BMC Bioinformatics

The Xenopus phenotype ontology: bridging model organism phenotype data to human health and development

Malcolm E. Fisher[1], Erik Segerdell[1], Nicolas Matentzoglu[2,3,4], Mardi J. Nenni[1], Joshua D. Fortriede[1], Stanley Chu[5], Troy J. Pells[5], David Osumi-Sutherland[4], Praneet Chaturvedi[1], Christina James-Zorn[1], Nivitha Sundararaj[1], Vaneet S. Lotay[5], Virgilio Ponferrada[1], Dong Zhuo Wang[5], Eugene Kim[5], Sergei Agalakov[5], Bradley I. Arshinoff[5], Kamran Karimi[5], Peter D. Vize[5] and Aaron M. Zorn[1]*

*Correspondence:
aaron.zorn@cchmc.org
[1] Division of Developmental
Biology, Cincinnati Children's
Hospital Medical Center,
Cincinnati, OH, USA
Full list of author information
is available at the end of the
article

Abstract

Background: Ontologies of precisely defined, controlled vocabularies are essential to curate the results of biological experiments such that the data are machine searchable, can be computationally analyzed, and are interoperable across the biomedical research continuum. There is also an increasing need for methods to interrelate phenotypic data easily and accurately from experiments in animal models with human development and disease.

Results: Here we present the *Xenopus* phenotype ontology (XPO) to annotate phenotypic data from experiments in *Xenopus*, one of the major vertebrate model organisms used to study gene function in development and disease. The XPO implements design patterns from the Unified Phenotype Ontology (uPheno), and the principles outlined by the Open Biological and Biomedical Ontologies (OBO Foundry) to maximize interoperability with other species and facilitate ongoing ontology management. Constructed in Web Ontology Language (OWL) the XPO combines the existing uPheno library of ontology design patterns with additional terms from the *Xenopus* Anatomy Ontology (XAO), the Phenotype and Trait Ontology (PATO) and the Gene Ontology (GO). The integration of these different ontologies into the XPO enables rich phenotypic curation, whilst the uPheno bridging axioms allows phenotypic data from *Xenopus* experiments to be related to phenotype data from other model organisms and human disease. Moreover, the simple post-composed uPheno design patterns facilitate ongoing XPO development as the generation of new terms and classes of terms can be substantially automated.

Conclusions: The XPO serves as an example of current best practices to help overcome many of the inherent challenges in harmonizing phenotype data between different species. The XPO currently consists of approximately 22,000 terms and is being used to curate phenotypes by Xenbase, the *Xenopus* Model Organism Knowledgebase, forming a standardized corpus of genotype–phenotype data that can be directly related to other uPheno compliant resources.

Fisher *et al. BMC Bioinformatics*   (2022) 23:99

Page 2 of 15

Background

Laboratory organisms, such as frogs, mice, fish, fruit flies and worms are essential to investigate conserved gene function and model human development, homeostasis and disease. In order for the experimental results in hundreds of thousands of published animal model papers to be computer readable, amenable to computational analysis and interrelated across species, the data must be curated using ontologies of controlled vocabulary describing genes, gene products, molecular and biological processes, anatomy, and phenotypes. Ontologies codify semantic relationships between biological concepts and are essential because natural language descriptions in publications are too variable, organism specific and cumbersome to be machine processed efficiently [1].

When biomedical ontologies initially began to proliferate the Open Biomedical Ontologies (OBO) consortium established a set of principles for ontology development to improve accessibility, specificity, and interoperability [2]. Curating with ontologies allows data from two papers using different phrases or synonyms to describe the same structure or phenotype (e.g. enlarged heart versus cardiac hypertrophy) to be annotated using the same ontology term and ID number. For a computer, the two experiments both use the same ID number and therefore contain data on the same anatomical structure. Ontologies also define the relationships between terms—for example the 'heart' is a part of the 'cardiovascular system' which also has its own unique ID number, thus papers on different parts of the cardiovascular system can be mapped to each other by a computer using this 'part_of' relationship ID. Ontologies can store many such relationships along with synonyms and cross-reference IDs to other ontologies, such as cellular components in an anatomy ontology being cross referenced to the equivalent Cell Component term in the Gene Ontology (GO) (RRID:SCR_002811) [3]. By curating data from thousands of publications with interconnected ontologies it is possible to build a web of knowledge that can be subjected to statistical and computational analyses. In the context of disease modeling, ontologies make it possible, in principle, to connect genotype and phenotype data from experiments in animal models with an understanding of pathological mechanisms associated with an orthologous human condition. To realize this potential, ontologies used by different biomedical research communities need to be interoperable and grounded in a common syntax.

Ontology based phenotype curation has traditionally taken one of two forms, either post-composed or pre-composed. In post-composed approaches two or more ontologies are used in combination, typically, an anatomical ontology term is used to define the entity (E) and this is combined with a quality ontology term (Q) to generate the entity-quality (EQ) phenotype description. For example, the entity may be 'heart' from an Anatomy Ontology and the quality may be 'decreased size' from the Phenotype and Trait Ontology (PATO) (RRID:SCR_004782) [4], these would be combined to generate the EQ statement: 'heart, decreased size'. The entity component can be complex with multiple independent entity terms joined by Relationship Ontology [5] terms, such as 'has_quality' or 'part_of', to give great flexibility in description. On the other hand, pre-composed approaches use a predefined phenotype ontology, where a single ontology term using a controlled syntax already exists, such as 'decreased size of the heart' (XPO:0103343).

Fisher *et al. BMC Bioinformatics*   (2022) 23:99

Page 3 of 15

While post-composed annotation allows for richer detailed descriptions, it has the drawback that different curators may select different combinations of terms to describe the same phenotype thus increasing variability. A second drawback is that the different component ontologies often diverge, through the natural course of different groups working independently on ontology development, making synchronization a challenge and requiring frequent re-annotation to already curated data. This variability and lack of synchrony has made it difficult to make post-composed based phenotype assertions interoperable between different model organisms and humans. While variability is less of a problem with the pre-composed approach, one drawback is the scale of pre-composed ontologies—groups of terms must be in place for every phenotype and anatomical structure: a 'small heart', 'small pancreas', 'small limb', 'small head' etc. This makes the ontology very large as when almost every anatomy ontology term gives rise to multiple phenotype terms the phenotype ontology must be several times larger than the associated anatomy ontology. This large size is not a major issue as it can be tackled programmatically, both for generation and management. A final challenge shared by both pre- and post-composed approaches is that anatomical structure, and terms commonly used to describe them in the literature, can be very species-specific.

