

CERTIFICATION PAGE

Certification for Authorized Organizational Representative or Individual Applicant:

By signing and submitting this proposal, the Authorized Organizational Representative or Individual Applicant is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding debarment and suspension, drug-free workplace, and lobbying activities (see below), nondiscrimination, and flood hazard insurance (when applicable) as set forth in the NSF Proposal & Award Policies & Procedures Guide, Part I: the Grant Proposal Guide (GPG) (NSF 09-29). Willful provision of false information in this application and its supporting documents or in reports required under an ensuing award is a criminal offense (U. S. Code, Title 18, Section 1001).

Conflict of Interest Certification

In addition, if the applicant institution employs more than fifty persons, by electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative of the applicant institution is certifying that the institution has implemented a written and enforced conflict of interest policy that is consistent with the provisions of the NSF Proposal & Award Policies & Procedures Guide, Part II, Award & Administration Guide (AAG) Chapter IV.A; that to the best of his/her knowledge, all financial disclosures required by that conflict of interest policy have been made; and that all identified conflicts of interest will have been satisfactorily managed, reduced or eliminated prior to the institution's expenditure of any funds under the award, in accordance with the institution's conflict of interest policy. Conflicts which cannot be satisfactorily managed, reduced or eliminated must be disclosed to NSF.

Drug Free Work Place Certification

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Drug Free Work Place Certification contained in Exhibit II-3 of the Grant Proposal Guide.

Debarment and Suspension Certification

(If answer "yes", please provide explanation.)

Is the organization or its principals presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded from covered transactions by any Federal department or agency?

Yes

No

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Debarment and Suspension Certification contained in Exhibit II-4 of the Grant Proposal Guide.

Certification Regarding Lobbying

The following certification is required for an award of a Federal contract, grant, or cooperative agreement exceeding \$100,000 and for an award of a Federal loan or a commitment providing for the United States to insure or guarantee a loan exceeding \$150,000.

Certification for Contracts, Grants, Loans and Cooperative Agreements

The undersigned certifies, to the best of his or her knowledge and belief, that:

- (1) No federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.
- (2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, "Disclosure of Lobbying Activities," in accordance with its instructions.
- (3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements and that all subrecipients shall certify and disclose accordingly.

This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, Title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than \$10,000 and not more than \$100,000 for each such failure.

Certification Regarding Nondiscrimination

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative is providing the Certification Regarding Nondiscrimination contained in Exhibit II-6 of the Grant Proposal Guide.

Certification Regarding Flood Hazard Insurance

Two sections of the National Flood Insurance Act of 1968 (42 USC §4012a and §4106) bar Federal agencies from giving financial assistance for acquisition or construction purposes in any area identified by the Federal Emergency Management Agency (FEMA) as having special flood hazards unless the:

- (1) community in which that area is located participates in the national flood insurance program; and
- (2) building (and any related equipment) is covered by adequate flood insurance.

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant located in FEMA-designated special flood hazard areas is certifying that adequate flood insurance has been or will be obtained in the following situations:

- (1) for NSF grants for the construction of a building or facility, regardless of the dollar amount of the grant; and
- (2) for other NSF Grants when more than \$25,000 has been budgeted in the proposal for repair, alteration or improvement (construction) of a building or facility.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE		SIGNATURE	DATE
NAME Laura J Jenski		Electronic Signature	Jul 23 2009 4:54PM
TELEPHONE NUMBER 605-677-5370	ELECTRONIC MAIL ADDRESS Laura.Jenski@usd.edu	FAX NUMBER 605-677-6387	

* EAGER - EARly-concept Grants for Exploratory Research

** RAPID - Grants for Rapid Response Research

**Directorate for Biological Sciences
Division of Biological Infrastructure
Research Coord Netwk in Bio Sc**

**Proposal Classification Form
PI: Mabee, Paula / Proposal Number: 0956049**

CATEGORY I: INVESTIGATOR STATUS (Select ONE)

- Beginning Investigator - No previous Federal support as PI or Co-PI, excluding fellowships, dissertations, planning grants, etc.
- Prior Federal support only
- Current Federal support only
- Current & prior Federal support

CATEGORY II: FIELDS OF SCIENCE OTHER THAN BIOLOGY INVOLVED IN THIS RESEARCH (Select 1 to 3)

- | | | |
|--|--------------------------------------|--|
| <input type="checkbox"/> Astronomy | <input type="checkbox"/> Engineering | <input type="checkbox"/> Psychology |
| <input type="checkbox"/> Chemistry | <input type="checkbox"/> Mathematics | <input type="checkbox"/> Social Sciences |
| <input checked="" type="checkbox"/> Computer Science | <input type="checkbox"/> Physics | <input type="checkbox"/> None of the Above |
| <input type="checkbox"/> Earth Science | | |

CATEGORY III: SUBSTANTIVE AREA (Select 1 to 4)

- | | | |
|--|---|---|
| <input type="checkbox"/> BIOMATERIALS
<input type="checkbox"/> BIOTECHNOLOGY
<input type="checkbox"/> Animal Biotechnology
<input type="checkbox"/> Plant Biotechnology
<input type="checkbox"/> Environmental Biotechnology
<input type="checkbox"/> Marine Biotechnology
<input type="checkbox"/> Metabolic Engineering
<input type="checkbox"/> CHROMOSOME STUDIES
<input type="checkbox"/> COMMUNITY ECOLOGY
<input type="checkbox"/> COMPUTATIONAL BIOLOGY
<input type="checkbox"/> CONSERVATION & RESTORATION BIOLOGY
<input type="checkbox"/> CORAL REEFS
<input checked="" type="checkbox"/> CURATION
<input checked="" type="checkbox"/> DATABASES
<input type="checkbox"/> ECOSYSTEMS LEVEL
<input type="checkbox"/> GENOMICS (Genome sequence, organization, function) | <input type="checkbox"/> Viral
<input type="checkbox"/> Microbial
<input type="checkbox"/> Fungal
<input type="checkbox"/> Plant
<input type="checkbox"/> Animal
<input checked="" type="checkbox"/> INFORMATICS
<input type="checkbox"/> MARINE MAMMALS
<input type="checkbox"/> Molecular Evolution
<input type="checkbox"/> Methodology/Theory
<input type="checkbox"/> Gene/Genome Mapping
<input type="checkbox"/> Natural Products
<input type="checkbox"/> NANOSCIENCE
<input type="checkbox"/> PHOTOSYNTHESIS
<input type="checkbox"/> PLANT BIOLOGY
<input type="checkbox"/> Arabidopsis-Related Plant Research
<input type="checkbox"/> POPULATION DYNAMICS & LIFE HISTORY | <input type="checkbox"/> POPULATION GENETICS & BREEDING SYSTEMS
<input type="checkbox"/> REPRODUCTIVE ANIMAL BIOLOGY
<input type="checkbox"/> Plant Pathology
<input type="checkbox"/> Coevolution
<input type="checkbox"/> Biological Control
<input type="checkbox"/> STATISTICS & MODELING
<input type="checkbox"/> Methods/ Instrumentation/ Software
<input type="checkbox"/> Modeling (general)
<input type="checkbox"/> Modeling of Biological or Molecular Systems
<input type="checkbox"/> Computational Modeling
<input type="checkbox"/> Statistics (general)
<input type="checkbox"/> STRUCTURAL BIOLOGY
<input checked="" type="checkbox"/> SYSTEMATICS
<input type="checkbox"/> Phenetics/Cladistics/ Numerical Taxonomy
<input type="checkbox"/> NONE OF THE ABOVE |
|--|---|---|

CATEGORY IV: INFRASTRUCTURE (Select 1 to 3)

- | | | |
|--|--|---|
| <input type="checkbox"/> COLLECTIONS/STOCK CULTURES
<input type="checkbox"/> Collection Enhancement
<input type="checkbox"/> Collection Refurbishment
<input type="checkbox"/> Living Organism Stock Cultures
<input type="checkbox"/> Natural History Collections
<input type="checkbox"/> DATABASES
<input type="checkbox"/> Database Initiation | <input type="checkbox"/> Database Enhancement
<input type="checkbox"/> Database Maintenance & Curation
<input checked="" type="checkbox"/> Database Methods
<input type="checkbox"/> FACILITIES
<input type="checkbox"/> Controlled Environment Facilities
<input type="checkbox"/> Field Stations
<input type="checkbox"/> Field Facility Structure | <input type="checkbox"/> Field Facility Equipment
<input type="checkbox"/> LTER Site
<input type="checkbox"/> GENOME SEQUENCING
<input type="checkbox"/> Other Plant Genome Sequencing
<input type="checkbox"/> INDUSTRY PARTICIPATION
<input type="checkbox"/> INSTRUMENTATION
<input type="checkbox"/> Instrument Development |
|--|--|---|

