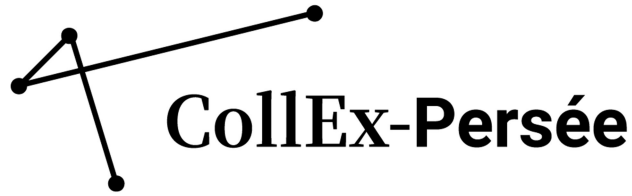


CollEx-Persée

Numérisation – Projet collaboratif



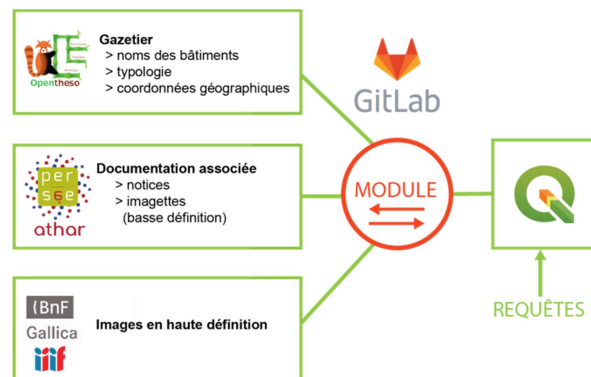
Rapport scientifique.

Projet : Géo-visualisation de contenus de la Perséide Athar : le cas du Caire moderne

Lauréat de l'appel à projet lancé le 5 avril 2018 par le GIS CollEx Persée pour susciter des initiatives en lien avec ses programmes de numérisation et le développement de services au chercheur, le projet intitulé « Géo-visualisation de contenus de la Perséide Athar : le cas du Caire moderne » visait à mettre à disposition sur ladite Perséide un corpus thématique outillé.

Le projet reposait sur trois objectifs. Il s'agissait de structurer au niveau infra-paginaire un ensemble de cinq albums déjà numérisés par Gallica, documentant la ville moderne du Caire au moyen de photographies originales ou issues de publications, ainsi que de coupures de presse. Cette structuration devait être enrichie par un référentiel des toponymes mentionnés en légende ou identifiés, qui devait servir de pivot pour alimenter un module de géo-visualisation relié à un système d'information géographique. L'enjeu documentaire et scientifique était de rendre compte du renouvellement urbain dans le temps de la ville du Caire moderne grâce à un fonds encore inexploré car assez atypique. Du point de vue technologique, l'ambition était de faire parler des données sémantisées, des images issues d'un server IIF et des données spatiales.

Ce projet reposait sur un partenariat engageant le laboratoire InVisu (CNRS/INHA), porteur du projet, l'UMS Persée, l'Institut français d'archéologie orientale – IFAO, et l'École normale supérieure de Géomatique - ENSG.



Objectifs et résultats

Le premier objectif du projet était de publier sur la perséide Athar dédiée aux corpus thématiques du laboratoire InVisu les cinq albums numérisés par Gallica.

The screenshot shows the 'Athar' website interface. At the top, there is a navigation menu with 'Persée', 'Portail Persée', 'Perséides', 'Data Persée', and 'Blog'. Below the menu is a search bar and a language selector (EN, FR). The main content area features a large image of an ornate architectural interior. Below this, there are several collection cards: 'Albums iconographiques du Caire moderne' (5 numbers, 508 documents, 1931-1960), 'Salon tunisien' (36 numbers, 294 documents, 1907-1979), and 'Publicatic' (10 num). A sidebar on the right displays '10 239 documents diffusés' and a section titled 'Qu'est-ce qu'une Perséide ?' explaining the project's goal of digitizing and structuring corpus. The footer includes 'Persée, InVisu © 2017-2020'.

Une fois les fichiers bruts des numérisations récupérés par Persée et intégrés à leur chaîne de traitement, la première tâche a été de structurer les albums via Jgalith, l'outil de segmentation développé par Persée. Afin de permettre de documenter chaque image et coupure de presse comme unité documentaire, Persée a dû développer un nouveau schéma de documentation spécialement pour le projet intégrant également la possibilité de dater la source du document. Il a également été nécessaire de développer un connecteur qui récupère pour chaque bâtiment identifié sur les images son url dans le référentiel OpenTheso. InVisu s'est ensuite chargé de traiter effectivement les albums (traitement achevé en octobre 2019), afin qu'ils puissent être publiés sur la Perséide (publication en janvier 2020) (<https://athar.persee.fr/collection/icocm>).

