REPUBLIKA HRVATSKA
Središnji državni ured za
razvoj digitalnog društva

Panel: Web archiving in Southeast Europe

Croatian public authorities' web resources archiving system

Tamara Horvat Klemen, *PhD*

Web Archiving Conference

Zagreb, 2019

Central State Office for the Development of the Digital Society

- central state administration body
- one of the tasks:
  - from 2013 - providing proactive disclosure of information, according to the Croatian Freedom of Information Act (Official Gazette 25/2013, 85/2015)
    - permanent storage and central access to the official documents, for the interested public, through **the Central Catalogue of the Official Documents of the Republic of Croatia**
  - before 2013 – providing a legal deposit copy of official publications, according to the Croatian Library Act (1997-2019)

Digital Archive of the Web Resources of the Republic of Croatia

- public authorities' web resources archiving system established in 2004
- in collaboration with the University of Zagreb University Computing Centre (Srce)
- to achieve the new quality in collecting official documentation

Digital Archive of the Web Resources of the Republic of Croatia

- 2004 the growth of the official web pages (public authorities' web resources)

=

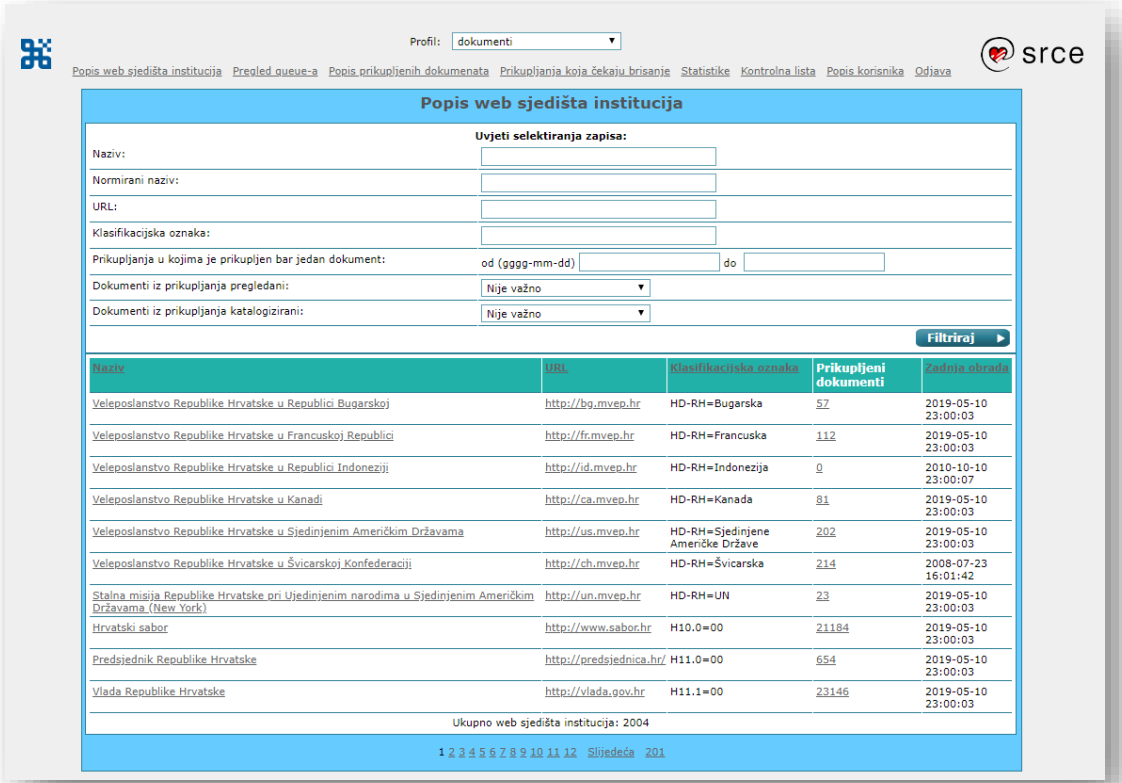
the increase number of the official documents published on the official web sites that should be permanently stored and permanently available to the interested public

