Universiteit Leiden
The Netherlands

Spatio-temporal framework for integrative analysis of zebrafish development studies

Belmamoune, M.

Citation
Belmamoune, M. (2009, November 17). *Spatio-temporal framework for integrative analysis of zebrafish development studies*. Retrieved from https://hdl.handle.net/1887/14433

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/14433

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 2

DEVELOPMENTAL ANATOMY ONTOLOGY OF ZEBRAFISH: AN INTEGRATIVE SEMANTIC FRAMEWORK

Based on:

M. Belmamoune and F.J. Verbeek.
Developmental Anatomy Ontology of Zebrafish: an Integrative semantic framework.
Journal of Integrative Bioinformatics, 4(3):65, 2007.

Partially published in:

Y. Bei, M. Belmamoune and F. J. Verbeek
Ontology and image semantics in multimodal imaging: submission and retrieval
Proc. of SPIE Internet Imaging VII, Vol. 6061, 60610C1 C12, 2006.

Abstract

Integration of information is quintessential to make use of the wealth of bioinformatics resources. One aspect of integration is to make databases interoperable through well annotated information. With new databases one strives to store complementary information and such results in collections of heterogeneous information systems. Concepts in these databases need to be connected and ontologies typically are providing a common terminology to share information among different resources.

Our focus of research is the zebrafish and we have developed several information systems in which ontologies are crucial. Pivot is an ontology describing the developmental anatomy, referred to as the Developmental Anatomy Ontology of Zebrafish (DAOZ). The anatomical and temporal concepts are provided by the zebrafish information network (ZFIN) and proven within the research community. We have constructed a 3D digital atlas of zebrafish development based on histology. The atlas is a series of volumetric models and in each instance every volume element is assigned to an anatomical term. Complementing the atlas we developed an information system with 3D patterns of gene expression in zebrafish development based on marker genes. The spatial and temporal annotations to these 3D images are drawn from the ontology that we have designed. In its design the DAOZ ontology is structured as a Directed Acyclic Graph (DAG). Such is required to find unique concept paths and prevent self referencing.

As we need to address the ontology in a direct manner, the DAG structure is transferred to a database. The database is used in the integration of our databases that share concepts at different levels of aggregation. In order to make sure that sufficient levels of aggregation for applications in mind are present, the original vocabulary was enriched with more relations and concepts. Both databases can now be addressed with the same unique terms and co-occurrence and co-expression of genes can be readily extracted from the databases. Integration can be further extended to the ZFIN resource and also by

including ontologies that relate to gene/gene expression (e.g. Gene Ontology). In this manner, interoperable information retrieval from heterogeneous databases can be accomplished. This greatly facilitates processing complex information and retrieving relations in the data through machine learning approaches.

2.1 Introduction

In the life sciences, data integration is one of the most challenging problems that bioinformatics is facing. In extending on new research results researchers in the life sciences have to interpret many different types of information from a variety of biological resources. Unfortunately, this information is not easy to identify and access, one of the reasons can be attributed to the semantic heterogeneity and data formats used by the underlying systems.

In this chapter, we present our approach to take up the challenge of data integration. The key is to describe and manage biological concepts into an integrated framework, leading to improved cooperation and thereby increasing scientific benefit (Baldock and Burger, 2005). In our work, we focus on the integration of data associated with the zebrafish model organism. The zebrafish (Danio rerio) is an important model organism in developmental and molecular genetics in the context of fundamental as well as disease studies. In zebrafish, experiments have produced a considerable range and huge amount of data. This fact in itself has been acknowledged by the zebrafish community and a dedicated resource, i.e. Zebrafish Information Network (ZFIN; http://zfin.org), is developed and maintained.

In the past years we have studied zebrafish development and in support of our research we have developed two important information systems. The first system is the 3D atlas of zebrafish development (3D atlas, in short); a digital atlas consisting of virtual models of *standard* zebrafish embryos at different but canonical stages of development (Verbeek et al, 1999, 2000 and 2002). The second is the Gene Expression Management System (GEMS) (Belmamoune and Verbeek, 2006). This system complements the 3D atlas by a collection of 3D patterns of gene expression of a broad range of marker genes.

The 3D atlas is the pivot in our work on developing a spatio-temporal framework for the zebrafish development; it serves as a reference for data submission and retrieval. A canonical number of developmental stages of the zebrafish are completely described as volumetric models in which every volume element is attributed to an anatomical structure. The atlas is built from serial sections portraying standard histology (Verbeek, et al, 2000 and 2002).

