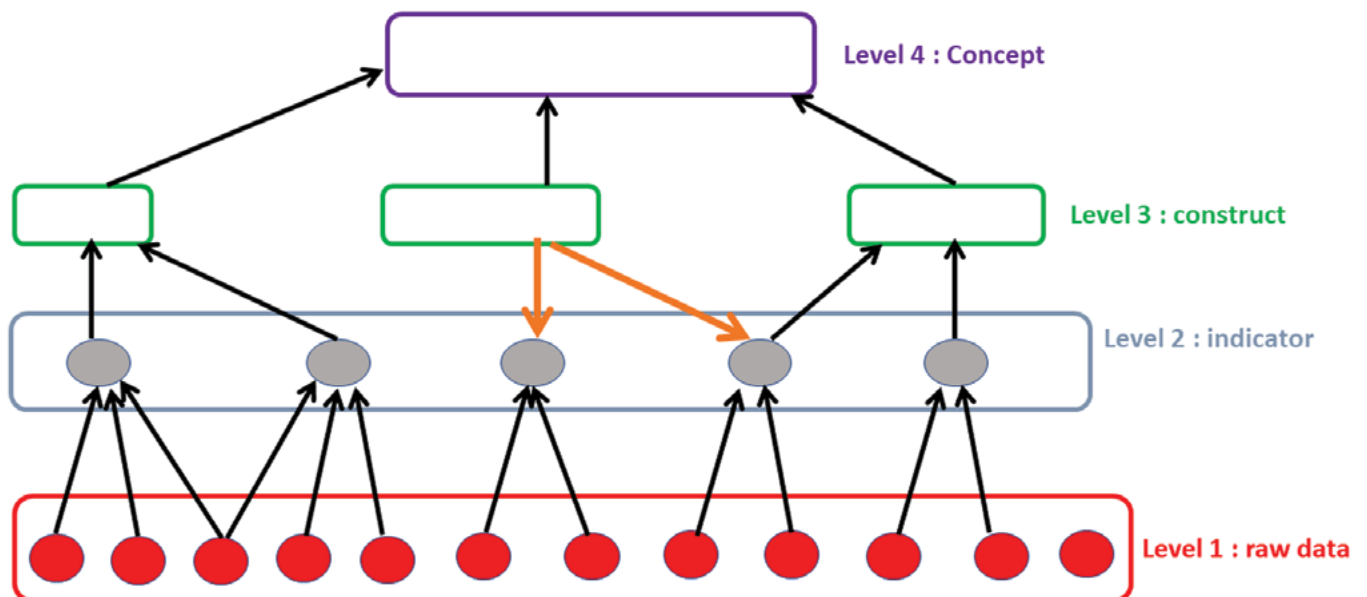


Big Data and Archaeology

edited by

François Djindjian and Paola Moscati



Big Data and Archaeology

Proceedings of the XVIII UISPP World Congress
(4-9 June 2018, Paris, France)

Volume 15

Session III-1

edited by

François Djindjian and Paola Moscati



ARCHAEOPRESS PUBLISHING LTD
Summertown Pavilion
18-24 Middle Way
Summertown
Oxford OX2 7LG

www.archaeopress.com

ISBN 978-1-78969-721-6
ISBN 978-1-78969-722-3 (e-Pdf)

© Archaeopress, UISPP and authors 2021

This book is available direct from Archaeopress or from our website www.archaeopress.com



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivatives 4.0 International License

UISPP PROCEEDINGS SERIES VOLUME 15 – Big Data and Archaeology

UISPP XVIII World Congress 2018

(4-9 Juin 2018, Paris)

Session III-1

VOLUME EDITORS: François Djindjian and Paola Moscati

SERIES EDITOR: The Board of UISPP

SERIES PROPERTY: UISPP – International Union of Prehistoric and Protohistoric Sciences

© 2021, UISPP and authors

KEY-WORDS IN THIS VOLUME: archaeology, archaeological theory, archaeological computing, Big Data

UISPP PROCEEDINGS SERIES is a printed on demand and an open access publication,
edited by UISPP through Archaeopress

BOARD OF UISPP: François Djindjian (President), Marta Arzarello (Secretary-General), Apostolos Sarris (Treasurer), Abdulaye Camara (Vice President), Erika Robrahn Gonzalez (Vice President). The Executive Committee of UISPP also includes the Presidents of all the international scientific commissions (www.uispp.org).

BOARD OF THE XVIIIe UISPP CONGRESS: François Djindjian, François Giligny, Laurent Costa, Pascal Depaepe, Katherine Gruel, Lioudmila Iakovleva, Anne-Marie Moigne, Sandrine Robert



Foreword to the XVIII UISPP Congress Proceedings

UISPP has a long history, starting 1865 with the International Congress of Prehistoric Anthropology and Archaeology (CIAAP), until 1931, date of the foundation in Bern of UISPP. In 1955, UISPP became member of the International Council of Philosophy and Human Sciences, associate of UNESCO.

UISPP is structured on more than thirty scientific commissions which a very representative network of worldwide specialist of prehistory and Protohistory, and covering all the specialties of archaeology: historiography, archaeological methods and theory; material Culture by period (palaeolithic, Neolithic, bronze age, Iron age) and by continents (Europe, Asia, Africa, Pacific, America), palaeoenvironment and palaeoclimatology; archaeology in specific environments (mountains, desert, steppes, tropical area), archeometry; Art and culture; Technology and economy; biological anthropology; funerary archaeology; Archaeology and societies.

The UISPP XVIII° world congress of 2018, in Paris, France with the strong support of all the French institutions related to archaeology, involved 122 sessions, over 1800 papers from scientist of almost 60 countries from all continents.

The proceedings, edited in this series but also as special issues of specialized scientific journals, will remain as the most important outcome of the congress.

L'UISPP a une longue histoire, à partir de 1865, avec le Congrès International d'Anthropologie et d'Archéologie Préhistorique (C.I.A.A.P.), jusqu'en 1931, date de la Fondation à Berne de l'UISPP. En 1955, l'UISPP est devenu membre du Conseil International de philosophie et de Sciences humaines, associée à l'UNESCO. L'UISPP repose sur plus de trente commissions scientifiques qui représentent un réseau représentatif des spécialistes mondiaux de la préhistoire et de la protohistoire, couvrant toutes les spécialités de l'archéologie : historiographie, théorie et méthodes de l'archéologie ; Culture matérielle par période (Paléolithique, néolithique, âge du bronze, âge du fer) et par continents (Europe, Asie, Afrique, Pacifique, Amérique), paléoenvironnement et paléoclimatologie ; Archéologie dans des environnements spécifiques (montagne, désert, steppes, zone tropicale), archéométrie ; Art et culture ; Technologie et économie ; anthropologie biologique ; archéologie funéraire ; archéologie et sociétés.

Le XVIII° Congrès mondial de l'UISPP en 2018, à Paris en France, avec le soutien de toutes les institutions françaises liées à l'archéologie, comportait 122 sessions, plus de 1800 communications de scientifiques venus de près de 60 pays et de tous les continents.

Les actes du congrès, édités par l'UISPP comme dans des numéros spéciaux de revues scientifiques spécialisées, restera comme le résultat le plus important du Congrès.

Marta Arzarello

Secretary-General /
Secrétaire générale UISPP

Contents

List of Figures	ii
Introduction au volume	iv
François Djindjian, Paola Moscati	
Mégadonnées et archéologie : une introduction	1
François Djindjian	
How Big Is Big Data?	8
Paola Moscati	
Les statistiques et l'analyse spatiale des sites archéologiques sont à notre portée	23
Olivier Buchsenschutz	
Innovative multidisciplinary method using Machine Learning to define human behaviors and environments during the Caune de l'Arago (Tautavel, France) Middle Pleistocene occupations	28
Sophie Grégoire, Nicolas Boulbes, Bernard Quinio, Matthieu Boussard, Caroline Chopinaud, Anne-Marie Moigne, Agnès Testu, Vincenzo Celiberti, Cédric Fontaneil, Christian Perrenoud, Anne-Sophie Lartigot Campin, Thibaud Saos, Tony Chevalier, Véronique Pois, Henry de Lumley, Marie-Antoinette de Lumley, Antoine Harfouche, Rolande Marciniack, Philippe Carrez, Thierry Hervé	
Cagny-l'Épinette (Somme Valley, France), Thirty Years of Mixed Data: Potential and Limits	48
Floriane Peudon, Éric Masson, Patrick Auguste, Agnès Lamotte, Anne-Marie Moigne, Alain Tuffreau	
Towards an Archaeological Information System: the evolution of Syslat, an archaeological data management software	62
Réjane Roure, Hakima Manseri, Sébastien Munos, Michel Py	
L'archéologie néoprocessuelle	71
François Djindjian	
Transcending 'Technocomplexes'. When French Empiricism calls for Hypothetico-deductive Method	83
Pascaline Gaussein	
List of Authors	93

List of Figures

P. Moscati: **How Big Is Big Data?**

Figure 1. J.-C. Gardin's and P. Braffort's paper presented at the UNESCO International Conference on Information Processing (Paris 1959).....	10
Figure 2. The Virtual Museum of Archaeological Computing, section Protagonists: interactive itinerary dedicated to Robert G. Chenhall.....	11
Figure 3. The Virtual Museum of Archaeological Computing, section Events: interactive itinerary dedicated to the European postgraduate course on Data Processing and Mathematics Applied to Archaeology (Valbonne-Montpellier 1983)	13
Figure 4. The Virtual Museum of Archaeological Computing, section Projects: interactive itinerary dedicated to the project 'Automatisation of Etruscan corpora'	14
Figure 5. The Virtual Museum of Archaeological Computing, section Institutions: interactive itinerary dedicated to the Centre de recherches sur les Traitements Automatisés en Archéologie Classique (TAAC)	16
Figure 6. Spatial distributions of the articles published in «Archeologia e Calcolatori» n. 2014-2017 on Periplo search engine.....	20

S. Grégoire *et al.*: **Innovative multidisciplinary method using Machine Learning to define human behaviors and environments during the Caune de l'Arago (Tautavel, France) Middle Pleistocene occupations**

Figure 1. The Caune de l'Arago cave (Tautavel, Pyrénées-Orientales ,France).	30
Figure 2. A conceptual model of research	31
Figure 3. Example of a part of the cognitive map of the occupation duration and site function scenario	31
Figure 4. 2D visualization of high dimensional data with TSNE.....	33
Figure 5. Biome's predictions in the stratigraphic levels of Caune de l'Arago cave.....	34
Figure 6. Variable importance plot for the level O	35
Figure 7. Visualization (force plot) of the prediction's explanation of the majority class from the level J.	36
Figure 8. Result of prediction classifying by short (C, <i>green</i>), not short (NC) output class. The training set has learnt on three expert labels: short, not short, and unknown (?) assigned and predicted on the whole layers (<i>red: concordant predictions</i>).....	40
Figure 9. Feature ranking and feature importance for the prediction of short and not short occupation duration.....	40
Figure 10. Principle of the triple loop learning in AI	41
Table 1. Example of variable selection to work on the question of occupation duration.....	37

F. Peudon *et al.*: **Cagny-l'Épinette (Somme Valley, France), Thirty Years of Mixed Data: Potential and Limits**

Figure 1. 3D-Model of the location of the site of Cagny-l'Épinette, Somme Valley, Northern France	49
Figure 2. Field recording protocol applied at the site of Cagny-l'Épinette from 1980 to 2010	51
Figure 3. Methodological protocol submitted to the archives of the site of Cagny-l'Épinette.....	53
Figure 4. Rule set applied to the digitized georeferenced field drawings with the OBIA-software eCognition.....	54
Figure 5. Overall digitalization of the archaeological vestiges of Level I1 overlapped by a corresponding scatter plot, both with their respective attribute tables linked through the ID Number	55
Figure 6. Integrity assessment of the spatial information of the Level I1 with three identified main degrees of integrity	56

R. Roure *et al.*: **Towards an Archaeological Information System: the evolution of Syslat, an archaeological data management software**

Figure 1. The Syslat software	63
Figure 2.1. Cover of Lattara 4.....	64

Figure 2.2. Cover of Lattara 10.....	65
Figure 3. A mind map of the Syslat architecture	66

F. Djindjian: **L'archéologie néoprocessuelle**

Figure 1. Evolution des procédés de taille des burins dans l'Aurignacien et le Gravettien du site paléolithique supérieur de la Ferrassie (Dordogne).....	75
Figure 2. Evolution des « cultures » du paléolithique supérieur européen depuis les débuts du stade isotopique 3 jusqu'à la fin du stade isotopique 2, illustrée par le modèle de Holling.....	79
Figure 3. Evolution des systèmes d'exploitation des territoires des groupes humains depuis les débuts du stade isotopique 3 jusqu'à la fin du stade isotopique 2, illustrée par le modèle de Holling.....	79
Figure 4. Evolution des procédés de fabrication de l'industrie lithique depuis les débuts du stade isotopique 8 jusqu'à la fin du stade isotopique 2, illustrée par le modèle de Holling.....	80
Figure 5. Processus concernés par la simulation globale d'une société complexe.....	81

P. Gaussein: **Transcending 'Technocomplexes'. When French Empiricism calls for Hypothetico-deductive Method**

Figure 1. Simplified framework for the interpretation of variability, homogeneity and change of material culture styles	86
Figure 2. Contextualized framework for the interpretation of variability, homogeneity and change of material culture styles	86
Figure 3. Schematic mapping of social unit territories depending on the application of human density.....	88

Introduction au volume

Le présent volume contient les articles de communications données dans deux sessions du XVIII^e Congrès mondial de l'UISPP, qui s'est tenu à Paris en juin 2018, organisées par la Commission « Méthodes et théorie de l'archéologie » :

- Session III-1 (CA) : Big data, databases and archaeology
- Session III-1 (T) : New advances in theoretical archaeology

Le volume contient six articles de la première session et deux articles de la deuxième session.

Après une introduction au thème des Big Data en archéologie donnée par François Djindjian, deux articles par Paola Moscati et Olivier Buchsenschutz retracent le contexte historiographique du thème et ses développements actuels. En suivant un laps du temps qui s'écoule des années 1950 à nos jours, le sujet est abordé d'un point de vue terminologique, conceptuel et technique, sous l'angle de la collection et du traitement de l'information et de l'interprétation et de la transmission de la connaissance.

Les articles de Sophie Grégoire (*et al.*) sur le site de la Caune de l'Aragon et de Floriane Peudon (*et al.*) sur le site acheuléen de Cagny l'Épinette, donnent des exemples complets d'application des méthodes informatiques (bases de données, systèmes d'information géographique) et d'algorithmes d'intelligence artificielle par apprentissage. Il en résulte une approche multidisciplinaire et intégrative à l'enregistrement, la gestion et l'exploitation des archives de fouilles et à la classification de la documentation, manuscrite et numérique, produite pendant une longue période durant laquelle l'archéologie a vu sa progressive informatisation.

L'article de Réjane Roure (*et al.*) concerne les plus récents développements du logiciel Syslat qui est un progiciel intégré d'acquisition et de gestion des données archéologiques, initialement développé dans les années 1980 pour le site de Lattes en Languedoc. Créé par Michel Py (CNRS), Syslat est aujourd'hui une suite d'outils complets et gratuits, destinée à l'enregistrement, l'exploitation et l'analyse documentaire des données de fouille et complétée par une application pour mobiles, une application web destinée à la consultation et un outil multiplateforme pour faciliter la saisie des données sur le terrain.

La deuxième session contient deux articles. Le premier, de François Djindjian, présente les caractéristiques générales d'une nouvelle plateforme théorique, l'archéologie néoprocessuelle, qui est une approche centrée sur le rôle des processus dans les principaux domaines de l'archéologie. Le second article, de Pascaline Gaussein, est une réflexion sur le concept de cultures archéologiques et sur les réalités humaines et sociales qui sous-tendent les cartes de répartition des traits culturels paléolithiques, en combinant les approches processuelle, contextuelle et empirique.

François Djindjian et Paola Moscati

Mégadonnées et archéologie : une introduction

François Djindjian

Université de Paris 1 Panthéon Sorbonne et CNRS UMR 7041

francois.djindjian@wanadoo.fr

Résumé

Dans cette introduction au thème des mégadonnées en archéologie, l'accent est mis sur le concept relatif de mégadonnées dans le temps de l'évolution technologique de l'informatique et dans l'histoire récente de l'archéologie computationnelle.

Mots-clés : Mégadonnées, archéologie, informatique

Abstract

In this introduction to the topic of Big Data in archaeology, emphasis is placed on the concept of Big Data in the time of the technological evolution of computing and in the recent history of computational archaeology.

Keywords: Big Data, archaeology, computer

1. Introduction

D'après Wikipedia (article Big Data), le volume des données stockées au niveau mondial est en pleine expansion : les données numériques créées dans le monde seraient passées de 1,2 zettaoctet par an en 2010 à 2,8 zettaoctets en 2012 et s'élèveront à 40 zettaoctets en 2020. À titre d'exemple, Twitter générerait en janvier 2013, 7 téraoctets de données chaque jour et Facebook 10 téraoctets. Ce sont les installations technico-scientifiques (météorologie, astronomie, CERN, etc.) qui produiraient le plus de données. Le radiotélescope « Square Kilometre Array » par exemple produira 50 téraoctets de données analysées par jour, tirées de données brutes produites à un rythme de 7000 téraoctets par seconde.

Le volume des données produit par l'archéologue n'est pas évidemment du même ordre de grandeur. Un chantier de fouilles archéologiques produira, après plus d'une dizaine d'années de campagnes annuelles, des données de l'ordre d'une ou plusieurs centaines de Moctets, compatibles avec les capacités de stockage des ordinateurs de bureau. Ces volumes sont dus pour l'essentiel à la numérisation des relevés et des documents photographiques. Mais, de plus en plus, sont utilisées en archéologie des applications grosses productrices de volume de données comme le 3D, le Lidar ou les analyses de laboratoire.

Par ailleurs, et indépendamment, se pose la question de l'archivage des données, dans un contexte institutionnel où les données archéologiques restent sous la responsabilité directe de l'archéologue, dans son environnement informatique individuel, et dont la sécurité de l'archivage n'est donc ni sûr ni pérenne.

2. Historique du concept de Big Data

A ses origines, l'archéologie a été une science de l'objet et les archéologues, eux-mêmes, se désignaient souvent comme antiquaires ou collectionneurs.

C'est à partir des années 1960, que l'archéologie devient progressivement une science de l'information des sociétés du passé : informations intrinsèques qui décrivent les artefacts de la culture matérielle et informations extrinsèques qui enregistrent le contexte de ces artefacts et leurs relations.

Ces informations étaient disséminées par le support écrit : livres, corpus, monographies, articles dans les revues académiques, consultables dans les bibliothèques institutionnelles et privées.

Les archéologues archivaient leurs documents de travail : carnets de fouilles, relevés stratigraphiques et planigraphiques, dessins d'objets, plans, photographies, inventaires, mensurations, mesures, notes, projets d'articles, tirés à part ainsi que les communications épistolaires entre archéologues dans le milieu académique. Tous ces documents faisaient l'objet au mieux d'un archivage institutionnel ou privé.

Le développement de l'informatique transforma progressivement le support papier en support électronique : systèmes bibliographiques de l'Information scientifique et technique, systèmes documentaires (« banques de données »), fichiers d'inventaires et de mesures.

La machine à écrire (comme la célèbre IBM à boule ou la petite portable), apparue dans la seconde moitié du XIX^e siècle, disparaît à la fin des années 1980 (tout comme le métier de dactylo), remplacée par le microordinateur et le logiciel de traitement de texte.

Le courrier devient messagerie mais le message n'est plus archivé sauf exception. L'historiographie perd ainsi les échanges privés entre chercheurs, souvent plus instructifs que les échanges officiels.

Les tirés à part d'articles publiés, ou leur photocopie, sont remplacés par des fichiers en format pdf. Ils sont échangés ou accessibles sur des sites en ligne en accès libre ou en vente sur les sites des éditeurs privés.

Le dessin, activité manuelle (les laboratoires employaient des ITA dessinateurs), devient dessin assisté par ordinateur (DAO) avec la fameuse suite Adobe : dessin vectoriel (Illustrator), composition (Pagemaker/Indesign), création et retouche d'image (Photoshop) et ses équivalents concurrents.

Puis, à partir des années 1990, la numérisation s'accéléra, qui engendra pour l'archéologie de nombreuses nouvelles données :

- les mesures physico-chimiques,
- la prospection géophysique (terrestre et maritime),
- Les données Lidar,
- la cartographie,
- le Système d'Information Géographique (SIG),
- la photographie numérique,
- la numérisation des photographies argentiques,
- le film numérique,
- la numérisation des relevés stratigraphiques et planigraphiques,
- et enfin le 3D avec la réalité virtuelle et la photogrammétrie numérique.

Dès lors, la question du Big Data s'imposa à tous.

3. Big Data : une longue histoire liée aux progrès de l'informatique

Le concept de « Big Data » est relatif. Il est lié aux problèmes que posent l'archivage et le traitement de grands volumes de données en rapport avec la disponibilité du hardware (stockage des données)

et des outils logiciels pour les rechercher (systèmes documentaires), les consulter, en extraire une partie, les visualiser (systèmes graphiques, SIG, 3D) et les traiter (visualisation graphique, analyse des données multidimensionnelles, modélisation, intelligence artificielle, etc.).

Le monde scientifique moderne aime à ressusciter les problématiques échouées face aux difficultés technologiques du moment par des nouveaux noms désignant les mêmes concepts. L'intelligence artificielle, ce grand mythe des temps modernes, en est un bon exemple : issue de la cybernétique d'avant-guerre, elle est née dans les années 1950 (perceptron de Rosenblatt) avec les premiers ordinateurs, et, elle se relance périodiquement sous différents noms : IA, apprentissage automatique, système expert, réseau de neurones, moteur de règles, et dernier en date, apprentissage profond. Ses applications les plus réussies se retrouvent dans la robotique, la traduction automatique, la reconnaissance de formes, l'aide au diagnostic, l'aide à la décision, le traitement des Big Data (où elle remplace le data mining des années 1990) et la plus médiatisée de toutes, les jeux (quand la machine bat l'humain : échecs, Go).

Le Big Data aussi a déjà une longue histoire. Il est lié à l'évolution de la taille des mémoires des ordinateurs et au volume de stockage des mémoires de masse (disques et bandes magnétiques). Dans les années 1960/70, les mémoires (à tores de ferrites) étaient limitées à plusieurs dizaines ou centaines de kilooctets. Les mémoires RAM actuelles (des circuits imprimés) font plusieurs à plusieurs dizaines de Gigaoctets, soit un million de fois plus ! Le stockage sur mémoire de masse a connu la même évolution technologique depuis les 2 Megaoctets du premier disque dur d'IBM en 1962, les 300 Megaoctets dans les années 1980, les 25 Gigaoctets en 1998 et plusieurs Teraoctets actuellement soit un million de fois plus aussi !

Les bandes magnétiques, organisées en baies de stockage qui peuvent contenir une dizaine ou une vingtaine de bandes magnétiques, peuvent atteindre une capacité totale jusqu'à plusieurs dizaines de téraoctets. Les bibliothèques de bandes sont donc le moyen le plus aisé d'assurer la sauvegarde et l'archivage de données volumineuses, comme pour les grandes fermes informatiques du Web ou le stockage institutionnel des organismes de recherches. Hélas la durée de vie d'une bande magnétique n'est que d'une vingtaine d'années !

Dans les années 1970, les banques de données (systèmes documentaires) et les grands tableaux étaient les « Big Data » de cette période. En France, c'est la grande période institutionnelle des systèmes documentaires mis en œuvre par le ministère de la Culture (musées, Inventaire général des monuments et richesses artistiques de la France, Carte archéologique), utilisant le logiciel Mistral de Bull. Mais les données de cette période sont du texte, les images étant stockées sur microfiche et consultables sur un lecteur installé à côté du terminal. C'est seulement à partir des années 1980, que les progrès technologiques des mémoires, des unités de stockage (disques magnétiques, vidéodisque puis disque optique numérique) et des réseaux ont vus arriver les premiers prototypes de serveur données/images/son qui deviennent opérationnels dans les années 1995, avec le développement d'Internet. Notons cependant que le système vidéotex, le précurseur d'Internet, a été opérationnel en France à partir de 1980 jusqu'en 2012.

Les grands tableaux, qui sont les données de base des archéologues pour la plupart des problématiques traitées (Djindjian 1991, 2011), faisaient l'objet de manipulations graphiques dans les années 1960, avant d'être traités par analyse des données multidimensionnelles dans les années 1975, malgré les limitations des ordinateurs en puissance de calcul et en mémoire centrale. A partir des années 1990, ces limitations ont disparues et ces traitements ont commencé à être effectués sur microordinateur.

Les années 1980 et 1990, qui sont les années du développement du microordinateur, des réseaux et des logiciels bureautique, voient l'archéologue s'approprier individuellement ces outils et l'institution se trouve alors en retrait sur des projets communautaires.

Les années 1990 voient l'arrivée d'un nouveau vocabulaire sinon d'une nouvelle approche, le Data mining (ou exploration de données) qui applique les techniques statistiques multidimensionnelles à de grand corpus de données comme ceux obtenus par les habitudes de consommation, de consultation de données sur Internet ou de questionnaires et qui permettent d'identifier des types de comportements de consommateurs (segmentation, scoring). Les techniques d'apprentissage font également leur apparition. Mais les méthodes de l'archéologie ne se sentent pas concernées par les intérêts essentiellement marketing du data mining.

Les années 2000 voient l'émergence du vocabulaire des Big Data (en français, mégadonnées), liée à la production massive (« orwellienne ») des données que le progrès technologique de l'informatique permet aujourd'hui de stocker, de communiquer par des réseaux, de visualiser et de traiter. Les organismes institutionnelles de la recherche commencent à s'émouvoir de la dispersion des données enregistrées par les chercheurs individuels (mais financés par l'institution) et qui se perdent quand le micro-ordinateur tombe en panne ou quand le chercheur part à la retraite, et tout particulièrement dans le domaine des Sciences humaines et sociales où le chercheur individuel prime sur le laboratoire.

En France, le CNRS lance le projet TGIR Huma-Num (www.huma-num.fr) du CNRS pour l'archivage des données numériques des Sciences humaines. Il s'agit d'une plateforme informatique permettant l'acquisition, le stockage, la dissémination, le traitement et l'archivage des données. Plusieurs laboratoires d'archéologie se sont regroupés au sein du consortium Masa (Mémoire des archéologues et des sites archéologiques) pour utiliser les services du TGIR Huma-Num. Il a pour objectif de proposer un accès unifié à des corpus variés de données et de documentations produites par les archéologues. Il développe des méthodes et des outils à destination de la communauté archéologique, en respectant les standards internationaux (<https://masa.hypotheses.org/>).

Au niveau européen, le projet Ariadne a lancé des coopérations entre les archéologies européennes sur des projets fédérateurs (notamment les thésaurus) et des plateformes de service, dont le sujet de l'archivage (<https://ariadne-infrastructure.eu/>).

4. Quelles mégadonnées archéologiques ?

Les mégadonnées archéologiques sont constituées d'un ensemble non limitatif de fichiers de taille, de format et de structure variable :

- Des bases de données, résultats de l'enregistrement des données de fouilles archéologiques (informations extrinsèques) et de la description des artefacts (informations intrinsèques). Ces données sont enregistrées dans des logiciels variés depuis les logiciels de traitements de texte, les tableurs jusqu'aux systèmes de gestion de bases de données.
- Des textes anciens dans leur écriture d'origine et leur traduction (philologie),
- Des banques de données créées avec des logiciels de recherche documentaire,
- Des documents numérisés : photos numériques, diapositives numérisées, relevés stratigraphiques et planigraphiques numérisés, films vidéo numérisés, 3D,
- Des documents graphiques vectoriels comme ceux créés par des logiciels de PAO ou des logiciels de système d'information géographique (SIG),
- Des tableaux de données quantitatives,
- Des fichiers de mesures comme ceux produits par des appareillages physico-chimiques : prospection géophysique, Lidar, spectrométrie variée, datations, etc.

5. Les fonctions d'un service « mégadonnées »

Les fonctions d'un service de mégadonnées ne sont pas limitées à l'archivage. Elles concernent l'ensemble de la chaîne depuis l'acquisition (*Submission information package*), le stockage, la

signalisation (c'est-à-dire l'indexation ainsi que la définition et la gestion des métadonnées qui décrivent les données), la diffusion (qui permet la consultation par Internet), l'archivage (suivant un format standardisé), la sélection (qui permet d'extraire et de formater les données pour un traitement) et le traitement.

La fonction de traitement est riche et variée et comprend tous les outils et logiciels utilisés depuis plus de cinquante ans : analyse lexicographique, statistiques, analyse des données multidimensionnelles, système d'information géographique, traitement d'image, modélisation, 3D, et plus récemment le retour de l'intelligence artificielle utilisant les techniques d'apprentissage automatique (deep learning) etc.

6. Les bonnes pratiques

Au-delà du plaisir de s'enivrer de mots à la mode, l'archéologue doit s'investir dans le domaine des projets, qui mêlent efficacement nouveautés techniques et pragmatisme. Les bonnes pratiques sont alors le meilleur garant d'un projet réussi.

Les métadonnées, qui sont les données qui décrivent les données regroupent deux ensemble : les métadonnées individuelles liées aux données produites par l'archéologue et les métadonnées communes, institutionnelles, globales et spécialisées, de plus en plus normalisées. Ces métadonnées institutionnelles sont issues en archéologie des projets documentaires des années 1970 (Ministère de la Culture, CNRS, Information scientifique et technique (INIST)) qui ont investi dans la réalisation des premiers grands thésaurus, qui sont les bases des métadonnées actuelles. En archéologie, le thésaurus de référence est Pactols développé à l'origine pour le projet signalétique Frantiq à la Maison de l'Orient Méditerranéen, qui possède 30 000 références (conforme à la norme ISO 25964 des thésaurus multilingues). Les thésaurus du Ministère de la Culture (Inventaire général, banques de données muséographiques) ont été regroupées sur la plate-forme Gincó.

Les normes qui homogénéisent les productions industrielles depuis plus de cinquante ans, concernent également progressivement l'archéologie, soit indirectement par des logiciels génériques, soit directement mais encore rarement par des normes dédiées à l'archéologie.

L'archivage (OAIS, Open Archival Information System) possède son propre standard, l'ISO 14721:2012. Dans cette norme, un « *paquet d'information* » contient les informations à archiver, à conserver ou à communiquer aux utilisateurs. Le paquet d'information contient toujours l'objet que l'on veut conserver, et les métadonnées nécessaires à sa préservation. Trois types sont définis :

- Le paquet d'information à verser (SIP): Produit par le dépositaire de l'archive, selon le modèle imposé par le gestionnaire de dépôt ;
- Le paquet d'information archivé (AIP): Contenus (*Content Data Objects*) et métadonnées. Produit par et pour le gestionnaire de dépôt ;
- Le paquet d'information diffusé (DIP): en fonction des droits de l'utilisateur qui effectue la requête et des droits de diffusion.

La norme CIDOC-CRM (ISO 21127:2014) est une norme propre au patrimoine culturel et à ce titre elle est concernée par le thème des Big Data et de l'archivage.

Les données élémentaires

Le paquet d'information à verser doit contenir les informations au niveau le plus élémentaire connu. Le système d'archivage doit posséder les fonctions de sélection, de filtrage et d'agrégation utiles à construire les données à n'importe quel niveau d'agrégation supérieur. Dans le cas contraire, les informations au niveau le plus élémentaire sont définitivement perdues.

Les données brutes

Le paquet d'information à verser doit contenir des données brutes (raw), avec la définition la meilleure possible, sans format ou traitement visant à diminuer le volume ou à modifier la donnée.

7. Les traitements

Il est illusoire de penser qu'une accumulation de données puisse sous l'action de quelques algorithmes, aussi puissants soient-ils, être capable, spontanément, de fournir des résultats : des connaissances ou des décisions.

L'exploration des données (c'est le nom donné aux différentes méthodes que l'on désignait dans les années 1970 sous le nom d'analyse des données multidimensionnelles) ne peut être efficace que dans le cadre d'une construction formelle qui puisse permettre à la fois de mettre en évidence une structure dans les données et de pouvoir la valider.

C'est sans doute cette trop grande confiance (ou une trop grande paresse) qui est à l'origine du désappointement dans l'utilisation de ces techniques, qui sont en retrait à partir des années 2000 avec le succès du post-modernisme.

L'intégration des techniques d'exploration des données dans un processus cognitif global nécessite une approche à plusieurs niveaux, comme celle que nous avons proposée, sous le titre de « le triplet systémique » (Djindjian 2002) à partir des résultats de notre thèse (Djindjian 1980, note 1). Ces processus, pour être réellement cognitifs, doivent intégrer explicitement des mécanismes d'apprentissage, que l'analyse des données réalise par le jeu des éléments supplémentaires et par l'itération sur les informations intrinsèques permettant une interaction archéologue-objet, mécanisme de l'apprentissage.

L'enjeu de l'intelligence artificielle, à travers les différents algorithmes qu'elle a développés depuis les années 1950, peut en fait se résumer dans le paradoxe suivant : utiliser la puissance de calcul toujours croissante des ordinateurs avec des algorithmes itératifs simples ou mettre en œuvre une construction formelle sophistiquée. L'analogie avec le jeu d'échec illustre bien ce paradoxe : soit calculer toutes les combinaisons possibles soit concevoir une stratégie de jeu qui réduit le calcul des combinaisons. La première option, dont le succès n'est dû qu'à l'amélioration des capacités de calcul de l'ordinateur, n'est qu'une étape préparant la seconde option, d'où le succès du terme et sans doute à terme des résultats du concept d'apprentissage profond, qui doit dépasser le fait d'être le mot à la mode.

8. Conclusions

Au-delà du terme « mégadonnées », se trouve en fait le rapport entre le chercheur scientifique et l'évolution fantastique de la technologie des ordinateurs dans la deuxième moitié du XX^e siècle. Plus cette technologie offre des moyens supplémentaires (capacité de calcul, volume de stockage, canal de communication), plus les besoins apparaissent (souvent plus avec l'aide d'un bon marketing des industriels qu'avec une attente exprimée des chercheurs). L'archéologie a suivi cette tendance avec des besoins certes incomparablement plus faibles, mais le développement de certaines méthodes (comme le 3D) comme la sociologie particulière de l'archéologue implique que l'institution se mobilise pour offrir des environnements, des standards et des services pour les mégadonnées archéologiques.

References

Djindjian, F. 1980. *Constructions de systèmes d'aides à la connaissance en archéologie préhistorique. Structuration et affectation: méthodes et algorithmes*. 2 volumes. Thèse 3eme cycle, Archéologie préhistorique : Paris I.

- Djindjian, F. and Ducasse, H. (eds) 1987. *Mathématiques et informatique appliquées à l'archéologie*. Conseil de l'Europe : PACT 16.
- Djindjian, F. 1991. *Méthodes pour l'Archéologie*. Paris : Armand Colin.
- Djindjian, F. 2002. Pour une théorie générale de la connaissance en archéologie, in *XIV Congrès International UISPP, Liège Septembre 2001. Colloque 1.3. Archeologia e Calcolatori* 13: 101-117.
- Djindjian, F. 2011. *Manuel d'Archéologie*. Paris : Armand Colin.

Note 1

Définition du triplet systémique S (O, I, E): les objets O, l'information intrinsèque I et l'information extrinsèque E.

- Etape 1: Définition du système S :
Le système S est lui-même défini, par un ensemble de valeurs constantes de E, comme par exemple les objets d'une même unité stratigraphique (ensemble clos), d'une même sépulture, les peintures d'une même grotte ornée, les outils d'une même structure d'habitat, les structures urbaines contemporaines d'un même territoire, etc. qui peuvent toutes être définies par un jeu de valeurs constantes d'informations extrinsèques de type T (temps), H (structure d'habitat), R (territoire), L (localisation), M (origine), EV (environnement), EC (économie), etc.
- Etape 2: Perception et description des informations intrinsèques I,
- Etape 3: Enregistrement des informations extrinsèques E,
- Etape 4: Formalisation de la structuration:
Structurer le système formalisé par le tableau Objets x Description des objets (O x I), qui fournit des structures de partitions (classifications ou typologies) ou des structures sérielles (sériations), donnant un nouvel ordre sur O, soit O+, et des corrélations sur I, soit I+. Le système passe alors de l'état cognitif S (O, I) à l'état S+ (O+, I+). Cette structuration est appelée structuration intrinsèque.
Structurer le système formalisé par le tableau d'occurrence (I x E), qui fournit des structures de correspondances entre les deux ensembles d'informations, structuration en faciès chronologiques pour E=T, structuration spatiale pour E=H ou L, déterminisme environnemental pour E = Ev, etc. Le système passe alors d'un état cognitif S (O, I, E) à un état cognitif S+ (O+, I+, E+). Cette structuration est appelée structuration extrinsèque.
- Etape 5: Application des techniques d'analyse des données sur les tableaux (O x I) ou (I x E),
- Etape 6: Rétroactions par retour sur I et E (c'est un mécanisme d'apprentissage),
- Etape 7: Enrichissements progressifs par intégration de nouveaux I et E ,
- Etape 8: Validation (sur un autre système de O, par une autre corrélation E, etc.).

How Big Is Big Data?

Paola Moscati

Istituto di Scienze del Patrimonio Culturale
CNR – Area della Ricerca di Montelibretti (RM)
paola.moscati@cnr.it

Abstract

The advent of Big Data is a recent and debated issue in Digital Archaeology. The term ‘Big Data’ was introduced about 15 years ago, as a consequence of the impact of the Internet’s expansion, which in turn inaugurated the era of Information Age. The shift in scale of data volume in archaeology and the need for new analytical methods and techniques are not a new phenomenon in the history of archaeological computing, which is scattered with periods of information growth coinciding with the introduction of both data dissemination strategies and new technologies. Several main periods can be identified, which correspond to a ten-year evolutionary logic. Theoretical, methodological and technical advances are illustrated through the words of the protagonists, the pages of the Journal «Archeologia e Calcolatori», now celebrating its 30th year of publication, and the information gathered in the Virtual museum of archaeological computing. Today, the role of Big Data methodologies in archaeology should be primarily that of i) combining and processing existing massive cultural datasets, ii) fostering the centralised aggregation of institutional datasets into commonly accessible resources, and iii) providing archaeological syntheses and new research strategies.

Keywords: Digital Archaeology, Databases, Open data, Big Data

Résumé

L’apparition des Big Data en archéologie numérique est un problème récent et débattu. Le terme Big Data a été introduit il y a une quinzaine d’années, en raison de l’expansion d’Internet, qui à son tour a inauguré une nouvelle ère de l’information. Cependant, le changement d’échelle du volume de données en archéologie et la nécessité de nouvelles méthodes et techniques d’analyse ne sont pas un phénomène nouveau dans l’histoire des applications informatiques à l’archéologie, qui est parsemée de périodes de croissance de l’information coïncidant avec l’introduction à la fois de stratégies de diffusion de données et de nouvelles technologies. On peut identifier plusieurs périodes principales, qui correspondent à une logique évolutive décennale. Les progrès théoriques, méthodologiques et techniques sont illustrées à travers les mots des protagonistes, les pages de la revue «Archeologia e Calcolatori», qui fête ses 30 ans de publication, et les informations recueillies au Virtual Museum of archaeological computing. Aujourd’hui, le rôle des mégadonnées en archéologie devrait être principalement celui de 1) combiner et traiter les ensembles de données culturelles massifs existants, 2) favoriser l’agrégation centralisée des ensembles de données institutionnelles en ressources communément accessibles, et 3) fournir des synthèses archéologiques et de nouvelles stratégies de recherche.

Mots-clés: Archéologie numérique, Bases de Données, Données ouvertes, Mégadonnées

1. A terminological introduction

The advent of Big Data is a recent and debated issue in Digital Humanities and, therefore, in Digital Archaeology. The term ‘Big Data’ was introduced in computing about 15 years ago, as a consequence of the impact of the Internet’s expansion, which in turn inaugurated the Information Age. A huge amount of data became then available, whose complexity and volume prevented its processing with traditional data management technologies, thus requiring a completely new range of analytical methods.

Technically speaking, the term ‘Big Data’ referring only to its size is diminishing, especially now when the computer-related connotation of ‘bigness’ has changed over time from megabytes to gigabytes, then evolving to terabytes, petabytes and so on. Following the methodological approach and the scientific framework of humanities research practice, the best definition of Big Data today is the one proposed in 2015 by the National Institute of Standards and Technology (NIST)¹ in which the accent is on four main features: volume, velocity, variety, and/or variability, to which we would willingly add interconnection. All these attributes require specific analytical tools to extract new insights and develop new interpretation methods. Such a process, in turn, requires new architectures for efficient data storage, manipulation, and analysis that traditional methods are no longer able to provide.

Before probing into the technical framework in which the new Big Data paradigm has grown in archaeological research, a 50-year historical retrospective would be helpful to identify its cultural roots. The shift in scale of data volume in archaeology and the resulting need for new analytical methods and techniques are not a new phenomenon. The history of archaeological computing is scattered with periods of information growth, which in general coincide with important stages for both data dissemination and the introduction of new technologies.

Several main periods can be identified, which correspond to a ten-year evolutionary logic. We will try to briefly illustrate these theoretical, methodological and technical advances through the words of the protagonists, who marked the development of contemporary digital archaeology, the pages of the Journal «Archeologia e Calcolatori» (<http://www.archcalc.cnr.it/>), now celebrating its 30th year of publication (Moscatti 2019), and the information we gathered in the ‘Virtual museum of archaeological computing’, an international ongoing exhibition project on the history of archaeological computing, jointly promoted by the National Research Council of Italy and the Accademia Nazionale dei Lincei (<http://archaeologicalcomputing.lincoln.it/>; on this project, see lastly Moscatti, Orlandi 2019).

2. The 1960s: the background for automatic documentation

Since the early Sixties, the interest in new documentation systems spread exponentially worldwide. The need for classification and automatic processing of scientific documentation is linked to two emerging themes in the twentieth century: military security and industrial development (Casolino 2010). In the 1960s, computer science had a strong impact on information policy and scientific documentation. The subject of ‘information retrieval’ soon affected documentation programs promoted by universities as well as by public and private research centres and academies.

Since the 1940s, IBM had played a prominent role in promoting this approach also in the Humanities, first of all, with the pioneering research work carried out by Father Roberto Busa and immediately afterwards in the interdisciplinary inspiring environment of the EURATOM Laboratories at Ispra (Italy). Here, thanks to the personal commitment of Paul Braffort and Jean-Claude Gardin, the foundations were laid for the development of automatic documentation and Information Retrieval Systems in the Humanities in general, and in Archaeology in particular.

«... C’est alors que M. Paul Braffort, Directeur du CETIS à EURATOM, a bien voulu nous demander d’étudier précisément ce problème de synthèse, à travers une expérience de documentation automatique que nous devons préparer et mener à bien en deux ans, dans le domaine le moins spécialisé et le plus hétérogène qui soit : les sciences humaines ... Le SYNTOL – tel est le nom donné au système en question (Syntagmatic Organization Language) – vient seulement d’être mis au point, après de longs essais sur IBM 7090...» (Gardin 1962, 29).

¹ <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf>. For an updated discussion on Big Data in Digital Humanities and Antiquity, see the debate in the Journal «Frontiers in Digital Humanities», and in particular the articles of Kaplan 2015 and Kaplan, di Lenardo 2017.

UNESCO/SS/ICIP/1
PARIS, 12 June 1959
Translated from the French

UNITED NATIONS EDUCATIONAL,
SCIENTIFIC AND CULTURAL ORGANIZATION

International Conference on Information Processing

Symposium on the use of automatic computation techniques in the social sciences, organized under the auspices of the Commission of the French Republic for Education, Science and Culture

16 and 18 June 1959, 9 p.m.

I

DETERMINATION AND MATHEMATICAL DEVELOPMENT OF THE STRUCTURES IN
HUMAN PHENOMENA

1. Anthropological angle
2. Mathematical angle

1. Anthropological angle

by J.C. GARDIN

The aim here will be to demonstrate, with the help of an actual example, the possibility of scientifically investigating certain human phenomena by methods comparable in rigour with those of the physical sciences without, however, resorting to any quantifications of the observations. The method proposed is based, firstly on analytical expression of the phenomena in terms of special descriptive "jargons", and secondly on the mathematical development of the data from these analyses. The procedure whereby the "jargons" are evolved consists in the breakdown of the overall "representations" in the relevant fields of reference into combinations of elements which the research-worker tries to reduce to the lowest possible number by ever more detailed analysis of the phenomena.

