MOTHER-DB: A Database for Sharing Nonhuman Ovarian Histology Images

Suzanne W. Dietrich, Wenli Ma, Yian Ding, Karen H. Watanabe, Mary B. Zelinski and James P. Sluka

***Abstract*—The goal of the Multispecies Ovary Tissue Histology Electronic Repository (MOTHER) project is to establish a collection of nonhuman ovary histology images for multiple species as a resource for researchers and educators. An important component of sharing scientific data is the inclusion of the contextual metadata that describes the data. MOTHER extends the Ecological Metadata Language (EML) for documenting research data, leveraging its data provenance and usage license with the inclusion of metadata for ovary histology images. The design of the MOTHER metadata includes information on the donor animal, including reproductive cycle status, the slide and its preparation. MOTHER also extends the ezEML tool, called ezEML+MOTHER, for the specification of the metadata. The design of the MOTHER database (MOTHER-DB) captures the metadata about the histology images, providing a searchable resource for discovering relevant images. MOTHER also defines a curation process for the ingestion of a collection of images and its metadata, verifying the validity of the metadata before its inclusion in the MOTHER collection. A Web search provides the ability to identify relevant images based on various characteristics in the metadata itself, such as genus and species, using filters.**

***Index Terms*—database, histology, metadata, ovary, female reproduction**

This work was supported in part by the National Science Foundation under Grant DBI-2054061 and P51 OD011092 (DPCPSI, ORIL, NIH to Oregon National Primate Research Center). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. *(Corresponding author: Suzanne W. Dietrich).*

Suzanne W. Dietrich is a Professor at Arizona State University, Phoenix, AZ 85069 USA (e-mail: dietrich@asu.edu).

Wenli Ma was a Master's student in the Biological Data Science program at Arizona State University, Phoenix, AZ 85069 USA. She is now with the Center for Translational Science, Florida International University, Port St. Lucie, FL 34987 USA (e-mail: wema@fiu.edu).

Yian Ding was a Master's student in the Biological Data Science program at Arizona State University, Phoenix, AZ 85069 USA.

Karen H. Watanabe is an Associate Professor at Arizona State University, Phoenix, AZ 85069 USA (e-mail: karen.watanabe@asu.edu).

Mary B. Zelinski is a Professor in the Division of Reproductive & Developmental Sciences, Oregon National Primate Research Center, Beaverton, OR 97006; and in the Department of Obstetrics and Gynecology, Oregon Health & Science University, Portland, OR 97239, USA (email: zelinski@ohsu.edu)

James P. Sluka is a Senior Scientist in the Biocomplexity Institute at Indiana University, Bloomington, IN 46202 USA (e-mail: jsluka@indiana.edu)

Details on the design of the database for MOTHER-DB can be found in supplemental materials for the article. Ovarian tissue specimens from nonhuman primates were obtained under the oversight of and approved by the Oregon Health & Science University Institutional Animal Care and Use Committee. Mouse ovaries were obtained with the approval of the Institutional Animal Care and Use Committee (IUCAC) at the University of Illinois Urbana-Champaign.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org

I. INTRODUCTION

THE goal of the Multispecies Ovary Tissue Histology Electronic Repository (MOTHER) project is to establish a shared collection of nonhuman ovary histology images for multiple species. Ovary histology slides, particularly in nonhuman primates, are expensive to obtain and currently there is no dedicated publicly available repository for the long-term storage of these valuable data artifacts. All types of ovary histology slides are encouraged for submission including field-collected specimens, laboratory studies of untreated and treated donors, and healthy and diseased animals. They may be full ovary cross-sections or specific parts of the ovary. Metadata included with each submission will facilitate the sharing and reuse of valuable histology data for broad applications, such as comparative analyses across species, development of predictive, biologically based models, and as an educational resource. Since some of the data deposited in MOTHER-DB will include pharmacological and toxicological studies, future researchers will be able to apply newer analysis tools, such as machine learning segmentation algorithms to extract information beyond what was done by the original researcher.

MOTHER adheres to the FAIR guiding principles of scientific data, providing findable, accessible, interoperable, and reusable data [1]. An important component of sharing scientific data is the inclusion of contextual metadata. The metadata must be downloadable with the data that it describes and provide the permissions for the use of the data. Additional aims for the metadata include support for data discovery and integration through the use of a formal language for knowledge representation [2]. MOTHER leverages the Ecological Metadata Language (EML) for documenting research data [3], [4]. EML is an open-source project that is managed by the National Center for Ecological Analysis and Synthesis (NCEAS) [5] and is used extensively by the Long Term Ecological Research (LTER) Network [6]. EML uses the eXtensible Markup Language (XML) and specifies an extensive XML Schema Definition (XSD) for the valid structure of an EML document. EML is designed for documenting ecological data sets, and includes substantive metadata information incorporating permissions, geographic and temporal coverage, taxonomic coverage, and data provenance. EML is extensible and is applicable beyond ecological data sets (e.g, Ocean Biodiversity Information

