

Datapoc 2.0 : un référentiel « Personnes » spécifique au Muséum national d'Histoire naturelle

Soutenu par CollEx-Persée, Datapoc 2.0 permettra d'accroître la performance des systèmes d'information pour les gestionnaires des bases de données, d'améliorer la qualité des réponses et l'accès aux données pour les utilisateurs.

Le projet Datapoc 2.0, lauréat de l'appel à projets CollEx-Persée 2019-2020, fait suite à Datapoc.mnhn.fr¹, sélectionné lors de l'appel à projets 2018-2019. L'idée a vu le jour en 2016 à l'initiative de la Direction des Bibliothèques du Muséum national d'Histoire naturelle (MNHN), avec la constitution d'un groupe réunissant des professionnels impliqués dans la construction et la gestion des systèmes d'information au Muséum. Ces systèmes très hétérogènes relèvent de chacune des missions fondamentales de l'établissement : collections naturalistes et documentaires, recherches fondamentales et appliquées, expertises institutionnelles, diffusion des connaissances et formations. Le cœur et l'objectif final des projets Datapoc successifs reposent sur l'alignement des différents gisements et types de données que possède l'établissement (données de collecte, données de taxonomie, données d'observation, données bibliographiques, données biographiques, données archivistiques), l'hétérogénéité de ces données étant marquée autant par la nature de leur contenu que par leurs formats. La seule donnée véritablement commune à tous ces systèmes est la personne qui collecte, observe, analyse, décrit, enregistre et illustre, rédige et publie. L'alignement est donc effectué à partir de cette donnée pivot et univoque : les noms de personnes. Le premier projet était avant tout un démon-

strateur de faisabilité, réalisé à partir d'un corpus limité de noms de naturalistes contemporains ou historiques (environ 500), liés à l'établissement. Il s'agissait de tester la performance de technologies et de méthodes de traitement et d'alignement automatique de données à partir des bases disponibles, sous la forme d'un prototype d'interface orienté chercheur pour la visualisation humaine et la réutilisation par des machines des résultats et des données exposées. Après un an de développements, ponctués de tests utilisateurs, la plateforme a été publiée². Considérant le vif intérêt des utilisateurs, il a été décidé de continuer l'aventure en suivant les préconisations émises en fin de projet.

UN RÉFÉRENTIEL « PERSONNES » SPÉCIFIQUE AU MUSÉUM

Datapoc 2.0 vise donc à consolider ces premiers résultats et notamment à améliorer la qualité des données par la construction d'une base de données de référence « Personnes » propre au Muséum, sur le modèle d'IdRef. L'objectif est que chaque gisement de données au Muséum puisse utiliser ce référentiel, à la manière dont les établissements du réseau de l'Abes utilisent à ce jour le référentiel IdRef.

Pourquoi, dans ce cas, ne pas utiliser directement les systèmes d'identifiants et les bases de données d'autorités personnes existantes dans les bases de données du Muséum ? Tout

simplement parce qu'elles ne se rapportent qu'aux « personnes » impliquées dans le processus de publication d'ouvrages, de revues ou de documentation. Il n'existe pas en revanche de liste d'autorités se rapportant à la fois à des personnes « publiantes » et à des personnes exerçant d'autres rôles tels que : observateurs naturalistes, collecteurs de spécimens, donateurs, etc. La richesse des personnes figurant dans les bases de données du Muséum tient dans cette extraordinaire diversité. Elle reflète la pluralité des missions de l'établissement ainsi que la pluralité des contributeurs à l'exercice de ses missions, autant dans la dimension historique de l'établissement que dans celle, plus contemporaine, des sciences participatives.

La création de ce référentiel « personnes » au Muséum permettra à la fois d'accroître la performance des systèmes d'information pour les gestionnaires des bases de données et d'améliorer la qualité des réponses et l'accès aux données pour les utilisateurs.

Grâce au soutien renouvelé du GIS CollEx-Persée, la dynamique engagée autour du projet a permis aux professionnels de l'IST, chercheurs et informaticiens de partager une même vision de ce que doivent être des données ouvertes. Pensé à l'origine comme une preuve de concept technique, le projet a surtout très vite révélé l'indispensable nécessité de porter collectivement des projets d'ampleur très structurants, puisque celui-ci trouve aujourd'hui sa place dans le projet global de refonte du système d'information du Muséum.

CHLOÉ BESOMBES

*Chef du projet, conservatrice à la direction générale déléguée aux collections du MNHN
chloe.besombes@mnhn.fr*

CÉCILE CALLOU

*Responsable scientifique du projet Datapoc, maître de conférences du MNHN
cecile.callou@mnhn.fr*

CINDY LIM

*Chargée de mission Datapoc 2.0
cindy.lim1@mnhn.fr*

[1] PoC : Proof of concept / People of Collections.
[2] <https://datapoc.mnhn.fr>

