

The Digital Archive

Customer

The National Archives is the national archive of the United Kingdom. It brings together all publicly available records from both the Government and the Law Courts and makes them available to anyone who wishes to view them. Their archives span an unbroken period from the 11th Century to the present day.

Business Problem

The National Archives wanted to create a digital archive for selected records from government departments. Due to the unique heritage of the nation, stringent conditions apply to ensuring integrity, survival and security of data. The digital archive required a secure store for electronic data in perpetuity with the ability to adapt to rapidly changing IT technology. 'Born digital' accessions needed to be preserved by archiving the original bitstream. In line with government policy the electronic Government Interoperability Framework (e-GIF) standards would apply to the design of the digital archive and its interaction with other systems both present and future.

Tessella Solution

Tessella proposed a technologically leading edge solution based on the use of Open Standards and "best of breed" components. These have been integrated in such a way as to provide The National Archives with functionality tailored to its specific, individual needs. This highly modular architecture allows the substitution of bespoke or Commercial Off The Shelf components. Thus, from time to time obsolete modules, such as the virus checkers could be replaced with new superior ones without having to reconstruct the whole Digital Archive. Similarly, although the archive must run using commercial operating systems, database and application servers, it is not intimately integrated to these. This means that The

National Archives can take advantage of the most powerful software available and will not be tied to any single supplier or proprietary product, either now or in the future. This advantage applies equally to the hardware components of the architecture as the software and is expected to enable migration to a new hardware platform.

The architecture of the digital archive has been built on Open Standards using the Java 2 Enterprise Edition (J2EE) layered architecture. Oracle 9iAS with native XML DB is the central database, supplying transaction management, the security framework, audit and searching capabilities. The management tools of the archive are defined by XML, SOAP and JDBC so that they can communicate with each other and other applications such as The National Archives' public online catalogue via an open protocol. The solution complies with the UK Government's e-GIF standard. The user interface component was developed by Tessella as bespoke software to ensure the fit with the users' requirements and capabilities.

Features

The Digital Archive has the following features:

- ❑ Two systems – One maintained inside an air-gap to ensure integrity and an open system available to users in the Reading Room at Kew. The former system has all the available functionality while the latter maintains only open records and allows users to search or browse for records and to download these.
- ❑ Loading – Submissions will arrive on removable media. Archivists then create the metadata and save it as an XML file. This can occur away from the main system. The submission can then be made to the air-gap system at which point

a number of consistency checks (including virus checking) are made.

- ❑ Storage – Metadata is stored in an Oracle database. Files are stored separately with access controlled by the database.
- ❑ Editing – Users with appropriate security can edit or delete data held in the archive.
- ❑ Search facilities – Users can find records by searching on a phrase (with optional date and department restrictions) or by references. There are also sophisticated browsing facilities.
- ❑ Download facilities – This allows users to retrieve digital objects.
- ❑ Synchronization tools – These ensure that the open system is updated with all changes made on the air-gap system.

Benefits

- ❑ Tessella's experience – Tessella has a track record of developing and supporting digital archives such as helping to support the scientific data of UKAEA JET over the past 20 years. This successful project has been achieved through careful architectural design and selection of appropriate technologies.
- ❑ Scalable solution – The component interfaces are designed using only Open Standards (XML and SOAP), therefore each component can easily be replaced.
- ❑ Substitutability – The use of a combination of Open Source Software and Commercial Off The Shelf components mean that The National Archives are not tied into any single supplier or proprietary product. Capabilities will not be restricted by a COTS system.
- ❑ Ability to preserve multiple file types – Many types of records can be stored including emails, attachments, files from legacy systems, official enquiry records,

web site snapshots, databases, text documents, images, video clips, sound files, multi-media files. These can be transferred into the archive from numerous media types old and new, including tapes, cartridges and floppy disks.

- ❑ Support for simultaneous users – Support for up to 100 users with 1TB of data rising to 100TB over five years.
- ❑ e-GIF compliant – This will allow the seamless flow of information between management tools and other applications, including The National Archives' public online catalogue, via an open protocol that complies with the electronic Government Interoperability Framework.

Fit to user requirements – The development of some bespoke software to interface the system to the user means that the users' capabilities will not be restricted by a COTS system.

Copyright © Tessella Support Services plc

V1.R1.M0

HEAD OFFICE

Tessella Support Services plc

3 Vineyard Chambers • Abingdon • Oxfordshire • OX14 3PX • England.

Telephone: +44 (0)1235 555511 • Facsimile +44 (0)1235 553301 • E-mail: info@tessella.com

Registered in England No. 1466429. VAT Registration No. GB 229 0489 52.