



HAL
open science

Enjeux pour la mise en réseau et l'analyse des connaissances archéologiques

Guillaume Reich, Sébastien Durost, Jean-Pierre Girard, Eric Lacombe, Miled Rousset

► **To cite this version:**

Guillaume Reich, Sébastien Durost, Jean-Pierre Girard, Eric Lacombe, Miled Rousset. Enjeux pour la mise en réseau et l'analyse des connaissances archéologiques. 16ème Conférence internationale TOTh – Terminologie & Ontologie: Theories and applications, Christophe Roche, Jun 2022, Chambéry, France. hal-04583684

HAL Id: hal-04583684

<https://hal.science/hal-04583684>

Submitted on 22 May 2024

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.

Enjeux pour la mise en réseau et l'analyse des connaissances archéologiques

Guillaume Reich*, Sébastien Durost**, Jean-Pierre Girard***
avec la collaboration d'Éric Lacombe**** et de Miled Rousset*****

* Bibracte - BIBRACTE EPCC - Centre archéologique européen
UMR 8546 CNRS-PSL-AOrOc - École normale supérieure
g.reich@bibracte.fr / guillaume.reich@ens.psl.eu

** Bibracte - BIBRACTE EPCC - Centre archéologique européen
s.durost@bibracte.fr

*** UMR 5133 Archéorient - CNRS, Université Lumière Lyon 2
jean-pierre.girard@mom.fr

**** UR 4426 MICA, Université Bordeaux Montaigne
eric.lacombe@eguilde.eu

***** FR 3747 Maison de l'Orient et de la Méditerranée Jean-Pouilloux
miled.rousset@mom.fr

Abstract. To identify, describe, classify, quantify, understand and explain ancient material realities and their connections in order to write the history of societies, archaeology manipulates graphic and textual descriptions. The restitution of a site exposes the data collected, their relations (spatial, temporal and/or thematic) and the nature of the reasoning used, while keeping the permanent concern to be fact-proof. The ISO 25964 standard (thesaurus management) is flexible enough to describe the different “*points of view*” of archaeologists, presiding over the elaboration of their specialized and evolving “*micro-languages*” and terminological concepts (e.g. typochronologies) and for logical modeling, as well as for structuring the data in addition to their metadata. Projected as computable graphs, the vocabulary, data and their mutual relationships then make it possible to characterize the processes of transformation and organization of the reasoning as of meaning of the concepts from specialized archaeological “*micro-languages*.”

L'archéologie consiste à collecter, identifier, décrire, classer, quantifier, puis comprendre et interpréter des réalités matérielles anciennes pour écrire de la façon la plus rigoureuse possible une histoire – matérielle et immatérielle – des sociétés humaines à partir de leurs traces conservées et des connaissances accumulées au sein de cette communauté scientifique. Les connaissances archéologiques sont obtenues initialement par des fouilles sur le terrain qui détruisent méthodiquement un site pour pouvoir l'observer tout en consignait rigoureusement les différentes étapes de cette opération. À l'issue du processus de fouille, la documentation enregistrée (Figure 1) et les vestiges archéologiques associés (artefacts et écofacts exhumés) deviennent les uniques témoins de la réalité physique du site et doivent rendre compte de deux temporalités interdépendantes :

- le temps de la fouille : une succession d'instantanés, appartenant à un passé très récent, correspondant aux méthodes employées lors de l'opération archéologique ;
- les différents événements observés et les divers vestiges (prélevés ou non) : succession d'instantanés chronologiquement cohérents regroupés en périodes, appartenant au passé étudié, bien antérieur à l'acte de fouille.

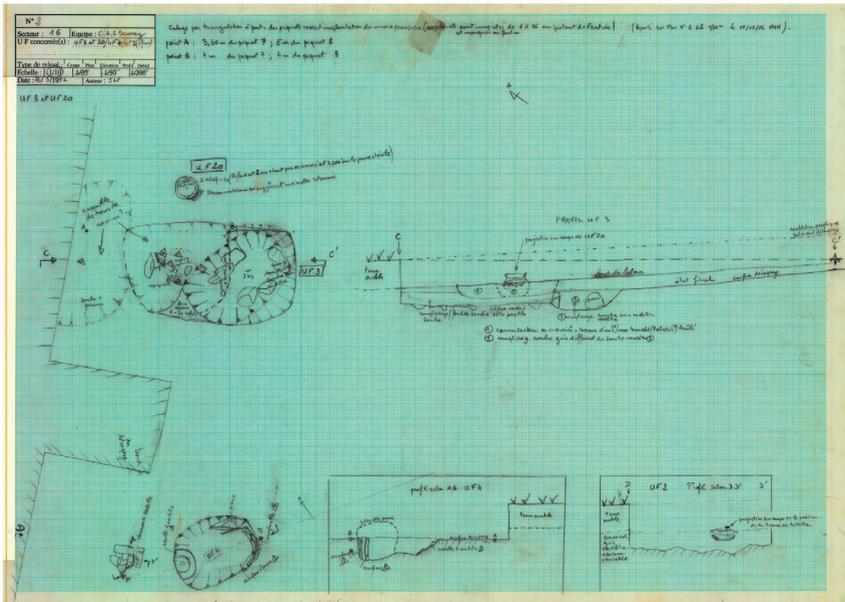


FIG. 1 – Minute (croquis de terrain) représentant des unités de fouille à Bibracte.

