

The National Archives' current contract for managing and developing the [UK Government Web Archive](#) and the [EU Exit Web Archive](#) is due to expire at the end of August, 2023. As such, we plan to issue an ITT in the New Year, inviting interested parties to submit bids for a replacement contract to commence in September, 2023.

*In advance of the ITT being published, we plan to arrange a supplier briefing session in January where potential bidders will have the opportunity to meet with us and to find out more about the requirement. If you would like to attend, please let us know by emailing procurement@nationalarchives.gov.uk by **5pm on Friday, December 23rd**.*

Please see below for a summary of the scope of the contract.

UK Government Web Archive (UKGWA)

www.nationalarchives.gov.uk/webarchive/

The Keeper of Public Records and Chief Executive of The National Archives (TNA), is appointed by the Secretary of State for DCMS under the Public Record Act 1958. The Keeper has a broad duty to take all practicable steps for the preservation of public records and also to provide reasonable facilities for the public to inspect and access public records.

Over the last 25 years an important part of the public record has been published on the web. TNA delivers a service to capture these records and make them available to the public. The operating model involves an in-house team alongside a contract for the provision of web archiving services. The contracted services involve crawling government websites, storing the contents in the UKGWA and making the contents available to the public. These records also provide valuable contextual information for other records when they are eventually made available by TNA.

EU Exit Web Archive (EEWA)

<https://webarchive.nationalarchives.gov.uk/eu-exit/>

The European Union (Withdrawal) Act 2018, Schedule 5, created a duty on the Queens Printer (also the Keeper and Chief Executive of TNA) to publish certain documents that were originally adopted as EU legislation, which have been retained and incorporated into UK domestic law. There is also a power to publish other documents, including the decisions of the European Court of Justice. To fulfil this duty, the EEWA serves as a comprehensive archive of EU Law, as it stood at Exit Day and later on Implementation Period Completion Day, as a permanent record of the law at those points.

For both web archives, the contractor will provide crawling services to harvest content from websites, replay services so users can access the archived content, search services, access management services, and hosting management services (the service will be hosted in TNA's AWS account but managed by the supplier).

The web archives rely on the standards that underpin the web: HTTP, HTML, etc. The content is stored as sequences of content blocks using the older ARC or more recent WARC (Web

ARChive) open standard format specifications. The WARC standard is a publicly documented open standard, developed by the International Internet Preservation Consortium (IIPC).

The contractor must demonstrate that they can use open source crawling technologies in a creative way. New content will be captured into WARCs by operating regular crawls of the sites designated by TNA. To run the crawls the contractor will almost certainly use the open source Heritrix, “industry-standard” crawler as well as deploying different crawlers, when appropriate. Each site requires a crawler configuration, framed and managed by the contractor to obtain a deep and accurate crawl of the content, which will be evidenced through a robust quality assurance process. The contractor will also need to demonstrate their capability in capturing content from various social media services (Twitter, YouTube and Flickr) and their ambition to develop their tools to capture other challenging services.

The contractor will provide a fully managed public-facing replay service of the archived content using the best possible open source software for this task, most likely the open-source Pywb system. To obtain best-in-class replay fidelity they will, where required, maintain and apply fixes such as custom redirects and a combination of server side and client side patches. They will also provide a powerful access management system as well as a workflow management portal that allows us to daily publish updates to the archive.

The contractor will use various products available in Amazon Web Services (AWS) and they will use TNA's AWS account to provide the majority of the services.

The latest crawled content will be transmitted to a separate TNA-owned AWS account and then exported via an AWS Snowball for long term preservation in our Digital Archive at Kew. We will verify the validity both of the content and the indexes (checking both fixity and validity of the transferred WARC and ARC files), the social media archives and all other data collected by the supplier and managed through the contract.

The archive will be searchable through an excellent cloud-based search service, most likely via OpenSearch or Apache Solr.

The services provided must follow the government's [Technology Code of Practice](#) and [Service Standard](#).