Sarah W. Carrier. The Dryad Repository Application Profile: Process, Development, and Refinement. A Master's paper for the M.S. in I.S. degree. April, 2008. 69 pages. Advisor: Jane Greenberg

This paper presents research and development work for the Dryad metadata application profile. Dryad is a digital repository for datasets underlying published works in evolutionary biology and related fields. The paper details the phased implementation of the repository and the corresponding modular application profile. The paper reviews the application profile methodology, reviews each element description, and describes how the schema supports the unique functionalities of each phase of Dryad. The approach presents a method for bringing the Level One application profile, which is currently being tested for Phase One of Dryad, into conformance with the Dublin Core Singapore Framework. The benefits of compliance with the Singapore Framework include maximum interoperability and long-term quality control of the schema. In addition, conformance will allow for the Dryad application profile to be utilized by other initiatives. Finally, this paper proposes a Level Two Dryad application profile and a means of implementation.

Headings:

      Application profile

      Metadata

      Dublin Core

      Evolutionary biology

      Interoperability

      Singapore Framework

THE DRYAD REPOSITORY APPLICATION PROFILE: PROCESS, DEVELOPMENT,
AND REFINEMENT

by
Sarah W. Carrier

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2008

Approved by

_____

Dr. Jane Greenberg

TABLE OF CONTENTS

TABLE OF FIGURES

Chapter 1: Introduction

Since October 2006, the Metadata Research Center (MRC)[1] at the School of

Information and Library Science (SILS)[2], UNC Chapel Hill and the National

Evolutionary Synthesis Center (NESCent)[3] have been collaborating to develop a digital

data repository that will host datasets underlying published research in evolutionary

biology and related fields. The repository, called Dryad[4], is intended to fulfill a research

and scholarly communication service to the evolutionary biology community by

providing a centralized means of data storage and access. Dryad will closely link

publications and the underlying data with a focus on "data-centric" services that would be

useful to the life science community. Specifically, the system will facilitate scholarly

research and collaboration through sophisticated data retrieval, reuse, and manipulation.

During the earliest stages of repository planning, the Dryad development team

established the functional requirements of the system (Dube, Carrier, & Greenberg,

2007). A chief component of the functional requirements focused on the role of metadata

and how the metadata architecture would be structured in order to function within the

larger system. As part of this work, the MRC developed an application profile of

metadata elements necessary to fully describe evolutionary biology datasets with a

---

[1] http://ils.unc.edu/mrc/
[2] http://sils.unc.edu/
[3] http://www.nescent.org/index.php
[4] http://dryad-dev.nescent.org/

consideration for preservation and retrieval. Application profiles, as defined by Heery and Patel, are "schemas which consist of data elements drawn from one or more namespaces, combined together by implementers, and optimized for a particular local application" (2000).

In order to ensure the maximum interoperability of the repository with other information systems, the Dublin Core[5] metadata standard was selected as the foundation scheme for the Dryad application profile. However, the unique "data-centric," domain-specific model of Dryad required the development team to incorporate other established metadata standards used by science communities, with particular attention to standards adopted by the life sciences. In addition, the close pairing of data objects with published documents required that the team consider namespaces that support the preservation of heterogeneous objects and those that facilitate digital publishing. Therefore, the namespaces considered by the team included PREservation Metadata Implementation Strategies (PREMIS)[6], Data Documentation Initiative (DDI)[7], Ecological Metadata Language (EML)[8], and Darwin Core[9].

The Dryad repository is being implemented in a phased fashion, with Phase One focusing on the capture, preservation, and simple retrieval of data objects, in addition to any article metadata with which they are associated. The Level One application profile represents the Phase One functionalities of Dryad. Metadata elements chosen for the Level One application profile support basic preservation, retrieval and reuse. Phase Two

---

[5] http://dublincore.org/
[6] http://www.loc.gov/standards/premis/
[7] http://www.icpsr.umich.edu/DDI/
[8] http://knb.ecoinformatics.org/software/eml/
[9] http://wiki.tdwg.org/twiki/bin/view/DarwinCore/WebHome

implementation will be initiated after proven success with Phase One. The functionality of this phase will include advanced preservation and retrieval, version representation and tracking, data manipulation and synthesis, and will focus greater attention on data life cycle management.

Dryad is currently in Phase One, and the development team is in the process of evaluating the Level One metadata structure. As we assess our Level One metadata structure, we are also conducting observations and planning for the development of Dryad's Level Two metadata structures, which will be expressed through an application profile that builds upon the first manifestation. As part of our metadata development plan, we will also bring the Level One application profile into conformance with the recently published Dublin Core Singapore Framework. This step will ensure the maximum interoperability and long-term quality control of the schema, and allow our application profile to be utilized by other initiatives.

Two key objectives underlie the pursuance and successful completion of the work outlined above and form the basis for the research detailed in this paper. They are: 1) To describe the development and implementation of the Level One Dryad application profile, including bringing the application profile into conformance with the Singapore Framework, and 2) To describe and propose the Level Two application profile, and to recommend means of implementation.

Chapter 2: Literature Review

The metadata standards community has published numerous guidelines describing

the design and development of an application profile. In addition, existing literature

points to the numerous purposes and functions of application profiles. Application

profiles have been developed by many initiatives, and reports and descriptions of these

projects have influenced the development of the Dryad schema. Literature in these areas

has informed the work presented in this research and will be described in the following

two sections.

2.1. The Application Profile

The concept of an "application profile" emerged from a need to expand Dublin

Core to accommodate domain- or community-specific semantics (Baker, Dekkers,

Fischer, & Heery, 2005). As stated above, an application profile combines elements from

existing namespaces in order to accommodate a purpose that is not sufficiently met by

one metadata schema. Makx Dekkers details the requirements of an application profile as

initially described by Thomas Baker (2001):

- A definition of entity classes described by the application profile and an identification of the functions that the schema is intended to support.
- A declaration of what elements are included in the schema and details about their semantics, rules for their usage, and allowable value data types.
- Details about which controlled vocabularies would be used to restrict the allowable values for particular fields.
- Human readable information about the schema and guidelines for use.

Application profiles serve numerous purposes for domain-specific collections. Such purposes, with particular importance to Dryad, include (Baker, Dekkers, Fischer, & Heery, 2005):

- Helping communities to harmonize metadata practices.
- Identifying domain-specific terminologies as candidates for formal standardization.
- Assisting in the guiding of semantic mappings and format conversions.

The Dublin Core community has developed standards for the development and documentation of application profiles. For example, the Dublin Core Application Profile (DCAP) Guidelines encourage interoperability based on Dublin Core and harmonization of usage and convergence (Baker, Dekkers, Fischer, & Heery, 2005). The DCAP provides guidelines for the structure and content of Dublin Core-based application profiles such as Dryad's. Included in the guidelines are instructions about identifying terms with appropriate precision, specifically through the citation of a Uniform Resource Identifier (URI), if available. Descriptions of term usage attributes are also required by the guidelines and include the identification of the term, definitional attributes, relational attributes, and any constraints on term usage. Constraints on term usage would include information about whether the term is required and the repeatability of the element (Baker, Dekkers, Fischer, & Heery, 2005). The DCAP guidelines stress the importance of human-readable descriptions, and therefore expressing schemas as text documents is emphasized. However, the guidelines do give recommendations for expressing application profiles in RDF and XML.

Another important standard upon which application profiles are built is the Dublin Core Metadata Initiative (DCMI) Abstract Model (Nilsson, Baker, & Johnston, 2008; Powell, Nilsson, Naeve, & Johnston, 2007; Nilsson, Johnston, Naeve, & Powell, 2006).

Application profiles should be based on the DCMI Abstract Model in order to be fully machine-processable (Nilsson, Miles, Johnston, & Enoksson, 2007). The DCMI Abstract Model describes the components of the Dublin Core schema and provides a structure for how they are related and combined. For the purposes of this research, the DCMI Description Set Model component of the Abstract Model is of particular interest. The Description Set Model specifies how information about resources is represented through Dublin Core. Each description within the set refers to a specific resource. Examples of a resource are anything that can be identified, which would include, for example, a document or an image. Statements about the resource include a URI for identification purposes and a value surrogate, which can either be literal or non-literal (Powell, Nilsson, Naeve, Johnston, 2007; Nilsson, Miles, Johnston, & Enoksson, 2007).

Application profiles have been used to ensure the interoperability of a metadata schema, and as an extension, provide an opportunity for collaboration. As Heery and Patel explain, "By defining application profiles and, most importantly by declaring them, implementers can start to share information about their schemas in order to inter-work with wider groupings" (2000). The importance of interoperability, and particularly how metadata can improve interoperability, has been recognized in numerous initiatives and endeavors that are similar to the Dryad project. For example, a statement produced by the *Scientific Data Chain Workshop* in January 2006 in Windsor, UK identified metadata, in addition to process documentation and ancillary data, as a key to reusability of data. The statement also emphasizes that data curation involves not only cataloging the object (the metadata), but also "expanding information about the data objects to meet the community needs" (Cox, Jones, Lawrence, Milic-Frayling, & Moreau, 2006). Similar

statements and reports regarding the significance of metadata's role in enhancing

interoperability have been issued as a result of task forces, workshops, and research

studies from multiple domains (e.g., Hunter & Lagoze, 2001; Chan & Zeng, 2006; Joint

Task Force on Library Support for E-Science, 2007; National Science Board, 2005;

Borgman, 2007).

    The publication of application profiles, which provide a structure in which domain-

specific metadata schemas can be developed, assist in serving community needs and

guarantee interoperability. In summary, De Roure and Hendler state:

> Achieving interoperable infrastructure requires the development of common
> vocabularies and metadata frameworks as the basis for description, discovery, and
> integration of the services, together with the use of domain-specific knowledge for
> problem solving in order to compose services (2004).

2.2. Related Work

    Application profiles have been used in numerous initiatives with success, and the

lessons learned in those endeavors have implications for the Dryad project, although

many are "document-centric" rather than "data-centric." The Joint Information Systems

Committee (JISC)[10], UKOLN[11], and the Eduserv Foundation[12] developed an application

for eprints (also known as "preprints" or scholarly works) (Allinson, Johnston, & Powell,

2007). The Eprints application profile employs the DCMI Abstract Model notion of

"description sets" where object metadata is represented in complex sets of entities. The

approach taken in this project is to represent five entities: ScholarlyWork, Expression,

Manifestation, Copy, and Agent. The concept of the five entities is drawn from the

---

[10] http://www.jisc.ac.uk/
[11] http://www.ukoln.ac.uk/
[12] http://www.eduserv.org.uk/

Functional Requirements for Bibliographic Records (FRBR) (IFLA Study Group on the

Functional Requirements for Bibliographic Records, 1998). The Eprints application

profile, therefore, represents the original scholarly work itself, any versions of the work

(reprints, copies, etc.), and the different actors (people, institutions, etc.) that are involved

with the various versions (Allinson, Johnston, & Powell, 2007). The Eprints application

profile, like Dryad's, is based on Dublin Core, but incorporates some community-specific

elements outside the scope of Dublin Core that are necessary for the full functionality of

the schema.

A similar example is the DiVA project application profile that is intended to

describe digital academic documents (Müller, Klosa, Andersson, & Hansson, 2003). The

inspiration for the DiVA application profile stemmed from a need to properly represent

the granularity in document description, the need to state relationships and hierarchies,

and to facilitate flexibility of format. DiVA was also derived from FRBR and

incorporates the concept of "manifestation."