The fact that different research communities use distinct approaches and different ontologies to curate phenotypes has been a challenge for cross-species comparisons. For example, human clinical disease phenotyping is mostly done with the pre-composed Human Phenotype ontology (HPO) (RRID:SCR_006016) [6, 7], the Mammalian Phenotype ontology (MP) (RRID:SCR_004855) [8, 9]. The Zebrafish Information Network (ZFIN) (RRID:SCR_002560) [10] uses a post-composed EQ approach combining the Zebrafish Anatomical Ontology (ZFA) (RRID:SCR_005887) [11], GO and PATO [4]. An approach to harmonize different phenotype ontologies and enable cross-species comparisons was recently established by the Monarch Initiative; a multi-species bioinformatic resource aggregating genotype and phenotype data from multiple model organisms to inform the genetic basis of human disease [12, 13]. The Monarch consortium and their collaborators, which include most of the major model organism knowledgebases, implemented the Unified Phenotype Ontology (uPheno) that uses 'bridging axioms' to equate terms from different species-specific ontologies. Monarch produces a knowledge graph based representation using data and ontologies loaded into a SciGraph database [13] where entities (from different ontologies) are represented by nodes connected by edges representing distinct relationships. uPheno allows connectivity and equivalences between the phenotype ontologies of multiple species. An important component of the uPheno plan was a community wide effort to reconcile and align different ontologies using a standard pre-composed template to maximize interoperability [14].

Leveraging these recent advances, we set out to build a uPheno-compliant *Xenopus* phenotype ontology (XPO) that Xenbase, the *Xenopus* model organism knowledgebase (RRID:SCR_003280) [15, 16], could use to curate phenotype data from *Xenopus* experiments with maximum interoperability to humans and other model organisms. The frog, *Xenopus*, is one of the leading vertebrate model systems and has been a major contributor to understanding fundamental biological processes such as cell division, cell differentiation, morphogenesis, organogenesis and neurobiology. It is also increasingly used to model human disease, particularly congenital conditions [17]. As a tetrapod,

Fisher *et al. BMC Bioinformatics*   (2022) 23:99

Page 4 of 15

Xenopus occupies a key evolutionary niche between fish and mammals. The large, abundant, externally developing *Xenopus* embryos have several unique features that lend themselves to functional genomics and disease modeling including; CRISPR gene editing, antisense morpholino knockdown, transgenics, experimental embryology, and live cell imaging. We estimate that there are ∼4000 publications with phenotype data from *Xenopus* experiments in Xenbase with more papers published every month, but up until recently there were no largescale efforts to curate this data and thus it was largely inaccessible to the wider biomedical community. Below we describe the development and implementation of a fully uPheno compliant XPO, which will facilitate access to *Xenopus* phenotypic data for the biomedical community.
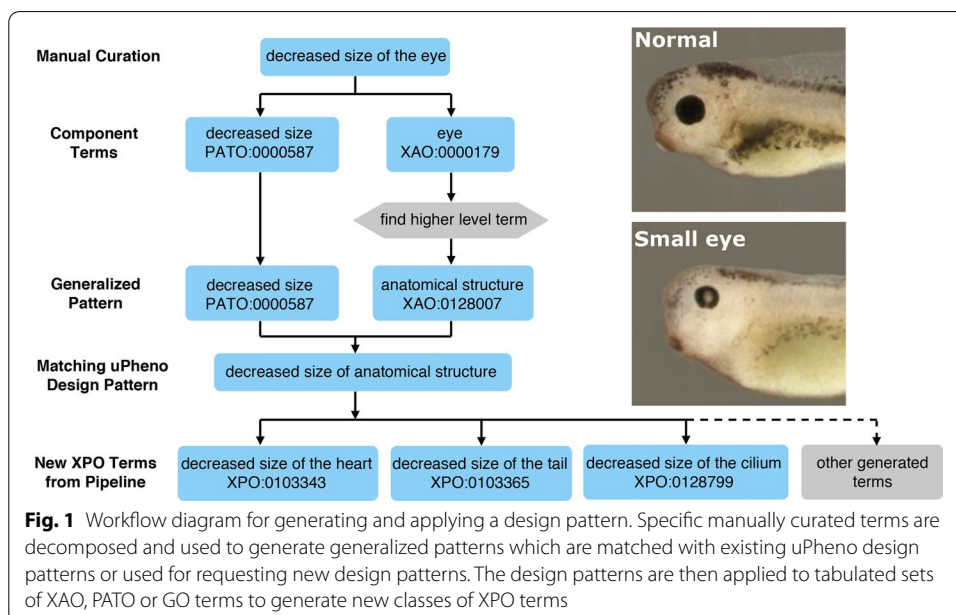
Methods

The XPO release pipeline takes uPheno design pattern files and compiles them into logical definitions and Web Ontology Language (OWL) files [18]. It uses an editors' version of the ontology–which includes ontology annotations such as the ontology definition and provides the root Xenopus phenotype class–and a definition file to merge and save a variety of release files in OBO, OWL, and JSON formats. The Ontology Development Kit (ODK) [19] provides a means for creating and managing the XPO project on GitHub. The current version used in the release pipeline is ODK 1.2.26 (https://github.com/INCATools/ontology-development-kit). It includes Makefiles that specify the release workflow, build ontology imports, run tests, and create quality control (QC) reports. The ODK configures GitHub Actions to check and test any pull requests using the ROBOT tool [20] designed for working with Open Biomedical Ontologies. It also specifies a standard directory layout, documentation, and additional file artifacts that make the XPO consistent with different ontology projects.