The GEMS is a database system for storage and retrieval of 3D spatio-temporal gene expression patterns in zebrafish including mechanisms for linking and mining. Detailed knowledge of both spatial and temporal expression patterns of genes is an important step towards analysis and understanding of complex networks governing changes during embryonic development (Meuleman et al, 2006). In our case, spatio-temporal gene expression patterns are generated through Fluorescent *In Situ* Hybridization (FISH) and whole-mount imaging (Welten et al, 2006) using the confocal laser scanning microscope (CLSM) resulting in 3D images.

For management, presentation and interoperability of the 3D images contained in the 3D atlas and GEMS, methodologies for integration need to be developed. Key is to be able come up with precise search phrases. In general, this problem is observed in an annotation phase where metadata is added to describe an object. If this, is not dealt with thoroughly, managing, mining and reasoning about information from databases will be seriously hampered. Thus, a common terminology for metadata is required. This problem is often solved with a controlled vocabulary, a series of unconnected standard concepts that is composed within a (research) community. Controlled vocabularies, however, have little to offer when it comes to reasoning by combining knowledge. It makes more sense to create agents that convey concept models with rich semantics. Ontologies are in the right position to address these issues. We have defined an approach for the annotation of our 3D images with a domain-specific ontology that implies data integration. To this end

we developed the Developmental Anatomy Ontology of the Zebrafish (DAOZ), a task-oriented ontology for annotation, retrieval and integration.

In life sciences quite a few ontologies have been developed in the model organism community. In parallel to these, the gene ontology (GO; http://www.geneontology.org/), supporting the annotation of attributes of gene products, was developed. Many of these ontologies are available from the Open Biological Ontologies resource (OBO; http://obo.sourceforge.net/) including comprehensive developmental and anatomical ontologies for many different model organisms as "Drosophila", "Arabidopsis thaliana", "Mouse" as well as an ontology for zebrafish development; i.e., the Zebrafish Anatomy Ontology (ZAO) (Sprague et al, 2006).

Our approach for handling developmental anatomy of zebrafish does not derogate the ZAO. It rather extends the ZAO with new some concepts and relationships. The DAOZ aims to provide conventions and a commonly accepted structured set of terms for annotating our research data; i.e., 3D images of *in situ* gene expression patterns. The DAOZ concepts and relationships have to supplement our 3D images with a structured annotation which is quintessential for data retrieval and mining. As a result, these annotations will enable additional comprehensive analysis of gene expression patterns during development.

Similar to ZAO, we initiated with standard anatomical vocabulary adapted from the staging series of Kimmel et al (Kimmel et al, 1995) as provided by ZFIN. The ZAO consists of two concepts types, i.e. anatomical structures and developmental stages. Anatomical structures are linked to developmental stages. In the temporal sense, each anatomical structure is defined within a time frame of start and end stage of development; this time frame records an anatomical structure as it appears and disappears during development. Anatomical structures can have relationships to each other in the ontology according to the following relationships: *is_a*, *part_of* and *develops_from*.

In the context of our work, the classes and relationships that the ZAO encapsulates are judged not sufficient to facilitate annotation, reasoning and analysis of our 3D images. The ZAO concepts and relationships limit the options for describing the inter- and intra-relationships of anatomical structures. This limitation of concepts and properties limits their use for annotation and comparative anatomical analyses. To that end, the original vocabulary has been adapted to our requirements and enriched with additional concepts and relationships. The new concepts and relationships are intended to enable descriptions of the anatomical structures in accordance with their spatial location and functional system. These concepts and their associated relations will help to structure the annotations and in that manner enabling to analyze the gene expression patterns in larger units. This is especially useful for reasoning with and mining of the data.

Similar to other ontologies, the DAOZ consists of concepts and a set of relationships. The DAOZ is organized as a directed acyclic graph (DAG); such is required to find unique concepts paths and to prevent self referencing. The nodes in the graph represent concepts and the edges joining the nodes represent relationships. Combining these relationships facilitates knowledge extraction and presentation. An important reason for using the DAOZ in annotation, apart from the consistency in the terminology for integration, is the structure in the concepts and the relations between the concepts. The relationships are intended to support retrieval of information and allow interpreting several gene expression patterns. Combining relationships also allows interpreting several gene expression patterns and obtaining information on co-localization and co-expression of genes within a common spatio-temporal framework. In this manner it can be possible to disclose “new” relations between genes.

