CIDOC CRM Translation Best Practices Guide

Version 1.0
Authors: Karine Léonard Brouillet, Philippe Michon
Reviewers: George Bruseker, Pat Riva
Intended audience: CIDOC CRM SIG
Created date: 2021-07-21
Last update: 2021-09-13

Abstract: Guide to best practices for small, medium, and large CIDOC CRM Translation Initiatives to establish efficient and consistent work processes.

Introduction

In order to make the documentation surrounding CIDOC CRM more accessible, the CRM SIG wishes to support the translation of its documentation into different languages. As such, a series of documents [ADD THE LINK TO THE CIDOC CRM WEBSITE] has been developed to support any person or group wishing to start or update a translation.

This document aims to support a translation initiative in setting up an infrastructure to simplify the translation process by reducing as much as possible

the iterations required to finalize translated sections, classes or properties while ensuring a quality translation.

The processes presented in this document can be adapted according to the specific constraints of the translation initiative (e.g. human resources, time, tools, etc.). However, some processes and documents presented in this guide could be used to officialize a translation with the CRM SIG. Information on the translation selection process, as well as the most relevant versions to translate, can be found in the Governance Guidelines [ADD THE LINK TO THE GOVERNANCE GUIDELINES].

Mandate

An initiative mandate should be written and transmitted to the CRM SIG when starting a translation. This is an important step in establishing the seriousness of the translation, especially when the CRM SIG cannot otherwise assess the quality of the translation due to language barriers.

The mandate should be written in the language of the translation team, but an English abstract is recommended as a way to describe the usefulness of the translation to the CIDOC CRM community at large.

A mandate should address the following elements provided they are relevant to the translation initiative:

- ⇻ Introduction
    - ⇻ Contextual information
    - ⇻ Purpose of the translation
    - ⇻ Targeted community
- ⇻ Translation plan
    - ⇻ Units subject to translation (e.g. introduction, classes and properties labels, definitions, supplementary materials)
    - ⇻ License chosen by the initiative (e.g. Creative Commons, copyrighted)
    - ⇻ Updating process, if there is one (e.g. latest version, latest ISO version)
- ⇻ Members and their responsibilities
    - ⇻ Working documents management
    - ⇻ Coordination with the CRM SIG
    - ⇻ Internal review process and governance if relevant
- ⇻ Participation process
    - ⇻ How to join the initiative
    - ⇻ Selection process
    - ⇻ Selection criteria
- ⇻ Workflows and tools

Several of these elements are addressed below.

Expertise and Skills

Although a translation can be done by one person working alone, the CRM SIG recommends carrying out the project with others in order to ensure an effective translation and revision process thanks to the varied expertise of team members. It should also be noted that a complete translation of CIDOC CRM is a long-term project and distributing the task often turns out to be wise. CIDOC CRM is an ontology that takes advantage of a wide variety of fields of expertise, each with its own specific terminology. As such, involving experts from different backgrounds will contribute favorably to the identification of the most suitable terms.

Here is an alphabetical non-exhaustive list of fields of expertise that might be relevant when translating CIDOC CRM:

- Archival Sciences
- Art History
- Computer Sciences
- Digital Humanities
- Ethnology
- History
- Library Sciences
- Life Sciences
- Linguistics
- Mathematics
- Museology
- Ontology
- Philosophy
- Terminology
- Translation

Expertise from such disparate domains may be called on at different points in the translation because of the subject matter and scope of the CIDOC CRM itself. As an intellectual tool aiming to provide a lingua franca between disparate domain specialists in cultural heritage, computer scientists, semantic modellers and machines, the standard covers a wide range of material and basic level concepts from different domains. Obviously, the majority of initiatives will not be in a position to bring together experts from all of these disciplines. Nonetheless, as much as possible, it makes sense to diversify the profiles of those involved in the translation in order to cover the broadest spectrum of relevant knowledge possible.

Translation Units

CIDOC CRM contains a lot of content over several hundred pages, especially when taking the official extensions into account. However, it is possible to segment the documentation into different units which can be translated individually. By identifying the needs to be filled and the existing translations, the translation

group will be able to determine which parts of which documents to translate first based on the needs of the group.

Firstly, the working group must determine which document(s) will be translated. If the main CIDOC CRM documentation (CRMbase) has not been translated into the target language yet, it is recommended to start with this document since all the extensions refer to it. If CRMbase has been translated, the translation group can look into translating a particular extension according to the needs of their target community or project.