The elements are artificial in the sense that they do not necessarily correspond to any direct item of ordinary observation; their value is purely operative, and they have no anthropological significance save on the plane of conceptualization whence they derive. Over the last few years, several systems of formal analysis have been worked out for archaeological documents - for tools, pottery forms, and figured representations. While the descriptive jargons are separate, they are identical in structure: they comprise firstly a vocabulary - i.e. a list of the nominal elements used in the analysis - and secondly a grammar - i.e. a body of procedures, variable from field to field, with which a simple definition of the often highly diversified relations between the nominal elements can be achieved. The "individuality" of each object is then expressed by a combination of the nominal and grammatical factors of the analysis.

WS/069.20

Figure 1. J.-C. Gardin's and P. Braffort's paper presented at the UNESCO International Conference on Information Processing (Paris 1959).
<<https://unesdoc.unesco.org/ark:/48223/pf0000157450?posInSet=3&queryId=5331e227-746e-401d-8db9-d395fb6d21e6>>.

3. The 1970s: the archaeological data banks

In the 1970s, mainframe computers housed the earliest data banks. They soon evolved beyond basic storage systems, in order to provide functions critical to the inventory of national sites and monuments records and to the cataloguing of archaeological artefacts and museum collections. A forceful distinction soon emerged between 'Information Retrieval', whose primary concern is the range of technical problems associated «with the physical transmission of meaningful information, not the content or meaning of the information transmitted», and 'System Analysis', which is vitally concerned «with what is communicated through information retrieval

The screenshot shows the website of the Virtual Museum of Archaeological Computing. The header includes the site name, logos for the Accademia Nazionale dei Lincei and the Consiglio Nazionale delle Ricerche, and navigation links for Home and email. The main banner features a portrait of Robert G. Chenhall, a woven bag, and a grid of small images. Below the banner, the section 'Itineraries - Protagonists' is displayed, with a sub-section 'ROBERT G. CHENHALL AND HIS "MUSEUM IN THE COMPUTER AGE"' by Stephen Fox. The text describes Chenhall's career, his founding of the 'Newsletter of Computer Archaeology' in 1965, and his participation in the 1968 Symposium 'Computers and their Potential Applications in Museums'. A 'Read more' link is provided. On the right, a sidebar contains links to 'A&C REPOSITORY', 'INSIGHTS', 'MULTIMEDIA', and a search icon. At the bottom, there are three smaller images: a book cover 'Celebrating 50 Years of Research', a graphic with the word 'COMPUTER', and a cartoon illustration of archaeological figures.

Figure 2. The Virtual Museum of Archaeological Computing, section Protagonists: interactive itinerary dedicated to Robert G. Chenhall.
<http://archaeologicalcomputing.isma.cnr.it/itineraries/>

techniques, as well as with why this communication takes place and what is accomplished by it» (Chenhall 1968, 77-78). In short, the first involves the technical side of the organisation of records and files, while the second deals with the final aims and objectives, i.e. the systematic modelling of automation processes that allow relations among the stored data to be made explicit.

In 1971 the first Conference on archaeological data banks was organised at the University of Arkansas (Chenhall 1971) and exactly ten years later Sylvia Gaines edited a volume on this specific theme (Gaines 1981). If midway through this decade we contemplate the Conference organised by Gardin and Borillo in Marseille (Borillo, Gardin 1974), we can get a fair picture of

the overall scenario, before it changed dramatically as a result of the introduction of personal computers.

In his introductory article to Gaines' book, Robert G. Chenhall – the founder in 1965 of the «Newsletter of Computer Archaeology» – summarises the reasons of the information explosion in some main points, world-wide in scope, which result from his museum curatorial perspective (Chenhall 1981, 2):

- 1) The technical development of computers both in terms of size and costs;
- 2) The gradual maturing of knowledge about how to effectively use computers in archaeological work;
- 3) The ever increasing demand for information than ever before, which caused an information explosion as great in archaeology as in other aspects of the contemporary society.

In short, technological development and information growth are the prerequisites for expecting machines to store large data files and retrieve information. At the same time – and more importantly – data banks can provide a new theoretical architecture capable of representing and enhancing the logical interrelationships between information that satisfy the needs of archaeological activities.

«If these activities are just a matter of accumulating, storing, and retrieving information, then a computerized data bank is probably not justified. On the other hand, the computer does permit one to accomplish various analyses, particularly with the ongoing, post-project activities that are not possible by any other means».

4. The 1980s: the quantification process

This brief historical excursus cannot do justice to the complexity of the situation, unless mention is made of the special contribution of the Quantitative Movement to the analysis of large archaeological datasets. From the 1950s onwards, Archaeology, as other Human and Social Sciences, witnessed the success of data quantification, which laid the foundations for the application of mathematical methods. However, elementary statistics techniques soon proved to be insufficient to analyse complex phenomena and the relationship between material, social and cultural aspects. This is the reason why the success and spread of Quantitative Archaeology was linked with the revolution in multidimensional data analysis.

A significant excerpt from François Djindjian's presentation during the Colloquium on the History of archaeological computing, held in Rome in 2008, can explain the role and importance of multivariate statistics as part of the archaeological methods:

«During the 1980s, the ability of multidimensional data analysis techniques to solve many archaeological methods was at the origin of its progressive success. Among the numerous classic or prototypal techniques of data analysis, Correspondence Analysis and Principal Component Analysis, associated with an appropriate Cluster Analysis, appeared to be robust and easy to use techniques, even to non-mathematician researchers. Since then, Archaeology has played a major role among all the Human and Social Sciences, in showing how to integrate statistics into archaeological methods. Quantification, statistics, data analysis were then embedded in archaeological methods as they were embedded in computer packages» (Djindjian 2009, 69).

I would also briefly recall what I wrote in the mid-1980s, while experiencing the use of multivariate statistical techniques for the classification of Etruscan mirrors, which offered a global overview of their production and made it possible to outline the artistic trends as well as the spatial and chronological distribution of ancient workshops.

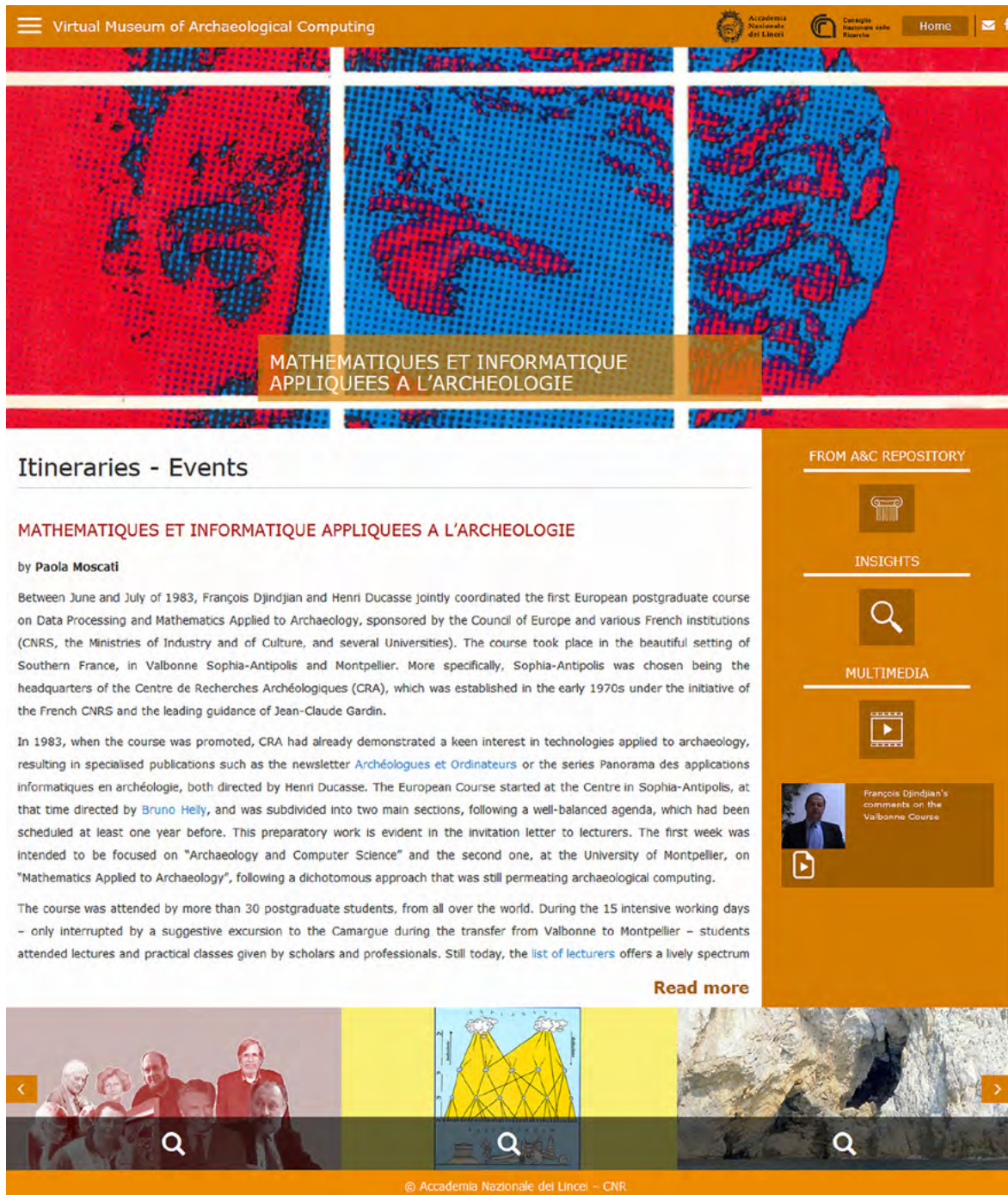


Figure 3. The Virtual Museum of Archaeological Computing, section Events: interactive itinerary dedicated to the European postgraduate course on Data Processing and Mathematics Applied to Archaeology (Valbonne-Montpellier 1983).
<http://archaeologicalcomputing.isma.cnr.it/itineraries/>.

«Questa nuova fase di studio intende sperimentare tecniche statistiche multivariate, le quali permettono, a differenza dei metodi unidimensionali o bidimensionali, di considerare simultaneamente un alto numero di variabili o attributi, siano essi di tipo quantitativo o qualitativo... Esse hanno offerto la possibilità di ottenere una sintesi globale del fenomeno relativo alla produzione degli specchi, delineandone al contempo i diversi filoni storico-artistici, determinati dalla varia dislocazione spazio-temporale delle officine artigianali» (Moscati 1986, 5).

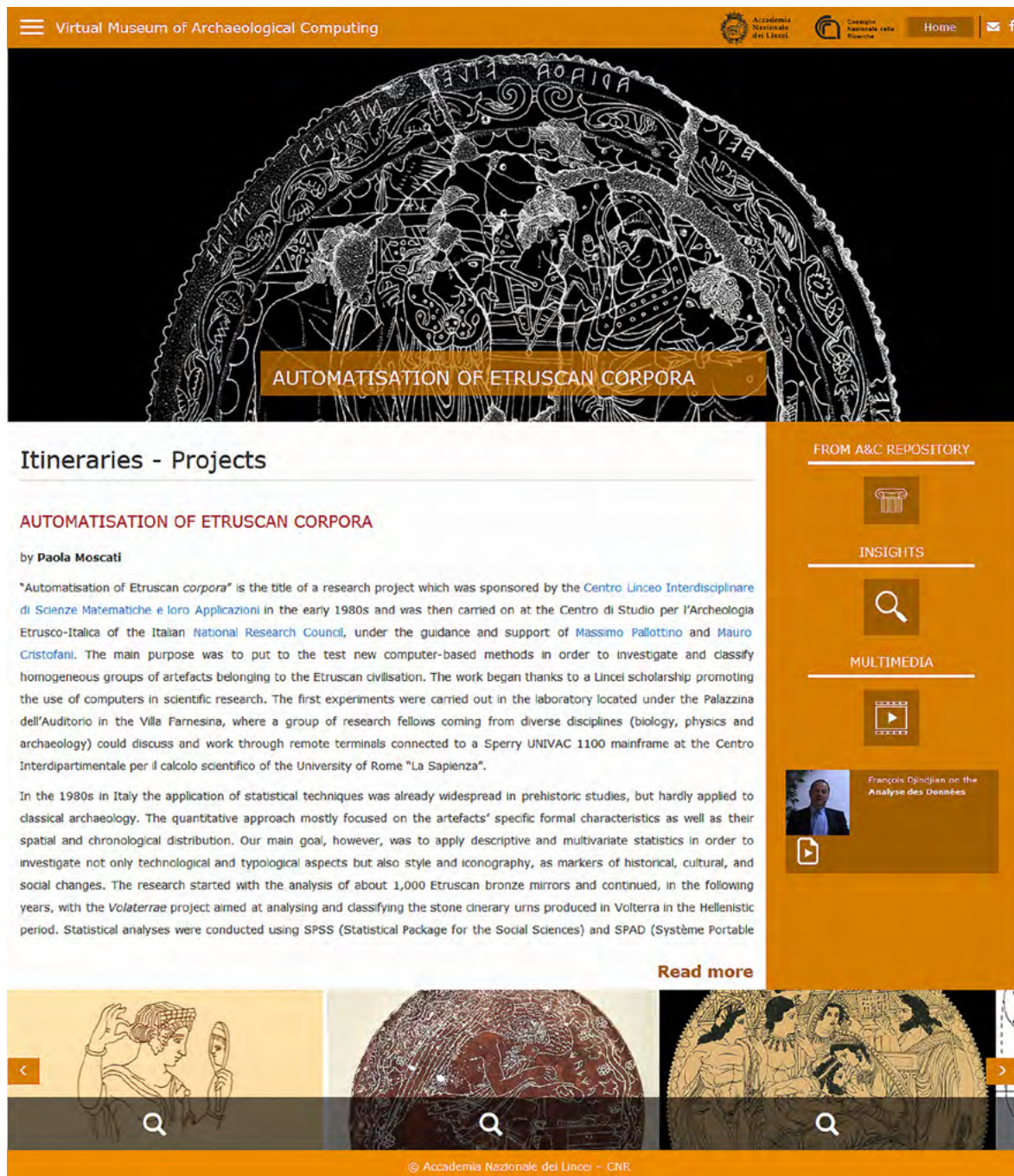


Figure 4. The Virtual Museum of Archaeological Computing, section Projects: interactive itinerary dedicated to the project 'Automatisation of Etruscan corpora'. <<http://archaeologicalcomputing.isma.cnr.it/itineraries/>>.

5. The 1990s: the advent of the Internet

The introduction in the 1980s of videotex in various European countries and the United States, which was based on computer communication on existing networks, played a major role in the evolution of the telecommunications policy, in particular in France, with the successful and lasting experience of the Minitel videotext system. Only ten years later, in the mid-1990s, the advent of the Internet, with its new communication philosophy, dramatically changed data dissemination paradigms and archaeology was faced with the new challenge of the 'Information Age'. The increasing availability of digital research data, recorded during long-lasting excavations or

collected in large databases, their integration into GIS platforms and their representation through 3D modelling techniques provided archaeologists with new insights.

In 1992, Sebastian Rahtz and Paul Reilly edited the book 'Archaeology and the Information Age: A Global Perspective'. The Internet is still not mentioned, but the role and potential of computer networks are already well defined.

«Traditional methods of making archaeological data available are becoming increasingly inadequate. Thanks to improved techniques for examining data from multiple viewpoints, archaeologists are now in a position to record different kinds of data, and to explore that data more fully than ever before. The growing availability of computer networks and other technologies means that communication should become increasingly available to international archaeologists. Will this result in the democratisation of archaeological knowledge on a global basis?» (Rahtz, Reilly 1992, xxi).

At the beginning of the 1990s, classical archaeologists were among the first to foster the wide use of networks. They immediately recognised in telecommunication services a powerful tool capable of reproducing the methodological design of the earliest computerised monumental corpora of archaeological collections. In 1992, in Ravello, John Boardman and Donna Kurtz held a Conference whose proceedings were published the following year in the journal «Archeologia e Calcolatori»: 'Data and Image Processing in Classical Archaeology' (Boardman, Kurtz 1993). In particular, Kurtz underlined that the Beazley Archive Pottery Database was intended to serve the needs of international scholarship and that the University of Oxford was already linked to national and international academic research networks, providing telecommunication services to foreign countries at low cost.

«The database is now in its thirteenth year and available on line for direct interrogation by remote users in continental Europe, North America and Australia.... The database runs on the University of Oxford's mainframe DEC VAX cluster to which the Archive offices in the Ashmolean Museum are linked on an internal network. Fibre optic cables will be added in the autumn of 1992. The university network is linked to national and international academic research networks which provide telecommunication services to foreign countries at low cost. From the outset the Beazley Archive database was intended to serve the needs of international scholarship. It was, therefore, swift to take advantage of international research networks such as the EEC's X.25 service and the North American NSFnet» (Kurtz 1993, 263).

A few years later, in the mid-1990s, the Internet was already gaining ground. In 1995, in Rome, for the first time an entire session of the international Symposium on Computing and Archaeology was entitled 'Networks, Museums and Education, Publications'. The introductory contribution by Anne-Marie Guimier-Sorbets – drawing upon the experience acquired since the late 1960s in developing archaeological information systems at the Centre de recherches sur les Traitements Automatisés en Archéologie Classique, directed by René Ginouvès – highlighted the need for disseminating information not only within academic circles but also towards a wider and more diversified audience (on the same subject, see also Guimier-Sorbets 1993):

«Nous ne pouvons que nous réjouir de cette floraison soudaine de produits d'information multimédia consacrés à la Préhistoire et à l'Antiquité et destinés à un vaste public, mais tachons de contrôler ces productions en participant à leur réalisation, dans la mesure du possible. J'ai exposé à plusieurs reprises l'intérêt qu'ils présentent pour les chercheurs, et il n'est pas utile d'y revenir. Simplement, gardons ces produits de vulgarisation, eux aussi, en synergie avec l'archivage et la publication scientifique» (Guimier-Sorbets 1996, 993).

The Internet started to take on a dominant role also on account of the new perspectives in electronic publication and dissemination of excavation archives. By capitalising on these assets, «Archeologia

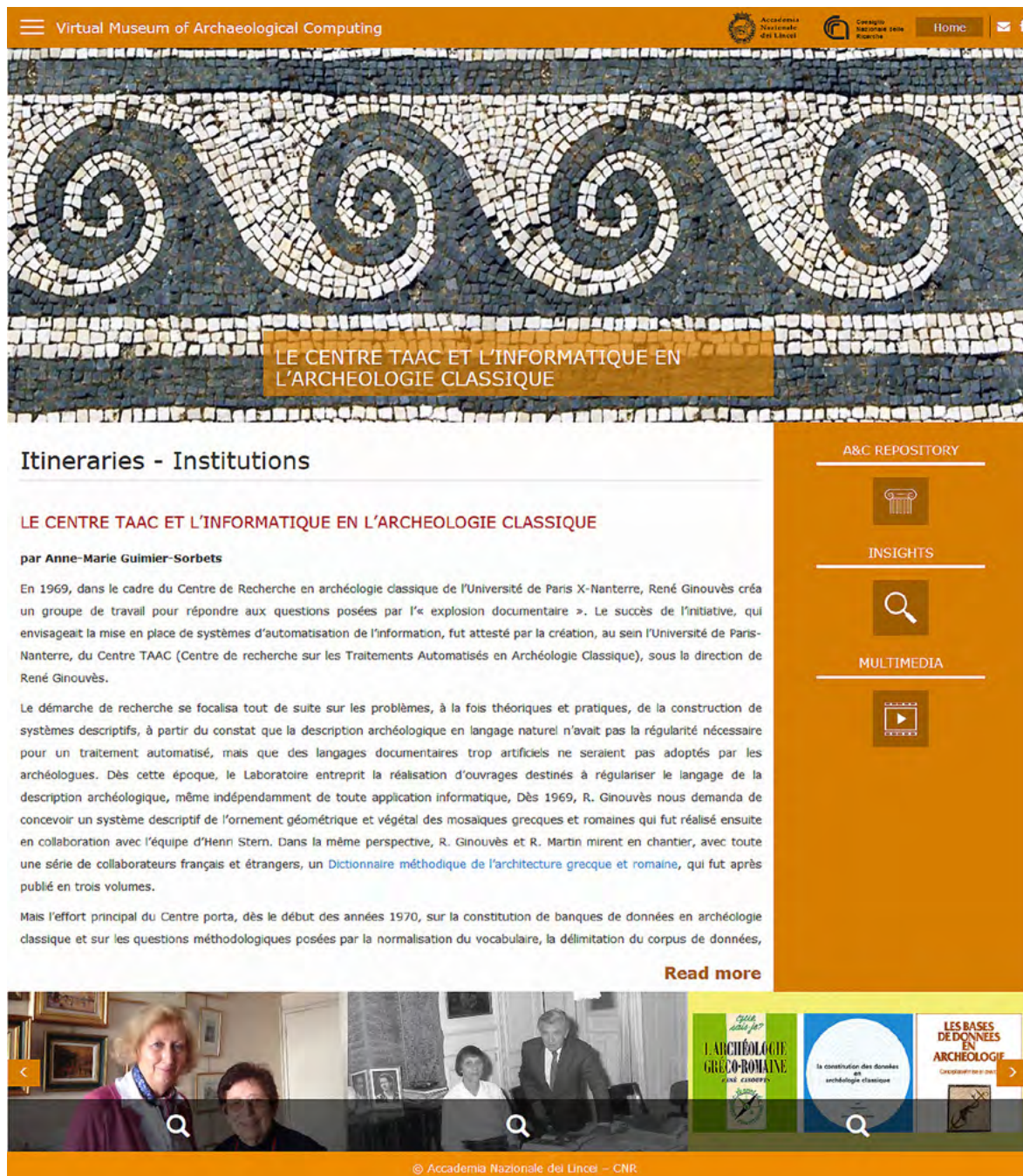


Figure 5. The Virtual Museum of Archaeological Computing, section Institutions: interactive itinerary dedicated to the Centre de recherche sur les Traitements Automatisés en Archéologie Classique (TAAC).
<http://archaeologicalcomputing.isma.cnr.it/itineraries/>.

e Calcolatori», in its seventh year of publication, created the Journal's website and the electronic Journal «Internet Archaeology» was presented to the public for the first time (Heyworth, Richards, Ross 1996), with the purpose of providing:

- A regular electronic journal,
- A detailed description of the process of establishing and managing an electronic journal,
- Definition of a suite of access and navigation tools that will allow the readers to use the journal,
- A contribution to cultural change through the increased use of electronic media.

In 1999, the Proceedings of the 25th CAA anniversary Conference – held in Birmingham two years before – were entitled ‘Archaeology in the Age of the Internet’, and a new page of history was written.

«The organisation and publication of the 1997 conference represents, in a sense, our entry into the ‘Information Age’... It is the first time the conference has been largely prepared, publicised and organized using Internet services such as E-mail and the WWW; it is also the first time that the proceedings have been published digitally on a CD-ROM» (Dingwall *et al.* 1999, i).

6. The new Millennium: archaeological e-infrastructures

This brings us to the new Millennium. The first two decades have been characterised by the open access movement and the explosion of available data that combined to foster the spreading of the Big Data approach, giving birth in archaeology to the so-called ‘data deluge’ question and its disciplinary impact (Bevan 2015). How can this be related to and instrumental in archaeology? In order to reply to this question, it is necessary first to address the issue of archaeological collaborative, distributed and data-intensive e-infrastructures, through which the concept of Big Data has disseminated.

Open digital data ignited numerous remarkable initiatives aimed at gathering existing datasets produced by several research teams, often covering a long time span, as well as ambitious projects to cross-search them within an international framework. Archaeological e-infrastructures, or cyberinfrastructures as they are often referred to in the United States, have today a point of reference in the ARIADNE initiative (the acronym for Advanced Research Infrastructure for Archaeological Dataset Networking in Europe), which was funded under the European Commission’s Seventh Framework Programme. In a recent collective article, all the actors involved in the project agree that:

«In the last 20 years, e-infrastructures have become ever more important for the conduct and progress of research in all branches of scientific enterprise... Moreover there is the expectation that with large datasets (‘big data’), e-infrastructure and advanced computing techniques, new scientific questions can be tackled» (Aloia *et al.* 2017).

But, one might ask, is there today such an archaeological project in which these questions have been raised, tackled and answered? And – as Big Data really requires – do these projects store data using, e.g., NoSQL or NewSQL databases or do they explore large and complex datasets using Big Data analytics processes and a series of techniques including, e.g., data mining? Or, instead, is it possible that in archaeology the use of the term Big Data only refers to ‘large integrated datasets’ being recorded over a long period of time and now stored in ad hoc digital repositories?

On a closer inspection, the expectations of the authors of the above-mentioned collective article correspond to those raised ten years earlier, in a farsighted manner, when the Archeology Data Service launched the Preservation and Management Strategies for Exceptionally Large Data Formats: ‘Big Data’ Project (<http://ads.ahds.ac.uk/project/bigdata/>), following the path opened by the Archaeotools project.

Thenceforth, Big Data has been languishing in the sphere of ‘new technological trends’ and, even in recent years, few academic papers are concretely dedicated to Big Data techniques in archaeology. A good reference sample can be retrieved from scholarly journals, which are the capillaries of scientific research and the most important medium of publication for well conducted and completed research.

For example, in «Archeologia e Calcolatori» – which has often served as a spring-board for future developments in archaeological computing – the theme of Big Data is rarely dealt with. It has

been mentioned after 2010, especially in thematic issues published in the Journal's Supplements. Often Authors focus on the definition of Big Data, on Big Data model and systems architectures and the techniques required to collect and analyse a wide range of data. When they approach the application phase, however, they often use the future tense. Even experts more closely involved with the Big Data revolution, like the members of the Mappa Project (Methodologies Applied to the Predictability of the Archaeological Potential: see e.g. Anichini *et al.* 2013), state that they are planning to give birth to the so-called Big Archaeological Data Project, in order to examine in depth the theoretical aspects of a Big Data approach before moving to Big Data analysis.

«MAPPA Laboratory of the University of Pisa is planning a research project that we decided to call the Big Archaeological Data Project (the BAD Project). This will examine in depth the theoretical aspects of a Big Data approach in archaeological research before moving to Big Data analysis of three different aspects: predictive modelling; the perception of archaeology; and historical/archaeological analysis» (Gattiglia 2015, 120).

The same authors acted as coordinators of the recent ArchAIDE Project that aims at creating a new platform for digital archaeology supporting archaeologists in recognising and classifying pottery sherds during excavation and post-excavation analysis. In 2018, they wrote that perhaps, this will be the right opportunity to move archaeology towards data-driven research and Big Data.

«Datafication fits a Big Data approach and promises to go significantly beyond digitisation. To datafy archaeology would mean to produce a flow of data starting from the data produced by the archaeological practice, for instance, locations, interactions and relations between finds and sites. The ArchAIDE project goes exactly in that direction » (Anichini, Gattiglia 2018, 22).

The situation is not that different in «Internet Archaeology», a Journal equally attentive to new theoretical questions raised by the computer analysis of archaeological data. Two articles are worthy of notice. In 2017, an interesting contribution by A. Cooper and C. Green (2017) evokes and updates Bevan's 'data deluge' overview to illustrate the English Landscape and Identities project (<https://england.wordpress.com/>), carried out between 2011 and 2016 as a concrete example of how significant information can be added to the interpretation of a large and complex dataset, combining archaeological landscape, features and finds.

More recently, a second article was published illustrating the 'Big Data on the Roman Table' (BDRT) network engagement in developing interdisciplinary, quantitative and qualitative approaches to material-cultural evidence (Allison, Pitts, Colley 2018). The authors state that the improvement of information standards in Roman pottery studies will surely have implications for future developments of the Big Data approach to investigate into such material signatures of socio-spatial behaviour in a multi-cultural world.

7. Are there Big Data in archaeology?

We could list other examples aplenty. Certainly, we can claim that numerous research teams worldwide, after integrating data into online open access databases and GIS, are now revealing the large volume of archaeological digital data generated in past and recent years. Databases can include primary finds and context data from excavations and topographic surveys, as well as files supporting secondary data analysis and synthesis; textual data and electronic bibliographies; archaeometric large datasets; geophysical and geographical data, also coming from remote sensing and underwater surveys, 3D laser scanning, DEM and visualization techniques.

As recently stated by Jeremy Huggett and Davide Tanasi in the Journal «Open Archaeology», it is now time to develop a series of 'grand challenges' which seek to generate novel approaches and methodologies in digital archaeology.

«In a digital environment increasingly characterised in terms of ‘big data’, cloud processing, crowd sourcing, social media, and intelligent computing, it is time to develop a series of ‘grand challenges’ which seek to generate genuinely novel approaches and methodologies in digital archaeology» (Huggett, Tanasi 2015).

Furthermore, recent Conferences dedicated to the Big Data approach in archaeology raise new questions. The 2017 Frison Institute Symposium on The Future of ‘Big Data’ in Archaeology, was organised, as part the Society for American Archaeology Meetings, in order to point out the need not only for new analytical tools but also for a comparative critical assessment of the nascent turn toward Big Data approaches at an international level. Likewise, the 2019 Conference on Big Data in Archaeology: Practicalities and Possibilities, organised by the McDonald Institute for Archaeological Research, Cambridge University, was conceived with the aim of demystifying the data science approach in the social sciences and critically examining concepts and methods associated with Big Data and data-intensive research approaches.

«What are the implications for archaeology of the interrelated concepts of datafication (an emphasis on quantification and automated data generation), dataism (a belief in the accuracy, completeness, and reliability of data), and data-centrism (an epistemological switch from theory-driven to data-driven research)? Big Data can be enthusiastically embraced for the transformative opportunities offered or seen as a socio-technical imaginary predicated on a belief in the pre-eminence of large datasets. Where does archaeology position itself?» (Huggett 2019).

8. Concluding remarks

To draw up some final considerations, I would like to highlight the rationale of our approach to open access data repositories that originated from the formation and analysis of the dataset being collected over thirty years since the scholarly Journal «Archeologia e Calcolatori» was launched (see lastly Moscati 2018, 2019). To date, textual resources exceed 15,000 pages and are described thanks to adequate standards and metadata. Moreover, the OAI-PMH protocol and the use of appropriate metadata mapping procedures have allowed us to integrate our dataset in Digital Heritage portals, such as Cultura Italia, in infrastructures for Linked Open Geodata, such as Pelagios, in international e-infrastructures supporting Open Science, such as OpenAIRE, and in institutional archives and repositories of open-access literature, such as the CNR SOLAR database (Piergrossi, Rossi 2019).

The textual dataset, although comparatively small in size, is currently being assessed through exploratory data analysis techniques and Social Network Analysis (SNA) and already indicates some basic guidelines to achieve a better understanding of the evolution of digital archaeology (Caravale, Cantone 2019). This approach fits well with the appropriate and punctual remarks of Julian Richards who maintains that archaeological data may not yet meet the criteria for Big Data, but that nonetheless, the growth of archaeological cyber-infrastructures is providing the foundations for Big Data research.

«The growth of archaeological cyber-infrastructures is providing the foundations for ‘big data’ research. Using digital repositories such as the ADS in the UK and tDAR in the USA, we have access to millions of records, from multiple resources. Data and text mining tools allow us to extract information from published and unpublished fieldwork reports, whilst the ability to create Linked Open Data or to integrate metadata via portals such as ARIADNE in Europe and DINAA in North America should allow us to break free from individual data silos. Nonetheless, the idea that ‘big data’ somehow carries an aura of truth, objectivity, and accuracy through size alone has been identified as a myth» (Richards 2017).

The Big Data approach in archaeology is an emerging field, which is based on the growing importance of massive and networked cultural datasets. It is becoming a well-structured field with

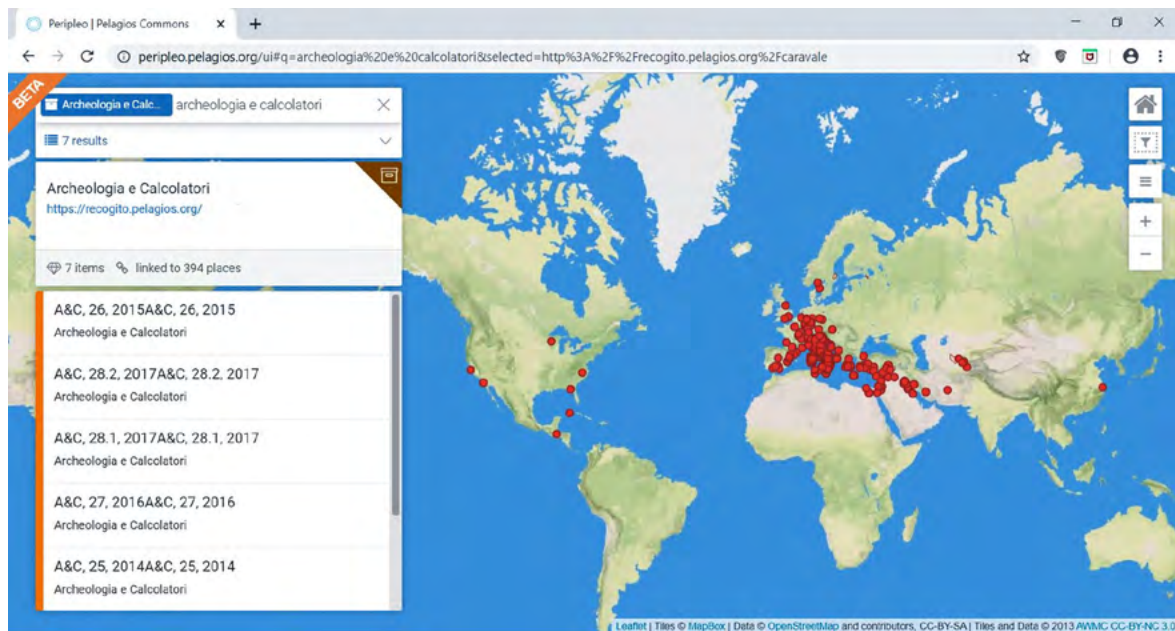


Figure 6. Spatial distributions of the articles published in
 «Archeologia e Calcolatori» n. 2014-2017 on Peripleo search engine.
<https://peripleo.pelagios.org/>.

specific objectives, but it is not the next scientific breakthrough in digital Humanities, as some wish to describe it. What we can say with certainty is that large databases and international digital repositories allow us today to extract information, to create Linked Open Data, or to map and integrate metadata via portals and e-infrastructures.

Therefore, following a solid epistemological approach, the role of Big Data methodologies in the Humanities, and therefore in Archaeology, should be primarily that of i) combining and processing already existing massive cultural datasets, ii) setting out re-documentation processes to foster the centralised aggregation of institutional datasets into commonly accessible resources, and iii) providing, at the same time, archaeological syntheses and new research strategies. This approach should prevent the risk of creating a gap between analytical research – that is increasingly fine-tuning its own data – and data sharing and dissemination – that is widening the network of knowledge.

References

- Allison, P.M., Pitts, M. and Colley, S. (eds) 2018. Big Data on the Roman Table: New approaches to tablewares in the Roman world. *Internet Archaeology*, Special volume 50. <<http://intarch.ac.uk/journal/issue50/index.html>>.
- Aloia, N. et al. 2017. Enabling European Archaeological Research: the ARIADNE e-Infrastructure. *Internet Archaeology* 43. <<https://doi.org/10.11141/ia.43.11>>.
- Anichini, F., Gattiglia, G., Gualandi, M.L. and Noti, V. 2013. MOD (MAPPA Open Data). Conservare, disseminare, collaborare: un archivio open data per l'archeologia italiana, in M. Serlorenzi (ed.) *ARCHEOFOSS Free, Libre and Open Source Software e Open Format nei processi di ricerca archeologica. Atti del VII Workshop (Roma, 11-13 giugno 2012)*. *Archeologia e Calcolatori*, Suppl. 4: 45-52. <http://www.archcalc.cnr.it/indice/Suppl_4/05_Anichini_et_al.pdf>.
- Anichini, F., Gattiglia, G. 2018. Big Archaeological Data. The ArchAIDE Project Approach, in *Conferenza GARR 2017. 'The data way to Science' (Venezia, 15-17 novembre 2017)*. <doi: 10.26314/GARR-Conf17-proceedings-03>.
- Bevan, A. 2015. The data deluge. *Antiquity* 89 (348): 1473-1484. <doi:10.15184/aqy.2015.102>.
- Boardman, J., Kurtz, D. (eds) 1993. *International Conference on Data and Image Processing in Classical Archaeology (Ravello, 3-4 April 1992)*. *Archeologia e Calcolatori* 4: 217-335.

- Borillo, M. and Gardin, J.-C. (eds) 1974. *Banques de données archéologiques. Actes du Colloque (Marseille, 12-14 juin 1972)*. Paris: CNRS.
- Caravale, A. and Cantone, F. 2019. «Archeologia e Calcolatori». Classificazione geografica e tematica per la condivisione della conoscenza, in Moscati 2019: 93-107.
- Casolino, E. 2010. Informazione scientifica, politiche di governi e di mercati, in *Sinergie invisibili, ricerca e informazione scientifica nell'economia della conoscenza*: 157-232. Roma: CNR.
- Chenhall, R.G. 1968. The Analysis of Museum Systems, in *Computers and their Potential Applications in Museums: A Conference Sponsored by the Metropolitan Museum of Art (New York April 15-17, 1968)*: 59-79. New York: Arno Press.
- Chenhall, R.G. 1971. The archaeological data bank conference. *Newsletter of Computer Archaeology* 6(4): 1-2.
- Cooper, A. and Green, C. 2017. Big Questions for Large, Complex Datasets: Approaching time and space using composite object assemblages. *Internet Archaeology* 45. <<https://doi.org/10.11141/ia.45.1>>.
- Djindjian, F. 2009. The Golden Years for Mathematics and Computers in Archaeology (1965-1985), in P. Moscati (ed.) *La nascita dell'informatica archeologica. Atti del convegno internazionale (Roma, Accademia Nazionale dei Lincei, 24 ottobre 2008)*. *Archeologia e Calcolatori* 20: 61-73. <http://www.archcalc.cnr.it/indice/PDF20/6_Djindjian.pdf>.
- Dingwall, L. et al. (eds) 1999. *Archaeology in the Age of the Internet*. CAA 97: Computer Applications and Quantitative Methods in Archaeology. *Proceedings of the 25th Anniversary Conference (Birmingham, April 1997)*. Oxford: Archaeopress.
- Gaines, S. (ed.) 1981. *Data Bank Applications in Archaeology*. Tucson: The University of Arizona Press.
- Gardin, J.-C. 1962. De l'archéologie à l'information automatique. *Bulletin Euratom* 4: 25-29.
- Gattiglia, G. 2015. Think big about data: Archaeology and the Big Data challenge. *Archäologische Informationen* 38: 113-124.
- Guimier-Sorbets, A.-M. 1993. Ouvrir à un large public l'accès à une information spécialisée. *Archeologia e Calcolatori* 4: 281-286.
- Guimier-Sorbets, A.-M. 1996. Le traitement de l'information en archéologie: archivage, publication et diffusion, in P. Moscati (ed.) *III Convegno Internazionale di Archeologia e Informatica - III International Symposium on Computing and Archaeology (Roma, 22-25 novembre 1995)*. *Archeologia e Calcolatori* 7: 985-995. <http://www.archcalc.cnr.it/indice/PDF7/84_Guimier-Sorbets.pdf>.
- Heyworth, M., Richards, J.D., Ross, S. and Vince, A. 1996. Internet Archaeology: An international Electronic Journal for Archaeology, in P. Moscati (ed.) *III Convegno Internazionale di Archeologia e Informatica - III International Symposium on Computing and Archaeology (Roma, 22-25 novembre 1995)*. *Archeologia e Calcolatori* 7: 1195-1206. <http://www.archcalc.cnr.it/indice/PDF7/103_Heyworth_et_al.pdf>.
- Huggett, J. 2019. Datafication, Dataism, Data-centrism: Examining Big Data in Archaeology, in *Big Data in Archaeology: Practicalities and Possibilities (Cambridge, March 27-28, 2019)*. Conference abstract. <https://docs.wixstatic.com/ugd/f855ce_dbf4556d21d247e0a2b481d46c519330.pdf>.
- Huggett, J. and Tanasi, D. (eds) 2015. Open Archaeology: Topical Issue on Challenging Digital Archaeology. *Open Archaeology* 1. <<https://www.degruyter.com/dg/page/topical-issue-archaeology-digital/open-archaeology-topical-issue-on-challenging-digital-archaeology>>.
- Kaplan, F. 2015. A Map for Big Data Research in Digital Humanities. *Frontiers in Digital Humanities* 2. <<https://doi.org/10.3389/fdigh.2015.00001>>.
- Kaplan, F. and di Lenardo, I. 2017. Big Data of the Past. *Frontiers in Digital Humanities* 4. <<https://doi.org/10.3389/fdigh.2017.00012>>.
- Kurtz, D.C. 1993. The Beazley Archive Database, in J. Boardman and D. Kurtz (eds) *International Conference on Data and Image Processing in Classical Archaeology (Ravello 3-4 April 1992)*. *Archeologia e Calcolatori* 4: 263-264.
- Moscati, P. 1986. *Analisi statistiche multivariate sugli specchi etruschi*. Contributi del Centro Linceo Interdisciplinare di Scienze matematiche e loro applicazioni 74. Roma: Accademia Nazionale dei Lincei. <http://archaeologicalcomputing.lincei.it/attachment/74_1986.pdf>.
- Moscati, P. 2018. Le rôle de l'open access dans la diffusion des résultats de la recherche scientifique: le cas de «Archeologia e Calcolatori». *UISPP Journal* 1: 27-41.

- Moscato, P. (ed.) 2019. *30 anni di «Archeologia e Calcolatori». Tra memoria e progettualità. Archeologia e Calcolatori* 30: 1-138.
- Moscato, P. and Orlandi, T. (eds) 2019. *Il Museo Virtuale dell'Informatica Archeologica. Una collaborazione tra l'Accademia Nazionale dei Lincei e il Consiglio Nazionale delle Ricerche. Atti della Segnatura (Roma 2017). Rendiconti dell'Accademia Nazionale dei Lincei* s. 9: 1-120.
- Piergrosso, A. and Rossi, I. 2019. «Archeologia e Calcolatori». *Accessibilità e diffusione della cultura scientifica*, in Moscati 2019: 75-92.
- Rahtz, S. and Reilly, P. (eds) 1992. *Archaeology and the Information Age. A Global Perspective*. London: Routledge.
- Richards, J. 2017. Size isn't everything: are our data good enough to be big?, in *The 81st Annual Meeting of the Society for American Archaeology (Vancouver, BC, March 29-April 2, 2017)*. Abstract (tDAR id: 429440).

Les statistiques et l'analyse spatiale des sites archéologiques sont à notre portée

Olivier Buchsenschutz

AOROC, UMR 8546 CNRS, Paris

olivier.buchsenschutz@ens.fr

Résumé

Aujourd'hui la connexion des serveurs et la constitution de « *big data* » ne pose plus de problèmes techniques. Il est urgent de connecter les données existantes, les résultats de l'archéologie de sauvetage et de l'archéologie programmée, des thèses et des informations qui papillonnent sur internet. Il ne manque plus qu'une décision politique, une distribution des tâches, et un financement pour mettre à la disposition de tous un outil sans lequel nos travaux reposent sur de petits nuages. Notre expérience des inventaires archéologiques informatisés et surtout spatialisés depuis 40 ans, de l'échelle de la fouille à celle de la France, montre qu'il faut réaliser des bases générales simples, dégagées des hésitations des spécialistes comme des préoccupations administratives, pour faire des atlas, des répertoires, des outils un peu ennuyeux mais utiles à tous.

Mots-clés : Analyse spatiale, archéologie nationale, archivage scientifique

Abstract

Today the connection of servers and the creation of 'big data' no longer poses technical problems, it is urgent to connect existing data, the results of rescue and programmed archeology, theses and information that flares on the internet. All that's missing is a political decision, a distribution of tasks, and funding to make available to everyone a tool without which our work rests on small clouds. Our experience of computerized and especially spatial archaeological inventories for 40 years, from the scale of the excavation to that of France, shows that it is necessary to carry out simple general bases, freed from the hesitations of specialists as from administrative concerns, to make atlases, directories, somewhat boring but useful tools for everyone.

Keywords: Spatial archaeology, national archeology, scientific archiving

Le thème des « *big data* », un nouveau nom pour les grands fichiers de référence, m'a remis en mémoire l'évolution des bases de données archéologiques spatiales sur lesquelles j'ai eu l'occasion de travailler un peu avant et depuis le colloque de l'UISPP de Nice en 1976. Elles ont précédé, suivi ou subi l'évolution des machines et des programmations informatiques, la numérisation des données cartographiques, et l'augmentation exponentielle en France du nombre de données. Nous avons essayé, pour rédiger cette année un chapitre dans « La protohistoire de la France » sur les habitats entre le IV^e et le I^{er} s. BC de réunir une documentation systématique sur les rapports de fouille concernant l'âge du Fer en France depuis 1976 : malgré la bonne volonté de nos collègues du Ministère de la Culture et de l'INRAP, il reste difficile actuellement de construire une statistique valable, voire un simple état des connaissances, à partir des fichiers disponibles dans ces institutions. C'est pour attirer l'attention sur ce problème que je voudrais rapidement rappeler l'historique des analyses spatiales en archéologie essentiellement – on voudra bien m'en excuser – à partir de mon expérience, sans ignorer bien sûr que de nombreux autres projets ont été développés en France ou ailleurs.

