**GLOBAL GENOME BIODIVERSITY NETWORK**

**Data Standards and Data Access for Genomic Samples Task Force**

**DARWIN CORE DNA AND TISSUE EXTENSION WORKING MEETING REPORT**

**TDWG 2013 Conference**

**Florence, Italy**

**Draft report date February 2014**

**Task group organizers and writing team: Katharine Barker, Smithsonian Institution; Gabriele Droege, Botanic Garden and Botanical Museum Berlin-Dahlem; Paul Flemons, Australian Museum.**

**GGBN TDWG Task Group Members:**

Katharine Barker, National Museum of Natural History, Smithsonian Institution

John Deck, University of California Berkeley

Gabriele Droege, Botanic Garden and Botanical Museum Berlin-Dahlem

Paul Flemons, Australian Museum

Éamonn Ó Tuama, Global Biodiversity Information Facility

John Wieczorek, University of California Berkeley

**Other Contributors:**

Markus Döring, Global Biodiversity Information Facility

Tim Robertson, Global Biodiversity Information Facility

Jamie Whitacre, National Museum of Natural History, Smithsonian Institution

**TABLE OF CONTENTS**

1. **INTRODUCTION AND OVERVIEW 3**
2. **DISCUSSION AND OUTCOMES 5**
3. **NEXT STEPS 10**
4. **INTRODUCTION AND OVERVIEW**

The Global Genome Biodiversity Network (GGBN) is a global consortium of organizations supporting the collection, maintenance and sharing of research-quality genomic specimens and samples. GGBN seeks to bring together the world’s leading collections of genome-quality specimens and samples representing Earth’s biodiversity to help accelerate and optimize research agendas everywhere.

In order to meet its goals, GGBN members need to publish information on their tissue and DNA holdings through a shared data portal. The upcoming GGBN data portal will be based on the DNA Bank Network’s data portal, and will work for both BioCASe/ABCD and IPT/DarwinCore Archive (DwC-A), and furthermore DiGIR providers (only voucher specimens in non-GGBN collections, not for DNA samples). DarwinCore serves as the data provision schema for many GGBN member institutions. However, DarwinCore lacks the DNA/Tissue Extension needed by GGBN members to publish DNA and Tissue data. In order to meet this need, the GGBN TDWG Task Group was formed in September 2012 to develop a DarwinCore DNA and Tissue Extension. Prior to the TDWG 2012 Conference, Task Group members first identified gaps in the existing standards based on GGBN use cases. Gaps in the existing ABCDDNA standard were filled and a preliminary list of required elements for the DwC-A extension was developed. Use cases and required elements were reviewed during the course of the TDWG 2012 Conference and a draft DwC-A DNA/Tissue Extension was developed (see figure 1).

Since the TDWG 2012 Conference the following progress has been made:

1. November 2012: Gaps in the ABCDDNA data standard, based on the required fields for the DarwinCore-A DNA/Tissue extension, were filled. Definitions were updated in DarwinCore.

2. December 2012: The term “Sample” was agreed on for the basis of record.

3. February 2013: The draft extension was implemented in GBIF’s Integrated Publishing Toolkit (IPT).

3. August 2013: IPT was tested and ways to include the draft DarwinCore DNA and Tissue Extension were identified.

4. October 2013: Both the draft DarwinCore DNA and Tissue Extension and revised ABCDDNA data standard (now referred to as GGBN data standard) were implemented and gaps were identified at the TDWG 2013 Conference.

