Mercury (Hg) Research Ontology: Employing Informatics in Geochemistry

Melinda Neville

Department of Ecology and Environmental Sciences, University of Maine, Orono ME, USA

ONTOLOGIES facilitate disparate data integration, dissemination and comparison for complex fields of study. The **Hg Research Ontology** developed in this project will place Maine legacy Hg data in the context of the Hg biogeochemical cycle and linked to contextual characteristics of the observation settings.

Introduction

THE mercury (Hg) monitoring legacy in the Northeastern United States spans three decades and many different ecosystem pools, but the diverse nature of Hg data makes comparison across studies difficult. The ability to measure whole ecosystem changes in Hg contamination in response to policies regulating Hg is often limited, and researchers have difficulties comparing results or utilizing the wealth of legacy Hg data more broadly. We are developing an informatics approach to facilitate Hg data reuse and modeling.

Objectives

THIS research aims to:

- Develop an ontology-based information system for Hg biogeochemistry and legacy research data integration.
- Evaluate and validate the competency of the ontology design with other researchers.
- Share with stakeholders to stimulate collaboration, information dissemination, and data integration.

The Data Problem

THE complexity of Hg biogeochemistry generates research methods and results that span academic disciplines, different spatial and temporal scales, orders of magnitude difference in concentration, and many varied landscape and climate contexts. In order for a cohesive picture of Hg fate and transport to emerge, methods are needed to make data comparable across projects. The Hg cycle (simplified in Figure 1) is inferred by many different sample media and models, as shown in the blue boxes.

THE research cycle (Fig 2) can be integrated with data management strategies that include metadata standards and ontologies. An ontology formally specifies representation of entities along with their properties and relations, and defines a common vocabulary that can be shared between researchers. Once formally defined using a computer-readable language, basic concepts and relations among them are searchable, reusable, aid new data additions to the existing knowledge framework, and support broader data access.

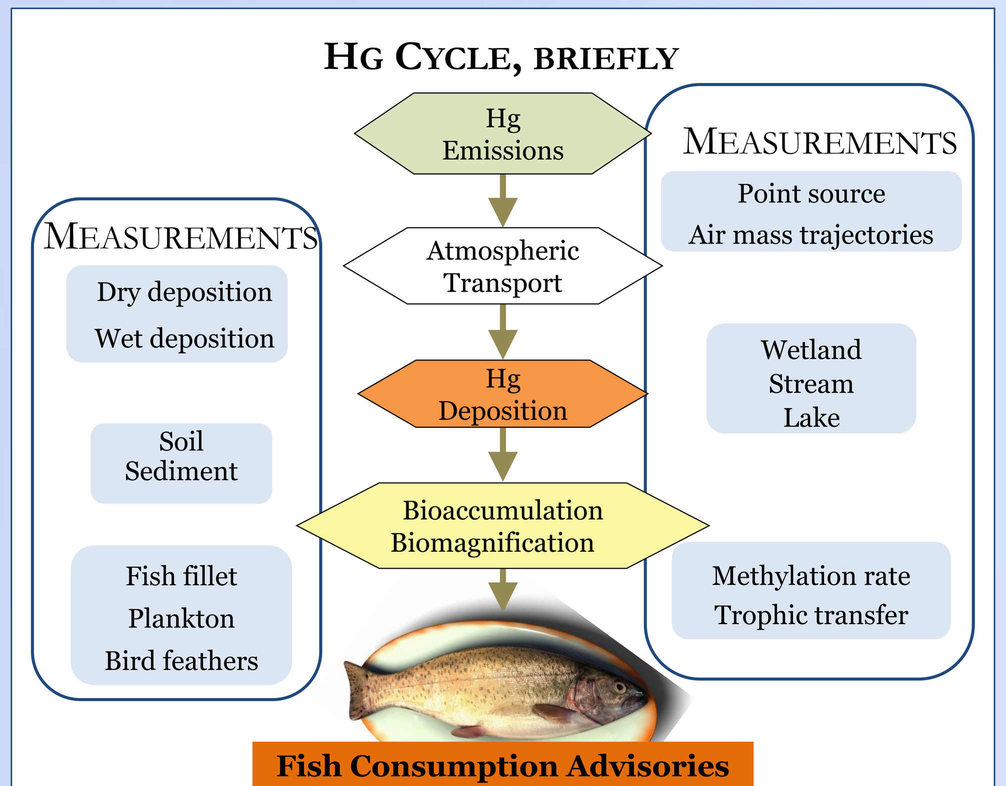

Fig 1. Example of the Hg cycle and some sample media and models used to discern fate and transport.

