

Schedule 2

Services Description

For the purposes of this Schedule 2 (Services Description), and in addition to any terms defined in Schedule 1 (Definitions), the following terms shall have the following meanings:

"Click-through"	when a User sees and clicks on an advertisement, leading the User to the Website;
"Image Page View"	an instance of a Library Archive Image being downloaded by a User;
"Implementation Progress Report"	the report to be delivered by the Contractor to the Library in accordance with, and containing the information detailed in, section F1.2 of this Schedule 2 (Services Description);
"Index Page View"	an instance of the index page being downloaded by a User;
"Page View"	an instance of a page on the Website being loaded by a browser. This includes HTML pages (.html, .htm, .shtml), script-generated pages (.cgi, .asp, .cfm, etc.) and plain-text pages. It also includes sound files (.wav, .aiff, etc.), video files (.mov, etc.) and other non-document files. Image files (.jpeg, .gif, .png), javascript (.js) and style sheets (.css) are not considered to be pages of the Website for this purpose;
"Payment Transaction"	a User connecting to the payment gateway and completing a purchase on the Website;
"Source of Traffic"	the origin of a referral to the Website. An example of a source is a visitor referred by an unpaid search engine listing, by a paid search engine listing, by typing the URL directly into a browser, or by linking from another website to the Website; and
"Unique Users"	the number of unduplicated (counted only once) visitors to the Website over a thirty (30) minute time period (determined by using cookies activated on or from the Website).

PART A – DIGITISATION SERVICES

1.1 Scanning Production

- 1.1.1 The Contractor shall scan pages from the Newspaper Archive Materials, in line with the Implementation Plan and in accordance with the terms of this Agreement on account of (and in order to fulfil) the prevailing Volume Commitment.
- 1.1.2 The Contractor shall handle all Newspaper Archive Materials in accordance with the Material Handling Standards set out in Schedule 4 (Standards).
- 1.1.3 The Contractor shall use [REDACTED] scanners to carry out the scanning element of the Services or such other number and type of scanning equipment as the Contractor determines necessary to fulfil its obligations under all Service Plans to the relevant Standards and in order to fulfil the Volume Commitment.
- 1.1.4 The Contractor shall undertake the Digitisation Services within the Designated Space(s) which in the first instance will be in the Colindale repository, a plan of which is attached in Annex A to this Schedule 2 (Services Description).

1.2 Scanning QC

- 1.2.1 The Contractor shall quality control the Library Archive Images to ensure that they meet the relevant Standards detailed in Schedule 4 (Standards).
- 1.2.2 The Contractor shall quality control a minimum of [REDACTED] of the total number of Library Archive Images during the provision of the Digitisation Services. This quality control shall be undertaken at a designated quality control workstation.
- 1.2.3 Any Library Archive Images not meeting the relevant Standards must be rescanned.

1.3 Library Archive Images

- 1.3.1 The Contractor shall produce the Library Archive Images which will conform to the relevant Standards in Schedule 4 (Standards).
- 1.3.2 The Contractor shall deliver the Library Archive Images to the Library for the purposes of Ingest in accordance with the requirements set out in Annex C.
- 1.3.3 Subject to Clause 16.1.2 and Clause 16.1.5 of the Agreement, the Contractor shall deliver the Acquired Images to the Library for the purposes of Ingest.
- 1.3.4 Subject to Clause 8.1.4, the Contractor shall, as a minimum, retain copies of Library Archive Images until they are delivered to the Library in accordance with Clause 8.1.4 (and receipt is acknowledged).

1.4 Library Provisions

- 1.4.1 The Library shall supply to the Contractor support resources as follows:

- (a) Security during normal working hours at the Library Premises containing the Designated Space. Security outside working hours may depend on shift patterns worked and the extent of supervision/supply agreed to be provided by Library;
- (b) Collection Care Services. The level of service provided by Collection Care Services to the Contractor will vary week-by-week depending on the level of such service requested by the Contractor. The Contractor will notify the Library of any such service from Collection Care Services as is required as soon as possible as part of the Access Procedures and Transfer Control Requirements to be developed by the parties pursuant to Part C of this Schedule 2 (Services Description); and
- (c) Contractor Personnel training on Materials Handling Standards as set out in Schedule 4 (Standards).

2.1 OCR Production

2.1.1 The Contractor shall produce the OCR Text to the relevant minimum Standards set out in Schedule 4 (Standards).

2.1.2 The production of OCR Text as part of the Digitisation Services shall be carried out in accordance with the Implementation Plan.

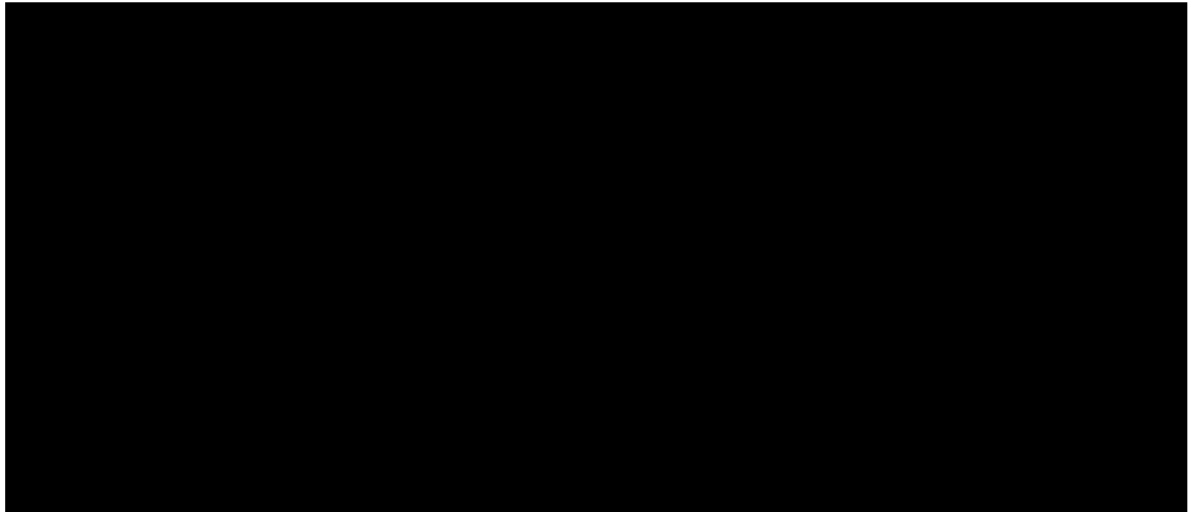
2.1.3

2.1.4

3. OCR Text and Metadata Quality Control

3.1.1 The Contractor shall supply the Metadata in accordance with the standards referred to in the Supplier Engagement Pack attached in Annex C to this Schedule 2 (Services Description).

3.1.2



PART B – ONLINE ACCESS SERVICE

OAS

- 1.1 The Contractor shall publish the Library Archive Images and OCR Text on the OAS following the OAS Go Live Date in accordance with the terms and subject to the conditions of this Agreement, including in particular Clause 8.3.1.
- 1.2 The OAS will be free to access from Library Premises as further detailed in Clause 17.1.

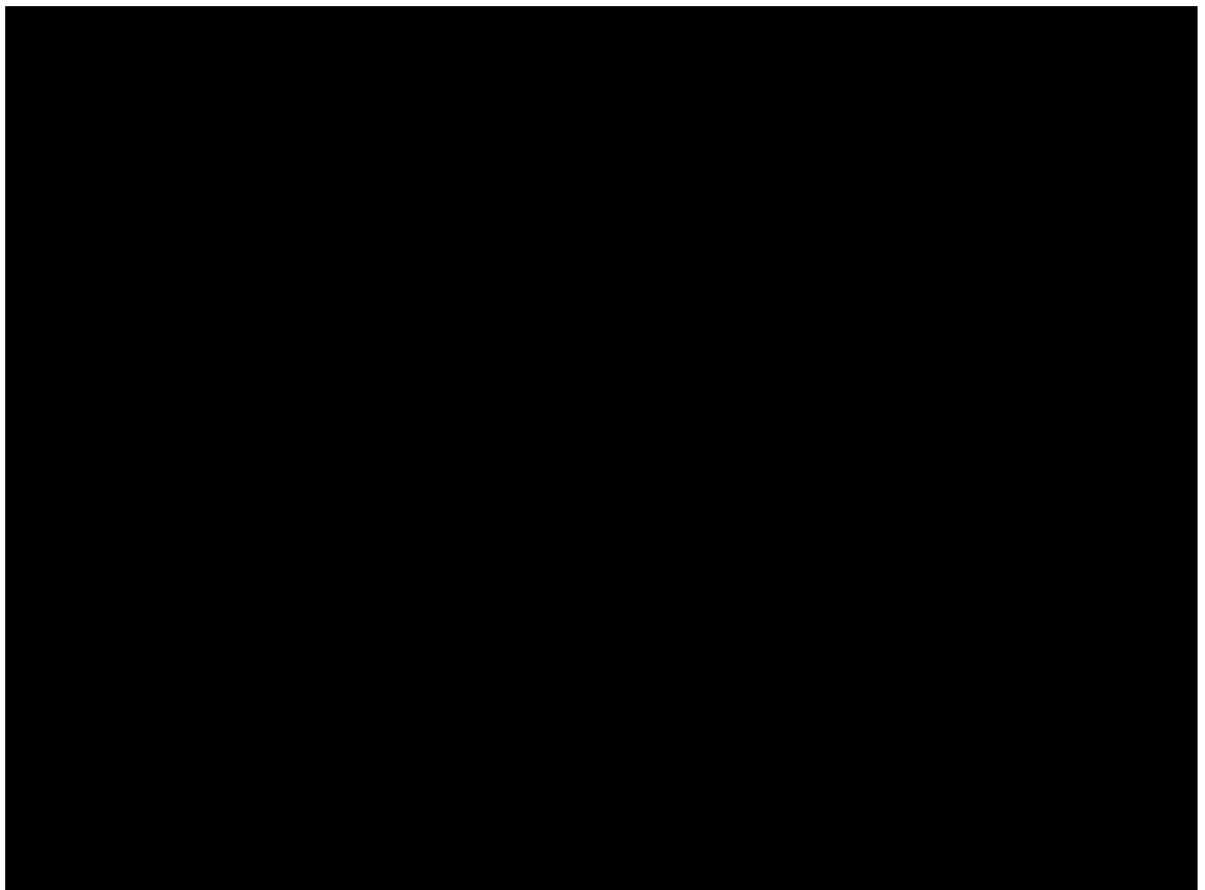
1.3

1.4

1.5

1.6

1.7



1.8

1.9

1.10

1.11

1.12

PART C – ACCESS PROCEDURES & TRANSFER CONTROL REQUIREMENTS

The Library and the Contractor shall develop, in accordance with the Implementation Plan, a suitable "Check In/Out Procedure" to control the flow of Newspaper Archive Materials into and out of control of the Contractor. This procedure shall identify how Newspaper Archive Materials are tracked when in the Contractor's control and how communications between the parties regarding Newspaper Archive Materials that are deemed to be of poor quality and possibly requiring Collection Care Services intervention should be initiated. The parties shall act reasonably in discussions relating to the development of such procedure.

PART D – DEVELOPMENT OF THE ONLINE ACCESS SERVICE

- 1.1. The OAS will be developed using the Online Delivery System.
- 1.2. The OAS will be ready for public access from the OAS Go Live Date. Actual dates to be agreed between the parties, and subject in any event to the Library issuing the Authority to Proceed Acceptance Certificate. Both parties shall act reasonably in seeking to agree such a date.
- 1.3. The Library OAS will be ready for access following the OAS Go Live Date. Actual dates to be agreed between the parties. Both parties shall act reasonably in agreeing such a date.

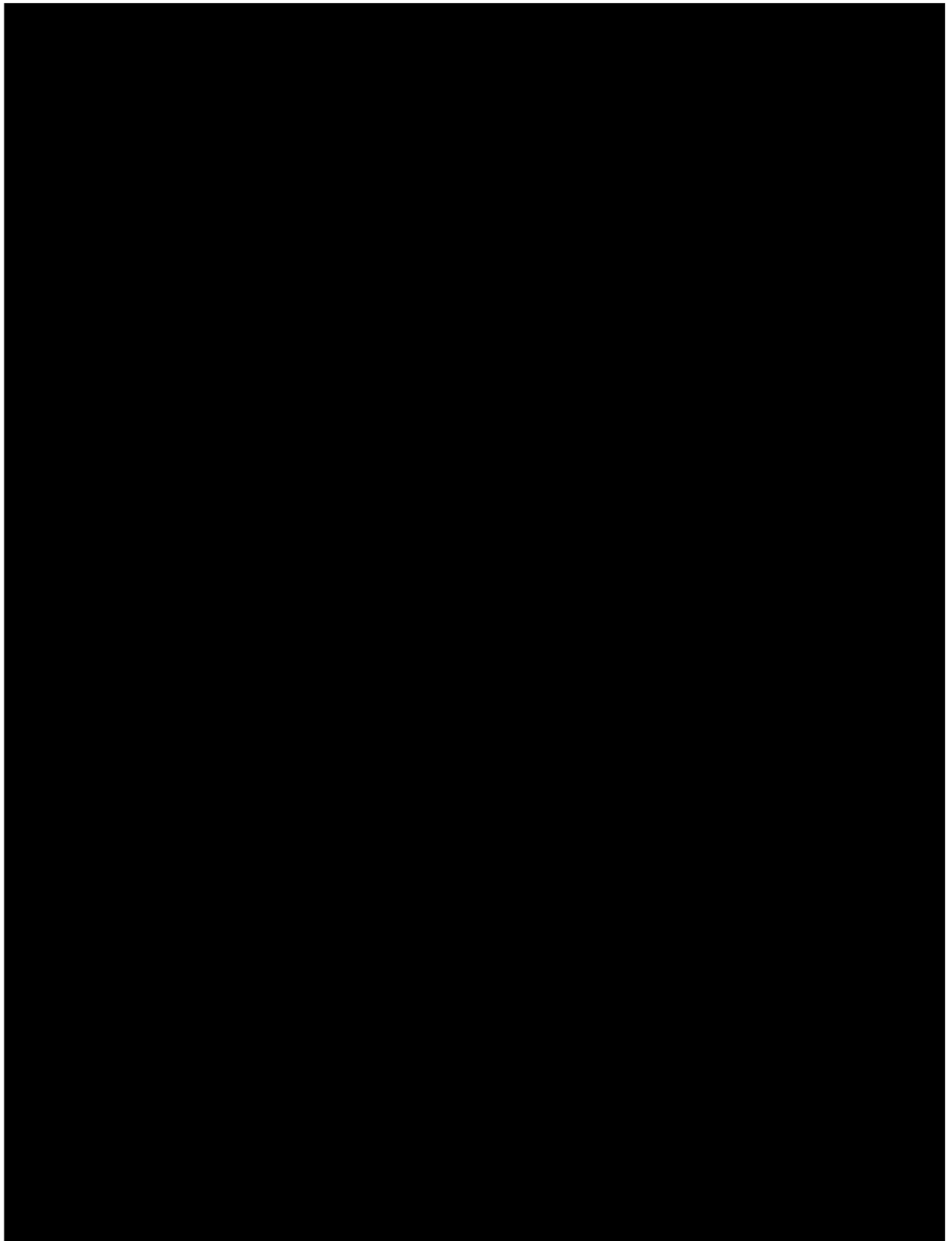
2. CHANGES TO THE ONLINE ACCESS SERVICE

2.1 Certain attributes of the Online Access Service are required by the Library. The materiality of these attributes varies according to the value the Library places on that attribute.

