

# UGESCO

## Tijdruimtelijke verrijking van (meta)data voor het optimaliseren van de exploratie en exploitatie van wetenschappelijke collecties

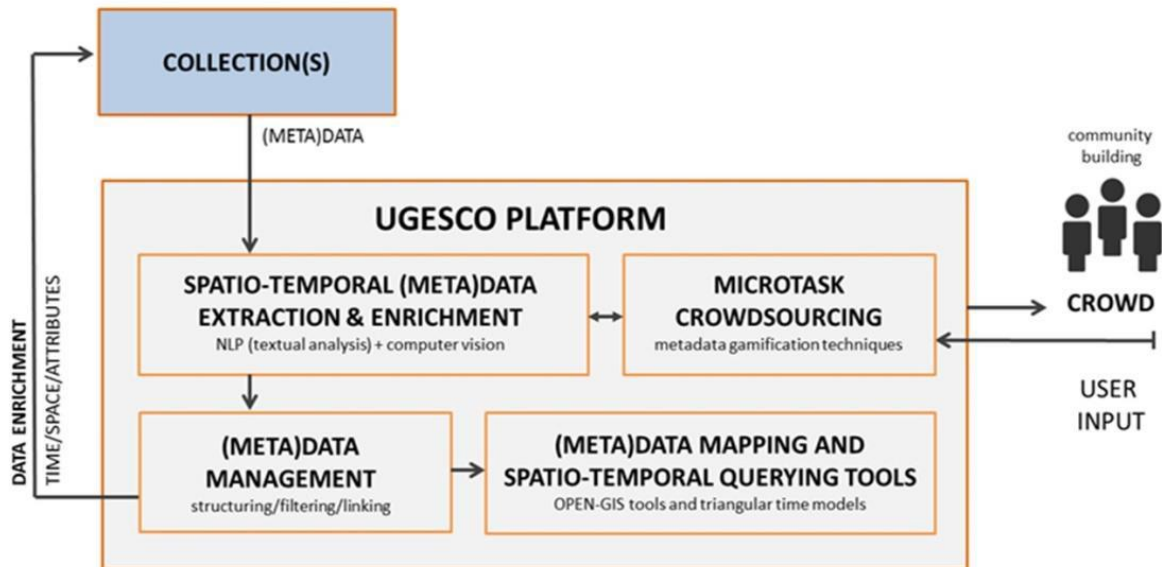
De temporele en spatiale **metadata** van de digitale collecties van de Federale Wetenschappelijke Instellingen (FWIs) zijn tot op heden nog vrij summier, wat de **exploratie, interpretatie en exploitatie van de FWI collecties bemoeilijkt**. Bovendien wordt de generatie van (wetenschappelijke) meerwaarde uit deze collecties bemoeilijkt doordat de coherentie tussen de verschillende items en hun doorzoekbaarheid te beperkt is. De huidige metadata scope van de fotoarchieven van CegeSoma, bijvoorbeeld, laat geen eenvoudige en adequate exploratie van hun collecties toe. Het **UGESCO** project (dat heeft gefocust op tijdruimtelijke verrijking van (meta)data voor het optimaliseren van de exploratie en exploitatie van wetenschappelijke collecties) heeft geotemporele metadata extractie en verrijkingstools ontwikkeld om de metadata van de **huidige collecties automatisch uit te breiden en het linken en tijdruimtelijk bevragen van collecties** te faciliteren. Om de kwaliteit van de automatisch gegenereerde “tijd en plaats” annotaties te verbeteren, is tevens ook onderzoek verricht naar de meerwaarde van participatieve validatie en correctie van de metadata via **microtask crowdsourcing**. Op deze wijze kan iedereen bijdragen tot het verbeteren van de collecties en kunnen we de “wisdom of the crowd” gebruiken om de gegenereerde metadata te verbeteren (in combinatie met automatische metadata management en filtering tools). Tot slot zijn ook tijdruimtelijke mapping services ontwikkeld om **de visualisatie en de bevraging van de data beter af te stemmen** op de noden van de eindgebruiker(s). Deze mappings laten toe om **tijdruimtelijke analyse over verschillende collecties heen uit te voeren, waardoor een bredere wetenschappelijke interpretatie van de verschillende databronnen mogelijk wordt**.

Het interdisciplinaire UGESCO project heeft onderzoek en ontwikkeling verricht in verscheidene wetenschappelijke domeinen: geografie, taalkundige analyse, multimedieverwerking, crowdsourcing en gebruikersonderzoek. Gedurende het project werd expertise opgebouwd en gedeeld omtrent **named entity recognition (NER)**, **semantisch-gebaseerde beeldanalyse**, **geographic information retrieval (GIR)** en **user-generated content (UGC)**. We evalueerden de impact van onze technologische innovaties op de Belgische Federale CegeSoma collecties. Hierbij dient echter te worden opgemerkt dat het voorgestelde raamwerk voor tijdruimtelijk verrijking breed toepasbaar is en dat onze open source bouwstenen voor extractie, verrijking, filtering en mapping van metadata hergebruikt en uitgebreid kunnen worden door andere FSIs om de digitale/mobiele toegang tot hun collecties te verbeteren.

### Techologische realisaties:

- 1) Verbeteren van de tijdruimtelijke metadata extractie/generatie door a) tekstuele analyse (voorbeeld: detectie van de plaats- en tijdsindicaties met named entity recognition) en b) clustering van beelden met computervisie gebaseerde technieken (voorbeeld: het gebruik van convolutionele neurale netwerken (CNNs) voor het herkennen van semantische klassen van objecten).
- 2) Ontwikkeling van microtask crowdsourcing tools voor de validatie en correctie/collectie van tijdruimtelijke metadata.
- 3) Optimaliseren van bestaande metadata management/filtering technieken en ontwikkeling van tools om de tijdruimtelijke similariteit tussen collecties en collectie-items te detecteren.

4) Spatio-temporele mapping van collectie items op basis van GIS-tools (geografische informatiesystemen) en ontwikkeling van een methodologie voor temporele exploratie van de collecties op basis van triangulaire modellen.



Voor elk van bovenstaande technologische realisaties zijn herbruikbare bouw-stenen/webservices ontwikkeld die naadloos met elkaar kunnen communiceren op basis van een eenvoudig standaard dataformaat. De resultaten uit het UGESCO project laten de metadata van de FWI collecties groeien en verbeteren, en de FWIs kunnen met behulp van deze tools hun content makkelijker verspreiden bij een breder publiek.

In het project hebben we hoofdzakelijk gebruik gemaakt van de fotografische collecties van CegeSoma (met een focus op WWII data), maar ook andere mediabronnen werden onderzocht om de herbruikbaarheid van het platform en zijn verschillende bouwstenen in andere contexten/domeinen te evalueren. Het academisch erfgoedarchief van Universiteit Gent, bijvoorbeeld, werd gebruikt worden om de cross-collection linking te testen en om tools te evalueren die we niet rechtstreeks op de CegeSoma content konden verifiëren.

De resultaten van het project werden gepresenteerd/gedemonstreerd op tal van (inter)nationale workshops en conferenties.

**Sleutelwoorden:** spatio-temporal verrijking, beeldanalyse, crowdsourcing, metadata filtering, mapping

## CONTACT INFORMATIE



### Coördinator

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

### Partners

Florence Gillet  
Algemeen Rijksarchief en het Rijksarchief in de provinciën (OD4) -  
Studie- en Documentatie-centrum Oorlog en Hedendaagse Maatschappij (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

### LINKS

[www.ugesco.be](http://www.ugesco.be)

---