



HAL
open science

Artefacts v2 : les enjeux d'une migration

Michel Feugère, Elise Vigier, Louis Charly Eyango, Alyssa Giraudo

► **To cite this version:**

Michel Feugère, Elise Vigier, Louis Charly Eyango, Alyssa Giraudo. Artefacts v2 : les enjeux d'une migration. 2017. <halshs-01541204>

HAL Id: halshs-01541204

<https://shs.hal.science/halshs-01541204v1>

Preprint submitted on 18 Jun 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.



HAL Authorization

M. Feugère*, E. Vigier**, L. Eyango***, A. Giraud****

L'équipe d'Artefacts (Univ-Lyon, UMR 5138 du Cnrs, ArAr - Archéologie et Archéométrie) a entrepris en 2017 un gros chantier destiné à transformer entièrement l'"Encyclopédie collaborative en ligne des objets archéologiques" (www.artefacts.mom.fr) en la maintenant à la pointe des outils numériques au service de l'archéologie. En quoi consiste cette migration et quels sont les enjeux d'une telle opération ?

Artefacts v1 : des fonctionnalités variées mais bridées

La version actuelle d'Artefacts résulte de l'évolution d'un projet dont les origines remontent au début des années 1990, le projet Syslat® conçu et mis en œuvre à Lattes, par M. Py et son équipe, à partir des années 1980 (Feugère 2010, 2011, 2015, 2016). Dans le projet tel qu'il a été conçu à Lattes (version β) et dans ses évolutions ultérieures à Lyon (version 1), plusieurs fichiers et "bibliothèques" de référence sont mis en relation par reconnaissance de "chaînes de caractères". Par exemple, les abréviations bibliographiques sous la forme usuelle "Nom date" sont détectées, ce qui permet d'afficher automatiquement la référence complète sous le texte qui contient l'appel bibliographique. De manière analogue, les musées déjà enregistrés peuvent être repérés dans les lignes d'attestations grâce à leur nom : il est ainsi possible de rechercher, à partir de la liste des

musées toutes les fiches qui mentionnent un musée spécifique.

Cette technique est également employée pour cartographier les attestations : les sites sont repérés par leurs noms, enregistrés avec leurs coordonnées dans la base. Bien qu'elle produise l'effet escompté, la recherche dans la base par reconnaissance de chaîne de caractères comporte plusieurs inconvénients. D'une part, les segments de texte doivent impérativement être correctement orthographiés à la lettre près pour être reconnus : la moindre erreur de saisie empêche l'identification du contenu. D'autre part, la correction d'un élément au sein d'une des bibliothèques (sites, musées ou bibliographie) doit être suivie par une correction manuelle et parfois fastidieuse de toutes les fiches comportant cet élément (Fig. 1).

* Chargé de Recherche au CNRS, Univ-Lyon, UMR 5138 (Arar) <michel.feugere@mom.fr>

** Doctorante, Univ-Lyon, UMR 5138 (Arar) <elise.vigier@univ-lyon2.fr>

*** Analyste-programmeur, Univ-Lyon, UMR 5138 (Arar) <louis.eyango@mom.fr>

**** Masterante, Univ-Lyon, UMR 5138 (Arar) <alyssa.giraud@univ-lyon2.fr>

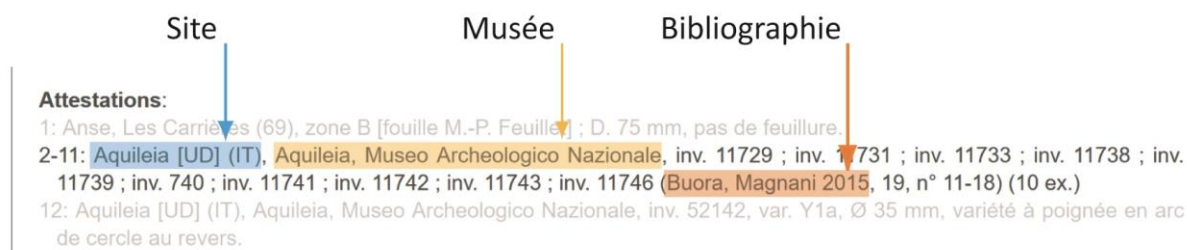


Fig. 1 — Capture d'écran de l'interface v1 actuelle. Les différents éléments sont saisis manuellement ou par copier-coller depuis les bibliothèques de références (sites, musées, bibliographie). Basé sur la reconnaissance de chaînes de caractères, ce système est, d'une part, cause de multiples erreurs de saisie ; d'autre part, les recherches sont lentes et les possibilités de tri limitées.