Cette segmentation a permis la création de 508 documents, 29 autorités « Personnes » (auteurs des clichés, des

articles ou des ouvrages cités) et 33 autorités « Organismes ».

Le corpus est désormais diffusé en accès libre et bénéficie des outils spécifiques de navigation et d'exploration mis en œuvre sur la plateforme :

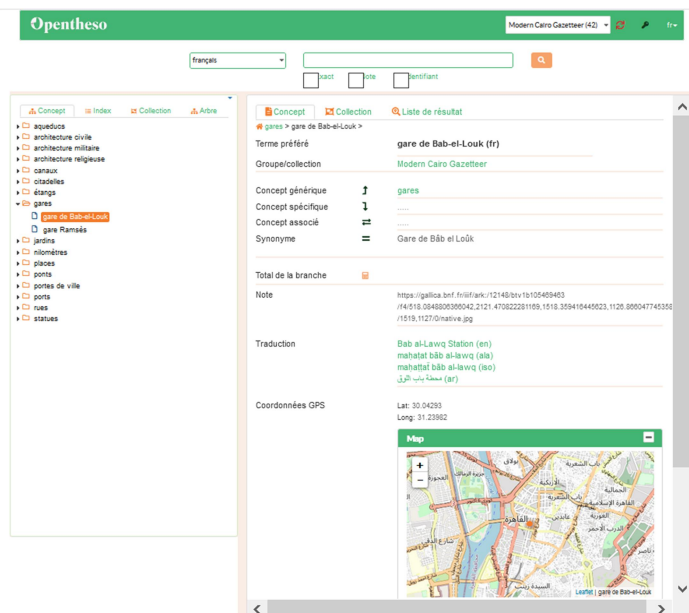
- un moteur de recherche dédié au corpus auquel sont associées des facettes et des fonctionnalités de tri des résultats
- un index « Personnes » qui permet de consulter soit les auteurs soit les personnes référencées dans les documents
- un index « Monuments » qui permet d'accéder, édifice par édifice, aux photographies et aux coupures de presse en dehors d'une navigation linéaire page à page des albums

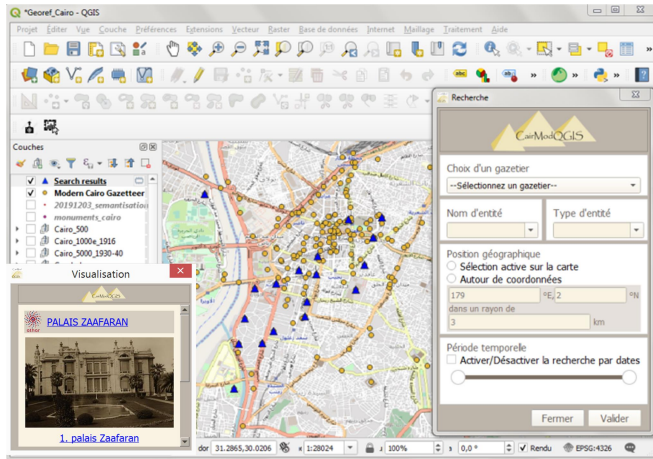
Cette première réalisation a été menée en parallèle de la construction d'un gazetier sémantique (au format SKOS). Ce gazetier vient enrichir la consultation des albums par une entrée via un index « Monuments », exposant les labels de références retenus pour chacun. Géré avec l'outil OpenTheso et hébergé par Huma-Num, ce gazetier est intitulé « Modern Cairo Gazetteer » (<https://opentheso.huma-num.fr/opentheso/index.xhtml>).

Référençant près de 300 toponymes en français, anglais, arabe, et plusieurs formes translittérées (norme ISO 233-2 1993, ALA LC, et formes libres issues des sources utilisées pour le projet), il rassemble l'ensemble des sites visibles dans le corpus iconographique. Son élaboration a impliqué d'importantes recherches afin d'identifier ou corriger les identifications fautives présentes en légende dans une partie des albums. La variabilité des désignations et des fonctions des constructions dans le temps, de même que celles des rues, ainsi que la rapidité du renouvellement urbain du Caire moderne soulèvent des difficultés particulières d'identification et de localisation qui ont dû être surmontées pour rendre cette indexation possible.

