Facilitating Learning Analytics in Histology Courses with Knowledge Graphs

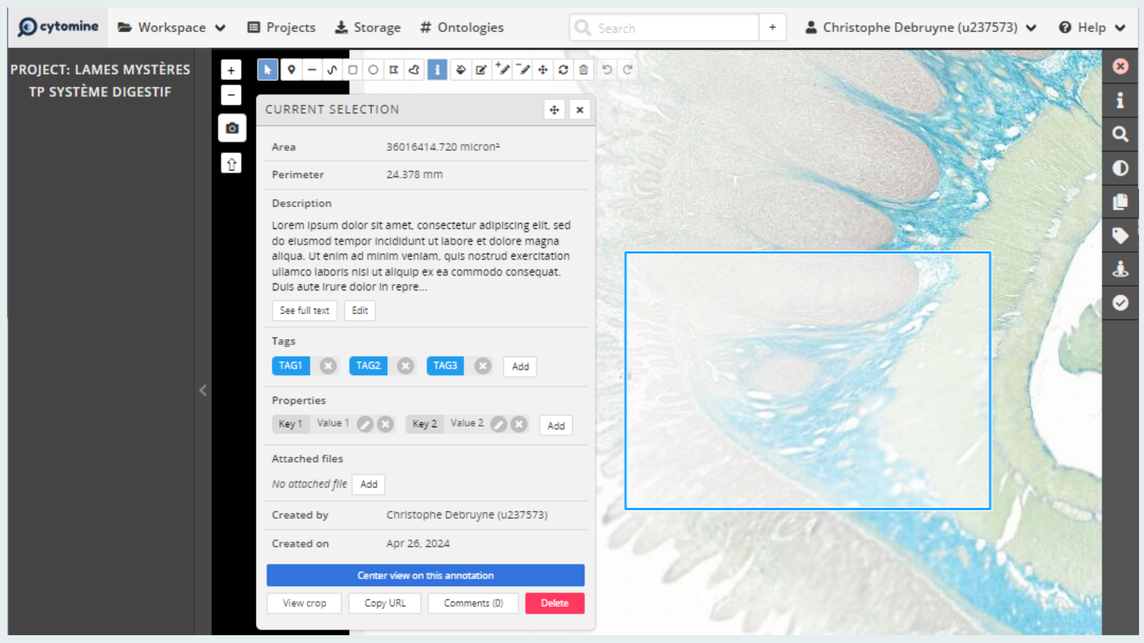
Jimmy Walraff[1] Andreas Coco[1] Guillaume Delporte[1] Merlin Michel[1] **Allyson Fries**[2] Valérie Defaweux[2] **Christophe Debruyne**[1]

[1] Montefiore Institute of Electrical Engineering and Computer Science, University of Liège, Belgium
[2] Department of Biomedical and Preclinical Sciences, Faculty of Medicine, University of Liège, Belgium

2024-09-18 at SEMANTiCS 2024

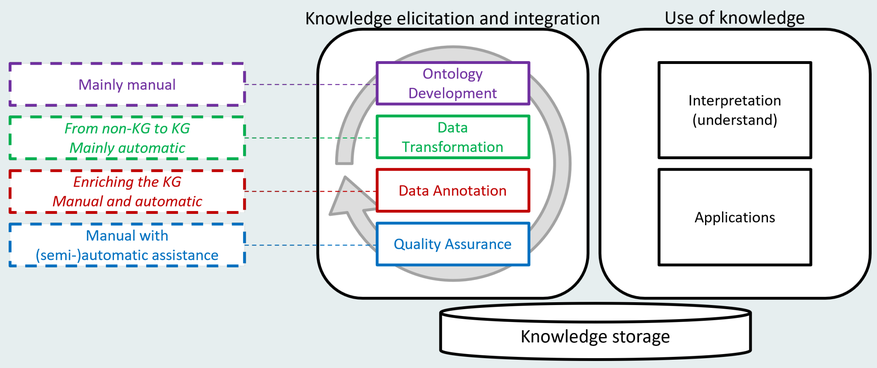
Introduction and Problem

- Cytomine [1] is a platform for collaborative exploration and analysis of large biological and medical image datasets used in histology courses at the University of Liège (and other universities).
- The document-oriented storage model of MongoDB used by Cytomine presents challenges for interconnected data analyses required in learning analytics. Prior learning analytics studies [2] relied on preprocessing pipelines to create CSV files for machine learning models, leading to various provenance issues.
- This study aims to investigate the suitability of KGs for learning analytics by making implicit relationships explicit and improving data retrieval for learning analytics.



Building CytoGRAPH

KG construction followed the Abstract Reference Architecture (ARA) [3].



Ontology Development The KG's ontology was engineered with a middle-out approach where entities in the data were identified and aligned with the UoD of domain experts and existing ontologies. We build upon PROV-O to represent the interactions between users and images and GeoSPARQL to represent the annotation's geometries, among others.