2.2

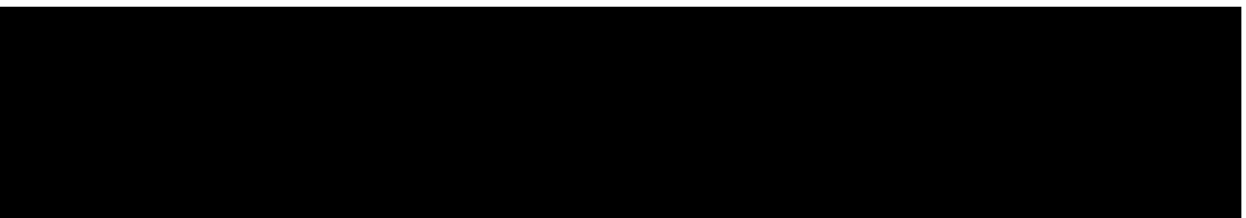
2.3

2.4



PART E – ONLINE DELIVERY SYSTEM

1.1



1.2

1.3

1.4

1.5

1.6

1.7 Online Delivery System Outline Infrastructure Diagram (for illustrative purposes) attached in Annex B.

PART F – GENERAL SERVICES

F1 REPORTING SERVICES

Implementation Reporting

F1.1 The Contractor shall provide the Library with an electronic copy of its Implementation Progress Report in accordance with section F1.2.

F1.2 Once a month, within five (5) Working Days of the end-of-month and covering the whole of the preceding month, the Contractor shall supply the Library with a monthly written Implementation Progress Report containing:

- (a) a brief descriptive account of progress made against the Implementation Plan (and Project Plan, to the extent applicable), including any variations to the Implementation Plan or the Project Plan and recovery plans associated with those variations;
- (b) descriptive updates on agreed actions and resolution of issues, including the date agreed for completion and the date completion is anticipated or actually occurred;
- (c) any issues for resolution by the Library;
- (d) required actions by, or dependencies on, the Library for each of the forthcoming three (3) months;
- (e) a summary list of all Deliverables due for completion for the month in which the report is generated, clearly differentiating those actually completed on time and those still to be delivered;
- (f) a summary list of Deliverables due for completion for the forthcoming month, containing and highlighting all Deliverables carried forward from earlier periods; and
- (g) a table of key dates containing a list of major Deliverables, dependencies, Key Milestones and key decision points from the Project Plan or Implementation Plan, together with associated Implementation Plan dates, estimated completion date and actual completion date as appropriate.

F1.3 Contemporaneously with the Implementation Progress Report to be submitted by the Contractor to the Library under section F1.2 above, the Contractor shall supply the Library with a monthly updated GANTT chart consistent with the descriptive and tabular Implementation Progress Report and showing:

- (a) the baseline Implementation Plan;
- (b) current re-plan if appropriate; and
- (c) current progress against the Implementation Plan.

Expenditure Reports

- F1.4 The Contractor shall provide the Library with regular updates of the anticipated expenditure profile identifying any changes to the Project Plan and/or Implementation Plan which have caused a material variation to the total anticipated expenditure profile.
- F1.5 The Contractor shall provide reports required under paragraph F1.4 against Commercial KPIs in accordance with Schedule 23 (Commercial KPIs).

Digitisation and OCR progress reporting

- F1.6 The Contractor shall provide the Library with a monthly report detailing and including the following in particular:
- F1.6.1 number of Pages digitised each week of the preceding month, reported against targets, and in total (total net of all rejections, with rescanning or scanning correction required) including title of publication;
 - F1.6.2 number of Pages with related OCR Text each week of the preceding month, reported against targets, and in total (total net of all rejections with re-OCR Text correction required); and
 - F1.6.3 monthly movements through scanning area in terms of number of volumes, binders etc in accordance with specified Access Procedures and Transfer Control Requirements,

that, for the avoidance of doubt, taking further account of and include the following:

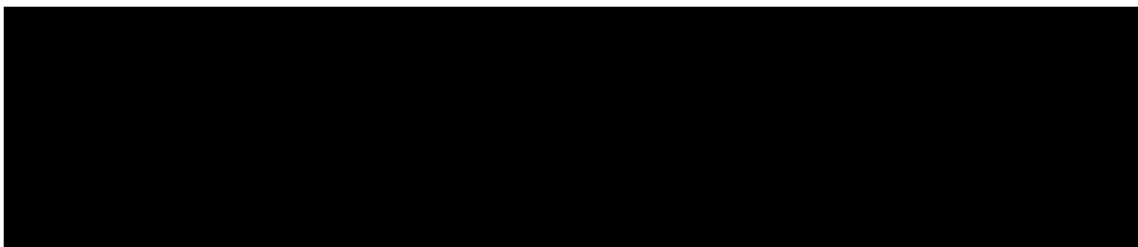
- (a) for each of the above reports (that is, for Digitisation Services and OCR Text progress reporting), planned figures for each week of the preceding month and in total at the then current stage;
- (b) for each of the above reports (that is, for Digitisation Services and OCR Text progress reporting), slippage calculated as the number of days difference between the end-of-week date and the date when that number of pieces should have been completed; and
- (c) if the difference indicates slippage of more than five (5) Working Days in total, a recovery plan detailing how the affected performance of the Digitisation Services will be absorbed into the Implementation Plan.

The parties shall agree the format of such monthly report (such agreement not to be unreasonably withheld).

- F1.7 The Contractor shall forward to the Library the weekly reports received from all Sub-contractors for information purposes where the Contractor has identified there is a problem meeting any Key Milestone date.

Online Access Service reporting

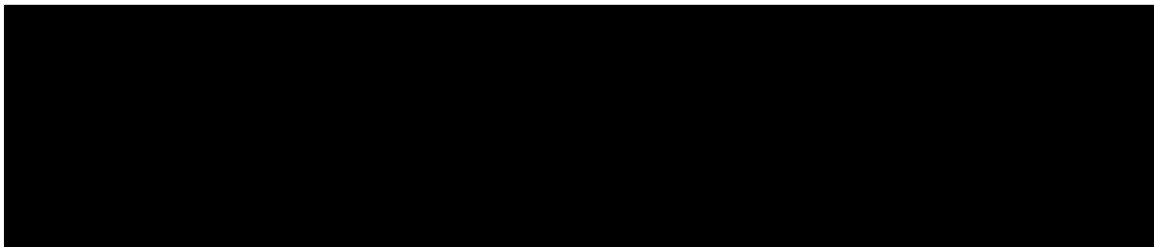
F1.8



- F1.9 The Contractor shall also record total Click-throughs from the Library's Website to the Online Access Service, the total number of Payment Transactions on the Online Access Service, and the top fifty (50) Sources of Traffic to the Online Access Service, on a daily, and monthly and rolling three (3) monthly cumulative basis. A written report summarising these statistics is to be supplied to the Library within thirty (30) Working Days of the end of each calendar month of operation.
- F1.10 The Contractor shall also supply to the Library upon request any other such usage reports and customer service/User statistics that it may have at its disposal and as are reasonably requested by the Library from time to time.
- F1.11 The detailed form of all such written statements and reports will be agreed between the parties prior to the OAS GO Live Date, or as reasonably required by the Library.

Library Online Access Service reporting

F1.12



- F1.13 The Contractor shall also record total Click-throughs from the Library's Website to the Library Online Access Service, on a daily, and monthly and rolling three-(3) monthly cumulative basis. A written report summarising these statistics is to be supplied to the Library within thirty (30) days of the end of each calendar month of operation.
- F1.14 The Contractor shall also supply to the Library upon request any other such usage reports and customer service/User statistics that it may have at its disposal and as are reasonably requested by the Library from time to time.
- F1.15 The detailed form of all such written statements and reports will be agreed between the parties prior to the OAS GO Live Date, or as reasonably required by the Library.

Parliamentary Questions

- F1.16 Where the Library informs the Contractor that a demand for any report detailed or referred to in this section F1 has been generated as a result of a Parliamentary Question, the Contractor shall use all reasonable endeavours to provide the relevant report in full, and in accordance with the requirements set out herein, within one (1) Working Day of the demand.
- F1.17 The Contractor shall use all reasonable endeavours to collaborate with the Library in providing more detailed responses to any Parliamentary Questions, including reasonable access to the Contractor's own relevant reports produced for its own purposes in managing the Services, should the Library request.

F2 CONTRACTOR MARKETING SERVICES

- F2.1 The Contractor shall appoint and brief a design agency and PR/marketing agency to plan the marketing of the Online Access Service in accordance with the Implementation Plan.

F2.2



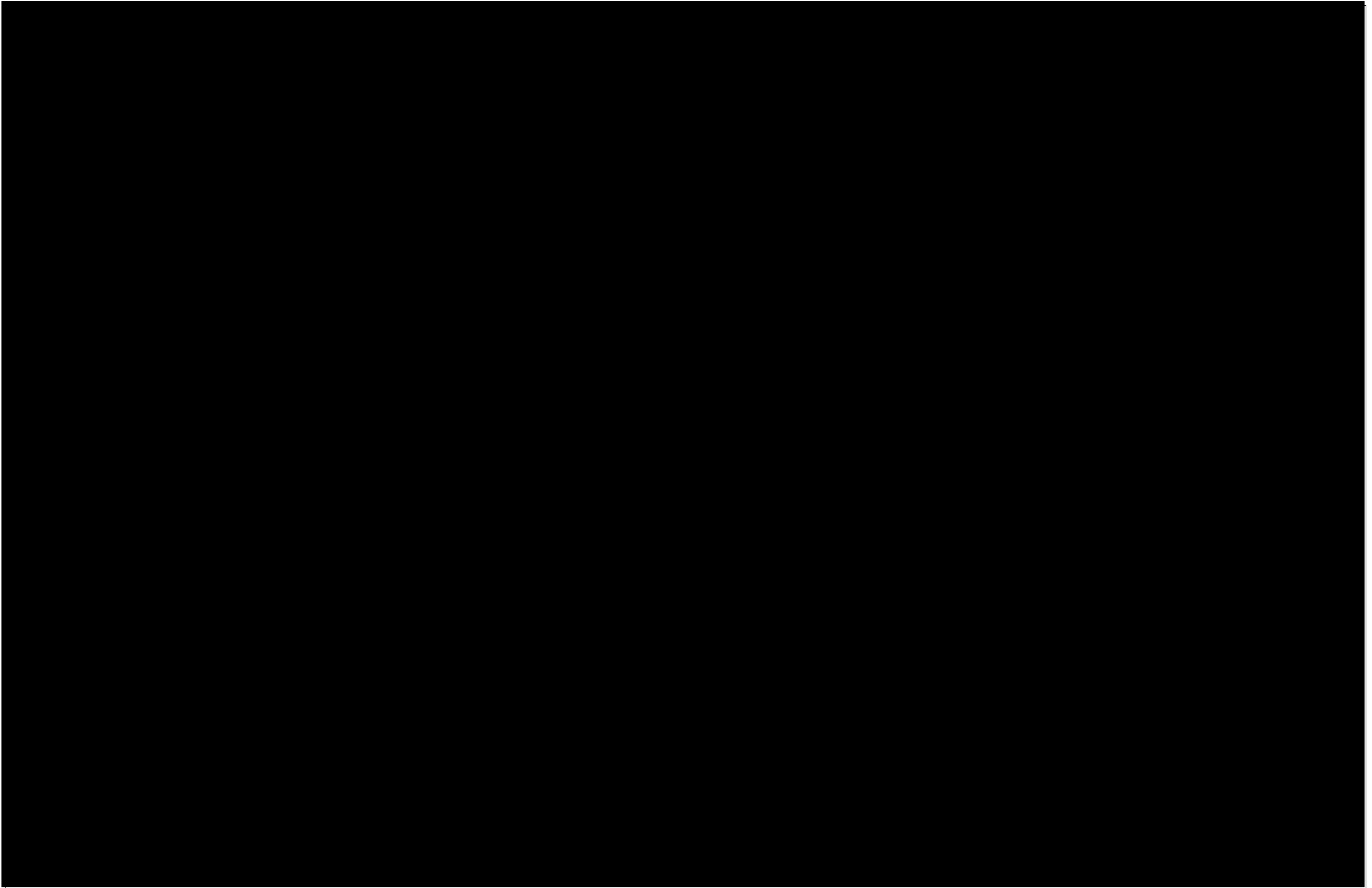
F3. EDUCATIONAL ACADEMIC AND OUTREACH SERVICES