There are two additional examples of application profiles based on Dublin core

that are relevant to the Dryad project. The first is the Government Application Profile

(Cumming, Aargaard, Dekkers, Murphy, & Borras, 2001). The Government Application

Profile is intended to clarify the use of Dublin Core in the government context. The

second relevant project from the government domain is the UK e-government metadata

standard application profile (Powell, 2000). The intention of this application profile is to

facilitate the development of UK e-government portals.

An important initiative in the context of the Dryad project is the eBank UK project

application profile (Koch, Duke, & Coles, 2005), particularly because data objects are

being linked with published articles in the field of chemistry (Lyon, 2003). The eCrystals

Open Repository utilizes the eBank schema (Lyon & Cole, 2008). The eCrystals

repository will support a federation of data repositories for crystallography. It functions

much like Dryad in that eCrystals makes available raw, derived, and results data.

Lessons learned in this initiative are directly applicable to the Dryad development

project, and specifically, the structure of the project's application profile is of great

interest in this research.

Chapter 3: Project Background

During the first stage of the project, the Dryad development team established the functional requirements, followed by the development of an application profile of metadata elements necessary to fully describe evolutionary biology datasets. The team sketched out a series of typical scenarios of use in order to detail the functional requirements (Dube, Carrier, & Greenberg, 2007). Two examples of typical scenarios that would be facilitated by the metadata structure are:

- A user depositing a dataset as a requirement for publication.
- A user searching for datasets that are applicable to their own research. This user could search by author, dataset description, species name, etc.

After consulting with Dryad team members and stakeholders in the project, the current Phase One Dryad workflow can be described as follows, with metadata being created in the third step:
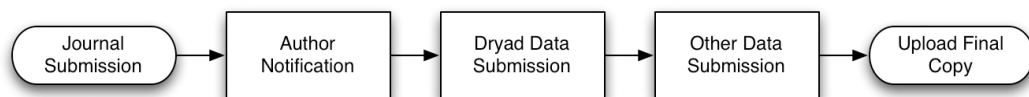

Figure 1: The Dryad Workflow

After considering the use case scenarios and the Dryad deposition workflow, the NESCent repository development team and the Metadata Research Center determined Dryad's long-term functional requirements (Jed, Carrier, & Greenberg, 2007). The team

concluded that the requirements had to support the following key aspects of Dryad's

functioning:

- Heterogeneous digital datasets
- Long-term data stewardship
- Tools and services available to repository users
- Incentives for use
- Reduced barriers for use, specifically reduced technical barriers
- Authors' intellectual property rights
- Datasets underlying published material

Another important consideration in developing the functional requirements was to

identify the object types and data types that would be hosted by Dryad, including

information about the life cycle of these object and data types (Carrier, Dube, &

Greenberg, 2007). Consideration of these issues was also essential to the decisions

informing the development of the application profile. The object types that would be

included in Dryad include:

- Publication (e.g., journal article, conference paper)
- Published piece of data in the publication (e.g. a table)
- Dataset behind the published data (e.g. supplemental data)
- Initial data source (e.g., American Ornithologists' Union checklist)
- Newly created data (e.g., data derived from any of above)

The data types that would be supported by Dryad include:

- Structured labeled data (e.g., tabular data with column and row headings)
- Structured unlabeled data (e.g., tabular data without column and row headings, or with undecipherable headings)
- Unstructured textual data (e.g., readable text)
- Unstructured non-textual data (e.g., maps, graphs, images)

Therefore, the long-term functional requirements for Dryad are as follows:

- Computer-aided metadata generation and augmentation.
- Specialized modules linking data submission and published material.
- Data and metadata quality control by integrating human and automatic techniques.
- Support for identity, authority and data security.

- Support for basic metadata repository functions, such as resource discovery, sharing, and interoperability.

After developing the above functional requirements, the team determined that the ideal implementation for Dryad would take a phased approach. Based in part on feedback given at a workshop that took place in December 2006, the user community established that the ideal implementation would involve an archival space being set up as soon as possible. This initial archival space would provide basic services and offer basic functionalities in order to address the data deluge taking place in the field of evolutionary biology. The workshop attendees determined that the second priority would be to incorporate more sophisticated functionalities after the archival space was in operation[13].

This determination affected the design of what became the "Level One" application profile. The repository development team concluded that the Level One application profile represents the Phase One functionalities of Dryad. Metadata elements chosen for the application profile provide basic preservation, retrieval and reuse. In addition, the team established that the linkages between publications and underlying data objects would best be represented in a modular fashion, with a "Publication" module that is related to the "DataObject" module. Therefore, the two entities supported by the Level One application profile are Publication and DataObject.

The Dryad team employed a multi-method approach to develop the application profile. In approaching the process, the team utilized the steps detailed by Makx Dekkers (2001):

1. Define metadata requirements.
2. Select most appropriate existing standard metadata element set.

---

[13] https://www.nescent.org/wg_digitaldata/Main_Page

3. Where possible, use standard elements for locally required elements, possibly narrowing semantics and adding local rules and vocabularies.
4. Define remaining elements in private namespace.

In addition, the methods used by the team included a *requirements assessment*, *content analysis*, and *crosswalk analysis* (Carrier, Dube, & Greenberg, 2007). The requirements assessment is described above and included an incorporation of typical use case scenarios, functional requirements, and stakeholder interests.

The next step in the process was content analysis as defined by Krippendorff (2004). The team first examined various metadata schemas and employed the content analysis methodology to identify relevant elements. For each schema being analyzed, the following questions were asked:

- Which schema is being analyzed and what elements are included?
- How is the schema defined?
- What are the recommended, mandatory, and optional elements?
- In what context was the schema designed, and how is it currently applied within the community for which it was developed?
- How does the context relate specifically to Dryad?

The results of the content analysis for the Level One Dryad application profile will be detailed in the next sections.

A Dublin Core-based application profile was chosen in order to assure full interoperability and compatibility with other systems. In addition, Dublin Core is an accepted standard within the metadata community, with enough flexibility to adapt to Dryad's needs. However, the objects stored in Dryad will be heterogeneous in nature, with only a fraction of the data in text form. Therefore, the metadata team decided to examine other namespaces designed for the sciences, the social sciences, and, specifically, evolutionary biology.

For each namespace the Dryad team considered required and recommended elements. The first step in the process was to look at the Dublin Core[14,15] standard and identify elements that are particularly applicable for a Phase One application. We sought to also include two well-known namespaces from the social science and science communities: the Data Documentation Initiative (DDI) for social science metadata and the Ecological Metadata Language (EML) for ecology and environmental research. Particular attention was given to elements that gave more detail to the description of the data object not accomplished by the Dublin Core. In addition, the PREservation Metadata Implementation Strategies (PREMIS) schema was considered by the team in order to ensure long-term data object preservation support. Finally, the Darwin Core schema, which is designed for metadata about collection specimens and the geographic occurrence of species, was included in the content analysis.

DDI 3.0 is scheduled for release in April 2008, although this new version will not supersede the previous Version 2.1. The DDI schema has five main sections with varying granularity of description:

1. Document Description
2. Study Description
3. Files Description
4. Data (Variables) Description
5. Other Related Materials

The Dryad application profile team decided that all levels from the DDI could be considered appropriate for inclusion in the Dryad schemas, both Level One and Level Two.

---

[14] http://dublincore.org/documents/dces/
[15] http://dublincore.org/documents/dcmi-terms/

EML Version 2.0.1 is a metadata schema developed for the field of ecology and is utilized by the Knowledge Network for Biocomplexity (KNB)[16]. The schema is of particular relevance to Dryad not only due to the domain-specificity, but also because EML is intended for the description of digital resources. EML elements are organized into hierarchical modules at various levels of granularity. The team was particularly interested in the "eml-dataset module" for dataset-specific information and the data organization modules that describe the structures of datasets. Another module that is of interest to the Dryad team is the "eml-software module," which describes offers great detail about the software used to generate a dataset, therefore ensuring that the data can not only be viewed, but that it can be processed and reused.

The PREMIS metadata model ensures the preservation of digital objects and was also considered for the Dryad application profile. The team found that PREMIS focuses mainly on specific technical metadata rather than descriptive metadata, agents, rights, or media/hardware details (PREMIS Working Group, 2005). Therefore, we reasoned that it was best to take advantage of the full scope of the PREMIS technical metadata elements in later levels of the application profile, and concentrate on mapping descriptive metadata for Level One of the application profile. We did, however, decide to include the element "fixity," which is a hidden value that tracks whether unauthorized changes have been made to the content.

Darwin Core is a known metadata standard within the evolutionary biology community. DwC 1.4 is currently a draft version under discussion, and the Dryad team considered elements from DwC 1.2 and 1.21. Of particular interest were elements

---

[16] http://knb.ecoinformatics.org/index.jsp

describing the collection specimens, with the determination that any geographical information about the specimen could be accomplished with Dublin Core "coverage." Elements from the Darwin Core schema are considered for both the Level One and Two application profiles.

After the completion of the content analysis, the final step in developing the Level One application profile was a crosswalk analysis. The crosswalk between the schemas enabled the team to compare and contrast the various namespaces and took the form of a spreadsheet. Included in the crosswalk were the element name and prefix, the element definition, and examples of use. Semantic overlaps in the crosswalk were "normalized" by the group. Since the DDI is partially derived from Dublin Core, equivalent elements were simple to identify. Similarly, the top-level structure of EML has been designed to be compatible with the Dublin Core syntax. The crosswalk informed the group as to which namespaces and elements supplemented the central Dublin Core elements. For example, we saw that it would be most appropriate to include the Darwin Core element "Species" to fill in a gap where other, more general namespaces could not offer an appropriate substitute. As a result of the crosswalk approach, the team decided to choose Dublin Core elements unless another namespace filled an obvious void. Where a namespace's element mapped directly to Dublin Core, the Dublin Core element was chosen.

In addition to the methods described above, another essential consideration in developing the application profile was to examine how the metadata architecture fit into the overall Dryad system design. The current Dryad architecture is modeled in the following figure:
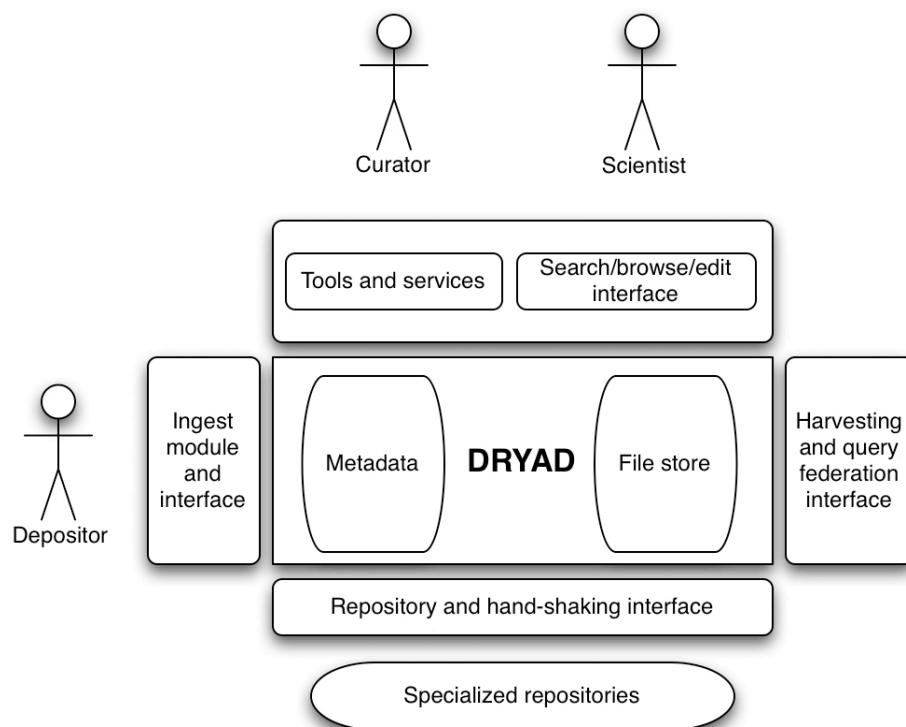
Figure 2: The Dryad System Architecture

When finalizing the Level One application profile, two influential articles regarding the ideal structure and functioning of digital repositories provided the context for evaluation. Altman and King (2007) propose a citation standard for quantitative data that includes six mandatory components: author, date, title, global unique identifier, a universal numeric fingerprint, and a bridge service. In addition, Jantz and Giarlo (2005) describe the required architecture and technology that would ensure the trustworthiness of a digital repository. The authors recommend that a trusted repository metadata scheme include descriptive, technical, source, rights, and digital provenance metadata. Both

articles provide valuable recommendations for the structure and design of the Level One

application profile.