Which text(s) to translate will depend on the translation team's community needs and what is the target for eventual semantic modelling and transformation. Elements that might be taken into consideration when determining this include:

- The selected extensions' domain should cover that of the dataset(s) or project(s) typical of the community;
- Actual data or projects might cover domains different from the primary subject typically associated with the community and should thus be taken into consideration as well (e.g. an archaeological project might need CRMarcheo but maybe CRMsci or CRMtex for its inscriptions as well);
- Information not yet covered by the community's typical projects but relevant to its members should be considered as well.

When approaching the translation of a particular CIDOC CRM document, it is possible to segment the task into different sections that make sense independently of the others. Here are three main approaches that require different levels of involvement:

- Full Translation
    - Translating the whole document requires the most work. Such a translation includes the information that is necessary to generate the RDF file, and thus will allow the users to export the data in the translated language, but also includes the introductory sections of the model document which contextualize it and can be used to foster understanding for beginner users. The end result should be a full text along with a list of stable labels and definitions (scope notes) that can be submitted to the official CIDOC CRM RDF file.
- Classes and Properties Translation
    - The body of the declared ontology is made up of a set of classes and properties. Translating only the declarations of the classes and properties, can be undertaken independently of the contextual text. Such a translation includes the information that is necessary to generate the RDF file, and thus will allow the users to approach the model properly in their own language. The end result should be a list of stable labels and definitions (scope notes) that can be submitted to the official CIDOC CRM RDF file.

➤ Labels Translation
    ➤ The body of the declared ontology is made up of a set of classes and properties identified at the first level by labels for ease of use / reference. The labels are the least expressive unit of information in CIDOC CRM, but nevertheless give a first access to the end user of the ontology in an interface. Translating the labels of classes and properties enables them to be added to the official RDF file and thus represented in a user interface (UI).

Style Guide

Before starting to translate, the translation initiative should, as a matter of good practice, put in place a reference document that will ensure consistency in the translation. This document will be the most effective means of demonstrating that the proposed translation is a quality one. The document allows an overview of the translation process which can be used to assess the overall approach to the translation conceptually with the SIG.

This document should be constantly updated during the translation process and the sections, classes or properties already translated should be reviewed to ensure relevant new rules are applied. It is important that all members adhere to the principles of this style guide and agree to abide by them, or formally propose changes so that the end result remains consistent.

The structure of the document can obviously vary according to the specific needs of each language, but the CRM SIG encourages translators to think about the following aspects:

➤ Writing Preferences
    ➤ Each language contains a variety of terms and wordings that can provide the same information. In order to ensure consistency of content and simplify reading of the document, the CRM SIG recommends identifying writing preferences. By way of example, it may be relevant to determine the verb tenses to prioritize, the use of capital letters, references to titles of works, quotes, translations of recurring phrases, spelling preferences, etc.
    ➤ The CRM SIG recommends respecting and taking into account local language sensitivities and best practices so that all readers are adequately represented by the translation (this can include, for example, adapting text so that it is formal, inclusive, or any other characteristic deemed relevant by the translation initiative).
    ➤ The CRM SIG recommends the use of a language level that is clear and understandable for the target audience of the translation. Linguistic variants of a language (e.g. dialects or regionalisms) can justify translations for local communities, for example, though it is also possible to offer a translation in an international style. This scope should be mentioned in the translation initiative's mandate.

- ➡ Reference Works (dictionaries, translation tools, etc.)
    - ➡ Especially when translating as part of a team, agreeing on primary reference material will ensure consistent understanding throughout the team and a more efficient revision process.
    - ➡ The CRM SIG recommends contacting the CRM SIG at the following address [email address] for any question concerning the intended meaning of a term or section of the CIDOC CRM.
- ➡ Translation of Recurring Terms and Sentences
    - ➡ Many formulations are recurrent in the CIDOC CRM documentation, the CRM SIG therefore recommends translating all of them before starting the translation. This list should not be taken as definitive, but can be useful to quickly identify all occurrences if the recommended translation must be changed.
    - ➡ Class and property labels are a specific case of recurring terms and their translation should be part of the first translation steps.
    - ➡ A list of classes and properties as well as recurring terms is available here [ADD LINK TO THE DOCUMENT].

