



Annex A: Terms of Reference

RFP/2018/1062

Request for Proposal

For the establishment of frame agreement for the provision of

Web Archiving Solution for UNHCR

Table of Contents

1	Introduction.....	4
1.1	Background	4
1.2	Statement of Purpose & Objectives.....	4
2	Requirements.....	6
2.1	Core requirement.....	6
2.2	Additional requirements.....	14
2.3	Customer Responsibilities.....	14
3	Content of the Technical Offer.....	15
3.1	Company Qualifications	15
3.2	Proposed Services.....	15
3.3	Personnel Qualifications	15
3.4	Vendor Registration Form.....	16
3.5	Applicable General Conditions	16
4	Evaluation	17
4.1	Technical Evaluation.....	17
4.2	Financial Evaluation.....	17
5	Key Performance Indicators	18
5.1	Performance Evaluation.....	18

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 six decades, the agency has helped tens of millions people to restart their lives. Today, UNHCR is one of the world's principal humanitarian agencies; its staff of more than 10,000 personnel is helping more than 65.3 million people in more than 128 countries. Staff members work in a diversity of locations and conditions including in our Geneva-based Headquarters (HQ). A massive 87 per cent of staff are based in the field, assisting the most vulnerable victims of displacement. For more information, please see <http://www.unhcr.org>.

1.2 Background

The Internet is one of the most important sources of information in UNHCR and 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 value to the organization and its partners, both for the present and in the near and long term. 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 for the near and long term future.

1.3 Statement of Purpose & Objectives

The purpose of this project is to enable the collection, preservation and long-term access of a rich body of Internet content and metadata from UNHCR for operational purposes and record of activity.

Archiving the Web requires a special attention in order to retain its value and ensure its greater fidelity.

Challenges come from the feature and richness of the Web: dynamics, volatility, variety of formats and Internet user contributions, all attributes are increasingly used.

UNHCR will enter into a contract with a provider for the harvesting of a selection of UNHCR's publicly available web pages that should be carried out on a regular basis in order to ensure that content and changes are archived adequately for ease of access by staff and researchers now and over time.

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.

Data and metadata will be stored by the provider in (ISO 28500:2009) WARC format to ensure long term preservation. UNHCR archives will link directly to the UNHCR branded pages where the archived sites will be made available for access and hosted by the service provider. There will be a full text index for search within it.

2 Requirements

2.1 Core requirements

1. Organizational requirements

A. The Supplier will demonstrate its experience working in collaboration with public non for profit organizations, providing references.
B. The Supplier should demonstrate research and development activities in the Web Archiving field, which may include membership in the International Internet Preservation Consortium.

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. This comprises the following websites:

Websites to be captured twice per year		
Title	URL	Main Language
UNHCR English website	http://www.unhcr.org/	English
UNHCR Spanish website	http://www.acnur.org/	Spanish
UNHCR Arabic website	http://www.unhcr.org/ar/	Arabic
UNHCR French website	http://www.unhcr.org/fr/	French
Innovation website	http://www.unhcr.org/innovation/	English
Data portal	http://data2.unhcr.org/en/situations	English
Reporting	http://reporting.unhcr.org	English
Global Camp Coordination and Camp Management Cluster	http://www.globalccmcluster.org/	English
Global Shelter cluster	https://www.sheltercluster.org/	English
Global Protection Cluster	http://www.globalprotectioncluster.org/	English
CRRF Global Digital Portal	crrf.unhcr.org	English
Websites to be captured once per year		

Title	URL	Main Language
UNHCR Niger blog: Life as a refugee in Niger	http://unhcrniger.tumblr.com/	English/French
Kora. Voices of Refugees in West and Central Africa	http://kora.unhcr.org	English
Help	http://help.unhcr.org/	English
Emergency	https://emergency.unhcr.org	English
Country website to capture once per year		
Title	URL	Main Language
UNHCR Sri Lanka website	http://unhcr.lk/	English
UNHCR Serbia	http://www.unhcr.rs/	Serbian
UNHCR Indonesia website	http://www.unhcr.org/id/	Indonesian/English
UNHCR Regional Representation for Central Europe website	http://www.unhcr.org/ceu/	English
UNHCR Regional Representation for Central Asia website	http://www.unhcr.kz/	Russian
UNHCR Belgium and Luxembourg website	http://www.unhcr.be/	French
UNHCR Ukraine website	http://unhcr.org.ua/uk/	Russian
UNHCR Netherlands website	http://www.unhcr.org/nl/	Dutch
UNHCR Turkey Website	http://www.unhcr.org/tr/	English/Turkish
UNHCR Germany website	http://www.unhcr.org/dach/de/	German
UNHCR Cyprus website	http://www.unhcr.org.cy	Turkish/Greece
UNHCR Japan website	http://www.unhcr.org/jp/	Japanese
UNHCR Austria website	http://www.unhcr.org/dach/at/	German
UNHCR Regional Representation for Northern Europe website	http://www.unhcr.org/neu/	English
UNHCR Canada website	http://www.unhcr.ca/	English
UNHCR Hong Kong website	http://www.unhcr.org/hk/	Chinese

