

# Translation of CIDOC CRM and XML integration

**FORTH**

E. Tzortzakakis

**European French Translation Initiative**

R. Krummeich, B. David, A. Guillem, M. Van Ruymbeke

## Related to

- **Issue 564 : Tools and formats relevant for the translation work (technical means of the exchange)**
- **Issue 528 : Guidelines and Protocols for Translating CIDOC CRM**

57th SIG Meeting, 9-13 october 2023, issue  
Tuesday October 10th, 2023

# CIDOC CRM SIG issues

## [528 & 564 short introduction to the XML translation interchange format](#)

**Issue 528: Guidelines and Protocols for Translating CIDOC CRM**

<https://cidoc-crm.org/Issue/ID-528-guidelines-and-protocols-for-translating-cidoc-crm>

**Issue 564: Tools and formats relevant for the translation work (technical means of the exchange)**

<https://cidoc-crm.org/Issue/ID-564-tools-and-formats-relevant-for-the-translation-work-technical-means-of-the-exchange>



# XML interchange format

XML file that is sufficient to represent the complete contents of each stable CIDOC-CRM Stable version. Images are handled as external links.

Goals:

1. Ability to provide online the most up-to-date translation for each new CIDOC-CRM stable version

## E22 Human-Made Object ([show all properties](#))

Version: 7.1.2

Also Defined in versions: [7.1.1](#), [7.1](#), [7.0.1](#), [7.0](#), [6.2.9](#), [6.2.8](#), [6.2.7](#), [6.2.6](#), [6.2.5](#), [6.2.4](#), [6.2.3](#), [6.2.2](#), [6.2.1](#), [6.2](#), [4.2.5a](#), [4.2.5](#), [4.2.4](#), [4.2](#), [4.1](#), [4.0](#)

Not Defined in versions: -

Class Name:

Human-Made Object

*unaltered since version 6.2.7 / October 2019*

*previous value in version [6.2.6](#)*

SubClass Of:

[E19](#) Physical Object *unaltered since version 4.0 / March 2004*

[E24](#) Physical Human-Made Thing *unaltered since version 4.0 / March 2004*

Navigate to a section

en

en | de

en | el

**en | fr**

en | pt

en | ru

en | zh

Man-Made Object

*English text for version: 6.2.1 / October 2015*

Objet fabriqué

*'French' translation created on June 2021 for version: 6.2.1 / October 2015*

**English text changed - translation needs updating. [View English text of version: 6.2.1](#)**

[E19](#) Objet matériel

[E24](#) Chose matérielle fabriquée



# XML interchange format

## 2. Minimise the effort for translation of a newer version based on a previous one (may require some alignment effort)

No assumptions are made as to the **7.2** re of this algorithm. "Without loss of meaning" signifies that designers and users of the system are satisfied that the data representation corresponds to the semantic definitions provided by this International Standard.

**5.0.4**

1.4 Terminology

Remove Equal

Toggle Differences Highlight

### 1.4 Terminology

The following definitions of key terminology used in this document are provided both as an aid to readers unfamiliar with object-oriented modelling terminology, and to specify the precise usage of terms that are sometimes applied inconsistently across the object-oriented modelling community for the purpose of this document. Where applicable, the editors have tried to consistently use terminology that is compatible with that of the Resource Description Framework (RDF), [3] a recommendation of the World Wide Web Consortium. The editors have tried to find a language, which is comprehensible to the non-computer expert and precise enough for the computer expert so that both understand the intended meaning.

**class** A class is a category of items that share one or more common traits serving as criteria to identify the items belonging to the class. These **properties** need not be explicitly formulated in logical terms, but may be described

### 5.1 Terminology

The following definitions of key terminology used in this document are provided both as an aid to readers unfamiliar with object-oriented modelling terminology, and to specify the precise usage of terms that are sometimes applied inconsistently across the object oriented modelling community for the purpose of this document. Where applicable, the editors have tried to consistently use terminology that is compatible with that of the Resource Description Framework (RDF) [3], a recommendation of the World Wide Web Consortium. The editors have tried to find a language which is comprehensible to the non-computer expert and precise enough for the computer expert so that both understand the intended meaning.

