

# Website archiving at the Rotterdam City Archive.



Concept and practice 2006 – present day.

**Stadsarchief Rotterdam**

André Skyaaasen.  
Digital archives and collections.



**Gemeente  
Rotterdam**

# Outline.

1. Introduction to Rotterdam City Archive.
2. Webarchiving concept at Rotterdam City Archive.
3. Practice 2006 - 2019.
4. Practice 2019 – present day.
5. Summary.
6. Questions.



An aerial photograph of Rotterdam, Netherlands, taken during sunset. The sky is a vibrant orange and yellow, with the sun low on the horizon. The city's skyline is visible in the background, featuring several modern high-rise buildings, including the Afdeling and the Rotterdam Tower. In the foreground, there is a dense residential area with traditional Dutch-style houses and a large, modern brick building with many windows. The word "Rotterdam." is overlaid in large white text across the center of the image.

# Rotterdam.

**Stadsarchief Rotterdam**



# Rotterdam City Archive.

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

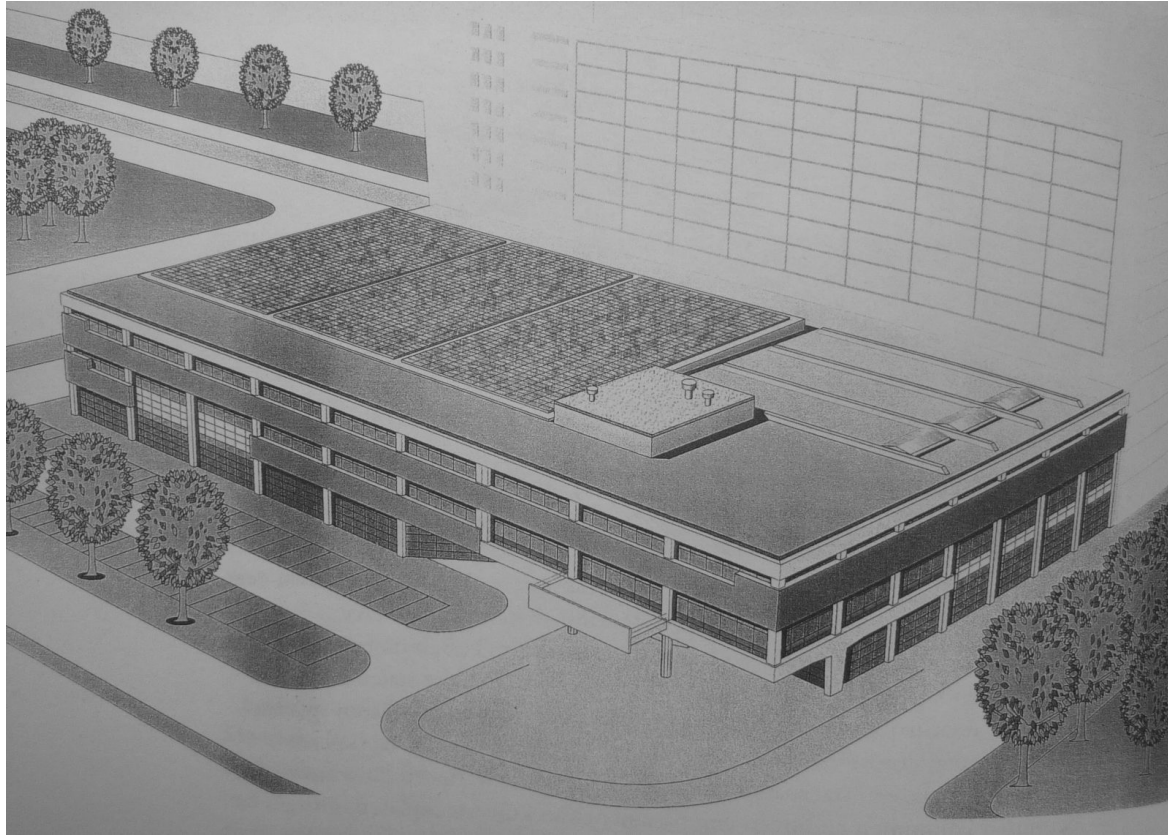
Municipality of Rotterdam built a dedicated archive building in  
1900. The building was destroyed during the Second World War in 1940.  
of the building was reconstructed in 1947.





