



DELIVERABLE

Project Acronym: thinkMOTION
Grant Agreement number: 250485
Project Title: Digital Mechanism and Gear Library goes Europeana

D5.3 - Final report on enhanced digitised and online available content accessible from Europeana

Revision: 1.1

Authors:
Erwin-Christian Lovasz (University Politehnica of Timisoara)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	x
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Organisation	Description
1.0	01.06.2013	E.-C. Lovasz	UPT	Final report
1.1	07.06.2013	E.-C. Lovasz	UPT	Update of figures

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

1	Introduction	4
2	Workflow.....	4
3	Results	7

1 Introduction

During the three years of the project, the activity in WP5 became more and more efficient so that the production of items available online assured the attaining of targets.

The objective of WP5 was to achieve a proper web-content representation in different levels of qualities and to make the content acceptable to a broad range of users by processing approved processing steps. The result of this work package is the online available content accessible through Europeana portal.

The tasks to carry out within WP5 are briefly described as follows:

- Task 5.1: OCR (Optical Character Recognition) of textual content
- Task 5.2: Quality improvement of digitized content
- Task 5.3: Import of digital existing content
- Task 5.4: Converting into web-compliant content
- Task 5.5: Integrating the content into the online portal of DMG-Lib accessible for Europeana.

WP5 performs activities which connect logically and coherently with tasks in other work packages. The dependencies of WP5 with other WPs are illustrated in Figure 1.

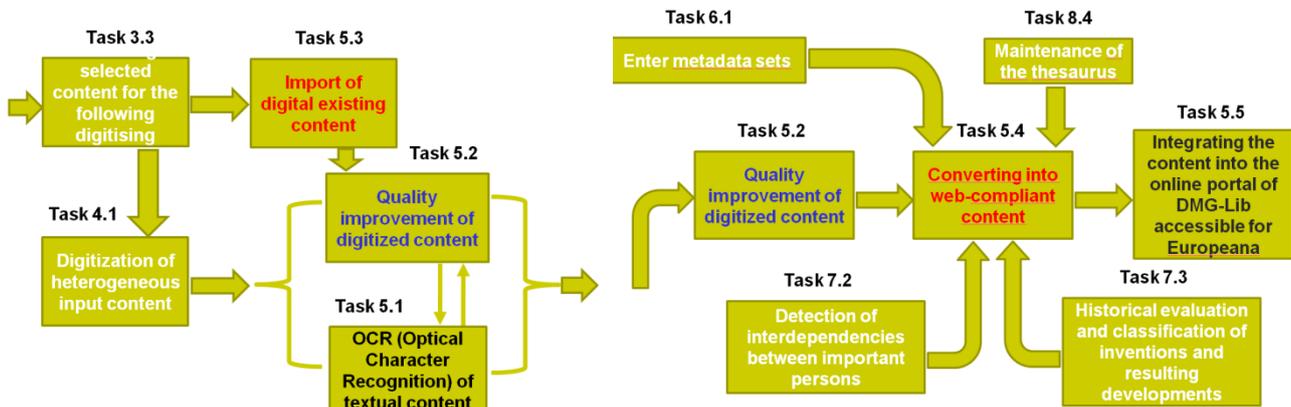


Figure 1. Dependencies of WP5 with other WPs

2 Workflow

Different classes of items, such as documents, images, animations, CAD applications and so on, are processed in WP5. Figure 2 illustrates, for example, the steps needed to obtain an item available online starting with an analogue document.

