



Annex A: Terms of Reference

2023\_RFP\_025

Request for Proposal

For the establishment of frame agreement for the provision of for the provision of  
Web and Social Media Archiving services

## Table of Contents

<b>1</b>	<b>Introduction.....</b>	<b>3</b>
1.1	Who we are.....	3
1.2	Background .....	3
1.3	Statement of Purpose & Objectives .....	3
<b>2</b>	<b>Requirements.....</b>	<b>4</b>
2.1	Core requirements .....	4
2.2	Additional requirements .....	8
2.3	Customer Responsibilities .....	8
<b>3</b>	<b>Content of the Technical Offer.....</b>	<b>9</b>
3.1	Company Qualifications.....	9
3.2	Proposed Services .....	9
3.3	Personnel Qualifications.....	9
3.4	Vendor Registration Form .....	10
3.5	Applicable General Conditions .....	10
3.6	Technical Evaluation .....	10
3.7	Financial Evaluation .....	10
<b>4</b>	<b>Key Performance Indicators .....</b>	<b>11</b>
4.1	Performance Evaluation .....	11

---

## **1 Introduction**

---

### **1.1 Who we are**

The Office of the United Nations High Commissioner for Refugees (UNHCR) was established by the U.N. General Assembly in 1950 to provide protection and assistance to refugees and Internally Displaced People (IDPs). In more than seven decades, the agency has helped tens of millions of people to restart their lives. Today, UNHCR is one of the world's principal humanitarian agencies; its staff of more than 18,500 personnel is helping more than 89 million forcibly displaced people in more than 135 countries. Staff members work in a diversity of locations and conditions including in our Geneva-based Headquarters (HQ). 91 percent of UNHCR staff are based in the field, assisting the most vulnerable victims of displacement. For more information, please see <https://www.unhcr.org>.

---

### **1.2 Background**

The Internet is one of the most important sources of information in UNHCR. It is, and will be, a key resource for the organization, now and in the future. Web resources are a crucial part of UNHCR's public information and play an important part in the mission of the organization holding information as well as multimedia content of both current and long-term value to the organization and its partners. Web pages are constantly changing: the average lifespan of a webpage is between 75 and 100 days. To make sure this content survives, it must be captured in real-time. It is the Records and Archives Section's responsibility to preserve UNHCR's digital web heritage and guarantee access to it in the short- and long-term.

---

### **1.3 Statement of Purpose & Objectives**

The purpose of the UNHCR web-archive is to enable collection, preservation, and long-term access to a rich body of internet content and metadata from UNHCR's web-estates, both for operational purposes and as a record of UNHCR's activities.

Archiving internet content requires special attention to maintain its value as a reliable and authentic source of information about UNHCR's operations and activities. Challenges come from the wide variety of features and possibilities of the web: dynamic content, a wide variety of formats and internet user contributions, changing business models of website and social media platform owners, all are factors that contribute to the internet being a volatile and fast-changing medium.

Data and metadata will be stored by the provider in WARC format (ISO 28500:2017) to ensure long-term preservation. UNHCR archives will link directly to UNHCR-branded pages where the archived sites will be made available for public access and hosted by the service provider. There will be a full-text index for search within the archived data. Captures and crawls of content should be carried out on a regular basis in order to ensure that selected content remains accessible to staff and researchers over time and irrespective of changes to websites and social media platforms.

UNHCR will enter into a contract with a provider for the capture and preservation of a selection of UNHCR's publicly available web pages and social media accounts. The successful bidder is expected to be awarded a frame agreement for a duration of 3 years, with an extension for a maximum of two periods of one year each (3+1+1 years). The awarded service provider will be requested to maintain the proposed pricing for the entire duration of the contract.

## 2 Requirements

### 2.1 Core requirements

#### 1. Organizational requirements

A. The Supplier will demonstrate its experience working in collaboration with public nonprofit organizations (NPOs), providing references.
B. The Supplier should demonstrate research and development activities in the Web Archiving field, which may include membership of the International Internet Preservation Consortium (IIPC).
C. The supplier will demonstrate their commitment to sustainability and reducing their carbon footprint in both their product delivery and working practices.