        According the DCAP guidelines, the Dryad application profile was described in a

table format, with each element described with sufficient detail to provide for human

understandability (Baker, Dekkers, Fischer, & Heery, 2005). This step ensures that the

application profile is text-based, human-readable, and conforms to Dublin Core

standards. A Descriptive Header is included to describe the entire application profile and

is based on Dublin Core. According to the guidelines, attributes of term usage are also

included. For each term, identifying attributes, definitional attributes, relational

attributes, and constraints are detailed. A representation of the Dryad Level One

Application Profile can be viewed in Appendix A.

Chapter 4: The Singapore Framework

The Singapore Framework was presented at the International Conference on Dublin Core and Metadata Applications in Singapore, September 2007 (Apps, 2007; Nilsson, Baker, & Johnston, 2008). In order to ensure the interoperability of Dublin Core application profiles, the Singapore Framework was developed as a standard for machine-understandable representations of metadata schemas (Nilsson, Baker, & Johnston, 2008). Description Set Profiles (DSPs) are the key aspect of the framework. Through an XML-based DSP, the structure of the application profile is represented in an interoperable, machine-readable format. Utilization of the DSP ensures the quality and long-term reusability of the schema (Nilsson, Miles, Johnston, & Enoksson, 2007).

This Master's paper research presents work undertaken to make the Dryad application profile compliant with the Singapore Framework. A driving motivation behind bringing the Level One project application profile into conformance with the Singapore Framework revolves around the concept of interoperability. The issue of interoperability is particularly important to Dryad as the system reaches its full functionality, which would include search, retrieval, "hand-shaking" with other repositories, exposing metadata for harvesting purposes, and web services.

As described in *The Singapore Framework for Application Profiles*, the document package for a Dublin Core Application Profile (DCAP) contains five components (Nilsson, Baker, & Johnston, 2008):

- *Functional requirements*, which describe the functions that an application profile is intended to support, plus functions that are not within the project's scope. Mandatory.
- *Domain model*, which defines the basic entities described by the application profile and their relationships and defines a basic scope. Mandatory.
- *Description Set Profile (DSP)*, which defines a set of metadata records that are valid instances of an application profile. Mandatory.
- *Usage guidelines*, which describe how to apply the application profile. Optional.
- *Encoding syntax guidelines*, which describe any application profile-specific syntaxes and/or guidelines. Optional.

Specifically, the DSP is an information model and XML expression of an application profile. A DSP is based on the DCMI Abstract Model, specifically the Description Set Model, and functions in the following ways (Nilsson, 2007):

- As a formal representation of the constraints of a DCAP.
- As a configuration for databases.
- As configuration for metadata editing tools.

There are two levels of templates in a DSP and reference the structure of the DCMI Abstract Model Description Set: a Description template and a Statement template. The Description template contains the Statement templates and refers to a particular identifiable resource. Statement templates include information about constraints, value strings, and vocabulary encoding schemes in reference to the particular resource (Nilsson, Baker, & Johnston, 2008; Nilsson, Miles, Johnston, & Enoksson, 2007; Nilsson, 2007).

Chapter 5: Methods for the Singapore Framework

This section will provide an overview of the methods utilized to bring the Dryad Level One application profile into conformance with the Singapore Framework. First, issues specific to the Dryad application profile will be reviewed and second, the approach for revising the application profile to comply with the Singapore Framework will be detailed. Some of the challenges involved with this procedure are also described.

5.1. Dryad Level One Application Profile Overview

Two unique aspects of the Dryad Level One application profile inform the formalization of the schema according to the Singapore Framework. Of particular importance is the module structure of the Dryad application profile, which contains a publication module and a data module. Second, it is important to note that the Level One functionalities of Dryad in Phase One implementation only include the simplest relationships between articles and the underlying data. This relationship can be described as a "one to many" relationship between publication and dataset, with many data objects associated with one publication. Versions of publications and versions of datasets are not accommodated in Level One, and therefore will not be represented in the Description Set Profile. In comparison, the Eprints DSP, which is currently the only example of a fully formed DSP, includes five modules or entities: Scholarly Work, Expression, Manifestation, Copy, and Agent (Allinson, Johnston, & Powell, 2007). Functionalities such as those exhibited by the Eprints model are not represented in the Level One

application profile; however, similar functionality will be accommodated in the Level Two application profile.

It is also of importance to reemphasize that Dryad is "data-centric" in focus, and that the storage and retrieval of a published article is not the focal point of the functioning of the repository. Therefore, the scope of the Level One application profile and its representation is more focused on how the published article relates to the data object, insofar as providing context for the hosted dataset.

5.2. Singapore Framework Procedures

As described above, the Singapore Framework has five components:

- Functional requirements
- Domain model
- Description Set Profile (DSP)
- Usage guidelines
- Encoding syntax guidelines

The scope of this research includes implementation of all of the components, excluding the optional "encoding syntax guidelines," which do not apply to the Dryad application profile. The encoding syntax guidelines would apply to an application profile that has schema-specific encoding rules.

The first step was to define the functional requirements of the project. The functional requirements were decided upon early in the repository development process and were detailed in the above sections. The functional requirements for the Dryad application profile are broken into four sections: scope, stakeholders and community, requirements gathering, and functional requirements specification. The structure of the functional requirements is based on the Eprints example. Please see Appendix B.1. for

the formal declaration of the Dryad functional requirements according to the Singapore

Framework guidelines.

The second step was to define the domain model. The domain model defines the

basic entities described in the application profile (Nilsson, Miles, Johnston, & Enoksson,

2007). Relationships between the entities are also described. In the case of the Level

One application profile for Dryad, there are two entities: the Publication and the

DataObject. The relationship between the Publication entity and the DataObject can be

described as "isSupplementedBy." In Level One, many DataObjects can supplement a

Publication. Not until Level Two will reuse be tracked in a way that there are multiple

publications associated with one DataObject.

In the context of the domain model, the Eprints project domain model example is

not directly applicable for the Level One application profile. As stated, the Phase One

functionality of Dryad will not represent manifestations or versions of the data objects

and the associated published articles. In addition, the Eprints model is document-centric

as it is based on the Functional Requirements for Bibliographic Records (FRBR). Due

the unique linking of publications and datasets, the roles of the entities described by the

application profile become more complicated, and also, therefore, the representations of

those relationships:

> Creating metadata for electronic "documents," such as prepublications,
> dissertations, and theses, is fairly straight forward, drawing from bibliographic
> control practices. This is demonstrated by the wide adoption of the OAI Protocol
> for Metadata Harvesting, based on the Dublin Core metadata standard. The
> metadata issues become more complicated, however, when a repository wants to
> include multiple object types, such as publications and data objects, and link them
> (Carrier, Dube, & Greenberg, 2007).

Please see Appendix B.2. for a representation of the Dryad domain model in UML

format.

The third step in the process was to represent the application profile as a

Description Set Profile (DSP). As mentioned above, the DSP defines a set of metadata

records that are valid instances of an application profile (Nilsson, Miles, Johnston, &

Enoksson, 2007; Nilsson, 2007). The Dryad DSP XML file was successfully validated

using the World Wide Web Consortium XML validation service[17]. The entire Dryad

Level One application profile DSP can be viewed in Appendix B.3.

A Description Set Profile does not, however, address human-readable

documentation, definition of vocabularies, or version control (Nilsson, 2007; Enoksson,

2007). Therefore, a supplemental step involved the parsing of the DSP into a human-

readable format for viewing online. A specialized Wiki-syntax for the Description Set

Profile was developed by the developers of the Singapore Framework to parse the XML

into a readable format. The syntax was developed specifically for the MoinMoin Wiki

engine[18] and is accomplished through a parser extension. An example of the output is

available and uses the Eprints application profile as a model[19]. In addition to the Wiki-

syntax developed by Enoksson and others (Enoksson, 2007; Nilsson, Miles, Johnston,

Enokkson, 2007), an XSLT can be used to accomplish the parsing for display online. At

the time of writing, the tool for generating Wikis is not yet installed for general use, and

---

[17]

http://www.w3.org/2001/03/webdata/xsv?docAddrs=http%3A%2F%2Fwww.ils.unc.edu
%2F%7Escarrier%2Fdryad%2FDSPLevelOneAppProf.xml&warnings=on&style=xsl#
[18] http://moinmo.in/
[19] http://knowware.nada.kth.se/DCWiki/EprintsApplicationProfile

therefore the human-readable version of the Dryad DSP is available online and is displayed using an XSLT[20].

The fourth step of the Singapore Framework was to formalize the application profile usage guidelines. The Dryad development team has defined the usage guidelines for both the Level One application profile and the Phase One of Dryad at the following URL:

https://www.nescent.org/wg_digitaldata/Public:Dryad_Level_One_Cataloging_Guidelines. The length of the guidelines precludes inclusion in this paper, and therefore, the content has been added to the Dryad project Wiki space for public viewing.

The final step of the Singapore Framework was to consider whether to include the optional encoding syntax guidelines. These guidelines define application profile-specific encoding and are not applicable in the Level One Dryad application profile.

In addition to the usage guidelines and the DSP that are available online, information for each component of the Dryad application profile has been organized into a project Wiki page for viewing[21].

---

[20] http://www.ils.unc.edu/~scarrier/dryad/DSPLevelOneAppProf.xml
[21] https://www.nescent.org/wg_digitaldata/Public:Level_One_Application_Profile

Chapter 6: Level Two Application Profile

As stated above, the Level One Dryad application profile includes what the Dryad team determined to be the minimum of requirements for deposition and is designed to collect basic metadata about the data object for the purposes of ingestion, archiving and access. This level, which is being implemented as Phase One of the Dryad repository, supports discovery, preservation, and encourages data use and understanding. In addition, the Level One application profile reflects the simplest relationships between data objects and the publication with which they are associated.

The Level Two application profile, which is under development, will include metadata that support more advanced functionalities such as sophisticated use/reuse, manipulation and synthesis. In addition, the Level Two application profile will support version tracking and more advanced data life cycle management. The Level Two application profile therefore reflects the requirements of the Phase Two Dryad repository as envisioned by the development team. Specific elements would support the following services for data objects in Dryad:

- Expanded metadata about preservation
- Enhanced granularity of data description
- More information about methodology and workflow
- More about known linkages (to publications, to other datasets, etc.)
- Tracking of use and reuse
- Provenance information

However, it should be noted that the modeling of complex digital objects will not be taking place in Phase Two of Dryad, and will be considered for later implementations.

As part of this research, I propose a Level Two application profile for Phase Two of Dryad. The same methods and procedures as undertaken for the Level One application profile development were employed for the Level Two application profile. The Level Two application profile builds upon the functionalities of the Level One schema. The same namespaces were consulted in order to accomplish the needs of the Phase Two Dryad repository.