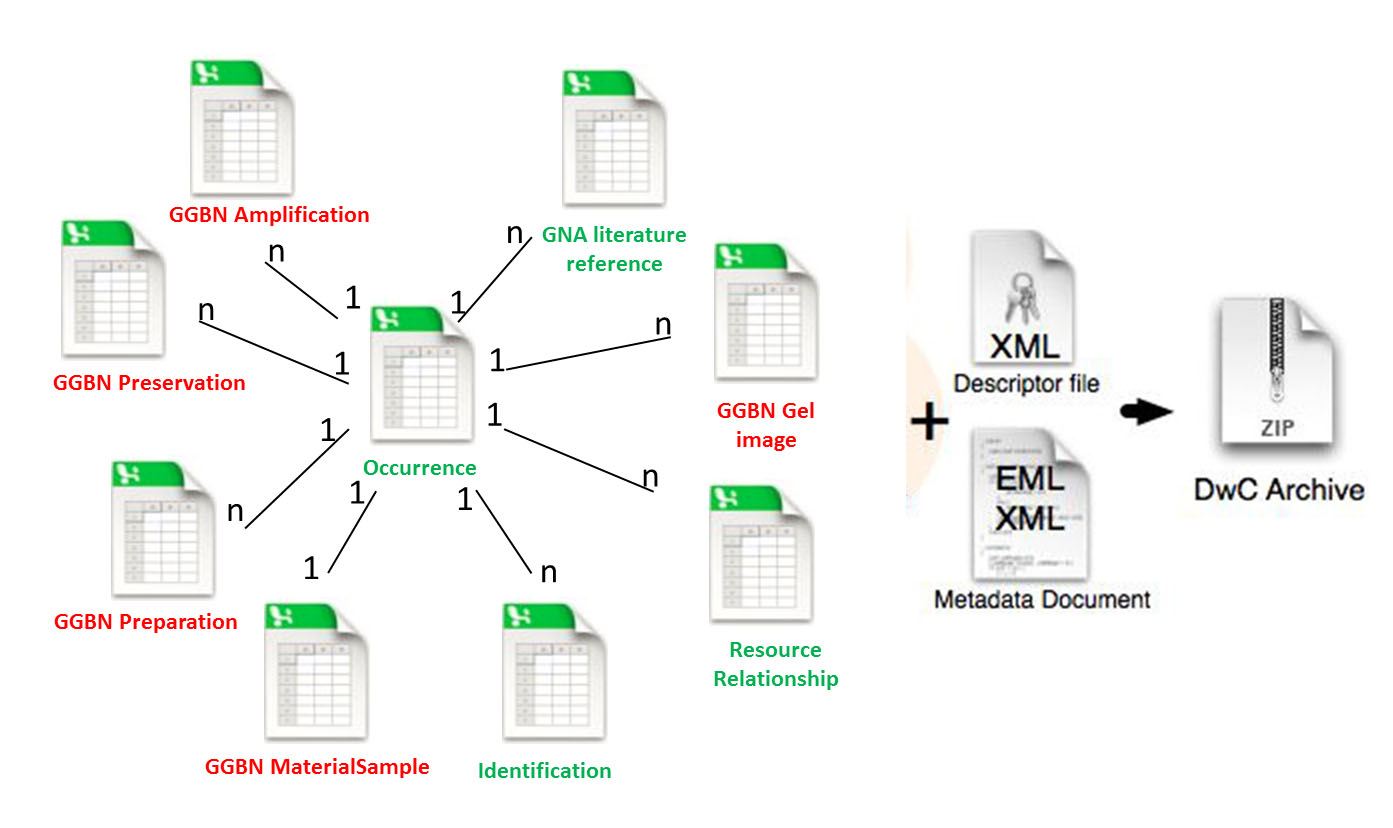


Figure 1. Red = new/renamed extension; Green = existing extension (Result from TDWG 2013 - Florence)

1. **DISCUSSION AND OUTCOMES**
   1. We agreed on using GGBN name spaces and terminology to enable correct citation and policy. After TDWG GGBN agreed to call the set of terms related to tissue and DNA samples “GGBN data standard”. ABCDDNA, DwC, and DwC-A are just implementations of the “GGBN data standard”.

GBIF should therefore use the following namespace: http://data.ggbn.org/schemas/ggbn/terms/

“see also” should point to: http://data.ggbn.org/schemas/ggbn/terms/

* 1. Four new extensions/classes were proposed based on gaps identified within the ABCDDNA and DwC-A DNA and Tissue Extension. These include Gel Image, Preparation, Preservation, and Permit (see below).

**1. GGBN Gel Image**

Relation to Core 1:1

|  |  |
| --- | --- |
| occurrence ID |  |
| identifier | Link to a DNA sample image (e.g. agarose gel image) |
| description | Remarks/Notes related to Gel |
| GelVoltage | Voltage of Gel (in V) |
| GelConcentration | Concentration of Gel |
| GelDuration | Duration of Gel |
| GelLadder | Ladder (Name of Ladder) |

**2. GGBN Preparation**

Relation to Core 1:n

|  |  |
| --- | --- |
| occurrenceID |  |
| preparationID |  |
| PreparationType | Description of preparation type, e.g. a DNA extraction method or a special tissue preparation |
| PreparationProcess | Process used in preparing the specimen or sample, can also be used to describe Phage/Plasmid propagation |
| PreparationMaterials | Materials and chemicals used in the preparation of the specimen or sample |
| PreparedBy | Person and/or institution responsible for or effecting the preparation |
| PreparationDate | The date of preparation |

**3. GGBN Preservation**

Relation to Core 1:n

|  |  |
| --- | --- |
| occurrenceID |  |
| preservationID |  |
| PreservationType | Description of preservation type, e.g. dried, silica gel, frozen |
| PreservationTemperature | Preferably a number - attribute gives temperature scale used |
| PreservationDateBegin | The date in text form when preservation begins |
| sequence | can be used to describe sequence of different preservations |

**4. GGBN Permit**

Relation to Core 1:n

|  |  |
| --- | --- |
| occurrenceID |  |
| PermitType | Description of permit type, e.g. Collecting permit, import permit, export permit |
| PermitStatus | Permit available, not required, not available (controlled vocabulary!) |
| PermitURI | A reference to the permit related to the sample or specimen |
| PermitText | The text of a permit related to the sample or specimen |

* 1. Changes to the existing extension were proposed based on gaps identified in the ABCDDNA and DwC-A DNA and Tissue Extension (see below).