<input type="checkbox"/> Instrument Acquisition <input type="checkbox"/> Computational Hardware Development/Acquisition <input type="checkbox"/> TOOLS DEVELOPMENT <input type="checkbox"/> Analytical Algorithm Development <input type="checkbox"/> Other Software Development <input checked="" type="checkbox"/> Informatics Tool Development	<input type="checkbox"/> Technique Development TRACKING SYSTEMS <input type="checkbox"/> Geographic Information Systems <input type="checkbox"/> Remote Sensing <input type="checkbox"/> TRAINING	<input type="checkbox"/> Multi-, Cross-, Interdisciplinary Training <input type="checkbox"/> Undergraduate Training <input type="checkbox"/> Predoctoral Training <input type="checkbox"/> Postdoctoral Training <input type="checkbox"/> NONE OF THE ABOVE
--	---	---

CATEGORY V: HABITAT (No selection required)

CATEGORY VI: GEOGRAPHIC AREA OF THE RESEARCH (No selection required)

CATEGORY VII: CLASSIFICATION OF ORGANISMS (Select 1 to 4)

<input type="checkbox"/> VIRUSES <input type="checkbox"/> Bacterial <input type="checkbox"/> Plant <input type="checkbox"/> Animal <input type="checkbox"/> PROKARYOTES <input type="checkbox"/> Archaeobacteria <input type="checkbox"/> Cyanobacteria <input type="checkbox"/> Eubacteria <input type="checkbox"/> PROTISTA (PROTOZOA) <input type="checkbox"/> FUNGI <input type="checkbox"/> LICHENS <input type="checkbox"/> SLIME MOLDS <input type="checkbox"/> ALGAE	<input checked="" type="checkbox"/> PLANTS <input type="checkbox"/> NON-VASCULAR PLANTS <input type="checkbox"/> VASCULAR PLANTS <input type="checkbox"/> GYMNOSPERMS <input type="checkbox"/> ANGIOSPERMS <input type="checkbox"/> Monocots <input type="checkbox"/> Dicots <input checked="" type="checkbox"/> ANIMALS <input type="checkbox"/> INVERTEBRATES <input type="checkbox"/> ARTHROPODA <input type="checkbox"/> Hexapoda (Insecta) (Insects) <input type="checkbox"/> VERTEBRATES <input type="checkbox"/> FISHES	<input type="checkbox"/> Chondrichthyes (Cartilaginous Fishes) (Sharks, Rays, Ratfish) <input type="checkbox"/> Osteichthyes (Bony Fishes) <input type="checkbox"/> AMPHIBIA <input type="checkbox"/> REPTILIA <input type="checkbox"/> AVES (Birds) <input type="checkbox"/> MAMMALIA <input type="checkbox"/> Primates <input type="checkbox"/> Humans <input type="checkbox"/> Rodentia <input type="checkbox"/> Marine Mammals (Seals, Walrus, Whales, Otters, Dolphins, Porpoises) <input type="checkbox"/> TRANSGENIC ORGANISMS <input type="checkbox"/> NO ORGANISMS
--	--	--

CATEGORY VIII: MODEL ORGANISM (Select ONE)

<input type="checkbox"/> NO MODEL ORGANISM MODEL ORGANISM (Choose from the list or input up to 9 characters) <input type="checkbox"/> Escherichia coli	FUNGAL PLANT <input type="checkbox"/> Mouse-Ear Cress (Arabidopsis thaliana)	<input checked="" type="checkbox"/> Fruitfly (Drosophila melanogaster) [Enter your own model organism - up to 9 characters] <input type="text"/>
--	--	--

PROJECT SUMMARY

We propose to establish a network of scientists who are developing phenotype ontologies and to use this network to enable and enhance the research of all of those involved. The goal is to create useful and scientifically accurate phenotype ontologies for general use in a variety of different types of biological investigations. Through a staged series of meetings, working groups and exchanges, we seek to put together the higher-level scaffolding of anatomy ontologies for three major taxonomic clades, to establish and publish a manual of standards and best practices for the collection, management, presentation and retrieval of these data, and to train and educate scientists in the methods for establishing ontologies across a broad range of phenotype data and using them in their research. Although there are over 1,000 databases for genomic, genetic and associated molecular and pathway data (Galperin and Cochrane 2009), many/most of which are supported by ontologies, there are few databases or ontologies developed for phenotypic data, though it is central to virtually all areas of biological inquiry. Genes, gene products, signaling networks, cell behavior and development are involved in producing the organism's phenotype – and yet establishing the infrastructure to support computational research into the relation of these to phenotypes is only in the earliest stages. Environment, ecology, and evolution also impact an organism's phenotype, and existing databases in these fields similarly require computation on phenotype to advance their research. The phenotype of an organism, which is broadly defined as any observable feature of an organism, is the result of genes and environment. Only by developing ontologies for phenotype and using them in databases, can 'systems level' computation and data mining be truly enabled in biological (and biomedical) fields.

We will bring together ontology developers and logicians with scientists from two different research backgrounds: (1) comparative evolutionary biologists, who focus on the variation of phenotypes across species, and (2) model organism biologists, who study the phenotype of a single 'model' species from the perspective of molecular biology, genetics, genomics and development. Individuals from these two distinct backgrounds share the research requirement to compare phenotype across species, and thus need to standardize methods of ontology development and alignment to enable interoperability across databases. Our primary phenotype focus will be on anatomy because these data comprise the most significant portion of the phenotype data collected in both evolutionary and model organism communities, and consequently this is where the nascent phenotype ontology efforts exist. Coordinating these efforts will eliminate redundancy, solidify standard methods and practices, and enable rapid progress toward addressing research needs.

Our specific goals are to: (1) Develop anatomical reference ontologies for three key taxonomic clades; (2) Align and synchronize anatomical ontologies using homology and various types of similarity relations; (3) Define, test and document anatomy ontology development best practices and standards; (4) Reach out to ancillary phenotype groups to share with them common concepts and practices; (5) Educate the community about the methods for developing ontologies and their importance and utility in research.

This work will result in a community of graduate students, postdocs, and scientists from diverse scientific and organizational backgrounds who will understand the benefits of database interoperability that result from representing data with ontologies. The expert knowledge that will be embodied in the coordinated development of reference anatomy ontologies is a lasting resource of central value to the many biological disciplines that seek to relate their data to an organism's phenotype.

TABLE OF CONTENTS

For font size and page formatting specifications, see GPG section II.B.2.

	Total No. of Pages	Page No.* (Optional)*
Cover Sheet for Proposal to the National Science Foundation		
Project Summary (not to exceed 1 page)	1	_____
Table of Contents	1	_____
Project Description (Including Results from Prior NSF Support) (not to exceed 15 pages) (Exceed only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)	15	_____
References Cited	3	_____
Biographical Sketches (Not to exceed 2 pages each)	10	_____
Budget (Plus up to 3 pages of budget justification)	8	_____
Current and Pending Support	7	_____
Facilities, Equipment and Other Resources	1	_____
Special Information/Supplementary Documentation	0	_____
Appendix (List below.) (Include only if allowed by a specific program announcement/ solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)	_____	_____
Appendix Items:		

*Proposers may select any numbering mechanism for the proposal. The entire proposal however, must be paginated. Complete both columns only if the proposal is numbered consecutively.

PROJECT DESCRIPTION

1. Establishing a Phenotype Ontology Coordination Network

Justification: We propose to establish a network of scientists that will enable and enhance the research of all of those involved as well as scientists in the wider research community. Using this network these scientists will create shared, standardized and scientifically accurate phenotype ontologies¹ for use in a broad variety of biological investigations. Through a staged series of meetings, working groups and exchanges, we seek to put together the high-level scaffolding of anatomy ontologies for three major taxonomic clades, to establish and publish a manual of standards and best practices for the collection, management, presentation and retrieval of data described via these ontologies, and to educate scientists in the methods for establishing ontologies across a broad range of phenotype data and exploiting them in their research. Although there are over 1,000 databases for genomic, genetic and associated molecular and pathway data (Galperin and Cochrane 2009), many of which are supported by ontologies, there are few databases, or ontologies, developed for well-structured phenotypic data, though it is central to virtually all areas of biological inquiry. Genes, gene products, signaling networks, cell behavior and development are involved in producing the organism's phenotype—and yet establishing the infrastructure to support computational research that can relate these to phenotypes is only in the earliest stages. Environment, ecology, and evolution also impact an organism's phenotype, and existing databases in these fields similarly require computation on phenotype to advance their research. The phenotype of an organism, broadly defined as any observable feature of an organism, is the result of genes and environment. Only by developing ontologies for phenotype and using them in databases, can 'systems level' computation and data mining be truly enabled in biological (and biomedical) fields.

