Background

**discussion tomorrow: how to filter suitable tissues in the portal??**

The following is based on the requirements set forth in the document “ABCDDNA\_Required fields for DwC-A-DNA.docx” dated 2012-10-06. It is an attempt to show how Darwin Core extensions could be used to publish DNA Tissue Sample and Amplification data in a Darwin Core Archive.

There are at least two new classes implied by the above-mentioned document, which will be set forth here as **DNASample** and **DNAAmplification**. The relationship between DNASample and DNAAmplification is necessarily one-to-many. It is also possible that the relationship between dwc:Occurrence and DNASample would be one-to-many. If so, a single extension combining DNASample and DNAAmplification properties could be used, with DNASample properties repeated in multiple rows of DNAAmplifications as necessary, as there is only one level of relationship in the star schema of a Darwin Core Archive.

From the Basis of Record, “sample”, we need three classes (Question 1: what is the difference between class and extensions): the first class is Sample extension, this includes sample type (e.g., tissue sample, DNA sample, RNA sample, environmental sample) and kind of unit (e.g., Blood, muscle, gDNA, nDNA, cpDNA, mtDNA, mRNA, rRNA, tRNA, soil, air, water), The second class is DNAAmplification (e.g., GenBank accession number, BOLD process id, amplification method, primers, barcode sequence, sequence length, amplicon). The third class is Associations (e.g., associated catalog number, institution code, collection code, access point, type of relation). Question 2: Do we need an extension for association, or is it already available -need three different elements not one (related catalog item)? Question 3: How do we deal with this third extension in DarwinCore?

The following proposal assumes that a single DNASample (and Tissue sample, Environmental sample) can be associated with a dwc:Occurrence and that multiple DNAAmplifications can be associated with each dwc:Ocurrence/DNASample. The proposal uses one extension for the DNASample and another for the DNAAmplification. This is meant to demonstrate the ways in which extension data can be mapped in a Darwin Core Archive.

Extensions for Darwin Core Archives

For the purposes of this exercise, let's suppose that the new terms used in the extensions are managed under the namespace

<http://rs.tdwg.org/dwc/extension/DNA/terms>

and the full URI of the two new classes of interest are:

http://rs.tdwg.org/dwc/extension/DNA/terms/DNASample

http://rs.tdwg.org/dwc/extension/DNA/terms/DNAAmplification

The proposed arrangement of properties from the requirements documents is as below, regardless of whether three extensions are used for them (one for Sample, one for DNAAmplification, one for Associations) or one extension is used (Sample extension).

**Class dwc:Occurrence properties**:

**dwc:catalogNumber** (Unit/UnitID) - will be in the associated Occurrence record. (add occuranceid to non-mandatory list)

**dwc:collectionCode** (Unit/SourceID) - will be in the associated Occurrence record.

**dwc:institutionCode** (Unit/SourceInstitutionID) - will be in the associated Occurrence record.

**dwc:occurrenceRemarks** (Unit/Notes) - dwc already has this term, no need to add it. If this is a distinct term, then recommend something like dNARemarks as another property in the DNASample class (We agree, use it)

**Class Associations (add from core ABCD)**

**AssociatedCatalogNumber** (AssociatedUnitID) - distinct from the dwc:catalogNumber for the specimen or (tissue) sample to which the DNA or the tissue is associated, which will be in the associated Occurrence record.

**AssociatedCollectionCode** (AssociatedUnitSourceID) - distinct from the dwc:collectionCode for the specimen to which it is associated, which will be in the associated Occurrence record.

**AssociatedInstitutionCode** (AssociatedUnitSourceInstitutionID) - distinct from the dwc:institutionCode for the specimen to which it is associated, which will be in the associated Occurrence record.

**associationType** - description of relation between the objects/records

**dcterms:references** (Associations/Unit Association/Comment) - the Dublin Core record level term is meant to contain the reference to the primary detailed record about the Occurrence (specimen). (John please clarify; we need element for “access point (wrapper url)”)

**Class Sample properties**:

**blockedUntilDate** (BlockedUntil) - renamed for consistency with DwC term naming pattern.

**blockedRemarks** (Blocked) - made an assumption here about content and renamed to be a little more explicit.

**concentration** - prefer a single term for concentration in one standard unit rather than separating terms.

**concentrationUnit** - combine unit with concentration if possible (e.g., concentration in µg/ml). should be separate elements