Pour produire cette documentation, l'archéologue choisit les unités taxinomiques qui lui semblent les plus adaptées pour rendre compte de ce qu'il fait (méthodes) et de ce qu'il observe (terrain ou vestige exhumé). L'analyse des relations physiques (stratigraphie pour la succession des couches et planimétrie pour la détermination de leur projection horizontale) et l'étude des caractéristiques communes (typologie) entre les unités taxinomiques permettent à l'archéologue de les hiérarchiser pour les intégrer à un réseau de connaissances multiscale en évolution permanente. Cette mise en réseau – que les archéologues nomment '*contextualisation*' – s'effectue à trois niveaux: microscopique (la compréhension des artefacts ou écofacts spatialisés), mésoscopique (l'histoire du site) et macroscopique (l'évolution de la société à

laquelle les vestiges du site sont rattachés). Ce travail constant de mise en contexte dépasse largement, en amont comme en aval, le temps de la fouille et de sa publication, notamment en raison du caractère partiel des informations recueillies et des informations nouvelles périodiquement apportées par l'exploration de nouveaux sites.

Après la fouille, ne sont donc manipulées que des réductions interprétatives du terrain, sous forme de descriptions graphiques (minutes dessinées, plans, croquis, dessins, orthophotographies, etc.) accompagnées de compléments textuels en langue vernaculaire plus ou moins structurés. Les vestiges archéologiques exhumés (artefacts, écofacts, échantillons de sédiment, etc.) font encore l'objet d'observations complémentaires en dehors de la fouille, par observation humaine directe (étude du mobilier) ou par le recours à des technologies d'investigation non mobilisables sur le site (analyses physico-chimiques, radiographies, expertises paléobotaniques, datations, *etc.*), elles aussi synthétisées par des données textuelles ou graphiques (Figure 2), calculables ou non (Gardin, Borillo 1970; Borillo 1978), et s'intégrant dans leur propre réseau de connaissances (palynologie, dendrochronologie, *etc.*).

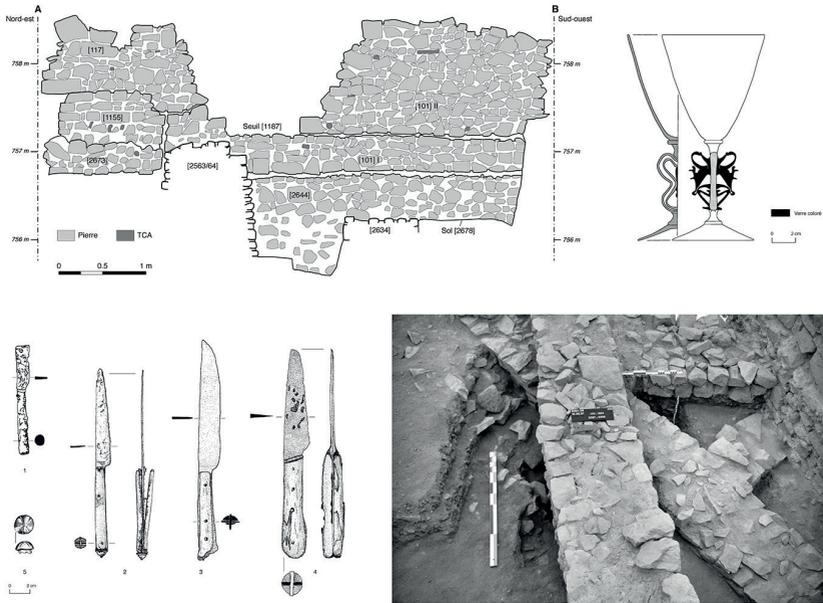


FIG. 2 – *Dessins techniques (post-fouille) et photographie de terrain (Bibracte).*

La restitution de l'histoire du site consiste à agencer les données collectées, c'est-à-dire les informations issues des données et de leurs relations, ainsi que la nature des raisonnements pour y parvenir, en gardant le souci permanent d'être à l'épreuve des faits archéologiques. Actuellement, cette restitution s'effectue par la diffusion institutionnalisée de publications écrites, sous forme de rapports scientifiques, d'articles ou de monographies, en langage vernaculaire. En France, la seule représentation structurée obligatoire des informations recueillies est le diagramme stratigraphique, qui représente la chronologie d'un site à partir de la succession de ses couches sédimentaires et anthropiques sous la forme d'un graphe calculable (Harris 1979) produit dans le cadre des documents administratifs que sont les rapports adressés

aux services de l'État; rares sont en revanche les publications de synthèse (monographies ou articles) exprimant l'ensemble de ces relations calculables. Les propriétés des relations ne relevant pas de la stratigraphie, quant à elles, ne sont généralement pas restituées et formalisées, de manière systématique, sous forme de représentations logiques du raisonnement mathématique, dans une instance permettant leur normalisation et leur manipulation en temps réel. Leur représentation est un enjeu scientifique majeur depuis quelques années en raison d'un fort accroissement du volume des données à traiter (naissance et développement de l'archéologie préventive), car elles rendent compte des interprétations et hypothèses (chronologiques, fonctionnelles, spatiales, anthropologiques, *etc.*) formulées à partir des vestiges observés et des connaissances accumulées par ailleurs.