**Class** A class is a category of items that share one or more common traits serving as criteria to identify the items belonging to the class. These **properties** need not be explicitly formulated in logical terms, but may be described

7.1.1		6.2.9	
<a href="#">Introduction</a>	<a href="#">comparable</a>	<a href="#">Introduction</a>	<a href="#">backwards comparable</a>
<a href="#">Objectives of the CIDOC CRM</a>	<a href="#">comparable (reworked)</a>	<a href="#">Objectives of the CIDOC CRM</a>	<a href="#">comparable</a>
<a href="#">Scope of the CIDOC CRM</a>	<a href="#">comparable (reworked)</a>	<a href="#">Scope of the CIDOC CRM</a>	<a href="#">comparable</a>
<a href="#">Compatibility with the CIDOC CRM</a>	<a href="#">comparable (identical)</a>	<a href="#">Terminology</a>	<a href="#">comparable</a>
<a href="#">Terminology</a>	<a href="#">comparable, moved. "symetry" added. "reflexivity" added</a>	<a href="#">Compatibility with the CIDOC CRM (the following text is taken from ISO21127:2014)</a>	<a href="#">"subproper added." "tra new and co</a>
<a href="#">Applied Form</a>	<a href="#">comparable, moved</a>	<a href="#">Property Quantifiers</a>	<a href="#">comparable</a>
<a href="#">Naming Conventions</a>	<a href="#">comparable, moved</a>	<a href="#">Naming Conventions</a>	<a href="#">comparable</a>
<a href="#">Inheritance and Transitivity</a>	<a href="#">comparable to "transitivity". title change. moved</a>	<a href="#">About the logical expressions of the CIDOC CRM</a>	<a href="#">new and co</a>
<a href="#">Shortcuts</a>	<a href="#">comparable</a>	<a href="#">Modelling principles</a>	<a href="#">comparable</a>

sum of  
class may  
properties  
properties

# XML schema Key-notes

- Images are included as links e.g.  
[https://cidoc-crm.org/html/version\\_images\\_full\\_text/images\\_7.1.2/image\\_04.png](https://cidoc-crm.org/html/version_images_full_text/images_7.1.2/image_04.png)
- The sources of the images for each version are available: <https://cidoc-crm.org/figures>
- Classes and properties declarations follow the exact same format as with the XML provided for declarations in the Encodings column of each stable version  
<https://cidoc-crm.org/versions-of-the-cidoc-crm>
- All other model documentation sections (apart from the classes and properties declarations) are compliant with the following schema where each section is split in title/contents/footnotes parts





```
<!--
```

```
section elements include child elements that require translation (title, contents, footnotes/note).
```

```
All attributes of section elements are just used for the representation of the XML in a format resembling the official CIDOC-CRM documentation.
```

```
@siblingSortOrder determines the relative order of the element among the elements with the same parentDescriptiveId. It is set to be optional as it may be preferable to be removed in order to facilitate comparison.
```

```
-->
```

```
<!ELEMENT section ((title*, contents*, footnotes?))>
```

```
<!ATTLIST section
```

```
    descriptiveId CDATA #REQUIRED
```

```
    parentDescriptiveId CDATA #REQUIRED
```

```
    siblingSortOrder CDATA #IMPLIED
```

```
    isHeaderSection (True | False) #REQUIRED
```

```
    isAutoText (True | False) #REQUIRED
```

```
>
```



# Envisioned workflow

- 1/Create and download an XML translation template by specifying
  - a Stable Cidoc-CRM version of interest e.g. 7.1.2
  - a language code tag for the translation e.g. el
  - a language name for the translation e.g. Ελληνικά
  - Optionally: upload the translated template file of a former version

Download template file for translation

Select version of XML template for translation:

Specify value of xml:lang tag:

Language description:

[DOWNLOAD template XML](#)



# Envisioned workflow

- 2/ Fill in the missing translation placeholders of the downloaded XML

Manual; semi-automated; automated

```
<contents xml:lang="el"/>
</section>
<section descriptiveId="[auto] table of contents" parentDescriptiveId=
  <title xml:lang="en">Table of Contents</title>
  <title xml:lang="el">Πίνακας περιεχομένων</title>
</section>
<section descriptiveId="introduction" parentDescriptiveId="" isAutoTex
  <title xml:lang="en">Introduction</title>
  <title xml:lang="el">Εισαγωγή</title>
  <contents xml:lang="en">&lt;p&gt;This document is the formal definit
  <contents xml:lang="el">&lt;p&gt;Αυτό το έγγραφο είναι ο επίσημος ορ
</section>
<section descriptiveId="objectives of the cidoc crm" parentDescriptive
  <title xml:lang="en">Objectives of the CIDOC CRM</title>
  <title xml:lang="el"/>
  <contents xml:lang="en">&lt;p&gt;The primary role of the CIDOC CRM i
  <contents xml:lang="el"/>
</section>
```





