A reproducible open platform for a collective process

Translating CIDOC CRM 7.1.2 in French

GUILLEM Anais, KRUMMEICH Raphaëlle, DAVID Bertrand, VAN RUIMBEKE Muriel
Useful links

Gitlab Link of the project:
https://gitlab.huma-num.fr/bdavid/doc-fr-cidoc-crm

Gitlab README of the translation project:

Gitlab wiki of the translation project:
https://gitlab.huma-num.fr/bdavid/doc-fr-cidoc-crm/-/wikis/home

Translation: you can find here the translation as it is continuously integrated
https://cidoc-crm-fr.mom.fr/

Zotero bibliography (shared Zotero bibliography on CIDOC CRM, ontologies, LOD etc):
https://www.zotero.org/groups/4641239/cidoc_crm_bibliographie
People and Institutions


+/- 15 researchers from different cultural and research institutions (CNRS, MOM, universities etc)

- Institutional support: MASA Consortium
- Technical support:
  - Huma-Num (Gitlab hosting);
  - Maison de l’Orient et de la Méditerranée (MOM) (translation site web)
Objectives: reproducible open platform and a collective process

- Full translation of the CIDOC CRM 7.1.2/3 (future ISO version), including introductive texts, figures, tables etc
- Project of translation as:
  - Findable
  - Accessible
  - Interoperable and Open
  - fully Reproducible and Reusable
- Building a CIDOC CRM user community among Digital Humanity research institutions
  - Translation as learning experience
  - Translation as collaboration project
  - Translation as a Digital Humanities research project
Method and workflow

Friday 3-5PM collective validation sessions

Collective synchronous work
- CIDOC translation issues
- Group’s methods & other issues

Individual asynchronous work
- Translations (non git or git users)
- Technical issues (git users)

Git push
CI process

https://cidoc-crm-fr.mom.fr/
Automated update of a static webpage with the actual translation
Continuous Integration Process

.gitlab-ci.yml:

2 Stages:

- build -> run a **pandoc** docker container to generate html from markdown files (Entities, Properties), generate and concat index and menus
- deploy -> adjust CSS settings, copy files on web-server
Gitlab- Huma-Num/Doc fr CIDOC CRM

files/texts to translate
project management
project info
technical elements
Gitlab issues and boards

Open general issues

To be translated
Translation ongoing
Translated, waiting for validation
Under review/validation
Once validated, issue closed

1 issue = 1 element to translate (ie. text of a class or a property)
Example issue #55: translation of E25

link to the file in repo

tags

current status

translation notes added as comments
Milestones’ construction: logical groups

Sources: class hierarchy & property-class hierarchy tables of Cidoc CRM reference Manual

Logical groups hierarchy for translation flow

Focus on the content of a logical group
Logical groups as “time-weighted” milestones

Source: XML (interchange format) file of Cidoc CRM reference Manual

- use of XML file to “weigh” classes and properties
  - XML automated parse of scopenote
  - Count words in scopenote
- group “weighted” concepts in logical groups
Translation flow labels inside logical groups

Source: GitLab Wiki link, ©Google follow-up spreadsheet from shared drive

- evaluate subgroups flows to set validation agenda
  - priority for asynchronous translation
  - creation of new issues
- quantify & qualify sessions needed to set validation of logical groups
  - provisional agenda
  - thematic sessions
Focus on logical groups and translation workflow
Focus on logical group E72 and validation workflow

Open issues: ongoing translation, or waiting for review

Closed issues: validated translations

Link to the calculated logical group graph md file in repo

Percentage of completion
Conclusion: a mean to build a community of CRM users

- CIDOC CRM ontology aims at facilitating Linked Open Data workflows. The translation in French is a mean to facilitate its intake in French speaking communities, not an end.

- The translation project is an opportunity for:
  - building a community of CIDOC CRM users, building on existing expertise and projects
  - open discussions on data integration among cultural heritage experts (from different institutions) through translation questions
  - share an understanding on CIDOC CRM ontology and (best) practices

- The translation project as experimental space in DH:
  - experimenting on open source tools, methods, workflows
  - (re)inventing how to do a translation: collective, peer-review, hybrid, open