OMExcavator: a tool for exporting and connecting domain-specific metadata in a wider knowledge graph

Stefan Dvoretskii[1,2,*], Michele Bortolomeazzi[3], Christian Schmidt[4], Josh Moore[5],
Klaus Maier-Hein[1,6] and Marco Nolden[1,2,6]

[1]German Cancer Research Center (DKFZ), Heidelberg, Division of Medical Image Computing, Germany
[2]Helmholtz Metadata Collaboration (HMC), Hub Health, Heidelberg
[3]German Cancer Research Center (DKFZ), Heidelberg, Single-cell Open Lab, Germany
[4]German Cancer Research Center (DKFZ), Heidelberg, Enabling Technology Department, Germany
[5]German BioImaging - Society for Microscopy and Image Analysis e.V, Konstanz, Germany
[6]Pattern Analysis and Learning Group, Department of Radiation Oncology, Heidelberg University Hospital, Heidelberg, Germany

Abstract

Bioimaging data volume has greatly increased in recent years, and this trend is paving the way for future important discoveries in Biology and Healthcare. Even though the volume of generated data is huge, its findability, interoperability, and general reusability needs to be improved. In Bioimaging, the most used RDM system is OMERO, which stores images and their accompanying metadata in the OME Model with a possibility of interoperable export. In this work, we developed a tool to make the metadata records of images from OMERO servers available for semantic exploration. We used domain-specific tools to generate a generic metadata representation to link with other resources, in this case, unHIDE an overarching knowledge graph of the Helmholtz Association as a part of a Helmholtz FAIR data space. This tool is a critical step for the relevant communities in the Bioimaging field in Germany and beyond. Furthermore, others may face similar challenges in their respective research domains.

Keywords

Bioimaging, Linked Data, Knowledge Graph

1. Introduction

The server-client software OME-Remote Objects (OMERO) is the most used research data management (RDM) system for bioimaging data [1] and already provides a possibility for the export of metadata records in interoperable formats, e.g., the resource description framework (RDF) [2]. The OME-Model defines the representation of the image-accompanying metadata in an open, well-structured way while also preserving the raw image metadata (e.g. vendor specific metadata). This metadata can be displayed in and extracted from OMERO, where additional semantically meaningful metadata can be added in the form of structured annotations. unHIDE [3] was an ideal destination for the export of our domain-specific bioimaging data, as it is a Knowledge Graph that brings Helmholtz digital assets together in a shared representation.

2. Methods

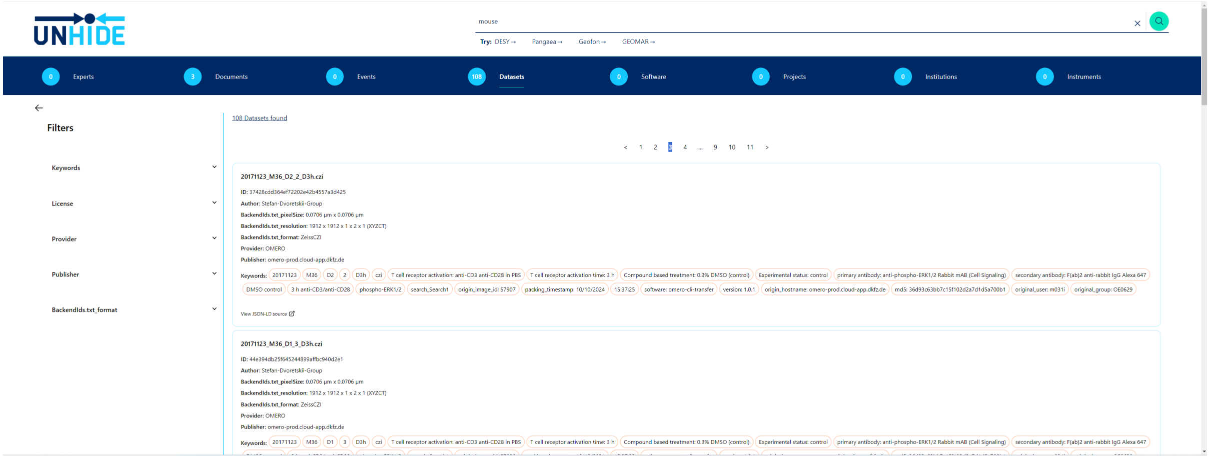
The server-client software OME-Remote Objects (OMERO) is the most used research data management (RDM) system for bioimaging data [1] and already provides a possibility for the export of metadata records in interoperable formats, e.g., the resource description framework (RDF) [2]. The OME-Model defines the representation of the image- accompanying metadata in an open, well-structured way while also preserving the raw image metadata (e.g. vendor specific metadata). This metadata can be

---

SWAT4HCLS 2025: 16th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences 2025

stefan.dvoretskii@dkfz-heidelberg.de (S. Dvoretskii)
0000-0001-7769-0167 (S. Dvoretskii); 0000-0001-5805-5774 (M. Bortolomeazzi); 0000-0001-9568-895X (C. Schmidt); 0000-0003-4028-811X (J. Moore); 0000-0001-9629-0564 (M. Nolden)

Figure 1: The unHIDE interface highlighting some images that match the “mouse” query. There are 3 additional search hits among Helmholtz digital documents (publications etc.)



