



HAL
open science

Recours au CIDOC-CRM pour évaluer l'interopérabilité de données archéologiques de terrain très variées : présentation des premiers résultats des tests effectués par l'Inrap

Christophe Tuffery, Emeline Le Goff, Julie Boudry, Federico Nurra

► To cite this version:

Christophe Tuffery, Emeline Le Goff, Julie Boudry, Federico Nurra. Recours au CIDOC-CRM pour évaluer l'interopérabilité de données archéologiques de terrain très variées : présentation des premiers résultats des tests effectués par l'Inrap. Spatial Analysis and GEomatics 2017, INSA de rouen, Nov 2017, Rouen, France. hal-01649750

HAL Id: hal-01649750

<https://hal.science/hal-01649750v1>

Submitted on 27 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Recours au CIDOC-CRM pour évaluer l'interopérabilité de données archéologiques de terrain très variées : présentation des premiers résultats des tests effectués par l'Inrap

Christophe Tufféry¹, Emeline Le Goff², Julie Boudry³, Federico Nurra⁴

1. Inrap

*Direction Scientifique et Technique
121 rue d'Alésia - CS 20007- 75685 Paris Cedex 14
christophe.tuffery@inrap.fr*

2. Inrap

*37 rue du Bignon - CS 67737
35577 Cesson-Sevigné
emeline.legoff@inrap.fr*

3. Inrap

*28 rue Robert Fulton
51689 Reims Cedex 2
julie.boudry@inrap.fr*

1. Inrap

*Direction Scientifique et Technique
Chargé de médiation scientifique du projet européen NEARCH
121 rue d'Alésia - CS 20007- 75685 Paris Cedex 14
federico.nurra@inrap.fr*

RESUME. Dans le cadre du projet européen ARIADNE, l'Institut National de Recherches Archéologiques Préventives (Inrap) s'est engagé dans un programme de tests de la norme CIDOC-CRM, et de son extension CRMarchaeo. Le modèle CIDOC-CRM intéresse l'Inrap dans la mesure où il permet d'évaluer les capacités d'interopérabilité de ses propres bases de données archéologiques de terrain, que ce soit pour l'échange entre systèmes ou pour la diffusion vers le Web sémantique. Les tests ont porté sur plusieurs données archéologiques de terrain acquises par l'Inrap sur diverses opérations (diagnostics et fouilles) avec des systèmes d'enregistrement numérique de terrain hétérogènes, s'appuyant sur divers Systèmes de Gestion de Bases de Données (SGBD). L'objectif était de faire correspondre les champs des

systèmes utilisés par l'Inrap et ceux du CIDOC-CRMarchaeo. Les premiers résultats des tests sont présentés, de même que l'application 3M (Memory Mapping Manager, développée par l'Institute of Computer Science (ICS)) qui a été utilisée pour réaliser ces tests.

ABSTRACT. Within the framework of the European ARIADNE project, the French National Institute for Preventive Archaeological Research (Inrap) has embarked on a test program of CIDOC-CRM standard and its CRMarchaeo extension. The CIDOC-CRM model is interesting for Inrap as it assesses the interoperability capabilities of its own archaeological field database, whether for inter-system exchange or for semantic Web delivery. The tests involved several archaeological field data acquired by Inrap on various operations (diagnostics on trial trenches and full excavations) with heterogeneous digital recording systems, based on various DataBase Management Systems (DBMS). The objective was to match the fields of the systems used by Inrap with those of CIDOC-CRM. The first results of the tests are presented in this paper as well as the 3M application (Memory Mapping Manager, developed by Institute of Computer Science (ICS)) which was used to realize these tests.

MOTS-CLES : CIDOC-CRM ; archéologie, BDD de terrain ; modélisation ; interopérabilité ; harmonisation

KEYWORDS: CIDOC-CRM, archaeology, field DBMS, modelization, interoperability

1. Rappel général sur la norme CIDOC-CRM et l'extension CIDOC-CRMarchaeo

Le CIDOC-CRM est un modèle conceptuel de référence (ou ontologie) de haut niveau pour décrire les données du domaine culturel. Le numéro de version actuel de la norme CIDOC-CRM est la version 5.0.4 publié comme norme ISO 21127:2014. Plusieurs extensions du CIDOC-CRM existent. Elles élargissent le domaine du CIDOC-CRM dans des domaines spécialisés¹.

Parmi ces extensions, CRMarchaeo est une extension spécialisée du domaine de l'archéologie, développée depuis 2013 par l'Institute of Computer Science de la Fondation Forth (Fondation pour la Recherche et la Technologie) en Grèce (Doerr M. et al. 2017). CRMarchaeo est une ontologie qui permet de désigner les entités d'une opération archéologique (prospection, diagnostic, fouille, etc.) et leurs relations ainsi que la documentation produite et le mobilier archéologique découvert.

L'objectif de cette extension est de fournir les moyens de documenter les fouilles pour permettre de :

- maximiser la capacité d'interprétation pendant ou après une opération ;
- formaliser les objectifs d'une opération par un questionnement scientifique précis ;
- réviser les connaissances après une opération ;
- comparer des opérations précédentes sur un même site ou dans un même territoire ;

¹ <http://www.cidoc-crm.org/collaborations>

– réaliser des études statistiques complètes de divers types.

2. Le travail d'appariement avec le modèle CIDOC-CRM

L'appariement est la mise en correspondance entre les entités d'un modèle de base de données à appairier et une classe du CIDOC-CRM considéré comme le modèle cible et les propriétés associées aux classes.

L'appariement des termes et des concepts est une étape essentielle pour vérifier la compatibilité des modèles de systèmes de description d'entités d'un domaine culturel avec le modèle cible du CIDOC-CRM et de ses extensions. Les principes à suivre pour effectuer ce travail sont décrits par Kondylakis H. et *al.* (2006).