1. Un projet de carte archéologique de la France

Cette question est ouverte depuis les années 1970, quand nous avons proposé de construire une base de données sur les sites archéologiques de la France pour obtenir des statistiques et dessiner

des cartes de répartition. Le projet développé à l'Université de Paris 1 et au CNRS avait été repris au Ministère de la Culture, où rapidement le volet administratif l'a emporté sur l'exploitation scientifique des données. Je me rappelle très bien qu'au colloque de l'UISPP de Nice en 1976, un journaliste qui interrogeait plusieurs chercheurs de la section consacrée aux bases de données nous avait fait prendre conscience de la mainmise des administrations sur les projets que nous avions élaborés à l'Université, dans une phase où l'intervention des chercheurs pour construire des bases efficaces était indispensable. J.-C. Gardin avait à juste titre attiré l'attention, en introduction à un colloque sur les bases de données en 1972 à Marseille, sur l'importance du volet institutionnel dans la constitution et surtout la pérennisation de ces bases, dont l'intérêt reste anecdotique si elles restent l'œuvre éphémère d'un petit groupe d'individus (Borillo et Gardin 1974, p. 25 ; Buchsenschutz *et al.* 1974b). Les administrations ont développé à cette époque des bases parce que l'informatisation était à la mode dans la foulée des travaux de l'Inventaire Général des Richesses Artistiques de la France. Les bases archéologiques ont vu leur volet scientifique rapidement s'étioler, au point même qu'on a essayé à un moment d'emprunter à l'Inventaire Général ses thésaurus, qui répondaient très mal aux problématiques archéologiques. Les bases archéologiques se sont contentées de gérer le volet administratif de l'information, et de produire des statistiques sur l'évolution et la nature des opérations par région. Le développement des fouilles de sauvetage a monopolisé les efforts de tous, imposé ses contraintes juridiques et son rythme de travail. La réalisation rétrospective de l'inventaire des sites fouillés et la mise à jour rapide sinon des publications ou des rapports complets, au moins d'une fiche signalétique, a été périodiquement remise sous la pile des dossiers à traiter.

Il a fallu attendre les années 1980 et le développement des micro-ordinateurs pour que les travaux statistiques des équipes de chercheurs reprennent, indépendamment cette fois des services informatiques des différents ministères concernés. On a vu fleurir de nombreuses bases de données spécialisées le plus souvent sur des objets ou des décors architecturaux (mosaïques, fibules, outillage), avec l'expérience des descriptions fines et structurées développées auparavant par J.-C. Gardin et son laboratoire. C'est dans cet esprit aussi que nous avons développé le « Code d'analyse pour une carte archéologique de la France » (Buchsenschutz *et al.* 1974 a) avec l'aide du CADA de Marseille. La syntaxe d'interrogation dans le logiciel « Satin » était souple et très fine, trop fine pour la pauvreté des données tirées des inventaires de sites et pour les questions posées par les archéologues.

Les questions n'étaient plus « *existe-t-il un site sur une pente ou un sommet avec un mur en pierre et un sol en brique contenant une sépulture à incinération etc. ?* », mais plutôt « *où sont tous les tumulus du Ve s. BC dans le département du Cher* ». Nous avons réalisé qu'il fallait simplifier le code et l'analyse linguistique et consacrer nos forces à développer le volet de l'analyse spatiale, mettre en réserve pour des bases plus spécialisées les travaux de J.-C. Gardin et développer les méthodes que nous avait enseignées J. Bertin (Bertin 1967) en traitement graphique de l'information.

2. Les cartes de répartition de l'échelle de la fouille à celle d'un pays

Les cartes de répartition, à l'échelle de la fouille, d'un département ou du pays tout entier, n'intéressaient pas du tout les informaticiens des années 1970, et les programmes et les machines disponibles à l'époque éditaient péniblement des nuages de points sur des tableaux à double entrée qui étaient supposés être des cartes. C'est sur un miniordinateur de l'Université de Paris I que nous avons édité des plans de répartition du mobilier sur les fouilles de Levroux grâce à Xavier Debanne (Buchsenschutz, Debanne 1978). Puis en 1980 G.-N. Lambert et M. Cartereau programmaient « Centrar » qui gérait les données textuelles, les inventaires et statistiques de fouille, et des nuages de points pour les cartes de répartition à n'importe quelle échelle (Buchsenschutz, Lambert, Cartereau 1985). Il faut croire que le traitement de l'espace était dans notre équipe une véritable obsession, puisque nous avons développé à partir de 1989 avec H. Murgalé, J.-M. Secondo et K. Gruel un système intégré de relevé de fouille et de traitement

spatial des données qui était déjà, en fait, un « Système d'Information Géographique » (Gruel, Buchsenschutz, Murgalé, Secondo 1992).

Le développement des SIG a remis en selle l'analyse spatiale et élargi son audience chez les archéologues. Les projets se multiplient à partir des années 2000, dans une vision presque toujours diachronique, à Tours à l'échelle du département, dans le projet Archéomedès sur des fenêtres de prospection sélectionnées (Archeomedès 1998). Nous l'avons utilisé à l'échelle d'une province, le Berry, pour produire un atlas suivi de plusieurs thèses et de la publication d'analyses spatiales thématiques, avec le concours de Christophe Batardy (Batardy *et al.* 2001 ; Gandini et Laüt 2013). Après une phase où chacun essaye d'appliquer sur un territoire les outils classiques des SIG – les pentes, les distances aux cours d'eau, les espaces visibles depuis un site donné etc. – et de profiter en parallèle de la numérisation des cartes de base – géologie, pédologie, végétation etc. – la tendance générale a été d'utiliser des outils spatiaux appliqués à des données statistiques sur les comportements humains. Les modèles ont été empruntés à toutes les disciplines, jusqu'aux programmes utilisés par les enquêtes commerciales. O. Nakoinz a développé dès cette période tout un panel de méthodes pour traiter statistiquement et spatialement une information archéologique préalablement filtrée pour n'en conserver que les éléments statistiquement valables (Nakoinz 2005, 2013). Si les SIG n'ont pas encore vraiment intégré les règles de la « *Sémiologie graphique* » de Jacques Bertin, l'idée que les représentations graphiques et cartographiques sont liées et qu'elles peuvent avoir valeur de preuve au même titre qu'une équation finit par faire son chemin (Bertin 1967).

3. De l'intérêt d'une base de données archéologiques générale

Les performances et la rapidité du développement de ces outils a ainsi suscité à juste titre l'enthousiasme des chercheurs et nourri de nombreuses thèses et publications. Ces travaux en général sont fondés sur des données fiables et significatives sélectionnées pour soutenir une théorie ou une hypothèse historique. L'idée d'appliquer ces méthodes à des inventaires systématiques des données archéologiques existantes sur une région ou une période est moins répandue. La qualité des données est en effet très inégale, leur quantité est de plus en plus importante, les obstacles d'ordre administratif nombreux. Mais les archéologues, particulièrement en France depuis quelque cinquante ans, sont responsables vis-à-vis des générations futures des milliers de sites qu'ils ont fouillés et donc détruits. S'ils ne peuvent pas tous faire l'objet d'une publication approfondie, au moins méritent-ils au-delà du rapport rédigé et déposé « à chaud », une interprétation explicite, scientifique, et une mise en inventaire. C'est là que les travaux d'analyse spatiale archéologiques rejoignent le domaine des « big data ». En effet la taille des fichiers, leur stockage, leur pérennité, exige un équipement et une coordination la plus large possible qui est au-delà des possibilités d'un laboratoire ou d'une institution qui n'a qu'accès qu'à une partie des données. Le travail de documentation n'est pas la mission prioritaire d'un chercheur, mais dans le domaine particulier de l'archéologie, il ne peut pas laisser de côté sa contribution à la conservation des données sans lequel son travail sera rapidement obsolète car les preuves, les données qui le fondent, auraient disparu. Le ministère de la Culture est avant tout chargé de la gestion et de la conservation des données archéologiques sur le territoire national, mais si leur exploitation scientifique n'est pas concrètement possible, à quoi bon conserver des millions de fragments de métal ou de céramique, dont les musées par ailleurs sont dans l'incapacité matérielle d'assurer la conservation, et encore moins l'étude ?

Nous avons fait l'expérience dans notre laboratoire (Aoroc, Cnrs, ENS, Paris) de développer pour l'âge du Fer une base sur l'hexagone. Il s'agit d'une base générale qui a pour but principal de construire des cartes de répartition des gisements (Buchsenschutz *et al.* 2015). La *Carte archéologique de la Gaule* du CNRS a été dépouillée systématiquement, ainsi que les thèses ou synthèses qui s'appuyaient sur une vaste espace géographique. La localisation des sites est au géocentre de la commune, d'après un fichier qui a été adapté d'une base européenne, une précision suffisante pour des cartes dont

l'échelle sera du 1/500.000^e au 5.000.000^e. La *BaseFer* a pour objectif de visualiser rapidement l'état des connaissances pour l'âge du Fer sur le territoire métropolitain. Elle est destinée à établir des cartes de répartition après un traitement statistique des données. Elle se limite actuellement à des gisements publiés et interprétés. Une fiche correspond à un seul site ayant la même fonction dans une occupation continue. Les données restent essentiellement qualitatives : la surface des habitats ou le nombre d'objets peuvent être signalés en commentaire, mais ils ne feront pas l'objet de calculs systématiques. De même la stratigraphie fine des sépultures ou les structures d'habitat ne sont pas prises en compte. Les quelque 250 critères descriptifs ont été retenus lorsqu'on estimait qu'ils seraient représentés dans la base par 50 à 500 individus pour que leur répartition soit significative (ex. « fibule », « amphore », « ossement animal »). Elle vise à prendre en charge et à faciliter les comparaisons entre elles. La saisie des fiches sur Filemaker © comme la gestion de la base en Mysql © facilitent les opérations répétitives et permettent de nombreux comptages et des recherches croisées. Un identifiant de gisement permet de relier cette base à des bases plus spécialisées sur la monnaie, les meules rotatives, l'*instrumentum*, les langues ou tout autre contenu plus approfondi. La base doit servir à documenter le contexte général de l'âge du Fer dans lequel s'inscrivent des bases plus spécialisées, qui la complètent. Le colloque de l'AFEAF de Chauvigny a montré que plusieurs bases spécialisées sur le mobilier ou les données environnementales s'étaient constituées (Bertrand *et al.* 2009a et b). La *Basefer* ne prétend pas se substituer à elles. Il existe une version en ligne simplifiée et accessible au grand public.² A la différence des nombreuses autres bases, qui cherchent surtout à répondre à la question « *Qu'y a-t-il à tel endroit ?* » avec un maximum de données, la *Basefer* cherche à répondre à la question « *Où sont tous les habitats de telle ou telle époque, toutes les mines, tous les bracelets en lignite etc.* » et à croiser ou modéliser ces cartes de répartition.

4. Des raisons d'espérer

Il existe quelques équipes en France et en Europe dont les objectifs sont assez proches, et avec lesquelles nous avons travaillé ces dernières années dans le cadre de l'ANR « Eurodigitmap » (ArkéoGis à Strasbourg, Topoi à Berlin etc.). Des séances de travail collectives ont montré que l'objectif était le même, avec des choix techniques différents, liés à la spécificité des corpus, mais complémentaires. Aucun obstacle théorique ne viendrait entraver la coopération de ces projets.

On pourrait objecter que la masse à traiter est déjà trop importante pour réaliser un projet de « big data » archéologique. L'exemple du Service archéologique de la région Bretagne, qui a mis en ligne tous les rapports de fouille depuis les origines et construit un système d'interrogation qui permet de faire des cartes thématiques, est là pour prouver le contraire. L'objection traditionnelle aux cartes de répartition archéologiques qui ne seraient que les cartes d'activité des chercheurs a été encore une fois démentie par le travail de S. Krausz dans la région Centre: les grandes phases de la protohistoire dessinent des répartitions différentes, clairement indépendante des contraintes de la recherche, du milieu naturel, des biais les plus divers (Krausz 2015) : les choix culturels sont bien visibles.

Le TGE Huma-num qui a été mis en place récemment assure aujourd'hui le stockage et la pérennité des données. Plusieurs moteurs de recherche sont disponibles ou en cours d'élaboration pour faciliter la saisie et le traitement statistique et cartographique de ces masses de données. Nous travaillons actuellement sur un logiciel permettant la saisie automatisée des inventaires, dont la fiabilité a déjà été testée (K. Gruel, M. Cartereau). Aujourd'hui la connexion des serveurs et la constitution de « big data » ne pose plus de problèmes techniques, il est urgent de connecter les données existantes, les résultats de l'archéologie de sauvetage et de l'archéologie programmée,

² Cet « Atlas de l'âge du Fer » a pour objectif de permettre au public de se faire rapidement une idée sur l'état de la connaissance sur un pays, une région, une sous-région ou département (**Où ?**), à une période donnée (**Quand ?**), à propos d'un thème et d'un sous-thème (**Quoi ?**). Atlas de l'âge du Fer, AOROC, <http://www.chronocarto.ens.fr/gcserver/atlas>. Le dépôt de ces cartes thématiques dans l'atlas se fait par l'intermédiaire de Chronocarto, le visualiseur en réseau de données géo-référencées d'AOROC, développé par la société Géocarta.

des thèses et des informations qui papillonnent sur internet. Il ne manque plus qu'une décision politique, une distribution des tâches, et un financement pour mettre à la disposition de tous un outil sans lequel nos travaux reposent sur de petits nuages.

Références

- Archeomedes 1998. *Des oppida aux métropoles*. Paris: Anthropos-Economica, 280 p.
- Batardy, C., Buchsenschutz, O. and Dumasy, F. (eds) 2001. *Le Berry antique : Atlas 2000*. Tours: RACF, Supplément à la Revue archéologique du centre de la France 21, 190 p.
- Batardy, Ch., Buchsenschutz, O., Cartereau, M., Gruel, K. and Levery, M. 2015. Une base pour l'élaboration de modèles de peuplement de l'Âge du Fer en France. *Archeosciences, Revue d'Archéométrie*, G.M.P.C.A/Presses universitaires de Rennes, 2015, 39 : 157-175.
- Bertin, J. 1967. *Sémiologie graphique*. Paris: Mouton Gauthier-Villars, 431 p.
- Borillo, M., and Gardin, J.-C. (eds) 1974. *Banques de données archéologiques. Actes du Colloque (Marseille, 12-14 juin 1972)*. Paris: CNRS (Colloques nationaux du Centre national de la recherche scientifique 932).
- Bertrand, I. et al. (eds) 2009. *Les Gaulois entre Loire et Dordogne. Actes du XXXI Colloque international de l'AFÉAF (Chauvigny, 17-20 mai 2007)*. Tome1, Thème régional (Mémoire-SRAC 34). Chauvigny: APC.
- Bertrand, I. et al. (eds) 2009. *Habitats et paysages ruraux en Gaule et regards sur d'autres régions du monde celtique. Actes du XXXI Colloque international de l'AFÉAF (Chauvigny, 17-20 mai 2007)*. Tome 2, Thème spécialisé (Mémoire-SRAC 35). Chauvigny: APC.
- Buchsenschutz, O., Dorion, J., Mennessier, C. and A. Querrien 1974a. *Code d'analyse pour une carte archéologique de la France*. Groupe de recherche Informatique et archéologie, CNRS-Université de Paris 1, UER d'art et archéologie. Dactylographié, 79.
- Buchsenschutz, O. and Mennessier, C. 1974b. *Projet d'exploitation automatique d'une carte archéologique de la France*, in Borillo, Gardin 1974: 301-312.
- Buchsenschutz, O. and Debanne, X. 1978. Quatre années de traitement informatique des données de fouilles à Levroux (Indre), in *Colloque français d'archéométrie (Rennes 1977)*, 2^e partie. *Revue d'archéométrie* 2: 19-27.
- Buchsenschutz, O., Cartereau, M. and Lambert, G.-N. 1986. CENTRAR: un système de gestion et de représentation graphique des données archéologiques, in *Applications informatiques en archéologie*, *Bulletin de la Société préhistorique française* 83, 10: 325-327.
- Gruel, K., Buchsenschutz, O., Murgalé, H., and Secondo, J.-M. 1992. *Arkéoplan, manuel d'utilisation*. Marseille: BEME, 225 p.
- Gandini, C. and Laiüt, L. (dir.) 2013. *Regards croisés sur le Berry ancien. Sites, réseaux et territoires*, 45^e Suppl. à la Revue Archéologique du Centre, Tours: ARCHEA/FERAC.
- Krausz, S. 2016. *Des premières communautés paysannes à la naissance de l'État dans le Centre de la France: 5000-50 A.C.*, *Scripta antiqua* 86, Bordeaux: 372 p.
- Nakoinz, O. 2005. *Studien zur räumlichen Abgrenzung und Strukturierung der älteren Hunsrück-Eifel-Kultur*, *Ur- Forschungen zur prähistorischen Archäologie* 118, Bonn: 363 p.
- Nakoinz O. 2013. *Archäologische Kulturgeographie der ältereisenzeitlichen Zentralorte Südwestdeutschlands*. *Universitätsforschungen zur prähistorischen Archäologie* 224, Bonn.

Innovative multidisciplinary method using Machine Learning to define human behaviors and environments during the Caune de l'Arago (Tautavel, France) Middle Pleistocene occupations

Sophie Grégoire¹, Nicolas Boulbes¹, Bernard Quinio², Matthieu Boussard³, Caroline Chopinaud³, Anne-Marie Moigne⁴, Agnès Testu¹, Vincenzo Celiberti¹, Cédric Fontaneil¹, Christian Perrenoud⁴, Anne-Sophie Lartigot Campin¹, Thibaud Saos¹, Tony Chevalier¹, Véronique Pois¹, Henry de Lumley⁵, Marie-Antoinette de Lumley⁵, Antoine Harfouche², Rolande Marciniack², Philippe Carrez⁶, Thierry Hervé⁷

¹ HNHP- UMR 7194, Université de Perpignan, CNRS, MNHN, Centre Européen de Recherches Préhistoriques, Tautavel. Corresponding author: gregoire@cerptautavel.com

² CEROS – EA 4429 -Université Paris Nanterre, UFR Segmi

³ Craft Ai. WeWork, Paris

⁴ HNHP- UMR 7194, Muséum national d'Histoire Naturelle, CNRS, UPVD, Centre Européen de Recherches Préhistoriques, Tautavel

⁵ Institut de Paléontologie Humaine, Paris

⁶ Immersion-Tools, Saint-Brieuc

⁷ 16 rue Schaeffer, 93300 Aubervilliers

Abstract

SCHOPPER ANR project aim to develop a method to test research hypotheses using Machine Learning and immersives virtual environments. After 54 years of excavations, the Caune de l'Arago cave delivers a large amount of raw data and many archaeometric databases were developed through the study of this Lower Paleolithic site. The current tests focused on the evaluation of paleoenvironmental conditions and behavioral ones in each archaeological levels between 560,000 years and 90,000 years BP. All archaeological and environmental variable providing information on these issues were collected and organized by the multidisciplinary team to be studied by Machine Learning approach able to build models based on current repositories or Expert hypothesis, to learn on these models and then classify the archeological levels according to these patterns. The different steps of this innovative method are presented in this paper.

Keywords: Pleistocene, Lower Paleolithic, Machine Learning, Caune de l'Arago

Résumé

Le projet ANR SCHOPPER vise à développer une méthode pour tester des hypothèses de recherche en utilisant les techniques de Machine Learning et des environnements virtuels immersifs. Des tests sur l'évaluation des conditions paléoenvironnementales et comportementales dans chaque niveau archéologique entre 560.000 ans et 90.000 ans BP. ont été menés. Toutes les variables archéologiques et environnementales fournissant des informations sur ces questions ont été collectées et organisées par l'équipe multidisciplinaire pour être étudiées selon une approche de Machine Learning capable de construire des modèles basés sur des référentiels actuels ou des hypothèses d'Experts, et de classer ensuite les niveaux archéologiques selon ces derniers. Les différentes étapes de cette méthode innovante sont présentées dans cet article.

Mots-clés : Pléistocène, Paléolithique inférieur, apprentissage automatique, Caune de l'Arago

Prehistoric archeology is a rapidly changing field thanks to the arrival of digital technologies which today allow the acquisition, management, conservation and valuation of data at undreamed

performance levels ten years ago. Multidisciplinary studies, more and more frequent in the field of prehistory, are beginning to provide a multitude of data that constitute a favorable substrate for the development of research tools using Machine Learning. In recent years, works based on these tools with various applications, related to taphonomic archaeological issues (Arriaza *et al.* 2016; Egeland *et al.* 2018; Domínguez-Rodrigo and Baquedano 2018; Byeon *et al.* 2019), Geochronology (Petrelli *et al.* 2017) or environmental ones (Sobol and Finkelstein 2018; Sobol *et al.* 2019; Žliobaitė 2019), and paleolithic art (Wang *et al.* 2010) are emerging.

In terms of human behavior, researches using ML are numerous and growing. Far from our prehistoric preoccupations, they concern for example the management of situations of conflicts or disasters (Provitolo *et al.* 2015). Regarding prehistoric behavior, research using Machine Learning is still rare (Brizi Godino *et al.* 2018), because it requires the use of many examples of clearly characterized behaviors, to feed the learning phases and because archeological data are still too few, heterogeneous and hypothetical. However, the need for analysis of multidisciplinary data on Paleolithic behaviors constitute a major research issue whereas classical methodological solutions (experimental archeology) are rapidly reaching their limits.

The SCHOPPER project aims to create a digital full use of databases through an innovative system able to test research hypothesis on Paleolithic human behavior, using Machine Learning (craft ai. company), and later, through simulations in virtual environment (Immersion Tools company), with the support of researcher on Knowledge Management (CEROS team of Nanterre Univ.). This program is based on the lower Palaeolithic site of the Caune de l'Arago, in the south of France. It exploits the field data collected for more than 50 years excavations and the results of multidisciplinary studies. These archaeological and environmental data, exceptionally rich and dense, are confronted with models of environmental processes and behaviors, developed thanks to different methods and tested according to the problematic. Virtual reconstructions, partly based on the predictions obtained, then allow the researchers to immerse themselves in past environments that favors formulations and tests of hypotheses (Grégoire *et al.*, submitted). This article presents and discusses the methodology of this new numerical research tool, particularly the ML approach, and the first results obtained in reconstructing paleoenvironments and Paleolithic behaviors.

1. Material and method

1.1. Site presentation

The pilot site to develop this new research is the Caune de l'Arago cave in Tautavel, in the south of France (Figure 1), which has benefited from 54 years of excavations (Lumley *et al.* 2014). This site contains nearly 600,000 coordinated objects, from 55 archaeological layers, recorded within a stratigraphic sequence of 15 meters thick, developed between 690,000 years and 90,000 years BP. (Perrenoud *et al.* 2016; Falguères *et al.* 2015; Falguères *et al.* 2004). The richness of this Paleolithic sequence, the quality of remains conservation, the exceptional excavation duration with a standardized and continuous data recording model, make it one of the best fields of development and application of the method presented in this article.

After 54 years of data collecting and multidisciplinary studies, a large amount of raw data was acquired through excavations and pluridisciplinary studies. While a monographic work of synthesis is completed (Lumley *et al.* 2015, in progress), archaeologists have enough perspective on their data to start the next step: using Ai to explore data.

1.2. The data organization phase

Because of its multidisciplinary nature, research in Prehistory makes it difficult to cross heterogeneous, often partial and for some of them, hypothetical data. A phase of Mind Mapping

	Complexes	Ensembles stratigraphiques	❖ Grandes unités archéostratigraphiques			
			▪ Unités archéostratigraphiques			
			◦ Sous-unités archéostratigraphiques			
			♦ Horizons			
			❖	▪	◦	*
	COMPLEXE SOMMITAL		RFB			
			RFO			
	COMPLEXE SUPÉRIEUR		A	A1 A2 A3 A4		
			B			
			C	Cs Cm Ci		
400 000 ans	COMPLEXE MOYEN	Ensemble stratigraphique III	D	Ds	Ds1 Ds2	
				Di	Dis	
					Dis3 Dis4	
					Dib5 Dib6	
			E	Es	Es1 Es2 Es3	
				Ei	Ei4 Ei5 Ei6	
			F	Fs1		
				Fm	Fm2 Fm3 Fm4	
				Fi ou Inter F/G	Fia inter F/G sup	
					Fia5 Fia6	
					Fib inter F/G inf	
					Fib7 Fib8	
450 000 ans		Ensemble stratigraphique II	G	Gs1		
				Gm	Gm2 Gm3	
				Gi4		
500 000 ans						
		Ensemble stratigraphique I	H	H1 H2 H3		
			I	I1 I2		
			J			
			K			
			L			
			M			
			N		N1 N2	
			O		O1 O2 O3	
			P		P1 P2 P3	
			Q		Q1 Q2 Q3 Q4	
550 000 ans						



Figure 1. The Caune de l'Arago cave (Tautavel, Pyrénées-Orientales, France).

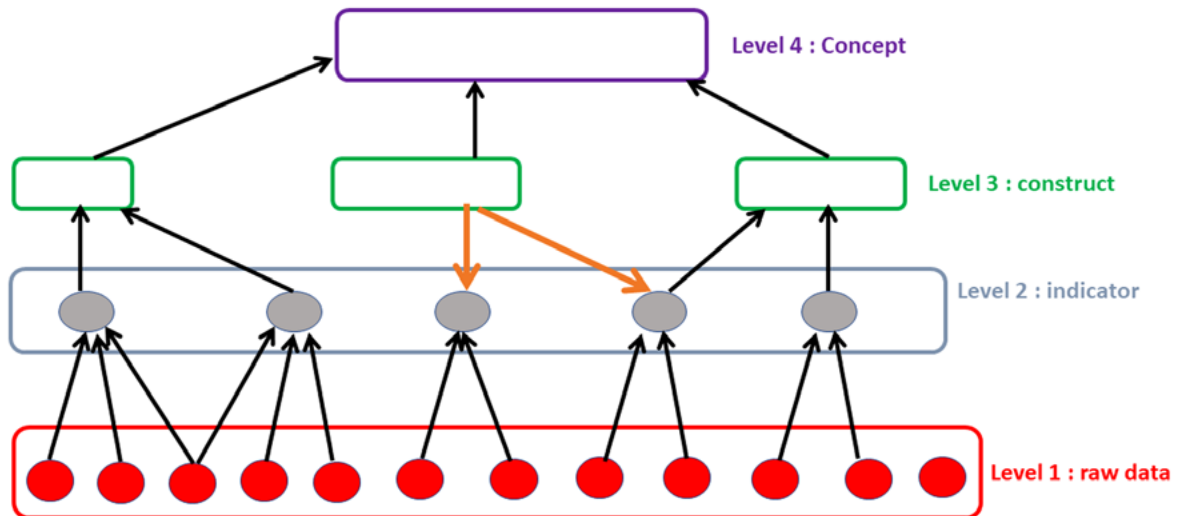


Figure 2. A conceptual model of research.

was used to structure the tacit and explicit knowledge related to each domain to hierarchize them from the initial raw data to the global concepts they allow to illustrate. Key concepts at the crossroads of disciplines were selected to develop sub-models of scenarios on which Machine Learning techniques could be applicable. The necessary and useful variables involved in a given scenario were then selected for each discipline. These data are extracted from a site-specific database called ‘matériel paléontologique et préhistorique’ augmented by researchers’ databases. These heterogeneous data are stored in different formats and computer media.

Researchers, from management to medical study, consider there are 4 classic levels of abstraction (MacKenzie *et al.* 2005). In the case of SCHOPPER, the levels can be considered as follows (Figure 2):

- Level 1: The first-level raw data captured in the central database is the palaeontological and lithic material descriptive and archaeometric characteristics. These first-level data are precise and specific to each piece or object but they do not yet make sense.
- Level 2: These data must be aggregated and interpreted to become indicators.
- Level 3: From indicators, formative or reflective constructs (Bedford *et al.* 2017) are developed.

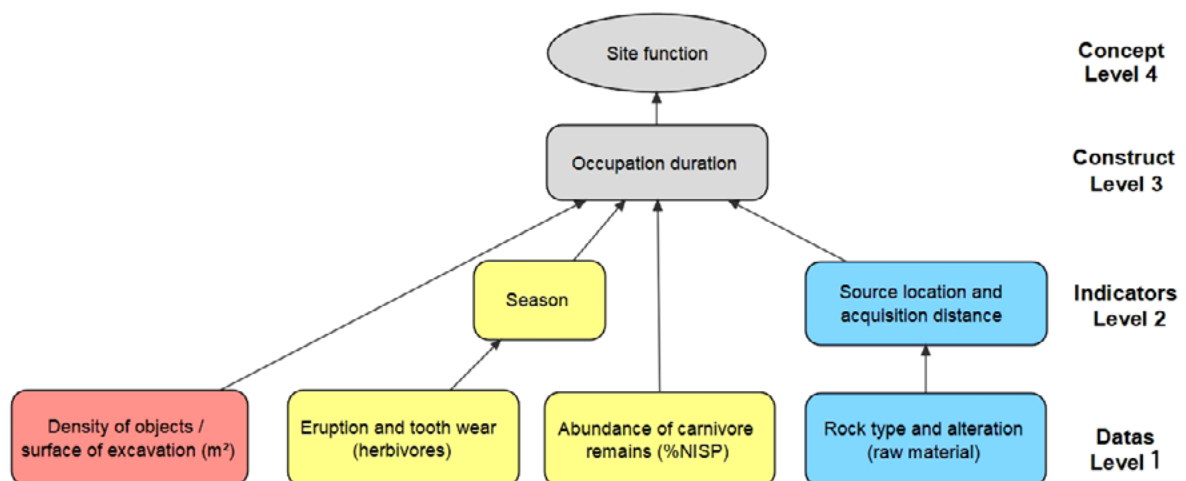


Figure 3. Example of a part of the cognitive map of the occupation duration and site function scenario (red: general archaeological considerations, yellow: faunal raw data and indicators, blue: lithics raw data and indicators, grey: constructs and concepts).

- Level 4: Finally these constructs participate to the definition of the terminal concepts which are the aim research questions which can only be treated in an interdisciplinary way.

Constructs and concepts are research hypotheses based on the analysis of multidisciplinary data and measured or calculated indicators. Cognitive maps by discipline were realized (Cosette 2008), then used to work on a construct. The elements of the map involved in the formation of a construct were put together in scenarios allowing to work on great concepts like mobility, occupation duration or site function (Figure 3).

2. Development of two methods of Machine Learning adapted to types of research hypotheses

Based on this first phase of data organizations and scenario definition, the second iterative stage led to the development of Machine Learning approaches for the generation of predictive models, ultimately allowing the testing of scientific hypotheses.

Supervised learning is a type of system in which both input and desired output data are provided. Input and output data are labelled for classification to provide a learning basis for future data processing.

In this context, two approaches appeared necessary to manage the different kinds of research hypothesis in the field of archeology.

Given that some research questions can benefit, for the learning step, of external repositories or not, we were led to develop two approaches, one for environmental questions and other for behavioral questions.

The former can rely on current ecological repositories according to the principle of actualism. The second one must necessary use prehistoric references and requires to work differently for the learning step. The principle is to use the most explicit and best characterized archeological data to classify the others thanks to algorithms tracking all the informative potential of raw data and indicators whose classical exploitation by researchers are still difficult.

These approaches could be qualified as ‘exogenous method’ for environmental questions using actualism principle and external repositories and as ‘endogenous method’ for behavioral questions using Expert Labeling. They are defined as follows:

2.1. Exogenous method: the example of biome prediction using fauna proxy

This method consists in generating predictive models learned on current and sub-current repositories (pollen spectra and open-source fauna database), following the principle of actualism, and then applied to the different levels of the Caune de l’Arago to deduce the paleoenvironments. This approach was used with two proxies related to paleoenvironments, pollen and fauna. The archeological pollen databases correspond to numerous samples throughout the entire Caune de l’Arago sequence (Renault-Miskovsky 1981; Lartigot-Campin 2007; Lartigot-Campin *et al.* in prep.). Each archaeological unit is therefore characterised by one or more pollen samples, broken down into the number of grains per taxon. Faunal proxies can also be used alone or combined with the pollen proxy. Faunal proxies is used here to describe an example of the work process.

2.1.1. Data preparation phase

Fauna database from Arago cave corresponds to a taxonomic inventory per archaeological unit of the entire community of vertebrates, large and small mammals, amphibians, reptiles and birds (Moigne *et al.* 2006; Desclaux 1992; Hanquet, Desclaux 2011; Manzano 2015; Magniez *et al.* 2013; Lebreton *et al.* 2016, 2017). The choice to work in ‘presence-absence’ is motivated by the

quantitative heterogeneity of different groups of vertebrates and the bias due to predation on the representativity of certain species. This is particularly the case of large herbivores hunted by Paleolithics and some rodents accumulated by the action of some specialized predators.

The external repository used for the learning step, is the Wild Finder database (World Wildlife Fund., 2006). It contains presence/absence data for the world's terrestrial amphibians, reptiles, birds, and mammals, by terrestrial ecoregion which represents as many datapoints. Ecoregions are defined as 'relatively large units of land that contains a distinct assemblage of natural communities and species, with boundaries that approximate the original extent of the natural communities prior to major land use change' (Olson *et al.* 2001). In the dataset, each ecoregion is characterized by a main biome (also called major habitat type), the output class chosen.

In order to prepare a training set as compatible as possible with the Caune de l'Arago faunal data, the data from the WWF dataset have been restricted to the biogeographic realm Palearctic and to the following eight biomes: Tundra; Montane Grasslands and Shrublands; Boreal Forests/Taiga; Temperate Coniferous Forests; Temperate Broadleaf and Mixed Forests; Temperate Grasslands, Savannas and Shrublands; Mediterranean Forests, Woodlands and Scrub; Deserts and Xeric Shrublands.

A current referent is assigned to the species from Arago cave. Three cases may occur: (1) the species is still present today, (2) the species is extinct, but a current taxon with closed ecological characteristics can be substituted for it, (3) the fossil species has no equivalent today and is not taken into account. Two tests were developed: the first on the vertebrate community of the Caune de l'Arago and the second only on mammals of the site but including list of species from other Middle Pleistocene localities (presented here).

2.1.2. Exploration

Programming is carried out from the notebook interface Jupyter with the language Python. Pandas library provides data processing structure and analysis. The reference dataset has first to be explored and analysed to detect potential issues like unbalanced classes, points of data with few observations, skewed distributions, missing values etc... We then use classical multivariate analysis tools like PCA and also tools to visualize high dimensional data, such as t-SNE (van der Maaten and Hinton 2008), which improves the visualization of complex point clouds (Figure 4).

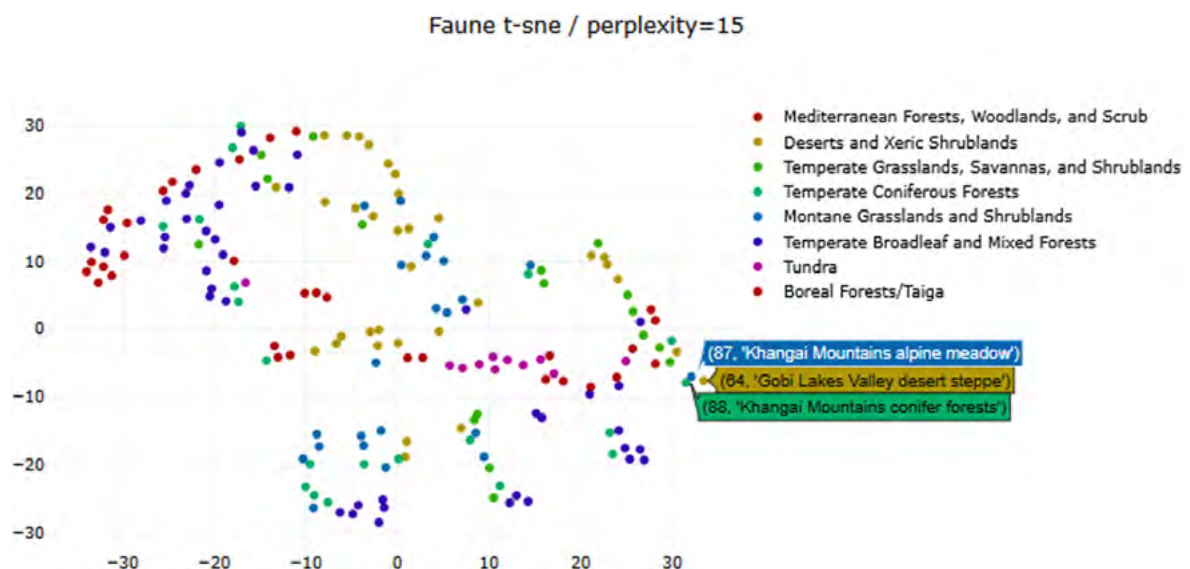


Figure 4. 2D visualization of high dimensional data with TSNE.

2.1.3. Modeling and validation

The machine learning algorithm chosen is ‘Extremely randomized trees’ (Geurtz *et al.* 2006) which is robust, available in the scikit learn library (Pedregosa *et al.* 2011). The modeling phase involves finding the optimal learning parameters and balancing the complexity and generalization capacity of the model. For a better suitability, two successive models are constructed: the first at the global scale of the reference frames so we can determine how the problem is difficult to classify knowing we have all taxa available. The second one, reduced to the taxa present at the Caune de l’Arago and in other localities, is used for predictions. The cross between the species from Middle Pleistocene sites including Arago cave and the WWF dataset occurs for more than 100 taxa of mammals (large mammals, lagomorphs, rodents, insectivores, bats). The validation of the models (k-fold Cross-validation with 8 fold) and the confusion matrix resulting from this step show that the classification errors are moderate (accuracy about 60%).

2.1.4. Predictions

Once the classifier is validated, it is applied to the faunal series of the different levels of the Caune de l’Arago. The predictions but also all the probabilities of belonging to each class are provided, which in this case, makes it possible to identify the biomes in competition in each level. Past communities in relation with the Pleistocene climatic variations (so-called ‘non-analogous’) results from biogeographical consequences, migration event and refuge (especially in mediterranean area). Biome competition predictions are regarding here as tendencies to climatic conditions and landscape, and should not be considered as aggregate or mozaic habitats.

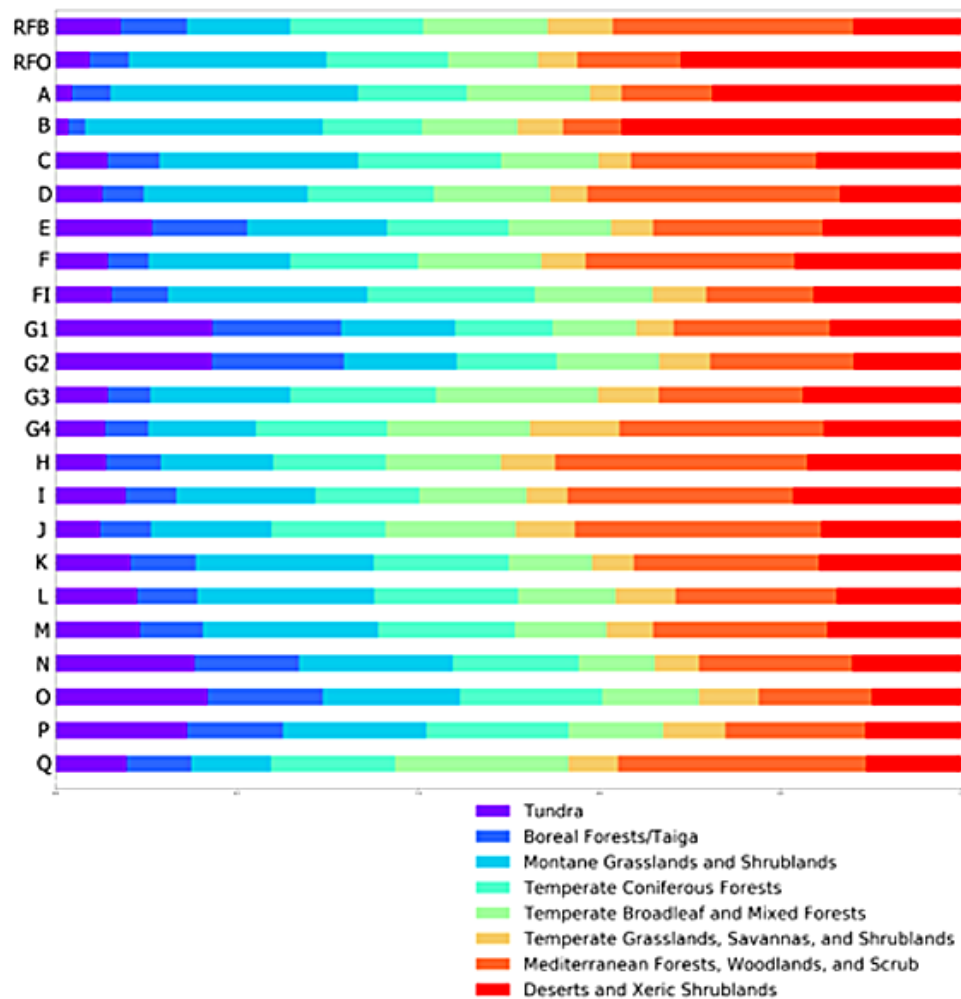


Figure 5.
Biome's
predictions in
the stratigraphic
levels of Caune
de l’Arago cave.

In the detail, from the bottom to top of the stratigraphy, we can describe the following climatic cycles. It is for example interesting to show the progressive general pattern evolution of the probabilities of discriminant and opposite classes as Mediterranean and Tundra (Figure 5). The main biomes in competition are temperate in the lower level Q (Mediterranean, Temperate Broadleaf and mixed forests). The composition is radically different in the level P, O and N where the probabilities of predictions of cold environments are greater (Tundra, Montane Grasslands, Taiga). Progressively, from level M to level H the Mediterranean class becomes largely majoritary before decreasing in UA G4 and G3, but temperate biomes still remain dominant. Levels G2 and G1 mark a major change: they are characterized by high probability of the biomes Tundra (majority class) and Taiga. The other levels from the upper part of the CM III are not very contrasted and balanced between temperate and cold climatic conditions. Deserts and Xeric Shrublands are for the first time the majority class in level B, the probability of this biome is also high in next level A and RFO. Finally mediterranean habitat is the most likely environment in the RFB layer.

2.1.5. Explanation and results analysis

To interpret the predictions, explanatory algorithms such Shap (Lundberg and Lee 2017) were used. This step is crucial, it makes it possible to identify the taxa that influence the prediction of such or such biomes on a given archeostratigraphic layer, facilitates the return to the data and assists in the interpretation of the results.

First representation provides a *global interpretability* of the model: the Shap values can show how much each predictor (species) contributes, either positively or negatively, to the model. The variable importance plot lists the most significant variables: in the example of the level O, top variable as the arctic fox (*Vulpes lagopus*) has a high predictive power (Figure 6).

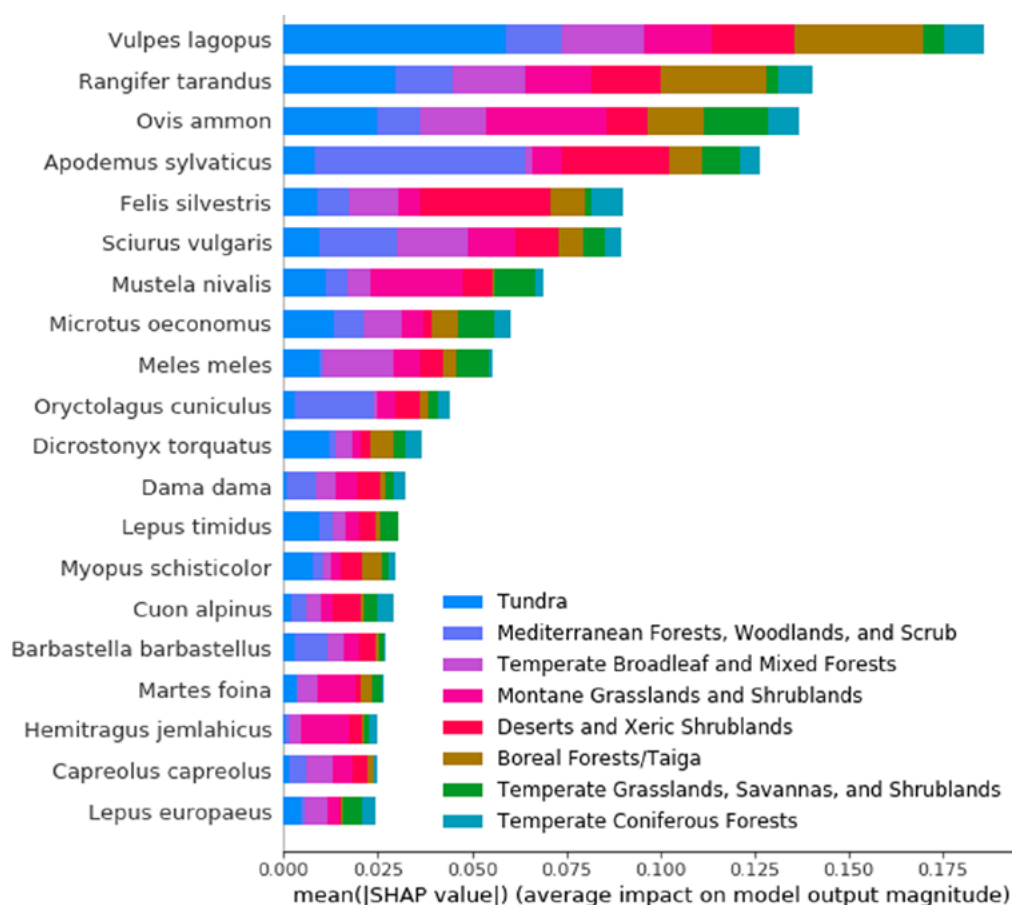


Figure 6. Variable importance plot for the level O.