System [7]). MOTHER chose to extend EML, adding the metadata for an ovary histology image. This paper describes the design of the MOTHER XSD for ovary histology images and its integration with EML. In addition, MOTHER leveraged the ezEML tool [8] for building a valid EML metadata document. The extended tool, ezEML+MOTHER, which is discussed in this paper, includes the description of the donor animal, the histology image, and any immunohistochemistry. To the best of our knowledge, MOTHER is the only project that has extended the ezEML metadata tool.

MOTHER-DB is a Web-searchable database of the MOTHER metadata. The ovary histology images are not stored in the database itself. The images are stored on a separate file system that is linked to the MOTHER-DB search engine. The MOTHER database does not share all of the extensive EML metadata but a substantial subset of it. This paper describes the design of the relational database for storing the EML and MOTHER metadata. MOTHER also defines a curation process for the ingestion of new images and their metadata, ensuring the validity of the metadata before its inclusion in the MOTHER collection. A Web search tool provides the ability to locate images based on characteristics in the metadata, such as genus and species among other filters.

This paper describes the database and computational processes used by the MOTHER team to share histology images of the ovary. This process is useful for others on a similar journey to share scientific data and their metadata. Specifically, Section II illustrates how MOTHER leverages both EML and its metadata tool, ezEML, extending both with the additional information needed for donor animals, the slide image, and immunohistochemistry. Section III describes the conceptual design of the database for storing the metadata and its mapping to an implementation in a relational database. The MOTHER curation process, in Section IV, discusses the various steps involved in data validation and creation for inclusion in the database. Section V briefly overviews the MOTHER database Web search, and Section VI elaborates on applications for using MOTHER and future directions.

II. Metadata

MOTHER leverages EML for the specification of metadata. EML uses XML for knowledge representation. The valid structure of an EML document is specified by an extensive XSD for documenting ecological data sets. EML is also extensible. MOTHER enhances EML with metadata on ovary histology images and extends its ezEML tool for the specification of metadata. This extended metadata includes information on the donor animal, histology slide image and staining and immunohistochemistry details.

A. Design

XML is a formal language for knowledge representation, which fulfills the interoperable aspect of the FAIR guiding principles for metadata. There are many tools available for searching and processing XML and its valid specification using XSD. XML is a textual format with tags, called element names, that a human can read, although it is typically meant to be processed by computers. An EML document contains the metadata about a dataset, including various people (contact, creator, metadata provider), intellectual rights, geographic coverage, temporal coverage, taxonomic coverage, and project. EML includes an additionalMetadata element for extensibility. Fig. 1 shows the outline of the structure of an EML XML document extended for MOTHER.

An XSD specifies the structure of a valid XML document, much like a schema defines the structure of a database. EML's XSD specifies its structure [4], and an XSD has been defined for MOTHER [9]. At a high level, an XSD specifies the tag names, the structure of its content and associated data type, and whether it is required or optional. An XSD can also specify valid values for a type, such as enumerated textual values.

Table I provides an overview of the required and optional metadata information with respect to the XSDs of EML and MOTHER. Note that an asterisk (*) on the topic in the first column of the table indicates whether EML or MOTHER requires that content. The donor and immunohistochemistry topics are specific to MOTHER, whereas the other topics are part of EML. The Image topic is stored in the XML reusing the features of an *other entity* in EML's data set. Although EML does not require intellectual rights and taxonomic coverage information, MOTHER requires these EML items. Thus, MOTHER's curation process must check that this metadata are included.

The MOTHER XSD includes the specification of the donor, slide, and its preparation. The valid values for the specification of the metadata, e.g. life stage and stage of cycle, use an accepted scientific vocabulary developed by the ovarian research community [10], [11]. Version 1.2 of the MOTHER XSD [9] includes the specification of menstrual and estrous stages of the reproductive cycle, based on *macaque* (monkey) histology contributions by the MOTHER project itself and species from early contributors, such as *mus musculus* (mouse). The MOTHER XSD uses abstract types [12] for the specification of the valid values of the menstrual and estrous

```
<eml>
  <dataset>
    …  <!-- metadata about the dataset -->
  </dataset>
  <additionalMetadata>
    <metadata>
      <mother …>
        … <!-- MOTHER metadata -->
      </mother>
    </metadata>
  </ additionalMetadata>
</eml>
```