# 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.

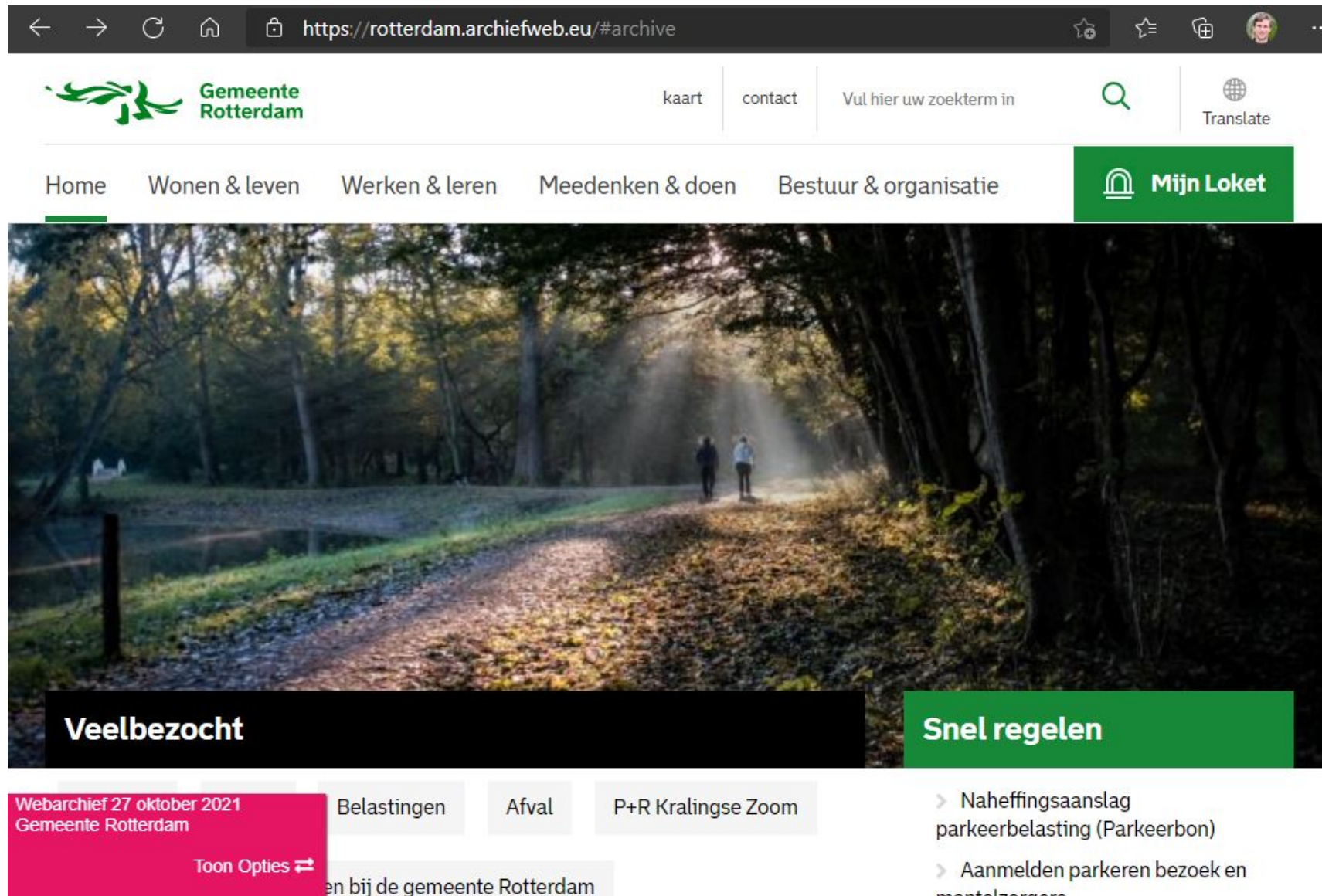
# Municipality websites.

