DELIVERABLE

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

D8.3 – Translation of metadata items

Revision: 2.0

Authors:

Jean-Christophe FAUROUX (IFMA)

Luc LETOFFE (IFMA)

Revision History

Revision	Date	Author	Organization	Description
1.0	27.05.2013	L. Letoffe	IFMA	First Draft
2.0	31.05.2013	J.-C. Fauroux	IFMA	Review, Second Draft

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

The thinkMOTION project[1] has the ambition to make the DMG-Lib portal[2] the major digital library of machines and mechanisms at the world level. As an international database containing a lot of textual content, thinkMOTION integrates a work package in charge of multilingual aspects. Although most researchers speak scientific English, it is important that the content remains perfectly understandable to a large audience, particularly the non-specialists, also concerned by this project.

2   Objectives

The targeted languages were defined to be English, German, French, Italian, Spanish and Romanian. Objectives are at two levels:

- Translating the portal and internal pages
- Translating the contents

Four tasks were defined in the work package WP8:

- 8.1 Improvement of translation workflow
- 8.2 Translation of webpages
- 8.3 Translation of metadata
- 8.4 Maintenance of a thesaurus

3   Main results on Task 8.3

3.1 Statistics

During the thinkMOTION project, Task 8.1 allowed to settle that the addressed contents for translation were mainly the titles and keywords. The summaries and longer descriptive contents could be optionally translated but the resources for this translation work are not available in this project. All the partners should generate metadata both in their native language and in English, which serves as a common language. It means that, at the end of the thinkMOTION project, all the contents exist in English and at least one supplemental language (Deliverable D8.1). Future translations in other languages can be considered for subsequent projects.

The translation process has been integrated since the beginning of the project. A monitoring system has been integrated at the end of Year 1 and is accessible to every partner from the startpage of ProDB (Figure 1). Statistics are available, such as "Absolute numbers of texts in certain languages" (Figure 2), "Absolute numbers of texts in certain languages where an English text is defined too" (Figure 3) and "Relative number (in %) of texts in certain languages where an English text is defined too" (Figure 4). These results were updated May 31st, 2013.

---

[1] thinkMOTION Project website, http://www.thinkmotion.eu
[2] DMG-Lib Portal, http://www.dmg-lib.org

Figure 1. Access to the statistic tool for translations

Key	de	en	fr	it	pt	ru	es	sh	zh	ro
d_AdvImgSeq.descrSID	1	888	0	0	0	0	161	0	0	0
d_AdvImgSeq.nameSID	1651	1054	0	0	0	0	161	0	0	0
d_DmgURI.comment	0	0	0	0	0	0	0	0	0	0
d_DmgURI.name	0	0	0	0	0	0	0	0	0	0
d_GeneralFBCElem.adiTxt	201	201	0	0	0	0	0	0	0	0
d_GeneralFBCElem.descrSID	1392	3267	360	158	0	0	420	0	0	0
d_GeneralFBCElem.nameSID	1826	3698	360	158	0	0	420	0	0	0
Image description	3906	23282	2387	6608	0	0	6800	0	0	0
Image name	11721	25523	5470	6622	1436	1437	8230	1436	1436	0
d_Model.application	511	7	1	0	0	0	1	0	0	0
d_Model.functionDesc	1832	103	2	1	0	0	1	0	0	0
d_Model.kinChainDesc	1507	61	1	0	0	0	0	0	0	0
d_Model.mechanismType	644	0	1	0	0	0	0	0	0	0
d_Model.publicComment	698	23	0	1	0	0	0	0	0	0
Person short descr.	758	1356	260	517	0	0	164	0	0	0
Person long descr.	312	1150	252	529	0	0	165	0	0	0
d_ThConceptInfo.data	10390	10284	10284	0	0	10287	0	0	0	0
VitaInfo activity	4674	5235	2646	2314	0	0	1011	0	0	0
Document title	13645	9720	1116	1053	4	82	2940	11	2	6694
Model title	3001	311	4	0	0	0	2	0	0	28
Keyword	5	260	1	10	0	0	0	0	0	4127
Language of Work (*)	14125	8158	713	875	4	186	3285	15	3	6639