Workflows

The CIDOC CRM as an ontological formulation of a diverse domain of information is a dense document and its translation will therefore generally have to be carried out over a period of time. As such, it is strongly recommended that the translation team sets up workflows that will ensure efficient translation processes and communication. The objective of this section is not to propose technological tools to implement these processes, but to identify such processes so that the translation group devises an efficient work system. It is important that the operating process is explained to members and that they formally adhere to its principles.

Technological tools are proposed in the document titled "Tools and Interchange Protocol(s)" [ADD THE LINK TO THE DOCUMENT].

Translation Status Workflow

If the translation group does not translate only the labels, it is preferable to divide the document into several parts to be translated: introductory parts can be divided into sections and classes and properties can be treated individually.

The translation process is normally divided into three stages:
- ➡ Translation: conversion of the contents into the target language.
- ➡ Revision: comparative review of the original and translated contents, as well as discussion and modification when applicable.
- ➡ Validation: approval of the translated contents by the initiative.

When several people are involved in the translation effort, it is essential to be able to quickly identify who is working on what and when another person needs to take over.

A calendar can be set up in order to reserve time slots dedicated to a particular activity (e.g. during this period the translated entities are under review). If doing so, generally speaking, revision takes twice as long as translation and validation do, so it is wise to plan accordingly (e.g. one week for translation, two weeks for revisions, one week for validation).

Translation Order

If the working group translates an entire document, the CRM SIG recommends starting with the introductory parts. This helps in:

- Having a general understanding of the model;
- Identifying the key concepts that will be part of the style guide.

Then, since the classes and properties of a semantic model are strongly interrelated, it is preferable to follow a translation order that differs from the numerical order used by CIDOC CRM, which is purely historical rather than semantic. Several combinations are possible, but the CRM SIG recommends following two main rules:
- Translate a class only when its super classes have been translated;
- Translate properties whose domain is an already translated class before translating the next subclass.

An example of a typical translation order is available here [ADD THE LINK TO THE DOCUMENT]. If you are working on a project that necessitates the prioritization of certain elements over others, you can also order your work accordingly (e.g. this order [ADD THE LINK TO THE DOCUMENT] based on an actors-focused project).

Communication Workflow

If the translation effort is carried out as a team, the implementation of a communication protocol can prove useful, especially if the people do not work within the same organization or in the same physical spaces. Internal communications must be taken into account, but external communications, particularly with the CRM SIG and other translation initiatives, if applicable, are also important.

Within the team itself, it is important to have an easy way of asking questions and discussing issues, but it is also necessary to plan regular meetings in order to:

- Review the issues that have arisen following the new translations;
- Make decisions about new stylistic choices;
- Validate the translated classes and properties.

If the translation committee has several participants or the number of participants is an even number, it may be useful to elect a decision-making committee who will decide when consensus cannot be achieved.

With regards to exchanges with the CRM SIG, it is important to keep the person in charge of translations informed, in particular concerning the following subjects:

- Questions on the intended meaning of terms, definitions, or units of text;
- Identification of errors in the original version;
- Upload or update of the official translated version on the CIDOC CRM website;
- Work in progress status.

Finally, it may be relevant to set up a communication flow with partners likely to contribute to the translation work. This could include an organization that uses CIDOC CRM, but also other translation initiatives. Indeed, it is possible that two translations in the same language are in progress at the same time, and it can be useful to liaise to ensure that the translation efforts are aligned.

Summary of Steps to Follow

1. Determine the scope of the translation (i.e. complete, class and property definitions, or labels).
2. Determine the workflow, tools, and license your team wishes to use and agree upon this, and decide which format the translation will take and how it will be made public.
3. Write your mandate and transmit to the CRM SIG so that it is aware that you might need support as part of your translation.
4. Select a list of reference materials the team can agree on (e.g. X dictionary will mostly be used, etc.). At this stage, if previous translations are available, assessing their quality and accessibility is recommended and the previous translation might become an important reference material.
5. Review and examine the list of recurring terms and phrases in the source material.
6. Based on the previous elements, establish a first version of your style guide.
7. Translate the labels of classes and properties as well as recurring terms and phrases first so that if changes need to be done to a unit, there is always a single one to search for and replace. Record these in your style guide.
8. Identify the translation order most meaningful to your initiative.
9. Translate and review per your workflows.

Contact Us

If you have any questions regarding this document or if you require special assistance, please do not hesitate to contact (TBD)