Suivant un protocole déjà expérimenté par InVisu avec d'autres gazetiers, tous les sites ont fait l'objet d'un alignement avec Geonames par la création dans cette plateforme libre de droit les points correspondants. Des alignements ont également été faits, quand cela était possible, avec Rameau, data.bnf, Wikidata et Archnet. Un lien avec l'image des bâtiments dans le server 3IF de Gallica a été inséré en note. Les coordonnées géographiques de chaque site ont été intégrées dans le gazetier, permettant la spatialisation de l'ensemble de ces données.

Dans un second temps, et grâce au recrutement d'un ingénieur de recherche géographe sur le financement du projet, le traitement spatial de ces informations a pu être entamé pour remplir le troisième objectif du projet. Un système d'information géographique (SIG) développé avec le logiciel libre QGIS a été construit afin de traduire l'évolution historique du tissu urbain cairote. Hébergé par Huma-Num, et maintenu par le service informatique de l'Ifao, ce SIG intègre un ensemble de 49 plans historiques de la ville (liste des plans classés par collection en bibliographie). Cette documentation est complétée par la spatialisation de tous les monuments référencés dans le gazetier, ainsi que par un index des noms de rue du Caire et leurs variantes au fil du temps, du XIXe siècle à nos jours. Ce SIG est mis à la disposition des chercheurs en local par l'Ifao.



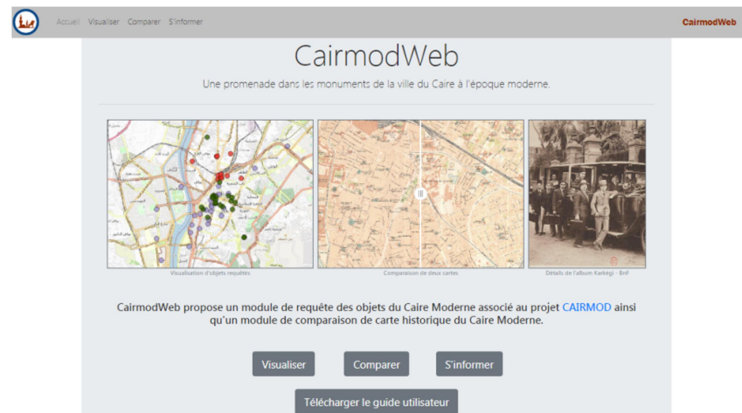


L'enjeu technologique du projet était de faire communiquer les différentes sources de données, la Perséide, le gazetier, le server 3IF, et le SIG. Pour répondre à cet objectif, une machine virtuelle a été développée ; elle permet de convertir les différents formats sources en couche d'information géographique de données. Un module de requête dans le logiciel QGIS peut ensuite interroger ces différentes bases de données et afficher ces résultats sur un fond cartographique

géré dans QGIS. Dans un second temps, une web application permettant de faire des recherches historiques à partir de ces mêmes données a été développée, en ayant cette fois pour objectif de permettre à des non-spécialistes d'utiliser cet outil en utilisant le SIG construit pour le projet.

La construction de ces deux outils est le résultat d'un partenariat fructueux avec les équipes pédagogiques de l'ENSG. Deux groupes d'étudiants de Master 2 TSI et de deuxième année (Ing2) et se sont succédés afin de développer le module de requête depuis QGIS (Master 2 TSI, projet de fin d'étude, promotion 2018-2019 puis mise à jour des briques technologiques, amélioration du plugin et rédaction de la documentation par les Ing2e année, projet tutoré, promotion 2019-2020). Une troisième équipe (Master 2 TSI, projet de fin d'étude, promotion 2019-2020) a travaillé au développement de la web application pour rendre le SIG accessible en ligne. Ces développements et leur documentation ont également fait l'objet d'un dépôt dans la forge GitLab¹.

La web application intitulée « [CairmodWeb](#) »² permet de faire des recherches historiques en ligne à travers la documentation traitée pour le projet grâce à une interface de recherche élaborée et simplifiée. S'appuyant principalement sur l'architecture et les entrées du thesaurus, elle permet de faire des



¹ <https://gitlab.com/dgricci/flask-collex-persee> (projet public, interface avec OpenTheso et le service de cartographie lié aux objets dans OpenTheso), <https://gitlab.com/cairmodweb/cairmodweb> (le client web qui affiche sur fond OpenStreetMap les objets du thesaurus, les cartes anciennes scannées et qui permet de faire des recherches et des affichages des fonds documentaires), et enfin <https://gitlab.com/Brun04/cairmodqgis> (client QGIS équivalent du client web précédent).