Figure 2. Statistics on "Absolute numbers of texts in certain languages" in ProDB

Key	de	en	fr	it	pt	ru	es	sh	zh	ro
d_AdvImgSeq.descrSID	1	888	0	0	0	0	161	0	0	0
d_AdvImgSeq.nameSID	1000	1054	0	0	0	0	161	0	0	0
d_DmgURI.comment	0	0	0	0	0	0	0	0	0	0
d_DmgURI.name	0	0	0	0	0	0	0	0	0	0
d_GeneralFBCElem.adiTxt	201	201	0	0	0	0	0	0	0	0
d_GeneralFBCElem.descrSID	1385	3267	359	158	0	0	419	0	0	0
d_GeneralFBCElem.nameSID	1819	3698	359	158	0	0	416	0	0	0
Image description	3338	23282	1827	2719	0	0	6657	0	0	0
Image name	6697	25523	4458	2074	1436	1437	8040	1436	1436	0
d_Model.application	3	7	0	0	0	0	1	0	0	0
d_Model.functionDesc	81	103	0	1	0	0	1	0	0	0
d_Model.kinChainDesc	32	61	0	0	0	0	0	0	0	0
d_Model.mechanismType	0	0	0	0	0	0	0	0	0	0
d_Model.publicComment	23	23	0	1	0	0	0	0	0	0
Person short descr.	233	1356	257	509	0	0	164	0	0	0
Person long descr.	142	1150	250	520	0	0	165	0	0	0
d_ThConceptInfo.data	10038	10284	10137	0	0	10115	0	0	0	0
VitaInfo activity	886	5235	545	2274	0	0	1011	0	0	0
Document title	959	9716	505	182	0	17	4	0	0	1485
Model title	272	311	0	0	0	0	2	0	0	28
Keyword	0	259	0	3	0	0	0	0	0	1
Language of Work (*)	0	0	0	0	0	0	0	0	0	0

Figure 3. Statistics on "Absolute numbers of texts in certain languages where an English text is defined too" in ProDB

Key	de	en	fr	it	pt	ru	es	sh	zh	ro
d_AdvImgSeq.descrSID	100	100	-	-	-	-	100	-	-	-
d_AdvImgSeq.nameSID	60	100	-	-	-	-	100	-	-	-
d_DmgURI.comment	-	-	-	-	-	-	-	-	-	-
d_DmgURI.name	-	-	-	-	-	-	-	-	-	-
d_GeneralFBCElem.adiTxt	100	100	-	-	-	-	-	-	-	-
d_GeneralFBCElem.descrSID	99	100	99	100	-	-	99	-	-	-
d_GeneralFBCElem.nameSID	99	100	99	100	-	-	99	-	-	-
Image description	85	100	76	41	-	-	97	-	-	-
Image name	57	100	81	31	100	100	97	100	100	-
d_Model.application	0	100	0	-	-	-	100	-	-	-
d_Model.functionDesc	4	100	0	100	-	-	100	-	-	-
d_Model.kinChainDesc	2	100	0	-	-	-	-	-	-	-
d_Model.mechanismType	0	-	0	-	-	-	-	-	-	-
d_Model.publicComment	3	100	-	100	-	-	-	-	-	-
Person short descr.	30	100	98	98	-	-	100	-	-	-
Person long descr.	45	100	99	98	-	-	100	-	-	-
d_ThConceptInfo.data	96	100	98	-	-	98	-	-	-	-
VitaInfo activity	18	100	20	98	-	-	100	-	-	-
Document title	7	100	45	17	0	20	0	0	0	22
Model title	9	100	0	-	-	-	100	-	-	100
Keyword	0	100	0	30	-	-	-	-	-	0
Language of Work (*)	0	0	0	0	0	0	0	0	0	0