# Envisioned workflow

- 3/ Upload back the XML file with the translations and download a generated HTML page HTML page that provides side-by-side comparison of the English version vs the Translated version

## Table of Contents

Automatically generated content

## Πίνακας περιεχομένων

### Introduction

This document is the formal definition of the CIDOC Conceptual Reference Model ("CIDOC CRM"), a formal ontology intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information and similar information from other domains, as further detailed below. The CRM is the culmination of more than two decades of standards development work by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM). Work on the CRM itself began in 1996 under the auspices of the ICOM-CIDOC Documentation Standards Working Group. Since 2000, development of the CRM has been officially delegated by ICOM-CIDOC to the CIDOC CRM Special Interest Group (SIG). The SIG, in turn, collaborates with the ISO working group ISO/TC46/SC4/WG9 to bring the CRM to the form and status of an International Standard. This set of collaborations has resulted in the production of ISO21127:2004 and ISO21127:2014, the ISO standard editions of the CIDOC CRM. This collaboration will be continued in order to support the the next update of the ISO standard edition. The present document belongs to the series of evolving versions of the formal definition of the CRM, which serve the ISO working group as community draft for the standard. Eventual minor differences, in semantics and notation, of the ISO standard text from the present, community CIDOC CRM version, which the ISO working group requires and implements, will be harmonized in the subsequent versions of the present, community CIDOC CRM formal definition document.

### Objectives of the CIDOC CRM

The primary role of the CIDOC CRM is to enable the exchange and integration of information from heterogeneous sources for the reconstruction and interpretation of the past at a human scale, based on all kinds of material evidence, including texts, audio-visual material and oral tradition. It starts from, but is not limited to, the needs of museum documentation and research based on museum holdings. It aims at providing the semantic definitions and clarifications needed to transform disparate, localised information

Navigate to a documentation section

[X]

en en | el

### Εισαγωγή

Αυτό το έγγραφο είναι ο επίσημος ορισμός του CIDOC Conceptual Reference Model ("CIDOC CRM"), μιας επίσημης οντολογίας που προορίζεται να διευκολύνει την ενσωμάτωση, τη διαμεσολάβηση και την ανταλλαγή ετερογενών πληροφοριών πολιτιστικής κληρονομιάς και παρόμοιων πληροφοριών από άλλους τομείς, όπως περιγράφεται περαιτέρω παρακάτω. Το CRM είναι το αποκορύφωμα των εργασιών ανάπτυξης προτύπων για περισσότερες από δύο δεκαετίες από τη Διεθνή Επιτροπή Τεκμηρίωσης (CIDOC) του Διεθνούς Συμβουλίου Μουσείων (ICOM). Οι εργασίες για το ίδιο το CRM ξεκίνησαν το 1996 υπό την αιγίδα της ομάδας εργασίας ICOM-CIDOC Documentation Standards Working Group. Από το 2000, η ανάπτυξη του CRM έχει ανατεθεί επίσημα από το ICOM-CIDOC στην Ομάδα Ειδικού Ενδιαφέροντος CIDOC CRM (SIG). Το SIG, με τη σειρά του, συνεργάζεται με την ομάδα εργασίας ISO ISO/TC46/SC4/WG9 για να φέρει το CRM στη μορφή και το καθεστώς ενός Διεθνούς Προτύπου. Αυτό το σύνολο συνεργασιών είχε ως αποτέλεσμα την παραγωγή των ISO21127:2004 και ISO21127:2014, των προτύπων ISO εκδόσεων του CIDOC CRM. Αυτή η συνεργασία θα συνεχιστεί προκειμένου να υποστηριχθεί η επόμενη ενημέρωση της τυπικής έκδοσης ISO. Το παρόν έγγραφο ανήκει στη σειρά εξελισσόμενων εκδόσεων του επίσημου ορισμού του CRM, που εξυπηρετούν την ομάδα εργασίας ISO ως κοινοτικό προσχέδιο για το πρότυπο. Τυχόν μικρές διαφορές, σε σημασιολογία και σημειογραφία, του προτύπου κειμένου ISO από την παρούσα κοινοτική έκδοση CIDOC CRM, την οποία απαιτεί και υλοποιεί η ομάδα εργασίας ISO, θα εναρμονιστούν στις επόμενες εκδόσεις του παρόντος, κοινοτικού επίσημου εγγράφου ορισμού CIDOC CRM.

