



HAL
open science

BDA : une Base de Données Archéologique collaborative en ligne

Thomas Perrin

► **To cite this version:**

Thomas Perrin. BDA : une Base de Données Archéologique collaborative en ligne. Bulletin de la Société préhistorique française, 2019, 116 (1), pp.159-162. 10.3406/bspf.2019.14988 . hal-02092025

HAL Id: hal-02092025

<https://hal.science/hal-02092025>

Submitted on 7 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BDA : une Base de Données Archéologique collaborative en ligne

Thomas PERRIN

Il ne faut guère de recul sur notre discipline et sur nos pratiques pour dresser le constat que nombre d'entre nous passent beaucoup de temps à élaborer des listes, des tableaux et parfois des bases de données recensant les sites de telle ou telle période dans telle ou telle région du monde. Ce travail, aussi fastidieux que nécessaire, permet à chaque chercheur de disposer d'un arsenal de données facilitant la mise en contexte de ses propres découvertes. Cette étape est devenue au fil des ans et de la multiplication des découvertes un rituel quasi obligatoire pour tout étudiant de master, la redondance de l'exercice le faisant aujourd'hui confiner à une sorte de passage initiatique ! De même, quel projet ANR ou autre ne comporte-t-il pas son traditionnel outil ou axe de constitution d'une base de données visant à établir l'état de l'art ? Toutes les bases de données ainsi constituées – construites plus ou moins dans les règles – restent ensuite pour la plupart propriété des animateurs des projets en question qui les utilisent pour eux-mêmes ou pour d'autres projets dans lesquels ils sont impliqués, mais sans que cela bénéficie directement à la communauté dans son ensemble. Et le manège est ainsi reparti pour un tour !... Le gâchis d'énergie, de temps et d'argent engendré par cette situation est d'autant plus regrettable que toutes ces bases partagent une grande part de données communes... À l'heure de la mise en avant de la science ouverte, de la publication *open access*, du partage de données et des Humanités Numériques, force est de constater que la masse de travail déployée pour construire et élaborer ainsi des outils restreints et redondants serait beaucoup plus pertinente à être concentrée sur les aspects réellement novateurs de chacun de ces programmes : les sites mobilisés seront toujours situés au même endroit, auront toujours le même historique ou la même stratigraphie, etc., et il devient inepte que la saisie de ce type d'informations mobilise encore des dizaines de chercheurs et d'étudiants chaque année.

Pendant ma formation universitaire, j'avais commencé, à partir de 1994, à construire une base de données de sites archéologiques dans le logiciel FileMaker Pro (alors en version 2). Elle avait pour vocation de réunir des informations essentielles sur les sites préhistoriques qui m'intéressaient (situation géographique, coordonnées, historique et type de fouilles, etc.), sur les occupations qui les composaient ainsi que sur les datations radiocarbone qui leur étaient liées et les industries lithiques qui en provenaient. J'ai poursuivi l'utilisation et l'amélioration de cette base de données, appelée BasePréRéc (Base de données des sites de la Préhistoire Récente), à partir de 1997 jusqu'à aujourd'hui. Différents modules ont été progressivement ajoutés, notamment une base liée à la lithologie des matières premières siliceuses (laquelle ne compte toutefois encore que peu d'enregistrements).