In addition to the namespaces already described, the MicroArray Gene Expression Markup Language (MAGE-ML)[22] has been consulted in order to increase the granularity of description options for data objects. MAGE-ML describes and communicates information about experiments based on DNA microarrays. Appendix C. includes a description of terms composing the proposed Level Two Dryad application profile.

---

[22] http://www.mged.org/Workgroups/MAGE/mage-ml.html

Chapter 7: Next Steps

The next steps involved with the Level One application profile implementation include continued testing and validation of the existing schema by the Dryad team. Phase One Dryad will be officially released to the public in the coming months, after which we will be seeking feedback from users regarding the deposition process, and be able to further examine the use and effectiveness of the Level One application profile and our overall metadata architecture.

Further steps include community acceptance, "hand-shaking" with other repositories, and full exposure of metadata to web services and metadata harvesters. The development team expects that the application profile in the Singapore Framework format will play a large role in this transition. Community acceptance will be furthered through the documentation of the development process and the web publication of the application profile based on the Singapore Framework. Once the mechanism is in place, another important step will be to convert the Level One DSP into the specialized Wiki-syntax for display online. The model for Description Set Profiles and the Singapore Framework itself is still evolving, and as the model advances, so will the Dryad application profile. Therefore, the Dryad schema is viewed as an ongoing, developing structure.

Next steps for the Level Two application profile involve:

1. Implementation
2. Testing
3. Feedback

4.   Reevaluation

The implementation of the Level Two schema will involve supplementing the current Level One application profile with the additional metadata elements. After successful completion of the listed next steps, the Level Two application profile will also be brought into conformance with the Singapore Framework.

Communication with other initiatives engaging in the development of application profiles will also be initiated, regardless of the domain focus of the project, but with a specific interest in the life sciences.

Chapter 8: Summary and Conclusions

This research described the development and implementation of the Level One Dryad application profile, which is currently in testing phase. In addition, this paper described the approach undertaken to bring the Level One application profile into conformance with the Singapore Framework. Finally, this research examined and proposed the Level Two Dryad application profile.

The research described in this paper will assist the Dryad project to move forward into the next phase of implementation. With Level One progress underway, a definition of the Level Two schema will move the repository into Phase Two. In addition, the public documentation and publishing of the Level One application profile will raise the profile of the project and increase community awareness and acceptance.

The work presented in this paper can assist other initiatives in developing application profiles that follow the Singapore Framework. The results of this process will provide additional context about a relatively new, expanding framework undertaken by the Dublin Core Metadata Initiative. Although the Dryad model is unique in its modular structure and phased implementation, lessons learned in the process of development have implications for other projects. In addition, initiatives outside the life sciences can draw from the Dryad experience, particularly those seeking to link publications or published documents with underlying data.

Chapter 9: Bibliography

Allinson, J., Johnston, P., & Powell, A. (2007). A Dublin Core Application Profile for Scholarly Works. *Ariadne*, *50*. Retrieved April 6, 2008, from http://www.ariadne.ac.uk/issue50/allinson-et-al/

Altman, M., & King, G. (2007). A Proposed Standard for the Scholarly Citation of Quantitative Data. *D-Lib Magazine*, *13*(3/4). Retrieved April 6, 2008, from http://www.dlib.org/dlib/march07/altman/03altman.html

Apps, A. (2007). DC2007 'Application Profiles: Theory and Practice'. *Ariadne*, *53*. Retrieved April 6, 2008, from http://www.ariadne.ac.uk/issue53/dc-2007-rpt/

Baker, T., Dekkers, M., Fischer, T., Heery, R. (2005). Dublin Core Application Profile Guidelines. Retrieved April 6, 2008, from http://dublincore.org/usage/documents/profile-guidelines/

Borgman, C. L. (2007). *Scholarship in the Digital Age: Information, Infrastructure and the Internet*. Cambridge, MA: The MIT Press.

Carrier, S., Dube, J., & Greenberg, J. (2007). The DRIADE Project: Phased Application Profile Development in Support of Open Science. In *Proc. Int'l Conf. on Dublin Core and Metadata Applications 2007*, 35-42. Retrieved April 6, 2008, from http://www.dcmipubs.org/ojs/index.php/pubs/article/viewFile/39/19

Chan, L. M., Zeng, M. L. (2006). Metadata Interoperability and Standardization–A Study of Methodology Part 1: Achieving Interoperability at the Schema Level. *D-Lib Magazine*, *12*(6). Retrieved April 6, 2008, from http://www.dlib.org/dlib/june06/chan/06chan.html

Cox, S., Jones, R., Lawrence, B., Milic-Frayling, N., & Moreau, L. (2006). Interoperability Issues in Scientific Data Management (Version 1.0). Technical report, The Technical Computing Initiative, Microsoft Corporation. Retrieved April 6, 2008, from http://download.microsoft.com/download/f/b/3/fb3d02b8-2210-4d0d-a747-9519eafae6c1/ScientificDataManagement4.18.07.pdf

Cumming, M., Aargaard, P., Dekkers, M., Murphy, P., & Borras, J. (2001). Government Application Profile. Retrieved April 6, 2008, from http://dublincore.org/documents/gov-application-profile/

De Roure, D., & Hendler, J. A. (2004). E-Science: The Grid and the Semantic Web. *IEEE Intelligent Systems*, *19*(1), 65-71.

Dekkers, M. (2001). Application Profiles, or How to Mix and Match Metadata Schemas. *Cultivate Interactive*, *3*. Retrieved April 6, 2008, from http://www.cultivate-int.org/issue3/schemas/

Dube, J., Carrier, S., & Greenberg, J. (2007). DRIADE: A Data Repository for Evolutionary Biology. In *Proceedings of the 2007 Conference on Digital Libraries* (Vancouver, BC, Canada, June 18 - 23, 2007). JCDL '07. ACM, New York, NY, 481.

Enoksson, F., ed. (2007). Draft: Wiki Format for Description Set Profile. Retrieved April 6, 2008, from http://dublincore.org/architecturewiki/DSPWikiSyntax

Heery, R., & Patel, M. (2000). Application Profiles: Mixing and Matching Metadata Schemas, *Ariadne*, *25*. Retrieved April 6, 2008, from http://www.ariadne.ac.uk/issue25/app-profiles/

Hunter, J. and Lagoze, C. (2001). Combining RDF and XML schemas to enhance interoperability between metadata application profiles. In *Proceedings of the 10th International Conference on World Wide Web* (Hong Kong, Hong Kong, May 01 - 05, 2001). WWW '01. ACM, New York, NY, 457-466.

IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). Functional Requirements for Bibliographic Records. Retrieved April 6, 2008, from http://www.ifla.org/VII/s13/frbr/frbr.htm

Jantz, R., & Giarlo, M. (2005). Digital Preservation: Architecture and Technology for Trusted Digital Repositories. *D-Lib Magazine*, *11*(6). Retrieved April 6, 2008, from http://www.dlib.org/dlib/june05/jantz/06jantz.html

Joint Task Force on Library Support for E-Science. (2007). Agenda for Developing E-Science in Research Libraries Final Report and Recommendations to the Scholarly Communication Steering Committee, the Public Policies Affecting Research Libraries Steering Committee, and the Research, Teaching, and Learning Steering Committee. Retrieved April 6, 2008, from http://www.arl.org/bm~doc/ARL_Escience_Final.pdf

Koch, T., Duke, M., & Coles, S. (2005). Metadata Application Profile: eBank UK project. Retrieved April 6, 2008, from http://www.ukoln.ac.uk/projects/ebank-uk/schemas/profile/

Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Lyon, L. (2003). eBank UK: Building the Links Between Research Data, Scholarly Communication and Learning. *Ariadne, 36*. Retrieved April 6, 2008, from http://www.ariadne.ac.uk/issue36/lyon/

Lyon, L., & Coles, S. (2008, January). *eCrystals Federation: Open Repositories for Global Open Science*. Presented at a DRIVER Summit, Göttingen, Germany. Retrieved April 6, 2008, from http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/cni-dec07-final.ppt

Müller, E., Klosa, U., Andersson, A., & Hansson, P. (2003). The DiVA Project - Development of an Electronic Publishing System. *D-Lib Magazine*, *9*(11). Retrieved April 6, 2008, from http://www.dlib.org/dlib/november03/muller/11muller.html

National Science Board. (2005). *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. Washington, D.C.: National Science Foundation.

Nilsson, M., Miles, A.J., Johnston, P., & Enoksson, F. (2007). Formalizing Dublin Core Application Profiles Description Set Profiles and Graph Constraints. In *Proceedings of the 2nd International Conference on Metadata and Semantics Research* (CD-ROM), Ionian Academy, Corfu, Greece.

Nilsson, M., ed. (2007). DCMI Description Set Profile Specification. Retrieved April 6, 2008, from http://dublincore.org/architecturewiki/DescriptionSetProfile

Nilsson, M., Baker, T., & Johnston, P. (2008). The Singapore Framework for Dublin Core Application Profiles. Retrieved April 6, 2008, from http://dublincore.org/architecturewiki/SingaporeFramework/

Powell, A. (2000). A UK e-Government Metadata Framework. Retrieved April 6, 2008, from http://www.ukoln.ac.uk/metadata/publications/e-gov-metadata/

Powell, A., Nilsson, M., Naeve, A., Johnston, P., & Baker, T. (2007). DCMI Abstract Model. Retrieved April 6, 2008, from http://dublincore.org/documents/abstract-model/index.shtml

PREMIS Working Group. (2005). Data Dictionary for Preservation Metadata. Retrieved April 6, 2008, from http://www.oclc.org/research/projects/pmwg/premis-final.pdf

Appendix A: The Dryad Level One Application Profile

Note: Documentation in "gray" includes recent suggestions for how to handle a metadata issue, and has not yet been reviewed by the Dryad development team.

A.1. Descriptive Header

Title	Dryad Level One Application Profile
Contributor	Metadata Research Center (MRC)
Contributor	National Evolutionary Synthesis Center (NESCent)
Coverage.spatial	USA
Date.issued	2008-04-07
Description	This document describes the Level One application profile designed by the MRC and NESCent for use with the Phase One implementation of Dryad.
Format	Text
Identifier	https://www.nescent.org/wg_digitaldata/Public:Level_One_Application_Profile
Language	Eng
Status	Version 1.0
Subject	Metadata

A.2. Term Usage – Dataset

Term URI	http://purl.org/dc/elements/1.1/identifier
Defined by	http://purl.org/dc/elements/
Name	Identifier
Label	Identifier
Source Definition	An unambiguous reference to the resource within a given context.
Source Comments	Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Example formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).
Local Definition	The unique identifier of the data object or dataset.
Local Comments	Will be assigned a Dryad-specific unique handle for identification and citation purposes.
Type of term	Element
Obligation	Mandatory
Occurrence	Non-repeatable
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/creator
Defined by	http://purl.org/dc/elements/1.1/
Name	Creator
Label	Author
Source Definition	An entity primarily responsible for making the content of the resource.
Source Comments	Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.