Annex A

Plan of Colindale Repository

[see overleaf]



Annex B

Online Delivery System Outline Infrastructure Diagram

[see overleaf]

Annex C

Supplier Engagement Pack

[see overleaf]



e-Strategy and Information Systems

Supplier engagement

Document history

Version	Date	Author	Status / Change
0.1	23/04/2009	Paul Clements	Draft
0.2	01/05/2009	Paul Clements	Incorporated changes from review by Steve Green and Martyn Lunn
0.3	11/05/2009	Paul Clements	After wider review
1.0	25/06/2009	Paul Clements	For approval by Newspaper Strategy Programme

Document Approval

Version	Date	Name	Role / Function
1.0	25/06/2009	Alasdair Bruce	

Key stakeholders and reviewers

Version reviewed	Name	Role / Function
0.2	Nicki Clegg	
0.2	Stephanie Cresswell	
0.2	Markus Enders	
0.2	David Golding	
0.1	Steven Green	
0.1+	Andy Hawkin	
0.1+	Martyn Lunn	
0.1+	Qasim Mahmood	
0.2	Sean Martin	
0.2	Rory McLeod	
0.2	Mike Ohren	
0.2	Robert Sheridan	
0.1+	Peter Stubbs	
0.2	Paul Wheatley	

Key reference documents

Version	Date	Author	Name of document
0.1	19/02/2009	Steve Green	Supplier Engagement
2.1	Unknown	Martyn Lunn	Newspapers Programme Digitisation Partnership – Competitive Dialogue - Standards
0.3	01/09/2008	Marcus Enders	METS Profile
0.2	18/01/2008	Victoria Nelson	Supplier Engagement
1.0	23/05/2008	National Library of Australia	Australian Newspapers Digitisation Program – Specification for Content Analysis and OCR
1.0	20/08/2008	Rose Holley, Australian Newspaper Digitisation Programme	Increasing the Accuracy of OCR
NA	6/04/09	IMPACT Conference	Presentations from conference

Acronyms

Acronym	Meaning
CA&D	Collection Acquisition and Description
DLS	Digital Library Store
METS	Metadata and Encoding Transmission Standard
OCR	Optical Character Recognition
PPI	Pixels per inch

Table of Contents

1	Introduction	6
1.1	Scope	6
1.2	Audience	6
1.3	Document review	6
2	The Engagement Process	7
2.1	Purpose	7
2.2	Key activities in the engagement process	8
2.3	Key milestones in the engagement process	10
3	The Content Capture Process	14
3.1	Introduction	14
3.2	Responsibilities	14
3.3	Content production	15
3.4	Conclusion	21
4	Conclusions	22
Appendix A.	Engagement Process Information Capture	23
Appendix B.	Report from Digitise on Demand Proof of Concept	30
Appendix C.	METS Schema with examples	31

1 Introduction

The Digital Library Programme is the British Library's strategic response to the long-term transition from print to digital information sources. The programme has been created to design, plan, develop and implement a secure, reliable and scalable management system for the Library's digital content assets. The resulting Digital Library System (DLS) will provide a single system to manage ingest, storage, preservation, discovery of and controlled access to the BL's digital content assets.

1.1 Scope

This document describes (a) the engagement process between the content supplier and the Library and (b) the preferred and minimum characteristics of the content being supplied to the Library. It makes the assumption that preliminary negotiation has taken place and that there is 'agreement in principle' that the content supplier will supply digital newspaper content to the Library.

This document then provides the detail on how this 'agreement in principle' can be progressed successfully to the point that the Library is ingesting material into the DLS from the content supplier.

Note the content supplier may be an internal or external supplier; the same processes should be carried out irrespective.

1.2 Audience

This document is principally for providers of newspaper digital content which is to be stored within the DLS. It may be used at the preliminary stage for expectation setting but is principally intended for those parties that will take the process through to go live.

1.3 Document review

This document will be reviewed on an annual basis or as a result on a change of the Library policy on how content is captured, processed or stored.

2 The Engagement Process

This section describes the process and its constituent activities which have to be carried out between a potential content supplier and the Library. Its aim is to provide the content supplier with a view on:

- Why there is a need for this process
- What are the key milestones during the engagement process?
- What are the likely timescales of the initial engagement?
- What are the main deliverables of the process from both the content supplier and Library side?

2.1 Purpose

The Library over the last 5-10 years has invested heavily in terms of both hardware and software to enable digital content to be stored in perpetuity and to allow access to this content by Library customers where the appropriate digital rights have been agreed with the content supplier. Whilst the solution built is flexible enough to deal with all types of content from sound archives to e-journals to scanned images, there is a degree of configuration that has to occur before the Library can ingest the content with confidence that:

- Content has been received correctly.
- Content is able to stored in accordance with International and Library standards and
- Content has sufficient meta data captured that it can be retrieved not only for access but for future preservation needs

Hence, the Library has developed a process which enables the Library and the content supplier to work together such that the above bullet points can be met with confidence. The following sections detail the key activities and milestones of a typical engagement process as shown in Figure 1.

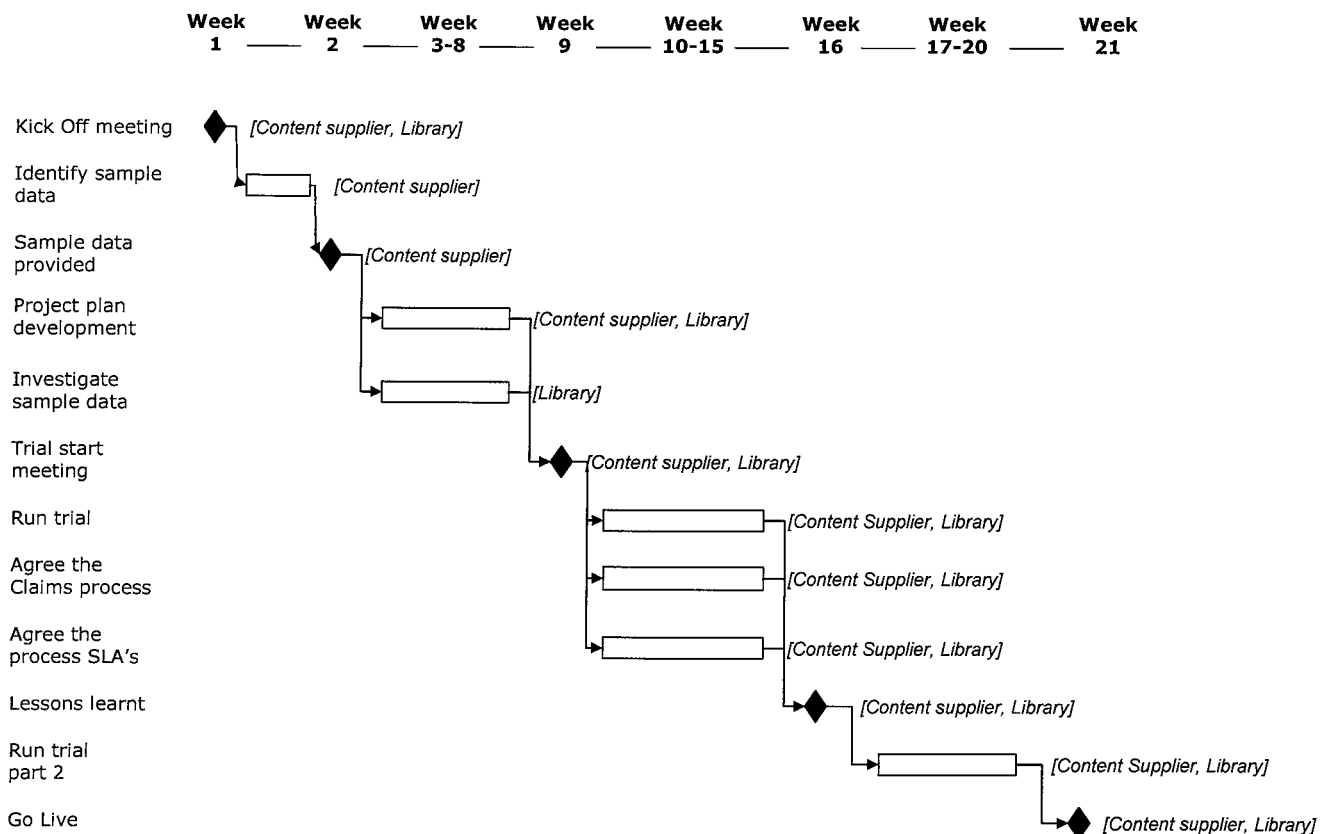


Figure 1 Example project plan for a typical engagement process

Note the time line is indicative of an engagement process which is building on previous ingest processes, the actual project plan will be developed during the project initiation phase.

2.2 Key activities in the engagement process

The following sections detail the key activities which need to be met in order to ensure that content can be ingested successfully into the Library's digital infrastructure.

2.2.1 Identify sample data

In order for the Library to identify and scope the amount of work that will be required to successfully ingest the content supplied into the DLS, the content supplier has to provide the Library with a small set of data which can be used as an initial proving exercise and that will feed into the specification of the trial. Hence this activity is all about the content supplier having to identify a representative set of content which will enable the Library to assess the best ingest process to use or what new configurations will be required.

2.2.2 Project Plan development

Once the Library has investigated the ingest of the sample data in to the DLS, then a more accurate project plan with associated activities can be developed building upon the framework plan shown in Figure 1.

2.2.3 Investigate sample data

This activity is all about determining the most appropriate process for ingesting the content into the DLS and proving on a small scale that the content can be successfully ingested into the DLS and where appropriate retrieved for viewing in the reading rooms.

2.2.4 Run trial

This is one of the main activities of the engagement process and builds upon the work undertaken during the sample data investigation. It includes ensuring that all aspects of the end to end process are functioning as envisaged, that any SLA's proposed can be met by both parties. It should run for sufficient time that both parties are confident that there will be no major surprises once the system moves into a live environment.

2.2.5 Agree the claims process

One of the key operational processes that has to be defined before any ingest content stream can go live is the claims process, The factors to be considered are (a) how does the Library claim for content that it believes has not been received by the Library (b) how long does it wait before storing the incomplete content into the DLS because the missing content will never be sent and (c) what frequency of claiming should be performed between the Library and the content supplier including number of times, duration between claims etc.

2.2.6 Agree the process SLA's

Another key operational process that has to be defined and associated with the above claims process is defining the agreed SLA's between the content supplier and the Library. These SLA's typically include (a) how fast the Library should claim for content it has deemed is missing (b) how quickly the content supplier or Library should respond to queries from the other party (c) how long the should be held at the Library before being ingested in to the DLS etc

2.2.7 Run trial part 2

This is an optional activity which occurs after the lessons learnt discussion and will normally be enacted due to some fine tuning being required either at the content supplier or the Library's end.

2.3 Key milestones in the engagement process

The following sections detail the key milestones which need to be met in order to ensure that content can be ingested successfully into the Library's digital infrastructure.

2.3.1 Kick Off Meeting

A half day workshop is held between the content supplier and the Library to initiate the engagement process and to ensure everyone understands their roles and responsibilities within the process and the various key milestones that need to be met.

Pre-requisites

- Content supplier has identified the key people that will be involved in the engagement process, typically the sponsor, a member of their content production team, a member of the content suppliers technical team and a project manager.
- Content supplier has identified the material that will be supplied on successful conclusion of the trial.
- Library has identified the technical project manager that will be the main contact between the content supplier and the Library during the engagement process.

Desired Outcomes

- There is agreement about the scope of the trial.
- There is agreement about the durations between the milestones described in these sections and also a high level project plan.
- There is draft agreement around the content capture and associated characteristic's, see section 3.3 for more detail.
- There is an agreed date when content can be supplied for the Library to trial the ingest process.

2.3.2 Sample data provided

Before running an end to end trial of the complete process, the Library require some representative sample data to look at the potential use of existing ingest processes or how an existing process could be modified in order to accept the material supplied by the content supplier.

Pre-requisites

- Library has identified the team that will look at the sample data and work on the most appropriate mechanism for ingesting this material.
- Obtain sample data

Desired Outcomes

- Library has a much clearer idea of how the end to end process from content being supplied through to storage in the DLS will work and where appropriate, how access within the reading rooms will also be accomplished.
- Library will have determined how long the trial part of the engagement process will be required to be executed for and what is required from the content supplier in order to meet these timescales.

2.3.3 Trial kick off

Typically, 6-8 weeks after the kick off meeting has taken place both parties should be in a position to commence an initial trial of the end to end process.

Pre-requisites

- Content supplier has identified the content which will form the basis of the trial.
- Content supplier has the process in place to deliver the files to the Library.
- Library has investigated the sample files and is confident that they can be handled by the current process engine.
- Library has made any necessary enhancements to the ingest process for the files being sent.
- Library has created the bibliographic information within the Integrated Catalogue and DLS. Note this is currently a manual process and will involve the serials cataloguing team to be part of this delivery.
- Content supplier and Library have checked that the delivery mechanism is working.

Desired outcomes

- Successful demonstration of the end to end process from content delivery through to ingest of content into the DLS and subsequent access.
- Green light to commence delivery of the content from the supplier to the Library.

2.3.4 Lessons learnt

Typically, 4 weeks after the trial starts a meeting is convened to consider the lessons learnt from the experience.

