

Cabinet Papers 1915-78

The National Archives

This is one of four case studies created to illustrate digital preservation practices within digitisation projects, sharing honestly the challenges of establishing digital continuity from a project of limited lifespan. It is selected from 16 projects which were funded by the JISC Digitisation Programme between 2004 and 2009. This programme has provided digital access to collections of international significance that would otherwise be inaccessible. In doing so it has also generated an opportunity to learn about the critical success factors in digitisation, such as ensuring long term access from short term projects. Each case study is based on an interview carried out on behalf of the JISC by the Digitisation Preservation Project.

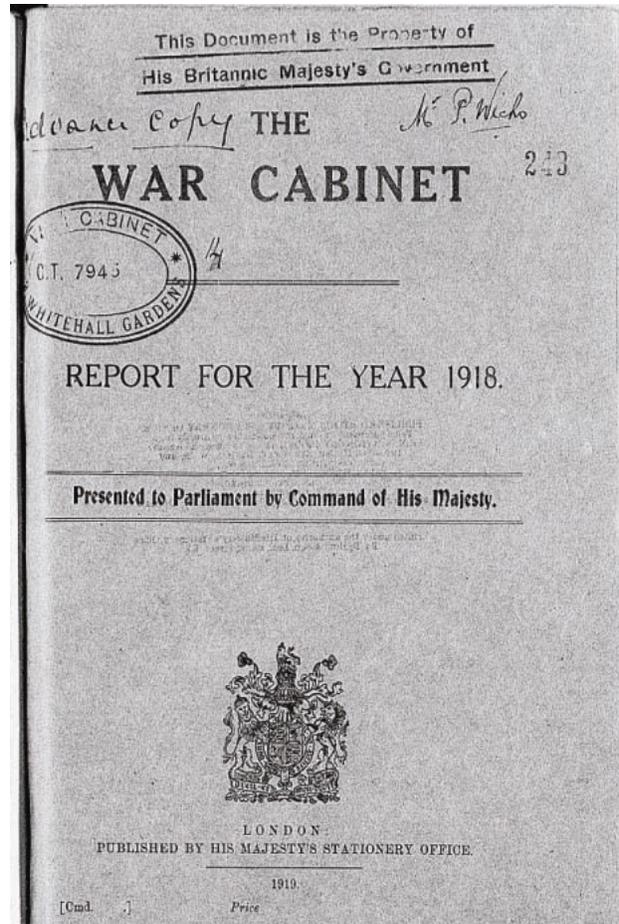
Introduction

The 'British Governance in the 20th Century Project' was a two year project led by The National Archives and partly sponsored by JISC. The project provides comprehensive online access to the papers of British Cabinet, between 1915 and 1978.

The National Archives has an impressive track record in digitisation, but this was their first project to receive JISC funding to produce digital content. The National Archives has been at the forefront of digital preservation practice in the UK. They have an exemplary digital preservation policy, have developed and implemented technology to support it. They have won the coveted Digital Preservation Award twice in recognition of this work. With such a firm foundation, the Cabinet Papers project had a robust preservation policy. Delivery and preservation of digital content have different requirements and therefore need related but distinct systems.

A key challenge in digital preservation is packaging content and metadata in a format that can be reused over and over. The National Archives' response to this challenge is very encouraging: they will be packaging preservation master images and its preservation master metadata together using *METS*, meeting one of the key requirements of preservation – the ability to move the master copies of content and metadata together from one organization or space to another.

Front page of the War Cabinet Report for 1918 (CAB 24/86) from the Cabinet Papers collection at the National Archives



Key Message

Digital preservation policies indicate whether an organization is committed to long-term access. Grant giving organizations should request copies of applicant's digital preservation policies when funding data creation.

Staff and Skills

Although led by The National Archives, elements of the project were executed by partners - Transmedia - who undertook digitisation and optical character recognition. Outsourcing can create tensions, but in this case there was a very well defined relationship between the project manager and the staff responsible for the preservation of the digital assets. The skills identified as necessary for such a project include planning, forward thinking, problem-solving and risk management. Knowledge and experience in digital preservation as well as knowing the current state of the art in relation to digital preservation were also seen as very important factors. Staff noted that it can be challenging to identify what other people are doing in similar fields, especially given the rapid developments in digital preservation in recent years.

Content

The project was concerned with the digitisation of minutes and memoranda from microfilm. These documents record the actions and decisions of the British government throughout the 20th century. They are a key source for a large constituency of researchers across a broad range of subject areas and disciplines including politics, economics, social science and international relations, as well as for students at all levels of study. The topics covered in the collections are an integral part of undergraduate, graduate and school programmes. Hitherto, usage of these collections has been limited to those who have the opportunity to visit The National Archives in Kew or whose institutions could afford to purchase the microfilm copies.