The name of the species corresponds to the current taxon used as referent.

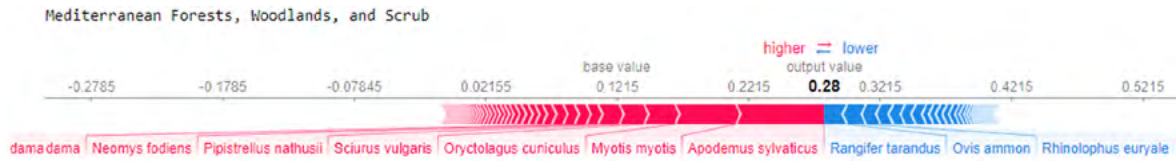


Figure 7. Visualization (force plot) of the prediction's explanation of the majority class from the level J.

The above explanation with Shap provides a *local interpretability*. It shows, for any level and any predictions, features each contributing to push the model output from the base value (the average model output over the training dataset we passed) to the model output. In the example below (Figure 7) level J belongs to the majoritary class 'Mediterranean' with a probability of 0.28 (base value $p=0.1215$). Features (taxa) pushing the prediction higher (in red) are the wood mouse (*Apodemus sylvaticus*), the greater mouse-eared bat (*Myotis myotis*), red squirrel (*Sciurus vulgaris*), European rabbit (*Oryctolagus cuniculus*), fallow deer (*Dama dama*). As opposite, species with a negative effect on prediction (in blue) are the reindeer (*Rangifer tarandus*), the argali (*Ovis ammon*).

Overall these results are in agreement with the climate and landscape reconstitutions based on other proxies. Especially the major cycle recognized on the Middle Complex correlated to glacial and interglacial phases (MIS 14-13-12) (Lumley *et al.* 2015; Perrenoud *et al.* 2016; Moigne *et al.* 2006). The difference of paleoenvironmental and climatic conditions within the large G unit was already demonstrated by recent study on small mammals (Lebreton *et al.* 2017). Levels G4 and G3 could correspond to a more temperate and humid climate with a more closed environment, consisting of temperate forests and Mediterranean vegetation. Levels G2 and G1 correspond to colder and open environments.

2.2. Endogenous method: the example of duration of occupation evaluation

The development of this approach is based on the observation that, to work on Paleolithic behaviors, large modernist ethnographic databases such as Lewis R. Binford one (Binford 2001) are hardly applicable as only reference in the chrono-cultural framework of the lower Paleolithic (Audouze 2013). On the other hand, it is difficult to set up benchmarks with a sufficient number of examples on fossil human population behaviors. This requires collecting examples of well-defined behaviors in publications with archaeological variables explicitly related to a particular behavior and constituting a reference dataset. The production of such a dataset, even if not easy, given the diversity of methodological approach in this field, would be necessarily subjective. So it is necessary to find a special approach to work on behavioral questions using Machine Learning.

2.2.1. Training set preparation with Expert Labeling

In this context, the exogenous method seems difficult to apply. Thereby, we have an obligation to learn from the internal database (i.e. endogenous method). Some work on the Caune de l'Arago was aimed at formulating hypotheses as to site function, occupation duration, activities, size and group composition (Lumley *et al.* 2004; Moigne *et al.* 2005; Barsky *et al.* 2005). These hypotheses were proposed on the multidisciplinary synthesis of 5 levels of occupation of the Caune de l'Arago (F, G, J, L and Inter-FG). These different occupation levels were best able to provide sufficiently contrasting archaeological data to indicate one or more behavioral changes from one level to the next. These first results constitute for the present work, a base to select the archaeological variables that affect the definition of the site function, duration and season occupation, and to qualify archaeological situations.

Once listed all the archaeological features (Table 1) that influence behaviors, the archaeologist, can propose a classification hypothesis based on an evaluation of multidisciplinary data and by using

Variables	Type	Interpretation ways	General and specific references
Density of material / surface excavation (m ²)	Numeric		Lumley <i>et al.</i> 2015
NISP lithic / (NISP lithic + fauna)	Numeric	Low density of lithic material suggests brief occupations during which predation/treatment animal resources dominates	Rendu <i>et al.</i> 2011
Number of species of large herbivores	Numeric	Varied herbivore spectrum could refer to short-term regular hunting camps	Daujeard, Moncel 2010
Shannon's diversity index (H)	Numeric	Indice used to determine the specialization of hunting. Ability to organize massive hunts on a taxon (over a short period of time) to build food reserves (stocks) for longer consumption on a base camp. Specialised hunting could also refer to long-term residential camps	Magniez 2010; Binford 1981; David et Enloe 1993; Daujeard, Moncel 2010; Lyman 1989
Abundance of carnivores	Numeric (%)	Abundance of carnivore: repeated passage of carnivores and therefore refers to frequent movements of human groups	Aura <i>et al.</i> 2002; Daujeard, Moncel 2010; Stiner 1991; Valensi 2000; Daujeard <i>et al.</i> 2011; Costamagno <i>et al.</i> 2006
Carnivore traces	Numeric (%)		
IFO index	Numeric	Measures the intensity of bone fragmentation (NISP complete long bones / NISP long bones). Could refer specialized activities and/or partial treatment of the carcass	Brugal, Patou-Mathis 1993
Anthropic traces	Numeric (%)	Cut marks and intentional bone fracturation. Intensity of exploitation/production of animal resources	Binford 1981; Lyman 1994; Moncel, Daujeard 2010
Anatomical connections	Numeric (%)	Large number of anatomical connections in an anthropogenic level implies a partial exploitation of the carcass rather linked to brief passages of human groups. Dispersion of remains	Moigne 1983; Rendu <i>et al.</i> 2011
Season	Categorical	Rhythmicity of the growth of cervid antlers and especially the age of young herbivorous. 1 season = short or repeated at the same season; several seasons = long occupation or frequent in different seasons	Chacon <i>et al.</i> 2015
Tooth ungulate microwear	Categorical	Allows to estimate the relative duration of the mortality event(s). (i) Season or short event; (ii) event longer than a season; (iii) separated events that occurred in different non-contiguous seasons. It can be used for expert labeling	Rivals <i>et al.</i> 2015
Manuports	Numeric (%)	A high % implies a long occupation. It could be interpreted as a raw material reserve	
Hammer	Numeric (%)	Indicate an <i>in situ</i> production rather than an introduction of finish tools. High % refers to a long occupation	
Milky quartz chaîne opératoire	Categorical	Local raw material / CO complete refers to long occupation	Barsky <i>et al.</i> 2005; Grégoire <i>et al.</i> 2006; Barsky 2013
Jasper chaîne opératoire		Semi-local raw material / CO incomplete refers to shorter occupation	
Miocene patinated flint chaîne opératoire		Allochthonous raw material / CO incomplete refers to shorter	

Table 1. Example of variables selection to work on the question of occupation duration.

Variables	Type	Interpretation ways	General and specific references
Cortical flakes	Numeric (%)	low % indicate preparation outside and bring on site = shorter occupation high % indicates introduction and preparation on site and refers to longer occupation	Dibble 1995; Shott 1996; Clark, Barton 2017; Bicho, Cascalheira 2018
Retouching flakes	Numeric (%)	high % assumes a longer utilization of tools and rather long occupation (edges modification and sharpening and curating processes edges). <i>But also reused tools can suggest short occupation</i>	
Sharpening flakes	Numeric (%)		
Multiples tools	Numeric (%)	specialized tools could refers short occupation and multiple tools longer one	
Recycled tool or blank	Numeric (%)	high % of recycling tools, blanksor matrix, reused tool and for another function / use of available resources rather thannew supply displacement suggest long duration occupation. <i>But depends of the function and raw materialused</i>	
Ebrechures	Numeric (%)	high % of trampling so rather long occupation / fresh edges short and fast occupation	
<i>In situ</i> raw material	Numeric (%)	4 concentric circular areas to give the information of origin of raw material high % of <i>in situ</i> raw material (zone 1) suggest long occupation	Grégoire <i>et al.</i> 2007
Local raw material		high % of local raw material (zone 2) suggest long occupation	
Semi-local raw material		> 20% of semi-local raw material (zone 3) suggest repeated displacement related to short duration occupation	
Allochthonous raw material		> 10% of allochthonous raw material (zone 4) suggest rather short occupation	

Table 1. Continued.

his explicit and tacit knowledge. On a given case, for a specific question, the expert will propose to classify the appraised occupation layer in a category called output class or label. They are defined during a researcher multidisciplinary discussion and then tested.

2.2.2. The example of occupation duration

The concept of duration is difficult to evaluate in prehistoric sites as the mechanisms from which it arises are complex and never produce the same effects (Sullivan 1992). Yet its evaluation is crucial for the knowledge of Paleolithic behaviors (Leonova 1993; Moncel, Rivals 2011; Clark, Barton 2017; Rusch et al. 2019). In the case of Caune de l'Arago cave, proposals have been made for certain occupations (*op. cit.*) and serve here as a calibration for assigning an output class to the other levels of the Caune de l'Arago. Those who do not benefit from all the necessary indicators, unlike the levels that could be qualified, can thus be classified by the algorithm in one or other of the categories. The purpose of these classifications by supervised learning is here to test the levels of occupation a priori none or less explicit, based on the first assignments with the so-called classical method (i.e. without the use of ML) and to implement evidence of the variables involved in discrimination between output classes.

2.2.3. Variables choice

It depends directly on the asked question and the archaeological material analysis level. In the case of the Caune de l'Arago, these variables are multidisciplinary, quantitative or qualitative and each of them provides information in favor of one or the other class of outputs, sometimes both (or more if there are several output classes). The data can be processed in presence/absence or quantification.

A dataset is constituted with all these variables for the Caune de l'Arago 55 archaeostratigraphic layers. Output classes are ultimately assigned by experts for archaeostratigraphic layers containing sufficient arguments to propose a classification hypothesis. Learning and modeling will be done primarily on these data, then the prediction will be done on the whole dataset.

2.2.4. Exploration and modelisation

Pandas profiling (open-source package, <https://pypi.org/project/pandas-profiling/>) is used for exploratory and quick data analysis. This module generates a complete report for the dataset: type of variables, unique and missing values, descriptive and quantile statistics, histogram for visualizing distributions, and multiple tests of correlation.

The algorithm used is a gradient boosting classifier (Friedman, Jerome 2000). It generates a succession of small decision trees, called weak classifiers, each one just aiming at doing better than a random classifier. The combination of all those weak classifiers, produces the final decision. This algorithm is very convenient since, with few parameters, it can produce high accuracy classifier, while being robust to variable dependencies or scaling issues.

The multidisciplinary variables listed in Table 1 were selected because of their significance for the duration of occupation evaluation. Some of them are generally used for the treatment of this question, others are specific to the Caune de l'Arago according to its position within the economic territory. Depending on the situation of each of these variables in a given archeological layer, an 'Expert label' can be assigned if the data are considered sufficient. The training set thus constituted allows to generate models and a first prediction on the whole dataset.

2.2.5. Prediction

The experts labeled the dataset, where they can tell if they think the sample is short or not short. Experts can also decide not to provide any label. Furthermore even if they gave a label, they can also decide not to learn on this point, due too much uncertainty. Finally we learn only on labeled point with high expert confidence. Then we predict on all data points, labeled, labeled but uncertain, and unlabeled.

The result obtained in Figure 8 enable to compare the Expert labeling and the prediction and point the contradictory results, the similar results and the decisions taken with expert label in order to understand why. For example, the results shows: 59% of decision similar to the expert label, 9% of decision opposed to the expert label and 32% of prediction where the expert did not have enough information to decide. In these cases it is important to know which variable influence the decision or which variable is missing or insufficient to make the expert decision possible. Particularly the 9% of contradictory decisions ask questions and implies that we analyse the prediction mechanisms. An explanation phase is necessary to understand them.

2.2.6. Explanation phase and result analysis

The gradient boosting offers the possibility to show the importance of each variable in the classification (Figure 9). In this case, the first eleven features are decisives for the classification and have each and together a signification in terms of duration. The meaning of each variable is known (Table 1), but the meaning of the set of variable chosen is unknown to the expert but defined by the algorithm. The percentage of cortical flakes indicate if the reduction process takes place into the cave or not. His variation value is significant for the classification between short or not short duration in the sense that if the value is high it involves a complete and longer reduction process for flakes and tools production or on the contrary it indicates that only the using step takes place into the cave, reducing the duration of the stay. The second feature is the percentage of carnivore wears on herbivore bones, usually correlated, if it's high, with frequent movement of

Unités archéostratigraphiques	densité d'objets par surface fouillée (m2)	MP éloignée %	% Manuports	CO complète SCH9	% éclats corticaux	% éclats de retouche	% outils multiples	Indice de diversité (Shannon H)	% IFO (Indice de fragmentation osseuse)	% traces anthropiques	% traces carnivore	Expert label	Prediction
RFB	1613,3	11,4%	0,6%		53,1%	0,3%	1,1%	1,87	2,6%	15,2%	1,4%	?	NC
RFO	849	7,1%	0,3%		42,7%	0,3%	1,1%	1,88	0,0%	17,6%	0,0%	?	NC
A1	95	10,3%	0,0%	n	13,3%	0,0%	6,7%			0,0%	0,0%	?	NC
A2	182,8571429	4,4%	0,0%	n	6,5%	1,1%	3,1%	0,00	0,0%	0,0%	0,0%	?	C
A3	362,5	2,0%	2,0%	n	18,7%	0,4%	1,2%	0,00	0,0%	0,0%	0,0%	?	NC
A4	175	4,1%	2,2%	n	18,4%	3,4%	2,8%	0,00	0,0%	0,0%	0,0%	?	NC
B	377	3,5%	1,1%	n	33,9%	1,0%	1,4%	1,58	0,0%	1,0%	0,0%	?	NC
C S	445,4	2,1%	0,5%	n	48,0%	0,3%	1,3%	1,12	0,0%	3,0%	0,0%	NC	NC
CM	115,5555556	3,9%	0,4%	o	32,2%	0,2%	2,3%	0,60	0,0%	10,4%	0,0%	?	NC
NC	35,6	7,0%	4,1%	o	21,6%	0,0%	0,7%	0,69	0,0%	12,5%	0,0%	?	NC
D1	34,03333333	8,1%	1,7%	o	53,2%	0,4%	1,2%	1,85	0,0%	9,5%	0,0%	?	NC
D2	43,38095238	7,1%	0,9%	n	53,2%	0,0%	1,0%	1,75	0,0%	3,3%	2,2%	NC	NC
D3	30,8	10,1%	0,8%	o	51,2%	0,1%	1,2%	1,73	18,2%	10,9%	4,3%	NC	NC
D4	15,38461538	6,2%	1,3%	o	33,5%	0,3%	0,9%	1,59	25,0%	13,0%	6,5%	?	NC
D5	48,75757576	5,8%	1,1%	n	36,9%	0,2%	1,6%	1,73	7,1%	16,8%	3,2%	C	NC
D6	40,02777778	6,2%	0,9%	n	34,7%	0,3%	1,0%	1,60	0,0%	10,1%	1,1%	C	NC
E1	47,21428571	4,0%	3,6%	o	24,8%	0,5%	1,6%	1,36	0,0%	15,2%	1,3%	NC	NC
E2	145,0714286	4,8%	3,1%	o	24,1%	1,0%	2,3%	1,46	1,2%	12,1%	0,7%	C	C
E3	63,76744186	6,0%	5,1%	o	23,9%	1,9%	3,5%	1,55	2,9%	14,8%	0,8%	NC	NC
E4	27,63636364	7,1%	4,3%	o	25,7%	2,7%	3,2%	1,29	0,0%	10,9%	1,5%	NC	NC
E5	15,13333333	4,5%	3,1%	n	27,4%	2,0%	2,3%	1,30	0,0%	19,6%	2,2%	C	C
E6	28,10869565	4,0%	8,1%	n	24,7%	1,7%	1,7%	1,37	0,0%	16,2%	2,4%	NC	NC
F0	16,18181818	2,7%	3,7%	n	23,8%	0,0%	2,7%	0,99	0,0%	26,3%	1,4%	C	C
F1	299,1276596	4,8%	4,1%	o	20,5%	3,1%	3,2%	1,28	0,5%	21,7%	2,5%	NC	NC
F2	569,9791667	5,3%	5,7%	o	21,1%	2,6%	2,6%	1,24	0,9%	21,5%	2,2%	NC	NC
F3	450,1020408	5,1%	7,6%	o	23,6%	2,6%	2,2%	1,29	1,4%	20,3%	3,0%	NC	NC
F1	74,79	2,9%	4,2%	n	25,5%	1,7%	2,4%	1,49	0,4%	18,5%	2,9%	NC	NC
G1	706,5961538	2,4%	7,0%	o	23,5%	2,1%	1,9%	1,75	1,8%	21,7%	2,9%	NC	NC
G2	866,1071429	3,8%	10,0%	o	21,9%	2,6%	2,7%	1,73	1,7%	21,6%	2,8%	NC	NC
G3	699,5333333	4,8%	9,8%	o	19,8%	3,2%	3,4%	1,49	0,9%	24,5%	3,3%	NC	NC
G4	580,3934426	7,2%	8,1%	o	22,9%	5,0%	3,4%	1,42	0,8%	23,9%	3,1%	NC	NC
H1	52,48076923	10,4%	7,1%	n	23,4%	4,4%	5,6%	1,48	0,0%	25,0%	2,5%	NC	NC
H2	41,73469388	10,2%	8,9%	o	22,6%	2,3%	3,1%	1,67	3,1%	31,0%	3,0%	NC	NC
H3	35,02	8,3%	4,6%	n	23,4%	1,4%	6,4%	1,78	0,0%	30,7%	4,1%	NC	NC
I1	69,13461538	3,5%	5,3%	n	13,1%	1,3%	2,3%	1,64	2,0%	32,0%	3,8%	C	NC
I2	24,94339623	7,3%	3,5%	n	23,6%	1,0%	3,4%	1,32	0,0%	31,2%	3,5%	NC	NC
J	285,056338	6,0%	1,3%	n	17,4%	0,7%	1,8%	1,19	0,3%	30,3%	3,5%	C	C
K	26,25	3,5%	1,8%	n	10,9%	1,1%	2,1%	1,56	1,6%	28,1%	8,9%	C	C
L	129,9142857	3,5%	3,5%	n	15,8%	0,7%	0,6%	0,65	10,9%	23,1%	13,0%	C	NC
M	21,5	2,1%	5,8%	n	18,8%	0,5%	2,4%	1,47	2,9%	31,3%	7,7%	C	C
N1	14,47297297	5,3%	5,5%	n	42,9%	1,9%	0,0%	1,18	7,7%	16,1%	9,5%	NC	NC
N2	24,94666667	6,2%	1,3%	n	50,0%	1,6%	0,8%	1,43	17,4%	12,1%	7,5%	NC	NC
O1	9,25974026	7,3%	1,8%	n	68,8%	3,8%	1,3%	1,67	0,0%	16,4%	4,5%	NC	NC
O2	8,064935065	13,3%	0,9%	n	76,4%	0,0%	3,6%	1,33	0,0%	17,1%	1,2%	C	C
O3	6,87012987	6,8%	0,9%	n	63,1%	2,7%	1,8%	1,41	0,0%	15,9%	9,8%	NC	NC
P2	32,72151899	2,2%	0,9%	n	54,4%	0,8%	1,8%	1,66	3,4%	17,1%	7,4%	C	NC
P3	108,5375	3,5%	0,6%	o	52,0%	0,9%	1,6%	1,36	1,7%	20,6%	6,4%	NC	NC
Q1	178,6811594	4,7%	0,6%	o	39,0%	1,6%	1,3%	1,48	1,9%	22,8%	4,7%	NC	NC
Q2	40,53571429	4,0%	0,3%	n	35,1%	1,6%	2,5%	1,24	7,7%	19,6%	4,1%	NC	NC
Q3	90,2962963	5,6%	0,2%	o	38,4%	2,8%	2,6%	1,11	0,0%	22,3%	3,1%	NC	NC
Q4	35,73333333	5,1%		n	22,8%	13,7%	2,2%	0,69	0,0%	9,5%	9,5%	?	NC

Figure 8. Result of prediction classifying by short (C, green), not short (NC) output class. The training set has learnt on three expert labels: short, not short, and unknown (?) assigned and predicted on the whole layers (red: concordant predictions). The figure presents only a selection of variables. The original dataset contains 27 variables.

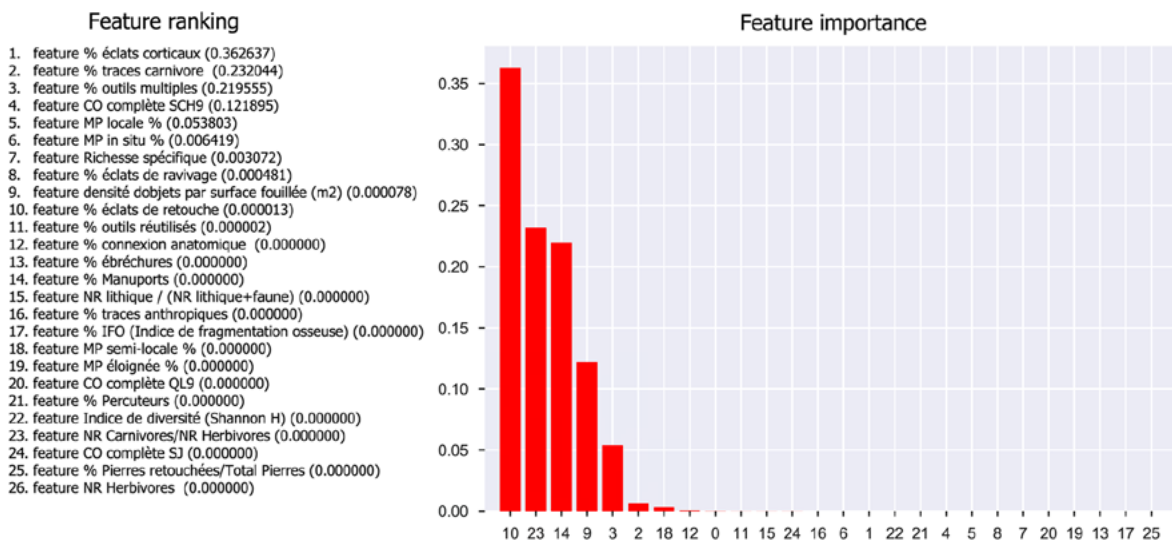


Figure 9. Feature ranking and feature importance for the prediction of short and not short occupation duration.

human groups and then with rather short-term occupation. The high value of the multiple tools seems to indicate a long duration since one blank could be optimized thanks to two or more active parts and having a long life and successive uses, and so on....

D5, D6, I1, L and P2 layers are predicted in opposition with the expert label. In each of them the expert had found indicators of short-duration, particularly on lithics indicators, considered less significant by the algorithm prediction. L is typically an equivocal layer because it is little extended, clearly stratigraphically delimited by sterile deposits, with indicators of short duration: carnivores action, low percentages of cortical flakes, retouching and sharpening flakes, seasonality hunting events during early winter on reindeer (Magniez *et al.* 2011), and microwear pattern (Rivals *et al.* 2015). Occupation was interpreted as short hunting camp (Lumley *et al.* 2004). Similar patterns (especially specialized hunting) are sometimes considered as a long term occupations (Deaujard, Moncel 2010; Betts, Friesen 2004). The prediction point out this ambiguity and leads to find new indicators to be more precise or to revise the thresholds determining the limit from one class to another.

2.3. For Exogenous and Endogenous method, a new way of learning was observed

Since the implementation of the regular tests, a triple learning loop adapted from that proposed by (Argyris and Schon 1996) have been observed. As a reminder, the Argyris and Schon organizational learning model describes two learning loops that make it possible to learn from an error or an offset between the expected result and the result obtained. The first is a simple adaptation to the difficulty encountered, it is fast and relies on routines. The second loop requires in-depth action that challenges the usual action patterns.

In the case of the Schopper project, the simple loop is similar to that described above (Figure 10). The data-scientist and the researcher will correct outliers of the data set or modify a parameter and redo the algorithm. The second loop concerns actions that challenge the data used by looking for other internal or external data. This loop can also lead to changing the nature of the algorithm used. Finally, the third loop questions the proposed conceptual infrastructure by questioning knowledge maps or scenarios. Indeed, if the results obtained by prediction appears aberrant or confused, it can lead to reconsider the output classes and to adjust them according to the degree of resolution reached on a given problematic. For example, the output classes to characterize the site occupation duration was initially: short, long or frequent and repeated. After a first series of

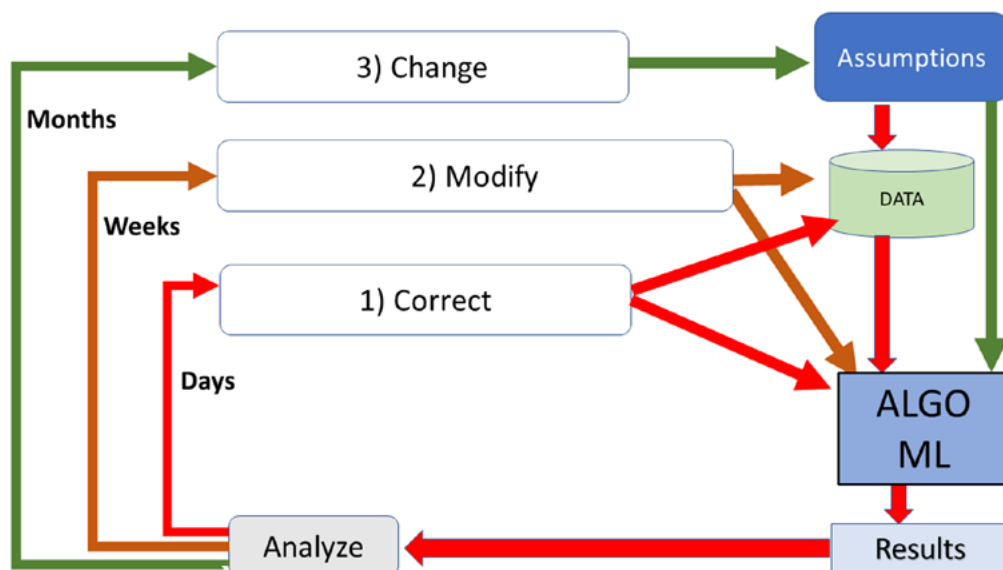


Figure 10. Principle of the triple loop learning in AI.

tests, the algorithm highlighted the unsuitability of the ‘frequent’ output class which corresponds to repeated occupations that can be confused with short. These first tests led to the conclusion that on the question of the occupation duration, the potentially identifiable results are: ‘short’ and/or ‘long’, ‘short’ or ‘not short’, or ‘long’ or ‘not long’ and the third output class was dropped to optimize learning and predictions. This approach led researchers to question the causal links that are usually used in their scientific reasoning and to identify new ones.

3. Discussion and conclusion

These first predictions demonstrated the potential of the Machine Learning algorithms both on environmental issues (exogenous method) and on behavioral ones (endogenous method). Their contribution is different depending on the type of hypothesis tested and the available repositories. This methodology requires a rigorous step of data preparation and the scientific care that must be provided on it. It requires the use of a new working method that combines the conceptual approach (conceptual model) with the analysis of data (Machine Learning).

The exogenous method allows to obtain usable results able to be published if they are interpreted and put in perspective as a part of the treated problematic. The example of biome reconstructions using actuals faunal repositories shows that it is possible to predict for the Caune de l’Arago archeostratigraphical layers the environment around the cave and consequently enable to reconstruct past landscapes and environmental and climatic conditions. These kinds of results can be reached thanks to other more specialized palaeological approaches for example dietary and ecological niche reconstitution (Rivals, Lister 2016) or studies based on separate different communities (Stoetzel, Montuire 2016; Magniez, Boulbes 2014; Hanquet *et al.* 2018 and references therein). The ML approach offers the possibility to generate automatically and quickly with the support of a robust frame of reference taking into account thousands of global data from different vertebrates communities. In this case the results are obtained without having recours of the expert labeling and strictly from a calculation and can be considered as more objectives. Moreover, once the training data set is done, it’s possible to test many archeological faunal spectra, from different sites and periods and obtain automatised predictions with systematic explanations. In order to enrich the results, the prediction can be done with combining faunal proxies and others environmental ones, like pollen or geological variables. This method can deal with a broad spectrum of environmental issues with training and modeling on Biomes, Ecoregions, Koppen climate classes... The repositories could also be improved by others Pleistocene contemporaneous available datas.

The endogenous method, using ‘Expert labeling’ allows the algorithm to compare the values of the Expert labels together and to explain the indicators that contributed the most to the algorithm classification and more implicitly to the Expert labeling. It is then possible to return to these indicators to study the particular role they played in the decision-making. In this method, the explanation phase is the most important benefit of the prediction rather than the result of the prediction itself. The identification of the variable weights help the archeologists to better organise his scientific reasoning and hypothesis and give strength to the future predictions. The prediction brings raw results unusable directly but interesting to highlight the significance and the importance of the data in resolving a given question. The successive iterations, enriched at each stage by an improvement of the dataset, taking into account the previous results, progressively allows to obtain a robust prediction. Above all, this iterative work process provide at each step the necessary explanation to demonstrate the path followed to reach the result. At this stage, this method provides the most important short-term development perspectives. For the occupation duration, new tests are in progress by Expert labeling with the result of tooth ungulate microwears (Rivals *et al.* 2015) and new datasets are prepared on a given great concept like ‘site function’, combining archeological repository (compilation of bibliographic data) and Expert labeling on a hybrid ‘exogenous-endogenous’ method.

References

- Argyrys, C. and Schon, D.A. 1996. *Organizational Learning: A Theory of Action Perspective*, Reading, Mass.: Addison Wesley, second edition, 344 p.
- Arriaza, M.C. and Domínguez-Rodrigo, M. 2016. When felids and hominins ruled at Olduvai Gorge. A machine learning analysis of the skeletal profiles of the non-anthropogenic Bed I sites. *Quaternary Science Reviews* 139: 43-52.
- Audouze, F. 2013. Les bases de données de Lewis R. Binford accessibles sur le serveur de la Maison de l'archéologie et de l'ethnologie René-Ginouvès. *Bulletin de la Société préhistorique française* 110(2): 353-355.
- Aura, J.E., Bonilla, V.V., Ripoll, M.P., Valle, R.M. and Calatayud, P.G. 2002. Big Game and Small Prey: Paleolithic and Epipaleolithic Economy from Valencia (Spain). *Journal of Archaeological Method and Theory* 9: 215-268.
- Barsky, D. 2013. The Caune de l'Arago stone industries in their stratigraphical context. *Comptes Rendus Palevol* 12(5): 305-325.
- Barsky, D., Grégoire, S. and Moigne, A.M. 2005. *Variabilité des types d'occupation et d'exploitation de territoires méditerranéens entre 600.000 ans et 300.000 ans*, in M.-H. Moncel and J.L. Monnier (eds) *Données récentes sur les premiers peuplements en Europe. Actes du colloque international, (Rennes 2003)*: 565-576. British Archeological Reports International Series 1364, Oxford: Archaeopress.
- Betts, M.W. and Friesen, T.M. 2004. Quantifying hunter-gatherer intensification: a zooarchaeological case study from Arctic Canada. *Journal of Anthropological Archaeology* 23: 357-384.
- Bicho, N. and Cascalheira, J. 2018. The use of lithic assemblages for the definition of short-term occupations in hunter-gatherer prehistory, in A. Picin and J. Cascalheira (eds) *Short-term Occupations in Paleolithic Archaeology. Interdisciplinary Contributions to Archaeology*. Springer.
- Binford, L.R. 2001. *Constructing Frames of Reference: an Analytical Method for Archaeological Theory Building using Hunter-Gatherer and Environmental Data Sets*. Berkeley: University of California Press, 563 p.
- Binford, L.R. 1981. *Bones: Ancient men and modern myths*. I ed., New York: Academic Press.
- Bedford, D.S. and Spekle, R.F. 2017. *Construct Validity in Survey-Based Management Accounting and Control Research*, Management Accounting Section (MAS) Meeting. <<https://ssrn.com/abstract=3011357>>.
- Briz i Godino, I., Ahedo, V., Alvarez, M., Pal, N., Turnes, L., Santos, J.I., Zurro, D., Caro, J. and Galan, J.M. 2018. Hunter-gatherer mobility and technological landscapes in southernmost South America: a statistical learning approach. *Royal Society Open Science* 5: 180906.
- Brugal, J.P. and Patou-Mathis, M. 1993. L'assemblage osseux de l'abri des Canalettes, in L. Meignen (ed.) *L'abri des Canalettes. Un habitat moustérien sur les grands Causses (Nant, Aveyron. Fouilles 1980-1986)*: 77-87. Paris : Éd. du CNRS 10.
- Byeon, W., Domínguez-Rodrigo, M., Arampatzis, G., Baquedano, E., Yravedra, J., Maté-González, M.A. and Koumoutsakos, P. 2019. Automated identification and deep classification of cut marks on bones and its paleoanthropological implications. *Journal of Computational Science* 32: 36-43.
- Chacon, M.G., Bargallo, A., Gabucio, M.J., Rivals, F. and Vaquero, M. 2015. *Chapter 12. Neanderthal behaviors from a spatio-temporal perspective: an interdisciplinary approach to interpret archaeological assemblages*, in N.J. Conard and A. Delagnes (eds) *Settlement Dynamics of the Middle Paleolithic and Middle Stone Age*. IV: 253-294. Tübingen: Tübingen Publications in Prehistory.
- Clark, G.A. and Barton, C.M. 2017. Lithics, landscapes & la Longue-durée – Curation & expediency as expressions of forager mobility. *Quaternary International* 450: 137-149.
- Cossette, P. 2008. La cartographie cognitive vue d'une perspective subjectiviste: mise à l'épreuve d'une nouvelle approche, *M@n@gement* 11(3): 259-281.
- Costamagno, S., Meignen, L., Beauval, C., Vandermeersch, B. and Maureille, B. 2006. Les Pradelles (Marillac-le-Franc, France): A Mousterian reindeer hunting camp? *Journal of Anthropological Archaeology* 25: 466-484.
- Daujeard, C., Moncel, M.H., Rivals, F., Fernandez, P., Aureli, D., Auguste, P., Bocherens, H., Crégut-Bonnoure, E., Debard, E. and Liouville, M. 2011. *Quel type d'occupation dans l'ensemble F de Payre (Ardèche, France)? Halte de chasse spécialisée ou campement de courte durée? Un exemple d'approche*

- multidisciplinaire, in F. Bon, S. Costamagno and N. Valdeyron (eds) *Haltes de chasse en Préhistoire. Quelles réalités archéologiques? Actes du colloque international (Toulouse, 13-15 mai 2009)*. *P@lethnologie* 3: 77-101.
- Daujeard, C., Moncel, M.H. 2010. On Neanderthal subsistence strategies and land use: A regional focus on the Rhone Valley area in southeastern France. *Journal of Anthropological Archaeology* 29: 368-391.
- David, F. and Enloe, J.G. 1993. L'exploitation des animaux sauvages de la fin du Paléolithique moyen au Magdalénien, in *Exploitation des animaux sauvages à travers le temps, XIIIes Rencontres d'Archéologie et d'Histoire d'Antibes, IVe colloque international de l'Homme et l'animal, Société de recherche interdisciplinaire (Antibes-Juan-les-Pins, 15-17 oct. 1992)*: 29-46. Antibes-Juan-les-Pins: Éd. APDCA.
- Desclaux, E. 1992. Les petits vertébrés de la Caune de l'Arago à Tautavel (Pyrénées-Orientales). Biostratigraphie, paléoécologie et taphonomie. *Bulletin du Musée d'Anthropologie Préhistorique de Monaco* 35: 35-64.
- Dibble, H. 1995. Middle paleolithic scraper reduction: background, clarification and review of the literature to date. *Journal of Archaeological Method and Theory* 2: 299-36.
- Domínguez-Rodrigo, M. and Baquedano, E. 2018. Distinguishing butchery cut marks from crocodile bite marks through machine learning methods. *Scientific Reports* 8(1): 5786.
- Egeland, C.P., Domínguez-Rodrigo, M., Rayne Pickering, T., Menter, C.G. and Heaton, J.L. 2018. Hominin skeletal part abundances and claims of deliberate disposal of corpses in the Middle Pleistocene, *PNAS* 115(18): 4601-4606.
- Falguères, C., Shao, Q., Han, F., Bahain, J.-J., Richard, M., Perrenoud, C., Moigne, A.-M. and Lumley, de H. 2015. New ESR and U-series dating at Caune de l'Arago, France: A key-site for European Middle Pleistocene. *Quaternary Geochronology* 30: 547-553.
- Falguères, C., Yokohama, Y., Shen, G., Bischoff, J.L., Ku, T.L. and Lumley, H. 2004. New U-series dates at the Caune de l'Arago, France. *Journal of Archaeological Science* 31: 941-952.
- Friedman, J. 2000. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29.
- Geurts, P., Ernst, D. and Wehenkel, L. 2006. Extremely randomized trees. *Machine learning* 63(1): 3-42.
- Grégoire, S., Moigne, A.M., Barsky, D. and Lumley, de H. 2007. Gestion et sélection des ressources au sein d'un territoire. Un exemple de comportement économique au Paléolithique inférieur dans le sud de la France: 1725-1727, BAR International Series 1364.
- Grégoire, S., Barsky, D. and Byrne, L. 2006. The Caune de l'Arago: an example of middle Pleistocene flint exploitation (Tautavel, Pyrénées-Orientales, France), in *Stone Age, Mining Age. Der Anschnitt, Beiheft 19, 2005-VIII International Flint Symposium (Bochum, 13-17 September 1999)*: 99-113. Bochum: Deutsches Bergbau-Museum.
- Hanquet, C., Stoetzel, E., Desclaux, E. and Bailon, S. 2018. Etat de la recherche sur les micromammifères et l'herpétofaune quaternaires en France, in *La Préhistoire de la France*: 99-102. Paris: Hermann.
- Hanquet, C. and Desclaux, E. 2011. Analyse paléoécologique des communautés de micromammifères de la Caune de l'Arago (Tautavel, France) dans le contexte des migrations de faunes en Europe méridionale au cours du Pléistocène moyen. *Quaternaire* 22(1): 35-45.
- Lartigot, A.S. 2007. *Taphonomie pollinique en grotte de sédiments détritiques et de spéléothèmes: Potentiels et limites pour la reconstitution de l'environnement végétal de l'homme préhistorique sur le pourtour nord-ouest méditerranéen. Application aux sites de la Caune de l'Arago (Tautavel, Pyrénées-Orientales), de la Baume Bonne (Quinson, Alpes-de-haute-Provence), de la grotte du Lazaret (Nice, Alpes-Maritimes) et de la grotte italienne de la Basura (Toirano, Ligurie)*. Thèse de doctorat, MNHN, Paris, 545 p.
- Lartigot-Campin, A.-S. et al. 2019. Prediction of past biomes from palynological data in archaeological context using Machine Learning: the sequence of Caune de l'Arago (Middle Pleistocene, South of France). In preparation.
- Lebreton, L., Desclaux, E., Hanquet, C., Cuenca-Besco, G., Moigne, A.M., Perrenoud, C. and Grégoire, S. 2017. Variations paléoenvironnementales au sein de l'Unité Archéostratigraphique G (UA G) de la Caune de l'Arago (Tautavel, France). Apport des paléocommunautés de rongeurs. *Quaternaire* 28(3): 313-321.

- Lebreton, L., Desclaux, E., Hanquet, C., Moigne, A.M. and Perrenoud, C. 2016. Environmental context of the Caune de l'Arago Acheulean occupations (Tautavel, France). New insights from micro vertebrates in Q-R levels. *Quaternary International* 411: 182-192.
- Leonova, N.B. 1993. *Criteria for estimating the duration of occupation at Paleolithic sites. An example from Kamennaia Balka II'*, in O. Soffer and N.D. Praslov (eds) *From Kostienki to Clovis. Upper Palaeolithic-Paleo-Indians Adaptations*: 149-157. New York: Plenum Press.
- Lumley, de H., Fontaneil, C., Grégoire, S., Batalla, G., Caumon, G., Celiberti, V., Chevalier, T., Deguillaume, S., Fournier, A., Lumley, de M.A., Magniez, P., Moigne, A.M., Notter, O., Perrenoud, C., Pois, V., Pollet, G. and Testu, A. 2015. *Caune de l'Arago Tome VI. Tautavel-en-Roussillon, Pyrénées-Orientales, France: Individualisation des unités archéostratigraphiques*. Paris: Éd. du CNRS, 641 p.
- Lumley, de H. (ed.) 2014. *La Caune de l'Arago Tome I, Tautavel-en-Roussillon, Pyrénées-Orientales, France*. Paris: Éd. du CNRS, 432 p.
- Lumley, de H., Grégoire, S., Barsky, D., Batalla, G., Bailon, S., Belda, V., Briki, D., Byrne, L., Desclaux, K., El Guenouni, K., Fournier, A., Kacimi, S., Lacombat, F., Lumley, de M.-A., Moigne, A.M., Moutoussamy, J., Paunescu, C., Perrenoud, C., Pois, V., Quilès, J., Rivals, F., Roger, T. and Testu, A. 2004. Habitat et mode de vie des chasseurs paléolithiques de la Caune de l'Arago (600,000-400,000 ans). *L'Anthropologie* 108: 159-184.
- Lundberg, S.M. and Lee, S.I. 2017. A unified approach to interpreting model predictions, in *31st Conference on Neural Information Processing Systems (NIPS 2017)*. Long Beach, CA, USA.
- Lyman, R.L. 1994. *Vertebrate Taphonomy*. Cambridge: Cambridge University Press.
- Lyman, R.L. 1989. Taphonomy of Cervids Killed by the 18 May 1980 Volcanic Eruption of Mount St. Helens, Washington, in R. Bonnicksen and M.H. Sorg (eds) *Bone Modification*: 149-167. Orono: Center for the Study of the First Americans, University of Maine.
- MacKenzie, S.B., Podsakoff, P.M. and Jarvis, C.B. 2005. The Problem of Measurement Model Mis-specification in Behavioral and Organizational Research and Some Recommended Solutions. *Journal of Applied Psychology* 90(4): 710-730.
- Magniez, P. 2010. *Etude paléontologique des Artiodactyles de la grotte Tournal. Etude taphonomique, archéozoologique et paléoécologique des grands Mammifères dans leur cadre biostratigraphique et paléoenvironnemental*. Thèse de Doctorat, Perpignan: Université de Perpignan, 916 p.
- Magniez, P., Moigne, A.M. and Lumley, de H. 2011. First reindeer exploitation by Acheulean people in South of France: Case study of the Lower Palaeolithic site of the Caune de l'Arago (Tautavel, Pyrénées-Orientales, France), in *Congrès International 'Deer and People' (Lincoln, UK, 10 September 2011)*. Abstract volume: 40.
- Magniez, P., Moigne, A.M., Testu, A. and Lumley, de H. 2013. Biochronologie des Mammifères quaternaires. Apport des Cervidae du site Pléistocène moyen de la Caune de l'Arago (Tautavel, Pyrénées-orientales, France). *Quaternaire* 24(4): 477-502.
- Magniez, P. and Boulbes, N. 2014. Environment during the Middle to Late Palaeolithic transition in southern France: The archaeological sequence of Tournal cave (Bize-Minervois, France). *Quaternary International* 337: 43-63.
- Manzano, A. 2015. *Les amphibiens et les reptiles des sites du Pléistocène moyen et supérieur du pourtour méditerranéen (Caune de l'Arago, grotte du Lazaret, Baume Moula-Guercy). Etude d'herpétofaunes et reconstitutions paléoclimatiques et paléoenvironnementales*. Thèse de Doctorat, Perpignan: Université de Perpignan, 560 p.
- Moigne, A.M. 1983. *Taphonomie des faunes quaternaires de la Caune de l'Arago, Tautavel*. Thèse 3ème cycle, Paris VI/Museum National d'Histoire Naturelle, 342 p.
- Moigne, A.M., Palombo, M.R., Belda, V., Heriech-Briki, D., Kacimi, S., Lacombat, F., Lumley, de M.A., Moutoussamy, J., Rivals, F., Quiles, J. and Testu, A. 2006. Les faunes de grands mammifères de la Caune de l'Arago (Tautavel) dans le cadre biochronologique des faunes du Pléistocène moyen italien. *L'Anthropologie* 110(5): 788-831.
- Moigne, A.M., Grégoire, S. and Lumley, de H. 2005. Les territoires de chasse et d'exploitation des matières premières des hommes préhistoriques de la Caune de l'Arago entre 600,000 ans et 400,000 ans, in J. Jaubert and M. Barbaza (eds) *Territoires, déplacements, mobilité, échanges durant la préhistoire. Actes des 126e Congrès national des sociétés historiques et scientifiques (Toulouse, 2001)*. Paris: Ed. du CTHS.

