UGESCO

Upscaling the Geo-temporal Enrichment, exploration and exploitation of Scientific COllections

| DURATION	BUDGET
15/12/2016 - 15/03/2019	**527 321 €**

PROJECT DESCRIPTION

The majority of digital collections at the FSIs (*Federal Scientific Institutions*) still have **(meta)data issues affecting the exploration, interpretation and exploitation of their content**.

The coherence between the collection items and their searchability is rather limited, which sometimes makes it difficult to generate scientific/added value out of it. The current metadata scope of the photo archives, for example, is too narrow and too high-level to allow easy and adequate exploration of the collection data. The UGESCO project will develop geo-temporal (meta)data extraction and enrichment tools to **extend and link the existing collection items** and facilitate **spatio-temporal collection mapping** for interactive querying. In order to optimize the quality of the temporal and spatial annotations that are retrieved by our automatic extraction/enrichment tools, the UGESCO project will also investigate the added value of **microtask crowdsourcing** in validating and improving the generated metadata. Finally, to **ensure optimal exploitation** of the generated content, metadata management/filtering techniques will be investigated to optimize the **quality and usability** of all this metadata and geo-temporal mapping services will be developed to visualize and query the data in an end user-oriented way. These mappings allow **cross-collection analysis in time and space facilitating scientific interpretation of collection items in a broader sense**.

The interdisciplinary UGESCO project concerns research and development activities in the domains of **named entity recognition** (NER), **semantics-driven image analysis**, **geographic information retrieval** (GIR) and **user-generated content** (UGC), and investigates how the Belgian Federal CegeSoma collections can benefit from technological innovations in these domains. The proposed **generic geo-temporal enrichment** framework, however, will be widely applicable and our **open source building blocks** for **extraction, enrichment, filtering and mapping of geo-temporal metadata** can be reused and extended by all FSIs/federal departments to enrich their data and to facilitate collection access (i.e., to improve the spatio-temporal exploration and linking to other collections).
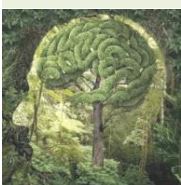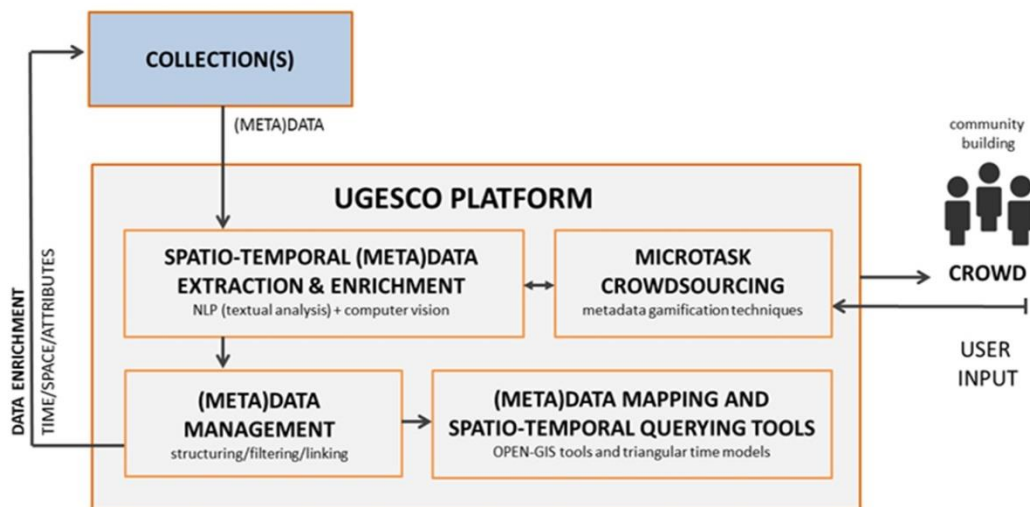
UGESCO

Technological Challenges of UGESCO project:

1) Improving spatio-temporal metadata extraction by textual analysis, e.g., named entity recognition (NER) of geo-locations / timestamps, and image clustering with computer vision techniques, such as convolutional neural networks (CNN) for object class recognition.

2) To develop crowdsourced microtasks to validate and collect geo-temporal metadata.

3) To optimize metadata management and filtering, and to provide tools for geo-temporal similarity detection.

4) Geo-temporal mapping of collection items using geographic information system (GIS) tools and an innovative approach for temporal data exploration based om triangular models.

For each of these technological challenges, a group of research partners will be responsible to create **(re)usable interoperable building blocks** that can seamlessly interact with each other and the collection data. UGESCO will allow collections to grow to improve, and to promote its content to a broader audience. In the end, our tools enable a wider exploitation of FSI collection data.

The main DATA focus will be on photographic collections (e.g. CegeSoma WWII data), but also other media-types will be investigated in the context of the (re)usability of the platform in other domains. Multimedia data of the academic heritage archives of Ghent University, for example, will be used to evaluate the cross-collection performance and to develop/test tools which cannot directly be evaluated on the CegeSoma collections. The dynamic collection linking of the CegeSoma photos and other FSIs' data with the collections of academic partners will show the strength of UGESCO's spatio-temporal enrichment process. Tests will also be performed on linking to collections hosted by partners of our follow-up committee, e.g. the photo collection of the Royal Institute for Cultural Heritage (KIK-IRPA), the CINEMATEK video collection, the collections of VIAA and ARhus Knowledge Centre, and the regional collections of cultural heritage clusters (such as Erfgoedcel Erfgoed Zuidwest).

Results of the project will be presented and demonstrated at different (inter)national workshops and conferences (e.g. Digital Dreams 2017, VISIGRAPP 2018 and Museums and the Web 2018) and will also be published in several journal articles by 2019.

CONTACT INFORMATION

Coordinator

Steven Verstockt, Sofie Van Hoecke, Samnang Nop
Universiteit Gent – imec, IDLab (UGent)
steven.verstockt@ugent.be

Partners

Florence Gillet, Mathieu Roeges
The State Archives and Archives in the Provinces (OD4)
Centre for Historical Research and Documentation on War and Society (CegeSoma)
florence.gillet@cegesoma.be
mathieu.roeges@ugent.be

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

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

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

LINKS

www.ugesco.be