We will bring together ontology developers with scientists from two different research backgrounds: (1) comparative evolutionary biologists, who focus on the variation of phenotypes across species, and (2) model organism biologists, who study the phenotype of a single 'model' species from the perspective of molecular biology, genetics, genomics and development. Individuals from these two distinct backgrounds share research requirement to compare phenotype across species, and thus need to standardize methods of ontology development and alignment to enable interoperability across databases. For historical reasons these scientists employ different terminologies and traditions of research. To achieve the necessary integration of results there is thus a need for them to meet face-to-face over several years in order to learn how to communicate effectively (to understand the respective vocabularies used in the different research areas) and thereby to establish an agreed method for reasoning across phenotype. Our primary phenotype focus will be on anatomy. Anatomical data (often synonymous with 'morphological data') comprise the most significant portion of the phenotype data collected in both evolutionary and model organism communities, and consequently this is where the nascent phenotype ontology efforts exist. Coordinating these efforts will eliminate redundancy, solidify standard methods and practices, and enable rapid progress toward addressing research needs.

Our specific goals are to:

- 1) Develop anatomical reference ontologies for three key taxonomic clades;
- 2) Align species-specific anatomical ontologies using homology and other similarity relationships;
- 3) Define, test and document anatomy ontology development best practices and standards;
- 4) Reach out to ancillary phenotype groups to share with them common concepts and practices
- 5) Educate the community in methods for developing ontologies and their utility in research

¹ *i.e.* structured vocabularies with logical relationships that can be computed on.

Uniting this unique community of diverse researchers bound by the goals of understanding the developmental and evolutionary bases of phenotype will result in novel synergies and networks of seemingly unrelated research paradigms. Just as the Gene Ontology (GO) Consortium (The Gene Ontology Consortium 2000, 2001, 2008) was formed to develop the shared structured vocabularies adequate for annotating molecular characteristics across organisms, an analogous Phenotype Ontology Network is required to coordinate and develop the ontologies critical for annotating phenotypic characteristics across species, with appropriate standards of logic and definitions needed for effective computation, and with the scientific standards of accuracy appropriate for evolutionary phenomena.

2. Introduction and background

Two of the broadest, most significant and most challenging research agendas in biology involve understanding the development and evolution of the phenotype. What are the genetic, genomic, and developmental factors underlying the development of an organism's phenotype? What are the patterns of phenotypic change across the evolution of species, and what are their genetic, genomic, and developmental bases? Though these outstanding research questions are being addressed by two disparate communities of scientists, each community requires that phenotypic data be integrated with other data types for research progress. Computing on phenotypic data, however, is a current challenge in bioinformatics, and basic methods of computational representation and standards are not well developed, consistently applied, or broadly understood. Querying phenotypes across databases is difficult or impossible because of the lack of terminological alignment, a lack which needs to be addressed by reference ontologies for the phenotype domain, which can combine high scientific accuracy with the sort of logical structure needed for computational integration of immense quantities of diverse data.

We propose to establish and maintain a communication network among evolutionary and model organism biologists, thus creating a new community where the development of phenotype ontologies is coordinated and methods to align and reason across them are standard and consistent. We propose to achieve this through a staged series of annual summit meetings, taxonomically focused working groups, an annual summer training course, outreach to ancillary phenotype groups, and an exchange program. The motivation behind this RCN proposal is very strong because if we can achieve interoperability, then there is tangible benefit in the form of shared access to the unique data store of each of these communities (e.g. genetic, genomic, developmental, and evolutionary data). The contributions and benefits of participating in this RCN for the two disciplines are as follows.

Evolutionary biology: Decades of comparative work in evolutionary biology have produced a rich body of data on the natural diversity of phenotypes. However, the non-computable free-text format in which these data are published poses a challenge to the use of these data outside the narrow scope of the original study in which they were collected (Mabee et al. 2007a; Mabee et al. 2007b; Dahdul et al. in press). Even a seemingly simple task, such as compiling a list of species that share a similar anatomical structure, requires substantial human time investment and expert knowledge of the literature. Large-scale analyses of the patterns of phenotypic evolution across multiple taxa are simply not tractable. Testing, for example, whether sets of anatomical characters vary in statistically similar ways across clades, or whether there is differential enrichment in the types of qualities (e.g., shape or size) among sets of anatomical features, can only be done if phenotype is represented in a format amenable to computation. The current format of these data not only precludes expert analysis for a 'big picture' view, but it also means that the data cannot be integrated with other data types. Moreover, they are inaccessible to scientists outside of the domain for use in interdisciplinary studies. These obstacles have been independently realized by the many investigators participating in the NSF AToL (Tree of Life) efforts (<http://atol.sdsc.edu/projects>) as they have aggregated the comparative phenotype data from their respective taxonomic groups. The AToL

participants contacted thus far are very interested in participating in this RCN (Table 2) because of the research potential that phenotype ontologies offer. Three of the NSF-funded programs of PI and coPIs on this proposal (Mabee, Phenoscope; Maglia, Amphibanat; Deans, Hymenoptera Anatomy Ontology) are focused on representing comparative evolutionary phenotypic data using ontologies in coordination with corresponding model organism biologists and informaticians. Many more of these efforts are needed if large-scale integration and comparisons of phenotype data are to be realized.

Model organism biology: Model organism biologists—*i.e.* experts studying zebrafish, *Drosophila*, *Arabidopsis*, corn, rice, mouse, *et cetera*—focus on phenotype as a way of assessing the genetic and developmental mechanisms underlying it. Particular phenotypes may be associated with particular genetic changes, for example, and through further experiments a causal link may be made. The data from genomics, genetics, developmental studies and experimental assays are rapidly proliferating and simultaneously being captured in databases and synthesized. Both genetic and phenotypic data are represented in these databases using ontologies (Table 1), and MODs have been the leaders in using ontologies to represent phenotype. Genetic data (along with genomic and developmental data) are curated in model organism databases using gene ontologies and are frequently associated with particular phenotypes. Phenotypic mutants are curated, often in relation to particular genes and gene expression patterns. Anatomy ontologies that are specific to the model organism species have been developed for the purpose of annotating gene expression, mutant phenotypes and associated data, thus rendering them computable. These databases, with ontology-annotations, provide model organism communities with an extensive range of interconnected data, including links between genes and mutant phenotypes. Despite this, however, few of these anatomy ontologies are interoperable with each other because of the lack of reference ontologies and a lack of standards in how to compare annotated phenotype data.

Description of phenotypes, in a manner that is amenable to systematic retrieval and comparison, is a much more challenging problem than that presented for molecularly oriented ontologies, because the scope of phenotypes is inherently much larger. For example, in the case of the GO there is a single type of entity, the gene product, and these entities were already well defined by their proxies in the sequence databases when the GO project began. In contrast, to describe phenotypes we must deal with a multitude of entities, and, even by restricting our scope to anatomical entities, considerable effort is entailed to first accurately define these as no analog to the sequence databases exist. Analogously, for the GO, it was straightforward to restrict the qualities that could be assigned to the entities (gene products) to three aspects: what can this entity do (*i.e.* molecular function), in what context might it carry this function out (*i.e.* biological process), and where can it do it (*i.e.* cellular component). Whereas to describe the morphological or developmental qualities of an anatomical entity will require many more complex aspects, such as shape and the spatial relationships between anatomical entities. It is interesting to note that it is only in the past year that the GO Consortium has initiated a protein-family-based strategy, similar to that proposed here for the reference anatomies, to eliminate the species-specific silos of curation used heretofore (Gaudet et al. 2009). Despite the challenges phenotypes present, considerable progress has been made in developing the necessary ontologies and formalisms over the past several years and the stage is now set for our proposed network which will address the driving research needs of this interdisciplinary community.

3. Specific Aims of the Phenotype Ontology Coordination Network

3.1 Aim 1: Develop reference anatomy ontologies for key taxonomic clades

As noted earlier, key ontologies for anatomy are lacking and, given existing ontologies, it is currently not possible algorithmically, to find ‘matching’ anatomical structures across taxa, e.g. between fly and wasp. Nor is it possible to explore the nature of the matches (homology, common function), or to build a

list of candidate genes from all (model) taxa possessing that structure to better understand the genetic and developmental bases of its evolution. Though the tools to conduct such a ‘phenoblast’ search have been at least partially developed within the model organism community (Washington et al. in review) and the Phenoscope group (kb.phenotype.org), the searches rely on ontologies that share a reference ontology for a particular broad taxonomic node and are aligned with respect to one another. For interoperability among ontologies and databases to be achieved, reference ontologies for the anatomy of broad taxonomic subsets, with an associated homology model are needed. These ontologies and associated homology mappings are required to address queries across data expressed using ontologies, for example, comparing the corresponding anatomical structure among *Drosophila*, wasps, and beetles. Therefore, we will develop three key reference ontologies for three higher-level taxonomic groups.

Among animals, single species anatomy ontologies have been developed for five model organisms (Table 1). Multispecies anatomy ontologies are developed or under development for four taxa, two vertebrate clades (teleost fishes, amphibians) and two arthropod clades (spiders, hymenopterans) (Table 1); broad swaths of animal diversity lack multispecies ontologies. Because reference anatomy ontologies for higher-level taxa (vertebrates, arthropods) are completely missing, lower-level anatomy ontologies (e.g. for spiders, hymenopterans) are effectively isolated from one another.