	2004	2018
public authorities' web pages	234	1112

Digital Archive of the Web Resources of the Republic of Croatia

Objective: to establish a system for downloading and archiving web-accessible documents from a selected set of web sites

- system for automatic downloading and archiving web-accessible documents
  - gatherer (crawler)
  - data centre
  - web interface for content managing and selection of documents



Digital Archive of the Web Resources of the Republic of Croatia

Profiles:

1. system for automatic gathering and permanent archiving of the public authorities' **web pages** (2006-2013, once a year, depth 3)

   In 2017 - handover it to the Croatian Web Archive in National and University Library in Zagreb

   - 1.540 titles
   - 6.640 copies
   - Each resource has a full level of description and is retrievable in the online catalogue of the National Library.

Hrvatski arhiv weba
Nacionalna i sveučilišna knjižnica u Zagrebu

Croatian Web Archive
National and University Library in Zagreb

Digital Archive of the Web Resources of the Republic of Croatia

2. system of automatic gathering and permanent archiving of the public authorities' documents (from 2004 - …)

- Additional modules:
  - Manual upload of documents
  - Gathering through lists

Each document has a full level of description and is retrievable in the Central Catalogue of the Official Documents of the Republic of Croatia (www.digured.hr)

System of the automatic gathering and the permanent archiving of public authorities' documents

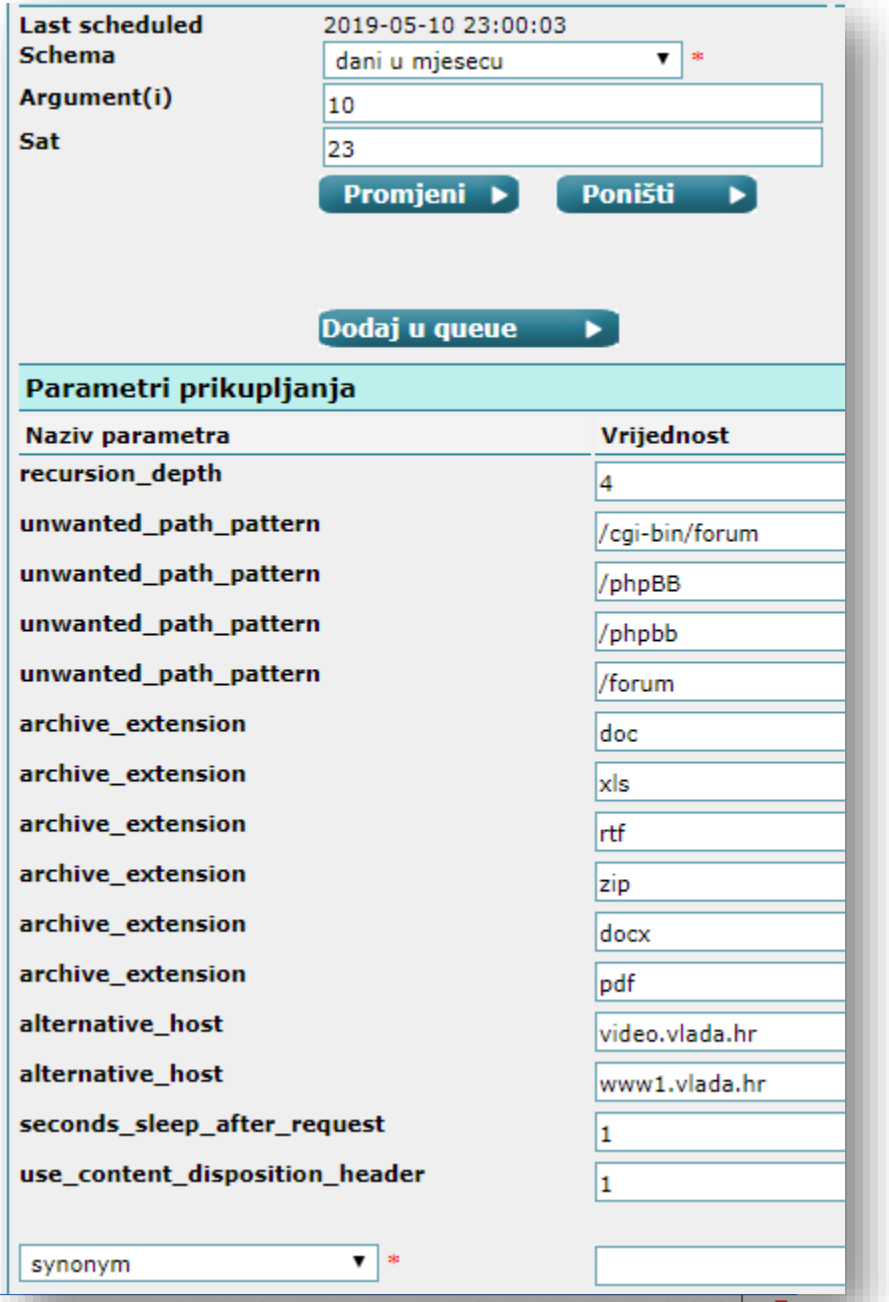
Long-lasting storage of individual documents from the web pages of the public authorities