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.





### 3. Website archiving practice 2006 – 2019.


External company: Capsis.


[Home](#) [Archief](#) [Taakplanner](#) [Control panel](#) [Help](#) [Uitloggen](#) capsis Presurf 2.5


**Archief**  
Bekijk gearcheerde websites in het archief

**Taakplanner**  
Plan taken in om websites te archiveren






**Wachtwoord wijzigen**  
Uw systeemwachtwoord wijzigen

**Control panel**  
Gebruikerbeheer, systeemstatus, instellingen

**Help**  
Wat is ... Hoe kan ik ...

**Uitloggen**  
Afmelden bij het systeem

**Laatste 5 snapshots**

	18-02-2020 / 06:16	Vereniging v... (26,65 MB)	Success
	15-02-2020 / 07:19	Rotterdamse ... (254,31 MB)	Success
	14-02-2020 / 11:49	Schiezicht (884,33 MB)	Success
	09-02-2020 / 06:31	Speelcentrum... (2,44 MB)	Success
	08-02-2020 / 06:46	Koor Nieuw Z... (11,03 MB)	Success

**Geen geplande taken gevonden**

Taalkeuze : [Nederlands](#) | [English](#)  
Capsis Presurf - © Capsis B.V. - Alle rechten voorbehouden

# 2006 – 2019 Capsis.

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

2006 – 2019 Capsis.

Small webarchiving department.

# 2006 – 2019 Capsis.

Small webarchiving department.

Created Snapshots, not necessarily WARCs.

# 2006 – 2019 Capsis.

Small webarchiving department.

Created Snapshots, not necessarily WARCs.

Problem handled 2014.

Solution provided by Capsis.

# 2006 – 2019 Capsis.

Small webarchiving department.

Created Snapshots, not necessarily WARCs.

Problem handled 2014.

Solution provided by Capsis.

Problem re-discovered in 2020/2021.

- WARCs do not conform to WARC format standard.
  - Major subsequent problems.



# 2006 – 2019 Capsis.

Small webarchiving department.

Created Snapshots, not necessarily WARCs.

Problem handled 2014.

Solution provided by Capsis.

Problem re-discovered in 2020/2021.

- WARCs do not conform to WARC format standard.
  - Major subsequent problems.

Integrity of some files remains uncertain.

# 2006 – 2019 Capsis.

Problematic, non-functional or partly functional  
WARCs.

Diagnosing the problems is challenging.

Questions:

What is the problem?

How do you find out?

What is the solution?

