A glance at the past, a look to the future: approaches to collecting Croatian web

Karolina Holub, Inge Rudomino, Draženko Celjak
National and University Library in Zagreb
SRCE - University of Zagreb, University Computing Centre
Web archiving in Croatia

- **1997 →** Law on Libraries (legal deposit)
- **1998 →** National and University Library in Zagreb started cataloguing online resources
- **2003 →** project Design of the System for Capturing and Archiving Legal Deposit of Croatian Web Publications
- September **2004 →** first archived web resource

*National and University Library in Zagreb*
selective archiving • since 2004

.hr harvesting • since 2011
• .hr, .from.hr, .com.hr

thematic collections • since 2011
• 8 collections
Selective approach

Selection criteria
http://haw.nsk.hr/en/selection-criteria

IN → news media, portals, web sites of institutions, associations, clubs, scientific and research projects, selected personal pages, blogs, books, journals, articles, etc.

OUT → search engines, games, advertising pages, pages of companies and businesses, mailing lists, resources distributed by e-mail, intranet, forums, digitised resources, other web archives

English title page of the Croatian Web Archive (HAW) http://haw.nsk.hr/en
What's behind selective archiving

- archiving system *DAMP* developed by SRCE - University of Zagreb, University Computing Centre
- identification (registration form) [http://haw.nsk.hr/en/registration_form](http://haw.nsk.hr/en/registration_form)
- cataloguing every title (MARC 21)
- daily data exchange between ILS and archiving system
- access → public/within the NUL
- full-text search (basic & advanced), browse by subject, alphabetically
- OAI-PMH interface for metadata harvesting (e.g. Europeana)
Domain distribution

- **hr**: 72.30%
- **com**: 18.25%
- **net**: 3.08%
- **org**: 3.02%
- **info**: 1.21%
- **eu**: 0.82%
- **other**: 1.32%

2017 Web Archiving Conference - London
.hr harvesting workflow

Novelty in 6th harvesting:
- single-letter .hr, e.g. http://6.hr
- diacritics, e.g. http://šktigrići.hr/
.hr harvesting data
The Croatian Web Archive periodically harvests websites related to topics and events of national importance. Please select a title of a particular thematic harvesting to get the list of archived websites.

<table>
<thead>
<tr>
<th>Thematic harvesting</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011 Croatian Parliament Elections</td>
</tr>
<tr>
<td>Referendum on the Accession of the Republic of Croatia to the European Union (22 January 2012)</td>
</tr>
<tr>
<td>2013 European Parliament Elections in Croatia</td>
</tr>
<tr>
<td>2013 Local Elections</td>
</tr>
<tr>
<td>The Accession of the Republic of Croatia to the European Union (1 July 2013)</td>
</tr>
<tr>
<td>Flood in Croatia 2014</td>
</tr>
<tr>
<td>2015 Croatian Parliament Elections</td>
</tr>
<tr>
<td>2016 Croatian Parliament Elections</td>
</tr>
</tbody>
</table>

Selective harvesting
6,410 catalogued & archived titles
47,578 archived instances
904 inactive titles
263,049,089 files

.hr harvesting
415 mil. Status 200

Thematic harvestings
1,256 web sites

40 TB
Google Analytics, 24.05.2016. – 24.05.2017.
A look to the future

- new web interface
- full-text search for domain (.hr) harvestings
- URN:NBN
- OpenWayback
- single access point to all digital collections

Digital Collections of the National and University Library in Zagreb [http://digitalna.nsk.hr/](http://digitalna.nsk.hr/)
Thank you!

http://haw.nsk.hr/en
haw@nsk.hr
damp@srce.hr

kholub@nsk.hr, irudomino@nsk.hr,
drazenko.celjak@srce.hr