Pre-requisites

- Content supplier has delivered the titles and issues which were agreed.
- Assurance has been carried out by the Digital Preservation Team to ensure material is captured as desired.
- Library has ingested the material into the DLS correctly.
- Any outstanding problems or issues have been resolved.

Desired Outcomes

- Content supplier is satisfied that the process is working as planned and is willing to move into normal live running.
- Any reporting requirements have been identified by the content supplier and Library has agreed to their production and schedule.
- A claims process has been agreed between the content supplier and the Library, see section 2.2.5.
- All SLA's have been agreed by both parties, see section 2.2.6.

2.3.5 Go Live

Some period after the conclusion of the trial and the lessons learnt meeting it is envisaged that the ingest process can move into a live environment

Pre-requisites

- All contractual matters have been resolved and signed-off
- Any digital rights have been agreed and signed-off
- Any required additions to the catalogue have been identified and actioned by CA&D.
- Handover complete within the Library.
- Processes agreed to handle changes after the Library has started to ingest live content, including:
 - Content provider to notify the Library of any changes to content and schedule
 - Content provider to approve any changes to how ingest is performed or usage of the ingested items
 - Changes to any licensing, IP etc.

Desired Outcomes

- An agreed schedule for when the content will be delivered to the Library.
- Content being streamed to the Library for ingest into the DLS.

3 The Content Capture Process

3.1 Introduction

In order to ensure that the Library is able to successfully ingest and potentially retrieve the newspaper content supplied to them by the content supplier, a number of content specific factors have to be agreed upon during the engagement process.

This section provides information that will guide this part of the engagement process and help to ensure that each party is aware of what is required and as importantly why these items are required.

3.2 Responsibilities

Once the engagement process has been completed and the content supplier is providing content to the Library, each of the parties will have ongoing responsibility for ensuring the successful ongoing operation of the service.

The content supplier will be responsible for:

- Ensuring the content is in line with their agreed specifications as defined during the engagement process
- Ensuring the content is delivered to the Library as agreed during the engagement process.
- Answering any questions by the Library about the content they have sent. The process for achieving this will have been defined during the engagement process.
- Notifying the Library of any new, changes or discontinuance to titles, publication pattern changes, supplement changes etc that would potentially impact the Library's correct storage of the material.

The Library will be responsible for:

- Ensuring that all content supplied is safe and secure during each stage of the ingest and retrieval process.
- Ensuring that the content delivered by the content supplier is not modified in any manner unless by prior agreement during the engagement process. For example, where the content supplier is unable to provide compressed files to the Library, the Library will as

part of ingest perform 'lossless' compression on the images and only those images will be stored.

- Ensuring that any reports agreed during the engagement process are delivered on schedule.
- Ensuring that any specific business rules identified during the engagement process are enacted within the ingest of the content.

3.3 Content production

The following sections provide information to content suppliers about information such as:

- Form and format of images supplied (3.3.1)
- Directory structure and file naming (3.3.2)
- File delivery (3.3.3)
- Optical Character Recognition (3.3.4)
- Descriptive and structural metadata (3.3.5)

Each of the sections highlight where appropriate the differences in information required dependent on the type of content being supplied i.e.:

- Scanned images from print copy or
- Born digital images

For each of the sections the document identifies two classifications:

- Preferred: This is the ideal for delivery to the Library and is anticipated to be the one that is used where the Library are entering a formal agreement with either an internal or external content supplier and
- Minimum: This is the least we would expect from a content supplier and is usually where the content supplier is voluntarily submitting content to the Library. Note in some of these cases even the minimum might not be able to be met.

3.3.1 Image information required

The following tables indicate the information required by content suppliers with respect to the page images to be supplied to the Library.

Table 1 details the information required from content suppliers who are scanning images for the Library and Table 2 details the information required from publishers who are supplying 'born digital' content.

	Preferred Option¹	Minimum
File Format	JPEG 2000 with lossless compression	A file type from one of the Library approved list.
Header information	All relevant fields populated.	Attributes: Date created
Resolution	400 ppi	300 ppi
Bit depth	24 bit colour	Bi-tonal
Scan area	De-skewed if greater than 1° rotation and cropped to page edge but no change in orientation	As captured
Master image	Full page including border	Cropped page to border

Table 1. Scanned Images

Preferred Option	Minimum
The files that were the ones used by the print process to produce the print copy	The files that initially were used to produce the layout of the printed copy

Table 2. Born Digital Images

3.3.2 Directory structure and file naming

In order to assist in the ingest process at the Library, the preferred structure and naming of files for delivery to the library is as shown in Figure 2.

¹ Detail to be confirmed following study that is taking place

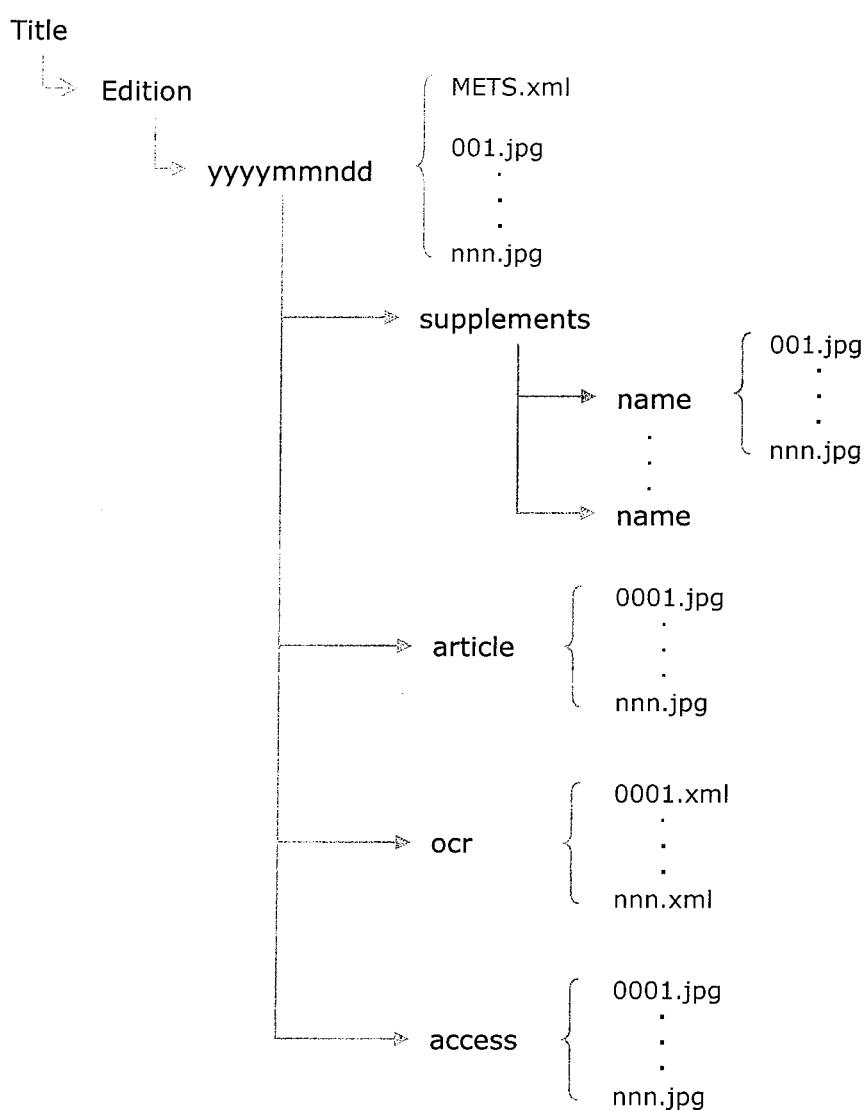


Figure 2 Directory structure and file naming for newspaper issues

Table 3² provides more information about each of the directories and the respective file-naming used within the directory

² [] indicates the directory is optional

File/Directory	Description
title	The agreed bibliographic title. Note during the initial engagement process this may be shortened to a unique identifier.
title/[edition]	If more than one edition is being supplied then the directory hierarchy should include the name of the edition. Note during the initial engagement process this may be shortened to a unique identifier.
title/[edition]/yyyymmdd	The date of the issue
title/[edition]/yyyymmdd/METS.xml	The METS file containing the basic descriptive data. See Appendix C for examples of how this is built up.
title/[edition]/yyyymmdd/ <i>nnn</i> .<image type>	Page images in the agreed format where <i>nnn</i> is the number of the page.
title/[edition]/yyyymmdd/[article]	Articles where available
title/[edition]/yyyymmdd/[article]/ <i>nnnn</i> [- <i>nn</i>].<image type>	Article files where appropriate. The '- <i>nn</i> ' is optional and can be used when an article spans more than one page. Note can accommodate 9999 articles per issue.
title/[edition]/yyyymmdd/[supplement]	Supplement where available
title/[edition]/yyyymmdd/[supplement]/name- <i>nnn</i> .<image type>	Note during the initial engagement process this may be shortened to a unique identifier.
title/[edition]/yyyymmdd/[ocr]	OCR files are available.
title/[edition]/yyyymmdd/[ocr]/ <i>nnn</i> .xml	OCR file where <i>nnn</i> is the number of the relevant page image. Should be one OCR file per page image.
title/[edition]/yyyymmdd/[ocr] / [supplement]/name- <i>nnn</i> .<image type>	OCR file for the supplements. Should be one OCR file per page image.
title/[edition]/yyyymmdd/[ocr]/[article]/ <i>nnn</i> .<image type>	OCR file for each article. Should be one OCR file per article image.
title/[edition]/yyyymmdd/[access]	Access copy or copies of the files are contained.
title/[edition]/yyyymmdd/[access]/ <i>nnn</i> .<filetype>	Access files where <i>nnn</i> is the number of the relevant page image.
or title/[edition]/yyyymmdd/[access]/issue.<filetype>	Access file for the complete issue
title/[edition]/yyyymmdd/[access]/[supplement]/ <i>nnn</i> .<filetype>	Access files where <i>nnn</i> is the number of the

>	relevant page image.
or	
title/[edition]/yyyymmdd/[access]/[supplement]/issue.<filetype>	Access file for the complete issue

Table 3. Preferred directory structure and naming conventions for files

The minimum level of structure and naming of files for delivery to the library of **an issue** is:

- No directory structure
- A naming convention for each file which consists of the following detail:
 - Title
 - Edition
 - Date of issue
 - Page number
- An issue manifest detailing the number of pages for the issue and any supplements.
Note the manifest would be named in accordance with the issue that it is describing.

Note in the event that a number of issues are put into one delivery package then it is envisaged that there would be a delivery package manifest which identifies the issues held in the package.

3.3.3 File delivery

In terms of how the scanned and 'born digital' files arrive at the Library, Table 4 describes the preferred and minimum mechanism.

	Preferred Option	Minimum
Granularity	Receive files as a complete package of content for an issue	On a page by page basis
Checksum	Yes	None
Mechanism	Content supplier pushes content onto a Library ftp server	Content supplier sends issues by post on a hard disk drive
Schedule	At the content suppliers convenience	

Manifest(s)	Issue level manifest conformant to Library METS Newspaper Profile Package level manifest detailing issues contained within package	Content supplier defined manifest.
-------------	---------------------------------------------------------------------------------------------------------------------------------------	------------------------------------

Table 4. Delivery mechanisms for files

No acknowledgement will be sent by the Library back to the content supplier that files have been successfully received by the Library. However, a report of issues successfully ingested into the DLS can be sent on a regular basis as determined during the engagement process, see section ??.

3.3.4 Optical Character Recognition

In order to open the collection up to a richer discovery experience for Library customers, it is requested that content suppliers provide us with a mechanism for enable searching within the text of the newspaper.

Table 5 describes the various characteristics to be considered for OCR of scanned images.

	Preferred Option	Minimum
Delivery mechanism	ALTO xml files for each page image	Proprietary XML file
Character accuracy	85%	70%
Word accuracy	80%	60%
Significant word accuracy	70%	55%
Words with capital letters	65%	50%
Word and character confidence factors	Yes	No
Word level position coordinates captured	Yes	No
Article level position coordinates captured	Yes	No

Table 5. OCR and scanned images

Table 6 describes for born digital content where the artefacts may be produced in a document format such as PDF rather than an image format. However, if delivered as a set of image formats then the OCR capability should be as Table 5.

Preferred Option	Minimum
Fully searchable	

Table 6. Searching and born digital content

3.3.5 Descriptive and structural meta data

In order for the Library to understand the content and the structure of the content being sent to them, it is necessary for some meta-data to be associated with each of the issues. Table 7 details the preferred and minimum requirements for this information.

Preferred Option	Minimum
<p>A populated METS³ file as per the Library standard, see Appendix C for the XML Schema.</p> <p>Note the intent is for the Library to supply the content provider with some mechanism whereby content suppliers do not have to hand craft the XML file.</p>	<p>A simple manifest detailing:</p> <ul style="list-style-type: none"> • The number of pages in the issue • Supplements in issue and number of pages • OCR files included • Articles and which pages they are on

Table 7. Meta data requirements

3.4 Conclusion

This section has detailed all the various aspects of the content which need to be considered before ingest in to the Library digital infrastructure can be started.

³ Metadata and Encoding Transmission Standard (<http://www.loc.gov/standards/mets/mets-profiles.html>)

4 Conclusions

Appendix A. Engagement Process Information Capture

This Appendix details the information that needs to be captured by the content supplier in order for the Library to be able to start to configure the ingest process.

Note this Appendix can be provided to the content supplier as an Excel spreadsheet if easier to complete.

B.1 Administrative Information

The following sections capture the various contact information that is required in order for the Library to manage the engagement process and the ongoing live operation.

