

Diplôme national de master

Domaine - Sciences humaines et sociales

Mention – Sciences de l'Information et des Bibliothèques

Parcours – Archives numériques

Les enjeux de l'interopérabilité dans la diffusion et la valorisation des données archéologiques

Pauline Vignaud

Sous la direction de Céline Guyon
Maître de conférences associé à l'ENSSIB

Remerciements

Je tiens à adresser mes remerciements à ma directrice de recherche Céline Guyon pour l'attention qu'elle m'a apporté tout au long de l'année et pour ses conseils de lecture avisés.

Mes remerciements s'adressent également à Emmanuelle Morlock (CNRS – Laboratoire HiSoMA) pour sa disponibilité, sa bienveillance à l'égard de mes recherches et nos divers échanges sur la question de l'interopérabilité, à Laure Bézard (archiviste de la Maison de l'Orient et de la Méditerranée à Lyon) pour son aide précieuse sur la question des archives de l'archéologie et à Elisabeth Bellon (responsable du service d'archives de la MAE de Nanterre) pour avoir bien voulu répondre à mes questions.

A ma famille, pour son soutien sans faille tout au long de cette année.

A Romain Bethoule, mon ami.

A Nicolas Grand, pour n'avoir jamais douté de moi.

Enfin, je tiens à remercier mes camarades de classe Cammie Talamona, Fanny Rebillard, Virginie Donat et Céline Barthélémy pour leurs encouragements durant la rédaction.

Résumé : Discipline historique et scientifique, l'archéologie a vu ses pratiques évoluées depuis l'arrivée du numérique. Dès lors, plusieurs problématiques se sont imposées aux archéologues notamment dans leur manière de diffuser et de valoriser leurs données. Dans ce contexte-là, des questions autour de l'interopérabilité ont émergé notamment les outils à développer (plateformes, applications, projets) et à mettre en place pour permettre le partage et la mise en valeur des données archéologiques. Ce mémoire propose d'explorer toutes les thématiques (jeux de données, réutilisation...) où l'interopérabilité intervient dans cet environnement scientifique comme un facteur favorisant - ou problématique dans la diffusion et la valorisation.

Descripteurs : métadonnées, archéologie, interopérabilité, données archéologiques, libres accès, ontologies, thésaurus, valorisation, diffusion, réutilisation, enjeux

Abstract: Archaeology is a field that concerns both history and science. As such, it's logical to observe an evolution of the practices related to the technological progress. Therefore, new problematics emerged, especially around the ways to promote and share the data. In this context, questions towards the interoperability emerged. How to develop and set up tools that will highlight the data and ease its distribution while facing the complexity of interoperability. This thesis will explore all the topics (datasets, reuse ...) in which interoperability is a major stake for the dissemination of archeological data.

Keywords: metadata, archaeology, interoperability, archaeological data, open access, ontology, thesaurus, promotion, diffusion, reuse, issues

Droits d'auteurs

Droits d'auteur réservés.

Toute reproduction sans accord exprès de l'auteur à des fins autres que strictement personnelles est prohibée.
--

Cette création est mise à disposition selon le Contrat : « **Paternité-Pas d'Utilisation Commerciale-Pas de Modification 4.0 France** » disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr> ou par courrier postal à Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

SOMMAIRE

<i>Sommaire</i>	5
Sigles et abréviations	6
INTRODUCTION	7
ARCHIVER, DIFFUSER ET VALORISER L'ENFOUI : ENJEUX, CONTRAINTES ET NOUVELLES PERSPECTIVES.....	11
Du sol au hors sol : archiver l'absence	11
Qu'est-ce qu'une donnée archéologique ?.....	11
Les archives des archéologues.....	14
Normes et réglementation.....	17
Diffuser, valoriser... et archiver	20
Diffusion et valorisation : enjeu, nécessité ou obligation ?.....	20
L'archivage des données archéologiques.....	23
<i>Open Access/Open Data</i> : les projets de l'archéologie.....	25
L'archéologie numérique : de nouvelles possibilités	27
UTILISER ET REUTILISER LES DONNEES ARCHEOLOGIQUES : UNE QUESTION D'INTEROPERABILITE ?.....	32
L'importance des métadonnées en archéologie	32
Des normes et des standards pour l'interopérabilité des données et des métadonnées	36
Le vocabulaire : du bon usage des thésaurus.....	36
Ontologie(s) : l'application du CIDOC-CRM	40
Bases de données et redocumentarisation.....	44
Réutiliser les données et les métadonnées : l'unique défi de la redocumentarisation ?	44
Les bases de données en archéologie : l'interopérabilité au service de la diffusion et de la valorisation	48
PRECONISATIONS ET BONNES PRATIQUES : DU PLAN DE GESTION DES DONNEES ET DES DIFFERENTS NIVEAUX DE L'INTEROPERABILITE A L'OPEN ACCESS.....	53
Créer un plan de gestion des données.....	53
Garantir les différents niveaux d'interopérabilité.....	58
Systématiser l'ouverture des données et le libre accès	62
CONCLUSION.....	65
BIBLIOGRAPHIE.....	67
ANNEXES.....	72
TABLE DES MATIERES.....	96

Sigles et abréviations

AAF	Association des archivistes français
CCSD	Centre pour la communication scientifique directe
CINES	Centre informatique national de l'Enseignement supérieur
COMETS	Comité d'éthique du CNRS
CMS	<i>Content management system</i>
CN2SV	Centre national pour la numérisation de sources visuelles
CNRS	Centre national de la recherche scientifique
DISIC	Direction interministérielle des systèmes d'informations et de communication
FRANTIQU	Fédération et ressources sur l'Antiquité
GED	Gestion électronique des données
HAL	Hyper article en ligne
INRAP	Institut national de recherche en archéologie préventive
ISO	<i>International Standard Organization</i>
LIMC	<i>Lexicon Iconographicum Mythologiae Classicae</i>
MAE	Maison de l'archéologie et de l'ethnologie René Ginouvès
MASA	Mémoires des archéologues et des sites archéologiques
MMSH	Maison méditerranéenne des sciences de l'homme
MOM	Maison de l'Orient et de la Méditerranée
MSH	Maison des sciences de l'Homme
OAI	<i>Open Archival Initiative</i>
OAIS	<i>Open Archival Information System</i>
OCDE	Organisation de coopération et de développement économiques
PGD	Plan de gestion des données
UMR	Unité mixte de recherche
RGI	Référentiel général d'interopérabilité
RDF	<i>Resource Description Framework</i>
SHS	Sciences humaines et sociales
SIA	Système d'information archéologique
SIG	Système d'information géographique
SKOS	<i>Simple Knowledge Organization System</i>
W3C	<i>World Wide Web Consortium</i>

INTRODUCTION

L'interopérabilité des données, permise par le développement du web sémantique, semble, depuis quelques années, amorcer un mouvement de convergence des pratiques dans la description des données. L'interopérabilité occupe une place importante dans les thèmes d'étude des recherches en Humanités numériques et de manière générale dans les Sciences humaines et sociales. « L'interopérabilité des images pour la recherche en Histoire de l'art » fut, d'ailleurs, le thème de la rencontre de la *Digital Art History* à l'Institut national d'histoire de l'art le 13 et 14 juin 2016 à Paris. Cette présentation lors de cette journée d'étude avait pour but de montrer l'évolution actuelle des différents types de ressources et de donner la parole à de jeunes chercheurs en histoire de l'art et en archéologie investis dans le domaine de la recherche numérique ; preuve que la question de la capacité des machines, logiciels ou formats à rendre les données interopérables pour être diffusées, valorisées, archivées voire même uniquement lisibles cristallise bien des enjeux. L'une des interventions, présentée par Clément Salviani (INHA) et Thomas Sagory (Musée d'Archéologie nationale - Domaine national de Saint-Germain-en-Laye), s'est articulée autour du thème suivant : « *Digital Archaeology* : les outils du digital et du numérique dans les processus métiers de l'archéologie, de l'acquisition de terrain à la conservation ultime des vestiges anciens », au cours de laquelle les deux intervenants soulèvent un point intéressant : quel nouveau regard le chercheur peut-il porter sur sa donnée ? Cette interrogation montre bien à quel point les enjeux recouverts par l'interopérabilité et l'informatisation des méthodologies de travail en archéologie préoccupent et laissent le sujet ouvert à de vastes questionnements.

Le Consortium MASA s'est aussi emparé du sujet en organisant des journées d'étude du 20 et 22 novembre 2017 à Tours, sur les « Interopérabilités ». Celles-ci sont le témoignage direct que les interopérabilités, et non pas seulement l'interopérabilité, constituent un enjeu de réflexion actuel important. L'interopérabilité permet l'échange d'informations ou de données d'un système à un autre sans aucune perte ou restriction. C'est la faculté pour plusieurs systèmes de communiquer entre eux sans aucun autre acteur particulier tout en se reposant sur la présence de standards ouverts et non propriétaires. Si aujourd'hui, l'interopérabilité se retrouve au centre des préoccupations de la communauté scientifique, et de surcroît, de la communauté archéologique c'est pour répondre à des enjeux de valorisation et de diffusion des données produites lors des campagnes de fouilles. L'utilisation des nouvelles technologies en archéologie est venue faire évoluer et modifier les pratiques des

professionnels et des chercheurs, au point que désormais numérique et archéologie semblent indissociables. Si autrefois la profession d'archéologue était envisagée – au sein même du corps de métier – comme incompatible avec les nouvelles technologies, celle-ci ne peut plus se détacher des nouveaux outils, de plus en plus perfectionnés, dont elle a besoin (instruments de mesure pour les relevés topographiques, appareil pour la photographie, ordinateur pour l'enregistrement des données, logiciels DAO pour le traitement et l'exploitation des données, LiDAR aéroporté par un drone, etc...). Ainsi, les nouvelles technologies ont directement influencé les méthodes d'acquisition des données par les archéologues. Ces données sont avant tout des informations qu'on cherche sans cesse à recontextualiser avec le plus d'exhaustivité possible pour comprendre un événement ou un fait établi et cela est d'autant plus dur que l'archéologie étudie directement le passé de l'humanité, des civilisations, des cités et de tous vestiges laissés par l'homme au cours des siècles.

Traiter le sujet des données archéologiques renvoie de manière intrinsèque au métier d'archéologue. L'archéologie est une discipline scientifique, dont les principaux objets sont les sociétés du passé, au travers de l'étude des vestiges matérielles de l'homme, découvertes essentiellement lors de fouilles ou de récupération. Trois outils conceptuels constituent les bases de l'archéologie : la chronologie, la typologie et la stratigraphie¹. Par définition, l'archéologie est une science destructrice, c'est l'une des rares sciences à détruire ses objets d'étude tout en les construisant. En effet, l'obtention de vestiges archéologiques nécessite une fouille stratigraphique provoquant la destruction systématique de couches de terre dans le sol, l'objectif étant d'exhumer les restes d'une occupation humaine. La construction d'une méthode efficace d'enregistrement des données produites et du contenu archéologique est au centre des préoccupations des archéologues et des professionnels de l'information spécialisés en archéologie. Au fil du temps, les nouvelles technologies se sont imposées comme les SIG, les croquis à main levée, la reconstitution 3D, les dessins de céramiques ou de petits mobiliers et la photographie. L'objectif de ces technologies est de garder une trace des structures au-delà des simples relevés de plans manuels. En fouillant, l'archéologue détruit et crée le contexte de prélèvement des données, sans aucun retour en arrière possible, que ce soit lors de fouilles archéologiques préventives ou programmées. La donnée est récupérée au

¹ La stratigraphie est la science étudiant les différentes couches de terre présentes dans le sol. Les unités stratigraphiques sont des découpages proposés pour définir des séquences de dépôt. Celles-ci s'évaluent par rapport à leur couleur, leur modalité (sec, humide...), leur porosité, leur texture, leur pourcentage en terre, calcaire, sodium... Les strates se superposent les unes aux autres dans un terrain. L'observation des agencements permet d'établir des rapports d'antériorité, de postériorité et de contemporanéité entre les différentes unités. L'objectif est de pouvoir proposer une chronologie relative des données rapportées.

moment même où l'archéologue la découvre, de ce fait la documentation et le support de récolte des données archéologiques sont malheureusement soumis à la dégradation du temps (cahiers de fouilles, plan, dessins, croquis, relevés au crayon à papier ou au stylo sur calque...). Cette documentation est, dans bien des cas, peu pérenne et sa consultation excessive la fragilise toujours un peu plus. Les chantiers de fouilles sont souvent des sites isolés, accessibles uniquement à pied ou en voiture, et bien souvent la documentation est entreposée sur place, restant sujette à de nombreux risques de dégradation. De ce fait, depuis les années 1980, les archéologues ont pris l'habitude de travailler avec des bases de données non-relationnelles afin de pouvoir répertorier et compiler leurs données. Cette pratique était toujours laissée au bon vouloir des procédés de l'archéologue sans réel protocole. Les choix méthodologiques dans leur constitution et leur documentation restaient et restent encore aujourd'hui subjectifs, posant de nombreux problèmes de compréhension. En effet, ce manque d'homogénéisation des pratiques en archéologie complique souvent la lecture des données. L'illisibilité de certaines données archéologiques perturbe les processus de travail des archéologues et des chercheurs, rendant complexe leur compréhension. Dans tout travail de diffusion et de valorisation des données, le renseignement des métadonnées est essentiel, sans lui les informations restent brutes et n'ont aucune vraie valeur.

La diffusion et valorisation des données archéologiques semblent revêtirent plusieurs enjeux. Enjeu réglementaire ou enjeu scientifique, valoriser et diffuser les données de l'archéologie constitue le centre des préoccupations des chercheurs. A l'heure des *Big Data* et de l'influence du numérique dans les pratiques, la diffusion et la valorisation de la production scientifique interrogent les professionnels de l'archéologie et des sciences de l'information. Programmes européens, plateformes de partage, de publication ou de dépôt, archives ouvertes ou solution d'archivage pérenne, les premières réponses face à un tel bouleversement se multiplient depuis une décennie. Nombreux sont ceux qui s'engagent dans ce problème et alimentent le débat, l'un de ces premiers acteurs est d'ailleurs l'Etat français lui-même. En effet, selon les instructions du ministère de la Culture et de la Communication, la diffusion est indissociable de la valorisation. Le partage des résultats de recherches avec la communauté scientifique et le grand public est obligatoire. Les différentes publications (ouvrages, articles, monographies, sites internet, etc...) éditées ou coéditées par les organismes de recherche (CNRS, INRAP...) et le ministère sont les premières formes de cette diffusion et de contribution à la valorisation. S'adaptant au public visé, la publication des données archéologiques dépendra de son lecteur. Ainsi, un atlas historique servira tout autant à un archéologue ou historien qu'à un gestionnaire du territoire ou un topographe ; un

colloque permettra aux chercheurs d'échanger et de présenter des résultats sur une étude ; et une exposition ou un ouvrage de vulgarisation favorisera la diffusion et la mise en valeur des découvertes archéologiques au grand public. Si la forme de publication influe directement sur les modes de partage des connaissances, diffusion et valorisation se complètent aisément.

Le développement des sciences de l'information et du web sémantique ouvre depuis quelques années un champ de perspectives encore inégalées jusqu'à aujourd'hui pour la recherche en archéologie en terme de volume de données compilées, étudiées et d'interopérabilité entre silos issus de différents projets ou institutions. Une discipline extérieure peut en influencer une autre sur ses méthodologies de travail et être porteuse de réflexions intéressantes. Interopérabilité et traitement des données archéologiques sont donc bien compatibles, mais dans quelles mesures les enjeux portés par l'interopérabilité peuvent-ils venir bouleverser l'archéologie dans sa capacité à diffuser et valoriser ses données ? Pourquoi l'adoption de formats standards représente-t-elle un enjeu si important pour la communauté archéologique ? En quoi la cristallisation de ces enjeux peut-il favoriser la diffusion des contenus ? Les enjeux recouverts par l'interopérabilité peuvent-ils être une source d'innovations dans les méthodologies de recherche en archéologie ?

Afin de répondre à ces questions, il semble essentiel de revenir, dans un premier temps sur les significations même de la valorisation, la diffusion et la documentation en archéologie. Sans ce premier développement, il semble que ce sujet ne puisse être abordé convenablement. A partir de ce point de départ découlera le fil conducteur suivant : démontrer pourquoi il est essentiel de rendre les données interopérables pour la compréhension des données archéologiques dans leurs principes de diffusion et de valorisation, et ce, peu importe la structure d'étude (institution, laboratoire de recherche, université...). Une deuxième partie, posera la question de la place de l'interopérabilité dans l'utilisation et la réutilisation des données, en abordant le thème des métadonnées, des ontologies, des thésaurus, de la redocumentarisation et des bases de données. Enfin, dans une dernière partie, plus exploratoire, nous tenterons de répondre à ces questions en proposant quelques préconisations et les bonnes pratiques à mettre en œuvre et les solutions apportées (plan de gestion des données, utilisation de normes et standards, ouverture des données et de leurs accès) nécessaires pour la diffusion et la valorisation.

ARCHIVER, DIFFUSER ET VALORISER L'ENFOUI : ENJEUX, CONTRAINTES ET NOUVELLES PERSPECTIVES

DU SOL AU HORS SOL : ARCHIVER L'ABSENCE

En archéologie, le double paradoxe de la destruction et de la création de la donnée et son contexte pose une première contrainte : leur traitement une fois la collecte terminée. Sur support physique ou sur support numérique, les données archéologiques suivent le même cycle de vie que n'importe quelle donnée, de leur création à leur fin d'usage courant en passant par leur utilisation et leur validation, le processus reste le même. Dès lors, avant de procéder à son archivage, une seule question se pose : qui a produit la donnée ? Archivistiquement, les données produites et collectées lors de chantiers de fouilles archéologiques dépendent du domaine du public puisque que son producteur, l'archéologue, constitue une personne physique rattachée à un centre de recherche ou un institut. Néanmoins, plusieurs difficultés, soulevées par ce double paradigme de donnée de la recherche, au statut public, découverte par une personne physique, soulèvent quelques questions autant d'ordre épistémologique que d'ordre normatif et législatif.

Qu'est-ce qu'une donnée archéologique ?

Une donnée archéologique est avant tout une donnée de la recherche mais selon les approches et les contextes, les définitions varient. Il y a une dizaine d'années, l'OCDE définissait les données de la recherche comme « [...] *des enregistrements factuels (chiffres, textes, images, sons) utilisés comme sources principales pour la recherche scientifique et généralement reconnus par la communauté scientifique comme nécessaires à la validation des résultats de recherche* »². Des principes et lignes directrices pour l'accès aux données financées sur fonds publics furent éditées. L'intérêt était sensiblement tourné vers l'amélioration du partage des données de la recherche. Cette première définition classait les données de la recherche comme des « [...] *enregistrements factuels* [...] », des captations de la réalité retranscrite sur différents supports : numérique (sons,

² OCDE. Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics [en ligne]. 2007, [consulté le 14 juin 2018]. Disponible sur le Web : < www.oecd.org/fr/sti/sci-tech/principesetlignesdirectricesdelocdepourlaccessauxdonneesdelarecherchefinancesurfondspublics.htm >.

images, relevés topographiques) et papier (fiches US, dessin de mobilier ou de tesson de céramique). Puis, quelques années plus tard en 2013, la loi sur l'Enseignement supérieur et la recherche vient modifier l'article L 112-1 du Code de la recherche en introduisant la notion d'« [...] organisation de l'accès libre aux données scientifiques³ ». Les données scientifiques et la question de leur libre accès étaient désormais inscrites dans la loi française. Reconnu par l'Etat, le statut des données de la recherche ne tarda pas à prendre place au sein de la loi pour une République numérique promulguée le 7 octobre 2016. Elle mentionne, à l'article 38 chapitre II *Economie du savoir*, en reprenant le Code de la propriété intellectuelle modifié, que les « [...] fichiers produits au terme des activités de recherche pour lesquelles elles ont été produites ; ces fichiers constituent des données de la recherche ». Cette définition est actuellement, l'une des rares mentions dans la loi française des données de la recherche en tant que telles. Celles-ci sont reconnues, bien qu'aucun autre texte ne spécifie encore de règles sur leur gestion, leur conservation ou leur traitement.

En Europe, le statut des données de la recherche fut défini par la Commission européenne. Dans la partie *Open Data and Data Management* du programme Horizon 2020⁴, l'accent est mis sur l'importance du renseignement des métadonnées :

« 1. The « underlying data » (the data needed to validate the results presented in scientific publications), including the associated metadata (i.e. metadata describing the research data deposited), as soon as possible;

2. Any other data (for instance curated data not directly attributable to a publication, or raw data), including the associated metadata, as specified and within the deadlines laid down in the DMP – that is, according to the individual judgement by each project/grantee⁵.

[...] In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recording and images. The focus is on research data that is available in digital form »⁶.

Cette définition, à l'échelle européenne, des données de la recherche rejoint celle donnée en 2017, par la section « Aurore » de l'AAF⁷. L'association décrit les données de la

³ Loi n° 2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche.

⁴ Horizon 2020 est un programme européen pour la recherche, le développement et l'innovation pour la période de 2014-2020.

⁵ European Commission. *H2020 Program Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, version 3.1, 25 août 2016, p.9.

⁶ Research & Innovation. Participant Portal H2020 Online Manual [en ligne]. [Consulté le 18 juin 2018]. Disponible sur le Web : < http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm >.

⁷ La section « Aurore » de l'Association des archivistes français regroupe les archivistes des universités, des rectorats, des organismes de recherche et mouvements étudiants.

recherche comme des informations, des spécimens, des matériaux produits, recueillis, documentés, collectés et exploités à des fins de recherche et de preuve par des chercheurs et des équipes, constituant, à ce titre, une partie des archives de la recherche⁸.

Selon François Djindjian, les données archéologiques sont « [...] *des informations variées résultant de la description d'objets ou d'ensembles d'objets archéologiques, ou d'enregistrements effectués à l'occasion d'une prospection ou d'une fouille archéologique, ou de compilations faites pour l'édition de grands corpus documentaires ou de sélections tirées de banques de données* »⁹. Ce sont donc des données de la recherche produites dans le cadre d'une activité de fouilles, sondages ou prospections archéologiques, menée par une personne physique (chercheur, archéologue) ou morale (institut, université, société d'archéologie préventive...). Néanmoins, les données archéologiques ne doivent pas être confondues avec les objets qu'elles décrivent, représentent ou imagent¹⁰. Il faut donc distinguer un mur de sa description, une céramique mise en sachet de sa représentation. A l'INRAP, on parle volontiers de données de terrain, de données de mobilier et de données documentaires. Les données de terrain décrivent toutes les données collectées sur un chantier de fouille archéologique, ce sont généralement les éléments observés par les archéologues et non conservés. Les données de mobilier désignent tous les éléments matériels recueillis lors d'une opération (ossement humain ou animal, céramique...). Elles font l'objet d'un catalogue où on procède parfois à un échantillonnage pour renseigner un lot. Enfin, les données documentaires sont l'ensemble des documents produits lors de la fouille archéologique (calque, plan, fiche US, fichier FileMaker...). Ces données possèdent toutes une part d'interprétation puisqu'elles ont été retranscrites de manière subjective, ces documents sont porteurs de données et de systèmes permettant d'exploiter les données¹¹. Difficiles à classer ou à nommer selon Anne Speller, les données archéologiques numériques ont semé la confusion dans la définition des archives des archéologues. Transmises uniquement sous forme documentaire (photographies, relevés topographiques, photogrammétrie¹², dessins,

⁸ Archivistes ! n°120 [en ligne]. 30 janvier 2017, [consulté le 14 juin 2018]. Disponible sur le Web : < www.archivistes.org >.

⁹ DJINDJIAN, François. *Manuel d'archéologie. Méthodes, objets et concepts*. Paris : Armand Colin, « U », 2011, ISBN : 9782200266769.

¹⁰ SPELLER Anne et COLIN Camille. *Quelles données pour quels projets : perspectives de l'Open Access pour l'archéologie préventive* [en ligne]. 23 mars 2017, [consulté le 21 juin 2018], mis à jour le 30 novembre 2017, 21:31 minutes. Disponible sur le Web : < <https://www.inrap.fr/quelles-donnees-pour-quels-projets-perspectives-de-l-open-access-pour-l-13180> >.

¹¹ En archéologie préventive, à l'INRAP notamment, le niveau de précision des données est placé sous la responsabilité d'un titulaire d'opération contrôlé par les services de l'Etat.

¹² La photogrammétrie est une technique permettant de déterminer les volumes et les dimensions d'un objet à partir de mesures effectuées sur des photographies montrant les perspectives de l'objet.

statistiques, notes, numérisations, modélisations 3D, bases de données, prélèvements, analyses de terrain...), les données archéologiques semblent pourtant être une partie intégrante des archives de l'archéologie.

Les archives des archéologues

Qu'est-ce qu'une archive d'archéologue ? Au sein de la communauté scientifique, le mot « archives » est au cœur de toutes les attentions. Tantôt évincée des conversations, tantôt usée de manière abusive, l'utilisation du mot « archives » ne semble pas être strictement représentative de la réalité et pose une certaine ambiguïté. En effet, une grande partie des chercheurs et les archéologues tendent à penser que l'intégralité de leur production documentaire constitue des archives définitives alors qu'en réalité, très peu d'entre elles atteindront cet âge-là. Emmanuelle Morlock, ingénieure d'étude CNRS rattachée au laboratoire HiSoMA et responsable de systèmes d'information documentaire, en charge du programme « Réflexions sur les pratiques d'édition et d'archivage » de l'axe transversal « Edition, archives, humanités numériques », parle plus volontiers de production documentaire que d'archives. Elle explique : « *Il existe un vrai problème avec ce mot. Les chercheurs l'utilisent massivement pour désigner uniquement les sources primaires, les sources « brutes » mais du point de vue archivistique, cela ne correspond pas. En réalité, les chercheurs produisent des documents, ils font principalement de la production documentaire et créent des archives courantes voire intermédiaires, [...] »*¹³. Sa réflexion vient confirmer les obstacles terminologiques auxquels se bloque la distinction entre données archéologiques et archives de l'archéologie.

Les archives de l'archéologie regroupent l'idée d'une production documentaire liée à une opération archéologique, préventive ou programmée et toutes pièces administratives s'y rapportant. Ainsi, les archives de l'archéologie sont des archives majoritairement publiques tout en étant soumises au droit d'auteur. La diversité de production, de supports et de cycle de vie, conjuguée à une gestion dépendant de nombreuses institutions et tutelles, compliquent un peu la prise en charge des fonds par des archivistes. Les archives de l'archéologie sont pourtant considérées comme des sources essentielles pour l'étude d'un site, d'un monument, d'une période historique mais aussi pour la révision des données.

¹³ Voir Annexes, Entretien avec Emmanuelle Morlock.

Définir les archives de l'archéologie est un exercice difficile auquel se sont adonnées en 2015, Marie Stahl et Lucie Schirr¹⁴ dans un article consacré à un état de l'art de la gestion des archives de l'archéologie en France. Leurs essais de définition permettent de mieux comprendre à quel type de documentation l'archiviste doit traiter. Elles définissent les archives de l'archéologie comme « [...] *une production individuelle (réalisée par un archéologue) dans un cadre collectif (pour un opérateur ou au sein d'une unité de recherche), la notion d'archives de l'archéologie peut être définie de manière différente d'une tutelle à l'autre, ou selon le cadre de réalisation de l'opération : préventif ou programmé.* ». Ainsi, les archives de l'archéologie sont des documents produits et reçus dans le cadre d'une recherche archéologique dont on doit contextualiser le fonds, de la création au versement, quelle que soit l'aire géographique, le type d'opération ou la tutelle dont dépend la structure bénéficiaire de la fouille. La tentative de définition des deux auteures s'inscrit dans la continuité des essais de Thérèse Charmasson, pionnière dans les archives scientifiques, qui sut différencier les archives des sciences des archives scientifiques en s'intéressant notamment aux archives des centres de recherche.

Les archives conservent l'idée de la production d'une structure. En effet, lors d'une opération archéologique (figure 1), les données mises à jour sont parfois détruites au fur et à mesure que la fouille ou que l'étude se poursuit. C'est donc le document qui constitue le seul lien, la seule information en lien avec les données récoltées. Cela donne aux archives de fouilles le statut *d'unica* c'est-à-dire que celles-ci sont uniques étant donné que le retour vers le terrain, détruit, est désormais impossible. Les données collectées sont généralement iconographiques et non explicites, elles nécessitent l'interprétation de l'archéologue pour être lisibles et compréhensibles, sans cela la donnée est juste une information.

L'idée d'archives de l'archéologie implique l'idée de document (support – papier ou numérique - sur lequel on inscrit une information) alors que la donnée évoque un caractère plus « brut » et recouvre l'idée de la description élémentaire d'une réalité (une observation,

¹⁴ STAHL Marie et SCHIRR Lucie. « Les archives de l'archéologie : définition, législation, état des lieux », in *Archimède : archéologie et histoire ancienne*, 2015, p.9-19.