Fig. 1. Structure of an EML document with MOTHER

TABLE I
METADATA OVERVIEW IN EML & MOTHER

Topic/Contents * indicates required	EML		MOTHER	
	Required	Optional	Required	Optional
Title*	✓			
Image*			Name, Image Type, Data Format, Image	Additional Info
People* *[MOTHER Recommended:* ORCID, last name, first name, organization, email] *[EML Required:* At least last name, organization, or position name]	Contact, Creator	Associated Parties, Metadata Provider		
Abstract and Keywords		Abstract, Keywords		
Intellectual Rights*			✓	
Geographic/ Temporal Coverage		Geographic, Temporal		
Taxonomic Coverage*			✓	
Methods				
Project	Project Title	Project Abstract		
		Personnel		
		Funding Awards		
		Related Projects		
Donor* *[NOTE: Compound, dose, route, duration is required IF the Experimental Treatment is Treatment or Control Mock Treatment]*			Donor ID	Years
			Sex	Days
			Life Stage	Day of Cycle
			Specimen Seq Number	Stage of Cycle
			Specimen Tissue	SectionSeqNumber
			Ovary Position	Microscope Maker
			Specimen Location	Microscope Model
			Slide ID	Microscope Notes
			Experimental Treatment	Compound, dose, route,duration
			Section Thickness	Other Pathology
			Section Units	
			Fixation	
			Stain	
			Magnification	
Immunohistochemistry [primary & secondary antibody]			Target Protein	RRID
			Detection Method	
			Target Species	
			Host Species	
			Dilution	
			Lot Number	
			Cat Number	
			Source Name, City, State	
			Clonality [primary]	

cycles in the XSD. This is similar to the use of abstract types/classes in object-oriented programming languages that force the instantiation of a concrete instance. In XML, abstract indicates that the XML data instance will specify the type of values to be validated, e.g. menstrual or estrous. Thus, as additional species are added to MOTHER, the validation can be expanded if the added species uses a different terminology for stages of the reproductive cycle.

Scientists can also explicitly enter the stage of the reproduction cycle, if the stage is not in the current version of the XSD. Fig. 2 shows snippets of XSD in part (a) to show the abstract definition of reproduction cycle stages and its extension for the menstrual cycle and part (b) shows its concrete type specification in the XML data instance. MOTHER encourages scientists' feedback via email for further extensions to support the validation of the data entered, which would also be incorporated in the metadata tool.

Fig. 2 shows that the options for the stages of the menstrual cycle type are predefined choices of XML elements rather than enumerated string content. This design decision was based on the need for additional information for some of the choices, e.g. follicular and luteal. These elements introduce a value attribute that contains the string representing the valid values. The MOTHER XSD applies this approach to ensure consistency across its design.

B. Tool

MOTHER extends the ezEML Web-based tool for entering the metadata. Similar to ezEML, ezEML+MOTHER requires a login that is email authenticated, enabling access to the system

a) XSD - definition of abstract for reproduction cycle stages

```
<xsd:complexType name="stageType">
  <xsd:sequence>
    <xsd:element ref="stageTypeAbstract" minOccurs="0" />
  </xsd:sequence>
</xsd:complexType>

<xsd:element name="stageTypeAbstract" abstract="true" />

<xsd:complexType name="menstrualStageType">
  <xsd:complexContent>
    <xsd:extension base="stageType">
      <xsd:choice>
        <xsd:element name="follicular" type="follicularType"/>
        <xsd:element name="pre-ovulatory" type="emptyElementType"/>
        <xsd:element name="ovulation" type="emptyElementType"/>
        <xsd:element name="luteal" type="lutealType"/>
        <xsd:element name="unspecified" type="emptyElementType" />
      </xsd:choice>
    </xsd:extension>
  </xsd:complexContent>
</xsd:complexType>

<xsd:complexType name="follicularType">
  <xsd:attribute name="value" type="follicularValues" use="required" />
</xsd:complexType>

<xsd:simpleType name="follicularValues">
  <xsd:restriction base="xsd:string">
    <xsd:enumeration value="early"/>
    <xsd:enumeration value="mid"/>
    <xsd:enumeration value="late"/>
  </xsd:restriction>
</xsd:simpleType>
```

b) XML - data instance

```
<mdb:stageOfCycle xsi:type="mdb:menstrualStageType">
   <mdb:follicular value="early" />
</mdb:stageOfCycle>
```

Fig. 2. Definition of abstract in XSD and use in XML

for saving documents across sessions and submitting the image and metadata. The ezEML+MOTHER tool guides the user through a sequence of Web forms to enter information into various categories, such as people, abstract and keywords, intellectual rights, geographic/temporal coverage, taxonomic coverage, methods and project. MOTHER modified the concept of an EML data set to an image and extended the tool to include Web forms for the donor animal with slide information and, if appropriate, immunohistochemistry. ezEML+MOTHER is available on the project's Web site [13].