B1.1 Publisher details

Publisher Name	
Address 1	
Address 2	
Town	
County	
Postcode	
Telephone	

B1.2 Main contact details

Name	
Role	
Address 1	
Address 2	
Town	
County	


Postcode	
Telephone	
Email address	

B1.2 Secondary contact details

Name	
Role	
Address 1	
Address 2	
Town	
County	
Postcode	
Telephone	
Email address	

B1.3 Other details

Anticipated start date	

	Supplier Engagement	Version 1.0	Date 25/06/2009
	e-Strategy and Information Systems		

B2.1 Publication information

For each of the titles published that will be part of this engagement please complete the following table, adding an additional row for each new title or edition.

Reference	Title	Edition	Main Edition	Frequency	Supplements	ISSN
Line number	Name of title	1 line per edition of title	Yes or No	When published	Supplements produced per publication day	

B3.1 File information

The following section provides essential information about the individual files, how they are received by the Library and if any batching occurs before delivery.

Can the preferred naming convention as described in section 3.3.2 be used. If no, please describe the naming conventions currently used.	Yes or No
What format will the page images be sent in?	JPEG JPEG2000 TIFF PDF Other
Will article files be available? If yes, in what format will they be sent?	Yes or No
Will OCR files be available? If yes, in what format will they be sent?	Yes or No
Will Access files be available? If yes, in what format will they be sent?	Yes or No
What will be the file delivery level to the Library be?	Individual pages Individual issue Batched issues of same titles Batched issues of various titles Other
What will the mechanism for getting the files to the Library be?	Push to Library FTP server Pull from Publisher server Email CD/DVD Hard disk Other
If the files are being batched, what will the batching mechanism be?	Individual issues Zip Tar Other

How regularly will the files be available for the Library?	<i>As produced</i> <i>After issue published</i> <i>Particular time of day/week/month/year</i> <i>Ad hoc</i> <i>Other</i>
------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------

B4.1 Meta Data Information

The following section provides essential information about the meta data that is transmitted with either the:

- Individual files or
- With the batch.

B4.1.1 Issue level metadata

Issue level metadata is all about providing the Library with a mechanism to determine that the content being stored within the DLS is correct and has nothing missing which would impact future use of the material by customers of the Library.

Can a METS file be produced for the issue? If No, then the following questions have to be answered.	<i>Yes or No</i>
Is it possible to produce a proprietary manifest?	<i>Yes or No</i>
If yes, what is the format and how can the Library decode what is being described?	<i>Describe the manifest in detail.</i>
If no, how can we deduce what files belong to what issues etc.	<i>Describe the mechanism for understanding which files belong to which issue.</i>

B4.1.2 Batch level metadata

Batch level metadata is all about providing the Library with a mechanism to determine what issues have been supplied within the package.

Does the batch have a manifest	<i>Yes or No.</i>
If yes, what is the format and how can the Library decode what is being described?	<i>Describe the manifest in detail.</i>

If no, how can we deduce what issues make up the batch

Describe the mechanism for understanding which issues are contained within the batch.

Appendix B. Report from Digitise on Demand Proof of Concept

In progress, results to follow.

Appendix C. METS Schema with examples

This Appendix takes the first two pages of the Pall Mall Gazette for 1st August 1867 and illustrates how the METS Profile for Newspapers would be created. It also provides examples of the naming conventions as described previously in section 3.3.2.

The next 3 pages show the following:

- Pall Mall Gazette for 1st August 1867 Page 1 master image
- Pall Mall Gazette for 1st August 1867 Page 2 master image

The two pages have then been decomposed in to the articles that make up the 2 pages.

- Article 1 image (blue) found on Page 1
- Article 2 image (red) found on Page 1 and continued on Page 2
- Article 3 image (green) found on Page 2
- Article 4 image (purple) found on Page 3

Following the guidelines laid down in section 3.3.2 these files would exist in the following directory structure with their respective names

PallMallGazette\18670801\001.tif

PallMallGazette\18670801\002.tif

PallMallGazette\18670801\article\0001.tif

PallMallGazette\18670801\article\0002-1.tif

PallMallGazette\18670801\article\0002-2.tif

PallMallGazette\18670801\article\0003.tif

PallMallGazette\18670801\article\0004.tif

PallMallGazette\18670801\ocr\0001.xml

PallMallGazette\18670801\ocr\0002.xml

PallMallGazette\18670801\ocr\0003.xml

PallMallGazette\18670801\ocr\0004.xml

PallMallGazette\18670801\presentation\001.tif

PallMallGazette\18670801\presentation\002.tif

[illegible]

Page 33 of 46

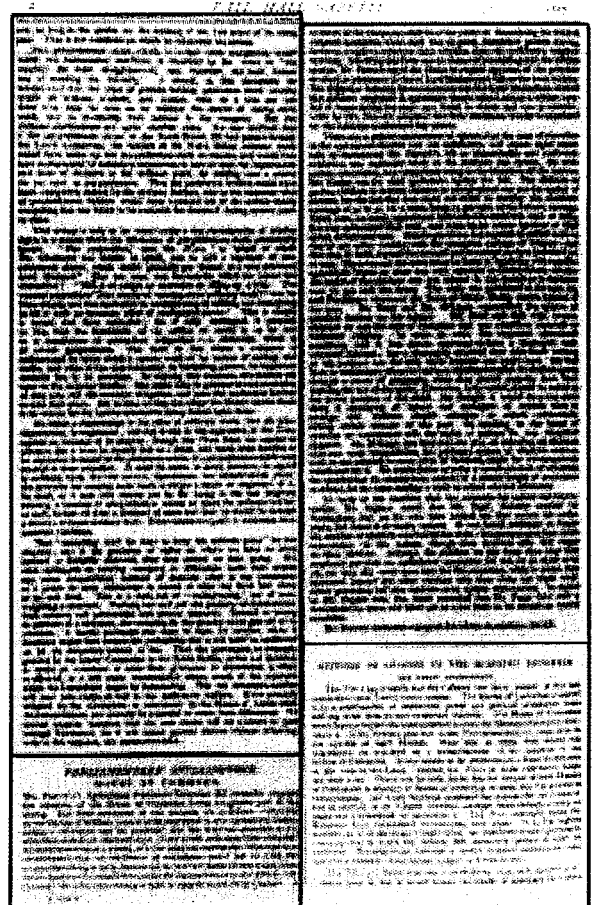
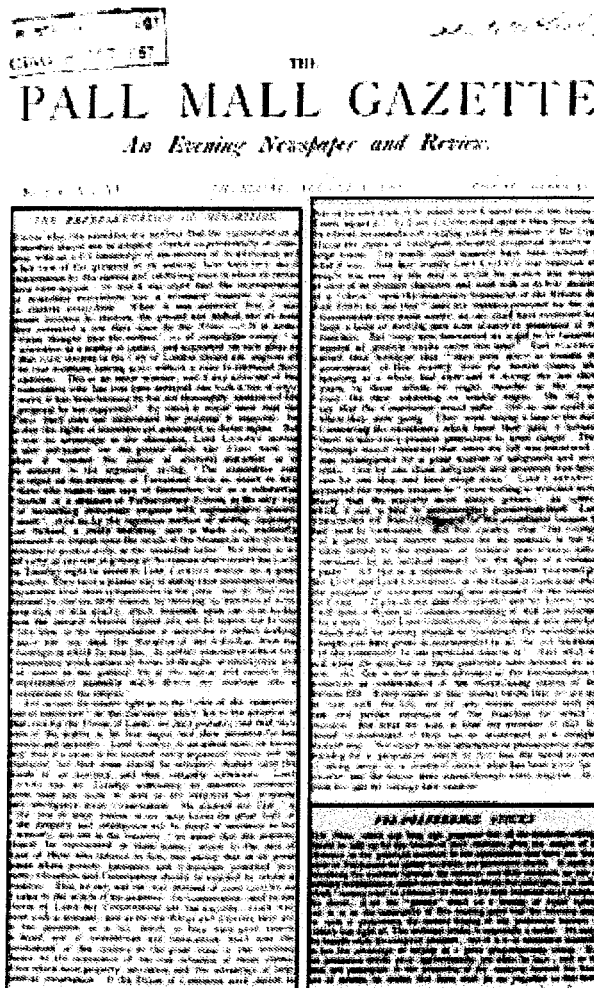


Figure 5 Article decomposition of Pages 1 and 2

C.1 Structure of a METS file

Before considering how the artefacts described in the introduction to this Appendix are incorporated into the METS files it is important to understand that the METS file consists of for this purpose 5 distinct logical sections, see Figure 6.

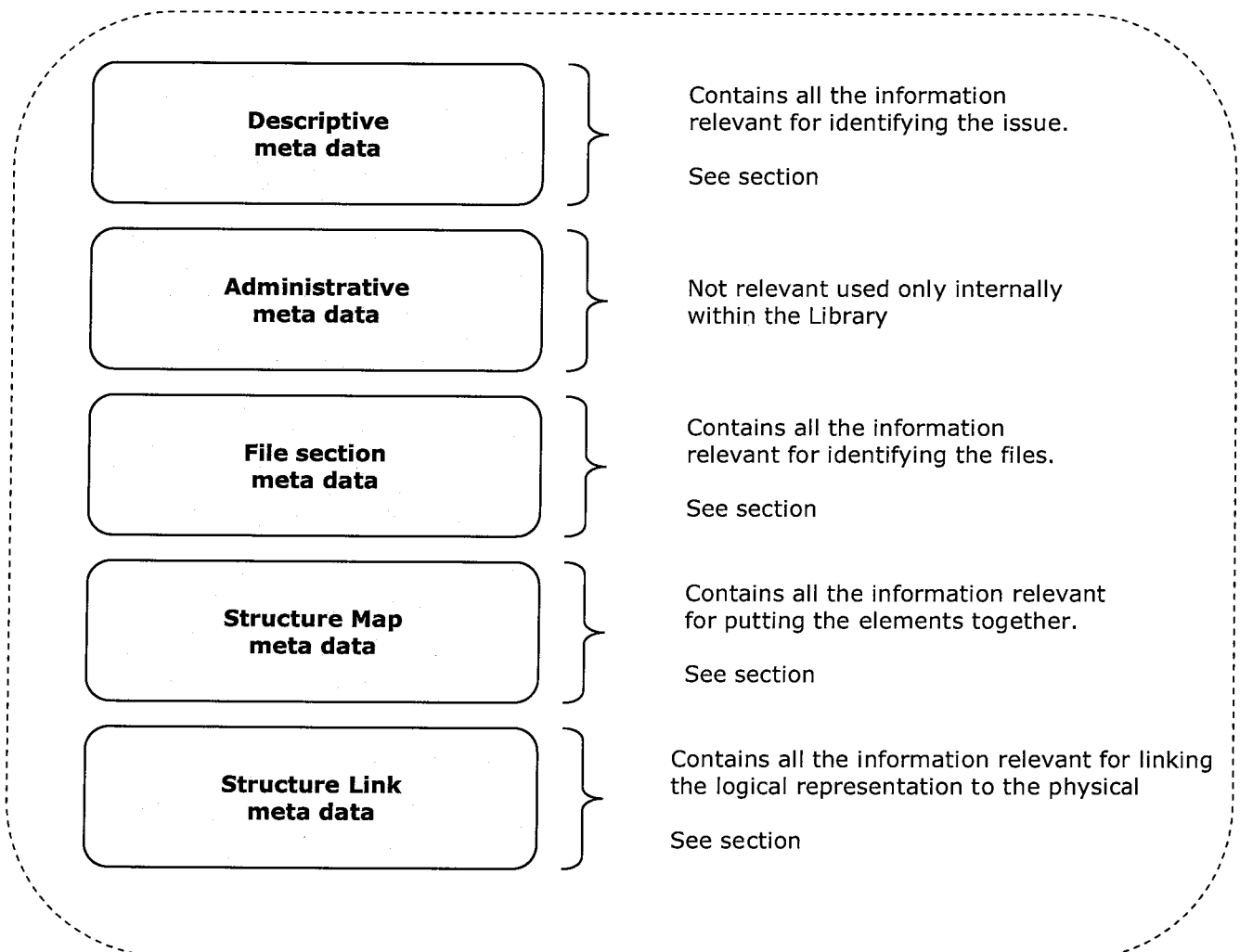


Figure 6 Breakdown of METS file

The following sections show how each of these sections are built up in order to provide a consistent and coherent METS file that can be utilised during the ingest process at the Library.

Step 1: Physical Structure

This section enables the physical structure of the paper to be captured, see Figure 7.

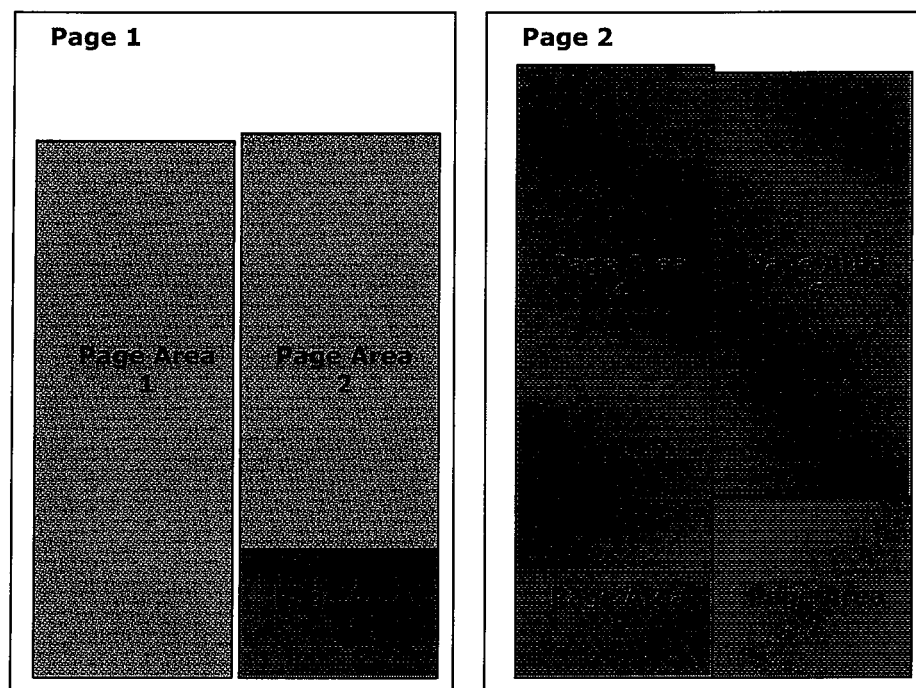


Figure 7 Physical structure of the example

In order to accomplish this representation, then a physical structure map is created within the METS file. Figure 8 details how the part of the METS file would be populated for our two page example.