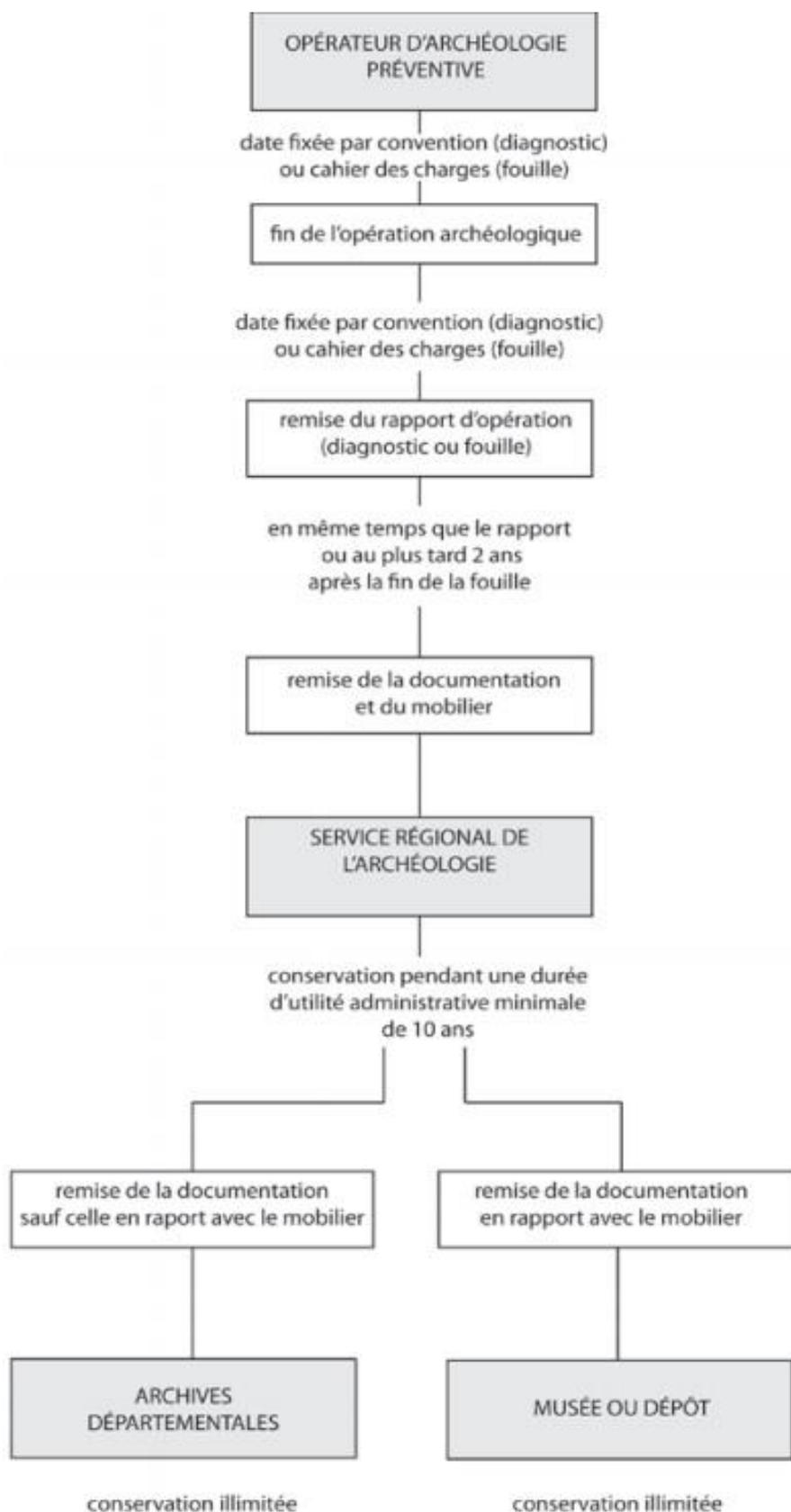


Figure 1 *Étapes du traitement des données archéologiques lors d'une fouille préventive.*
 Source : STAHL Marie et SCHIRR Lucie. *Les archives de l'archéologie : définition, législation, état des lieux, Archimède : archéologie et histoire ancienne, 2015, p.13.*

une mesure). Il est intéressant de questionner le terme de « donnée » et de l'opposer à d'autres terminologies afin de comprendre pourquoi cela pose problème en archéologie. Certains spécialistes parlent plus aisément de « construits »¹⁵ ou même de « capta »¹⁶. Ainsi le « construit » résulte d'une construction de l'esprit face à une information et le « capta » d'une information capturée parce qu'elle est conforme aux règles et aux hypothèses établies pour l'expérience, alors que la donnée est considérée comme une information objective. Ainsi, l'information archéologique n'est pas une donnée, mais le résultat d'un processus interprétatif. Pourquoi la donnée archéologique ne pourrait-elle pas plutôt être une « capta » ? Celle-ci est bien recueillie conformément à des règles (dans des unités stratigraphiques sur un chantier de fouilles) et vient le plus souvent confirmer qu'infirmer des hypothèses... Pourtant, la donnée archéologique est bien une donnée et non une « capta » et revêt même plusieurs catégories, ainsi la donnée archéologique peut être : une donnée d'observation (fouilles archéologiques) et/ou une donnée dérivée ou compilée (on entend ici les bases de données compilées, les corpus de sources textuelles et en soi un ensemble d'inscriptions peut en contenir des sources textuelles). Mais la donnée archéologique n'est en aucun cas une donnée reproductible, donc elle n'est pas une donnée expérimentale, ni une donnée computationnelle ou de simulation et ni une donnée canonique ou de référence. La donnée suit donc un continuum, un cycle de vie bien précis et est agrémentée de métadonnées. Leur gestion et leur conservation sont d'ailleurs réglementées par des textes.

Normes et réglementation

La réglementation relativement récente (moins de 30 ans) sur les archives de fouilles semble avoir impacté la manière dont la documentation archéologique est appréhendée et de ce fait archivée ou conservée. Les archives de l'archéologie dépendent encore une fois de leur producteur, l'archéologue qui gère, fouille, étudie et met en valeur des vestiges et des

¹⁵ C'est le sociologue des sciences Bruno Latour qui propose ce terme. Selon lui, les objets d'études scientifiques sont « socialement construits » dans les laboratoires ou centres de recherche et n'ont aucune existence en dehors des instruments qui les lisent et des spécialistes qui les interprètent. Ainsi le relevé topographique effectué par un topographe, interprété par un archéologue, n'a pas de vraie valeur sans une machine et une conscience réflexive capable de donner une signification. Cela ferait des données des objets interprétés de manière subjective et donc biaisés puisqu'on leur cherche sans cesse du sens.

¹⁶ Johanna Drucker met ici en opposition le terme « capta » et « data ». Le « capta » est la captation d'une information tandis que les données sont supposées « être données » et donc enregistrées et observées. Il faut, selon elle, reconnaître le caractère localisé, partiel et constitutif de la production du savoir et de la construction de la connaissance. Une donnée est par essence, une représentation naturelle d'un fait préexistant. Or en archéologie, une donnée est nécessairement une interprétation du réel par un archéologue dans un contexte de découverte défini.

sites archéologiques conformément à une réglementation et aux politiques d'aménagement et de développement culturel du territoire. De ce fait, l'archéologie est une discipline du service public, travaillant sur des vestiges du patrimoine historique, les archives des archéologues – bien que soumises au droit d'auteur pour certains documents (plans, cartes...) – sont des archives publiques d'abord encadrées par plusieurs textes législatifs :

- le Code du Patrimoine, Livre II ;
- la circulaire du 2 novembre 2001 traitant de la gestion des archives dans les services et établissements publics de l'Etat ;
- la circulaire DGP/SIAF/2010/020 du 25 novembre 2010 sur le contrôle et la collecte des archives des opérateurs de l'Etat ;
- la circulaire MCC1106465C du 1^{er} avril 2011 accompagnée de la note du 15 avril 2011 relative à la fonction archives.

Les textes législatifs les plus importants restent ceux concernant la gestion des archives issues de l'archéologie préventive :

- le Code du Patrimoine, livre V (articles L523-11, L523-12 et R523-66 à 5523-66) ;
- la circulaire du 5 février 1993 sur les obligations liées à l'achèvement d'une fouille archéologique préventive, notamment sur la documentation et le document final de synthèse (DFS) ;
- l'instruction DPACI/RES/2004/019 du 21 septembre 2004 sur le tri et la conservation des archives des Directions régionales des affaires culturelles (DRAC) ;
- l'instruction DAF/DPACI/RES/2009/028 du 16 décembre 2009 relative au traitement et à la conservation des archives de l'INRAP.

Quant aux archives de l'archéologie produites dans un établissement d'enseignement supérieur et de la recherche (dans les services universitaires notamment), celles-ci sont régies par une législation encore différente de celle de l'archéologie préventive :

- l'instruction DPACI/RES/2007/002 du 15 janvier 2007 relative au traitement et à la conservation des archives des délégations du CNRS et des archives des unités de recherche et de service (laboratoires de recherche en SHS, UMR...);
- l'instruction DPACI/RES/2005/003 du 22 février 2005 relative au tri et à la conservation des archives reçues et produites par les services et établissements concourant à l'éducation nationale.

De manière générale les archives de l'archéologie, qu'elles soient produites ou reçues dans le cadre d'un chantier de fouille programmée ou préventive, sont de nature administrative, publique et par conséquent communicable. Néanmoins, les archives archéologiques de la recherche étant constituées, pour certaines, dans le cadre d'un travail personnel fourni par le chercheur, celles-ci sont tout de même soumises et protégées par le droit d'auteur. Le rôle des producteurs est donc à prendre en compte afin de distinguer les droits moraux et les droits patrimoniaux qui sont, eux, liés à la paternité du document reçu ou créé. Ainsi, le Code de la propriété intellectuelle s'applique aussi aux archives de l'archéologie. Le service public donne un véritable cadre en ce qui concerne les possibilités de retrait et de modification apportés aux documents. Les droits patrimoniaux quant à eux s'appliquent la plupart du temps au personnel de recherche de l'enseignement supérieur (chercheurs, archéologues, responsables de secteur, etc...). De ce fait, bien que parfois méconnues chez les professionnels des archives, les archives de fouilles répondent à une législation belle et bien existante.

Les détenteurs d'archives archéologiques sont donc soumis à différentes actions, souvent négligées : un devoir d'inventaire et une obligation de versement. Mais bien qu'il existe toute une législation, les archives de l'archéologie restent une documentation très particulière. Cela réside sans doute dans leur caractère variable, du fait que chaque documentation est propre à chaque site, à chaque situation, à chaque contexte géographique, climatique et historique dans lequel il a été fouillé ainsi qu'à la personnalité des responsables de la fouille et des fouilleurs, voire même de l'organisme qui a été chargé des opérations.

Quant aux normes appliquées à la documentation de l'archéologie, elles sont peu nombreuses. Les arrêtés MCCB0400762A et MCCB0400702A du 16 septembre 2004 et 27 septembre 2004 sont en réalité deux normes bibliographiques des Documents d'archéologie française (DAF) portant sur le contenu et la présentation des rapports d'opérations

archéologiques. Ces normes sont surtout utiles pour la publication des données et par extension leur diffusion auprès du public ou de la communauté scientifique.

La documentation des données archéologique que ce soit lors de leur archivage, leur diffusion et leur valorisation est primordiale pour garantir, par la suite, leur interopérabilité. Chacune de ses étapes ne peut se dissocier l'une de l'autre puisque chacune d'entre elles nécessite, à terme, l'action de l'autre pour pouvoir se réaliser : valoriser, c'est contribuer à la diffusion de contenus et diffuser, c'est permettre la mise en valeur d'un résultat.

DIFFUSER, VALORISER... ET ARCHIVER

La diffusion et la valorisation des résultats de recherche répondent, la plupart du temps, à des enjeux et éléments d'analyses juridiques, éditoriaux, réglementaires et de politique scientifique. Diffuser et valoriser répondent donc à des enjeux stratégiques de partage du savoir. Dans un contexte archivistique strict, la valorisation et la diffusion viennent clore le processus d'archivage. Cependant, en archéologie, ces deux étapes interviennent souvent avant. Le chercheur diffuse généralement ses résultats de recherche avant d'envisager – souvent plusieurs années après – l'archivage de ses données.

En 2013, l'Université Paris Sud ouvrait une formation aux doctorants sur la « Diffusion et la valorisation des productions de la recherche doctorale » avec pour objectif de sensibiliser les jeunes chercheurs à penser le mode de diffusion et de valorisation de leur production scientifique. Priorités de la communauté scientifique, la diffusion et la valorisation des données se feraient-elles au dépend de leur archivage ?

Diffusion et valorisation : enjeu, nécessité ou obligation ?

La diffusion des résultats de la recherche scientifique constitue bel et bien une obligation réglementaire, morale et déontologique pour tout chercheur ou archéologue. Le rapport final d'opération est le premier document à devoir être diffusé au sein de la communauté scientifique de la région où les fouilles ont eu lieu. Ce document administratif, scientifique et patrimonial présentant les résultats d'une opération et marquant son achèvement est réglementairement structuré par l'arrêté du 24 septembre 2004 qui prévoit

une structure en trois sections¹⁷. Premier document à devoir être rendu public et diffusable, sa publication est souvent suivie d'autres documents tout aussi importants. Les monographies ou articles de recherche à destination d'amateurs ou de professionnels de centres de recherche interrégionaux forment, à la suite, un premier prolongement à ce rendu. Ces documents aident ensuite à la publication de synthèses dans des revues ou sur des sites d'intérêt national ou international. Ces publications de synthèse sont généralement coéditées en partenariat avec le CNRS, le ministère chargé de l'Enseignement Supérieur et de la Recherche, l'INRAP et la fondation de la MSH. Pour la diffusion des données archéologiques à destination du grand public, le département de l'Enseignement supérieur de la recherche et de la Technologie du ministère de la Culture et de la Communication collabore avec les Editions du Patrimoine à la publication de guides d'archéologie. La diffusion des résultats de recherche répond d'abord à une obligation réglementaire inscrite dans la loi française.

De manière générale, un chercheur peut diffuser ses données de plusieurs façons. La première consiste à publier ses travaux de recherche dans une revue scientifique avec un comité de lecture ou dans un ouvrage chez un éditeur. Par la suite, la publication est faite sur un support numérique et un support papier. La diffusion est ensuite laissée au libre choix du chercheur qui le destine à un large public ou à une circulation plus restreinte. La deuxième voie de diffusion est celle d'Internet avec les revues en libre accès et les archives ouvertes. Depuis quelques années, la tendance est à la publication en ligne des données via les archives ouvertes thématiques (ArXiv) et pluridisciplinaires (HAL)¹⁸. Dans ce cas précis, les archives ouvertes ne sont pas considérées comme un processus d'archivage, mais comme un mode de diffusion. L'avantage est de pouvoir rendre communicables, accessibles gratuitement et rapidement les données ; l'inconvénient est la pauvreté des métadonnées renseignées. On distingue alors les post-publications correspondant à des travaux validés et relus et le dépôt de préparations de publications, des compte-rendus de colloques ou des avis. Les archives ouvertes sont considérées, par le COMETS, comme jouant un « [...] *un rôle essentiel dans la diffusion totalement gratuite des résultats de la recherche ou ouvrant d'une part un grand espace de liberté [...]. Elles présentent le grand avantage de permettre un accès rapide aux connaissances et aux résultats récents pour l'ensemble des communautés, sans limitation de type financier, et peuvent favoriser la diffusion de la production scientifique [...]* ». Il ne faut donc pas confondre les archives ouvertes permettant la diffusion des données de la recherche et les archives institutionnelles, lieu de dépôt légal des archives d'institut de recherche ou

¹⁷ INRAP. Rapport d'opération [en ligne]. 12 janvier 2016, mis à jour le 2 décembre 2016, [consulté le 18 juin 2018]. Disponible sur le Web : < <https://www.inrap.fr/rapport-d-operation-9855> >.

¹⁸ COMETS. Avis Réflexion éthique sur la diffusion des résultats de recherche, 1^{er} mars 2007, p. 8-9.

d'université. La plateforme MédiHAL¹⁹ est un des exemples les plus concrets de l'utilisation des archives ouvertes comme mode de conservation, valorisation et diffusion des données iconographiques scientifiques. La création de cette plateforme fut une première réponse de l'archéologie face aux différentes problématiques auxquelles elle faisait face : *Big Data*, droits d'auteur, documentation, formats de conservation, interopérabilité et diffusion. MédiHAL est une archive ouverte publique utilisant HAL et le principe de *l'Open Access* afin de créer un espace de dépôt pour les chercheurs dans le respect des droits d'auteur. Elle propose aussi un stockage illimité avec six licences *Creatives commons* facilitant le partage des données. Géo-référencement, compatibilité avec le logiciel bibliographique Zotero et récupération automatique des métadonnées sont les avantages qu'offre MédiHAL, néanmoins, seules les images issues de travaux ou projets de recherche officiels sont autorisés à être déposées. Une fois déposées dans MédiHAL, les données sont archivées à long terme par le CINES selon le modèles OAIS. L'autre exemple de l'appropriation des archives ouvertes comme solution de diffusion et valorisation des données archéologiques est la plateforme HAL utilisée par l'université de Grenoble. Selon Lucia Albaret de la direction des services à la recherche, service inter-établissement de la documentation de l'université Grenoble Alpes, le but était de proposer un outil de gestion des publications de tous types pour permettre le signalement, la valorisation, l'exploitation et enfin l'archivage et la conservation²⁰. Cet outil commun à plusieurs établissements a été pensé de manière à ce que la priorité soit donnée à la qualité des métadonnées descriptives et à l'interopérabilité avec les systèmes d'informations locaux et nationaux. HAL est considérée comme la plateforme nationale des archives ouvertes et une infrastructure de recherche indexée par tous les moteurs de recherche. Le dépôt y est, depuis plusieurs années, fortement encouragé. Bien que son appropriation est très aléatoire au niveau des SHS, les historiens et les archéologues sont les premiers devant les chercheurs en droit a utilisé la plateforme.

Enfin, la vulgarisation est une autre forme de diffusion et de valorisation des données archéologiques. Généralement destinée à rendre accessible, au plus large public possible, les résultats de recherche ou enjeux défendus par la communauté scientifique, cette action se présente même comme une nécessité pour l'avancée de la recherche. Utile à la production du savoir pour le CNRS, la diffusion par la vulgarisation est reconnue comme l'une des

¹⁹ MédiHAL est créé le 3 février 2010 par le CCSD en coopération, de 2010 à 2013, avec le CN2SV du CNRS.

²⁰ ALBARET Lucie. Un portail HAL à l'université Grenoble Alpes : un projet au service des chercheur... et du libre accès ! [en ligne]. 23 mars 2017, mis à jour le 30 novembre 2017, [consulté le 20 juin 2018], 24:33 minutes. Disponible sur le Web : < <https://www.inrap.fr/un-portail-hal-l-universite-grenoble-alpes-un-projet-au-service-des-chercheurs-13181> >.

premières missions du chercheur. Rendre public, accessible et compréhensible des projets ou résultats de recherche par le biais de différents supports (cours d'universités, expositions, ouvrages de référence, expositions muséales, participations à des émissions de télévision, articles de magazines de vulgarisation scientifique...) est l'un des enjeux du chercheur et de tout institut de recherche. D'une certaine manière, cette vulgarisation participe à la conservation de la production scientifique puisque celle-ci s'inscrit dans la défense et la sauvegarde du « bien commun »²¹. La diffusion par la vulgarisation, c'est mettre à disposition des autres des informations réfléchies, codifiées, interprétées et documentées. En 2007, le CNRS a même fourni plusieurs recommandations à destination des chercheurs et des instituts de recherche sur la nécessité de diffuser les résultats de recherche. Ces réflexions et recommandations présentées sous la forme d'un avis²² avaient pour objectif d'accompagner tout chercheur ou institut dans son choix de modèle de diffusion des données. Concentré sur les publications scientifiques, cet avis préconise l'ouverture des données en proposant de diversifier les modes de publication : l'*open source* y est l'une des solutions privilégiées.

Diffuser est donc considéré, d'une certaine manière au sein de la communauté scientifique, comme un processus d'archivage. La prolifération des plateformes proposant simultanément diffusion et archivage le prouve. Tendance à deux niveaux, cette étape dans le cycle de vie des données n'est pourtant pas ressentie comme primordiale.

L'archivage des données archéologiques

En plus de représenter une ambition importante dans la carrière d'un chercheur, l'archivage des données archéologiques est, depuis 2012, l'un des premiers enjeux défendus par le consortium MASA. Ce premier enjeu se décline en plusieurs objectifs dont l'archivage, l'interopérabilité des bases de données, le traitement des archives anciennes et la mise en place de procédures pour leur gestion font partis²³. L'enjeu de l'archivage des données archéologiques s'est présenté comme un impératif face à l'obsolescence des

²¹ Le bien commun est une notion philosophique, théologique, juridique et socio-politique, désignant l'idée qu'un bien patrimonial peut et doit être partagé par toute une communauté (ici, le grand public). Le mot « bien » signifie aussi bien au sens pratique que spirituel, le bien matériel dont on dispose ou possède.

²² Voir Annexe n°1 – AVIS DU COMETS – RÉFLEXION ETHIQUE SUR LA DIFFUSION DES RESULTATS DE RECHERCHE.

²³ HURLET Frédéric et RODIER Xavier Rodier. Archivage, publication et mise à disposition de données archéologiques [en ligne]. 26 et 27 septembre 2016 [consulté le 27 juin 2018]. Disponible sur le Web : < <https://vimeo.com/189300888> >.

machines et des logiciels. En archéologie, l'archivage est une action mal comprise des chercheurs. L'étape de la publication y est d'ailleurs considérée purement et simplement comme de l'archivage, c'est un « paradigme » traditionnel dénoncé par Bertrand Müller²⁴. Les données archéologiques, partie intégrante des archives de la recherche, constituées sous la forme de fonds d'archives de chercheurs sont généralement versées aux Archives départementales ou aux Archives Nationales. Néanmoins, si les archives archéologiques - sur support papier - sont relativement bien prises en charge, les solutions d'archivage ou de pré-archivage des données nativement numériques restent minoritaires.

Lors de sa création en 2002, l'INRAP s'est rapidement questionnée sur le devenir de ses données et sur leur archivage à long terme. Pour répondre à cette problématique, elle s'est dotée d'un portail HAL. Depuis 2010, les données intégrées directement dans HAL après validation sont archivées quotidiennement par le CINES. Le choix des archives ouvertes permet non seulement une première réponse à un archivage à long terme mais aussi à une diffusion et valorisation, non-négligeable, des données. Pour les données archéologiques issues des fouilles programmées, leur archivage est laissé au libre-arbitre des établissements de recherche qui les hébergent (instituts, centres de recherche, laboratoire...) ou au directeur de la mission archéologique. Dans le cas où aucune personne physique ou morale n'a été sensibilisé aux enjeux de l'archivage, les données se retrouvent stockées plusieurs années sur des supports amovibles (clef USB, disque dur externe...), voire oubliées. De plus, face à l'absence de service d'archives, les données sont souvent récupérées par le service informatique ou le service documentation pour finalement ne jamais être traitées. A cela s'ajoute l'absence totale de métadonnées descriptives rendant les données inexploitable. Ce vaste problème provient avant tout du manque de sensibilisation des chercheurs aux archives. Dans l'imaginaire collectif l'archivage des données débute seulement après leur publication et les archéologues sont les seuls propriétaires de leurs données. Hors, bien que soumises au droit d'auteur et au Code de la propriété intellectuelle, les données archéologiques sont des données publiques et des solutions d'archivage spécifiques existent.

La TGIR Huma-Num²⁵ est l'une des seules infrastructures à l'échelle européenne à proposer une solution d'archivage pérenne couplée à une solution de diffusion des données

²⁴ MÜLLER Bertrand. Archiver les données en Sciences humaines et sociales. Informations, positions de problèmes et propositions [en ligne]. 2012, [consulté le 27 juin 2018]. Disponible sur le Web : < <http://www.cnrs.fr/inshs/informations-DU/docs/du2013/archives-shs.pdf> >.

²⁵ La TGIR Huma-Num est une Très Grande Infrastructure de Recherche en Humanités numériques mettant en œuvre des dispositifs technologiques et humains pour répondre aux besoins des chercheurs en SHS en Europe

de la recherche en SHS. S'appuyant essentiellement sur les infrastructures du CINES, Huma-Num accompagne les chercheurs et leurs données tout au long du processus d'archivage. Les données présentant une valeur scientifique et historique reconnues par la communauté scientifique sont soumises à la validation d'un conseil scientifique et doivent respecter quelques recommandations : avoir un format libre et ouvert, des métadonnées descriptives renseignées, être préalablement publiées et recevables par le CINES. Par exemple, la MMSH d'Aix-Marseille archive ses données dans la plateforme PAC du CINES, via la TGIR Huma-Num et le laboratoire HiSoMA à Lyon verse, via l'utilisation de l'Huma-Num Box, les données de certains projets à la TGIR. D'une autre manière, certaines solutions GED, comme le logiciel Nuxeo, sont utilisées par les UMR pour diffuser, valoriser et pré-archiver leurs données. Le service PSIR de la MOM permet par exemple aux différents laboratoires qu'elle abrite de bénéficier de la GED Nuxeo. Cette gestion électronique des données est avant tout considéré comme une plateforme de diffusion et de valorisation avant d'être utilisé pour sa fonction archivage. Pour la MOM, l'avantage de Nuxeo réside dans sa capacité à produire un site internet à partir des données et métadonnées intégrés dans la GED, possible grâce à l'interopérabilité des systèmes. La mission archéologique française de Kition et Salamine de Chypre est l'une des premières du laboratoire HiSoMA à s'être appropriée la solution de pré-archivage pour diffuser et valoriser ses données archéologiques²⁶. Ainsi, l'interopérabilité intervient directement dans la valorisation et la diffusion de données archéologiques. D'autres solutions ou plateformes d'archivage/valorisation/diffusion des données (IR Numedif, OPIDoR, Parthenos ou Dariah-eu), dont l'utilisation est encouragée par les établissements de recherche, sont aussi mises à disposition des chercheurs.

L'archivage des données archéologiques dépend donc essentiellement des solutions choisies par le chercheur. Quant aux modes diffusion et de valorisation, ils sont fortement influencés par *l'Open Access*.

Open Access/Open Data : les projets de l'archéologie

Depuis quelques années, la diffusion des données est rendu possible par l'intermédiaire de *l'Open Access*. « Actualité de l'Open Access en archéologie française » fut, par ailleurs,

et en France. Elle les accompagne en leur proposant des solutions de valorisation, de diffusion, de signalement, de stockage, de traitement et d'archivage des données de la recherche.

²⁶ Le site de la mission archéologique française de Kition et Salamine de Chypre est disponible sur le Web au lien suivant : < <http://chypre.mom.fr/KitionSalamine/home> >.

l'objet d'un séminaire organisé par l'INRAP et avec le soutien du consortium MASA, le 23 et 24 mars 2017 à Paris dans le cadre de l'appel à projet « *Open Access week 2016-2017* ». Au sein de ce séminaire, plusieurs thématiques liant archéologie et *Open Access* ont été abordés : le rôle de *l'Open Access* dans la diffusion des résultats de recherche scientifique, les perspectives de *l'Open Access* pour l'archéologie préventive et une réflexion sur l'ouverture et le partage des données archéologiques. L'ouverture et le libre accès ont véritablement bouleversé les modes de diffusion et de valorisation.

Qu'est-ce que *l'Open Access* ? Le libre accès, en français, signifie, selon Paola Moscati (éditrice de la revue *Archeologia e Calcolatori*, de l'Institut des études sur la Méditerranée antique (ISMA – CNR) : « [...] [la mise] à disposition gratuite sur l'Internet public de la littérature de recherche validée par les pairs, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou créer un lien vers un texte intégral de ces articles à des fins non-commerciales, à condition qu'un contrôle sur l'intégrité des travaux et leur droit à être correctement reconnus et cités soient garantis aux auteurs »²⁷. *L'Open Access* est considéré comme un mode de diffusion et, par extension, de valorisation.

Parmi les premiers impacts notables du rôle de *l'Open Access* dans la diffusion des données archéologiques, il y eut la revue italienne *Archeologia e Calcolatori*, qui en 2005 décida de rejoindre *l'Open Archives Initiative*. Désormais soumise au renseignement des métadonnées pour la description des ressources, au protocole OAI-PMH et à sa contribution, la revue s'est donnée pour mission de diffuser le plus largement possible les données publiées. A partir de cette année-là, la mise en ligne de leurs articles et le renseignement de leurs métadonnées permirent au lecteur de passer d'une couche d'information à une autre, d'une table des matières au contenu issu d'un référentiel où les ressources décrites par des métadonnées respectaient le Dublin Core²⁸, tout en offrant une possibilité de téléchargement gratuit. *L'Open Access* a ouvert de manière évidente de nouvelles perspectives à la recherche en archéologie. L'INRAP a notamment abordé cette question il y a une dizaine d'années en mettant en œuvre la plateforme DOLIA (catalogue de ressources documentaires) signalant et donnant accès à un grand nombre de rapports²⁹. Tout en souhaitant garder un droit d'accès

²⁷ MOSCATI Paola. Archéologie et informatique. Le rôle de *l'Open Access* dans la diffusion des résultats de recherche scientifique [en ligne]. 23 mars 2017, mis à jour le 30 novembre 2017, [consulté en ligne le 20 juin 2018], 23:41 minutes. Disponible sur Web : < <https://www.inrap.fr/archeologie-et-informatique-le-role-de-l-open-access-dans-la-diffusion-des-13176> >.

²⁸ Le Dublin Core est un schéma de métadonnées générique permettant la description de ressources tout en établissant des relations avec d'autres ressources. Comprenant 15 éléments de description formels et thématiques relatifs à la propriété intellectuelle, le Dublin Core est une norme internationale ISO 15836.

²⁹ SPELLER Anne et COLIN Camille. Quelles données pour quels projets : perspectives de *l'Open Access* pour l'archéologie préventive [en ligne]. 23 mars 2017, mis à jour le 30 novembre 2017, [consulté le 22 juin

sur ses données en les ouvrant au public, elle s'est tenue à une rigueur d'enregistrement et de présentation des données produites et collectées sur les chantiers de fouilles archéologiques. En 2017, près de 30 000 notices étaient ainsi mises à disposition du grand public, des collectivités et des chercheurs au travers de la plateforme DOLIA et d'autres plateformes du gouvernement comme Open Data ou le portail d'échange au niveau européen ARIADNE.

Open Access et jeux de données permettent ainsi que se mettent en place, au niveau national et européen, des projets d'ouverture et de diffusion des données archéologiques. Fort de cette expérience, Anne Speller envisageait, en 2017, une ouverture des données confidentielles de l'INRAP (rapports de spécialistes ou rapports issus de programmes collectifs de recherche). La diffusion et la valorisation des données archéologiques, à l'échelle régionale, s'est aussi faite en Bretagne par l'intermédiaire de *l'Open Access*. Sortie d'une logique de conservation pure, la DRAC Bretagne, après avoir numérisé quelques-unes de ses archives, a choisi de les diffuser en ligne *via* une bibliothèque numérique spécialisée sur les rapports d'opération archéologique et mises en lien avec une base de données géographiques : GéoBretagne³⁰. Chaque rapport numérisé a été associé à des métadonnées renseignées selon le Dublin Core. Le choix du logiciel CMS s'est porté sur Omeka et ses *pluggings* (recherche, moissonnage, géolocalisation, etc...) notamment pour la structure de l'architecture favorisant l'interopérabilité des données³¹. La plateforme assure ainsi la diffusion et la valorisation de données géographiques et archéologiques.

Le mouvement *Open Access* a non seulement modifié les modes de diffusion de la culture scientifique, mais aussi les pratiques. Ce mouvement a en souvent accompagné un autre, l'archéologie numérique, ouvrant de nouvelles possibilités.

L'ARCHEOLOGIE NUMERIQUE : DE NOUVELLES POSSIBILITES

2018], 21:31 minutes. Disponible sur le Web : < <https://www.inrap.fr/quelles-donnees-pour-quels-projets-perspectives-de-l-open-access-pour-l-13180> >.

³⁰ GéoBretagne est un géoportail de l'urbanisme en lien avec les infrastructures de données géographiques de la région bretonne.

³¹ MENEZ Yves, LORHO Thierry et PINEL Marie-Dominique. Une expérience d'open access/open data : la DRAC Bretagne – service régional de l'archéologie met en ligne ses données archéologiques [en ligne]. 23 mars 2017, mis à jour le 1^{er} décembre 2017, [consulté le 22 juin 2018], 30:27 minutes. Disponible sur le Web : < <https://www.inrap.fr/une-experience-d-open-accessopen-data-la-drac-bretagne-met-en-ligne-ses-donnees-13179> >.

Tout au long des opérations de fouille, les archéologues produisent des données : lors de la phase de terrain, lors de la post-fouille et lors de la synthèse des données jusqu'à la diffusion des résultats sous forme de rapports d'activité ou de publications. Cette diffusion résulte généralement d'un processus d'enregistrement, d'un traitement, d'une gestion et d'une mise en page des données collectées et produites.