- Moncel, M.H. and Rivals, F. 2011. On the question of short-term Neanderthal site occupations: Payre, France (MIS 8-7), and Taubach/Weimar, Germany (MIS 5). *Journal of Anthropological Research* 67(1): 47-75.
- Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'Amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P. and Kassem, K.R. 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* 51(11): 935-938.
- Pedregosa, F. et al. 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12: 2825-2830.
- Perrenoud, C., Falguères, C., Moigne, A.M., Testu, A., Magniez, P., Boulbes, N., Lebreton, L., Hanquet, C., Desclaux, E., Lartigot-Campin, A.S., Celiberti, V., Grégoire, S., Lumley, de H., Viallet, C., Chevalier, T., Lumley, de M.A., Viallet, A., Fontaneil, C., Pois, V., Hu, H.M., Shen, C.C. and Michel, V. 2016. Diversité des occupations humaines en contexte glaciaire à la Caune de l'Arago, in *Colloque AFEQ CNF-INQUA Q10 (Bordeaux, 16-18 février 2016)*. Bordeaux.
- Petrelli, M., Bizzarri, R., Morgavi, D., Baldanza, A. and Perugini, D. 2017. Combining machine learning techniques, microanalyses and large geochemical datasets for tephrochronological studies in complex volcanic areas: New age constraints for the Pleistocene magmatism of central Italy. *Quaternary Geochronology* 40: 33-44.
- Provitolo, D., Dubos-Paillard, E., Verdière, N., Lanza, V., Charrier, R., Bertelle, C. and Aziz-Alaoui, M.A. 2015. Les comportements humains en situation de catastrophe: de l'observation à la modélisation conceptuelle et mathématique. *Cybergeog: European Journal of Geography, Systèmes, Modélisation, Géostatistiques*, document 735. <<http://journals.openedition.org/cybergeog/27150>> [Accessed October 26, 2019].
- Renault-Miskovsky, J. 1981. Etude palynologique du remplissage de la Caune de l'Arago à Tautavel. Signification chronologique, paléoclimatique et palethnographique des flores, in H. de Lumley and J. Labeyrie (eds) *Datations absolues et analyses isotopiques en Préhistoire-Méthodes et limites*: 253-258. Paris: CNRS, Prétirage.
- Rendu, W., Bourguignon, L., Costamagno, S., Meignen, L., Soulier, M.-C., Armand, D., Beauval, C., David, F., Griggo, C., Jaubert, J., Maureille, B. and Park, S.J. 2011. Approche interdisciplinaire et réflexions méthodologiques sur la question des haltes de chasse au Moustérien, in Fr. Bon, S. Costamagno and N. Valdeyron (eds) *Haltes de chasse en Préhistoire. Quelles réalités archéologiques?, Actes du colloque international (Toulouse 13-15 mai 2009)*. *P@lethnologie* 3: 61-76.
- Rivals, F. and Lister, A.M. 2016. Dietary flexibility and niche partitioning of large herbivores through the Pleistocene of Britain. *Quaternary Science Reviews* 146: 116-133.
- Rivals, F., Prignano, L., Semperebon, G.M., Lozano, S. 2015. A tool for determining duration of mortality events in archaeological assemblages using extant ungulate microwear. *Scientific Reports* 5: 17330.
- Rusch, L., Grégoire, S., Pois, V. and Moigne, A.M. (accepted). Neanderthal and Carnivore Occupations in Unit II from the Upper Pleistocene Site of Ramandils Cave, (Port-la-Nouvelle, Aude, France). *Journal of Archaeological Science: Reports* 102038 PII S2352-409X(19)30374-8.
- Shott, M.J. 1996. An exegesis of the curation concept. *Journal of Anthropological Research* 52: 259-280.
- Sobol, M.K. and Finkelstein, S.A. 2018. Predictive pollen-based biome modeling using machine learning. *PLoS ONE* 13(8): e0202214.
- Sobol, M.K., Scott, L. and Finkelstein, S.A. 2019. Reconstructing past biomes states using machine learning and modern pollen assemblages: A case study from Southern Africa. *Quaternary Science Reviews* 212: 1-17.
- Stiner, M.C. 1991. Food procurement and transport by human and non-human predators. *Journal of Archaeological Science* 18: 455-482.
- Stoetzel, E. and Montuire, S. 2016. Les rongeurs, indicateurs des paléoclimats: application aux assemblages de trois sites du nord de la France. *Quaternaire* 27(3): 227-238.
- Sullivan, A. 1992. Investigating the Archaeological Consequences of Short-Duration Occupations. *American Antiquity* 57(1): 99-115.
- Toniolo, A.M. 2009. Le comportement: entre perception et action, un concept à réhabiliter. *L'Année psychologique* 2009/1, 109: 155-193.

- Valensi, P. 2000. The archaeozoology of Lazaret Cave (Nice, France). *Interntional Jorrnal of Osteoarchaeology* 10: 357-367.
- Van der Maaten, L.J.P. and Hinton, G.E. 2008. Visualizing High-Dimensional Data Using t-SNE, *Journal of Machine Learning Research* 9: 2579-2605.
- Wang, J.Z., Ge, W., Snow, D.R., Mitra, P. and Giles, C.L. 2010. Determining the Sexual Identities of Prehistoric Cave Artists Using Digitized Handprints: A Machine Learning Approach, in *Proceedings of the ACM International Conference on Multimedia (Florence, Italy, ACM, October 2010)*: 1325-1332.
- World Wildlife Fund 2006. WildFinder: Online database of species distributions. <www.worldwildlife.org/WildFinder>[ver. Jan-06].
- Zliobaite, I. 2019. Concept drift over geological times: predictive modeling baselines for analyzing the mammalian fossil record. *Data mining and knowledge discovery* 33(3): 773-803.

Cagny-l'Épinette (Somme Valley, France), Thirty Years of Mixed Data: Potential and Limits

Floriane Peudon¹, Éric Masson², Patrick Auguste³, Agnès Lamotte¹,
Anne-Marie Moigne⁴, Alain Tuffreau¹

¹ HALMA – UMR 8164, CNRS, University of Lille

² SHS – Laboratory Territoires, Villes, Environnement et Société (TVES) EA 4477,
University of Lille

³ UMR 8198, CNRS – Unit EVO ECO PALEO, University of Lille

⁴ UMR 7194, CNRS, Centre Européen de Recherches Préhistoriques de Tautavel

floriane.peudon@gmail.com (corresponding author)

eric.masson@univ-lille.fr

patrick.auguste@univ-lille.fr

agnes.lamotte@univ-lille.fr

anne-marie.moigne@cerptautavel.com

alain.tuffreau@univ-lille1.fr

Abstract

The Acheulean site of Cagny-l'Épinette (Somme Valley, France) has been excavated for thirty years, from 1980 to 2010, a period during which archaeological data recording and management were concerned by a gradual computerization. The constant evolution of practices has yielded a large amount of analogue and digital archaeological documents. Through the instance of the site of Cagny-l'Épinette, this paper proposes a method to manage, consolidate and analyse diverse, sometimes incomplete, data, collected through different survey protocols. For the very palaeo-ethnological purpose of spatially identifying activities in the Acheulean levels, a methodology combining a Geographic Information System software (ArcGIS) with an Object Based Image Analysis software (eCognition®) has been applied to the comprehensive catalogue of available documents. This research constitutes the first known application of the (Archae)OBIA methodology on archaeological handwritten archives such as planimetric field drawings of remains.

Keywords: Digital Archaeology, Field Drawings, Object Based Image Analysis, GIS, Cagny-l'Épinette

Résumé

Le site acheuléen de Cagny-l'Épinette (Vallée de la Somme, France) a été fouillé pendant trente ans de 1980 à 2010, période durant laquelle l'archéologie s'est vue informatisée concernant le protocole d'enregistrement et de traitement des données. La constante évolution des pratiques a produit une quantité conséquente de documents, manuscrits et numériques, archéologiques. À travers l'exemple du site de Cagny-l'Épinette, cet article propose une pensée méthodologique concernant les possibilités de gestion, d'enrichissement et d'exploitation de données aussi diversifiées, parfois incomplètes, obtenues par des protocoles différents. Dans la finalité ultime paléothnologique d'identification spatiale d'activités dans les niveaux acheuléens, une méthodologie combinant un logiciel Système d'Information Géographique (ArcGIS) avec un logiciel d'Analyse d'Image Orientée Objet (eCognition®) a été appliquée à l'intégralité des données disponibles. Ce travail constitue la première application de la méthodologie OBIA sur des archives manuscrites telles que les relevés planimétriques de terrain de vestiges.

Mots-clés : Archéologie numérique, relevés de terrain, analyse d'image, SIG, Cagny-l'Épinette

Introduction

The site of Cagny-l'Épinette belongs to the fourth alluvial terrace of the Somme Valley in Northern France (Figure 1) (Commont 1911; Bourdier *et al.* 1974; Haesaerts *et al.* 1984; Sommé *et al.* 1984;

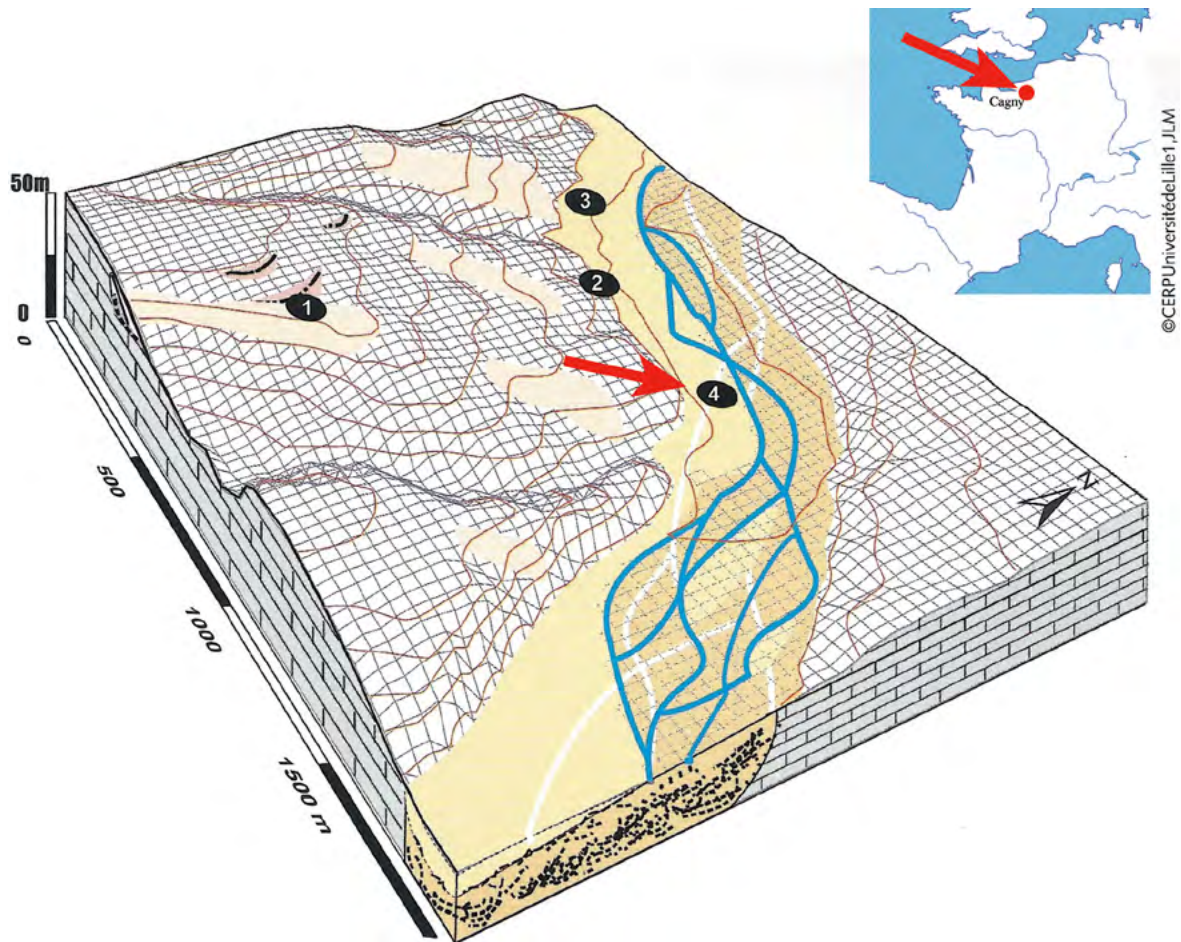


Figure 1. 3D-Model of the location of the site of Cagny-l'Épinette, Somme Valley, Northern France (Map: Modified after d-map.com; 3D-Model: Modified after Marcy in Tuffreau *et al.* (2001)).

Antoine 1990, 2010). It has provided numerous archaeological levels in the alluvial deposits and in the loess cover (Antoine in Tuffreau *et al.* 1995). The levels I, I0, I1, I2 and J include Acheulean assemblages embedded in the alluvium (MIS 9 and MIS 10) with lithic industries and very well preserved faunal remains due to the calcareous silts while the Middle Palaeolithic industries from the levels C1, C2, E1, E2, F, G and H belong to the loess cover (from MIS 5e to MIS 8) (Haesaerts *et al.* 1984; Balescu 1987; Balescu *et al.* 1988; Bates 1993; Laurent *et al.* 1994; Tuffreau *et al.* 1995; Balescu and Tuffreau 2004).

From 1980, the year of the first test trenches, to the last planned campaign in 2010, this site has been excavated for a period of thirty years, during which the field of archaeology was subject to the emergence of computerization of data recording and management. Such an adjustment in the protocol has yielded a large amount of archaeological data, both manual and digital, consisting of manual field drawings (planimetric maps of the vestiges and stratigraphic profiles), film and digital photography, manual coordinate measurements and Total Station recordings to which the databases of the lithicians and archaeozoologists are added (Tuffreau 1989; Tuffreau *et al.* 1995; Tuffreau 2010).

Given the abundance of well preserved faunal remains, the site of Cagny-l'Épinette has been chosen to carry out spatial analyses in the context of a Ph.D. work to understand human activities in the period around 300 ka. A specific methodology involving the combination of a Geographic Information System software (ArcGIS) with an Object Based Image Analysis software (eCognition®)

has been applied to optimize the process of gathering, homogenizing and enriching the data and link the spatial information with the qualitative data. In recent years, as detailed in the review of Davis (2018), the OBIA methodology has mainly been used in archaeology in landscape studies on remote-sensing imageries such as LiDAR, for diverse purposes (Verhagen and Drăguț 2012; De Guio *et al.* 2015; Cerrillo-Cuenca 2017), but it has also been applied to surface magnetic surveys imagery (Pregeßbauer *et al.* 2014). Furthermore, OBIA has been of great assistance in automating geometrical and spectral measurements from photos, 3D-models or even thin sections of archaeological objects (Loriot *et al.* 2007; Hofmann *et al.* 2011; Lamotte and Masson 2016; Masson and Lamotte 2018). This methodological success led Lamotte and Masson (2016) to propose the concept of ArchaeOBIA. However, the (Archae)OBIA methodology has never been used on archaeological handwritten archives such as planimetric field mapping of lithic or faunal remains, making the protocol used on the site of Cagny-l'Épinette its first known application of this kind.

1. A thirty year long excavation: current state of analogue and digital data

1.1. Thirty years of data recording and management

Given the open air nature of the site, an open-area type excavation was conducted at Cagny-l'Épinette, following progressive one-meter wide strips in a metric grid (Tuffreau *et al.* 1995; Tuffreau 2010). Like in all excavations, once the archaeological remains were unearthed, a specific protocol was applied to record the information before removing and packing them (Figure 2). This protocol included vertical photo coverage, field mapping and measurement of the coordinates (x, y, z). The vertical photo coverage was recorded for each square meter by film photography in the 1980s and then by digital photography from the beginning of the 1990s (Figure 2A). Regarding the field drawings, one-tenth scale field maps of the remains were composed either on graph paper (1980-1996, 2000-2002) or directly on the vertical photographs with a clean twin drawing on tracing paper (1997-1999) depending on the time period (Figure 2C). This field drawing protocol was sustained per square meter until 2002 and was then discontinued from 2003 to 2010. The Cartesian coordinates (x, y) and the altitudes (z) were first manually measured according to an origin specific to each square meter within a metric grid in an orthonormal coordinate system and a base altitude (Figure 2B). Each vestige belonging to the time period 1980-1990 is thus spatially documented by one set of coordinates (x, y, z) measured at its centre.

In 1991 the excavation started being computerized with the introduction of a Leica Electronic Theodolite in the field as part of a collaboration with an American team from the Department of Anthropology of the University of Pennsylvania (Dibble and McPherron in Tuffreau *et al.* 1995; Dibble *et al.* 1997). Since then, data acquisition has been carried out according to a global metric grid. The process of measurement was henceforth changed for all the remains with a longitudinal axis with three sets of coordinates (x, y, z) obtained at each tip and at the bottom of the vestige. For compact shaped remains only one set of coordinates was measured at the bottom (Tuffreau *et al.* 1995).

1.2. Thirty years of specialized studies: on the way to forty

Several specialists have contributed to the multidisciplinary study of the site which includes field data recording, quaternary deposit studies and dating, bioclimatic studies, archaeological vestige studies and computer data analyses. The field data recording and management, manually organized in the 1980s, became the main goal of a French-American project started in 1991, through which they were integrated in a dual Protocol: measurement computerization and manual drawings (Dibble and McPherron in Tuffreau *et al.* 1995; Dibble *et al.* 1997). The quaternary deposits have been subject to chronostratigraphic and micromorphological analyses to understand the formation mechanisms of the alluvial terraces along with the local neotectonic features specific to the site (Antoine 1990; Van Vliet-Lanoë 1989; Van Vliet-Lanoë *et al.* 2001). Furthermore, three

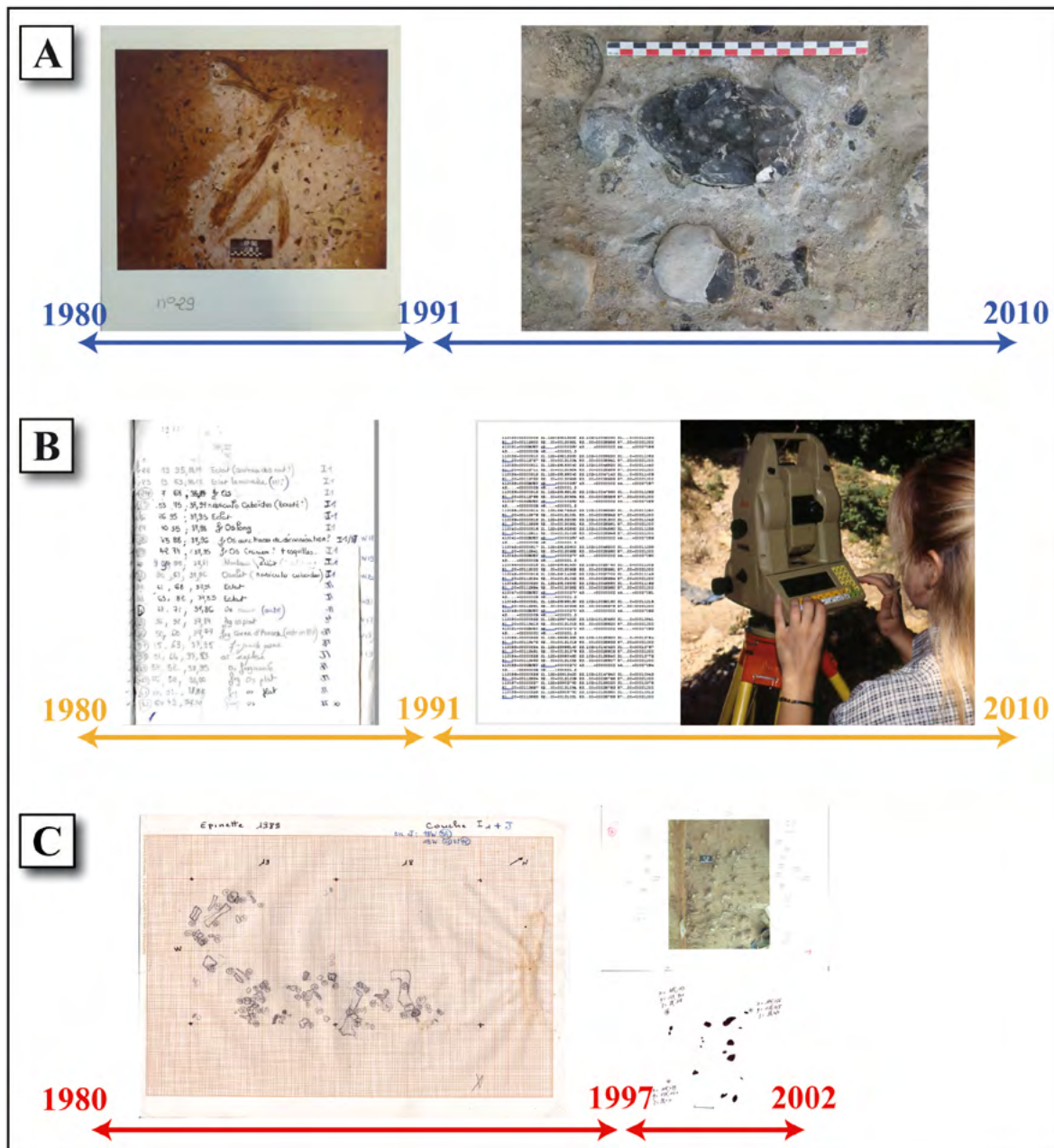


Figure 2. Field recording protocol applied at the site of Cagny-l'Épinette from 1980 to 2010. A – Photo coverage by film photography in the 1980s, then by digital photography, B – Manual coordinate measurements in the 1980s, then replaced with a Total Station, C – One-tenth scale drawings on graph paper or vertical photographs (Field Photos: A. Tuffreau; Field drawings & inventories: Archives; Scan & CAD: A. Lamotte, F. Peudon).

different dating methods have been applied at Cagny-l'Épinette on either the loess cover (TL Dating on sediments and ESR/U-Th Method on horses' teeth) and/or the fine grained silts (ESR on bleached quartz and ESR/U-Th Method on horses' teeth) (Balescu 1987; Laurent *et al.* 1994; Bahain, Falguères in Tuffreau 2010). Regarding the bioclimatic approach, palynological analyses have been carried out on the alluvium (Levels I and I1) and at the bottom of the loess cover (Level H) while studies have focused on micromammals and malacological remains of the same levels (Munaut 1989; Cordy 1989; Van Kolfschoten in Tuffreau *et al.* 1995; Rousseau in Tuffreau 1989; Limondin-Lozouet in Tuffreau 2010).

These results have been complemented by a palaeontological study of the large mammal bones from the levels belonging to the coarse silts, fine alluvium and upper loess sediments (Moigne in Tuffreau *et al.* 1995; Auguste in Tuffreau 2004). The archaeological and archaeozoological studies concerning the lithic and faunal remains have taken place in the established chronostratigraphic and bioclimatic frame (Moigne 1989; Lamotte and Tuffreau 2001; Auguste in Tuffreau 2010). They have consisted of expert lithic typological and faunal anatomical assessments together with usewear and anthropogenic mark observations. This process has been finalized in the context of this Ph.D. work through a review of the faunal collections aiming to research new specific functional marks (Moigne 1989; Moigne and Chase in Tuffreau *et al.* 1995; Léopold 1997; Lamotte and Tuffreau 2001; Auguste in Tuffreau 2010). Finally, computer data analyses have been progressively performed since the middle of the 1980s to apprehend some taphonomical aspects and the palethnological features of the site (Bouchet in Tuffreau 1989; Dibble and McPherron in Tuffreau *et al.* 1995; Dibble *et al.* 1997; Marcy and Lamotte in Tuffreau 2010).

In thirty years of excavation, these analyses had always focused on some distinct sections of the data set; all of the available documents – manuscripts and digital – having never been gathered. For this reason, 25 square meters were chosen in the frame of a Master's Dissertation in 2013, first attempt of elaboration of a consistent methodological protocol aiming to gather and analyse the data of a specific section of the site, whose results have been integrated in the overall view of the site (Peudon 2013; Lamotte *et al.* 2014). The site of Cagny-l'Épinette is currently the subject of a Ph.D. work, initiated in 2016, based on two main directions. The first one, a methodological issue, concerns the application of a single optimised and consistent protocol of data management, leading to appropriate data for spatial analyses (Peudon *et al.* 2018a). This question has been considered to be useful to other archaeological sites with complex and diversified documentation obtained over a long period of time. The second one, a taphonomic and palethnological issue, aims to provide some results about the settlement organisation of a camp around 300 ka through quantitative geostatistical analyses (Peudon *et al.* 2018b).

2. Data Processing

2.1. Materials and overall methodological protocol

These studies, together with the evolution of the recording protocol in the field, have yielded very abundant and diverse archives. The field recordings include annual stratigraphic profiles and hundreds of manual planimetric drawings of the vestiges for the entire period of excavation. Film or digital photos, with overall and detailed shots, have systematically been taken of the excavated square meters and the profiles. The manual coordinate measurements have been registered in annual handwritten field books and/or retyped in an Excel spreadsheet, while the original yearly Total Station recordings have been saved in Excel or Word format and partially gathered in an Excel spreadsheet. The results of the lithic and archaeozoological typological studies of the archaeological material are archived in several handwritten books and/or typewritten in Excel spreadsheets. The aim of this Ph.D. work is then to prepare these diversified documents, in other terms to gather, homogenize and enrich them, so that they can be exploited through spatial analyses. A two-fold methodological approach has been chosen here to obtain two final kinds of data, linked to each other, to which quantitative analyses could be applied. One is a pair of two final databases, for the lithic and faunal remains respectively, in which all the attributes, that is, typological, geometrical and spatial features, are gathered. The other is the digitized and georeferenced maps of the vestiges merged into one overall map for each archaeological level.

The protocol to obtain said data involves several steps, the first of which have now become a standard protocol in archaeology (Figure 3). Regarding the databases, all the available data (analogue and digital) have been manually gathered and homogenized in two Excel files. As for

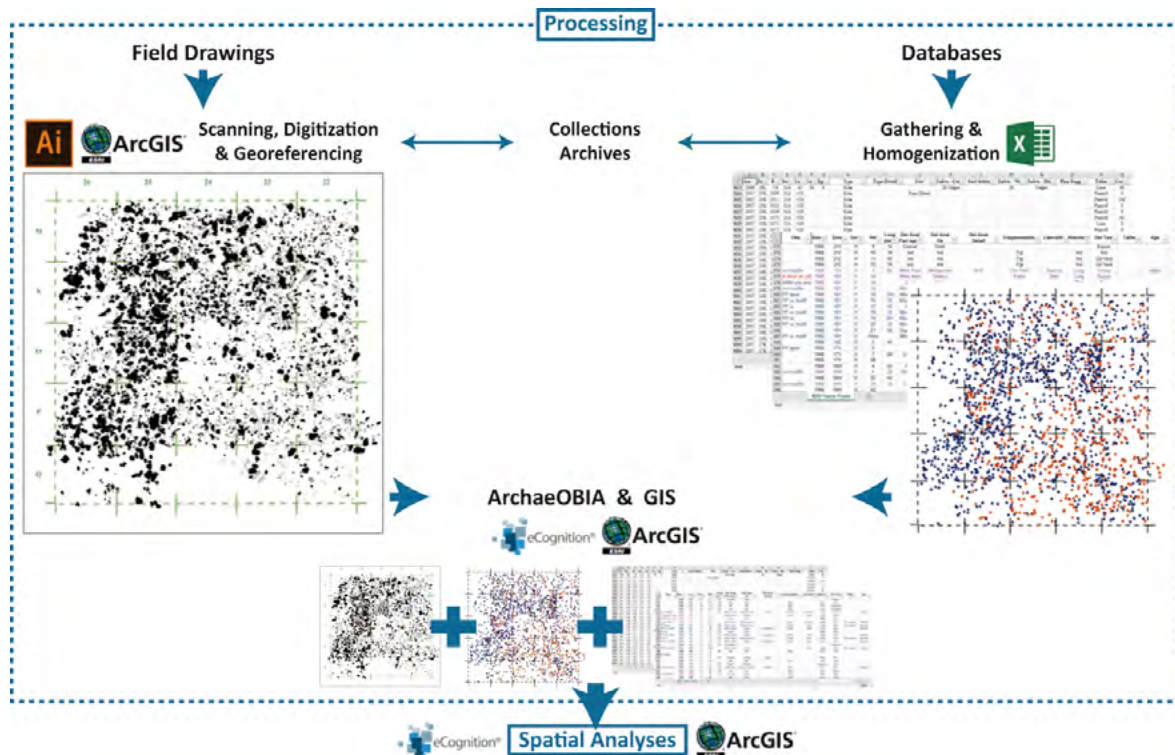


Figure 3. Methodological protocol submitted to the archives of the site of Cagny-l'Épinette (Database: A. Lamotte, A.-M. Moigne, P. Auguste; Field drawings & CAD: F. Peudon).

the field drawings, they have been scanned, digitized and georeferenced by combining a Computer Assisted Drawing (CAD) software with a GIS one. The use of a CAD tool instead of directly digitizing with a GIS software has been preserved for a portion of the maps to investigate the methodological possibilities when one is dealing with CAD documents in the archives without having to digitize them again. During this phase it has been necessary to visit the collections and the archives multiple times to review the materials and complete the databases. After this process, the two kinds of data are still separate and the CAD digitized georeferenced maps are still not quantitatively exploitable with a GIS software. Furthermore, errors relating to the spatial information, i.e. errors of location on the maps and missing coordinates in the databases, needed to be corrected. The last steps of the protocol have solved these problems and substantially enriched the data by combining the GIS software ArcGIS with the Object Based Image Analysis software eCognition®.

2.2. Data Extraction and Fusion

The OBIA methodology aims to treat and analyse raster images. Depending on the features of its pixels, such as their position and their layer values, an image is divided into several components that are then gathered into specific meaningful groups – i.e. representative of real phenomena – called image objects. These image objects are then easily exploitable with a GIS (Figure 4) (Batz and Schäpe 2000; Blaschke and Strobl 2001; Blaschke 2003; Hay *et al.* 2005; Hay and Castilla 2006; Blaschke 2010; Trimble 2016). The software eCognition was chosen for this work, and is based on segmentation and classification algorithms that are specifically chosen and organized in a rule set according to a cyclical ‘Segmentation-Classification-Segmentation’ pattern, repeated as needed (Batz *et al.* 2008; Trimble 2016). Once the rule set is developed, it can then be applied automatically to the complete data set that shares the same features. This Ph.D. work has been expanded in the frame of the concept of ArchaeOBIA, patented by A. Lamotte and E. Masson, which constitutes an application of the Object Based Image Analysis methodology in archaeology (Lamotte and Masson 2016; Masson and Lamotte 2018).

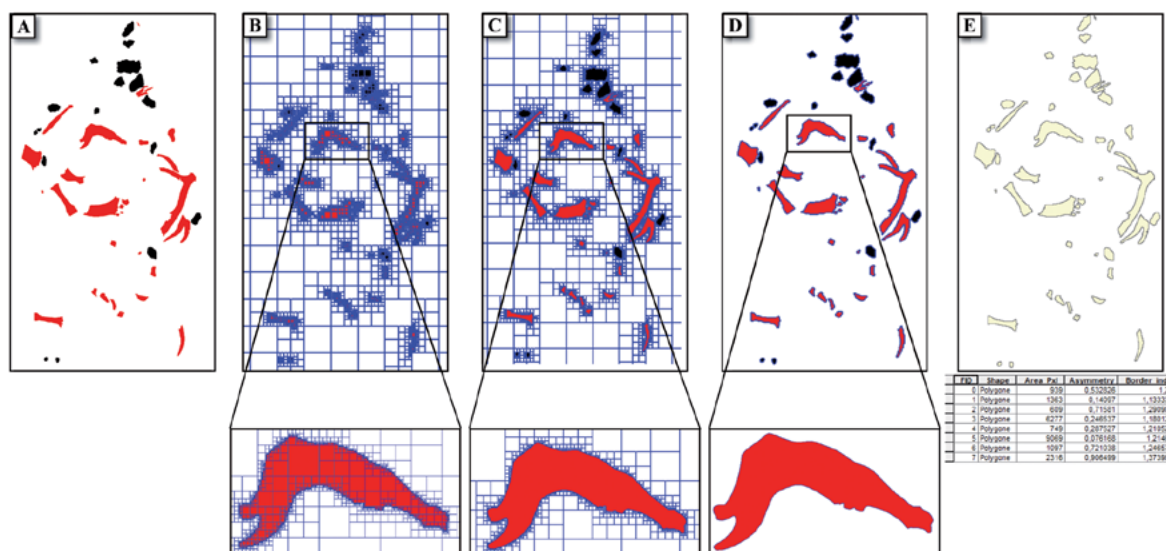


Figure 4. Rule set applied to the digitized georeferenced field drawings with the OBIA-software eCognition: A – Raster digitized georeferenced field drawing, B – Initial Segmentation, C – Classification and Objects Merging, D – Removal of the landmarks, Background Merging and their classification into the ‘No Data’ Class, E – Export as a shapefile with its new attributes within its generated attribute table (F. Peudon).

First applied to hand-axes to automate and tweak the measurement of their geometrical features by Lamotte and Masson (2016), its application in the context of this work constitutes the first use on documents such as field drawings of remains. The software eCognition has thus been used to automate the process of data enrichment from the digitized georeferenced maps, and transform the CAD plans into shapefiles so that they can be analysed in ArcGIS. To do so, a specific rule set has been developed in this project according to the features of the digitizations, through which the software eCognition can automatically transform each digitized shape into an image object and automatically classify these image objects into Faunal, Lithic or No Data (i.e. the background of the image) classes. Moreover once the image objects are created, eCognition can automatically measure new features such as their geometry, their position, their layer values, their texture and can automatically generate an attribute table. The configuration of the rule set allowing such image processing is as follows, the features of each algorithm being documented from Trimble (2016):

- 1) Initial segmentation of the raster image through the ‘Quadtree based Segmentation’ algorithm (Figure 4B). Based on the values of the pixels, this algorithm divides each non-homogeneous mesh into smaller quadrants, to the pixel scale, until each one of them are homogeneous, regardless of their dimensions.
- 2) Classification of the obtained homogeneous quadrants through the Basic ‘Classification’ algorithm (Figure 4C). For this purpose, this algorithm uses the pre-determined features of each class, which are in this work the pixel values 227, attributed to the faunal remains (Red); 0, to the lithic artefacts (Black); and 255, to the background (White). The attribution of specific and unique colours to the different types of remains during the digitalisation has allowed a simple classification based only on the RGB layer values of the image.
- 3) Merging of the similar contiguous quadrants into an image object within the ‘Faunal’ and ‘Lithic’ Classes through the Basic Object Reshaping ‘Merge Region’ algorithm (Figure 4C). A faunal or a lithic archaeological vestige now corresponds to each image object.
- 4) Classification of the landmarks, used during the georeferencing of the CAD plans, and their attribution to the ‘No Data’ Class by application of the Basic Object Reshaping ‘Remove Objects’ algorithm (Figure 4D). The image objects whose features have been selected are sorted into the appropriate Class; in this case, specific geometrical features of the landmarks (index of compactness and dimensions) and the pixel value 0 (Black) have been used. All the

components belonging to the 'No Data' Class (background and landmarks) have then been merged into one Image Object through the Basic Object Reshaping 'Merge Region' algorithm.

- 5) Export of the plan as a shapefile by application of the 'Export Vector Layer' algorithm (Figure 4E). This algorithm allows the multiple selection of geometrical (e.g. the length, width, orientation, indexes of compactness, of roundness and so on), spatial (e.g. coordinates (x, y) of the centre, the maximal and minimal coordinates (x, y)) and RGB layer values features, which are then automatically measured for each image object or for each image object Class of a plan when exported.

After the eCognition image processing, the plans have been imported back into ArcGIS, with their new attribute table, where they have been merged into one overall map for each archaeological level. Each digitized shape is then given its archaeological ID number obtained from the archaeological databases, which is then used as a common attribute between the archaeological databases and the eCognition output so that they can finally be linked by an attribute join (Figure 5).

3. Results

3.1. Enriched Data and Data Integrity

Two types of maps are provided for each archaeological level on the 277 square meters excavated, a scatter plot and the digitized representation of the remains. Each geometrical representation of a vestige – with its new geometrical and spatial features measured by eCognition – is thus linked to its point representation with the archaeological typological attributes (Figure 5). The Arché-OBIA methodology combined with the GIS software ArcGIS has hence provided enriched data. Firstly, these maps are linked to a third attribute table from eCognition containing additional measurements of the artefacts, such as the orientation and indexes of compactness or roundness, which can be relevant for subsequent analyses, particularly regarding taphonomic concerns. Another key point to consider is that it allows the completion of some the missing data of the databases, especially the Cartesian coordinates x and y. Additionally, this combined methodology

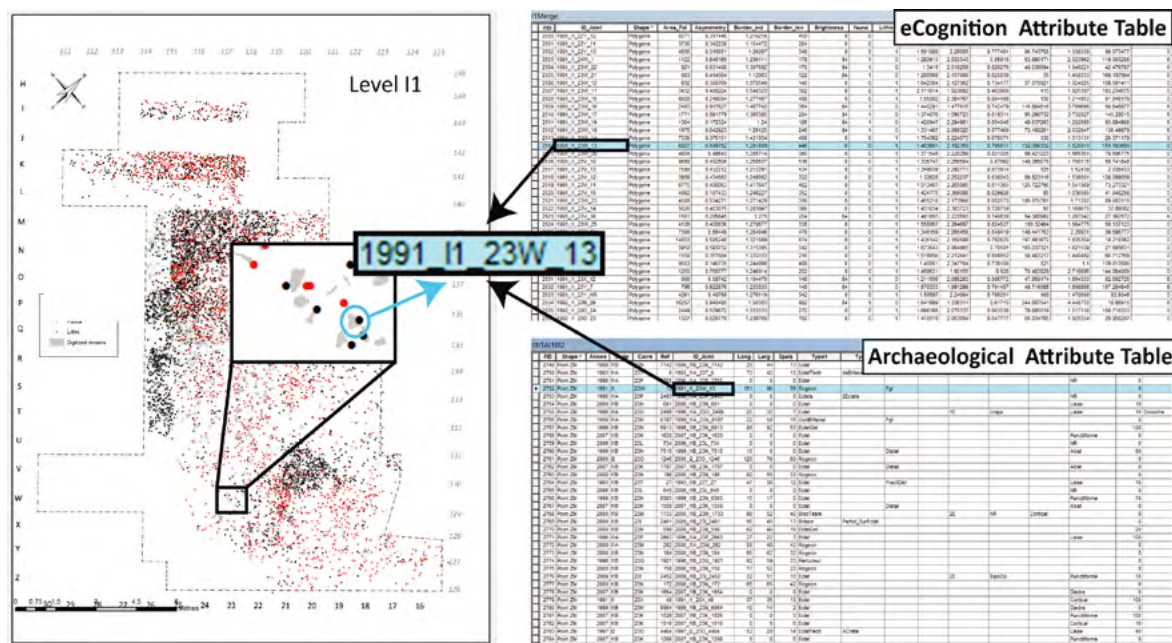


Figure 5. Overall digitalization of the archaeological vestiges of Level I1 overlapped by a corresponding scatter plot, both with their respective attribute tables linked through the ID Number. Grey: Digitized remains (Lithic and faunal remains), Red Dots: Faunal vestiges, Black Dots: Lithic artefacts (F. Peudon).

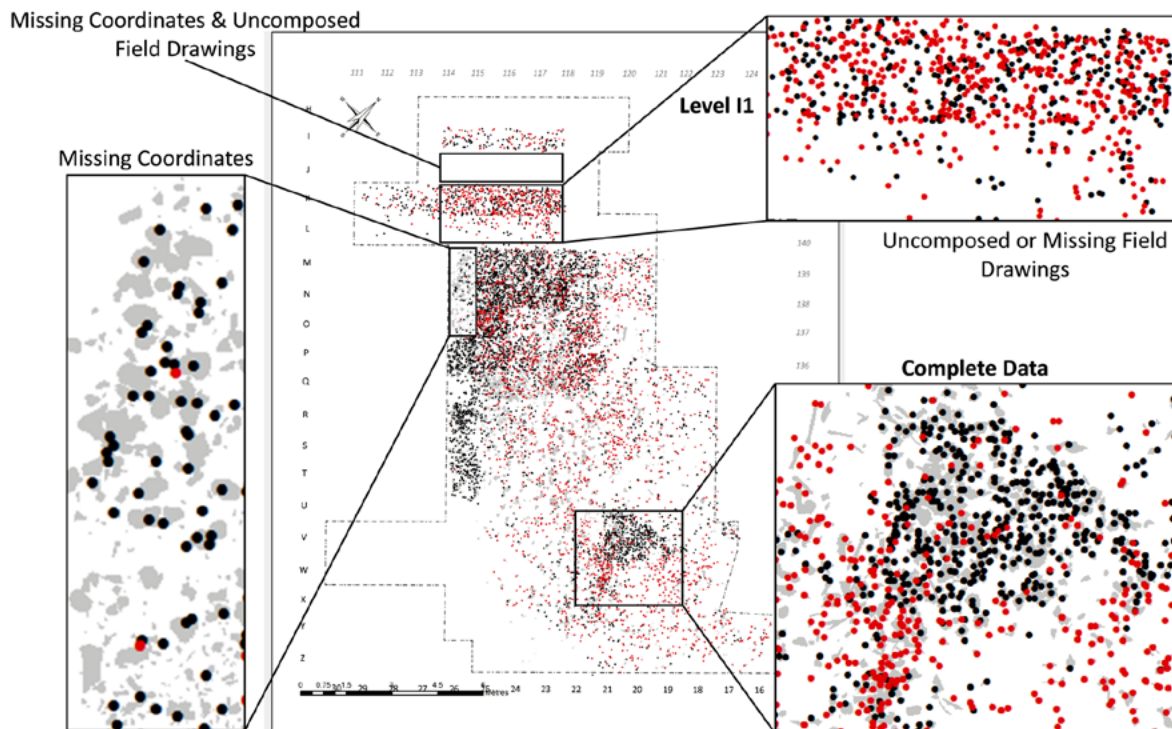


Figure 6. Integrity assessment of the spatial information of the Level I1 with three identified main degrees of integrity: areas with complete spatial data (field drawings and coordinates (x, y, z)), areas with either field drawings or coordinates (x, y, z) and the areas with no field data recording except the square meter affiliation. Grey: Digitized vestiges (Lithic and faunal remains), Red Dots: Faunal remains, Black Dots: Lithic artefacts (F. Peudon).

allows the data integrity review process to be semi-automated, which is essential before starting any spatial analysis.

When a vector does not match any coordinates of the databases – in other terms, when the eCognition coordinates do not match those from the field recording – it suggests either a potential field drawing or measurement mistake, a remain without a field drawing and/or missing coordinates. Once automatically highlighted, these vestiges can then be classified into distinct levels of integrity that will be taken into account during spatial analyses. Regarding the spatial information contained in the archives of Cagny-l'Épinette, three main degrees of integrity have been observed (Figure 6). Most of the site, documented by complete recordings (coordinate measurements and field mapping), belongs to the first degree. Some zones illustrate the second degree with only one kind of recording left, either the coordinates (x, y, z) or the field drawings. The third degree denotes sections for which no spatial data has been retrieved except for the square meter affiliation.