Missing 'title' translation in 'Greek' language

Missing 'HTML content' translation in 'Greek' language

# Useful Links

Gitlab issues:

<https://gitlab.huma-num.fr/gt-cidoc-crm/gt-traduction-cidoc-crm-fr/doc-fr-cidoc-crm/-/issues/289>

<https://gitlab.huma-num.fr/gt-cidoc-crm/gt-traduction-cidoc-crm-fr/doc-fr-cidoc-crm/-/issues/285>

Readme:

<https://gitlab.huma-num.fr/gt-cidoc-crm/gt-traduction-cidoc-crm-fr/doc-fr-cidoc-crm/-/blob/translate/xml/Readme.md>

[https://gitlab.huma-num.fr/gt-cidoc-crm/gt-traduction-cidoc-crm-fr/doc-fr-cidoc-crm/-/blob/translate/stats/issues\\_flow\\_metadata/Readme.md](https://gitlab.huma-num.fr/gt-cidoc-crm/gt-traduction-cidoc-crm-fr/doc-fr-cidoc-crm/-/blob/translate/stats/issues_flow_metadata/Readme.md)

FORTH translations Website:

[https://cidoc-crm.org/html/cidoc\\_crm\\_v7.1.2\\_with\\_translations.html#E53](https://cidoc-crm.org/html/cidoc_crm_v7.1.2_with_translations.html#E53)

# Problem addressed and Objectives

- Automatic integration of the markdown files of the Fr-Fr translation into the XML file for publication on CIDOC CRM website
- Facilitation of the navigation in the CIDOC CRM documentation (versions, languages etc)
- Integration of the translation information (metadata & paratext)

