HAL
open science

Harmonisation de l'acquisition des données d'opérations d'archéologie préventive. Retours d'expériences et perspectives à partir de l'application EDArc

Christophe Tuffery, Stéphane Augry

▶ **To cite this version:**

Christophe Tuffery, Stéphane Augry. Harmonisation de l'acquisition des données d'opérations d'archéologie préventive. Retours d'expériences et perspectives à partir de l'application EDArc. Atelier DAHLIA DigitAL Humanities and cuLtural herITAge : data and knowledge management analys, Jan 2019, Metz, France. hal-02472817

HAL Id: hal-02472817

https://hal.science/hal-02472817v1

Submitted on 10 Feb 2020

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.

Harmonisation de l'acquisition des données d'opérations d'archéologie préventive. Retours d'expériences et perspectives à partir de l'application EDArc

Christophe Tufféry* et Stéphane Augry**

* Direction Scientifique et Technique
Inrap - 121 rue d'Alésia - CS 20007 - 75685 Paris Cedex 14
christophe.tuffery@inrap.fr
http://www.inrap.fr
** Inrap, 4, rue du tertre 44470 Carquefou
stephane.augry@inrap.fr
http://www.inrap.fr

Résumé. Les activités de l'Inrap en matière d'acquisition de données archéologiques ont conduit à concevoir et développer depuis 2015 une application qui permet de tendre vers l'harmonisation des pratiques d'enregistrement numérique sur le terrain. Baptisée EDArc, cette application utilise plusieurs normes et thésaurus qui permettent d'inscrire ce projet dans une démarche d'interopérabilité technique et sémantique. Le déploiement de cet outil est en cours depuis trois 2015, et uniquement sur la base du volontariat des utilisateurs. La mise en œuvre d'EDArc et d'une tablette numérique sur une opération de fouille au Mans a permis d'observer l'usage de ces outils numériques auprès d'une équipe d'agents opérationnels, plus ou moins à l'aise avec les outils numériques. Ainsi cette expérience et d'autres permettent de s'interroger sur l'impact de dispositifs numériques sur les pratiques et l'identité professionnelle de leurs utilisateurs et de ne pas se contenter d'une vision trop idyllique de l'utilisation d'outils informatiques.

1. Enregistrement des données archéologiques de terrain, pourquoi choisir EDArc ?

L'Inrap (Institut National de Recherches Archéologiques Préventives), est un établissement public à caractère administratif, sous la double tutelle du Ministère de la Culture et de celui de l'Enseignement Supérieur, de la Recherche et de l'Innovation. En application de la loi sur l'archéologie préventive de 2001 modifiée en 2003 puis en 2016, tout projet d'aménagement du territoire doit donner lieu à une phase préalable de diagnostic du potentiel archéologique qui peut être réalisé par l'Inrap ou les services compétents et habilités de collectivités territoriales. Si le diagnostic est positif, c'est-à-dire qu'il révèle la présence de vestiges archéologiques à l'emplacement du projet d'aménagement, une fouille prescrite par le Service Régional de l'Archéologie, qui dépend de la Direction des Affaires Culturelles peut ensuite avoir lieu, parfois dans le cadre d'une mise en concurrence entre l'Inrap et des opérateurs publics ou privés respectivement agréés (ou habilités) par l'Etat.

Lors de toute opération archéologique, l'ensemble des observations et des découvertes archéologiques sont enregistrées dans des fiches ou carnets papier et, depuis quelques

années, dans des systèmes d'enregistrement de terrain numériques, fonctionnant sur micro-ordinateurs et, plus récemment, sur tablettes.

Actuellement, de nombreux systèmes d'enregistrement sont utilisés par les archéologues de l'Inrap (Tufféry et al. 2017a, 2017b, Badey 2017). Une étude comparative a été réalisée entre ces différents systèmes. A quelques nuances près, ils utilisent tous les mêmes types d'entités puisque l'Archéologie est basée sur des protocoles communs : emprise, tranchée, sondage, unités stratigraphiques, faits archéologiques, ensembles, structures, mobiliers, prélèvements, documentation, etc.

Toutefois, le vocabulaire en usage peut varier d'une équipe à l'autre et chaque chantier peut développer des problématiques propres nécessitant un enregistrement particulier.

