Apiary Project
HIGH-THROUGHPUT WORKFLOW FOR COMPUTER-ASSISTED HUMAN PARSING OF BIOLOGICAL SPECIMEN LABEL DATA

William E. Moen—william.moen@unt.edu
Jane Huang—janehuang@my.unt.edu
Melody McCotter—melody.mccotter@unt.edu
Jason H. Best—jbest@brit.org
Amanda K. Neill—aneill@brit.org

AN APPLICATION PROFILE USING DARWIN CORE RENDERED IN THE NEW DUBLIN CORE APPLICATION PROFILE FRAMEWORK

www.apiaryproject.org

Dublin Core Application Profile (DCAP)

Defines metadata records that meet specific application needs while providing semantic interoperability with other applications on the basis of globally defined vocabularies and models.

Components of DCAP:

A DCAP is a set of documents addressing the following components:

- **Functional Requirements (mandatory):** Describes the functions that the application profile (AP) is designed to support, as well as functions that are out of scope.
- **Domain Model (mandatory):** Describes the objects metadata will describe, and the relationships between those objects.
- **Description Set Profile (DSP) (mandatory):** Defines a set of metadata records that are valid instances of an AP.
- **Usage Guidelines (optional):** Describes how to apply the AP, how the used properties are intended to be used in the application context, etc.
- **Encoding Syntax Guidelines (optional):** Describes AP-specific syntaxes and/or syntax guideline, if any.

(Source: http://dublincore.org/documents/singapore-framework/)

DCAP Framework
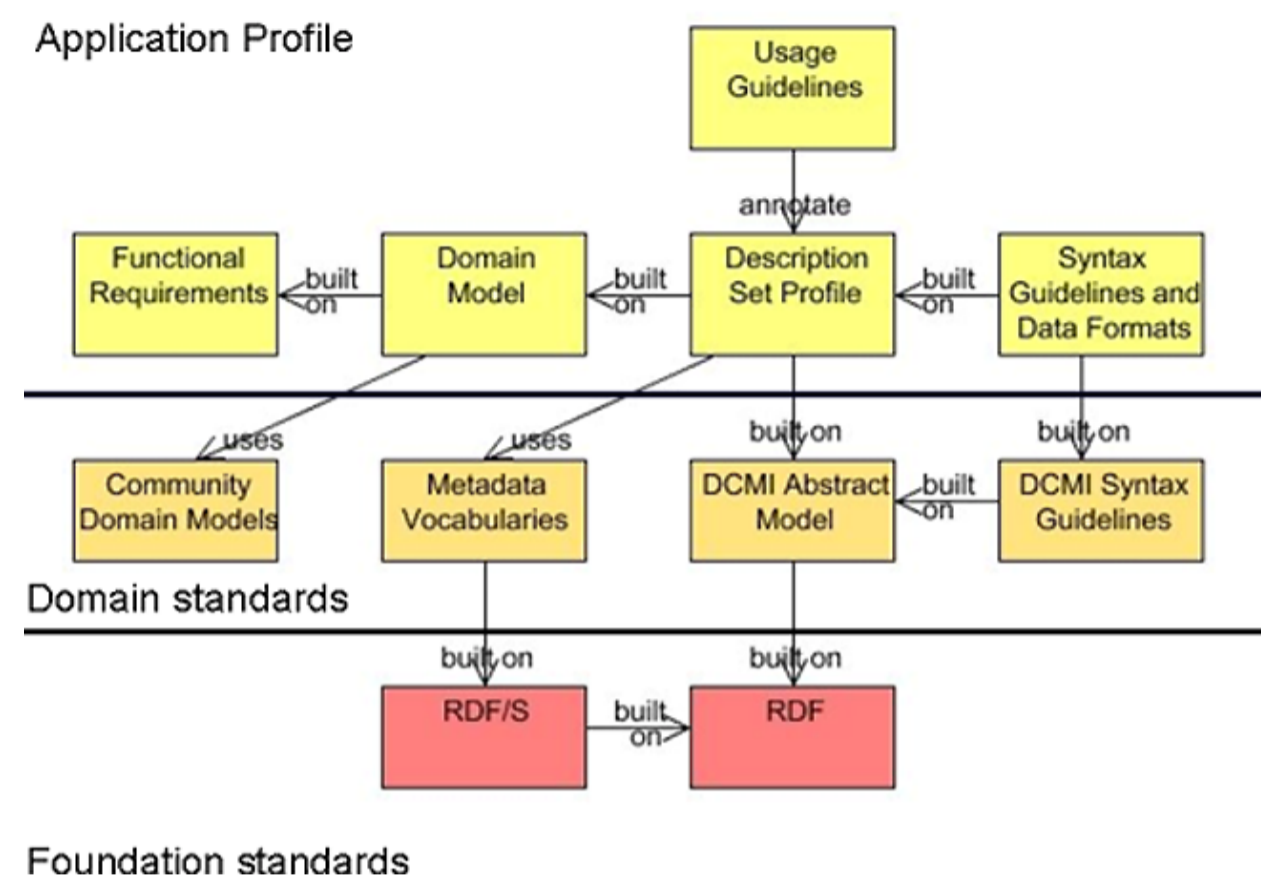
- **Community Domain Model:** A domain model which can be used in a community, such as the Functional Requirements for Bibliographic Records (FRBR) model for the library community.