# Integration Workflow for XML

- **Integration for HTML publication**

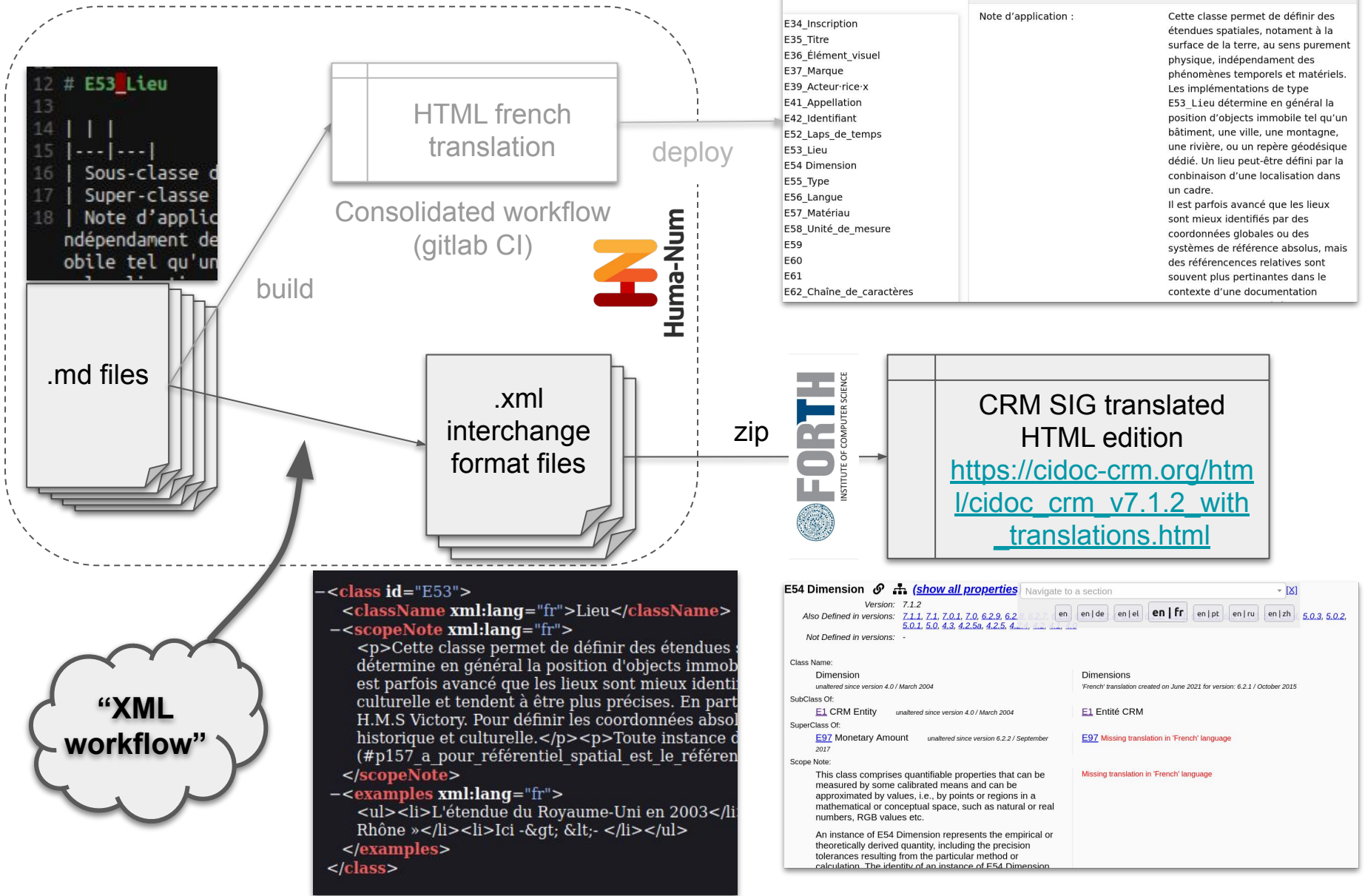
We have previously set up a Continuous Integration (YAML) workflow to build the HTML webpage of the french translation from markdown working files.

- **New : translated text and metadata integration for XML**

We define some script to do similar tasks to convert markdown to XML.

- Step 1/ extract text from .md files and convert in XML (issue #289)
- Step 2/ extract paratext & metadata from issues' documentation (status, issue #285, title, labels, translation notes etc.)

# Workflow for XML schema



```

12 # E53_Lieu
13
14 | | |
15 |---|---|
16 | Sous-classe de
17 | Super-classe
18 | Note d'applic
    ndépendant de
    obile tel qu'un
  
```

HTML french translation

Consolidated workflow (gitlab CI)



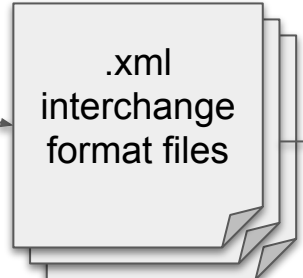
- E34\_Inscription
- E35\_Titre
- E36\_Élément\_visuel
- E37\_Marque
- E39\_Acteur-riche-x
- E41\_Appellation
- E42\_Identifiant
- E52\_Laps\_de\_temps
- E53\_Lieu
- E54 Dimension
- E55\_Type
- E56\_Langue
- E57\_Matériau
- E58\_Unité\_de\_mesure
- E59
- E60
- E61
- E62\_Chaine\_de\_caractères

## E53\_Lieu

Sous-classe de	E1_Entité_CRM
Super-classe de	
Note d'application :	Cette classe permet de définir des étendues spatiales, notamment à la surface de la terre, au sens purement physique, indépendamment des phénomènes temporels et matériels. Les implémentations de type E53_Lieu déterminent en général la position d'objets immobile tel qu'un bâtiment, une ville, une montagne, une rivière, ou un repère géodésique dédié. Un lieu peut-être défini par la combinaison d'une localisation dans un cadre. Il est parfois avancé que les lieux sont mieux identifiés par des coordonnées globales ou des systèmes de référence absolus, mais des références relatives sont souvent plus pertinentes dans le contexte d'une documentation