Figure 4. Statistics on "Relative number (in %) of texts in certain languages where an English text is defined too" in ProDB

3.2 Special case of CAD pictures

Some types of items, such as CAD pictures and videos, can be also automatically entitled and translated. An example of pictures can be seen in Figure 5. Four pictures were generated by CAD, corresponding to four views of a mechanism created by J. Dupond, respectively Front / Left / Top / ISO views. Let's consider the ISO view. Its filename has been constructed according to naming conventions that describe its content:

Example: image file name: CFAO_LD_Iso_P0_jdupond.png

Where

- CFAO means Computer Aided Design
- LD means low definition (some pictures can exist in various definitions)
- Iso means ISO view but could be also another keyword such as: Front, Left, Right, Back, Top, Bottom, 4views, Iso, Expl (exploded view) or Di (detailed view number i)
- P0 means position 0 of the mechanism (i.e. initial position)
- jdupond is the login of the creator
- png is the file format

Following these naming conventions, the corresponding title to the ISO view should be:

- **In English:** ISO perspective view of the initial position of the CAD model made by J. Dupond.
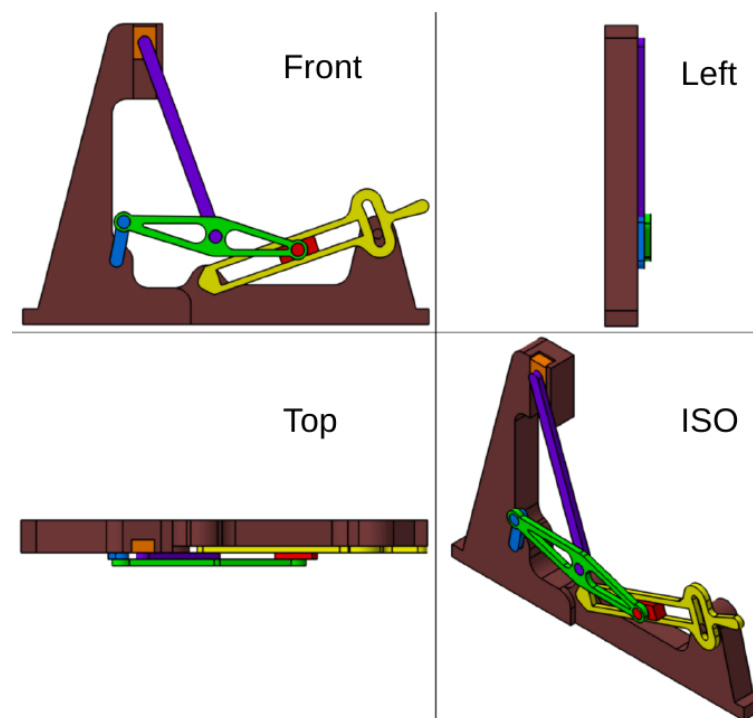- **In French:** Vue en perspective ISO et position initiale du modèle CAO créé par J. Dupond.

Figure 5. A sample of mechanism with four attached CAD pictures, each one depending on the angle of view

Thanks to automatic macros created by Ilmenau and IFMA, the tedious part of translation could be avoided and automatized for the particular case of CAD pictures. This process was also partially extended to the case of real machines, where each machine has also been photographed along several views.

4   Conclusion

The translation process was an important challenge in the thinkMOTION project. After having improved the translation workflow (Deliverable D8.1), having translated the webpages, both of the portal DMG-Lib and the internal database ProDB (Deliverable D8.2), the contents themselves (focused on tiles and keywords only), initially provided in the native language of the team, were translated systematically in English.

Thanks to this work, thinkMOTION can now provide to Europeana big quantities of high quality data relative to the science of machines, and that are accessible, indexable and understandable easily by anyone in Europe and in the rest of the world.