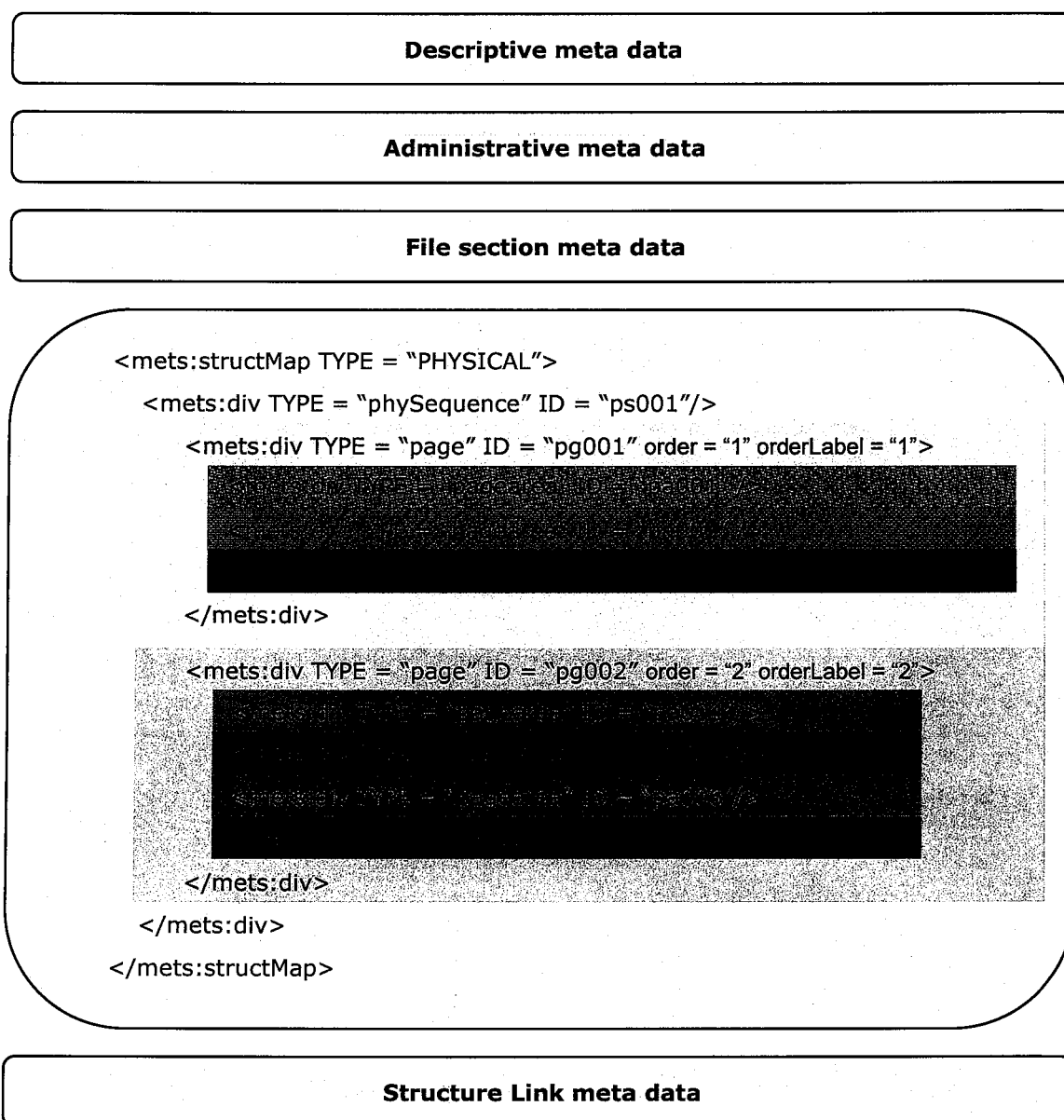


Figure 8 METS fragment for mapping the physical structure of the newspaper

Step 2: Logical Structure

This section enables the logical structure, in this case, an article view of the newspaper to be captured. If we consider our example 2 pages then there are 4 articles which need to be defined logically, see Figure 9.

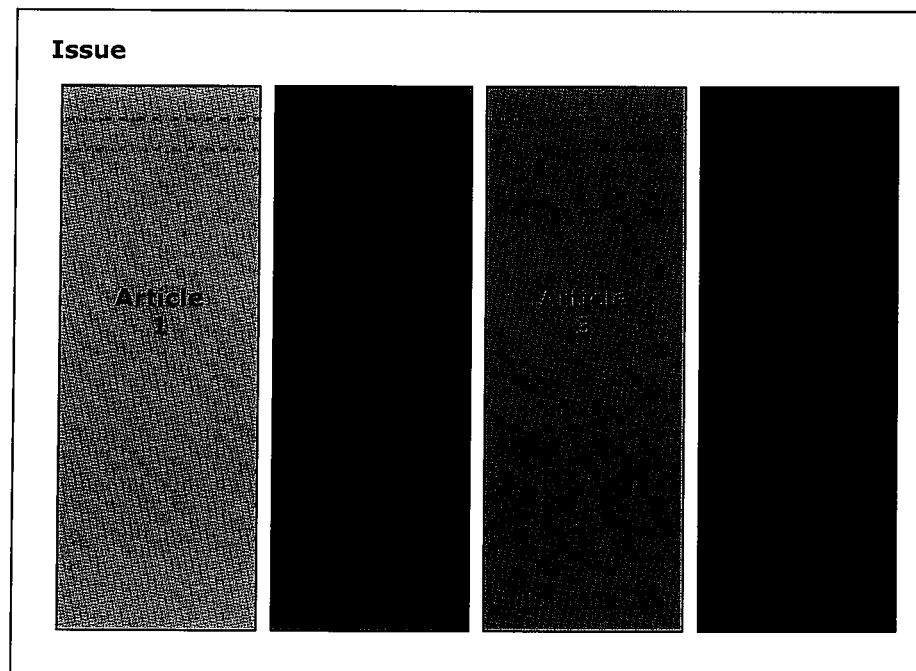


Figure 9 Logical structure of the newspaper

In order to accomplish this representation, then a logical structure map is created within the METS file. Figure 10 details how the part of the METS file would be populated for our two page example.

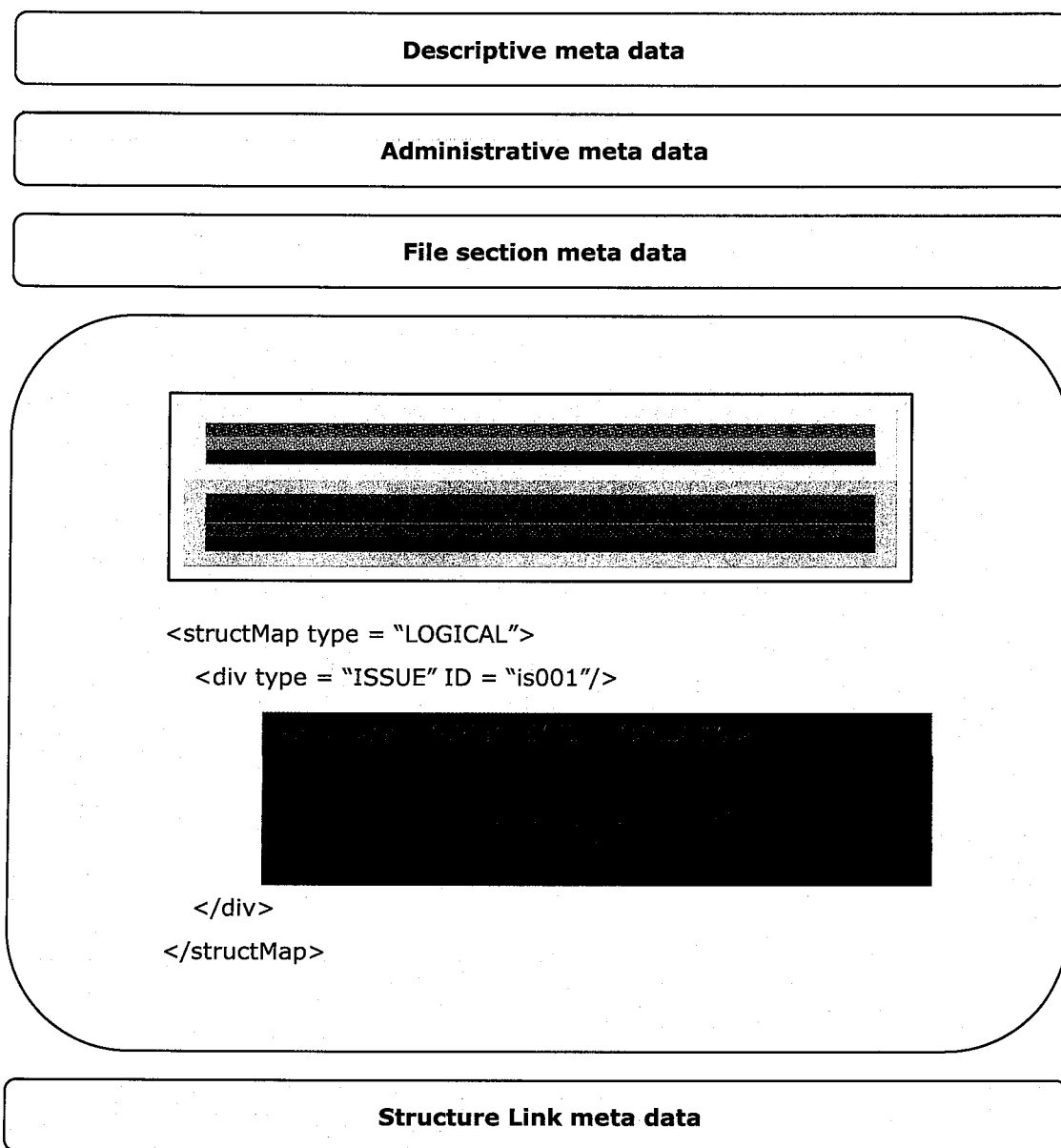


Figure 10 METS fragment for mapping the logical structure of the newspaper

C.3 Link Logical to Physical

Once the physical and logical elements of the newspaper have been defined, and then the next task in the process is to link the logical elements with the physical elements, see Figure 12.

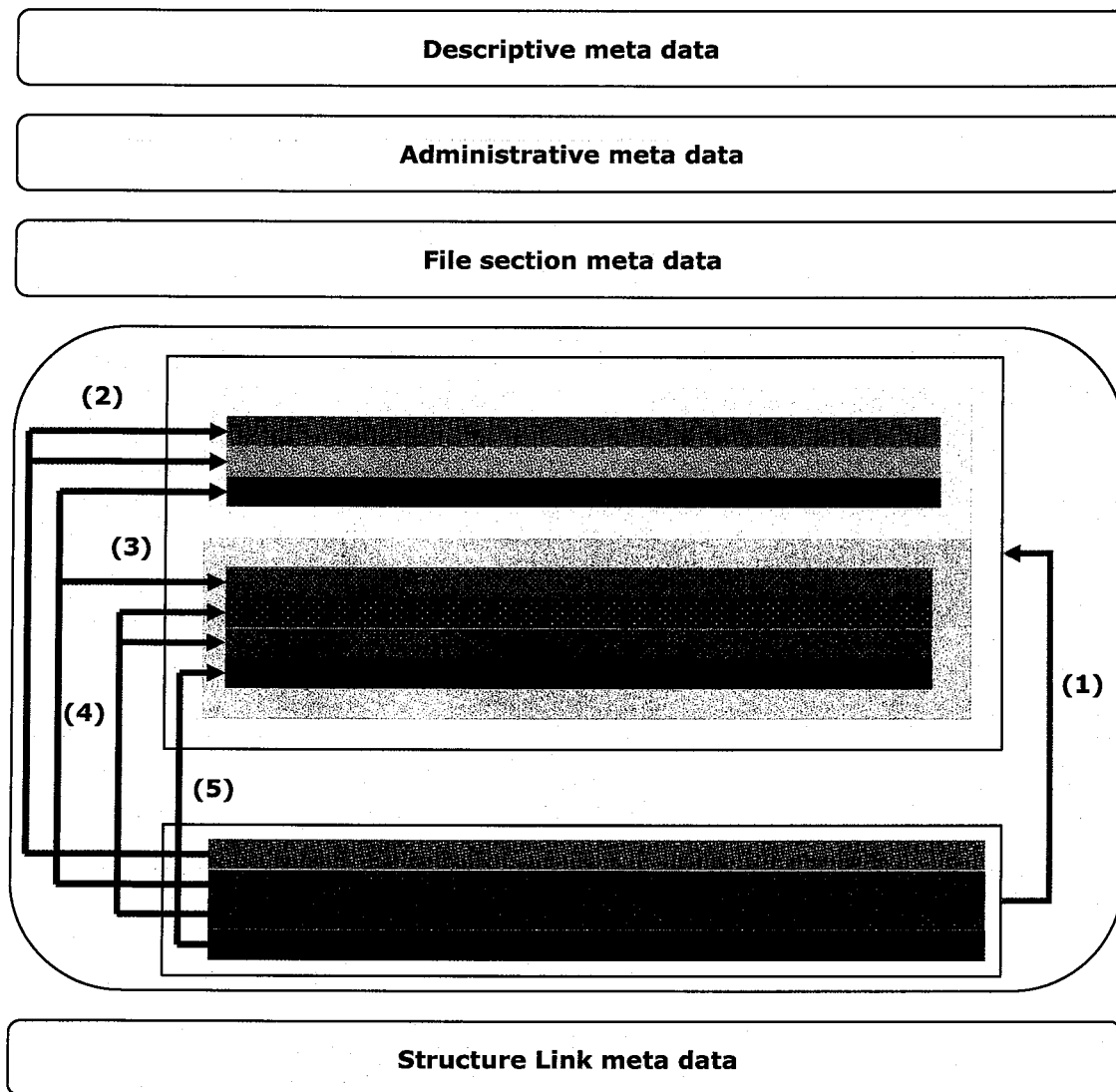


Figure 11 Linking between the logical and physical elements previously defined

In order to accomplish this representation, then a structure link is created within the METS file. Figure 12 details how the part of the METS file would be populated for our two page example.