Fig. 3 shows the donor form. Fields marked with an asterisk are required. The donor ID must be a unique designation of the donor animal for the ovary image across all donors. Thus, this is suggested to be an encoding of a Lab designator (initials or abbreviation), the genus and species (first letter of genus and first letter of species) and the scientist's unique animal id, which may be a simple number or the date with a sequence number that uniquely identifies the donor. The uniqueness of the donor ID is verified during the curation process. Additional information about the donor includes age and life stage. A specimen refers to the ovary from which the slide's tissue section was obtained. The specimen sequence number is typically 1 unless there are images being shared from multiple ovaries for the same donor. The position of the ovary indicates whether it was the left or right

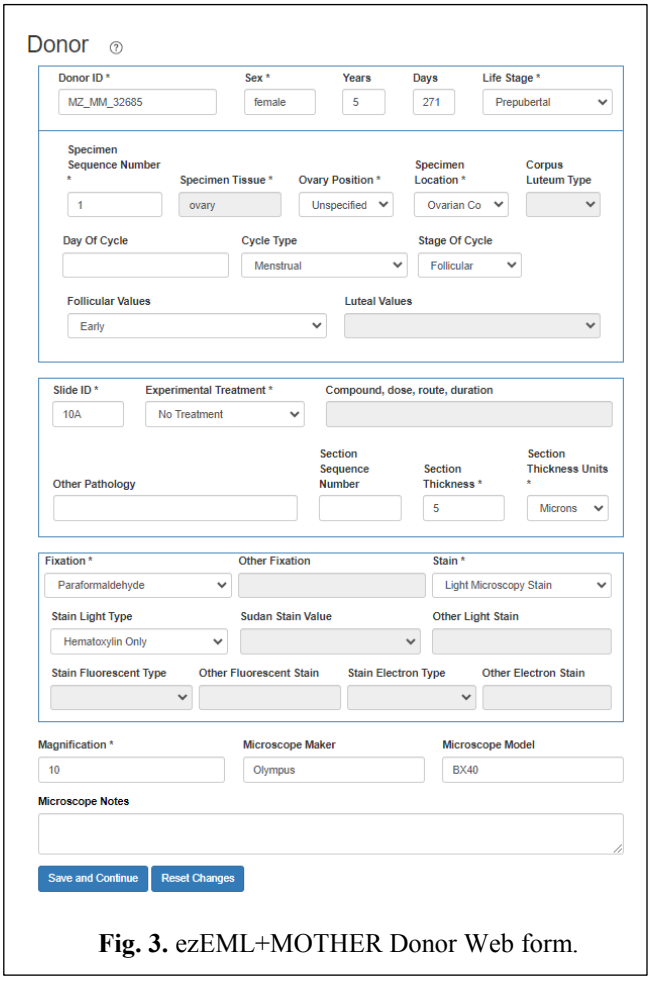

Fig. 3. ezEML+MOTHER Donor Web form.

ovary or unknown. The location choices include whether the specimen was the whole ovary or located within the ovary, such as cortex, medulla, or corpus luteum. The metadata also allows for the specification of the day of cycle or stage of the reproductive cycle, if known, for the donor at the time that the specimen was collected. Each slide for a donor must have an identifier that is unique for the donor specimen. If multiple sections on a slide are shared, the slide id must include the slide number and section designation. MOTHER also explicitly asks for whether the donor animal was involved in an experiment with the following choices: no treatment, control no treatment, control mock treatment, and treatment. If control mock treatment or treatment, the 'Compound, dose, route, duration' field is required to save this information. In addition, the optional 'Other Pathology' field records whether the donor animal had any other pathologies due to diet, infection, or genetic background. The fixation and stain are required along with the section thickness and units. The magnification of the digitized image is also required. If the slide preparation used immunohistochemistry, then the details about the primary and secondary antibodies are specified using the immunohistochemistry form.

The Check Metadata category validates the information entered. There is a circle symbol next to the category title that summarizes the status of the XML document, where red signals that an error exists, orange-yellow means that there are warnings, and green indicates that there are no errors or warnings. Selecting the Check Metadata category displays a screen with any errors or warnings. Once the errors are fixed and warnings reviewed, the user can submit the metadata and image to MOTHER. Once the user has shared all their images and the associated metadata, they should contact MOTHER for curation.

The tool also facilitates uploading slides for the same donor animal. The user can download the completed XML for the first slide of a donor, and then upload that XML into a new document for any subsequent slide from the same donor. This populates all fields for the new slide with the previous information except for the title, image, and slide id, since these fields are unique for each image.