Les actualisations du discours archéologique s'inscrivent dans un processus itératif permanent (Moberg 1969) oscillant (*feed-back*) entre les étapes de l'archéographie (demander / observer (archéométrie, archéoscopie) / décrire (ordonner, classifier, trier, généraliser)) et celles de la production du discours sur le passé (archéologie: *Ἀρχαῖος*, 'ancien', et *Λόγος*, 'discours') à proprement parler (analyser, associer, interpréter, comprendre, *etc.*). Cependant, un tel dispositif heuristique est handicapé par une restitution très incomplète des informations. Cette faiblesse méthodologique a très tôt interpellé la communauté archéologique (Gardin 1971; Tchernia *et al.* 1984), mais les solutions envisagées pour y pallier se heurtaient à une double barrière: l'impossibilité technologique de la numérisation massive des données pour un traitement algorithmique et la contrainte financière du coût de leur reproduction exhaustive dans une publication imprimée. Avec la démocratisation de l'informatique à partir des années 1990 et la mise en réseau de documents numériques accessibles par l'Internet dans les années 2000, la communauté scientifique aurait pu s'attendre à la résorption rapide de ce problème méthodologique par l'émergence de nouveaux dispositifs, ainsi qu'à la démocratisation de la calculabilité des données archéologiques.

Les technologies numériques actuelles ont sans doute atteint la maturité nécessaire pour répondre à cette question technico-méthodo-

logique, mais cela contraindrait en l'état la communauté archéologique à conditionner sa pratique heuristique à l'usage d'outils conceptualisés en dehors de son périmètre intellectuel, ce qui n'est guère souhaitable pour des raisons épistémologiques (leur logique lui échapperait) et pratiques (elle ne les maîtriserait pas). Pourtant, ces solutions numériques permettraient de réduire la part trop grande laissée aux capacités déductives cumulatives issues de l'expérience de chacun et de la culture commune de la discipline, qui donnent trop d'importance à l'implicite et à la capacité individuelle de mémorisation. Comment passer, avec les archéologues (co-construction), de restitutions en langage vernaculaire, fluides mais déconnectées des données, à des encodages numériques fidèles aux discours comme aux données et permettant de les relier (interopérabilité), pour finalement leur permettre de déléguer à la machine une partie de leurs capacités de mémorisation et d'exploitation (classement et calcul) ?

1. Modalités de la mise en réseau des données numériques en archéologie

1.1. Le vocabulaire

La première condition est de formaliser les vocabulaires mobilisés pour la description des données par un chercheur, c'est-à-dire d'exposer la singularité de son point de vue et la richesse de son vocabulaire, d'être mis en capacité de comprendre d'autres paradigmes et de pouvoir comparer sa terminologie à l'espace sémantique de ses collègues. Cela suppose de définir clairement chacun des termes mobilisés. Il est très fréquent dans la discipline qu'un vocabulaire soit créé spécifiquement ou réorganisé en fonction d'un cas d'étude particulier (les typologies, par exemple) et qu'un même terme renvoie, selon le contexte (culturel, chronologique, *etc.*) et les points de vue, à des réalités différentes, évolutives, sans que les spécialistes en aient toujours conscience. Ce problème de polysémie rejaillit sur la gestion de la documentation, puisque, selon ses usagers, un terme utilisé comme *mot-clé* dans un catalogue ou dans une base de données ne renvoie pas nécessairement

aux mêmes réalités matérielles ou conceptuelles. Cela rend également incertaine la comparaison de corpus archéologiques, car le sens des vocabulaires mobilisés pour indexer les données diffère souvent d'un jeu de données (base de données ou publications) à l'autre.

En parallèle de la publication physique du volume 31 de la collection *Bibracte* portant sur *La vaisselle céramique de Bibracte* (Barrier, Luginbühl 2021), la formalisation du vocabulaire spécialisé permettant de caractériser et de décrire les fragments de poterie retrouvés sur l'*oppidum* gaulois éponyme a été expérimentée avec *Opentheso* (<https://opentheso.huma-num.fr/opentheso/>), un gestionnaire de thésaurus conforme à la norme ISO 25964 (<https://www.iso.org/fr/standard/53657.html>) qui cadre l'élaboration des vocabulaires contrôlés (Durost *et al.* 2022).

L'une des règles de la norme établit l'impossibilité d'exposer dans un même thésaurus deux termes strictement identiques en leur associant des espaces sémantiques différents (position dans la hiérarchie des termes, relations, définition). Une telle polysémie est habituellement éclairée dans les publications scientifiques par la référence bibliographique mentionnant le patronyme de son producteur et l'année de création du terme – que nous appelons '*millésime*'. Ce recours au millésime est apparu comme une solution évidente et fonctionnelle dans l'exposition de vocabulaires contrôlés concurrents et contextualisés par l'ajout au libellé d'un appel de citation bibliographique: nom(s) + année. Découlant de ce choix, la compréhension d'un terme s'inscrit dans ce même contexte bibliographique *millésimé*. La définition (au sens de la norme ISO 1087-2019: "*representation of a concept by an expression that describes it and differentiates it from related concepts*") permet d'individualiser le concept en qualifiant ses propriétés. Elle précise le paradigme du chercheur pour appréhender son sujet d'étude qui, par le jeu des alignements entre termes de thésaurus différents, peut être confronté à la multiplicité des points de vue exposés dans ceux-là et s'inscrire ainsi dans un réseau de connaissance. Par ailleurs, pour un même type d'artefact, les méthodes d'étude et d'analyse peuvent changer en fonction de la problématique, de la personnalité du chercheur et des conditions taphonomiques (c'est-à-dire de tous les pro-

cessus physico-chimiques qui interviennent après l'abandon de l'objet). Dès lors, l'archéologue crée son propre vocabulaire et/ou puise tout ou partie de sa terminologie dans un référentiel déjà utilisé par ailleurs pour l'adapter à l'expression de ses besoins.

