The relational (table) data model is best-known among illustrators how these terms may be packaged and shared in an Darwin Core with additional input from a working group of over disciplines. Here, we present a draft vocabulary of metadata standards for primatologists and other behavior-oriented researchers. The Ethoinformatics project is a collaborative effort to create data standards in the terminologies, structures, and technologies used on data from various researchers and research groups.

Introduction

Primatology is inherently comparative and therefore dependent on data from various researchers and research groups. Comparison, however, is often hampered by a lack of formal standards. Here, we present a draft vocabulary of metadata standards for primatologists and other behavior-oriented researchers. The Ethoinformatics project is a collaborative effort to create data standards in the terminologies, structures, and technologies used on data from various researchers and research groups.

Data Models

The relational (table) data model is best-known among primatologists and other behavior-oriented disciplines. Here, we present a draft vocabulary of metadata classes and associated terms for this community. The vocabulary, which we call EthoCore, builds on related initiatives such as Darwin Core with additional input from a working group of over 50 primatologists from diverse areas of primatology. We also present an abstract data model, which we call EthoGrammar, that illustrates how these terms may be packaged and shared in an open and consistent manner.

EthoCore

A standard vocabulary for primatology and behavior, modeled on Darwin Core

EthoCore terms encompass classes and properties of information. EthoCore is not a comprehensive dictionary of the vernacular used by primatologists.

These terms can be flexibly used in various contexts including as columns in a table, keys in a document-based or key-value database, and predicates in a linked data triplestore.

The EthoCore vocabulary currently comprises 166 terms in 11 classes:

- Observation
- MaterialSample
- Event
- Activity
- Location
- Identification
- Taxon
- ResourceRelationship

The EthoCore documentation can be accessed at http://ethoinformatics.org/ethocore

Design Principles

1. Complex data structures can be deconstructed into atomic elements such as key-value pairs.

2. Terms in the vocabulary should be designed for packaged data structures that are as generalized as possible.

The Event and Location classes in Darwin Core, for instance, are useful for describing time and space in various contexts while the new Observation class in EthoCore describes observational information in a generic manner.

Design Principles, continued

3. Relationships among resources (e.g., tables or documents) can themselves be represented as resources.

4. A document-based data design holds considerable appeal for work in primatology. Apart from desirable features for mobile-and cloud-computing, it is a promising intermediary from table-based (relational) data designs to the global linked data ecosystem described using vocabularies such as EthoCore.

Discussion

We have found through this project that the ubiquity of tables in primatology is a major obstacle to comparison. As a community, it is important to break free of our “table-thinking” and to focus instead on commonalities in our data models.

New tools are already enabling better software for various applications by veering away from the relational model. These “NoSQL” technologies have particular advantages for mobile and cloud applications that are increasingly salient for primatologists. Transitioning to these technologies also affords an auspicious opportunity to incorporate standard terms and shared constraints into our research data. As such, we strongly encourage colleagues to explore these novel tools for their work.

Acknowledgments

We thank Jane Phillips-Conroy, Jennifer Moore, Cynthia Hudson, Aaron Addison, Denné Reed, and the Ethoinformatics Working Group. This work is supported by NSF SMA 1338524, 1338467, and 1338452.

Kenneth L. Chiou, Anthony Di Fiore, Robyn Overstreet, Mike Chevett, Tom Igoe

Department of Anthropology, Washington University in St. Louis, Department of Anthropology, University of Texas at Austin, Interactive Telecommunications Program, New York University, Freelance software developer.