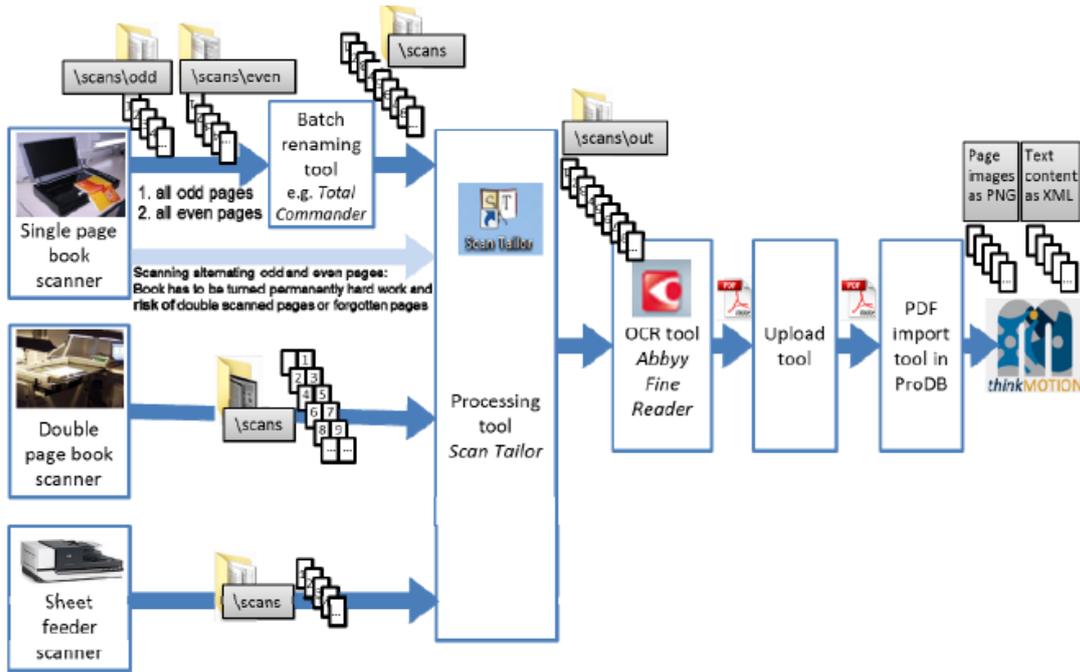


Figure 2. General workflow for processing documents

Figure 3 presents a raw image obtained through scanning and the final look of the processed initial file.

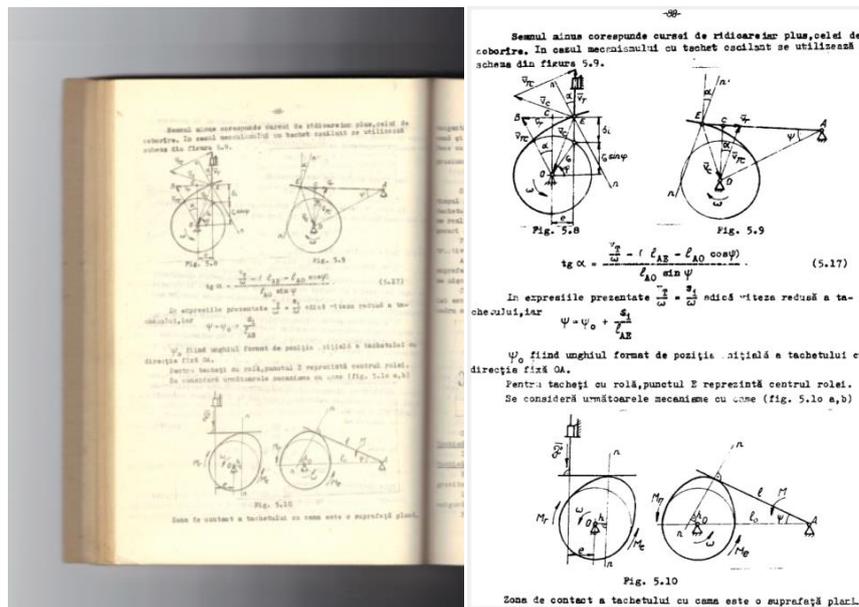


Figure 3. Raw scanned image of a page (left) and its final digitized look online (right)

The scanned raw image of the analogue document results with unavoidable flaws (tilt of written content, non-uniform contrast etc.) and requires an improvement of quality. There are special software tools, such as Scan Tailor, to perform actions such as: splitting the two pages of a book, de-skewing the useful content (Figure 4), cutting the useless margins (Figure 5), resolution and colour mode setting etc.