The DAOZ incorporates terminology of anatomical structures and developmental stages identical to the ZAO. The developmental stages are the temporal concepts by which anatomical structures are organized according to appearance and disappearance during the development. In DAOZ we subsequently augment the anatomical terms conceptual

schema with additional top level concepts i.e. functional system and spatial location aspects. The concepts *functional system* and *spatial location* provide these supplementary levels of abstraction extending the data semantic and subsequently encapsulating its functional and spatial conceptual model. These concepts enable to structure anatomical terms in units using a functional system and spatial location. Searching in the ontology for concepts to annotate data is, therefore, facilitated. The annotated images are structured in the same way as their ontological metadata. This structure enables to process the 3D images in larger units which is considered useful in reasoning and mining.

To manage and use the DAOZ in a context of integration, we designed and built an ontology database. In this chapter, this database is further referred to as DAOZ. It was considered necessary to facilitate data annotation in both the 3D atlas and GEMS. Our task-oriented ontology enables interoperability and data sharing between our information system databases while cross-referencing to the ZAO is provided. Consequently, DAOZ permits integration of different information in the context of the embryonic development of the zebrafish, facilitating data analysis and knowledge extraction for presentation. The DAOZ is accessible through a user-friendly java applet.

The remaining part of this chapter is structured as follows: section 2 contains a detailed description of the adapted methods to develop the DAOZ. In section 3 conclusions and discussions are presented. Finally, section 4 describes our future work.

2.2 Methods

The major function of the DAOZ is to provide conventions and a commonly accepted structured set of terms for annotating research data; therefore, we started with the 44 staging series provided by ZFIN. This anatomical nomenclature is understandable and used by the research community and thus establishes an ideal starting point for an integrative terminology between researchers.

In this section we will describe the framework for the development of DAOZ, including conceptualization of the ontological model, relationships specification, knowledge acquisition, formal description and the subsequent choices of implementation, presentation and integration tools.

2.2.1 Conceptualization

The conceptualization phase involves identification of the key concepts in the ontology. First, we considered the anatomical structures as extracted from the staging series as our primary concepts. Second, we use temporal concepts i.e. development stages, to define anatomical terms within a range of developmental stages. For our research, however, we required an ontology that embodies more information about anatomical structures at varying degrees of granularity. Different levels of granularity enable organization of anatomical structures in units. Such organization permits integration of concepts and the objects that they describe at various levels of resolution. For this purpose, each of the anatomical terms is being evaluated and a number of paths to a certain term have been conceptualized. Two additional concepts were specified. First, specialization of functional system concepts that describe anatomical structures in relation to their functionality; e.g. 'eye' is described as a member of a functional system: 'the visual system'. Second, the spatial location has been conceptualized to organize anatomical structures within a common spatial framework. This conceptualization describes the location of each anatomical domain; e.g. 'eye' could be described by its location in the head region. These two concepts enable to capture function and location of an anatomical structure and, as such, provide extra levels of representation for both anatomical structures as well as for our annotated images.

We further note that the scope of the ontological concepts can always be extended by adding new concepts as well as new granularities.

2.2.2 Relationships specification

We start by two hierarchical relations that were specified to describe the relationships between the various DAOZ concepts: generalization, i.e., '*is_a'* relationship and aggregation, i.e., '*part_of*' relationships (Patrick et al, 2006). The *is_a* relation specifies a generalization hierarchy between a child and its parent; e.g. 'somite 5'*is_a* 'stage of development'. With this relation a child term is linked to a broader concept. The *is_a* relationship is characterized by the fact that each child term has a transitive relationship with its parents and children, that is, properties are inherited from parents to children downstream the hierarchy, but separate properties attributed to a child term are not propagated upstream the hierarchy.

The *part_of* relationship specifies an aggregation; the idea of this relation is that individual parts are brought together into a hierarchy to construct a more generic concept. In DAOZ, we used the *part_of* relationship in two different ways. (1) "*part_of*" is used to link entities of spatial locations, functional systems or temporal concepts; in this case it does not take time constraint into consideration. For example, it always holds that 'central nervous system' is *part_of* 'nervous system'. (2) The parenthood of an anatomical structure may change over time during development (cf. Figure 1). Therefore, the *part_of* relation has been modified to incorporate temporal arguments when invoked in linking anatomical structures with each other. For example at stage '75% epiboly' (time 1) 'the presumptive brain' is *part_of* 'the ectoderm', while at stage '1 somite' (time 2) 'the presumptive brain' is *part_of* 'the presumptive central nervous system'. In both case (1) and (2) of using '*part_of*" it concerns a transitive relationship between parent and children. Such transitivity is for example expressed in a one day old zebrafish embryo where the 'retina' is *part_of* 'optic vesicle' and 'optic vesicle is *part_of* 'eye' consequently 'retina'is also *part_of* 'eye'