# Warc analysis.

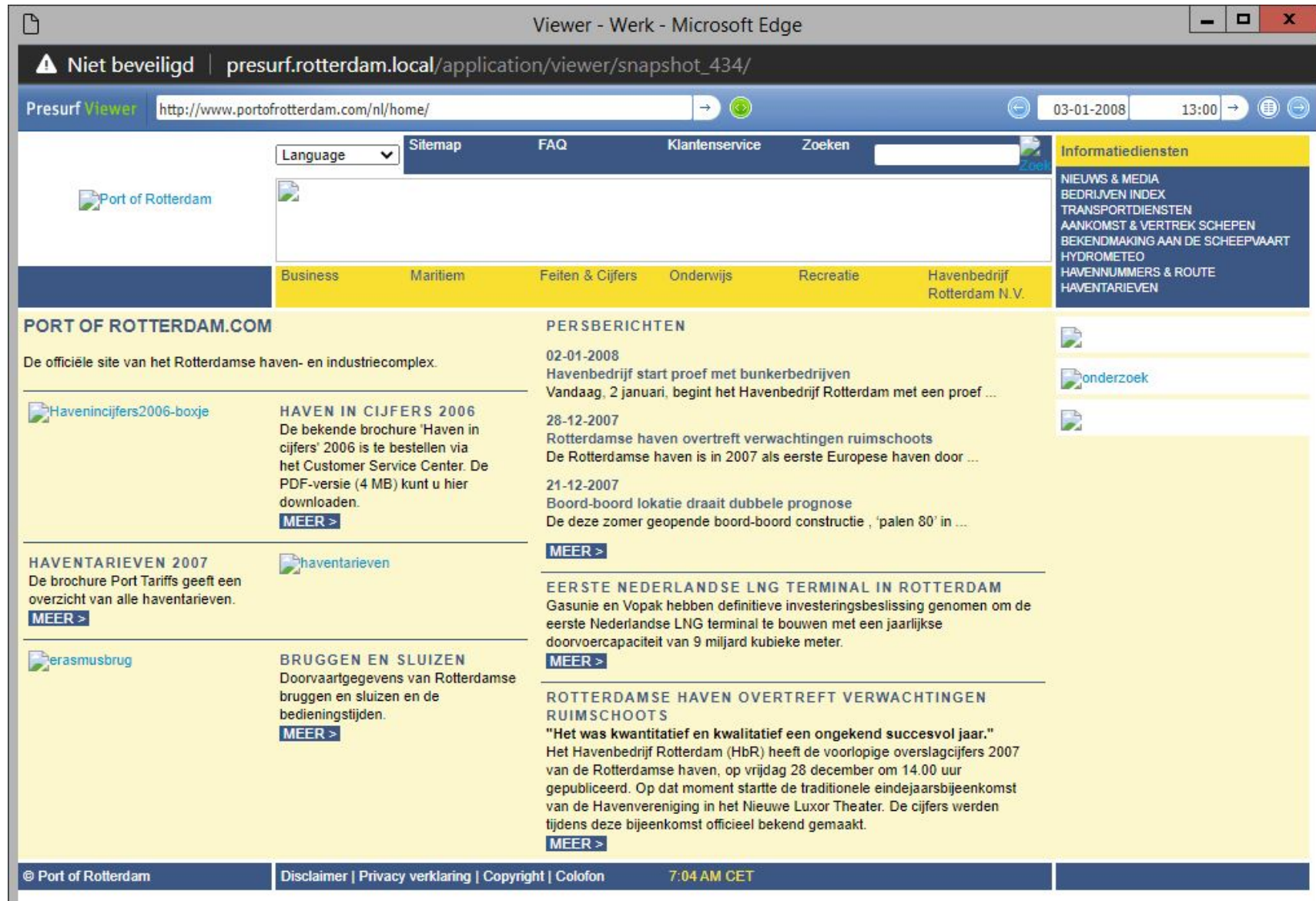
## Quality control tools:

- Notepad.
- DROID.
- JHOVE.
- Warcat.
- Warcio.
- More...

## Third party WARC viewers:

- Replayweb.page.
- Conifer.
- Preservica.

# Capsis PreSurf.



# Capsis PreSurf.

**preview.preservica.com vertelt**

Unable to Render webpage. If this problem persists, please contact your System Administrator.