Pour connaître avec précision les vocabulaires créés ou convoqués par les auteurs, il est nécessaire d'avoir accès au(x) corpus sur le(s) quel(s) ils se basent. Leur mise en relation, pour des comparaisons, est un besoin fondamental qui s'exprime, à l'heure du numérique, par la nécessité d'une mise en réseau technologique des connaissances. Un gestionnaire de thésaurus est tout à fait capable d'initier cette dernière par l'intermédiaire de ses relations internes (associations entre concepts d'un même thésaurus) et externes (alignements entre des concepts issus de thésaurus différents), pour peu que l'organisation et la mise en réseau de ces vocabulaires soient prolongées par l'accès aux données enregistrées et/ou définies à partir des réalités matérielles ou immatérielles. En effet, seul le retour aux données permet d'apprécier le sens du vocabulaire et la pertinence des alignements avec un vocabulaire de comparaison et – donc – son jeu de données associé.

1.2. La grammaire

La deuxième condition est de rendre comparable la grammaire des données archéologiques pour mieux expliciter les vocabulaires de la recherche et ce sur quoi ils se fondent. La spécificité des problématiques et les différentes disciplines mobilisées par l'archéologie conduisent à structurer les données dans des bases modélisées selon des logiques adaptées, comme autant de prémices d'ontologies d'application (Guarino 1998): la multiplication des données, la part d'inconnu propre à l'investigation archéologique, les progrès méthodologiques et le renouvellement constant des questionnements anthropologiques favorisent ainsi la structuration des données selon des logiques singulières et conduisent généralement au fonctionnement en silo des bases de données. La gouvernance individualisée des données est alors maximale.

Il semble judicieux d'envisager *a contrario* un écosystème informationnel favorisant et mettant en réseau tant la diversité des données et des ressources que la pluralité assumée des points de vue. En l'état, la non-formalisation et le non-partage des modèles logiques d'enregistrement des données – qui doivent décrire la structure des données sollicitées sans faire référence au langage de programmation utilisé pour les produire – par les concepteurs des bases de données conduit à l'absence de syndication entre ces dernières. Ce n'est qu'en exposant et en explicitant les modèles logiques, c'est-à-dire en confrontant perpétuellement les divers paradigmes des chercheurs, qu'une synthèse collégiale est possible à un instant T et peut se renouveler à la mesure de l'évolution des corpus archéologiques. En effet, si tous les avis ne se valent pas, tous doivent être pris en compte pour une discussion. Loin de la création artificielle d'un *esperanto des données*, ni d'une *tour de Babel* où plus personne ne se comprendrait et ne prendrait la peine de dialoguer, nous voulons que l'écosystème informationnel intègre les différentes réflexions et toutes les données, expose les divers vocabulaires, idées et corpus produits par la communauté archéologique. Cela donnerait aux archéologues la possibilité technologique d'exprimer, avec mesure, divergences et convergences paradigmatiques/idéologiques, le tout pour tendre vers une connaissance plus affirmée et transparente, sans amoindrissement des richesses sémantiques (méta-référentiel illusoire) dans ce domaine.

Les besoins d'exposition et d'explicitation des modèles logiques sont comparables à ceux des vocabulaires de la recherche; les solutions aussi. Pour expérimenter cette approche, une partie des modèles logiques de la base de gestion de la documentation de Bibracte *bdB* développée par Raphaël Moreau (Guichard 2000; Chaillou 2003, 2004; Bibracte 2006), de l'outil de saisie de terrain *EDArc* conçu par Christophe Tufféry pour l'Inrap (Tufféry, Augry 2019) et de la base d'enregistrement *ODS* créée par Bertrand Bonaventure (<https://www.9heuresprecises.com/>) ont été décrits sous forme de thésaurus.

L'organisation des données dans les bases, en tables, en champs et en lignes, a fait l'objet d'une transposition en '*concepts*' définis et structurés sur la base de la documentation fournie par les concepteurs des

modèles ; les relations affichées entre les données (liens entre les tables) ont été exprimées par les principes de l'association et de la polyhiérarchie. Le test d'alignement entre ces trois modèles logiques, en cours, ouvre la voie à la migration de données d'une structure logique à une autre (Figure 3). Les concepts décrivant les champs de *bdB* appelant du vocabulaire contrôlé ayant été mis en relation avec les concepts des vocabulaires spécifiques de la recherche à Bibracte, formalisés dans un thésaurus nommé *Bibracte_Thesaurus*, il devient alors théoriquement possible de mettre en relation l'ensemble des vocabulaires contrôlés et de les utiliser indépendamment du modèle logique (Figure 4). La démocratisation d'un tel écosystème numérique de coopération entre partenaires d'un domaine permettrait de résoudre les conflits formels (formats, structures, modèles logiques) et sémantiques lors de l'échange et de la mise en commun de données issues de différents systèmes.

