
Aller au-delà des données agrégées dans ArkeoGIS : premiers essais de graphes au sein d'IndexMed

Loup Bernard^{*1}, Romain David^{†2}, Jean-Pierre Féral^{‡2}, Thierry Tatoni^{§2}, Cyrille Blanpain^{¶3}, Julien Lecubin^{||3}, and Michelle Leydet^{**2}

¹Université de Strasbourg (UMR 7044) – université de Strasbourg – MISHA 5 allée du Gal Rouvillois CS 50008 67083 Strasbourg cedex, France

²Institut méditerranéen de biodiversité et d'écologie marine et continentale (IMBE) – INEE, Université d'Avignon et des Pays de Vaucluse, Institut de recherche pour le développement [IRD] : UMR237, Aix Marseille Université, CNRS : UMR7263, INSB, INSU – Aix Marseille Université, Campus Etoile, Faculté St-Jérôme case 421 Av. . escadrille Normandie-Niemen 13397 MARSEILLE CEDEX 20, France

³OSU Pytheas – INSU – France

Résumé

Le point commun des études en archéologie, biodiversité ou sur les systèmes sociaux est que la production de données est à la fois coûteuse et peu automatisée. Les suivis de longues séries temporelles et/ou à larges emprises spatiales sont difficiles à mener, dès lors qu'il faut recourir à plusieurs observateurs. La robustesse et la reproductibilité de l'observation est aussi plus difficile à obtenir.

Dans un cadre de production de données multi-sources, l'équivalence des systèmes d'observations et l'inter-calibration d'observateurs deviennent cruciales. Des approches intégratives plurivoire transdisciplinaires deviennent nécessaires, dans l'étude de systèmes où la production de données dans chaque discipline est discontinue, peu précise et mal répartie. Pourtant, toutes les variables (cartographie d'installations humaines, caractérisation des activités économiques, études des productions, recensements d'objets, données biotiques, abiotiques, cartographies des pressions anthropiques et naturelles, services rendus et ressentis, image sociétale,...) de ces systèmes interagissent dans le temps et à chaque échelle spatiale.

Après quelques années d'existence, ArkeoGIS permet d'agréger à ce jour plus de 60 bases de données représentant plus de 50 000 objets (sites, analyses). Fort de cette normalisation de l'information archéologique et paléo-environnementale, il nous a semblé important d'utiliser de nouvelles méthodes de fouille de données afin de voir si des données " connexes " peuvent être reliées à ces jeux de donnée en archéologie. Le lien entre ArkeoGIS et EPD (*european pollen database*) nous a permis de mettre en place une requête croisée et de tester cette possibilité au sein d'un prototype développé par le consortium IndexMed. Ce prototype, en open source, permet la mise en place de liens entre objets de bases de données différentes.

*Intervenant

† Auteur correspondant: romain.david@imbe.fr

‡ Auteur correspondant: jean-pierre.feral@imbe.fr

§ Auteur correspondant: thierry.tatoni@imbe.fr

¶ Auteur correspondant: cyrille.blanpain@osupytheas.fr

|| Auteur correspondant: julien.lecubin@osupytheas.fr

** Auteur correspondant: michelle.leydet@imbe.fr

Le consortium IndexMed a pour objectif d'identifier puis de lever les verrous scientifiques liés à la qualité des données et leur hétérogénéité. L'utilisation de graphes permet de les considérer malgré leur disparité et sans les hiérarchiser, et améliore l'aide à la décision en utilisant des méthodes émergentes de fouille de données (clustering collaboratif, machine-learning, fouille de graphes, représentation de connaissances) ; adapter ces méthodes à l'archéologie nous permet d'aller au-delà de la " simple " agrégation de données.

L'objectif : une meilleure compréhension globale des interactions historiques entre l'homme et la biodiversité qui sera permise par la construction et le test de méthodes de co-interprétation de ces données hétérogènes. Les méthodes de fouille de données apporteront de nouvelles perspectives aux recherches disciplinaires qui étudient en fin de compte des objets intimement liés (Lien entre données archéologiques et chimie environnementale, génomique, transcriptomique, métabolomique, écologie des peuplements/des paysages, systèmes socio-écologiques).
L. Bernard, R. David, J.-P. Féral, T. Taroni, C. Blanpain, A. Dias, J. Lecubin, M. Leydet, C. Surace

Mots-Clés: archéologie, data mining, arkeogis, indexmed, graphes