“...PROVIDE A LASTING LEGACY FOR GLASGOW AND THE NATION”

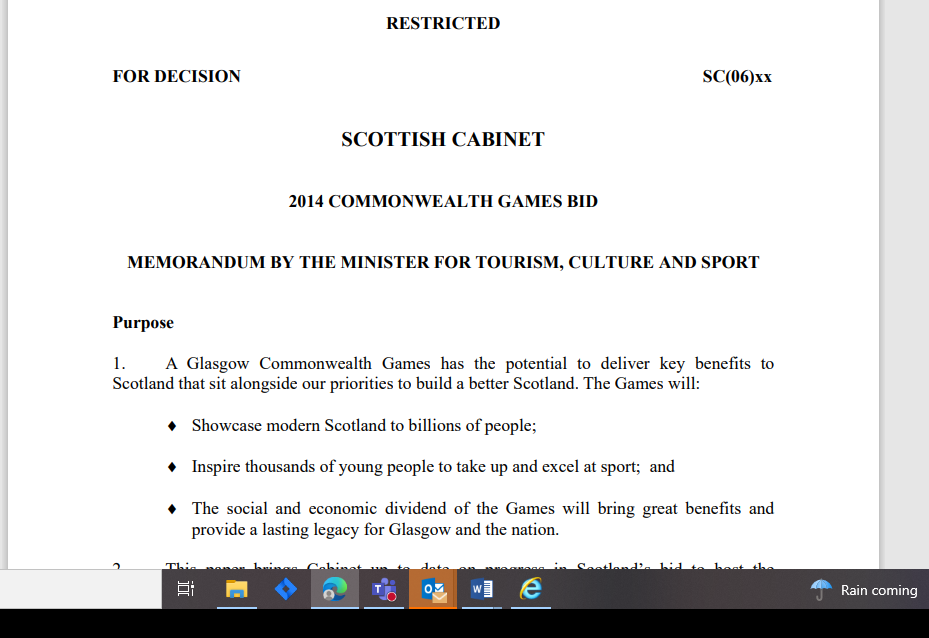
Two years of transferring Scottish Cabinet records to National Records of Scotland

|  |  |  |
| --- | --- | --- |
| **Garth Stewart** |  |  |
| *National Records of Scotland*  *UK*  *Garth.stewart@nrscotland.gov.uk* |  |  |
|  |  |  |

**Abstract – This paper outlines the recent transfer of a series of high-level Scottish Government records to National Records of Scotland (NRS). The records needed to be made publicly available almost as soon as they arrived at our archive, and the transfers prompted NRS to consider how we could provide online access to born-digital records for two distinct use cases. This paper outlines; the background to this task, how NRS evaluated the solutions that could be used to deliver online access, and lessons learned from an enterprise carried out exclusively within the challenging context of the Covid-19 pandemic**

**Keywords – Transfer, Access, Re-use of technologies, collaboration, Public records**

**Conference Topics – Innovation; Resilience**



Extract from records of the Scottish Cabinet, Minister for Tourism, Culture and Sport, Draft: 2014 Commonwealth Games Bid (NRS Reference SCR14/26/5) available at <https://www.scotlandspeople.gov.uk/>

# **Introduction**

This paper outlines the recent transfer of a series of high-level Scottish Government records to National Records of Scotland (NRS). The records needed to be made publicly available almost as soon as they arrived at our archive, and the transfers prompted NRS to consider how we could provide online access to born-digital records for *two* distinct use cases. This paper outlines; the background to this task, how NRS evaluated the solutions that could be used to deliver online access, and lessons learned from an enterprise carried out exclusively within the challenging context of the Covid-19 pandemic.

1. **What is NRS?**

NRS is a non-Ministerial Department of the Scottish Government (SG) - a devolved administration of the UK. NRS carries out a range of nationally-important statutory functions[[1]](#footnote-1), including that it holds, preserves and makes available the national archive collection of Scotland.[[2]](#footnote-2)

NRS has a wide and highly diverse range of depositors; from the SG and numerous public authorities, to the Scottish Courts and a large array of private depositors.

NRS’s archiving function extends to all record formats – from parchment to born-digital. NRS has been collecting born-digital records since 1998, and we have progressively developed policies and processes which combine the corporate tools and services at our disposal with the very latest in digital preservation good practice.

In the year that iPres comes to Scotland, the paper highlights the themes of community, innovation and resilience, which were paramount to delivering brand-new forms of access to one of Scotland’s national collections.