En 2008, l'obtention d'une allocation de recherche postdoctorale de la fondation Fyssen (Perrin *et al.*, 2009) m'avait amené à fortement réorganiser cette base en fusionnant notamment les multiples fichiers (correspondant à autant de tables) en un seul, ce qui n'était techniquement pas possible auparavant. Ce fut notamment à cette occasion que cette base de données put être utilisée de façon collective par les différents membres de ce programme. BasePréRéc fut pour l'occasion renommée BDA (Base de Données Archéologique) puisqu'elle fut modifiée pour pouvoir être utilisée largement au-delà de la seule Préhistoire récente. Cependant, les outils de travail collaboratif et de partage en ligne offerts par le logiciel à l'époque n'étaient guère conviviaux et cet effort d'ouverture s'est rapidement heurté à des difficultés techniques. Mais la volonté de partager cette base beaucoup plus largement restait présente et les efforts ont alors été orientés dans ce but. En 2014, la base a été mobilisée dans le cadre de l'ANR Procome¹ (dir. C. Manen, CNRS ; Perrin *et al.*, 2018). Une tentative de portage au format MySQL/PHP ne s'est, à l'époque, guère avérée convaincante. Une solution temporaire fut de mettre la base en question dans un dossier partagé dans le *Cloud*, mais sans que cela soit non plus satisfaisant, du fait de l'impossibilité d'accès multiples. Faute de mieux, c'est toutefois la solution qui fut également retenue en 2016 lors de la mobilisation de cette même base dans le cadre du programme IDEX MeNeMOIA², durant lequel de nombreux sites nord-africains furent ajoutés. À partir de la mi-2017, une solution plus viable fut trouvée grâce à la TGIR Huma-Num qui propose l'hébergement de bases de données collaboratives développées sous FileMaker Pro (grâce au recours à la version Server du logiciel). La base BDA est depuis hébergée sur ce site et accessible aisément à tout un chacun (<https://bda.huma-num.fr/>). La solution n'est évidemment pas encore parfaite, car elle reste tributaire d'un logiciel propriétaire au langage spécifique, et en conséquence elle ne peut que difficilement s'interfacer avec d'autres logiciels tels les SIG ou les logiciels de statistiques. Dans le cadre de l'ANR MONUMEN en cours³ (dir. Vincent Ard, CNRS), une transposition de cette base au format PostgreSQL/PHP va être tentée, et cette nouvelle version sera éventuellement amenée à remplacer l'actuelle si elle s'avère pertinente et efficace. En attendant, la version FileMaker Pro hébergée chez Huma-Num est fonctionnelle et permet d'envisager la mise en place d'un travail collaboratif à l'échelle au moins nationale, si la communauté juge cet outil pertinent.

À l'heure actuelle, BDA est une base de données relationnelle composée d'une trentaine de tables constituant un système cohérent visant à recenser les informations essentielles sur les sites archéologiques, leurs occupations

The screenshot shows a web browser window with the URL '1' and '2 / 4673 Trouvés (Non triés)'. The page title is 'Sites'. The main content area is titled 'Basi' and contains several sections:

- Appellation:** Basi
- Pays:** France
- Région:** Corse
- Département:** Corse-du-Sud (2A)
- Commune:** Serra-di-Ferro
- Type:** plein air
- Opé.:** fouille
- Fiabilité:** bonne
- Index:** à revoir, doublon, à supprimer
- Précision géographique:** site (carte)
- Coordonnées UTM décimales (datum : WGS84):** Latitude: 41.7473128 N, Longitude: 8.8467404 E, Altitude: 80
- Projections carto. nationales:** Système de coordonnées, X, Y
- Situation géographique:** Site de plein air dans la vallée du Taravo sur une petite terrasse sur une butte à 6 km du rivage actuel
- Historique et extension des fouilles:** Site découvert en 1965 par R. Grosjean et E. Atzeni à l'occasion d'une ouverture de carrière de granite. Fouilles effectuées en conditions de sauvetage par G. Bailloud entre 1968 et 1971 (à 3 personnes) lors de l'extension d'une carrière. Les fouilles se sont limitées à environ 13 m². L'essentiel du site a été détruit par des engins mécaniques. Puis diagnostic INRAP Anne Hasler avril 2011.
- Biblio.:** Tykot 1995 p.504 ; Costa 2004 p. 153 ; Hasler et al. 2014
- Lieu de dépôt:** Musée de Sartène
- Stratigraphie générale:** 7 niveaux reconnus par Bailloud du NA au Bze. Caridal classique / récent. Présence potentielle d'Impressa, à vérifier (qq tessons seraient présents). c.1-3 : Torréen remanié. présence de Terrinien dans les vitrines de Filitosa...
- Commentaires:**
- Occupations:**

	Age du Bronze	Bronze moyen	récent	-1650 / -1200	□ / 3	
				-1850 / -1450	□ / 2	2
	Néolithique final	Terrinien		-3000 / -2000	□ / 3	
	chantier 1, c.4	Age du Bronze	Torréen	-3500 / -2500	□ / 2	
	chantier 1, c.5	Néolithique récent	Basien	-4300 / -3700	□ / 2	3

At the bottom, there are buttons for 'Liste des sites', 'Liste des phases', 'Liste des dates', 'Dates liées', and 'Synthèse site'. A footer note reads: 'Fiche créée le 14/01/2014 par Thomas Perrin dernière modification le 12/02/2019 par Thomas Perrin'.

Fig. 1 – Copie d'écran de la base sites, au 12 février 2019.

et leurs datations, qui composent les trois modules principaux.