The pre-processing pipeline is run after pull requests and changes have been merged into a local copy of the XPO repository. A bash script wraps a call into the ODK and runs an automated pipeline that downloads the current release version of the Xenopus Anatomy Ontology (XAO) and, if necessary, adds abnormal, abnormalMorphology, and other phenotypes for all XAO classes except for classes and branches which have been exclusion listed in a configuration file. In the TSV pattern files, the pipeline creates unique XPO Internationalized Resource Identifiers (IRIs) where they are missing, i.e. where new items have been added, and converts the TSV templates to OWL modules ("DOSDP templates"). The PURLs of design patterns being used for the first time are added to a text file which triggers the pipeline to pull the relevant YAML [21] specification files from the uPheno repository. The release pipeline then is run with another call to ODK and leverages ROBOT to carry out an automated series of tasks, including updates of upstream ontology imports, SPARQL queries for QC violation checks, classification of terms with the ELK reasoner, and the assembly of the main release files and exports. The top level files in the GitHub repository are the release products. Prior to release we inspect the xpo.owl in an ontology editor and xpo.obo in a text editor, e.g. to make sure no terms have unexpectedly disappeared. The latest version of the XPO can always be found at: http://purl.obolibrary.org/obo/xpo.owl, which points to the xpo.owl file in the top-level directory of the GitHub repository (https://github.com/obophenotype/xenopus-phenotype-ontology).

Results

XPO design strategy

We wanted to design the XPO to capture the breadth of experimental phenotypes in the *Xenopus* literature which range from disrupted gastrulation to congenital malformations and limb regeneration in adult frogs. To assess this phenotypic range a team of four expert Xenbase curators annotated phenotypes from 200 *Xenopus* papers with a free form Entity-Quality (EQ) syntax in the Phenote software package [1] using the *Xenopus* Anatomy Ontology (XAO) (RRID:SCR_004337) [22], PATO [4], GO [3], Basic Formal Ontology (BFO) [23] and the Relations Ontology (RO) [5]. This initial set of 1078 EQ phenotype statements served as a seed to develop the XPO. From the EQ definitions we extracted combinations of high level XAO, GO and PATO terms that we wanted to incorporate into the XPO such as "anatomical structure", "cell part", "morphology", "localised" and "anatomical space" and mapped these to existing uPheno patterns. Figure 1 shows an example of how an EQ curation, for an image from Gouignard et al. [24] summarized as 'decreased size of the eye', was used to construct a generalized design pattern of 'decreased size of anatomical structure', which was then applied to generate new 'decreased size' XPO terms for each appropriate anatomical entity in the XAO. This process, identified 13 frequently used patterns (Additional file 1: Table S1) that were then submitted as new class requests to the ongoing phenotype ontologies reconciliation effort (PORE) [25]. For example, several patterns were developed relating to cilium motility in various ciliated tissues. Once these new patterns were validated and added to uPheno by the ontology development team, we implemented them in the XPO. In this way Xenbase curators contributed to the definition of phenotype patterns that are now reused across many other domains. This reiterative approach ensures that the XPO remains in synchrony with uPheno.

Fig. 1 Workflow diagram for generating and applying a design pattern. Specific manually curated terms are decomposed and used to generate generalized patterns which are matched with existing uPheno design patterns or used for requesting new design patterns. The design patterns are then applied to tabulated sets of XAO, PATO or GO terms to generate new classes of XPO terms

Generating XPO terms using uPheno design patterns

To build a uPheno compliant XPO we used standard tools such as the Ontology Development Kit (ODK) [19] and OBO Tool (ROBOT) [20] to generate an ontology in the W3C Web Ontology Language (OWL) format [18], a semantic language designed for complex knowledge and relationships. By using uPheno design patterns as templates we were able to efficiently construct a pre-composed phenotype ontology incorporating terms from existing ontologies, such as the XAO, PATO, and molecular functions and biological processes from GO. uPheno design patterns prescribe a statement syntax which takes variables from a tab separated value (TSV) file containing a table of component terms to produce multiple appropriately formatted terms. Figure 2 shows a conceptual diagram of the pipeline including a partial example of a design pattern YAML file [21]. The pattern pipeline fills in new IRIs in all TSV files corresponding to
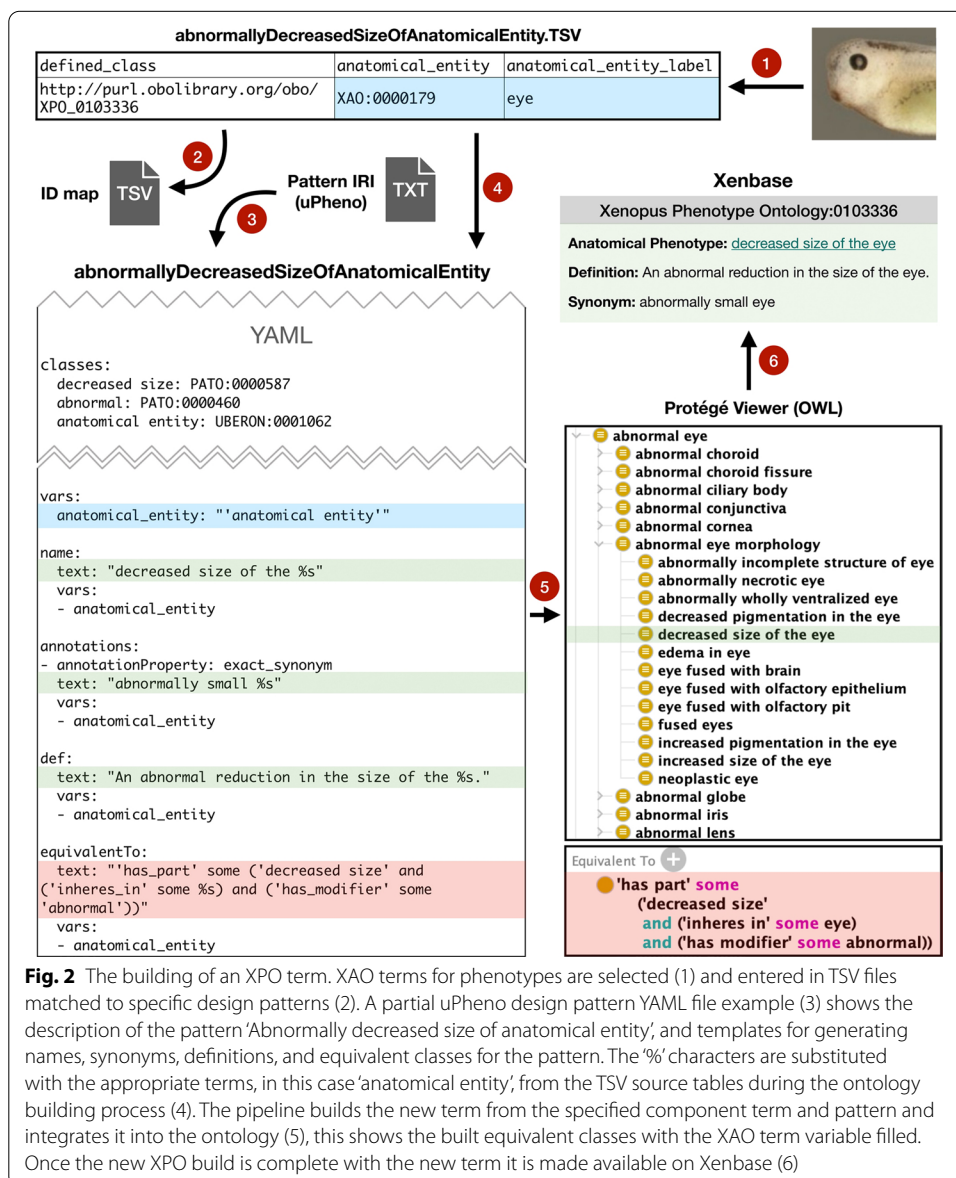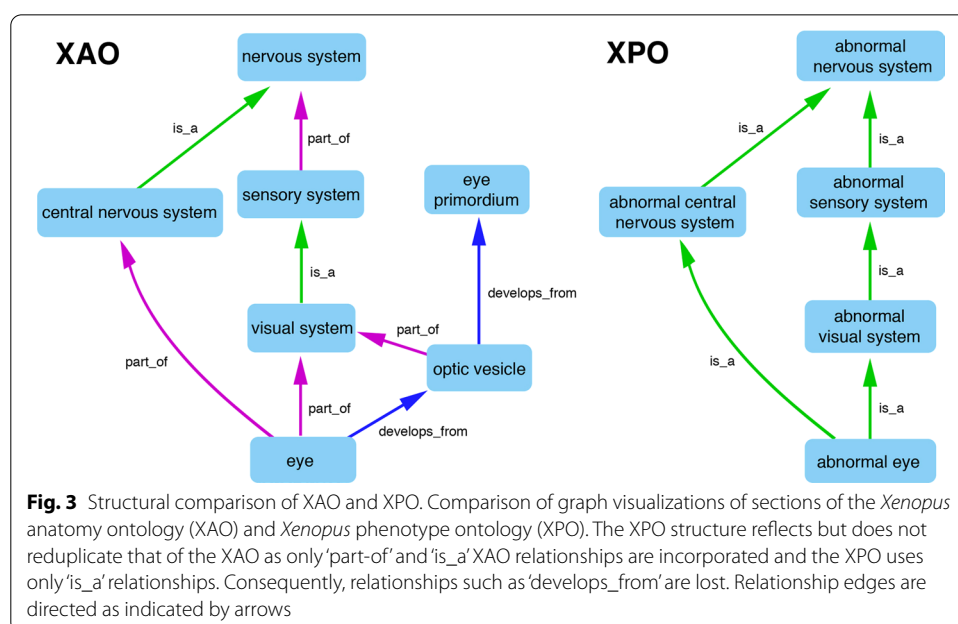

