



GéoDOAD à mi-parcours

Michel Feugère, Alyssa Giraudo, Miguel Rodriguez

► **To cite this version:**

Michel Feugère, Alyssa Giraudo, Miguel Rodriguez. GéoDOAD à mi-parcours. 2019. halshs-02418628

HAL Id: halshs-02418628

<https://halshs.archives-ouvertes.fr/halshs-02418628>

Submitted on 19 Dec 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.

GéoDOAD à mi-parcours

[Michel Feugère](#), [Alyssa Giraud](#), [Miguel Rodriguez](#)

Dans un précédent billet, M. Feugère et E. Vigier ont présenté ici-même les objectifs du programme GéoDOAD (voir [ici](#)), qui a pour principal objectif de visualiser des données chronologiques issues des contextes archéologiques, en associant une cartographie dynamique à une courbe statistique pondérée. L'enjeu du projet est de prendre en compte les très nombreuses datations apportées chaque année à la connaissance des objets archéologiques par les fouilles préventives, des données très dispersées que les spécialistes ont bien du mal à retrouver dans les rapports de fouilles ou même dans les publications de sites. Faciliter un accès synthétique à ces données éparses est donc le premier objectif de notre projet, le second étant de les visualiser.

Mais intégrer ces informations dans un processus de traitement statistique et de visualisation cartographique pose un certain nombre de questions méthodologiques qu'il nous a fallu, dans un premier temps, discuter entre nous et avec divers collègues. Ce billet délivré à mi-parcours (GéoDOAD doit être achevé en janvier 2020) établit un point d'étape sur le travail en cours, tant sur la réflexion théorique que sur la mise en œuvre de ses différents éléments.

GéoDOAD se propose donc de rationaliser l'exploitation des données chronologiques d'une forme archéologique, où les données de départ consistent, pour un type donné, en plages chronologiques proposées par le responsable d'une opération pour la datation d'un niveau de fouille. Un objet provenant d'un niveau daté des années 40/60 a donc un certain nombre de chances d'avoir été produit / utilisé pendant cette plage de temps, et en cumulant des données de ce type, on espère approcher la réalité de la "vie du type" avec ses phases successives : apparition, diffusion, abandon.

Bien sûr, les archéologues produisant des données sur des fouilles différentes proposent des fourchettes de datation de durée inégale. L'ambition de GéoDOAD est de les utiliser pour fournir, sous forme de courbe, une image statistique de cette durée. On considère que si on divise la plage indiquée en tranches (d'un an ou d'une décennie), l'objet a autant de chances d'appartenir à l'une ou l'autre des tranches, mais cette probabilité diminue en fonction de la longueur de la plage ; si on cumule des plages chronologiques de durée inégale, il faut donc évidemment pondérer chaque occurrence en fonction de la durée de la plage chronologique. Le module s'appuiera donc sur la méthode de **datation probabiliste pondérée** qui permet de lisser les datations, en privilégiant les datations précises : plus la plage chronologique est courte, plus la datation est fiable (et *vice versa*).

Cette courbe se construit à partir du calcul du nombre de fois où l'unité choisie est présente dans la fourchette de datation d'un objet, afin de définir son poids par unité de temps. Le résultat nous donnera alors un taux permettant d'exprimer la fiabilité de la datation par rapport à sa durée, le poids total d'un objet étant arbitrairement fixé à 1 (**fig. 1**).



Figure 1 – Exemple de courbe de datation probabiliste pondérée commenté (DAO E. Vigier)

Une fois les taux obtenus et regroupés dans un tableau répartis par tranches chronologiques, la somme de ces coefficients nous donne un indice d'effectif de l'objet par unité de temps. Ce dernier cumule alors les datations contextuelles sur une même base comparative, et permet de rendre compte de la durée de vie d'un objet.

Questions théoriques

Avant d'utiliser les données fournies par les archéologues, nous avons cherché à en apprécier la signification. Sur une fouille stratigraphique, les fourchettes de datation déterminées par le responsable de fouille ne signifient pas toujours la même chose : une couche datée des années 50/80 peut traduire une imprécision, mais il peut aussi s'agir d'un remblai dont la constitution couvre effectivement ces trente années. Dans tous les cas, les dates indiquées correspondent à l'enfouissement des objets, leur fabrication pouvant être bien antérieure.