In plants, several ontologies have been developed to describe plant anatomy and phenotype, ranging from the general (the PO Plant Ontology anatomy and developmental stage ontologies) to the relatively species-specific (the TO Trait Ontology for cereal plants) (Table 1). Some preliminary efforts have begun to align terms in different ontologies, however, the ontology development and alignment efforts to date have centered on the angiosperm clade and the needs of a somewhat limited set of research areas, including model organism researchers and plant breeders, and even within these groups the existing ontologies have not been widely adopted. Population geneticists cataloguing natural variation, evolutionary biologists, researchers working with species outside of angiosperms, and other systematists have not yet contributed their knowledge or requirements to these existing efforts. With the recent development of high throughput plant phenotyping facilities and projects in Australia, Germany and France there is a strong need to ensure that the large new datasets produced by these facilities will be available in a widely accepted phenotype representation useful for a variety of research purposes.

We propose three taxonomically focused working groups, each led by an evolutionary morphologist who is an expert in anatomy and systematics of that clade and a model organism biologist whose species is a member. The initial focus of each group will be communication and alignment of both single species and multispecies ontologies, particularly at the higher nodes. We feel that we will be most successful by facilitating the development of common anatomy reference ontologies and phenotype ontology alignments within relatively similar taxonomic groups such as vertebrates and arthropods and vascular plants, thereby taking advantage of taxonomic “comfort zones” and existing collaborative attempts to align phenotype ontologies (such as that currently being attempted for the Teleost Anatomy Ontology and Amphibian Anatomy Ontology). And by facilitating epistemological and alignment-process discussions among the knowledge workers representing the highly diverse groups working to align their ontologies in tandem, we will be able to identify and proactively address a more comprehensive suite of alignment issues that will arise when attempting to align phenotype ontologies at more inclusive levels (e.g., Metazoa, Eukaryota, etc.).

- **Arthropods:** including the Hymenoptera Anatomy Ontology, the Spider Comparative Biology Ontology, the Lepidoptera AToL and its associated morphology ontologies, FlyBase and its associated *Drosophila* development and anatomy ontologies. We will contact heads of the ticks and mosquitoes ontologies as well as those interested in developing ontologies for *Apis* (honey bee) and *Nasonia*. This will be expanded to include the nematode anatomy ontology and other protostome ontologies as

they develop during the course of this five-year RCN.

- **Vertebrates:** Teleost Anatomical Ontology, the Zebrafish Information Network (ZFIN) and its associated anatomical ontology, the Amphibian Anatomical Ontology, and Xenbase and its associated *Xenopus* embryo anatomy ontology. This group builds on the momentum existing within the Phenoscape and AmphibAnat groups.
- **Plants:** the Plant Ontology and its associated plant growth and developmental stages and plant structure ontologies, which was built with the input of the *Arabidopsis*, maize and rice communities. Participants will be recruited from these groups as well as from morphologists (such as those involved in Tree of Life work).

These taxonomic exemplars reflect a broad swath of phenotypic diversity, but point to the challenge of aligning phenotypes as taxonomic groups become more inclusive. A previous top-down attempt has been made to produce a very high-level common reference ontology for all anatomy (Haendel et al. 2008); however, this effort was developed largely from aligning a small set of model organism ontologies, and faced the challenges of aligning extremely diverse organisms such as a fish, a human, and a fly. As ontologies have become larger and represent more diversity, the challenge of aligning anatomical ontologies, including the abilities to reflect biological reality in a usable and computable framework, will grow. We suggest that the existing ontology-building progress and phenotypic diversity reflected in the taxonomic groups listed above provides an ideal framework for developing reference anatomy ontologies using a bottom up approach, i.e. working from the terminal taxa to intermediate taxonomic nodes, as we work toward common high level terms in CARO.

We will develop these reference anatomy ontologies by establishing three, clade-based working groups that will meet twice a year to discuss specific anatomy ontology nodes (e.g. ‘circulatory system’ or ‘skeletal system’) and produce a set of strictly defined terms held in common across taxa within the clade. It is envisioned that this may be a small number (<100) of high-level terms for each of these three groups. However, if these terms are defined in such a way that can be agreed by these different communities, they will be the foundation of reasoning at that taxonomic node. These high-level anatomy ontologies will be deposited to the OBO Foundry (<http://www.obofoundry.org>). The OBO Foundry (Smith et al. 2007) supports community members who are developing and publishing ontologies in the biological domain. By promoting common design principles, these ontologies will be interoperable and “form a single, consistent, cumulatively expanding, and algorithmically tractable whole” (Smith et al. 2007).

The co-leaders of each group will ensure the involvement of both model organism and evolutionary biologists and will open the working groups to new participants annually. We will work with an OBO Foundry coordinator (e.g. Barry Smith) so that we always have an ontologist/logician present (at least one person to help the three groups). This will aid clarity of thinking about required relationships among terms. Additionally, we will invite an ontology consumer/user, e.g. someone who has an application that relies on anatomy ontologies (e.g. Greg Riccardi, who is leading efforts to incorporate ontologies into image annotation). We are requesting the partial funding of a curator to carry out the anatomy reference ontology building, deposition of reference ontologies in OBO Foundry (year 1) and frequent updates, and necessary modifications to existing ontologies so that they will be in alignment.

Table 1. Anatomy ontologies for single model species and multiple species (*) in the Open Biological Ontologies Foundry (<http://obofoundry.org/>) or under development (as of 15 July 2009)

Taxon	Ontology	Reference(s)	Associated database or source
Vertebrates			

Teleosts*	Teleost Anatomy Ontology (TAO)	(Dahdul et al. in press)	http://kb.phenoscape.org/
Zebrafish (<i>Danio rerio</i>)	Zebrafish Anatomy (ZFA)	(Sprague et al. 2001; Sprague et al. 2003; Sprague et al. 2006)	http://zfin.org/
Amphibia*	Amphibian Anatomical Ontology (AAO)	(Maglia et al. 2007)	http://www.amphibanat.org/
<i>Xenopus</i>	<i>Xenopus</i> Anatomy (XAO)	(Segerdell et al. 2008)	http://xenbase.org/
Mouse (<i>Mus</i>)	Mouse Adult Gross Anatomy (MA)	(Bult et al. 2008)	http://www.informatics.jax.org/
Arthropods			
Spiders*	Spider Ontology (SPD)	(Ramirez et al. 2007)	http://research.amnh.org/atol/files/
Ticks*	Tick Gross Anatomy		AnoBase / Vectorbase (http://www.anobase.org)
Fruit fly (<i>Drosophila</i>)	<i>Drosophila</i> Gross Anatomy (FBbt)	(Wilson et al. 2008)	http://flybase.org/
Hymenoptera *	Hymenoptera Anatomy Ontology (HAO)	(in prep.)	(In OBO Foundry after August 1) http://hymao.org/
Mosquito	Mosquito gross anatomy		AnoBase / Vectorbase (http://www.anobase.org)
Plants			
Plants*	Plant Ontology (PO)	(Jaiswal et al. 2005; Ilic et al. 2007; Avraham et al. 2008; Ilic et al. 2008)	http://www.plantontology.org/
Cereal Plants*	Trait Ontology (TO)	(Jaiswal et al. 2002; Yamazaki and Jaiswal 2005)	http://www.gramene.org/plant_ontology/ontology_browse.html#to

3.2 Aim 2: Align and synchronize anatomical ontologies using homology and other similarity:

Ontology alignment is the process of determining correspondences between terms (and less often relations) between two or more ontologies. This process, an important step in any comparative analysis using ontologies, poses a significant challenge for both single and multispecies anatomy ontologies. That said, single species ontologies that are developed by model organism communities such as fly, zebrafish, and mouse or *Arabidopsis*, rice and maize are commonly developed in collaboration with each other to promote cross-species comparisons (Sprague et al. 2006). Still, given the rapid development of new anatomy ontologies and the growing importance of multispecies ontologies, alignment of terms and relations has become a significant problem. Alignment is key, however, to enable automatic searches for similar phenotypes across species, such as provided by tools like phenoblast.

We will foster best practice and encourage: 1) alignment and synchronization of model organism ontologies with their corresponding multispecies ontologies; 2) alignment of the above with the common

reference anatomy ontologies for arthropods, vertebrates, or plants; 3) work toward alignment among model and multispecies ontologies by community evaluation of existing methods, and formalizing the concepts and standards for alignment. We will discuss issues that arise in summit meetings as a group.

