Website archiving at the Rotterdam City Archive.

Concept and practice 2006 – present day.

André Skyaasen.
Digital archives and collections.
Outline.

1. Introduction to Rotterdam City Archive.
2. Webarchiving concept at Rotterdam City Archive.
4. Practice 2019 – present day.
5. Summary.
6. Questions.
Rotterdam City Archive.

Municipality began collecting archives and curiosities in 1851. Archive established in 1857.

Municipality of Rotterdam built a dedicated archive building in 1900 at the Mathenesserlaan. Survived the bombing in 1940 because of...
Rotterdam City Archive.

From 1998 onwards at the Hofdijk.
Rotterdam City Archive.

One of the oldest and largest local archives in the Netherlands.
Rotterdam City Archive.

One of the oldest and largest local archives in the Netherlands.

Collection of 23 kilometres of documents, photos, maps, posters, blueprints, audio visual material and more.
Rotterdam City Archive.

One of the oldest and largest local archives in the Netherlands.

Collection of 23 kilometres of documents, photos, maps, posters, blueprints, audio visual material and more.

Oldest piece is from 1201.
Rotterdam City Archive.

One of the oldest and largest local archives in the Netherlands.

Collection of 23 kilometres of documents, photos, maps, posters, blueprints, audio visual material and more.

Oldest piece is from 1201.

City charter from 1340.
Rotterdam City Archive.

One of the oldest and largest local archives in the Netherlands.

Collection of 23 kilometres of documents, photos, maps, posters, blueprints, audio visual material and more.

Oldest piece is from 1201.

City charter from 1340.

Websites and other material from 2021. E-depot with 300 TB of data.
2. Webarchiving concept in Rotterdam municipality.
2. Webarchiving concept in Rotterdam municipality.

   Two trajectories:
2. Webarchiving concept in Rotterdam municipality.

Two trajectories:

Legally required to archive Rotterdam municipality websites.
- Performed by separate departments within the municipality.
2. Webarchiving concept in Rotterdam municipality.

Two trajectories:

Legally required to archive Rotterdam municipality websites.

- Performed by a separate department within the municipality.

Curated selection according to an acquisition policy.
Municipality websites.
Municipality websites.

Archived according to legal requirements for government bodies for:

Accountability.

Evidence.
Municipality websites.
Curated selection.
Curated selection.

Goal: Build a collection of web archives that represents websites of social and cultural aspects within the municipality.
Curated selection.

Goal: Build a collection of web archives that represents websites of social and cultural within the municipality.

Similar to collecting programmes, folders and flyers.
- Cultural organisations and
- Social clubs and groups.
- Website development over
Curated selection.

Look-and-feel, not primarily content.

- Replicate the communication to users.
- Deliberately not server-side method.
- No software archiving.
Curated selection.

Look-and-feel, not primarily content.
- Replicate the communication to users.
- Deliberately not server-side method.
- No software archiving.

Issues and essential aspects:
Curated selection.

Look-and-feel, not primarily content.
• Replicate the communication to users.
• Deliberately not server-side method.
• No software archiving.

Issues and essential aspects:
• How much to archive relative to how much there is available.
Curated selection.

Look-and-feel, not primarily content.
• Replicate the communication to users.
• Deliberately not server-side method.
• No software archiving.

Issues and essential aspects:
• How much to archive relative to how much there is available.
• Frequency: How often to archive? Useful?
Curated selection.

Look-and-feel, not primarily content.
- Replicate the communication to users.
- Deliberately not server-side method.
- No software archiving.

Issues and essential aspects:
- How much to archive relative to how much there is available.
- Frequency: How often to archive? Useful?
- Costs.
  - External company.
  - Personnel.
  - Storage.
Curated selection.

Look-and-feel, not primarily content.
• Replicate the communication to users.
• Deliberately not server-side method.
• No software archiving.

Issues and essential aspects:
• How much to archive relative to how much there is available.
• Frequency: How often to archive? Useful?
• Costs.
  • External company.
  • Personnel.
  • Storage.
• Relevance of the task.
  • What is worthwhile?
  • What is representative?
Curated selection.

Look-and-feel, not primarily content.
- Replicate the communication to users.
- Deliberately not server-side method.
- No software archiving.

Issues and essential aspects:
- How much to archive relative to how much there is available.
- Frequency: How often to archive? Useful?
- Costs.
  - External company.
  - Personnel.
  - Storage.
- Relevance of the task.
  - What is worthwhile?
  - What is representative?
- Selection of sites.
Webarchiving setup at RCA.
Webarchiving setup at RCA.