#### 2. Technical requirements

##### I. Capturing data and remote harvesting

- A. The Supplier will provide for the capture of periodic snapshots of the UNHCR web estate. UNHCR currently performs 62 website crawls, and 33 social media account captures on a yearly recurring basis, spread over 3 key dates in the UNHCR calendar and consisting of core sites from UNHCR's web estate.

UNHCR may also require captures in addition to the usual yearly crawls. These would be performed on an ad hoc basis, usually in response to emergencies and other world events of concern to UNHCR and as these events unfold. Extra crawls may also be required in order to create a snapshot of a website before it is taken offline. The supplier should be able include these extra crawls, at short notice if needed, and should provide a pricing list for ad hoc crawls.

Please note that UNHCR websites are very different in sizes and can be in a wide range of languages and scripts/alphabets. Suppliers will be able to capture and replay all websites and social media accounts, regardless of language or script (e.g. Latin, Cyrillic, Kanji, etc.). Languages include but are not limited to; English; French; Spanish; Arabic; Russian; Chinese.

Examples of UNHCR websites and their sizes are:

Title	URL	Approximate current size
UNHCR English website	<a href="https://www.unhcr.org/">https://www.unhcr.org/</a>	110 GB
UNHCR Spanish website	<a href="https://www.acnur.org/">https://www.acnur.org/</a>	13.5 GB
Global Protection Cluster	<a href="https://www.globalprotectioncluster.org/">https://www.globalprotectioncluster.org/</a>	6.1 GB
UNHCR Sri Lanka website	<a href="https://www.unhcr.org/en-lk/">https://www.unhcr.org/en-lk/</a>	12.4 GB
UNHCR Serbia	<a href="https://www.unhcr.org/rs/">https://www.unhcr.org/rs/</a>	546 MB
UNHCR Indonesia website	<a href="http://www.unhcr.org/id/">http://www.unhcr.org/id/</a>	207 MB
UNHCR Belgium and Luxembourg website	<a href="https://www.unhcr.org/be/">https://www.unhcr.org/be/</a>	4.3 GB
UNHCR Cyprus website	<a href="https://www.unhcr.org/cy/">https://www.unhcr.org/cy/</a>	2.5 GB
UNHCR Regional Representation for Northern Europe website	<a href="https://www.unhcr.org/neu/">https://www.unhcr.org/neu/</a>	3.7 GB

UNHCR Canada website	<a href="https://www.unhcr.ca/">https://www.unhcr.ca/</a>	7.4 GB
UNHCR Malta website	<a href="https://www.unhcr.org/mt/">https://www.unhcr.org/mt/</a>	437 MB
UNHCR Emergency Handbook	<a href="https://emergency.unhcr.org/">https://emergency.unhcr.org/</a>	1.1 GB
UNHCR Hong Kong website	<a href="https://www.unhcr.org/hk/">https://www.unhcr.org/hk/</a>	2.8 GB
UNHCR Ukraine website (Ukrainian version)	<a href="https://unhcr.org/ua/">https://unhcr.org/ua/</a>	2 GB
Regional Safe Spaces Network	<a href="https://rssn-americas.org">https://rssn-americas.org</a>	269 MB

  

B. As part of the RFP process, all bidding suppliers are required to provide test crawls of the following website sections:

- <https://data.unhcr.org/en/situations/ukraine>
- <https://www.unhcr.org/ar/about-unhcr>
- <https://emergency.unhcr.org/protection>

and the following social media accounts:

- <https://twitter.com/Refugees>
- <https://www.facebook.com/UNHCR/>
- <https://www.youtube.com/UNHCR>

After the tender submission deadline, the test crawls will be pre-planned with the suppliers so that only one crawl is done at any given time. The Records and Archives Section will act as focal point and will nominate one person who will co-ordinate the crawls with the suppliers. The crawls will be evaluated based on completeness of the crawl. Suppliers are required to set up a temporary portal and provide the results to UNHCR for review.

C. Crawls will be domain or path level, as appropriate, and they are expected to be complete.

D. The required method of capture for websites will be by remote harvesting. Social media accounts should be captured remotely as well, without requiring authentication from the account owner where possible.