Website of the Representative Office of the UNHCR in the Russian Federation	http://www.unhcr.ru/	Russian
UNHCR Thailand website	https://www.unhcr.or.th/	Thai/English
UNHCR Spain website	http://www.acnur.es/	Spanish
UNHCR Italy Website	http://www.unhcr.it/	Italian
UNHCR Malta website	http://www.unhcr.org.mt/	English
UNHCR Pakistan website	http://unhcrpk.org/	English
UNHCR Brazil website www.acnur.org/portugues	http://www.acnur.org/portugues /	Portuguese
UNHCR South Korea website	http://www.unhcr.or.kr/unhcr/main/index.jsp	Korean
UNHCR Greece Website	http://www.unhcr.gr/	Greek
UNHCR Australia Website	http://www.unhcr.org/en-au/	English
UNHCR United Kingdom website	http://unhcr.org/uk/	English
UNHCR Ireland website	http://www.unhcr.org/en-ie/	English
UNHCR Switzerland website	http://www.unhcr.org/dach/ch-de/	German
UNHCR Switzerland (in French)	http://www.unhcr.org/dach/ch-fr/	French
UNHCR Jordan	http://www.unhcr.jo	
UNHCR Kenya	http://www.unhcr.org/ke/	English
UNHCR Rwanda	http://www.unhcr.org/rw/	English
UNHCR Philippines	http://www.unhcr.org/ph/	English
UNHCR Lebanon	http://www.unhcr.org/lb/	English
UNHCR Syria	http://www.unhcr.org/sy/	English
Social Media account to capture twice per year		
UNHCR Official Twitter account	https://twitter.com/refugees	
Social Media account to capture once per year		
Twitter		
UNHCR Official Twitter account in Arabic	http://twitter.com/UNHCR_Arabic	
Official UNHCR Twitter account in French	https://twitter.com/Le_HCR	
Melissa Fleming Twitter account	https://twitter.com/melissarfleming	

Filippo Grandi Twitter account	https://twitter.com/RefugeesChief
Refugee media	https://twitter.com/refugeesmedia
U Refugee gender twitter account	https://twitter.com/refugee_gender
UNHCR Egypt Twitter account	https://twitter.com/unhcregypt
UNHCR Lebanon twitter account	https://twitter.com/unhcrlebanon
UNHCR Greece Twitter account	https://twitter.com/unhcrgreece
UNHCR Regional Representation for Central Europe Twitter account	https://twitter.com/refugeesce
UNHCR Syria Twitter Account	https://twitter.com/unhcrinsyria
UNHCR Jordan	https://twitter.com/unhcrjordan
UNHCR Zaatari	https://twitter.com/zaataricamp
UNHCR Somalia Twitter account	https://twitter.com/unhcrsom
UNHCR Kenya Twitter account	https://twitter.com/unhcr_kenya?lang=it
UNHCR Cameroun Twitter account	https://twitter.com/RefugeesCmr
Twitter Account for the Regional Representation for West Africa	https://twitter.com/unhcrwestafrica
Twitter account of the Regional Office for Southern Africa (Rosa)	https://twitter.com/unhcrrosa
UNHCR Yemen Twitter account	https://twitter.com/unhcryemen
UNHCR Rwanda Twitter account	https://twitter.com/refugeesrwanda
UNHCR Lybia Twitter account	https://twitter.com/unhcrlibya
UNHCR Afghanistan Twitter account	https://twitter.com/UNHCRAfg
UNHCR Bangladesh Twitter account	https://twitter.com/UNHCR_BGD
UNHCR Chad Twitter account	twitter.com/unhcrTchad
UNHCR Uganda Twitter account	https://twitter.com/UNHCRuganda
UNHCR Nigeria Twitter account	https://twitter.com/unhcrnigeria
Facebook	
UNHCR DRC Facebook page	https://www.facebook.com/UNHCR-République-Démocratique-du-Congo-111965425530257/
UNHCR South Sudan Facebook page	https://www.facebook.com/SouthSudanUNHCR
UNHCR Mali Facebook page	https://it-it.facebook.com/UNHCR-Mali-1433966120156823/
YouTube	
UNHCR Youtube channel	https://www.youtube.com/user/unhcr

The estate may change due to unforeseen emergencies which may arise and for which new websites may be created or existing websites not on the list may be updated and added to the list.