- **Metadata Vocabularies:** One or more standard vocabularies that are used in the AP.

- **DCMI Abstract Model:** Specifies the components and constructs used in Dublin Core metadata, and defines the nature of those components and how they are combined to create information structures.

- **Usage Guidelines:** Provide instructions to create values for metadata terms.

- **DCMI Syntax Guidelines:** Provide instructions on implementation syntax for developers to turn APs into software applications. DCMI provides various encoding guidelines [DCMI-ENCODINGS].

- **Resource Description Framework (RDF):** Provides a way to express vocabulary terms and their semantics (http://www.w3.org/RDF/); a foundation which provides a robust flexible architecture for processing metadata RDF/S indicates an RDF Schema.

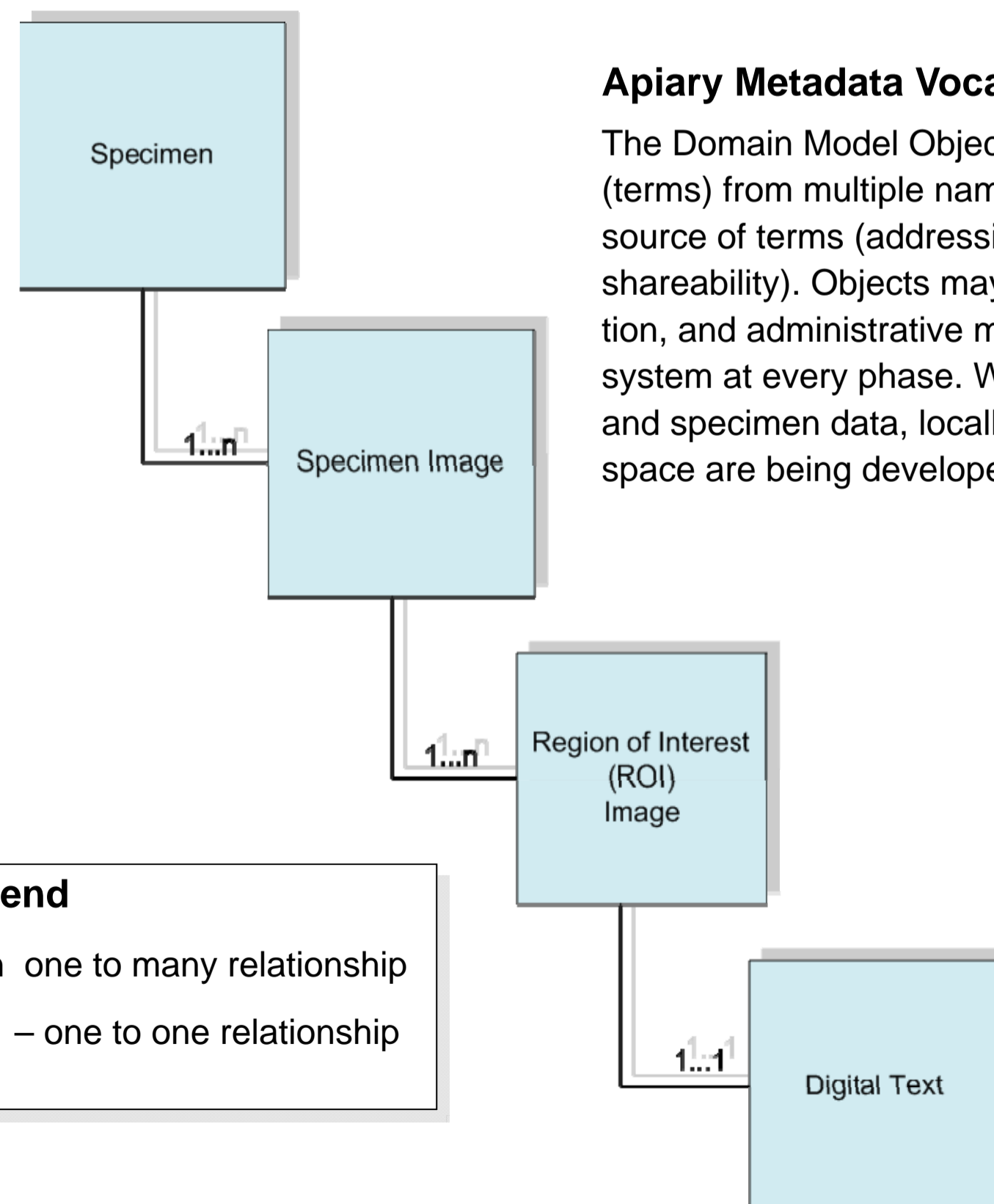---

Apiary Functional Requirements

The Apiary Project's functional requirements include high-level systems requirements and goals (e.g., optimizing the workflow, system integration, and reusability of code) as well as more detailed requirements, especially in terms of metadata needed for various objects that move through the system. We used process-centered use case modeling to identify key objects, processes, and tasks. Specific metadata requirements identified related to: types of metadata; standard vocabularies (e.g., DwC); interoperability/shareability; consistency and comprehensiveness; granularity; reusability; and specific constraints on metadata terms.

Use Case Modeling Example
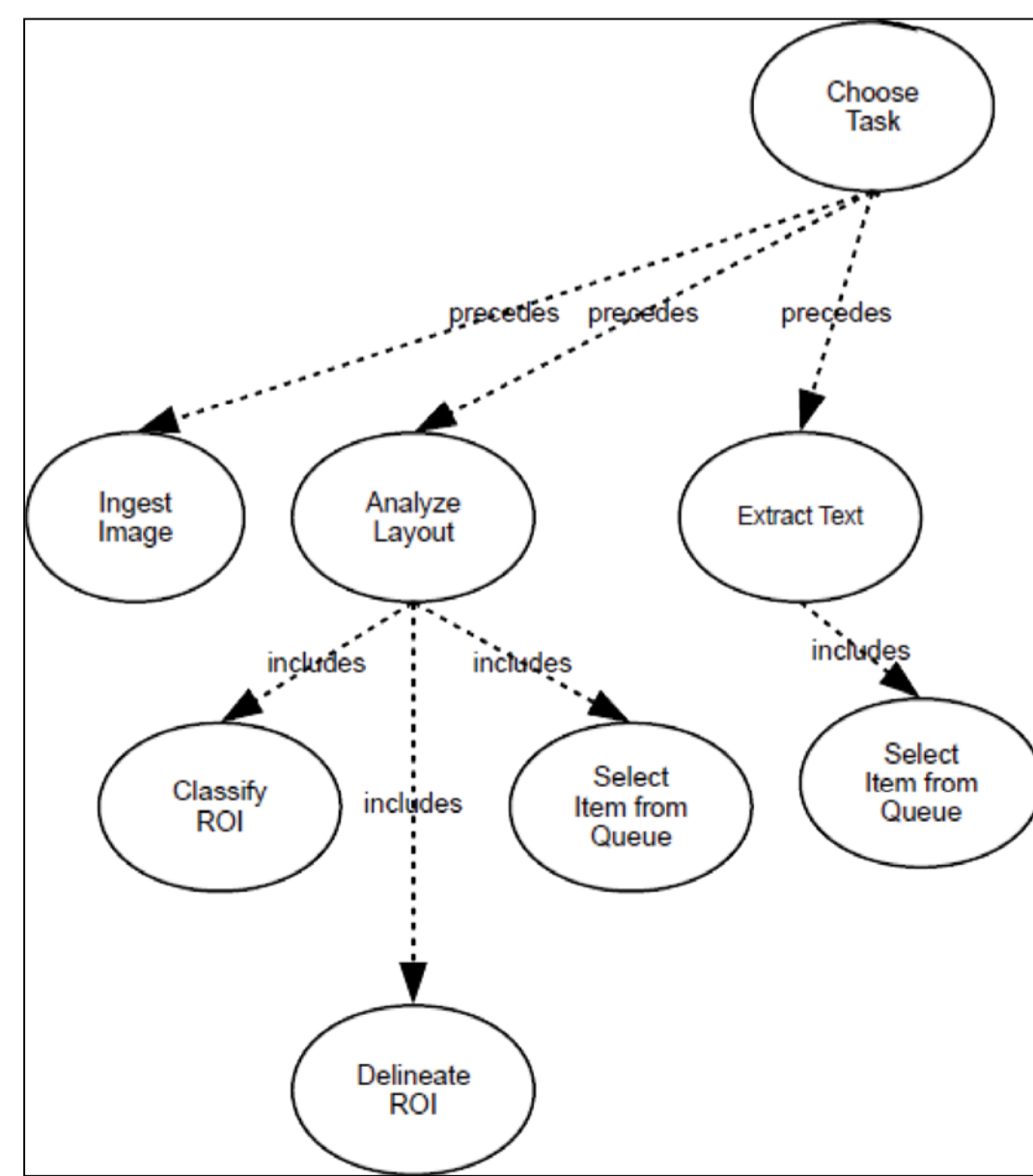
Apiary Domain Model