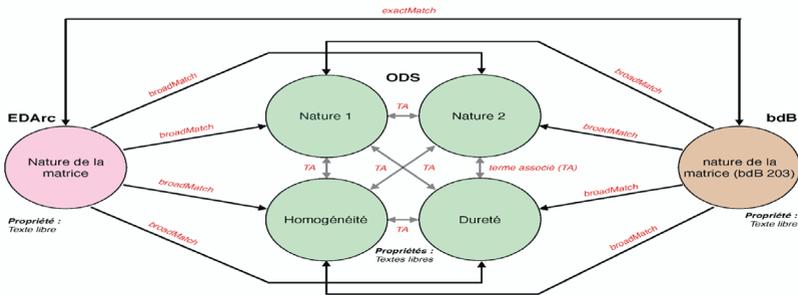


FIG. 3 – Description et alignements des modèles logiques de trois bases de données en archéologie à granularités variables.

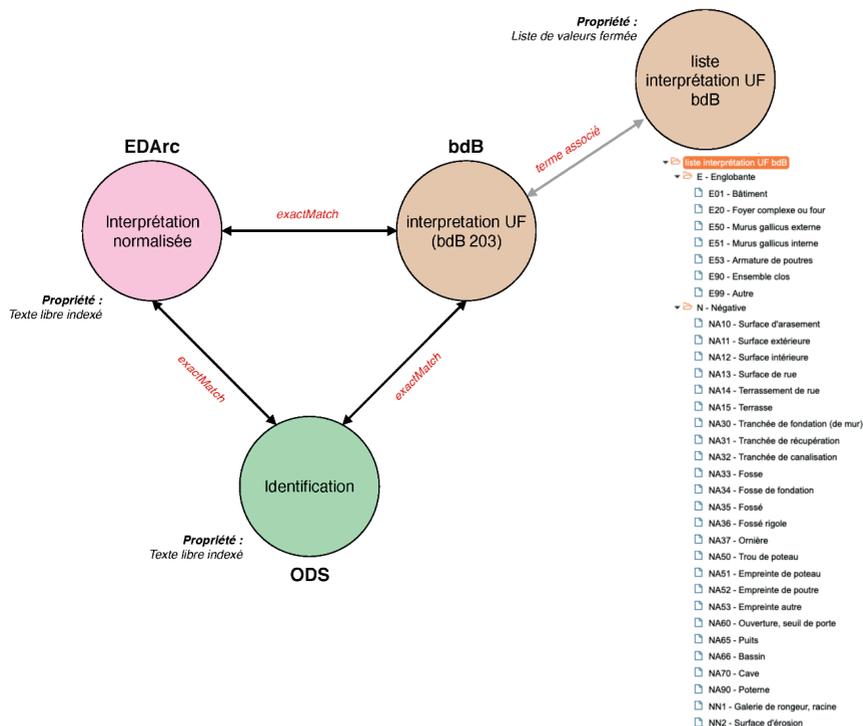


FIG. 4 – Mise en relation d'un vocabulaire structuré en thésaurus avec les modèles logiques de trois bases de données en archéologie.

1.3. Le raisonnement

La troisième condition est de décrire et de codifier la (ou les) nature(s) de chacune des relations (abduction, induction, déduction, observation, hypothèse, expérimentation, comparaison/analogie, *etc.*) de façon à formaliser et expliciter les raisonnements archéologiques. Une tentative est en cours sur un jeu de données de Bibracte à travers l'infrastructure logicielle *Neo4j*, un système de gestion de base de