1. **Digital Archiving at NRS**

Our process for handing born-digital records is straightforward: our Archive Depositor Liaison (ADL) team work with the depositor ahead of transfer to agree precisely which records are to be deposited, and prepare these for transfer. Often colleagues from the Digital Records Unit (DRU) support this engagement and advise on format, metadata, export etc. Sensitivity review pre-transfer is undertaken by the depositor, who remains the owner and Data Controller for the records. NRS may ask for test data to be submitted to clarify technical matters, after which a ‘transfer package’ – records, manifest, and associated metadata - is transferred on encrypted USB Hard Drive. For further information see our [Depositor Guidance for the Transfer of Archival Born Digital Records](https://www.nrscotland.gov.uk/files/record-keeping/depositor-guidance-for-the-transfer-of-archival-born-digital-records-may-2020.pdf).

Once at NRS, the transfer package is scanned for malware, and undergoes a series of ingest processes (fixity, characterisation, completeness checks)[[3]](#footnote-3), before being uploaded to the NRS Digital Repository. Until recently, all of these processes took place onsite.

Storage for the Digital Repository is built on existing NRS ICT infrastructure, ensuring we have a high level of support to maintain our security, maintenance, and storage capacity needs. Multiple copies of records are kept on different media types and geographical locations.

Generally digital records are only catalogued to accession package level, however we are starting to catalogue more at a more granular level. Until recently, access to digital records could only be provided on-site on a standalone PC, or via provision of copies.

1. **Challenging our practice: Scottish Cabinet Records**

Two recent transfers of significant government records challenged our access procedures.

The Scottish Cabinet (SCab) is the group of senior Ministers, including the First Minister of Scotland, which is responsible for SG policy. It came into existence in 1999, following the establishment of the Scottish Parliament and the devolved Scottish administration. The records document senior governmental decisions and policy matters in Scotland; from infrastructure and social policies, to global summits and health services.



Photograph of the 2005 Scottish Cabinet with the SG Permanent Secretary inside Bute House, Edinburgh  
Crown copyright, National Records of Scotland, SCR14/6

To support government transparency and openness, SG records – including SCab - are transferred to NRS after 15 years[[4]](#footnote-4) – at which point they fall ‘open’ and are made available for public access by NRS, unless exempt under freedom of information legislation.[[5]](#footnote-5) Records of the SCab have been created and held entirely digitally from 2005, with records from that year scheduled for transfer to NRS in late 2020.

For at least 40 years, NRS has provided a ‘media preview’ to government records (including SCab), whereby journalists are invited to see selected records ahead of their public release date (normally on or around 1st January). Until December 2019, journalists attended our search rooms in-person to see/preview paper records, and the public could consult the paper files from the following January onwards.

In 2020 – and again in 2021 - this solution became untenable, due to Covid-19 restrictions. Nevertheless, the 2005 and 2006 records still had to be transferred and made publicly available – this time *online*. The show must go on, but how? The fact that the pandemic coincided with the first tranche of fully born-digital SCab files meant we *had* to undertake almost this entire operation remotely.

1. **Part 1: Preservation**

**Transfer and ingest**

The selection, transfer and ingest of the 2005 and 2006 SCab records went reasonably smoothly. SG uses [Objective ECM](https://www.objective.co.uk/products/objective-ecm) as its main records management system, and all SCab records were drawn from this source. NRS also use this system for records management: this familiarity helped us understand the records we received.

SG carried out sensitivity review of records pre-transfer, leading to the generation of redacted sets for public access, and un-redacted sets (which remain ‘Closed’ until FOISA exemptions lapse). The files were exported from eRDM, and the export included manifest CSV files containing essential metadata such as original eRDM locations, Object IDs, file sizes and checksums.

Once transferred, ingest into the NRS Digital Repository was straightforward. Both record sets were successfully scanned for malware, profiled using DROID, and verified upon transfer using checksums contained in the manifest files.

1. **‘You may have forgotten to attach a file’**

Email preservation presented a specific challenge. Upon comparing the 2006 record sets, we noticed that an identical number of records was contained within the ‘Open’ and ‘Closed’ sets. This was in contrast to 2005, where individual email attachments had been taken *out* of emails, redacted and saved as *separate* objects. For 2006, it was evident that redacted attachments had been *added back into* the email files.

The Open record set contained 80 Outlook email files (x-fmt/430) which, in effect, were acting as container files for 265 attachment files (mainly Word document and pdf)! Future work will be undertaken to manage this preservation risk.

1. **Part 2: Access**

**Unlocking access**

Approaching transfer day in late 2020, NRS had *no* obvious means of providing online access to journalists or the wider public. In Autumn 2020 a project team was established to tackle this challenge. In both cases, we needed solutions which would be cheap (or free!) to procure, quick to install (given our deadlines), and easy for users to operate. In the case of the journalists, we also required a system which could be more strictly controlled, so that copies of records were not shared ahead of their release date.