Depuis plusieurs décennies, l'archéologie interroge ces pratiques évoluant au grès des avancées technologiques. Selon Paola Moscati, l'archéologie numérique est ce qui a guidé l'archéologie depuis trois décennies ouvrant des voies alternatives d'investigations scientifiques avec pour conséquence un impact profond sur les pratiques. Les premières expérimentations de l'archéologie numérique remontent aux années 1990 avec les premiers projets soutenus par l'Istituto per l'Archeologia Etrusco-Italica du CNR (aujourd'hui Istituti di Studi sul Mediterraneo Antica), sur le classement et le traitement automatique des corpus d'objets archéologiques de la culture étrusque³². Quelques années plus tard, les premiers encodages de carnets de fouilles sont lancés avec la possibilité de consultation dans un SIG. Alexia Dozon présente d'ailleurs le SIG en archéologie comme une extension de l'utilisation de l'informatique en le justifiant par « [...] leur capacité à gérer massivement différents types de données et par leur interopérabilité »³³.

Le numérique en archéologie intervient donc très tôt dans le processus-métier. Dès le début des opérations, lors de l'enregistrement des données de terrain, l'archéologue ou le responsable d'opérations charge un topographe (cela peut être un tiers extérieur ou un des membres de l'équipe) de cartographier le site de fouille et d'effectuer des relevés. Ensuite on utilise des SIA destinés à l'enregistrement, l'exploitation et l'analyse documentaire des données. Le premier objectif est de gérer des bases de données SQL hébergées sur un serveur (à distance, en local ou sur un poste fixe) puis de transférer les données à un opérateur DAO-SIG³⁴ qui se chargera de la vectorisation, la numérisation et du traitement des images produites. Enfin, l'opérateur PAO corrigera les éventuelles erreurs et organisera la mise en page des données. Ce processus de traitement des données (figure 2) peut être réalisé par une seule et même personne si les ressources humaines et financières accordées à la fouille d'un

³² MOSCATI Paola. Archéologie et informatique. Le rôle de l'Open Access dans la diffusion des résultats de recherche scientifique [en ligne]. 23 mars 2017, mis à jour le 30 novembre 2017, [consulté en ligne le 20 juin 2018], 23:41 minutes. Disponible sur le Web : < <https://www.inrap.fr/archeologie-et-informatique-le-role-de-l-open-access-dans-la-diffusion-des-13176> >.

³³ DOZON Alexia. Archéologie numérique [en ligne]. 17 juin 2016, [consulté le 26 juin 2018]. Disponible sur le Web : < <https://archnum.hypotheses.org/40> >.

³⁴ La DAO (dessin appliqué par ordinateur) et les SIG sont les principaux outils numériques des archéologues.

site ne sont pas suffisantes. S'il n'est pas bien respecté, ce processus peut altérer le traitement des données endommageant la documentation archéologique produite.

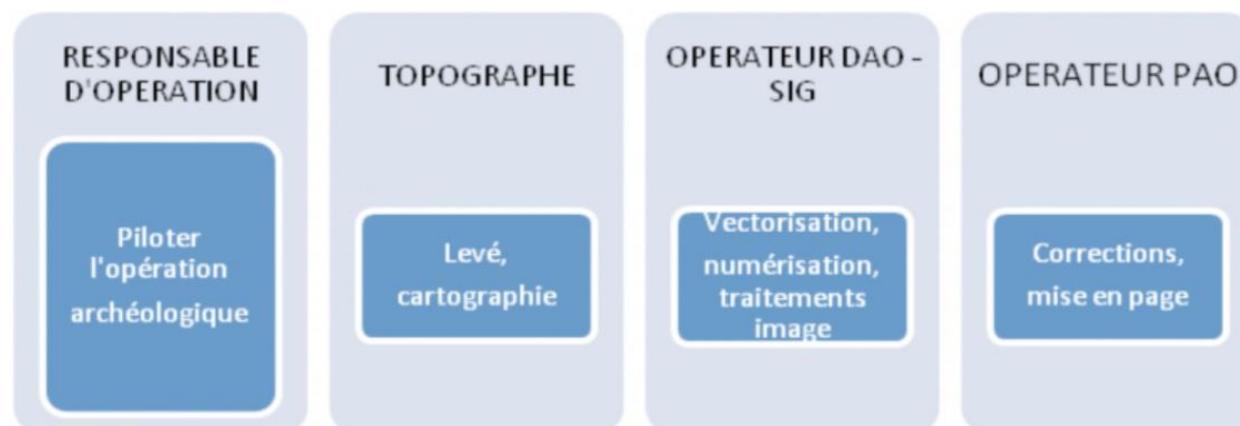


Figure 2 Opérations de traitement et de gestion de la documentation archéologique par SIG.
Source : DOZON Alexia. « La chaîne opératoire » du traitement et de la gestion de la documentation archéologique par SIG : une modélisation des données archéologiques vers un modèle d'information archéologique (projet « CSNE »), 22 juin 2016, in Archéologie numérique.

Dans les nouvelles perspectives ouvertes par l'archéologie numérique, le relevé archéologique sous forme numérique se présente comme l'un des outils les plus emblématiques. Le relevé archéologique numérique est le résultat d'une démonstration archéologique s'exprimant au travers de plusieurs méthodes d'acquisition tel que le LIDAR, la photogrammétrie 3D et la modélisation 3D. Le Li.D.A.R (Light Detection and Ranging) est une technique de télédétection aéroportée permettant les relevés topographiques. Cette méthode de cartographie élaborée à partir d'un référentiel de vestiges archéologiques nécessite un traitement exigeant tel que la représentation de certaines interprétations par des graphiques. Modéliser l'information archéologique a pour objectif, à terme, de permettre la comparaison des modèles mais, cela n'est possible que si les données sont structurées.

Ainsi, la structuration des données est essentielle si l'on veut accroître l'interopérabilité. Une modélisation 3D de l'existant peut se faire à partir d'un relevé ou de la photogrammétrie. La photogrammétrie architecturale est une méthode récemment utilisée en archéologie permettant d'effectuer des mesures en utilisant la parallaxe³⁵ obtenue par des

³⁵ La parallaxe est l'effet du changement de position de l'observateur sur ce qu'il perçoit.

images acquises selon des points de vue différents³⁶. L'utilisation de cette méthode fait partie des nouvelles formes d'acquisition des données de terrain. L'acquisition 3D acquise par la photogrammétrie vient compléter les données archéologiques produites, cela permet de documenter de manière la plus exhaustive possible une fouille. La photogrammétrie permet aussi l'élaboration de procédures d'accès à l'information, l'incorporation de l'image et d'une représentation graphique - élément de gestion de l'information - et ouvre un nouveau champ de possibilités. Cette nouvelle manière d'acquérir l'information semble plus neutre et objective laissant de côté la libre interprétation des chercheurs. Cela relèverait plus de la modélisation de l'information vers la donnée qu'une simple acquisition de terrain. Précieux vecteur de communication, la photogrammétrie permet au plus grand public de se représenter un objet ou monument antique. Le numérique vient directement influencer les modes de partage et de valorisation des données archéologiques récoltées. Le développement de l'archéologie numérique amène avec elle tout un débat autour des SI et notamment des SIG.

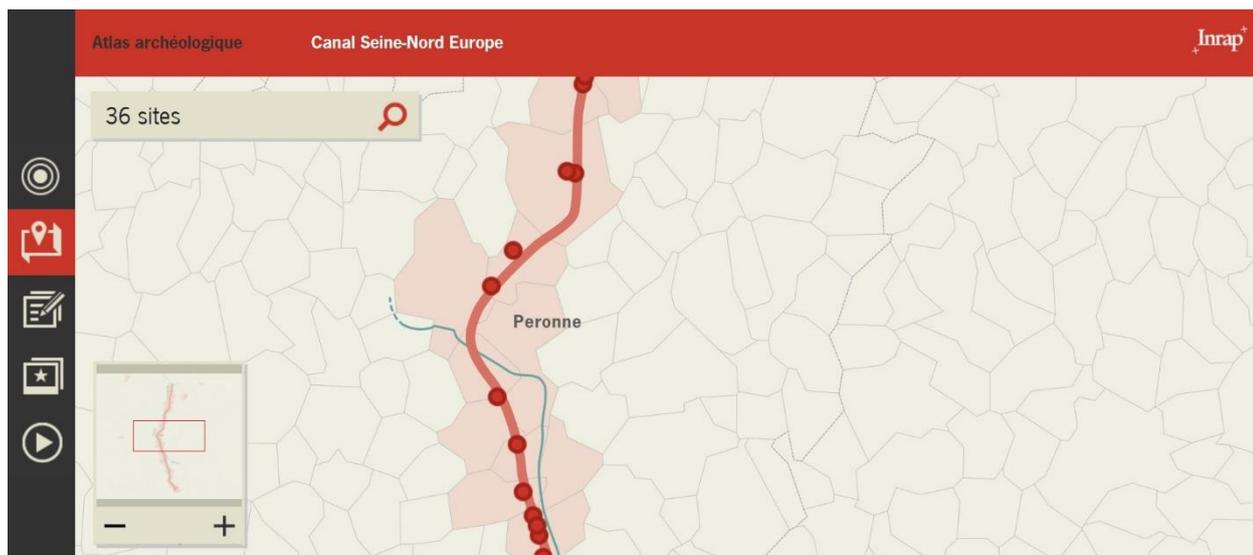


Figure 3 Atlas archéologique en collaboration avec Canal-Seine Nord Europe et l'INRAP.

Le projet collaboratif CSNE (Canal Seine Nord Europe) et de l'INRAP par exemple, est l'un des premiers en France à aborder la question de la modélisation des données archéologiques grâce à la mise en place d'un SIG. Le but de cet atlas archéologique est de croiser des jeux de données pour permettre de créer une carte en répertoriant et en localisant les sites archéologiques du nord de la Seine en Europe (figure 3). Cette chaîne opératoire est

³⁶ BELARBI Mehdi, RAYMOND Pascal, SAULIERE Nicolas et TOUQUET Régis. L'acquisition 3D par photogrammétrie en archéologie [en ligne], juillet 2012, mis en ligne le 1^{er} juillet 2012, [consulté le 24 juillet 2018]. Disponible sur le Web : < <http://journals.openedition.org/archeopages/415> >.

présentée par Alexia Dozon comme « [...] *une logique de production et de gestion de contenu, sous forme de chaîne éditoriale, définie par un processus métier, un modèle d'information archéologique et un logiciel type SIG que ce projet a voulu mettre en œuvre pour [...] faciliter certaines communications. [...] Dans une démarche de recueil et de traitements de ces données, nous avons eu à définir une méthode d'archivage de ces données.* »³⁷. Ce principe de modélisation 3D est repris par une plateforme d'annotation sémantique 3D pour la documentation d'objets patrimoniaux : AĬOLI. L'objectif de cette plateforme est de créer un lien entre l'objet antique étudié avec l'information transmise, l'application génère une représentation 3D de l'objet pouvant être enrichi *via* des annotations sémantiques ou des ressources complémentaires. Cet outil vise essentiellement le renouvellement des méthodologies de travail.

L'apport de ces nombreux outils et techniques de relevé archéologique font émerger de nombreux questionnements dont un nous intéresse plus particulièrement : une fois les données archéologiques annotées, indexées, signalées, traitées, modélisées, documentées, comment les diffuser et les valoriser sans perte d'informations ? C'est précisément au cœur de cette problématique que les enjeux de l'interopérabilité interviennent. Exerçant une vraie influence sur la manière d'utiliser et de réutiliser les données, les enjeux de l'interopérabilité sont le point d'ancrage de plusieurs débats animés où se rencontrent numérique, sciences de l'information et archéologie, faisant émerger de nouvelles perspectives d'étude et d'évolution, notamment autour de l'utilisation, la réutilisation et la redocumentarisation des données.

³⁷ DOZON Alexia. « La chaîne opératoire » du traitement et de la gestion de la documentation archéologique par SIG : une modélisation des données archéologiques vers un modèle d'information archéologique (projet « CSNE ») [en ligne]. 22 juin 2016, Archéologie numérique, [consulté le 3 juillet 2018]. Disponible sur le Web : < <https://archnum.hypotheses.org/189> >.

UTILISER ET REUTILISER LES DONNEES ARCHEOLOGIQUES : UNE QUESTION D'INTEROPERABILITE ?

Afin de pallier aux différents problèmes liées aux environnements informatiques hétérogènes, les producteurs de systèmes d'exploitation cherchent depuis plusieurs décennies à répondre de nouveaux enjeux tel que la portabilité des programmes informatiques et l'interopérabilité. Tandis que la portabilité désigne la migration d'une application de plateforme vers une autre, l'interopérabilité consiste à permettre à des systèmes, plateformes, composants ou applications à se connecter et à échanger entre elles.

L'interopérabilité des données, d'un point de vue essentiellement technologique, peut être assurée de plusieurs manières : par le développement de logiciels interopérables, par le licenciement et la publication de formats non-propriétaires, par l'établissement de collaborations spécifiques avec concurrents, clients ou partenaires et par la mise en œuvre de standards ouverts. D'un point de vue sémantique, l'interopérabilité est favorisée par l'usage de normes, de métadonnées, d'ontologie et de vocabulaire précis permettant une connexion entre les données sollicitées. Présentée dans les années 1990 comme la solution aux problèmes d'échanges entre les ordinateurs et les machines, l'interopérabilité est devenue, aujourd'hui, un enjeu pour la communication et la réutilisation des données.

L'IMPORTANCE DES METADONNEES EN ARCHEOLOGIE

L'archéologie s'appuie, depuis une décennie, sur la collecte et la production de données nativement numérique de plus en plus complexe. Ces données sont très souvent accompagnées de données sur les données que l'on désigne communément par le terme de métadonnées. Selon les différentes communautés professionnelles, le terme de métadonnées n'a pas la même signification. Chez les archivistes et les bibliothécaires, les métadonnées désignent l'ensemble des informations appliquées – et applicables – à un document. Ces informations peuvent être produites manuelles par les professionnels et réunis sous forme de notice bibliographie ou extraites automatiquement.

L'ajout manuel d'informations est désigné par Muriel Amar comme des informations « extrinsèques » permettant une indexation interprétative³⁸ en opposition aux métadonnées « intrinsèques » que l'extraction ne fait qu'isoler, sans aucune interprétation ou analyse. Chez les techniciens et les informaticiens, les métadonnées désignent les données produites automatiquement par les systèmes informatiques tel que l'emplacement de fichiers, le nommage, la taille, le format ou encore l'horodatage³⁹. En archéologie, les métadonnées ont une toute autre importance.

Par exemple, lors de l'archivage de données archéologiques modélisées en 3D (figure 4) et déposées au CINES, les métadonnées sont généralement réparties en trois grandes catégories⁴⁰ :

- les métadonnées administratives : elles décrivent des droits et sont relatives à l'ensemble du dépôt ;
- les métadonnées archéologiques : elles décrivent l'époque, le lieu, les auteurs et sont aussi relatives à l'ensemble du dépôt ;
- les métadonnées techniques : elles décrivent chaque donnée individuellement, pour les sources, les photographies, les archives utilisées, les stations lasers et pour le modèle 3D final et les fichiers 3D (position, nom et aperçu).

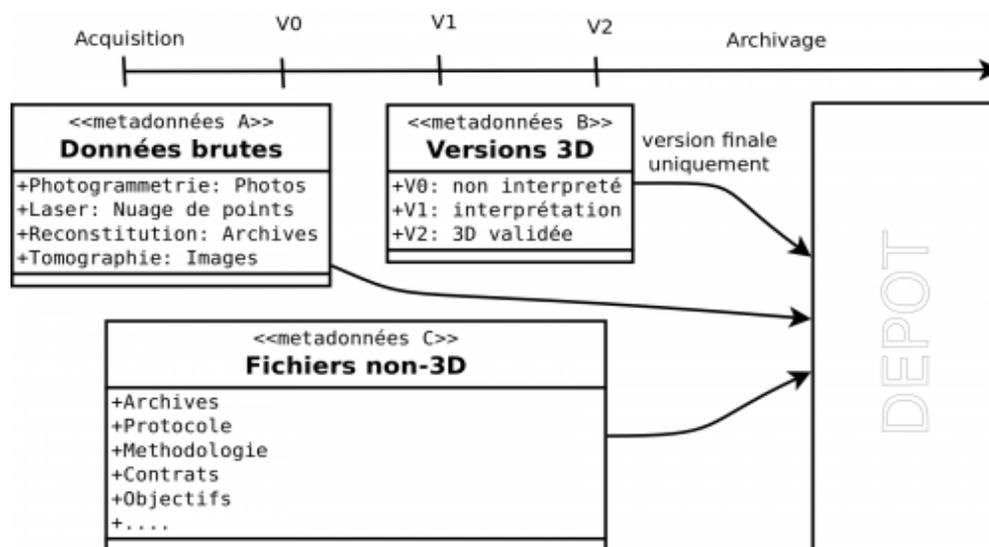


Figure 4 Schéma représentant la nature et le type de données 3D et métadonnées à archiver lors du cycle de production
Sources QUANTIN Matthieu. Archivage de modèles 3D.

³⁸ AMAR, Muriel. *Les fondements théoriques de l'indexation : une approche linguistique*, Paris : ADBS Éditions, 2000, 355 pages.

³⁹ BROUDOUX Evelyne et SCOPSI Claire. L'enjeu des métadonnées dans un contexte de "redocumentarisation", in *Etudes de communication*, [en ligne], 2011, [consulté le 30 juillet 2018]. Disponible sur le Web : < https://www.researchgate.net/publication/281156482_L%27enjeu_des_metadonnees_dans_un_contexte_de_redocumentarisation >.

⁴⁰ QUANTIN Matthieu. Archivage de modèles 3D, [en ligne], 15 septembre 2016, [consulté le 5 août 2018]. Disponible sur le Web : < <https://shs3d.hypotheses.org/2354> >.

La variété de la nature des métadonnées produites pour l'archivage, découle généralement de la nature du projet auquel elles sont rattachées. Ces métadonnées ne sont pas les mêmes lorsqu'il s'agit de les diffuser ou de les valoriser. En effet, les formats pour l'archivage des données⁴¹ ne sont pas les mêmes que ceux pour la diffusion ou la valorisation, il en va de même pour les métadonnées. Emmanuelle Morlock appuie ce point-là. « *Selon moi on ne dépose pas les mêmes données sur une plateforme de partage que sur une plateforme d'archivage. La plateforme de partage est là pour partager, diffuser des données à une communauté tandis que la plateforme d'archivage est choisie pour un rôle de conservation des données avec une logique de fonds et de durée* »⁴². L'ingénieure d'étude au CNRS défend son idée en évoquant la finalité du projet. Lors de projets de diffusion ou de valorisation, les formats requis pour déposer sur des plateformes de diffusion les données et les métadonnées sont des formats facilitant l'interopérabilité.

Depuis quelques années, l'un des enjeux les plus importantes dans l'exploitation de la documentation archéologique est la normalisation des schémas de métadonnées. Elle joue un rôle notable dans l'organisation de l'accès aux données. En archéologie, la standardisation des schémas ou formats de métadonnées rencontre quelques difficultés. L'impossible consensus autour de la caractérisation pertinente des métadonnées en archéologie freine considérablement le développement du processus d'interopérabilité, que ce soit dans les SIA, les SIG, sur les plateformes et les GED. L'accès partagé aux données et métadonnées archéologiques est un deuxième problème rencontrés ces dernières années. Le manque de développement de schémas de standards de métadonnées facilitant la collecte, la réutilisation et l'interprétation empêche l'interopérabilité bloquant parfois les programmes de valorisation et de diffusion. Depuis quelques années, les archéologues et les professionnels de la documentation archéologique tentent de saisir certains schémas et modèles pour les transposer à leurs systèmes d'informations. Le Dublin Core⁴³, le modèle OAIS⁴⁴ et

⁴¹ Le CINES a d'ailleurs mis en place une veille pour l'utilisation des bons formats en créant un répertoire via la plateforme FACILE, service de validation de formats pour permettre de vérifier l'éligibilité des données et documents pour un archivage pérenne. Disponible sur le Web : < <https://facile.cines.fr/> >.

⁴² Voir Annexes, Entretien avec Emmanuelle Morlock

⁴³ Le Dublin Core est un standard, un ensemble de recommandations pour la description des métadonnées. Il établit 15 éléments de base pour la description bibliographique des données, des documents et des objets se trouvant sur le web, que ce soit dans une GED, une bibliothèque numérique ou une plateforme quelconque. Tandis que certains éléments décrivent le contenu (titre, sujet, description, source, langue, relation, couverture) d'autres décrivent la propriété intellectuelle (créateur, éditeur, contributeur et les droits associés) et certaines particularités (date, type, format, identifiant). Le Dublin Core peut être exprimé en HTML, en XML et même en XML/RDF. On parle d'ailleurs plus souvent de schéma de métadonnées.

⁴⁴ Le modèle OAIS est un modèle conceptuel et fonctionnel pour la pérennisation des données numériques. C'est une norme depuis 2003 prenant en considération la qualité des données produites, les cycles de vie et la pérennisation. Le modèle OAIS sert généralement à modéliser un système de pérennisation des données.

les projets d'archives OAI⁴⁵ sont les concepts et modèles les couramment utilisés. Les métadonnées ont un but précis et sont standardisées depuis plus d'une décennie. Elles ont acquis un véritable statut à part entière, ce ne sont pas de simples renseignements sur les données mais des ressources⁴⁶. Ce statut de ressources nécessite une normalisation dans les usages et de respecter certaines règles. A ce jour en archéologie française, les métadonnées ne sont pas systématiquement renseignées, cristallisant un grand nombre de frustrations autant chez les professionnels de l'information et de la documentation que chez les archéologues. Malheureusement trop peu de chercheurs y ont recours pour décrire leurs métadonnées, ils justifient ce rejet par le manque de caractérisation des différents vocabulaires contrôlés. Pourtant les impératifs de la recherche en archéologie font que la précision des données collectées et produites par le renseignement des métadonnées demeure un enjeu fondamental bien que la communauté archéologique n'est pas encore donné d'avis tranché dans la résolution de ce problème de forme et de fond. En archéologie, les exigences de précision dans la définition des données doivent être comprise par les professionnels comme une réelle priorité. La complexité des données traitées implique une rigueur telle que l'interprétation libre ne peut malheureusement pas être la forme à privilégier.

L'usage des métadonnées en archéologie est d'ailleurs l'un des trois grands enjeux défendus par le consortium MASA. Face à la multitude de données traitées et produites, les archéologues ont recours depuis les années 1970 aux bases de données non-hétérogènes, conçues souvent sans protocole ou modèle méthodologique et technique. Ce manque d'organisation nécessite généralement une reprise du travail déjà effectué en amont pour rendre les bases de données interopérables et normalisées. Afin de faciliter l'interopérabilité des bases de données, le consortium MASA recommande d'utiliser certains schémas ou modèles de représentation interopérables comme SKOS⁴⁷, RDF⁴⁸ et le Dublin Core. L'usage de référentiels communs permet d'envisager une mise en lien des données et métadonnées dans le cadre du web des données liées. Cet enrichissement est rendu possible par la redocumentarisation, qui, en intégrant de nouvelles métadonnées dans le

⁴⁵ Le projet OAI, projet visant à faciliter l'échange et la valorisation des archives numériques, permet le moissonnage des métadonnées *via* des sites de fournisseurs de données. De nombreuses plateformes et sites internet sont actuellement dotés du modèle OAI. Ce protocole est très utile pour la création d'outils de recherche fonctionnant simultanément à partir de plusieurs bases de données.

⁴⁶ Data Archive : documentation et métadonnées [en ligne], 11 août 2011, [consulté le 3 août 2018]. Disponible sur le Web : < <https://archishs.hypotheses.org/532> >.

⁴⁷ SKOS n'est pas uniquement une recommandation formulée par le consortium MASA. SKOS constitue aussi, depuis 2009, recommandée par le W3C pour la représentation des thésaurus, des classifications et vocabulaires contrôlés. L'intérêt premier de SKOS est d'être interopérable avec le plus de standards possibles. Basé sur le modèle RDF, SKOS permet de faciliter les publications de vocabulaires structurés lors de leur utilisation dans le cadre du Web sémantique. En archéologie, SKOS intervient essentiellement dans la création d'ontologies.

⁴⁸ RDF est un modèle de graphe pour la description des ressources Web et leurs métadonnées, permettant, par extension, leur traitement automatique. RDF est le langage basique du Web sémantique, plus connu sous le nom de RDF/XML. Un document RDF repose sur le renseignement d'un ensemble de triplet (nom, prédicat, objet). Le sujet et l'objet peuvent parfois être comprises comme des ressources.

but de compléter les données, enrichit et facilite l'accès tout en assurant la pérennité et l'évolutivité des données⁴⁹. Le traitement, massif, des données archéologiques est un travail long et complexe, pour répondre à cet enjeu de taille, le consortium MASA cherche à donner une dimension nationale aux enjeux qu'il défend et la valorisation de données archéologiques passe indéniablement par la gestion des métadonnées associées. Les métadonnées et la façon dont les bases sont renseignées livrent des informations parfois inattendues sur la densité de la recherche dans un thème précis, se révélant un outil précieux pour les chercheurs. Sans métadonnées, les données archéologiques restent non seulement inutilisables scientifiquement parlant, mais aussi inexploitablement techniquement.

Le bon usage des métadonnées n'est pas sans rappeler l'importance de respecter certaines normes et standards. En effet, la définition de vocabulaires de données communs apporte une base solide pour garantir l'interopérabilité des systèmes d'information en archéologie, des bases de données ou tous autres jeux de données liées des plateformes de diffusion ou de valorisation.

DES NORMES ET DES STANDARDS POUR L'INTEROPERABILITE DES DONNEES ET DES METADONNEES

Dans un contexte d'interopérabilité, les normes et les standards jouent le rôle de garant de l'unicité des syntaxes et des vocabulaires dans les échanges entre systèmes. Peu importe les technologies utilisées, leur rôle est d'unifier la manière d'exprimer la donnée. La favorisation de la diffusion des données sur le Web et leur structuration constituer un réseau d'informations interrogeable propre au Web des données.

Tout comme dans d'autres sciences humaines et sciences sociales, l'utilisation et le développement des vocabulaires contrôlés et d'ontologies en archéologie semblent essentiels pour pouvoir permettre différents niveaux d'interopérabilités.

Le vocabulaire : du bon usage des thésaurus

En interopérabilité, la problématique générale est une problématique de protocole d'échange ; la question du développement des protocoles pour l'échange des données est donc inhérente à celle des enjeux recouverts par l'interopérabilité. Pour répondre à ce

⁴⁹ HURLET Frédéric. Les enjeux du consortium MASA, [en ligne], 10 août 2014, [consulté le 5 août 2018]. Disponible sur le Web : < <https://masa.hypotheses.org/27> >.

problème, il est recommandé de s'appuyer sur des formats et des protocoles répandus et communs à tous afin d'avoir le plus de chance possible d' « être compris » et « traduis »⁵⁰. Les standards du W3C qui forment les fondamentaux du Web, ont servi de base aux professionnels de l'information et de la documentation des sciences de l'Antiquité, qui ont dû s'adapter pour proposer des modèles de référence pour l'interopérabilité des données.

Le thésaurus est un mode d'indexation, issu du monde des bibliothèques, que les professionnels de l'archéologie se sont donc appropriés afin de répondre aux nouvelles problématiques amenées par le numérique. Le premier thésaurus pour l'archéologie, créé en 1987, né de la nécessité d'organiser les mots-clefs utilisés pour l'indexation des notices du Catalogue Collectif Indexé (CCI) du réseau FRANTIQ. Désormais et depuis les années 1990, les thésaurus PACTOLS, qui regroupent six microthésaurus (Peuples, Anthroponymes, Chronologie, Toponymes, Œuvres, Lieux et Sujets) sont utilisés par quelques archéologues, documentaliste et archivistes dans divers contextes, plateformes et logiciels.

Les thésaurus PACTOLS concernent toutes les époques, de la Préhistoire à l'époque contemporaine. Avec plus de 30000 descripteurs, les PACTOLS sont répandus *via* le logiciel OpenTheso développé par les réseaux documentaires de la MOM à Lyon afin d'offrir de nouvelles fonctionnalités et un outil de recherche multilingue (français, anglais, allemand, espagnol et italien) avancé⁵¹. Le pôle informatique et réseau est d'ailleurs l'un des premiers, a utilisé, de manière systématique, les thésaurus PACTOLS dans la description de leurs ressources documentaires et de leurs données intégrées dans la GED Nuxeo. Plateformes et laboratoire de recherche en histoire et archéologie font donc bon usage des thésaurus. En octobre 2015, le LAT (Laboratoire Archéologie et Territoire) de la MSH du Val de Loire à Tours, s'est réuni avec le réseau FRANTIQ afin de définir les prérequis pour l'alignement des vocabulaires utilisés dans leur base de données ArSol (Archives du Sol) avec les thésaurus multilingues de PACTOLS. Les responsables du projet souhaitaient relier les termes utilisés dans ArSol avec les identifiants ARK du vocabulaire des PACTOLS pour permettre une interaction, une communication entre ArSol et OpenTheso. Cette intervention montre bien la nécessité de développer les thésaurus pour garantir une interopérabilité.

Outre les bases de données et les GED, l'utilisation des thésaurus intervient aussi dans la diffusion d'informations *via* les plateformes de recherche en SHS. Isidore, par exemple,

⁵⁰ FRANCART Thomas. Ontologies et thésaurus, l'interopérabilité grâce au web des données, [en ligne], 21 novembre 2017, [consulté le 8 août 2018], 32 :3. Disponible sur le Web : < https://www.canal-u.tv/video/msh_val_de_loire/ontologies_et_thesaurus_l_interoperabilite_grace_au_web_de_donnees_thomas_francart_sparna_expert_en_technologies_semantiques.39735 >.