MOTHER added a general feature to ezEML+MOTHER to view an XML document generated by the tool. The FAIR principles promote the sharing of the metadata along with the scientific data itself. The shared XML metadata file can be examined using a text editor, but this approach is not user friendly. ezEML+MOTHER allows the View option under the Upload/Download/View feature so that the user can view the metadata directly in the tool. A "View Next" button appears to navigate the metadata, and all edit features of the tool are disabled.

There is an alternate pathway for sharing a collection of images with MOTHER. A user can establish a project on osf.io [14] to store their images and the corresponding XML metadata that they create using ezEML+MOTHER. The user then contacts MOTHER, providing access to the OSF project for curation. Establishing a project on osf.io provides an additional method of discovery of the user's images. A user may also want to choose this method if they have additional information about their images that they want to share, which goes beyond the metadata stored by MOTHER. There is an additional information field for an image in which the scientist can include their OSF project link and other information not explicitly collected by MOTHER, such as a publication citation.

III. Database

The initial release of MOTHER-DB consists of over 300 digitized histology slides from three macaque species (rhesus, japanese, and cynomolgus) contributed by the MOTHER project itself, with a first contributor sharing 12 digitized mouse slides. There are many scientists who have committed to contributing slides to MOTHER, who had to wait until the ezEML+MOTHER tool was available for the creation of the accompanying metadata. Thus, MOTHER-DB will continue to expand as collaborators contribute their histology slides and metadata, which are curated as part of MOTHER.

This section elaborates on the design of the database for storing the metadata of the ovary histology images. First, an Entity-Relationship (ER) Diagram captured the conceptual design of the database, which was then mapped to its implementation in a relational database [15]. There are numerous database texts available that provide more information on ER Diagrams and the process for mapping ER Diagrams to a relational schema, e.g. see [16].

A. Conceptual Design

ER Diagrams provide information on the entities or concepts of importance in the database, and relationships

describe the associations between these concepts. A conceptual design of the EML metadata was designed and then extended to include the donor, slide, and immunohistochemistry information where some of the slide metadata overlaps with that of the dataset from EML. The supplementary materials provide a color-coded view of the ER diagram with green symbols representing concepts leveraged from EML, blue symbols indicating additions for MOTHER and red symbols meaning that it is not part of the metadata but generated by the database itself [15]. For example, when a slide is processed for incorporation into the database, an accession number is generated to uniquely identify that slide in the format MDB#######. Another example is the reuse of the unique id generated for a person.

B. Database Schema

The conceptual design of the database was then mapped to a relational database schema using established heuristic approaches. The supplementary materials provide a document that includes each table along with the attributes and their description, along with a textual summary of the database schema [15]. This presentation uses the same color coding scheme of green for EML reuse, blue for MOTHER, and red for database only. Fig. 4 provides an abstraction of the database schema that shows the table names and their primary and foreign keys. A primary key is the attribute or combination of attributes that uniquely identify a row in the table. A foreign key is the appearance of a primary key from a table that is included in another (foreign) table to relate the information. For example, the primary key of 'slide' is the 'slide_accession_number', which appears as foreign keys in the tables: 'ihc', 'metadata_provider' and 'funding_source'.

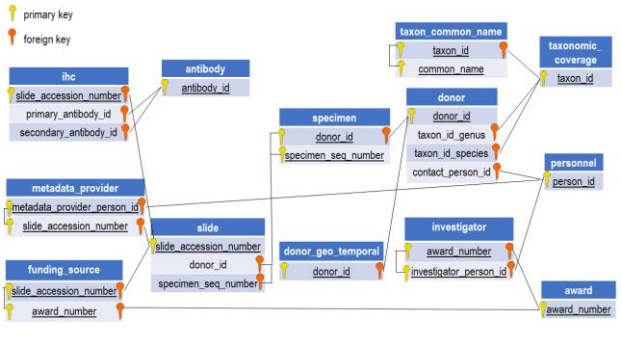

Fig 4. Table and Keys in MOTHER-DB

IV. CURATION

A. Overview

MOTHER defines a process for the curation of a collection of images and its metadata, verifying the validity of the metadata before its inclusion in MOTHER-DB. Specifically, the collection of images and metadata are in a project folder either from uploads to the ezEML+MOTHER tool or through an OSF project. The MOTHER team wrote a series of Python scripts for curation that are run against the collection. If errors are found during the validation and verification phase for the images and metadata submission, then the MOTHER team will coordinate with the submitter to address the issues before they can be incorporated into MOTHER-DB.