NRS is a large government organisation with numerous statutory obligations and processes, and it has a large amount of digital assets, licences and services in operation at any one time. The project team took advantage of this by evaluating and creatively engineering the ‘re-use’ of existing tools for born-digital access, rather than buying anything ‘new’.

1. **Media preview**

For journalists, a solution which would render a copy of the records in their original formats as far as possible was required. [Objective Connect](https://www.objective.co.uk/products/objective-connect) (OC) – a secure external file sharing application licensed for use by SG and NRS – was selected given its supporting features:

* a ‘private’ virtual reading-room could be created on OC for users, who would register/log-in to review content
* a folder structure could be created to replicate the original storage hierarchy of the records. Folders were labelled with NRS archive references and descriptions to facilitate browsing
* Each folder and document has a unique url, meaning that journalists could be provided with a hyperlinked index allowing them to click to directly access particularly significant records identified by NRS
* Once within the reading room, documents could be accessed concurrently, and users would be unable to identify others ‘in the room’. This supported the control and confidentiality of the process
* Access to every document could be tracked and audited by NRS
* Digital objects would be ‘viewed’ (referred to as ‘Preview Mode’) rather than downloaded by users: this mitigated the risk of copies being shared improperly before public release
* In practical terms, this was a familiar tool for NRS/SG colleagues, with no need for further investment, licensing or training

This solution worked very well – journalists received registration instructions in advance, as well as contextual information about the records, a detailed catalogue to enable browsing, and an extensive Index with hyperlinks to significant documents. An unexpected benefit of using OC was access audit trail functionality – this provided archive colleagues with data on what records were viewed the most and for how long – analytical data impractical to obtain in the paper world.

One issue did relate to our old friend, email files. OC’s Preview Mode did not allow for embedded files (e.g. email attachments) to be opened. To get around this, NRS migrated copies of the email files from Outlook into eml format and used a simple Linux utility called ‘[munpack’](https://linux.die.net/man/1/munpack) to extract attachments. This enabled us to create a folder structure for each email file so that the original email files could be reviewed alongside the attachments. This was a significant lesson learned, and for future accessions we intend to conduct this attachment extraction at point of receipt.



Munpack image taken from <https://www.youtube.com/watch?v=SKkvb5JF0Qshttps://www.youtube.com/watch?v=SKkvb5JF0Qs>  
Copyright Kris Occhipinti

1. **Public Access**

If delivering one born-digital access solution wasn’t hard enough, how about designing a second for the wider public?!

OC would not suffice for this, given its limited scalability, but a parallel NRS project provided an option. The [ScotlandsPeople](https://www.scotlandspeople.gov.uk/) website is NRS’s primary platform for sharing digitised images of our collections. Used mostly for genealogical research, users can search record indexes for free (e.g. births marriages and deaths statutory registers, wills, valuation rolls etc.) and use a pay-per-view service to download images of relevant records.

During the early stages of the pandemic NRS was considering how to share more records online, via a new component on ScotlandsPeople called ‘Virtual Volumes Online’ (VVO). A large series of digitised church records were being prepared for release in early 2021, and this platform seemed suitable for SCab records too, given that it was scalable, secure, and soon to become part of the recognised ScotlandsPeople service offer. Again, no additional licensing would be required and this was selected for our use case.

As with any prototypical service, there were some snags! Unlike OC – which could render most formats - VVO had been set up to host image formats only: JPEG and TIFF. This configuration would present issues when converting formats such as email and Powerpoint into images for access. An alternative ‘render’ format for VVO was required. PDF was chosen, for a number of business reasons:

* Web-friendly: generally PDFs would be smaller files than the originals so would be easier to download and make available on the Web
* Information security: As PDFs are a fairly simple file format, they had a lower risk of containing or hiding ‘nasties’ such as malware
* Information fixity: NRS wanted to ‘fix’ the content of records as much as possible. We didn’t want to make content available that would be easy to alter and pass off as something that it wasn’t. PDF offered this aspect
* Accessibility and obsolescence: As the original records would have been created in a 15+ year old version of MS Office, there could be greater format obsolescence risk. PDFs would be more backwards compatible for people to read.
* Efficiency and management: generating and tagging the metadata of JPEG images for each ‘page’ of transferred records would have been highly laborious. Using PDF simplified this process, whereby multiple documents could be appended together with the same metadata – ensuring provenance and original order was maintained - and presented as larger files e.g. one PDF per meeting of the Scottish Cabinet with all related documentation combined (be it Word, powerpoint, Excel etc.).

Adobe Acrobat was used to do the bulk of conversion (with licenses borrowed from another NRS team), and generally this pre-access processing approach worked. Such is life, there were still a few technical issues to troubleshoot.