3.2. Multiscale Data and Optimal Resolution

With such materials, quantitative and thematic spatial analyses can now be considered – in a GIS interface – at different scales (square meter scale, area scale or site scale) with the appropriate degree of resolution based on the degree of data integrity. In other terms the 'Optimal Definition' of Gowlett (1997) is sought for each square meter to distinguish zones on which adequate geostatistical analyses can be applied. The spatial distribution of the artefacts according to their orientation, their size or even geometrical indexes is now conceivable. The areas devoid of archaeological remains and those that contain them can be quantified, as can the overall spatial extent of the vestiges. The spatial relations between specific categories of vestiges can be subjected to geostatistical analyses,

for instance distance analyses between bones bearing anthropogenic marks and lithic tools or blanks. Spatial analyses can be sophisticated with this approach.

3.3. Methodological Reflections

This ArchaeOBIA/GIS methodology is a powerful tool to automate and optimize the extraction of information from substantial, mixed, and sometimes incomplete archives, and also serves to merge and link them. In a single GIS protocol, the extraction of the digitized shapes and each kind of measurement would be the result of individual applications of extraction, classification and measurements. Sometimes it is possible in such a protocol to apply functions in chain (e.g. Image Analyst of the professional version of ArcGIS), but without the work-flow of the eCognition configuration containing all the needed rules in one set simultaneously applicable to the raster images. Thus, as already demonstrated in many methodological papers and case studies, this OBIA approach offers the opportunity to automate the simultaneous and rapid image objects extraction and classification along with measurement of a wide range of object-related, layer-related and class-related features (Baatz and Schäpe 2000; Blaschke and Strobl 2001; Blaschke 2003; Blaschke *et al.* 2014; Trimble 2016). Moreover, the elaboration of the eCognition rule set has proved to be a relative simple task in the case of pre-digitized georeferenced field drawings when compared to the complex rule sets required to process remote-sensing or surface magnetic survey imageries as well as photos and thin sections of archaeological vestiges, which require combinations of geometrical, textural and spectral features in the segmentation and classification algorithms (Baatz and Schäpe 2000; Blaschke and Strobl 2001; Blaschke 2003; Hay *et al.* 2005; Blaschke 2010; Hofmann *et al.* 2011; Verhagen and Drăguț 2012; Pregeßbauer *et al.* 2014; De Guio *et al.* 2015; Cerrillo-Cuenca 2017; Lamotte and Masson 2016; Masson and Lamotte 2018). And, as mentioned by Blaschke (2003), once the conception of the rule set is achieved, it can be used again on other data with similar features without having to re-develop it. In this work, the chain-application of the rule-set to all the raster plans took less than one minute per image.

Finally, this work shows that the use of the software eCognition allows, without too much added time and effort, the automation of the process of transforming raster images into shapefiles when documents have already been digitized with a CAD tool. It can also be applied to shapefiles digitized in a GIS, exported as raster images into eCognition, and imported back into the GIS as shapefiles. Thus, this protocol provides a flexible manner to process the data. In this methodology, it is apparent that the handwritten field drawings are important in the sense that they constitute archives about the spatial attributes and geometrical features for each archaeological remain. This is important in cases of remains that are not registered with a Total Station or not systematically measured during old excavations. These results follow the statement of Blaschke and Strobl (2001) – working on remote-sensing imageries – according to which ‘[...] images often are legitimately seen as the most information-rich base data available [...]’. In our case, the field drawings provide an important source of information that can be referred to if one is dealing with old archives, or in the event of digital file loss due to computer failure.

However, this methodology presents some limits. First, overlapping remains – or vestiges too close to each other – are a noteworthy problem, as the eCognition software is unable to distinguish two components of the same exact colour and would merge them into one single image object. This can be a major problem when working on former handwritten archives or already digitized documents and would probably only be solvable by manually (re-)digitizing the elements before submitting them to the OBIA process. Furthermore, the overlay of two or more shapes affects the integrity of the geometrical information for at least one of them. Regarding the data enrichment abilities, if the OBIA software eCognition allows the retrieval of some of the missing or originally unmeasured data, such as the orientation or the Cartesian coordinates, the altitudes can only be approximately recovered if the altitudes of their belonging spits are informed. Moreover, the measurement of the dip of the vestiges can unfortunately not be obtained after the field campaign. Finally, the

results of such methodology can be affected by errors on the field drawings and/or in the original databases, which can only be noticed and corrected if the two kinds of data are available for a cross-comparison.

4. Conclusion and Perspectives

The site of Cagny-l'Épinette represents 30 years of annual excavation and specialized studies involving 20 specialists, during which 277 square meters have been opened on 7 archaeological levels from which 12,112 lithic artefacts and 5,649 faunal remains have been unearthed and hand-drawn on 654 field drawings. The unusually high density of vestiges and the presence of faunal bones requiring specific mending treatment explains the excavated surface area, which could be considered restricted for an open-air palaeolithic site in comparison with the thousands of square meters opened during each single rescue archaeology operation.

Cagny-l'Épinette belongs to the extensive list of archaeological sites that have yielded a considerable amount of rich, mixed and complex archives and has offered a great opportunity to test the ArchaeOBIA/GIS methodological approach aiming to constitute adequate substance to the subsequent spatial analyses in a single GIS interface. This protocol has allowed us to extract all of the available information from this diversified data and to turn them into an homogenized and enriched data bank in the most optimal manner possible. Work involving remote-sensing imageries, for which the degree of resolution is a permanent concern, has led some authors to consider that 'One limitation of OBIA is that it requires very high-resolution datasets to work effectively' (Davis 2018). However, the presented work shows on the contrary that the application of OBIA on handwritten archives, with all the biases that it implies regarding precision and resolution, can still be considered as a relevant, very helpful and resourceful tool.

All this work has been carried out so that specific questions could be answered. Taphonomic issues concern the alluvial sequence that yielded several homogeneous archaeological levels, between which lithic refittings and faunal connections have been identified. The objective here would be to determine how many levels are embedded in the fine-grained alluvium. Secondly, as no obvious spatial features appeared during the excavation, a palaeo-ethnological approach has to be considered to attempt to distinguish potential activity areas and to identify the number of implements.

Acknowledgement

We thank to the entire scientific team and all the students who have contributed to the production of the spatial data. We are grateful to Armelle Masse, who has given us access to the collections (Centre de Conservation et d'Étude archéologiques, Dainville, Pas-de-Calais, France). We thank to Tom Mather for his review concerning the English Grammar.

References

- Antoine, P. 1990. *Chronostratigraphie et environnement du paléolithique du Bassin de la Somme*. Villeneuve d'Ascq: Publications du CERP 2, 231 p.
- Antoine, P., Auguste, P., Bahain, J.-J., Chaussé, C., Falguères, C., Ghaleb, B., Limondin-Lozouet, N., Locht, J.-L. and Voinchet, P. 2010. Chronostratigraphy and palaeoenvironments of Acheulean occupations in Northern France (Somme, Seine and Yonne valleys). *Quaternary International* 223: 456-461. <doi:10.1016/j.quaint.2009.07.014>.
- Baatz, M., Hoffmann, C. and Willhauck, G. 2008. Progressing from object-based to object-oriented image analysis, in T. Blaschke, S. Lang, S. and G.J. Hay (eds) *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*: 29-42. Berlin, Heidelberg: Springer.
- Baatz, M. and Schäpe, A. 2000. Multiresolution Segmentation – An optimization approach for high quality multi-scale image segmentation, in J. Strobl, T. Blaschke and G. Griesebner (eds)

- Angewandte geographische Informationsverarbeitung XII: Beiträge zum AGIT-Symposium (Salzburg 5-7 Juli 2000)*: 12-23. Heidelberg: Herbert Wichmann Verlag.
- Balescu, S., Dupuis, C. and Quinif, Y. 1988. TL stratigraphy of pre-Weichselian loess from NW Europe using feldspar coarse grains. *Quaternary Science Reviews* 7(3-4): 309-313. <doi:10.1016/0277-3791(88)90020-0>.
- Balescu, S. and Tuffreau, A. 2004. La phase ancienne du Paléolithique moyen dans la France septentrionale (stades isotopiques 8 à 6): apports de la datation par luminescence des séquences loessiques. *Archeologicheskii almanach, Donetsk* 16: 5-22.
- Bates, M.R. 1993. Quaternary aminostratigraphy in Northwestern France. *Quaternary Science Reviews* 12(9): 793-809. <doi:10.1016/0277-3791(93)90018-H>.
- Blaschke, T. 2003. Object-based contextual image classification built on image segmentation, in *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003*: 113-119. Greenbelt, Maryland, USA: IEEE.
- Blaschke, T. 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65(1): 2-16. <doi:10.1016/j.isprsjprs.2009.06.004>.
- Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F. and Tiede, D. 2014. Geographic object-based image analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 180-191. <doi:10.1016/j.isprsjprs.2013.09.014>.
- Blaschke, T. and Strobl, J. 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GeoBIT/GIS* 14(6): 12-17.
- Bourdier, F., Chaline, J., Munaut, A.-V. and Puisségur, J.-J. 1974. Le complexe Mindelien: II – La moyenne terrasse de l'Avre. *Quaternaire* 11(3): 168-180. <doi:10.3406/quate.1974.2127>.
- Cerrillo-Cuenca, E. 2017. An approach to the automatic surveying of prehistoric barrows through LiDAR. *Quaternary International* 435: 135-145. <doi:10.1016/j.quaint.2015.12.099>.
- Commont, M.V. 1911. Les terrasses fluviales de la vallée de la Somme. *Bulletin archéologique, Géologie*: 173-195.
- Cordy, J.-M. 1989. Les micromammifères de Cagny-l'Épinette (rapport préliminaire), in *Livret-guide de l'excursion dans la vallée de la Somme. Colloque L'Acheuléen dans l'Ouest de l'Europe (Saint-Riquier, 6-10 June 1989)*: 80-81. Lille: Publications du CERP.
- Davis, D.S. 2018. Object-based image analysis: a review of developments and future directions of automated feature detection in landscape archaeology. *Archaeological Prospection*: 1-9. <doi:10.1002/arp.1730>.
- De Guio, A., Magnini, L. and Bettineschi, C. 2015. GeOBIA approaches to remote sensing of fossil landscapes: two case studies from Northern Italy, in A. Travaglia (ed.) *Across Space and Time. Papers from the 41st Conference on Computer Applications and Quantitative Methods in Archaeology (Perth, 25-28 March 2013)*: 45-53. Amsterdam: Amsterdam University Press.
- Dibble, H.L., Chase, P.G., McPherron, S.P. and Tuffreau, A. 1997. Testing the reality of a 'living floor' with archaeological data. *American Antiquity* 62(4): 629-651. <doi:10.2307/281882>.
- Gowlett, J.A.J. 1997. High definition archaeology: ideas and evaluation. *World Archaeology* 29(2): 152-171. <doi:10.1080/00438243.1997.9980371>.
- Haesaerts, P., Balescu, S., Dupuis, C. and Van Vliet, B. 1984. Contribution à la stratigraphie des gisements paléolithiques de Cagny (Somme), in « Numéro édité à l'occasion du 22 Congrès Préhistorique de France (Lille-Mons, 2-7 septembre 1984) ». *Cahier de Géographie Physique. Travaux du Laboratoire de Géomorphologie et d'Etude du Quaternaire Villeneuve d'Ascq* 5: 77-94.
- Hay, G.J. and Castilla, G. 2006. Object-based Image Analysis: Strengths, Weaknesses, Opportunities and Threats (Swot), in 'Proceedings of the 1st International Conference on Object-based Image Analysis. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences' XXXVI-4/C42.
- Hay, G.J., Castilla, G., Wulder, M.A. and Ruiz, J.R. 2005. An automated object-based approach for the multiscale image segmentation of forest scenes. *International Journal of Applied Earth Observation and Geoinformation* 7(4): 339-359. <doi:10.1016/j.jag.2005.06.005>.
- Hofmann, P., Marschallinger, R., Unterwurzacher, M. and Zobl, F. 2011. Designation of marble provenance: State-of-art rock fabric characterization in thin sections by object based image

- analysis. Presented at the 2011 *Annual Conference of the International Association for Mathematical Geosciences, Salzburg, Austria*.
- Lamotte, A. and Masson, E. 2016. Arché-OBIA: un concept d'analyse quantitative d'images numériques appliqué aux bifaces du gisement de Gouzeaucourt (Nord, FR). *Notae Praehistoricae* 36: 121-130.
- Lamotte, A. and Tuffreau, A. 2001. Les industries lithiques de la séquence fluviatile fine de Cagny-l'Épinette (Somme): 113-134. Villeneuve d'Ascq: Publications du CERP 6.
- Lamotte, A., Tuffreau, A., Auguste, P. and Peudon, F. 2014. Is it possible to recognize structured living-places and activities at Cagny-l'Épinette, Somme Basin?, Presented at the *XVII Mundial Congress of Prehistoric and Protohistoric Sciences 2014, Burgos, Spain*.
- Laurent, M., Falguères, C., Bahain, J.-J. and Yokoyama, Y. 1994. Géochronologie du système de terrasses fluviatiles quaternaires du bassin de la Somme par datation RPE sur quartz, déséquilibres des familles de l'uranium et magnétostratigraphie. *Comptes rendus hebdomadaires des séances de l'Académie des sciences* 318(II): 521-526.
- Léopold, J. 1997. Les industries lithiques de Cagny-L'Épinette, Cagny-Cimetière et Cagny-la-Garenne dans le contexte Nord-Ouest européen. *L'Anthropologie* 101(4): 639-669.
- Loriot, B., Fougereolle, Y., Sestier, C. and Seulin, R. 2007. 3D acquisition and modeling for flint artefacts analysis. Presented at O3A: *Optics for Arts, Architecture, and Archaeology*. International Society for Optics and Photonics, 66180G.
- Masson, E. and Lamotte, A. 2018. Arché-OBIA: vestiges lithiques au Paléolithique et exploitation quantifiée de leurs images numériques. *CRAHN-SNEPH, Hors-série* 1: 153-165.
- Moigne, A.-M. 1989. Cagny-l'Épinette: la grande faune, in *Livret-guide de l'excursion dans la vallée de la Somme, Colloque 'L'Acheuléen dans l'Ouest de l'Europe' (Saint-Riquier, 6-10 June 1989)*: 82-87. Lille: Publications du CERP.
- Munaut, A.-V. 1989. Cagny-l'Épinette: analyses palynologiques, in *Livret-guide de l'excursion dans la vallée de la Somme. Colloque 'L'Acheuléen dans l'Ouest de l'Europe' (Saint-Riquier, 6-10 June 1989)*: 75-79. Lille: Publications du CERP.
- Peudon, F. 2013. *Le gisement acheuléen de Cagny-l'Épinette de la quatrième terrasse alluviale de la vallée de la Somme (Picardie, France): contribution à la mise en application d'une méthodologie d'analyse spatiale sur les niveaux archéologiques de la séquence fluviatile*. Master's dissertation 2. Université de Bourgogne, Dijon.
- Peudon, F., Masson, E., Auguste, P., Lamotte, A., Moigne, A.-M. and Tuffreau, A. 2018a. Archaeological Data and Archaeology of Data. Presented at the *Jour de Contact Préhistoire FNRS 2018, Tongeren, Belgium*.
- Peudon, F., Lamotte, A. and Tuffreau, A. 2018b. Spatialisation des chaînes opératoires de façonnage et de débitage du gisement acheuléen de Cagny-l'Épinette (MIS 9-10) (Somme, France). Presented at the *Jour de Contact Préhistoire FNRS 2018, Tongeren, Belgium*.
- Pregesbauer, M., Trinks, I. and Neubauer, W. 2014. An object oriented approach to automatic classification of archaeological features in magnetic prospection data. *Near Surface Geophysics* 12(5): 651-656. <doi:10.3997/1873-0604.2014014>.
- Sommé, J., Fagnart, J.-P., Léger, M., Munaut, A.-V., Puisségur, J.-J. and Tuffreau, A. 1984. Terrasses fluviatiles du Pléistocène moyen en France septentrionale: signification dynamique et climatique. *Quaternaire* 21(1): 52-58. <doi:10.3406/quate.1984.1480>.
- Trimble 2016. *eCognition® Developer User Guide 9.2*. Trimble Documentation, München, Germany.
- Tuffreau, A. (ed.) 1989. *Fouille programmée de Cagny-l'Épinette (Somme). Rapport de synthèse (1987-1989)*. Villeneuve d'Ascq: Université des Sciences et Techniques de Lille Flandres Artois. Rapport de synthèse.
- Tuffreau, A. (ed.) 2004. *Fouille programmée de Cagny-l'Épinette (Somme). Rapport de synthèse (2002-2004)*. Villeneuve d'Ascq: Laboratoire de Préhistoire et Quaternaire, Université des Sciences et Technologies de Lille 1. Rapport de synthèse.
- Tuffreau, A. (ed.) 2010. *Fouille programmée de Cagny-l'Épinette (Somme). Rapport de synthèse (2008-2010)*. Villeneuve d'Ascq: Laboratoire de Préhistoire et Quaternaire, Université des Sciences et Technologies de Lille 1. Rapport de synthèse.
- Tuffreau, A., Antoine, P., Chase, P.G., Dibble, H.L., Ellwood, B.B., van Kolfschoten, T., Lamotte, A., Laurent, M., McPherron, S.P., Moigne, A.-M. and Munaut, A.-V. 1995. Le gisement acheuléen de

- Cagny-l'Épinette (Somme). *Bulletin de la Société préhistorique française* 92(2): 169-192. <doi:10.3406/bspf.1995.10004>.
- Tuffreau, A., Lamotte, A. and Marcy, J.-L. 2001. *Les gisements acheuléens de Cagny (Somme)*. Série Archéologie en Picardie 18. Amiens: DRAC Picardie.
- Van Vliet-Lanoë, B. 1989. Observations paléopédologiques sur les gisements de Cagny-la Garenne, Cagny-l'Épinette et Saint-Acheul. In *Livret-*
- Van Vliet-Lanoë, B., Laurent, M. and Beun, N. 2001. Évolution néogène et quaternaire de la vallée de la Somme: 19-27. Villeneuve d'Ascq : Publications du CERP 6.
- Verhagen, P. and Drăguț, L. 2012. Object-based landform delineation and classification from DEMs for archaeological predictive mapping. *Journal of Archaeological Science* 39(3): 698-703. <doi:10.1016/j.jas.2011.11.001>.

Towards an Archaeological Information System: the evolution of Syslat, an archaeological data management software

Réjane Roure^{1,2}, Hakima Manseri¹, Sébastien Munos²,
Michel Py³

¹ Archéologie des Sociétés Méditerranéennes (ASM) – Université Paul Valéry – Montpellier III,
CNRS: UMR5140, Ministère de la Culture

² Labex Archimède « Archéologie et Histoire de la Méditerranée et de l’Égypte anciennes »
(PIA-ANR-11-LABX-0032-01)

³ Directeur de recherche honoraire au CNRS, Archéologie des Sociétés Méditerranéennes (ASM)
rejane.roure@univ-montp3.fr
hakima.manseri@cnrs.fr
sebastien.munos@univ-montp3.fr
michel-py@sfr.fr

Abstract

Our laboratory is strongly involved in the development and the continuation of the Syslat software, created by Michel Py (CNRS) and the research team of Lattes in the 1980s. Designed as a software tool for managing and exploiting data from archaeological excavations, Syslat is evolving into an Archaeological Information System (AIS), suitable for any type of sites, vestiges or periods. The many versions of this program since its inception allowed to take up one of the main challenges raised by the growth of archaeology since the 1980s, that is of the management and normalisation of the recording of archaeological data, which are becoming exponential, in all their diversity and complexity. The evolution of Syslat into an AIS will allow the integration of the latest development in data recording and exploitation procedures while pursuing the effort on the normalisation front and the production of references based on the dictionaries and libraries attached to the software.

Keywords: data, archaeological excavations, exploitation, recordings, information system

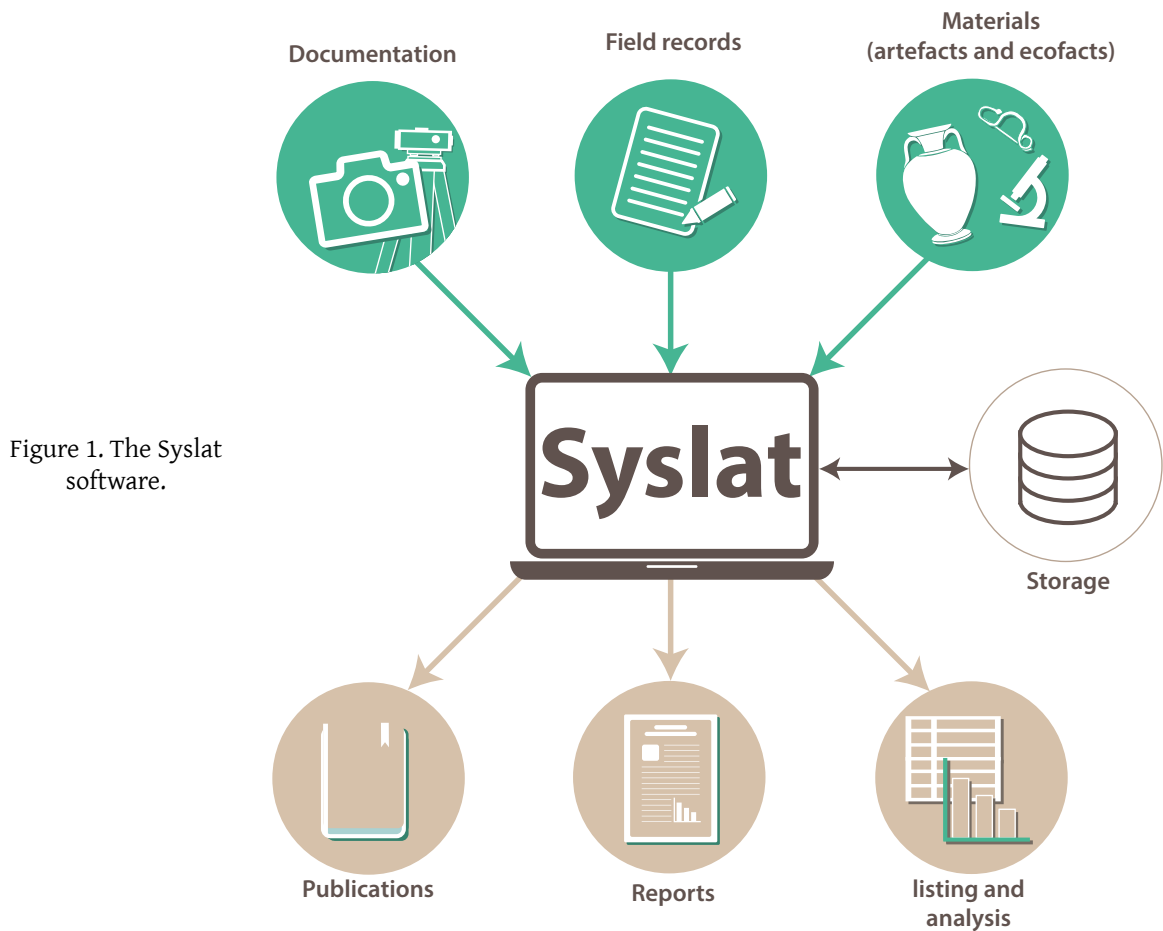
Résumé

Notre laboratoire est fortement impliqué dans le développement et l’évolution du logiciel Syslat, créé par Michel Py (CNRS) et l’équipe de recherche de Lattes dans les années 1980. Conçu comme un outil logiciel pour la gestion et l’exploitation des données provenant de fouilles archéologiques, Syslat évolue vers un système d’information archéologique (SIA), adapté à tout type de sites, vestiges ou périodes. Les nombreuses versions de ce programme depuis sa création ont permis de relever l’un des principaux défis soulevés par la croissance de l’archéologie depuis les années 1980, c’est-à-dire la gestion et la normalisation de l’enregistrement des données archéologiques, qui deviennent exponentielles, dans toute leur diversité et leur complexité. L’évolution de Syslat en AIS permettra l’intégration des derniers développements dans les procédures d’enregistrement et d’exploitation des données tout en poursuivant l’effort sur le front de la normalisation et la production de références basées sur les dictionnaires et les bibliothèques attachés au logiciel.

Mots-clés : données, fouilles archéologiques, enregistrement, exploitation, système d’information

1. Introduction

Syslat (Syslat 2018a) is a software suite dedicated to the recording, management and exploitation of scientific data from archaeological excavations (Figure 1). It was created by Michel Py (CNRS)



in the ASM laboratory (UMR5140), in collaboration with numerous archaeologists associated with the project. Initiated in the 1980s., the software has undergone several technological and scientific developments and, thanks to the support of the Labex Archimede (PIA-ANR-11-LABX-0032-01), managed to get closer to becoming an Archaeological Information System (AIS): by developing the software and improving its quality, we aim at guaranteeing the integrity and quality of the archaeological data.

The development of this software participates in the movement of standardization and sharing of these data among the community of archaeologists in the broad sense (Universities, CNRS, local authorities, Inrap,...). The project's progress is presented in the research blog dedicated to Syslat (Syslat 2018b).

In this article, after describing the Syslat project background, we move onto the presentation of the software itself and conclude with the future developments in the works.

2. History of the development

2.1. The beginnings

The research team located in Lattes (Hérault, France) has been using computers for archaeological research since 1981, thanks to a partnership with the French laboratory LISH (Laboratoire d'Informatique pour les Sciences Humaines) based in Marseille. In 1988, the French Ministry of Culture set up an important call for projects regarding the creation of a software dedicated to the recording and storing of archaeological data, called 'Archives de fouilles', when computing

devices started to increase in all scientific works. Archaeology was one of the first scientific field among the *Humanities* to include computers in its daily practices. Many laboratories presented a project and a dozen obtained a grant to develop such a software. The research team of Lattes used the grant to buy computing equipment in order to improve the way archaeological data were recorded. It is important to notice that since 1983, the research team working on the antic and protohistoric settlement of *Lattara* had built an important protocol to record archaeological data, including not only the data from the field (walls, floor, structures) but also all the materials (ceramic, metal, glass, coins, etc.) and – for quite the first time – environmental data (archeozoology, carpology, etc.). Their protocol was based on a modified version of Harris method (Harris 1979): it included new approaches and a deep reflection on archaeological fields and methods. This work was driven by Michel Bats, Jean-Luc Fiches and Michel Py, all of them belonging to the CNRS, and published in 1986 (Bats, Py, Fiches 1986). The paper forms that had been initially created were computerised and many tools to process these data were added (graphs, summary tables, etc.). At first, these simple files were processed with File (Microsoft) then Filemaker (Claris) both running on Macintosh devices (Apple Macintosh Plus, then Macintosh II). Then, the different and independent files were linked thanks to an interface using Hypercard. After many trials, the results of this first stage of the Syslat development were published in 1991, in the issue number 4 of the *Lattara* series (Py 1991), which had been sold out in less than a year (see Figure 2.1).

2.2. Syslat 3.1

The progress of archaeological research field on one hand and of computing devices on the other hand required to update the software constantly. The team of Lattara and all the members of the Joint Research Unit (CNRS-Ministry of Culture) of Lattes, settled near the archaeological site and the local museum, organised many conferences, seminars and collaborations, in particular with the so-called ‘rescue archaeology’ and the members of the AFAN (French Association for National Archaeology). These helped define the news needs of the archaeologists and therefore, adapt Syslat to those needs.

On the computing front, the development of Hypercard by all the user of this tool all over the world convinced the Syslat team to switch to it for the development of Syslat. The software had been re-written and the user interface had been normalized. This new version of the software integrated new tools to manage, search, find and process all the recorded data. After six years of brainstorming and testing, an almost stable version of Syslat had been released, Syslat 3.1, with its user book *Lattara10* (Py *et al.* 1997, see Figure 2.2). The general organisation of the software and its data was fixed: field, materials, environmental samples, documentation and tools.



Figure 2.1. Cover of Lattara 4.

2.3. Syslat X

The 2000s marked a turning point as Macintosh developed the System 10 on (Os X). Hypercard was bound to disappear for Apple did not want to put it on System 10 despite thousands of protests around the world. Syslat was very dependent on Hypercard, so many users were worried about the future of their data and their work habits. It was therefore necessary to make a new version (Syslat X) based on a new engine. The Revolution™ software was the natural choice since it was the heir to Hypercard, Supercard and Metacard (from which came part of the codes) and offered the best porting possibilities on various environments (especially on Os 10 and on Windows). On this occasion, the Syslat system was simplified, its main files and tools improved in their presentation and operation and the user interface revised, broadening its use to archaeological sites other than Lattara-like digs.

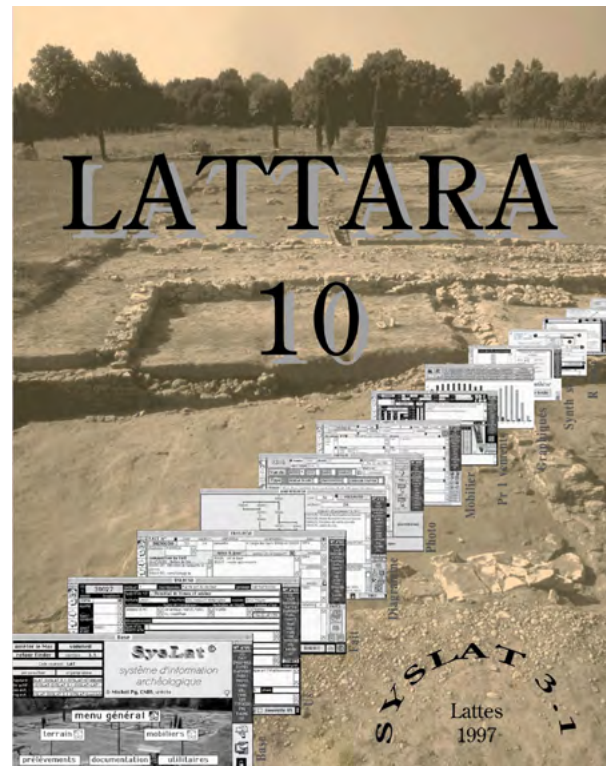


Figure 2.2. Cover of Lattara 10.

2.4. Syslat-Terminal and beyond

Just a few years later, Syslat climbed another big step through the integration of the potentialities of internet. Indeed, Syslat-Terminal corresponds to the latest evolution of the Syslat software, developed on LiveCode. The data are now stored and managed through MySQL, a DataBase Management System. With the database and the digital documentation available on a remote server, Syslat-Terminal allows the sharing of all the information between an unlimited number of collaborators by simply indicating the code identifying the excavation site and a password identifying the privileges of each user. Syslat-Terminal is still usable as a standalone application with no internet connection, which is useful on isolated archaeological sites. A mobile application named iSyslatPro has been released, allowing the users to access their data through their mobile device (Android and iOS).

The importance of the project has since been recognized as the development of Syslat and its scientific by-products (dictionaries, catalogues) are supported by the Labex Archimede (PIA-ANR-11-LABX-0032-01). Of similar importance is the release of the latest version, Syslat-Terminal 5, in April 2017: on this occasion, the Syslat data server was transferred to the TGIR HumaNum, a French Very Large Research Infrastructure ('Très Grande Infrastructure de Recherche', TGIR) with international reach devoted to Social Sciences and Humanities and supported by the CNRS, Aix-Marseille University and the Campus Condorcet. This ensures the preservation and accessibility of the data acquired by the archaeologists through Syslat.

Syslat is now a multiplatform software suite: Syslat-Terminal works on most computers (OS(X), Windows, Linux) while mobile devices (Android and iOS) are not forgotten with iSyslatPro and iSyslight, the latest product of the Syslat family geared towards data recording on the field. New

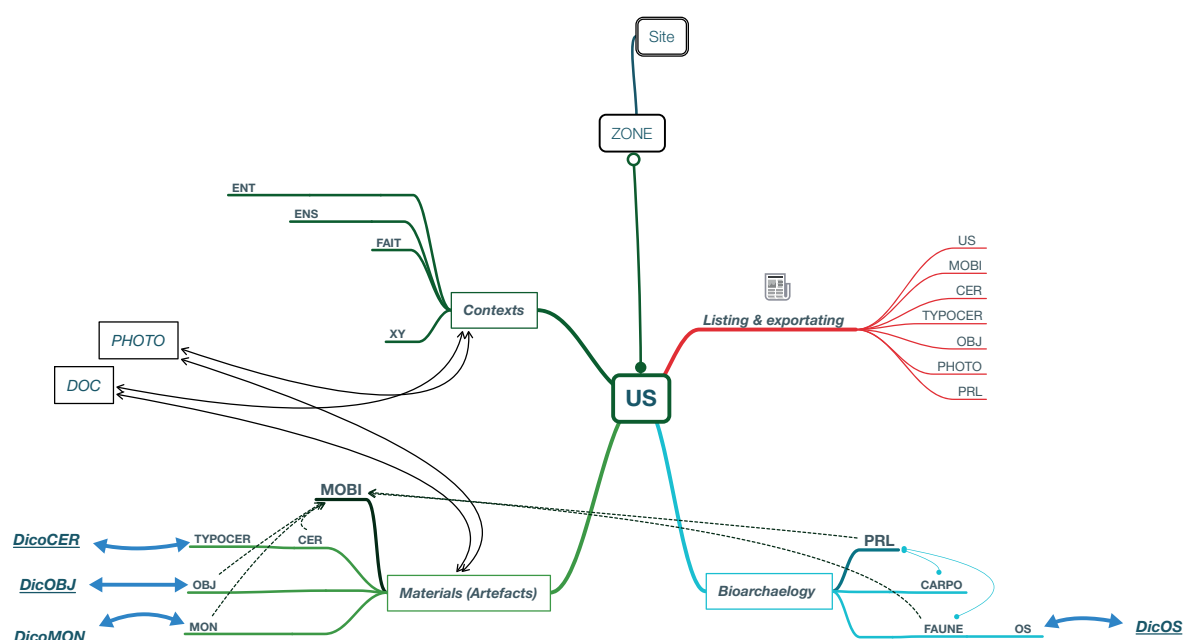


Figure 3. A mind map of the Syslat architecture.

functionalities are designed and implemented, improving the software, its performance and its adaptability.

3. The Syslat software suite

3.1. Description

The Syslat software suite is designed to follow the process of recording and exploiting the archaeological data as usually done by the archaeologists on the field and beyond. To do so, the Syslat data are organised in what are called *files*: for example, the file FAUNE contains all the records regarding the fauna. We can group the files in four categories:

- field context: US, FAIT, ENS, ENT, ZONE, SITE
- artefacts: CER, TYPOCER, OBJ, MOBI
- ecofacts: OS, FAUNE, CARPO, PRL
- documentation: PHOTO, DOC

As shown in Figure 3, Syslat offers a wide range of files that cover most of the data that can be acquired and produced during an archaeological excavation. There are files adapted to registration and operation, whether for field data, materials, samples or any type of documentation (pictures, drawings, etc.). The software allows each specialist to access a view adapted to his work, while ensuring dynamic links between the files.

The main component of the software suite is Syslat-Terminal, a desktop application running on Mac OS(X), Windows and Linux. Syslat-Terminal allows its user to make the most out of the data. The main functionalities are described later (see 3.3).

Mobility is not forgotten with iSyslatPro and iSyslight, two mobile applications available for iOS and Android. While iSyslatPro is a mobile version of Syslat-Terminal, iSyslight is a lighter, highly adaptable version which focuses on recording the data directly on the field using tablets: the scope is narrower than for iSyslatPro but by using only the necessary files, it is far more efficient while recording.

3.2. Software architecture

The Syslat software suite follows a classical client-server architecture: the clients are written using the Livecode development framework while a database management system DBMS, coupled with file transfer capability, acts as the server.

The notion of server is relative: for example, the MySQL DBMS can be installed on a remote server but can also be located on a simple computer allowing Syslat-Terminal to work as a standalone application, independent from network constraints. The same is possible with the mobile clients by using SQLite as a DBMS. Synchronization is possible between the data located on the computer or tablet and a remote server.

Each site has a dedicated database created on the server with a unique alphanumerical code. Access to said database is subjected to authentication using the code and a password corresponding to a certain level of authorisation (consultation, modification, administration).

3.3. Functionalities: recording, retrieving and processing data

The recording of field data relies on different files related to the management of space as well as to contexts and stratigraphic relationships. Six files are available: US, FACT, ENSEMBLE, ENTITY, ZONE and SITE. Each record can be automatically associated with its documentation and linked to other files, especially those related to materials.

Syslat also integrates a series of files dedicated to the management and the exploitation of the different archaeological materials, as varied as they are. The implementation allows the specific recordings and processing inherent in each type of material.

The MOBI file makes it possible to synthesize and inventory on the same page all the furniture of a US, while specifying their physical location. More specifically, files are available for the different types of furniture. Artefacts could be recorded on three files: MON is intended for the registration and description of coins; CER and TYOCER are dedicated to first the inventory and counting of ceramics and then to the typological specifications for the Iron Age and Roman periods (in the present state of the program); OBJ is for the registration and description of all the other kind of objects (registration by individual): glass, metal, stone, etc.

The data belonging to Bioarchaeology can be recorded thanks to three files: PRL, which allows specialists to have a tool for managing environmental samples from the archaeological excavation; CARPO for the study of seeds; OS and FAUNE concern archaeozoology (inventory and counting of bones for each US and files for each specific bone).

To manage the documentation, the software integrates two dedicated files: the PHOTO and DOC files. The first makes it possible to record all the photographs, whether it is pictures of excavation or materials, and there are automatic links with the other files (especially US and FACT). Regarding the graphic documentation, the DOC file takes care of inventorying and integrating into the system all the sketches, the drawings of materials.

In addition to the data recording, Syslat offers data retrieval functions in order to search, manage, select and export all the data. A comprehensive search module covers all fields (simple search, multi-criteria search, search/replace, select and search in all files and set, etc.). It is possible to make accurate record selections with fast and convenient procedures. The search results can be used to quickly generate tabular files or a complete textual synthesis.

Syslat-Terminal also offers several analysis modules integrated to the various files to quickly exploit data. Associated with the synthesis module, it makes it possible to edit numerous pages of data in an elaborate form (texts, tables) and to be integrated easily in a report or a publication. Other modules provide real-time textual or graphical information on many aspects of the search: chronology, distribution maps, statistics, etc.

3.4. *Controlled vocabularies and shared libraries*

One of the fundamental aspects of the software is the standardisation and the optimisation of the recording. In Syslat-Terminal, iSyslight and iSyslatPro, most input fields are augmented with drop-down menus powered by a set of libraries, which enable efficient data entry. Most of the input can be done through that drop down and hierarchical menus, that have the huge advantage of optimising the recording and offering all users a controlled vocabulary. The software offers a set of standard libraries, and users are free to customise them or create new ones for each site account. So it is possible to create and edit site or period specific libraries and even share them with other site accounts. These personal libraries can be edited via a dedicated interface in Syslat-Terminal. Obviously, all these useful functions do not prevent to make the choice to enter the data freely in all the fields if the needed words are not available in the controlled vocabulary currently in use.

3.5. *Around Syslat: dictionaries, catalogs and documentation*

As stated before, the software heavily relies on controlled vocabularies which are partly based on published typological dictionaries: Dicocer, dictionary of ancient ceramics (7th BC-7th AD in the North Mediterranean Sea); Dicomon, dictionary of the coins of Mediterranean Gaul; Dicobj, dictionary of Iron Age objects of Mediterranean Gaul. Those dictionaries are themselves the results of the analysis of archaeological data recorded using Syslat. One using the content of the other, they feed and enrich each other, creating a virtuous circle and helping the dissemination of the typologies and therefore, improving the standardisation of the data. Syslat offers a direct access to the content of these dictionaries as references.

In the same manner, the osteological reference catalogue of the ASM laboratory, aptly named Dicos, is accessible through Syslat: the description and pictures of thousands of bones are available as a reference for osteological studies.

Last but not least, Syslat comes with a manual in PDF format of 300 pages. Though this manual is only in French, a translation in five languages (French, English, Italian, Spanish and Catalan) is available directly in Syslat-Terminal and on-line as well (Syslat 2018a).

4. Conclusion

First designed with Iron Age and Antic settlement sites in mind, the Syslat software is a working tool for research in archaeology, used by archaeologists in the field as well as by specialists, from the excavation to the publication. Syslat is currently used for more than 900 archaeological operations, in France and abroad. Several archaeological departments use the Syslat software suite on their internal network (Mediterranean Inrap for example). The Syslat family is used in the South of France mainly, but also in Catalunya, South of Italy, Western France. Anywhere someone has started to use it, he or she keeps it and go on to record his data with it.

Indeed, Syslat is a software designed to collect, store, process, analyze and present all types of data necessary for the collaborative exploitation of an archaeological dig. The use of such a

system involves the standardization of record and exploitation procedures, which improves the conservation, transmission and sharing of archaeological data: field observations, excavation documentation and collected materials, according to criteria of spatial (coordinates, area, site) and temporal (context, stratigraphy). This AIS thus contributes to the management of the archaeological dig from the fieldwork to the archiving of the resulting data, including its study. By these aspects, it represents a useful tool for the restitution of the history of an archaeological site, the long-term archiving of its data and their dissemination. Indeed, the data are stored on a standardized and exploitable support (archiving, consultation, extraction, etc.); it allows fast and efficient data exploitation and plural and adaptable extractions. Exploitation of some information can feed another part (relative/absolute chronology, phasing, etc.) because cross analysis between the different recordings are very simple. Cross analysis with other systems (comparison of data, establishment of regional chronologies, reference dictionaries, etc.) are possible too through exports or multi-base search; it allows to have a global as well as partial approach. Syslat is interfaced with GIS (QGIS) and integrates all the advantages of a database (search, fast access to information, etc.), with dissemination facilities (other systems, open data). Open access is already available by giving the lecture password to all the people who want it, nevertheless the Syslat team thinks about a dedicated website for the open access to the data recorded on Syslat, for the archaeologists who want to share their data.

Syslat is still a work in progress. The software will be enriched in the next years with several completely new modules and an update of some of the tools associated with the software. A module dedicated to glassware is in the work, comprising two record files (VERRE and TYPOVERRE), associated with a dedicated typological dictionary: DICOVERRE and all the tools for data exploitation. Libraries specialized in the archaeology of buildings have been updated and a dictionary-corpus of fauna is in development (Pégase 2018). A module dedicated to archeo-thanatology is also developed with some specialists, to integrate data from funerary archaeology and biological anthropology, which require specialised records.

In the next few years, we hope Syslat will be moving from a software suite to a fully functioning Archaeological Information System, ensuring the preservation, integrity and reuse of all the kind of archaeological data acquired, promoting and facilitating collaborative work among the archaeologist community.

Acknowledgment

This work was supported by the Labex Archimède «Archéologie et Histoire de la Méditerranée et de l’Égypte anciennes» [ANR-11-LABX-0032-01].

References

- Bats, M., Fiches, J.-L. and Py, M. (eds) 1986. *Enregistrer la fouille archéologique: le système élaboré pour le site de Lattes (Hérault)*. Lattes.
- Dicocer 2017. *Dicocer, dictionnaire des céramiques anciennes*. <<http://dicocer.cnrs.fr/>> [Accessed May 15, 2018].
- Harris, E.-C. 1979. *Principles of Archaeological Stratigraphy*. London: Academic Press, 136 p.
- Pégase 2018. *Pégase, Référentiel ostéologique de comparaison*. <<http://pegase.cnrs.fr/>> [Accessed May 15, 2018].
- Py, M. (ed.) 1991. *Système d’enregistrement, de gestion et d’exploitation de la documentation issue des fouilles de Lattes*. Lattara 4. Lattes, 224 p.
- Syslat 2018a. *Syslat SIA – Système d’Information Archéologique*. <<http://www.syslat.fr/>> [Accessed May 15, 2018].

Syslat 2018b. *Système d'Information Archéologique - Syslat Horizon 2020*. <<https://syslat.hypotheses.org/>> [Accessed May 15, 2018].

L'archéologie néoprocessuelle

François Djindjian

Université de Paris 1 Panthéon Sorbonne et CNRS UMR 7041 Arscan

francois.djindjian@wanadoo.fr

Résumé

L'archéologie néoprocessuelle est une nouvelle plate-forme théorique, dont l'objectif et les méthodes sont de renouveler le cadre théorique de l'archéologie après trente années de post-modernisme. L'archéologie néoprocessuelle s'intéresse à définir et caractériser les processus métiers de l'archéologie, les processus taphonomiques et les processus systémiques des sociétés du passé. Dans cet article, ont été particulièrement approfondis les processus taphonomiques, lato sensu, et deux processus systémiques liés aux approches typologiques du XIX^e siècle : types et « cultures » préhistoriques et protohistoriques, pour lesquels l'approche par les processus montre sa capacité à identifier, caractériser et reconstituer les systèmes techniques et les systèmes « culturels ».

Mots-clés : archéologie théorique, processus, système

Abstract

Neoprocessual archaeology is a new theoretical platform, whose objective and methods is to renew the theoretical framework of archaeology after thirty years of post-modernism. Neoprocessual. archaeology is interested in defining and characterizing the business processes of archaeology, the taphonomic processes and the systemic processes of past societies. In this article, the taphonomic processes, lato sensu, and two systemic processes related to the typological approaches of the nineteenth century were particularly detailed: prehistoric and protohistorical types and 'cultures', for which the approach by the process demonstrates its ability to identify, characterize and reconstruct technical and 'cultural' systems.