In order to describe anatomical structures with properties associated with spatial location, functional system and temporal conceps, we specified four associative relationships: i.e., the *located_at*, *belongs_to*, *starts_development_at* and *ends_development_at* relationships. These relationships are used to describe an anatomical term with its spatial location, functional system and developmental stages respectively. We defined each anatomical structure within a range of the appropriate developmental stages. To that end, temporal relations like *starts_development_at* and *ends_development_at* have been defined to specify time-point at which an anatomical structure appears and disappears from the process of development, respectively. Additionally, we exploit these temporal relationships to code the chronological lineage of anatomical structures during development. An anatomical structure may have several anatomical parents during its lifespan (cf. Figure 1) and therefore we coded the chronological lineage progress of each anatomical structure during its occurrence. Consequently, each anatomical term has been linked to a stage of development when it appears the first time as well as each time when its parent changes. Tracking the chronological changes over time allows following the lineage path of anatomical structures. Moreover, it enables additional reasoning about anatomical structures as well as the objects they describe.

The *part_of* relationship links two anatomical structures with each other; it attributes a specific spatial description at a fine level of granularity. We introduced the *located_at* relationship to associate anatomical terms with a spatial description at a gross level of granularity. As such each anatomical structure is associated with a spatial location concept allowing for divide and conquest strategies. For example, specifically 'retina' is *part_of* 'eye' but more generally, retina could be described by its location in head: 'retina' *located_at* 'head'. Finally, the *belongs_to* relationship is used to associate an anatomical structure with a functional system; e.g. 'retina' *belongs_to* 'visual system'.

The associative relationships also imply inheritance, so that any attribute associated with a concept describing an anatomical structure is propagated downstream by this structure;

e.g. ‘brain’ *belongs_to* ‘the central nervous system’ and ‘the central nervous system’ is *part_of*’ ‘the nervous system’ then ‘brain’ *belongs_to* ‘the nervous system’ too.

The associative relationships have been specified in order to describe properties associated with various anatomical concepts. Furthermore, the aggregation (*part_of*), generalization (*is_a*) and the associatives relationships are binary relationships that imply irreflexivity i.e. no term has a relationship with itself; and asymmetry i.e. if ‘retina’ is *part_of* ‘optic vesicle’ then ‘optic vesicle’ is not *part_of* ‘retina’ (cf. 2.4.2), this corresponds to a DAG.

The aggregation, generalization as well as the associative relationships aims to capture the form and the dynamic development of an anatomical structure in addition to its location and functional system.

Using DAOZ in image annotation implies that these images could later be accessed from different perspectives, amongst other things; using the anatomical structure name and also the characteristics that this structure may have: i.e. developmental stage, spatial location and functional system. Some users would use the precise term, e.g. ‘diencephalon, whereas others would use a less specific terms such as ‘brain’, ‘head’ or ‘nervous system’ to retrieve the images. Therefore, the DAOZ structure enables users to search for large data units from general concepts e.g. brain, head, and central nervous system or specifically for records from an anatomical structure name e.g. ‘diencephalon (cf. Figure 2).

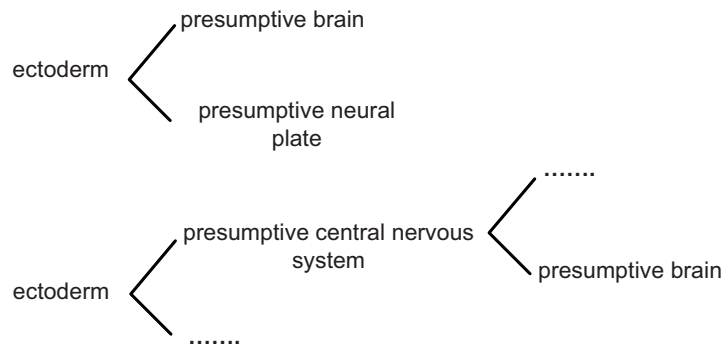

Figure 1: At '75% epiboly' is the presumptive brain part of the ectoderm, while at stage '1 somite' it becomes part of the presumptive central nervous system.

2.2.3 Knowledge acquisition

We start by the anatomical and temporal concepts as well as their relationships. The anatomical structures and stages of development nomenclature were extracted from the staging series. Information describing anatomical structures by their relationship *part_of*, *starts_development_at* and *ends_development_at*, was also extracted from the staging series. The concepts of spatial location and functional system were defined in close collaboration with domain experts. With the help of experts we established a list of attributes for the spatial locations and their relationships with anatomical structures. Concerning functional system attributes and their relationships with the anatomical structures, these have been extracted from the staging series as well as defined from both literature and domain experts. For correctness, the ontology was verified extensively.