3.2.1 *Model organism—Multispecies synchronization and alignment:*

As multispecies anatomy ontologies are built, concepts within the single-species anatomy ontologies of the model organisms that are phylogenetically contained within the broader set of taxa will logically become subtypes. In the case of the Teleost Anatomy Ontology, Zebrafish Anatomy Ontology concepts became the subtypes. Frequent synchronization between subtype and parent is critical for effective information sharing and interoperability. Most new terms added to the ZFA (zebrafish) are added to the TAO, and new terms in the TAO for entities that are found in zebrafish are added to the ZFA. The tool developed by Phenoscape to partially automate the synchronization process, i.e. the Synchronization Tool, a plug-in for OBO-Edit (Day-Richter et al. 2007) that is publicly available from SourceForge (http://sourceforge.net/project/showfiles.php?group_id=224046), can be generalized to be used by all of the MOD-Multispecies ontology pairs. The Synchronization Tool aids in keeping two ontologies aligned by checking for missing cross-references between identically named terms, conflicting data between cross-referenced terms in the two ontologies, terms present in one but missing from the other ontology, and structural differences such as differences in parent terms of cross-referenced terms. Similar functionality has been developed in “*mx*” by the Hymenoptera project.

3.2.2 *Alignment with common reference anatomy ontologies:*

Alignment is much more than just matching terms with the same name. Even apparently similar strings may have very dissimilar semantics, e.g., the term *upper jaw* represents a cluster of different bones in teleosts vs. amphibian anatomy ontologies. As we noted, although tools for ontology alignment exist (see list at <http://ontology.matching.org/projects.html>), none are fully reliable and automatic.

One approach to alignment is to group synonyms and identifiers from anatomy ontologies into high-level categories (Bard et al. 2008), creating a minimal anatomy ontology. Another approach is to create a cross-species anatomy ontology based on summation of terms and relations held in common across ontologies, e.g., "UBERON", see http://bioontology.org/wiki/index.php/UBERON:Main_Page; (Mungall 2004). UBERON groups similar structures in different organisms based on any kind of similarity, while maintaining cross-references to the contributing anatomy ontologies. This approach is useful because it may uncover homology at different levels of granularity, but the number of shared terms across ontologies may be so few that it may not be informative. A third approach is to represent the anatomy of the most recent common ancestor, thereby explicitly formalizing homology between structures within the ontology. This would involve creating a separate higher level anatomy ontology for the common ancestor of the taxa represented in the anatomy ontologies, and then use *derives_from* to relate descendant structures to ancestral structures. Although this approach could be advantageous for the model organisms to link homologous structures (for example, zebrafish heart and mouse heart) it would preclude testing alternative phylogenetic views if only a single *derives_from* relationship could be captured for any one structure. The variety of possible methods for aligning multispecies ontologies, and lack of commonly accepted methodologies, demonstrates that the alignment of separately growing anatomy ontologies will require new approaches, new software tools, and coordination across communities.

3.3 Aim 3: Define phenotype ontology development best practices and standards;

Best practices and standards for the alignment and comparison of phenotypes across diverse and often very different body plans are incompletely developed and require high-level community consideration before adoption and implementation. It is vital to interoperability among data types and databases, indeed to addressing the most basic use cases, that we carefully define and rigorously formalize the logic of these

relations and specify the type of evidence and attribution required.

3.3.1 *Formalize similarity relationships (homology, homoplasy) of phenotype entities among taxa:*

Perhaps the most essential, misunderstood and hotly debated relationship among phenotypes is that of homology, i.e. the similarity in cross-species characteristics due to common ancestry. The lack of a homology model was identified by the community as a critical obstacle to data exchange and data integration. How should homology of phenotypic entities be represented? Should a relationship such as *homologous_to* be part of the ontology or should an anatomical ontology, e.g., have a purely structural axis? The recommendation of some is to maintain a clear separation of homology from its ontological and solely structural representation of anatomy (Dahdul et al., in press). However, an argument can be made that ontology term names convey homology and that synonyms can be used to represent homologues (e.g. Plant Structure Ontology, Illic et al., 2008). This assumes that anatomical structures in different taxa that bear the same name are homologous. There are many exceptions to this however (e.g. leg or eye of fly vs. human), and based on current phylogenies, it is likely that the leg of a fly is *not homologous_to* the leg of a human. Searching on semantic similarity alone (i.e. on 'leg') would thus result in finding similarity due to common ancestry (homology) lumped with similarity that has evolved independently (homoplasy, analogy). Although this type of search may be most appropriate, e.g., if one is searching for common gene signaling pathways among structures that are judged 'similar' by any criterion, a user may well want to filter results by, e.g., homology. Clearly statements concerning the evolutionary derivation of structures in different taxa that are identically named or not, require well-considered relations (i.e. phylogenetic trees). The cross linkages among ontologies which seek to incorporate concepts of homology are further complicated because phylogenetic trees are relatively fluid hypotheses of relationship. The ontology relations *homologous_to* and *not_homologous_to* are in the process of being formalized (a working definition is available from the Relations Ontology wiki page, http://www.bioontology.org/wiki/index.php/RO:Main_Page), but it requires further input, examples, and discussion from the evolutionary community. The ontology relations *homologous_to* and *not_homologous_to* are in the process of being formalized², but require further input from the evolutionary community. The phenotype community recognizes that homology issues are not isolated to phenotype (though it was in consideration of anatomical features that the concept of homology was conceived), but apply to every aspect of the organism, including molecular data. The RNA Ontology Research Coordination network has recently addressed this issue (Brown et al. In press) and we plan to work with them to formalize these concepts (Leontis pers. comm. has agreed to participate in these meetings). Marc Robinson-Rechavi (University of Lausanne, Switzerland), is working to clarify and formalize the various concepts which are related to homology (*BGee* project), and his group is interested in participating as well.

3.3.2 *Assess adequacy of evidence codes and attribution.*

Homology assertions are hypotheses, sometimes opposing, that require evidence and attribution. Standard lines of evidence that are used to assess homology of phenotypes *a priori* include similarity in shape and size, topographic position, complexity, and development (Remane 1952; Roth 1984; Patterson 1988). Homology can be tested *a posteriori* by the distribution of character states on phylogenies resulting from character state analysis (Mayden and Wiley 1992). Phenoscape followed OBO community standards for establishing and using evidence codes to annotate homology of anatomical structures (Dahdul et al., in press) and is implementing this approach. Thus users can view competing hypotheses where they occur, or those supported by a particular type of evidence. The Phenoscape database can

serve as a model for the broader phenotype community to assess the completeness of evidence codes and to consider how this approach might be extended to critical use cases.

3.3.3 *Standardize representation of serial homologues (metamerism).*

Representing serial homologues in a multi-species anatomy ontology requires accommodation of different naming methods used in the literature for elements of the series. Serial homologues are usually given names based on ordinal position (e.g., vertebra 1), but some elements, frequently the first or last of a series, may also be referred to by a structural name. Consequently, a term in an ontology may have its name also designated as a synonym of another term. For example, the infraorbital series is a set of bones that typically lie immediately below and posterior to the eye in fishes. The last bone of the series is considered homologous across many fishes because of similarity in position and structure. It is either named the dermosphenotic or it is named with a number according to its terminal position in the series. In some cypriniform fishes, the dermosphenotic is the terminal of five bones and it is also termed "infraorbital 5". In some characiform fishes, however, there are six bones in the infraorbital series: the terminal bone is referred to as the dermosphenotic, and it is also termed "infraorbital 6". To represent these alternative names for dermosphenotic in TAO, the term dermosphenotic is given the synonyms infraorbital 5 and infraorbital 6. However when a query is made on infraorbital 5, the data from non-homologous bones, i.e. the dermosphenotic in cypriniforms and the infraorbital 5 in characiforms will be returned together. Filtering these results through a homology table that specifies that the dermosphenotic [synonym: infraorbital 5] in cypriniforms is homologous to the dermosphenotic [synonym: infraorbital 6] in characiforms for these data, however, will effectively restrict the search outcome to homologues.

3.3.4 *Representation of taxonomically variable structures in a multispecies ontology.*

Different species possess different collections of phenotypic features. When features are aggregated in a multispecies ontology, however, some of the relations among them may become taxonomically inappropriate. For example, vertebra 1, vertebra 2, vertebra 3 and vertebra 4 are subtypes (is_a relation) of Weberian vertebra in otophysan fishes but not in other fishes, which do not have Weberian vertebrae (vertebrae 1--4 in other fishes are subtypes of vertebra). Solutions to this general issue and their ramifications for reasoning need to be explored.

3.4 Aim 4: Reach out to ancillary phenotype groups to share with them common concepts and practices

We will promote the development of ontologies that are required for annotation of data related to anatomy. In both evolutionary and model organism research, phenotypic variation involves changes in not only anatomy, but also in behavior and function. We propose to initiate and or foster the development of ontologies in these areas that are critical for the comprehensive annotation of the research data, by inviting one 'ancillary' group to meet with the RCN at the summit meeting each year (beginning year 2) and mentoring them with additional training through the summer course or exchange program. We will also mentor the development of new ontologies, helping researchers develop them in a coordinated manner and connecting them to best practices established for this community.