⁵¹ NOUVEL Blandine. Des outils d'enrichissement documentaire multilingues pour l'archéologie, [en ligne], 14 décembre 2015, [consulté le 7 août 2018]. Disponible sur le Web : < <https://masa.hypotheses.org/218> >.

l'utilise pour l'indexation des notices de la revue Archéologie de la France-Information et la revue Archéologie Médiévale y a récemment eu recours pour le développement de son projet Archéomed, où l'objectif était le même⁵². Ce travail portant sur les notices de fouilles est important, il permet non seulement de diffuser de manière structurée et compréhensible les données archéologiques, mais aussi de les rendre accessibles à toute la communauté scientifique et au grand public. Ce gain de visibilité permet alors une vraie valorisation. De plus, la ressource libre supportant les PACTOLS, OpenTheso, aborde plusieurs fonctionnalités importantes (gestion avancée de l'arborescence du thésaurus et des concepts, support de caractères Unicode, import/export de concepts ou de branches en CSV, XML-SKOS, JSon-LD, identifiants pérennes types ARK...) dont les alignements – conformes à l'ISO 25964-1 (2011) : Thésaurus et interopérabilité avec d'autres vocabulaires – favorisent

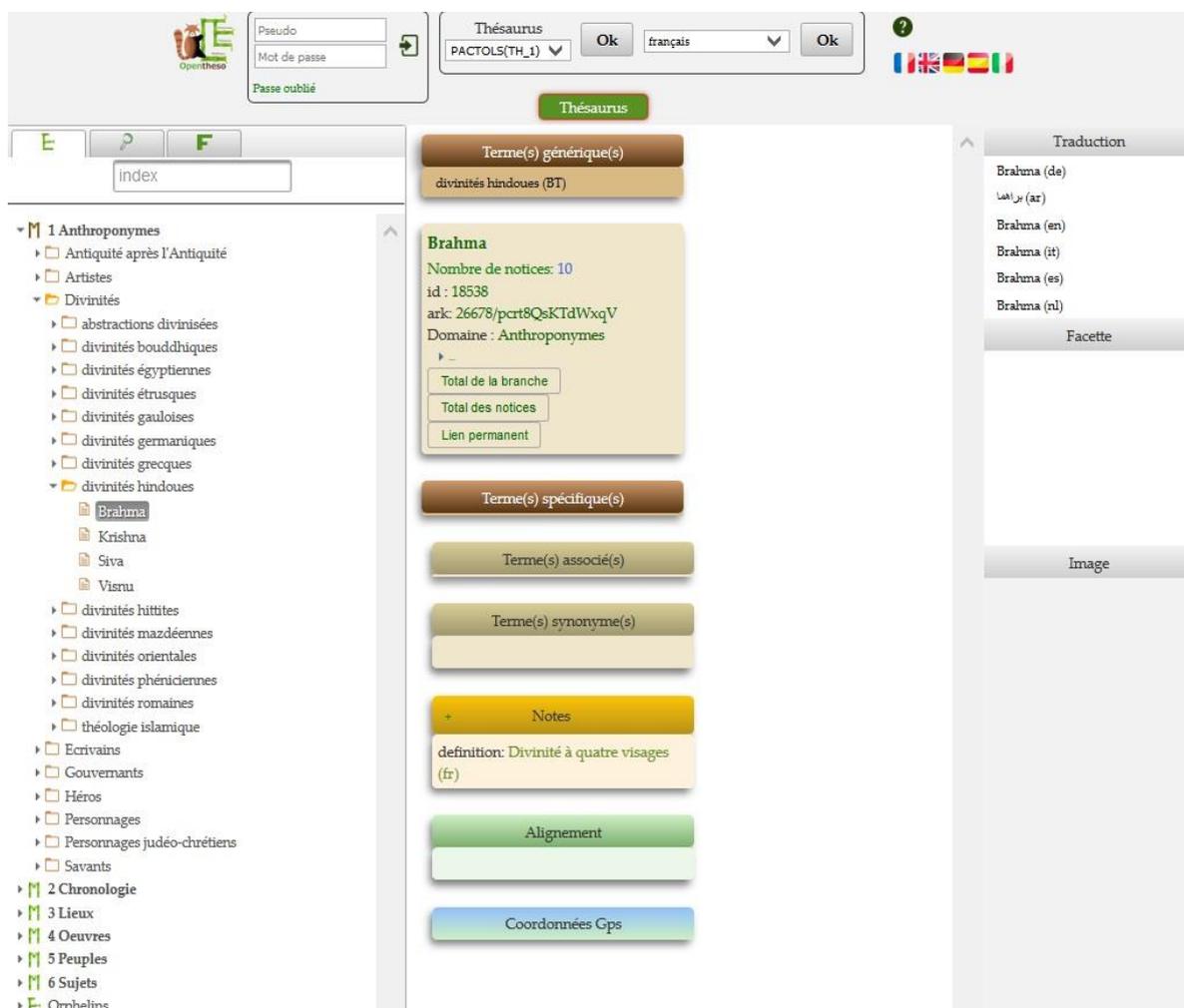


Figure 5 Exemple d'utilisation du microthésaurus Anthroponymes des thésaurus PACTOLS.
 Source : <http://pactols.frantiq.fr/opentheso/>

⁵² ALAINGUILLAUME Micael. Projet Archéomed – 50 ans d'archéologie en ligne [en ligne], 29 septembre 2017, mis à jour le 21 décembre 2017, [consulté le 7 août 2018]. Disponible sur le Web : < <https://craham.hypotheses.org/1356> >. ISSN : 2552-3139.

l'interopérabilité avec d'autres vocabulaires. Hormis l'indexation, les thésaurus PACTOLS (figure 5) participent à l'enrichissement des données et des métadonnées et font office de modèle.

Néanmoins, depuis quelques années, la pertinence des thésaurus PACTOLS est contestée. Emmanuelle Morlock démontre que, malgré la spécificité scientifique légitime des PACTOLS, la caractérisation des thésaurus en archéologie ne relève peut-être pas du travail des documentalistes. Le détail de description et la sélection de termes précis doivent faire l'objet d'un vrai consensus commun, mené, par exemple, par le réseau FRANTIQ et le consortium MASA puis validé par les archéologues eux-mêmes. Selon Emmanuelle Morlock, ce premier travail permettrait ainsi de refonder les thésaurus PACTOLS et de les appréhender différemment : « *A mon sens les thésaurus PACTOLS doivent se penser comme des outils permettant l'interopérabilité et la découvrabilité. Les vocabulaires scientifiques, les consensus, les négociations et les débats terminologiques sont trop nombreux, il faudrait un décret ou une instruction ministérielle pour répondre définitivement à ce problème* »⁵³. Une norme serait donc une des solutions à privilégier et bien d'autres thésaurus, spécifiques à l'archéologie, soient encore en cours d'élaboration, de relecture et adaptés aux standards pour l'interopérabilité, d'autres, à vocation plus universelle, se développent⁵⁴.

Backbone Thesaurus, un thésaurus global pour les sciences humaines est actuellement en cours de développement par le groupe Maintenance de Thesaurus⁵⁵. Leur objectif est de pouvoir identifier des concepts de haut niveau, dans le but de trouver une base commune à toutes les sciences humaines, étant donné que l'un des principaux problèmes reprochés aux PACTOLS, est leur manque de précision. Néanmoins, les archéologues ne semblent pas réellement se préoccuper de leurs développements et leurs enrichissements, affichant même, selon Emmanuelle Morlock, un réel désintérêt pour leurs usages : « *Il faut donc un consensus général dès le départ car les chercheurs ne respectent pas l'utilisation des thésaurus PACTOLS, ils le laissent de côté et produisent leur propre vocabulaire. Il vaut mieux se restreindre à utiliser un vocabulaire commun et contribuer à son enrichissement que chercher à créer le sien dans son coin* »⁵⁶. L'idée de la généralisation d'un thésaurus pour toutes les sciences humaines s'oppose à une volonté de précision et de spécificité, propre à cette discipline qu'est l'archéologie. Cependant, la volonté interdisciplinaire de

⁵³ Voir Annexes, Entretien avec Emmanuelle Morlock.

⁵⁴ On peut par exemple citer le thésaurus Antiquité TheA qui est toujours en cours de développement par le LIMC-France et Ortolang (Open Ressources and TOols for LANGuages), un vocabulaire comprenant 1960 entrées couvrant les domaines scientifiques en art et archéologie.

⁵⁵ Disponible sur le Web : < <http://backbonethesaurus.eu/> >.

⁵⁶ Voir Annexes, Entretien avec Emmanuelle Morlock.

BackBone Thesaurus, repose essentiellement sur l'affranchissement de la hiérarchisation des thésaurus classiques (tel que les PACTOLS). Ce nouveau thésaurus tend donc vers une forme plus ontologique – basée sur le CIDOC-CRM – offrant ainsi plus de possibilités, de précisions et de contextualisation. Le consortium MASA présente ce projet comme un futur modèle vers lesquels les thésaurus PACTOLS pourraient s'orienter. Le modèle de l'ontologie comme garantie de l'interopérabilité est une solution plus en plus privilégiée.

Ontologie(s) : l'application du CIDOC-CRM

« Pour garantir l'interopérabilité il faut des vocabulaires partagés, contrôlés et des références de haut-niveau tel que le CIDOC-CRM »⁵⁷. Lors d'un entretien accordé à Emmanuelle Morlock, l'ingénieure du CNRS tenait à rappeler que garantir l'interopérabilité des données en archéologie, ne passait pas uniquement par l'utilisation de vocabulaires contrôlés. Elle recommandait l'usage de norme de référence de haut niveau tel que le CIDOC-CRM (*International committee for documentation - Conceptual Reference Model*)⁵⁸. Depuis quelques années, l'utilisation d'interprétation commune de descriptions s'appuyant sur des vocabulaires partagés est une solution de plus en plus privilégiée par les professionnels de la Culture et du Patrimoine. Le CIDOC-CRM est un modèle conceptuel de référence et une norme internationale (ISO 21127:2014 - *Information and documentation — A reference ontology for the interchange of cultural heritage information*) pour l'information muséographique, publié pour la première fois en 1998. Ce modèle sémantique de référence constitue une ontologie de l'information relative au patrimoine culturel. Elle a pour vocation de formaliser les relations unissant des concepts fondamentaux d'information patrimoniale⁵⁹, tout en fournissant des définitions et une structure formelle pour décrire des relations implicites ou explicites utilisées dans la documentation sur le patrimoine culturel.

Bien que la polysémie du mot « ontologie » recouvre plusieurs significations, sa compréhension reste la même chez les professionnels de l'information et les archéologues. Une ontologie désigne à la fois, un vocabulaire commun aux chercheurs ayant besoin de partager l'information dans une discipline et une description formelle de concepts d'un domaine d'un discours, par des propriétés, des caractéristiques et des attributs. Tandis qu'en informatique et en sciences de l'information, une ontologie est un ensemble structuré de concepts représentant le sens d'un champ

⁵⁷ Voir Annexes, Entretien avec Emmanuelle Morlock.

⁵⁸ What is the CIDOC-CRM? [en ligne], [consulté le 10 août 2018]. Disponible sur le Web : < <http://www.cidoc-crm.org/> >.

⁵⁹ BnF. Modèle CIDOC-CRM [en ligne], 18 août 2017, [consulté le 9 août 2010]. Disponible sur le Web : < http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.modele_cidoc_crm.html >.

d'informations (métadonnées, éléments d'un domaine de connaissances...), elle constitue un modèle de données représentatif d'un ensemble de termes et de relations entre concepts dans un même domaine. Devenue une composante centrale pour un grand nombre d'applications, l'ontologie joue un rôle considérable dans le développement du Web des données⁶⁰.

Le consortium MASA, pionnier dans le développement de solutions adaptées aux enjeux recouverts par l'interopérabilité des données en archéologie, s'interroge depuis quelques années sur la structuration entre bases de données et ontologie. L'idée privilégiée du « mapping » serait de mettre en correspondance, de rattacher les champs d'une structure à une propriété de l'ontologie permettant de se référer à une structure commune avec un vocabulaire unique. Ainsi, l'interrogation des bases de données se ferait *via* l'ontologie permettant au moteur de recherche de retranscrire la requête formulée par l'utilisateur sur la base de données. Enfin, les résultats seraient renvoyés au demandeur tout en respectant l'organisation de l'ontologie. L'interrogation simultanée de plusieurs bases de données, possible par l'utilisation d'une ontologie, est le plus haut degré d'interopérabilité attendue par les membres du consortium MASA⁶¹. Depuis 2014, des développements sont en cours, comme au LAT de la MSH Val-de-Loire à Tours, pour permettre d'interroger la base de données ArSol à partir de l'ontologie du CIDOC-CRM (figure 6).

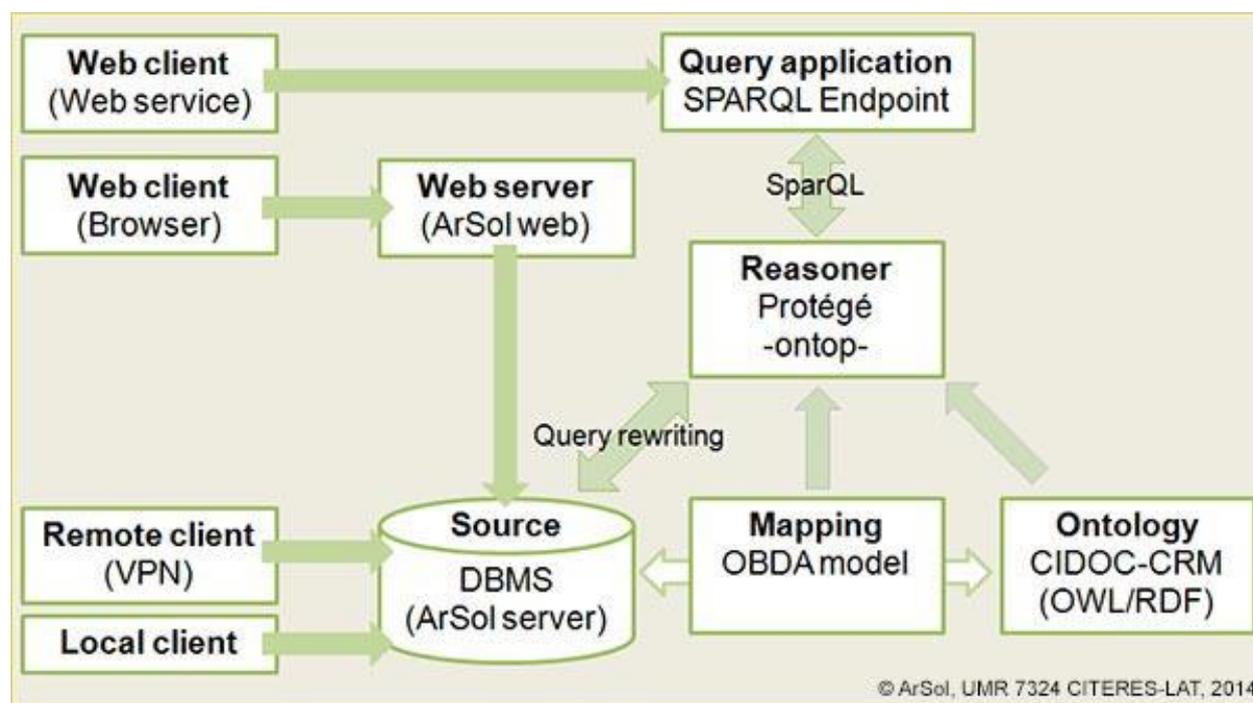


Figure 6 Schéma du principe de fonctionnement web de la base de données ArSol et son système d'interrogation via l'intermédiaire du CIDOC-CRM.

Source : MARLET Olivier. Une ontologie pour rendre interopérables des bases de données en archéologie ? in *Carnet de recherche du consortium MASA*, 19 août 2014.

⁶⁰ BnF. Elaboration d'ontologies [en ligne], 5 février 2015, [consulté le 9 août 2010]. Disponible sur le Web : < http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.elaboration_ontologies.html >.

⁶¹ MARLET Olivier. Une ontologie pour rendre interopérables des bases de données en archéologie ? [en ligne], 19 août 2014, [consulté le 9 août 2018]. Disponible sur le Web : < <https://masa.hypotheses.org/60> >.

La fonction même d'une ontologie, en informatique, est de fournir une structure à l'intérieur de laquelle des jeux de données différents peuvent interagir. Initialement, le CIDOC CRM a été créé pour donner un cadre favorable à l'harmonisation des données, permettant aux institutions culturelles de rendre compatibles leurs documentations sans perte d'informations et de précision. L'utilisation de l'ontologie CIDOC-CRM exige une certaine qualité de l'information traitée, les domaines concernés par son application sont uniquement les informations et objets relatifs au patrimoine culturel dont l'archéologie, au même titre que l'ethnographie et les arts, fait partie. La large couverture du CRM permet d'envisager cet outil comme le moyen idéal pour rassembler, de manière normée, des données archéologiques de toutes formes⁶². Le recours à l'application du CIDOC-CRM de manière systématique, donnerait non seulement un modèle commun à tous les archéologues mais aussi une couverture, non-négligeable, pour toutes les données produites lors d'une fouille. Le modèle anglais, CRM-EH, conceptualisé et développé lors de l'élaboration du projet STAR (*Semantic Technologies for Archeology Resources*), conçu pour rendre les ressources archéologiques interopérables, se présente comme une extension de cette ontologie. Le CIDOC-CRM permet alors de créer et de faciliter des points de jonction dans les bases de données, de véritablement modéliser les jeux de données archéologiques. Les jeux de données organisent une structure, cette même structure contient des données parfois bien différentes les unes des autres (lieu de fouille, objet, date de découverte, période chronologique estimée, dimension...). Le formatage de données et de métadonnées, c'est le propre des bases de données, on peut donc considérer qu'une base de données est un jeu de données, une collection de plusieurs unités de données organisées entre elles et accessibles par une plateforme ou un moteur de recherche.

Lors de son intervention aux journées consacrées à l'interopérabilité des données archéologiques, le 20, 21 et 22 novembre 2017 à Tours au consortium MASA, Thomas Francart a rappelé que les ontologies n'étaient pas uniquement des modèles de vocabulaires partagés pour structurer les données mais pour les partager⁶³. En insistant sur ce point, Thomas Francart s'attarde sur le caractère participatif des ontologies au web des données, vecteur essentiel du développement et de la diffusion des données. Comme nous l'avons exposé, en archéologie, le partage des données passe sous plusieurs formes : plateformes, bibliothèques numériques, catalogues en ligne, sites

⁶² SZABADOS Anne-Violaine et LETRICOT Rosemonde. L'ontologie CIDOC CRM appliquée aux objets du patrimoine antique, in 3e Journées d'Informatique et Archéologie de Paris - JIAP 2012, [en ligne], juin 2012, [consulté le 9 août 2018]. Disponible sur le Web : < https://halshs.archives-ouvertes.fr/file/index/docid/752996/filename/SZABADOS_LETRICOT_ontologie_CIDOC_CRM_appliquee_aux_objets_du_patrimoine_antique_2012.pdf >.

⁶³ FRANCAERT Thomas. Ontologies et thesaurus, l'interopérabilité grâce au web des données, [en ligne], 21 novembre 2017, [consulté le 8 août 2018], 32 :38 minutes. Disponible sur le Web : < https://www.canal-u.tv/video/msh_val_de_loire/ontologies_et_thesaurus_l_interoperabilite_grace_au_web_de_donnees_thomas_francart_sparna_expert_en_technologies_semantiques.39735 >.

internet et rapports d'opération sont les modes de diffusion favorisés. Depuis quelques années, l'INRAP publie ses données et résultats de recherche sous forme de rapport sur la plateforme DOLIA (catalogue de fonds documentaires et bibliothèque numérique). Chaque ressource numérique publiée dans DOLIA est associée de métadonnées permettant d'enrichir les données archéologiques présentées dans les publications. Toujours en quête de développement, l'INRAP travaille depuis 2015 sur le déploiement de l'application EDArc pour une approche générique des données de terrain. L'objectif est de renouveler la manière d'acquérir les données de terrain et de les enregistrer lors des fouilles archéologiques. L'INRAP tenait à ce qu'EDArc fasse appel à plusieurs normes et standards relatifs aux modèles de représentation de SIA. Pour satisfaire ce critère, l'application utilise des métalangages comme XML (*Extensive Markup Language*), des thésaurus comme PACTOLS, des normes dès l'enregistrement des métadonnées de terrain comme l'ISO 15836-1:2017 (*Information and documentation – The Dublin Core metadata element set – Part 1 : Core elements*) et a été apparié au CIDOC-CRM, dans le but de « [...] de publier les données de terrain dans le Web sémantique dès lors que des décisions seront prises sur les modalités de la publication des données d'archéologie préventive en données ouvertes »⁶⁴. Cette innovation permettrait ainsi de travailler directement depuis l'étape de l'acquisition des données de terrain. Tournée vers une volonté d'*Open Data*, la mise en marche de cette application pourrait bien venir bouleverser, une nouvelle fois, les méthodologies de travail en archéologie préventive. Enfin, l'utilisation de normes pour garantir l'interopérabilité des données en archéologie peut aussi s'appliquer au domaine documentaire : le LIMC et le portail Claros en sont les meilleurs exemples. Tandis que le LIMC a recours à une adaptation sémantique du CIDOC-CRM-France⁶⁵ pour la facilitation de l'interopérabilité et la mise en relation de son information scientifique avec d'autres sources d'informations, Claros l'utilise pour son système d'interrogation des informations mises en ligne sur différents sites internet. L'interrogation porte essentiellement sur les données et métadonnées correspondant aux « classes » (concepts) et « propriétés » (interactions possibles entre les concepts) du CIDOC-CRM. Ainsi, les données relatives à un artefact archéologique peuvent être réparties en informations : l'objet (type, matériau, dimensions, lieu de découverte), ornementation

⁶⁴ SZABADOS Anne-Violaine. L'usage de normes et de thésaurus pour les données archéologiques et les métadonnées, de la production sur le terrain à la publication des rapports d'opération : l'expérience de l'Inrap, [en ligne], 23 mai 2018, [consulté le 10 août 2018]. Disponible sur le Web : < <https://archeonum.hypotheses.org/738> >.

⁶⁵ SZABADOS Anne-Violaine. L'usage de normes pour l'interopérabilité des données sur l'Antiquité : l'exemple des outils documentaires du LIMC, in Journée d'études pour les professionnels de la BAPF en SHS, Paris : 25 novembre 2010.

(iconographie...), interprétation (date de création, origine...), documentation (publications, références, archives...) et conservation (inventaire, localisation actuelle ou passée...)⁶⁶. Claros utilise ainsi des équivalences de l'ontologie CIDOC-CRM pour permettre des interrogations.

L'homogénéisation sémantique, possible par la structuration de l'information (CIDOC-CRM) et par la standardisation du vocabulaire (PACTOLS), permet le développement d'une interopérabilité à plusieurs niveaux. Réutilisable dans un contexte de diffusion et de valorisation des données archéologiques, les normes et les standards ne sont pas là uniquement pour privilégier l'échange et la connexion entre les données, en effet : « [...] *l'interopérabilité des données n'est pas la seule raison motivant l'utilisation d'ontologies ; elles peuvent également s'insérer à différents stades d'un projet de création de systèmes d'informations, que ce soit pour établir une nomenclature commune entre différentes disciplines ou pour servir de socle lors de comparaison de jeux de données pour un même domaine* »⁶⁷. Car en archéologie, les jeux de données (bases de données, inventaires...) et l'utilisation de la redocumentarisation, semblent peut-être cristalliser de nombreux points communs aux enjeux recouverts par l'interopérabilité.

BASES DE DONNEES ET REDOCUMENTARISATION

Réutiliser les données et les métadonnées : l'unique défi de la redocumentarisation ?

La redocumentarisation peut-elle être un des enjeux recouverts par l'interopérabilité dans le cadre de la réutilisation des données ? Quel impact la redocumentarisation peut-elle avoir dans leur diffusion et leur valorisation ?

Défi pour les sciences de l'information, réutilisation et redocumentarisation cristallisent certains questionnements sans apporter de réelles réponses. Leur point de convergence, la réutilisation et la redocumentarisation le trouvent en archéologie dans le traitement accordé

⁶⁶ SZABADOS Anne-Violaine. Du système documentaire du LIMC au portail Claros. Interopérabilité et optimisation de l'information archéologique grâce à l'usage de normes, in *Archeologia e Calcolatori Supplemento 3*, [en ligne], 2012, [consulté le 10 août 2018]. Disponible sur le Web : < <https://docplayer.fr/2313143-Du-systeme-documentaire-du-limc-au-portail-claros-interoperabilite-et-optimisation-de-l-information-archeologique-grace-a-l-usage-de-normes.html> >.

⁶⁷ SZABADOS Anne-Violaine et LETRICOT Rosemonde. L'ontologie CIDOC CRM appliquée aux objets du patrimoine antique, in 3e Journées d'Informatique et Archéologie de Paris - JIAP 2012, [en ligne], juin 2012, [consulté le 9 août 2018]. Disponible sur le Web : < https://halshs.archives-ouvertes.fr/file/index/docid/752996/filename/SZABADOS_LETRICOT_ontologie_CIDOC_CRM_appliquee_aux_objets_du_patrimoine_antique_2012.pdf >.

des données. Tandis que la réutilisation permet de regrouper différents jeux de données, d'analyser des données déjà produites, de vérifier la qualité et l'exploitabilité des données dans un souci de transparence de la recherche scientifique, d'innover, de mutualiser l'utilisation des données déjà produites, la redocumentarisation vise une sorte de restitution des données, de la manière la plus exhaustive qui soit. La reprise de données déjà exploitées implique généralement une préparation scientifique stricte ; essentielle si l'on veut pouvoir migrer d'un support à un autre et proposer une diffusion et une valorisation par l'intermédiaire de l'interopérabilité. Manuel Zacklad impacte à le phénomène de la redocumentarisation un devoir de documentarisation renouvelé, permettant le réagencement de contenus sémiotiques dans une dimension interne (extraction de séquences, annotations en marge d'un livre...) ou dimension externe (organisation d'une collection, d'une archive, d'un catalogue croisant des ressources selon une nouvelle logique d'association)⁶⁸. Quant à Jean-Michel Salaün, il définit la redocumentarisation comme l'action de « [...] *traiter à nouveau des documents traditionnels qui ont été transposés sur un support numérique en utilisant les fonctionnalités de ce dernier.* »⁶⁹. Selon lui, la redocumentarisation suit la même logique que la documentarisation, qui, contrairement à la documentation d'objets, consiste à « [...] *optimiser l'usage du document en permettant un meilleur accès à son contenu et une meilleure mise en contexte.* ». Le numérique intervient donc dans le processus de redocumentarisation : le support du document, devenu instable, évolue. Ce renouvellement du traitement documentaire suppose, par essence, une relecture, une réinterprétation entraînant une réutilisation des données et métadonnées.

La redocumentarisation est un terme récemment théorisé par le collectif Roger T Pédaque proposant trois facettes constitutives à un document : la forme (objet matériel ou immatériel, structure manipulable), le texte (le sens) et le médium (statut social du document, trace d'une communication)⁷⁰. Cette approche permet de pouvoir mieux appréhender la transformation qu'a connu le document ces dernières décennies lors du passage au numérique. L'objectif est de faciliter la compréhension et l'utilisation des nouvelles possibilités offertes par la dimension numérique du document. Bien que la forme change selon les supports de lecture, l'information, le contenu et donc la donnée reste la même. Tout en s'inscrivant dans un système documentaire classique (classement, indexation, accès), la redocumentarisation

⁶⁸ Manuel Zacklad est l'un des premiers à théoriser en France le processus de redocumentarisation.

⁶⁹ SALAÜN, Jean-Michel. La redocumentarisation, un défi pour les sciences de l'information, in *Études de communication* [en ligne], 2007, mis en ligne le 1^{er} octobre 2009, [consulté le 25 juillet 2018]. Disponible sur le Web : < <http://journals.openedition.org/edc/428> >.

⁷⁰ PÉDAUQUE Roger T. Le document à la lumière du numérique : forme, texte, médium : comprendre le rôle du document numérique dans l'émergence d'une nouvelle modernité. Caen : C&F éd, 2006.

s'applique, quant à elle, uniquement au document nativement numérique. L'importance de l'information d'une information – métadonnées – réside dans sa capacité d'enrichissement du contenu et la simplification de son accès. De manière générale, la redocumentarisation permet de réarticuler des contenus selon leur interprétation et leur usage. Elle offre aux données, la possibilité de pouvoir être clairement et rapidement identifiables. Tandis que l'indexation, la classification et les thésaurus documentarisent les données, les protocoles web (HTML, URL), les ontologies et le web sémantique redocumentarisent. Le numérique, venu bouleverser les pratiques, influe directement sur la manière de concevoir l'information et de la renseigner. Redocumentariser n'implique donc pas uniquement de transposer des données d'un support à un autre, mais de stabiliser et de maintenir un accès. Cette autre manière de relier des contenus, d'enrichir d'annotations, de différencier les métadonnées associées ou de séquencer les données d'une même série, c'est offrir un nouvel accès, une nouvelle porte sur d'autres résultats. La réflexion autour de la redocumentarisation montre bien que désormais, le centre d'intérêt n'est plus le document, mais les données qu'il porte.

En archéologie, la redocumentarisation intervient surtout dans le cadre de la réutilisation des données et métadonnées et notamment dans le domaine de la cartographie et de la géolocalisation. L'atlas archéologique de Touraine en est un bon exemple⁷¹. Ce projet propose deux volets, une publication électronique faisant un état des lieux des connaissances actuelles du territoire de Touraine, de la Préhistoire à nos jours et une plateforme de *webmapping*, permettant de croiser les données présentes dans les notices pour produire de nouvelles cartes. Par la redocumentarisation, les données extraites de différents documents (cartes anciennes, carnets de fouilles, relevés topographiques, documents d'archéozoologie) sont transposées dans un environnement numérique et réutilisées dans le cadre de nouvelles actions, ici la cartographie. Pour la publication électronique, l'objectif est de pouvoir présenter des corpus de sources et d'évaluer leur représentativité spatiale⁷². Quant à la plateforme de *webmapping*, les couches superposables correspondent à l'ensemble des informations de l'atlas. Selon la recherche formulée, sont simultanément interrogées la base de données et l'analyse des données permettant de produire une nouvelle carte. De ce fait, ce ne sont pas uniquement les données qui sont réutilisées mais aussi les métadonnées. Cette réutilisation massive des données et métadonnées archéologiques permet, non seulement une réouverture de certaines données mais aussi le croisement avec d'autres données, issues de

⁷¹ Atlas archéologique de Touraine. Disponible sur le Web : < <http://a2t.univ-tours.fr/> >.

⁷² ZADORA-RIO Elisabeth. Atlas Archéologique de Touraine. *Supplément à la Revue Archéologique du Centre de la France*, 53, 2014, FERACF, Tours.

travaux extérieurs au domaine de l'archéologie. La réutilisation des données archéologiques est très prisée dans toutes les démarches de géolocalisation et spatialisation du territoire ou de l'espace urbain et périurbain. Les données issues des fouilles menées par l'INRAP sont d'ailleurs exploitées par le gouvernement français et mises à disposition des utilisateurs via des plateformes internet⁷³. Néanmoins, la mise à bien de ce genre de projet n'est pas toujours possible. En effet, la multitude de données archéologiques et ces spécificités impliquent une certaine rigueur en terme d'organisation. Bien souvent, avant d'étudier les données produites et collectées, il faut les rassembler et les identifier afin de pouvoir les inventorier : ce sont les inventaires réglementaires. La réutilisation des données des inventaires réglementaires (unités stratigraphiques, prélèvements, données issues de documents graphiques...) pose, selon Anne Speller, certaines difficultés : si les inventaires réglementaires ont la capacité de recenser des données descriptives, des données brutes et des données primaires, le manque d'accès à ces données empêche malheureusement leur réutilisation (figure 7)⁷⁴.

Données		Mode	Format	Réf. Carto	Date	Echelle	Résolution
SCAN25	Lille	Raster	TIF/shp	Lambert 1 (Carto)	1993	1/25 000	2,5 m
	Carvin				1993		
BD Carto		Vecteur	Arc info	Lambert 2 étendu	2000	1/100 000	Précision décamétrique
BD Alt		Vecteur	shp	Lambert 2 étendu	1999	-	Grille de pas de 50 m Précision métrique
Carte géol.	Lille	-	papier	Lambert 1 (Carto)	1967	1/50 000	50 m
	Carvin				1968		
SPOT 4		Raster	DAT	-	Av. 2002	-	20 m
Orthophotos coul.		Raster	TIF/shp	Lambert 1 (Nord)	2001	-	20 cm
		Raster	TIF/shp	Lambert 1 (Nord)	Jlt. 1998	-	50 cm
Photographies Infrarouge coul.		Raster	ECW	-	2001	-	50 cm
Cadastré numérisé		Vecteur	shp	Lambert 1 (Nord)	2001	-	parcellaire

Figure 7 Exemples de données archéologiques produites, collectées et réutilisées selon leur mode, leur format et leur nature.