OK



# Warc analysis.

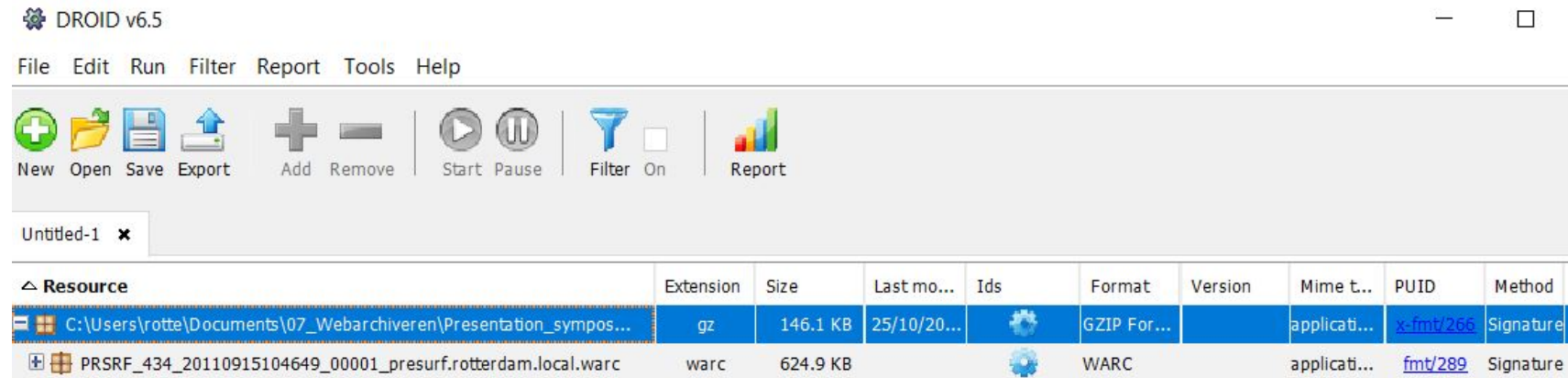
snapshot_431	10-6-2017 07:20
snapshot_433	10-6-2017 07:27
snapshot_434	10-6-2017 07:38
snapshot_435	10-6-2017 07:43
snapshot_436	10-6-2017 07:53

presurf_data	13-5-2017 11:15
snapshot_data	11-5-2015 14:34
warc_data	25-11-2012 16:33

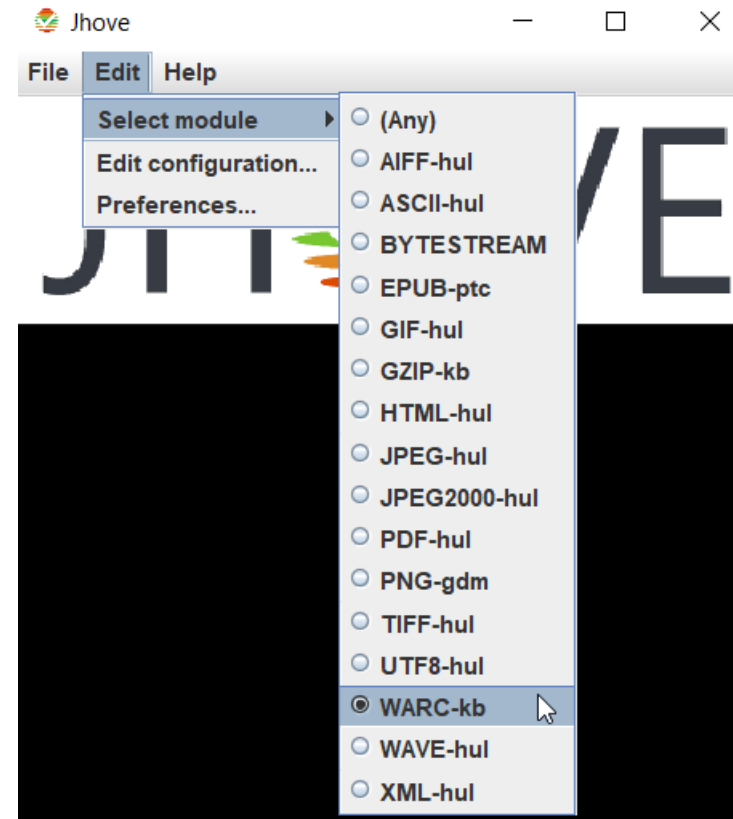
Naam	Gewijzigd op	Type	Grootte
 PRSRF_434_20110915104649_00001_presurf.rotterdam.local.warc.gz	15-9-2011 12:43	7-Zip bestand	147 kB