The Apiary Domain Model defines four objects within the workflow that require metadata and shows the relationships/derivations of the separate objects. For example, technical and administrative metadata may be needed to manage the objects in the workflow. The Specimen Object will have DwC metadata that is a result of the entire Apriary workflow. The Specimen Image Object is a scan of the herbarium specimen sheet and the source from which the Region of Interest (ROI) Objects are derived; ROIs may include the primary label, first annotation, and other textual or graphical information on the herbarium sheet. The Digital Text Object results from OCR on the ROI or manual transcription of data from an ROI. Relationships between these objects can be one to one (1…1) or one to many (1…n) as indicated.

Apiary Metadata Vocabulary Terms

The Domain Model Objects have associated metadata properties (terms) from multiple namespaces. DwC serves as the primary source of terms (addressing the requirement for interoperability/shareability). Objects may also have associated technical, preservation, and administrative metadata to serve the needs of the Apiary system at every phase. Where essential to the Apiary application and specimen data, locally-defined elements in an Apiary namespace are being developed.

Legend

1…n one to many relationship

1…1 – one to one relationship

Description Set Profile -

A DSP defines an information model and can be expressed in XML. The Apiary DSP formalizes the Apiary metadata following guidelines for DSPs (http://dublincore.org/documents/dc-dsp/).

```
<DescriptionSetTemplate>
    <DescriptionTemplate ID="ApiarySpecimenCoreMetadata" maxOccur="1" minOccur="1" standalone="no">
        <ResourceClass>http://rs.brit.org/ap/objects/SpecimenMetadata</ResourceClass>

        <StatementTemplate type="literal">
        <Property>http://rs.brit.org/ap/terms/barcode</Property>
            <LiteralConstraint>
                <SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>
                <LanguageOccurrence>optional</LanguageOccurrence>
            </LiteralConstraint>
        </StatementTemplate>

        <StatementTemplate type="nonliteral">
        <Property>http://rs.tdwg.org/dwc/terms/#scientificName</Property>
            <NonliteralConstraint>
                <VocabularyEncodingSchemeOccurrence>optional</VocabularyEncodingSchemeOccurrence>
                <VocabularyEncodingSchemeURI>http://www.ipni.org/</VocabularyEncodingSchemeURI>
                <ValueStringConstraint maxOccur="1">
                <SyntaxEncodingSchemeOccurrence>disallowed</SyntaxEncodingSchemeOccurrence>
                <LanguageOccurrence>optional</LanguageOccurrence>
                </ValueStringConstraint>
            </NonliteralConstraint>
        </StatementTemplate>
    </DescriptionTemplate>
</DescriptionSetTemplate>
```

The Utility of Application Profiles Rendered in DCAP

"An application profile describes the set of guidelines, description rules, and constraints used in creating a specific *set of metadata records*…application profiles are about providing high-level syntactic or structural interoperability in addition to the semantic interoperability" (http://dublincore.org/documents/singapore-framework) By developing an Apiary AP, and especially the DSP, it can support one or more of the following purposes:

- as a formal representation of the constraints of a Dublin Core Application Profile
- as configuration for databases
- as configuration for metadata editing tools.

DCMI has also issued a document, Interoperability Levels for Dublin Core Metadata (http://dublincore.org/documents/interoperability-levels/), that assists developers in indicating conformance to specifications resulting in likely interoperability. The Apiary AP efforts target Level 4 Description Set Profile Interoperability. This should enable the production of shareable and interoperable metadata records, as well as a basis for the validation of such records. Implementations of DCAPs in the biodiversity community should offer new and interesting opportunities for data integration.

References

DCMI Abstract Model http://dublincore.org/documents/abstract-model/

Singapore Framework for Dublin Core Application Profile http://dublincore.org/documents/singapore-framework/

Interoperability Levels for Dublin Core Metadata http://dublincore.org/documents/interoperability-levels/

Description Set Profiles: A constraint language for Dublin Core Application Profiles http://dublincore.org/documents/2008/03/31/dc-dsp/

Guidelines for Dublin Core Application Profiles http://dublincore.org/documents/2009/05/18/profile-guidelines/