Keywords: archaeological theory, process, system

1. Introduction

Depuis les années 1990, l'archéologie, comme toutes les sciences humaines et sociales, est sous l'influence du post-modernisme, malgré le fait que la plupart des apports méthodologiques et théoriques depuis le début des années 1960 soient issues des sciences « exactes » (physique, chimie, mathématiques et informatique), des sciences biologiques (paléogénétique), des sciences de la terre (géoarchéologie, paléontologie, archéobotanique, archéozoologie, paléoclimatologie).

Presque trente ans après, la contribution du post-modernisme à l'archéologie s'est concrétisée dans la déconstruction (à tort ou à raison), et dans l'émergence d'une archéologie militante et relativiste, hostile à la quantification et à la rigueur de raisonnements explicites, démonstratifs et réfutables, et faisant plonger l'archéologie dans les enfers de l'herméneutique.

Certes, les réalisations de la « New Archaeology » anglo-saxonne des années 1960, n'avaient pas été à la hauteur de ses ambitions. La « New Archaeology » était un mélange hétérogène et parfois contradictoire de plusieurs approches scientifiques : un positivisme logique, le recours aux modèles hypothético-déductifs et l'utilisation systématique des données quantitatives, sans appareil critique associé. Les paradigmes préférés des adeptes de la « New Archaeology » étaient l'anthropologie (« *l'Archéologie est Anthropologie ou rien* »), l'évolutionnisme culturel, le fonctionnalisme (Binford), l'écologie culturelle (Steward) et la théorie des systèmes (Flannery). Cette nouvelle approche aurait donc permis d'identifier des processus culturels, à l'origine du

mot « archéologie processuelle », qui aurait une valeur universelle ou locale. De ce fatras mal digéré de tendances scientifiques diverses nés du dynamisme exceptionnel des universités américaines dans l'après deuxième guerre mondiale, aurait pu naître réellement une nouvelle archéologie si la jeune génération d'anthropologues qui l'avait prônée, avait possédé un background scientifique suffisant pour éviter les pièges de la démesure voire même d'une ivresse dans l'utilisation de concepts et de mots mal compris, et à des résultats caricaturaux faisant la joie des archéologues restés classiques et permettant au post-modernisme d'en prendre la suite. Il n'en reste pas moins que l'aventure de la « New Archaeology » restera un moment clé dans le développement de l'archéologie mondiale.

Face à ces deux écoles trentenaires (1960-1990 pour la première ; 1990-2020 pour la deuxième), dont nous connaissons bien aujourd'hui les apports et surtout les faiblesses, il est indispensable, pour continuer à construire une archéologie du futur, de proposer une nouvelle approche, qui par l'effet bien connu de l'alternance, visera à relancer une archéologie processuelle, débarrassée, autant que faire se peut, de ses faiblesses. Cette nouvelle approche, nous l'appellerons ici l'archéologie néoprocessuelle, non pas pour en faire une relance de l'archéologie processuelle de la « New Archaeology », qui était dans la réalité si peu processuelle, mais pour démontrer que l'archéologie du XXI^e siècle est déjà engagée naturellement dans une archéologie des processus (que ce soit des processus métiers, des processus taphonomiques ou des processus systémiques).

2. L'archéologie néoprocessuelle

L'archéologie néoprocessuelle est une approche théorique de l'archéologie qui privilégie l'identification, la caractérisation et l'étude des changements des processus qui sont le cœur du fonctionnement des systèmes. A ce titre, l'archéologie néoprocessuelle est une archéologie systémique, comme l'était la « New Archaeology ». Mais elle ne fait aucune réduction systémique comme le fonctionnalisme (Binford), l'écologie culturelle (Steward) ou l'archéologie évolutionniste (Dunnell, Shennan).

Dans cette approche, l'accent est mis sur les changements d'un système et sur les variables qui sont à l'origine de ces changements. Les données archéologiques enregistrées sont considérées comme un état du système, à un instant *t* : système en naissance, en développement, en stabilisation, en régression ou en effondrement.

En archéologie, plusieurs types processus peuvent être considérés : les processus métiers de l'archéologie, les processus taphonomiques et les processus systémiques.

3. Les processus métiers de l'archéologie

Les processus métiers de l'archéologie définissent l'organisation et les fonctions du métier d'archéologue dans ses différents organismes et institutions : chercheur et ITA, enseignant-chercheur, conservateur (patrimoine, musée ou archéologie préventive (publique, institutionnelle, privée)). C'est la méthode aujourd'hui universellement appliquée du BPM (*Business Process Management*) qui, appliquée aux métiers de l'archéologie, devient « *Archaeological Process Management* ». Nous renvoyons le lecteur à des précédentes publications (Djindjian 2002, 2012, 2014) qui décrivent l'application de cette approche formelle et informatique et nous n'y reviendront pas ici.

4. Les processus taphonomiques

La taphonomie est l'étude des processus que suit une entité biologique depuis la biosphère de son vivant, jusqu'à la lithosphère au moment de sa découverte à l'état fossile (Chaix, Meniel, 2001 ; Denys *et alii*, 2014). Dans l'article Wikipedia, il est précisé : « *Le concept a été défini par le*

paléontologue Efemov en 1940 ». Cette définition a été précisée par Behrensmeyer et Kidwell en 1985 : « *the study of the processes of preservation and how they affect information in the fossil record* », et par Koch (1989) : « *a process or chain of events that begins just before the death of an organism, and proceeds through decomposition, disarticulation, burial, fossilization, exposure, and collection. At each stage of this process, various taphonomic agencies intervene to obscure, bias, and add to the information that is available from the resulting collection* ».

Plusieurs remarques peuvent être faites ici. Tout d'abord, les définitions données mettent bien en évidence que les phénomènes observés sont des processus. Ensuite, le terme taphonomique, à l'origine paléontologique, possède une généralité bien plus grande, applicable à l'ensemble des entités biologiques animales et végétales (pollens, charbons de bois, macro-restes végétaux, molaires de rongeurs, coprolithes, etc.), et bien au-delà, à toutes les sciences de la terre, et plus particulièrement en archéologie dans les processus post-dépositionnels de remplissage de grottes et d'abris sous roche (remplissage karstique, érosions, lessivages, dépôts clastiques, spéléothèmes, etc.), et de sédimentation en plein air (alluvionnement, colluvionnement, dépôt éolien, ensablement, dépôt de pente, phénomènes périglaciaires, plages fossiles, dépôts d'origine volcanique), sans oublier les activités animales et végétales et les bouleversements anthropiques.

Mais, bien plus, et comme il a été dit précédemment, la taphonomie met en relation l'échantillon enregistré et la population cible de l'étude, et, dès lors, l'étude des processus taphonomiques peut permettre de mettre en évidence les biais entre échantillon et population. Nous retrouvons ici les objectifs de la « *Behavioural Archaeology* » de B. Schiffer, qui, en 1976, mettait en évidence les biais apportés à la reconstitution archéologique par les processus de formation, d'abandon et de conservation des sites archéologiques ce qui l'amenait à distinguer la culture systémique (système culturel réel) de la culture archéologique (système reconstitué à partir du seul enregistrement des données archéologiques par les archéologues). L'étude des processus de formation, d'abandon et de conservations des sites archéologiques permet aujourd'hui de mieux déterminer les biais et donc d'éviter les pièges de non représentativité que dénonce le paradigme de la « *Behavioural archaeology* ».

En archéozoologie, quelques exemples peuvent être donnés de ces processus taphonomiques :

- processus d'origine des accumulations d'ossements, qu'elles soient uniques ou répétitives naturelle (dépôt de charriage, piège, catastrophe naturelle), ou anthropique (chasse, piégeage, feu, etc.),
- processus de dépeçage des animaux (boucherie),
- processus de transport de parties de carcasse,
- processus d'exploitation des carcasses (fragmentation, matériaux de construction, fabrication d'objets, combustible),
- processus d'altération à l'air libre,
- processus d'actions des carnivores,
- processus d'enfouissement,
- processus de fossilisation (et de conservation différentielle).

Il en est de même en géoarchéologie pour l'étude des processus post-dépositionnels qui peuvent être décomposés en processus d'enfouissement, processus de conservation des objets enfouis (milieu basique, milieu acide, milieu désertique, milieu humide, milieu gelé) et processus de perturbation des sites et des objets enfouis : actions anthropiques postérieures (piétinements, fosses, fondations, etc.), actions des insectes, des vers de terre, des animaux fouisseurs (blaireaux, taupes, lapins), actions des animaux de cavernes (ours, hyène, loup, renard, félins, etc.), actions de gel et de dégel en surface (fentes de gel et sols polygonaux), actions de cryoturbation, actions de solifluxion, actions de ruissellement, actions d'érosion, actions de sous-tirage en grotte karstique, actions liées à l'infiltration de l'eau et à ses pouvoirs de solvant (percolation, carbonatation, oxydation),

réactions chimiques (oxydations, patines, migrations d'éléments chimiques), disparition des matières organiques, fossilisation.

Certains phénomènes peuvent détruire le niveau d'occupation archéologique en provoquant des remaniements complets (mélange de niveaux, inversion stratigraphique, vidanges déplaçant le niveau sur la pente d'un versant). D'autres provoquent des remaniements partiels dont il importe de déterminer l'amplitude par des analyses statistiques fines en étudiant les localisations, les orientations et les pendages d'objets dans un niveau archéologique (Bertran, Lenoble 2002).

C'est également le cas pour la question de la découverte et de la conservation des sites archéologiques, qui doit faire l'objet de l'étude de plusieurs processus :

- un processus de choix de la localisation du site et de son urbanisation,
- un processus de formation et de développement du site (continûment ou par phase dans un temps défini),
- un processus d'entretien (structures d'habitat, structures funéraires, infrastructures comme des canaux d'irrigation ou des routes, paysages cultivés, etc.),
- un processus de réparation ou de rénovation (aménagement successifs),
- un processus de transformation (en cas de modification de son état et de sa fonction),
- un processus de réemploi (particulièrement pour des matériaux de construction mais aussi pour des bâtis, des structures funéraires, etc.),
- un processus d'abandon ou de destruction d'origine naturelle ou anthropique,
- un processus d'enfouissement,
- un processus de conservation post-dépositionnel,
- un processus de découverte (prospection archéologique ou découverte fortuite liée à des travaux).

C'est dans ce contexte que la stratigraphie confirme être bien la méthode de base de la science archéologique, que ce soit des remplissages d'origine essentiellement naturels ou des remplissages essentiellement anthropiques (archéologie urbaine). La représentation stratigraphique peut être formalisée par un graphe, dont les matrices de Harris sont un bon exemple (Harris *et alii* 1993). La méthode stratigraphique est bien un processus : formalisation par un graphe et réorganisation automatique par un algorithme d'analyse des sonnées (Desachy, Djindjian 1990 ; Djindjian 2011, chapitre 5).

5. Les processus systémiques des sociétés du passé

Il faut insister sur le fait que le but de l'archéologie, la redécouverte du passé, n'est autre que la reconstitution des systèmes que les sociétés du passé ont développés, systèmes que nous étudions pour chercher à les caractériser et à en étudier les changements. Ces systèmes sont définis et caractérisés par des processus, dont les plus critiques nous intéressent en priorité, et parmi ceux-ci, ceux qui ont laissé suffisamment de vestiges matériels pour permettre de les reconstituer :

- Procédés de fabrication (systèmes technologiques),
- Procédés de construction (habitations, monuments, infrastructure),
- Processus de gestion des ressources alimentaires,
- Processus d'approvisionnement en matières premières,
- Processus économique (systèmes d'échanges et de commerce),
- Processus d'occupation du territoire,
- Processus de changement culturel,
- Processus d'adaptation à l'environnement et au changement climatique,
- Processus fonctionnels,
- Processus de gouvernance sociétale,
- Processus administratifs,

- Processus d'organisation sociale,
- Processus d'attitude sociétale,
- Processus de sacralisation de la société (religions, tabous et croyances populaires).

Certains de ces processus sont plus faciles à étudier (notamment les systèmes technologiques liés à l'abondance de la culture matérielle), d'autres sont plus faciles à quantifier (ceux notamment liés à l'environnement et au changement climatique). Les processus que l'on classe dans les sciences humaines et sociales sont les plus difficiles à étudier car ils laissent peu de vestiges matériels. Il existe évidemment une différence considérable dans l'étude des sociétés préhistoriques (limitée par la capacité cognitive des vestiges matériels à caractériser le système étudié) et l'étude des sociétés historiques (limitée par la véracité des textes disponibles).

5.1. Les procédés de fabrication

Nous avons détaillé, dans deux articles précédents (Djindjian 2012, 2013), la formalisation de bout en bout d'un procédé de fabrication, et nous n'y reviendrons donc pas ici.

Un procédé de fabrication est un processus qui peut être décrit en identifiant les gestes et la séquence des gestes de l'artisan (exactement de la même façon que pour un robot). Elle implique en conséquence la création d'un vocabulaire identifiant et décrivant précisément ces différents gestes.

Cette séquence de gestes peut être représentée par un graphe. Le réseau de Petri est le graphe le mieux adapté à la formalisation d'un procédé de fabrication. Les graphes de type réseau de Petri peuvent être quantifiés, transformés en un tableau de Burt et traités par des techniques d'analyse multidimensionnelle des données.

Un exemple d'application de cette méthode formelle est l'étude des évolutions techniques des procédés de fabrication des burins sur la séquence de l'abri de La Ferrassie en Périgord (Djindjian 1980, 1996).

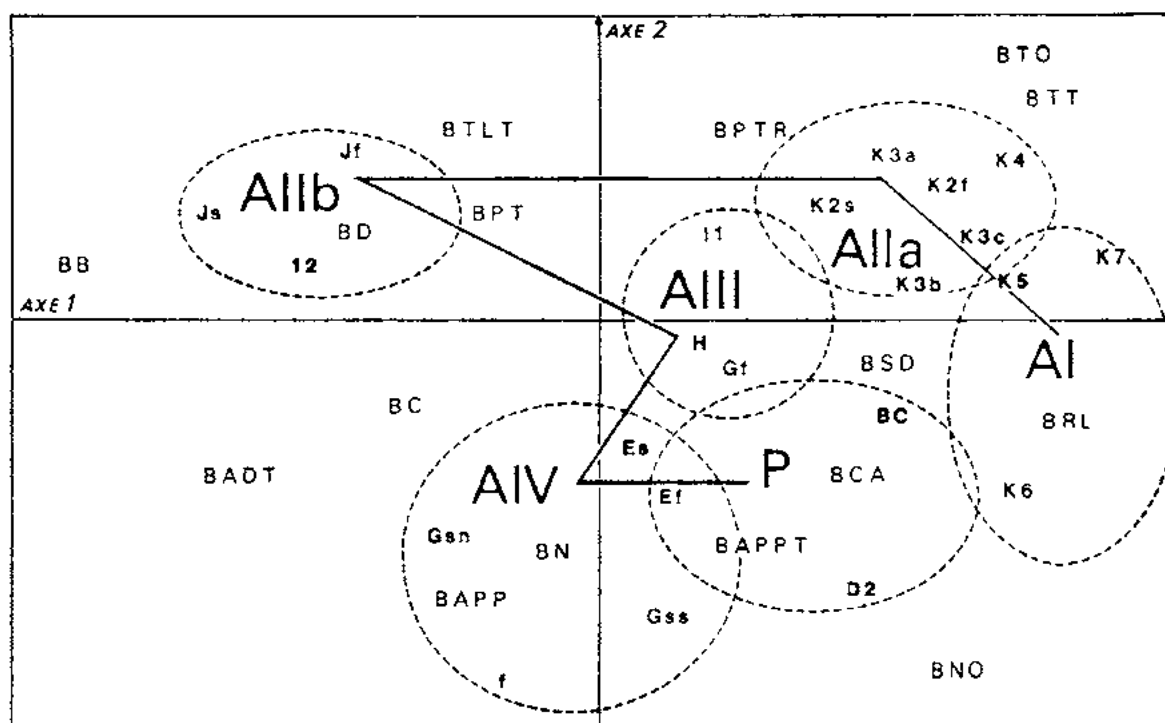


Figure 1. Evolution des procédés de taille des burins dans l'Aurignacien et le Gravettien du site paléolithique supérieur de la Ferrassie (Dordogne)

5.2. Les « cultures » préhistoriques

Le concept de culture est une conséquence directe de l'archéologie typologique définie par Montélius au XIX^e siècle, en nommant des assemblages typologiques similaires et en les situant dans un espace-temps connu qui leur est propre. Les noms, issus généralement des premiers sites publiés où ces assemblages ont été identifiés (éponymes), sont le plus souvent restés dans la littérature, et, par exemple pour le Paléolithique supérieur, sont toujours utilisés aujourd'hui, même si leur contenu sémantique a varié dans le temps jusqu'à ne plus devenir à tort pour certains qu'une commodité d'écriture (Djindjian *et alii* 1999).

Que signifient donc ces « cultures » ? Ces « cultures » du paléolithique supérieur sont des processus d'adaptations du système de chasse et de cueillette des groupes humains dans un environnement lié à une géographie physique et un climat. Ces « cultures » se particularisent, se développent, régressent et se transforment sous le double effet de la tradition qui perpétue le système et des changements d'environnement qui font évoluer le système. Ces systèmes sont caractérisés par des processus dont l'identification devient alors l'objet principal de l'étude de ces systèmes : mobilité des groupes, superficie des territoires de déplacements, accessibilité des gîtes d'approvisionnement en matière première, disponibilité de ressources alimentaires dans le cycle annuel, zoocénoses chassées, stratégies de gestion du territoire, saisonnalité des habitats, densité démographique, etc.

Plusieurs grandes catégories de processus peuvent être ainsi mises en évidence : stabilité, différenciation, uniformisation, expansion géographique, cloisonnement, effondrement, redéploiement, etc., dont les passages de l'un à l'autre sont principalement liés dans les sociétés de chasseurs-cueilleurs à des variations climatiques entraînant des changements de végétation et de faune tout en conservant par tradition une résilience du processus précédent.

Les groupes de chasseurs-cueilleurs avaient colonisé l'Europe à partir du dernier million d'années en profitant des épisodes climatiques interglaciaires (notamment les stades isotopiques 15, 13 et 11). L'avant-dernière glaciation (stades isotopiques 8 à 6), qui a été le plus froid épisode glaciaire de dernier million d'années, a vu les groupes humains rester en Europe et, par un processus d'adaptation au climat glaciaire, changer morphologiquement, devenant ce que les préhistoriens ont appelé l'homme de Neandertal.

Le stade isotopique 5, il y a entre 115 000 et 75 000 ans, qui a été une période interglaciaire, a vu en Europe l'homme de Neandertal, profiter de cet environnement plus favorable, pour coloniser les latitudes plus septentrionales de l'Europe et à l'Est jusqu'en Asie. C'est un processus de développement liée à une expansion géographique entraînant une croissance démographique, avec une dynamique d'innovations technologiques avec l'invention du débitage laminaire, qu'accompagne le développement du débitage Levallois et du débitage bifacial, associé à une plus grande mobilité des groupes, dont nous ne connaissons pas encore les modalités de gestion de leur territoire.

Le stade isotopique 4, il y a entre 75 000 et 55 000 ans, voit le retour d'une phase glaciaire. En Europe, les groupes humains refluent vers la moitié sud du continent, quand ils ne migrent pas vers un Moyen-Orient plus favorable. C'est un processus de stagnation / récession. Le type morphologique néanderthalien s'accroît. Les groupes humains, cloisonnés par la géographie montagneuse de l'Europe, ont une mobilité faible limitée à un territoire d'environ 1 000 km², avec des approvisionnements opportunistes locaux inférieurs à 30 km, et une chasse ubiquiste privilégiant les embuscades. L'industrie sur éclat, se limite pour l'essentiel à des encoches, denticulées et racloirs, le débitage Levallois diminue fortement et le débitage laminaire disparaît.

Le réchauffement climatique progressif de la fin du stade isotopique 4, il y a 45 000 ans, voit l'émergence d'un processus de diversification technique, avec la réapparition de pièces bifaciales, de

couteaux, des pointes Levallois laminaires. Le débitage Levallois se redéveloppe, tandis qu'apparaît le débitage lamellaire et le débitage laminaire. Le cloisonnement géographique de l'Europe, encore présent, entraîne une différenciation régionale de cette diversification technique, qui a été identifiée sous de nouveaux noms : moustérien de tradition acheuléenne en Europe occidentale, Néronien dans la vallée du Rhône, Micoquien en Europe centrale et orientale, Bohunicien/Kremenicien en Europe centrale et orientale, etc.

Le début du stade isotopique 3, il y a 40 000 ans, voit le retour d'un climat plus tempéré, interpléniglaciaire, qui va favoriser la transformation de la culture matérielle des groupes humains en Europe, avec l'apparition d'industries, appelées à tort « industries de transition », mais qui sont en fait déjà des industries de type paléolithique supérieur, caractérisées par l'usage prépondérant d'un débitage prismatique laminaire et lamellaire. La multiplicité de ces industries marque un processus de polymorphisme régional. Les groupes humains exploitent encore des territoires restreints avec des approvisionnements en matière première et des ressources alimentaires locales : Châtelpéronien en Europe occidentale, Uluzzien en Italie, Aurignacien 0 dans la haute vallée du Danube, Szélétien en Europe centrale, Strélétien en Europe orientale, Ouralien dans le Nord-est de l'Europe orientale, Jerzmanowicien/Lincombien/Ranisien en Europe septentrionale. La perdurance d'un Moustérien dans les régions méditerranéennes (sud de la péninsule ibérique, Crimée) et l'auteur de ces industries (supposé être Neandertal) reste encore un sujet de controverses.

Le stade isotopique 3 est une période interpléniglaciaire à oscillations tempérées (anciennement nommées Les Cottès, Arcy, Maizières). La période entre 35 000 et 29 000 BP, voit le développement de l'atlantique à l'Oural et au Moyen-Orient d'une industrie unique : l'Aurignacien, qui marque un processus d'uniformisation. Les territoires de déplacements des chasseurs-cueilleurs se sont agrandis progressivement jusqu'à 500 000 km², révélant une croissance démographique, une plus grande mobilité et une exploitation du territoire caractérisés par des approvisionnements lointains pour des gîtes de bonne matière première en silex et des chasses saisonnières spécialisées de troupeaux migrateurs. Les oscillations climatiques voient des variations entre débitage lamellaire (épisodes plus tempérés) et débitage laminaire (épisodes plus froids).

Le stade isotopique 2 correspond au retour de la dernière glaciation connue. Ses débuts, de 29 000 à 26 000 BP, appelé souvent le pléniglaciaire supérieur ancien, correspond à un processus d'adaptation progressive au refroidissement. L'industrie de l'Aurignacien se transforme en Gravettien ancien, dans la haute vallée du Danube, par des innovations techniques comme des pointes en silex variées : fléchettes, pointes à pédoncules (pointes de la Font-Robert), pointes à emmanchement latéral (pointes de la Gravette, microgravettes, lamelles à dos) et des burins pour produire des lamelles en série (Raysse, Noailles). Les territoires étendus de déplacement, l'approvisionnement lointain en matière première, la chasse saisonnière aux troupeaux migrateurs, la mobilité restent identiques.

La suite du stade isotopique 2, entre 26 000 et 22 000 BP, à l'approche du dernier maximum glaciaire, voit la différenciation régionale des industries du Gravettien moyen et du Gravettien récent en Europe occidentale et du Gravettien oriental, qui correspond à un processus de différenciation régional par cloisonnement géographique.

Le début du dernier maximum glaciaire vers 22 000-21 000 BP voit un changement brutal, correspondant à un processus d'effondrement (collapse) du système précédent. Les industries associées sont le Gravettien final (Protomagdalénien de D. Peyrony en France), les industries aurignacoïdes, le Solutréen inférieur (en Europe occidentale). Les groupes humains refluent vers les zones méditerranéennes, abandonnant l'Europe moyenne, entraînant l'abandon des systèmes de subsistance et des approvisionnements en silex et en matières dures animales (bois de renne, ivoire). Les territoires de déplacement redeviennent restreints avec une chasse aux mammifères grégaires et un approvisionnement opportuniste local en matière première.

Au cours du dernier maximum glaciaire, un épisode climatique plus humide (mais pas plus tempéré), autour de 20 000 BP, crée une situation plus favorable, permettant d'enclencher un processus de redémarrage, basé sur des déplacements estivaux en Europe moyenne à partir des refuges méditerranéens. Il permet de retrouver les sites d'approvisionnement en bon silex. Ce processus correspond au Solutrén récent en Europe occidentale et à l'Épigravettien ancien qui se différencie régionalement de la Provence à la Mer noire. Des innovations substitutives font apparaître des pointes foliacées en silex (feuille de laurier, feuille de saule, pointe à cran, pointe arénienne, etc.).

A la fin du dernier maximum glaciaire, entre 18 500 et 17 000 BP, un autre épisode climatique plus humide, permet cette fois de démarrer un processus de réoccupation permanente, accompagnées d'incursions estivales vers le Nord. C'est le cas du Badegoulien en Europe occidentale en Aquitaine, de la culture de Zamiatnine en Europe orientale (Don moyen) et du Sagvarien en Europe centrale (plaine de Pannonie). La disparition des pointes en silex remplacées par des pointes en bois de renne, révèle le retour de la chasse au renne et à l'approvisionnement de ses bois. Le retour du débitage sur éclat et d'un outillage frustre est à noter. Parallèlement, le peuplement en région méditerranéenne, par un processus de perduration, se maintient par des industries épisolutréennes et épigravettiennes.

La fin du stade isotopique 2, à partir de 17 000 BP jusque vers 12 500 BP, se traduit par un climat froid et sec, le pléniglaciaire supérieur récent, qui avec un réchauffement progressif et un retour de l'humidité, va permettre le démarrage d'un processus de recolonisation de l'Europe moyenne puis de l'Europe septentrionale. C'est le cas en Europe occidentale du Magdalénien, à partir d'un foyer aquitaino-cantabrique, qui atteint rapidement la vallée de la Loire puis pénètre en Europe centrale jusqu'en Pologne. Pendant l'épisode de Bölling, il recolonisera l'Europe septentrionale avec un processus de différenciation régionale (Magdalénien supérieur, Hambourgien, Cresswellien). En Europe orientale, le phénomène est analogue avec le Mézinien, dans le bassin moyen et supérieur du Dniepr.

La fin du stade isotopique 2 (épisode tempéré et humide d'Alleröd) révèle un changement rapide de la végétation et des zoocénoses animales, à l'origine d'un processus déjà nommé d'azilianisation, qui est un nouveau processus d'uniformisation qui concerne toutes les industries européennes (faciès à pointes à dos courbe) : Azilien, Épigravettien final. L'occupation humaine est marquée par un retour à des territoires restreints avec une gestion des ressources diversifiée et opportuniste pour l'alimentation et les matières premières.

Le court épisode froid du Dryas récent, avant le début de l'Holocène (stade isotopique 1) marque un processus de repli et de réadaptation au froid avec des différenciations régionales importantes (en France, Ahrensbourgien, Belloisien, Laborien, Montadien).

Le début de l'Holocène, vers 11 000 BP, avec le grand développement de la forêt tempérée, voit l'adaptation des groupes humains à la chasse et à la cueillette en milieu forestier, à la pêche et au ramassage de coquillages en milieu côtier, qui se traduit au niveau des industries par un processus de microlithisation des armatures, pour l'usage de l'arc. Les territoires sont restreints. Une différenciation locale se traduit dans la morphologie des microlithes.

Cette approche par les processus retrouve la dynamique naturelle de l'adaptation et de l'évolution des systèmes humains, qui était perdu par l'approche typologique figée des « cultures », dont le principal inconvénient était de pousser le préhistorien à identifier ces « cultures » à la façon de Kossinna, d'imaginer des migrations, des contemporanéités et donc d'empêcher de voir les changements culturels, dont la rapidité (quelques générations) est largement inférieure à la précision des datations radiocarbone (aujourd'hui 500 à 1 000 ans).

Elle cherche en outre à trouver des explications aux choix techniques et fonctionnels qui caractérisent les assemblages d'objets au lieu de les considérer a priori et sans justification comme des marqueurs « culturels ». Cette approche a vu des travaux pionniers, comme ceux de G. Laplace (1966) qui a notamment proposé l'hypothèse du synthétype aurignacien et le processus d'azilianisation, travaux influencés par N. Vavilov (1935) et sa théorie des centres d'origine des plantes cultivées.

L'utilisation du modèle de résilience écologique de C.S. Holling (1978) possède une vertu pédagogique plus que démonstrative. Ce modèle figure, d'une façon très robuste mais illustrative, l'évolution d'un système dans ses différentes phases successives : croissance, stabilisation, régression et effondrement, renouveau. La figure 2 montre ainsi une illustration de l'évolution des « cultures » du paléolithique supérieur européen depuis les débuts du stade isotopique 3 jusqu'à la fin du stade isotopique 2. La figure 3 montre une illustration de l'évolution des systèmes d'exploitation des territoires des groupes humains depuis les débuts du stade isotopique 3 jusqu'à la fin du stade isotopique 2. La figure montre une illustration de l'évolution des procédés de fabrication de l'industrie lithique depuis les débuts du stade isotopique 8 jusqu'à la fin du stade isotopique 2.

Figure 2. Evolution des « cultures » du paléolithique supérieur européen depuis les débuts du stade isotopique 3 jusqu'à la fin du stade isotopique 2, illustrée par le modèle de Holling.

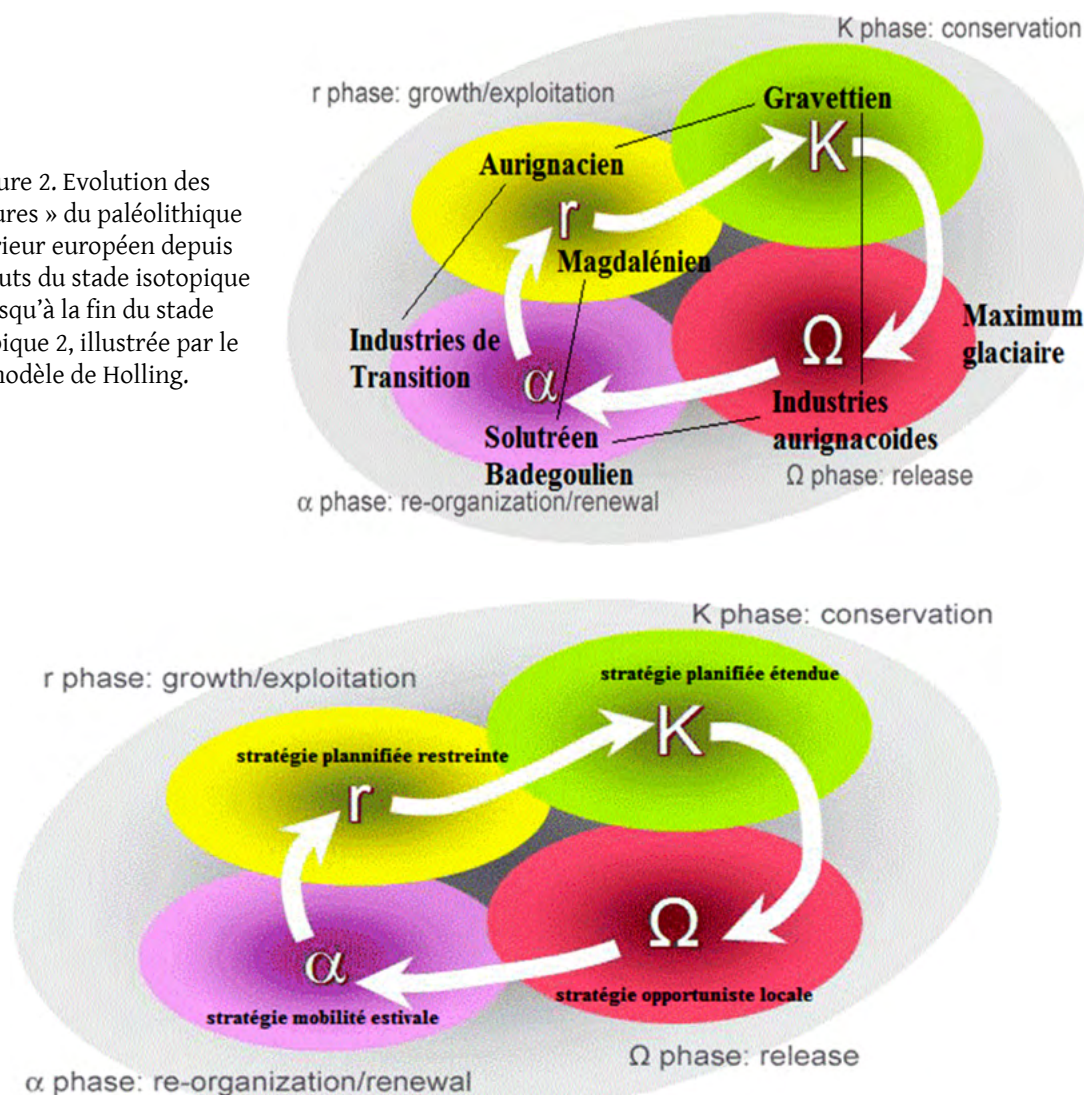


Figure 3. Evolution des systèmes d'exploitation des territoires des groupes humains depuis les débuts du stade isotopique 3 jusqu'à la fin du stade isotopique 2, illustrée par le modèle de Holling.

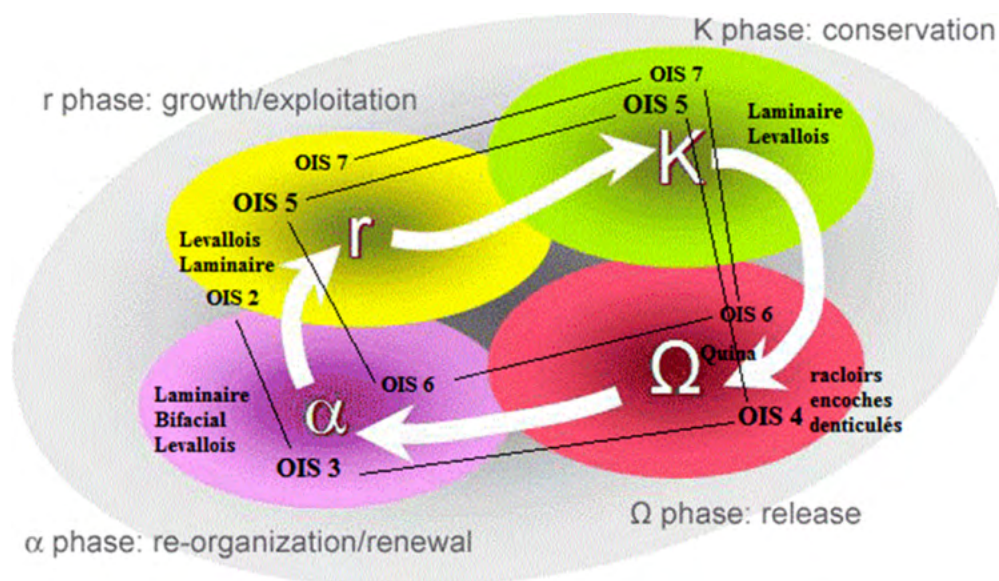


Figure 4. Evolution des procédés de fabrication de l'industrie lithique depuis les débuts du stade isotopique 8 jusqu'à la fin du stade isotopique 2, illustrée par le modèle de Holling.

5.3. Processus d'adaptation à l'environnement et aux changements climatiques

Les processus d'adaptation des sociétés de chasseurs-cueilleurs et des sociétés d'agriculteurs-éleveurs au changement de l'environnement ou au changement climatique, ont fait l'objet de nombreuses études récentes utilisant la simulation informatique par les systèmes multi-agents. En effet, les variables d'environnement peuvent être assez aisément enregistrées et quantifiées : changement de température, changement dans les précipitations (aridité, humidité), érosion des sols, réduction de la végétation, catastrophes naturelles. Ces données seront mises en relation avec les données des productions humaines : gestion des ressources alimentaires dans le cycle annuel, rendement des champs cultivés, spécialisations agricoles, assolements, épuisement des sols, aménagements du paysage (terrasses, irrigation), épidémies humaines, animales et végétales, etc.

Le projet EOS (Doran *et alii* 1994), le « Village Ecodynamics Project » (Kohler *et alii* 2000, 2012), le projet Anasazi (Dean *et alii* 2000) ou le projet Enkimdu en Mésopotamie (Wilkinson *et alii* 2007) sont de bons exemples d'applications de cette approche. Mais il faut signaler un danger de ces études qui serait celui de se limiter aux seules variables environnementales et économiques, qui amèneraient à des conclusions mécaniquement malthusiennes (cf. infra §6).

6. Processus SHS

Dans deux récents articles (Djindjian 2012, 2014), nous avons développé l'importance du rôle des processus des sciences humaines et sociales dans l'étude des sociétés du passé :

- Processus d'organisation sociale,
- Processus de gouvernance,
- Processus d'attitudes sociétales,
- Processus de sacralisation de la société.

Le tableau de la figure 5 montre les différents sous-systèmes concernés par une étude de modélisation globale d'une société, que ce soit par des systèmes multi-agents ou par d'autres méthodes. Il est enfin un dernier processus difficile à formaliser, que les historiens connaissent

Figure 5. Processus concernés par la simulation globale d'une société complexe.

Systèmes environnementaux		Systèmes de production systèmes agro-alimentaires systèmes d'artisanat systèmes énergétiques systèmes comerciaux systèmes de distribution systèmes de transport systèmes financiers	
Organisation territoriale			
Administration Défense Police Diplomatie Taxation	Systèmes de gouvernance		Tribus Butins Esclaves
Savoirs Savoir-faire	Croyances Religions	Règles sociales	Attitudes sociétales

bien, c'est celui de la gestion de crise, quand un évènement (d'origine naturelle ou humaine) survient et fait basculer le sens de l'histoire dans un sens qui aurait pu être l'inverse.

7. Conclusions

Trois ensembles de processus concernent l'archéologue :

- Les processus métiers de l'archéologie,
- Les processus taphonomiques, sensu lato, qui jouent un rôle considérable dans la mise en relation entre les données archéologiques enregistrées et les populations cibles étudiées,
- Les processus systémiques des sociétés du passé.

Ce trop rapide parcours à travers les différents processus qui sous-tendent le fonctionnement des systèmes qui constituent les sociétés du passé n'a pour but ici que de montrer l'énorme potentialité de cette approche, qui, à nos yeux, représente une révolution aussi importante que celle de l'approche typologique des archéologues scandinaves du XIX^e siècle.

Sans ce nouveau cadre théorique, l'exploitation des données nouvelles, que permettent les développements de la méthode archéologique, n'apporteraient pas de notables résultats nouveaux et ne permettrait pas de résoudre des questions restées sans solutions jusqu'à aujourd'hui.

L'exemple de la « chaîne opératoire » est caractéristique de cette difficulté. Cette approche, qui essentiellement s'intéresse à étudier les procédés de fabrication dans le débitage lithique (plus variés d'ailleurs dans la paléolithique moyen que dans le paléolithique supérieur), a certes abouti à mettre en évidence des types de débitage nouveaux, voire à mieux les caractériser, remplaçant les types de façonnage des listes typologiques des années 1960. Mais la variabilité des industries lithiques du paléolithique moyen reste toujours une question sans solution, que ne peut résoudre la seule approche typologique du débitage, comme n'a pu le résoudre l'approche par les types de façonnage (Bordes) ou l'approche fonctionnelle (Binford). Une approche par les

processus apporte pourtant la voie d'une solution quels sont en effet les processus responsables de la variabilité des industries lithiques du paléolithique moyen : qualité variable des matières premières, faible mobilité des groupes humains, facteurs chronologiques liés aux variations climatiques, etc.?

Bibliographie

- Behrensmeyer, A., Kidwell, S. 1985. Taphonomy's contributions to Paleobiology. *Paleobiology*, 11, 1: 105-119
- Bertran, P., Lenoble, A. 2002, Fabrique des niveaux archéologiques: méthode et premier bilan des apports à l'étude taphonomique des sites paléolithiques, *Paleo*, 14, 13-28.
- Chaix, L., Meniel, P. 2001. Archéozoologie. Paris : Errance.
- Dean, J.S., Gummerman, G.J., Epstein, J.M., Axtell, R.L., Swedlund, A.C., Parker, M.T., McCarroll, S. 2000. Understanding Anasazi Culture Change Through Agent-Based Modeling, in 'A. Timothy Kohler, G.J. Gummerman (éds.), Dynamics in Human and Primate Societies, Agent-Based Modeling of social and spatial processes'. New York-Oxford: Oxford University Press, 179-207.
- Denys, Ch., Patou-Mathis, M. eds. 2014. Manuel de taphonomie, Paris: Errance, coll. « Archéologiques ».
- Desachy, B., Djindjian, F. 1990. Sur l'aide au traitement des données stratigraphiques des sites archéologiques, *Histoire et mesure*, 5, 1-2, 51-88.
- Djindjian, F. 1980. Faciès chronologiques aurignaciens et périgordiens à La Ferrassie (Dordogne), in L'analyse des objets archéologiques et les procédés statistiques d'interprétation, *Les Dossiers de l'Archéologie*, 42, Dijon : Archéologia, 70-74.
- Djindjian, F. 1996. Histoires de burins, *Bulletin du Centre Genevois d'Anthropologie*, 1993-4, 3-21.
- Djindjian, F. 2002. Pour une théorie générale de la connaissance en Archéologie, *Archeologia e Calcolatori*, 13, 101-117.
- Djindjian, F. 2011. *Manuel d'archéologie*. Paris : Armand Colin.
- Djindjian, F. 2012. L'approche par les processus en archéologie. Actes des 2^e Journées Informatique et Archéologie de Paris – JIAP 2010, Paris, 11-12 Juin 2010. *Archeologia e Calcolatori*, Supplément 3, 2012, 279-300.
- Djindjian, F. 2013. Us et abus du concept de chaîne opératoire dans l'archéologie française contemporaine. In 'Archéologie de l'âge du Fer en Europe. Mélanges offerts à Olivier Buchsenschutz ; Krausz, S., Colin, A., Gruel, K., Ralston, I., Dechezleprêtre, Th. Edts'. Bordeaux : Ausonius, collection Mémoires n°32, 93-107
- Djindjian, F. 2014. Simuler une 'Artificial Society': organisation sociale, gouvernance et attitudes sociétales. Actes des 3^e Journées Informatique et Archéologie de Paris – JIAP 2012, Paris, Juin 2012. *Archeologia e Calcolatori*, Supplément 5, 2014, 18-33.
- Djindjian, F., Kozłowski, J., Otte, M. 1999. *Le Paléolithique supérieur en Europe*. Paris : Armand Colin.
- Doran, J., Palmer, M., Gilbert, N., Mellars, P. 1994. The EOS Project: Modeling Upper Palaeolithic Social Change, in 'N. Gilbert, J. Doran (eds.), *Artificial Societies*'. London: UCL Press, 195-221.
- Efremov, I.A. 1940. Taphonomy: a new branch of paleontology, *Pan-American Geology*, 74: 81-93.
- Harris, E.C., Brown, M.R. III, Brown, G.J. (eds.) 1993. *Practices of Archaeological Stratigraphy*. London: Academic Press.
- Koch, C.P. 1989. *Taphonomy: a bibliographic guide to the literature*. Orono: University of Maine.
- Kohler, T.A., Gumerman, G.J. (eds.) 2000. *Dynamics in Human and Primate Societies: Agent- Based Modelling of Social and Spatial Processes*. Oxford: Oxford University Press.
- Kohler, T.A., Varien, M.D. (eds.) 2012. *Emergence and Collapse of Early Villages: Models of Central Mesa Verde Archaeology*. Berkeley: University of California Press.
- Laplace, G. 1966. *Recherches sur l'origine et l'évolution des complexes leptolithiques*. Paris : de Boccard, Mélanges d'archéologie et d'histoire, École Française de Rome, Suppléments, 4.
- Wilkinson, T.J., Christiansen, J.H., Ur, J., Widell, M., Altaweel, M. 2007. Modeling Settlement Systems in a dynamic Environment. Case Studies from Mesopotamia, in 'T.A. Kohler, E.S. van der Leeuw (eds.), *The Model-Based Archaeology of Socio-natural Systems*'. Santa Fe: School for Advance Research Press, 175-208.

Transcending ‘Technocomplexes’ When French Empiricism calls for Hypothetico-deductive Method

Pascaline Gaussein
PhD, UMR 7055 PréTech
pascaline.gaussein@gmail.com

Abstract

Besides ‘techno-complexes’, often designated as archaeological ‘cultures’, what does a map of archaeological features distribution reveal about Palaeolithic human and social realities? In order to address this fundamental issue, archaeological methods needed to be re-evaluated to question the anthropological meaning of archaeological cultural features’ variability, processes of homogenization and change in space and time. It relies on the postulate, which appears to be an anthropological fact, that material culture can be actively invested in social practices as its styles (ways of doing) depend on interaction modalities, in relation to natural and social environment. Consequently, a framework and interpretative keys inspired by anthropological references are called for to allow the diversification and enlightenment of the interpretative potential of large scale and detailed empirical databases. This method is quite innovative as it requires a combination of processual, contextual and empirical approaches.