20171123_M36_D2_2_D3h.czi

ID: 37428cdd364ef72202e42b4557a3d425
Author: Stefan-Dvoretskii-Group
BackendIds.txt_pixelSize: 0.0706 µm x 0.0706 µm
BackendIds.txt_resolution: 1912 x 1912 x 1 x 2 x 1 (XYZCT)
BackendIds.txt_format: ZeissCZI
Provider: OMERO
Publisher: omero-prod.cloud-app.dkfz.de
Keywords: 20171123 | M36 | D2 | 2 | D3h | czi | T cell receptor activation: anti-CD3 anti-CD28 in PBS | T cell receptor activation time: 3 h | Compound based treatment: 0.3% DMSO (control) | Experimental status: control | primary antibody: anti-phospho-ERK1/2 Rabbit mAB (Cell Signaling) | secondary antibody: F(ab)2 anti-rabbit IgG Alexa 647 | DMSO control | 3 h anti-CD3/anti-CD28 | phospho-ERK1/2 | search_Search1 | origin_image_id: 57907 | packing_timestamp: 10/10/2024 | 15:37:25 | software: omero-cli-transfer | version: 1.0.1 | origin_hostname: omero-prod.cloud-app.dkfz.de | md5: 36d93c63bb7c15f102d2a7d1d5a700b1 | original_user: m031i | original_group: OE0629

20171123_M36_D1_3_D3h.czi

ID: 44e394db25f645244899affbc940d2e1
Author: Stefan-Dvoretskii-Group
BackendIds.txt_pixelSize: 0.0706 µm x 0.0706 µm
BackendIds.txt_resolution: 1912 x 1912 x 1 x 2 x 1 (XYZCT)
BackendIds.txt_format: ZeissCZI
Provider: OMERO
Publisher: omero-prod.cloud-app.dkfz.de
Keywords: 20171123 | M36 | D1 | 3 | D3h | czi | T cell receptor activation: anti-CD3 anti-CD28 in PBS | T cell receptor activation time: 3 h | Compound based treatment: 0.3% DMSO (control) | Experimental status: control | primary antibody: anti-phospho-ERK1/2 Rabbit mAB (Cell Signaling) | secondary antibody: F(ab)2 anti-rabbit IgG Alexa 647

displayed in and extracted from OMERO, where additional semantically meaningful metadata can be added in the form of structured annotations. unHIDE [3] was an ideal destination for the export of our domain-specific bioimaging data, as it is a Knowledge Graph that brings Helmholtz digital assets together in a shared representation.

3. Results

OMExcavator makes metadata records of images stored on the OMERO servers more accessible to a wider scientific community, thus promoting scientific data sharing and scientific collaboration. The Web UI search interface allows the users to browse the bioimaging data JSON-LD records in a user-friendly manner and provides record text search capabilities. Additionally, bioimaging metadata records are put in context of the other harvested metadata records - such as relevant publications and experts. This provides added value by linking bioimaging metadata records and other relevant research assets in the Helmholtz data space.

We have highlighted some of the bioimaging records metadata fields in bold for user comfort considering the Recommended Metadata for Bioimaging (REMBI) guideline [4].

4. Conclusion & Outlook

We created a tool to export bioimaging metadata records and embed them in a wider knowledge graph. We plan to increase the connectivity of the underlying metadata records and aim for a preciser control of the access scope for the metadata records to ensure the privacy requirements of data owners can be met with more granularity. The UX will be further improved, e.g. by structuring keywords better and creating a dedicated area for bioimaging annotations.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] C. Schmidt, J. Hanne, J. Moore, C. Meesters, E. Ferrando-May, S. Weidtkamp-Peters, n. null, Research data management for bioimaging: the 2021 nfdi4bioimage community survey [version 2; peer review: 2 approved], F1000Research 11 (2022). doi:10.12688/f1000research.121714.2.

[2] J. Moore, A. Waagmeester, K. Hettne, K. Wolstencroft, S. Kunis, There and back again: RDF as a bridge to domain-platforms like OMERO, in: M. Roos, A. Jacobsen, A. Splendiani, M. S. Marshall, A. Waagmeester, L. J. G. Castro, K. Wolstencroft, K. Hettne, R. Vos (Eds.), 15th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences (SWAT4HCLS), number 3890 in CEUR Workshop Proceedings, Aachen, 2024, pp. 198–199. URL: http://ceur-ws.org/Vol-3890/#paper-40.

[3] Public instance of unhide can be accessed at https://search.unhide.helmholtz-metadaten.de/, 2025. (note: no Bioimaging records yet as of 01.2025).

[4] U. Sarkans, W. Chiu, L. Collinson, M. C. Darrow, J. Ellenberg, D. Grunwald, J.-K. Hériché, A. Iudin, G. G. Martins, T. Meehan, K. Narayan, A. Patwardhan, M. R. G. Russell, H. R. Saibil, C. Strambio-De-Castillia, J. R. Swedlow, C. Tischer, V. Uhlmann, P. Verkade, M. Barlow, O. Bayraktar, E. Birney, C. Catavitello, C. Cawthorne, S. Wagner-Conrad, E. Duke, P. Paul-Gilloteaux, E. Gustin, M. Harkiolaki, P. Kankaanpää, T. Lemberger, J. McEntyre, J. Moore, A. W. Nicholls, S. Onami, H. Parkinson, M. Parsons, M. Romanchikova, N. Sofroniew, J. Swoger, N. Utz, L. M. Voortman, F. Wong, P. Zhang, G. J. Kleywegt, A. Brazma, REMBI: Recommended metadata for biological images-enabling reuse of microscopy data in biology, Nat. Methods 18 (2021) 1418–1422.