Fig. 2 The building of an XPO term. XAO terms for phenotypes are selected (1) and entered in TSV files matched to specific design patterns (2). A partial uPheno design pattern YAML file example (3) shows the description of the pattern 'Abnormally decreased size of anatomical entity', and templates for generating names, synonyms, definitions, and equivalent classes for the pattern. The '%' characters are substituted with the appropriate terms, in this case 'anatomical entity', from the TSV source tables during the ontology building process (4). The pipeline builds the new term from the specified component term and pattern and integrates it into the ontology (5), this shows the built equivalent classes with the XAO term variable filled. Once the new XPO build is complete with the new term it is made available on Xenbase (6)

the selected terms. A second major step in the workflow, an automated release preparation pipeline, checks for updates to uPheno patterns and ontology, makes subclass assertions and generates uPheno-compliant logical definitions, flags errors such as duplicate equivalent classes and term labels, and builds the OWL files that comprise a new XPO release. Before making the release official, curators can inspect the ontology in an editor such as Protege to ensure that changes appear as expected. Curators may occasionally need to edit ontology metadata such as the ontology description by opening an "editor's" OWL file; otherwise, routine class requests and updates are handled exclusively in the TSV tables and configuration file.

The initial TSV lists for each pattern were generated based on the higher-level ontological classes from our previous review phase by identifying appropriate terms and selecting all their children with specific relationships. To enable more precise control over these automatically generated classes and prevent creation of phenotypes that make little or no biological sense some subclass terms and their children were excluded, for instance XAO terms that were children of 'anatomical space' were excluded from lists for the generic pattern 'biological process in location' for where the processes were the GO terms 'cell population proliferation' or 'apoptotic process', as these cellular processes do not occur in acellular anatomical spaces. Only certain relationships, such as 'is_a' and 'part_of' but not 'develops_from', are used when traversing from higher level to lower level XAO terms when selecting terms to be used in the XPO. 'Develops from' is not used as it would lead to the propagation of phenotypes in developed tissues to their precursor structures, which is not a given. While all 'abnormal eyes' are part of an 'abnormal visual system' an 'abnormal eye' does not necessarily imply that it developed from an 'abnormal optic vesicle'. This leads to the structure of the XPO reflecting but not duplicating the XAO (Fig. 3).