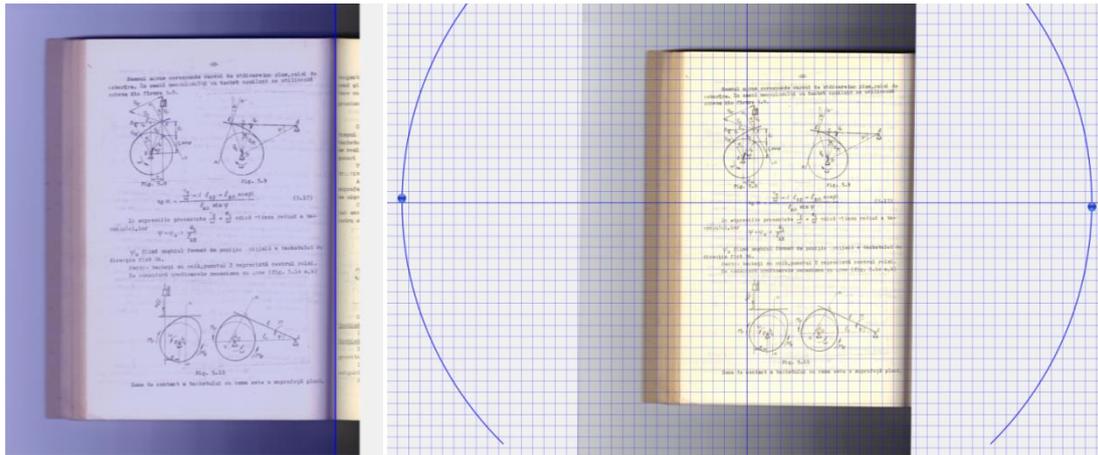


Figure 4. Splitting of pages (left) and automatic de-skewing of one page (right)

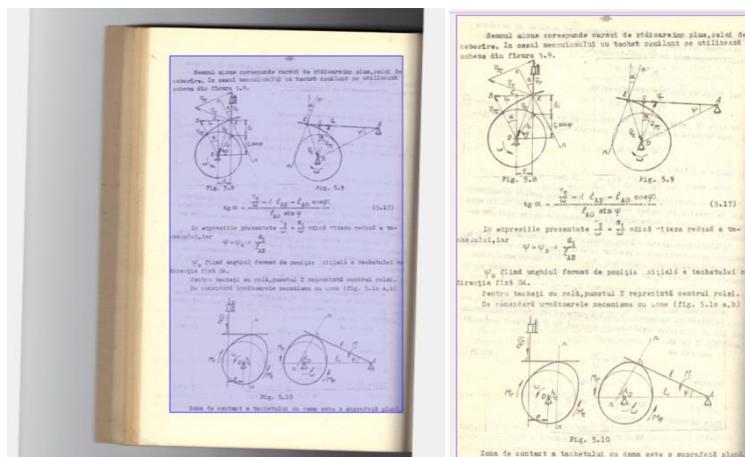


Figure 5. Selection of content (left) and setting margins of the page (right)

If the image quality is set to the desired quality level, then the file is converted to a searchable PDF file. At this stage of processing, WP5 used professional software – ABBYY FineReader. Figure 6 shows a screenshot during the operating with this powerful tool.

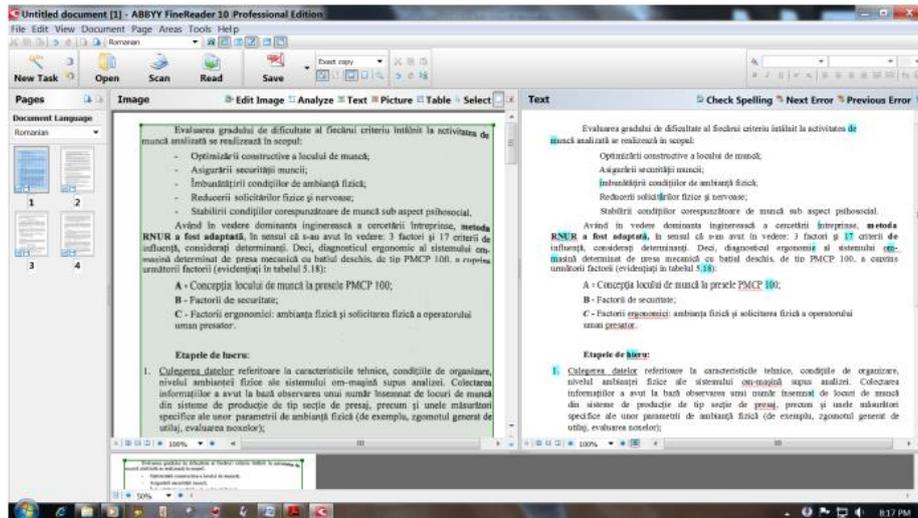


Figure 6. Optical recognition of characters with ABBYY FineReader

Images and photos can be enhanced using specific software such as Adobe Photoshop or IrfanView.

Content published in the last few years often exists in digital form such as PDF or MS Word formats. In this case the step of digitization is skipped and the content is passed on to advanced processing steps.