Afin de répondre à plusieurs enjeux liés à l'interopérabilité entre ces divers systèmes et aux principes FAIR (Facilement Trouvable, Accessible, Interopérable et Réutilisable) des données produites dès le terrain, la Direction Scientifique et Technique et la Direction des Systèmes d'Information de l'Inrap ont développé depuis 2015 l'application EDArc (Enregistrement de Données Archéologiques),

Cet article, à travers l'expérience d'un chantier emblématique, permet de faire un bilan critique de l'introduction de nouveaux outils informatiques dans la chaine opératoire de l'archéologie préventive.

2. Présentation de l'application

EDArc est une application logicielle fonctionnant en local avec un navigateur Web tel que Google Chrome, une base de données, un ensemble de pages.html, des fichiers JavaScript et CSS. EDArc n'a pas besoin d'une connexion Internet. Le navigateur Web sert à afficher les interfaces utilisateur et à saisir les données de terrain. Utilisé sur des tablettes durcies sous Windows lors des opérations archéologiques de l'Inrap, les données produites avec l'application EDArc sont enregistrées dans une base de données SQLite installée en local sur le poste de l'utilisateur. Cela permet, si besoin, d'afficher et d'interroger les données directement dans un logiciel SIG tel que QGis, en liaison avec des données géoréférencées.

A chaque opération l'archéologue produit une base de données terrain particulière. Néanmoins, il est possible d'embarquer dans EDArc pour une opération de fouille, les données préalablement saisies lors de l'opération de diagnostic, permettant ainsi de disposer dans un même outil de l'ensemble des données d'un site archéologique.

Le modèle conceptuel de données d'EDArc est relativement simple. Il s'appuie sur le modèle générique (Desachy 2008) qui distingue principalement les entités de contexte archéologique (US, faits, structures, etc.), les entités de documentation (photos, minutes, etc.), les entités d'élément mobilier recueilli (mobiliers, prélèvements, etc.).

Le modèle physique de données est constitué d'une série de tables de données (entre 7 et 13 selon les versions et les besoins des utilisateurs) qui comprennent des données générales administratives sur l'opération archéologique et des données brutes et interprétées sur les observations et découvertes archéologiques : Opération, Métadonnées, Unités Stratigraphiques (US), Fait, Ensemble, Tranchée, Sondage, Prélèvement, Mobilier, Minute et Photo, etc. A ces tables s'ajoutent plusieurs tables de jointures qui permettent d'interroger les liens (stratigraphiques et chronologiques) entre certaines tables de données et d'exploiter ainsi les avantages du modèle relationnel.

Plusieurs champs des tables s'appuient sur des listes de valeurs dont certaines sont issues de vocabulaires contrôlés (application Patriarche du Ministère de la Culture) ou de thesaurus

(micro-thésaurus Sujets des thesaurus PACTOLS[1]). En revanche, il n'existe pas de vocabulaires de référence ni *a fortiori* de thesaurus pour toutes les données archéologiques enregistrées dans EDArc. Quant aux métadonnées, elles décrivent la ressource documentaire qu'est la base de données EDArc de l'opération. Ces métadonnées s'appuient sur la norme internationale ISO 15836 dite Dublin Core. Elle comprend 15 descripteurs formels (titre, créateur, éditeur...), thématiques (description, langue...) et relatifs à la propriété intellectuelle (droit, auteur...). Ces descripteurs ont été retenus par l'Inrap pour permettre le versement de ses rapports d'opérations archéologiques auprès du CINES (Centre National Informatique de l'Enseignement Supérieur), qui en assure l'archivage pérenne.

A ce jour, EDArc est le seul système d'enregistrement de terrain à avoir pris en compte la nécessité d'utiliser des normes tels que le format XML pour importer et exporter des données, ce qui le rend directement interopérable et compatible avec de nombreux systèmes d'information respectant ces mêmes formats normés.

De plus, EDArc offre la possibilité d'exporter les données nécessaires pour *Le Stratifiant*, une application utilisant le logiciel Excel. (Desachy 2008). Grâce aux trois fichiers au format .xls que permet d'exporter EDArc vers *Le Stratifiant* (ExportUS, ExportRelations, ExportSynchros), cette application permet ensuite aux utilisateurs de contrôler la cohérence des données et de leurs relations de chronologie relative entre les US et de dessiner le diagramme stratigraphique en tenant compte des phases de regroupement des US associées et des éventuels éléments de datation absolue disponibles. Si le contrôle de cohérence révèle des erreurs dans la saisie des données sur les US, celles-ci sont identifiées précisément dans *Le Stratifiant*. L'utilisateur peut alors revenir dans EDArc, corriger toutes les erreurs révélées avant de réimporter dans *Le Stratifian*t les trois fichiers corrigés et de pouvoir construire le diagramme stratigraphique avec des données de meilleure de qualité.

