



“We’ve got a secure system which will store and preserve digital records, and continue to make them readable for a long time into the future, and I think that’s a really important tool to have available to us.”

Dr David Thomas
Director of Technology
The National Archives

The National Archives

Leading archival institution securely stores and preserves digital records using best-in-class archiving solution developed through an innovative public-private partnership.

Background

The National Archives’ collection is one of the largest in the world, with 11 million records, from the Domesday Book and the Magna Carta to modern government papers. It has been managing paper documents since it was founded in 1838, but the advent of digital technology has created a whole new challenge – that of storing and preserving digital records. To meet this challenge, The National Archives has developed a new storage and preservation system, built in partnership with Tessella, that not only allows it to efficiently store, manage and, where necessary, migrate its current collection of digital files, but also provides the headroom to significantly expand its electronic archive.

Business Situation

Based in Kew, West London and employing 600 people, The National Archives is the UK Government’s official archive, documenting almost 1,000 years of history, with records ranging from parchment and paper scrolls through to digital files and archived websites.

At one time, paper was its stock-in-trade but as it has moved into the digital age, electronic records have become a growing part of its records library. So, beginning in 2000, The National Archives embarked on a series of projects to create a system for managing the growing volume of digital files and records under management. The key impetus for the project was the fact that the UK Government had started creating all of its documents digitally. “All government documents are electronic these days so we had to have a system for storing and preserving them,” explains Dr David Thomas, Director of Technology at The National Archives.

Digital documents present a special challenge in that, unlike paper records which can last hundreds or even thousands of years if properly maintained, digital objects are extremely vulnerable to changing technology, which can make the record unreadable and therefore worthless. Addressing this vulnerability would have to be a key part of any digital archiving system that was developed.



Project

The National Archives realised it would need outside help to develop the system it wanted and in 2001 awarded Tessella a contract to register the various file formats within its digital repository. Next, it commissioned Tessella to build a straightforward digital archive that would take in diverse files, such as emails and their attachments, website images, video clips, and sound files. The system, which came on-stream in March 2003, was based on a detailed specification provided by The National Archives. FileTek StorHouse storage management software was used to manage the storage of the digital files in the repository, which were archived on high-speed, high-capacity tape cartridges. The system was designed to scale to 100 terabytes of storage or even more if required.

More was to come. Later in 2003, The National Archives initiated a major project called 'Seamless Flow'. The idea of Seamless Flow was that a record would move smoothly from point of creation within a government department, through its history and use in that department, to the point where it became an inactive record and then was passed through to The National Archives, providing it got selected for permanent retention.

Seamless Flow comprised eight sub-projects, the majority of which were done in-house at The National Archives. Two of them were awarded to Tessella, following a tender process.

Thomas explains why: "Tessella is very familiar with the technologies and approaches of archiving historical documents. It's very important to work with a company with experience of working in this space – that was the main attraction of working with them."

The objective of the first project, called 'Preservation & Maintenance', was to store the digital records permanently in a secure environment and give other parts of the Archives controlled access to those records. The second project, 'Technology Watch', involved understanding the formats and technology used to create the record in the first place, and then enabling the migration of the record to a new format if the technology for reading the original files became obsolete. The migration feature was necessary to ensure that records could be read even if technologies moved on.

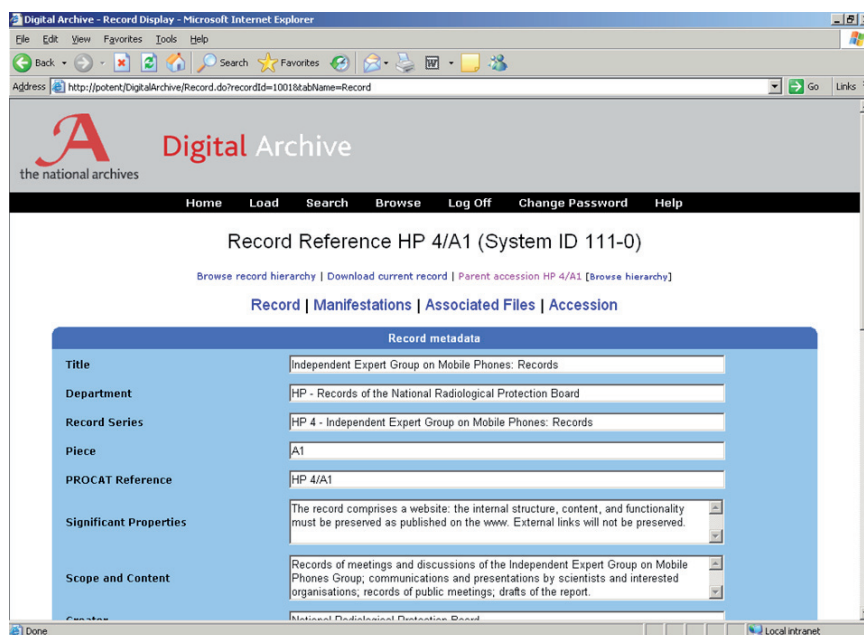
"We needed a system which would take in digital records, store them and, once they became unreadable using current desktop technology, migrate them to something which could be read using current technology," Thomas explains.

The other key requirement was that the system would integrate closely with The National Archives' own systems, particularly its catalogue and online delivery applications.