The output of task 5.2 and task 5.3 must be converted into a web-compliant version. In this processing step the colour depth and the resolution will be reduced and the content is stored in a web-compliant format. For textual content PNG file format is used. To allow a full text search and highlighting of found words in the online portal, the full text must be stored with information about the position of each word in XML-files. The digitized pictures and slides are stored in JPG file format. The taken image sequences of the physical models are converted into video files (MPEG format) and into an interactive animation file format playable by a Java applet inside the DMG-Lib portal.

The last step in the workflow is uploading and integrating the digitized and enhanced content into the DMG-Lib portal. After this step the content is online and accessible from the DMG-Lib portal and also from the Europeana portal, if the metadata pass the internal quality check.

3 Results

During the third year of the project, the activity in WP5 was much more efficient than in the first two years and resulted in production of a large amount of items available online. Figure 7 provides a sequence of the interactive animation created for a physical model and its CAD representation.

The following notices describe the way the teams accomplished the work:

- The target of 62.000 items/project was achieved
- All types of items as planned in the Grant Agreement were created
- The Europeana quality standards were fulfilled
- In order to perform basic processing and quality enhancement the use of a wide range of professional software tools was required, such as ABBYY FineReader v.10, Scan Tailor v.9.9.2, IrfanView v.10, Nitro PDF v.6.2, Adobe Acrobat Pro v.10.4, CATIA, ADAMS,

ProEngineer, SolidWorkd, Cinderella, Kosim, Camstudio, different scanning software and ProDB

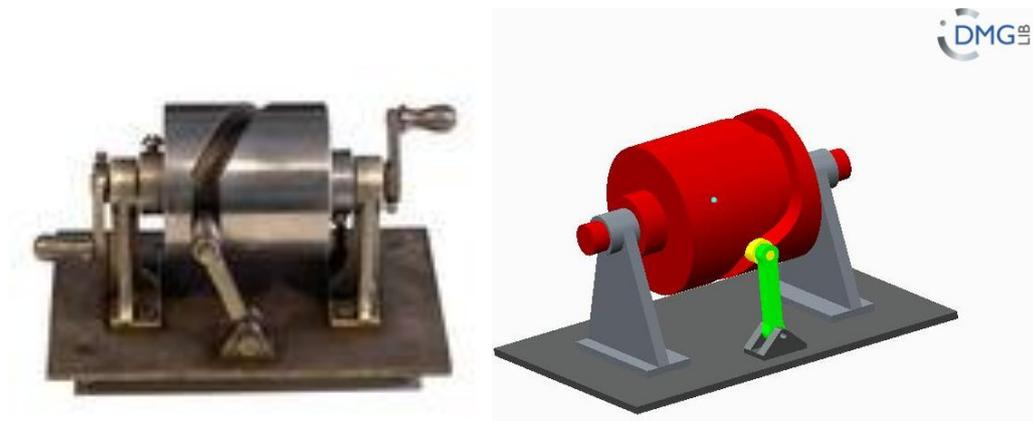


Figure 7. Still image of the physical model (left) and of the CAD model (right)

Significant results may be highlighted, as follows:

- A very large amount of items was processed up to different stages of work; very few content existed in digital form; most items needed performing of all steps in the workflow
- The quality of digitized content was pursued as well as the scientific value of items
- The workflow proved to be correctly conceived so that efficiency of work increased
- The CAx models were generated and processed using advanced complex knowledge and appropriate software
- The partners approached various classes of items
- A series of valuable and rare historical books and journals from the beginning of the twentieth century were digitized (Figure 8)
- Complex software, which is able to synthetize images and produce digital models, was implemented

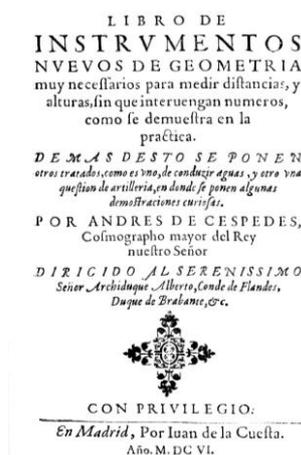


Figure 8. Andrés de Céspedes. Instrumentos nuevos de geometría muy necesarios para medir distancias, y alturas, sin que intervengan números , como se demuestra en la práctica. Imprenta Juan de la Cuesta, Madrid, 1606