La première façon de rationaliser le fonctionnement de la base est d'indexer ces différentes données alphanumériques par des identifiants numériques uniques. Cette évolution structurelle importante sera transparente pour l'utilisateur qui verra toujours à l'affichage le contenu en langage naturel. Sans entrer dans les détails techniques, cette modification réduit considérablement le temps de calcul des requêtes et augmente le potentiel de développement futur. Mais l'évolution des pratiques nécessite aussi d'autres transformations.

D'abord caractérisé par la mise à disposition massive de données 'mortes', l'internet

a été une première fois renouvelé par le "web 2.0", grâce auquel les utilisateurs peuvent intervenir eux-mêmes sur les données traitées et affichées en ligne. Cette révolution a démultiplié la production de données, une tendance qui s'accroît aujourd'hui avec les objets connectés et la récolte souvent invisible des données personnelles issues de nos comportements. Une telle masse de données représente une richesse considérable, à condition qu'elle soit accessible : les données structurées seront, à terme, les seules utilisées puisque les autres ne fourniront pas les éléments nécessaires à une exploitation assez fine ou assez rapide.

Artefacts 2.0 : vers une structuration enrichie

La version 2 d'Artefacts doit naturellement résoudre des questions d'ergonomie et d'architecture, les données devant, dans bien des cas, être saisies et enregistrées de manière plus commode, par exemple en utilisant la *complétion automatique* pour les champs renvoyant à une bibliothèque normalisée. Mais c'est surtout au niveau de l'enrichissement de la base et de l'exploitation des données que ces changements devraient représenter un saut qualitatif.

Pour Artefacts, cette structuration des données suppose une *segmentation* de champs jusqu'ici uniques. Au cœur du projet, les fiches typologiques sont évidemment impactées par cette révolution. Rappelons que chaque fiche

regroupe sur une page de synthèse les informations liées à chaque forme d'objet - le type - et ses attestations. Actuellement, une ligne d'attestation peut signaler plusieurs exemplaires retrouvés sur un site, mais peut également contenir d'autres éléments. Ces précisions sont de nature et de format variables (musée, numéro d'inventaire, dimensions, contexte, datation archéologique, variante, etc.) et ne sont pas systématiquement enregistrées. L'utilisateur ne dispose que d'une liste d'attestations, dans l'ordre alphabétique des communes, et ne sait pas si l'objet qui l'intéresse est illustré ou non. Dans la version 1, la seule façon d'établir un lien entre le texte et l'image, qui illustre un ou plusieurs

exemplaires, est la légende, entrée manuellement, qui recopie tout ou partie de la ligne d'attestation selon les cas.

Traiter chaque site à partir d'une association de quatre champs (commune, lieu-dit, région ou département, pays) et non plus d'un seul permettra par exemple de classer les listes d'attestations, à la demande, par département ou par pays, et non uniquement par nom de commune comme c'est le cas aujourd'hui. Les recherches typologiques pourront être limitées à une région (Land allemand, province italienne...) pour rechercher la présence d'une forme dans une zone géographique spécifique, ce qui n'est pas possible actuellement.

Surtout, la segmentation des lignes d'attestations en autant de sous-lignes consacrées à un seul objet, chaque sous-ligne pouvant être associée à une ou plusieurs images, ouvre la voie à l'édition automatique de catalogues illustrés. Il deviendra en effet possible d'exporter les objets d'un musée ou d'un site, par exemple, pour éditer un catalogue au format

choisi. Nul doute que cette nouvelle fonctionnalité sera grandement appréciée des conservateurs de musées et responsables de collections qui, actuellement, n'ont pas cette possibilité, sauf à parcourir une par une toutes les fiches où apparaissent des objets du site ou du musée.