build



zip



CRM SIG translated HTML edition  
[https://cidoc-crm.org/html/cidoc\\_crm\\_v7.1.2\\_with\\_translations.html](https://cidoc-crm.org/html/cidoc_crm_v7.1.2_with_translations.html)



```

-<class id="E53">
  <className xml:lang="fr">Lieu</className>
  -<scopeNote xml:lang="fr">
    <p>Cette classe permet de définir des étendues spatiales, notamment à la surface de la terre, au sens purement physique, indépendamment des phénomènes temporels et matériels. Les implémentations de type E53_Lieu déterminent en général la position d'objets immobile tel qu'un bâtiment, une ville, une montagne, une rivière, ou un repère géodésique dédié. Un lieu peut-être défini par la combinaison d'une localisation dans un cadre. Il est parfois avancé que les lieux sont mieux identifiés par des coordonnées globales ou des systèmes de référence absolus, mais des références relatives sont souvent plus pertinentes dans le contexte d'une documentation</p></scopeNote>
  -<examples xml:lang="fr">
    <ul><li>L'étendue du Royaume-Uni en 2003</li><li>Rhône</li><li>Ici -&gt; &lt;-</li></ul>
  </examples>
</class>
  
```

E54 Dimension [\(show all properties\)](#)

Version: 7.1.2  
 Also Defined in versions: 7.1.1, 7.1, 7.0.1, 7.0, 6.2.9, 6.2.8, 6.2.7, 6.2.6, 6.2.5, 6.2.4, 6.2.3, 6.2.2, 6.2.1, 6.2, 6.1, 6.0, 5.0.1, 5.0, 4.3, 4.2.5a, 4.2.5, 4.2.4, 4.2.3, 4.2.2, 4.2.1, 4.2, 4.1, 4.0, 3.0, 3.0.1, 3.0, 2.0, 2.0.1, 2.0, 1.0, 1.0.1, 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.0, 0.0.1, 0.0.0

Not Defined in versions: -

Class Name: Dimension unaltered since version 4.0 / March 2004	Dimensions "French" translation created on June 2021 for version: 6.2.1 / October 2015
SubClass Of: E1 CRM Entity unaltered since version 4.0 / March 2004	E1 Entité CRM
SuperClass Of: E97 Monetary Amount unaltered since version 6.2.2 / September 2017	E97 Missing translation in 'French' language
Scope Note: This class comprises quantifiable properties that can be measured by some calibrated means and can be approximated by values, i.e., by points or regions in a mathematical or conceptual space, such as natural or real numbers, RGB values etc.	Missing translation in 'French' language
An instance of E54 Dimension represents the empirical or theoretically derived quantity, including the precision tolerances resulting from the particular method or calculation. The identity of an instance of E54 Dimension	



# Starting point : .md file with markdown syntax

<https://gitlab.huma-num.fr/-/ide/project/gt-cidoc-crm/gt-traduction-cidoc-crm-fr/doc-fr-cidoc-crm/edit/translate/-/entities/e53.md>

english

```

1 > # E53 Place
2
3 > | | |
4 > | --- | --- |
5 > | Subclass of: | `E1 CRM Entity`
6 > | Superclass of: |
7 > | Scope note: | This class comprises extents in space, in particular on the surface of the earth, in the pure sense of physics: independent f
rom temporal phenomena and matter.<br>The instances of `E53 Place` are usually determined by reference to the position of "immobile" objects su
ch as buildings, cities, mountains, rivers, or dedicated geodetic marks, but may also be determined by reference to mobile objects. A Place ca
n be determined by combining a frame of reference and a location with respect to this frame.<br>It is sometimes argued that instances of `E53 Pl
ace` are best identified by global coordinates or absolute reference systems. However, relative references are often more relevant in the conte
xt of cultural documentation and tend to be more precise. In particular, we are often interested in position in relation to large, mobile objec
ts, such as ships. For example, the Place at which Nelson died is known with reference to a large mobile object - H.M.S Victory. A resolution o
f this Place in terms of absolute coordinates would require knowledge of the movements of the vessel and the precise time of death, either of w
hich may be revised, and the result would lack historical and cultural relevance.<br>Any instance of `E18 Physical Thing` can serve as a frame
of reference for an instance of `E53 Place`. This may be documented using the property `P157 is at rest relative to (provides reference space f
or)`.
```

français

# Step 1 : transformation flow from md files to XML

- Identification of sed REGEX to collect text and create XML tags
  - for classes and properties
  - for texts
- Integration of the transformation in the consolidated workflow (gitlab CI)
  - Shell script with sed operations
  - Zip file building
- Dialog with FORTH routine for CRM SIG HTML semi-automated integration