3.4.1 *Taxonomy group (year 2):*

A taxonomy ontology is required for annotation of cross-species phenotype data. In the evolutionary community, particular phenotypes are borne by particular species, and thus a species name from a taxonomy ontology is required for association with the particular phenotype. Although the National Center for Biological Information (NCBI) provides a taxonomy for navigating and locating species for which it has data deposited, it is missing extinct taxa and extant taxa for which genomic data are not available. Coordination of how to add these species to existing taxonomies for the groups is required for data exchange, however, and we propose to discuss this issue with the experts in this field. Taxonomy

ontologies are not interchangeable with phylogenies; the evolutionary community ultimately desires a system where taxa (bearing their phenotype annotations) can be viewed on multiple phylogenies.

To date, individual evolutionary databases have developed local taxonomy ontologies. For example, the Teleost Taxonomy Ontology (TTO) was constructed from the expert taxonomic database, the Catalog of Fishes (Eschmeyer 1998) and extinct fishes added as required by literature annotation. The AmphibAnat group developed the Amphibian Taxonomy Ontology (ATO), closely following the most widely accepted amphibian taxonomy (AmphibiaWeb portal, <http://www.amphibiaweb.org>). A broader solution is required, with integration with TDWG, EoL, and NCBI. Participants from these groups will be included.

3.4.2 Behavioral ontology group (year 3):



The best established of the independent behavior ontologies is the Animal Behavior Core (ABOCore) ontology maintained and used by the Ethosource Project and by David Shotton's group at Oxford. There are a number of other projects using either independent behavior ontologies (e.g. Christopher D. Smith, UCSF has a social insect behavior ontology) or contributing behavior terms to the biological process tree in the Gene Ontology (GO). The ABOCore has a behavioral ecology focus, whereas most of the GO-related behavior work has involved behavior genetics. Facilitating communication and working towards interoperability among these behavior ontology efforts will be the primary goal of this group. Of particular interest to the behavior group will be the way that the anatomy groups represent developmental time, because behavioral phenotypes are frequently sequences of events during a temporal period. This group will be led by Peter Midford, and participants may include Anne Clark, Sue Margulis, David Shotton, Ed Scholes, and Judy Blake (representing GO Process).

3.4.3 Functional ontology group (year 4):

Ontologies of functional terms, are a critical complement to existing phenotype ontologies. These terms might be added to existing ontologies (e.g., the GO) or require the development of new ontologies. Function terms are appropriate to annotating both structure (e.g., anatomy) as well as process (e.g., biomechanics, behavior, physiology). Frequently, structure and process data will share functional annotations, allowing the generation of hypotheses (e.g., existence of selective forces). By bringing together anatomy and process ontologies with developers of a whole organism functional ontology, the RCN could facilitate a unique contribution. Leadership of this group has not been determined, but evolutionary biologists who have expressed interest in this area include Betsy Dumont, Mark Westneat, Peter Wainwright, Marvalee Wake, Kathrine Loudon, and Janice Voltzow. Additionally, philosopher and logician Barry Smith and Japanese scientist Riichiro Mizoguchi are also interested in participating.

3.4.4 New anatomy ontology groups (year 5):

We will seek out and include participants in this group who are interested in building new ontologies for model species or multiple species (clade). We will foster the development of these anatomy ontologies by helping them initialize their ontology and providing support through the part-time ontology curator (see budget justification). In this last year of the RCN we will have a nearly final version of the textbook, and some of these people will have participated in the summer course.

3.5 Aim 5: Educate the community about the methods for developing ontologies and their importance and utility in research

At a Phenoscope sponsored workshop in April 2009, participants identified a substantial lack of understanding of ontologies and their applications by the scientific community. The group felt the dearth was owing to several factors, including: a lack of trained ontology workers, sociological aversions to adopting standardized vocabularies, and a general unfamiliarity with ontologies and their philosophical basis. The workshop participants stated that their peers do not understand what an ontology is, are not familiar with the ontology-related resources that are available to them, and most importantly, do not know

how ontologies are used to enhance science and address novel research questions. Community training is critical to the successful incorporation of ontologies into the study of phenotype, and consequently, our ability to integrate diverse phenotype data. A strong training plan must include a variety of instructive and informative modalities, must focus on overcoming the sociological issues that arise as communities adopt new technologies and ideas, and must involve comprehensive, structured documentation of tools, data, and annotation best practices.

3.5.1 *RCN meetings and working groups*

Annual Summit Meeting. We will hold an annual group summit of all stakeholders (including PIs, Advisory board members, and working group participants) that will focus one or more specific conceptual issues surrounding our goals. Meetings will be held at NESCent, and will include break-out sessions for groups working on the aims outlined in Section 4. In addition to work accomplished and issues resolved, a major goal of the summit is to facilitate communication and networking among diverse stakeholders. In the first two years, summit meetings and taxonomic working groups will focus on alignment and synchronization issues within taxonomic groups. Participants will include relevant core members plus additional participants with a strong interest in ontology development within that taxonomic group.

Bi-annual Working Group Meetings. We will hold two working group meetings for each working group in years 1—3 and annually thereafter. One of these meetings will be held in conjunction with the annual Summit meeting at NESCent and the other will be held at a time chosen by group members.

3.5.2 *Presentations at national/international meetings*

We will support the travel of 2 participants per year to a national or international meeting to present ontology-based work. Funding decisions will be made by the Advisory board, with priority going to those presenting to the broadest audience and also funding prioritized to students and postdoctoral fellows. The goal is to broadly disseminate information to the larger scientific community regarding phenotype ontologies and their utility and involve new participants in the RCN.

3.5.3 *Collaborative exchange opportunities*

We will implement a program to facilitate collaborations among network participants; wherein students, postdocs, faculty, and/or small working groups can apply for travel funds to support collaborations related to one of the project aims. We envision these fellowships supporting exchanges such as a model organism focused student working for a month in a evolutionary ontology lab, a scientist visiting the RCN ontology coordinator for a few days to discuss and resolve alignment issues, or an experienced ontology-builder visiting a neophyte to help him/her set up a framework for their ontology.

3.5.4 *Summer short course; 3 student stipends*

We will hold an annual summer short course beginning in year 2. We will offset the costs three students (@ \$1,000 each) each year (funding decisions made by Advisory board). We will have two concurrent sessions: 5 days of introductory/beginning topics, and 5 days of advanced topics (with a possibility of receiving 2 graduate credits). The course will include intensive training in ontology development and software, curation, structure, best practices, and uses, with focus on phenotype and multispecies ontologies. Much of the underlying information for the course will be made available by the PIs and interested working group members. Maglia will be responsible for coordination of the course though it is expected that the other PIs will contribute.

3.5.5 *Textbook/handbook*

As a tangible outcome of the summit meetings, short course, and symposia, we will develop a handbook of anatomy ontology standards and best practices, as defined by the participant community. This book will be designed to address both conceptual and practical issues so that the process of ontology

building becomes transparent, consistent, and readily practicable by members of the scientific community.

4. Timeline of Activities

Year 1 (1 March 2010-28 Feb 2011) --*Summit meeting*: Reference anatomy ontologies

Working groups: Reference anatomy building at high-level nodes - begin work within taxonomic groups; discuss and generalize ontology building issues. Identify domain experts for future participation (e.g., which anatomists are the best qualified for resolving some tricky problems)

Year 2 (1 March 2011-28 Feb 2012)--*Summit meeting*: Homology, similarity, alignment and integration

Ancillary working group: Taxonomy

Working groups: Align higher level ontology terms; discuss and generalize ontology building issues

Short course (June 2011)

Year 3 (1 March 2012-28 Feb 2013)--*Summit meeting*: Annotation and query issues, best practices

Ancillary working group: Behavior

Working groups: annotate queries across higher levels, integrate with GO

Short course (June 2012)

Year 4 (1 March 2013-28 Feb 2014)--*Summit meeting*: Best practices, standards; Visualization methods

Ancillary working group: Function

Working groups: build/align higher level taxonomic ontology

Short course (June 2013)

Year 5 (1 March 2014-28 Feb 2015)--*Summit meeting*: Identify and discuss integration of community/existing data resources and use cases, propose best practices; finalize handbook/textbook

Ancillary working group: New ontologies

Working groups: develop key use cases, publish results

Short course (June 2014)

5. Management Plan and Responsibilities

The coordination network comprises a large international group of stakeholders with varied levels of responsibility and commitment, as follows.

5.1 Participants

A large initial group of international working group participants have identified themselves as interested in participating in this RCN. They come from diverse organizations and represent the diverse areas of biology in which understanding phenotype is a key aspect of the research (Table 2). These researchers come from three traditions: 1) model organism biology, whose focus is on the genomic, genetic, and developmental basis of phenotype; 2) evolutionary biology/systematic biology, whose focus is on understanding the relationships of extant and extinct species based on phenotypic and molecular data; and 3) ontology/logic/philosophy/informatics, who work closely with biological researchers to develop the databases, interfaces, and analytical tools required to compute on phenotype. Some of these participants are leading or heavily involved with existing efforts that involve using ontologies to represent phenotype, and thus are also included as important members of the working groups. We plan to advertise the RCN meetings through listserves for these communities (e.g. NESCent, Evoldir, Biocurator, OBO lists, Devo-evo SICB). The advisory board members, in consultation with working group leaders, will select from among those recruited so as to ensure new participants and broad community training and yet maintain continuity with regard to ontology development within each group. The inclusion of new researchers, post-docs, graduate students, and undergraduates will be prioritized. The institutions represented in the proposal have existing, well-designed strategies to recruit individuals from underrepresented groups (including minorities and women); we will embrace these practices in our recruitment efforts. The PI's institution, e.g., has several programs to recruit Native American students.