Le cas de Pompéi est emblématique. Si la destruction de la ville est bien datée de 79 ap. J.-C., de quand date le mobilier de ce site ? Faut-il l'attribuer à l'année précédant l'éruption, à la décennie, au quart de siècle précédent, ou davantage ? Chacun comprend que l'archéologue / historien peut adopter une valeur "raisonnable" : si certains objets de Pompei peuvent remonter à l'époque augustéenne, voire avant, il est certain que bien plus de mobilier date de dernières années avant le drame. Attribuer le mobilier de Pompéi aux années 70/79, par exemple, est donc une approximation statistiquement recevable, même si des cas particuliers sont avérés.

Traduire en données numériques des datations exprimées en langage naturel pose un autre problème, à la fois philosophique et statistique. GéoDOAD va devoir traduire une datation donnée par le fouille, par exemple "début Ier s.", en nombre (n. de décennies concernées) : faut-il comprendre "1/10 ap. J.-C.", "1/15" ou 1/30" ? Certes, l'archéologue qui a daté un contexte "début Ier s." et non pas dans la fourchette "1/10" tient à cette imprécision, qui exprime un doute méthodologique. Il nous a semblé que GéoDOAD devait traduire cette imprécision par une fourchette large, soit dans ce cas "1/30", la valeur qui sera la plus diluée par rapport à une fourchette plus courte : si l'archéologue veut être plus précis et que le résultat proposé par défaut par GéoDOAD ne lui convient pas, il peut tout à fait donner des dates plus resserrées, qui seront alors prises telles quelles.

Le biais technique ne peut être totalement évité : quand ils datent des contextes en indiquant des dates, les archéologues n'utilisent généralement que des valeurs rondes qui satisfont leur besoin d'imprécision : ainsi, sur les stèles funéraires romaines, les âges de décès sont plus souvent 25, 40 ou 45 ans que des âges intermédiaires. En exprimant une approximation, on arrondit à des valeurs repères. Nous essayons de tenir compte de ces constantes pour ne pas faire apparaître, sur les courbes de GéoDOAD, de pic ou de creux artificiel.

Étapes de travail

Comme expliqué précédemment, la courbe pondérée se construit à partir de datations numériques. Il a donc fallu réfléchir à une manière de traiter automatiquement les datations textuelles pour qu'elles correspondent au format désiré.

Au début du projet, la base était constituée de 15 370 attestations datées dont 4488 formes différentes. Cela représente moins de 10% des objets présents dans la base (à savoir, 169 978 objets au 30/11/19). Ainsi, l'un des enjeux de GéoDOAD est d'augmenter ces données chronologiques afin de disposer d'un maximum d'objets datés. Aujourd'hui, à mi-parcours du projet, 810 objets datés ont été saisis (pour 102 formes différentes). Les graphies des éléments chronologiques se répartissent actuellement selon plusieurs cas de figure :

- des datations traitables en l'état (ex. : 15/60) ;
- des siècles avec (ou non) une précision (ex. : début du Ier s. ap. J.-C.) ;
- des périodes chronologiques ;
- des noms d'empereurs ;
- des datations uniques (ex. : TAQ 56) ;
- d'autres informations qui ne sont pas liées à la datation (ex. : intrusion).

Après réflexion, nous avons convenu de traiter ces différentes graphies par lots, à l'aide d'expressions régulières et de dictionnaires de conversion. Ces derniers contiendront les éléments permettant de convertir les données textuelles en numérique, telles que les noms des périodes chronologiques et d'empereurs, mais aussi les siècles et les abréviations.

Un nettoyage des données sera toutefois nécessaire afin d'évacuer dès le début les lignes qui ne pourront être prises en compte par le script, et ainsi éviter de l'alourdir inutilement. Dans certains cas, en effet, le champ "datation" contient seulement un espace ou un "/". De même, lors de la migration vers la v2, des dimensions se sont retrouvées par erreur dans le champ "datation". Il faudra prévoir que le programme aille chercher dans le reste de la ligne de potentielles datations. S'il n'y en a pas, la ligne sera mise de côté ; s'il y en a, un contrôle utilisateur sera demandé afin de les valider. Si elles ne sont pas corrigées par l'utilisateur, ces lignes ne seront pas prises en compte par GéoDOAD, tout comme les lignes ne donnant que d'une seule date, par exemple un TAQ (*terminus post quem*).