En 2015, dans le cadre du programme de formation Trans National Access du projet européen ARIADNE, l'Inrap s'est engagé dans un premier test d'appariement entre les classes et propriétés de l'extension CIDOC-CRM*archaeo* et les tables de plusieurs systèmes numériques d'acquisition de données archéologiques de terrain ⁽²⁾. Une présentation des résultats de ces tests a eu lieu lors du CAA2016 (Tufféry et *al.*, 2016). D'autres tests d'appariement ont été entrepris au cours de l'année 2017.

Pour effectuer ce travail d'appariement, l'Inrap a utilisé l'application 3M (Memory Mapping Manager), développée par l'Institute of Computer Science (ICS). Cet outil permet d'appairier des données archéologiques avec les classes et propriétés du modèle CIDOC-CRM et des extensions CRM*sci*, CRM*archaeo*, CRM*ba* spécifiques au domaine de l'archéologie, et d'autres extensions comme CRM*geo* pour les données spatiales, CRM*dig* pour les métadonnées de représentations 2D, 3D ou animées.

Concrètement, l'appariement avec 3M exploite un fichier .xml pro provenant des tables de données des systèmes d'enregistrement de terrain étudiés. Le passage des fichiers sources à leurs versions au format .xml s'appuie sur une exportation des tables au format .csv puis au format .xml avec un éditeur XML. Une validation de la syntaxe .xml est effectuée avec un outil en ligne³. Une fois le fichier .xml chargé dans 3M, il est possible d'appairier les champs des tables avec les classes et les propriétés adaptées du CIDOC-CRM et des extensions utilisées.

L'une des premières conclusions des tests est qu'il n'existe pas une description unique possible des entités archéologiques et de leurs relations selon le modèle CIDOC-CRM. Cette diversité tient aussi bien aux choix faits pour désigner les entités archéologiques et les relations entre entités, qu'à la diversité des types d'opérations archéologiques.

² Ce travail a été réalisé avec les systèmes d'enregistrement de terrain ArchoDB, utilisant MS Access, et CADoc, utilisant FileMaker Pro, respectivement développés par N. Holzem et T. Guillemard, tous les deux techniciens de fouille de l'Inrap que nous tenons à remercier.

³ <http://www.xmlvalidation.com>

Ainsi, une unité stratigraphique peut être désignée par la seule classe A8 (Stratigraphic Unit) ou par l'usage combiné des classes A3 (Stratigraphic Interface) et A2 (Stratigraphic Volume Unit), si on souhaite détailler les entités composant une unité stratigraphique. Un fait archéologique peut être décrit de la même façon que des unités stratigraphiques ou en décomposant les unités qui le constituent. Mais un fait peut aussi être décrit en utilisant les classes S20 (Physical Feature) et S10 (Material Substance) provenant du modèle de l'extension CRM_{sci}.

Par ailleurs, les notions temporelles de chronologie absolue des unités stratigraphiques peuvent être décrites dans le modèle CRM_{archaeo} en utilisant diverses classes (E50Date, E63Beginning Of Existence, E52Time-Span, E64End of Existence, E92SpaceTimeVolume). Les notions de chronologie absolue utilisent la classe E49(TimeAppellation). Les composantes spatiales ont été décrites en utilisant des classes propres à la localisation absolue (E53Place et E27Site) ou les propriétés des relations de chronologie relative comme AP11(hasPhysicalrelation) et AP13(hasStratigraphicrelation (is stratigraphic relation of)).

Le modèle CIDOC-CRM et les extensions utiles pour le domaine de l'archéologie semblent donc adaptés aux types de données produites en archéologie préventive. Le fait qu'il peut exister plusieurs appariements pour un ensemble de données indique une marge de liberté dans le choix du mode de description des données provenant de divers systèmes d'enregistrement de terrain, comme cela a déjà été décrit par d'autres auteurs (E. Le Goff et *al.*, 2015, M. Doerr et *al.* 2017).

Les tests d'appariement seront poursuivis en élargissant le nombre et les types de bases de données de terrain testées afin d'étendre la représentativité des résultats des premiers tests et de pouvoir confirmer la robustesse du modèle CIDOC-CRM pour un grand nombre de données archéologiques de l'Inrap sur diverses opérations (diagnostics et fouilles).

Bibliographie

- Doerr M., Felicetti A., Hermon S., Hiebel G., Kritsotaki A., Masur A., May K., Ronzino P., Schmidle W., Theodoridou M., Tsiafaki, D. and others. (2017). *Definition of the CRM_{archaeo} An Extension of CIDOC CRM to support the archaeological excavation process. Proposal for approval by CIDOC CRM-SIG. Version 1.4. February 2016* : http://www.ics.forth.gr/isl/CRMext/CRM_{archaeo}/docs/CRM_{archaeo}_v1.4.pdf
- Kondylakis H., Doerr M., Plexousakis D. (2006) *Mapping Language for Information Integration. Technical Report 385, ICS-FORTH, December 2006* : http://www.cidoc-crm.org/sites/default/files/Mapping_TR385_December06.pdf
- Le Goff E., Marlet O., Rodier X., Curet S., Husi P. (2015). Interoperability of the ArSol (Archives du Sol) database based on the CIDOC-CRM ontology. *Actes du colloque CAA 2014*, Paris, pp. 179-186
- Tufféry C., Felicetti A., Jard P., Holzem N., Guillemard T. (2016). An essay of mapping archaeological land-record systems used by Inrap with CIDOC-CRM and CIDOC-CRM_{archaeo} extension using 3M on-line tool. *Actes du colloque CAA 2016, S11-03*, Oslo