# Warc analysis.



# Warc analysis.



# Warc analysis.

ReplInfo

File

Documents

- C:\Users\rotte\Documents\07\_Webarchiveren\Presentation\_symposium\WARC's\_for\_presentation\warc\_data\_434 Port of Rotterdam\PRSRF\_434\_20110915104
- Module
- ReplInfo
  - URI: C:\Users\rotte\Documents\07\_Webarchiveren\Presentation\_symposium\WARC's\_for\_presentation\warc\_data\_434 Port of Rotterdam\PRSRF\_434\_20110915104
  - LastModified: Mon Oct 25 16:59:59 CEST 2021
  - Size: 149647
  - Format: WARC
  - Version: 1.0
  - Status: Not well-formed
  - Messages
    - ErrorMessage: INVALID\_EXPECTED
      - ID: NO-ID
      - SubMessage: Entity: 'WARC-Record-ID' value, 4832e9b0-9d7d-43d0-851f-1d2ad16d017d, Missing encapsulating '<' and '>' characters
    - ErrorMessage: INVALID\_EXPECTED
      - ID: NO-ID
      - SubMessage: Entity: 'WARC-Record-ID' value, 4832e9b0-9d7d-43d0-851f-1d2ad16d017d, Invalid URI - relative URIs not allowed: 4832e9b0-9d7d-43d0-851f-1d2ad16d017d
    - ErrorMessage: INVALID\_EXPECTED
    - ErrorMessage: REQUIRED\_INVALID
    - ErrorMessage: REQUIRED\_INVALID
    - ErrorMessage: INVALID\_EXPECTED
  - MimeType: application/warc
  - Records

# Warc analysis.

```
PRSRF_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-1d2ad16d017d
WARC-Date: 2008-01-05T07:01:58+01:00
Content-Type: application/warc-fields
Content-Length: 406
WARC-Filename: PRSRF_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 / NEN-ISO-28500
description:
conformsTo: http://www.nen.nl/web/Normshop/Norm/NENISO-285002009-en.htm
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-1d2ad16d017d
WARC-Payload-Digest: sha1:c27f55e6894c7270bcc9035d2680e6ecbed8010e
WARC-Block-Digest: sha1:c27f55e6894c7270bcc9035d2680e6ecbed8010e
WARC-IP-Address: 10.33.40.13
WARC-Record-ID: <urn:uuid:54b495c7-2bf2-4eaf-ad65-6e78102ffb55>
Content-Type: application/http;msgtype=response
WARC-Identified-Payload-Type: text/html
```

# 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.

Stadsarchief Rotterdam

mei 2021

04

04 mei 2021

Vul hier uw zoekterm in... [geavanceerd zoeken](#)

**CityLab010**

CityLab010 is een ondersteuningsprogramma van de gemeente Rotterdam voor initiatiefnemers - bewoners, ondernemers en organisaties - die een oplossing hebben voor maatschappelijke issues in Rotterdam.

**GROEN LICHT VOOR VRIJE ENERGIE.**

**CITYLAB010 ONDERSTEUNT EN STIMULEERT DE ENERGIE VAN AMBITIEUZE STADSDROMERS, ZODAT HUN DROMEN VOOR EN IN ROTTERDAM REALITEIT WORDEN.**