Le premier développement est disponible à <https://gitlab.com/tsi2019/cairmodweb>.
² Dans l'attente de sa migration, la web application est consultable sur <http://217.160.173.154/cairmodweb/visualisation>. Sa documentation technique est disponible sur <http://217.160.173.154/api/doc/>

requêtes par bâtiment avec possibilité d'auto-complétion, par zone et par date. L'application permet d'afficher les données raster et vecteur, et affiche les informations disponibles sur les bâtiments. Il est possible de gérer les différentes couches du SIG et de faire des comparaisons entre les plans à différentes dates. Enfin, il est possible d'exporter les cartes générées au format pdf.

L'extension pour QGIS³ permet quant à elle de sélectionner des gazetiers gérés avec OpenTheso, en l'occurrence le Modern Cairo Gazetteer, et de visualiser dans QGIS les sites qui sont indexés. Elle permet également de visualiser les informations liées à chacun de ces sites, provenant de Persée et d'OpenTheso. Tout comme la web application, elle permet de faire des requêtes spécifiques dans ces données à partir de fourchettes temporelles, de zones spatiales ou des labels référencés dans le thesaurus. Il est possible de faire des recherches croisées afin de faire apparaître les informations les plus précises sur des couches temporaires dans le SIG qui peuvent être enregistrées afin d'être réutilisées.

Ce module est susceptible d'être utilisé pour d'autres corpus publiés par Persée, en lien avec d'autres gazetiers gérés avec OpenTheso et d'autres fonds iconographiques gérés avec le protocole 3IF.

Diffusion et pérennité

L'ensemble du projet jouit d'une importante diffusion sur internet grâce à ses différents points d'accès, pérennes et maintenus par des équipes dédiées. Le traitement des données respecte les bonnes pratiques sémantiques et les contraintes des données FAIR, **F**aciles à (re)trouver, **A**ccessibles, **I**nteropérables et **R**éutilisables.

Chaque élément dans Persée bénéficie de la création d'un identifiant pérenne, tant pour les albums, les unités documentaires, les autorités « Personnes » et « Monuments » et il en va de même pour les concepts du thesaurus. Dans Persée, les autorités « Personnes » sont alignées sur IdRef.

Persée dispose d'une infrastructure dédiée pour assurer la sauvegarde et la diffusion de ces documents, et est hébergée par l'ENS de Lyon. Les publications sont actuellement accessibles via l'OAI du portail, les données et métadonnées de la perséide seront prochainement intégrées au graphe de Persée et interrogeables via le Triplestore data.persee.fr (à l'horizon 2022). D'ores et déjà, l'ensemble des autorités mises en œuvre dans ce projet sont accessibles via le web service Authority de Persée (<http://ws.persee.fr/authority/>) qui permet de récupérer grâce à l'identifiant de l'autorité un ensemble de métadonnées qui lui sont associées: liste de ses identifiants (Persée et autres sources alignées), rôle (sujet ou contributeur d'un document), liste des collections, nombre de références, liste des références

³ Dépôt sur OSGEO.

Le gazetier est hébergé et maintenu par Huma-Num, et administré par InVisu. Chaque label bénéficie d'une URI pérenne.

Le SIG ainsi que son versant web sont également hébergés par Huma-Num et maintenus par le service informatique de l'Ifao (en cours de déploiement).

Enfin, le module développé pour connecter l'ensemble de ces données, ainsi que le server 3IF de Gallica a fait l'objet d'un dépôt et d'une documentation dans GitLab de manière à être partagé et réemployé par l'ensemble de la communauté.

Communication et impact

Le corpus et ses métadonnées hébergés et diffusés sur la perséide Athar bénéficient d'une grande visibilité : la perséide fait l'objet à ce jour de plusieurs milliers de pages consultées chaque mois. Dès le premier mois de sa diffusion, le corpus a ainsi été visionné plus de 260 fois et comptabilise aujourd'hui 2638 documents consultés (PDF et version en ligne) et 3454 vues uniques.