B. Curation Process

Fig. 5 provides a flowchart that overviews the processing in the curation pipeline, assuming that the collection of images and metadata are in a folder that can possibly have subfolders. Interactions with files and the database are not explicitly shown on the diagram to increase readability. There are several Python scripts in the curation pipeline:

1) XMLfileValidation
   Validates the XML document against the EML and MOTHER XML Schema Definitions.
2) VerifyProject
   Verifies that there is an XML document for each image and an image for each XML document.
3) ReducedThumbnails
   Generates both a reduced-size image and a thumbnail image of the shared histology image.
4) XMLtoDB
   Processes the valid XML file and generates an SQL file of insert commands to add the metadata about the image into MOTHER-DB.
5) DBscript
   Processes the SQL file of insert commands generated by XMLtoDB to update the database.

These scripts are called by an auto-curation script [17], which is given the folder of images and metadata, coordinating the calls to the other scripts in the process. If the XMLfileValidation script determines that there is an invalid XML file or if VerifyProject determines that there is a missing image/metadata file, then the process terminates, resulting in MOTHER contacting the metadata provider. Note that the XMLfileValidation script also includes checking MOTHER-specific requirements for the EML, such as requiring the inclusion of intellectual rights and taxonomic coverage, which are optional in EML. For a validated project, the ReducedThumbnails script verifies that a valid image file was uploaded and generates a thumbnail image and a reduced-size image of the original histology image. (Note that the quality of the image is checked manually during the curation process.) The Web search displays the thumbnail image on the Web search page. MOTHER decided to also generate a reduced-size image that may be suitable for some users, e.g. educators. An original histology image may be several hundred megabytes in size, suitable for scientific applications. However, some users may not need the original full-size image and may choose to download the reduced-size image, which is about 10% of the original size. These thumbnail and reduced-size images are added to the original project folder with an appropriate suffix, e.g. _thumbnail or _reduced, added to the name of the shared image.

The next step in the curation pipeline is to process the metadata. Recall that MOTHER-DB stores the metadata about the histology images. The current version of the database must be

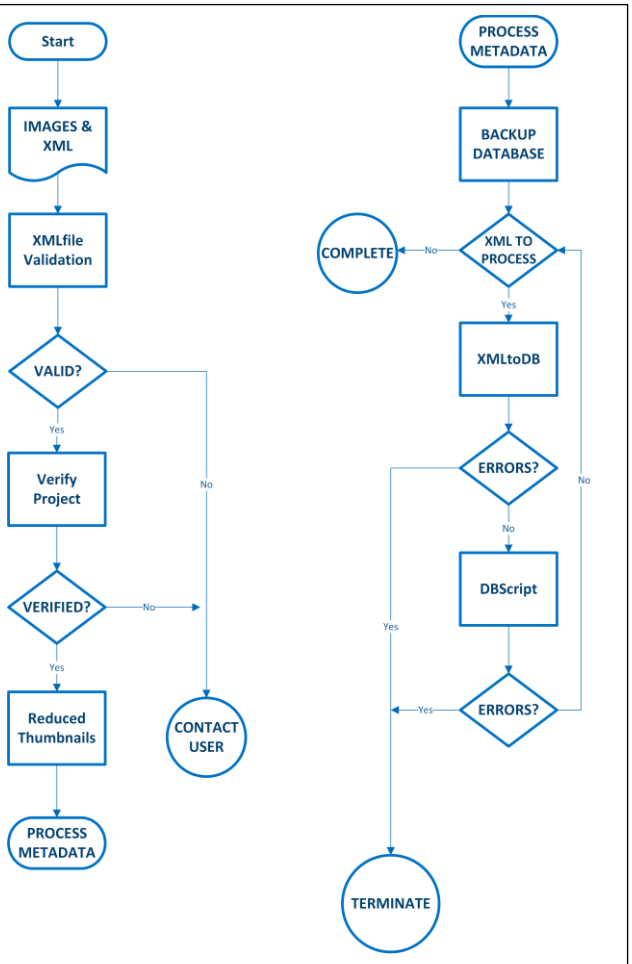

Fig. 5. Overview of the MOTHER Curation Process

backed up before processing the metadata files for the collection. The XMLtoDB script is called for each file in the collection. For repeatability, the metadata files are processed in sorted order by name. XMLtoDB queries the database during its processing to generate an SQL script for incorporating the metadata for that image into the database. Each histology image is assigned a unique accession number for MOTHER in the format MDB########. Queries to the database include determining the next available accession number and checking whether a person is already in MOTHER-DB. The script also adds the association of a histology image with its MOTHER accession number to a file for later use in moving the files into a Web-accessible location for searching and downloading. If no errors occurred, then the SQL script generated by XMLtoDB is executed by DBscript before processing the next metadata file. If all files in the collection are processed successfully, then the updated database can be published for searching.