Local Definition	The entity or entities responsible for the creation and development of the dataset.
Local Comments	Should be the same as the associated publication, unless a different set of authors is explicitly stated. Currently, authority control is manual.
Type of term	Element
Occurrence	Repeatable
Obligation	Mandatory
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/contributor/
Defined by	http://purl.org/dc/elements/1.1/
Name	Contributor
Label	Coauthor
Source Definition	An entity responsible for making contributions to the resource.
Source Comments	Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
Local Definition	The entity or entities responsible for contribution to the creation and development of the data set.
Local Comments	Coauthors would be included as contributors.
Type of term	Element
Occurrence	Repeatable
Obligation	Recommended
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/title
Defined by	http://purl.org/dc/elements/1.1/
Name	Title
Label	Title
Source Definition	A name given to the resource.
Local Definition	Descriptive title of the dataset.
Local Comments	Human-readable description of the dataset. Should not be more than 100 characters. If the author does not provide any additional information, we will use the filename as the title, and assume that the contents of the file are obvious to anyone who reads the associated article.
Type of term	Element
Occurrence	Non-repeatable
Obligation	Mandatory
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/rights
Defined by	http://purl.org/dc/elements/1.1/
Name	Rights
Label	Rights Statement
Source Definition	Information about rights held in and over the resource.
Local Definition	Statement regarding rights held in and over the resource.
Local Comments	A short human-readable phrase describing the access rights, which may also be machine-readable. For example: • CreativeCommons license (CC-BY)

- Public Domain
- Copyright held by publisher

A blank value indicates that the dataset is a "normal" status item.

Type of term	Element
Occurrence	Non-repeatable
Obligation	Mandatory
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/description
Defined by	http://purl.org/dc/elements/1.1/
Name	Description
Label	Description
Source Definition	An account of the resource.
Local Definition	Human-readable description of the dataset.
Local Comments	Can contain much more detail than the title. Any description that seems too long to put in this element (e.g., more than one page of text) should be placed in a separate file, which will be a supplemental data stream of this object. It will be given a name of the form READMEx.yyy, where x is a sequence number (omitted if only one documentation file is submitted) and yyy is the file extension of the original (documentation) file.
Type of term	Element
Occurrence	Non-repeatable
Obligation	Optional
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/subject
Defined by	http://purl.org/dc/elements/1.1/
Name	Subject
Label	Subject
Source Definition	The topic of the resource.
Source Comments	Typically, a Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.
Local Definition	Dataset keywords.
Local Comments	Keywords from the publication will be attached to datasets only when it is obvious that they apply. Other keywords may be manually applied to datasets.
Type of term	Element
Occurrence	Repeatable
Obligation	Recommended
Datatype	String

Term URI	http://purl.org/dc/terms/issued
Defined by	http://purl.org/dc/terms/
Name	Date of Issue
Label	Date of Issue
Source Definition	Date of formal issuance (e.g., publication) of the resource.

Local Definition	Publication date.
Local Comments	If you don't choose "this has been published before", automatically filled with the current date. Otherwise specify the date on which it was previously published.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/date
Has Encoding Scheme	http://purl.org/dc/terms/W3CDTF
Constraints	The value must always be taken from the specified encoding scheme.
Obligation	Mandatory
Occurrence	Non-repeatable
Datatype	Date

Term URI	http://purl.org/dc/terms/modified
Defined by	http://purl.org/dc/terms/
Name	Date Modified
Label	Date Modified
Source Definition	Date on which the resource was changed.
Local Definition	Date on which the dataset was changed
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/date
Has Encoding Scheme	http://purl.org/dc/terms/W3CDTF
Constraints	The value must always be taken from the specified encoding scheme.
Obligation	Mandatory
Occurrence	Non-repeatable
Datatype	Date

Term URI	http://purl.org/dc/terms/available
Defined by	http://purl.org/dc/terms/
Name	Date Available
Label	Embargo Date
Source Definition	Date (often a range) that the resource became or will become available.
Local Definition	A date after which the dataset will be made public.
Local Comments	This is only used for datasets under embargo.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/date
Refines	http://purl.org/dc/terms/date
Has Encoding Scheme	http://purl.org/dc/terms/W3CDTF
Obligation	Optional
Occurrence	Non-repeatable
Datatype	Date

Term URI	http://purl.org/dc/elements/1.1/type
Defined by	http://purl.org/dc/elements/1.1
Name	Type
Label	Type

Source Definition	The nature or genre of the content of the resource.
Local Definition	The type of resource.
Local Comments	Choose an appropriate type, most likely "Dataset" or "Image".
Type of term	Element
Obligation	Mandatory
Occurrence	Non-repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/temporal
Defined by	http://purl.org/dc/terms/
Name	Temporal Coverage
Label	Date Range
Source Definition	Temporal characteristics of the resource.
Local Definition	The temporal description of the data set including start date and end date of the collection/creation of the data set.
Local Comments	Temporal period may be a named period, date, or date range. Textual description of the time span covered by the dataset.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/coverage
Has Encoding Scheme	http://www.w3.org/TR/NOTE-datetime
Has Encoding Scheme	http://dublincore.org/documents/dcmi-period
Constraints	Values must be always be taken from the specified encoding scheme.
Obligation	Optional
Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/spatial
Defined by	http://purl.org/dc/terms/
Name	Locality
Label	Locality
Source Definition	Spatial characteristics of the intellectual content of the resource.
Local Definition	The spatial description of the data set specified by a geographic description and geographic coordinates.
Local Comments	Textual description of the geographic area covered by the dataset. Spatial topic may be a named place or a location specified by its geographic coordinates.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/coverage
Has Encoding Scheme	http://dublincore.org/documents/dcmi-point
Has Encoding Scheme	http://dublincore.org/documents/dcmi-box
Has Encoding Scheme	http://www.getty.edu/research/tools/vocabulary/tgn/index.html
Has Encoding Scheme	http://obo.cvs.sourceforge.net/obo/obo/ontology/environmental/gaz.obo?view=log
Constraints	Values must be taken from an encoding scheme. Other encoding schemes may be used where appropriate.
Obligation	Optional

Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/extent
Defined by	http://purl.org/dc/terms/
Name	Extent
Label	Extent
Source Definition	The size or duration of the resource.
Local Definition	The size of the file storage.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/format
Obligation	Mandatory
Occurrence	Non-repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/format
Defined by	http://purl.org/dc/terms/
Name	Format
Label	File Format
Source Definition	The file format, physical medium, or dimensions of the resource.
Local Definition	The format in which the data set is stored.
Local Comments	Code indicating the type of file. This is automatically detected by DSpace, but can be modified manually.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/format
Obligation	Mandatory
Occurrence	Non-repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/isPartOf
Defined by	http://purl.org/dc/terms/
Name	Is Part Of
Label	Is Part Of
Source Definition	A related resource in which the described resource is physically or logically included.
Local Definition	Identifier of the published article with which data set is associated.
Local Comments	The identifier of the publication.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/relation
Has Encoding Scheme	http://purl.org/dc/terms/URI
Has Encoding Scheme	http://www.isbn.org/standards/home/index.asp
Has Encoding Scheme	http://www.issn.org/
Constraints	It is recommended that values be taken from an encoding scheme. Other encoding schemes may be used where appropriate.
Obligation	Required

Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/language
Defined by	http://purl.org/dc/elements/1.1/
Name	Language
Label	Language
Source Definition	A language of the resource.
Local Definition	The language of the data file.
Local Comments	If the data file includes human-readable text, choose an appropriate language.
Type of term	Element
Has Encoding Scheme	http://purl.org/dc/terms/ISO639-2
Has Encoding Scheme	http://purl.org/dc/terms/RFC3066
Occurrence	Non-repeatable
Obligation	Optional
Datatype	String

Term URI	http://wiki.tdwg.org/twiki/bin/view/DarwinCore/SpecificEpithet
Defined by	http://digir.net/schema/conceptual/darwin/manis/1.21/darwin2.xsd
Name	Species
Label	Species
Source Definition	The phylogenetic specific epithet of the cataloged item.
Local Definition	The specific epithet of the scientific name applied to the organism.
Local Comments	As DarwinCore moves to version 1.4, "Species" will be replaced with "SpecificEpithet."
Type of term	Element
Has Encoding Scheme	http://www.itis.gov/
Has Encoding Scheme	http://www.ubio.org/
Occurrence	Repeatable
Obligation	Optional
Datatype	String

Term URI	http://www.ddialliance.org/cocoon/DDI/LIBRARY/Version2-1.xsd?element-definition=contactType&reps=*
Defined by	http://www.ddialliance.org/DDI/dtd/version2-1-all.html
Name	Contact
Label	Contact
Source Definition	Names and addresses of individuals responsible for the work. Individuals listed as contact persons will be used as resource persons regarding problems or questions raised by the user community. The URI attribute should be used to indicate a URN or URL for the homepage of the contact individual. The email attribute is used to indicate an email address for the contact individual.
Local Definition	The individuals responsible for the creation of the data or dataset and their contact

	information.
Type of term	Element
Occurrence	Repeatable
Obligation	Optional
Datatype	String

Term URI	http://www.ddialliance.org/cocoon/DDI/LIBRARY/Version2-1.xsd?element-definition=depositrType&reps=*
Defined by	http://www.ddialliance.org/DDI/dtd/version2-1-all.html
Name	Depositor
Label	Depositor
Source Definition	The name of the person (or institution) who provided this work to the archive storing it.
Local Definition	The name of the person who deposited the dataset in the repository.
Type of term	Element
Occurrence	Repeatable
Obligation	Optional
Datatype	String

Term URI	http://www.loc.gov/standards/premis/v1/Object-v1-1.xsd
Defined by	http://www.oclc.org/research/projects/pmwg/premis-final.pdf
Name	Fixity
Label	Fixity
Source Definition	Information used to verify whether an object has been altered in an undocumented or unauthorized way.
Local Definition	Information used to verify whether an object has been altered in an undocumented or unauthorized way.
Local Comments	Automatically generated by Dryad.
Type of term	Element
Occurrence	Non-repeatable
Obligation	Optional
Datatype	String

Term URI	http://knb.ecoinformatics.org/software/eml/eml-2.0.1/eml-software.html
Defined by	http://knb.ecoinformatics.org/software/eml/eml-2.0.1/eml-software.html
Name	Software
Label	Software
Source Definition	The software element contains general information about a software resource that is being documented. This field is intended to give information for software tools that are needed to interpret a dataset, software that was written to process a resource, or software as a resource in itself. It is based on eml-resource and Open Software Description (OSD) a W3C submission. There can be multiple implementations within a software package because a physical software package can run on multiple hardware and/or operating systems.
Source Comments	The eml-software module contains general information that describes software resources. This module is intended to fully document software that is needed in order to view a resource (such as a dataset) or to process a dataset. The software module is also imported into the eml-methods module in order to document what software was used to process or perform quality control procedures on a dataset.
Local Definition	Software used to produce the data.

