



DELIVERABLE

Project Acronym: thinkMOTION

Grant Agreement number: 250485

Project Title: Digital Mechanism and Gear Library goes Europeana

D4.2 - Intermediate report on digitised input content

Revision: 1.0

Authors:

Veit Henkel (Ilmenau University of Technology)

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	30.06.2012	V. Henkel	IUT	Final release

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This deliverable contains original unpublished work except where clearly indicated otherwise.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 and dependencies on other WPs.....	4
3	Results	10
4	Outlook for the next project year.....	10

1 Introduction

In June 2010 partners from six European universities started the project thinkMOTION with the main objective of providing content from the field of motion systems via the Europeana online portal.

The located content, which is proved to be relevant, will be digitised, processed and presented in the multilingual interactive online portal DMG-Lib (www.dmg-lib.org) accessible also via Europeana and by this way also accessible for a wide range of European user groups like interested laymen, engineers, scientists, lecturers and students. The provided interactive material leads to a deeper understanding and motivates to learn more about the scientific and technical background in a European society of lifelong learning. Very different types of content like textual sources, physical demonstration models and drawings will be collected by partners in several European countries. In the library, new ways of content representation, information retrieval (supported by a multilingual semantic network) and cross-linking are supported.

The thinkMOTION project is divided into ten work packages, which depend partly from each other.

Work package 4 (abbreviation: WP4) with the title “Digitising heterogeneous input content” has the aim to digitise heterogeneous content sources using different special work places in a high quality and quantity. This paper is an intermediate report about the work in WP4, which was done in the second project year. The results from the first project year are published in the deliverable D4.1.

2 Workflow and dependencies on other WPs

There are two main tasks in WP4, which have to be done:

- Task 4.1: Digitisation of heterogeneous input content
- Task 4.2: Summary of the collected experiences

In the first project year, the focus of WP4 was on digitising of paper based documents such as books, journals, journal articles, patents, theses, proceedings, technical drawings, drawings from printed documents and teaching material.

These works were continued in the second project year. An additional point in the second year was the digitisation of images and slides of mechanisms and the digitisation of physical demonstration models in motion.

These physical demonstration models are normally used for the education in lectures. In addition to a still image, the movement of the mechanism is important for the understanding of its function. For digitising physical demonstration models in motion, the project partner IUT has extensive experiences and has supported the other partners.

Within the project, the work in WP4 depends on other work packages and tasks. The proven project workflow (see Figure 1), which was used in the first project year, was used unchanged in the second year. According to this workflow, WP4 receives input from WP3 (Locating and providing relevant sources and clarification of rights of use). In WP3 the located and selected content will be registered in the DMG-Lib database, the Intellectual Property Rights (IPR) for these input contents will be clarified and necessary rights of use will be obtained. The work in WP3 (see deliverables D3.1 und 3.2) provides the selected content for the digitising process in WP4. And WP4 delivers input for

WP5 (Processing of digitised content and integration into DMG-Lib) as described in deliverables D5.1 and D5.2.

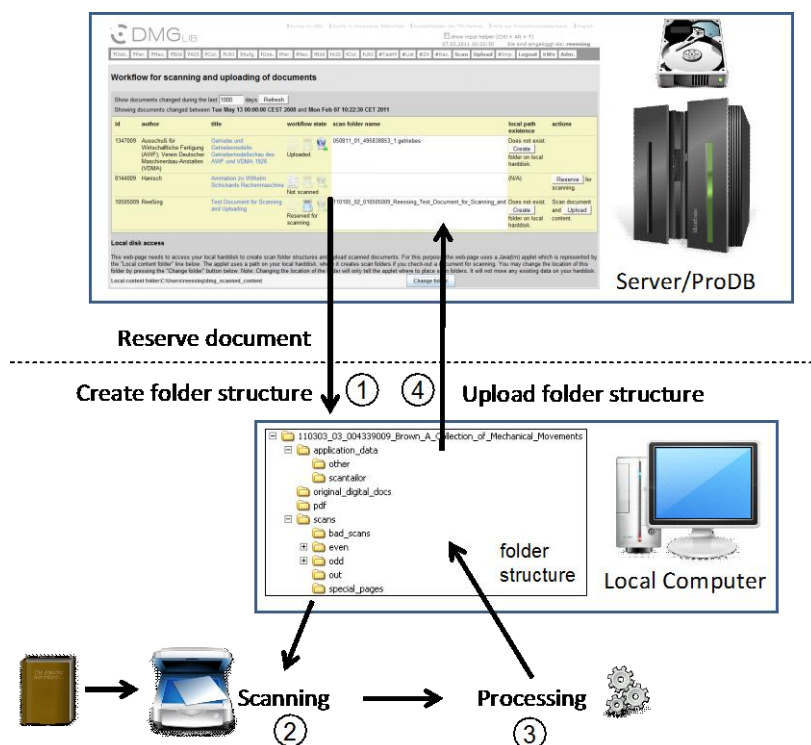


Figure 1. General workflow for reserving and uploading documents (Figure contains tasks of WP4 and WP5)

The regional digitisation service centers, which were established at all thinkMOTION partner institutions in the first project year, are working now very efficiently. Although at some project partners the staff for digitising was changed at the beginning of the second year, the new staff could be trained by using the training materials and the tutorials, which were produced in the first project year. These experiences are published in the deliverables D4.1 and can be used by other European digitisation projects (Task 4.2).