V. Web Search

The Web search over MOTHER-DB is available on the project's Web site [13], with a screenshot shown in Fig. 6. The site provides filtering for various features within the metadata, such as the genus/species, life stage, stage of cycle, slide fixation, and stain. The values shown for filtering under each category

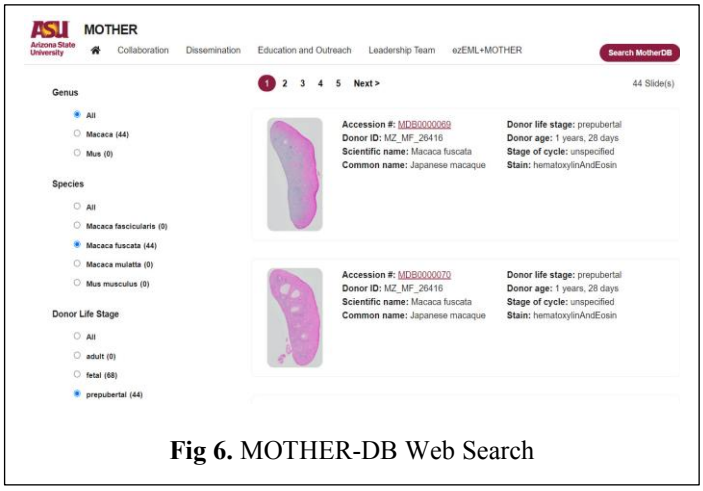

Fig 6. MOTHER-DB Web Search

are based on the histology images that are currently part of the MOTHER collection. Thus, these values will expand as MOTHER-DB continues to grow. In addition, there are future plans to continue to update the Web search and download capabilities based on ongoing research.

VI. Conclusion

The MOTHER data sets will facilitate comparisons across species and allow analyses of ovarian morphology that can lead to further insights into cell and tissue states. The inclusion of samples from pharmacological and toxicological studies will provide a valuable resource for future analysis and comparisons of reproductive effects across species. In addition, the data in MOTHER can form the basis for developing species-specific computational models of ovarian function. We are developing automatic follicle identification and counting processes using artificial intelligence and machine learning (AI/ML), as well as image download tools that will allow other researchers to bulk-download images for their own work. Finally, MOTHER captures these well-documented, high-value images that will be an important resource for future research.

Acknowledgment

The MOTHER team thanks the many students who assisted in the development, testing, and documentation of the ezEML+MOTHER tool: Javier Galaviz, Nick Mertz, Pierce Tyler, William Baird, Justin Raymer, Steven Anderson, Gerry Fernandez, and Michael Barden. Thanks are due to Jeremy Juve for the refinement of the auto curation pipeline to include a framework for unit testing, which is beyond the scope of this paper. We also appreciate the members of ASU's research computing services who assisted in the design and development of MOTHER's Web presence and database search capabilities: Nathan Rollins, Matthew Thompson, Justin Holloway, and Walter McConnell.

References

[1] J. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," *Sci Data,* vol. 3, article 160018, 2016, doi: https://doi.org/10.1038/sdata.2016.18.