Local Comments	A Dryad-specific controlled vocabulary may be developed to populate this field.
Type of term	Element
Occurrence	Repeatable
Obligation	Optional
Datatype	String

A.3. Term Usage – Publication

Term URI	http://purl.org/dc/elements/1.1/identifier
Defined by	http://purl.org/dc/elements/
Name	Identifier
Label	Identifier
Source Definition	An unambiguous reference to the resource within a given context.
Source Comments	Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Example formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).
Local Definition	The Digital Object Identifier of a journal article.
Comments	Select URI and enter the DOI of the publication in URL form, if available. Otherwise, use the most "permanent" URL available that represents the publication.
Type of term	Element
Refined by	http://purl.org/dc/terms/bibliographicCitation
Obligation	Mandatory
Occurrence	Non-repeatable
Datatype	URI

Term URI	http://purl.org/dc/terms/bibliographicCitation
Defined by	http://purl.org/dc/terms/
Name	Bibliographic Citation
Label	Bibliographic Citation
Source Definition	A bibliographic reference for the resource
Source Comment	Recommended practice is to include sufficient bibliographic detail to identify the resource as unambiguously as possible, whether or not the citation is in standard format.
Local Definition	The citation information for the journal article.
Local Comments	A plain-text citation. Currently, copied from the publisher's site if available. Some attempt should be made to normalize case (don't include all caps). In the future, this may be automatically generated, to provide consistent formatting.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/identifier
Occurrence	Repeatable
Obligation	Optional
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/creator
Defined by	http://purl.org/dc/elements/1.1/
Name	Creator

Label	Author
Source Definition	An entity primarily responsible for making the content of the resource.
Source Comments	Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
Local Definition	Author(s) of the article.
Local Comments	List the full names of authors. Do not just copy abbreviated names from a citation, try to find the actual names. Currently, authority control is manual. Author/contributor names will typically be formatted as "Lastname, Firstname" OR as "Lastname, A. B.", depending on the text received from the publisher. We will optimize for searches on lastname only, knowing that Firstname may often only be available as initials. We will store email addresses for disambiguation. Initially, we won't track email addresses in normal metadata, just letting DSpace track the submitter in the provenance.
Type of term	Element
Occurrence	Repeatable
Obligation	Mandatory
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/contributor
Defined by	http://purl.org/dc/elements/1.1/
Name	Contributor
Label	Coauthor
Source Definition	An entity responsible for making contributions to the resource.
Source Comments	Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
Local Definition	Coauthor(s) of the article.
Type of term	Element
Occurrence	Repeatable
Obligation	Recommended
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/title
Defined by	http://purl.org/dc/elements/1.1/
Name	Title
Label	Title
Source Definition	A name given to the resource.
Local Definition	Title of the article.
Type of term	Element
Occurrence	Non-repeatable
Obligation	Mandatory
Datatype	String

Term URI	http://purl.org/dc/terms/issued
Defined by	http://purl.org/dc/terms/
Name	Date of Issue
Label	Date of Issue
Source	Date of formal issuance (e.g., publication) of the resource.

Definition	
Local Definition	Date of publication.
Local Comments	The official date of publication. Year is required. Include month and day if possible.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/date
Has Encoding Scheme	http://purl.org/dc/terms/W3CDTF
Constraints	The value must always be taken from the specified encoding scheme.
Obligation	Mandatory
Occurrence	Non-repeatable
Datatype	Date

Term URI	http://purl.org/dc/elements/1.1/rights
Defined by	http://purl.org/dc/elements/1.1/
Name	Rights
Label	Rights Statement
Source Definition	Information about rights held in and over the resource.
Local Definition	Statement regarding rights held in and over the resource.
Local Comments	A short human-readable phrase describing the access rights, which may also be machine-readable. For example: • Creative Commons license (CC-BY) • Public Domain • Copyright held by publisher A blank value indicates that the dataset is a "normal" status item.
Type of term	Element
Occurrence	Non-repeatable
Obligation	Mandatory
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/language
Defined by	http://purl.org/dc/elements/1.1/
Name	Language
Label	Language
Source Definition	A language of the resource.
Local Definition	The language of the text.
Local Comments	Choose an appropriate language.
Type of term	Element
Has Encoding Scheme	http://purl.org/dc/terms/ISO639-2
Has Encoding Scheme	http://purl.org/dc/terms/RFC3066
Occurrence	Non-repeatable
Obligation	Recommended
Datatype	String

Term URI	http://purl.org/dc/terms/abstract
Defined by	http://purl.org/dc/terms/
Name	Abstract
Label	Abstract
Source Definition	A summary of the resource.
Local Definition	The abstract from the publication.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/description
Refines	http://purl.org/dc/terms/description
Occurrence	Non-repeatable
Obligation	Required
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/subject
Defined by	http://purl.org/dc/elements/1.1/
Name	Subject
Label	Subject Keywords
Source Definition	The topic of the resource.
Source Comments	Typically, a Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.
Local Definition	Article keywords.
Local Comments	Initially, only explicitly stated keywords will be cataloged as such. In the future, we hope to perform more automatic keyword extraction. Species/taxa names will be cataloged as such, and will not be replicated as keywords (though they will be searchable as keywords).
Type of term	Element
Occurrence	Repeatable
Obligation	Required
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/type
Defined by	http://purl.org/dc/elements/1.1
Name	Type
Label	Type
Source Definition	The nature or genre of the content of the resource.
Local Definition	The type of resource.
Local Comments	Choose "Article."
Type of term	Element
Obligation	Mandatory
Occurrence	Non-repeatable
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/publisher
Defined by	http://purl.org/dc/elements/1.1
Name	Publisher

Label	Publisher
Source Definition	An entity responsible for making the resource available.
Local Definition	Journal publisher.
Local Comments	The original publisher of the article. Note: This should be a publishing company, which is normally different than the journal name.
Type of term	Element
Obligation	Mandatory
Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/hasPartOf
Defined by	http://purl.org/dc/terms/
Name	Has Part
Label	Has Part
Source Definition	The described resource includes the referenced resource either physically or logically.
Local Definition	The identifier of the dataset(s) that underlie the publication.
Local Comments	The DOI for the publication.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/relation
Has Encoding Scheme	http://purl.org/dc/terms/URI
Has Encoding Scheme	http://www.isbn.org/standards/home/index.asp
Has Encoding Scheme	http://www.issn.org/
Constraints	It is recommended that values be taken from an encoding scheme. Other encoding schemes may be used where appropriate.
Obligation	Required
Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/isPartOf
Defined by	http://purl.org/dc/terms/
Name	Is Part Of
Label	Is Part Of
Source Definition	A related resource in which the described resource is physically or logically included.
Local Definition	Digital Object Identifier of the published article with which data set is associated.
Comments	Example - urn:ISSN:0740-8188
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/relation
Has Encoding Scheme	http://purl.org/dc/terms/URI
Has Encoding Scheme	http://www.isbn.org/standards/home/index.asp
Has Encoding Scheme	http://www.issn.org/
Constraints	It is recommended that values be taken from an encoding scheme. Other encoding

	schemes may be used where appropriate.
Obligation	Required
Occurrence	Repeatable
Datatype	String

Appendix B: Singapore Framework Components – Level One Application Profile

B.1. Functional requirements

1. Scope

    a. Metadata:

        i.   In scope: Dublin Core elements and any additional elements from domain-specific namespaces, or namespaces that perform required functions, or provide required services.

        ii. Out of scope: Metadata formats that do not meet the stated requirements.

    b. Identifiers:

        i.   In scope: Use of identifiers to link related resources, use of identifiers for the description itself.

        ii. Out of scope: Other uses of identifiers.

    c. Controlled vocabularies:

        i.   In scope: Ensuring that the application profile supports various means of access, that the process of deposition and metadata creation is eased by the assistance of controlled vocabularies, and that they quality of the metadata is controlled using existing terminologies. Controlled vocabularies within the scope include: classification schemes, controlled vocabularies, and name authority lists.

        ii. Out of scope: Permanent decisions concerning terminology solutions.
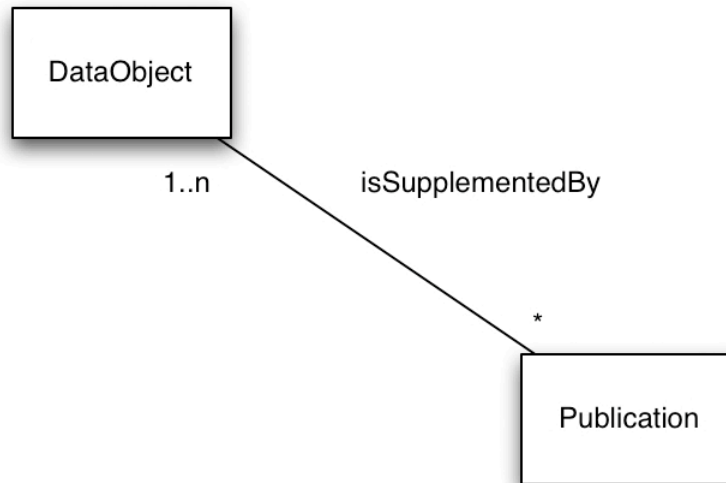
    d. Complex objects:

        i.   In scope: Being aware of current work being undertaken in this area, and using existing work to formulate requirements.

        ii. Out of scope: Decisions on how to model complex objects.

    e. Citations and references

        i.   In scope: Bibliographic citations for published articles with underlying datasets hosted by Dryad.

        ii. Out of scope: Citation analysis, complex bibliometrics.

2. Stakeholders and designated community

    a. Designated community: Researchers in the field of evolutionary biology who are generating data and reusing data for their own projects.

    b. Stakeholder community: Evolutionary biologists, journal publishers in the field of evolutionary biology, professional societies in evolutionary biology, and NESCent—a research center for synthetic research addressing fundamental questions in evolutionary biology.

3. Requirements gathering

    a. Methodology: The needs and goals of these individuals and groups identified as stakeholders and community members were identified in a workshop held in December 2006 at NESCent in Durham, North Carolina. Among initial questions addressed at the workshop were: What is the minimum number of metadata elements required? What functions will the Dryad scheme support? Answers to these questions have informed the development of Dryad's functional requirements and the metadata framework. In addition, the repository development team is currently undertaking two studies intended to assess data sharing attitudes and behaviors: a use case study and an online survey.

    b. Scenarios and use case:

        i. A user submitting a dataset as a requirement for publication.

        ii. A user searching for datasets that are applicable to their own research, or for a particular author, in order to use the information for their own project.

4. Functional requirements specification

    a. Computer-aided metadata generation and augmentation.

    b. Specialized modules linking publications and underlying datasets.

    c. Data and metadata quality control through the integration of manual and automatic techniques.

    d. Support for identity, authority and data security.

    e. Support for basic metadata repository functions, such as resource discovery, sharing, and interoperability.