Cette procédure de migration passera par une vérification manuelle de toutes les lignes et sous-lignes existantes grâce à un module de transfert semi-automatisé (Fig. 2). Un script repère automatiquement à l'intérieur des lignes différents contenus et les répartit dans les champs adéquats. L'opérateur doit cependant vérifier que les contenus correspondent aux champs proposés, et corriger si besoin. Cette opération sera d'autant plus rapide qu'elle bénéficiera d'un nombre important de collaborations. Une fois que tout le contenu d'une fiche a été transféré, les nouvelles attestations peuvent être ajoutées au format v2. Le fait que des champs plus nombreux et au contenu mieux identifié soient proposés à la saisie devrait inciter l'utilisateur à étoffer le contenu renseigné pour chaque objet.

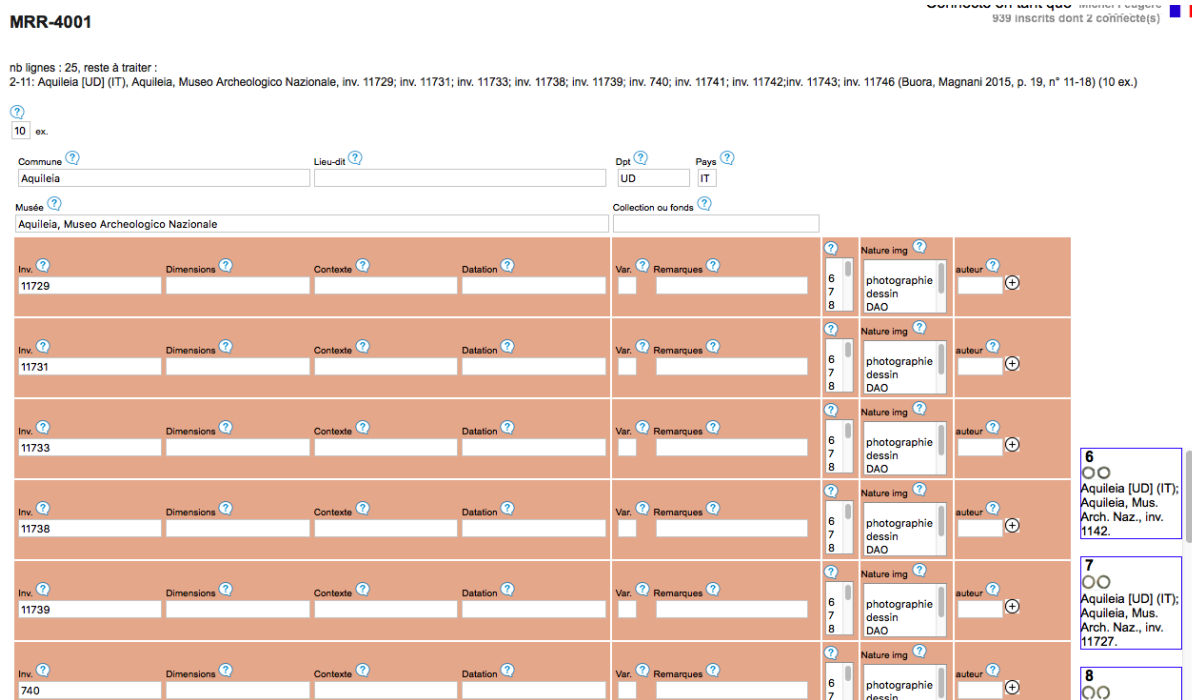


Fig. 2 — Capture d'écran du module de migration, ici pour une ligne comportant au départ 10 objets (MRR-4001). Le module repère que ces dix objets du même site doivent faire l'objet de 10 sous-lignes chacune liée à un objet identifié par un numéro d'inventaire, des mensurations et peut-être des variantes différentes; chaque image doit être affectée à l'une de ces sous-lignes, avec ses métadonnées (copyright, etc.).