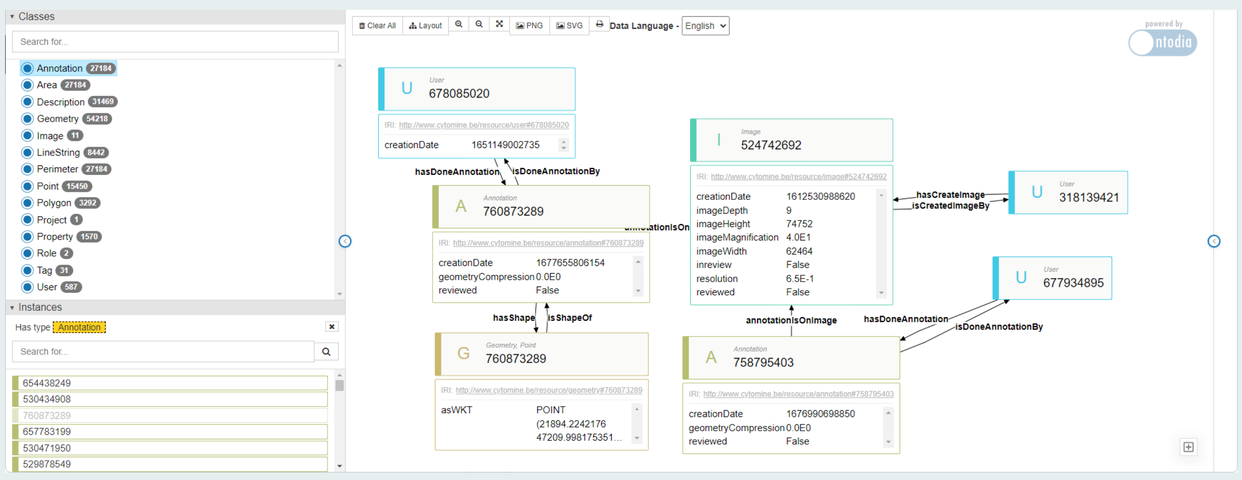
Data Transformation We had no access to Cytomine's MongoDB instance, though we could download the data via its API. The data of one project consisting of 11 images, 588 users (pseudonymized), and 27185 annotations, 1571 properties, and 31507 descriptions. We used RML to generate RDF from the data.

Data Annotation In ARA, this activity links concepts and data with other relevant sources. While we have yet to create links to other datasets and even other institutional repositories (e.g., the e-learning platform), we have decided to represent geometries using `geo:wktLiterals` so that we can retrieve activities from certain areas on the images. As such, we enriched the data with a geometric dimension.

Quality Assurance We validated the KG via competency questions formulated by the domain experts.

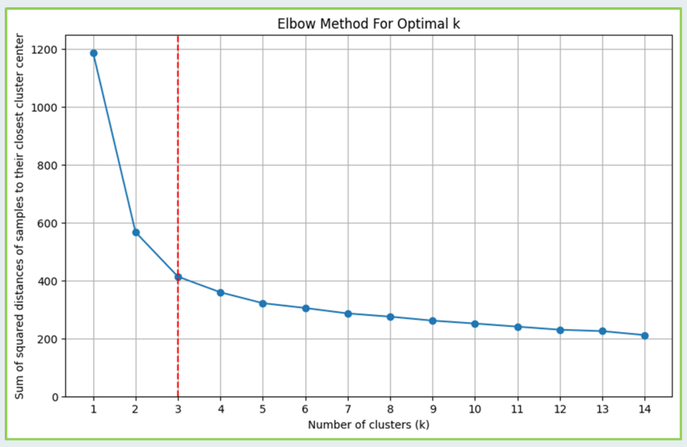
Results

The KG can be explored with tools such as Ontodia [4] and currently contains information on over 27K annotations made by 587 users over one decade, which is for the sole project to which we have access.



To demonstrate that one could engage with the KG for learning analytics, we created a Jupyter Notebook that retrieved the number of annotations per contributor and used this to determine the optimal number of clusters using the Elbow Method.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX cyt: <http://www.cytomine.be/ont#>
PREFIX rol: <http://www.cytomine.be/resource/role>
SELECT ?userID ?imageID (COUNT(*) AS ?annotationCount)
WHERE {
  ?annotation cyt:creationDate ?date ;
              cyt:isDoneAnnotationBy ?user ;
              cyt:annotationIsOnImage ?image .
  ?user cyt:userID ?userID ;
        cyt:hasRole rol:contributor .
  ?image cyt:imageID ?imageID .
  FILTER (?date >= "2022-09-15T00:00:00"^^xsd:dateTime &&
          ?date <= "2023-09-14T23:59:59"^^xsd:dateTime)
} GROUP BY ?userID ?imageID
```



A limitation beyond our control is the inability to transform the data stored in MongoDB. Moreover, Cytomine's API is fairly restricted, allowing us to retrieve data when sufficient restrictions are placed.

Conclusions

- Demonstrated the feasibility of creating a KG from Cytomine data for learning analytics.
- Demonstrated potential for learning analytics through KG exploration and Jupyter Notebooks.
- Future work includes improving data transformation methods and exploring the KG's evolution.

References

U. Rubens, R. Hoyoux, L. Vanosmael, M. Ouras, M. Tasset, C. Hamilton, R. Longuespée, and R. Marée, "Cytomine: Toward an open and collaborative software platform for digital pathology bridged to molecular investigations," *PROTEOMICS – Clinical Applications*, vol. 13, no. 1, p. 1800057, 2019.

A. Fries, M. Pirotte, L. Vanhee, P. Bonnet, P. Quatresooz, C. Debruyne, R. Marée, and V. Defaweux, "Validating instructional design and predicting student performance in histology education: Using machine learning via virtual microscopy," *Anatomical Sciences Education*, vol. 17, no. 5, pp. 984–997, 2024.

R. Denaux, Y. Ren, B. Villazón-Terrazas, P. Alexopoulos, A. Faraotti, and H. Wu, "Knowledge architecture for organisations," in *Exploiting Linked Data and Knowledge Graphs in Large Organisations*, pp. 57–84, Springer, 2017.

D. Mouromtsev, D. S. Pavlov, Y. Emelyanov, A. V. Morozov, D. S. Razdyakonov, and M. Galkin, "The simple web-based tool for visualization and sharing of semantic data and ontologies," in *ISWC 2015 Posters & Demonstrations co-located with ISWC-2015, Bethlehem, PA, USA, October 11, 2015*, vol. 1486 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2015.

LIÈGE université Sciences Appliquées

LIÈGE université Médecine