Fig. 3 Structural comparison of XAO and XPO. Comparison of graph visualizations of sections of the *Xenopus* anatomy ontology (XAO) and *Xenopus* phenotype ontology (XPO). The XPO structure reflects but does not reduplicate that of the XAO as only 'part-of' and 'is_a' XAO relationships are incorporated and the XPO uses only 'is_a' relationships. Consequently, relationships such as 'develops_from' are lost. Relationship edges are directed as indicated by arrows

Fisher *et al. BMC Bioinformatics*   (2022) 23:99

Page 8 of 15

The XPO is the only phenotype ontology to date whose classification is purely driven by logical definitions and fully uPheno compliant. Subclass assertions, the defining of hierarchical parent child relationships between terms, a process that is known to be error-prone and incomplete, do not need to be made manually. This significantly reduces maintenance effort and increases interoperability of the XPO. We extended this novel approach to automatically construct phenotype terms from other ontologies, most importantly the XAO but also GO and NBO. For example, instead of having to curate a new phenotype term for "abnormal structure-*X*" whenever a new anatomy term is added to the XAO, the XPO pipeline automatically scans the XAO for new terms which are then used to automatically generate new XPO terms according to the specific predesignated design patterns (Table 1), these 14 standard patterns account for ~96% of terms in the XPO. This significantly reduces the effort in maintaining an up-to-date phenotype ontology and maintains synchrony with the anatomy ontology. We estimate that these patterns are likely to be used for comparatively large and diverse sets of anatomical entities; by applying these patterns to almost the whole XAO and populating the XPO with these classes up front, our hope is to streamline curation efforts by reducing the frequency of new term requests over time.

In the course of developing the XPO pipeline we introduced some novel components to the ODK framework [19] approach including a system for automatically generating phenotypes from component ontologies that keeps them synchronized with the most updated patterns. If a pattern changes (e.g. a definition is updated) it is automatically updated and the ontology stays conformant. For example: the phenotype ontology reconciliation effort consortium recently decided to use the PATO class "mass density" instead of "mass". Xenbase curators were not required to edit the XPO directly to keep it in sync with this decision, it was automatically updated by the system. In many cases in the past ontologies would employ patterns but in different ways, as illustrated by the equivalent classes for the human and mammalian (mouse) "unilateral deafness" phenotypes (Fig. 4). Even the common elements are framed distinctly with the 'unilateral' class

Table 1 Automatically applied class design patterns for new XAO terms (X)

Class pattern
Abnormal X
Absent X
Abnormal morphology of X
Abnormal development of X
Mislocalized X
Necrotic X
Abnormally increased size of X
Abnormally decreased size of X
Abnormally increased number of X
Abnormally decreased number of X
Increased apoptosis in X
Decreased apoptosis in X
Increased cell population proliferation in X
Decreased cell population proliferation in X

HPO: Unilateral deafness
 has part some (
• 'lacking processual parts'
• and (towards some 'sensory perception of sound')
• and 'has modifier' some abnormal
• and 'has modifier' some unilateral)

MP: Unilateral deafness
 has part some (
• 'unilateral'
• and 'inheres in' some MPATH:deafness
• and 'has modifier' some abnormal

Fig. 4 Differences between definitions of equivalent classes. Comparison of the equivalent classes for the 'Unilateral deafness' phenotype class in the human (HPO) and mammalian (MP) phenotype ontologies

being treated as a modifier only in the HPO, although it is still one of the equivalent classes in the MP.

The design patterns, pipeline, and release process guarantee that ontologies developed using this system are always fully interoperable and aligned with ongoing reconciliation efforts. Consequently, the XPO and any other uPheno compliant ontologies should have identically structured equivalent classes varying only in the species-specific anatomy ontology terms used. The current XPO.v1.1 consists of approximately 22,000 terms built from a set of 80 design patterns (Additional file 2: Table S2).

Ongoing XPO maintenance

XPO curators or community members may make requests for new XPO terms by providing the variables specified in the design pattern corresponding to the new term, in the form of IDs and labels from the relevant ontologies, as a GitHub 'issue'. In many cases this requires only a single entity from the XAO, GO, or Neuro Behavior Ontology (NBO) [26], for instance in the "abnormalBehavior" or "edematousAnatomicalEntity" tables. More complex patterns include "in location" and "by type" components that allow the construction of phenotypes such as "abnormal axon regeneration in the optic nerve" or "Y-shaped femur in the regenerating hindlimb" without requiring the limiting and potentially time-consuming process of creating or requesting specific new classes for GO or the XAO. Additionally, TSV files are used to manage obsoleted terms and to specify which terms they have been "replaced by", this information is also handled automatically by the pipeline to update or obsolete existing terms. New design patterns can also be requested but these are subject to the wider PORE review process and may take longer to be incorporated. There is no set release cycle for new versions of the XPO, new releases are produced when the developers consider a significant body of new terms will be generated or there is a curator need for specific terms.

Accessing the XPO

The file structure for generating the XPO is hosted on GitHub (https://github.com/obophenotype/xenopus-phenotype-ontology), as well as scripts and makefiles for building the XPO from source. In addition on the XPO GitHub repository wiki we provide a brief text description of the procedure for adding new terms and running the XPO build pipeline (https://github.com/obophenotype/xenopus-phenotype-ontology/wiki/

Curation-and-processing-pipeline). The XPO v1.1 is available for download on Xenbase in owl (http://ftp.xenbase.org/pub/Ontologies/XPO/XPO_1.1/XPO_1.1.owl) and obo formats (http://ftp.xenbase.org/pub/Ontologies/XPO/XPO_1.1/XPO_1.1.obo) and in the GitHub repository. The XPO is browsable at various online endpoints such as the European Bioinformatic Institute's Ontology Lookup Service (OLS) (https://www.ebi.ac.uk/ols/ontologies/xpo) and Ontobee (http://www.ontobee.org/ontology/xpo). The XPO is licensed under a Creative Commons CC BY 3.0 license (https://creativecommons.org/licenses/by/3.0/). The specification of the XPO in line with the 'minimum information for the reporting of an ontology' (MIRO) guidelines [27] is available in Additional file 3: Table S3.

Application of the XPO for phenotype curation

Xenbase has begun routine curation of published *Xenopus* research articles using the XPO. The curation system allows either direct association with 'target' genes or indirect association through mutant or transgenic lines and reagents with existing gene associations (Fig. 5).

The given example, for an image from Naert et al. [28], shows an experiment associated with two CRISPR targets captured as distinct guide RNA (gRNA) reagents. These gRNAs are associated in the database with the Xenbase genes they target, in this case rbl1 and rb1. After association of the experimental description with an image and an assay type the combined elements, stored as XB-PHENO entities, can then be associated with XPO terms such as the 'neoplastic eye' term for the retinoblastoma phenotype in Fig. 5, or with human disease terms from the Human Disease Ontology (DO) (RRID:SCR_000476) [29]. The use of a controlled vocabulary allows the common variability in author descriptions of phenotypes to be accounted for by curator expertise so that similar or identical phenotypes are all identified with the same term. This allows the common phenotypes to be identified where a simplified text matching approach might fail.
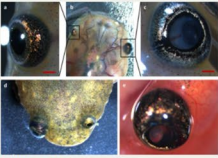
Cross-species comparisons

Curation with the XPO allows us to link Xenopus phenotypes with and those of human, mouse, and others using the uPheno bridging ontology (Fig. 6) but the linkage is currently limited by non uPheno pattern compliant terms in ontologies of other species.