L'interface d'EDArc et l'usage de la tablette présentent beaucoup de similitude avec les smartphones aujourd'hui dans les poches de nombreuses personnes y compris des plus réfractaires ou des moins habiles avec l'outil informatique. L'application est appréciée pour sa facilité d'utilisation, sa simplicité de mise en œuvre puisqu'elle ne nécessite l'installation d'aucun programme informatique particulier. L'utilisation possible de l'application sur divers types de tablettes sous Windows et prochainement sous Android est aussi appréciée.

La possibilité d'accéder directement dans un logiciel de SIG comme QGis aux données de la base de données SQLite dans laquelle les données d'EDArc sont enregistrées, est souvent souligné par les utilisateurs. Cela évite de devoir procéder par des exports et des imports entre EDArc et ces types de logiciels.

3. L'exemple du chantier des jardins de la cathédrale du Mans

A ce jour, une vingtaine d'opérations archéologiques de l'Inrap ont utilisé EDArc pour l'enregistrement des données de terrain.

Parmi les opérations importantes ayant adopté l'application, l'opération de fouille des jardins de la cathédrale du Mans dans la Sarthe apparaît emblématique à plus d'un titre. Cette fouille a été réalisée, de septembre 2017 à juillet 2018, aux abords du chevet gothique de la cathédrale Saint-Julien, au Mans, à l'occasion de l'aménagement des jardins, programmé par

[1] https://www.frantiq.fr/fr/thesaurus et https://pactols.frantiq.fr/opentheso/index.xhtml

la mairie[2]. La fouille a porté sur une surface de plus de 2 000 m² et concerne une stratigraphie pouvant atteindre plusieurs mètres d'épaisseur. Cette fouille est également caractérisée par les nombreux axes de recherches induits par les découvertes même si ces dernières se focalisent autour du fait urbain, l'étude de la ville, au sens général.

EDArc a été utilisé sur une tablette numérique. L'enregistrement se concentre autour de la notion d'unité stratigraphique (US) dans le sens de plus petite unité d'enregistrement de terrain. Il s'agit d'une notion primordiale au cœur de la démarche de terrain. Ici l'ensemble de l'équipe a procédé à l'enregistrement de ces US. Ce point mérite d'être souligné car le chantier réunissait plusieurs générations d'archéologues aux compétences variés. La possibilité de générer aisément les données au format XML garantit la disponibilité des données dans un formalisme normé, auquel se conforment de plus en plus de logiciels et d'applications utilisés par les agents de l'Inrap tout au long du cycle de vie des données archéologiques.

L'application permet également une veille concernant la complétude des fiches et un contrôle supplémentaire sur la réalisation des différentes étapes de la fouille (prise de vue, levers topographiques en particulier).

Pour la partie étude, aujourd'hui en cours, le lien avec chaque base de données propre aux différents spécialistes est très simple grâce aux possibilités d'export. C'est à souligner, d'autant que leur nombre est relativement important et qu'ils n'appartiennent pas tous forcément à l'Inrap. Le « phasage », (la mise en cohérence chronologique sur laquelle s'appuie le discours archéologique) des données géoréférencées est en cours de réalisation mais l'outil informatique permet d'affronter la masse de données générées. L'étape qui consistait à saisir les fiches papiers est supprimée. Le risque d'erreur de saisie est ainsi réduit et les manipulations en vue du travail de synthèse sont plus facilement accessibles.

Par ailleurs, EDArc permet aux responsables scientifiques d'opérations archéologiques de produire relativement simplement les inventaires scientifiques prévus en fin des rapports d'opérations (inventaires réglementaires de la section 3) tels qu'ils sont définis par l'arrêté du 27 septembre 2004 portant définition des normes de contenu et de présentation des rapports d'opérations archéologiques. Notons ici que la masse de données produites peut rapidement devenir exponentielle. A titre exemple, sur le chantier du Mans les unités stratigraphiques se comptent en milliers. Le mobilier archéologique est très varié et est contenu sur plusieurs mètres linéaires d'étagère. On dénombre 250 relevés manuels sur calque A3 et une quantité innombrable de clichés issus de quatre appareils photos numériques différents.

La réalisation des inventaires peut donc rapidement se relever fastidieuse.