2.2.4 Formal description

To give a more precise description of the ontology semantics, we define the concept of order (cf. 2.4.1). The concept of order is used to specify how to line up the ontology elements. Furthermore, we use 9 axioms to formalize the current representation of the DAOZ. These axioms are required as rules to check for the consistency of the ontology upon changes; as such these rules can be integrated in automated agents for ontology update (cf. 2.5).

The DAOZ consists of concepts and relationships that are organized as a DAG structure (cf. axioms 1; figure 2). In the DAG, nodes (concepts; cf. axiom 2) are linked by directed edges (relationships; cf. axiom 3). All relations imply asymmetry (cf. axioms 4) and irreflexibilty (cf. axiom 5). The *part_of* and *is_a* relationships are defined to link only attributes of the same concept type (cf. axioms 6) which means that two different attributes of different concept types could never be linked by a relationship like aggregation (*part_of*) or generalization (*is_a*). The *part_of* relationship has been modified to include time arguments in its usage to link anatomical structures concepts. (cf. axiom 7).

In a DAG each term could be linked to several parents. Therefore, each anatomical structure could be linked to other concept types thereby having more than one occurrence in the hierarchy. Anatomical structures could be associated to spatial locations, functional systems and developmental stages using the *located_at*, *belongs_to*, *starts_development_at* and *ends_development_at* relations; respectively (cf axiom 8, 9).

Definition for order in ontology

A partial order on a set S is a binary relation $< \subseteq S \times S$:

1. $\forall\, d \in S$, not $d < d$ ($<$ is irreflexive).

2. $\forall\ d_1, d_2, d_3 \in S$, if $d_1 < d_2$ and $d_2 < d_3$, then $d_1 < d_3$ (< is transitive).

Axioms underlying DAOZ

1. DAOZ is an ontology having a DAG structure.
2. A DAG G consists of two components: $G = SN, SE$ with SN is the set of nodes of G and SE its set of edges ($SE \subseteq SN \times SN$), such that for no node $n \in SN$, there are edges in SE forming a path from n to n.
3. SN consists of four mutually disjoint subsets: $SN = SA, ST, SL, SFs$. Here SA is the set of anatomical term concepts, ST is the set of temporal concepts a.k.a. developmental stages, SL is the set of spatial locations and SFs is the set of functional systems.
4. SE consists of 6 types of edges a.k.a. relationships, where $SE = is_a \cup part_of \cup belongs_to \cup starts_development_at \cup ends_development_at \cup located_at$.
    a. $\forall\ n_1, n_2 \in SN$ and $e \in SE$ if $n_1\ e\ n_2$, then never $n_2\ e\ n_1$. This means that all relations imply asymmetry. For example: if 'optic vesicle' is *part_of* 'eye' then never 'eye' is *part_of* 'optic vesicle'
    b. $\forall\ n \in SN$ and $e \in SE$ then never $n\ e\ n$. This means that all relations imply irreflexibility such that no concept has a relationship with itself.
5. $\forall\ n_1, n_2 \in SN_1$ with $SN_1 = SA, ST, SL$ or SFs (n_1 and n_2 are two concepts of the same subset) if $n_1\ e\ n_2$ with $e \in SE$ then $e \in part_of \vee e \in is_a$.

This means that the *part_of* and *is_a* are the only relationships linking two concepts of the same type (implying that an ordering between these exists). Consider two functional system concepts: the nervous system and the central nervous system; they only should be linked by the *part_of* relation such that 'the central nervous system' is *part_of* 'nervous system'.

6. $\forall\ n_1, n_2 \in SA$, if $n_1\ e\ n_2$ and $e \in SE \wedge e \in part_of$ then $\exists t \in ST$ such that $n_1\ e'\ t$ with $e' \in starts_development_at$.

If there is a *part_of* relation between two anatomical structures we need to incorporate the time constraint since parenthood of anatomical structure may change over time during development.

7. Let SN_1, SN_2 be *SA, ST, SL* or *SFs* such that $SN_1 \neq SN_2$. $\forall\ n_1 \in SN_1$ if $\exists n_2 \in SN_2$ such that $n_1\ e\ n_2$, where $e \in SE \wedge e \notin part_of' \wedge e \notin is_a$.