Each project partner is digitising in the own house with the own staff. Additionally, some of the partners give limited digitisation tasks to external companies or institutions by using of subcontracting. Reasons are e.g. the availability of old valuable books in the own house and problems to ship these or to less digitising resources in the own house.

Scanner hardware and software used in the thinkMOTION project

The scanners for literature, which were used in the first project year, are also used in the second year. One of the used scanners *Plustek OpticBook A300* got a total technical fault at the end of the second year. A further self-made V-shape book scanner based on digital photo cameras was completed and started its work successfully at IFMA, the French thinkMOTION partner, (see Figure 3). The partner UBC from Spain bought a new scanner Epson GT 20000.

Different types of digitising equipment are used in the thinkMOTION project. Figure 2 to Figure 8 give an overview over this equipment.



Figure 2. Self-made V-shape book scanner “tm-books”, developed by the Romanian partner UPT



Figure 3. Self-made V-shape book scanner based on double EOS600D (18MPixel each) developed by the French partner IFMA. CAD-Model and prototype



Figure 4. Various Zeutschel book scanners with a special book-cradle



Figure 5. Low-cost book scanner Plustek OpticBook A300 with a special book-edge



Figure 6. Digitising equipment for slides



Figure 7. Various flatbed book scanners for journal articles or loose leaf collections



Figure 8. Digitisation workstations for physical demonstration models at German partner IUT

For each scanner hardware an own scanner software is delivered or offered. Usually these software tools work probably and provide the main functionality of the hardware. Most of the project partners use those software tools or alternatively, picture viewer or desktop publishing software with a batch scanning function via TWAIN interface. (Details see deliverable D4.1). For the self-made scanners, the software that was delivered with the used cameras, was modified and adapted for the purpose in the project.

Correlation of digitising in WP4 and processing in WP5

Figure 9 to Figure 11 illustrate the output of the digitisation process in WP4 and the necessity of processing and quality improvement steps in WP5. Digitisation of literature in WP4 delivers only raw images of the documents. The detailed workflow and the parameters are described in the deliverable D4.1. These raw images are normally double pages, with a black border and are often a little distorted (Figure 9). These images have to be processed and quality improved in WP5.

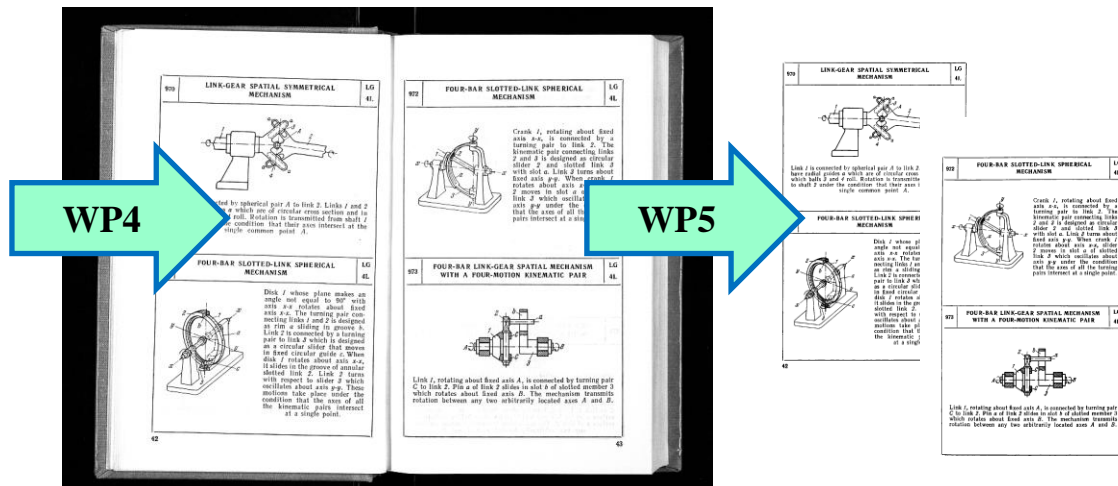


Figure 9. WP4 output of digitising process for documents

The digitisation of slides in WP4 delivers only raw images of the slides, often with a broad black border and often a little distorted (Figure 10). These raw images have to be processed and quality improved in WP5.