External web archiving service provider.
Webarchiving setup at RCA.

External web archiving service provider.

Four co-workers who are involved in small capacities.

- Librarian.
- Digital information advisor.
- Information manager.
- Digital archives and collections specialist.
Webarchiving setup at RCA.

External web archiving service provider.

Four co-workers who are involved in small capacities.
  • Librarian.
  • Digital information advisor.
  • Information manager.
  • Digital archives and collections specialist.

Limited capacity.
Webarchiving setup at RCA.

External web archiving service provider.

Four co-workers who are involved in small capacities.
  • Librarian.
  • Digital information advisor.
  • Information manager.
  • Digital archives and collections specialist.

Limited capacity.

Important **unanswered** questions.
Consent.
Consent.

**Opt in.**

- Proved too complicated.
- Too much e-mail work.
Consent.

Opt in.
- Proved too complicated.
- Too much e-mail work.

Issues:
- Fear of outdated versions online.
- Content rights.
- Design copyrights.
Consent.

Opt in.
- Proved too complicated.
- Too much e-mail work.

Issues:
- Fear of outdated versions online.
- Content rights.
- Design copyrights.

Opt out.
- No reply = consent.

External company: Capsis.

<table>
<thead>
<tr>
<th>Gemeentearchief Rotterdam</th>
<th>Archive size</th>
<th>2,95 TB (53128466 files)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collecties van websites van het Gemeentearchief Rotterdam (GAR).</td>
<td>Number of groups</td>
<td>52</td>
</tr>
<tr>
<td>View and edit advanced information</td>
<td>Number of collections</td>
<td>1201</td>
</tr>
<tr>
<td></td>
<td>Number of snapshots</td>
<td>4654</td>
</tr>
<tr>
<td></td>
<td>First snapshot</td>
<td>02-14-2006 / 14:51</td>
</tr>
<tr>
<td></td>
<td>Last snapshot</td>
<td>02-18-2020 / 06:06</td>
</tr>
</tbody>
</table>

Small webarchiving department.

Small webarchiving department.

Created Snapshots, not necessarily WARC's.

Small webarchiving department.

Created Snapshots, not necessarily WARCs.

Problem handled 2014.

Solution provided by Capsis.

Small webarchiving department.

Created Snapshots, not necessarily WARC.

Problem handled 2014.

Solution provided by Capsis.

Problem re-discovered in 2020/2021.

• WARC’s do not conform to WARC format standard.
  • Major subsequent problems.

Small webarchiving department.

Created Snapshots, not necessarily WARC's.

Problem handled 2014.

Solution provided by Capsis.

Problem re-discovered in 2020/2021.

• WARC’s do not conform to WARC format standard.
  • Major subsequent problems.

Integrity of some files remains uncertain.

Problematic, non-functional or partly functional WARC.s.

Diagnosing the problems is challenging.

Questions:

What is the problem?

How do you find out?

What is the solution?
Warc analysis.

Quality control tools:

• Notepad.
• DROID.
• JHOVE.
• Warcat.
• Warcio.
• More...

Third party WARC viewers:

• Replayweb.page.
• Conifer.
• Preservica.
Capsis PreSurf.
Capsis PreSurf.
Warc analysis.
Warc analysis.
Warc analysis.
Warc analysis.
Warc analysis.

```
[ ] PRSFR_434_20110915104649_00001_presurf.rotterdam.local.warc - Notepad
File Edit Format View Help

WARC/1.0
WARC-Type: warcinfo
WARC-Record-ID: 4832e9b0-9d7d-43d0-851f-1d2ab16d017d
WARC-Date: 2008-01-05T07:01:58+01:00
Content-Type: application/warc-fields
Content-Length: 406
WARC-Filename: PRSFR_434_20110915104649_00001_presurf.rotterdam.local.warc.gz

operator: Beheerders
software: Capsis Presurf/2.5.1 http://www.capsis.nl/
robots: Ignore
hostname: presurf.rotterdam.local
ip: 10.33.40.13
http-header-user-agent: Mozilla/4.0 (compatible; MSIE 6.0; GAR)
http-header-from: Noreply@isr.rotterdam.nl
format: WARC File Format 1.0 / EN ISO 28500
description:
isPartOf: Port of Rotterdam

WARC/1.0
WARC-Type: response
WARC-Target-URI: http://www.portofrotterdam.com/nl/home/
WARC-Date: 2008-01-03T07:02:00+01:00
WARC-Warcinfo-ID: 4832e9b0-9d7d-43d0-851f-1d2ab16d017d
WARC-Payload-Digest: sha1:c2f55e6894c7270bcc9e35d26b0e6ecbed08010e
WARC-Block-Digest: sha1:c2f55e6894c7270bcc9e35d26b0e6ecbed08010e
WARC-IP-Address: 10.33.40.13
WARC-Record-ID: <urn:uuid:54db495c7-2bf2-4eaf-ad65-6e78102fbb55>
Content-Type: application/http;msgtype=response
WARC-Identified-Payload-Type: text/html
```