Please note that UNHCR websites are very different in sizes. Here are some examples:

Title	URL	Approximate current size
UNHCR English website	http://www.unhcr.org/	221 GB
UNHCR Spanish website	http://www.acnur.org/	30 GB

Global Protection Cluster	http://www.globalprotectioncluster.org/	4.25 GB
UNHCR Niger blog: Life as a refugee in Niger	http://unhcrniger.tumblr.com/	375 MB
Kora. Voices of Refugees in West and Central Africa	http://kora.unhcr.org	464 MB
UNHCR Sri Lanka website	http://unhcr.lk/	68 MB
UNHCR Serbia	http://www.unhcr.rs/	89 MB
UNHCR Indonesia website	http://www.unhcr.org/id/	110 MB
UNHCR Belgium and Luxembourg website	http://www.unhcr.be/	231 MB
UNHCR Cyprus website	http://www.unhcr.org.cy	511 MB
UNHCR Regional Representation for Northern Europe website	http://www.unhcr.org/neu/	872 MB
UNHCR Canada website	http://www.unhcr.ca/	1.4 GB
UNHCR Malta website	http://www.unhcr.org.mt/	2.6 GB
UNHCR Pakistan website	http://unhcrpk.org/	2.9 GB
UNHCR South Korea website	http://www.unhcr.or.kr/unhcr/main/index.jsp	4.2 GB
UNHCR United Kingdom website	http://unhcr.org/uk/	32 GB
UNHCR Ireland website	http://www.unhcr.org/en-ie/	33 GB

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

- <https://data2.unhcr.org/en/situations>
- <http://www.unhcr.org/ar/>
- <http://www.acnur.org/>
- <https://twitter.com/refugees>

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.

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

D. The required method of capture will be by remote harvesting.
E. The politeness setting for the crawler will be agreed prior to crawling.
F. Archive crawler will identify itself using UNHCR Records and Archives agent. E.g. using 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: <ul style="list-style-type: none"> • 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. The schedule of crawls will be determined by UNHCR. Crawls may be added/alterd in response to particular events. The Supplier will be required to adjust its service to allow for any changes. The estate may change due to unforeseen emergencies which may arise and for which new websites may be created or existing websites not on the list may be updated and added to the list.
I. The Supplier will use their own crawler and access tools in combination with open source tools (such as for example Heritrix, Webrecorder)
J. The Supplier will demonstrate its capacity to Archive social media. The minimum required is the capacity of archiving UNHCR Twitter 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 and Instagram. 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 online.

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 branded (with UNHCR page banner) access portal, via a link from the UNHCR Records and Archives website.
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 an agreed UNHCR sub-domain e.g. http://webarchive.archive.unhcr.org/
D.	The content should be accessible via a URL of the form: http://webarchive.UNHCR/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 3 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.
F.	The Supplier will host the archive in an approved web archiving preservation format (ISO 28500:2009 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.
G.	The Supplier will demonstrate capacity and scalability in terms of size e.g. 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 the hosted version, it must not be removed from the underlying ARC or WARC files, to be transferred to UNHCR for preservation purposes.
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.

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 associate metadata, in WARC format. The Supplier should ensure the sustainability of the archive.

<p>B) Each WARC file must relate to only one website snapshot, including WARCs created during patch and fixing. A crawl of one website may consist of multiple WARCs, however. 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.</p>
<p>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.</p>

3. Support service and reporting

<p>A) The Supplier will provide technical support during UNHCR working hours (Monday to Friday, 09:00 - 18:00 CET)</p> <ul style="list-style-type: none"> • Timely response: 100% within 5 working days • Fix time for queries/issues: To be specified by the Supplier.
<p>B) The Supplier will provide project management and technical support with a dedicated team for UNHCR Web Archiving project. This team will include developers, crawl engineers, quality assurance and system administrators.</p>
<p>C) The Supplier will attend service review meetings (monthly and annual, either in person or via teleconferencing).</p>
<p>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</p>
<p>E) For each issue affecting crawl quality, reporting must include:</p> <ul style="list-style-type: none"> • 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
<p>F) The Supplier will provide:</p> <ul style="list-style-type: none"> • annual user statistics reports, • annual downtime statistics, • annual hosted data volumes
<p>G) The Supplier will provide scheduled downtime notification a minimum of two weeks in advance.</p>
<p>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.</p>
<p>I) All communications between the Supplier and UNHCR will be in English.</p>
<p>J) The Supplier will guarantee transparency regarding costs, providing UNHCR with tools for monitoring cost.</p>
<p>K) The Supplier will support UNHCR with adequate guidance and training (i.e. guidelines on how to initiate the crawl)</p>

-
- | |
|---|
| L) The Supplier's approach to problems will be pro-active 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 social media (Twitter, YouTube, Facebook, Instagram) |
| 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 at some point 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.

If 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 NGO's should be included, please specify the number of years working with public non for profit organizations;
- 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 the 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.I.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.

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 hqepu@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 with your Technical Proposal the Vendor Registration Form (Annex C).

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.

4 Evaluation

4.1 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 non for profit organizations	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 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	6
	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.

4.2 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.

5 Key Performance Indicators

5.1 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