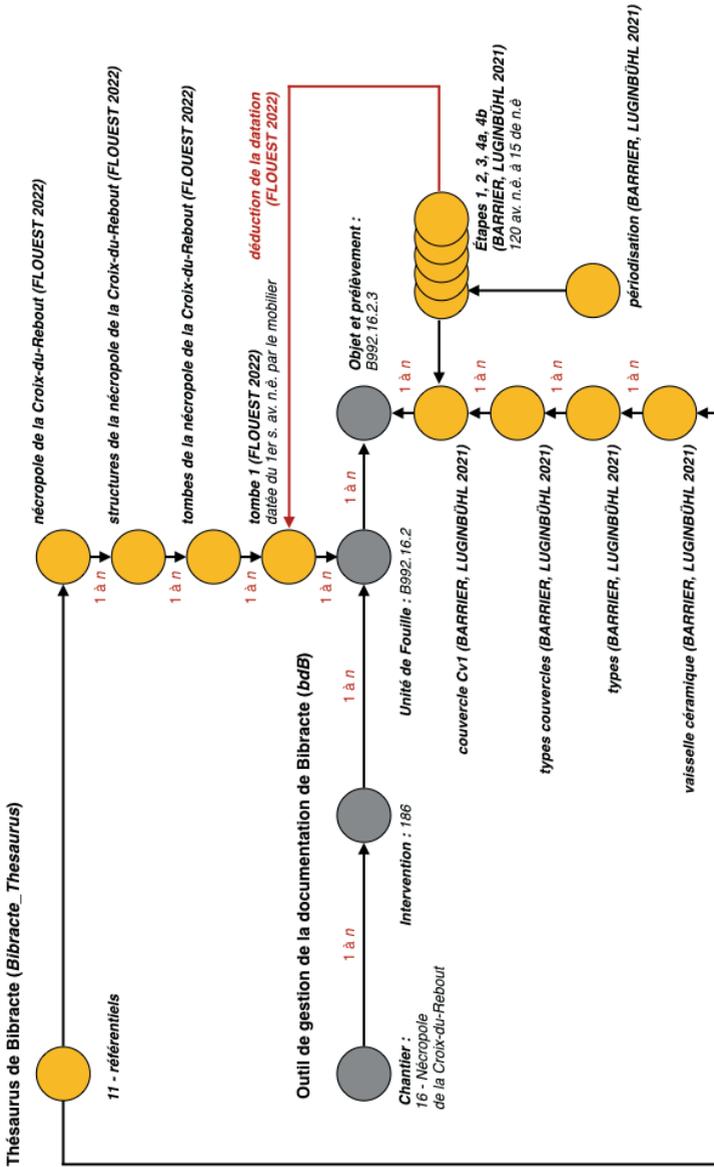


FIG. 5 – Modélisation de l’articulation logique entre thésaurus et système d’enregistrement pour la création de la connaissance archéologique (exemple d’une fouille de Bibracte).

données orientée graphe au code source libre, couplée au gestionnaire de thésaurus *Opentheso*. Une des finalités heuristiques de cette expérimentation numérique est de tester le traçage historiographique de la compréhension et de la réappropriation par un chercheur A du paradigme d'un chercheur B (Figure 5).

À Bibracte, l'outil de gestion de la documentation (*bdb*) permet de gérer la documentation produite par les fouilles jusqu'au niveau de l'objet ou prélèvement. Dans notre exemple, l'artefact *B992.16.2.3* a été collecté dans l'unité de fouille (UF) *B992.16.2*, lors de l'intervention *186* (fouille préventive du 23/03/1992 au 31/12/1992 sous la responsabilité de Jean-Loup Flouest), dans le cadre du chantier *16* qui regroupe toutes les interventions sur le secteur de la Croix-du-Rebout de l'*oppidum* de Bibracte.

La poterie *B992.16.2.3*, recouvrant un autre récipient (*B992.16.2.2*), a fait l'objet d'une étude complémentaire spécialisée par un céramologue qui a identifié cet artefact comme étant un couvercle de type *Cv1*, en usage entre 120 av. n.è et 15 de n.è, servant à couvrir un récipient pour la cuisson ou le stockage des aliments, d'après la typologie de *La vaisselle céramique de Bibracte* publiée par Sylvie Barrier et Thierry Luginbühl en 2021. Cette typologie a été construite à partir des céramiques trouvées en différents secteurs de l'*oppidum*, notamment dans l'habitat, contexte ordinaire et principal de l'utilisation de cette vaisselle. Elle les classe selon leurs usages (servir, conserver, stocker, cuire, etc.) et les date en fonction de critères intrinsèques (archéométrie) et extrinsèques (comparaisons avec d'autres sites).

Par le jeu des observations et des raisonnements, une douzaine de types de structures a pu être identifiée (aires de crémation, dépôts de restes de bûchers, dépôts d'offrande, etc.) sur l'emprise de la fouille. L'UF *B992.16.2* a été interprétée comme faisant partie d'une tombe à incinération (*tombe 1*) par ses différentes caractéristiques (notamment un creusement dans la terre avec du mobilier mêlé à des cendres humaines). Elle contenait le couvercle *B992.16.2.3*, qui a permis de dater par déduction cette sépulture du 1^{er} s. av. n.è. et d'établir un *terminus post quem* (c'est-à-dire que le couvercle ne peut pas avoir été

déposé avant sa (période de) fabrication). D'autres éléments de datation ont été pris en compte, mais ne sont pas détaillés dans cet exemple. On voit ici que seule une partie des propriétés de la typologie de la vaisselle céramique de Bibracte a été réutilisée pour comprendre la place de cette sépulture dans l'histoire du site. La fonctionnalité de ce type de couvercle est ici partiellement respectée : l'objet sert bien à couvrir un récipient en vue de conserver son contenu, mais dans un contexte non alimentaire. La présence d'ossements humains calcinés permet de définir le couple d'objets comme une urne funéraire avec son couvercle. Ainsi, une même réalité matérielle peut être définie différemment selon son contexte d'utilisation. Pour l'heure, la diversité et l'enchaînement de raisonnements parfois complexes qu'opèrent les archéologues sont seulement retranscrites en langage vernaculaire ou restent implicites.

Pourtant, cette opération de justification des étapes clés du raisonnement permettrait de pallier les faiblesses humaines (individuelles comme collectives) dans la compilation et l'utilisation de données, et autoriserait, par leur explicitation, à en visualiser, voire à en calculer les propriétés en temps réel sur la base des corpus mobilisés. Par sa logique itérative et cumulative, cette approche méthodologique permettrait de limiter les raisonnements circulaires induits par l'utilisation de jeux de données restreints dont la pertinence – c'est-à-dire la représentativité et la volumétrie – par rapport à l'ensemble des données collectées par la communauté ne peut être mesurée. Dans cette expérimentation, qui se base sur la théorie des graphes, l'enjeu est de qualifier et de millésimer les raisonnements permettant de relier et d'exploiter des données numériques, qui sans cela ne sont que des corpus juxtaposés et dénués de sens (orientation et signification). Il deviendrait alors possible de cheminer dans les données en suivant schématiquement les raisonnements mobilisés, en parallèle des discours en langage vernaculaire.