B.2. Domain model

B.3. Description Set Profile (DSP)

```
<?xml version="1.0"?>

<DescriptionSetTemplate>

<DescriptionTemplate ID="DataObject" maxOccur="1" minOccur="1">

<StatementTemplate minOccur="1" maxOccur="1" type="literal">
<Property>http://purl.org/dc/elements/1.1/type</Property>
</StatementTemplate>

<StatementTemplate minOccur="1" maxOccur="1" type="literal">
<Property>http://purl.org/dc/elements/1.1/title</Property>
</StatementTemplate>

<StatementTemplate minOccur="1" type="nonliteral">
<Property>http://purl.org/dc/elements/1.1/subject</Property>
<NonliteralConstraint>
<ValueURIOccurrence>optional</ValueURIOccurrence>
```

```
<VocabularyEncodingSchemeOccurrence>optional</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint maxOccur="1">
<SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>
<LanguageOccurrence>optional</LanguageOccurrence>
</ValueStringConstraint>
</NonliteralConstraint>
</StatementTemplate>

<StatementTemplate minOccur="1" maxOccur="1" type="literal">
<Property>http://purl.org/dc/elements/1.1/identifier</Property>
<LiteralConstraint>
<SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence>
<SyntaxEncodingScheme>http://purl.org/dc/terms/URI</SyntaxEncodingScheme>
<SyntaxEncodingScheme>http://www.isbn.org/standards/home/index.asp</SyntaxEncodingScheme>
<SyntaxEncodingScheme>http://www.issn.org/</SyntaxEncodingScheme>
</LiteralConstraint>
</StatementTemplate>

<StatementTemplate minOccur="1" type="nonliteral">
<Property>http://purl.org/dc/elements/1.1/creator</Property>
<NonliteralConstraint>
<VocabularyEncodingSchemeOccurrence>optional</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint maxOccur="1">
<SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>
<LanguageOccurrence>optional</LanguageOccurrence>
</ValueStringConstraint>
</NonliteralConstraint>
</StatementTemplate>

<StatementTemplate minOccur="0" type="nonliteral">
<Property>http://purl.org/dc/elements/1.1/contributor</Property>
<NonliteralConstraint>
<VocabularyEncodingSchemeOccurrence>optional</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint maxOccur="1">
<SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>
<LanguageOccurrence>optional</LanguageOccurrence>
</ValueStringConstraint>
```

```
</NonliteralConstraint>
</StatementTemplate>

<StatementTemplate minOccur="1" maxOccur="1" type="literal">
<Property>http://purl.org/dc/terms/available</Property>
<LiteralConstraint>
<SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence>
<SyntaxEncodingScheme>http://purl.org/dc/terms/W3CDTF</SyntaxEncodingScheme>
</LiteralConstraint>
</StatementTemplate>

<StatementTemplate minOccur="1" maxOccur="1" type="literal">
<Property>http://purl.org/dc/terms/issued</Property>
<LiteralConstraint>
<SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence>
<SyntaxEncodingScheme>http://purl.org/dc/terms/W3CDTF</SyntaxEncodingScheme>
</LiteralConstraint>
</StatementTemplate>

<StatementTemplate maxOccur="1" type="nonliteral">
<Property>http://purl.org/dc/elements/1.1/description</Property>
</StatementTemplate>

<StatementTemplate minOccur="1" maxOccur="1" type="literal">
<Property>http://purl.org/dc/elements/1.1/rights</Property>
<NonliteralConstraint>
<ValueURIOccurrence>mandatory</ValueURIOccurrence>
<VocabularyEncodingSchemeOccurrence>disallowed</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint maxOccur="0"></ValueStringConstraint>
</NonliteralConstraint>
</StatementTemplate>

<StatementTemplate minOccur="1" maxOccur="1" type="nonliteral">
<Property>http://purl.org/dc/elements/1.1/language</Property>
<NonliteralConstraint>
<ValueURIOccurrence>disallowed</ValueURIOccurrence>
<VocabularyEncodingSchemeOccurrence>disallowed</VocabularyEncodingSchemeOccurrence>
```

```xml
<ValueStringConstraint maxOccur="1">
<SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence>
<SyntaxEncodingScheme>http://purl.org/dc/terms/ISO639-2</SyntaxEncodingScheme>
<SyntaxEncodingScheme>http://purl.org/dc/terms/RFC3066</SyntaxEncodingScheme>
<LanguageOccurrence>disallowed</LanguageOccurrence>
</ValueStringConstraint>
</NonliteralConstraint>
</StatementTemplate>


<StatementTemplate minOccur="1" type="nonliteral">
<Property>http://purl.org/dc/terms/isPartOf</Property>
<NonliteralConstraint>
<ValueURIOccurrence>mandatory</ValueURIOccurrence>
<VocabularyEncodingSchemeOccurrence>disallowed</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint maxOccur="1">
<SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence>
<LanguageOccurrence>optional</LanguageOccurrence>
</ValueStringConstraint>
</NonliteralConstraint>
</StatementTemplate>


<StatementTemplate minOccur="1" maxOccur="1" type="literal">
<Property>http://purl.org/dc/terms/modified</Property>
<LiteralConstraint><SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence>
>
<SyntaxEncodingScheme>http://purl.org/dc/terms/W3CDTF</SyntaxEncodingScheme>
</LiteralConstraint>
</StatementTemplate>


<StatementTemplate minOccur="1" maxOccur="1" type="nonliteral">
<Property>http://purl.org/dc/elements/1.1/format</Property>
<NonliteralConstraint>
<ValueURIOccurrence>disallowed</ValueURIOccurrence>
<VocabularyEncodingSchemeOccurrence>disallowed</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint
maxOccur="1"><SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence>
<SyntaxEncodingScheme>http://purl.org/dc/terms/IMT</SyntaxEncodingScheme>
<LanguageOccurrence>disallowed</LanguageOccurrence>
```

```
</ValueStringConstraint></NonliteralConstraint></StatementTemplate>


<StatementTemplate minOccur="1" maxOccur="1" type="literal">
<Property>http://purl.org/dc/terms/extent</Property>
</StatementTemplate>


<StatementTemplate minOccur="0" type="literal">
<Property>http://purl.org/dc/terms/spatial</Property>
<LiteralConstraint><SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence
>
<SyntaxEncodingScheme>http://www.getty.edu/research/tools/vocabulary/tgn/index.html</SyntaxEncodin
gScheme>
<SyntaxEncodingScheme>http://obo.cvs.sourceforge.net/obo/obo/ontology/environmental/gaz.obo?view=l
og</SyntaxEncodingScheme>
<SyntaxEncodingScheme>http://dublincore.org/documents/dcmi-box</SyntaxEncodingScheme>
<SyntaxEncodingScheme>http://dublincore.org/documents/dcmi-point</SyntaxEncodingScheme>
</LiteralConstraint>
</StatementTemplate>


<StatementTemplate minOccur="0" type="literal">
<Property>http://purl.org/dc/terms/temporal</Property>
<LiteralConstraint><SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence
>
<SyntaxEncodingScheme>http://www.w3.org/TR/NOTE-datetime</SyntaxEncodingScheme>
<SyntaxEncodingScheme>http://dublincore.org/documents/dcmi-period</SyntaxEncodingScheme>
</LiteralConstraint>
</StatementTemplate>


<StatementTemplate minOccur="0" type="literal">
<Property>http://wiki.tdwg.org/twiki/bin/view/DarwinCore/SpecificEpithet</Property>
<LiteralConstraint><SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence
>
<SyntaxEncodingScheme>http://www.ubio.org/</SyntaxEncodingScheme>
<SyntaxEncodingScheme>http://www.itis.gov/</SyntaxEncodingScheme>
</LiteralConstraint>
</StatementTemplate>


<StatementTemplate minOccur="0" type="nonliteral">
```

```
<Property>http://www.ddialliance.org/cocoon/DDI/LIBRARY/Version2-1.xsd?element-
definition=contactType&amp;reps=*</Property>

<NonliteralConstraint>

<VocabularyEncodingSchemeOccurrence>optional</VocabularyEncodingSchemeOccurrence>

<ValueStringConstraint maxOccur="1">

<SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>

<LanguageOccurrence>optional</LanguageOccurrence>

</ValueStringConstraint>

</NonliteralConstraint>

</StatementTemplate>


<StatementTemplate type="nonliteral">

<Property>http://www.ddialliance.org/cocoon/DDI/LIBRARY/Version2-1.xsd?element-
definition=depositrType&amp;reps=*</Property>

<NonliteralConstraint>

<VocabularyEncodingSchemeOccurrence>optional</VocabularyEncodingSchemeOccurrence>

<ValueStringConstraint maxOccur="1">

<SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>

<LanguageOccurrence>optional</LanguageOccurrence>

</ValueStringConstraint>

</NonliteralConstraint>

</StatementTemplate>


<StatementTemplate type="literal">

<Property>http://www.loc.gov/standards/premis/v1/Object-v1-1.xsd</Property>

</StatementTemplate>


<StatementTemplate type="nonliteral">

<Property>http://knb.ecoinformatics.org/software/eml/eml-2.0.1/eml-software.html</Property>

<NonliteralConstraint>

<VocabularyEncodingSchemeOccurrence>optional</VocabularyEncodingSchemeOccurrence>

<ValueStringConstraint maxOccur="1">

<SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>

<LanguageOccurrence>optional</LanguageOccurrence>

</ValueStringConstraint>

</NonliteralConstraint>

</StatementTemplate>
```

```
</DescriptionTemplate>


<DescriptionTemplate ID="Publication" maxOccur="1" minOccur="1">

<StatementTemplate minOccur="1" maxOccur="1" type="literal">
<Property>http://purl.org/dc/elements/1.1/type</Property>
</StatementTemplate>

<StatementTemplate minOccur="1" maxOccur="1" type="literal">
<Property>http://purl.org/dc/terms/issued</Property>
<LiteralConstraint>
<SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence>
<SyntaxEncodingScheme>http://purl.org/dc/terms/W3CDTF</SyntaxEncodingScheme>
</LiteralConstraint>
</StatementTemplate>

<StatementTemplate minOccur="1" type="literal">
<Property>http://purl.org/dc/terms/bibliographicCitation</Property>
<LiteralConstraint>
<SyntaxEncodingSchemeOccurrence>optional</SyntaxEncodingSchemeOccurrence>
</LiteralConstraint>
</StatementTemplate>

<StatementTemplate minOccur="1" maxOccur="1" type="literal">
<Property>http://purl.org/dc/elements/1.1/identifier</Property>
<LiteralConstraint>
<SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence>
<SyntaxEncodingScheme>http://purl.org/dc/terms/URI</SyntaxEncodingScheme>
<SyntaxEncodingScheme>http://www.isbn.org/standards/home/index.asp</SyntaxEncodingScheme>
<SyntaxEncodingScheme>http://www.issn.org/</SyntaxEncodingScheme>
</LiteralConstraint>
</StatementTemplate>

<StatementTemplate type="nonliteral">
```

```
<Property>http://purl.org/dc/terms/abstract</Property>
</StatementTemplate>


<StatementTemplate type="nonliteral">
<Property>http://purl.org/dc/elements/1.1/publisher</Property>
<NonliteralConstraint descriptionTemplateID="agent">
<VocabularyEncodingSchemeOccurrence>optional</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint maxOccur="1">
<SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>
<LanguageOccurrence>optional</LanguageOccurrence>
</ValueStringConstraint>
</NonliteralConstraint>
</StatementTemplate>


<StatementTemplate minOccur="1" maxOccur="1" type="literal">
<Property>http://purl.org/dc/elements/1.1/title</Property>
</StatementTemplate>


<StatementTemplate minOccur="1" type="nonliteral">
<Property>http://purl.org/dc/elements/1.1/subject</Property>
<NonliteralConstraint>
<ValueURIOccurrence>optional</ValueURIOccurrence>
<VocabularyEncodingSchemeOccurrence>optional</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint maxOccur="1">
<SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>
<LanguageOccurrence>optional</LanguageOccurrence>
</ValueStringConstraint>
</NonliteralConstraint>
</StatementTemplate>


<StatementTemplate minOccur="1" maxOccur="1" type="nonliteral">
<Property>http://purl.org/dc/elements/1.1/rights</Property>
<NonliteralConstraint>
<ValueURIOccurrence>mandatory</ValueURIOccurrence>
<VocabularyEncodingSchemeOccurrence>disallowed</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint maxOccur="0">
</ValueStringConstraint>
```

```
</NonliteralConstraint>
</StatementTemplate>

<StatementTemplate minOccur="1" type="nonliteral">
<Property>http://purl.org/dc/elements/1.1/creator</Property>
<NonliteralConstraint>
<VocabularyEncodingSchemeOccurrence>optional</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint maxOccur="1">
<SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>
<LanguageOccurrence>optional</LanguageOccurrence>
</ValueStringConstraint>
</NonliteralConstraint>
</StatementTemplate>

<StatementTemplate minOccur="0" type="nonliteral">
<Property>http://purl.org/dc/elements/1.1/contributor</Property>
<NonliteralConstraint>
<VocabularyEncodingSchemeOccurrence>optional</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint maxOccur="1">
<SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>
<LanguageOccurrence>optional</LanguageOccurrence>
</ValueStringConstraint>
</NonliteralConstraint>
</StatementTemplate>

<StatementTemplate minOccur="1" maxOccur="1" type="nonliteral">
<Property>http://purl.org/dc/elements/1.1/language</Property>
<NonliteralConstraint>
<ValueURIOccurrence>disallowed</ValueURIOccurrence>
<VocabularyEncodingSchemeOccurrence>disallowed</VocabularyEncodingSchemeOccurrence>
<VocabularyEncodingScheme></VocabularyEncodingScheme><ValueStringConstraint maxOccur="1">
<SyntaxEncodingSchemeOccurrence>mandatory</SyntaxEncodingSchemeOccurrence>
<SyntaxEncodingScheme>http://purl.org/dc/terms/ISO639-2</SyntaxEncodingScheme>
<SyntaxEncodingScheme>http://purl.org/dc/terms/RFC3066</SyntaxEncodingScheme>
<LanguageOccurrence>disallowed</LanguageOccurrence>
</ValueStringConstraint></NonliteralConstraint>
</StatementTemplate>
```

```
<StatementTemplate minOccur="1" type="nonliteral">
<Property>http://purl.org/dc/terms/hasPart</Property>
<NonliteralConstraint>
<ValueURIOccurrence>mandatory</ValueURIOccurrence>
<VocabularyEncodingSchemeOccurrence>disallowed</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint maxOccur="1">
<SyntaxEncodingSchemeOccurrence>optional</SyntaxEncodingSchemeOccurrence>
<LanguageOccurrence>optional</LanguageOccurrence>
</ValueStringConstraint>
</NonliteralConstraint>
</StatementTemplate>

<StatementTemplate minOccur="1" type="nonliteral">
<Property>http://purl.org/dc/terms/isPartOf</Property>
<NonliteralConstraint>
<ValueURIOccurrence>mandatory</ValueURIOccurrence>
<VocabularyEncodingSchemeOccurrence>disallowed</VocabularyEncodingSchemeOccurrence>
<ValueStringConstraint
maxOccur="1"><SyntaxEncodingSchemeOccurrence>optional</SyntaxEncodingSchemeOccurrence>
<LanguageOccurrence>optional</LanguageOccurrence>
</ValueStringConstraint></NonliteralConstraint></StatementTemplate>

</DescriptionTemplate>

</DescriptionSetTemplate>
```