En conclusion sur l'opération du Mans, les gains qualitatifs et quantitatifs ont été notables sans avoir pour autant mis de côté une partie de l'équipe, comme cela peut être le cas lors du déploiement de nouveaux protocoles ou de nouvelles technologies qui modifient parfois en profondeur les pratiques professionnelles. Ainsi une partie de l'enregistrement archéologique se retrouve « dématérialisé » ; les tâches laborieuses de saisie sont été écartées et l'édition des inventaires est grandement automatisée. La qualité de l'enregistrement (complétude et fiabilité des fiches) a également connu une amélioration même si ce volant de l'expérimentation n'a pas été confronté à une évaluation détaillée. Enfin le format numérique se prête bien à l'archivage même si des protocoles nouveaux sont probablement à mettre en place en lien avec une formation minimale des équipes archéologiques. En dernier lieu, il convient d'insister sur les facilités et les possibilités crées entre les données de terrain brutes

[2] https://www.inrap.fr/les-fouilles-des-jardins-de-la-cathedrale-du-mans-13569

et les données géoréférencées ou encore les données issues des différentes études spécialisées.

4. Suites envisagées et perspectives

Depuis 2015, des travaux d'appariement entre EDArc et le CIDOC CRM, norme ISO depuis 2006, ont été réalisés, d'abord dans le cadre du programme européen ARIADNE (Tufféry et al. 2016) puis poursuivis depuis trois ans (Tufféry et al. 2017c, Le Goff 2018). Ces résultats vont pouvoir contribuer prochainement au groupe de travail Web des données du Consortium MASA, soutenu par la Très Grande Infrastructure de Recherche HumaNum. L'intérêt de ces travaux est de faciliter la publication dans le Web Sémantique des données de terrain produites à l'aide d'EDArc, sous la forme de triplets sémantiques au format Resource Description Framework (RDF). Une fois l'appariement effectué, les données pourront être publiées sur le web sémantique depuis un End point SPARQL. Ces données seront alors des ressources documentaires que les machines pourront mieux exploiter à l'aide de langages de requêtes et de plateformes d'interrogation adaptées à l'exploitation des données archéologiques publiées dans le Web des données.

L'une des retombées de la mise en œuvre de l'application EDArc est le fait de démontrer la possibilité d'un ensemble de données minimales à enregistrer sur le terrain qui soit à la fois admissible par l'ensemble des archéologues de l'Inrap et en même temps qui respectent la diversité des pratiques actuelles et des problématiques scientifiques qui peuvent varier d'une opération à l'autre. De ce point de vue, EDArc présente une souplesse de modification et d'adaptation à des opérations, des contextes archéologiques et des types de données spécifiques à prendre en compte.

Il faut aussi souligner que l'introduction sur les opérations archéologiques d'une application comme EDArc s'inscrit, pour l'Inrap et au-delà pour la communauté archéologique, dans le cadre du déploiement de techniques et dispositifs numériques, initié depuis une dizaine d'années avec l'utilisation des premières tablettes numériques. A ce jour une centaine de ces équipements ont été acquis par l'Inrap et utilisés sur plusieurs centaines d'opérations.

EDArc modifie considérablement la répartition traditionnelle du temps consacré à la saisie numérique des données, habituellement réalisée après la phase terrain, à partir des enregistrements réalisés sur supports papier pendant la fouille. Ainsi l'utilisation d'EDArc sur le terrain laisse davantage de temps au traitement des données après la fouille puisque les données sont produites nativement sous forme numérique dès le terrain. Dès lors, les pratiques et les logiques de travail évoluent, de même que la répartition des tâches au sein des équipes ou encore les compétences numériques des agents. Cette évolution multiple se traduit par des situations très différentes selon les agents. Si pour une partie d'entre eux, l'introduction d'outils de travail numériques sur le terrain constitue une opportunité pour faire évoluer leurs pratiques et leurs compétences dans un sens positif, pour d'autres agents cela représente une évolution contrainte, qui modifie en profondeur et de façon très rapide leurs manières habituelles d'exercer leur activité et peut aller jusqu'à perturber leur identité sociale et professionnelle (Dubar 1991).