5.2 Working groups

The three major taxon-based working groups (Vertebrates, Arthropods, Plants) will meet twice a year throughout the project, with the goal of establishing communication about phenotype representation between model organism and evolutionary scientists and working to align particularly the high-level nodes in their respective phenotype ontologies, generalize the issues, and develop the textbook/manual. One of the meetings will be held at NESCent in association with the Summit meeting in October and the other will be held at a time and location chosen by the participants.

Working groups will be led or co-led by the PI, a co-PI, or Advisory board member, and they will include model organism and evolutionary biologists who are engaged in ontology development. The vertebrate group will be led by PI Paula Mabee and Advisory board member Monte Westerfield. The arthropod working group will be led by co-PI Andy Deans and Advisory board member Michael Ashburner. The plant working group will be led by co-PI Eva Huala and Advisory board member Peter Stevens. Each working group will consist of 6-7 members; at least a third of the members will change each year (turnover of at least 2 members each year). Initial working group participants, i.e. those who have expressed interest and commitment to this project, are listed in Table 2. At each meeting we will involve an 'ontologist', i.e. an expert in the logic, philosophy, and practical aspects of developing and using ontologies (Table 2). The OBO Foundry has agreed to help us with this function. We will also invite an ontology 'consumer', i.e. an applications-based person with a different view of ontologies. Additionally, the ontology coordinator hired by this grant (part time) will be in attendance at each meeting to provide help with ontology development issues.

For ancillary phenotype working groups, the behavior group will be led by Advisory board member Peter Midford. The leadership of the function group will be determined in year 2, but will likely be Mark Westneat or Betsy Dumont. The leadership of the taxonomy group will be determined in year 3; possibilities include someone from NCBI, Peter Midford, or Encyclopedia of Life person.

5.3 Principle investigators

The PI and co-PIs are actively involved in ontology-based research, ontology development, curation, and community outreach and education. Together they comprise a team that can coordinate the needs of the disparate communities involved (e.g. standards and best practices) and yet allow each sub-community to determine its own requirements and needs as well. PI Paula Mabee is the contact person for the group and will provide leadership in fully coordinating and integrating the activities of the group. Mabee and co-PIs will share the responsibilities of coordinating the research network activities. Mabee, under the direction of the Advisory board and assistance of co-PIs, will organize the annual summit meetings and help organize the mid-year working groups and also provide the logistic support for setting up these meetings). She will direct the development of the RCN wiki/webpage, lead the vertebrate working group, and edit the textbook/manual. Co-PI Huala will be lead the plant working group; co-PI Deans will lead the arthropod group; co-PI Maglia will coordinate teaching the summer course; co-PI Lewis will organize and lead the logicians and philosophers at the summit meetings to ensure best practices.

5.4 Advisory board

The Advisory board is made up of an invited group of invested participants who are stakeholders in database efforts to develop or use phenotype ontologies, and represent a broad range of subdisciplines and expertise. They have committed to meeting yearly at the summit, will make recommendations and advise the PIs in the oversight of the grant, and will review and make recommendations for funding of student stipends, exchange proposals. They will advise the PI on ways to maintain openness and involve the community, and they will select among interested participants in working groups (in the event that advertising the RCN meetings leads to more interest than funding). Additional board members may be

added as required and/or several members rotated off the board in year 3, depending on participation.

5.5 Supporting organizations/staff

5.5.1 Organizations

NESCent (National Evolutionary Synthesis Center): NESCent can provide facilities for some of the annual summits, working group meetings, and the summer course. The center is also willing to host the part-time ontology coordinator/facilitator as a staff member or a visiting scientist (see letter).

5.5.2 Part-time ontology coordinator/facilitator

The ontology coordinator will use OBO-Edit software to develop the anatomy reference ontologies in relation to the terms, definitions, and relations specified by the 3 taxon-based working groups. He/she will deposit the reference ontologies in OBO foundry <http://www.obofoundry.org> and update them on a regular basis. They will provide a ‘help desk’ function for new groups developing ontologies, help facilitate the summit meeting and working groups and also help teach and assist in the summer course. They will also be available to assist the participants and working groups with issues and best practices.

Table 2. Initial set of participants in the Phenotype Ontology Coordination Network. All of these people have been contacted (or contacted us) and have agreed to their initial role(s) in the RCN

Participant	Role	Expertise	Affiliation
Mabee, Paula	PI	Morphology, vertebrates, ToL, Phenoscape	University of South Dakota
Deans, Andy	Co-PI	Arthropods, Hymenoptera, Morphology, insects	Department of Entomology, North Carolina State University
Huala, Eva	Co-PI	Plants, MOD (The Arabidopsis Information Resource)	Carnegie Institution for Science
Lewis, Suzanna	Co-PI	Ontologies, MOD (FlyBase)	Berkeley Bioinformatics and Ontology Project
Maglia, Anne	Co-PI	Vertebrates, AmphibAnat	Missouri University of Science and Technology
Ashburner, Michael	Advisory board	Arthropods, Ontologies, morphology, insects, (FlyBase)	Department of Genetics, University of Cambridge, UK
Midford, Peter	Advisory board	Behavior, arthropods, ontologies, Phenoscape	Univ. Kansas and NESCent
Stevens, Peter	Advisory board	Plants, ToL	Missouri Botanical Garden
Westerfield, Monte	Advisory board	Vertebrates, MOD, Zebrafish Information Network (ZFIN)	Institute of Neuroscience, University of Oregon
Westneat, Mark	Advisory board	Vertebrates, EoL	Field Museum
Sutherland, David	Working group	Arthropods; <i>FlyBase</i>	University of Cambridge, UK
Yoder, Matt	Working group	Arthropods; Hymenoptera Tree of Life; Informatics, <i>Mx</i>	Entomology, NCSU
Weller, Susan	Working group	Arthropods; Lepidopteran Tree of Life	University of Minnesota
Ronquist, Fred	Working group	Arthropods, Informatics, <i>Morphbank</i>	Swedish Museum of Natural History, Sweden
Coddington, Jonathan	Working group	Arthropods; Spider Tree of Life	Entomology, Smithsonian Institution

Ramirez, Martin	Working group	Arthropods; Spider Tree of Life	Museo Argentino de Ciencias Naturales, Buenos Aires
Riccardi, Greg	Working group	Informatics, <i>Morphbank</i>	Florida State University
Lapp, Hilmar	Working group	Informatics, <i>Phenoscape</i>	NESCent
Stoltzfus, Arlin	Working group	Informatics; Evolutionary Informatics Working Group	Center for Advanced Research in Biotechnology, U Maryland
Vogt, Lars	Working group	Informatics; ontologies; invertebrates	Berlin
Parr, Cyndy	Working group	Ontologies, EOL; Arthropods; Lepidoptera Tree of Life	Smithsonian Institution
Cowell, Lindsay	Working group	Ontologies	Duke University Medical Center
Mungall, Chris	Working group	Ontologies	Berkeley Bioinformatics and Ontology Project
Smith, Barry	Working group	Ontologies; philosophy; logic	Natl Ctr Biomedical Ontologies (Buffalo); OBO Foundry
Leontis, Neocles	Working group	Ontologies; RNA ontology RCN	Bowling Green University
Parkinson, Helen	Working group	Ontology (Array Express)	EMBL
Mueller, Lukas	Working group	Plants, Sol Genomics Network, Solanaceae Phenotype ontology	Boyce Thompson Institute, Cornell
Shaeffer, Mary	Working group	MaizeGDB Phenotypic Controlled Vocabulary	USDA ARS, University of Missouri
Shoemaker, Randy	Working group	Plants, SoyBase, soybean ontology	Iowa State University
Mast, Austin	Working group	Plants, Morphbank	Florida State University
Vision, Todd	Working group	Plants, Ontology consumer (Phenoscape)	Univ. NC; NESCent
Blake, Judy	Working group	Vertebrates, <i>Mouse Genome Informatics</i>	The Jackson Laboratory
Dahdul, Wasila	Working group	Vertebrates; <i>Phenoscape</i>	NESCent
Dettai, Agnes	Working group	Vertebrates	Muséum National d'Histoire Naturelle, Paris
Haendel, Melissa	Working group	Vertebrates; <i>ZFIN</i> , Common Anatomy Reference Ontology	Zebrafish Information Network (ZFIN); University of Oregon
O'Leary, Maureen	Working group	Vertebrates; <i>Morphobank</i> ; Mammal Tree of Life	Stony Brook
Shearman, Rebecca	Working group	Vertebrates; <i>AmphibAnat</i>	Wesleyan University
Vize, Peter	Working group	Vertebrates; <i>Xenbase</i>	University of Calgary
Wake, Marvalee	Working group	Vertebrates; <i>AmphibAnat</i> , Amphibian Tree of Life	UC Berkeley