Keywords: Palaeolithic demography, Empiric-inductive method, Hypothetic-deductive method, Social dynamics, style analysis

Résumé

Au-delà des « techno-complexes », également appelés « cultures archéologiques », quelles réalités humaines et sociales sous-tendent nos cartes de répartition des traits culturels paléolithiques ? Pour répondre à cette problématique, nos outils méthodologiques ont dû être révisés pour interroger le sens de ces traits culturels, leur variabilité, les phénomènes d’homogénéisation et de changement dans l’espace et dans le temps. Cette méthode repose sur le postulat, qui semble être un fait anthropologique, que la culture matérielle peut être activement investie dans les pratiques sociales et que ses styles (manières de faire) dépendent des modalités d’interaction, en fonction de leur environnement naturel et social. C’est pourquoi un cadre de réflexion et des clés interprétatives extraits de références anthropologiques ont dû être mobilisés, pour permettre de diversifier et d’enrichir le potentiel interprétatif de bases de données empiriques conséquentes. Cette méthode peut être considérée comme innovante puisqu’elle requiert la combinaison d’approches processuelle, contextuelle et empirique.

Mots-clés : Paléogéographie, Méthode empirico-inductive, Méthode hypothético-déductive, techno-complexes, Dynamiques sociales, Analyse stylistique

1. A Fundamental Issue in its Context

This paper addresses the issue of the human and social meaning underlying material cultural features distribution. Besides ‘archaeological cultures’, carefully renamed ‘techno-complexes’, how do we go further typo-technologies? And consequently, what did Palaeolithic societies look like?

1.1. A (too) long-lasting taboo

As predicted by epistemologists, empiric-inductive approach begins to unveil its limits and produces intellectual frustration within French Palaeolithic researchers. Indeed, to fully understand the place of ways of doing (styles), and notably of art, in a foreign society which concepts and ideology

are unknown, we need an enhanced knowledge of contextualized human behaviours in its diversity. An anthropological referential of stylistic behaviours is needed to elucidate the contexts and conditions according to which cultural and social features were produced, reproduced, innovated, modified, borrowed or influenced. However, in western European Palaeolithic and in art study in particular, even though we try to apply concepts borrowed to anthropology, it is usually in a way simpler version, a position justified by the highly vestigial state of the data we work on. But because of this simplifying, we end up with almost empty concepts we finally fill in with the most logic meaning, our own.

This reflects the evolution of the study of Palaeolithic art in France. Indeed, it is roughly at the beginning of the 21st century that anthropological social issues reappeared in art study, after having been banned for almost half a century as advised by ethnologist André Leroi-Gourhan (1990 [1964]), as a reaction to an inhibited use of ethnological comparatism and actualism by the first prehistorians (Gallay 2015).

1.2. *Style (still) lacks meaning*

Nowadays, when anthropology is called for, it is usually applied with high caution, mostly to material structures interpretation or technic understanding, and to try understanding very global concepts. It is still too scarcely used to go further in art interpretation (Sauvet *et al.* 2009; Djindjian 2012; Fuentes *et al.* 2019). Considered as tangible symbolic expression ordered by the ideology of the cultures, 'art' is usually employed as a major cultural clue and even a cultural emblem *per se*: specific features, themes or even artefacts, when carefully selected, appear as identifying items (Conkey 1978; Bourdier 2010; Sauvet *et al.* 2014; Feruglio, Averbough 2016). But identifying what? Indeed, styles tend to lack meaning and we can't actually know what we are pointing out in terms of social entities when drawing maps of features distribution: which features can be considered as identity's indicators? Identifying Cultures? Human groups? Or are we still focused on delineating subdivisions of techno-complexes? There is a clear lack of anthropological background to apprehend human behaviours in their whole complexity. So we mostly hang on technological, stylistic and thematic analysis and their distribution.

Conversely, anthropologic knowledge of human behaviours has been called for without much caution too, usually to confirm an univocal controversial hypothesis through the analysis of a single category of artefacts, from which results black and white approaches: egalitarian versus aristocratic, nomadic vs sedentary societies etc., quite often applied to a whole period – as Upper Palaeolithic that lasts more than 10.000 years – and even to a whole continent – as Europe for instance –. We may recall, however, that such hypothesizes are usually suggested as a dare, consciously or not, in order to shake profoundly rooted popular beliefs and postulates (Delage 2012; Guy 2017).

Consequently, even though data keeps accumulating, interpretative emulation struggles: data won't talk by itself, interpretative hypothesizes are usually too global and not specific enough, and traditional paradigms would not allow giving sense to data analyses anymore, even though getting more and more detailed and high-definition recorded. That is why it appeared clear to us that if we were willing to rethink our ambitions for Palaeolithic material culture interpretation, the paradigms, models and methods applied had to be re-evaluated: the long-lasting French taboo of using hypothetic-deductive approach in Palaeolithic research may be slowly breaking apart.

2. A Re-evaluation of Concepts and Postulates

2.1. *Defining a Sociocultural Scale to the Study*

For starters, the concept of 'culture' is fundamental in human sciences and in archaeology in particular as it is the first level of ancient human population archaeologists try to delineate. The

regularities and significant indicators observed and selected from archaeological data are expected to help retrace '*stable association of shared, learnt behaviours*' in a given time and space (Perlès 2013, 293). But we have to recall that in the anthropological point of view, 'culture' appear more likely as a conceptual tool study than a reality for the studied population (Csonka 1992; Izard 2000; Amselle 2010; Gallay 2011). Resulting from a long and complex history of interaction, influences, innovations and borrowing of innovations, a cultural identity is a dynamic and situational phenomenon. What is more, there are several identity degrees and categories for each individual, the most redundant being at the scales of the family group, and of the endogamous group. Its change results from the manipulation of identity, according to social, economic and political interactions and needs (Jones 1997; Mac Eachern 1998; Gosselain 2000; Godelier 2007; Amselle 2010). Indeed, there are numerous reasons for maintaining more or less extended networks (biological viable reproduction, safety networks preventing shortages, political agreement to prevent tensions, etc.). Consequently, it appears that the material culture tends to better trace back interaction processes than identifying cultural identities (Jones 1997; Mac Eachern 1998), and finally, the 'cultures' we are searching for through archaeological material culture, do not exist *per se*.

That is why it seems more appropriate to work on more tangible scales of population organization. Elementary social structures, also approached by prehistorians who work on territorial and resources management, appeared the most relevant. On the one hand, we consider the minimum band, the basic productive and residential unit gathering an average number of 25 individuals; on the other hand, we also consider the Mating network, mainly endogamous network gathering at least 175 to 500 individuals for a viable endogamous population (also known as the 'maximum band' and 'marriage universe') (Wobst 1974; Gamble 1986; Burch 1998; Testart 2000; Marlowe 2005; Whallon 2006; Hill *et al.* 2011).

2.2. Material Culture Stylistic Processes

In our work we chose not to focus on what prehistoric material culture is not and what we can't do about it – it is not a docile mirror of the whole prehistoric culture and we can't treat Prehistory study as a fully ethnological science –, but on what it actually is: the product of dynamic and complex societies of a modern human species (*Homo sapiens sapiens*). Indeed, as previously suggested, the keystone to the 'techno-complex' issue relies on the anthropological fact that material culture can be actively invested in social practices. Its styles are considered carrying information, being impacted as well as playing a part, willingly or not, in the complex machinery that production and reproduction of society and its identities are (Wobst 1977; Conkey 1978; Wiessner 1983, 1990; Gamble 1986; de Boer 1990; Barton *et al.* 1994; Gosselain 2000; Vialou 2009; Schwendler 2012; Sauvet *et al.* 2014; Fuentes *et al.* 2019, amongst others). What is more, its agency (Gell 1998; Villeneuve and Hayden 2007; Descola 2009), its capability to trigger an action is at stake as it can play a part in these sociocultural processes, allowing identification, showing a social status, being part of alliance genesis and perpetuation as gift and exchange goods, etc.

Some tendencies encountered in anthropological literature help have a better understanding of the causality between social process and styles' variability, distributions, innovation and disappearance, and could be used as a referential to our research (Figures 1 and 2). First, we have to be aware of the passivity of part of the styles, which belong to a 'stylistic universe' reproduced by habit. It remains quite stable over long periods of time and in very wide areas even though populations don't maintain interactions anymore (Conkey 1978; Wiessner 1990; Gosselain 2000, 2011). The rhythm of change for passive styles is continuous, following the gradual modifications of imperfect transmission along generations (de Boer 1990; Lycett 2013). An overall change can occur in times of major crisis.

On the contrary, some stylistic aspects and categories of items are said to be 'active', because manipulated and affected by conscious communication of social information. Diversity, quantities,

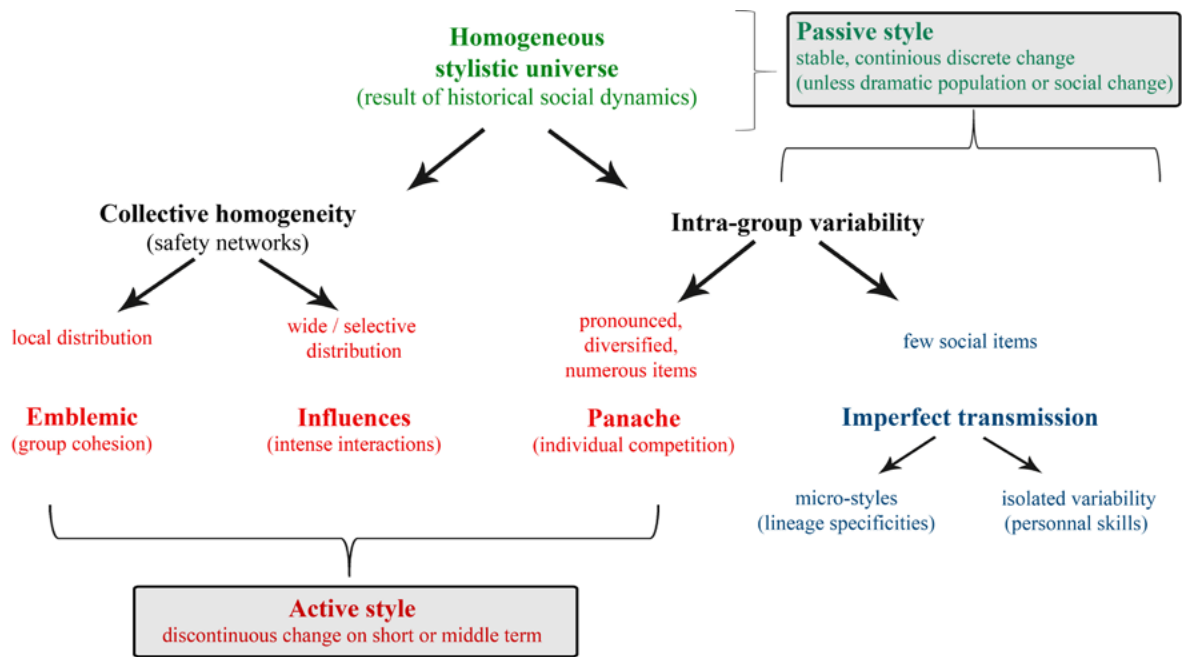


Figure 1. Simplified framework for the interpretation of variability, homogeneity and change of material culture styles. After Wobst 1977; Wiessner 1986, 1990; Boer (de) 1990; Macdonald 1990; Gosselain 2000, 2011. CAD P. Gaussein.

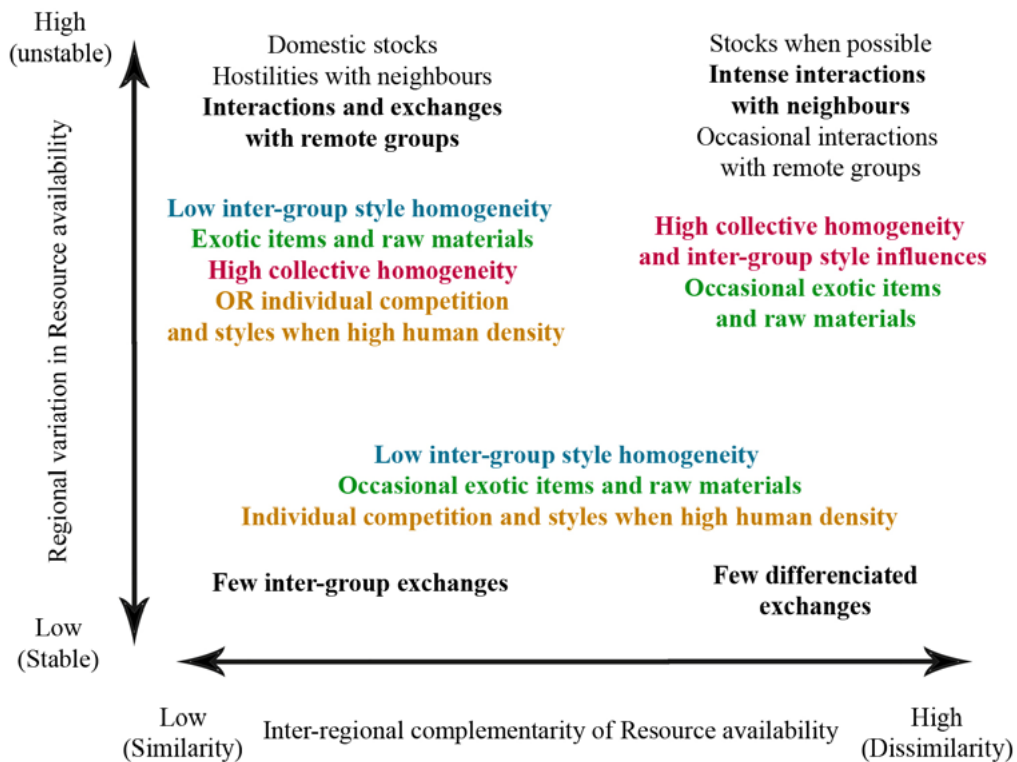


Figure 2. Contextualized framework for the interpretation of variability, homogeneity and change of material culture styles. After Wiessner 1990; Whallon 2006; Kelly 2013. CAD P. Gaussein.

styles and part played by those artefacts can change with a discontinuous rhythm, over short (crisis) or middle term (demographic or economic circumstances) (Perlès 2013; Roux, Courty 2013). In a context of environmental stress, collective unity for mutual aid prevails over individuality

(Figure 2). As a result, normative 'emblemic' styles are shared within social groups, and even between allied groups (Macdonald 1990; Wiessner 1990), and some specific status based on skills or occasional leadership stays materially discrete (Damas 1969; Darmangeat 2013; Kelly 2013). Intense interactions can take place between neighbouring or remote groups when their territories yield complementary resources (Whallon 2006; Kelly 2013). In that case, they tend to produce similar styles (Jones 1997). Conversely, interactions are more occasional when mutual help is logistically impossible. Safety networks are also rare in the context of more stable resources (Whallon 2006; Kelly 2013). Local identities tend to strengthen, stocks can be made and human density tends to increase. It brings social competition and with it the need to be individually distinct from each other, implying aesthetic innovations, and individual stylistic variability (Wiessner 1986, 1990; Macdonald 1990; Schwendler 2012).

3. A Methodological Essay

3.1. *At the Crossroads of Three Major Approaches*

Traditional empirical scientific approach directly questions data within quite a vague anthropological background: it usually analyses the most relevant cultural features distribution in space and time, in order to deduce cultural areas and interactions. The approach suggested in this paper to address the 'techno-complex issue' is the other way round to a certain extent: it works back and forth between empiric-inductive and hypothetic-deductive processes, and at the crossroads of processual and contextual archaeology.

On the one hand, a sharp empirical analysis constitutes the very foundation of our approach, originally funded in Palaeolithic art analysis (Gaussein 2017, 2019). The whole data must be accounted for in order to circumvent the selective bias of electing 'relevant' features. Indeed, by considering supposedly non-specific items and stylistic details, a quantitative and descriptive analysis of numerous artefacts in a contextualized perspective can help extract the main social dynamics as well as discrete organizations and processes within the social system (imperfect transmission, individuality, exceptional events, etc.) (Wiessner 1990).

On the other hand, given the complexity and situational nature of the uses of style underlined by anthropologists (Wiessner 1983, 1990; de Boer 1990; Macdonald 1990; Mac Eachern 1998; Gosselain 2011) and our scarce knowledge of Palaeolithic territories, we must find a way to submit stable social model(s) to the material culture studied. The tendencies previously described in this paper – that consider identification and interaction processes' impact on material culture (styles) depending on natural and demographic contexts – become a keystone to our understanding of prehistoric societies.

As such, in order to reach prehistoric populations and to investigate their organization, their cultural and social identities and dynamics through artefacts, we propose to submit archaeological data to various plausible demographic and sociocultural contexts. For each tested context, if data does not behave accordingly to the tendencies evidenced by anthropologists (our keystone referential), that is to say sociocultural processes are not intended to produce and impact material culture's features as data shows in such context, this very context and all its sociocultural implications are likely to be dismissed. Complexity of human behaviours and the vestigial state of archaeological data rises the necessity of interpretative caution, that is why we could be expecting several hypotheses to be considered as plausible. Thus, contrary to the 'confirmation epistemology' described by Stoczkowski (1992), this research leads an epistemology of contradictory confirmation. It allows opening the door to diversified hypothesis to work on, but also to multiple more detailed interpretative theories, factually approved until new data tell otherwise. The more detailed models will be, the more easily it will get dismissed ('falsified', *sensu* Karl Popper: Chalmers 1987).

In fine, a more detailed approach of prehistoric populations could be reached, as every thematic, stylistic, technical and material variation in space and time could actually carry social information, which we must investigate: does it underlie different local groups' identities? Mating networks? Or individual specificities? Lineage apprenticeship? High definition data should help retrace the complex history of transmissions, interactions, influences and identification of prehistoric populations.

3.2. How to Apply This Method

Palaeolithic social and territorial contexts are hard to reach. However, as human density is considered having a major influence over social (territories and network scales) and economic parameters (residential core, foraging and logistical zone) it constitutes a major keystone for this framework. As suggested in other studies (Wobst 1977; Whallon 2006; Banks *et al.* 2011), density associated with anthropologic magic numbers makes possible, thanks to a simple Rule of Three, to draw maps of human territories at the elementary social units scale: the band and the mating network. Several human densities have been hypothesized for the same periods of Upper Palaeolithic times (Delpech 1999; Bocquet-Appel *et al.* 2005; Whallon 2006; French 2015): each will be submitted to archaeological data in order to evaluate the plausibility of different social and

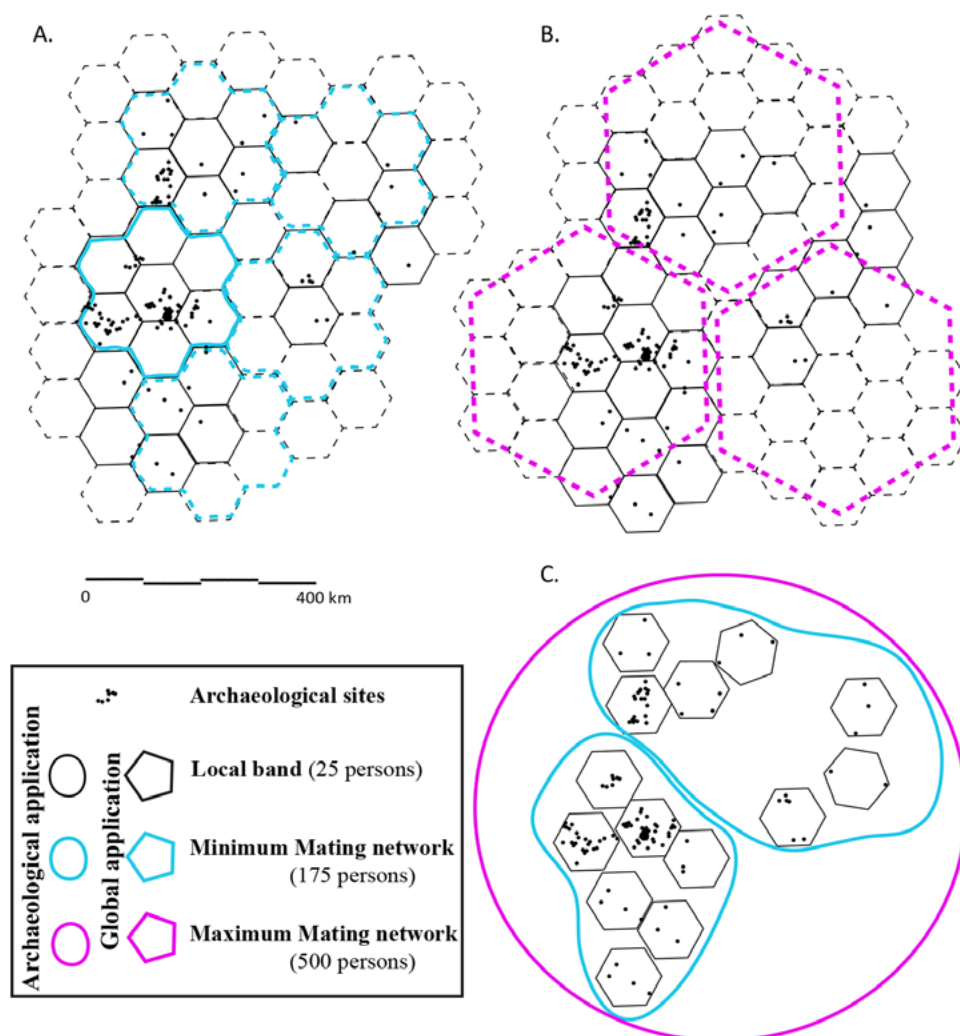


Figure 3. Schematic mapping of social unit territories depending on the application of human density: a. and b. 'global application', archaeological map none taken into account; c. 'archaeological application' according to known archaeological settlements. CAP P. Gaussein.

demographic models. Density must be applied following the known archaeological settlements (theoretical territories only drawn according to the site dispersion; Figure 3, elliptic shapes). But should then be extrapolated to the whole area studied, with archaeological gaps not taken into account so archaeological bias can be partly accounted for (Figure 3).

As far as archaeological data is concerned, the analysis of the artefacts relies on basic statistics applied to a data base of qualitative and quantitative criteria describing the artefact in itself, the technic used to craft it, decorate it, the depicted themes and the stylistic features of each image. These criteria are examined by categories (technic, themes, shape of each detail etc.) through mapped diagrams (*QuantumGis*©), allowing the illustration of the proportion of stylistic and technic choices for each settlement. The observations made on each category of criteria is crossed and associated with a more global discussion on stylistic practices for each settlement and for each region. These observations confronted to our stylistic behaviour referential allows determining what social and demographic organizations can be considered plausible or should be dismissed.

When archaeological data is submitted to the demographic models: for each human density version, the distribution of material features is instantly provided with a sociological meaning. Finally, one or several models can be confirmed and will be available to submit to new data (re-evaluation of old collections, new findings...).

4. Discussion on the Methods' Limits

Certain difficulties intrinsic to anthropological use and to archaeological data necessarily alter the analysis conducted through this method. For starters, behavioural tendencies underlined in anthropology and essential to prehistoric archaeology cannot take into account the whole diversity and complexity of human behaviours. For instance, our analysis usually avoid calling for incoherent economic and social behaviours, even though it is encountered in ethnographic studies. We should keep completing and tempering our stylistic behaviour referential with the diversification of ethnographic references.

Concerning archaeological bias, the map of known archaeological settlements quite imperfectly mirrors the actual prehistoric human settlements, and consequently affects the original distribution of artistic and technic features. Consequently, our social and demographic interpretation of Magdalenian data stays at a hypothetical level. What is more, the precise chrono-stratigraphic attribution and dating of numerous artefacts and settlements could not be obtained, mainly because of their ancient excavation (19th to mid-20th centuries for most of them). This bias leads to the major difficulty to reach a thin chronological resolution of environmental, social, political events and stylistic rhythm of change. This compaction of (pre)historic time alters our interpretation of stylistic and sociocultural behaviours as we could not decompose the different stages of a cultural feature's life: innovation date and location; spreading within a group's transmission lineage; period of interaction leading from progressive gifts, influences or borrowing to the feature's homogenization; then its loss of conscious sociocultural sense (passive features reproduced by habit, forgetting the reason it has entered groups' customs); and finally the feature's disappearance while identities, traditions are changing (Gaussein 2017, 2019).

What is more, we will need to deal with the pluri-disciplinary nature of this approach in depth and at a more local as well as more inter-regional scale. It will help reach a better understanding of the spatial and temporal organization and annual viability of our models for each local bands' territory. How local and regional raw material procurement has been dealt with? How individual and group movement were practicable in the local and regional topography over the ice age seasons (frozen rivers, swampy periods of snow and ice melting...)? Clarifying, if possible, the benefits and adaptability of the different bone and lithic industries, and the complexity or easiness of technical

and stylistic influences and borrowing (Brun *et al.* 2005; Perlès 2007) would also help define the social dynamics responsible for the homogenization of some of artistic and technical features in each tested context.

5. Conclusion

This methodological essay relies on a combination of processual, contextual and empirical approaches, quite innovative for south-western Europe Palaeolithic researchers: the vestigial and biased nature of archaeological data paralysing our study is left behind in order to focus on human realities underlaid by the excavated material culture. Moreover, it is applicable to the re-evaluation of numerous ancient excavations series of various categories of artefacts, and to other 'archaeological cultures'.

This reorientation of empirical analyses by calling for diversified anthropological models leads, on the one hand, to more precise and more human-like interpretative hypotheses of Palaeolithic social organization, and, on the other hand, to the diversification of the hypotheses we work on and falsify (*sensu* K. Popper), preventing the usual epistemology of confirming univocal dichotomous hypothesis (Stoczkowski 1992).

References

- Amselle, J.-L. 2010. *Logiques métisses*. Paris: ed. Payot.
- Banks, W.E., Aubry, T., d'Errico, F., Zilhão, J., Lira-Noriega, A. and Townsend Peterson, A. 2011. Eco-cultural niches of the Badegoulian: Unraveling links between cultural adaptation and ecology during the Last Glacial Maximum in France. *Journal of Anthropological Archaeology* 30: 359-374.
- Barton, C.M., Clark, G.A. and Cohen, A.E. 1994. Art as information: Explaining Upper Palaeolithic art in western Europe. *World Archaeology* 26: 185-207.
- Bocquet-Appel, J.-P., Demars, P.-Y., Noiret, L. and Dobrowsky, D. 2005. Estimates of Upper Palaeolithic meta-population size in Europe from archaeological data. *Journal of Archaeological Science* 32: 156-168.
- Boer (de), W.R. 1990. Interaction, Imitation, and communication as expressed in style: the Ucayali experience, in Conkey, Hastorf 1990: 82-104.
- Bourdier, C. 2010. *Paléogéographie symbolique au Magdalénien moyen. Apport de l'étude des productions graphiques pariétales des abris occupés et sculptés de l'Ouest français (Roc-aux-Sorciers, Chaire_à_Calvin, Reverdit, Cap-Blanc)*. PhD dissertation, Université Bordeaux I, Bordeaux.
- Brun, P., Averbouh, A., Karlin, C., Méry, S. and Miroschedji, P. (de) 2005. Les liens entre la complexité des sociétés traditionnelles et le niveau de spécialisation artisanale: bilan et perspectives, in A. Averbouh, P. Brun, C. Karlin, S. Méry and P. Miroschedji (de) (eds), *Spécialisation des Tâches et Sociétés. Actes de La Table Ronde Annuelle de l'UMR ArScan (MAE, Nanterre)*: 325-347. Paris: Ed. CNRS; Nanterre: Éditions de la Maison des sciences de l'Homme (Techniques et Cultures 46-47).
- Burch, E.S. 1998. Boundaries and Borders in Early Contact North-Central Alaska. *Arctic Anthropology* 35: 19-48.
- Chalmers, A.F. 1987. *Qu'est-ce que la science? Récents développements en philosophie des sciences: Popper, Kuhn, Lakatos, Feyerabend*. Paris: La Découverte.
- Conkey, M.W. 1978. *An Analysis of Design Structure: Variability Among Magdalenian Engraved Bones from North Coastal Spain*. Ph.D. dissertation. University of Chicago, Chicago.
- Conkey, M.W. and Hastorf, C.A. (eds) 1990. *The Uses of Style in Archaeology*. Cambridge, New York, Portchester, Melbourne, Sydney: Cambridge University Press.
- Csonka, Y. 1992. Identification et définition d'entités socio-démographiques: concepts des Inuits et classements des ethnologues, in *Ethnoarchéologie. Justification, Problèmes, Limites. Actes des XIIe Rencontres Internationales d'Archéologie et d'Histoire d'Antibes (Juan-Les-Pins, 17-19 Octobre 1991)*: 171-180. Antibes: APDCA.
- Damas, D. (ed.) 1969. *Contributions to Anthropology Band Societies, Conference on Band Organization*: 172-179. Anthropological Series 84. Ottawa: Bull. of the National Museums of Canada.

- Darmangeat, C. 2013. *Conversation sur la naissance des inégalités*. Marseille: Agone (Passé & présent).
- Delage, C. 2012. De la 'pointe de sagaie' à la 'culture de Lussac-Angles', il y a plus qu'un pas. *Argumentaire. Bulletin de la S.E.R.P.E.* 62: 23-48.
- Delpech, F. 1999. Biomasse d'Ongulés au Paléolithique et inférences sur la démographie. *PALEO* 11: 19-42.
- Descola, P. 2009. L'Envers du visible: ontologie et iconologie, in *Histoire de l'art et anthropologie, Les actes*. Paris: INHA/Musée du quai Branly.
- Djindjian, F. 2012. Contacts et déplacements des groupes humains dans le Paléolithique supérieur européen: les adaptations aux variations climatiques des stratégies de gestion des ressources dans le territoire et dans le cycle annuel, in M. Otte and F. Lebrun-Ricalens (eds), *Modes de Contacts et de Déplacements Au Paléolithique Eurasiatique. Actes Du Colloque International de La Commission 8 (28-31 Mai 2012)*: 645-678. Liège: ERAUL 140; Luxembourg: ArchéoLogiques (5).
- Feruglio, V. and Averbouh, A. 2016. Identification d'un marqueur identitaire des groupes du Magdalénien par l'approche technique, thématique et chronologique: l'exemple des baguettes demi-rondes à tubérosité, in J.-J. Cleyet-Merle, J.-M. Geneste and E. Man-Estier (eds), *L'art Au Quotidien - Objets Ornés du Paléolithique Supérieur. Actes du Colloque International (Les Eyzies-de-Tayac, 16-20 Juin 2014)*. *PALEO*, numéro spécial: 199-215.
- French, J.C. 2015. The demography of the Upper Palaeolithic hunter-gatherers of Southwestern France: A multi-proxy approach using archaeological data. *Journal of Anthropological Archaeology* 39: 193-209.
- Fuentes, O., Lucas, C. and Robert, E. 2019. An approach to Palaeolithic networks: The question of symbolic territories and their interpretation through Magdalenian art. *Quaternary International* 503: 233-247.
- Gallay, A. 2011. *Pour une ethnoarchéologie théorique. Mérites et limites de l'analogie ethnographique*. Paris: Errance (Hespérides).
- Gallay, A. 2015. André Leroi-Gourhan et l'ethnologie : gérer un héritage, in P. Soulier (ed.), *André-Leroi-Gourhan : 'l'Homme Tout Simplement'*: 47-57. Travaux de La Maison de l'Archéologie et de l'Ethnologie, René-Ginouvès. Paris: de Boccard.
- Gamble, C. 1986. *The Palaeolithic Settlement of Europe*. Cambridge: Cambridge University Press.
- Gaussein, P. 2017. *Pour une anthropologie de l'art mobilier: Identités et réseaux entre Loire et Dordogne*. PhD dissertation. Université Paris Nanterre, Nanterre.
- Gaussein, P. 2019. À la recherche des identités et dynamiques sociales paléolithiques. Investigation anthropologique sur l'art mobilier du Magdalénien moyen ancien dans le centre-ouest de la France, in E. Caron-Laviolette, N. Matomou-Adzo, C. Millot-Richard and B. Ramé (eds), *Biais, hiatus et absences en archéologie*. Paris: Éditions de la Sorbonne (Archéo.doct 12). <<https://books.openedition.org/psorbonne/19342>>.
- Gell, A. 1998. *Art and Agency: Towards a New Anthropological Theory*. Oxford: Clarendon Press.
- Godelier, M. 2007. *Au fondement des sociétés humaines: ce que nous apprend l'anthropologie*. Paris: Albin Michel (coll. Idées).
- Gosselain, O.P. 2000. Materializing Identities: an African Perspective. *Journal of Anthropological Archaeology* 7: 187-217.
- Gosselain, O.P. 2011. Pourquoi le décorer? Quelques observations sur le décor céramique en Afrique. *Archaeological Research in Africa* 46: 3-19.
- Guy, E. 2017. *Ce que l'art préhistorique dit de nos origines*. Paris: Flammarion.
- Hill, K.R., Walker, R.S., Božičević, M., Eder, J., Headland, T., Hewlett, B., Hurtado, A.M., Marlowe, F., Wiessner, P. and Wood, B. 2011. Co-Residence Patterns in Hunter-Gatherer Societies Show Unique Human Social Structure. *Science* 331: 1286-1289.
- Izard, M. 2000. Culture, 1. Le problème, in P. Bonte and M. Izard (eds) *Dictionnaire de L'ethnologie et de L'anthropologie*: 190-192. Paris: Quadrige/PUF.
- Jones, S. 1997. *The archaeology of ethnicity: constructing identities in the past and present*. London, New York: Routledge.
- Kelly, R.L. 2013. *The Life ways of Hunter-Gatherers: the Foraging Spectrum*. Cambridge: Cambridge University Press.
- Leroi-Gourhan, A. 1990. *Les religions de la préhistoire: Paléolithique*. Paris: PUF (3rd édition).

- Lycett, S.J. 2013. Cultural Transmission Theory and Fossil Hominin Behaviour: A Discussion of Epistemological and Methodological Strengths, in R. Ellen, S.J. Lycett and S.E. Johns (eds) *Understanding Cultural Transmission in Anthropology: A Critical Synthesis, Methodology & History in Anthropology*: 102-130. New York/Oxford: Berghahn Books.
- Mac Eachern, S. 1998. Scale, Style and Cultural Variation: Technological Traditions in the Northern Mandara Mountains, in M.T. Stark (ed.) *The Archaeology of Social Boundaries*: 107-131. Washington, London: Smithsonian Institution Press.
- Macdonald, W.K. 1990. Investigating style: an exploratory analysis of some Plains burials, in Conkey, Hastorf 1990: 52-60.
- Marlowe, F.W. 2005. Hunter-gatherers and human evolution. *Evolutionary Anthropology: Issues, News, and Reviews* 14: 54-67.
- Perlès, C. 2007. Synthèse. Diffusions, emprunts, refus d'emprunts: les acteurs humains, in P. Rouillard et al. (ed.) *Mobilités, Immobilismes. L'emprunt et son refus. Actes du 3e Colloque de la Maison René-Ginouvès*: 319-326. Paris: De Boccard.
- Perlès, C. 2013. Tempi of Change: When Soloists don't play Together. Arrhythmia in 'Continuous' Change. *Journal of Archaeological Method and Theory* 20: 281-299.
- Roux, V. and Courty, M.-A. 2013. Introduction to Discontinuities and Continuities: Theories, Methods and Proxies for a Historical and Sociological Approach to Evolution of Past Societies. *Journal of Archaeological Method & Theory* 20: 187-193.
- Sauvet, G., Fritz, C., Fortea Pérez, J. and Tosello, G. 2014. Fluctuations des échanges symboliques au Paléolithique supérieur en France et dans le nord de l'Espagne, in J. Jaubert, N. Fourment and P. Depaepe (eds) *Transitions, Ruptures et Continuité En Préhistoire. Actes Du XXVIIe Congrès Préhistorique de France (Bordeaux-Les Eyzies, 31 Mai-5 Juin 2010)*: 403-415. Paris: Société préhistorique française.
- Sauvet, G., Layton, R., Lenssen-Erz, T., Taçon, P. and Włodarczyk, A. 2009. Thinking with animals in Upper Palaeolithic rock art. *Cambridge Archaeological Journal* 19: 319-336.
- Schwendler, R.H. 2012. Diversity in social organization across Magdalenian Western Europe ca. 17-12,000 BP. *Quaternary International* 272-273: 333-353.
- Stoczkowski, W. 1992. Préhistoire, ethnologie et approche prédictive: la tentation d'une épistémologie spontanée, in F. Audouze (ed.) *Ethnoarchéologie. Justification, Problèmes, Limites. Actes des XIIe Rencontres Internationales d'Archéologie et d'Histoire d'Antibes-Juan-Les-Pins, 17-19 Octobre 1991*: 33-44. Antibes: APDCA.
- Testart, A. 2000. Chasseurs-cueilleurs, in Bonte, P. and M. Izard (eds) *Dictionnaire de l'ethnologie et de l'anthropologie*: 135-138. Paris: Quadrige/PUF.
- Vialou, D. 2009. L'image du sens, en préhistoire. *L'Anthropologie* 113: 464-477.
- Villeneuve, S. and Hayden, B. 2007. Nouvelle approche de l'analyse du contexte des figurations pariétales, in S. De Beaune (ed.) *Chasseurs-cueilleurs. Comment vivaient nos ancêtres du Paléolithique Supérieur. Méthodes d'analyse et interprétation en Préhistoire*: 151-159. Paris: CNRS.
- Whallon, R. 2006. Social networks and information: Non-'utilitarian' mobility among hunter-gatherers. *Journal of Anthropological Archaeology* 25: 259-270.
- Wiessner, P. 1983. Style and Social Information in Kalahari San Projectile Points. *American Antiquity* 48: 253-276.
- Wiessner, P. 1990. Is there a unity to style? in Conkey, Hastorf 1990: 105-112.
- Wobst, H.M. 1974. Boundary Conditions for Paleolithic Social Systems: A Simulation Approach. *American Antiquity* 39: 147-178.
- Wobst, H.M. 1977. Stylistic Behavior and Information Exchange, in C.E. Cleland (ed.) *For the Director: Research Essays in Honor of James B. Griffin*: 317-342. Ann Arbor, Michigan: Museum of Anthropology, University of Michigan (Anthropological Papers).

List of Authors

Auguste, Patrick

UMR 8198, CNRS – Unit EVO ECO PALEO, University of Lille
patrick.auguste@univ-lille.fr

Boulbes, Nicolas

HNHP- UMR 7194, Université de Perpignan, CNRS, MNHN, Centre Européen de Recherches Préhistoriques, Tautavel

Boussard, Matthieu

Craft Ai. WeWork, Paris

Buchsenschutz, Olivier

AOROC, UMR 8546 CNRS, Paris
olivier.buchsenschutz@ens.fr

Carrez, Philippe

Immersion-Tools, Saint-Brieuc

Celiberti, Vincenzo

HNHP- UMR 7194, Université de Perpignan, CNRS, MNHN, Centre Européen de Recherches Préhistoriques, Tautavel

Chevalier, Tony

HNHP- UMR 7194, Université de Perpignan, CNRS, MNHN, Centre Européen de Recherches Préhistoriques, Tautavel

Chopinaud, Caroline

Craft Ai. WeWork, Paris

de Lumley, Henry

Institut de Paléontologie Humaine, Paris

de Lumley, Marie-Antoinette

Institut de Paléontologie Humaine, Paris

Djindjian, François

Université de Paris 1 Panthéon Sorbonne et CNRS UMR 7041
francois.djindjian@wanadoo.fr

Fontaneil, Cédric

HNHP- UMR 7194, Université de Perpignan, CNRS, MNHN, Centre Européen de Recherches Préhistoriques, Tautavel

Gaussein, Pascaline

UMR 7055 PréTech
pascaline.gaussein@gmail.com

Grégoire, Sophie

HNHP- UMR 7194, Université de Perpignan, CNRS, MNHN, Centre Européen de Recherches Préhistoriques, Tautavel.
gregoire@cerptautavel.com

Harfouche, Antoine

CEROS – EA 4429 -Université Paris Nanterre, UFR Segmi

Hervé, Thierry

16 rue Schaeffer, 93300 Aubervilliers

Lamotte, Agnès

HALMA – UMR 8164, CNRS, University of Lille
agnes.lamotte@univ-lille.fr

Lartigot Campin, Anne-Sophie

HNHP- UMR 7194, Université de Perpignan, CNRS, MNHN, Centre Européen de Recherches Préhistoriques, Tautavel

Manseri, Hakima

Archéologie des Sociétés Méditerranéennes (ASM) – Université Paul Valéry – Montpellier III, CNRS: UMR5140, Ministère de la Culture
hakima.manseri@cnrs.fr

Marciniack, Rolande

CEROS – EA 4429 -Université Paris Nanterre, UFR Segmi

Masson, Éric

SHS – Laboratory Territoires, Villes, Environnement et Société (TVES) EA 4477, University of Lille
eric.masson@univ-lille.fr

Moigne, Anne-Marie

HNHP- UMR 7194, Muséum national d'Histoire Naturelle, CNRS, UPVD, Centre Européen de Recherches Préhistoriques, Tautavel
anne-marie.moigne@cerptautavel.com

Moscatti, Paola

Istituto di Scienze del Patrimonio Culturale
CNR – Area della Ricerca di Montelibretti (RM)
paola.moscatti@cnr.it

Munos, Sébastien

Labex Archimède « Archéologie et Histoire de
la Méditerranée et de l’Egypte anciennes »
(PIA-ANR-11-LABX-0032-01)
sebastien.munos@univ-montp3.fr

Perrenoud, Christian

HNHP- UMR 7194, Muséum national d’Histoire
Naturelle, CNRS, UPVD, Centre Européen de
Recherches Préhistoriques, Tautavel

Peudon, Floriane

HALMA – UMR 8164, CNRS, University of Lille
floriane.peudon@gmail.com

Pois, Véronique

HNHP- UMR 7194, Université de Perpignan,
CNRS, MNHN, Centre Européen de Recherches
Préhistoriques, Tautavel

Py, Michel

Directeur de recherche honoraire au CNRS,
Archéologie des Sociétés Méditerranéennes
(ASM)
michel-py@sfr.fr

Quinio, Bernard

CEROS – EA 4429 -Université Paris Nanterre,
UFR Segmi

Roure, Réjane

Archéologie des Sociétés Méditerranéennes
(ASM) – Université Paul Valéry – Montpellier
III, CNRS: UMR5140, Ministère de la Culture
Labex Archimède « Archéologie et Histoire de
la Méditerranée et de l’Egypte anciennes »
(PIA-ANR-11-LABX-0032-01)
rejane.roure@univ-montp3.fr

Testu, Agnès

HNHP- UMR 7194, Université de Perpignan,
CNRS, MNHN, Centre Européen de Recherches
Préhistoriques, Tautavel

Saos, Thibaud

HNHP- UMR 7194, Université de Perpignan,
CNRS, MNHN, Centre Européen de Recherches
Préhistoriques, Tautavel

Tuffreau, Alain

HALMA – UMR 8164, CNRS, University of Lille
alain.tuffreau@univ-lille1.fr

Big Data and Archaeology presents the papers from two sessions of the 18th UISPP World Congress (Paris, June 2018): Session III-1 (CA): 'Big data, databases and archaeology', and Session III-1 (T): 'New advances in theoretical archaeology'. The advent of Big Data is a recent and debated issue in Digital Archaeology. Historiographic context and current developments are illustrated in this volume, as well as comprehensive examples of a multidisciplinary and integrative approach to the recording, management and exploitation of excavation data and documents produced over a long period of archaeological research. In addition, specific attention is paid to neoprocessual archaeology, as a new platform aimed at renewing the theoretical framework of archaeology after thirty years of post-modernism, and to the refinement of the concept of archaeological cultures, combining processual, contextual and empirical approaches.

François Djindjian is *ancien professeur* at the University of Paris I Pantheon Sorbonne (chair of archaeological methods and theory) and associate member of the CNRS UMR 7041. He is President of the International Union for Prehistoric and Protohistoric Sciences (UISPP), member of the executive committee of the International Council for Philosophy and Human Sciences (CIPSH) of Unesco, and delegate member of the International Academic Union (UAI).

Paola Moscati is research director at the Institute of Heritage Science of the National Research Council of Italy. As an archaeologist, specialised in computer applications in archaeology, she is Vice President of the UISPP Commission IV, editor in chief of the international journal 'Archeologia e Calcolatori' and scientific coordinator of the international project 'The Virtual Museum of Archaeological Computing', jointly promoted with the Accademia Nazionale dei Lincei.