E. The politeness setting for the crawler will be agreed prior to crawling.

F. Any archive crawler will identify itself using a UNHCR Records and Archives agent. E.g., using [archives@unhcr.org](mailto:archives@unhcr.org)

G. The Supplier will generate metadata for each snapshot and for each digital object derived from a URL. The minimum information required for snapshot and WARC file is:

- Snapshot name
- Seed URL
- Time and date of capture
- Size (no of objects and volume)
- Content type (i.e. format)
- Object (e.g., ARC or WARC file)
- URL retrieved
- It must be clear which objects relate to which snapshots
- Response code
- HTTP response header

H. UNHCR will determine the scheduling of the regular crawls. UNHCR will communicate crawl lists and requirements to the supplier in advance of the crawls. Extra crawls may be required in response to world events affecting UNHCR and the supplier should be able to accommodate these extra crawls as needed.

I. The Supplier will use their own crawler and access tools in combination with open-source tools (such as Heritrix, Conifer, etc.)

J. The Supplier will demonstrate its capacity to archive social media. The minimum required is the capacity of archiving UNHCR Twitter and YouTube accounts which are already included in the

agency's Web Archive. UNHCR also wishes to include social media other than Twitter in its archive, in particular Facebook, Instagram and LinkedIn, or other platforms in future. Consequently, the Supplier will have to demonstrate its capacity to archive these social media, and/or its R&D activities in this field.

## **II. Quality Assurance (QA) requirements**

- A. The Supplier will provide quality assurance for all harvested content. Issues will be identified within 2 weeks of crawl by the supplier. The Supplier will agree a resolution process with UNHCR. Once the Supplier has undertaken QA, UNHCR will then undertake its own QA. Subject to satisfactory resolution of all identified issues, UNHCR will subsequently sign off each snapshot.
- B. Outside the formal QA process, the Supplier will continue to resolve further issues which may be identified at any time. Further issues may involve broken links, provision of video and audio content which may require manual intervention to capture.
- C. The Supplier will provide a tracking system to perform the QA, preferably via online portal.

## **III. Access and hosting**

- A. Once content has passed the quality assurance process, the Supplier will host the content for public access for the lifetime of the contract. The content will be made available online, on a customized and UNHCR branded access portal, via a link from the UNHCR Records and Archives website. The supplier should be able to provide the access portal interface in different languages (ideally English, French, Arabic, Russian, Chinese, Spanish) or demonstrate willingness to develop this feature if requested.
- B. The Supplier is required to ensure a minimum of 99.9 % uptime for the online access system. The Supplier must provide a process to notify UNHCR immediately of any scheduled downtime and to resolve the issue and restore access.
- C. The content will be hosted under UNHCR's domain: <https://webarchive.archive.unhcr.org/>
- D. The content should be accessible via a URL of the form: <https://webarchive.archive.unhcr.org/datestamp/original> URL
- E. The Supplier will be required to host the existing web archive information from all previous crawls. The current size of the UNHCR Web Archive is around 7 TB. At the end of the contract all archive content will be transferred by the Supplier to the new provider for the Web Archiving service. The Supplier is required to describe in its offer its takeover and handover strategies. The supplier will also be expected to transfer content to UNHCR for long-term preservation at agreed points during the contract.
- F. The Supplier will host the archive in an approved web archiving preservation format (ISO 28500:2017 WARC) in a secure data repository. At least two copies of the harvested content must be held, in geographically separate locations. Appropriate business continuity and disaster recovery facilities must be in place and regularly tested. The supplier will also be expected to meet UNHCR data security requirements and will be subject to review by UNHCR IT Security and Architecture Review Board.
- G. The Supplier will demonstrate capacity and scalability in terms of size e.g., they should have the ability to archive Terabytes of data per month
- H. The Supplier will provide full index search and retrieval for all harvested content. As a minimum, this should include all content in HTML, XML, PDF, RTF, DOC and DOCX. It should be possible to search across collections as well as limit searches to specific snapshots.
- I. The Supplier must offer a mechanism for the takedown of web archive content from public view, at the request of UNHCR. This might be either a complete snapshot, or pages within a snapshot. While the content must be removed from public view/access, it must not be removed from the underlying ARC or WARC files.
- J. The online access system should dynamically rewrite all hyperlinks present in archived content, to point to the relevant archived version of the link target.
- K. Suppliers will provide an architecture diagram for how the content will be stored, replayed and

backed up.