The new resource consists of approximately 50,000 digital objects equivalent to around 800 gigabytes of disk space. It is very closely aligned with other thematically related digitized collections focusing on the political landscape of Britain in the 20th century. Collections relevant to this project include like data sets like DocumentsOnline, The Macmillan Cabinet Papers 1957-63 Online, BOPCRIS and Newsfilm Online.

Content Management and Workflow

The National Archives manually coordinates the movement of content through digitisation, delivery and preservation processes using the *PRINCE2* project management method. The workflow tends to be manual. Microfilm is sent to the contractor where it is scanned and 'OCR'd'. Technical metadata on image creation is described using MIX. The bibliographic metadata is described by using Dublin Core (DCMI) standards. The two resources are "wrapped" in METS to create a set of

single packages (one for each archival series) which describe the resource in a standardized way.

Extract from War Cabinet Report for 1918 (CAB 24/46)

INTRODUCTION.

The record herewith submitted is of a period to which no immediate survey can do justice. The magnitude and intensity of the operations, the swift reversal of the fortunes of war, and the political upheavals following the military disaster among the Central Powers, mark the year of climax. In the rapid succession of striking events, developments which in normal times would have attracted general attention passed almost unnoticed. But, deeply significant as have been many of the steps in national and international organisation, the downfall of the military power of the Central Empires stands out as the central theme of 1918.

Surrogates are stored on the National Archives' preservation server. Delivery versions of the whole (PDFs of the text and catalogue entries for the metadata) are derived from the masters.

Manual processes like this leave room for manual errors and make coordination of delivery and preservation objects more difficult through time. Consequently there is scope for The National Archives to define more rigorously how the delivery and preservation systems are coordinated and how updates and content will flow from one system to the other. It is often simpler if any changes are synchronized automatically, however a well-defined manual process with quality control checks is sufficient. Alternatively it is also possible to define a preservation policy that defends why updates at the delivery or preservation site do not need to be propagated to the other copy of this content. Such an approach will be necessary in many cases.

Description and metadata

Metadata standards used include *METS*, *MIX* and Dublin Core.

Manual processes leave room for manual errors and make coordination of delivery and preservation objects more difficult through time.

Digital Preservation:

As noted, The National Archives brings much of its experience with the preservation of print materials to its

treatment of digital materials. They have a digital preservation policy in place and in 2007 developed a specific policy for digital surrogates. The preservation approach for this data is based on migration of formats and refreshing of hardware. Masters are backed up on servers and DVD, tape copies are kept off-site and at The National Archives. This is also supplemented by a detailed disaster recovery plan and regular ongoing risk assessments. Given the role and expertise of the The National Archives, it is hardly surprising to find that these interlocking and complimentary policies are already in place. Other institutions wanting to maintain digital content in the long term may want to develop similar policies. Funders may wish to take account of the existence of such policies before awarding grants for digitisation.

Preservation of Teaching Materials

For the Cabinet Papers project, The National Archives has developed a series of study and contextualization packages. These are backed up on web servers along with the rest of the National Archives website and are also backed up to tapes which are stored off-site. The cabinet papers web pages are a part of the institution's disaster recovery plan for the web if these circumstances were to occur.

Summary

The project is a good example of a large national institution managing a project such as this to a high standard. It can often be assumed that such a large institution has the infrastructure, resources and skills which have made this project work have been planning and collaboration, i.e. good management. The National Archives have been active in the area of digital preservation for many years and the preservation plans they have in place are strong. It is interesting to note the wider policy framework within which digital preservation sits. In addition to a digital preservation policy, there is a specific strategy for digital surrogates, and an extensive

and appropriate set of disaster recovery plans. This wider framework lends credibility to the digital preservation policy. In addition, the project and The National Archives have demonstrated productive teamwork and a willingness to share experience.

Given the role and expertise of The National Archives it is hardly surprising to find that interlocking and complimentary policies are already in place. Other institutions wanting to maintain digital objects in the long term may want to develop similar policies. Funder may wish to take account of them.

The recommendations regarding the preservation of these digital images created by the project should only serve to increase the profile and quality of this successful project in relation to its sustainability and longevity.



Cabinet papers are online at:
<http://www.nationalarchives.gov.uk/cabinetpapers/>

This case study was prepared by Patricia Sleeman of the University of London Computer Centre, with the assistance of Portico and the Digital Preservation Coalition. It was made possible with funding from JISC and the support of The National Archives. Produced September 2009 for publication in 2010