# Step 1 : `sed` REGEX substitutions

## Examples with CIDOC Classes

Labels	Scope note
--------	------------

# E53\_Lieu

markdown syntax

| Note d'application : | Cette classe permet de définir des étendues spatiales, notamment à la surface de la terre, au sens purement physique, indépendamment des phénomènes temporels et matériels.<br>Les implémentations de...

```
/^#/s/_/ /g;
```

sed rules

# E53 Lieu

```
s/^# \([E0-9]+\)$/<class id="\1"/>;
s/^# \([E0-9]+\) \(.*\)/<class id="\1">\n\t:
<className xml:lang="fr">\2</className>;
```

```
s/^| Note.d['']application[ ]*: *|[ ]*\(.*\) *|/\t<scopeNote
xml:lang="fr">\&lt;p\&gt;\1\&lt;/p\&gt;</scopeNote>/g;
```

```
1 <class id="E53">
2   <className xml:lang="fr">Lieu</className>
3   <scopeNote xml:lang="fr">&lt;p&gt;Cette classe permet d
s.&lt;/p&gt;&lt;p&gt;Les implémentations de type `E53_Lieu`
u peut-être défini par la combinaison d'une localisation da
bsolus, mais des références relatives sont souvent plus
positions d'objets très grand, en mouvement tel qu'un navir
faudrait connaître le mouvement du navire et l'heure préci
instance de [E18_Chose_matériel](#e18_chose_matériel) p
iel_spatial_de) &lt;/p&gt;</scopeNote>
4   <examples xml:lang="fr">&lt;ul&gt;&lt;li&gt;L'étendue d
'expression : « poissons pêchés à trois miles au nord du co
5 </class>
```

# Step 1 : REGEX operations in the XML workflow

Example of sed instructions (.sed) for properties items operated with bash script command (.sh) in XML workflow

```

5 s/^# \([P[0-9]\|+\)\$/<class id="\1"/>;
6 s/^# \([P[0-9]\|+\) \(.*\) (\(.*\))/<class id="\1">\n\t<directName xml:lang="fr">\2</directName>\n\t<inverseName xml:lang="fr">\3</inverseName>;
7 s/^# \([P[0-9]\|+\) \(.*\)\$/<class id="\1">\n\t<directName xml:lang="fr">\2</directName>;
8 /^| *| */d;
9 /^[ -]*|[ -]*|/d;
10 /\.omaine./d;
11 /^[ ]*Supe./d;
12 /^[ ]*Sou./d;
13 /^| Port./d;
14 #/^| Super-pro.*|[ ]*\(.*\) *|<superPropertyOf id="\1"/>/g;
15 s/\.Quantification.*|[ ]*\(.*\) *[]*/<quantification xml:lang="fr">\&lt;p\&gt;\1\&lt;\/p\&gt;</quantification>/g;
16 s/^[ ]+\+Note.d[']application[ ]*: *|[ ]*\(.*\) *[]*/<scopeNote xml:lang="fr">\&lt;p\&gt;\1\&lt;\/p\&gt;</scopeNote>/g;
17 /scopeNote/s/<br>/\&lt;\/p\&gt;\&lt;p\&gt;/g;
18 s/^[ ]+\+Exemples[ ]*: [ ]*\(.*\)[]*/<examples xml:lang="fr">\&lt;ul\&gt;\&lt;li\&gt;\1\&lt;\/li\&gt;\&lt;\/ul\&gt;</examples>/g;
19 /examples xml/s/<br>/\&lt;\/li\&gt;\&lt;li\&gt;/g;
20 /^[ ]+\+Logique du premier ordre./d;
21 /^[ ]+\+Propriétés./d;
22 /^[ ]*$/d;
23 s|//g;
  
```



## Step 2 : Translation information & metadata

For each item or subitem, the transformation to XML may benefit **information on translation process**.