This implies that the aggregation (*part_of*) and generalization (*is_a*) relations do not link concept types with other concept types. Thus an anatomical term can be linked to another concept type using only one of the associative relationships. For example, the only relation that links 'head' (a spatial location concept) and 'eye' (an anatomical structure) is the *located_at* relationship.

8. $\forall n_1 \in ST, SL$ or SFs and $n_2 \in SA$, $\neg\exists\ e \in SE$ such that $n_1\ e\ n_2$.

Any anatomical term concept can be linked to another concept type using one of the associative relations. But there is no relation that links both concepts the other way around. The relations (edges) are always directed. For example we have 'eye' is *located_at* 'head' but never 'head' is *located_at* 'eye'.

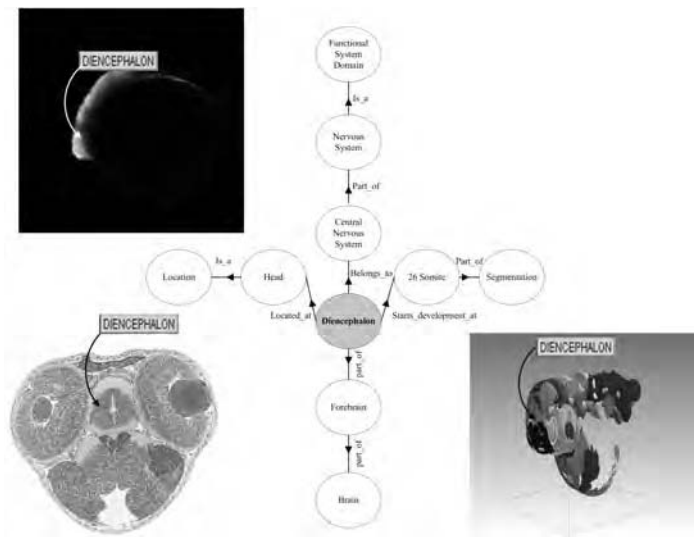

Figure 2: The diencephalon hierarchical organization to show the DAG structure of the anatomy ontology. This structure is inherited by the annotated images, e.g. top left: *msxb* gene expression pattern in a 24 hours post fertilization (hpf) zebrafish embryo, 2D projection of a 3D CLSM image. 3D model from the atlas (lower left: 2D view; lower right: 3D view of a 48 hpf. zebrafish embryo).

2.3 Implementation

To date, the most common procedure for constructing ontologies is by using tools such as DAG-Edit (http://amigo.geneontology.org/dev/java/dagedit/docs/index.html) or Protégé (http://protege.stanford.edu/). Using these tools one starts with a root term and continues adding sub-terms via connecting relationships until the ontology appears to be complete (Bard et al, 2005). In the context of our work however, we considered this an inefficient procedure. First, the DAOZ has a complex data structure with a wide range of terms and relationships, thus adding term by term will be laborious. Second, the specific aim of the DAOZ is to derive the annotation for data within other database resources. The use of the anatomy ontology in this context requires a well-designed and well-defined format that

could be easily linked to other systems and should enable complex queries to be performed to facilitate data extraction for annotation. The ontology format also should provide sufficient flexibility to permit regular updating without a need to modify the hierarchy. We therefore concluded that the anatomy ontology should be stored directly in a database, i.e. the DAOZ database.

The design of the DAOZ as DAG with a set of concepts and binary irreflexive relationships was translated to a database (cf. Figure 3). For each concept type and relationship separate tables have been designed and we assigned to each concept a unique identifier. The DAOZ database is currently implemented using the MySQL database management system.

The specific aim of the DAOZ database is to provide a common semantic framework for the annotation of our data. Therefore, it is directly linked to the 3D atlas and the GEMS to offer a common terminology for spatio-temporal data annotation in these systems. Both databases can be addressed with the same unique terms; as direct result, the 3D patterns of gene expression of the GEMS are spatially mapped onto the 3D atlas and vice versa (Belmamoune et al, 2006). Moreover, using terms from the DAOZ to annotate our biological objects means that the latter will inherit all characteristics and relationships that their annotations might embody. Henceforth, data is hierarchically organized exactly as their *ontological* metadata which is quintessential for retrieval, reasoning and mining (cf. Figure 2). Therefore, 3D images could be retrieved by anatomical structure name, as well as spatial, functional and temporal characteristics of an anatomical structure. To increase search result precision, combinatorial relationships could also be performed. For example 3D gene expression patterns annotated with DAOZ terms could be retrieved by queries in the form of “what patterns are expressed in location X” or “what patterns are expressed at time X in structures *part_of* Y”.