2. Possibilités théoriques de l'analyse de la transformation des connaissances

La réflexion et l'expérimentation directement issues du terrain, exposées ci-dessus, constituent, en effet, des tentatives de réponse à

des questions épistémologiques inhérentes à la discipline archéologique. L'utilisation d'un thésaurus construit à partir de l'observation pour référencer puis comparer les vocabulaires documente la réponse à la première des deux questions constitutives de la pratique disciplinaire (Boissinot 2015): *Qu'est-ce qu'il y a ici?* L'interprétation va ensuite s'appuyer sur ce référentiel pour tenter de répondre à la deuxième question: *Que s'est-il passé là?* Ce passage des données (*ce qu'il y a ici*) aux connaissances (*ce qui s'est passé là*) fait appel à des raisonnements (par inférence, déduction et/ou abduction) et des processus (physiques et mentaux), générant des constructions intermédiaires, avec des allers-retours entre des connaissances tacites, autrement dit implicites, et des connaissances explicites. Le langage et l'écriture, en tant qu'énonciation, sont des actes de conversion des connaissances tacites en connaissances explicites (Nonaka, Takeuchi 1997; Emig 1983). Ce qu'exprime ainsi Lev Semionovitch Vygotsky (2012 [1934]): «*Le discours intérieur est un discours condensé, abrégé. Le discours écrit est déployé dans toute son ampleur, plus complet que le discours oral. Le discours intérieur est presque entièrement prédictif car la situation, le sujet de la pensée, est toujours connue du penseur. Le discours écrit, au contraire, doit expliquer la situation de manière complète pour être intelligible. Le passage du discours intérieur le plus compact au discours écrit le plus détaillé exige ce que l'on pourrait appeler une sémantique délibérée, c'est-à-dire une structuration délibérée du réseau de signification*». Le discours, construit par tissage d'expressions linéaires du langage verbal, se heurte en effet, comme on l'a vu précédemment, à la polysémie. Pour y échapper, sans évidemment renoncer à la catégorisation liée au raisonnement par inférence, la prise en compte d'une vision systémique évolutive appelle une logique étendue.

La recherche de l'ordre et de la structure des choses est aujourd'hui actualisée par les sciences informatiques et cognitives (Chazal 2000). Aussi, aux deux systèmes d'écriture traditionnels, la langue (verbe) et le nombre (mathématique), peut-on en ajouter un troisième: le code (Herrenschmidt 2007)? La langue qualifie le réel, observable selon différents points de vue et donc source de multiples interprétations; elle jongle entre termes (polysémiques) et concepts (monosémiques). Le

nombre quantifie le réel, par comptage de la discontinuité et mesure de la continuité. Le code s'appuie conjointement sur la langue et le nombre, pour décrire et incarner une organisation et les lois de sa transformation et ainsi exprimer un mécanisme de structuration dynamique de l'information : un système de connaissance. Code génétique du vivant, code juridique qui fixe explicitement les règles de la vie en société, code informatique qui pilote nos systèmes socio-techniques et fixe implicitement d'autres règles, complémentaires et parfois contradictoires : "*Code is law*" (Lessig 2000). Dans ce cadre, connaître consiste à décrypter le code ; la connaissance archéologique n'échappe pas à cette logique.

De nombreux travaux ont emprunté cette voie pour s'efforcer d'appliquer à l'archéologie le formalisme conceptuel des ontologies, qu'elles soient 'de haut niveau' (tel le CIDOC-CRM) ou '*de domaine*' (par exemple : Liuzzo, Evangelisti 2021). Une telle approche a néanmoins pour inconvénient (outre sa complexité d'application qui l'éloigne inévitablement du terrain) d'aboutir implicitement à normaliser, donc effacer, *de facto*, l'originalité de point de vue exprimée par l'infinie variété d'un langage de description du réel, tel qu'observé... et transcrit (Iacovella *et al.* 2006). En s'appuyant sur une unité élémentaire de sens, appelée '*graine d'information*' (Lacombe 2021), qui regroupe un jeu minimal et cohérent de données en seulement dix espèces différentes, Éric Lacombe tente actuellement de modéliser ce processus à partir des données de Bibracte. L'objectif de cette analyse est de mesurer l'adéquation entre le modèle logique d'enregistrement des données de Bibracte, les vocabulaires et les processus mis en place à Bibracte pour produire et organiser ses données. Autrement dit, comment influe le temps de la fouille sur la représentation du passé par les archéologues.

La conception effective par la communauté archéologique d'un tel écosystème numérique permettrait de caractériser, hiérarchiser et partager des données diverses au sein d'un réseau de connaissance multi-scalaire (microscopique, mésoscopique et macroscopique) en évolution permanente et de rendre plus efficient le partage de données ouvertes, de fortifier, grâce à l'outil numérique, la mise en récit scientifique du processus de création de connaissance bâti à partir des faits matériels

observés par les archéologues, et de consolider épistémologiquement la constitution progressive d'un discours sur les sociétés humaines passées.