Figure 1:
Web-based
user interface
to search for
digital objects



Figure 2:
Web-based
user interface
summarising
digital record



Solution

Tessella worked closely with The National Archives to help them build the world's first purpose-built national digital archive which incorporates a unique approach to the preservation of digital objects. The system is built using Java and runs on a Windows Server. There is a second physical server running Solaris that powers an Oracle database and interfaces with the software element of the FileTek bulk storage system via a virtualised directory.

The digital archive communicates with other National Archives systems; specifically Procat, the catalogue application that records the holdings of The National Archives, and the online delivery or presentation system, which is the interface through which members of the public conduct online searches of archived information.

A key feature of the digital archive provided by Tessella is its 'active preservation' technology. This means that the information contained within a record can be preserved even if the record is migrated to a new technology format.

Tessella director Jon Tilbury notes that digital obsolescence has become a critical issue for many large organisations. "Once they have put their data into an organised data structure, the big problem they have is digital obsolescence – the fact that the data is no longer readable. A good example is Microsoft Project 4, which was popular in the 1990s for doing project planning. There is no currently supported Microsoft software that will read Microsoft Project 4 files. People are constantly coming across

problems like this, which are a creation of the digital age."

Tessella's solution, Tilbury says, is to "have a flexible framework that allows the most common preservation technique – file migration to newer formats – to be done in a way that is intelligent and managed. So you can migrate to new formats that are appropriate; you can confirm that process works; and you can cope with complex file relationships."

According to Rob Sharpe, Head of Archiving Solutions at Tessella, a key feature of the active preservation technology is its tight control of the migration process. "This is done by measuring the properties of those records before migration, migrating them, and then measuring their properties again when they've been imported into the new technology. We then compare the two sets of properties to ensure no information has been lost during the migration."

Deployment

The project team consisted of 11 people in all – six core Tessella personnel working with an internal team of five people. The work began in mid-2006 and the system delivered in late 2007. There then began an extended commissioning phase where data was imported into the archive and various components of the system were evaluated and fine-tuned.



Customisation

To be effective, the digital archive could not exist in isolation but needed to interact with existing systems at the Archives. This required a significant amount of integration work on Tessella's part, particularly in terms of interfacing with the cataloguing and presentation systems.

It was essential that these three systems could speak to each other, as Tim Gollins, Head of Digital Preservation at The National Archives, explains.

"It was critical to achieve this integration because when members of the public key in a search request, they are not searching the archive as such, they are searching the catalogue, so it was necessary to export the critical facts about the archive somewhere where the public was allowed to search."

There were also special requirements. The public is not permitted to search the archive directly, and therefore, there needed to be a physical separation between the archive and the cataloguing and presentation system.

Adds Gollins: "Further to that, we also needed a secure system internally within the archive that would keep certain material separate from other material. This was a one of our key requirements and Tessella were able to deliver for us."

The most challenging part of the project from Tessella's perspective was fine-tuning the process to migrate a record from one technology format to another.

"Measuring properties of a record and comparing the before and after turned out to have more complications than either party envisaged," recalls Sharpe. "We had to come up with a framework for measuring properties that didn't rely on the same file structure, before and after."

A second challenge was that because records are hierarchical – i.e. they contain parent records and sub-records – this hierarchy had to be recognised in the migrated files too.

"What we realised was that if you do a migration of any hierarchical structure you need to tell its parent, and its parent's parent, of this new version. That was something we hadn't realised from the outset and again we had to come up with something that made consistent, logical sense," Sharpe explains.

According to Sharpe, the close working relationship that Tessella and The National Archives had built up over a number of years was invaluable in overcoming these and other challenges. "We were talking the same language," he says simply.

"There is a very strong and successful working relationship between the two sides," agrees Gollins. "We have been working together for a number of years now and it's true to say that there is a real feeling of togetherness."

The research partnership between Tessella and The National Archives has resulted in the Tessella Safety Deposit Box (SDB). Sold by Tessella under a commercial exploitation agreement, SDB has been sold to several other national archives and libraries in Europe and beyond.

Result

This Digital Archiving project is an excellent example of a true public-private partnership. The National Archives has a world-leading facility that cements their position as leaders in their field. Tessella has a system that can be re-used for other clients and an arrangement to feed the benefits back to The National Archives. The world of archiving has gained access to SDB, a technology that will help preserve the world's knowledge for many years to come.

Tessella plc 26 The Quadrant, Abingdon Science Park, Abingdon, Oxfordshire OX14 3YS, UK
T: +44 (0)1235 555511 | F: +44 (0)1235 553301 | E: info@tessella.com

Tessella Inc 233 Needham Street, Suite 300, Newton, MA 02464, USA
T: 1 617 454 1220 | E: info@tessella.com

Tessella – successfully delivering IT and consulting services to world leaders in R&D, science and engineering.

For decades, Tessella has been successfully delivering IT and consulting services to world leaders in R&D, science, and engineering. Through the application of scientific methods and rigorous quality procedures, we enable clients in life sciences, energy, the public sector, and consumer industries to achieve a wide range of objectives, including, forecasting floods, developing fusion power, enhancing military sensor capability, improving drug discovery and development efficiency, and reducing risk to health and the environment in the extraction and production of oil and gas. With offices in Europe and North America, global companies rely on Tessella for business critical assignments.

Copyright © Tessella plc 2009, all trademarks acknowledged. Issue: V1.R0.M0 | Sept-09



www.tessella.com

