

GENERAL SPECIFICATION

Geoinformation

GS EP GIN 103

Dematerializing geoscience paper documents

02	10/2016	Typing correction (new Managing entity code, new template)
01	06/2015	Complete revision (new title, document format and content)
00	09/2011	Update of scanning and profiles, addition of the rule about bookmarks. Cancels and replaces GS EP SUG 103
Rev.	Date	Purpose of the revision

Owning entity: EXPLO/GTS	Managing entity: EXPLO/GTS/IGR
Other Approving entity(ies):	

General Specification		GS EP GIN 103
Dematerializing geoscience paper documents		
Rev.: 02	Effective date: 10/2016	Page: 2 of 19

Contents

1. Scope	3
2. Reference documents.....	3
3. Definitions.....	4
4. Roles and responsibilities.....	4
4.1 Document and Content Manager	4
4.2 Operator.....	5
5. Dematerialization process.....	5
5.1 Step 1 - Scanning	5
5.2 Step 2 - Generating the PDF file: generator settings	5
5.3 Step 3 - OCR processing	6
5.4 Step 4 - Structuring the PDF file.....	6
5.5 Step 5 - Quality control.....	7
Appendix 1 Image type settings	9
Appendix 2 PDF generator settings.....	10
Appendix 3 Document properties and footer settings	18
Appendix 4 Bookmark settings.....	19

General Specification		GS EP GIN 103
Dematerializing geoscience paper documents		
Rev.: 02	Effective date: 10/2016	Page: 3 of 19

1. Scope

This General Specification defines the requirements for dematerializing paper documents and converting them to PDF format.

Its objective is to guarantee the integrity of the dematerialized document, its readability and print quality, and uniformity of its PDF layout.

This General Specification applies to any person, at headquarters or in an affiliate, company or contractor employee, who generates PDF files from paper documents.

It applies to all the documents intended for archiving, as defined in [GS EP GIN 305](#).

2. Reference documents

The reference documents listed below form an integral part of this General Specification.

External Documents

Unless otherwise stipulated, the applicable version of these documents, including relevant appendices and addendums, is the latest revision published at the effective date of this document.

Reference	Title
Not applicable	

Total General Specifications

Unless otherwise stipulated, the applicable version of these documents, including relevant appendices and addendums, is the latest revision published.

Reference	Title
GS EP GIN 305	Indexing of corporate technical documents

General Specification		GS EP GIN 103
Dematerializing geoscience paper documents		
Rev.: 02	Effective date: 10/2016	Page: 4 of 19

3. Definitions

The following definitions serve for the purposes of this document only.

Acrobat Distiller: Adobe Acrobat software program used for generating files in PDF format.

Acrobat Distiller Profiles: models containing the settings for conversion into PDF format.

Body of the document: substantial part of the document that presents the development and conclusions. It may include illustrations, such as figures, tables, photo plates, etc.

Bookmarks: digital table of contents of a PDF file. Bookmarks make it easier to navigate in a document and search through it.

Dematerialization: actions and techniques used to transform printed documents into digital format.

Document: independent unit comprising an information media of any type and the data recorded on it.

Document appendices: text, figures, tables and plates added to the main body of the document.

DPI (Dot Per Inch): measurement unit used to define the resolution of an image, a scanner or a printer.

Figures: illustrations (drawings, maps, diagrams, etc.) of different sizes included in the text or in A4/A3 format and bound after the text.

Integrity: state of a document that has preserved its content, qualities and original state, without any alteration.

Metadata: data that describes or defines the information about the document.

OCR (Optical Character Recognition): capacity of a software program to recognize previously scanned printed characters and to convert them into digital text format. This OCR processing makes it possible to carry out a full-text search inside a document.

PDF (Portable Document Format): universal file format that retains the fonts, formatting, colors and images of the source documents, regardless of the application or platform used to create it. PDF files are compact and can be shared, viewed and printed using Adobe Acrobat Reader or an equivalent software program.

Plates: all the oversize drawings, maps, diagrams, etc., folded into A4 and placed after the text.

Scanning: action of converting a paper document into a digital format (image).

Table of contents: list at the beginning or sometimes at the end of a document with the titles of its different parts. Any list containing the parts of a document (Table of contents, List of..., Index, Chapters, etc.) is considered as a table of contents.

4. Roles and responsibilities

4.1 Document and Content Manager

The Document and Content Manager (DCM):

- Provides the operator with the documents to be scanned.

General Specification		GS EP GIN 103
Dematerializing geoscience paper documents		
Rev.: 02	Effective date: 10/2016	Page: 5 of 19

- Controls the quality of the PDF files converted by the operator.
- Sends the PDF documents to the entity responsible for archiving them (see [GS EP GIN 305](#)).

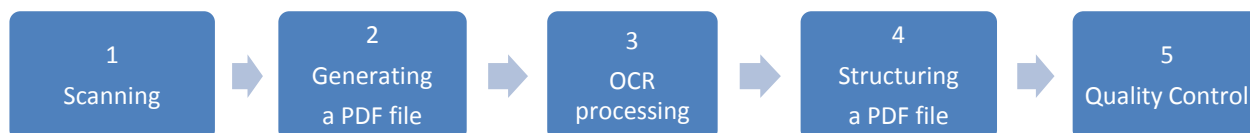
4.2 Operator

The operator:

- Scans the paper documents.
- Runs the OCR processing.
- Generates the files in PDF format.
- Structures the PDF files created.
- Controls the quality of the files before delivery.
- Sends the PDF files created to the DCM.

5. Dematerialization process

The process below describes the main steps for converting paper documents into PDF files.



5.1 Step 1 - Scanning

{Rule 1:

Operators shall scan paper documents (pages with text and illustrations) at a 300 DPI resolution.}

{Rule 2:

Operators shall respect pre-defined settings according to the image type.}

Pre-defined settings concern text pages (background, text, elements and illustrations included) and illustrations (plates, maps, photos, etc.). The required settings are detailed in Appendix 1.

5.2 Step 2 - Generating the PDF file: generator settings

{Rule 3:

The settings pre-defined by the Company shall not be modified irrespective of the PDF generator used.}

The Acrobat Distiller "Total Standard quality" and "Total Photo quality" profiles have been pre-defined to guarantee the integrity of the document, its readability and print quality. The profiles are detailed in Appendix 2.

5.3 Step 3 - OCR processing

{Rule 4:

The software program used for OCR processing shall guarantee an optical recognition rate higher than 80%.}

Note: The optical recognition rate is calculated according to the number of recognized words.

OCR processing concerns pages with text and illustrations in the following sizes:

<i>Elements</i>		<i>Size</i>	<i>Type</i>	<i>OCR</i>
Text pages		A4/A3	Standard*	Yes
Illustrations	Plate, Figure, Map, Log, etc.	A4/A3	Standard*	Yes
			Technical**	
		Larger than A3 (A2, A1, A0, etc.)	Standard*	No
			Technical**	

* documents containing mainly text and tables.

** documents composed of maps, logs, seismic cross-sections, photos, graphic elements, etc.

{Rule 5:

The specific "scan" (image format) and "OCR" (text format) layers shall be included in the PDF file.}

Document content can be viewed on the specific "scan" layer in image format, the full-text search is done on the specific "OCR" layer in text format.

5.4 Step 4 - Structuring the PDF file

5.4.1 Document structure

It may be necessary to limit file size to 500 Mo, if conditions for viewing documents are poor (network quality, etc.), in which case, the document may need to be divided into several files. It is advisable to respect the document structure (body of text, tables, figures, plates, appendices, etc.). Documents divided into several files are searchable using the bookmarks.