Sources : FRANCHOMME Magalie, SALVADOR Pierre-Gil et KERGOMARD Claude. Conception et utilisation d'un SIG pour l'étude des données (paléo-) environnementales et archéologiques dans le bassin versant de la haute Deûle », *Revue du Nord-Archéologie*, 85 (353), 2003, p. 35-45.

⁷³ La plateforme data.gouv met à disposition un jeu de données sur la localisation des sites de fouilles archéologiques de l'INRAP. Disponible sur le Web : < <https://www.data.gouv.fr/fr/datasets/localisation-des-sites-de-fouille-archeologiques-de-l-inrap-idf/> >.

⁷⁴ SPELLER Anne et COLIN Camille. Quelles données pour quels projets : perspectives de l'Open Access pour l'archéologie préventive, [en ligne], 23 mars 2017, [consulté le 1^{er} août 2018], mis à jour le 30 novembre 2017, 21:31 minutes. Disponible sur le Web : <https://www.inrap.fr/quelles-donnees-pour-quels-projets-perspectives-de-l-open-access-pour-l-13180>

Le défi de la redocumentarisation ouvre pourtant quelques perspectives intéressantes, mais encore peu exploitées en matière de valorisation et de diffusion des données archéologiques. La capacité de la redocumentarisation à prendre des informations existantes et à les transcoder sous une nouvelle forme (numérique) nécessite une ouverture accès. Ce retour incessant à l'*Open Access* n'est pas sans rappeler le besoin de rendre les données, même lors d'une réutilisation, interopérables. Cependant l'archéologie, en tant que science humaine et sociale, a encore de nombreux progrès à faire afin de pouvoir faire intervenir pleinement le processus de redocumentarisation dans sa démarche de valorisation et de diffusion des données. La question de l'interopérabilité peut se poser dans le cadre d'une redocumentarisation notamment en archéologie à cause de la nature première des données. Leur mode de collecte et de production est nettement désavantagé par rapport aux sciences dites « dures » où l'encodage de données liées se fait automatiquement à partir des appareils utilisés. En archéologie seules les données issues des relevés topographiques ont le privilège d'être nativement numériques, quant aux autres données, elles subissent nécessairement un traitement numérique avant d'être exploitées. La redocumentarisation, tout en renouvelant le traitement accordé aux données, crée, par extension, un nouvel ensemble prêt à être réexploité. Ce nouvel ensemble peut parfaitement désigner une réutilisation de données dont l'un des premiers enjeux est de créer des corrélations et des connaissances nouvelles⁷⁵.

Alors que la réutilisation des données repose essentiellement, en archéologie, sur la constitution de bases de données, leur utilisation n'est possible que par l'interopérabilité.

Les bases de données en archéologie : l'interopérabilité au service de la diffusion et de la valorisation

En archéologie et dans toutes sciences sociales ou humaines faisant appel au numérique dans le traitement, la gestion et l'acquisition de leurs données, l'interopérabilité et les jeux de données représentent des enjeux importants. En effet, la création d'un modèle commun de jeu de données *minimum* pour faciliter l'interopérabilité des bases de données en

⁷⁵ BROUDOUX Evelyne et SCOPSI Claire. L'enjeu des métadonnées dans un contexte de "redocumentarisation", in *Etudes de communication*, [en ligne]. 2011, [consulté le 2 août 2018]. Disponible sur le Web : <https://www.researchgate.net/publication/281156482_L%27enjeu_des_metadonnees_dans_un_contexte_de_redocumentarisation>.

archéologie est un enjeu essentiel. D'ailleurs, l'archéologie est le secteur scientifique ayant le plus recours aux bases de données, au point que l'estimation de leur nombre et leur variété est inquantifiable. On identifie les bases de données opérationnelles (chaque opération de fouille archéologique possède sa propre base de données), chronologiques (par période historique), thématiques (compilation et étude de la céramique, de la numismatique, du petit mobilier, de l'archéozoologie...), administratives (on recense une base de données pour une équipe de recherche) et documentaires (documents bibliographiques, photographies, photocopies d'ouvrages...)⁷⁶. De manière générale, il existe autant de bases de données qu'il existe de période chronologique, de lieu d'étude et de laboratoires/équipes/projets de recherche. Cet outil de travail est aussi le premier outil de valorisation des données archéologiques : les bases de données sont les premiers modes interactifs mobilisés par le grand public et la communauté scientifique pour la recherche d'informations. Les archéologues l'utilisent avant tout pour répondre à une nécessité réglementaire (nécessité pour les fouilles d'archéologie préventive de publier et diffuser leurs résultats de recherche) et le grand public y a recours pour s'informer et développer de nouvelles connaissances.

Avant les années 2000, le développement des bases de données comme outils de travail puis de diffusion et de valorisation des données occupait déjà l'esprit des archéologues. En 1990, René Ginouvès, spécialiste de l'étude de l'architecture grecque et précurseur des nouvelles méthodes de traitement de l'information en archéologie, posait déjà la question du futur des banques de données et des problèmes liés à leur interrogation. Auparavant, les données étaient compilées dans une banque, sans être reliées entre elles et interrogeables⁷⁷. Même si leur multiplication se présentait comme une avancée importante dans la pratique de l'archéologie, certaines barrières, comme la consultation, empêchait une utilisation optimale. Heureusement, les bases de données relationnelles assez rapidement développées, remplacèrent les banques de données. Dans une base de données relationnelle, l'information est organisée dans des tableaux aux dimensions multiples appelées relations ou tables. En archéologie, ce système facilitant la mise en liant des données, permet de pouvoir faire apparaître plusieurs éléments simultanément. Les archéologues utilisent les bases de données relationnelles quotidiennement malgré certaines contraintes (absence de vrais référentiels, thésaurus peu pertinents...), leur choix se porte d'ailleurs presque exclusivement sur le

⁷⁶ THIVET Matthieu. Bases de données archéologiques. Problématiques, contraintes, enjeux, [en ligne], [consulté le 12 août 2018]. Disponible sur le Web : < <https://docplayer.fr/6369056-Base-de-donnees-archeologiques.html> >.

⁷⁷ GINOUVES René. Des banques de données pour l'archéologie ? *in* Traitement de l'information en archéologie, Brises : 1990, CNRS, p.97-107, [en ligne], [consulté le 12 août 2018]. Disponible sur le Web : < <http://www.mae.u-paris10.fr/ginouves/texteimpr/AR16.html> >.

logiciel FileMakerPro pour leur conception. Néanmoins, ce recours systématique ne facilite pas la diffusion, la valorisation et l'archivage des données à long terme, elle ne fait que compiler les données produites. Anne-Marie Guimier-Sorbets reprochait déjà, à la fin des années 1990, le manque de connexion des banques ou bases de données en archéologie, conçues plus comme des « manuels » et des « dictionnaires » que comme des entités capables de s'interconnecter. Ces banques de données ont d'ailleurs, toujours été élaborées à part des données qu'elles recevaient : « *Les concepteurs et les chercheurs qui consacrent du temps à la constitution de ces outils n'ont pas besoin d'innover à tout prix pour justifier leur travail. Ce sont les données qui comptent avant tout, le support vient après.* »⁷⁸. Malheureusement, ce modèle a longtemps persisté nuisant à leur performance et aujourd'hui encore, une grande partie des bases de données en archéologie ne respectent pas les normes de l'interopérabilité : formats ouverts (.pdf/A, .tif...), standards (Dublin Core), indexation à l'aide des thésaurus (PACTOLS), normes pour les métadonnées (ISO 19115, ISO 19139) ...

Pourtant, depuis les années 1990, la valorisation du patrimoine archéologique revêt un enjeu politique et sociétal. De ce fait, la mise en valeur et la constitution de répertoire identifiant les bases de données archéologiques en Europe et en France se sont présentées comme des objectifs de premier ordre pour le ministère de la Culture jusqu'à être inscrit dans le Code du Patrimoine⁷⁹ à l'article L 522-5 (livre V, chapitre 2, section 1) : « [...] *Avec le concours des établissements publics ayant des activités de recherche archéologique et des collectivités territoriales, l'Etat dresse et met à jour la carte archéologique nationale. Cette carte rassemble et ordonne pour l'ensemble du territoire national les données archéologiques disponibles.* ». Cette compétence, inscrite depuis 2004 dans le Code du Patrimoine montre bien la préoccupation de l'Etat à mettre en valeur les données archéologiques et cela passe notamment par le recensement des différentes bases de données.

L'un des plus grands paradoxes des bases de données en archéologie réside dans leur propre spécificité : les données mises en valeur le sont généralement au détriment d'autres. Les bases de données isolent certaines données tout en y donnant accès ; les données sont donc sélectionnées indépendamment des autres. Le programme européen ARIADNE est l'un des meilleurs exemples de diffusion et de valorisation des données archéologiques. Ce

⁷⁸ BUCHSENSCHUTZ, Olivier. Anne-Marie Guimier-Sorbets, Les bases de données en archéologie, Conception et mise en œuvre (compte-rendu), in *Histoire & Mesure*, [en ligne], 1994, [consulté le 13 août 2018], n°3-4. Disponible sur le Web : < http://www.persee.fr/doc/hism_0982-1783_1994_num_9_3_1713 >.

⁷⁹ Le site officiel du ministère de la Culture a mise en ligne une liste (non exhaustive) des bases de données en archéologie : < <http://www.culturecommunication.gouv.fr/Thematiques/Archeologie/Ressources/Bases-de-donnees> >.

programme dédié à leur traitement numérique propose plusieurs services et outils en ligne aux utilisateurs tel que des portails d'accès aux données, des images RTI, de la consultation des données en 3D, recensement des vocabulaires...). Ce programme a pour objectif de diffuser, de rendre accessible certaines données de l'archéologie mais surtout, d'accompagner le travail des archéologues et de faciliter l'interopérabilité, la gestion et la conservation des données archéologiques à l'échelle européenne⁸⁰. L'avantage de ce programme réside dans sa faculté à pouvoir avoir les moyens suffisants pour pouvoir rassembler des collections de données et divers éléments associées entre eux et accessibles individuellement ou de façon combinée. Quelques projets de bases de données valorisant les données archéologiques et facilitant l'interopérabilité ont vu le jour depuis 2010 :

- le projet ISTHAR : outil de gestion et d'inventaire en ligne du mobilier archéologique permettant un enregistrement et une consultation des données de terrain pour les archéologues, les professionnels des musées et des laboratoires. Ce projet, archivé depuis quelques années, avait pour objectif de donner plus de visibilité à la documentation archéologique en offrant un accès direct aux étudiants, aux chercheurs ou à toute personne s'intéressant au patrimoine archéologique ;
- le projet Epicherchel : projet d'édition d'inscriptions antiques de Césarée de Mauritanie dont la base de données est accessible depuis un site. Cette base de données, accompagnant un projet de publication de corpus d'inscriptions, a encore une fois pour objectif de réunir et mettre à disposition de la communauté scientifique les inscriptions étudiées et les photographies associées⁸¹.

En archéologie, la gestion des bases de données représente un véritable challenge que la communauté scientifique cherche à tout prix à relever, les bases de données sont le point d'ancrage de plusieurs enjeux. Normes, standards, thésaurus, ontologies, métadonnées, redocumentarisation et réutilisation, tout est mis en œuvre depuis ces dix dernières années pour permettre l'interopérabilité des données. De plus en plus de documents d'accompagnement à l'ouverture des données ont été promulgués par de grands organismes

⁸⁰ Les outils mis à disposition par le programme ARIADNE sont disponibles *via* le lien suivant : < <http://www.ariadne-infrastructure.eu/fre> >.

⁸¹ BAUDOIN Bruno et SATRE Stéphanie. Base de données de recherche et plateforme d'exposition de corpus épigraphiques : EpiCherchel et CoReA, in *Interopérabilités – Journées MASA*, [en ligne], 21 novembre 2017, [consulté le 29 août 2018]. Disponible sur le Web : < https://www.canal-u.tv/video/msh_val_de_loire/base_de_donnees_de_recherche_et_plateforme_d_exposition_de_corpus_epigraphiques_epicherchel_et_corea_stephanie_satre_et_bruno_baudoin_centre_camille_jullian_umr_7299_et_masa.39751 >.

préconisant certaines pratiques à adopter pour les équipes de recherche cherchant à formaliser des processus relatifs aux données tout en garantissant l'interopérabilité.

PRECONISATIONS ET BONNES PRATIQUES : DU PLAN DE GESTION DES DONNEES ET DES DIFFERENTS NIVEAUX DE L'INTEROPERABILITE A L'OPEN ACCESS

En appliquant certaines recommandations, l'interopérabilité des données est un facteur que l'on peut facilement garantir. Aujourd'hui, il n'existe pas en France de référentiels ou de principes présentant les bonnes pratiques à adopter pour garantir l'interopérabilité des données en archéologie. Que ce soit en archéologie préventive ou en archéologie programmée, chaque équipe ou laboratoire de recherche, chaque chercheur, chaque archéologue fonctionne indépendamment, sans chercher à vouloir regrouper les pratiques. Sans groupe reconnu donnant un cadre institutionnel ou normatif, il est assez difficile de regrouper les méthodologies de travail, de dialoguer sur les pratiques de chacun et de tenter d'homogénéiser les techniques. Bien que le consortium MASA – organisme pionnier sur la réflexion des pratiques en archéologie – organise régulièrement des journées d'étude thématique, les regroupements se font rares et les procédures évoluent lentement.

Pourtant, en tenant compte de certaines recommandations, il est possible de pouvoir suivre un cadre garantissant au mieux l'interopérabilité des données. Ce cadre, on peut le trouver si les archéologues se donnent la peine de suivre certaines pratiques novatrices impulsées dans différents instituts de recherche. Plan de gestion de données, principes FAIR, *Open Access* et réflexion autour des différents de niveaux d'interopérabilités se présentent comme de véritables règles de l'art.

CREER UN PLAN DE GESTION DES DONNEES

La recherche est valorisée seulement si elle est organisée. Malheureusement, il n'existe pas en archéologie, comme dans l'administration, de référentiel - tel que le RGI - fixant des recommandations de normes ou de standards à utiliser pour favoriser l'interopérabilité des données ou des systèmes d'information. Il est pourtant essentiel de déterminer des objectifs à atteindre pour permettre aux acteurs de la recherche en archéologie, d'interagir, de s'interconnecter ou même d'aller au-delà de simples arrangements informatiques bilatérales. Organiser, structurer la manière dont l'archéologue va traiter, diffuser, mettre à disposition les données depuis leur création c'est l'un des enjeux fixés par un document unique : le plan de gestion des données.

Le plan de gestion des données se définit comme : « [...] *un document formalisé explicitant la manière dont seront obtenues, documentées, analysées, disséminées et utilisées les données produites au cours et à l'issue d'un processus ou d'un projet de recherche. Le DMP s'appuie sur le*

cycle de vie des données/documents et décrit les choix réalisés en termes de normes de métadonnées, formats des bases de données, méthodes et sécurité d'accès, durée d'archivage, ainsi que les coûts associés à la gestion des données. »⁸². Le principal enjeu du plan de gestion des données est donc de permettre la valorisation et l'accessibilité des données produites au cours d'un projet de recherche. Initié et réfléchi au début d'un projet de recherche, il aborde divers éléments : dans une première partie, les utilisateurs doivent renseigner plusieurs informations relatives à l'administration du projet (responsable scientifique, nom de l'appel à projet, programme de recherche, titre du projet, convention de financement, propriété des données...) et ce n'est que dans une seconde partie – sous forme de tableur – que le plan de gestion s'intéresse aux différents éléments relatifs aux jeux de données :

- **la description** : les données collectées/produites et les jeux de données doivent faire l'objet d'une description précise (nature et nom des jeux de données, identifiant, méthode de production des données, réutilisation des données existantes, formats des données) ;
- **le stockage** : les données sauvegardées pendant et après la phase de projet sont soumises à diverses descriptions (support, volumétrie et type d'hébergement des données) ;
- **la sécurité** : les données produites doivent faire l'objet d'un minimum de sécurité (risques ou menaces sur les données, garantie de confidentialité des données, garantie d'intégrité et de traçabilité des données) ;
- **l'accès** : les données produites ou récoltées doivent pouvoir être accessibles (lecture des données, garantie de disponibilité, gestion des accès, échanges et partage) ;
- **les métadonnées** : les données seront logiquement accompagnées de métadonnées, le plan de gestion devra répondre à certains éléments (standards et formats disciplinaires des métadonnées, mode de production et responsabilité, arborescence de classement, règles de nommage des jeux de données et documentation associées) ;
- **la dissémination** : la dissémination implique les modes de partage de diffusion et de réutilisation des données (principe général de diffusion, type de licence, potentiel de réutilisation, existence de publications associées aux données, identification du lieu de dépôt et de dissémination des données) ;

⁸² CARTIER Aurore, MOYSAN Magalie et REYMONET Nathalie. Réaliser un plan de gestion des données [en ligne], 9 janvier 2015, [consulté le 20 août 2018]. Disponible sur le Web : < http://cache.media.education.gouv.fr/file/Comment_Participer/36/0/realiser_un_dmp_406360.pdf >.

- **la protection** : afin de répondre à des questions juridiques, éthiques, techniques et financières, les données peuvent nécessiter d'une protection échappant ainsi aux principes ou canaux habituels de diffusion (identification des jeux de données sensibles, justification du principe d'exception aux conditions générales de diffusion, embargo, mesures de protection) ;
- **l'archivage** : les données diffusées ou non à la fin ou au cours du projet de recherche doivent néanmoins faire l'objet d'une procédure d'archivage à court ou à long terme (sort des données à l'issue du projet, sélection des données, volume final, durée de conservation préconisée et choix de la plateforme ou solution d'archivage).

Dans un plan de gestion des données, la notion d'archivage est assez secondaire, même si cet aspect n'est pas occulté, l'accent est davantage mis sur le partage et la diffusion des données. Le plan de gestion des données est un document présentant essentiellement des informations sur la collecte, mode de collecte/création des données, les contraintes éthiques et légales, la documentation des données, l'organisation des fichiers, les conventions de nommage, les procédures de stockage, de sélection, de préservation et de modalités d'archivage des données, du niveau d'ouverture et de partage et les responsabilités dans la gestion des données et le coût. Bien que le PDG soit créé par le chercheur et se présente surtout comme une garantie pour le financeur du projet, cet outil d'aide à la décision se veut transparent au niveau scientifique, facilitant la réutilisation des données. Le cycle de vie est anticipé et l'histoire des données est documentée. Le PDG permet de montrer de quelle manière un projet de recherche gère ses données et métadonnées et comment les rendre facilement trouvables, accessibles et identifiables, quels standards, normes, formats ouverts et modes de diffusion vont être utilisés. En archéologie, le PDG a, par exemple, accompagné le projet ArchAIDE fondé et financé par le programme européen H2020. Le projet ArchAIDE⁸³ est un système de reconnaissance automatique de tessons de céramique. L'accent est mis sur l'accessibilité et la mise à disposition de données archéologiques spécifiques (ici la céramique) tout en créant de nouvelles connaissances grâce aux jeux de données.

L'établissement d'un plan de gestion de données dans le milieu de la recherche répond essentiellement aux demandes relatives au programme européen pour la recherche et l'innovation Horizon 2020. En effet, désormais tous les projets de recherche pilotes pour le libre accès des données et de la recherche devront se soumettre à l'élaboration d'un plan de gestion de données. Obligatoire pour les projets de recherche financés sur fonds publics, le plan de gestion reste

⁸³ Le projet est disponible sur le Web à l'adresse suivante : < <http://www.archaide.eu/> >.

cependant optionnel pour les autres. Il reste néanmoins préconisé. De plus, dans l'intérêt de garantir l'interopérabilité des données, il est conseillé aux équipes de portant un projet de recherche de conduire un plan de gestion de données respectant le principe FAIR.

Le principe FAIR (*Findable, Accessible, Interoperable, Reusable*) est défini par Emmanuelle Morlock comme : « [...] *le meilleur outil systématique pour garantir l'interopérabilité des données. Le principe fonctionne par grade, c'est un fonctionnement dans la progression, qui ne prend pas uniquement en compte la dimension purement technique de l'interopérabilité. Le principe FAIR s'attarde sur les moyens organisationnels à mettre en œuvre pour garantir l'interopérabilité, il faut donc des processus contre qualité pour pouvoir valider ce principe.* »⁸⁴. La notion de *FAIR data* renvoie directement au contexte de production et de diffusion des données ancré dans les dynamiques actuelles que sont *l'Open Science*, les *Big Data*, l'interopérabilité et l'accessibilité. L'interopérabilité s'y présente comme un facteur favorisant et complémentaire à l'ouverture, la réutilisation et la recherche (identification) des données. Dans le contexte du principe FAIR, l'interopérabilité est envisagée comme une faculté sémantique et syntaxique impliquant l'utilisation de métadonnées précises, de vocabulaires, de standards et de formats ouverts et internationaux.

Récemment, l'ENSSIB (Ecole des Sciences de l'Information et des Bibliothèques), l'Université Paris-Diderot 7 et l'Université Paris Descartes-Paris 5 ont réalisé en étroite collaboration, un document aidant à la rédaction d'un plan de gestion de données intégrant le principe FAIR. Il prend notamment en compte les travaux du collectif FORCE11⁸⁵ qui propose de décrire les données de façon FAIR en indiquant le traitement des données durant et après le projet, leur organisation, leur conservation, leur mode de partage, leur accessibilité, les normes appliquées et le type de données produites/collectées. Le principe FAIR permet ainsi au projet de répondre à certaines questions – parfois nécessaires – sur l'évolution du projet. L'objectif est de guider le responsable ou l'équipe projet à se poser les bonnes questions sur la garantie de l'interopérabilité, par exemple:

- les données sont-elles interopérables (capables d'être lues, échangées, exploitées par différents systèmes, plateformes ou applications) ?
- les données peuvent-elles être échangées et (ré)utilisées avec d'autres organismes ou pays ?
- les données sont-elles conformes aux formats préconisés, compatibles avec des logiciels ouverts, combinables avec des données provenant de différentes sources ?

⁸⁴ Voir Annexes, entretien avec Emmanuelle Morlock.

⁸⁵ FORCE11 est une communauté d'académiques, d'archivistes, de bibliothécaires et d'éditeurs organisés en comité pour faciliter la création et le partage des connaissances.

- quelles méthodologies et quels standards de données et métadonnées utiliserez-vous pour rendre les données interopérables ?
- utiliserez-vous des vocabulaires standards pour tous les types de données présents dans le jeu de données afin de permettre une interopérabilité interdisciplinaire ?
- dans le cas où vous devrez utiliser des ontologies ou des vocabulaires non standards, prévoyez-vous de faire des liens vers des ontologies ou des vocabulaires communs ?

Toutes ces questions offrent un véritable cadre de réflexion au responsable ou à l'équipe projet, propice à l'élaboration d'une ligne directrice de travail dans laquelle, les processus d'archivage, de diffusion, de valorisation des données produites et leur interopérabilité pourront être garantis. Afin de favoriser au mieux l'interopérabilité, ce document recommande :

- **documenter les données produites tout au long de leur cycle de vie** avec, par exemple, des métadonnées descriptives, des métadonnées de gestion et des métadonnées d'archivage structurées respectant certains formats (Dublin Core, DataCite Metadata Schema) ;
- **privilégier des formats libres et non propriétaire** (XML, CSV, PNG, X3D, PDF, ODS...) que ce soit dans le cadre de la réutilisation, de l'archivage ou de la diffusion ;
- **créer un format de métadonnées spécifiques**, si et seulement si celui-ci est décrit et documenté avec une grande précision en s'appuyant de préférence sur le langage XML ;
- **privilégier l'utilisation de licences libres** ou libre de diffusion⁸⁶.

Si l'interopérabilité est un facteur garanti, cela permettra d'assurer la pérennité des métadonnées, une utilisation plus efficace et pertinente des données et une facilitation dans les processus de réutilisation. En archéologie, l'utilisation de plan de gestion de données n'est pas encore un réflexe dans les méthodologies de travail. Cependant, cette tendance semble s'inverser, en effet, le 21 juin 2018 un groupe de travail s'est réuni à Lyon, au sein de la MOM, pour travailler sur un plan de gestion de données type pour les projets de recherche en archéologie.

⁸⁶ REYMONET Nathalie, MOYSAN Magalie, CARTIER Aurore, DELEMONTEZ Renaud. Réaliser un plan de gestion de données " FAIR " : modèle, [en ligne], 2018, [consulter le 21 août 2018]. Disponible sur le Web : < https://archivesic.ccsd.cnrs.fr/sic_01690547/document >.

Bien que l'archéologie tente de travailler à l'élaboration de méthodes adaptées à la spécificité des données qu'elle produit, plusieurs améliorations restent encore envisageables. Emmanuelle Morlock concluait d'ailleurs son interview sur le fait que l'archéologie se doit sans de : « [...] *tester l'organisation et la pertinence des données et de leurs descriptions, il faut réussir à organiser la description des données au moment de la collecte et les renseigner par lot ou par dossier tout en callant la description sur les processus de production. En archéologie, on est contraint par la synchronisation des méthodes, des fiches descriptives : il faut tout tester et ne pas envisager indépendamment un contexte de production. Par exemple : est-ce la donnée est interoperable dans l'écosystème actuel de valorisation et de diffusion ? Il faut mettre en regard le système de référence de l'interopérabilité et l'état de l'art aujourd'hui* »⁸⁷. Dès lors, un projet peut vouloir viser plusieurs niveaux d'interopérabilité afin de pouvoir garantir au maximum sa capacité à pouvoir diffuser et valoriser, le plus efficacement possible, ses données.

GARANTIR LES DIFFERENTS NIVEAUX D'INTEROPERABILITE

Depuis quelques années, l'usage et le développement d'Internet semblent avoir considérablement élargi la conception de l'interopérabilité. Clef de voûte de nombreux systèmes, l'interopérabilité est là pour maximiser l'efficacité des échanges et la circulation des informations. Grâce à Internet et à la mutualisation des données, les plateformes de diffusion et de valorisation, mises à disposition des chercheurs et des archéologues, fleurissent. Des ressources en ligne sont proposées afin de faciliter le partage et la mise en valeur optimale des données déposées. L'ouverture de la diffusion des données a ainsi permis à des établissements à l'échelle européenne, nationales et locales de développer leurs propres plateformes, enrichissant considérablement les savoirs. La mise à disposition d'un large panel de connaissances au public pose alors la question de l'interopérabilité. En effet, désormais il ne suffit plus de garantir les échanges d'informations entre les machines, logiciels, applications ou plateformes, il faut les rendre accessibles à la compréhension de tous et organiser la mise en commun des ressources, bien trop souvent éparpillées. Syntaxe, lexique et sémantique trouvent alors leur place au sein des systèmes informatiques.

De récents travaux en informatique et en science de l'information laissent alors penser qu'il serait plus aisé de parler des interopérabilités et non de l'interopérabilité. Les « Interopérabilités » ont d'ailleurs été le thème abordé par l'une des journées d'études du consortium MASA les 20, 21

⁸⁷ Voir Annexes, entretien avec Emmanuelle Morlock.

et 22 novembre 2017 à Tours, preuve que la communauté scientifique archéologique pose une vraie question sur le paradoxe du terme. Lors d'une interview avec Emmanuelle Morlock, l'ingénieure préférait déjà parler des interopérabilités : « *Selon moi, il y a deux interopérabilités : tout d'abord il y a l'interopérabilité selon les formats syntactiques et sémantiques.* »⁸⁸. Il existe donc plusieurs niveaux d'interopérabilité plus ou moins important en fonction du niveau visé. Les chercheurs poussent leur réflexion plus loin et parlent alors de niveaux d'interopérabilité :

- **un premier niveau d'interopérabilité « technique »** permettant de rendre possible la circulation, la communication des données entre les systèmes, les logiciels et les machines grâce à la définition de formats et d'interfaces standardisées ;
- **un deuxième niveau d'interopérabilité « syntactique »** concerne cette fois-ci les formats de niveau dit « inférieur » comme le formatage de données en XML ou SQL ;
- **un troisième niveau d'interopérabilité « lexicale » et « sémantique »** s'attachant à rendre interprétables les données échangées, on a alors recours aux standards et aux formats afin de caractériser les contenus sous forme de métadonnées⁸⁹. On utilise souvent le Dublin Core qui liste une série de descripteurs et sert de modèle.

Les données doivent ainsi pouvoir être facilement identifiables, partageables, interrogeables, tout restant intelligibles pour les machines et pour l'Homme. Les institutions de recherche, culturelles et patrimoniales ne cessent depuis des années de développer et de travailler sur les conditions d'échanges des données et métadonnées. L'identification et la description sont d'ailleurs au cœur de ces prérogatives. Les projets de valorisation et de diffusion des données, notamment en archéologie, mettent donc un point d'honneur à vraiment travailler sur les différentes facettes de l'interopérabilité. Bien que l'interopérabilité technique soit de plus en plus maîtrisée, celle-ci évolue chaque jour laissant les chercheurs et professionnels de l'information archéologique dépassées par les nouvelles avancées. Les institutions de recherche cherchent donc à optimiser et assurer au maximum les couches de l'interopérabilité, chacune représentant un enjeu. En l'absence de référentiel pour l'archéologie, les chercheurs peuvent s'appuyer sur le RGI.

Le RGI (version 1.9.7) présente le niveau d'interopérabilité technique en catégories (session, application, réseau et transport de l'information) et sous-catégories (transfert, exploitation,

⁸⁸ Voir Annexes, entretien avec Emmanuelle Morlock.

⁸⁹ FAVIER Laurence, MUSTAFA EL HADI Widad et VINCK Dominique. Présentation, Interopérabilité culturelle, in *Communication*, [en ligne], vol. 34/1 | 2016, mis en ligne le 17 août 2016, consulté le 27 août 2018. Disponible sur le Web : < <http://journals.openedition.org/communication/6614> >.

multimédia, services web, orchestration des services, accès et authentification) et liste les standards qu'il est préférable d'adopter (la liste présente ici n'est pas exhaustive) :

- pour le réseau, sont recommandés le IPv6 ou IPSec ;
- pour le transport, sont recommandés RTP, UDP, NTP et TCP ;
- pour la session, est recommandé, SSH ;
- pour le service, sont recommandés, SAMLv2, Open ID Connect ou WSDL ;
- pour l'application, sont recommandés et HTTP, HTTPS, DNS, FTP, CSW ou SFTP⁹⁰.

On notera que les protocoles libres sont généralement privilégiés.