**BEKIJK ALLE INITIATIEVEN**

**HEB JIJ EEN IDEE VOOR DE STAD?**

**ZO DOE JE MEE**

Stadsarchief Rotterdam

Verberg Opties

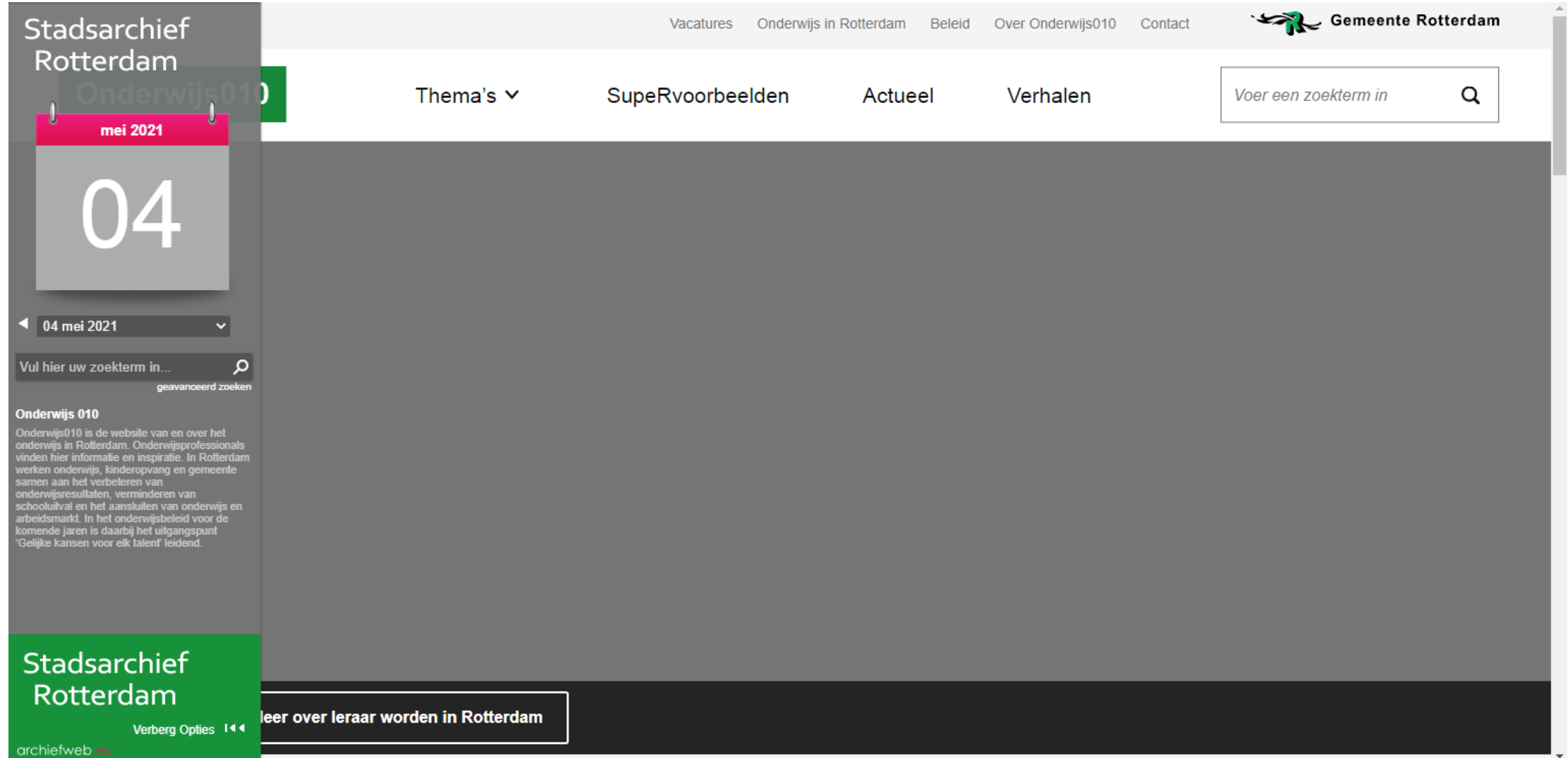
archiefweb

VOOR INITIATIEFNEMERS

INLOGGEN

CONTRAST

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.