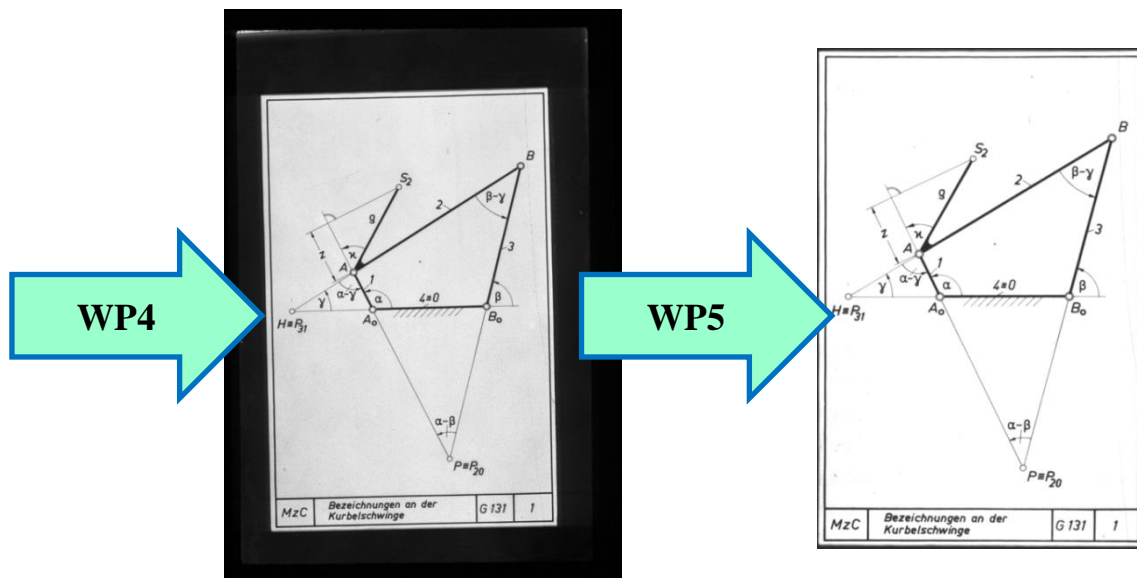


Figure 10. WP4 output of digitising process for slides

The physical demonstration models, are digitised as an image sequence step by step with a photo camera. For this process, the models are driven by a stepper motor and after each step an image is taken. The output of this digitisation process in WP4 delivers only raw images of the physical demonstration models (Figure 11), which have to be processed and quality improved in WP5. Normally, 400 images per turn for each model are taken. From each image sequence a video file, an

interactive animation and still images of representative positions of the mechanism are generated in WP5.

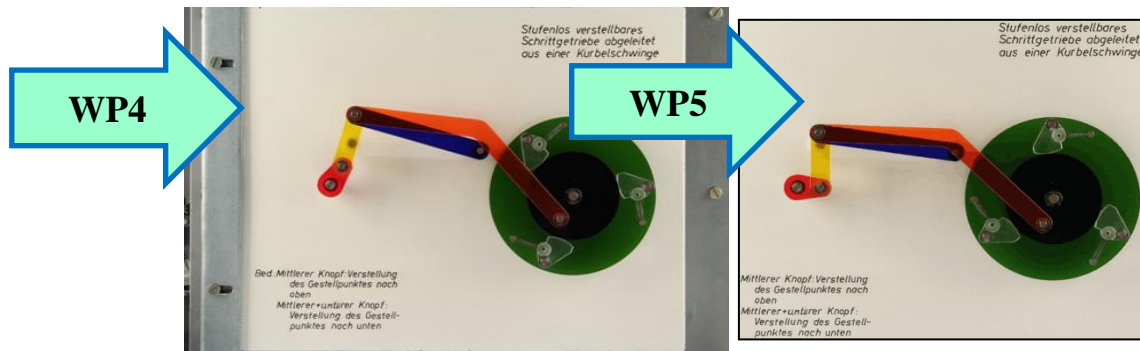


Figure 11. WP4 output of digitising process for physical demonstration models

3 Results

In the first project year, regional digitisation service centres were established in all partner countries. It was necessary to perform training courses with the staff to practice the workflow, the handling with the database and to point to possible sources of errors and problems. In preparation to these training days a lot of training materials and a 50-page tutorial was written, which explain all necessary steps of the workflow for digitising of paper- based documents in detail. This tutorial summarises the experiences in a catalogue of rules, which can be used by other European digitisation projects (see Annex I of deliverable D4.1).

All project partners have appropriate scanning devices. Two partners have built their one book scanner. The others have bought scanning devices or share devices with their libraries. With these equipment, with the proven workflow and with the well trained staff all thinkMOTION partners work together as a European digitisation network for this project and in the future for further European digitisation projects.

By now, the following content were digitised and will be made accessible via the Europeana online portal step by step during the last project year:

- Nearly 12,000 books, journal articles, teaching materials with approximately 300,000 pages
- About 7,500 images digitised from photos; slides, physical demonstration models, machines, devices, portraits, technical drawings or drawings from books, journals, articles etc.)
- Approximately 190,000 single images taken as image sequences from mechanism and machine models in motion.

For content which already exists in digital form (e.g. proceedings of conferences of the last few years, which are available as PDF or MS Word file format, digital photo files or video files) the workflow steps in WP4 can be skipped. Such type of content will be processed only in WP5 – “Processing of digitised content and integration into DMG-Lib”.

4 Outlook for the next project year

In the third project year, the number of digitised content will be increased to reach the overall project aim of 62,000 Europeana items. The scanning equipment and the digitisation workflow will be used in the proven way of the first two years. Additionally, the scanning equipment will be enhanced by enlarging the scanning area of one of the self-made scanners for scanning larger drawings.