Le module « Sites » comporte actuellement près de 4700 gisements, concentrés essentiellement autour du bassin méditerranéen (de par les thèmes de recherche du concepteur ! ; fig. 1 et 2). Le concept sous-jacent à ce module est d'en rassembler les données génériques. Chaque site est ainsi localisé d'un point de vue administratif, grâce à des listes de choix normalisées : la localisation d'un site passe ainsi d'abord par le choix d'un pays, dont on propose ensuite le choix d'une région, qui ne donne alors accès qu'aux départements de ladite région, qui permet ensuite de choisir uniquement parmi les communes de ce département, etc. Le choix d'une commune entraîne ensuite l'inscription des coordonnées du centroïde géographique de la commune dans les cases latitude / longitude (pour la France uniquement). Les types de sites (plein air, grotte, dolmen, sépulture...) et d'opération (prospection, sondage, fouilles...) doivent également être renseignés. Les coordonnées sont par défaut saisies en UTM WGS84 de façon à s'affranchir des systèmes nationaux (tel le Lambert) inappropriés pour des projets à large échelle ; ces coordonnées nationales peuvent tou-

tefois être également mentionnées dans des champs spécifiques. La localisation géographique du site peut également être précisée dans un champ descriptif textuel. Un module de visualisation cartographique rapide (basé sur Google Map©) est également disponible (par un clic sur le planisphère), mais son résultat est devenu très aléatoire depuis la version 16 du logiciel. L'historique des fouilles, leur étendue ainsi que les caractères généraux du remplissage sont également précisés dans ce premier module. Enfin, un champ de type texte permet également de faire figurer les principales références bibliographiques. Celles-ci n'apparaissent que sous forme d'appels, dans un souci de gain de place et de lisibilité. L'essentiel de ces références se retrouve dans des groupes Zotero libres d'accès, développés par l'auteur et les collaborateurs de la base, dénommés « BibPreRec » (plus de 13 000 références saisies)⁴ et « Préhistoire récente de l'Afrique du Nord » (environ 500 références)⁵, groupes auxquels chacun est également libre de contribuer pour le bien de tous !

Le module « Occupations » a pour vocation de donner des informations plus détaillées sur les différentes couches, niveaux et occupations des sites en question.