Descriptive meta data

Administrative meta data

File section meta data

Structure map meta data

<structLink>

```
<smLinkGrp>
  <smLocatorLink xlink:href="#I001" xlink:label="issue"/>
  <smLocatorLink xlink:href="#ps001" xlink:label="physical sequence"/>
  <smArcLink xlink:type="arc" xlink:from="issue" xlink:to="physical sequence"
    ARCTYPE="logicalphysical"/>
</smLinkGrp>

<smLinkGrp>
  <smLocatorLink xlink:href="#ar0001" xlink:label="article1"/>
  <smLocatorLink xlink:href="#pa001" xlink:label="pageArea1"/>
  <smLocatorLink xlink:href="#pa002" xlink:label="pageArea2"/>
  <smArcLink xlink:type="arc" xlink:from="article1" xlink:to="pageArea1" ARCTYPE="logicalphysical"/>
  <smArcLink xlink:type="arc" xlink:from="article1" xlink:to="pageArea2" ARCTYPE="logicalphysical"/>
</smLinkGrp>
```

</structLink>

Figure 12 METS fragment for linking the logical elements to the physical elements of the newspaper

C.4 File definition

Once the overall structure has been defined for the newspaper the next task in the process is to define the actual files that represent the newspaper which has been digitised.

The first part of accomplishing this is to (a) identify the class of files i.e. are the files for preservation, display, OCR etc and (b) where can these files be found in relationship to the METS file location.

In order to accomplish this definition then a set of file groups are created within the METS file. Figure 13 details how the part of the METS file would be populated for our two page example.

The second part of accomplishing this is to add to the previously created physical structure map a file pointer which maps to the underlying physical file as defined in the previous step.

In order to accomplish this definition then a set of file groups are created within the METS file. Figure 14 details how the part of the METS file would be populated for our two page example.

Descriptive meta data
Administrative meta data
<pre><fileGrp use = "DigitalManifestation"> <fileGrp use = "master"> <file ID = "01" mimeType = "image/TIF" size = "5646" checksum = "" checksumtype = "slew-512"> <fileLoc locType = "URL" xlink:href = "/001.tif"/> </file> <file ID = "02" mimeType = "image/TIF" size = "10666" checksum = "" checksumtype = "slew-512"> <fileLoc locType = "URL" xlink:href = "/002.tif"/> </file> </fileGrp> <fileGrp use = "presentation"> <file ID = "01" mimeType = "image/TIF" size = "717" checksum = "" checksumtype = "slew-512"> <fileLoc locType = "URL" xlink:href = "/presentation/0001.tif"/> </file> <file ID = "02" mimeType = "image/TIF" size = "763" checksum = "" checksumtype = "slew-512"> <fileLoc locType = "URL" xlink:href = "/presentation/0002.tif"/> </file> </fileGrp> <fileGrp use = "derivatives"> <file ID = "01" mimeType = "image/TIF" size = "137" checksum = "" checksumtype = "slew-512"> <fileLoc locType = "URL" xlink:href = "/derivatives/0001.tif"/> </file> <file ID = "02" mimeType = "image/TIF" size = "137" checksum = "" checksumtype = "slew-512"> <fileLoc locType = "URL" xlink:href = "/derivatives/0002.tif"/> </file> <file ID = "03" mimeType = "image/TIF" size = "137" checksum = "" checksumtype = "slew-512"> <fileLoc locType = "URL" xlink:href = "/derivatives/0003.tif"/> </file> <file ID = "04" mimeType = "image/TIF" size = "137" checksum = "" checksumtype = "slew-512"> <fileLoc locType = "URL" xlink:href = "/derivatives/0004.tif"/> </file> </fileGrp> </fileGrp></pre>
Structure map meta data
Structure link meta data

Figure 13 METS fragment for identifying the various files that have been created during the digitisation process

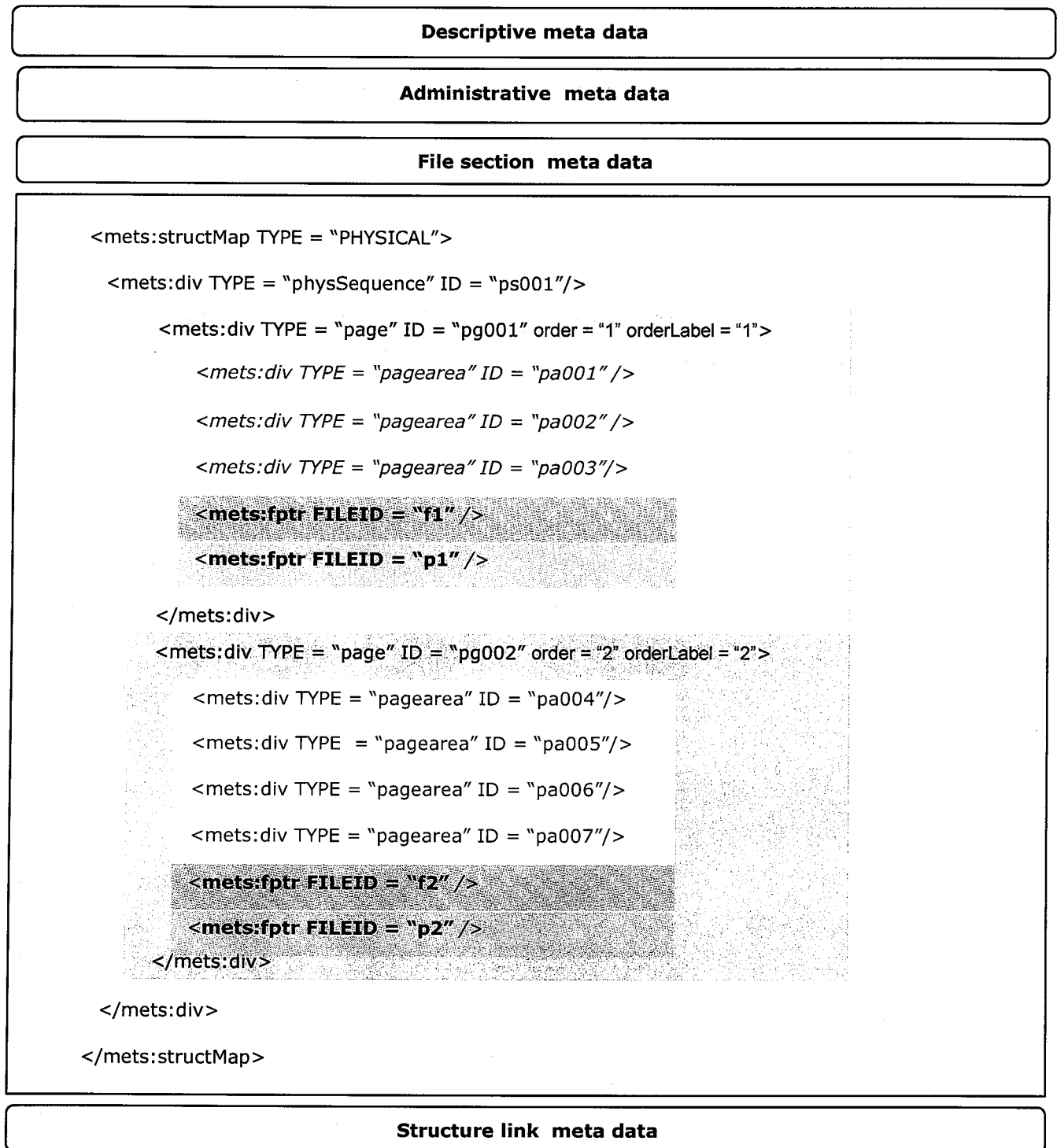


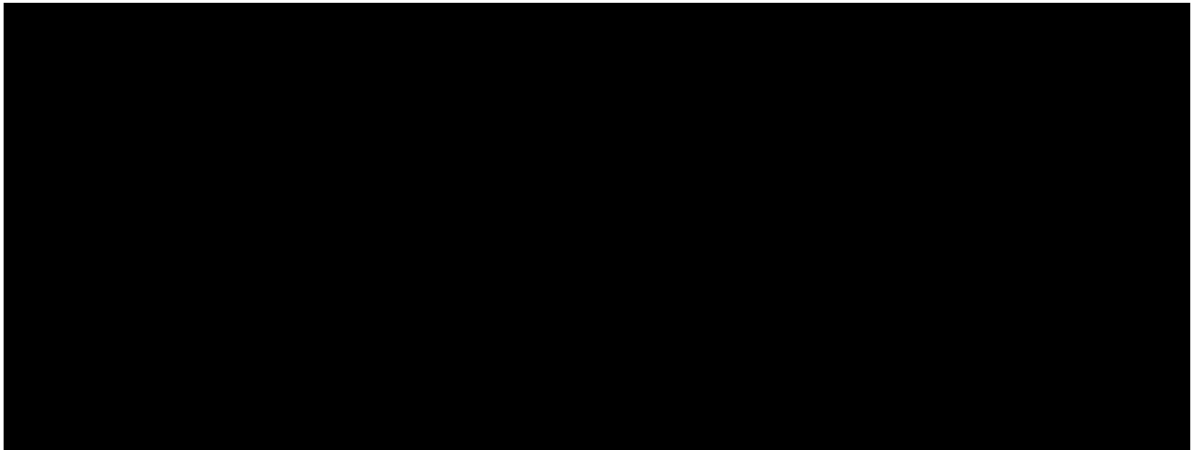
Figure 14 METS fragment for linking the various files to the physical structure map

C.5 Descriptive metadata

Schedule 4

Standards

1. **DEFINITIONS**



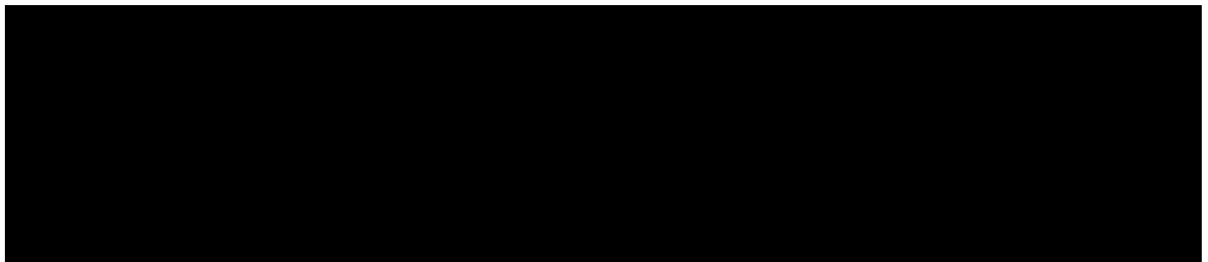
2. **STANDARDS FOR THE DIGITISATION SERVICES - GENERAL**