Pour le niveau d'interopérabilité syntactique, c'est le même schéma de présentation que pour l'interopérabilité technique, le RGI l'évoque sous forme de catégories (encodage, document, web, structuration de données, traitement des données structurées, multimédia et signature) et en sous-catégories (caractère, impression, identifiant, géospatial, conteneur, image, audio, vidéo et calendrier). Là aussi, plusieurs standards sont recommandés :

- pour l'encodage, sont recommandés UTF-8, LMZA, ZIP, 7z et TAR ;
- pour le document, sont recommandés TXT, ODF, PDF et PDF/A ;
- pour le Web, sont recommandés HTML, XHTML, CSS, Javascript, ATOM, APP et CMIS ;
- pour la structuration des données, sont recommandés ARK, URI, GeoJSON, GeSpatial-Metadata, XQuery, XML Signature, OWS Context et SLD ;
- pour le multimédia, sont recommandés, MKV, TIFF, PNG, JPEG, GIF, MP3, Vorbis, MPEG-2 et iCal ;
- pour la signature, sont recommandés, ASIC, CADES, XAdES et PAdES.

Certaines recommandations présentées ici sont directement promulguées par des organisations internationales tel que l'ISO et le W3C.

⁹⁰ Référentiel Général d'Interopérabilité. Standardiser, s'aligner et se focaliser pour échanger efficacement, [en ligne], mars 2015, [consulté le 27 août 2017]. Disponible sur le Web : < http://references.modernisation.gouv.fr/sites/default/files/Referentiel_General_Interoperabilite_V1.9.7-8.pdf >.

Enfin, le niveau d'interopérabilité sémantique qui présente cette fois-ci plusieurs définitions importantes tel que l'agent, l'autorité administrative, la donnée, les données de référence, le fournisseur d'identité, le fournisseur d'authentification, le fournisseur de données, le fournisseur de données ou agrégateur de données (hub de données), fournisseur de service, personne, service, unité légale (entité légale), utilisateur et usager. La présentation de chaque terme montre bien que ce niveau d'interopérabilité s'applique essentiellement sur le lexique utilisé. Pour les travaux de modélisation, seulement un langage (UML) et une notation (CPMN) sont recommandés. Les formats pivots (formats sémantiques et syntaxiques) interviennent essentiellement dans les échanges.

En archéologie, les différents niveaux d'interopérabilité (figure 8) dépendent de la nature des entités traitées et exploitées : modèles, vocabulaires, notices d'autorité et notices.

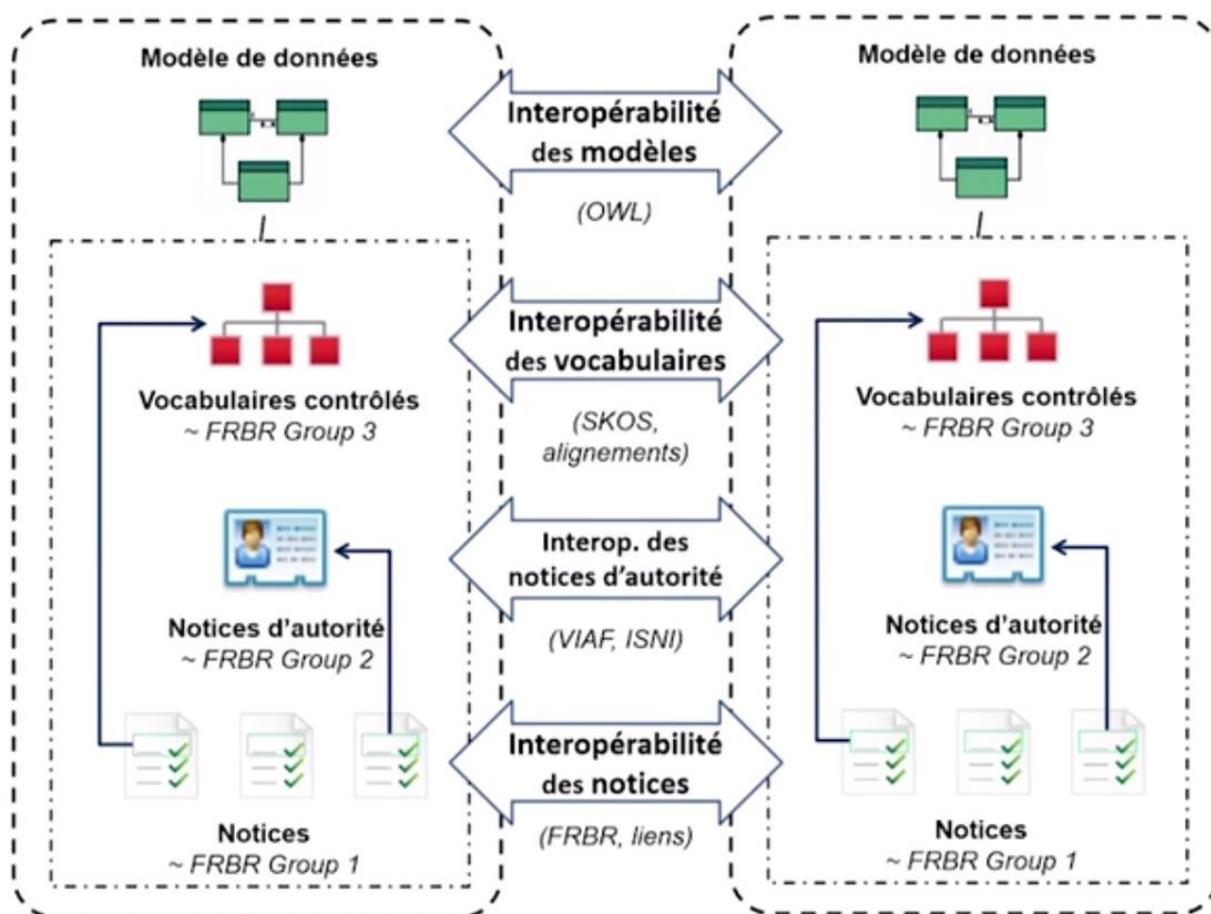


Figure 8 Schéma simplifié des différents enjeux de l'interopérabilité d'un système documentaire archéologique.

Source : FRANCAERT Thomas. *Ontologies et thesaurus, l'interopérabilité grâce au web des données*, in *Interopérabilités, Journées d'études du Consortium MASA*.

De plus, l'interopérabilité représente un véritable enjeu mondial et plusieurs organisations se sont constituées : la coalition *Warrior Interoperability*, le W3C (il développe un Web sémantique), l'ISO (elle fixe des définitions et des normes) ou encore *l'Interoperable Delivery of Pan-*

European eGovernment Services to Public Administrations. Elles fixent un cadre et des règles pour la garantie de l'interopérabilité et s'accordent toutes à préconiser des protocoles, standards et logiciels libres de droits, toujours dans l'objectif de promouvoir l'ouverture de l'information.

SYSTEMATISER L'OUVERTURE DES DONNEES ET LE LIBRE ACCES

Bien que l'ouverture et le libre accès aux données soient considérés comme l'un des projets de l'archéologie pour les années à venir, plusieurs réflexions émergent autour de la légitimité de la « libération » de l'information archéologique et les bouleversements que *l'Open Access* opère.

Lors du colloque « Actualité de l'Open Access en archéologie française » organisé par l'INRAP le 23 mars 2017, Stéphane Deschamps, conservateur général du patrimoine du SRA Île-de-France, débutait son propos sur les possibilités offertes par *l'Open Access* dans la diffusion des savoirs et interrogeait les pratiques de partage, d'archivage et de valorisation⁹¹. Depuis quelques années, les efforts sont essentiellement concentrés sur modalités de transmission amenant les archéologues à se poser la question de la légitimité de leur partage et par extension du recours aux licences libres de droits. Au cours d'une table ronde menée par Xavier Rodier, certains professionnels du patrimoine désapprouvaient par exemple le partage de certaines informations propres aux sites archéologiques comme leur géolocalisation. Cette pensée nuit considérablement à l'avancée de l'archéologie en matière de diffusion et de valorisation des données, freinant les projets de certaines institutions de recherche à partager. *L'Open Access* impulsé, en archéologie, pour ouvrir la voie pour la mise à disposition de ressources numériques tel que les rapports d'opération ou pose problème. Aux yeux de certains professionnels, *l'Open Access* presse l'archéologie à s'ouvrir plus rapidement qu'elle ne le peut. Pourtant, l'heure est bien au libre accès et pour le besoin de tout à chacun, il est important d'ouvrir l'accès aux données archéologiques.

Afin de contrer les courants de pensées associant *l'Open Access* à une forme de « pillage » du savoir, Stéphane Deschamps présente la diffusion comme une forme de protection. En effet selon lui, en dissimulant l'existence d'un quelconque patrimoine archéologique à une commune par exemple, il n'est pas possible de demander aux élus de le protéger et par extension de le valoriser. Ce schéma de réflexion est le même dans le monde numérique. En refusant d'ouvrir les données et en les dissimulant sous couvert de protection, c'est l'effet qui pourrait se produire : les

⁹¹ RODIER Xavier, DESCHAMPS Stéphane, FUCHS Matthieu, ROURE Réjane et TUFFERY Christophe. Ouverture et partage des données de l'archéologie, [en ligne], 72 : 02 minutes, 23 mars 2017, mis à jour le 30 novembre 2017, [consulté le 27 août 2018]. Disponible sur le Web : < <https://www.inrap.fr/table-ronde-ouverture-et-partage-des-donnees-de-l-archeologie-13183> >.

données pourraient être diffusées en masse sans réel cadre ou structure. Il faut donc amener les chercheurs, les instituts ou centres de recherche à diffuser leur savoir convenablement et pour répondre à cela, Anne Speller présente plusieurs propositions pour l'ouverture des données archéologiques, elle conseille notamment, en plus du respect des formats et standards ouverts, de :

- **dissocier les champs interprétatifs des champs descriptifs** ;
- **cadrer des rubriques** permettant de définir :
 - la forme, les dimensions observées et la localisation ;
 - la nature et les composants ;
 - les modalités d'observations des données archéologiques ;
- **faire le lien** entre les données textuelles et les données iconographiques⁹².

Systématiser l'ouverture et le libre accès aux données c'est donc avant tout avoir recours à modes de diffusion, de traitement, de gestion, de dépôt libres et ouverts et à des licences tel que *Creative commons*. C'est aussi avoir le réflexe de déposer ces données ou ces résultats de recherche sur des plateformes spécifiques destinées à partager l'information archéologique. À l'échelle internationale et européenne, les initiatives se multiplient pour permettre l'ouverture des données. Certains groupes réfléchissent par exemple à des outils pour *l'Open licensing*, à des plateformes qui proposent les inventaires des dépôts en *Open Access* : on peut citer DOAR (*Directory of Open Access Repositories*). Mais actuellement, cela ne suffit pas pour faire de l'archéologie une science « ouverte » et libre dans ses modes de diffusion et de valorisation.

Aujourd'hui, l'importance des chantiers de fouilles archéologiques et des données produites et récoltées imposent la nécessité de réfléchir et, peut-être, de repenser à la manière de publier, de diffuser et de valoriser. Pour chaque projet de diffusion, il faut prendre en compte l'accumulation de données, au fil des ans et des compétences humaines, institutionnelles et financières en présence. Le SRA (service régional d'archéologie préventive) Alsace est l'exemple, par excellence, de la collectivité territoriale souhaitant restituer à l'ensemble des publics les données archéologiques de la région et les porter à la connaissance de la communauté scientifique. Dans un contexte de partage et de valorisation du patrimoine ré-

⁹² SPELLER Anne et COLIN Camille, *Quelles données pour quels projets : perspectives de l'Open Access pour l'archéologie préventive* [en ligne]. 23 mars 2017, mis à jour le 30 novembre 2017, [consulté le 21 juin 2018], 21:31 minutes. Disponible sur le Web : < <https://www.inrap.fr/quelles-donnees-pour-quels-projets-perspectives-de-l-open-access-pour-l-13180> >

gionale, elle a souhaité rendre accessible non seulement l'ensemble des données archéologiques et rapports d'activité, mais aussi les plateformes sur lesquelles ils étaient diffusés.

Fort est de constater, au vu de ces expériences, que plusieurs actions peuvent être mises en place pour permettre la systématisation de l'ouverture des données, tel que :

- **instaurer une politique** pour la diffusion et valorisation des données ;
- **sensibiliser** les chercheurs et les archéologues ;
- **mettre en place une grille de critères** afin d'automatiser l'utilisation des standards et des formats non-proprétaires ;
- **(ré)utiliser** des données déjà ouvertes dans d'autres projets de recherche ;
- **aligner les jeux de données** ;
- **déposer les données** sur des plateformes spécifiques.

L'ouverture des accès aux données de l'archéologie est aussi un élément important à prendre en compte. Il existe encore, malheureusement, trop de plateformes inaccessibles au grand public où seul le chercheur peut s'authentifier et par extension accéder aux données. Ce blocage des accès ne favorise pas la valorisation et la diffusion convenable.

L'organisation de la gestion des données, la garantie de l'interopérabilité appliquée pour chaque niveau et l'ouverture des données restent des problématiques récurrentes en archéologie. Face à cela, on aura de cesse de recommander un accompagnement autant technique qu'humain auprès des professionnels de l'archéologie afin d'amener, indubitablement, à un renouvellement des méthodes de travail.

CONCLUSION

L'archéologie est une science des pratiques et non d'une pratique. Science source par excellence, elle est le reflet des préoccupations les plus contemporaines de notre époque dans la recherche scientifique et historique⁹³. Son ancrage entre passé humain et technologie de pointe fait d'elle une discipline de tous les enjeux, autant techniques que syntactiques ou méthodologiques. Les données archéologiques font partie du patrimoine de l'humanité qu'il est impératif de diffuser au plus nombre et cela passe indéniablement par la valorisation.

Aujourd'hui, le numérique influe directement sur les méthodes de production, de gestion, de traitement, d'archivage et de partage des données archéologiques. Face à toutes les réflexions que cela inclue, l'interopérabilité y tient une place non-négligeable. Pour faciliter cette interopérabilité à plusieurs niveaux, de nombreux facteurs sont requis : choix de termes communs approuvés et reconnus par tous, utilisation de formats et standards ouverts et renseignement de métadonnées. Néanmoins, bien que toutes ces initiatives – révolutionnaires dans les méthodes de partage de l'information archéologique – pleines de bon sens et en faveur de l'ouverture des données, trouvent, progressivement, leur place, elles cristallisent aussi d'autres difficultés. En effet, le fait de recourir aux mêmes termes et descripteurs ne rencontre pas l'adhésion de l'ensemble de la communauté scientifique et ne semble pas se présenter comme une garantie suffisante dans la capacité à réutiliser les données, ni à optimiser la compréhension. En archéologie, en tant que science historico-scientifique, un terme identique – et défini – ne mène pas nécessairement à une même interprétation, engendrant parfois confusion. Cette confusion peut alors se ressentir au plan technique et technologique et l'interopérabilité comme enjeu primordial dans un contexte de diffusion et de valorisation, peut être menacée. Cela se traduit aussi par une prolifération de thésaurus et d'ontologies, qui essayent, tant bien que mal, de rendre possible une interopérabilité sémantique ; alors que les différentes plateformes ou applications existantes à ce jour laissent à penser que l'archéologie tendrait vers une harmonisation voire une uniformisation. S'étant inspirée et s'inspirant encore, à tort peut-être, des autres sciences humaines, l'archéologie tente de faire évoluer ses pratiques et méthodologies de travail.

De manière générale, l'axe organisationnel consacré à la normalisation et à la structuration des métadonnées en archéologie, est l'un des éléments les plus importants à

⁹³ Archéologie et Antiquité, Mésopotamie, Egypte, Grèce, Rome, les Gaules : méthodes et outils pour comprendre le monde antique, [en ligne], décembre 2016, [consulté le 6 août 2018]. Disponible sur le Web : < http://www.bnf.fr/fr/collections_et_services/sc_humaines_philo_religions/a.archeologie_antiquite.html >.

développer dans les années à venir. Sans cet investissement de la part des professionnels et un consensus commun, les engagements de différentes institutions pour la garantie de l'interopérabilité seront nul. Au sein de la communauté scientifique, l'interopérabilité des données permet essentiellement de lier les sources primaires et publiées via des plateformes (Nakala, Isidore, etc..) et leur développement laisse espérer une valorisation et une diffusion harmonisée des données archéologiques rendue possible par l'interopérabilité, dont l'objectif est de faciliter l'accès et la restitution des données au public et aux archéologues. De plus, l'enjeu de l'utilisation des normes ne porte pas uniquement sur l'interopérabilité des données mais aussi sur leur pérennité⁹⁴. Le 4 août 2018 le gouvernement français lançait, pour la première fois, un plan national pour la science ouverte. Ce plan rend obligatoire l'accès ouvert pour les publications et pour les données issues de recherches financées sur projets, tout en soutenant les initiatives de structuration de ces résultats⁹⁵. Ainsi, les projets archéologiques financés sur fonds publics devront se soumettre à ce plan national pour la science ouverte, permettant de les faire sortir de leur environnement confiné. L'ouverture à la diffusion se présenterait alors comme un des principaux facteurs de la valorisation.

Pour les années à venir, on peut espérer que l'archéologie s'appuie sur les projets mis en place par d'autres structures, tel que la création d'un modèle de diffusion des données. Ce type d'initiative prend généralement appui lors des regroupements de professionnels de l'archéologie. Le 24 septembre 2018, une première réunion du groupe de travail sur l'utilisation du CIDOC-CRM aura lieu à Lyon, afin de se concerter sur la modélisation des jeux de données archéologiques, ouvrant la voie à d'autres réflexions similaires.

⁹⁴ SZABADOS Anne-Violaine. Du système documentaire du LIMC au portail Claros. Interopérabilité et optimisation de l'information archéologique grâce à l'usage de normes, in *Archeologia e Calcolatori Supplemento* 3, [en ligne], 2012, [consulté le 10 août 2018]. Disponible sur le Web : < <https://docplayer.fr/2313143-Du-systeme-documentaire-du-limc-au-portail-claros-interoperabilite-et-optimisation-de-l-information-archeologique-grace-a-l-usage-de-normes.html> >.

⁹⁵ Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. Le plan national pour la science ouverte : les résultats de la recherche scientifique ouverts à tous, sans entrave, sans délai et sans paiement, [en ligne], le 4 août 2018, mis à jour le 12 août 2018, [consulté le 29 août 2018]. Disponible sur le Web : < <http://m.enseignementsup-recherche.gouv.fr/cid132529/www.enseignementsup-recherche.gouv.fr/cid132529/le-plan-national-pour-la-science-ouverte-les-resultats-de-la-recherche-scientifique-ouverts-a-tous-sans-entrave-sans-delai-sans-paiement.html> >.

BIBLIOGRAPHIE

Les ouvrages, articles, conférences filmées, billets de blog, lois et normes ayant servi pour la rédaction de ce mémoire sont classés par thème :

- **sur l'archéologie :**

Archéologie et Antiquité, Mésopotamie, Egypte, Grèce, Rome, les Gaules : méthodes et outils pour comprendre le monde antique, [en ligne], décembre 2016, [consulté le 6 août 2018]. Disponible sur le Web : <http://www.bnf.fr/fr/collections_et_services/sc_humaines_philo_religions/a.archeologie_antiquite.html>.

ALAINGUILLAUME Micael. Projet Archéomed – 50 ans d'archéologie en ligne [en ligne], 29 septembre 2017, mis à jour le 21 décembre 2017, [consulté le 7 août 2018]. Disponible sur le Web : <<https://craham.hypotheses.org/1356>>. ISSN : 2552-3139.

DJINDJIAN, François. *Manuel d'archéologie. Méthodes, objets et concepts*. Paris : Armand Colin, « U », 2011, ISBN : 9782200266769.

FRANCHOMME Magalie, SALVADOR Pierre-Gil et KERGOMARD Claude. Conception et utilisation d'un SIG pour l'étude des données (paléo-) environnementales et archéologiques dans le bassin versant de la haute Deûle, in *Revue du Nord-Archéologie*, 85 (353), 2003, p. 35-45.

INRAP. Rapport d'opération [en ligne]. 12 janvier 2016, mis à jour le 2 décembre 2016, [consulté le 18 juin 2018]. Disponible sur le Web : <<https://www.inrap.fr/rapport-d-operation-9855>>.

ZADORA-RIO Elisabeth. Atlas Archéologique de Touraine. *Supplément à la Revue Archéologique du Centre de la France*, 53, 2014, FERACF, Tours

- **sur l'interopérabilité :**

FAVIER Laurence, MUSTAFA EL HADI Widad et VINCK Dominique. Présentation, Interopérabilité culturelle, in *Communication*, [en ligne], vol. 34/1 | 2016, mis en ligne le 17 août 2016, consulté le 27 août 2018. Disponible sur le Web : <<http://journals.openedition.org/communication/6614>>.

Référentiel Général d'Interopérabilité. Standardiser, s'aligner et se focaliser pour échanger efficacement, [en ligne], mars 2015, [consulté le 27 août 2017]. Disponible sur le Web : <http://references.modernisation.gouv.fr/sites/default/files/Referentiel_General_Interoperabilite_V1.9_7-8.pdf>.

- **sur l'Open Access :**

ALBARET Lucie. Un portail HAL à l'université Grenoble Alpes : un projet au service des chercheurs... et du libre accès ! [en ligne]. 23 mars 2017, mis à jour le 30 novembre 2017, [consulté le 20 juin 2018],

24:33 minutes. Disponible sur le Web : < <https://www.inrap.fr/un-portail-hal-l-universite-grenoble-alpes-un-projet-au-service-des-chercheurs-13181> >.

MENEZ Yves, LORHO Thierry et PINEL Marie-Dominique. Une expérience d'open access/open data : la DRAC Bretagne – service régional de l'archéologie met en ligne ses données archéologiques [en ligne]. 23 mars 2017, mis à jour le 1^{er} décembre 2017, [consulté le 22 juin 2018], 30:27 minutes. Disponible sur le Web : < <https://www.inrap.fr/une-experience-d-open-accessopen-data-la-drac-bretagne-met-en-ligne-ses-donnees-13179> >.

MOSCATI Paola. Archéologie et informatique. Le rôle de l'Open Access dans la diffusion des résultats de recherche scientifique [en ligne]. 23 mars 2017, mis à jour le 30 novembre 2017, [consulté en ligne le 20 juin 2018], 23:41 minutes. Disponible sur Web : < <https://www.inrap.fr/archeologie-et-informatique-le-role-de-l-open-access-dans-la-diffusion-des-13176> >.

RODIER Xavier, DESCHAMPS Stéphane, FUCHS Matthieu, ROURE Réjane et TUFFERY Christophe. Ouverture et partage des données de l'archéologie, [en ligne], 72 : 02 minutes, 23 mars 2017, mis à jour le 30 novembre 2017, [consulté le 27 août 2018]. Disponible sur le Web : < <https://www.inrap.fr/table-ronde-ouverture-et-partage-des-donnees-de-l-archeologie-13183> >.

SPELLER Anne et COLIN Camille. Quelles données pour quels projets : perspectives de l'Open Access pour l'archéologie préventive [en ligne]. 23 mars 2017, [consulté le 21 juin 2018], mis à jour le 30 novembre 2017, 21:31 minutes. Disponible sur le Web : < <https://www.inrap.fr/quelles-donnees-pour-quels-projets-perspectives-de-l-open-access-pour-l-13180> >.

- *sur les technologies ou l'informatique appliquée à l'archéologie :*

BELARBI Mehdi, RAYMOND Pascal, SAULIERE Nicolas et TOUQUET Régis. L'acquisition 3D par photogrammétrie en archéologie [en ligne], juillet 2012, mis en ligne le 1^{er} juillet 2012, [consulté le 24 juillet 2018]. Disponible sur le Web : < <http://journals.openedition.org/archeopages/415> >.

DOZON Alexia. Archéologie numérique [en ligne]. 17 juin 2016, [consulté le 26 juin 2018]. Disponible sur le Web : < <https://archnum.hypotheses.org/40> >.

- *sur le traitement et la gestion des données archéologiques :*

CARTIER Aurore, MOYSAN Magalie et REYMONET Nathalie. Réaliser un plan de gestion des données [en ligne], 9 janvier 2015, [consulté le 20 août 2018]. Disponible sur le Web : < http://cache.media.education.gouv.fr/file/Comment_Participer/36/0/realiser_un_dmp_406360.pdf >.

DOZON Alexia. « La chaîne opératoire » du traitement et de la gestion de la documentation archéologique par SIG : une modélisation des données archéologiques vers un modèle d'information archéologique (projet « CSNE ») [en ligne]. 22 juin 2016, Archéologie numérique, [consulté le 3 juillet 2018]. Disponible sur le Web : < <https://archnum.hypotheses.org/189> >.

REYMONET Nathalie, MOYSAN Magalie, CARTIER Aurore, DELEMONTÉZ Renaud. Réaliser un plan de gestion de données " FAIR " : modèle, [en ligne], 2018, [consulter le 21 août 2018]. Disponible sur le Web : < https://archivesic.ccsd.cnrs.fr/sic_01690547/document >.

- *sur l'archivage et les archives de l'archéologie :*

HURLET Frédéric et RODIER Xavier Rodier. Archivage, publication et mise à disposition de données archéologiques [en ligne]. 26 et 27 septembre 2016 [consulté le 27 juin 2018]. Disponible sur le Web : < <https://vimeo.com/189300888> >.

MÜLLER Bertrand. Archiver les données en Sciences humaines et sociales. Informations, positions de problèmes et propositions [en ligne]. 2012, [consulté le 27 juin 2018]. Disponible sur le Web : < <http://www.cnrs.fr/inshs/informations-DU/docs/du2013/archives-shs.pdf> >.

STAHL Marie et SCHIRR Lucie. « Les archives de l'archéologie : définition, législation, état des lieux », Archimède : archéologie et histoire ancienne, 2015, p.9-19.

QUANTIN Matthieu. Archivage de modèles 3D, [en ligne], 15 septembre 2016, [consulté le 5 août 2018]. Disponible sur le Web : < <https://shs3d.hypotheses.org/2354> >.

- *sur la diffusion et la valorisation des données :*

COMETS. Avis Réflexion éthique sur la diffusion des résultats de recherche, 1^{er} mars 2007.

OCDE. Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics [en ligne]. 2007, [consulté le 14 juin 2018]. Disponible sur le Web : < www.oecd.org/fr/sti/sci-tech/principesetlignesdirectricesdelocdepourlaccessauxdonneesdelarecherchefinancesurfondspublics.htm >.

- *sur les ontologies et les thésaurus :*

BnF. Modèle CIDOC-CRM [en ligne], 18 août 2017, [consulté le 9 août 2010]. Disponible sur le Web : < http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.modele_cidoc_crm.html >.

BnF. Elaboration d'ontologies [en ligne], 5 février 2015, [consulté le 9 août 2010]. Disponible sur le Web : < http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.elaboration_ontologies.html >.

FRANCART Thomas. Ontologies et thesaurus, l'interopérabilité grâce au web des données, [en ligne], 21 novembre 2017, [consulté le 8 août 2018], 32 :3. Disponible sur le Web : < https://www.canal-u.tv/video/msh_val_de_loire/ontologies_et_thesaurus_l_interoperabilite_grace_au_web_de_donnees_thomas_francart_sparna_expert_en_technologies_semantiques.39735 >.

MARLET Olivier. Une ontologie pour rendre interopérables des bases de données en archéologie ? [en ligne], 19 août 2014, [consulté le 9 août 2018]. Disponible sur le Web : < <https://masa.hypotheses.org/60> >.

NOUVEL Blandine. Des outils d'enrichissement documentaire multilingues pour l'archéologie, [en ligne], 14 décembre 2015, [consulté le 7 août 2018]. Disponible sur le Web : < <https://masa.hypotheses.org/218> >.

SZABADOS Anne-Violaine et LETRICOT Rosemonde. L'ontologie CIDOC CRM appliquée aux objets du patrimoine antique, in 3e Journées d'Informatique et Archéologie de Paris - JIAP 2012, [en ligne], juin 2012, [consulté le 9 août 2018]. Disponible sur le Web : < https://halshs.archives-ouvertes.fr/file/index/docid/752996/filename/SZABADOS_LETRICOT_ontologie_CIDOC_CRM_appliquee_aux_objets_du_patrimoine_antique_2012.pdf >.

What is the CIDOC-CRM? [en ligne], [consulté le 10 août 2018]. Disponible sur le Web : < <http://www.cidoc-crm.org/> >.

- **sur les normes et leur usage en archéologie :**

SZABADOS Anne-Violaine. L'usage de normes pour l'interopérabilité des données sur l'Antiquité : l'exemple des outils documentaires du LIMC, in Journée d'études pour les professionnels de la BAPF en SHS, Paris : 25 novembre 2010.

SZABADOS Anne-Violaine. Du système documentaire du LIMC au portail Claros. Interopérabilité et optimisation de l'information archéologique grâce à l'usage de normes, in *Archeologia e Calcolatori Supplemento 3*, [en ligne], 2012, [consulté le 10 août 2018]. Disponible sur le Web : < <https://docplayer.fr/2313143-Du-systeme-documentaire-du-limc-au-portail-claros-interoperabilite-et-optimisation-de-l-information-archeologique-grace-a-l-usage-de-normes.html> >.

SZABADOS Anne-Violaine. L'usage de normes et de thésaurus pour les données archéologiques et les métadonnées, de la production sur le terrain à la publication des rapports d'opération : l'expérience de l'Inrap, [en ligne], 23 mai 2018, [consulté le 10 août 2018]. Disponible sur le Web : < <https://archeo-num.hypotheses.org/738> >.

- **sur la redocumentarisation :**

BROUDOUX Evelyne et SCOPSI Claire. L'enjeu des métadonnées dans un contexte de "redocumentarisation", in *Etudes de communication*, [en ligne], 2011, [consulté le 30 juillet 2018]. Disponible sur le Web : < https://www.researchgate.net/publication/281156482_L%27enjeu_des_metadonnees_dans_un_contexte_de_redocumentarisation >.

SALAÛN, Jean-Michel. La redocumentarisation, un défi pour les sciences de l'information, in *Études de communication* [en ligne], 2007, mis en ligne le 1^{er} octobre 2009, [consulté le 25 juillet 2018]. Disponible sur le Web : < <http://journals.openedition.org/edc/428> >.

- **sur les banques et bases de données en archéologie :**

BAUDOUIN Bruno et SATRE Stéphanie. Base de données de recherche et plateforme d'exposition de corpus épigraphiques : EpiCherchel et CoReA, in *Interopérabilités – Journées MASA*, [en ligne], 21 novembre 2017, [consulté le 29 août 2018]. Disponible sur le Web : < https://www.canal-u.tv/video/msh_val_de_loire/base_de_donnees_de_recherche_et_plateforme_d_exposition_de_corpus_epigraphiques_epicherchel_et_corea_stephanie_satre_et_bruno_baudoin_centre_camille_jullian_umr_7299_et_masa.39751 >.

BUCHSENSCHUTZ, Olivier. Anne-Marie Guimier-Sorbets, Les bases de données en archéologie, Conception et mise en œuvre (compte-rendu), in *Histoire & Mesure*, [en ligne], 1994, [consulté le 13 août 2018], n°3-4. Disponible sur le Web : < http://www.persee.fr/doc/hism_0982-1783_1994_num_9_3_1713 >.