Depuis le début de l'introduction d'EDArc et de tablettes sur les opérations archéologiques de l'Inrap, nous avons veillé à prendre en compte la diversité des réactions, des capacités, et parfois de difficultés incontestables des agents de l'Inrap face à ce type de dispositif. Ainsi, des actions de formation et de tutorat ont été mises en œuvre. Sans ce genre d'accompagnement, le risque est grand de voir apparaître des fractures entre d'un côté les

agents les plus à l'aise, les primo-adoptants, et de l'autre côté ceux pour lesquels le recours au numérique est synonyme de difficulté, voire de souffrance au travail. Accompagner ainsi l'introduction de dispositifs numériques dans un champ professionnel comme celui de l'archéologie, permet de veiller à ne pas confondre « l'informatique comme pratique et comme croyance », comme cela a déjà été proposé depuis une vingtaine d'années (Gollac M. et al. 2000).

Actuellement plusieurs évolutions d'EDArc sont en cours pour répondre encore mieux aux besoins des utilisateurs et pour tenir compte des retours d'expériences.

EDArc est maintenant présentée dans les actions de formation à l'enregistrement de terrain qui ont lieu à raison de trois sessions par an, soit environ 25 agents formés par an, lesquels peuvent ensuite facilement transférer leur maîtrise de l'application à leurs collègues sur le terrain. Ainsi si les outils numériques entraînent une modification profonde de la nature du travail archéologique, les formations permettent, si ce n'est d'écarter, au moins d'atténuer les aspects négatifs. Des retours d'expérience et des bilans réguliers s'avèrent également nécessaires.

5. Références

Badey S. et Moreau A. (2017), Teaching Archaeology or Teaching Digital Archaeology: Do We Have to Choose ?, in M. Matsumoto, E. Uleberg (eds.). In CAA 2016. Proceedings of the 44th Annual Conference on Computer Applications and Quantitative Methods in Archaeology - Oceans of Data. Oslo, Museum of Cultural History, 30/08-03/09/2017, 10p.

Desachy B. (2008), De la formalisation du traitement des données stratigraphiques en archéologie de terrain. Sciences de l'Homme et Société. Université Panthéon-Sorbonne - Paris I, 2008. Français. https://tel.archives-ouvertes.fr/tel-00406241

Dubar C. (1991), La socialisation, construction des identités sociales et professionnelles, Paris, Armand Collin, Collection U, 256 pages

Le Goff E. et Tufféry C. (2018), Inrap's experience in the digital acquisition of field archaeological data and associated metadata compliant with CIDOC CRM and Dublin Core standards. Conference CIDOC 2018 Heraklion, Crete, Grèce, 2 octobre 2018

Gollac M. et Kramarz F. (2000) L'informatique comme pratique et comme croyance. In: Actes de la recherche en sciences sociales. Vol. 134, septembre 2000. L'informatique au travail. pp. 4-21

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-CRMarchaeo extension using 3M on-line tool. CAA 2016. Proceedings of the 44th Annual Conference on Computer Applications and Quantitative Methods in Archaeology - Oceans of Data. Oslo, Museum of Cultural History, 30/08-03/09/2017, Session #11: Supporting researchers in the use and re-use of archaeological data: continuing the ARIADNE thread

Tufféry C., Le Goff E., Boudry J., Nurra F. (2017a), Using CIDOC-CRM for exchanging field data: the point of view of the National Institute for Preventive Archaeological Research (Inrap), EAA, Session #251, Maastricht, 30/08-01/09/2017

Tufféry C., Le Goff E., Boudry J., Nurra F. (2017b), 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, in Spatial Analysis and GEOmatics SAGEO, Rouen, 6-9 novembre 2017, 4 pages. https://hal.archives-ouvertes.fr/hal-01649750/document

Tufféry C. et Le Goff E. (2017c), Contribution de l'Inrap à l'usage du CIDOC-CRM pour les données archéologiques d'enregistrement de terrain. Participation à la table-ronde des Journées MASA Interopérabilités, 20-22 novembre 2017, Tours : http://masa.hypotheses.org/430

Summary

Inrap's activities in the field of archaeological data acquisition have led to the design and development since 2015 of an application that aims to harmonize digital recording practices in the field. Called EDArc, this application uses several standards and thesauri that allow this project to be part of a technical and semantic interoperability approach. The deployment of this tool has been ongoing since three years 2015, and only on a voluntary basis. The implementation of EDArc and a digital tablet on a search operation in Le Mans made it possible to observe the use of these digital tools with a team of operational agents, more or less at ease with digital tools. Thus, this and other experiences allow us to question the impact of digital devices on the practices and professional identity of their users and not to be satisfied with an overly idyllic vision of the use of computer tools.