

# UGESCO

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

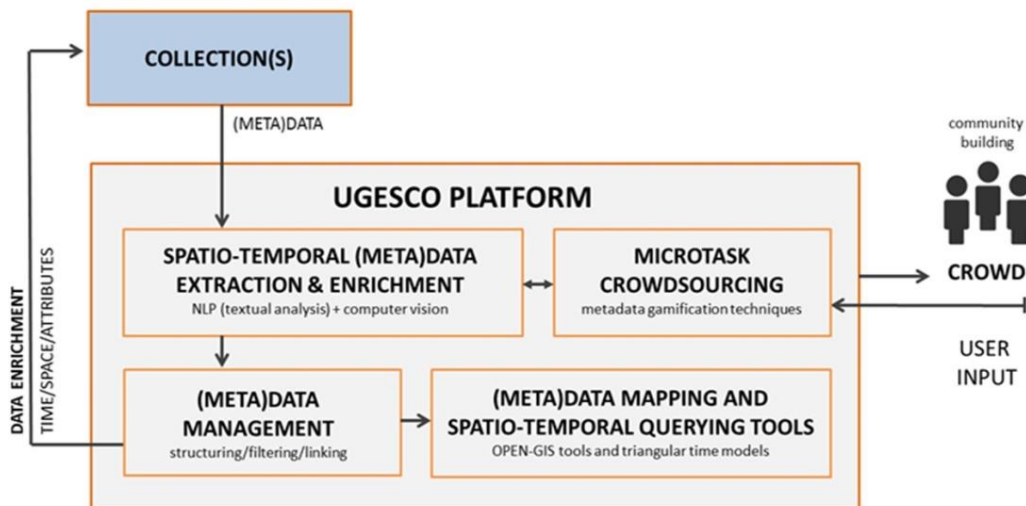
**DUUR**  
 15/12/2016 - 15/03/2019

**BUDGET**  
 527 321 €

### PROJECT BESCHRIJVING

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 focust op tijdruimtelijke verrijking van (meta)data voor het optimaliseren van de exploratie en exploitatie van wetenschappelijke collecties) zal geo-temporele metadata extractie en verrijkingstools ontwikkelen 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, zal tevens ook onderzoek verricht worden 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 zullen ook tijdruimtelijke mapping services ontwikkeld worden om **de visualisatie en de bevraging van de data beter af te stemmen** op de noden van de eindgebruiker(s). Deze mappings zullen ook **de tijdruimtelijke analyse over verschillende collecties heen faciliteren, waardoor een bredere wetenschappelijke interpretatie van de verschillende databronnen mogelijk wordt**.

Het interdisciplinaire UGESCO project verricht onderzoek en ontwikkeling in verscheidene wetenschappelijke domeinen: geografie, taalkundige analyse, multimedieverwerking, crowdsourcing en gebruikersonderzoek. Gedurende het project wordt expertise opgebouwd omtrent **named entity recognition (NER)**, **semantisch-gebaseerde beeldanalyse**, **geographic information retrieval (GIR)** en **user-generated content (UGC)**. We evalueren 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.



# UGESCO

## Techologische uitdagingen in het UGESCO project:

- 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 uitdagingen zullen de betrokken onderzoeksgroepen enkele **herbruikbare bouwstenen/webservices** ontwikkelen die naadloos met elkaar kunnen communiceren op basis van een eenvoudig standaard dataformaat. De resultaten uit het UGESCO project zullen de metadata van de FWI collecties laten groeien en verbeteren, en de FWIs toelaten om hun content makkelijker te verspreiden bij een breder publiek.

In het project maken we hoofdzakelijk gebruik van de fotografische collecties van CegeSoma (met een focus op WWII data), maar ook andere mediabronnen zullen onderzocht worden om de herbruikbaarheid van het platform en zijn verschillende bouwstenen in andere contexten/domeinen te evalueren. Het academisch erfgoedarchief van Universiteit Gent, bijvoorbeeld, zal gedurende het project ook gebruikt worden om de cross-collection linking te testen en om tools te evalueren die we niet rechtstreeks op de CegeSoma content kunnen verifiëren. Het dynamisch linken van CegeSoma foto's en data van andere FWIs, alsook met data van de academische partners, zal de kracht/meerwaarde van UGESCO's tijdruimtelijke verrijking illustreren. Daarenboven zullen ook tests worden uitgevoerd met datasets van de partners uit het follow-up committee (zoals de fotocollecties van KIK-IRPA, de CINEMATEK videocollectie en de collecties van VIAA, ARHus en Erfgoedcel Erfgoed Zuidwest.)

De resultaten van het project zullen ook worden gepresenteerd/gedemonstreerd op tal van (inter)nationale workshops en conferenties (zoals Digital Dreams 2017, VISIGRAPP 2018 en Museums and the Web 2018). Bovendien zullen we tegen 2019 ook enkele wetenschappelijke artikels trachten te publiceren (in internationale tijdschriften) over onze vernieuwende aanpak voor het tijdruimtelijk verrijken van archiefcollecties.

## CONTACT INFORMATIE

### Coördinator

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

### Partners

Florence Gillet, Mathieu Roeges  
Algemeen Rijksarchief en het Rijksarchief in de provincie (OD4) -  
Studie- en Documentatie-centrum Oorlog en Hedendaagse Maatschappij (CegeSoma)  
[florence.gillet@cegesoma.be](mailto:florence.gillet@cegesoma.be)  
[mathieu.roeges@ugent.be](mailto:mathieu.roeges@ugent.be)

Seth Van Hooland, Ettore Rizza  
Information and Communication Science department (ULB)  
[ettorepizza@gmail.com](mailto:ettorepizza@gmail.com)

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

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

## LINKS

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