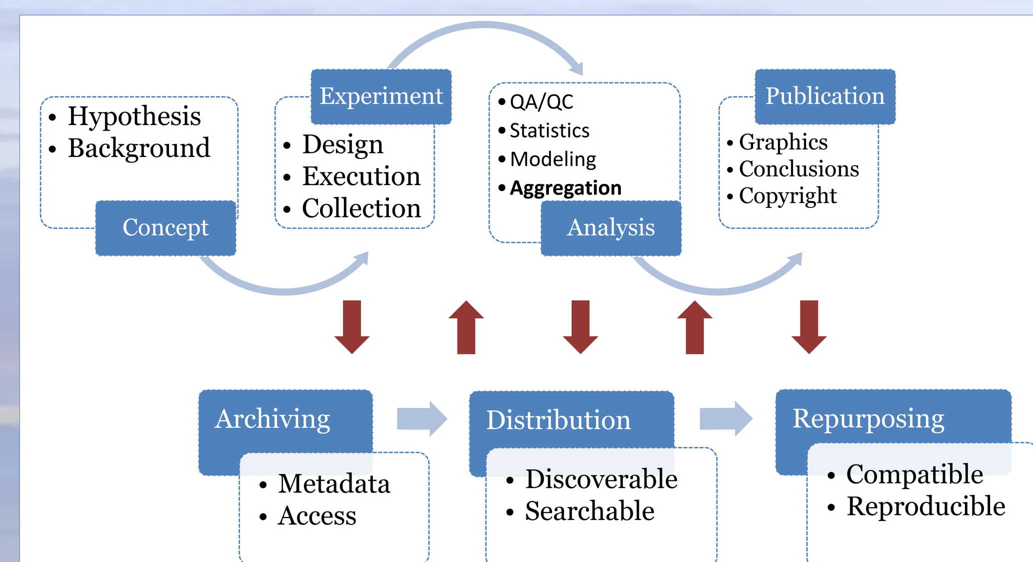

Fig 2. Data reusability necessitates planning at all stages of the research cycle.

Role of an Ontology

Using a programming language like OWL (Web Ontology Language) and an ontology editor (protege.stanford.edu), the ontology takes shape with specification of basic concepts. We start with specification of recognized entities within the Hg research:

> **Biomagnification** is a process, is cumulative, occurs in biota, has a rate, and is controlled by bioaccumulation.
>
> **Stream Hg sample** is an observation, has a measured concentration, is located within a catchment, and has a specific timestamp.

Specification of these entities, while simplified, provide a framework that data observations on Hg can be associated. Many of the relationships between measurements, natural phenomena and the inferences within Hg biogeochemical research are implicit (Fig 3), but an ontological framework makes such relationships explicit (Fig 4).

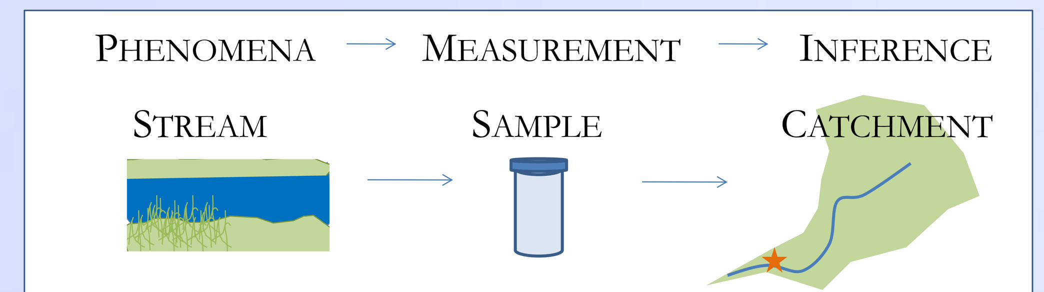

Fig 3. Simple measurements commonly have complex spatiotemporal contexts.

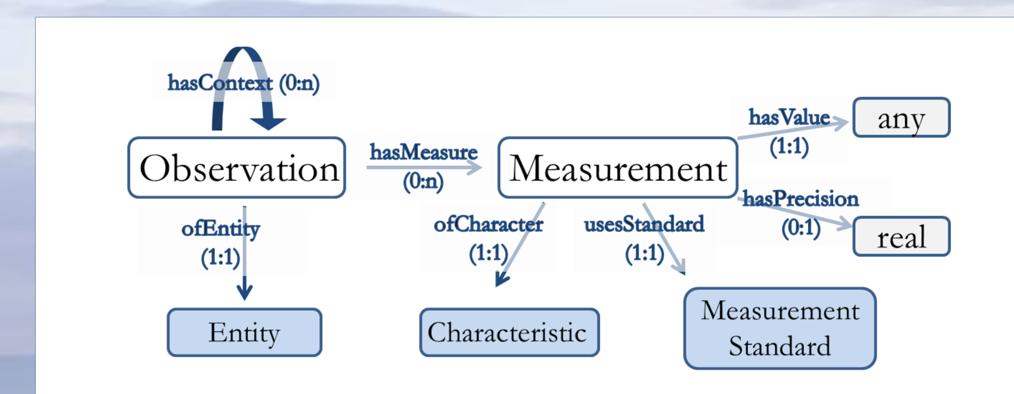

Fig 4. An ontology of entities and their relationships can explicitly define spatiotemporal contexts. (*Adapted from Madin et al 2007. An ontology for describing and synthesizing ecological observation data. Ecological Informatics 2(3):279-296*)

Development

THE design of an ontology is based on a series of competency questions that the ontology must be able to answer upon its completion. Some of the informal competency questions that we are using to test the **Hg Research Ontology** are:

- **Temporal**: Given *x* data, captured at *t* time, what process, event, or state in the Hg cycle can this observation attach to?
- **Spatial**: Given *x* data, collected at *s* location using *z* protocol, what spatial contexts in the Hg cycle can this observation attach to?
- **Monitoring**: "What *y* events (regulatory, climatic) affect *x* measurements?"
- **Efficiency**: "What *x* measurement maximizes our knowledge of change in *y*?"

Outcomes

AS this research progresses, the **Hg Research Ontology** will support several key activities and goals, including:

- Stakeholders, researchers, policy makers, and non-profits (lake associations, schools) will have an information 'blueprint' to support data collection and reporting. The ontology will incorporate data sourcing, thus citizen science can be cataloged and compared with more traditional Hg data sources.
- The use of common terms and controlled vocabularies by researchers and stakeholders will assist in identifying data gaps, results of specific regulations, and potential areas of concern that warrant further data collection and analysis to inform decision making.

Acknowledgements: THIS work has been generously funded by the National Science Foundation IGERT Program under Grant 054494 and the Maine Water Resources Research Institute Grants Program.