GINOUVES René. Des banques de données pour l'archéologie ? in *Traitement de l'information en archéologie*, Brises : 1990, CNRS, p.97-107, [en ligne], [consulté le 12 août 2018]. Disponible sur le Web : < <http://www.mae.u-paris10.fr/ginouves/texteimpr/AR16.html> >.

THIVET Matthieu. Bases de données archéologiques. Problématiques, contraintes, enjeux, [en ligne], [consulté le 12 août 2018]. Disponible sur le Web : < <https://docplayer.fr/6369056-Base-de-donnees-archeologiques.html> >.

Sitographie

Site Archéologie numérique : retours d'expériences : <https://archnum.hypotheses.org/>

Site officiel de l'INRAP : <https://www.inrap.fr/>

Site de la MAE (Maison Archéologie Ethnologie – René-Ginouvès) Paris-Nanterre : <http://www.mae.parisnanterre.fr/>

Site du consortium MASA : <http://masa.hypotheses.org/>

Site officiel du ministère de la Culture – Archéologie – Bases de données : <http://www.culturecommunication.gouv.fr/Thematiques/Archeologie/Ressources/Bases-de-donnees>

Site officiel du projet ARIADNE : <http://ariadne-portal.dcu.gr/>

Normes

ISO 15836-1:2017 *Information and documentation – The Dublin Core metadata element set – Part 1: Core elements*

ISO 19115:2014 - *Geographic information -- Metadata -- Part 1: Fundamentals*

ISO 19139:2007 - *Geographic information -- Metadata -- XML schema implementation*

ISO 21127:2014 - *Information and documentation — A reference ontology for the interchange of cultural heritage information*

ISO 25964-1:2011 - *Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval*

ANNEXES

Table des annexes

ANNEXE N°1 - AVIS DU COMETS – REFLEXION ETHIQUE SUR LA DIFFUSION DES RESULTATS DE LA RECHERCHE P.73

ANNEXE N°2 – ENTRETIEN AVEC EMMANUELLE MORLOCK p.89

ANNEXE N°3 – RECOMMANDATIONS DU CONSORTIUM MASA POUR LA PUBLICATION ET LE PARTAGE DES JEUX DE DONNEES p.93

**ANNEXE N°1 – AVIS DU COMETS - REFLEXION ETHIQUE SUR LA
DIFFUSION DES RESULTATS DE LA RECHERCHE**



Comité d'éthique du CNRS (COMETS)

**AVIS
REFLEXION ETHIQUE
SUR LA DIFFUSION DES RESULTATS DE LA RECHERCHE**
(rendu le 1er mars 2007)

Site Internet : <http://www.cnrs.fr/fr/presentation/ethique/comets/index.htm>

RESUME

Faire connaître les résultats de la recherche est une des missions du chercheur et des institutions qui le financent. Or, le contexte actuel de la diffusion des résultats de la recherche fait l'objet d'évolutions rapides liées à la place de plus en plus grande prise par la numérisation des données et par Internet dans le fonctionnement de nos sociétés. Le présent avis, centré sur la publication scientifique, vise à faire émerger les enjeux éthiques de cette évolution et à formuler quelques recommandations à leur égard.

I. Mise en perspective. Les règles de la diffusion organisée du savoir, établies depuis le XVII^e siècle, n'ont pas fondamentalement changé, mais la professionnalisation de la recherche scientifique en a modifié profondément le contexte, par l'évolution des revues, l'impératif de publier et les pressions qu'il engendre.

II. La nature des vecteurs de diffusion. Cet avis est centré sur les publications, et ne traite ni des brevets, ni des bases de données, deux sujets essentiels qui pourront être abordés ultérieurement. Aux circuits classiques des revues et des archives institutionnelles s'ajoutent ceux, en rapide expansion des revues en libre accès et des archives ouvertes. Chacun de ces vecteurs doit satisfaire à l'excellence du travail diffusé, à la rapidité de la diffusion et à l'amplitude du champ de celle-ci : ces exigences ne sont pas toujours satisfaites.

III. Réflexions sur les modalités. Les revues en libre accès et les archives ouvertes sont des étapes vers un modèle optimisé, encore à créer, où les impératifs de libre diffusion, les préoccupations éthiques, les critères de qualité et ceux d'équilibre économique seraient mieux conjugués, peut-être par auto-régulation entre auteurs et lecteurs. L'existence de filtres scientifiques, nécessaires, ne garantit pas l'absence de dérives. Les modalités de signature des publications (nombre, ordre), celles de leur référence aux organismes dont dépendent les chercheurs sont insatisfaisantes. Les modalités électroniques de large diffusion facilitent la mise à disposition rapide d'un immense volant de données d'origines diverses par des moteurs de recherche. Cette nouvelle approche, séduisante et efficace, peut s'accompagner d'inconvénients en ce qui concerne les *méta-données* (référencement, indexation par le contenu). Cet aspect crucial implique une coopération entre des documentalistes de nouvelle génération et les chercheurs. La construction d'un espace européen de la recherche offre la possibilité d'évolutions significatives, en particulier à court terme pour favoriser l'accessibilité aux résultats de recherches financées par le 7^{ème} programme-cadre.

IV. Les enjeux éthiques. Faciliter pour le plus grand nombre l'accès aux connaissances, acquises grâce à l'effort financier public, en les considérant comme un bien universel, au sein des pays les plus développés mais aussi au-delà de leurs frontières, relève d'un choix éthique majeur, sans ignorer leurs conséquences pour l'homme et les sociétés. Une politique commune, qui pourrait être d'inspiration européenne, des établissements de recherche et des éditeurs apparaît indispensable, reposant sur la répartition des responsabilités et la définition d'aides institutionnelles. Aux positions dominantes de certains journaux s'ajoutent aujourd'hui de nouvelles formes d'édition scientifique. L'existence d'un nombre assez restreint de moteurs de recherche peut conduire à un monopole de la diffusion des connaissances sur la Toile. L'utilisation d'une langue unique peut aboutir alors à un appauvrissement dans l'expression des résultats de la recherche. Le chercheur doit s'interroger sur l'opportunité et la pertinence de sa décision de communiquer ses résultats. Les évaluateurs et comités d'évaluation doivent être attentifs à ne pas succomber aux facilités associées à une exploitation exclusive de paramètres liés à la nature des revues ou aux fréquences de citation.

RECOMMANDATIONS

- Identifier et à soutenir des systèmes de publications accessibles et de qualité, afin d'assurer une diffusion du savoir la plus large possible aux bénéficiaires de la science et de la société.
- Sensibiliser les acteurs de la recherche de façon claire, honnête et complète aux différentes modalités de communication de leurs résultats ; faciliter la compréhension et les avantages des systèmes en libre accès. Dans le cas des archives ouvertes, les institutions doivent clairement souligner, par-delà l'intérêt général, leurs intérêts spécifiques dans le suivi et l'évaluation de la production scientifique. Elles doivent aussi prodiguer aux chercheurs les accompagnements nécessaires, concernant les aspects techniques et réglementaires des nouvelles procédures. Ces attitudes sont à même d'effacer la tiédeur, voire l'inertie, des milieux de la recherche vis-à-vis de ces évolutions et les conduire à se les approprier. Si l'incitation au dépôt des résultats dans des archives ouvertes peut convenir dans un premier temps, c'est l'obligation qui donnera au système toute son efficacité. Cette obligation sera d'autant mieux acceptée que l'explicitation aura été satisfaisante.
- Assurer au niveau des organismes et institutions un soutien éditorial significatif pour que la qualité et la large dissémination des données archivées (archives ouvertes) soient assurées.
- Inciter les établissements à développer une concertation et des partenariats avec les acteurs de l'édition, publique et privée, sous toutes ses formes pour aboutir à des solutions équilibrées prenant en compte les intérêts de chacun.
- Réfléchir à corriger les abus de situation dominante ou de monopole. Il est difficile d'inverser l'hégémonie exercée par des revues prestigieuses, celle de nouveaux systèmes efficaces de diffusion électronique ou de moteurs de recherche. Il demeure possible, par des initiatives, des soutiens financiers, des codes de conduite dans l'évaluation, de prendre des mesures atténuant des situations éthiquement discutables.
- Envisager des formules pour contrebalancer la prédominance de l'anglais dans les systèmes de diffusion de connaissance en favorisant en particulier le bilinguisme et éventuellement dans certaines disciplines la traduction assistée par ordinateur. Ces démarches peuvent paraître vaines mais elles s'inscrivent dans le souci, qui doit demeurer permanent, de la protection et de la promotion de la diversité des expressions culturelles (Convention Unesco 2005).
- Sensibiliser celui qui produit les connaissances à sa responsabilité éthique, que ce soit dans les publications institutionnelles du CNRS ou dans les recommandations faites aux chercheurs entrants ; identifier une structure de contact, à laquelle le chercheur pourrait se référer si nécessaire pour la diffusion des résultats de sa recherche. L'acte de publication doit être raisonné, la contribution doit être appropriée, équilibrée, justifiée en dépassant les seules motivations de progression de carrière ou de présence dans la compétition internationale.
- S'agissant de l'évaluation, veiller à fonder le jugement sur des indicateurs diversifiés, prenant en compte l'originalité, l'innovativité des résultats, sans se limiter à des critères uniquement quantitatifs tels que facteurs d'impact des revues ou index de citations, dont la grande facilité d'utilisation peut conduire à ne fonder l'évaluation que sur des appréciations de type bibliométrique.

SOMMAIRE

Réflexion éthique sur la diffusion des résultats de la recherche

Préambule

Introduction

I. Mise en perspective historique

II. La nature des vecteurs de diffusion des résultats de la recherche

Cadrage général

Les principaux vecteurs de diffusion aujourd'hui

III. Quelques réflexions liées aux différentes modalités de diffusion des connaissances

- a) la dynamique du libre accès
- b) l'absence ou la présence d'un filtre scientifique
- c) sur le référencement des publications
- d) l'accélération de la transmission des connaissances
- e) une réflexion active à l'échelle européenne

IV. Enjeux éthiques dans le domaine de la diffusion des connaissances

- a) Le problème général d'un large partage des connaissances versus une attitude élitiste de restriction de la diffusion des savoirs
- b) les antagonismes et conflits d'intérêts dans l'édition scientifique
- c) les notions de monopole ou d'impérialisme dans la diffusion des connaissances
- d) les dimensions éthiques associées à la décision de publication scientifique et à l'exploitation des publications dans l'évaluation

V. En forme de conclusion provisoire

Recommandations

Réflexion éthique sur la diffusion des résultats de la recherche

Préambule :

Le contexte général de la diffusion des résultats de la recherche fait l'objet d'évolutions rapides, liées à la place de plus en plus grande prise par la numérisation des données et par Internet dans le fonctionnement de nos sociétés.

Cet avis est centré sur une des formes ouvertes du savoir, la publication scientifique qui est ici l'objet d'une réflexion prenant en compte à la fois des positions de principe et de profonds changements dans les techniques de diffusion. D'autres aspects comme le traitement et le stockage de données primaires (banques de données, résultats intermédiaires) ou le problème de la propriété intellectuelle et de la brevetabilité des résultats ne sont pas abordés. Ils mériteraient d'être considérés en détail plus spécifiquement.

Quatre problèmes majeurs semblent associés à la problématique générale de la diffusion des résultats de la recherche publique dans la perspective de l'émergence du tout numérique :

- les résultats ont vocation à être diffusés très largement, rapidement et gratuitement, en particulier via les moyens électroniques ;
- ces objectifs ne doivent pas être atteints au détriment de la qualité des productions publiées ;
- il faut trouver des modèles économiques consensuels qui conjuguent les intérêts des différents protagonistes : chercheurs, laboratoires, organismes et maisons d'édition ;
- en regard de leurs avantages évidents les formules du tout numérique présentent au moins deux types de risque :
 - asphyxie du système par un foisonnement d'informations mal gérées et mal exploitées ; l'intérêt d'un système cohérent de méta données est à ce niveau évident ;
 - appauvrissement de la qualité de l'écrit et de la publication dans un contexte d'accélération de la parution.

L'entrée dans une économie de la connaissance ouverte implique plus que jamais que soient pris en compte des principes éthiques qui, en dépit de la révolution des moyens de production et de circulation des documents, demeurent constants dans leurs grandes lignes : exigence de qualité, responsabilité des producteurs de documents, transparence des sources et accessibilité des objets publiés.

Le fait de ne pas avoir approché le champ de la façon la plus large possible n'altère cependant pas la façon dont les considérations éthiques introduites dans cet avis s'appliquent au processus en cours, et donc n'invalide pas les recommandations faites. L'évolution rapide de la situation a incité à limiter notre approche aux publications scientifiques avec la volonté de prendre position suffisamment vite pour que nos analyses et recommandations puissent avoir un impact avant que le processus ait rendu complètement obsolète un certain nombre de pratiques. Ces pratiques jouent un rôle important dans la régulation de la vie scientifique et servent de repères majeurs, bien qu'à des degrés divers d'une discipline à l'autre. Il faut en effet souligner que le rôle et l'exploitation des publications sont de facto très différents dans l'ensemble du champ scientifique, phénomène dont il est important de prendre bien conscience quand on se préoccupe de la diffusion des résultats scientifiques.

Introduction :

Faire connaître les résultats de la recherche est une des missions du chercheur et de l'institution qui le finance. Communiquer et partager les connaissances qu'il a contribué à développer est donc une dimension significative de son activité. Peut-on imaginer la science sans diffusion de résultats ?

Cette diffusion est aussi pour le chercheur un moyen d'accéder aux résultats des autres membres de la communauté scientifique dans un processus d'échanges qui contribue à faire avancer la connaissance. Ce qu'Isaac Newton exprimait de façon éloquente quand il écrivait en 1676, s'inspirant de Bernard de Chartres : « *If I have seen further than others, it is by standing on the shoulders of giants* ». Cette diffusion par des vecteurs reconnus représente de plus, dans les milieux de la recherche, un élément crucial de la reconnaissance scientifique et du processus d'évaluation.

Le présent avis a pour objectif de faire émerger quelques-uns des enjeux éthiques de la diffusion des résultats de la recherche, principalement par les publications, et les différentes modalités pratiques qui y sont associées. Il s'agit d'un secteur en évolution rapide dont un des enjeux éthiques, largement partagé, a été concrétisé par la déclaration de Berlin⁹⁶ en 2003 : celui d'une très large distribution du savoir, rendue possible à grande échelle par les moyens électroniques.

Après une mise en perspective historique, nous examinerons successivement la nature des vecteurs de diffusion des résultats de la recherche en prenant en compte les tendances actuelles, les problèmes soulevés par les évolutions récentes et, finalement, les enjeux éthiques suivis de recommandations.

I – Mise en perspective historique

La diffusion des connaissances, rendue possible grâce à l'imprimerie et aux correspondances savantes, est devenue une exigence au XVII^e siècle. C'est ce qui conduit à discréditer le savoir hermétique des alchimistes et qui pousse Galilée et Descartes à troquer le latin pour les langues vernaculaires afin d'assurer une diffusion plus large de certains de leurs écrits. Mais c'est avec la création des académies au fil du XVII^e siècle, que la diffusion organisée du savoir devient une norme. Les expériences doivent être conduites devant un public de témoins « actuels » qui attesteront des faits observés ; et grâce à des comptes rendus d'expérience dûment rédigés, les lecteurs deviennent témoins « virtuels », susceptibles de reproduire l'expérience pour la vérifier. Le régime académique dans lequel se développe la science moderne repose sur trois principes : libre exercice de l'esprit critique entre membres des académies, évaluation par le tribunal des pairs, et attribution de la découverte à un auteur. La communauté académique fonctionne donc comme une « république des sciences » autonome, indépendante de l'Église, de l'opinion publique, et du monarque, bien qu'elle soit sous son patronage.

La professionnalisation de la recherche scientifique au cours du XIX^e siècle a certes affaibli les académies et instauré un nouveau type de sociabilité entre spécialistes d'une même discipline, mais elle n'a pas fondamentalement changé les règles du jeu académique. Les revues attachées aux sociétés savantes fonctionnent avec un comité de rédaction, un comité de lecture, et des experts.

Après 1945, les revues évoluent en raison de l'augmentation rapide de la taille des communautés de recherche ; un seul éditeur en chef ne pouvant assurer toute la responsabilité d'une revue, c'est à cette époque qu'on assiste à une généralisation des comités de lecture. Cette augmentation considérable du nombre d'articles scientifiques publiés alourdit la gestion, et fait que les revues deviennent de plus en plus chères, et que nombre d'entre elles passent sous le contrôle de grands éditeurs commerciaux. La revue devient parfois un investissement à visée lucrative. A cause de l'impact de la guerre en Europe,

⁹⁶ http://www.zim.mpg.de/openaccess-berlin/BerlinDeclaration_fr.pdf

le centre de gravité scientifique se déplace aux États-Unis et on assiste dans certaines disciplines à une baisse d'influence progressive des revues européennes au profit des revues américaines.

En même temps, pour répondre au développement de la recherche et au besoin de publier engendré par les impératifs de carrière des chercheurs, les revues se multiplient, tandis que l'inventivité de certains et la pression du système conduisent au développement des index-citations.

II – La nature des vecteurs de diffusion des résultats de la recherche

Cadrage général :

Avec des cibles et des moyens de diffusion variés, l'acte de diffusion, la décision de publier, de communiquer correspondent le plus souvent à une volonté de faire connaître, de partager des connaissances et des idées, et de valoriser son travail.

Le plus classiquement un chercheur diffuse ses résultats à l'attention de la communauté scientifique dans le cadre d'obligations devenues statutaires afin de prendre date et de contribuer à l'enrichissement du patrimoine des savoirs dans une optique de large partage des connaissances. On peut aussi vouloir transmettre ses connaissances à des publics non spécialistes et plus largement à l'ensemble de la société.

Enfin, dans un souci de valorisation industrielle, c'est une protection des connaissances qui sera envisagée par une prise initiale de brevet, suivie éventuellement d'une publication pour favoriser la communication.

La prise en compte de chacun de ces objectifs a une incidence sur le mode de diffusion et le document diffusé, l'approche rédactionnelle étant à l'évidence différente dans une revue grand public et dans une revue scientifique destinée aux collègues de la même communauté.

Par ailleurs la nature du matériel publié, et secondairement la durée de vie et de pertinence d'un résultat, dépendent des spécificités disciplinaires. A titre d'exemple, en mathématiques a contrario de ce qui se passe dans d'autres disciplines, la norme est de publier le plus souvent des résultats totalement aboutis dont la pertinence peut demeurer intacte après plusieurs siècles. Cependant dans le cas des articles originaux et quelles que soient les spécificités disciplinaires, on peut identifier des caractéristiques largement partagées au niveau de l'objet publié. Il s'agit des notions de nouveauté et d'originalité, des critères de traçabilité et de reproductibilité conférant toute leur validité à la publication, des impératifs de restitution exacte des observations et de leur interprétation non tendancieuse, de précision sur la paternité des faits relatés et de citation correcte des contributeurs antérieurs. Ces caractéristiques correspondent à des règles implicites, à de bonnes pratiques qui sont très généralement observées et qui figurent en partie de façon plus ou moins formalisée dans les grilles d'évaluation des journaux à comité de lecture.

Dans la diffusion des connaissances, on inclut aussi naturellement, la dissémination de données intermédiaires : composantes de bases de données (y compris les bases de données génétiques), recueil de données primaires (matériel de fouilles, enregistrements sonores ou tout autre corpus...). Une attention particulière doit être portée à l'exploitation des outils de diffusion électronique assortie des meilleures techniques d'indexation pour distribuer le plus largement possible ces données, savoirs, expériences parfois ponctuelles... qui contribuent de façon importante à notre patrimoine scientifique et culturel.

Cependant cet aspect ne sera pas abordé en détail dans le présent avis ; il faut dire qu'il requiert une analyse complémentaire approfondie. Cet avis est centré sur les questions posées par la diffusion de tous les types de résultats destinés aux milieux scientifiques par le biais des publications.

Les principaux vecteurs de diffusion aujourd'hui

Trois qualités peuvent être *a priori* attendues d'un vecteur de diffusion : tout d'abord l'excellence du travail diffusé, ensuite la rapidité de la diffusion, enfin l'amplitude du champ de la diffusion. Nous verrons que ces exigences ne sont pas toujours satisfaites.

Deux circuits de diffusion des résultats peuvent être distingués :

- Le premier consiste à publier dans une revue à comité de lecture ou dans un ouvrage publié chez un éditeur qui le fait valider par un comité de lecture, la publication étant faite sur un support papier et de plus en plus simultanément sur support numérique (version imprimée et version électronique). Ce mode traditionnel de publication répond le plus souvent au critère de qualité attendue mais il ne permet pas toujours une publication rapide (en particulier pour les ouvrages). Quant à la diffusion, elle est tributaire de celle du vecteur choisi (une revue à circulation confidentielle ou une grande revue), et de la qualité de l'éditeur.

- Le second circuit de diffusion, qui se développe rapidement, s'inscrit dans le mouvement général du libre accès et repose sur les transmissions électroniques rendues possibles grâce au développement d'Internet. Il prend lui-même deux formes :

- 1) les revues en libre accès (de type PLoS Biology, Biomed Central) où l'auteur (ou son institution de rattachement) devient le payeur, permettant à la communauté scientifique d'accéder aux résultats de la recherche sans payer d'abonnement et souvent plus rapidement que dans des revues de type traditionnel. Ce type de revue fonctionne la plupart du temps avec un comité de lecture. Cette formule, qui se développe rapidement, mérite d'être soutenue comme cela a été le cas jusqu'à ces dernières années par la mise à disposition de moyens financiers, attribués aux auteurs par les organismes. Il est à noter que les travaux publiés avec ce système ont souvent un impact significativement plus important que lorsqu'ils sont publiés dans des revues payantes. La publication en libre accès est évidemment compatible avec la formule des archives ouvertes.

- 2) les archives ouvertes, thématiques (Physique : ArXiv) ou pluridisciplinaires (HAL), sont des réservoirs où le chercheur dépose les résultats de ses recherches pour les communiquer aux collègues du monde entier, immédiatement et gratuitement. Dans ce cas, l'évaluation lors du dépôt n'est que beaucoup plus primitive et la responsabilité des auteurs est grande, mais l'avantage majeur réside dans la rapidité de communication, la très large diffusion de l'information et la gratuité totale des opérations. Au niveau des archives ouvertes, on peut schématiquement distinguer d'une part la mise à disposition des post-publications correspondant à des travaux acceptés et validés par ailleurs dans des systèmes à comité de lecture (revues traditionnelles, revues en libre accès), et d'autre part, le dépôt spontané et individuel de prépublications, comptes rendus de colloques, documents, avis... ne relevant strictement que de la responsabilité de leur auteur. Dans ce contexte, il faudrait organiser un affichage précis du statut de ces deux grandes catégories de productions permettant de les distinguer aisément dans un souci de clarté et de transparence. Par ailleurs, dans le cas de propositions

individuelles non validées antérieurement et, même si la démarche de contrôle est certainement lourde et complexe, un filtre minimal est mis en place par les responsables de l'archive (on parle généralement de « contrôles scientifiques élémentaires » pour éliminer des contributions non pertinentes qui pourraient décrédibiliser le système).

Les archives ouvertes peuvent jouer un rôle essentiel dans la diffusion totalement gratuite des résultats de la recherche en ouvrant d'une part un grand espace de liberté aux résultats insolites ou « hors modes », et d'autre part en contournant l'impérialisme de certains vecteurs de publications. Elles présentent le grand avantage de permettre un accès rapide aux connaissances et aux résultats récents pour l'ensemble des communautés, sans limitation de type financier, et peuvent favoriser la diffusion de la production scientifique des pays émergents dans un système plus « accueillant » que celui, très sélectif et onéreux, représenté par les revues. Toutefois le caractère élémentaire de validation des articles peut demeurer un frein à l'essor de ce mouvement en conduisant à une masse d'informations difficilement exploitables par le chercheur.

Par ailleurs, il existe un système d'archives institutionnelles, lieu de dépôt officiel de publications (ou autres informations) de tout un organisme, une université, exploité par les comités d'évaluation et les établissements dans le cadre de leur politique et de leurs choix stratégiques, et utilisable par les chercheurs pour leur compte rendu annuel d'activité.

III– Quelques réflexions liées aux différentes modalités de diffusion des connaissances

a) La dynamique du libre accès

Aucune bibliothèque ne peut acquérir et archiver la masse d'informations délivrées régulièrement par les quelques 20 000 revues à comité de lecture qui existent dans toutes les disciplines. Il est encore plus difficile de prendre en compte les productions individuelles plus confidentielles.

Ainsi les structures pratiquant le libre accès aux résultats de la recherche se trouvent actuellement dans une phase transitoire encore expérimentale. Cependant ce mouvement apparaît déjà comme irréversible avec pour résultat la facilitation de l'accès du lecteur à l'information. Des formules nouvelles apparaissent. L'éditeur scientifique américain à but non lucratif *Public library of science* (PLOS) vient ainsi de franchir une nouvelle étape de son développement en lançant très récemment une nouvelle revue en libre accès baptisée PLOS One qui vise à publier sur une base hebdomadaire un très grand nombre d'articles. Après une présélection par un comité de lecture sur des critères techniques et objectifs, les travaux publiés sont, dès leur publication en ligne, soumis aux commentaires, aux annotations et à l'évaluation de la communauté scientifique. La validation scientifique par les pairs (*peer review*) se pratique ainsi avant la publication mais également après la publication. Ce qui représente une formule intermédiaire entre les précédentes versions thématiques de PLOS et les archives ouvertes de type HAL.

Beaucoup d'institutions, particulièrement parmi celles qui sont basées en France et en Europe, s'engagent dans le soutien à cette politique de libre accès, à travers les archives ouvertes. Elles doivent cependant adopter un langage convaincant vis-à-vis des « producteurs » de connaissances mettant en avant les avantages d'une large diffusion, les effets de solidarité et finalement d'efficacité pour la recherche de ces nouveaux dispositifs. Elles doivent de plus apporter un soutien éditorial significatif à ces initiatives pour que la qualité des données archivées soit assurée.

Les archives ouvertes peuvent contrecarrer les intérêts des maisons d'édition et progressivement entraîner la disparition des revues imprimées traditionnelles.

Il semble donc que les deux formules actuelles du libre accès : revues en libre accès sur le principe auteur-payeur et archives ouvertes représentent des étapes vers un modèle optimisé encore à créer où les impératifs de libre diffusion des savoirs, les préoccupations éthiques et les critères de qualité et d'équilibres économiques seraient mieux conjugués.

b) L'absence ou la présence d'un filtre scientifique

Ces enjeux concernent le chercheur comme l'institution.

(1) L'absence de filtre scientifique.

Cette situation concerne principalement les archives ouvertes, mais le même raisonnement peut être tenu pour des revues sans comité de lecture ou des éditions d'ouvrages sans validation scientifique. L'absence de filtre peut provoquer des effets pervers : encombrement avec des textes médiocres ou insuffisamment mûris et soignés ; trop-plein d'informations non contrôlées et non hiérarchisées ; passage d'objectifs qualitatifs à des objectifs quantitatifs. Se pose alors le problème de la responsabilité de ceux qui mettent à disposition l'information :

- d'une part le chercheur qui dépose une production scientifique assume, ainsi que l'équipe et le laboratoire qui lui apportent leur caution, les conséquences de la diffusion de sa publication. Cette responsabilité est renforcée par le caractère définitif de son dépôt. Il peut revenir sur son texte et présenter une nouvelle version de son travail, il peut reconnaître ses erreurs, mais il ne peut retirer le texte qui a été déposé. En mettant à disposition du public son travail sur un site officiel (à distinguer d'une page personnelle), le chercheur – en connaissance de cause – abandonne sa possibilité de retrait, de la même manière que dans une revue papier. Sa production scientifique s'inscrit dans le processus du savoir et rentre dans une mémoire collective que l'on ne peut effacer ;

- d'autre part, l'organisme qui met à disposition et organise ce mode de diffusion et qui contrôle superficiellement l'introduction des textes dans la base de données engage aussi sa responsabilité : en l'absence de filtre scientifique, les utilisateurs peuvent interpréter le statut des institutions comme une garantie de validation.

(2) La présence d'un filtre scientifique :

Même si la validation par les pairs est considérée comme une sécurité, certaines dérives peuvent concerner chacun des acteurs du processus.

Tout d'abord, l'auteur engage sa responsabilité dans le choix de la revue en fonction du type de production à publier. Il peut aussi avoir tendance à faire évoluer son travail pour rentrer dans le cadre tacite ou imposé de la revue. Ce formatage, s'il est excessif, nuit à l'originalité et à l'inventivité. L'éditeur de la revue engage sa responsabilité dans le choix des rapporteurs. Ceux-ci ont à respecter les règles relatives aux conflits d'intérêts ; on attend d'eux : sérieux, rigueur et honnêteté dans l'évaluation et ils doivent se garder de profiter de leur accès privilégié – avant les autres - au contenu des articles (voir sur ces points la recommandation sur l'évaluation).

Les avantages d'une validation par les pairs sont certains mais le formalisme des revues peut aussi

entraîner des effets pervers. Si on n'y prend pas garde, le système peut contribuer à instaurer un conformisme scientifique refusant tout ce qui se démarque des principes reconnus. Par ailleurs la pression à publier, que le système fait peser sur les chercheurs, peut être un frein à une recherche originale.

Enfin, on ne saurait trop recommander la prudence dans l'exploitation systématique des indicateurs (facteurs d'impact des revues, indice de citation des auteurs). Le risque est encore plus grand si ceux-ci ne sont pas suffisamment diversifiés. La simplicité d'exploitation de ces formules peut séduire de façon abusive les évaluateurs et les pousser à donner trop d'importance à ces indicateurs monotones parfois inadaptés en raison de leur conception et de leurs formalismes.

Au final les variantes dans les modalités de diffusion électronique des résultats se différencient considérablement. Si les revues en « libre accès » représentent une innovation intéressante (gratuité pour le lecteur), elles ne correspondent qu'à un aménagement de l'existant (auteur payeur) avec les mêmes effets de monopole, d'exclusivité et de freins financiers pour la diffusion des connaissances.

Les archives ouvertes représentent davantage une véritable révolution avec une réappropriation de l'information scientifique par les scientifiques et une modification profonde de l'appréciation du travail effectué qui devrait passer, si le système évolue favorablement, d'une évaluation *a priori* (rapporteurs, facteur d'impact du journal) à une évaluation *a posteriori* (celle de la communauté scientifique exprimée par les taux de citation).