**extractionDate (DNA samples only)** -

**extractionMethod (DNA samples only)**-

**extractionStaff (DNA samples only)**-

**preservation** - use as concatenated field

**providedBy** - (should be moved to core in the future)

**quality** - used for both DNA and tissue

**qualityCheckDate** - used for both DNA and tissue

**ratioOfAbsorbance260\_230 (DNA only)**-

**ratioOfAbsorbance260\_280 (DNA only)**-

**sampleType** (Tissue) - changed name to be more explicit.

**storageCondition**- use as concatenated field

**collectingPermit (will be moved to core in future)**-

**collectingPermitURL (will be moved to core in future)**-

**accessBenefitSharing** (there were two terms in the document having this same name, one boolean and one text) **(will be moved to core in future)**

**fieldPreservationMethod** (PreservationField) - changed the name to be more explicit (Possible fourth extension=preservation history? Options are multiple terms or piping).

**QualityCheckedMethod** - used for both DNA and tissue

**sourceGone** -

**molecularWeight** (MolWeight) - changed the name to be explicit.

(@Unit) - do not understand the role of this term **(DNA only).**

**Class Amplification properties**:

**dNAAmplificationID** - proposed identifier for the DNAAmplification class.

**geneticLocus** -

**geneticAccessionNumber** (GenBankNumber) - recommend more generic name.

**geneticAccessionURI** (GenBankNumber-URI) - recommend more generic name.

**SequenceLength**

**barcodeSequence** -

Class Reference check back with John W.

**referenceCitation** - not sure why this isn't part of DNA/Amplification in "ABCDDNA\_Required fields for DwC-A-DNA.docx".

**referenceURL** - not sure why this isn't part of DNA/Amplification in "ABCDDNA\_Required fields for DwC-A-DNA.docx".

**John: Do we have a multimedia extension, if so we could use the gelimpagepath there, otherwise we need another extension**

**gelImagePath** - the requirements document says that this is one-to-many and an attribute of the DNA, not of the amplification. If so, it could only be captured in concatenated list of images in a single field, or it would have to be in its own extension and linked to the coreID in the Occurrence, which would force Occurrence and DNA Sample to be in a one-to-one relationship.

Darwin Core Archive Example

Following is an example Darwin Core Archive meta.xml file connecting DNA Amplification data from a file "DNAAmplifications.csv" with a combined specimen and DNA Sample file "specimens.txt". Another way to do this would be to have all DNASample and DNAAmplification information in one file with DNASample properties repeated for every DNA Amplification record. That method would require just one extension, while the one below requires two extensions

<?xml version="1.0" encoding="UTF-8"?>

<archive xmlns="http://rs.tdwg.org/dwc/text/"

xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

xmlns:xs="http://www.w3.org/2001/XMLSchema"

xsi:schemaLocation="http://rs.tdwg.org/dwc/text/ http://rs.tdwg.org/dwc/text/tdwg\_dwc\_text.xsd">

<core encoding="UTF-8" fieldsTerminatedBy="\t" linesTerminatedBy="\n" ignoreHeaderLines="1"

rowType="http://rs.tdwg.org/dwc/terms/Occurrence">

<files>

<location>specimens.txt</location>

</files>

<id index="0" />

<field index="0" term="http://rs.tdwg.org/dwc/terms/occurrenceID"/>

<field index="1" term="http://purl.org/dc/terms/modified"/>

<field index="2" term="http://purl.org/dc/terms/references"/>

<field index="3" term="http://rs.tdwg.org/dwc/terms/institutionCode"/>

<field index="4" term="http://rs.tdwg.org/dwc/terms/collectionCode"/>

<field index="5" term="http://rs.tdwg.org/dwc/terms/catalogNumber"/>

<field index="6" term="http://rs.tdwg.org/dwc/terms/scientificName"/>

<field index="7" term="http://rs.tdwg.org/dwc/terms/occurrenceRemarks"/>

</core>

<extension encoding="UTF-8" fieldsTerminatedBy="\t" linesTerminatedBy="\n" ignoreHeaderLines="1"

rowType="http://rs.tdwg.org/dwc/extension/DNA/terms/DNASample">

<files>

<location>specimens.txt</location>

</files>

<coreId index="0" />

<field index="8" term="http://rs.tdwg.org/dwc/extension/DNA/terms/dNASampleID"/>

<field index="9" term="http://rs.tdwg.org/dwc/extension/DNA/terms/dNAInstitutionCode"/>

<field index="10" term="http://rs.tdwg.org/dwc/extension/DNA/terms/dNACollectionCode"/>