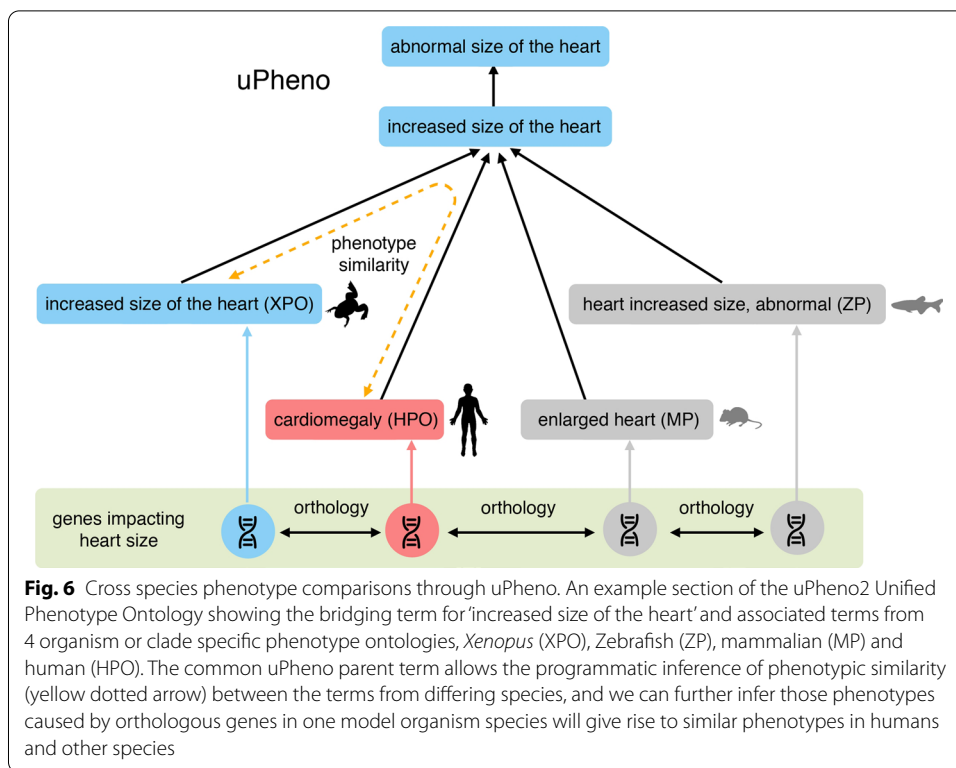**Fig. 5** Phenotype curation using the XPO on Xenbase. A basic phenotype curation in Xenbase. The record has a brief precis of the experiment, reagents, and assay details, including background *Xenopus* strain, for the observed phenotypes as well as disease associations

Fig. 6 Cross species phenotype comparisons through uPheno. An example section of the uPheno2 Unified Phenotype Ontology showing the bridging term for 'increased size of the heart' and associated terms from 4 organism or clade specific phenotype ontologies, *Xenopus* (XPO), Zebrafish (ZP), mammalian (MP) and human (HPO). The common uPheno parent term allows the programmatic inference of phenotypic similarity (yellow dotted arrow) between the terms from differing species, and we can further infer those phenotypes caused by orthologous genes in one model organism species will give rise to similar phenotypes in humans and other species

Mapping between non-compliant terms is not impossible, using logical or lexical mapping approaches [30], but is more challenging and often incorporates some fuzziness. Once various species phenotype data are stored in a common framework such as the Monarch SciGraph database and built using a common syntax such as the design patterns from uPheno the ease of inferring cross-species comparisons should be greatly improved.

Discussion

Over the years a variety of ways of codifying certain phenotypic spectra in *Xenopus* have been put forward. These include the index of axis deficiency (IAD) [31] and the related dorso-anterior index (DAI) [32], which gave numerical values for specific degrees of axis-perturbation, and the widely used Frog Embryo Teratogenesis Assay-Xenopus (FETAX) system [33, 34], which is employed in testing the developmental toxicity of compounds and uses a standardized form to identify malformations in specific embryonic tissues and of specific types such as edema or hemorrhage. None of these existing systems provide a suitable system for general phenotype curation, either through having too narrow a focus, such as the DAI which spans a range of 0 to 10, or too shallow a capture of broader phenotypes, such as in FETAX. The new XPO provides broad coverage, incorporating several basic PATO terms for 96% of terms from the XAO, and depth down to individual cell types and subcellular components. While the XPO provides a crucial resource for internal Xenbase phenotype curation we hope it will also be a resource for researchers as a standard reference set for categorizing phenotypes, in line with this we have already produced curation

for one of the broadest existing phenotypic screens in *X. tropicalis* that described and categorized phenotypes for ~ 136 morpholino knockdowns [35].

By building the XPO based on uPheno design patterns it is consistent with current best practices advocated by the phenotype reconciliation effort consortium to maximize interoperability. The XPO can serve as a model for ongoing efforts to integrate ontology based phenotype curation across different species and can be used as a template and workflow for the development of new phenotype ontologies [36, 37]. Refining and improving such cross-species mappings will require continued ongoing discussion between various model organism knowledgebases.

The design pattern based approach also allows the XPO to be highly responsive with new XAO terms being integrated into the XPO shortly after release. Managing new ontology requests through GitHub allows for transparency and anyone can submit requests allowing the research community to help direct development into areas to benefit active research. Ongoing development is in line with our initial approach of basing our core terms and classes of terms for XPO development on review and curation of existing phenotype papers to reflect the spectrum of *Xenopus* research. This research led approach reduces the likelihood of bloating the ontology as opposed to just taking every uPheno design pattern taking an anatomical term and applying it to all XAO terms and its descendants, even with basic logical restrictions on certain classes of terms as discussed previously this would still lead to rampant term proliferation.