A new tag or attribute **“status”** may support information using Gitlab API :

- “iid” issue number (#163 for P20, see infra) to get access to **translation notes**
- “state” of the issue (open or close) respective to the **state of the translation adoption flow**
- “title” of issue to set values to proper item CIDOC label
- “labels” relative to the **translated version** for example (CIDOC CRM version, translation steps etc.)



# Step 2 : `curl` with GitLab API (json)



GitLab



```
curl --header "PRIVATE-TOKEN: abc-123-def-456"  
"https://gitlab.huma-num.fr/api/v4/projects/380/issues/285" | jq
```

```
curl --header "PRIVATE-TOKEN: abc-123-def-456" "https://gitlab.huma-num.fr/api/v4/projects/380/issues/285"  
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current  
           Dload  Upload   Total      Spent    Left  Speed  
100 2010    100 2010     0     0   3976      0  --:--:--  --:--:--  --:--:--  3980  
{  
  "id": 10307,  
  "iid": 285,  
  "project_id": 380,  
  "title": "Construction version XML",  
  "description": "Peut-on inclure l'état des tickets ? pour savoir si l'entité, la propriété est en co  
  "state": "opened",  
  "created_at": "2023-05-19T14:21:45.464Z",  
  "updated_at": "2023-05-26T18:07:27.655Z",  
  "closed_at": null,  
  "closed_by": null,  
  "labels": [  
    "Techniques en coulisses"  
  ],  
  "milestone": null,  
  "assignees": [],  
  ...  
}
```

# Step 2 : Information scrapped from gitlab

Access to part of the “whole” story of the translation

issue id, state, id class (or property) name (inversename), gitlab labels

```
curl -s --header "PRIVATE-TOKEN:$TOKEN_READ_API" "$REPO_URL_API?per_page=100"
| jq '.[]\
| "\(.iid):\(.state):\(.title):\(.labels)"'\
| grep [EP][[:digit:]]\
| sed 's/...
```

```
doc-fr-cidoc-crm/-/issues/163;opened;P20 had specific purpose;3-Traduit en attente de revision,Propriété,v7.1.1
doc-fr-cidoc-crm/-/issues/162;opened;P19 was intended use of;3-Traduit en attente de revision,v7.1.1
doc-fr-cidoc-crm/-/issues/161;opened;P17 was motivated by (motivated);3-Traduit en attente de revision,Propriété,v7.1.1
doc-fr-cidoc-crm/-/issues/160;opened;P15 was influenced by;3-Traduit en attente de revision,v7.1.1
doc-fr-cidoc-crm/-/issues/159;opened;P14 carried out by;3-Traduit en attente de revision,Propriété,v7.1.1
doc-fr-cidoc-crm/-/issues/157;closed;P57 has number of parts;5-Valide,Propriété,v7.1.1,v7.1.2
doc-fr-cidoc-crm/-/issues/156;closed;P56 bears feature;5-Valide,Propriété,v7.1.1,v7.1.2
doc-fr-cidoc-crm/-/issues/155;closed;P55 has current location;5-Valide,Propriété,v7.1.1,v7.1.2
doc-fr-cidoc-crm/-/issues/154;closed;P54 has current permanent location;5-Valide,Propriété,v7.1.1,v7.1.2
doc-fr-cidoc-crm/-/issues/153;opened;P109 has current of former location;3-Traduit en attente de revision,v7.1.1
doc-fr-cidoc-crm/-/issues/152;closed;P65 shows visual item (is shown by);5-Valide,v7.1.1,v7.1.2
doc-fr-cidoc-crm/-/issues/151;closed;P62 depicts (is depicted by);5-Valide,v7.1.1,v7.1.2
```

translation notes :  
Gitlab issue URL

state :  

- adopted (closed)
- in progress (opened)

labels :  

- steps of translation (3 - Traduit ..., 5 - validé etc.)
  - type of item (entité, propriété, texte...)
- actual CIDOC version under translation (latest has to be considered)



# Conclusions

## Main results

- Semi-automated workflow to push adopted french translation on different platform, notably CIDOC CRM website [Work in Progress]
- Tracking the information from the French initiative translation platform
- Cooperative process that benefits both partners

## Further work

- Building translation metadata and further informations (translation notes e.g.) as attribute or tag instances of XML interchange format

-> *to be discussed/evaluated*

- Towards a fully automated workflow between fr-fr translation initiative WG to FORTH website