Le SIG est en cours de déploiement sur les servers d'Huma-Num sous la responsabilité de l'IFAO.

Parallèlement, l'équipe du comité de pilotage a régulièrement communiqué afin de faire connaître le projet et son avancement par le biais de publications sur des blogs scientifiques, ainsi que par sa participation à plusieurs rencontres professionnelles et scientifiques. Ainsi, outre la page de présentation du projet visible sur le site du GIS CollEx Persée⁴, on pourra lire deux billets sur le blog maintenu par la Duke University (Durham, Caroline du Nord) dans le cadre du projet de recherche « La fabrique du Caire moderne » : Bulle Tuil Leonetti, « Our geovisualization project / Notre projet de géo-visualisation » publié en août 2019⁵ et Julie Erismann « Visualisation du Caire moderne dans le temps et l'espace » publié en février 2020⁶. En avril 2019, la réflexion et le travail sur la sémantisation des toponymes dans le gazetier a fait

⁴ Voir la page dédiée au projet sur le site du GIS CollEx Persée :

<https://www.collexpersee.eu/projet/cairmod/>

⁵ Bulle Tuil Leonetti, « Our geovisualization project / Notre projet de géo-visualisation », <https://sites.duke.edu/cairemoderne/2019/08/25/our-geovisualization-project-notre-projet-de-geovisualisation/>

⁶ Julie Erismann, « Visualisation du Caire moderne dans le temps et dans l'espace », <https://sites.duke.edu/cairemoderne/2020/02/21/visualisation-du-caire-moderne-dans-le-temps-et-l'espace-%d8%aa%d8%b5%d9%88%d9%8a%d8%b1-%d8%a7%d9%84%d9%82%d8%a7%d9%87%d8%b1%d8%a9-%d8%a7%d9%84%d8%ad%d8%af%d9%8a%d8%ab%d8%a9-%d9%85%d9%86/>

l'objet d'une présentation dédiée lors des journées professionnelles de l'ADBU : Bulle Tuil Leonetti et Hélène Begnis, « Polygraphie et référentiels multilingues. Le cas du projet Cairmod »⁷. Enfin, la dimension cartographique du projet et les développements ad-hoc ont été présentés lors des journées « L'interopérabilité des données de la recherche : textes, images, bases de données » organisées au Caire par l'IFAO : Julie Erismann et Hélène Begnis, « Apports de la cartographie à l'exploitation d'un corpus iconographique sur la ville du Caire »⁸.

La communication autour du projet Cairmod a également pris la forme de participations à des projets de formation pour faire avancer l'ensemble de la communauté travaillant sur l'aire culturelle Moyen-Orient / Mondes musulmans. En novembre 2019, une session de formation spécialement dédiée à l'utilisation de QGIS pour créer et utiliser un système d'information géographique a été proposée à l'IFAO, et suivie par 8 participants. En juin 2020, la présentation du projet a pris la forme d'un retour d'expérience sur les problématiques de traitement d'un corpus iconographique dans une perspective historique et cartographique lors d'une conférence donnée dans le cadre de l'Action Nationale de Formation « Digital Areal 2020 »⁹.

Bibliographie

Plans historiques géoréférencés par collection

Plans cadastraux du Caire, Survey of Egypt, 1:500° : 5 feuilles

Goad plan, 1905, 1:600° : 14 feuilles

Cairo, Survey of Egypt, 1910-1932, 1:1000° : 13 feuilles

Cairo, Survey of Egypt, 1940-1970, 1:5000° : 13 feuilles

Plans anciens, 1874-1916 : 4 feuilles

- Plan Grand Bey, 1874
- Nouveau plan du Caire, 1886
- Cairo Masr el-Kâhira, 1895
- Cairo, 1916

⁷ Bulle Tuil Leonetti et Hélène Begnis, "Polygraphie et référentiels multilingues. Le cas du projet Cairmod", <https://adbu.fr/retour-sur-les-journees-professionnelles-collex-persee-adbu/>.

⁸ : Julie Erismann et Hélène Begnis, "Apports de la cartographie à l'exploitation d'un corpus iconographique sur la ville du Caire"

(<https://www.ifao.egnet.net/recherche/manifestations/ma1270/>)

⁹ Julie Erismann, "Apport de la cartographie à l'exploitation d'un corpus iconographique sur la ville du Caire". (<http://majlis-remomm.fr/63436>)