



JISC Final Report

Cabinet Papers 1915-1978 (British Governance in the 20th Century)

Laura Withey, Jane Gray, 20/01/2009

Table of Contents

Contents

JISC Final Report	1
Title Page	Error! Bookmark not defined.
Table of Contents	1
Contents	1
Acknowledgements	1
Executive Summary	1
Background	2
Aims and Objectives.....	3
Methodology.....	4
User Engagement	6
Implementation.....	7
Outputs and Results.....	9
Outcomes	10
Conclusions.....	14
Implications	14
Recommendations (optional)	15
References.....	15
Appendixes (optional).....	15

Acknowledgements

The Cabinet Papers Project was funded by the Joint Information Systems Committee as a part of the JISC Digitisation Programme and led by the National Archives. With special thanks to Paola Marchionni, Alastair Dunning, Wave, Dr Ed Hampshire, Dr Nigel Ashton, Dr Martin Doherty, Margaret Jones, Dr Chris Murphy, Dr Philip Murphy, Dr Glen O'Hara, Dr Helen Parr, Dr Hugh Pemberton, Dr Richard Toye, Esther Arnott, Professor Saki R Dockrill, Professor Bernard Harris, Professor Peter Hennessy, Professor Rodney Lowe, Reuben Moore, Tessa Stirling, Transmedia Technology, Alison Webster, Caroline Williams, Mike Waugh, Gemma Richardson, Amanda Russell, Emma Allen, Angela Mullen, Vera Mehta, Nicola Coulson, Roumiana Kurktchian, Ben Walsh, Dr. Duncan Redford, Dr. Dan Gilfoyle, Hannah Edwards, Elizabeth Loving, Martin Bazley and all the many others who kindly contributed their time to this project.

Executive Summary

From its outset the Cabinet Papers Project aimed to provide an online resource for learning and research that would make The National Archives' holdings of the Cabinet

Papers available to the public, and in particular to Higher Education and A-level students. The aims and objectives of this project have remained consistent and The National Archives is now able to provide an entire collection of searchable digitised images of the Cabinet Papers to users of the site as well as providing relevant information and study guidance for both teachers and students of British 20th-century history.

This project broke new ground for The National Archives in the use of Optical Character Recognition (OCR) technology. The ease of access that the OCR provides to the documents is of great value to research, however this project has also discovered the limitations of OCR in its inability to recognise images, foreign text etc. and the use of manually transcribed metadata is useful here in ensuring access to the document. The sheer quantity of digitised content used by the Cabinet Papers project has precipitated a move forwards in The National Archives' processes for the handling and loading of documents for digital storage and presentation. New processes have been developed for Quality Assurance (QA) and correction of OCR and for managing document releases within the collection.

The Cabinet Papers site has been tailored to the needs of its projected users in a drive to make the content, not only available and accessible to the public, but useable and relevant for the intended audience. Part of this user centred design process has been the creation of the study packages and educational information available on the site. The large amounts of additional content involved in this project has allowed The National Archives to increase its experience of using internal content management systems and the Cabinet Papers site now provides a structured set of informative pages that can be added to and expanded in the future.

Background

The Cabinet Papers project is lead by The National Archives and was funded under the Joint Information Systems Committee (JISC) digitisation programme. JISC is a joint committee of the UK Further and Higher Education funding bodies, and is responsible for supporting the innovative use of Information and Communication Technology (ICT) to support learning, teaching, and research. Its overall aim was to provide researchers, learners and teachers with an online resource discovery and contextualisation package with comprehensive access to the papers of Cabinet, throughout the 20th Century.

This project provides online access to the minutes and memoranda of Cabinet meetings from 1915 – 1978. It offers a comprehensive online learning resource and opens up this extremely valuable collection in new ways to people all around the world.

This collection, which includes some of the seminal decisions of the 20th Century, is of relevance not only to those researching modern British History but to many people who have lived through, or been affected by, the events covered in the records.

The Minutes and Memoranda constitute a fascinating record of the way in which the British government grappled with the events of 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 National Archives has developed the web pages' content which is aimed at A-level, Further and Higher Education students and many of the topics included are key parts of the A-level examinations syllabus and university programmes. By digitising these

records, The National Archives is making them available to all, not just to those that are able to travel to Kew, and so filling a gap in the availability of primary source material for students from A-level through to Higher Education.

The new resource has been digitised from microfilms held at The National Archives that contain a large collection of volumes of documents. It builds upon and adds value to other thematically related digitised collections focusing on the political landscape of Britain in the 20th Century, which have already been made available. These services are frequently used by researchers, students and teachers, for example: DocumentsOnline, The Macmillan Cabinet Papers 1957-63 Online, BOPCRIS and Newsfilm Online.

Aims and Objectives

The project aims were:

- Create a facility for free text and category searching based on both the study packages and the online collection
- Create a set of study packages for A-Level up to post-graduate on a number of relevant topics and link these to the content presentation system
- To provide both the collections and 'study packages' free of charge to schools for 10 years and free of charge to higher education for 5 years

The Project Objectives were:

- To create over half a million searchable digitised images from the Cabinet papers held at The National Archives and make them available to all through The National Archives website.
- To develop a sophisticated online resource discovery and learning packages that open up research opportunities that paper and microform-based finding aids cannot deliver.
- To provide access to material for use in A-level and Higher Education courses with or without the contextualisation material also offered by The National Archives.
- To undertake evaluation of the 'added value' which this service provides with schools, higher and further educational institutions.

None of the aims or objectives were changed throughout the project, with the exception that category based searching was originally envisioned to be provided via topic mapping, and is now provided via the free text and metadata search facility, in combination with the browse sections of the website.

The latter of the above listed objectives, evaluation, will not be fully realised for some time after the completion of the project as the service will need time to "bed in", or become established with, students and lecturers, especially as the service launched part way through the school year. A formal summative evaluation will, however, be made within the project closure. This will look at: the success of the project overall, the development process, how useful the service is, lessons learned and suggestions for

further development. The report will provide the basis for further reviews of the service to be carried out by the Education department after the product is bedded in.

Methodology

Overall Approach

The project followed PRINCE2 methodology. This approach is widely used within Government and also the private sector, and provides a structure which staff at The National Archives are experienced at working within. The Strategic Development Department provided overall project management for the project and relevant Departments or teams within The National Archives managed their individual elements and work-packages. Each area was represented on a project board with the Head of Strategic Development acting as Senior Responsible Officer and chair of the Board.

The Project Milestones were as follows:

- Completion of scanning - Feb 2008
- Completion of data creation - May 2008
- JISC Progress reports (3 per year) sent end September 2007, end Jan 2008, end May 2008, end September 2008, final report Feb 2009
- Completion of A level study packages which receive positive comments for the schools community, match with curricula and complement The National Archives' existing educational packages - September 2008
- Completion of HE contextualised packages which encourages use of this online resource in HE institutions - October 2008
- Formal evaluation of both HE and A-level study packages and content - Feb 2009

All milestones up to the formal evaluation have been met successfully and the formal evaluation target is on track to be met comfortably.

The project benefited greatly from the additional structure provided by the Advisory Panel (previously called the steering group) and the Peer Review Board of Academics, who provided guidance and reviews of the site content. It also benefited from following the National Curriculum in development of the A-level content, which allowed for review and guidance from A-level examination boards and teachers.

An overall work-flow diagram and detailed work flows are attached in Appendix A

Detailed work strands

Image creation, loading and image delivery:

The project created digital images scanned from film and loaded them into the DocumentsOnline system. This used pre-existing processes and procedures already well established at The National Archives, including, scanning of the film, QA and inserting the most up to date versions of the documents, all of which functioned well. New issues did arise, around versioning (where a previously redacted image is now un-redacted) and robust new processes (see Appendix A) were developed for handling these, within the existing management structure. There was also a need to develop a new process for load of information that could be indexed by the search facility (XML files of OCR) which is not normally held in the DocumentsOnline system. In order to do this a new database had to be created which linked the references and information in DocumentsOnline to the XML file references.

Many of the PDF images were also loaded directly into the web pages. This was a relatively new approach to providing access to records and as such raised some issues around bandwidth, due to the size and quantity of the files. Using an existing approach at The National Archives (trialled via the provision of the UFO's records) these issues were addressed by providing the PDFs through externally hosted servers.

OCR and Metadata:

An OCR service was procured and Adobe Acrobat was used to produce PDF images for delivery, in line with the other collections in the hosting system. Although Adobe is propriety software, licenses are free to use and it is widely used. Document descriptions or metadata, were also transcribed manually using standard processes at The National Archives and QA'd to ensure that this was to a minimum 98% accurate standard.

Following trials of the OCR process, The National Archives developed a two part system for improving and correcting the quality of text produced via the OCR, first by identifying keywords which were frequently miss-read, by comparing the text against dictionaries and then automating the correction of these. We have looked at the overall accuracy of our collection and identified that those files which were classified in the 0-5% category for accuracy, actually contain pages which are blank, or are graphics or foreign text that could not be read by the OCR, and hence gave a misleading accuracy rating. As a result of the auto-correction, and after deducting the 0-5% accuracy statistics, the average word accuracy rating of the collection is now 96%. Following on from this some of the least accurate OCR text (which was identified as English prose) has been manually re-transcribed up to the 90% marker.

Search facilities:

The National Archives has used the IDOL 7.0 search engine supplied by Autonomy to power searching of the Cabinet Papers. This software provides the facility for both free-text and metadata based searching. Testing of the service was used to ensure that the free text searching was effective when used in conjunction with OCR of PDF scanned images of the Cabinet Papers features.

Editorial:

The editorial process was developed in such a way that each batch of content was reviewed by academic peers and amended as appropriate. This dialogue was continued for each batch until the content was approved by the peer reviewer. The content was then reviewed by the web editor to ensure a consistent "voice", migrated into the content management system and then put through a final review by the web team editors.

Testing:

The project followed a process of on-going iterative testing throughout its duration. User testing was out-sourced to an external supplier while functional testing and security checks were carried out in-house. The full user testing and research process is detailed below under 'User Engagement'. Functional and security testing took a structured approach under the direction and control of the Test Manager and according to a test plan drawn up prior to the delivery of products.

Preservation of images and technical metadata:

Delivery versions and preservation versions of the images and metadata have been created. The preservation images are large high quality TIFFS and preservation

metadata is being converted to MIX v2.0 for images and METS encapsulating metadata mapped to Dublin Core resource descriptions. Preservation images are stored within the systems used for digital surrogate storage at The National Archives and actively maintained under the digital surrogates policy.

User Engagement

At the start of the project, the first scoping carried out was to assess the needs of the intended audiences identified in the bid. One to one interviews were carried out with key providers in the academic arena and these were used to concept shape the potential structure and shape of the study packages to be provided.

Iterative user testing has been used throughout the project, which was based on a user centred design approach, to ensure that users needs have remained central to the project. Testing has included class room testing in schools, one to one interviews, trials of the content and expert reviews. An external company, ICT4Learning, was used to carry out the usability testing and formal evaluation of the website. The results of each round of testing were fed back into the development of both the website and the content.