Une fois ce premier traitement réalisée, le plus simple est de traiter les dates qui sont déjà au format numérique. Pour les autres formats, le principe général de traitement sera plus ou moins le même : le programme ira chercher les correspondances dans le dictionnaire spécialement créé pour traduire les dates textuelles en numérique. Par exemple, la fourchette "Claude-Néron" sera transcrite "41/68" pour que GéoDOAD puisse comprendre la ligne, et ensuite arrondie à 40/70 pour les besoins du script. Il restera toutefois des éléments qui ne pourront pas être traités automatiquement. Dans ces cas-là, nous avons opté pour un contrôle

utilisateur : ce dernier devra saisir manuellement les dates dans le bon format, ou alors décider que GéoDOAD ne prendra pas la ligne en compte.

Une fois le module de conversion des données lancé en production, il a fallu réfléchir à la modélisation des interfaces. Plusieurs points nous paraissent importants quant à la conception de ces interfaces, à savoir la *transparence* et la *paramétrabilité* du module. En effet, GéoDOAD propose une visualisation de données plus élaborée qu'une simple carte : il est donc important que l'utilisateur ait accès aux éléments ayant permis de construire la courbe (et qu'il ait aussi un contrôle sur les données de départ). Ainsi, les dictionnaires de convention seront consultables, tout comme le tableau contenant les datations de départ et leur transcription par le script.

L'objectif est de proposer un outil où les calculs et la visualisation sont automatisés, mais validés par l'utilisateur, non une fonctionnalité presse-bouton. Il est donc important que l'outil s'adapte aux besoins de chacun, et non l'inverse.

Progression du développement

Alors que le temps prévu pour le développement arrive à son terme, l'interface web de GéoDOAD adopte progressivement son design définitif. Dans sa configuration actuelle, l'environnement graphique proposé à l'utilisateur est constitué de cinq modules : une carte dynamique, un graphique en courbe, un graphique en anneau, un compteur et un tableau (fig. 2). Le travail récent effectué sur l'ergonomie acte la fin de l'intégration de l'essentiel des fonctionnalités primaires prévues dans le cahier des charges du projet. Dans le détail, chaque module possède ses propres spécificités techniques.