References

- Barrier, S. & Luginbühl, T. (2021). *La vaisselle céramique de Bibracte : de l'identification à l'analyse*. Glux-en-Glenne : Bibracte, 316.
- Bibracte (2006). *Gestion de la documentation scientifique et des mobiliers issus des opérations archéologiques dans le cadre de la réglementation actuelle. Séminaire 25-27 septembre 2006 – Bibracte*. Glux-en-Glenne : Bibracte.
- Boissinot, P. (2015). *Qu'est-ce qu'un fait archéologique ?* Paris : EHESS, 366.
- Borillo, M. (1978). *Archéologie et calcul*. Paris : Union générale d'éditions, 246.
- Chaillou, A. (2003). *Nature, statut et traitements informatisés des données en archéologie : les enjeux des systèmes d'informations archéologiques*. Thèse de doctorat, Lyon.
- Chaillou, A. (2004). Présentation de bdB, base de données utilisée à BIBRACTE, centre archéologique européen. *Cahier des thèmes transversaux ArScAn*, CNRS - UMR 7041 (Archéologie et Sciences de l'Antiquité - ArScAn), 130-133.
- Chazal, G. (2000). *Les réseaux du sens. De l'informatique aux neurosciences*. Seyssel : Champ Vallon, 286.
- Durost, S., Reich, G. & Girard, J.-P. (2022). Terminologies, modèles de données archéologiques et thésaurus documentaires : réflexions à partir d'une typologie de céramique. *Humanités numériques*, 6, [En ligne] <https://doi.org/10.4000/revuehn.3119>
- Emig, J. (1983). *The Web of meaning: essays on writing, teaching, learning, and thinking*. Boynton: Cook Publishers, 178.
- Gardin, J.-C. (1971). *UNISIST, Study report on the feasibility of a World Science Information System*. Paris : UNESCO, 159.
- Gardin, J.-C. & Borillo, M. (1970). *Archéologie et calculateurs : problèmes sémiologiques et mathématiques. Acte du colloque interna-*

- tional du Centre National de la Recherche Scientifique, Marseille, 1969.* Paris: Ed. du CNRS, 371.
- Guarino, N. (1998). *Formal Ontology and Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998.* Amsterdam: IOS Press, 3-15.
- Guichard, V. (2000). *La gestion de la documentation archéologique au Centre Archéologique Européen du Mont Beuvray.* Bourges, Assises nationales de la conservation archéologique, Bourges, 26-28 novembre 1998.
- Harris, E. C. (1979). *Principles of archaeological stratigraphy.* London: Academic press, 136.
- Herrenschmidt, C. (2007). *Les trois écritures. Langue, Nombre, Code.* Paris: Gallimard, 510.
- Iacovella, A., Bénel, A., Pétard, X. & Helly, B. (2006). *Corpus scientifiques numérisés : savoirs de référence et points de vue des experts.* In: Pédaque, R.T. (2006). *La redocumentarisation du monde.* Toulouse: Cépaduès Édition, 213.
- Lacombe, E. (2021). *Transformation numérique des organisations en réseau : les potentiels d'une schématisation dynamique de l'information.* Université Michel de Montaigne - Bordeaux III, thèse de doctorat. [En ligne] <https://www.theses.fr/2021BOR30016>
- Lessig, L. (2000). *Code is Law.* Harvard Magazine, [En ligne] <https://lessig.org/product/code>
- Liuzzo, P. M. & Evangelisti, S. (2021). Modeling execution techniques of inscriptions. *Semantic Web*, vol. 12, n°2, 181-190.
- Moberg, C.-A. (1969). *Introduktion till arkeologi: jämförande och nordisk fornkunskap.* Stockholm: Natur och kultur, 228.
- Nonaka, I. & Takeuchi, H. (1997). *La connaissance créatrice.* Paris: De Boeck Université, 304.
- Tchernia, A., Gardin, J.-C., Morel, J.-P., Gros, P., Thébert, Y., Guzzo, P. G., Picard, O., Ozanam, D., Vincent, B. & Pietri, C. (1986). La publication en archéologie. *Mélanges de l'École française de Rome. Antiquité*, 98, n°1, 359-386.
- Tufféry, C. & Augry, S. (2019). *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 [En ligne] <https://hal-inrap.archives-ouvertes.fr/hal-02472817/>

Vygotsky, L. S. (1934). *Thought and Language*. Cambridge : MIT Press (rééd. 2012), 307.

Résumé

Pour identifier, décrire, classer, quantifier, comprendre et interpréter des réalités matérielles anciennes et leurs connexions en vue d'écrire l'histoire des sociétés, l'archéologie manipule des descriptions graphiques et textuelles. La restitution d'un site expose les données collectées, leurs relations (spatiales, temporelles, thématiques) et la nature des raisonnements opérés pour leur analyse, en gardant le souci permanent d'être à l'épreuve des faits. S'agissant du texte, la norme ISO 25964 (gestion des thésaurus) est suffisamment souple pour décrire les différents '*points de vue*' présidant à l'élaboration des '*micro-langages*' spécialisés et évolutifs des archéologues, à la création de concepts terminologiques (par exemple les typo-chronologies) et pour modéliser leur logique, ainsi que pour exposer les données en plus de leurs métadonnées. Projetés sous forme de graphes calculables, le vocabulaire, les données et leurs relations permettent alors de caractériser les processus de transformation et d'organisation des raisonnements et la signification des concepts issus des '*micro-langages*' archéologiques spécialisés.