<field index="11" term="http://rs.tdwg.org/dwc/extension/DNA/terms/dNACatalogNumber"/>

<field index="12” term="http://rs.tdwg.org/dwc/extension/DNA/terms/associationType”/>

<field index="13" term="http://rs.tdwg.org/dwc/extension/DNA/terms/blockedUntilDate"/>

<field index="14" term="http://rs.tdwg.org/dwc/extension/DNA/terms/blockedRemarks"/>

<field index="15" term="http://rs.tdwg.org/dwc/extension/DNA/terms/concentration"/>

<field index="16" term="http://rs.tdwg.org/dwc/extension/DNA/terms/concentrationUnit"/>

<field index="17" term="http://rs.tdwg.org/dwc/extension/DNA/terms/extractionDate"/>

<field index="18" term="http://rs.tdwg.org/dwc/extension/DNA/terms/extractionMethod"/>

<field index="19" term="http://rs.tdwg.org/dwc/extension/DNA/terms/extractionStaff"/>

<field index="20" term="http://rs.tdwg.org/dwc/extension/DNA/terms/preservation"/>

<field index="21" term="http://rs.tdwg.org/dwc/extension/DNA/terms/providedBy"/>

<field index="22" term="http://rs.tdwg.org/dwc/extension/DNA/terms/quality"/>

<field index="23" term="http://rs.tdwg.org/dwc/extension/DNA/terms/qualityCheckDate"/>

<field index="24" term="http://rs.tdwg.org/dwc/extension/DNA/terms/ratioOfAbsorbance260\_230"/>

<field index="25" term="http://rs.tdwg.org/dwc/extension/DNA/terms/ratioOfAbsorbance260\_280"/>

<field index="26" term="http://rs.tdwg.org/dwc/extension/DNA/terms/tissueType"/>

<field index="27" term="http://rs.tdwg.org/dwc/extension/DNA/terms/preservationDNA"/>

<field index="28" term="http://rs.tdwg.org/dwc/extension/DNA/terms/storageConditionDNA"/>

<field index="29" term="http://rs.tdwg.org/dwc/extension/DNA/terms/storageConditionTissue"/>

<field index="30" term="http://rs.tdwg.org/dwc/extension/DNA/terms/collectingPermit"/>

<field index="31" term="http://rs.tdwg.org/dwc/extension/DNA/terms/collectingPermitURL"/>

<field index="32" term="http://rs.tdwg.org/dwc/extension/DNA/terms/accessBenefitSharing"/>

<field index="33" term="http://rs.tdwg.org/dwc/extension/DNA/terms/fieldPreservationMethod"/>

<field index="34" term="http://rs.tdwg.org/dwc/extension/DNA/terms/tissueQuality"/>

<field index="35" term="http://rs.tdwg.org/dwc/extension/DNA/terms/barcodeSequence"/>

<field index="36" term="http://rs.tdwg.org/dwc/extension/DNA/terms/dNAQualityCheckedMethod"/>

<field index="37" term="http://rs.tdwg.org/dwc/extension/DNA/terms/sourceGone"/>

<field index="38" term="http://rs.tdwg.org/dwc/extension/DNA/terms/molecularWeight"/>

</extension>

<extension encoding="UTF-8" fieldsTerminatedBy="," linesTerminatedBy="\n" fieldsEnclosedBy='"' ignoreHeaderLines="1"

rowType="http://rs.tdwg.org/dwc/extension/DNA/terms/DNAAmplification">

<files>

<location>DNAAmplifications.csv</location>

</files>

<coreId index="0" />

<field index="0" term="http://rs.tdwg.org/dwc/terms/occurrenceID"/>

<field index="1" term="http://rs.tdwg.org/dwc/extension/DNA/terms/dNAAmplificationID"/>

<field index="2" term="http://rs.tdwg.org/dwc/extension/DNA/terms/geneticLocus"/>

<field index="3" term="http://rs.tdwg.org/dwc/extension/DNA/terms/geneticAccessionNumber"/>

<field index="4" term="http://rs.tdwg.org/dwc/extension/DNA/terms/geneticAccessionURI"/>

<field index="5" term="http://rs.tdwg.org/dwc/extension/DNA/terms/referenceCitation"/>

<field index="6" term="http://rs.tdwg.org/dwc/extension/DNA/terms/referenceURL"/>

<field index="7" term="http://rs.tdwg.org/dwc/extension/DNA/terms/gelImagePath"/>

</extension>

</archive>