Throughout the project, the contextual content being created was peer reviewed by key academic contacts who were able to ensure that a balanced approach was taken in creating the content and that content was consistent with current approaches to teaching and lecturing. This was important in meeting user needs, both from the point of view of the user group represented by the Peer Reviewers (Academics and lecturers) and from the point of view of usefulness to students – if the content did not fit with the teaching of their course content it would quickly become less relevant to the users.

The A-level section of the website was also reviewed and approved by the examination boards Edexcel and AQA. The reference to the site can be seen from the Edexcel website here: <http://www.edexcel.com/notices/Pages/history-website.aspx>
This ensures that the content meets the National Curriculum, and is therefore able to meet the needs of A-level teachers and students, as identified in the initial scoping.

The National Archives already has systems in place for receiving feedback from users on web services after they have been launched, either online or through our contact centre, or through the reading room user forum and online user panels. As well as receiving positive user feedback from all of these forums the Cabinet Papers site received positive coverage in the press following its launch (3rd December, 2008). Articles were picked up in the national, regional and trade press, and in a variety of online publications. In terms of newspaper coverage, we picked up 14 articles (3 national, 11 regional daily press, 2 regional weekly press, 1 trade press). This coverage represented an AVE of £26,961 and circulation coverage of 7,756,000. In terms of online coverage, the launch of the website was covered on a number of online news websites (including MSN, Telegraph.co.uk and SecEd), and was also featured on a number of university and academic sites and blogs (see Appendix C for some examples).

In order to encourage usage and “bedding in” of the website, we are also producing promotional posters and postcards which will be sent to schools and institutions for display. We have also disseminated information about the site at the BETT educational show, will be liaising with the NEN via our Education Department and have some visits to Higher Education institutions organised, where the Cabinet Papers will be discussed

as a part of the services provided by the National Archives. In addition, the Cabinet Papers will be included in the ongoing marketing work for the National Archives website.

Implementation

The project has been conducted under the terms agreed with JISC. There have been no alterations to these terms and only minor alterations to the agreed implementation, which have been agreed with JISC. These are:

- Minor changes in the budget profiling (capital/recurrent split)
- Change to the start date (agreed by 10 Jan 2007)
- Removal of CAB 22 from series to be scanned (due to poor legibility of paper/text/images)
- Inclusion of additional content from recent years (1976,77 and latterly, 1978) and from the Cabinet Secretary's notebooks and the precedent books

Scanning.

The scanning and cropping of documents was achieved according to the project plan and was complete by May 08. Early on in the project The National Archives carried out some trial runs with OCR software and identified that the microfilm was of sufficient quality to be suitable for scanning to a quality that could be used for OCR. As the content of the records often has partial releases of previously redacted content, the process developed for updating this into the files had to be carefully considered in order to ensure the correct file and place were found for insertion of the content and that the system was able to allow this. The workflow for the scanning and redactions processes are included in Appendix A.

OCR.

All records had been OCR-ed by May 08. The standard of OCR was generally very good, however, the QA process identified a number of files with pages of poor OCR within them. Although this was sometimes caused by the state of the original files (e.g. thin paper used during the war years) or the microfilm quality (despite master versions being scanned for the best possible clarity), there was often no apparent indication as to why one page had been read clearer than another by the process. In order to provide the best possible search therefore, The National Archives improved the quality of the OCR as much as possible across the board. As indicated above, this was done by focussing on keywords rather than accuracy ratings. Often the accuracy ratings for individual pages were not representative because the page contained either foreign text, graphics or was blank. Following on from this a further amount of manual re-transcription of OCR was done to bring the majority of cases of poor OCR of English prose up to the 90% accuracy mark.

Transcription

Early on in the project it was agreed that document descriptions should be manually transcribed, as for other digitisation projects at The National Archives, in order to mitigate the risk posed by the OCR. It was felt that if the OCR had been of a quality which did not deliver satisfactory search results, there would still be the metadata to fall back on for document accession. As the project developed and we found that the OCR was of a high enough quality to search, in fact often many results were returned, but it was, nevertheless, useful to have the metadata as a means of narrowing the search. This was implemented by adding an option to the search, based on user testing, which allows users to narrow the search to "document descriptions only", so reducing the number of results returned. The transcription of the metadata was outsourced, following

standard processes at The National Archives, and QA'd in house to a level of 98% accuracy.

Content and Records

Once the structure of the A Level and HE study packages had been agreed content was written in batches, and as described above, peer reviewed by contemporary academics. Once each batch had received approval from the board, content was passed through a thorough editorial and proofing process whilst also being migrated into the Red Dot Content Management system.

Throughout the project a number of enhancements were made to the systems for hosting the collection, in order to improve the efficiency of the Cabinet Papers site as a learning and research resource. In particular, the way in which DocumentsOnline interacted with the Autonomy search function has been addressed, simplifying the route to the images from the search result. The web pages hosting images directly, as opposed to via the DocumentsOnline delivery system, have also been developed.

New processes have been developed (detailed above and in Appendix A) to facilitate the load of documents. Improvement has been made in the interface between the DocumentsOnline system, the web pages and the search facility. Issues arose around this which were related to the complexity of the back-end software/system which runs the search. It was difficult to set up the functioning search facility and several bugs were successfully found via testing and logged as issues. They were then corrected via the Development Team. As a result of the complexity of developing the search, it was made clear early on in the project that Search as a whole needed a product owner as well as to have technical specialist knowledge. This role is now filled and there has been a gradual increase in knowledge shared about the tool amongst The National Archives' staff.

Building the Site

A site map was completed early on in the project and a high quality final product was received from our supplier for the development of our online designs and interactive elements. We identified that the project would be developing alongside the re-development of The National Archives website, and as such, ran the risk of clashing in resource availability and of design clash, particularly as the Cabinet Papers designs were beginning to be developed slightly earlier than the work on The National Archives' homepage. In order to mitigate these risks, after completing preliminary design and structure work, we outsourced the template design in order to relieve pressure on internal resources. We liaised closely with the Web Team as the plans for the website redesign developed and ensured that as soon as guidelines were created these were passed to our suppliers so that they could be incorporated. This has resulted in web pages which fit well with the overall site structure and look and feel.

The design work for the look and feel of the site was complete by Oct 08 although some development work did continue on the interactive applications up to the launch date (Dec 3rd 2008).

Testing

The testing and user testing approaches are detailed above. Issues faced arising from the testing, which was carried out iteratively throughout the development process, mainly stemmed from a delay in the delivery of some of the products from developers.

This delay was a result of the number of changes that were made and led to time pressure in carrying out the testing and fixing of issues. For future projects, it would be recommended to plan in a long testing period between final delivery of the products and the launch date. Having said this, all fixes and work on the project were carried out within the timeframe available at The National Archives for publishing the website, and several months prior to the final date for completion required by JISC.

Outputs and Results

The main deliverable of the project is the Cabinet Papers site at the following address:

www.nationalarchives.gov.uk/cabinetpapers

The entire collection of records is searchable from the web pages and the records can be downloaded from The National Archives' DocumentsOnline service. Key documents have also been selected to illustrate major themes and topics and are available direct from the web pages at the above address.

This online resource makes use of a host of innovative features such as optical character recognition, which means that the entire contents of the collection are fully searchable. Of the records themselves, we have produced scanned large TIFFs for preservation purposes, and delivery sized PDFs for presentation. All the images are scanned to 300dpi greyscale TIFF. The PDFs have a layer of text embedded within them, which means that it is possible to use the Adobe software functionality to search within a PDF. The OCR process also allowed us to produce XML files containing the transcript of the text of each image. This is the text which is searched by our search engines in combination with manually transcribed descriptive metadata which is also plugged in to DocumentsOnline.

Other features of particular interest on the website include a tool that allows users to pan around world maps and see the changes in geo-political boundaries over time. This has been developed with world-view and zoom perspectives, and with information "hotspots" which provide additional information about key events or decisions discussed within the Cabinet Papers. Some of these information points are classed as "resource links" and give access to visual extracts, such as maps, photos, and statistics, from the original documents.

The writing frame is the other key application on the site, which aims to support students in their understanding of writing essays based around primary historical sources, such as the Cabinet Papers.

As a part of developing the web pages we have also obtained licenses for the use of various cartoons, images and videos on the site for the next 10 years. These help to highlight the key points from the records, helping students to engage with the content, and to create a visual snapshot of the period. We have created two "starter activities" for A-level students based around these as well as detailed teacher's guides which help aid teachers in using the content in schools.

As a part of the project we have successfully passed through three individual evaluations or "health-checks" which ensured the project was running satisfactorily and three successful reports and action plans were produced as a result. The "health-checks" were based on the Office of Government Commerce's Gateway approach, which stipulates key points at which to run these checks and covers areas such as: the

procurement carried out within the project, the issues considered prior to launch and readiness to launch. The site has launched successfully and site statistics for December have been positive, with 10,235 visits in the month.

Enhancements to the advanced search have been completed so that users are now able to search web pages themselves as well as the documents.

Outcomes

The project has successfully achieved its aims in the developed site. The entire collection is digitised and available through The National Archives website and the DocumentsOnline delivery system. Free text searching of the entire content, as detailed above, has been made fully available for all the content and functions to a high level. In combination with this and the browse facilities provided, the site opens up research opportunities that paper and microform-based finding aids cannot deliver.

The key documents under a particular theme have been picked out ready for the researcher and then keywords can be taken from these, and plugged into the free text search to find other related documents that would have been very difficult to identify otherwise. Because of this we expect the site to be of huge benefit to anyone studying or researching the Cabinet meetings and courses of action that arose out of them, from researchers to students and lecturers.

The site provides contextual information, in the form of study packages for A-Level up to post-graduate level on many different topics and links these back to the records themselves in a meaningful way. This content, as well as the access to the records themselves, is provided free of charge to schools for 10 years and to Higher Education institutions for 5 years. To ensure that this service also provides an added value to these customers, we tested and reviewed it thoroughly with these audiences as we developed and are making an initial evaluation of the value in the final evaluation report currently underway.

In addition, the service is being provided free to all for a minimum of 18 months from launch (Dec 3rd 2008). This will be of benefit to the general public in helping to bring history to life for our users and allowing them access to records of discussions that took place on the key decisions which have shaped Britain and the wider world.

Throughout the course of the project, we encountered a number of issues, which we had to overcome or manage in order to successfully complete, and a number of areas where we felt we had particularly succeeded. A full log of these was kept and there are a number of lessons that we have learned from these which may be beneficial to other projects in the future. The key points are laid out below:

1. **Project Methodology, Management, Planning and Strategy:** The project methodology was successful. There were a number of minor issues in planning of the project, particularly in the bid stages, which TNA successfully overcame and have led to changes in TNA's system for bidding and should be of benefit to future projects. For example, the timescale for the set up of the project was underestimated and there was a failure to fully establish early on whether in-house skills and resources would be sufficient for project needs. Closer communication with resource providers and time dedicated at this early stage to the establishment of clear timescales should, in future projects, eliminate the problems encountered due to this. In particular there should be an awareness of

interfaces and dependencies between project elements and other projects. Despite the minor issues detailed above, the project was able to successfully resource gaps in available in-house skills by outsourcing some elements, e.g. web-design.