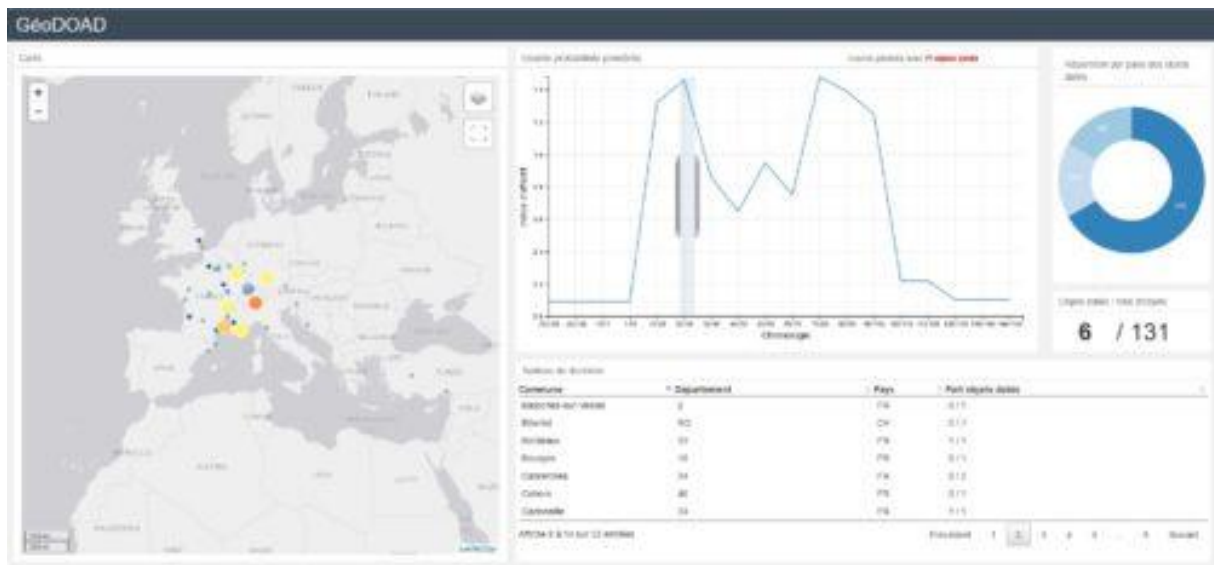


Figure 2 -Environnement graphique de GéoDOAD (M. Rodriguez)

- La carte dynamique

Elle permet l'affichage de quatre fonds de carte :

- une orthophotographie satellite
- un fond Open Street Map
- un MNE

- un fond vectoriel

L'architecture de l'application intègre par ailleurs la possibilité d'importer des jeux de données spatiales via des flux WFS/WMS et propose, à titre d'exemple, une couche représentant les provinces romaines (fournie par le Digital Atlas of Roman and Medieval Civilizations). La capacité de projeter des jeux de données internes (comme l'actuel fond de carte Artefacts) est également intégrée grâce à la mise en place d'un serveur cartographique. Les données relatives aux objets (datés et non datés) sont projetées en deux temps. Une première couche, non dynamique, permet la visualisation de l'ensemble des attestations d'un type d'objet, avec un affichage en cercles proportionnels illustrant la quantité d'attestations par commune. La seconde couche (dynamique) ne montre que les objets datés et propose un affichage catégorisé en fonction de la valeur de l'indice d'effectif propre à chaque attestation.

- Le graphique en courbe

Élément central de l'application, ce graphique (qui représente la courbe probabiliste pondérée) gère l'interactivité entre les différents modules. Il est équipé d'un curseur dont le déplacement le long de la courbe filtre les données et modifie leur affichage (en l'état actuel, seul le tableau de données n'est pas interactif). L'axe des abscisses est programmé pour s'adapter dynamiquement à la plage chronologique des attestations d'un type d'objet tandis que l'axe des ordonnées fournit des indications sur la valeur de l'indice d'effectif. L'en-tête du graphique donne également des indications sur la valeur statistique des données.

- Le graphique en anneau

L'intérêt de ce module dépasse le simple choix des données représentées (la répartition par pays des objets datés). Sa valeur réside dans sa capacité à illustrer les possibilités offertes par GéoDOAD en termes de visualisation interactive de données.

- Le compteur

Également dynamique, ce petit module permet à l'utilisateur de connaître le nombre d'objets datés dans la plage chronologique choisie par rapport au nombre total d'attestations d'un type d'objet.

- Le tableau

Il s'agit de l'unique module non interactif de l'application (pour des raisons de structure des données). Son rôle est de fournir, en format alphanumérique et de manière facilement lisible, un accès complet au jeu de données utilisé pour générer les visualisations des autres modules.

Afin d'achever la maturation de ces fonctionnalités, il est indispensable de mettre en ligne une démo de l'application. Celle-ci permettra de profiter de retours utilisateurs, toujours bienvenus en phase de développement. Dans le même temps, deux fonctionnalités importantes doivent encore être intégrées à GéoDOAD : le changement de pas, qui permettra de pondérer les indices d'effectifs en fonction de l'étendue de la plage chronologique, et l'export des données.

Fin du projet

Dans les six dernières semaines du projet, les efforts seront concentrés sur les aspects informatiques. En effet, il sera nécessaire de travailler sur la production des données de manière automatique en faisant tourner le script réalisé par les informaticiens du laboratoire pour la conversion des données chronologiques textuelles en données numériques. Enfin, il s'agira d'implanter GéoDOAD dans Artefacts de manière pérenne, afin que les auteurs de l'encyclopédie en ligne puissent l'utiliser, et nous faire part de leurs retours.

Citer ce billet : Michel Feugère, Alyssa Giraudo, Miguel Rodriguez, "GéoDOAD à mi parcours." *Le Fil d'ArAr*, 19/12/2019, <https://lefilдарar.hypotheses.org/1603>.