Hotel+Mediod%C3%ADa,+Plaza+del+Emperador+Carlos+V,+8,+28012+Madrid/Biblioteca+Nacional+de+España,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+España/@40.4283965,3.6971013,17z/data=!3m1!4b1!4m14!5i0!7i11019!8i29169

- **Hotel NH Collection Madrid Colón****
  https://www.nh-hotels.es/hotel/nh-collection-madrid-colon
  How to get: https://www.google.com/maps/dir/Hotel%20NH%20Collection%20Madrid%20Col%C3%B3n,+Calle+Marqu%C3%A9s+de+Zurgena,+Madrid/Biblioteca+Nacional+de+Esp%C3%A1n,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+Esp%C3%A1n/@40.4283965,3.6971013,17z/data=!3m1!4b1!4m14!5i0!7i11019!8i29169

- **Hotel NH Madrid Alonso Martínez**
  https://www.nh-hotels.es/hotel/nh-madrid-alonso-martinez
  How to get: https://www.google.com/maps/dir/Hotel+NH+Madrid+Alonso+Mart%C3%ADnez,+Calle+de+Santa+Engracia,+Madrid/Biblioteca+Nacional+de+Esp%C3%A1n,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+Esp%C3%A1n/@40.4283965,3.6971013,17z/data=!3m1!4b1!4m14!5i0!7i11019!8i29169

- **Hotel Mora**
  http://www.hotelmora.com/
  How to get: Bus 27 https://www.google.com/maps/dir/Hotel+Mora,+Paseo+del+Prado,+Madrid/Biblioteca+Nacional+de+Esp%C3%A1n,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+Esp%C3%A1n/@40.4283965,3.6971013,17z/data=!3m1!4b1!4m14!5i0!7i11019!8i29169
• Hotel One Shot Recoletos 04****
  http://www.hotelonesthotrecoletos04.com/

  How to get: https://www.google.com/maps/dir/Hotel+One+Shot+Recoletos+04,+Calle+Salustiano+O%27%3B3zagora,+Madrid
  /Biblioteca+Nacional+de+Espa%C3%B1a+B1a,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+Espa%C3%B1a+B1a@40.422246,-3.6928704,17z/data=!3m1!4b1!4m1!4m11!1e1!1s0xd4228907e039667!0xd5b5764d8a0a53f72m2!d-3.6893424!2d40.4235049!3e2

• Hotel Ibis Styles Madrid Prado***

  How to get: https://www.google.com/maps/dir/Hotel+ibis+Styles+Madrid+Prado,+Calle+del+Prado,+Madrid/Biblioteca+Nacional+de+Espa%C3%B1a,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+Espa%C3%B1a+B1a@40.428549,-3.7033754,16z/data=!3m1!4b1!4m11!1e1!1s0xd4228907e039667!0xd5b5764d8a0a53f72m2!d-3.6893424!2d40.4235049!3e2

• Hotel Leonardo City Center
  https://www.leonardo-hotels.es/leonardo-hotel-madrid-city-center

  How to get: Tube Line 4
  https://www.google.com/maps/dir/Hotel+Leonardo+City+Center,+Calle+de+Alberto+Aguilera,+Madrid
  /Biblioteca+Nacional+de+Espa%C3%B1a+B1a,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+Espa%C3%B1a+B1a@40.428549,-3.7033754,16z/data=!3m1!4b1!4m11!1e1!1s0xd4228907e039667!0xd5b5764d8a0a53f72m2!d-3.6893424!2d40.4235049!3e2

• Hostal Pizarro

  How to get: https://www.google.com/maps/dir/Hotel+Hostal+Pizarro,+Calle+de+Fuencarral,+95,+28004+Madrid/Biblioteca+Nacional+de+Espa%C3%B1a,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+Espa%C3%B1a+B1a@40.422246,-3.6928704,17z/data=!3m1!4b1!4m11!1e1!1s0xd4228907e039667!0xd5b5764d8a0a53f72m2!d-3.6893424!2d40.4235049!3e2

• Hostal Residencia Don Diego***
  http://www.hostaldondiego.com

  How to get: https://www.google.com/maps/dir/Hotel+Hostal+Residencia+Don+Diego,+Calle+de+Vel%C3%A1zquez,+Madrid
  /Biblioteca+Nacional+de+Espa%C3%B1a+B1a,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+Espa%C3%B1a+B1a@40.422246,-3.6928704,17z/data=!3m1!4b1!4m11!1e1!1s0xd4228907e039667!0xd5b5764d8a0a53f72m2!d-3.6893424!2d40.4235049!3e2

• Hostal Gallardo***
  www.hostalgallardo.com

  How to get: https://www.google.com/maps/dir/Hotel+Gallardo,+Calle+de+Fuencarral,+95,+28004+Madrid/Biblioteca+Nacional+de+Espa%C3%B1a,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+Espa%C3%B1a+B1a@40.422246,-3.6928704,17z/data=!3m1!4b1!4m11!1e1!1s0xd4228907e039667!0xd5b5764d8a0a53f72m2!d-3.6893424!2d40.4235049!3e2

• Hostal Retiro
  http://www.hostalretiro.com

  How to get: https://www.google.com/maps/dir/HOSTAL+RETIRO,+Calle+de+O'Donnell,+Madrid/Biblioteca+Nacional+de+Espa%C3%B1a,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+Espa%C3%B1a+B1a@40.422246,-3.6928704,17z/data=!3m1!4b1!4m11!1e1!1s0xd4228907e039667!0xd5b5764d8a0a53f72m2!d-3.6893424!2d40.4235049!3e2

• Hostal Salamanca
  http://hostalsalamanca.com

  How to get: Tube lines 2 - 4 https://www.google.com/maps/dir/Hotel+Hostal+Salamanca,+Calle+de+Jos%27C3%A9+Ortega+y+Gasset,+89,+28006+Madrid/Biblioteca+Nacional+de+Espa%C3%B1a+B1a,+Paseo+de+Recoletos,+20-22,+28001+Madrid,+Espa%C3%B1a+B1a@40.422246,-3.6928704,17z/data=!3m1!4b1!4m11!1e1!1s0xd4228907e039667!0xd5b5764d8a0a53f72m2!d-3.6893424!2d40.4235049!3e2