#### IV. Preservation requirements

- A) The web content captured through the service will form part of the permanent collection of the UNHCR. The service must support its long-term preservation. The harvested content will be preserved in WARC (ISO 28500:2017) format. The Supplier will provide for the export of all harvested content, and associated metadata, in WARC format. The Supplier should ensure the sustainability of the archive.
- B) Each WARC file must relate to only one website snapshot, including WARCs created during patch and fixing, however, a crawl of one website may consist of multiple WARCs. The filename, or other metadata, must clearly identify each WARC with its source seed URL (i.e., the URL of the website being harvested) and crawl date. There must be no “orphan” files, every WARC must be linked to its source website.
- C) The Supplier will provide UNHCR with the possibility of carrying out at any time an export of all or part of the archived content and metadata, hosted by the Supplier. The Supplier will describe how this export will be facilitated.

#### 3. Support service and reporting

- A) The Supplier will provide technical support during UNHCR working hours (Monday to Friday, 09:00 – 18:00 CET)
  - Timely response: 100% within 5 working days
  - Fix time for queries/issues: To be specified by the Supplier.
- B) The Supplier will provide project management and technical support with a dedicated team for UNHCR's Web Archiving project. This team will include developers, crawl engineers, quality assurance and system administrators.
- C) The Supplier will attend service review meetings (monthly and annual, either in person or via teleconferencing).
- D) The Supplier will provide a mechanism for reporting on the progress of the crawl and will guarantee that UNHCR is kept informed of current crawl status
- E) For each issue affecting crawl quality, reporting must include:
  - Nature of the issue
  - Whether it relates to the crawl process or to the live site
  - Advice on changes to the live site which will solve the issue
  - Whether a fix to the crawl is possible in the future
- F) The Supplier will provide:
  - annual user statistics reports,
  - annual downtime statistics,
  - annual hosted data volumes
- G) The Supplier will provide scheduled downtime notification a minimum of two weeks in advance.
- H) The Supplier will report immediately to UNHCR any unscheduled downtime, and details of these type of incidents will be included in the annual report.
- I) All communications between the supplier and UNHCR will be in English.

J) The Supplier will guarantee transparency regarding costs, providing UNHCR with tools for monitoring cost.
K) The Supplier will support UNHCR with adequate guidance and training (i.e., guidelines on how to initiate the crawl or to use management portals)
L) The Supplier's approach to problems will be proactive instead of reactive. The Supplier will guarantee transparency regarding any issue that might arise in the provision of the Web Archiving service.
M) The Supplier will demonstrate commitment to continuous improvement.

---

## **2.2 Additional requirements**

---

A. The Supplier will provide a service which enables UNHCR to archive dynamic content from a target website, and/or will demonstrate their commitment to development in the field of archiving dynamic content.
B. The Supplier will demonstrate their commitment to development in the field of archiving websites, social media and other web-based content.
C. At one point in the future, UNHCR may wish to implement a redirection service, to redirect users to the archived version of a removed page instead of a 404-error message. In the event that UNHCR chooses to implement this option during the lifetime of the contract, the Supplier will be able to support this redirection.
D. The Supplier might be requested to directly liaise with webmasters to discuss problems that might arise during the archiving process and improve crawl quality.

---

## **2.3 Customer Responsibilities**

---

UNHCR Records and Archives Section (RAS) will provide project management for the Web Archiving project.

UNHCR RAS will undertake its own quality assurance (QA) on the harvested content. UNHCR RAS will aim to conduct its QA within two weeks.

Where needed, UNHCR RAS will facilitate the communication between the Supplier and other UNHCR Divisions and Sections managing the UNHCR web estate, for the purpose of improving the crawl quality.



---

### **3 Content of the Technical Offer**

Your technical proposal should be concisely presented and structured in the following order to include, but not necessarily be limited to, the following information:

---

#### **3.1 Company Qualifications**

- A description of your company with evidence of your company's capacity to perform the services required, including company profile, registration certificate, security certificates, and last audit reports
- Year founded
- If a multi-location company, please specify the location of the company's headquarters, and the branches that will be involved in the project work with founding dates.
- Experience in the business or the number of similar and successfully completed projects
- Any relevant experience working with UNHCR, other UN agencies and NGOs should be included. Please specify the number of years working with public NPOs.
- Number of similar projects currently underway. Please elaborate on these projects and provide proof.
- Number of research and development projects in the web archiving field.
- Total number of clients. Please provide a list.
- You are requested to provide three (3) references that we may contact from your current client list

Provide any other information that will facilitate our evaluation of your company's substantive reliability, financial and managerial capacity to provide these services. The bidders should demonstrate an in-depth understanding of the project, from an operational and strategic point of view.