- Automatic gathering once a month,

(only documents of the selected extensions that have changed compared to the previous gathering)

- Selected documents are linked to the information systems and published in the Central Catalogue of the Official Documents

Possibility of setting the parameters (format, depth…), time and frequency of gathering…



Numbers

	Used space	Free space
Space on disk (Srce, 31/03/2019)	2.171 GB	911 GB

Total number of documents	Total size (30/05/2019)
2.081.275	1.676 GB

	31/03/2019
Total number of web profiles	1.989
Total number of ingathering	173.221
Total number of selected documents	226.326

Initial gathering: November 2004	Regular monthly gathering: March 2019
8.422 documents	16.893 documents

Advantages

- faster processing
- simpler approach
- wider availability
- unified information in one place
- long-term maintenance
- availability even when it no longer exists at the original address
- increased transparency of public authorities

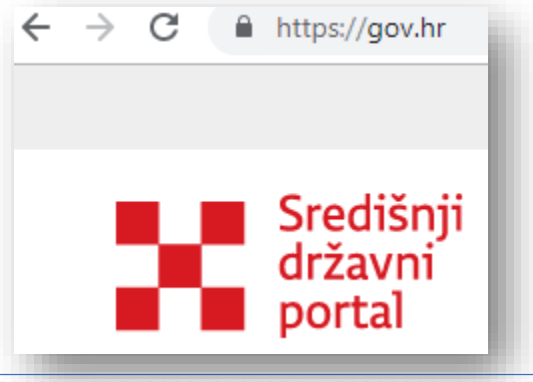
15 years of harvesting Croatian official web resources

Problems encountered:

- collected documents do not comply with the definition of the official document
- insufficient use of metadata by the creator (public authorities)
- non-standard formats for the document publishing
- non-standard web site design solutions - difficult access to the documents

The need for standardization:

- the format and structure of the document and its accompanying elements for its identification
- data display and permanent storage of e-documents
- graphics display and the structure of the web pages of the public authorities



Digital Archive of the Web Resources of the Republic of Croatia

- Before:

It represented the way of collecting content during the period of unordered use of the web technology and the lack of understanding of the consequences of non-compliance with standards

- Now:

The auxiliary tool for automatically retrieving and archiving web-accessible official documents. To ensure the availability of documents of those public authorities that do not fulfill their legal obligation to deliver documents to the Central Catalogue.

Also, the place of the permanent storage of all documents that we are obligated to provide to the general public.

…with the desire to permanently save at least one part of the official documents published online...


REPUBLIKA HRVATSKA
Središnji državni ured za razvoj digitalnog društva
Središnji katalog
Adresari i imenici
Političke stranke i izbori
Naslovnica » Središnji katalog
Središnji katalog službenih dokumenata RH
Pravni propisi
Međunarodni ugovori
Službena glasila tijela lokalne samouprave
Dokumenti i publikacije

Thank you!