As a part of the project methodology, regular team meetings were invaluable throughout the project, providing a forum for discussion and for the exchange of knowledge and expertise. There was a general impetus within the Project Team to learn new skills and share knowledge, which was fostered by this methodology and resulted in widespread upskilling, with staff learning new techniques and sharing training with others. This has been of great benefit to the organisation as a whole, and of particular benefit to the Web Team who have advanced their team knowledge of the skills required for using content management systems; in the case of this project, a system called RedDot.

The project also benefited from the establishment of an Advisory Panel, comprising senior representatives from education and research and a panel of Peer Reviewers from the same sectors. They provided The National Archives with an objective view and constructive criticism on its plans and contributed to the content, lending the project authority and transparency. However, in particular relation to the Peer Reviews, this did slow the editorial process down, with return of completed work often being un-predictable and this affected the ability to deliver the content for migration into the content management system in a structured way. The project would therefore have benefited from devising a more structured workflow for the writing, editing and uploading of content in ordered sections.

The project would have benefited from having specific and consistent points of contact in the Web and Content/Editorial Teams as there was sometimes confusion over the progress of work, or work flow, due to information being communicated from multiple individuals and perspectives. There was also a relatively high turnover of staff working on the project, and this did cause some confusion over roles.

2. User Centred Design: The project greatly benefited from early and continuing iterative user testing to create the site around a user centred design principle. Early informal discussions with primary user groups were critical in ensuring that comments and views were incorporated at an early high level design stages. However, towards the end stages of the project, there was an issue with the integration between the continuing requirements for change from the iterative user testing being carried out, and the needs of the organisation, the flexibility of the system that the site was being developed in, the timeframe available for making changes, the available budget and the change management processes. There is a very real need for any future projects using this approach to consider, before they embark on any testing, at what point a line should be drawn under the user testing and how good is "good enough" for the users. This is particularly true when it comes to site structure, which needs to be fixed early in order to allow for the development work to take place. Projects should also be encouraged to consider what other requirements will need to be integrated into the site development as well as the outcomes of testing, at the earliest possible stage of the project. At The National Archives, it is recommended that the web-

team user specialist is closely involved throughout future projects, even where user testing is outsourced.

3. Scanning and OCR: OCR was a new technology for use on this scale at The National Archives. The OCR was completed very successfully, after automated correction of keywords, using the method of working out explained above under "Methodology", the average overall word accuracy rating across the collection is 96% (once blank pages, graphics etc are discounted). The full data and calculations on OCR statistics are included in Appendix B. We have also carried out some testing using the search facility and found that generally the OCR provides a good level of search ability and returns the documents expected for the keyword searched.

Although we have a good overall result for accuracy, this does not mean that all pages of OCR are as highly accurate as each other. We have found that files may contain many pages of very accurate text and then a page of much poorer text. As detailed above, this can be due to the quality of the original file or microfilm, or due to marks on the page, even though the best available quality microfilm has been used for scanning to give the best possible recognition from the OCR. It can also be a misleading accuracy rating caused by the content of the page (e.g. a blank page). Equally however, in a small number of cases there may be no obvious cause for the sudden relative loss in OCR quality. As some pages contain poorer text, and a small section of OCR text is displayed in the search results, we took the decision to also manually re-transcribe a majority of the English prose that falls under the 90% accuracy mark. Much of the lower accuracy pages are in French or German, blank, or contain images/graphics and so these have not been re-transcribed.

Issues with OCR accuracy which did occur, were due in large part to the complexity of the Cabinet Papers (foreign text etc. etc.). This complexity was only unveiled over time so establishing a definitive accuracy level to aim for across the board was somewhat arbitrary. With OCR in the future it is therefore difficult to say what an "acceptable" level of accuracy actually is. Each project will be distinctive and the accuracy is dependent on the actual quality, content and consistency of the documents and scanning. In general however, it is preferable to establish a guaranteed achievable OCR accuracy rating, where possible, at the outset, through a greater amount of preliminary sampling (e.g. identify how many have graphics that can't be read), preparatory work, OCR trialling, identifying potential problems etc and coordinate resources accordingly.

We have learnt a great deal from attempting to provide the best possible quality OCR, as detailed in other sections of this report from the processes used, i.e. auto-correction, focussing on keywords and scanning quality required to provide an acceptable OCR result. In dealing with these issues, the project has found successful solutions, however, this has been on a case-by-case basis as each issue was encountered. This approach was mainly caused by a lack of prior experience with OCR, although there is also the perception that OCR technology has not kept up with improvements in scanning and imaging technology. The National Archives is therefore continuing to monitor and document OCR developments and emerging technologies in the same way as imaging, to identify new developments and any impact they have had. For The National Archives, this in combination with the project Lessons Learned Report, will be critical for

any future projects using this technology in order to avoid re-invention of solutions that may already be available.

For any future projects intending to use OCR, it would be beneficial to ensure that OCR scans are carried out with UK English programmes, rather than US English as this undermines the accuracy of word searches in the documents. If possible it would also be beneficial to have foreign language scans run so that text, often in French and German, can be recognised as valid. This would raise accuracy ratings, but whether this is necessary will be dictated by the content itself. It is likely that future OCR projects using more recent documents will improve in their usefulness and accuracy as the original documents become of a higher standard – e.g. The text is clearer.

We have also found that since inputting the OCR text into our entire website search engine, that search results can sometimes be swamped by the quantity of Cabinet Papers text or other full text documents, over the metadata provided by other parts of the National Archives website. A project is being carried out to correct this by weighting different types of information provided to the search engine. It may be useful for other institutions to be aware that this kind of data can swamp their other search results, and may therefore need to be weighted accordingly.

4. Web development: Some difficulty was caused by the time allowance given for review and sign off of web designs or applications created by external providers. Significant time needs to be built into the project plan for this. External parties should be given notice of approaching reviews and The National Archives' teams carrying responsibility should have reviews booked into their calendars. Also testing timescales between the final delivery of some products and the proposed launch date were tight at the end of the Project. Thorough testing and fixing of problems arising would be aided in future if more significant amounts of time could be built into the project plan after the last products have been delivered and before the designated project completion date.

The project benefited greatly from the services of the Test Manager who coordinated all the functional testing and carried out technical testing. The Test Manager and those owning responsibility for possible changes arising from testing must however, establish a clear understanding on how testing issues are to be logged and dealt with in order to improve the overall feeling of coherency at the time of launch.

Some issues were encountered with separating the documents themselves, situated in the DocumentsOnline delivery system, and the contextualisation, situated on separate web pages. The link between the two is provided by the search function and this arrangement was necessary because the amount of contextualisation and interactive tools provided, did not sit with the other content in DocumentsOnline. The sheer quantity of records, search functions and options for payment models meant that DocumentsOnline was required for delivery of the majority of the records. However, the gap between the two types of content means that there are several clicks in journeying between the two areas and this does not provide an ideal user experience. This issue was mitigated by improving the journey as much as possible and providing key documents, relating to each study package or subject theme, direct from the web pages as well as through

DocumentsOnline. Future projects would, however, be advised to consider these issues carefully before beginning development and to consider the impact this separation can have.

5. Creation and Upload of PDFs: OCR work on the scale of this project has never before been carried out by The National Archives and it was found as a result that manual loading processes previously employed were inadequate for project needs. As a consequence new automated processes have been created. The quantity of digitised records is ever increasing and these new systems will be invaluable to The National Archives in the future. Future projects should however be aware that loading large numbers of XMLs and PDFs puts a strain on some computer systems and in this project actually caused a hard drive to blow. The National Archives still has considerable scope for improving its capacity to work with XMLs in the technical capabilities of the systems available for handling the files and will be looking to make strides towards this in the future.

Conclusions

The Cabinet Papers project has been an extremely successful project, delivering a useful, useable and relevant site, which is of benefit to a wide range of our users and opens up access to these records to everyone. We are confident that the usage of the site will in fact increase over time as teachers and lecturers pick it up as a part of taught courses, and the availability of the site becomes more well-known. A formal evaluation report will be provided to JISC at the final completion of the project.

The challenges that the project has overcome have been of long-term benefit to The National Archives, with skills and knowledge of all kinds of systems and processes improving across the board. We have gained a wealth of valuable experience through the course of the project, particularly in the area of OCR, and are now looking forward to the next strides that we can take forward in the delivery of excellence in our online services.

In particular, interesting conclusions can be drawn from our experience with OCR. For example, factors which can influence the quality of OCR are as often related to the content of the file or the language of the OCR program used as the quality of the image itself. It will be interesting to follow developments in the field of OCR correction to see whether more joined up approaches and solutions are developed, or whether the focus in this area is on quantity of data over quality.

Implications

From the focus in this report placed upon the OCR technology, there are clearly still developments to be made in this area. Given the difficulties faced at The National Archives in providing accurate OCR text to the user, it would be interesting to see whether commercial development will allow for the continuing improvement of either the creation or correction of OCR text, or whether instead the focus should be on providing a quantity of OCR direct to users. Should we in point of fact be allowing the correction to be carried out by the users themselves, rather than providing very accurate content ourselves? This is something that is already being carried out by the National Library of Australia with OCR from newspapers and The National Archives will be following their work in this area. There would be several issues for us to tackle if we were to contemplate this approach, for example, would we be potentially degrading the veracity of the content of our documents in the eyes of our users? If the OCR content is changed by an individual user to say something other than the text displayed in the image itself,

would the reliability and authority of our content be compromised, i.e. would the content we are providing then become misleading for our users? Would we wish to vet the content being corrected and would we have the resource to do so given the potential quantity that there may be available for users to correct? This is an issue which would bear a lot more investigation.

The Cabinet papers site is providing a valuable research tool to researchers and the academic sector, it will be interesting in the future to see how it is used and whether particular areas of the site are in fact more valuable than others. This is something that will be monitored by our trends analysis tools, and would be taken into account in any further expansion of the site. We have already identified that there are clearly opportunities to expand the site, for example there are cabinet committee records which could be added if the funding were available, and the site could be expanded to appeal to GCSE students as well as A-level and Higher Education. We will also be looking into text mining techniques available with full text data as a whole at The National Archives, and what steps could be taken forwards here. However any future development will be dependent on the available funding.