Les auteurs et les lecteurs sont en prise directe et l'auteur comprendra qu'en publiant dans ce système, il archive en même temps ses résultats « pour toujours ». Cette énorme responsabilité non partagée (avec des rapporteurs, des éditeurs...) devrait jouer en faveur d'une certaine autorégulation.

c) Sur le référencement des publications

Le référencement des publications par les institutions qui financent les auteurs signataires de la publication est un problème sensible. Depuis des décennies, les spécialistes de la bibliométrie savent qu'il n'existe aucun outil permettant d'extraire automatiquement la production d'un laboratoire, encore moins celle d'un organisme de recherche. Cela est d'autant plus sensible en France, où les chercheurs appartiennent très souvent à des laboratoires ayant plusieurs tutelles. Si tous les établissements de rattachement ne sont pas mentionnés, les manquants seront pénalisés, comme peut l'être une université dans le classement dit « de Shanghai » (qui toutefois ne dépend que peu des publications, puisqu'il donne le poids le plus grand aux prix Nobel). Mais les politiques éditoriales des journaux, dont le but est scientifique et pas du tout bibliométrique, ne permettent pas de multiplier les affiliations, qui sont très souvent supprimées lors de la publication. Le problème est difficile, et l'Observatoire des sciences et des techniques ainsi que beaucoup d'organismes français se penchent actuellement sur cette question. La situation actuelle est cacophonique, chaque établissement cherchant à promouvoir une politique qui ne prend en compte que son propre intérêt. Le CNRS, toutefois, a publié sur son site des recommandations de rédaction d'adresses qui prennent soigneusement en compte celles des universités et sont compatibles avec les contraintes éditoriales des revues scientifiques.

Il est clair également que le nombre parfois excessif de contributeurs à une publication et leur positionnement relatif dans la liste des auteurs, variable selon les disciplines, peut entraîner des difficultés dans l'évaluation. Ces aspects mériteraient d'être normalisés.

d) L'accélération de la transmission des connaissances par voies électroniques et via Internet

Comparées aux formes traditionnelles de communication orale ou écrite, les nouvelles modalités électroniques de large diffusion des connaissances, représentent une rupture importante facilitant la mise à disposition rapide d'un immense volant de données d'origines diverses.

Nous avons vu l'innovation apportée par les revues en libre accès et les archives ouvertes dans les milieux de la recherche. Au-delà, du chercheur au collégien, en passant par le simple citoyen, les moteurs de recherche permettent d'accéder à des informations nombreuses et diversifiées sur un sujet donné par utilisation de mots-clefs adaptés.

Cette nouvelle approche extrêmement séduisante et efficace peut s'accompagner d'inconvénients. Les moteurs de recherche délivrent des messages dans des champs ou selon des hiérarchies qui leur sont propres. L'investigateur, aspiré par la facilité de la démarche, peut perdre son sens critique et répercuter les différentes données sans analyse personnelle. Enfin il serait très intéressant d'avoir la possibilité, lors d'une prospection sur un problème spécifique, de collecter des combinaisons pertinentes de ressources issues de champs scientifiques différents dans une optique pluridisciplinaire.

Dans le domaine des publications, et des banques de données au sens large, la transition vers une numérisation accrue des savoirs implique de porter une attention soutenue à la définition et la pertinence des « méta données » : référencement, indexation par le contenu, liens... qui permettront de localiser, identifier les résultats concernés. En effet un travail de qualité non repérable devient sans intérêt. Cet aspect crucial implique une coopération entre les documentalistes de nouvelle génération préparés à ces fonctions spécifiques et les chercheurs aptes à bien décrire et à déposer les données de leur recherche.

e) Une réflexion active à l'échelle européenne :

Des initiatives se développent à l'échelle européenne visant à stimuler la réflexion sur les futures politiques de publication des résultats de la recherche en prenant compte les retombées de la révolution numérique :

- récente pétition demandant une garantie d'accès public aux résultats des recherches financées sur fonds publics peu de temps après leur publication ;
- conférence dédiée "*Scientific publishing in the European area – Access, dissemination and preservation in the digital age*" (Bruxelles, 15 février 2007).

Il est clair que des modalités de publication des résultats techniquement efficaces, facilitant un large accès aux utilisateurs et prenant en compte les intérêts des différents protagonistes représentent une composante significative d'un système global de recherche efficace et équilibré.

Les deux questions principales posées par la Communauté européenne concernent :

- la mise à disposition des scientifiques de moyens rapides et largement ouverts de diffusion des résultats ;
- les moyens de concilier ces exigences avec la juste rémunération des éditeurs scientifiques qui investissent dans les mécanismes d'évaluation par les pairs et de circulation de l'information.

La confrontation actuelle des points de vue prépare sans doute des évolutions significatives sont prévisibles au niveau européen visant, en particulier à court terme, à favoriser l'accessibilité des résultats des recherches financées dans le cadre du 7ème programme-cadre.

V. Enjeux éthiques dans le domaine de la diffusion des connaissances

a) Le problème général d'un large partage des connaissances versus une attitude élitiste de restriction de la diffusion des savoirs

Le système actuel de diffusion du savoir, segmenté, et restrictif, peut se révéler un frein à l'efficacité de la recherche et à l'utilisation de ses bénéfices par la société. Les scientifiques ont donc un accès inégal et toujours partiel à ce qui constitue pourtant le produit de leur travail. Le problème est particulièrement aigu dans les pays en développement mais aussi au niveau des communicateurs scientifiques des pays développés et *a fortiori* du grand public.

Se pose donc le problème général du droit d'accès aux résultats de la recherche. Procurer au plus grand nombre un accès grandement facilité d'une connaissance acquise grâce à un effort financier public en la considérant comme un bien universel largement distribué, relève d'une préoccupation éthique majeure. Cependant il conviendrait de mieux coordonner les efforts de diffusion des résultats à une analyse de leurs conséquences pour l'homme et la société.

b) Les antagonismes et conflits d'intérêts dans l'édition scientifique

Aujourd'hui le débat n'est plus uniquement celui d'une opposition schématique entre diffusion électronique libre et revues traditionnelles ; il s'inscrit plus globalement dans une réflexion sur le modèle économique dans lequel devrait s'inscrire l'édition scientifique. Il y a donc nécessité d'anticiper et de négocier une politique commune, qui pourrait être d'inspiration européenne, des établissements de recherche et des éditeurs reposant sur la répartition des responsabilités et la définition d'aides institutionnelles (aide à la publication, à la diffusion, à la reprographie ou à la copie numérique à des fins de recherche).

c) Les notions de monopole ou d'impérialisme dans la diffusion des connaissances

Les positions dominantes de certains journaux, qui ont toujours existé, trouvent un prolongement maintenant au niveau des nouvelles formes d'édition scientifique. Le fait que les revues en libre accès soient payantes conduit à des tensions puisque les acteurs scientifiques doivent payer cher pour publier les travaux dont ils sont les auteurs, dans le cadre d'un système auquel ils contribuent également en tant qu'évaluateurs.

L'existence d'un nombre assez limité de moteurs de recherche réellement consultés conduit potentiellement à un monopole de la diffusion des connaissances sur Internet susceptible de contrôler abusivement la diversité des informations ou leur hiérarchie. On peut souligner ici un abus de situation dominante.

Des problèmes de langue se posent également. Si le français ou l'allemand ont encore le droit d'existence dans certains domaines scientifiques, ce n'est pas le cas pour de nombreuses disciplines où l'anglais est devenu la référence, qu'il faut utiliser si l'on veut faire connaître son travail. Les Anglo-Saxons sont donc avantagés, puisqu'ils peuvent publier dans une langue qu'ils maîtrisent parfaitement.

La tentation est forte de ne plus publier qu'en anglais, mais cela aura des conséquences importantes pour la langue française, qui ne se développera plus sur le terrain scientifique. Enfin, toute langue est porteuse d'une référence culturelle et d'une démarche de pensée qu'il importe de conserver. L'utilisation d'une langue unique aboutit donc à un appauvrissement intellectuel dans l'expression des résultats de la recherche.

d) Les dimensions éthiques associées à la décision de publication scientifique et à l'exploitation des publications dans l'évaluation

Tout en mettant en œuvre les règles de bonne pratique et de la déontologie en vigueur dans les centres de recherche et dans les systèmes de publication (citation des travaux antérieurs, implication correcte des contributeurs, vérification rigoureuse et reproductibilité des résultats...), le chercheur doit s'interroger sur l'opportunité et la pertinence de sa décision de communiquer ses résultats. Le travail est-il suffisamment abouti ? Quelles seront les répercussions de sa publication ? Le choix de la revue est-il totalement fondé et n'est-il pas le résultat d'un effet de mode ? Dans le cas d'archives ouvertes peu sélectives, les résultats proposés sont-ils réellement significatifs et ne vont-ils pas encombrer un système déjà très lourd ?

Les évaluateurs et comités d'évaluation doivent être attentifs à ne pas succomber aux facilités associées à une exploitation exclusive de paramètres liés à la nature des revues ou aux fréquences de citation. Ces systèmes ont leurs finalités qui ne correspondent pas forcément aux objectifs d'une évaluation équilibrée de l'opération de recherche considérée.

VI. En forme de conclusion provisoire

Diffuser ses connaissances et ses résultats correspond à une des missions du chercheur au bénéfice des milieux scientifiques et de la société.

La diffusion organisée des résultats de la recherche représente une activité obéissant à des codes et des modalités qui ont progressivement évolué. Cette évolution s'accélère considérablement avec l'exploitation croissante des moyens électroniques pour la diffusion de résultats, d'informations et de données à travers différentes formules : revues en libre accès avec comité de lecture, archives ouvertes, banque de données (ce dernier point n'a pas été considéré dans ce rapport et méritera une analyse particulière).

Ces nouvelles formules, qui concurrencent les traditionnelles revues papier à comité de lecture peuvent contribuer à l'effondrement du rayonnement des revues et corrélativement des bibliothèques. Ces points méritent d'être évalués et anticipés.

La transition vers le tout numérique, si elle offre indiscutablement des avantages en terme de rapidité d'obtention et d'accessibilité gratuite à l'information pour le lecteur, présente également des inconvénients qu'il faudra maîtriser : caractère très onéreux pour les laboratoires des revues en libre accès sur le principe auteur-payeur, hétérogénéité potentielle et insuffisance des filtres de sélection pour le matériel publié librement dans les archives ouvertes.

L'accélération généralisée des opérations de publication liée à ces nouveaux moyens peut conduire à une insuffisante responsabilisation des auteurs au niveau de la qualité des résultats publiés et secondairement de la qualité de l'écrit. Le système des index-citations qui s'amplifie peut par ailleurs

prendre une importance exagérée dans l'évaluation en raison de sa facilité d'utilisation, évaluation dont il est indispensable de diversifier les critères.

Enfin la masse accrue d'informations n'accroît pas forcément la capacité de se former un jugement tant qu'on n'a pas les moyens de discriminer, identifier et obtenir les données pertinentes. D'où des effets potentiellement négatifs d'un accroissement du volume des informations, un petit nombre d'initiés continuant d'échanger de façon élitiste et privilégiée les données les plus significatives.

Dans ce contexte général, il semble donc que des progrès doivent être réalisées dans plusieurs directions. Les différentes formules du libre accès représentent sans doute des étapes intermédiaires dans l'émergence de solutions concertées plus satisfaisantes, qui conjugueraient les avantages et minimiseraient les inconvénients identifiés, en prenant en compte de façon équilibrée les dimensions économiques associées. De plus, un effort particulier doit être porté aux infrastructures et au soutien éditorial qui doivent accompagner le système des archives ouvertes pour leur pleine efficacité. Enfin une attention soutenue doit être portée à la création, gestion et exploitation des méta données qui conditionnent une recherche optimisée de l'information pertinente.

Ces différentes dimensions concourent à l'évidence à la satisfaction d'une préoccupation éthique majeure, à savoir la distribution la plus large, la plus efficace et la plus accessible de résultats de qualité au bénéfice des sociétés humaines.

Recommandations :

- Veiller à identifier et à soutenir des systèmes de publications des résultats de la recherche largement accessible et de qualité, de manière à assurer une diffusion du savoir la plus large possible aux bénéficiaires de la science et de la société.
- Sensibiliser les acteurs de la recherche de façon claire, honnête et complète aux avantages/inconvénients des différentes modalités de communication de leurs résultats et faciliter la compréhension et les avantages des systèmes en libre accès. Dans le cas des archives ouvertes, les institutions doivent en particulier, très clairement souligner, par-delà l'intérêt général, leurs intérêts spécifiques au niveau du suivi et de l'évaluation de la production scientifique. Dans cette période intermédiaire ces institutions doivent aussi prodiguer aux chercheurs les accompagnements nécessaires concernant les aspects techniques et réglementaires des nouvelles procédures. Ces attitudes sont à même d'effacer la tiédeur et parfois même l'inertie des milieux de la recherche vis-à-vis de ces évolutions et les conduire à s'approprier cette nouvelle culture de la diffusion. Par ailleurs, dans cette période intermédiaire, on peut comprendre qu'au niveau des organismes soient pour le moment adoptées de simples positions d'incitation à déposer les travaux dans des archives ouvertes. Cependant, à terme, c'est l'obligation qui donnera au système toute son efficacité. Cette obligation sera d'autant mieux acceptée que l'explicitation du système aura été satisfaisante.
- Assurer au niveau des organismes et institutions un soutien éditorial significatif pour que la qualité et la large dissémination des données archivées (archives ouvertes) soient assurées.
- Inciter les établissements à développer une concertation et des partenariats avec les acteurs de l'édition, publique et privée, sous toutes ses formes pour aboutir à des solutions équilibrées prenant en compte les intérêts de chacun.
- Réfléchir à corriger les abus de situation dominante et de monopole. Il est difficile d'inverser l'hégémonie exercée par des revues prestigieuses ou de nouveaux systèmes efficaces de diffusion électronique ou moteurs de recherche. Il est cependant possible par des initiatives, des soutiens financiers, des codes de conduite dans l'évaluation, de prendre des mesures atténuant des situations éthiquement discutables.
- Envisager des formules pour contrebalancer la prédominance de l'anglais dans les systèmes de diffusion de connaissance en favorisant en particulier le bilinguisme et éventuellement dans certaines disciplines la traduction assistée par ordinateur. Ces démarches peuvent paraître vaines mais elles s'inscrivent dans le souci, qui doit demeurer permanent, de la protection et de la promotion de la diversité des expressions culturelles (Convention Unesco 2005).
- Sensibiliser « le producteur de connaissances » à sa responsabilité éthique dans les publications institutionnelles du CNRS, dans les recommandations aux chercheurs entrants et identifier une « structure contact » à laquelle le chercheur pourrait se référer en cas de besoin, dans le contexte de la diffusion des résultats de sa recherche. L'acte de publication doit être raisonné, la contribution doit être appropriée, équilibrée, justifiée en dépassant les seules motivations de progression de carrière ou de présence dans la compétition internationale.
- Veiller au niveau de l'évaluation à se former un jugement sur des indicateurs diversifiés prenant en compte l'originalité, l'inventivité des résultats, sans se limiter à des critères uniquement quantitatifs tels que les facteurs d'impact des revues et les index de citations, dont la grande facilité d'utilisation peut conduire à ne fonder l'évaluation que sur des appréciations de type bibliométrique.

ANNEXE N°2 - ENTRETIEN AVEC EMMANUELLE MORLOCK

Vendredi 30 mars 2018, laboratoire HiSoMA, 11h30 à 13h00.

Emmanuelle Morlock est ingénieure d'étude au CNRS, responsable de systèmes d'information documentaire, en charge du programme « Réflexions sur les pratiques d'édition et d'archivage » de l'axe transversal *Edition, archives, humanités numériques* du laboratoire HiSoMA, elle est auteure d'un billet de veille sur les Humanités numériques.

Pauline Vignaud : Bonjour Emmanuelle Morlock. Je m'appelle Pauline Vignaud. Dans le cadre de ma formation, je rédige un mémoire de recherche sur les enjeux de l'interopérabilité dans la diffusion et la valorisation des données archéologiques. Par ailleurs, je m'intéresse aux archives de l'archéologie : comment les définiriez-vous ?

Emmanuelle Morlock : « C'est une très bonne question. Je souhaitais justement revenir sur le mot « archive » qui me dérange un peu car il semble, de manière générale, être utilisé de façon un peu abusive par les chercheurs. Il existe un vrai problème avec ce mot. Les chercheurs l'utilisent massivement pour désigner uniquement les sources primaires, les sources « brutes » mais, du point de vue archivistique cela ne correspond pas. En réalité, les chercheurs produisent des documents, ils font principalement de la production documentaire et créent des archives courantes voire intermédiaires. La première mission des chercheurs est de publier des articles de recherche, ils ne sont pas archivistes. Il existe un vrai conflit dans la relation archiviste/chercheur. L'idée de mettre en place des plans de gestion des données est là pour venir, en quelque sorte, réconcilier le chercheur et l'archiviste. C'est important de montrer au chercheur qu'il a une vraie responsabilité dans sa production documentaire qui n'a pas vocation à être gardé dans un coin reclus de tous. Tout ce qui sera partagé aura plus de chance d'être gardé. Les choses sont préservés parce qu'il y a un intérêt à les sauvegarder. Je trouve que le numérique, impacte réellement les pratiques dans le monde de la recherche. Il faut vraiment réfléchir aux exploitations scientifiques, c'est la responsabilité du chercheur d'intellectualiser sa production documentaire, de la réfléchir, ce n'est pas le rôle de l'informaticien de choisir les outils nécessaires à cette production et cette diffusion. A mon sens, une archive n'est pas une archive mais une documentation primaire scientifique. « Archiver » veut dire sélectionner, trier, décrire et préserver, selon des critères, selon la connaissance d'un fonds, etc... Il y a un problème et même temps une mutation avec ce mot, c'est très complexe à appréhender. Dans cette vision

« archives », il y a une chose qui n'est pas dans le spectre des chercheurs : la gestion de fichiers dont les chercheurs veulent extraire les données, c'est assez faussé comme vision. Il faut rester prudent et avoir des vraies discussions où personne ne se laisse piéger sur ses points de vue. »

Pauline Vignaud : Que signifie, selon vous, diffuser et valoriser ? Quelles définitions mettez-vous derrière ces deux verbes ?

Emmanuelle Morlock : « Pour vous, que signifie valoriser et diffuser ? Existe-il une réelle différence entre ces deux mots ? La valorisation signifie : « donner de la valeur économique à quelque chose », d'où les brevets et la publication, c'est mettre en valeur, c'est la valorisation au sens économique du terme. Puis il y a le verbe « valoriser » qui implique de mettre des critères pour obtenir des données factuelles. Les datapapers ont un rôle à jouer sur le téléchargement, la réutilisation et la traçabilité des données. On peut aussi considérer que la diffusion est une forme de valorisation. Il y a la diffusion traditionnelle sur support papier avec la publication qui inclut la conception, la fabrication et diffusion – j'entends par là, la vente - avec les démarches d'auto publication. Selon moi, publier et diffuser c'est la même chose avec une spécialisation différente, la diffusion c'est le fait de mettre en œuvre certaines actions pour obtenir un résultat et la publication c'est le résultat d'une chaîne opératoire. Les questions de dépôt sont problématiques pour les chercheurs, certains pensent que diffuser les données sur une plateforme c'est publier mais ce n'est qu'une forme de publication, parmi tant d'autres, non conventionnée et non répertoriée. On peut considérer le data publication as curation, c'est-à-dire qu'on intègre les données dans un concept de publication, par exemple la plateforme OpenContext DINAA⁹⁷. C'est une plateforme de données archéologiques s'appliquant sur le modèles des data curation pour les services numériques. Cette plateforme de preview en archéologie est peu connue mais c'est un interlocuteur clef dans le domaine dont la communauté est très ouverte. »

⁹⁷ Le DINAA signifie littéralement *Digital Index of North America Archeology* (Indice numérique de l'archéologie nord-américaine). Cette plateforme regroupe plusieurs ensembles de données archéologiques et historiques produites au cours du XIX^{ème} siècle à partir de nombreuses sources dont les gouvernements étatiques et fédéraux des Etats-Unis. DINAA est un projet collaboratif impliquant des chercheurs de plusieurs universités américaines (Tennessee, Indiana, Oak Ridge National Laboratory), OpenContext et des partenaires gouvernementaux. La DINAA se concentre sur la publication de documents archéologiques sur les « sites files ». Chaque « site file » décrit un site archéologique enregistré par des organismes gouvernementaux fédéraux et étatiques qui appliquent les lois fédérales de protection historique. De nombreux bureaux gouvernementaux ont directement fourni des données à la DINAA. Ces données sont expurgées d'informations confidentielles comme les informations de localisation afin qu'elles puissent être rendues publiques.

Pauline Vignaud : *Pourquoi selon vous est-il nécessaire de strictement séparer plateforme de partage des données et plateforme d'archivage des données ?*

Emmanuelle Morlock : *« Chaque plateforme ne répond pas à un même modèle. Les plateformes d'archivage pérenne répondent, elles, au modèle OAIS avec des paquets comme la TGIR Huma-num, intermédiaire pour le CINES, qui récupère les données ayant une valeur patrimoniale importante pour les archiver, les diffuser, les indexer, les valoriser, la TGIR Huma-Num propose de nombreux services. Par exemple, leur plateforme Nakala, ce n'est pas une plateforme d'archivage mais une plateforme de partage et de diffusion des données de la recherche. Chaque plateforme a ses propres demandes, cela dépend sensiblement de comment on perçoit et définit l'archivage pérenne. Je ne sais pas si vous connaissez le projet ERC dans Horizon 2020⁹⁸, c'est un programme récompensant l'excellence scientifique, j'ai en tête un exemple bien précis, celui d'un projet sur des données géographiques extraites d'un manuscrit du IV^{ème} siècle, cela permet de se faire une idée de la représentation du monde à cette époque. Ce type de projet pose la question de sa perspective d'étude, de sa durée de validité scientifique différente de sa durée d'utilité administrative. A ce moment-là, la bonne stratégie est de parler de l'intérêt patrimonial à long terme et de la durée de responsabilité engagée pour assurer le projet, ici le minimum c'est de 3 à 5 ans.*

Nous parlions des plateformes de partage des données tout à l'heure et à mon sens il y a une vraie différence avec les plateformes d'archivage, la notion de critères, de normes est très importante. Selon moi on ne dépose pas les mêmes données sur une plateforme de partage que sur une plateforme d'archivage. La plateforme de partage est là pour partager, diffuser des données à une communauté tandis que la plateforme d'archivage est choisie pour un rôle de conservation des données avec une logique de fonds et de durée. L'idée de la GED comme plateforme de pré-archivage me gêne un peu, est-ce utile ? Je vous invite à lire un article d'Archimag sur le sujet expliquant que la GED n'a plus de beaux jours devant elle. Lorsqu'on installe une GED dans un laboratoire comme HiSoMA, il faut aider à accompagner le changement, il faut un personnel mutualisé, il faut faire des choix car certaines données ont plus de valeur que d'autres. La politique des laboratoires de recherche est de conditionner le soutien des ingénieurs. Les décisions politiques sont difficiles à prendre, investir sur des ressources – humaines ou financières – pour archiver les données c'est un travail lourd et à long terme dont

⁹⁸ Le projet ERC signifie littéralement *European Research Council*. C'est un programme « scientifique blanc » dédié à la recherche exploratoire, dont le critère de sélection est l'excellence scientifique.

on ne perçoit pas directement les bénéfiques. La Maison de l'Orient et de la Méditerranée doit responsabiliser les chercheurs en montrant les bonnes pratiques à adopter, montrer que l'archivage des données a un sens, que leur réutilisation dans les décennies à venir importe. Il faut sensibiliser c'est un élément clef favorisant le contexte d'un bon archivage. Cela rejoint les questions de l'interopérabilité. »

Pauline Vignaud : Comment définiriez-vous l'interopérabilité ? Quels problèmes cela soulève en archéologie ?

Emmanuelle Morlock : « *Selon moi, il y a deux interopérabilités : tout d'abord il y a l'interopérabilité selon les formats syntactiques et sémantiques. Prenons par exemple un article rédigé il y a plusieurs années, cet article est tellement hétérogène qu'il n'a aucune unité sémantique et aucune interopérabilité sémantique, il est donc impossible de le lire ou de le partager. Pour garantir l'interopérabilité il faut des vocabulaires partagés, contrôlés et des références de haut-niveau tel que le CIDOC-CRM. Néanmoins, le problème de ces vocabulaires c'est qu'ils ont été conçus comme des thésaurus pour les bibliothèques. Prenons par exemple les thésaurus PACTOLS que vous connaissez, ils ont une spécificité scientifique légitime. Le travail du documentaliste n'est pas d'aller dans le détail des descriptions des métadonnées mais de signaler, identifier et retrouver un document qu'un institut ou une bibliothèque possède. Les chercheurs sont souvent perdus face aux enjeux recouverts par les métiers de l'information, ils ne peuvent pas comprendre pourquoi les thésaurus PACTOLS ne sont pas assez précis. A mon sens les thésaurus PACTOLS doivent se penser comme des outils permettant l'interopérabilité et la découvrabilité. Les vocabulaires scientifiques, les consensus, les négociations et les débats terminologiques sont trop nombreux, il faudrait un décret ou une instruction ministérielle pour répondre définitivement à ce problème. Je vais vous donner un autre exemple : l'épigraphie. En épigraphie, les projets de recherche ont des matériaux, des inscriptions sur pierre, chaque projet aura besoin d'être affiné. Il faut donc un consensus général dès le départ car les chercheurs ne respectent pas l'utilisation des thésaurus PACTOLS, ils le laissent de côté et produisent leur propre vocabulaire. Il vaut mieux se restreindre à utiliser un vocabulaire commun et contribuer à son enrichissement que chercher à créer le sien dans son coin. Cela fonctionne parfaitement en TEI. Ici la véritable problématique c'est la publication des productions scientifiques. D'un autre côté, le danger d'avoir un vocabulaire précis est de ne pas avoir de données interopérables et de ne pas pouvoir les diffuser. La bonne méthode serait de faire son vocabulaire particulier et de l'aligner grâce à OpenTheso, il faut choisir ce qu'il est pertinent d'exposer, le travail avec*

l'informaticien doit être au centre des priorités du chercheur s'il souhaite diffuser ses données. Aujourd'hui on publie essentiellement pour les machines : le chercheur doit, par exemple, comprendre l'expressivité du modèle RDF et ses limites. A l'aide d'un thésaurus aligné, on peut travailler sur un vocabulaire « maison » au moment de la publication et l'intégrer. Intégrer l'interopérabilité c'est tester des scripts pour générer une automatisation. Le plan de gestion de données est un outil essentiel pour l'enrichissement des données et le mieux est de respecter les principes FAIR. »

Pauline Vignaud : Comment selon vous garantir l'interopérabilité des données ?

Emmanuelle Morlock : « *En définissant clairement le type d'interopérabilité souhaité et le niveau visé. Le principe FAIR est le meilleur outil systématique pour garantir l'interopérabilité des données. Le principe fonctionne par grade, c'est un fonctionnement dans la progression, qui ne prend pas uniquement en compte la dimension purement technique de l'interopérabilité. Le principe FAIR s'attarde sur les moyens organisationnels à mettre en œuvre pour garantir l'interopérabilité, il faut donc des processus contre qualité pour pouvoir valider ce principe. Il faut tester l'organisation et la pertinence des données et de leurs descriptions, il faut réussir à organiser la description des données au moment de la collecte et les renseigner par lot ou par dossier tout en callant la description sur les processus de production. En archéologie, on est contraint par la synchronisation des méthodes, des fiches descriptives : il fut tout tester et ne pas envisager indépendamment un contexte de production. Est-ce la donnée est interopérable dans l'écosystème actuel de valorisation et de diffusion ? Il faut mettre en regard le système de référence de l'interopérabilité et l'état de l'art aujourd'hui. L'utilisation du Dublin Core, de l'OAIPHM, l'alignement des vocabulaires de description des données, le CIDOC-CRM présentent un avantage : la relation mot-clef/objet qui crée l'interopérabilité et un lien sémantique. L'interopérabilité est une chose vers laquelle tous systèmes ou SI tendent. L'interopérabilité n'est cependant pas un objectif. On a besoin de connaître l'état de l'art d'une discipline scientifique pour pouvoir travailler efficacement, la personne révisant notre projet de publication ou de diffusion par exemple aura une certaine idée de notre positionnement via l'état de l'art. aujourd'hui être interopérable c'est être Dublin Core, pourtant si on est interopérable sémantiquement pas au niveau des fichiers, cela ne sert à rien, cela donne un sentiment de fausse sécurité ».*

ANNEXE N°3 - RECOMMANDATIONS DU CONSORTIUM MASA POUR LA PUBLICATION ET LE PARTAGE DES JEUX DE DONNEE



Publish and share archaeological data sets



Metadata and identification

- Accessibility and metadata referencing (Huma-Num Nakala service).
- Unique persistent identifier ARK to provide permanent access to resources (URI).



Data exposure on the web



ArSol (archaeological excavation archives of Tours, France), MSH Val de Loire



EpiCherchel, ancient inscriptions of Cherchel (Algeria), MMSH Aix-en-Provence



Gaul Sanctuaries, MAE Nanterre



Document library M. De Boüard, MRSH Caen

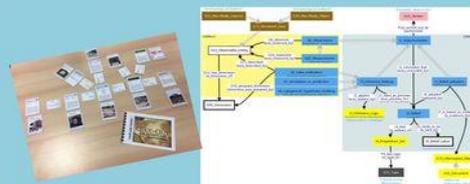


Reference vocabulary



- Data enrichment with PACTOLS, multilingual thesaurus (ISO 5964) aligned in the web.
- Evolution of Opentheso, free thesaurus manager.

CIDOC-CRM ontology (ISO 21127)



- Collaboration with ICS-FORTH / CRM-SIG.
- Initiation to the use of the CIDOC with cards game.
- Training to map data with CIDOC.



Platforms integration

- Creation of an interface to query archaeological data sets.
- Supplying the platform of the European ARIADNE programme.



<https://masa.hypotheses.org>



TABLE DES MATIERES

INTRODUCTION.....	7
ARCHIVER, DIFFUSER ET VALORISER L'ENFOUI : ENJEUX, CONTRAINTES ET NOUVELLES PERSPECTIVES.....	11
Du sol au hors sol : archiver l'absence.....	11
Qu'est-ce qu'une donnée archéologique ?	11
Les archives des archéologues	14
Normes et réglementation	17
Diffuser, valoriser... et archiver.....	20
Diffusion et valorisation : enjeu, nécessité ou obligation ?.....	20
L'archivage des données archéologiques	23
<i>Open Access/Open Data</i> : les projets de l'archéologie	25
L'archéologie numérique : de nouvelles possibilités	27
UTILISER ET REUTILISER LES DONNEES ARCHEOLOGIQUES : UNE QUESTION D'INTEROPERABILITE ?.....	32
L'importance des métadonnées en archéologie.....	32
Des normes et des standards pour l'interopérabilité des données et des métadonnées.....	36
Le vocabulaire : du bon usage des thésaurus	36
Ontologie(s) : l'application du CIDOC-CRM.....	40
Bases de données et redocumentarisation	44
Réutiliser les données et les métadonnées : l'unique défi de la redocumentarisation ?.....	44
Les bases de données en archéologie : l'interopérabilité au service de la diffusion et de la valorisation	48
PRECONISATIONS ET BONNES PRATIQUES : DU PLAN DE GESTION DES DONNEES ET DES DIFFERENTS NIVEAUX DE L'INTEROPERABILITE A L' <i>OPEN ACCESS</i>	53
Créer un plan de gestion des données	53
Garantir les différents niveaux d'interopérabilité	58
Systématiser l'ouverture des données et le libre accès	62
CONCLUSION.....	65
BIBLIOGRAPHIE	67
ANNEXES	72
TABLE DES MATIERES	96
Table des matières	96