---

#### **3.2 Proposed Services**

Please submit a written proposal outlining:

- How you intend to develop the above-mentioned services, as per requirements stated in section 2.1 and 2.2),

As mentioned in point 2.1.B of the Requirements and as part of the technical evaluation, all bidding suppliers are required to perform test crawl of some UNHCR websites.

After the tender submission deadline, the crawls will be pre-planned with the suppliers so that only one crawl is done at any given time. The Records and Archives Section will act as focal point and will nominate one person who will co-ordinate the crawls with the suppliers. The crawls will be evaluated based on completeness of the crawl. Suppliers are required to set up a temporary portal and provide the results to UNHCR for review.

Shortly after the tender submission deadline suppliers will be contacted by UNHCR to arrange for the test crawls. Bidders are requested to provide contact details for the focal point in their organization for the crawls.

All bidding suppliers will be requested to identify themselves before performing the test crawls, to avoid the harvesting being blocked due to security concerns. The supplier can email [hgepu@unhcr.org](mailto:hgepu@unhcr.org) for assistance and to identify themselves prior to test crawls.

---

#### **3.3 Personnel Qualifications**

Please provide CVs of personnel who would be involved in the project. Please note that bidders will not be assessed on this as a criterion, it is required for informative purposes.

### 3.4 Vendor Registration Form

If your company is not already registered with UNHCR, please complete, sign, and submit the Vendor Registration Form (Annex C) with your Technical Proposal.

### 3.5 Applicable General Conditions

Please indicate your acknowledgement of the UNHCR General Conditions of Contract for the Provision of Services by signing this document (Annex D) and including it in your submitted Technical Proposal.

### 3.6 Technical Evaluation

The **Technical offer** will be evaluated using inter alia the following criteria and percentage distribution: **70%** from the total score.

Criteria	Sub-criteria	Weight
1) Organizational requirements (Please refer to point 1 of the section "2.1 Core requirements")	1.1. Demonstrated experience working in collaboration with public NPOs.	3
	1.2. Research and development projects in the Web Archiving field	4
2) Technical Requirements (Please refer to point 2 of the section "2.1. Core requirements")	2.1. Proposed solution for the capture of data and remote harvesting, including the results of the test crawls.	11
	2.2. Proposed solution for the Quality Assurance process.	9
	2.3. Proposed solution for access and hosting, including the takeover and handover strategy	9
	2.4. Proposed architecture, backup, and preservation solution	9
3) Support service and reporting (Please refer to point 3 of the section "2.1. Core requirements")	3.1. Proposed solution for technical support and project management	6
	3.2. Proposed solution for the provision of training and guidelines	6
	3.3. Proposed solution for the provision of exhaustive reports	6
4) Additional services (Please refer to point 4 of the section "2.1. Core requirements")	4.1. Proposed solution for archiving social networks and dynamic contents, or demonstrated R&D activities in these fields	5
	4.2 Demonstrated capability of liaising with webmasters and web developers to implement web archiving solutions	2

The minimum passing score is 46 of 70 points; if a bid does not meet this minimum, it will be deemed technically non-compliant and will not proceed to the financial evaluation.

### 3.7 Financial Evaluation

The Financial offer will be evaluated using the Financial Offer form attached under Annex B and the percentage distribution will be **30%** from the total score.

## **4 Key Performance Indicators**

---

### **4.1 Performance Evaluation**

---

UNHCR expects to monitor the performance of the selected Supplier.

UNHCR will consider a performance satisfactory if it meets the following criteria:

- Completeness of captures
- Harvested content is stored in WARC format in a secure data repository
- The archived content is hosted in a customized online portal providing a high-performance full text search
- The up time of the system is 99,9%
- The reports are exhaustive
- Efficient and transparent communication