1. **Troubleshooting**

Email files would not convert to PDF via Adobe Acrobat. Instead our IT colleagues used, [BitRecover PST to PDF Wizard](https://www.bitrecover.com/pst-to-pdf/), to successfully convert all Outlook files into PDF. Pleasingly, whilst IT colleagues pioneered the use of this tool for the 2005 records, they were able to pass on this learning to archive staff, who successfully converted 2006 emails using the same method.

It was difficult on some occasions to convert information stored in Excel files into PDF – if this presents an issue to users, we will provide a copy of the original Excel file via a suitable transfer mechanism.

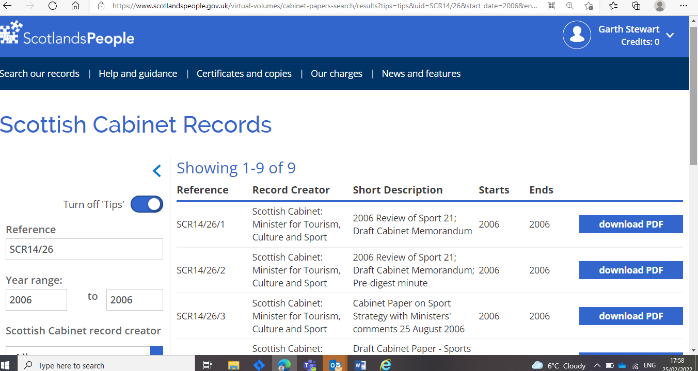
File names also proved quite painful: all files names needed to have an NRS reference added to them, in order to be uploaded to VVO. To do this at scale, the team used a Powershell command line script, which proved effective.

However, we did encounter issues likely familiar to digital preservation colleagues:

* File names sometimes exported in different orders depending on what tool was used for this (e.g. Windows file explorer, Powershell).
* Some file names were also uncomfortably long for our systems to cope with before they were converted to PDF, which obstructed file conversion.
* We also needed to ensure that the metadata populated for the access copy of the files corresponded with how they were catalogued within our central catalogue

Fixing these issues required a lot of data merging and time-consuming manual checks.

Lastly, in order to publish these records on ScotlandsPeople, new web pages, search functionality, and browsing features needed to be incorporated into the site by our third party provider. This meant that the 2005 records were not made fully available until [October 2021](https://www.scotlandspeople.gov.uk/article/news-article-scottish-cabinet-records-released-scotlandspeople), although the 2006 set were successfully published on 6 January 2022.



How the SCab papers are presented on ScotlandsPeople

1. **Conclusions, lessons, and the future**

Preparing these born-digital records for access was a fairly herculean effort. NRS’s examples may not be suitable or feasible for other archive services who similarly need to develop online access to born-digital records from scratch.

What the examples do show however is what can be achieved in a short period of time, with the right people ‘in the room’, and with a defined common goal in mind. All collaboration needed for this enterprise – from initial conversations to project planning, to format conversion and record upload – took place entirely virtually – whether on MS Teams or in one of NRS’s virtualised desktop environments (Adobe Acrobat and Powershell tasks, for example, were conducted in virtualised environments). The project was also delivered ‘on a shoestring budget’, and was supported by identifying alternative uses for corporate IT tools and services.

NRS has a long way to go in its digital archiving journey. We will need to take stock of the ScotlandsPeople platform (especially its ‘normalisation’ workflow) and receive user feedback on its strengths and limitations. The processing work required to process the records was also beset by particular inefficiencies, which we will work to overcome.

One thing is certain: NRS will continue to receive born-digital records, and be expected to provide public access to these – starting with the 2007 SCab records which will arrive later this year. With our Objective Connect and VVO prototypes successfully established, and with ideas on how to innovate our practice (e.g. use of further automation to handle ingest, use of natural language processing to query records for sensitivity and topic modelling), we have manufactured a much stronger foundation of digital archiving practice to confront our next set of challenges.

#### ACKNOWLEDGMENT

The author wishes to thank all colleagues at NRS who helped provide text, edit, or supply resources for this paper, which covers what was a cross-departmental, highly innovative success for our institution.

1. For a full range see <https://www.nrscotland.gov.uk/about-us/what-we-do> [↑](#footnote-ref-1)
2. For further information see: <https://www.nrscotland.gov.uk/research/research-guides/historical-records-an-overview> [↑](#footnote-ref-2)
3. DROID, CSV Validator and Teracopy are now used for these processes [↑](#footnote-ref-3)
4. Reduced from 30 years in Scotland in 2009 [↑](#footnote-ref-4)
5. [The Freedom of Information (Scotland) Act 2002 (Historical Periods) Order 2013](https://www.legislation.gov.uk/ssi/2013/365/made) (FOISA) confirmed that records become ‘historical’ after 15 years. This is slightly earlier than other UK administrations. [↑](#footnote-ref-5)