There is still plenty of scope for expansion and refinement of the XPO using this focused approach and Xenbase is in active collaboration with the uPheno team to establish new design patterns that will enable wider curation of Xenopus research. For example, we are assessing the addition of selected GO metabolism terms and cell types into XPO to better accommodate single cell, toxicological, and immunological data. Some more fundamental structural factors under discussion with uPheno for future development are using more relationships in the XAO to inform our XPO classes, such as whether the 'develops from' (RO:0002202) relationship (Fig. 3) can be used in an inverse manner, does abnormal development of an eye primordium imply abnormal eye development and are limits needed on logical propagation of such implied phenotypes?

The increased ability to perform cross-species phenotype comparisons should enhance the utility of *Xenopus* as a disease model [17]. Both the new uPheno compliant XPO and the ongoing work of projects such as Monarch to associate human diseases with Human phenotype ontology (HPO) [6] should help identifying phenotypes associated with human disease associated variants. *Xenopus* provides a rapid and flexible system for studying human sequence variants as the mRNA for a potential causative variant can be directly injected into the developing embryo [38–40] in large numbers allowing a quick survey of phenotypic effects. In addition to these forward genetics approaches, phenotypes derived from perturbing novel or under investigated genes, either by overexpression or knockdown using CRISPR or morpholinos, should be more amenable to identifying equivalent disease associated phenotypes in humans [41].

Fisher *et al. BMC Bioinformatics*   (2022) 23:99

Page 13 of 15

Conclusion

This new *Xenopus* phenotype ontology, along with developments throughout the biocuration community for disease and phenotype curation, will allow *Xenopus* to continue as one of the major model organisms for the study of vertebrate development and human developmental disorders and diseases.

Abbreviations

BFO: Basic formal ontology; DAI: Dorso-anterior index; DO: Human disease ontology; EQ: Entity-quality; FETAX: Frog embryo teratogenesis assay-Xenopus; GO: Gene ontology; HPO: Human phenotype ontology; IAD: Index of axis deficiency; IRI: Internationalized resource identifiers; MP: Mammalian phenotype ontology; NBO: Neuro behavior ontology; OBO: Open biomedical ontologies; ODK: Ontology development kit; OWL: W3C web ontology language; PATO: Phenotype and trait ontology; RO: Relations ontology; TSV: Tab separated values; XAO: Xenopus anatomy ontology; XPO: Xenopus phenotype ontology; ZFA: Zebrafish anatomy and development ontology; ZP: Zebrafish phenotype ontology.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-022-04636-8.

> **Additional file 1: Table S1.** 13 design patterns proposed to PORE.
>
> **Additional file 2: Table S2.** 80 design patterns used in XPO.
>
> **Additional file 3: Table S3.** MIRO report for Xenopus phenotype ontology.

Availability of data and materials

The initial curation datasets used during the current study are available from the corresponding author on reasonable request. The data files and code to build the XPO from source are available from GitHub at https://github.com/obophenotype/xenopus-phenotype-ontology.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing Interests.

Author details

[1]Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [2]Monarch Initiative, London, UK. [3]Semanticly Ltd, London, UK. [4]European Bioinformatics Institute (EMBL-EBI), Hinxton, UK. [5]Department of Biological Science, University of Calgary, Calgary, AB, Canada.