An ontology is never complete as knowledge progresses continuously. The organization of the DAOZ ontological concepts into a database enables updating without altering the ontology hierarchy. The actual anatomical structures of ZFIN are subject to a constant update by a consortium of researchers. We are aware that the DAOZ as well has to be validated constantly against the ZFIN nomenclature in order to improve its comprehensibility and accuracy. To this end, we developed a number of agents to maintain and update the DAOZ on the fly.

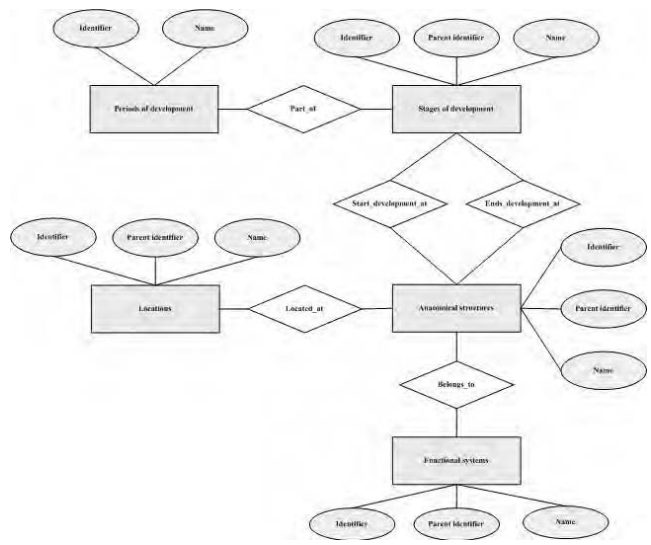

Figure 3: The entity-relationship diagram illustrates the logical structure of the DAOZ database.

2.3.1 Standalone presentation of DAOZ

In order to access the ontology, we have developed a browser: i.e., the 'AnatomyOntology'. The 'AnatomyOntology' is a java applet connected to the ontology database. The applet has been developed to enable navigation and querying anatomical terms through a pre-defined query interface (cf. Figure 4). The applet offers reasoning

possibilities; it provides users with various inference abilities to deduce implicit knowledge from the explicit represented data. The "AnatomyOntology' applet is available online (http://bio imaging.liacs.nl/liacsontology.html). In addition to the applet, on the level of database administration there is always the possibility for free-form SQL queries.

From the DAOZ database, the ontological concepts could always be represented in several common formats such as GO flat file, OBO as well as XML/RDF and OWL. To generate the DAOZ in an OBO format an additional java application, the 'OntologyGenerator', has been designed and developed. As a result, anatomical terms as present in the OBO flat file could be loaded and handled by the DAG-Edit module which offers an additional means of visualization of the data organization.

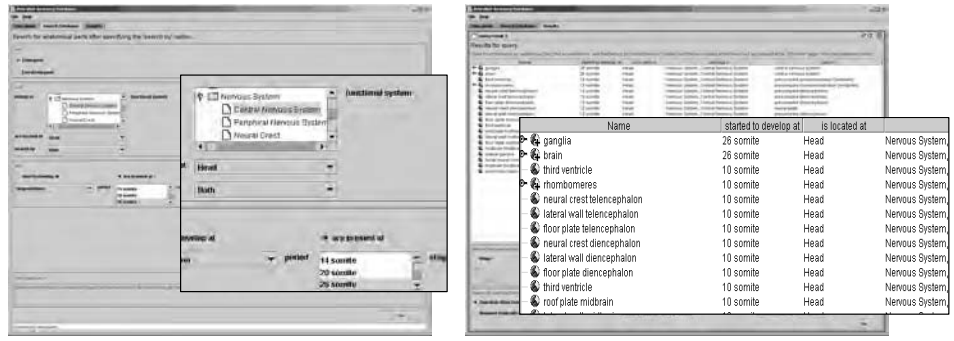

Figure 4: (Left) The applet to query the ontology database. Through this applet users are able to construct a query and submit it to the database to generate on the fly a search result. In this example we constructed the following query: 'search for all anatomical tissues present at '26 somite', belong to the central nervous system and located in head'. (Right) The result screen shows the query result with anatomical structures and their relationships.