[2] R. Wittner et al., "Towards a common standard for data and specimen provenance in life sciences," *Learning Health Systems,* e10365, Apr. 2023, doi: https://doi.org/10.1002/lrh2.10365.
[3] E. H. Fegraus, S. Andelman, M. B. Jones and M. Schildhauer, "Maximizing the Value of Ecological Data with Structured Metadata: An Introduction to Ecological Metadata Language (EML) and Principles for Metadata Creation," *Bulletin of the Ecological Society of America*, vol. 86, no. 3, pp. 158-168, Jul. 2005, doi: https://www.jstor.org/stable/bullecosociamer.86.3.158
[4] M. B. Jones et al., "Ecological Metadata Language version 2.2.0," KNB Data Repository, 2019, doi:10.5063/F11834T2
[5] National Center for Ecological Analysis & Synthesis, "National Center for Ecological Analysis & Synthesis (NCEAS) - Home Page," Jun. 24, 2024. [Online]. Available: https://nceas.ucsb.edu/
[6] Long Term Educational Research Network, "HOME - LTER," Jun. 24, 2024. [Online]. Available: https://lternet.edu/
[7] Ocean Biodiversity Information System, "Ocean Biodiversity Information System," Jun. 24, 2024. [Online]. Available: https://obis.org/
[8] K. Vanderbilt et al., 2022, "Publishing Ecological Data in a Repository: An Easy Workflow for Everyone," *Bulletin of the Ecological Society of America,* vol. 103, no. 4, Aug. 2022, doi: https://doi.org/10.1002/bes2.2018
[9] K. H. Watanabe, S. W. Dietrich, M. B. Zelinski and J. P. Sluka, Nov. 2023. "MOTHER XSD, version 1.2," Accessed: Dec. 21, 2023. Available: https://resources.mother-db.org/xml/1.2/mdb.xsd
[10] Tsui, E. L. et al., 2023, "Creating a common language for the subanatomy of the ovary", *Biology of Reproduction*, vol. 108, pp. 1-4, doi:10.1093/biolre/ioac199.
[11] K. E. O'Neill et al., "Anatomic nomenclature and 3-dimensional regional model of the human ovary: call for a new paradigm," *Am J Obstet Gynecol*, Mar. 2023, vol. 228, no. 3, pp. 270-275, doi: 10.1016/j.ajog.2022.09.040.
[12] P. Walmsley, *Definitive XML Schema*, 2nd. ed, Prentice Hall, 2012.
[13] K. H. Watanabe, S. W. Dietrich, M. B. Zelinski and J. P. Sluka, "Multispecies Ovary Tissue Histology Electronic Repository," Dec. 21, 2023. [Online]. Available: https://mother-db.org
[14] Center for Open Science, "Open Science Framework (OSF)," *J Med Libr Assoc*., vol. 105, no. 2, Apr. 2017, pp. 203-206, doi: 10.5195/jmla.2017.88
[15] W. Ma, "Computational Approaches for the Multispecies Ovary Tissue Histology Electronic Repository (MOTHER) Curation Pipeline, Master's of Biological Data Science Applied Project Report, ASU, Phoenix, AZ, May 2022.
[16] S. W. Dietrich, *Understanding Databases: Concepts and Practice*, Wiley, 2021.
[17] Y. Ding, "Automating the MOTHER Computational Curation Pipeline and Classifying Ovarian Follicles using Machine Learning," Master's of Biological Data Science Applied Project Report, ASU, Phoenix, AZ, May 2023.

Suzanne W. Dietrich received the Ph.D. degree in computer science from Stony Brook University, Stony Brook, NY, in 1987.

She is a Professor in the School of Mathematical and Natural Sciences, Arizona State University, Phoenix, AZ. She is the author of several database textbooks, including *Understanding Databases: Concepts and Practice* (Wiley, 2021). Her current research interests include interdisciplinary applications of databases and computer science education.

Dr. Dietrich is recognized by the ACM as a Distinguished Educator for her impact and significant accomplishments in the field of computer science education.

Wenli Ma received the M.S. degree in biological data science from Arizona State University in 2022. She is currently a Research Scientist at the Center of Translational Science, Florida International University. Her current research interests include translational bioinformatics on human diseases.

Yian Ding received the M.S. degree in biological data science from Arizona State University in 2023. Her Master's project (MOTHER) focused on leveraging programming languages to develop a curation pipeline and identify different types of follicles, thus advancing research at the intersection of these disciplines. She is enthusiastic about applying her knowledge to tackle challenging problems in biology and making a meaningful impact in the scientific community.

Karen H. Watanabe received a Ph.D. degree in Mechanical Engineering from the University of California, Berkeley, CA.

Her current research focuses on the development of mathematical/computational models of how chemicals affect living organisms primarily with respect to reproductive effects.

Mary B. Zelinski received the M.S. and Ph.D. degrees in Animal Science from Oregon State University, Corvallis, OR. Her current research focuses on preserving fertility in female cancer survivors using ovarian tissue cryopreservation and transplantation as well as in vitro follicle maturation. She recently received a Senior Scholar Award from the Global Consortium on Reproductive Longevity and Equality to investigate interventions for ovarian aging in macaques.

She has received continuous funding from the National Institutes of Health since 1998. She has been an invited speaker at many national and international meetings, has numerous manuscripts in peer-reviewed journals, and served as Secretary, Program Chair and on the Board of Directors of the Society for the Study of Reproduction from which she received the Distinguished Service Award. She is currently a co-Editor-in-Chief of Biology of Reproduction. She is passionate about bringing science to the public wherein she directs and participates in many educational outreach activities for adults, high school, and middle school students.

James P. Sluka received his Ph.D. in chemistry from California Institute of Technology, Pasadena California, USA in 1988.

He worked as Pharmaceutical Chemist for Eli Lilly and Company for ten years. In 2001 he started his own bioinformatics company, InPharmix, which provided text mining software to the pharmaceutical industry and the US government. In 2010 he joined the Biocomplexity Institute at Indiana University in Bloomington Indiana as a Senior Scientist. His research interests include computational modeling of pharmaceutical and toxicological processes in humans and animals and artificial intelligence and machine leaning applications in biological modeling.

Dr. Sluka is an active member of the International Organization of Standards (ISO) where he contributes to standards in biotechnology and biomedicine. He is an award winning author in toxicological modeling and is a member of the Society of Toxicology.