Artefacts au futur

Bibliographie : vers plus d'interopérabilité

Prenons maintenant l'exemple de la bibliographie : Artefacts utilise jusqu'ici seulement deux champs pour désigner une référence : une abréviation normalisée (nom date) et un champ de texte libre dans lequel se retrouvent toutes les données d'une référence bibliographique classique : les noms et prénoms de tous les auteurs, le titre, le nom du volume ou de la revue dans laquelle s'insèrent le titre, la ville et l'année de publication, la pagination. Aux données de ce champ s'ajoutent des métadonnées pour caractériser le contenu (période[s] concernée[s], thème[s] abordé[s]) ainsi que des liens pour trouver le document en ligne, l'acheter ou en lire les comptes-rendus publiés.

Issue des pratiques académiques classiques (les chercheurs ont appris à construire leurs bibliographies exactement de cette manière), cette structuration des données n'est nullement satisfaisante d'un point de vue informatique. D'abord, les données du premier champ, le nom de l'auteur et la date de publication, se retrouvent dans le second : chacun voit bien que la redondance pourrait et devrait être évitée. Mais surtout, différentes informations ne sont insérées que de manière indirecte. Le fait que la publication soit un article dans un périodique, dans les Actes d'un colloque ou dans un recueil thématique, par exemple, n'est exprimé que par des normes, du reste variables selon les pays et les pratiques (adjonction de "in", "dans", etc.). Un premier enjeu, de l'ordre de la rationalisation, impose donc de compléter les métadonnées associées à chaque référence.

Mais si cette manière de présenter la bibliographie sur un champ unique est peu logique, voire peu cohérente, il faut surtout

souligner qu'elle est un obstacle à une tendance lourde des bases de données actuelles : l'interopérabilité. Ce concept apparu dès les premières communications entre bases de données¹ prend aujourd'hui une importance cruciale. Il consiste à permettre à des ensembles de données de communiquer entre eux, voire de s'enrichir mutuellement en récoltant, à travers des concepts et des filtres bien définis ("moissonnage"), des données apparues sur un point ou un autre du 'world wide web'.

La question du formatage prend ici tout son sens : il est en effet extrêmement difficile, voire impossible, de collecter des données qui ne répondent pas à un format sinon commun, du moins reconnu. Dans le cas de la bibliographie, chaque élément utilisé doit pouvoir être identifié et donc, idéalement, enregistré dans un champ distinct. L'enjeu est la communication d'Artefacts avec les bases de données couvrant des domaines similaires, comme Frantiq (www.frantiq.fr), ou les outils de gestion bibliographique, comme Zotero (www.zotero.org). Par communication, nous entendons ici les procédures semi-automatiques d'import, mais aussi d'export entre Artefacts et des fichiers personnels ou des bases de données en ligne.

Avec un format d'enregistrement normalisé, Artefacts pourra par exemple importer de Frantiq des références qui lui sont utiles ou exporter des bibliographies vers Zotero. Mais comme le programme comprend une action de dépouillement systématique des revues, nous pourrions aussi, de manière symétrique, proposer à Frantiq ou aux membres d'Artefacts utilisant Zotero des listes de références bibliographiques assorties de métadonnées et classées.

¹ https://books.google.com/ngrams/graph?content=interoperability&year_start=1950&year_end=2017&corpus=15&smoothing=3&share=&direct_url=t1%3B%2Cinter%20-%20operability%3B%2Cc0

Artefacts comme communauté : outils de sociabilité

Nous pensons ensuite, dans une version 2.1, travailler sur les outils de sociabilité visant à renforcer le fonctionnement du site en tant que lieu de rencontre d'une communauté. Le chercheur qui travaille sur Artefacts est en effet isolé : il ne communique avec les personnes qui peuvent partager des mêmes centres d'intérêt qu'à travers le forum, c'est-à-dire en lecture différée. Une amélioration très sensible serait donc apportée par l'adjonction d'une messagerie instantanée permettant des échanges ponctuels, avec un fonctionnement plus immédiat. Il sera également utile d'ajouter à chaque fiche typologique un module de discussion permanente en ligne qui sera consacré à toutes les questions liées à ce type particulier : définition morphologique, relations typologiques avec les formes connexes, datation, état d'avancement, etc.