B.4. Usage guidelines

Cataloging and usage guidelines can be found at the following URL:

https://www.nescent.org/wg_digitaldata/Public:Dryad_Level_One_Cataloging_Guidelines

B.5. Encoding syntax guidelines

Not applicable.

Appendix C: The Dryad Level Two Application Profile – Proposal

Note: Documentation in "gray" includes recent suggestions for how to handle a metadata issue, and has not yet been reviewed by the Dryad development team.

C.1. Descriptive Header

Title	Dryad Level Two Application Profile
Contributor	Metadata Research Center (MRC)
Contributor	National Evolutionary Synthesis Center (NESCent)
Coverage.spatial	USA
Date.issued	2008-04-07
Description	This document describes the Level Two application profile designed by the MRC and NESCent for metadata used with the Phase Two implementation of Dryad.
Format	Text
Language	Eng
Status	Version 2.0
Subject	Metadata
Subject.category	Information management

C.2. Term Usage – Dataset

Term URI	http://purl.org/dc/terms/isPartOf
Defined by	http://purl.org/dc/terms/
Name	Is Part Of
Label	Is Part Of
Source Definition	A related resource in which the described resource is physically or logically included.
Local Definition	Identifier of the published article with which data set is associated.
Local Comments	The identifier of the publication. In this implementation, multiple data objects can be associated with multiple publications.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/relation
Has Encoding Scheme	http://purl.org/dc/terms/URI
Has Encoding Scheme	http://www.isbn.org/standards/home/index.asp
Has Encoding Scheme	http://www.issn.org/
Constraints	It is recommended that values be taken from an encoding scheme. Other encoding schemes may be used where appropriate.
Obligation	Optional
Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/isVersionOf
Defined by	http://purl.org/dc/terms/
Name	Is Version Of

Label	Is Version Of
Source Definition	A related resource of which the described resource is a version, edition, or adaptation.
Local Definition	Identifier of a version or adaptation of a dataset.
Local Comments	
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/relation
Has Encoding Scheme	http://purl.org/dc/terms/URI
Has Encoding Scheme	http://www.isbn.org/standards/home/index.asp
Has Encoding Scheme	http://www.issn.org/
Constraints	It is recommended that values be taken from an encoding scheme. Other encoding schemes may be used where appropriate.
Obligation	Optional
Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/isFormatOf
Defined by	http://purl.org/dc/terms/
Name	Is Format Of
Label	Is Format Of
Source Definition	A related resource that is substantially the same as the described resource, but in another format.
Local Definition	Identifier or name of a dataset in another format.
Local Comments	An acceptable field entry could be a name or title.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/relation
Obligation	Optional
Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/references
Defined by	http://purl.org/dc/terms/
Name	References
Label	References
Source Definition	A related resource that is referenced, cited, or otherwise pointed to by the described resource.
Local Definition	Identifier or name of a dataset that is referenced by the described dataset.
Local Comments	An acceptable field entry could be a name or title.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/relation
Obligation	Optional
Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/isReferencedBy
Defined by	http://purl.org/dc/terms/
Name	Is Referenced By
Label	Is Referenced By

Source Definition	A related resource that references, cites, or otherwise points to the described resource.
Local Definition	Identifier or name of a dataset that reference the described dataset.
Local Comments	An acceptable field entry could be a name or title.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/relation
Obligation	Optional
Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/requires
Defined by	http://purl.org/dc/terms/
Name	Requires
Label	Requires
Source Definition	A related resource that is required by the described resource to support its function, delivery, or coherence.
Local Definition	Identifier or name of a dataset that is required to understand the described dataset.
Local Comments	An acceptable field entry could be a name or title.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/relation
Obligation	Optional
Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/isRequiredBy
Defined by	http://purl.org/dc/terms/
Name	Is Required By
Label	Is Required By
Source Definition	A related resource that requires the described resource to support its function, delivery, or coherence.
Local Definition	Identifier or name of a dataset that requires the described dataset.
Local Comments	An acceptable field entry could be a name or title.
Type of term	Element refinement
Refines	http://purl.org/dc/elements/1.1/relation
Obligation	Optional
Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/terms/provenance
Defined by	http://purl.org/dc/terms/
Name	Provenance
Label	Provenance
Source Definition	A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation.
Source Comments	The statement may include a description of any changes successive custodians made to the resource.
Local Definition	A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation.
Type of term	Property
Obligation	Optional
Occurrence	Repeatable

Datatype	String

Term URI	http://purl.org/dc/terms/rightsHolder
Defined by	http://purl.org/dc/terms/
Name	Rights Holder
Label	Rights Holder
Source Definition	A person or organization owning or managing rights over the resource.
Local Definition	A person or organization owning or managing rights over the resource.
Local Comments	Entries for this field would include those who have been identified as the "owners" of the data, and would not necessarily be the authors and coauthors.
Type of term	Property
Obligation	Optional
Occurrence	Repeatable
Datatype	String

Term URI	http://purl.org/dc/elements/1.1/source
Defined by	http://purl.org/dc/elements/1.1/
Name	Source
Label	Data Source
Source Definition	A related resource from which the described resource is derived.
Source Comments	The described resource may be derived from the related resource in whole or in part. Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.
Local Definition	A related resource from which the dataset is derived.
Local Comments	Entries for this field would include those who have been identified as the "owners" of the data, and would not necessarily be the authors and coauthors.
Type of term	Property
Obligation	Optional
Occurrence	Repeatable
Datatype	String

Term URI	http://knb.ecoinformatics.org/software/eml/eml-2.0.1/eml-protocol.html
Defined by	http://knb.ecoinformatics.org/software/eml/eml-2.0.1/eml-protocol.html
Name	Protocol
Label	Protocol
Source Definition	The EML Protocol Module is used to define abstract, prescriptive procedures for generating or processing data. Conceptually, a protocol is a standardized method.
Source Comments	Eml-protocol resembles eml-methods; however, eml-methods is descriptive (often written in the declarative mood: "I took five subsamples...") whereas eml-protocol is prescriptive (often written in the imperative mood: "Take five subsamples..."). A protocol may have versions, whereas methods (as used in eml-methods) should not.
Local Definition	Protocol used to generate the data.
Type of term	Element
Occurrence	Repeatable
Obligation	Optional
Datatype	String

Term URI	http://knb.ecoinformatics.org/software/eml/eml-2.0.1/eml-methods.html

Defined by	http://knb.ecoinformatics.org/software/eml/eml-2.0.1/eml-methods.html
Name	Methods
Label	Methods
Source Definition	The eml-methods module describes the methods followed in the creation of the dataset, including description of field, laboratory and processing steps, sampling methods and units, quality control procedures.
Source Comments	The eml-methods module is used to describe the actual procedures that are used in the creation or the subsequent processing of a dataset. Likewise, eml-methods is used to describe processes that have been used to define / improve the quality of a data file, or to identify potential problems with the data file. Note that the eml-protocol module is intended to be used to document a prescribed procedure, whereas the eml-method module is used to describe procedures that were actually performed. The distinction is that the use of the term "protocol" is used in the "prescriptive" sense, and the term "method" is used in the "descriptive" sense. This distinction allows managers to build a protocol library of well-known, established protocols (procedures), but also document what procedure was truly performed in relation to the established protocol. The method may have diverged from the protocol purposefully, or perhaps incidentally, but the procedural lineage is still preserved and understandable.
Local Definition	Methods used to generate the data.
Type of term	Element
Occurrence	Repeatable
Obligation	Optional
Datatype	String

Term URI	http://www.loc.gov/standards/premis/v1/Object-v1-1.xsd
Defined by	http://www.oclc.org/research/projects/pmwg/premis-final.pdf
Name	Original Name
Label	Original Name
Source Definition	The name of the object as submitted to or harvested by the repository, before any renaming by the repository.
Source Comments	The name used within the preservation repository may not be known outside of the repository. A depositor might need to request a file by its original name. Also, the repository may need to reconstruct internal links for dissemination.
Local Definition	The name of the object as submitted to or harvested by the repository, before any renaming by the repository.
Type of term	Element
Occurrence	Non-repeatable
Obligation	Optional
Datatype	String

Term URI	http://www.loc.gov/standards/premis/v1/Object-v1-1.xsd
Defined by	http://www.oclc.org/research/projects/pmwg/premis-final.pdf
Name	Dependency
Label	Dependency
Source Definition	Information about a non-software component or associated file needed in order to use or render the representation or file, for example, a schema, a DTD, or an entity file declaration.
Local Definition	Information about a non-software component or associated file needed in order to use or render the representation or file, for example, a schema, a DTD, or an entity file declaration.
Local Comments	This field serves the important purpose of providing documentation needed to understand how to process or understand a data object. Includes dependencyName

	and dependencyIdentifier.
Type of term	Element
Occurrence	Non-repeatable
Obligation	Optional
Datatype	String

Term URI	http://wiki.tdwg.org/twiki/bin/view/DarwinCore/Genus
Defined by	http://digir.net/schema/conceptual/darwin/manis/1.21/darwin2.xsd
Name	Genus
Label	Genus
Source Definition	The name of the genus in which the organism is classified.
Local Definition	The name of the genus in which the organism is classified.
Local Comments	This field will be populated by a yet-to-be-determined controlled vocabulary.
Type of term	Element
Occurrence	Repeatable
Obligation	Optional
Datatype	String

Term URI	http://www.omg.org/cgi-bin/doc?formal/03-02-03
Defined by	http://www.omg.org/cgi-bin/doc?formal/03-02-03
Name	BioSequence
Label	Gene
Source Definition	Specifies classes that describe the sequence information for a BioSequence.
Source Comments	Describes a known gene or sequence. A BioSequence is a representation of a DNA, RNA, or protein sequence. It can be represented by a Clone, Gene, or the sequence.
Local Definition	Specifies classes that describe the sequence information for a BioSequence.
Type of term	Element
Has Encoding Scheme	This field will be populated by a yet-to-be-determined controlled vocabulary.
Occurrence	Repeatable
Obligation	Optional
Datatype	String