References

1. Washington NL, Haendel MA, Mungall CJ, Ashburner M, Westerfield M, Lewis SE. Linking human diseases to animal models using ontology-based phenotype annotation. PLoS Biol. 2009;7(11):e1000247.
2. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, et al. The OBO foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007;25(11):1251–5.
3. The Gene Ontology. http://purl.obolibrary.org/obo/go.obo. Accessed 26 Aug 2021.
4. PATO—The Phenotype And Trait Ontology. http://purl.obolibrary.org/obo/pato.obo. Accessed 26 Aug 2021.
5. The Relations Ontology. http://purl.obolibrary.org/obo/ro.obo. Accessed 26 Aug 2021.
6. The Human Phenotype Ontology. http://purl.obolibrary.org/obo/hp.owl. Accessed 26 Aug 2021.
7. Robinson PN, Mundlos S. The human phenotype ontology. Clin Genet. 2010;77(6):525–34.
8. Smith CL, Eppig JT. Expanding the mammalian phenotype ontology to support automated exchange of high throughput mouse phenotyping data generated by large-scale mouse knockout screens. J Biomed Semant. 2015;6:11.
9. Smith CL, Goldsmith CA, Eppig JT. The Mammalian phenotype ontology as a tool for annotating, analyzing and comparing phenotypic information. Genome Biol. 2005;6(1):R7.
10. Ruzicka L, Howe DG, Ramachandran S, Toro S, Van Slyke CE, Bradford YM, Eagle A, Fashena D, Frazer K, Kalita P, et al. The Zebrafish information network: new support for non-coding genes, richer gene ontology annotations and the alliance of genome resources. Nucl Acids Res. 2019;47(D1):D867–73.
11. Van Slyke CE, Bradford YM, Westerfield M, Haendel MA. The zebrafish anatomy and stage ontologies: representing the anatomy and development of Danio rerio. J Biomed Semant. 2014;5(1):12.
12. McMurry JA, Kohler S, Washington NL, Balhoff JP, Borromeo C, Brush M, Carbon S, Conlin T, Dunn N, Engelstad M, et al. Navigating the phenotype frontier: the monarch initiative. Genetics. 2016;203(4):1491–5.
13. Shefchek KA, Harris NL, Gargano M, Matentzoglu N, Unni D, Brush M, Keith D, Conlin T, Vasilevsky N, Zhang XA, et al. The monarch initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucl Acids Res. 2020;48(D1):D704–15.
14. Matentzoglu NB, James P, Bello SM, Boerkoel CF, Bradford YM, Carmody LC, Cooper LD, Grove CA, Harris NL, Köhler S, Laporte M-A, Laulederkind SLF, Lee R, Mazandu GK, McMurry JA, Mungall C, Osumi-Sutherland D, Pilgrim C, Rageth K, Robb SMC, Robinson PN, Segerdell E, Thessen A, Vasilevsky N, Zhang XA, Haendel MA. Phenotype ontologies traversing all the organisms (POTATO) workshop aims to reconcile logical definitions across species. 2018.
15. Fortriede JD, Pells TJ, Chu S, Chaturvedi P, Wang D, Fisher ME, James-Zorn C, Wang Y, Nenni MJ, Burns KA, et al. Xenbase: deep integration of GEO & SRA RNA-seq and ChIP-seq data in a model organism database. Nucl Acids Res. 2020;48(D1):D776–82.
16. James-Zorn C, Ponferrada V, Fisher ME, Burns K, Fortriede J, Segerdell E, Karimi K, Lotay V, Wang DZ, Chu S, et al. Navigating Xenbase: an integrated *Xenopus* genomics and gene expression database. Methods Mol Biol. 2018;1757:251–305.
17. Nenni MJ, Fisher ME, James-Zorn C, Pells TJ, Ponferrada V, Chu S, Fortriede JD, Burns KA, Wang Y, Lotay VS, et al. Xenbase: facilitating the use of Xenopus to model human disease. Front Physiol. 2019;10:154.
18. Web Ontology Language (OWL). https://www.w3.org/OWL/. Accessed 29 Oct 2021.
19. Ontology Development Kit. https://doi.org/10.5281/zenodo.5564481. Accessed 29 Oct 2021.
20. Jackson RC, Balhoff JP, Douglass E, Harris NL, Mungall CJ, Overton JA. ROBOT: a tool for automating ontology workflows. BMC Bioinform. 2019;20(1):407.
21. YAML™ Specification Index. https://yaml.org/spec/. Accessed 25 Oct 2021.
22. The Xenopus Anatomy Ontology http://purl.obolibrary.org/obo/xao.obo. Accessed 26 Aug 2021.
23. The Basic Formal Ontology. http://purl.obolibrary.org/obo/bfo.obo. Accessed 26 Aug 2021.
24. Gouignard N, Maccarana M, Strate I, von Stedingk K, Malmstrom A, Pera EM. Musculocontractural Ehlers–Danlos syndrome and neurocristopathies: dermatan sulfate is required for *Xenopus* neural crest cells to migrate and adhere to fibronectin. Dis Models Mech. 2016;9(6):607–20.
25. Phenotype Ontologies Reconciliation Effort. https://github.com/obophenotype/upheno/wiki/Phenotype-Ontologies-Reconciliation-Effort. Accessed 26 Aug 2021.
26. The Neuro Behavior Ontology. http://purl.obolibrary.org/obo/nbo.owl. Accessed 26 Aug 2021.
27. Matentzoglu N, Malone J, Mungall C, Stevens R. MIRO: guidelines for minimum information for the reporting of an ontology. J Biomed Semant. 2018;9(1):6.
28. Naert T, Colpaert R, Van Nieuwenhuysen T, Dimitrakopoulou D, Leoen J, Haustraete J, Boel A, Steyaert W, Lepez T, Deforce D, et al. CRISPR/Cas9 mediated knockout of rb1 and rbl1 leads to rapid and penetrant retinoblastoma development in Xenopus tropicalis. Sci Rep. 2016;6:35264.
29. Bello SM, Shimoyama M, Mitraka E, Laulederkind SJF, Smith CL, Eppig JT, Schriml LM. Disease ontology: improving and unifying disease annotations across species. Dis Models Mech. 2018;11(3):1–9.
30. Oellrich A, Gkoutos GV, Hoehndorf R, Rebholz-Schuhmann D. Quantitative comparison of mapping methods between Human and Mammalian Phenotype Ontology. J Biomed Semant. 2012;3(Suppl 2):S1.
31. Scharf SR, Gerhart JC. Determination of the dorsal-ventral axis in eggs of *Xenopus laevis*: complete rescue of uv-impaired eggs by oblique orientation before first cleavage. Dev Biol. 1980;79(1):181–98.
32. Scharf SR, Gerhart JC. Axis determination in eggs of *Xenopus laevis*: a critical period before first cleavage, identified by the common effects of cold, pressure and ultraviolet irradiation. Dev Biol. 1983;99(1):75–87.
33. Dumont JN, Schultz TW, Buchanan MV, Kao GL. Frog embryo teratogenesis assay Xenopus: FETAX—a short-term assay applicable to complex environmental mixtures. In: Sandhu SS, Lewtas J, Claxton L, Chernoff N, Nesnow S, editors. Symposium on the application of short-term bioassays in the analysis of complex environmental mixtures: III. New York: Springer; 1983.
34. Mouche I, Malesic L, Gillardeaux O. FETAX assay for evaluation of developmental toxicity. Methods Mol Biol. 2017;1641:311–24.

35. Rana AA, Collart C, Gilchrist MJ, Smith JC. Defining synphenotype groups in *Xenopus tropicalis* by use of antisense morpholino oligonucleotides. PLoS Genet. 2006;2(11):e193.
36. Kohler S, Doelken SC, Ruef BJ, Bauer S, Washington N, Westerfield M, Gkoutos G, Schofield P, Smedley D, Lewis SE, et al. Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. F1000Res. 2013;2:30.
37. Kohler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM, et al. The Human Phenotype Ontology in 2021. Nucleic Acids Res. 2020;49:D1207–17.
38. Shah AM, Krohn P, Baxi AB, Tavares ALP, Sullivan CH, Chillakuru YR, Majumdar HD, Neilson KM, Moody SA. Six1 proteins with human branchio-oto-renal mutations differentially affect cranial gene expression and otic development. Dis Models Mech. 2020;13(3):1–14.
39. Li J, Zhang J, Tang W, Mizu RK, Kusumoto H, XiangWei W, Xu Y, Chen W, Amin JB, Hu C, et al. De novo GRIN variants in NMDA receptor M2 channel pore-forming loop are associated with neurological diseases. Hum Mutat. 2019;40(12):2393–413.
40. Ott T, Kaufmann L, Granzow M, Hinderhofer K, Bartram CR, Theiss S, Seitz A, Paramasivam N, Schulz A, Moog U, et al. The frog *Xenopus* as a model to study joubert syndrome: the case of a human patient with compound heterozygous variants in PIBF1. Front Physiol. 2019;10:134.
41. Rosenthal SB, Willsey HR, Xu Y, Mei Y, Dea J, Wang S, Curtis C, Sempou E, Khokha MK, Chi NC, et al. A convergent molecular network underlying autism and congenital heart disease. Cell Syst. 2021;12:1094–107.

Publisher's Note