Dès maintenant (LE, AG) nous mettons en place les champs et les scripts qui vont permettre de retracer l'historique des comportements de nos auteurs : une banque d'informations indispensable à notre perception de ce qui motive les visiteurs, les auteurs et des procédures suivies pour parvenir à l'information souhaitée. Mieux connaître le comportement des utilisateurs est la condition indispensable à toute recherche d'une évolution qui voudrait répondre de manière plus fine aux souhaits de nos usagers. Plus de 3000 marqueurs sont désormais régulièrement mis à jour pour analyser et caractériser l'évolution d'Artefacts et l'activité de ses visiteurs. Il s'agit par exemple de connaître quelle période d'Artefacts intéresse les visiteurs de tel ou tel pays. On peut donc imaginer, dans l'avenir (et comme le font dès aujourd'hui un grand nombre de sites commerciaux), de voir apparaître dans Artefacts des "recommandations". A celui qui a consulté par exemple un

grand nombre d'objets du début du Moyen Âge, on pourra signaler les nouveaux titres bibliographiques récemment enregistrés pour cette période dans Artefacts.

Cartographie

Un certain nombre de perfectionnements pourront ensuite être apportés, au sein d'une version ultérieure, à tout ce qui concerne la cartographie d'Artefacts. Le premier projet concerne le choix des fonds de cartes, actuellement limité à 3 par défaut (France et zones limitrophes, Europe, Europe et Méditerranée). Nous envisageons dans un premier temps de multiplier ces fonds et de donner la possibilité à chaque utilisateur d'afficher, par défaut, les trois qu'il utilise le plus fréquemment, après avoir choisi celles qui conviennent le mieux à sa recherche en cours. Le module de création d'un espace géographique personnel, actuellement sur les pages typologiques, pourrait prendre place sur cette future page de sélection de cartes (Fig. 3).

A plus long terme, nous avons commencé à réfléchir à l'utilisation d'un thésaurus permettant d'entrer dans le "web sémantique", où les recherches ne se font plus sur un terme spécifique mais sur des associations de termes reliés entre eux par différentes hiérarchies. Dans la foulée des bibliothécaires, les archéologues commencent eux aussi à constituer des thésaurus spécialisés. L'enjeu est de taille, car ces listes multilingues doivent être alignées sur différentes langues académiques et sur l'existant (Pactols). Nous pensons donc que la démarche devrait être coordonnée avec les programmes similaires qui existent, notamment, en Grande-Bretagne et en Allemagne.

artefacts
Encyclopédie collaborative en ligne des objets archéologiques

UNIVERSITÉ LYON 2

Connecté en tant que Michel Feugère (Administrateur)
919 inscrits dont 1 connecté(s)

[Vue liste](#) [Vue planche vignettes](#)

Gestion des fonds de cartes

Artefacts affiche par défaut 3 fonds de cartes utilisables pour vos cartes automatiques, mais vous pouvez à tout moment modifier cette sélection et choisir les fonds de cartes que vous utiliserez le plus souvent, ou même créer votre propre fond de carte personnalisé.

Sélectionnez vos fonds de cartes favoris en sélectionnant 3 cartes ci-dessous :

France et ses marges Europe moyenne Europe et Méditerranée Italie

Europe, Méditerranée, limes c. 120 Europe et Méditerranée (bichromie) Europe et Méditerranée (N&B) Créer sa propre zone

Fig. 3 — Projet de bibliothèque de fonds de cartes : les fonds géo-référencés sont proposés dans un espace personnel, où chaque utilisateur choisit à la demande les fonds qui conviennent le mieux à son espace de réflexion.

Bibliographie

Feugère 2010 : M. Feugère, The Artefacts Project : An Encyclopaedia of Archaeological Small Finds. *Lucerna (The Roman Finds Group Newsletter)* n°39, Sept. 2010, 4-6. [Lien](#)

Feugère 2011 : M. Feugère, Artefacts : encyclopédie en projet, outil d'aujourd'hui. *Instrumentum* n°33, juin 2011, 24-27. [Lien](#)

Feugère 2015 : M. Feugère, Bases de données en archéologie : de la révolution informatique au changement de paradigme. *Cahiers Philosophiques* n°141, 2015, 139-147. [Lien](#)

Feugère 2016 : M. Feugère, Artefacts en 2015-2016. *Instrumentum* 43, juin 2016, 54-57. [Lien](#)