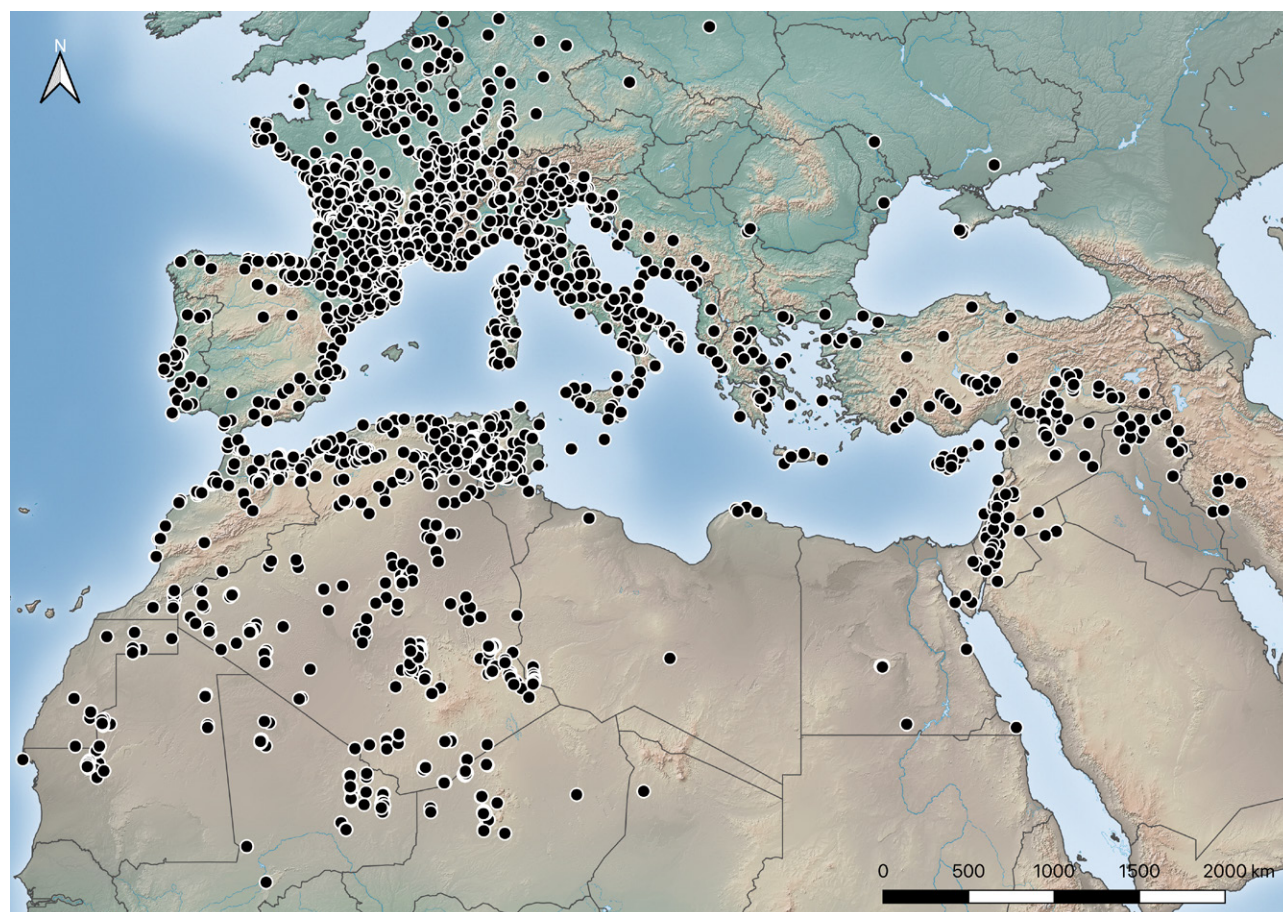


Fig. 2 – Carte des 4 673 sites présents dans la base de données au 12 février 2019.

Pour chacune de ces occupations, les données stratigraphiques et taphonomiques peuvent être précisées, de même que leur datation et la description des aménagements anthropiques. Cette dernière se fait de deux manières. Il faut tout d'abord préciser l'attribution chronoculturelle de l'occupation en question, avec plus ou moins de précision : cela peut aller de la simple période (Néolithique par exemple) jusqu'à la phase (Chasséen récent par exemple). Dans tous les cas, des fourchettes chronologiques larges sont alors automatiquement entrées dans les champs TPQ/TAQ. Ces deux valeurs sont bien entendu modifiables et peuvent (doivent) être affinées en fonction des datations absolues éventuellement disponibles ou des données typo-chronologiques régionales. D'autres onglets permettent de détailler avec plus ou moins de finesse les productions matérielles, les ressources alimentaires et les données environnementales. De par mes centres de recherche et les projets que j'ai dirigés, c'est l'onglet consacré aux industries lithiques qui est actuellement le plus détaillé, mais l'idée d'un travail collaboratif est justement que chaque expert dans son domaine puisse apporter sa pierre à la communauté et enrichir ainsi la base de ses propres connaissances et en développant de nouvelles rubriques au sein d'éventuels nouveaux onglets plus spécifiques. Actuellement, ce sont près de 6 200 occupations qui sont référencées et affectées d'un indice de fiabilité (de 1 à 3) en fonction de la cohérence et de la précision des informations disponibles.

Enfin, la situation précise de chacune de ces occupations dans le temps étant une donnée fondamentale et préalable à tout raisonnement archéologique, le module « Dates » recense toutes les datations radiocarbone liées aux occupations, elles-mêmes liées aux sites. Les critères descriptifs de ces dates sont ceux de la date elle-même (code laboratoire, mesure, écart-type, $\delta^{13}C$, année de mesure, méthode de mesure, logiciel et courbe de calibration...) et de son contexte (localisation stratigraphique, nature de l'échantillon, référence, évaluation de son lien avec l'événement daté et de la qualité de sa représentativité...). Une mention sur le programme ayant financé la mesure est également possible. Ce sont plus de 5 600 dates qui sont ainsi aujourd'hui recensées. Là encore, leur valeur informative est estimée par la valeur de l'échantillon daté (selon l'échelle proposée par Chapman et Müller, 1990) et également par une estimation subjective de leur fiabilité (douteuse, possible, fiable) basée sur la cohérence de la date en fonction des autres données disponibles sur le site ou régionalement.