Recommendations (optional)

Recommendations for future projects have been integrated into the lessons learned above under “Outcomes”

References

The cabinet papers site developed by The National Archives is available at:
www.nationalarchives.gov.uk/cabinetpapers

The details of the National Library of Australia’s Newspapers digitisation project are at:
http://www.nla.gov.au/ndp/project_details/documents/ANDP_User_interaction_ideas.pdf
And their beta service allowing users to correct the text is at:
<http://ndpbeta.nla.gov.au/ndp/del/home>

Edexcel website: <http://www.edexcel.com/notices/Pages/history-website.aspx>

Appendixes (optional)

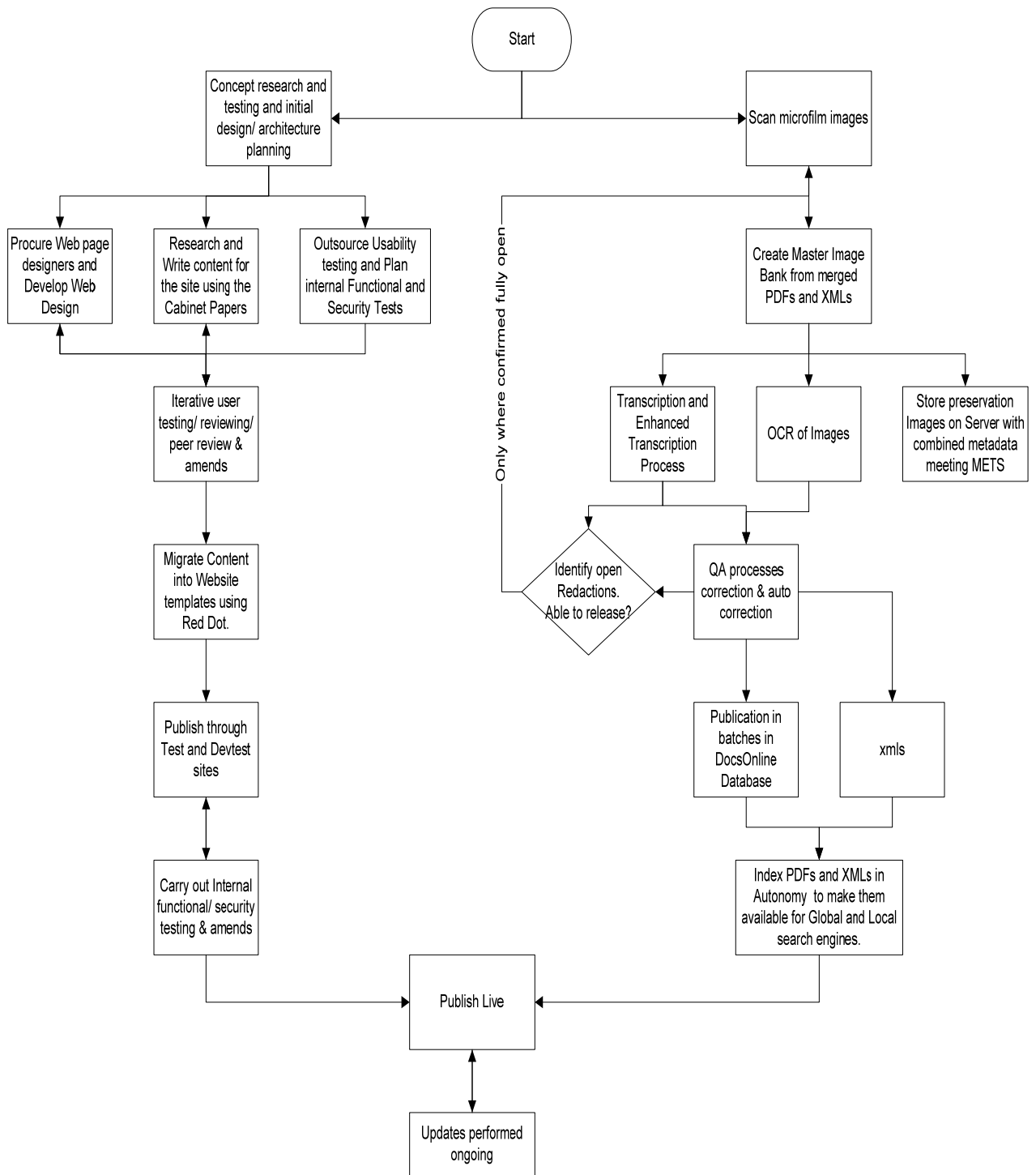
Appendix A – Workflows

Appendix B – OCR accuracy statistics

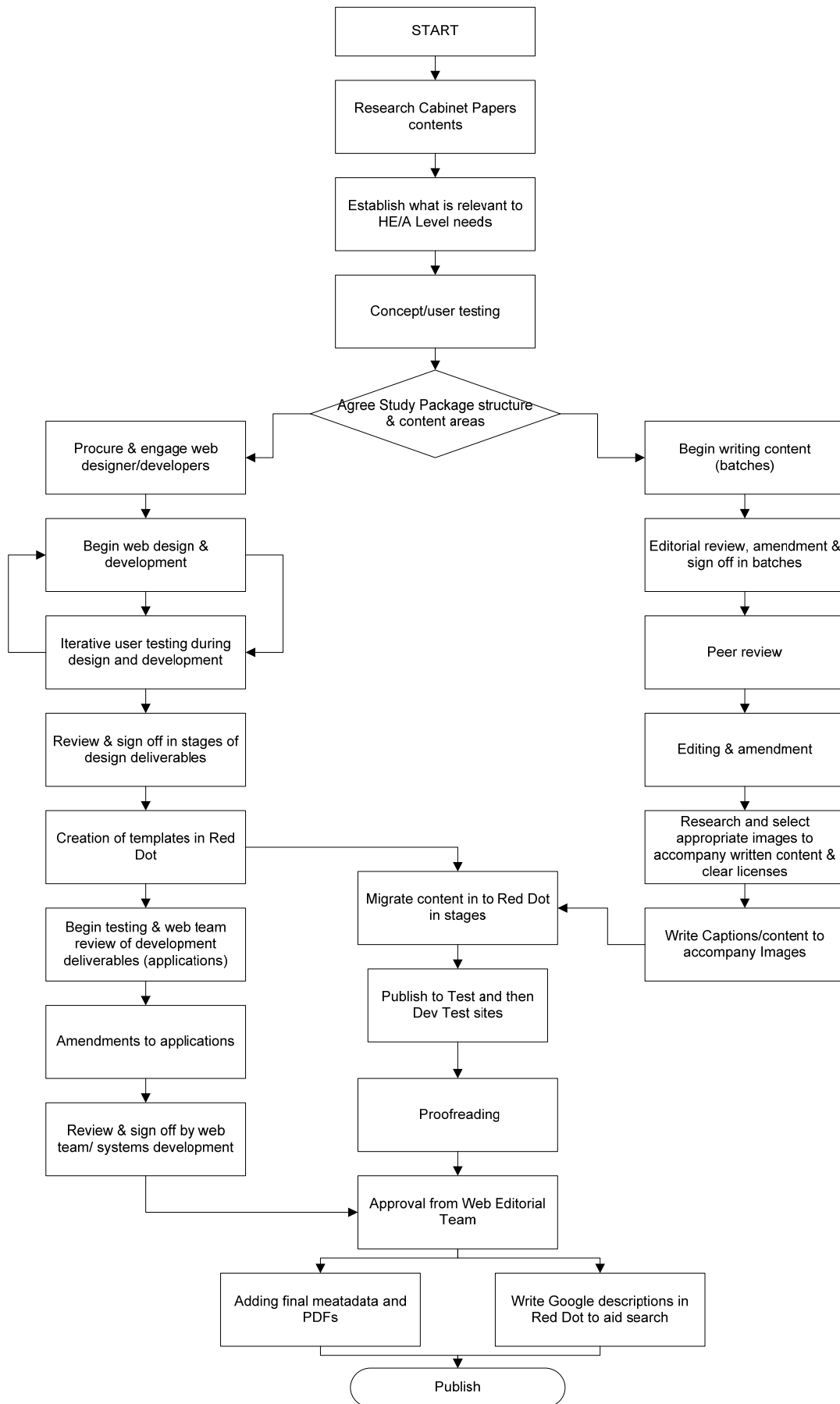
Appendix C – Example press coverage

Appendix A

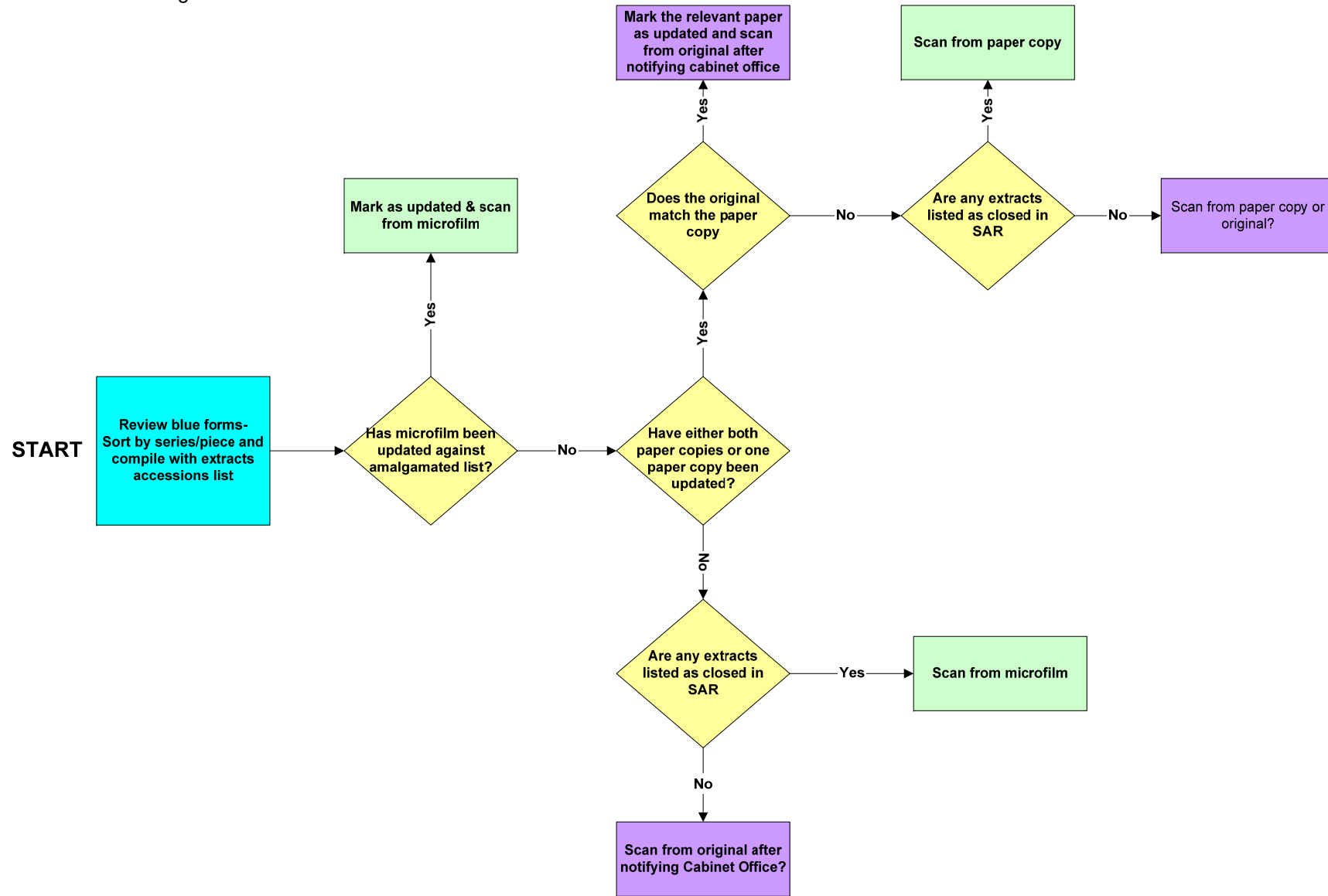
Overall Project Workflow



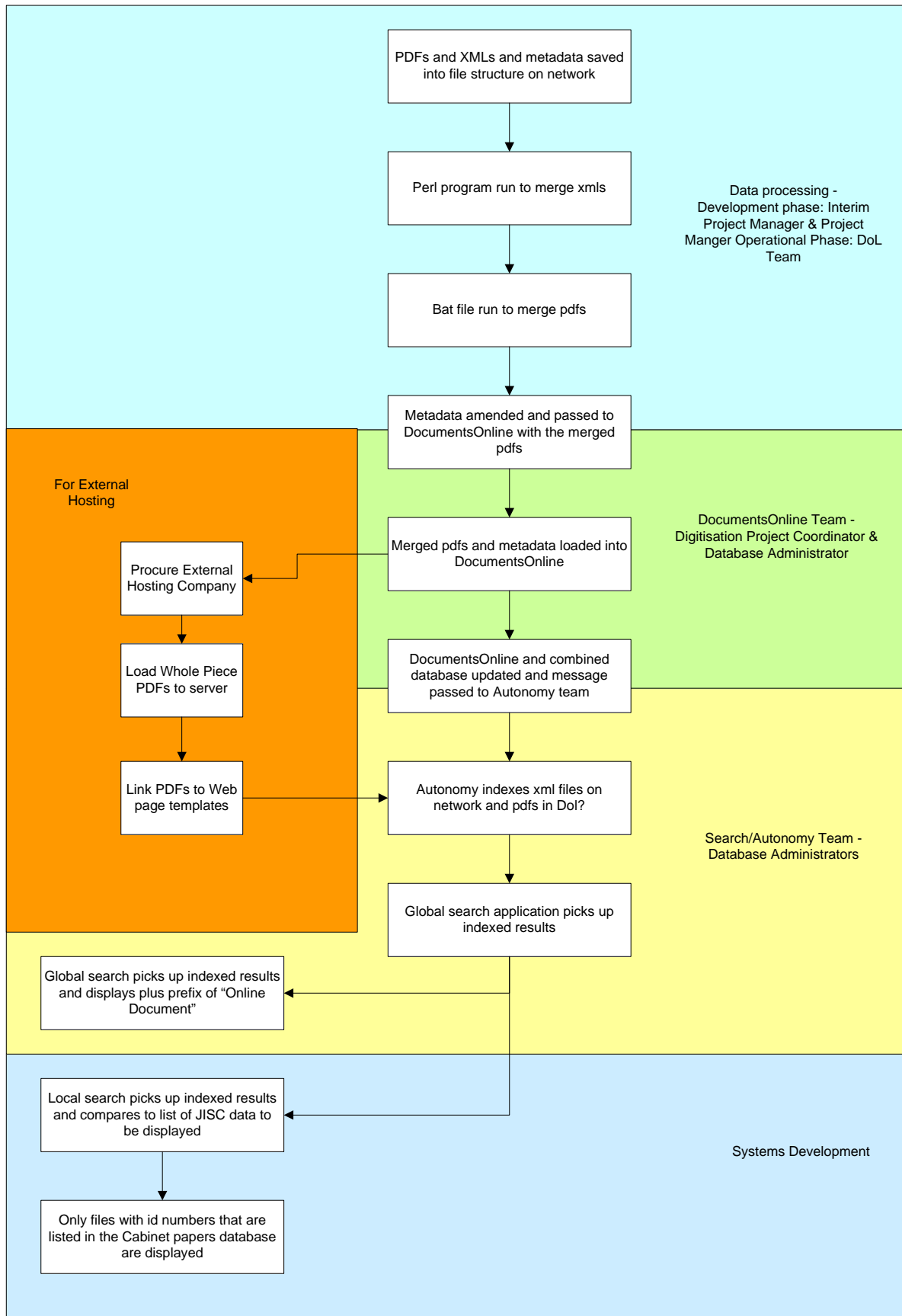
Editorial Process Workflow



Redactions Checking Workflow



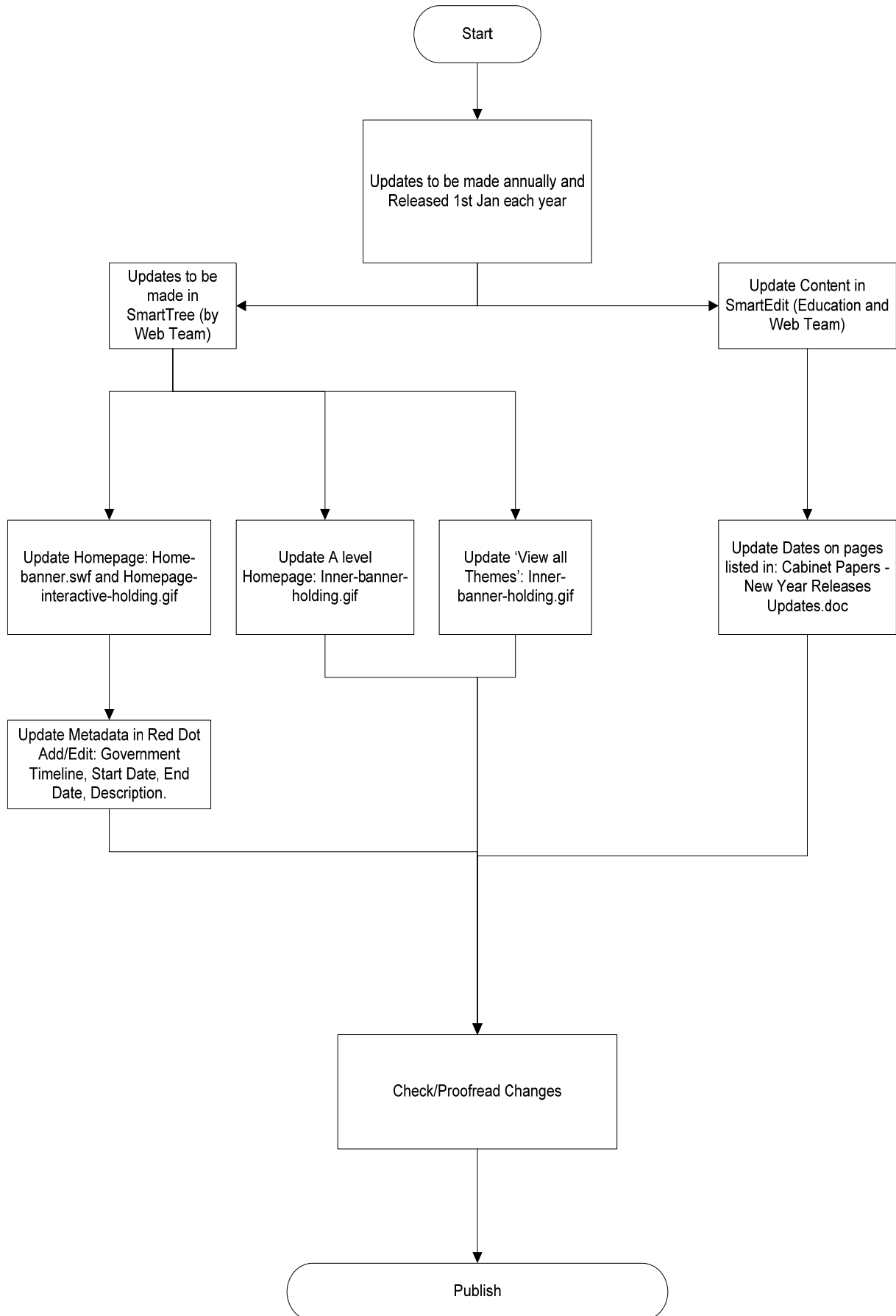
Load of Documents Workflow



Project Acronym:
Version:
Contact:
Date:

Web
changesWorkflow

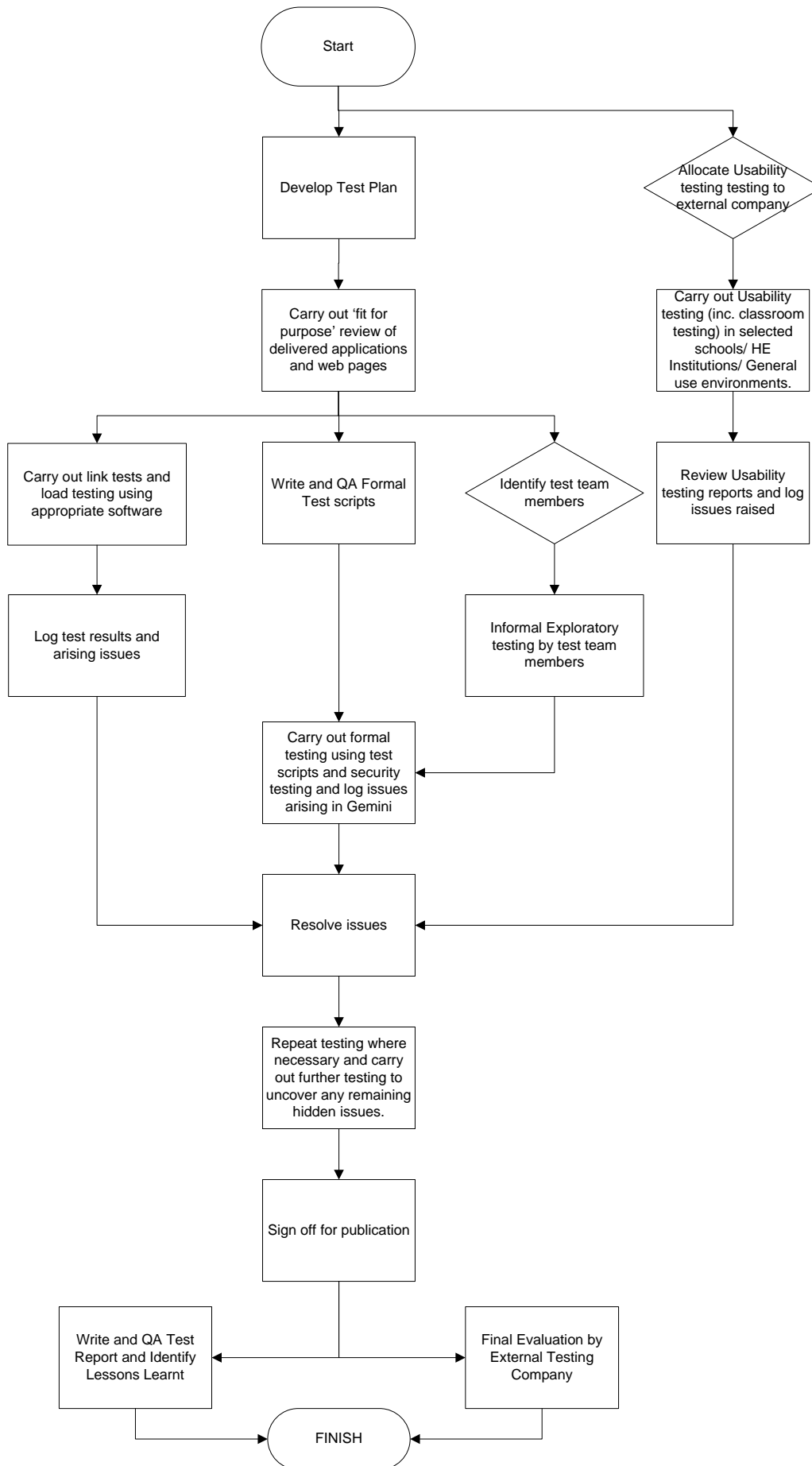
Project Acronym:
Version:
Contact:
Date:



Project Acronym:
Version:
Contact:
Date:

Testing Process Workflow

Project Acronym:
Version:
Contact:
Date:



Project Acronym:
Version:
Contact:
Date:

OCR Accuracy results analysis – (After correction of keywords)

The figures given in Tables 1 and 2 represent the OCR accuracy stats for the majority of the Cabinet Papers collection. At the time of sampling some of the files from later years were being used, checked and worked on for other purposes and could not be included in the report. Cabinet Secretary's notebooks are not included in these details as they are handwritten and were therefore not submitted for OCR.

The OCR accuracy program places files into accuracy rating bands. These bands are displayed as percentages of accuracy and are shown in the left hand column of Table 1 below, next to the classification series and number of files that fall into each band. The program does not give an individual accuracy rating for each file, however the next three columns represent the minimum, median and maximum accuracy ratings for those files, i.e. the two files from the series CAB 24 which are in the 16-20.9% band have a minimum accuracy rating of 16% a median rating of 18.45% and a maximum accuracy of 20.9%..

The three columns on the right hand side of the table correspond to the previous three and show the minimum, mid and maximum accuracy aggregated for the number of files in each band and series. For example, the aggregated minimum accuracy of the 2 files in CAB 24 which are in the 16 – 20.9% band is 32/200, while the aggregated minimum accuracy for the 8 files in the 21-25.9% band from CAB 24 is 168/800.

In order to assess the overall accuracy of the collection, we have therefore calculated a minimum, maximum and mid average % accuracy rating, which is given at the bottom of Table 1. We have selected the middle value as being most representative of the overall collection, which gives us an average accuracy of 96.17%. The collection is no less accurate than 93.72% and no more accurate than 98.62%

There were no files which fell into the accuracy bands between 6% and 15.9% so these bands are not displayed. Files which fall in to the 0-5.9% accuracy band represent files which contain foreign text, graphics or, in the majority of cases, are simply blank. These are provided in a separate table below. Because the ratings for this band are misleading, they have been removed from the overall accuracy assessment. These files are however of relevance since they give an indication for future projects of the likely amounts of OCR unreadable graphics, blanks foreign text etc which may be found in a collection of this type. The figures in Table 2 indicate that approximately 13% of the collection was unreadable by OCR. It also indicates the amount that the overall accuracy would be incorrectly skewed by if these figures were included – the minimum would become 81.49%, the middle value 84% and the maximum 86.52%

Graph 1 below shows the distribution of number of files across accuracy ratings. This shows how the majority of the files are in the top accuracy bands while there is a small peak in the lowest accuracy band where blank pages etc could not be recognised.

Graph 2, below, shows the average word accuracy by series, after the values for blanks, graphics etc have been removed and assuming the mid range value is used. This shows that there is a general trend for more recent documents to return higher accuracy ratings, CAB 128 and 129 series being the more recent and the CAB 23 and 24 series being older. The exception to this is CAB 68, which is a smaller series and so more subject to variation. It is also a memoranda series and contains a higher quantity of assorted additions, such as graphics and blanks. The number of OCR unreadable files in each series is shown in Graph 3 below. It is interesting to see that unlike the other accuracy stats there is no general improvement over time, i.e. the earlier series such as cab 128 and 129 are not more frequently in this band than the later series. This is because unlike the other stats which reflect the clarity of text in the documents being used (a factor which contributes to high OCR accuracy ratings), the 0-5.9% band contains all the blank pages, graphics etc that could not be read no matter when they were put into the files and regardless of the clarity of the text included in them.

Table 1

Project Acronym:
Version:
Contact:
Date:

OCR Accuracy Bands and calculations, 6-100%

Band	Series	No. files	Min accuracy value	Mid accuracy value	Max accuracy value	Min Av. Accuracy	Mid Av. Accuracy	Max. Av Accuracy
16.0-20.9	CAB 24	2	16	18.45	20.9	32	36.9	41.8
16.0-20.9	CAB 66	1	16	18.45	20.9	16	18.45	20.9
16.0-20.9	CAB 129	2	16	18.45	20.9	32	36.9	41.8
21.0-25.9	CAB 23	1	21	23.45	25.9	21	23.45	25.9
21.0-25.9	CAB 24	8	21	23.45	25.9	168	187.6	207.2
26.0-30.9	CAB 23	1	26	28.45	30.9	26	28.45	30.9
26.0-30.9	CAB 24	16	26	28.45	30.9	416	455.2	494.4
26.0-30.9	CAB 24	2	26	28.45	30.9	52	56.9	61.8
26.0-30.9	CAB 66	1	26	28.45	30.9	26	28.45	30.9
31.0-35.9	CAB 23	16	31	33.45	35.9	496	535.2	574.4
31.0-35.9	CAB 24	6	31	33.45	35.9	186	200.7	215.4
31.0-35.9	CAB 24	2	31	33.45	35.9	62	66.9	71.8
31.0-35.9	CAB 65	4	31	33.45	35.9	124	133.8	143.6
31.0-35.9	CAB 66	2	31	33.45	35.9	62	66.9	71.8
36.0-40.9	CAB 23	19	36	38.45	40.9	684	730.55	777.1
36.0-40.9	CAB 24	4	36	38.45	40.9	144	153.8	163.6
36.0-40.9	CAB 24	9	36	38.45	40.9	324	346.05	368.1
36.0-40.9	CAB 65	5	36	38.45	40.9	180	192.25	204.5
36.0-40.9	CAB 66	1	36	38.45	40.9	36	38.45	40.9
36.0-40.9	CAB 129	2	36	38.45	40.9	72	76.9	81.8
41.0-45.9	CAB 23	23	41	43.45	45.9	943	999.35	1055.7
41.0-45.9	CAB 24	58	41	43.45	45.9	2378	2520.1	2662.2
41.0-45.9	CAB 24	120	41	43.45	45.9	4920	5214	5508
41.0-45.9	CAB 65	4	41	43.45	45.9	164	173.8	183.6
41.0-45.9	CAB 66	6	41	43.45	45.9	246	260.7	275.4
41.0-45.9	CAB 67	3	41	43.45	45.9	123	130.35	137.7

Project Acronym:
Version:
Contact:
Date:

Band	Series	No. files	Min accuracy value	Mid accuracy value	Max accuracy value	Min Av. Accuracy	Mid Av. Accuracy	Max. Av Accuracy
41.0-45.9	CAB 68	1	41	43.45	45.9	41	43.45	45.9
41.0-45.9	CAB 129	22	41	43.45	45.9	902	955.9	1009.8
46.0-50.9	CAB 23	25	46	48.45	50.9	1150	1211.25	1272.5
46.0-50.9	CAB 24	269	46	48.45	50.9	12374	13033.1	13692.1
46.0-50.9	CAB 24	270	46	48.45	50.9	12420	13081.5	13743
46.0-50.9	CAB 65	11	46	48.45	50.9	506	532.95	559.9
46.0-50.9	CAB 66	14	46	48.45	50.9	644	678.3	712.6
46.0-50.9	CAB 67	7	46	48.45	50.9	322	339.15	356.3
46.0-50.9	CAB 68	1	46	48.45	50.9	46	48.45	50.9
46.0-50.9	CAB 129	89	46	48.45	50.9	4094	4312.05	4530.1
51.0-55.9	CAB 23	15	51	53.45	55.9	765	801.75	838.5
51.0-55.9	CAB 24	344	51	53.45	55.9	17544	18386.8	19229.6
51.0-55.9	CAB 24	187	51	53.45	55.9	9537	9995.15	10453.3
51.0-55.9	CAB 65	24	51	53.45	55.9	1224	1282.8	1341.6
51.0-55.9	CAB 66	1	51	53.45	55.9	51	53.45	55.9
51.0-55.9	CAB 67	5	51	53.45	55.9	255	267.25	279.5
51.0-55.9	CAB 68	1	51	53.45	55.9	51	53.45	55.9
51.0-55.9	CAB 129	147	51	53.45	55.9	7497	7857.15	8217.3
56.0-60.9	CAB 23	22	56	58.45	60.9	1232	1285.9	1339.8
56.0-60.9	CAB 24	177	56	58.45	60.9	9912	10345.7	10779.3
56.0-60.9	CAB 24	79	56	58.45	60.9	4424	4617.55	4811.1
56.0-60.9	CAB 65	13	56	58.45	60.9	728	759.85	791.7
56.0-60.9	CAB 66	7	56	58.45	60.9	392	409.15	426.3
56.0-60.9	CAB 67	1	56	58.45	60.9	56	58.45	60.9
56.0-60.9	CAB 128	2	56	58.45	60.9	112	116.9	121.8
56.0-60.9	CAB 129	51	56	58.45	60.9	2856	2980.95	3105.9
61.0-65.9	CAB 23	31	61	63.45	65.9	1891	1966.95	2042.9

Project Acronym:
Version:
Contact:
Date:

Band	Series	No. files	Min accuracy value	Mid accuracy value	Max accuracy value	Min Av. Accuracy	Mid Av. Accuracy	Max. Av Accuracy
61.0-65.9	CAB 24	98	61	63.45	65.9	5978	6218.1	6458.2
61.0-65.9	CAB 24	49	61	63.45	65.9	2989	3109.05	3229.1
61.0-65.9	CAB 65	43	61	63.45	65.9	2623	2728.35	2833.7
61.0-65.9	CAB 66	10	61	63.45	65.9	610	634.5	659
61.0-65.9	CAB 67	1	61	63.45	65.9	61	63.45	65.9
61.0-65.9	CAB 68	1	61	63.45	65.9	61	63.45	65.9
61.0-65.9	CAB 129	25	61	63.45	65.9	1525	1586.25	1647.5
66.0-70.9	CAB 23	73	66	68.45	70.9	4818	4996.85	5175.7
66.0-70.9	CAB 24	158	66	68.45	70.9	10428	10815.1	11202.2
66.0-70.9	CAB 24	101	66	68.45	70.9	6666	6913.45	7160.9
66.0-70.9	CAB 65	40	66	68.45	70.9	2640	2738	2836
66.0-70.9	CAB 66	15	66	68.45	70.9	990	1026.75	1063.5
66.0-70.9	CAB 67	5	66	68.45	70.9	330	342.25	354.5
66.0-70.9	CAB 68	6	66	68.45	70.9	396	410.7	425.4
66.0-70.9	CAB 128	4	66	68.45	70.9	264	273.8	283.6
66.0-70.9	CAB 129	50	66	68.45	70.9	3300	3422.5	3545
71.0-75.9	CAB 23	248	71	73.45	75.9	17608	18215.6	18823.2
71.0-75.9	CAB 24	321	71	73.45	75.9	22791	23577.5	24363.9
71.0-75.9	CAB 24	277	71	73.45	75.9	19667	20345.7	21024.3
71.0-75.9	CAB 65	84	71	73.45	75.9	5964	6169.8	6375.6
71.0-75.9	CAB 66	21	71	73.45	75.9	1491	1542.45	1593.9
71.0-75.9	CAB 67	3	71	73.45	75.9	213	220.35	227.7
71.0-75.9	CAB 68	19	71	73.45	75.9	1349	1395.55	1442.1
71.0-75.9	CAB 128	3	71	73.45	75.9	213	220.35	227.7
71.0-75.9	CAB 129	162	71	73.45	75.9	11502	11898.9	12295.8
76.0-80.9	CAB 23	645	76	78.45	80.9	49020	50600.3	52180.5
76.0-80.9	CAB 24	554	76	78.45	80.9	42104	43461.3	44818.6

Project Acronym:
Version:
Contact:
Date:

Band	Series	No. files	Min accuracy value	Mid accuracy value	Max accuracy value	Min Av. Accuracy	Mid Av. Accuracy	Max. Av Accuracy
76.0-80.9	CAB 24	392	76	78.45	80.9	29792	30752.4	31712.8
76.0-80.9	CAB 65	107	76	78.45	80.9	8132	8394.15	8656.3
76.0-80.9	CAB 66	59	76	78.45	80.9	4484	4628.55	4773.1
76.0-80.9	CAB 67	8	76	78.45	80.9	608	627.6	647.2
76.0-80.9	CAB 68	40	76	78.45	80.9	3040	3138	3236
76.0-80.9	CAB 128	11	76	78.45	80.9	836	862.95	889.9
76.0-80.9	CAB 129	171	76	78.45	80.9	12996	13415	13833.9
81.0-85.9	CAB 23	1605	81	83.45	85.9	130005	133937	137870
81.0-85.9	CAB 24	1364	81	83.45	85.9	110484	113826	117168
81.0-85.9	CAB 24	807	81	83.45	85.9	65367	67344.2	69321.3
81.0-85.9	CAB 65	201	81	83.45	85.9	16281	16773.5	17265.9
81.0-85.9	CAB 66	138	81	83.45	85.9	11178	11516.1	11854.2
81.0-85.9	CAB 67	20	81	83.45	85.9	1620	1669	1718
81.0-85.9	CAB 68	100	81	83.45	85.9	8100	8345	8590
81.0-85.9	CAB 128	30	81	83.45	85.9	2430	2503.5	2577
81.0-85.9	CAB 129	310	81	83.45	85.9	25110	25869.5	26629
86.0-90.9	CAB 23	2863	86	88.45	90.9	246218	253232	260247
86.0-90.9	CAB 24	4339	86	88.45	90.9	373154	383785	394415
86.0-90.9	CAB 24	2624	86	88.45	90.9	225664	232093	238522
86.0-90.9	CAB 65	441	86	88.45	90.9	37926	39006.5	40086.9
86.0-90.9	CAB 66	540	86	88.45	90.9	46440	47763	49086
86.0-90.9	CAB 67	70	86	88.45	90.9	6020	6191.5	6363
86.0-90.9	CAB 68	321	86	88.45	90.9	27606	28392.5	29178.9
86.0-90.9	CAB 128	1267	86	88.45	90.9	108962	112066	115170
86.0-90.9	CAB 129	923	86	88.45	90.9	79378	81639.4	83900.7
91.0-95.9	CAB 23	10044	91	93.45	95.9	914004	938612	963220
91.0-95.9	CAB 24	12761	91	93.45	95.9	1161251	1192515	1223780

Project Acronym:
Version:
Contact:
Date:

Band	Series	No. files	Min accuracy value	Mid accuracy value	Max accuracy value	Min Av. Accuracy	Mid Av. Accuracy	Max. Av Accuracy
91.0-95.9	CAB 24	13328	91	93.45	95.9	1212848	1245502	1278155
91.0-95.9	CAB 65	2849	91	93.45	95.9	259259	266239	273219
91.0-95.9	CAB 66	4226	91	93.45	95.9	384566	394920	405273
91.0-95.9	CAB 67	517	91	93.45	95.9	47047	48313.7	49580.3
91.0-95.9	CAB 68	978	91	93.45	95.9	88998	91394.1	93790.2
91.0-95.9	CAB 128	1612	91	93.45	95.9	146692	150641	154591
91.0-95.9	CAB 129	5907	91	93.45	95.9	537537	552009	566481
96.0-100.9	CAB 23	20946	96	98.45	100.9	2010816	2062134	2113451
96.0-100.9	CAB 24	23585	96	98.45	100.9	2264160	2321943	2379727
96.0-100.9	CAB 24	59501	96	98.45	100.9	5712096	5857873	6003651
96.0-100.9	CAB 65	7445	96	98.45	100.9	714720	732960	751201
96.0-100.9	CAB 66	15671	96	98.45	100.9	1504416	1542810	1581204
96.0-100.9	CAB 67	2949	96	98.45	100.9	283104	290329	297554
96.0-100.9	CAB 68	2159	96	98.45	100.9	207264	212554	217843
96.0-100.9	CAB 128	18609	96	98.45	100.9	1786464	1832056	1877648
96.0-100.9	CAB 129	60538	96	98.45	100.9	5811648	5959966	6108284
Total		287526				93.7212	96.1712	98.6212

Table 2

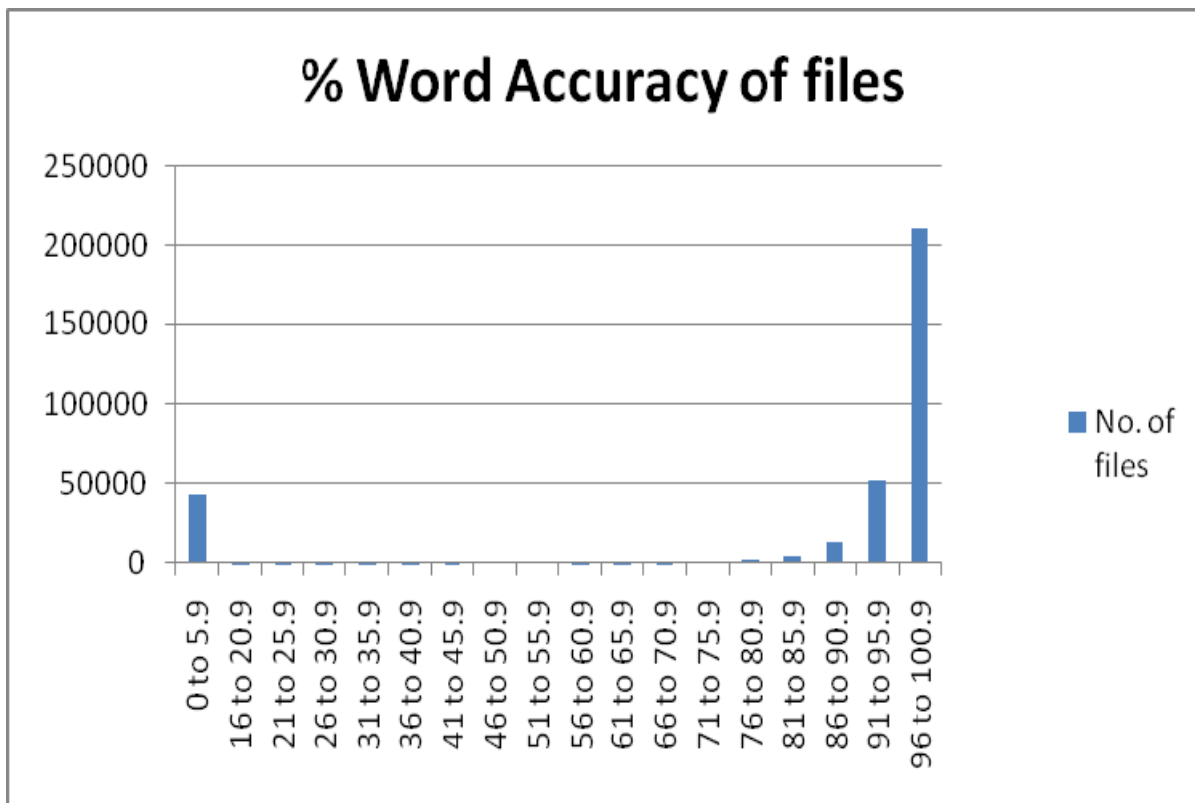
OCR Accuracy ratings and Bands 0-5.9%

Band	Series	No. files	Min accuracy value	Mid accuracy value	Max accuracy value	Min Av. Accuracy	Mid Av. Accuracy	Max. Av Accuracy
0.0-5.9	23	2106	0	2.95	5.9	0	6212.7	12425.4
0.0-5.9	24	274	0	2.95	5.9	0	808.3	1616.6
0.0-5.9	24	4	0	2.95	5.9	0	11.8	23.6
0.0-5.9	65	7428	0	2.95	5.9	0	21912.6	43825.2
0.0-5.9	66	15011	0	2.95	5.9	0	44282.5	88564.9

