

New Digital Archive at The National Archives

In 1999, the British government launched an initiative to modernise its departments and streamline business processes by taking advantage of new and emerging technologies. A major part of that initiative mandated the *electronic* storage and retrieval of all new records by 2004. Recognizing that it would be responsible for preparing the long-term storage for many of these diverse 'born digital' records, The National Archives has implemented a system capable of meeting these requirements.

Preservation at the National Archives

As it had to address the secure storage and operational management of records in a service environment, The National Archives decided to focus on building a digital *storage* repository. This system would preserve the digital objects' bitstreams and any subsequent manifestations. The archive was also to hold information about the records and audit trails from the processes used to manage the records. In March 2002, it requested bids from technology companies to build such an archive while adhering to its specialized provisions. There was a requirement for a modular configuration so that components could be easily changed as technology advanced. The system had to provide absolute secure access. And most important, the archive had to be extremely scalable for future needs.

Methodologies

The National Archives used proven software engineering (XML, Java) and project management methodologies (PRINCE 2, UML) to ensure that the project was delivered on time. A document detailing the requirements of the system was drafted in March 2002. This was reviewed internally, and circulated externally to a number of expert groups (including JISC, NDAD and the British Library). A formal request for proposals was then published in the Official Journal of the European Communities (OJEC).

Many responses were received and, after a robust scoring process, these were whittled down to 7 companies that were invited to submit tenders. In August 2002, following thorough assessment and auditing, Tessella were selected to develop the digital archive system.

The Solution

Working with Tessella, Functional Specification and Use Cases documents were completed before programming started in November 2002.

The system, compliant with the e-gif standard for XML based open source software development by UK government departments is defined in the Architectural Design Document. It comprises:

- A Java applet for metadata compilation and file association (See Image 1)
- An XML Web browser interface for loading, administration, browsing and searching (See Image 2), supported by an Oracle Application Server
- An Oracle 9i XML database running on Unix (Sun Solaris) for storing the metadata
- A Compaq RAID server for processing the digital objects
- Filetek's Storhouse/RFS application to manage the tape library system where the objects are transferred in batch mode from the Compaq server
- A Storagetek L20 tape library with 2 LTO tape drives (expandable to 100TB)

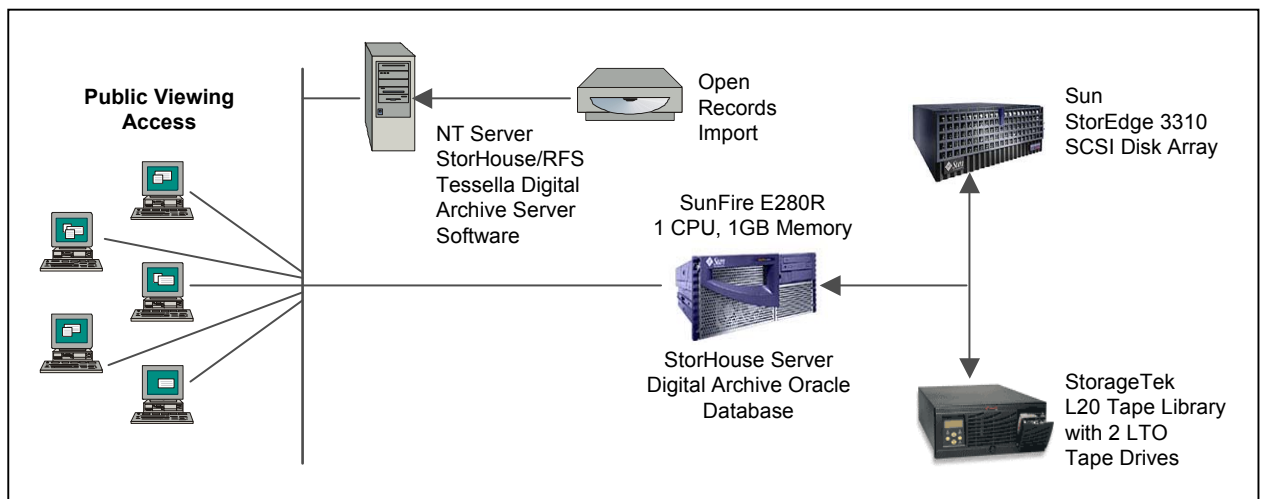
The metadata loaded by government departments as part of the transfer process reflects best current practice. It is a superset of EAD, Dublin Core, PRO ERMS, VERS and OCLC metadata schema. The metadata, search and browse schema are available from the UKGOVTALK website (www.govtalk.gov.uk/).

An intense development period by six programmers was completed during the winter of 2002-03. At the same time, two sets of hardware and COTS software were delivered and installed at Kew for the 'air-gapped' master system and the public access system. TNA testing commenced in mid March 2003. By the end of the month, the access system containing records open to the public could be accessed in the search rooms.

The future

Work will begin on the second stage of the project in 2004: the development of a presentation system that will provide a range of delivery channels across the Internet and onsite at Kew.

Concurrently, PRONOM (file format reference database see www.pro.gov.uk/pronom) will be developed as a key component of our digital archiving work and drive our migration strategies. It will be released on the web in October 2003.



The National Archives' Digital Archive Solution