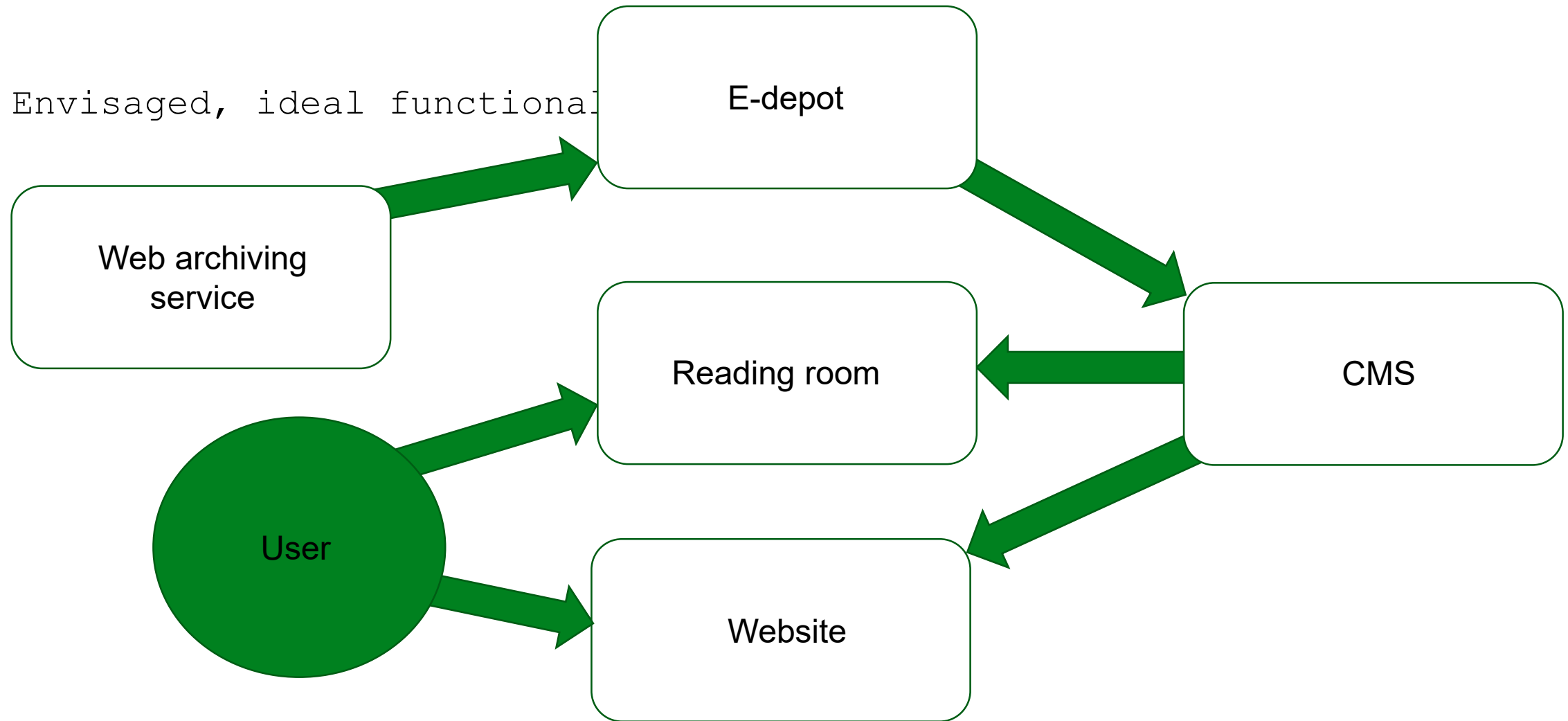
# Public access.

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.



## 5. Summary of website archiving practices at RCA.

Concept:

- Legally required municipality websites.
- Curated selection.

External service provider 2006-2019.

- 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](mailto:a.skyaasen@rotterdam.nl).

---

**Stadsarchief Rotterdam**



**Gemeente  
Rotterdam**