Website archiving at the RCA| André Skyassen | 04.11.2021 Vreden.
Warc analysis.

Checks to investigate:

Integrity.

Authenticity.
Warc analysis.

Checks to investigate:

Integrity.

Authenticity.

We cannot always guarantee this for webarchives from 2006 – 2019.
Warc analysis.

Checks to investigate:

Integrity.

Authenticity.

We cannot always guarantee this for webarchives from 2006 – 2019.

Solution:
• Negotiate a solution with Capsis based on a contract.
Warc analysis.

Checks to investigate:

Integrity.

Authenticity.

We cannot always guarantee this for webarchives from 2006 – 2019.

Solution:
• Negotiate a solution with Capsis based on a contract.
• Ask Archiefweb.eu to make a solution.
Warc analysis.

Checks to investigate:

Integrity.

Authenticity.

We cannot always guarantee this for webarchives from 2006 – 2019.

Solution:

• Negotiate a solution with Capsis based on a contract.
• Ask Archiefweb.eu to make a solution.
• Find a solution via our network.
4. Website archiving practice 2019 – present day.
4. Website archiving practice 2019 – present day

External company: Archiefweb.eu.
4. Website archiving practice 2019 – present day.

External company: Archiefweb.eu.

The issues of 2006 – 2019 were part of the decision to make a change.

Other unsolved technical issues.
4. Website archiving practice 2019 – present day.

External company: Archiefweb.eu.

The issues of 2006 – 2019 were part of the decision to make a change.

Other unsolved technical issues.

Challenging middle man: Technical management was run by municipality IT department.
- Always a different case handler.
- Lack of familiarity with web archiving.
4. Website archiving practice 2019 – present day.

External company: Archiefweb.eu.

The issues of 2006 – 2019 were part of the decision to make a change.

Other unsolved technical issues.

Challenging middle man: Technical management was run by municipality IT department.
- Always a different case handler.
- Lack of familiarity with web archiving.

Archiefweb.eu has webarchiving as their core business.
4. Website archiving practice 2019 – present day.

External company: Archiefweb.eu.

The issues of 2006 – 2019 were part of the decision to make a change.

Other unsolved technical issues.

Challenging middle man: Technical management was run by municipality IT department.
  - Always a different case handler.
  - Lack of familiarity with web archiving.

Archiefweb.eu has webarchiving as their core business.

SaaS solution.
2019 – present day.
2019 – present day.
2019 - present day.
2019 – present day.

Issues:
2019 – present day.

Issues:

YouTube videos cannot be harvested as embedded content since 2020.
2019 – present day.

Issues:

YouTube videos cannot be harvested as embedded content since 2020.

Partly failed harvests.
2019 – present day.

Issues:

YouTube videos cannot be harvested as embedded content since 2020.

Partly failed harvests.

Stored externally.
Public access:

Public access:
Public access.

Public access:

Not currently being made accessible.

Would like to make everything accessible.

Would like to be able to invest more time into resolving these issues.
Public access.

Reading room, not website.

Complicated and unsolved legal questions concerning stakeholders.

- Web design rights.
- Content rights.
  - Text.
  - Images.
- Protection of privacy.
  - Comments.
  - Posts.
- Social media.
  - Links.
  - Widgets.
2019 - present day.

Envisaged, ideal functionality:

- **Web archiving service**
- **E-depot**
- **Reading room**
- **CMS**
- **Website**
- **User**
5. Summary of website archiving practices at RCA.

Concept:
• Legally required municipality websites.
• Curated selection.

• Several inherited problems.

Unsolved legal questions on national and international level regarding public access.

External service provider 2019 – present day.

Regularly archive 5-600 websites.
6. Questions.

Does anyone have any questions?
Thank you for your attention.

André Skyaasen.
Digital archives and collections.
a.skyaasen@rotterdam.nl.