Validating Histology Metadata for MOTHER
Multispecies Ovarian Tissue Histology Electronic Repository

Maxx A. Mudd, Suzanne W. Dietrich, Karen H. Watanabe
School of Mathematical and Natural Sciences

Introduction

MOTHER is an open-access resource that shares digitized microscope slides of ovarian tissue along with relevant specimen metadata that provides contextual information about the histology image, supporting future research in reproductive biology.

MOTHER adheres to the FAIR principles (Wilkinson et al., 2016) of sharing scientific data to ensure its longevity and usability:

- ❑ Findability – Unique identifiers, searchable index
- ❑ Accessibility – Human and machine readable
- ❑ Interoperability – Standardized structure for different systems
- ❑ Reusability – Usage licenses and data provenance

ezEML+MOTHER tool is an extension of ezEML, which is a web application designed to streamline the creation of metadata (Ide & Servilla, 2021). This modified version includes additional features for donor and immunohistochemistry information. The donor section is shown in Fig. 1.

Fig. 1. Donor information form provided by the ezEML+MOTHER tool.

ezEML+MOTHER helps maintain data provenance by documenting the origin of the digitized microscope slides to ensure transparency and authenticity. Metadata are stored using the Ecological Metadata Language (EML), a standard based on the Extensible Markup Language (XML), to provide a structured format for organizing ecological data (Jones et al., 2019). Once metadata has been submitted through ezEML+MOTHER, it is subject to curation, which is the process of submitting the digitized microscope slides and associated metadata into the MOTHER database, illustrated in Fig. 2.
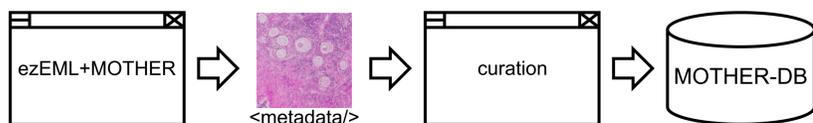

Fig. 2. Data flow from user input in ezEML+MOTHER to final upload in the MOTHER database (MOTHER-DB).

Objectives

MOTHER has specific requirements for metadata, including data type, formatting, file type, and character limits. ezEML+MOTHER uses input validation mechanisms that provide immediate feedback to the user, thus ensuring errors are identified prior to curation.

ezEML+MOTHER is built with Python and utilizes Flask, a framework for developing lightweight web applications, along with WTForms, a framework for managing web forms. This combination enables the creation of dynamic, customizable web forms and allows creation of custom input validation methods that suit the needs of MOTHER.

Methods

ezEML+MOTHER features a series of categorized web forms for entering information about slides, including origin, donor characteristics, and immunohistochemistry data. Submissions produce an XML document formatted for MOTHER, a sample of which is shown in Fig. 3.


```
<mdb:sampleProcessing>
    <mdb:fixation>
        <mdb:bouins/>
    </mdb:fixation>
    <mdb:stain>
        <mdb:lightMicroscopyStain>
            <mdb:hematoxylinAndEosin/>
        </mdb:lightMicroscopyStain>
    </mdb:stain>
</mdb:sampleProcessing>
```

Fig. 3. Selected Fixation and Stain values with the resulting XML snippet produced by ezEML+MOTHER.

XML documents must contain valid data before the submission can be curated. Input validation was put in place across fields throughout ezEML+MOTHER as a protection layer that identifies errors prior to the curation stage. To meet the requirements for MOTHER, input validation checks for issues including invalid values, improper formatting, missing fields, and character length, pictured in Fig. 4.

Fig. 4. Input validation detects issues with user input and gives immediate feedback to the user.

Fig. 5 shows input validation will enable or disable fields based on prior selection, preventing overlapping and inconsistent data.

Fig. 5. Sudan Stain Value is only enabled when Sudan is selected for Stain Light Type.

Both ezEML+MOTHER and MOTHER have constraints on file types. Validation was added that checks if a file's MIME type and its file extension match upload field requirements, shown in Fig. 6.

Fig. 6. ezEML+MOTHER rejecting an invalid file, with feedback displayed to the user.

Conclusion

Support from input validation mechanisms reduced the potential for errors that could reach the curation stage. Feedback provided to the user not only notifies them of a needed correction, but the mechanism prevents the user from submitting until the correction is made.

Validating fields within ezEML+MOTHER has improved user experience by providing instruction when invalid fields are detected, taking the burden of work away from the MOTHER team to notify users manually when a submission contains errors.

Future Work

Future validation work on MOTHER will focus on the curation pipeline. Currently, there are some unit tests in place, which are tests run to ensure individual components of a system are functioning properly. Each unit test includes a sample submission that contains rule violations or valid content. These unit tests are in place to verify that specific violations of the database requirements produce expected errors, and that valid submissions are processed correctly without producing errors.

The future of MOTHER includes creating comprehensive unit tests of the curation process. Unit tests will cover a large number of test cases based on fields that populate the database and the combinations in which they can be entered.

The goal of this unit testing is to incorporate a comprehensive validation mechanism in the curation pipeline to ensure the correctness of the system.

Acknowledgements

References

- ❑ Dietrich, S. W., et al. (2024). MOTHER-DB: Sharing nonhuman ovarian histology images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 21*(6), 2598–2603.
- ❑ Ide, J. & Servilla, M. (2021). *DATABITS: ezEML — A do-it-yourself form-based editor for creating metadata documents in EML, the Ecological Metadata Language*. https://lternet.edu/stories/databits-ezeml-ecological-metadata-language LTER Network. Retrieved: March 18, 2025.
- ❑ Jones, M. B., et al. (2019). *Ecological Metadata Language version 2.2.0*. https://eml.ecoinformatics.org/ KNB Data Repository. Retrieved: March 18, 2025.
- ❑ Watanabe, K. H., et al. (2024). Overview of the MOTHER repository. *Biology of Reproduction, 111*(3), 512–515.
- ❑ Wilkinson et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*. Article 160018. https://www.nature.com/articles/sdata201618 Retrieved: March 18, 2025.
- ❑ Zelinski, M. (2023). 'Zelinski Lab: Japanese macaque Fetal 20478-F1-09 Unspecified Ovary 003d'. https://mother-db.org/slide/mdb0000003/ MOTHER. Retrieved: March 16, 2025.
- ❑ Zelinski, M. (2023). 'Zelinski Lab: Rhesus macaque 19006 Unspecified Ovary 008a'. https://mother-db.org/slide/mdb0000203/ MOTHER. Retrieved: April 4, 2025.