2.3.2 Integration with other resources

The DAOZ terminology is used to annotate objects in both the 3D Atlas and the GEMS. Both databases can now be addressed with the same unique concepts and co-occurrence and co-expression of genes can be readily extracted from the databases. Another important requirement for DAOZ is to establish interoperability with other biological resources; ZFIN in particular. Anatomical terms of the DAOZ are identical to those present in ZAO; the zebrafish community ontology (ZFIN). Therefore, an object annotated with DOAZ ontological concepts can be linked straightforwardly to ZFIN which is interconnected with other database resources such as GO and the National Center for Biotechnology Information (NCBI). This means that through ZFIN, objects in our databases are integrated with others. Integration with resources such as GO and NCBI, enables our data to be presented into a large integrated research network.

GO is developed by the gene ontology consortium, and is an evolving structured and standardized vocabulary of nearly 16,000 terms in the domain of biological function (Camon et al, 2004)). GO is widely used for annotation of entries in biological-databases and in biomedical research in general.

NCBI provides an integrated approach to the use of gene and protein sequence information, the scientific literature (MEDLINE), molecular structures, and related resources, in biomedicine. Cross-references of our information systems with, but not restricted to, GO and NCBI implies integration with a wealth of bioinformatics databases leading to an increase of scientific benefit of our data.

2.4 Conclusion and Discussion

We have developed an ontology that describes the zebrafish anatomy during development based on a vocabulary established and approved by the zebrafish community. The

ontology uses several concepts and relationships for anatomical structures description which attribute numerous levels of representation. Specification of concepts and relationships has been achieved in close collaboration with experts in the field of embryology and developmental biology. As a result, the ontology provides an approved specification of domain information representing consensual agreement on concepts and relationships. Moreover, our relationships have been formally defined in order to give them uniform definitions to improve ontological consistency and to approach a maximum consistency with other ontologies; the Relation Ontology (RO) (Smith et al, 2005) especially, as it provides additional tools for relation consistency.

DAOZ is a task-oriented ontology that has been designed to annotate biological data such as 3D images of patterns of gene expression and 3D models of zebrafish embryos: i.e. the typical data in our information systems (http://bio-imaging.liacs.nl/atlasbrowserstart.html and http://bio-imaging.liacs.nl/gems/) (Bei et al, 2006). We considered it a crucial step to our efforts to implement the ontology into a well structured database that could easily be linked to other databases for data annotation. The ontology database is how we use DAOZ in applications. The structure of the ontology database is derived from the ontology DAG representation. In this database, anatomical concepts are described by unique identifiers, their anatomical, temporal, spatial and functional properties. The ontology database holds information about anatomical structures at varying degrees of granularity which enables concepts integration and descriptions at different levels of resolution; therefore complex queries could be performed against the ontological concepts to annotate data of the 3D atlas and the 3D patterns of gene expression. Moreover, powerful and complex search queries against the annotated data can be performed. The ontology is made available through a user-friendly web interface.

The DAOZ ontological concepts enable to group the annotated data in larger units. For example, the organization of spatio-temporal images with DAOZ concepts allows retrieval and integration of the relevant "*in situ*" patterns as well as obtaining information

on co-localization and co-expression of genes. This feature is very important for reasoning and mining in such data.

The DAOZ provides a common semantic framework for gene expression and phenotype annotation thus providing an integrative framework between these two types of data usually employed to study and analyze development. DAOZ improves integration and data sharing between our information systems and ZFIN as well as cross-references to other external resources, i.e. not species specific, such as GO and NCBI.

2.5 Future work

An ontology provides the conceptual framework that is used to capture knowledge in a specific domain. DAOZ concepts enable anatomical terms representation at different level of abstraction with a complex data structure. The anatomical structures are queried through a pre-defined query interface: the "AnantomyOntlogy" browser applet. This applet offers a 2D representation of the hierarchical data structure of the DAOZ. Allowing possibility of free queries as well as enabling better visualization and understanding of the ontology components and their relationships, an new improved interface to the ontology database is the route to take. Currently, we are working on the release of an interface that supports free search and allows visualization of ontological concepts and their relationships using 3D visualization. This interface is a java applet that offers a dynamic interaction with the ontology in a 3D space which will give users new insights in ontological data.

The actual ontology satisfies our requirements. However, an ontology is never complete; it can always be extended with new concepts and relationships. The RO will be extensively taken into account when new relationships will be defined in order to improve DAOZ interoperability with other ontologies. As part of the ontology ongoing development, the spatial granularity is being extended. This extension is intended to

further enrich the ontology conceptual schema. Moreover studies are in progress to realize cross-species interoperability with our ontology. A development in these ongoing studies is the recent Common Anatomy Reference Ontology (CARO) (Haendel et al, 2007). CARO is being developed to facilitate interoperability between existing anatomy ontologies for different species; this will be extremely useful in linking data between developmental model systems.