References

- Avraham, S., C.-W. Tung, K. Ilic, P. Jaiswal, E. Kellogg, S. McCouch, A. Pujar, L. Reiser, S. Rhee, M. Sachs, M. Schaeffer, L. Stein, P. Stevens, V. Leszek, F. Zapata, and D. Ware. 2008. The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Research* 36:D449-D454.
- Bard, J. B. L., J. Malone, T. F. Rayner, and H. Parkinson. 2008. Minimal Anatomy Terminology (MAT): a species-independent terminology for anatomical mapping and retrieval. *Bio-Ontologies Meeting*, Toronto, Canada.
- Brown, J. W., A. Birmingham, P. E. Griffiths, F. Jossinet, R. Kachouri-Lafond, R. Knight, B. Franz Lang, N. Leontis, G. Steger, J. Stombaugh, and E. Westhof. In press. The RNA Structure Alignment Ontology *Journal of the RNA Society* <http://rnajournal.cshlp.org/content/early/recent>.
- Bult, C. J., J. T. Eppig, J. A. Kadin, J. E. Richardson, J. A. Blake, and members of the Mouse Genome Database Group. 2008. The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Research* 36:D724-D728.
- Dahdul, W. M., J. G. Lundberg, P. E. Midford, J. P. Balhoff, H. Lapp, T. J. Vision, M. A. Haendel, M. Westerfield, and P. M. Mabee. in press. The Teleost Anatomy Ontology: Anatomical representation for the genomics age. *Systematic Biology*.
- Day-Richter, J., M. Harris, M. Haendel, T. G. O. O.-E. w. Group, and S. Lewis. 2007. OBO-Edit - an ontology editor for biologists. *Bioinformatics Applications Note* 23:2198-2200.
- Galperin, M. Y., and G. R. Cochrane. 2009. *Nucleic Acids Research annual Database issue and the NAR online Molecular Biology Database Collection in 2009*. *Nucleic Acids Research* 37:D1-D4.
- Gaudet, P. and the Reference Genome Group of the Gene Ontology Consortium. 2009. The Gene Ontology's Reference Genome Project: a unified framework for functional annotation across species. *PLoS Comput Biol.* 2009 Jul;5(7):e1000431.
- Haendel, M. A., F. Neuhaus, D. S. Osumi-Sutherland, P. M. Mabee, J. L. V. Mejino, C. J. Mungall, and B. Smith. 2008. CARO—the common anatomy reference ontology. Pp. 327-349 in A. Burger, D. Davidson, and R. Baldock, eds. *Anatomy Ontologies for Bioinformatics: Principles and Practice*. Springer.
- Ilic, K., E. A. Kellogg, P. Jaiswal, F. Zapata, P. F. Stevens, L. P. Vincent, S. Avraham, L. Reiser, A. Pujar, M. M. Sachs, N. T. Whitman, S. R. McCouch, M. L. Schaeffer, D. H. Ware, L. D. Stein, and S. Y. Rhee. 2007. The plant structure ontology, a unified vocabulary of anatomy and morphology of a flowering plant. *Plant Physiol* 143:587-599.
- Ilic, K., S. Y. Rhee, E. A. Kellogg, and P. F. Stevens. 2008. Plant Structure Ontology (PSO) -- A Morphological and Anatomical Ontology of Flowering Plants in A. Burger, D. Davidson, and R. Baldock, eds. *Anatomy Ontologies for Bioinformatics: Principles and Practice*. Springer.
- Jaiswal, P., S. Avraham, K. Ilic, E. Kellogg, S. McCouch, A. Pujar, L. Reiser, S. Rhee, M. Sachs, M. Schaeffer, L. Stein, P. Stevens, L. Vincent, D. Ware, and F. Zapata. 2005. Plant Ontology (PO): a controlled vocabulary of plant structures and growth stages. *Comparative and Functional Genomics* 6:388-397.

- Jaiswal, P., D. Ware, J. Ni, K. Chang, W. Zhao, S. Schmidt, X. Pan, K. Clark, L. Teytelman, S. Cartinhour, L. Stein, and S. McCouch. 2002. Gramene: development and integration of trait and gene ontologies for rice. *Comparative and Functional Genomics* 3:132-136.
- Mabee, P., M. Ashburner, Q. Cronk, G. Gkoutos, M. Haendel, E. Segerdell, C. Mungall, and M. Westerfield. 2007a. Phenotype ontologies: the bridge between genomics and evolution. *Trends in Ecology & Evolution* 22:345-350.
- Mabee, P. M., G. Arratia, M. Coburn, M. Haendel, E. J. Hilton, J. G. Lundberg, R. L. Mayden, N. Rios, and M. Westerfield. 2007b. Connecting evolutionary morphology to genomics using ontologies: A case study from cypriniformes including zebrafish. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution* 308B:655-668.
- Maglia, A. M., J. L. Leopold, L. A. Pugener, and S. Gauchi. 2007. An anatomical ontology for amphibians. *Pacific Symposium for Biocomputing* 12:367-378.
- Mayden, R. L., and E. O. Wiley. 1992. The fundamentals of phylogenetic systematics. Pp. 114-185 in R. L. Mayden, ed. *Systematics, Historical Ecology, and North American Freshwater Fishes*. Stanford University Press, Stanford, CA.
- Mungall, C. 2004. Obol: integrating language and meaning in bio-ontologies. *Comparative and Functional Genomics* 5:509-520.
- Patterson, C. 1988. Homology in classical and molecular biology. *Molecular Biology and Evolution* 5:603-625.
- Ramirez, M., J. Coddington, W. Maddison, P. Midford, L. Prendini, J. Miller, C. Griswold, G. Hormiga, P. Sierwald, N. Scharff, S. Benjamin, and W. Wheeler. 2007. Linking of Digital Images to Phylogenetic Data Matrices Using a Morphological Ontology. *Systematic Biology* 56:283-294.
- Remane, A. 1952. *Die Grundlagen des naturlichen Systems der vergleichenden Anatomie und der Phylogenetick*. Geest und Portig, Leipzig.
- Roth, V. L. 1984. On homology. *Biological Journal of the Linnean Society* 22:13-29.
- Segerdell, E., J. Bowes, N. Pollet, and P. Vize. 2008. An ontology for *Xenopus* anatomy and development. *BMC Developmental Biology* 8:92.
- Smith, B., M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. Goldberg, K. Eilbeck, A. Ireland, C. Mungall, T. O. Consortium, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S. Sasone, R. Scheuermann, N. Shah, P. Whetzel, and S. Lewis. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* 25:1251-1255.
- Sprague, J., L. Bayraktaroglu, D. Clements, T. Conlin, D. Fashena, K. Frazer, M. Haendel, D. Howe, P. Mani, S. Ramachandran, K. Schaper, E. Segerdell, P. Song, B. Sprunger, S. Taylor, C. Van Slyke, and M. Westerfield. 2006. The Zebrafish Information Network: the zebrafish model organism database. *Nucleic Acids Research* 34:D581-D585.
- Sprague, J., D. Clements, T. Conlin, P. Edwards, K. Frazer, K. Schaper, E. Segerdell, P. Song, B. Sprunger, and M. Westerfield. 2003. The Zebrafish Information Network (ZFIN): the zebrafish model organism database. *Nucleic Acids Research* 31:241-243.

- Sprague, J., E. Doerry, S. Douglas, and M. Westerfield. 2001. The Zebrafish Information Network (ZFIN): a resource for genetic, genomic and developmental research. *Nucleic Acids Research* 29:87-90.
- The Gene Ontology Consortium. 2000. Gene Ontology: Tool for the unification of biology. *Nature Genetics* 25:25-29.
- The Gene Ontology Consortium. 2001. Creating the gene ontology resource: Design and implementation. *Genome Research* 11:1425-1433.
- The Gene Ontology Consortium. 2008. The Gene Ontology Project in 2008. *Nucleic Acids Research* 36:D440-D444.
- Washington, N. L., M. A. Haendel, C. J. Mungall, M. Ashburner, M. Westerfield, and S. E. Lewis. in review. Linking human diseases to animal models using ontology-based phenotype annotation. 46 pages.
- Wilson, R. J., J. L. Goodman, V. B. Strelets, and the FlyBase Consortium. 2008. FlyBase: integration and improvements to query tools. *Nucleic Acids Research* 36:D588-D593.
- Yamazaki, Y., and P. Jaiswal. 2005. Biological ontologies in rice databases. An introduction to the activities in Gramene and Oryzabase. *Plant Cell Physiol.* 46:63-68.