

UGESCO

Enrichissement Spatio-temporel des métadonnées pour améliorer l'exploration et l'exploitation des collections scientifiques

La majorité des collections digitalisées des ESF (Etablissements Scientifiques Fédéraux) fait encore l'objet **de problèmes de (méta)données qui entravent l'exploration, l'interprétation et l'exploitation de leurs contenus**. La correspondance entre les objets de collection et leur niveau d'ouverture à la recherche est trop limitée, compliquant ainsi la possibilité de créer de la valeur ajoutée ou de générer de nouveaux contenus scientifiques. Par exemple, le cadre actuel des métadonnées des archives photographiques est d'une part trop restreint et d'autre part trop complexe que pour permettre une exploration adéquate des collections. Le projet **UGESCO** a développé des outils d'extraction et d'enrichissement de (méta)données géo-temporelles dans le but de **décloisonner ces collections, de les mettre en relation avec d'autres contenus** et d'en faciliter une **modélisation spatio-temporelle** afin de permettre l'exploitation interactive de ces contenus. Dans le but d'optimiser la qualité des annotations temporelles et géographiques produites par nos outils d'extraction/enrichissement automatique, le projet UGESCO mènera également des recherches sur les possibilités offertes par le **crowdsourcing de micro-tâches** en matière de validation et d'amélioration des données générées. Enfin, pour assurer l'utilisation optimale des contenus générés, le projet étudiera les techniques de management/filtrage des données pour **l'amélioration de la qualité et de l'utilisabilité** de toutes ces métadonnées. Des services de cartographie géo-temporelle seront également développés afin de permettre aux utilisateurs finaux de visualiser les données et d'interroger les contenus au moyen de requêtes. Ces modélisations ont pour objectif de permettre des **analyses croisées dans le temps et dans l'espace entre collections** et ainsi de faciliter l'interprétation scientifique de ces objets dans un cadre plus large.

Le projet interdisciplinaire UGESCO a réalisé des activités de recherche et développement en matière de **reconnaissance d'entités nommées (NER)**, d'**analyse sémantique automatisée de l'image**, de **recupération d'informations géographiques (GIR)** et de **contenus générés par l'utilisateur (UGC)**. Nous avons évalué comment les collections fédérales du CegeSoma peuvent bénéficier des innovations en ces domaines. La **structure générique pour l'enrichissement spatio-temporel des collections** a cependant pour objectif d'être applicable pour une large gamme de collections et d'institutions. Les multiples composantes/modules du projet pour l'extraction, l'enrichissement, le filtrage et la modélisation des métadonnées spatio-temporelles seront **open source** et pourront être réutilisés et adaptés pour l'ensemble des ESF et des services fédéraux afin d'enrichir leurs données et améliorer l'accès digital/mobile à leurs collections.

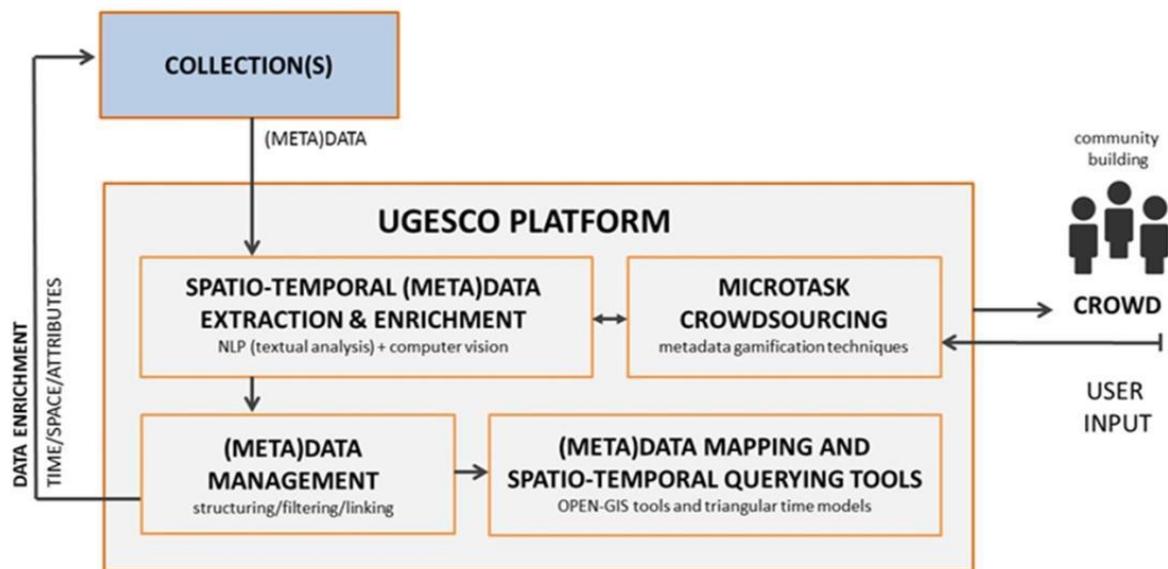
Réalisations technologiques:

- 1) Améliorer l'extraction de données spatio-temporelles au moyen analyse automatisée du langage naturel, en particulier par la reconnaissance d'entités nommées (NER) de lieux permettant la géolocalisation et de balises temporelles ainsi que par des procédés de vision computationnelle permettant la segmentation de l'image comme les réseaux neuronaux convolutifs (CNN) pour la reconnaissance de classes sémantiques d'objets.
- 2) Outils de crowdsourcing de micro-tâches pour la validation et correction/collecte de métadonnées spatio-temporelles.
- 3) Amélioration de métadonnées existantes au moyen de techniques de management/filtrage et par le développement d'outils de détection de similarités géo-temporelles au sein des collections.
- 4) Modélisation spatio-temporelle des objets de collections sur base d'outils GIS (Systèmes d'Informations Géographiques) et développement d'une méthodologie pour l'exploration temporelle des collections sur base de modèles triangulaires.

Pour chacun des défis technologiques cités, les groupes de travail concernés ont développé des modules et/ou services web réutilisables et interopérables permettant d'interagir entre eux sur base d'un format de données simple et standardisé. Les résultats du projet UGESCO permettront d'enrichir les métadonnées des ESF et d'améliorer leur qualité tout en leur offrant des possibilités de diffuser facilement le contenu de leurs collections patrimoniales auprès d'un large public.

Le projet a essentiellement utilisé des collections photographiques du CegeSoma (avec un focus sur la Seconde Guerre mondiale) mais d'autres sources également été testés afin d'évaluer les possibilités de réutilisation de la plateforme et des différents modules dans d'autres contextes et pour d'autres domaines. Les archives académiques de l'Universiteit Gent, seront par exemple utilisées pour tester les corrélations inter-collections et évaluer les outils ne pouvant être validés sur base du travail effectué sur les seules collections du CegeSoma.

Les Résultats du projet ont été présentés lors de divers workshops et conférences.



Mots-clés: enrichissement géo-temporelles, analyse d'images, crowdsourcing, filtrage de métadonnées, cartographie

COORDONNEES



Coordinateur

Steven Verstockt
Universiteit Gent – imec, IDLab (UGent)
steven.verstockt@ugent.be

Partenaires

Florence Gillet
Les Archives générales du Royaume et Archives de l'État dans les Provinces (OD4)
Centre d'Etudes et de Documentation Guerre et Sociétés contemporaines (CegeSoma)
florence.gillet@cegesoma.be

Seth Van Hooland, Ettore Rizza
Information and Communication Science department (ULB)
ettorerizza@gmail.com

Philippe De Maeyer, Nico Van de Weghe
CARTOGIS, Department of Geography (UGent)
nico.vandeweghe@ugent.be

Hans Paulussen, Frederik Cornillie
Imec - ITEC - KU Leuven, Faculteit Letteren, Campus Kulak Kortrijk (KUL)
hans.paulussen@kuleuven.be

LIENS

www.ugesco.be