2.1



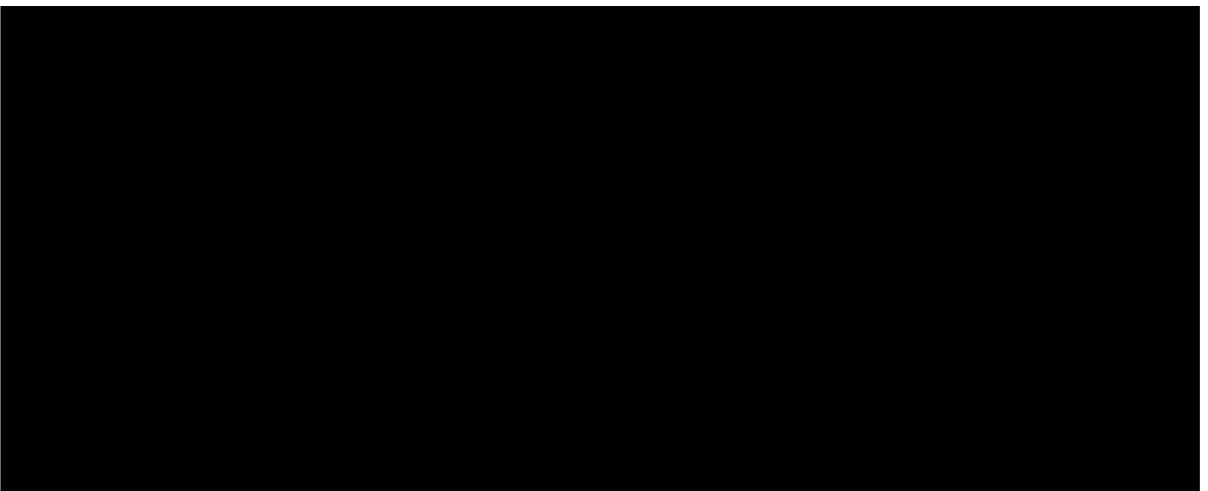
Confidentiality

2.2



Compliance with the Standards

2.3



2.4

2.5

Staff Supervision

2.6

Staff Training

2.7

2.8

2.9

2.10

2.11

3. MATERIALS HANDLING STANDARDS

3.1 General Guidelines on Document Handling when Copying/Scanning

3.2 Responsibilities

3.3 Working with Documents

3.3.2

3.3.3

3.3.4

3.3.5

3.3.6

3.3.7

3.3.8

3.3.9

3.3.10

3.3.11

3.3.12

3.3.13

3.3.14

3.3.15

3.3.16

3.3.17

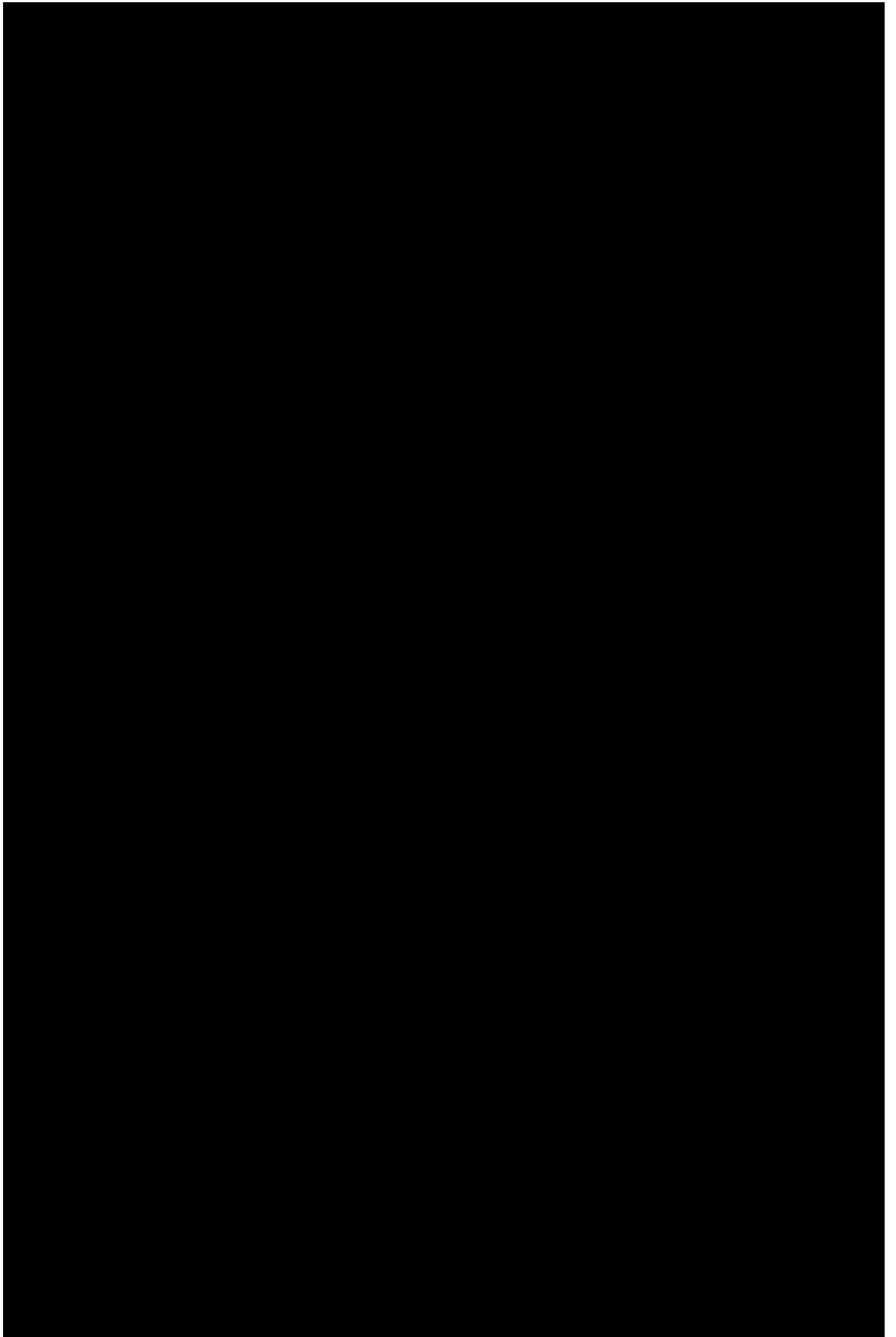
3.3.18

3.3.19

3.3.20

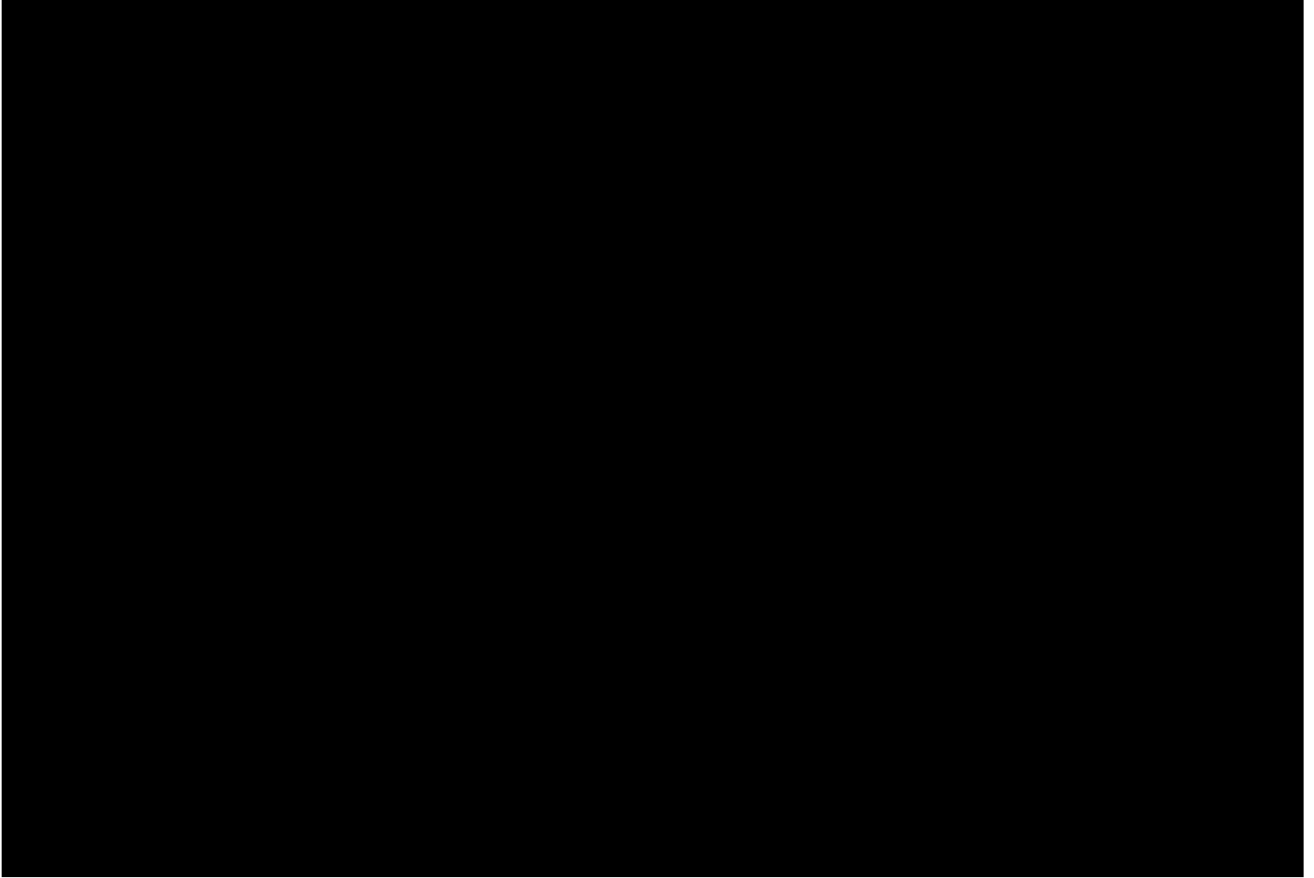
3.4 **Cleaning and Maintenance**

3.4.1

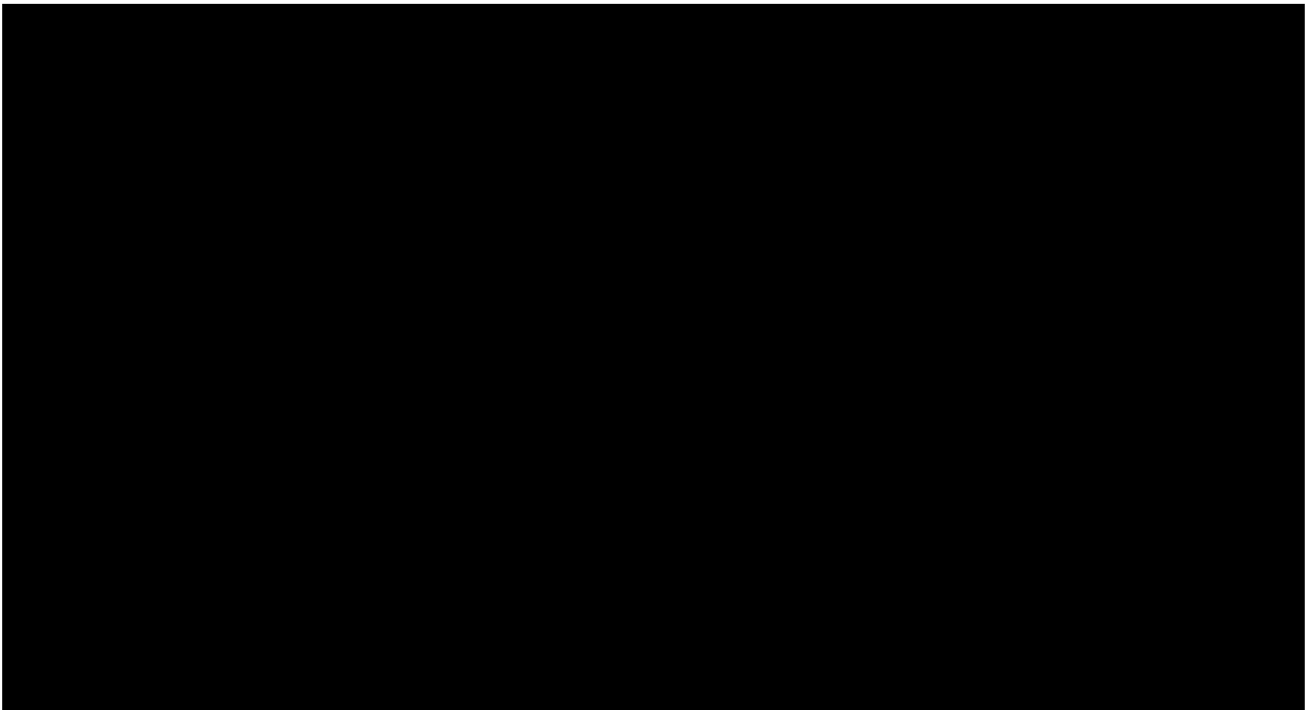


4. IMAGE STANDARDS

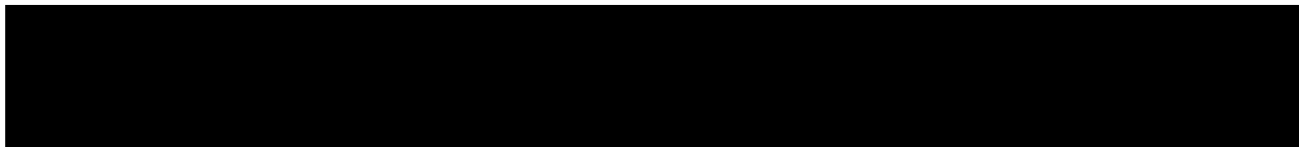
Part A: Image Format:



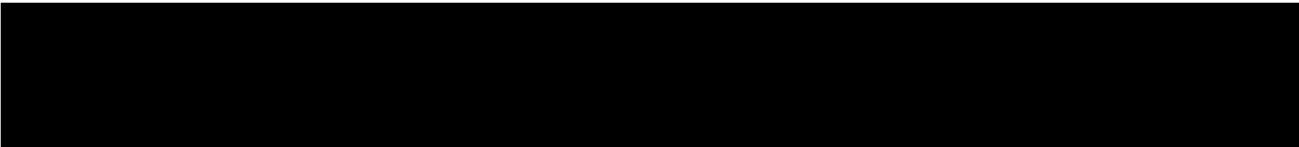
Part B: Metadata Format:



Part C: Ingest process:

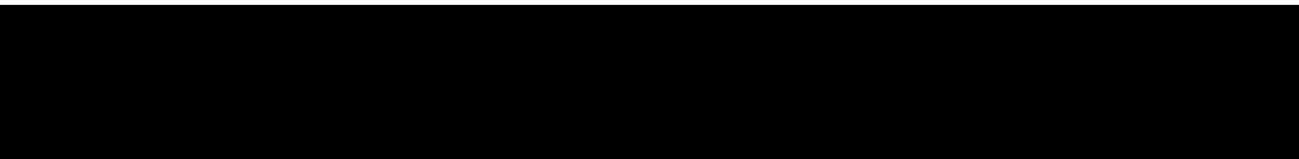


Part D: Microfilm Scanning



5. OCR STANDARDS

Accuracy and rules



6. INGEST STANDARDS & REQUIREMENTS



7. STANDARDS FOR THE ONLINE ACCESS SERVICE

Accessibility