À partir de toutes ces informations et de ces trois tables, il est possible d'effectuer de nombreuses recherches, plus ou moins complexes, croisant les attributions chronoculturelles aux régions, aux fourchettes de datation, aux données typologiques, etc. Toutes ces données peuvent être exportées pour produire des cartes dans des SIG ou autres, pour des traitements bayésiens des datations, etc. C'est un outil dont l'utilité me paraît évidente et dont il

me semble que l'ensemble de la communauté archéologique pourrait bénéficier en s'en emparant et en l'enrichissant de manière collective et constructive. De ce point de vue, la base est aujourd'hui placée sous licence Open Database Licence (qui ne prendra toutefois pleinement son sens qu'une fois le processus de migration vers une solution libre achevé), chaque contributeur restant toutefois responsable des données qu'il y saisit. De manière générale, seules les données publiées ou accessibles y sont mentionnées.

Trois niveaux d'accès sont possibles pour l'utilisation de BDA. Tout un chacun est libre de la consulter (niveau de compte « invité ») et d'y faire des recherches, sans toutefois pouvoir en exporter les résultats ou les modifier. De même, à ce niveau d'accès le plus simple, les coordonnées des sites ne sont visibles qu'à une décimale près, soit une précision d'environ 10 km. Cette limitation a pour but de permettre une cartographie à petite échelle tout en garantissant la protection des sites archéologiques eux-mêmes. Les chercheurs et étudiants (sur recommandation de leur encadrant) peuvent bénéficier d'un compte « lecteur » (sur demande explicite), leur permettant d'accéder à l'ensemble des données et des fonctionnalités (d'export notamment). Il est ensuite possible d'accéder au niveau « contributeur », pour ceux qui le souhaitent et qui s'inscrivent dans la philosophie générale de ce projet, et d'enrichir les enregistrements voire d'en créer de nouveaux. La suppression d'enregistrements ou la modification de la base ne peuvent se faire qu'au niveau « administrateur ».

Plus de vingt ans après sa création, il est aujourd'hui temps que cet outil soit mis à disposition de tous. Nous avons tous à gagner à ne plus perdre de temps à dépouiller inlassablement les mêmes ouvrages pour en établir des listes plus ou moins complètes et exactes. L'usage collectif et l'enrichissement constructif de cet outil nous offrent, je crois, la possibilité de nous inscrire dans une démarche de partage scientifique et surtout de concentrer nos efforts et notre temps de travail sur le cœur de nos problématiques.

NOTES

- (1) « ANR-13-CULT-0001-01 » : <http://www.agence-nationale-recherche.fr/Projet-ANR-13-CULT-0001> et <https://traces.univ-tlse2.fr/accueil/navigation/la-recherche/prbm-prehistoire-recente-du-bassin-mediterraneen/anr-procome-prolongements-continentaux-de-la-neolithisation-mediterraneenne-545072.kjsp?RH=1405424730228>
- (2) <https://menemoia.hypotheses.org/>
- (3) « ANR-17-CE27-0009 » : <http://www.agence-nationale-recherche.fr/Projet-ANR-17-CE27-0009> et <http://monumen.huma-num.fr/>
- (4) <https://www.zotero.org/groups/154609/bibprerec> ; une très grande partie de ces références sont issues de la base de références développée dans Endnote© par Dominique Commelin dans le cadre de son travail de documentaliste au sein de l'UMR 7269 LAMPEA et que je remercie d'en avoir toujours accepté le partage.
- (5) https://www.zotero.org/groups/2095865/prhistoire_recente_de_la_frique_du_nord
- (6) <https://opendatacommons.org/licenses/odbl/1.0/>

RÉFÉRENCES BIBLIOGRAPHIQUES

- CHAPMAN J., MÜLLER J. (1990) – Early farmers in the Mediterranean Basin: the Dalmatian evidence, *Antiquity*, 64, 242, p. 127–134.
- PERRIN T., MANEN C., VALDEYRON N., GUILAINE J. (2018) – Beyond the sea... The Neolithic transition in the Southwest of France, in J. F. Gibaja et al. dir, *Expansion of the Neolithic in Central and Western Mediterranean*, Quaternary International, p. 318–332.
- PERRIN T., MARCHAND G., ALLARD P., BINDER D., COLLINA C., GARCÍA PUCHOL O., VALDEYRON N. (2009) – Le second Mésolithique d'Europe occidentale : origines et gradient chronologique (The late Mesolithic of Western Europe: origins and chronological stages), *Annales de la Fondation Fyssen*, 24, p. 160–169.

Thomas PERRIN
CNRS UMR5608 TRACES
tperrin@univ-tlse2.fr