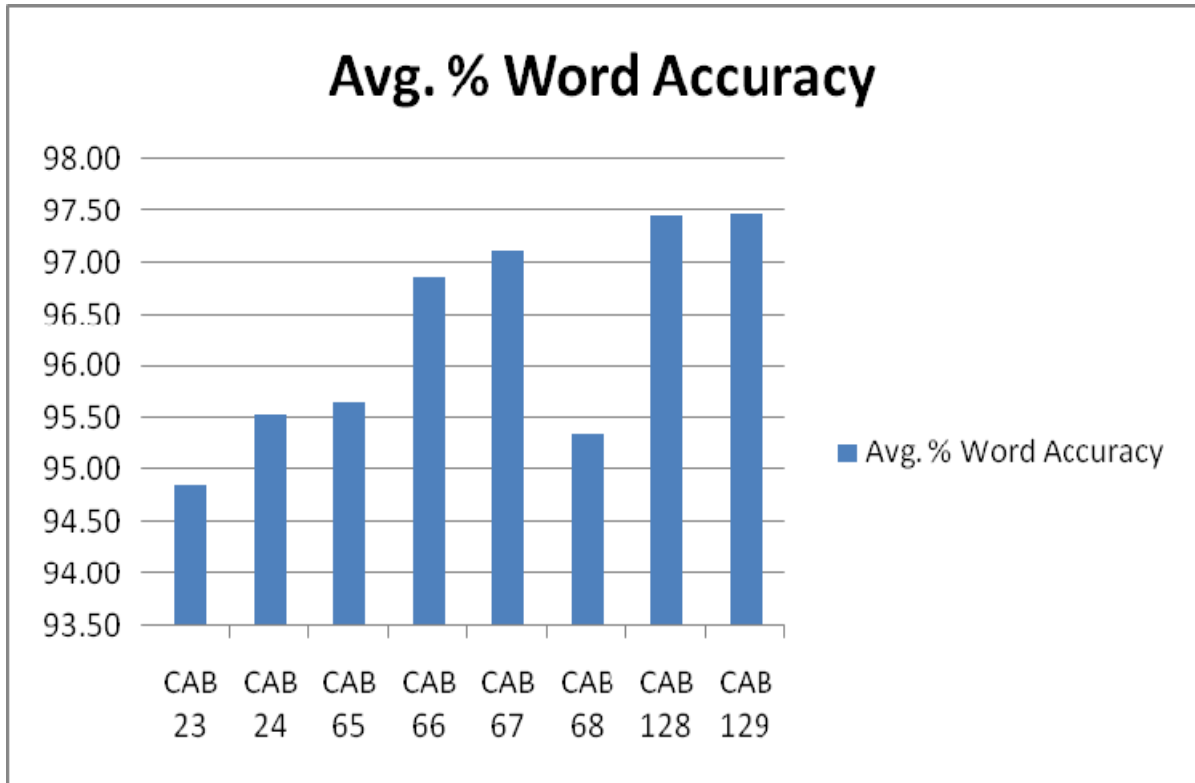
Project Acronym:
 Version:
 Contact:
 Date:

0.0-5.9	67	2667	0	2.95	5.9	0	7867.65	15735.3
0.0-5.9	68	2557	0	2.95	5.9	0	7543.15	15086.3
0.0-5.9	128	1974	0	2.95	5.9	0	5823.3	11646.6
0.0-5.9	129	11131	0	2.95	5.9	0	32836.5	65672.9
Total		43152				0	2.95	5.9

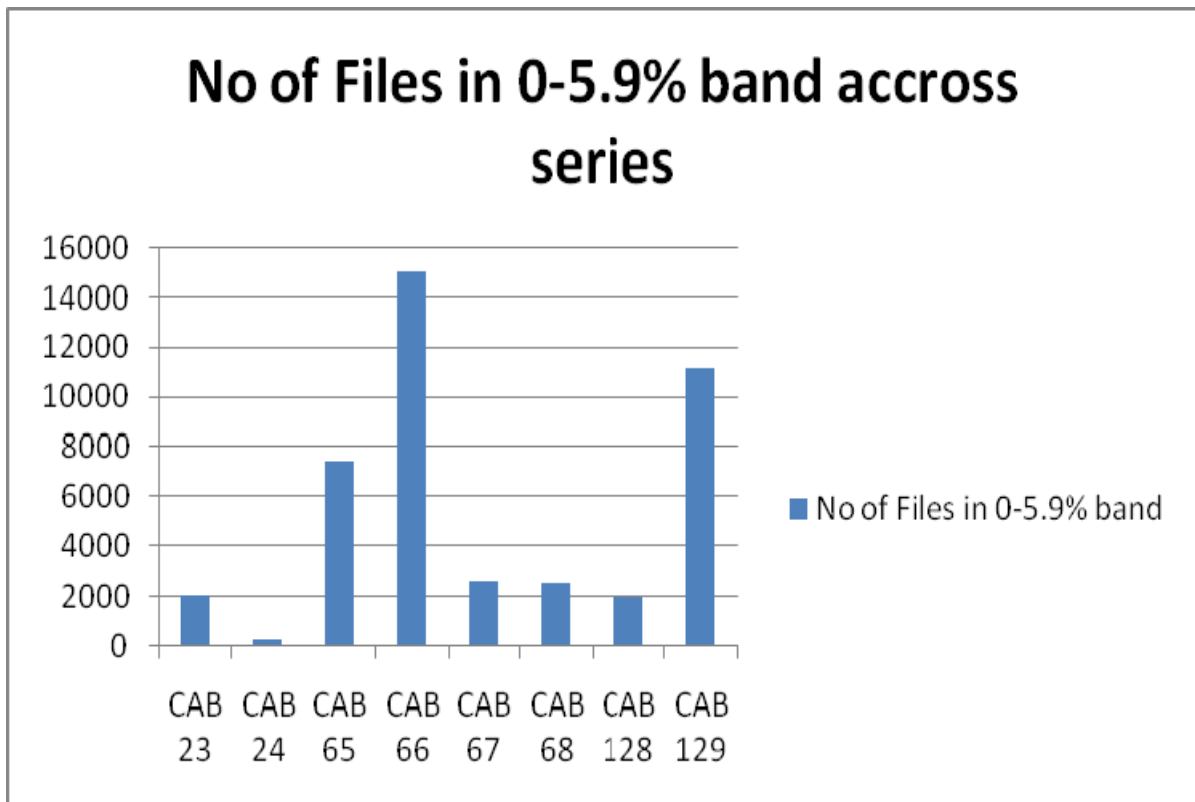
Graph 1



Graph 2



Graph 3



Project Acronym:
Version:
Contact:
Date:

Appendix C - Press Acknowledgement of Cabinet Papers Launch

Channel 4 Website Published 04/12/08

Science & Technology

Cabinet papers on the web

Last Modified: 04 Dec 2008
Source: PA News

More than half a million pages of government documents are available for the very first time on the internet.

Thanks to a project launched by The National Archives on its website, Cabinet papers from a 60-year period have been digitised and can now be accessed for free by the public.

Now anyone interested in the decisions made at the highest levels of Government can search for them online as part of the Cabinet Papers, 1915-1977 project.

It will provide an opportunity for historians, students and those interested in political history to delve into key Cabinet discussions from the last century.

Users will be able to search the entire collection, opening up a vast amount of information to people all over the world.

The website will include a series of teaching aids, such as interactive maps and images. Officials hope it will become a valuable resource for students, covering topics which form an integral part of exam syllabuses.

Edward Hampshire, contemporary records expert at The National Archives said: "This is an extremely valuable collection in terms of modern British history, and for the first time we have been able to open it up to everyone, not just those who are able to come to Kew.

"While we have created some fantastic resources for students, we hope the tools will help to open up the files to everyone with an interest in British history and politics.

"After all, this is recent history, and many of us will have lived through, and been affected by, the momentous issues and decisions recorded in these records."

The records will be free to access and download in schools and colleges teaching A-Levels in England and Wales for 10 years, universities and colleges in England and Wales for five years, and to everyone else for 18 months. The papers can be found at www.nationalarchives.gov.uk/cabinetpapers.

These news feeds are provided by an independent third party and Channel 4 is not responsible or liable to you for the same.

Cabinet papers online

The National Archives has launched a website providing free online access to more than 60 years of Cabinet papers from between 1915 and 1977. It is aimed at A-level and higher education students and contains more than half a million pages of Cabinet Minutes and Memoranda.

The Times Educational Supplement (Published 05/12/2008)

Project Acronym:
Version:
Contact:
Date:

60 years of Cabinet papers free online

BEN PADLEY

MORE than half-a-million pages of government documents will become available on the Internet for the first time today.

Thanks to a project launched by the National Archives on its website, cabinet papers from a 60-year period have been digitised and can now be accessed for free.

Now anyone interested in the decisions made at the highest levels of government can search for them online as part of the cabinet papers 1915-1977 project.

It will provide an opportunity for historians, students and those interested in political history to delve into key cabinet discussions.

The website will include a series of teaching aids, such as interactive maps and images. Officials hope it will become a valuable resource

for students, as it covers topics that form an integral part of school syllabuses.

Edward Hampshire, contemporary records expert at the National Archives, said: "This is an extremely valuable collection in terms of modern British history and for the first time we have been able to open it up to everyone, not just those who are able to come to (the archives in) Kew."

"While we have created some fantastic resources for students, we hope the tools will help to open up the files to everyone with an interest in British history and politics."

The records, at www.nationalarchives.gov.uk/cabinetpapers, will be free to access and download for schools and colleges teaching A-Levels in England and Wales for 10 years, for universities and colleges in England and Wales for five years, and for everyone else for 18 months.

The Herald (Published 04/12/2008)

Project Acronym:
Version:
Contact:
Date:

The Press Association

Cabinet papers on the web

19 hours ago

More than half a million pages of government documents are available for the very first time on the internet.

Thanks to a project launched by The National Archives on its website, Cabinet papers from a 60-year period have digitised and can now be accessed for free by the public.

Now anyone interested in the decisions made at the highest levels of Government can search for them online as part of the Cabinet Papers, 1915-1977 project.

It will provide an opportunity for historians, students and those interested in political history to delve into key Cabinet discussions from the last century.

Users will be able to search the entire collection, opening up a vast amount of information to people all over the world.

The website will include a series of teaching aids, such as interactive maps and images. Officials hope it will become a resource for students, covering topics which form an integral part of exam syllabuses.

Edward Hampshire, contemporary records expert at The National Archives said: "This is an extremely valuable collection of terms of modern British history, and for the first time we have been able to open it up to everyone, not just those who have to come to Kew."

"While we have created some fantastic resources for students, we hope the tools will help to open up the files to a wider interest in British history and politics."

"After all, this is recent history, and many of us will have lived through, and been affected by, the momentous issues and decisions recorded in these records."

The records will be free to access and download in schools and colleges teaching A-Levels in England and Wales, universities and colleges in England and Wales for five years, and to everyone else for 18 months. The papers can be accessed at www.nationalarchives.gov.uk/cabinetpapers.

Hosted by 

Copyright © 2008 The Press Association. All rights reserved.

The Press Association (Published 04/12/2008)

Project Acronym:
Version:
Contact:
Date:

NewsManager™

Title Number of cuttings per Type of Media for the "Cabinet Papers" feed

Description A Report outlining the Value of cuttings (AVE) per Type of Media for the "Cabinet Papers" feed

Media Type	AVE
National Press	£8,488
Regional Daily Press	£16,510
Regional Weekly Press	£1,313
Trade Press	£652

Total 26,963