**1. Sample, should be renamed into GGBN MaterialSample**

Relation to Core 1:1 or use new MaterialSampleCore instead

|  |  |  |
| --- | --- | --- |
| occurrenceID |  | ok |
| sampleID |  | ok, rename into materialSampleID |
| sampleType | e.g. “DNA”, “tissue” | ok, rename into materialSampleType |
| *sourceType* |  | *can be deleted (following the discussion from Florence) all sort of sample/source types can be handels with materialSampleType* |
| *sampleInstitutionID* | *for a sample record the InstitutionID of occurrence core will be used* | *can be deleted* |
| *sampleCollectionID* | *for a sample record the CollectionID of occurrence core will be used* | *can be deleted* |
| *sampleInstitutionCode* | *for a sample record the InstitutionCode of occurrence core will be used* | *can be deleted* |
| *sampleCollectionCode* | *for a sample record the CollectionCode of occurrence core will be used* | *can be deleted* |
| *sampleCatalogueNumber* | *for a sample record the CatalogueNumber of occurrence core will be used* | *can be deleted* |
| blockedUntilDate | documentation ok | ok, rename into BlockedUntil |
| Blocked | True if sample is blocked in general for thirds parties. | new |
| sourceGone | documentation ok | can be deleted |
| *blockedRemarks* | *use disposition instead with recommended value* | *rename to LoanConditions* |
| concentration | documentation ok | ok, rename into Concentration |
| concentrationUnit | documentation ok | ok, rename into ConcentrationUnit |
| MethodDeterminationConcentrationAndRatios | Description of method used for concentration measurement | new |
| Volume | DNA volume value | new |
| VolumeUnit | Unit attribute, e.g. µl | new |
| Weight | DNA weight value | new |
| WeightUnit | Unit attribute, e.g. ng | new |
| MethodDeterminationWeight | Description of method used for weight measurement | new |
| *exctractionDate* | *preparation extension should be used instead* | *can be deleted* |
| *extractionMethod* | *preparation extension should be used instead* | *can be deleted* |
| *extractedBy* | *preparation extension should be used instead* | *can be deleted* |
| purificationMethod | documentation ok | ok, rename into PurificationMethod |
| providedBy | in ABCD Unit/SpecimenUnit/Acquisition/AcquiredFrom can be used  documentation ok | ok, rename into ProvidedBy |
| *quality* |  | *can be deleted* |
| qualityCheckDate | documentation ok | ok, rename into QualityCheckDate |
| qualityCheckMethod | documentation ok | ok, rename into QualityCheckMethod |
| ratioOfAbsorbance260\_230 | documentation ok | ok, rename into RatioOfAbsorbance260\_230 |
| ratioOfAbsorbance260\_280 | documentation ok | ok, rename into RatioOfAbsorbance260\_280 |
| *collectingPermit* | *Permit extension required* | *can be deleted* |
| *collectingPermitURL* | *Permit extension required* | *can be deleted* |
| *accessBenefitSharing* | *Permit extension required* | *can be deleted* |
| *accessBenefitSharingURL* | *Permit extension required* | *can be deleted* |
| *molecularWeight100kb* | *seems to be unnecessary* | *can be deleted* |

**2. GGBN DNA Amplification**

|  |  |  |
| --- | --- | --- |
| occurrenceID |  | new |
| amplificationID | documentation ok | ok, rename into AmplificationID |
| geneticLocus | documentation ok | rename into Marker |
| geneticAccessionNumber | documentation ok | rename into GeneticAccessionNumber |
| geneticAccessionURI | documentation ok | rename into GeneticAccessionURI |
| sequenceLength | documentation ok | rename into SequenceLength |
| barcodeSequence | documentation ok | rename into BarcodeSequence |

1. **NEXT STEPS**

1. January 2014: Gaps identified at the TDWG 2013 Conference will be filled.

2. March 2014: The updated ABCDDNA and DwC extension will be implemented in GBIF’s Integrated Publishing Toolkit (IPT).

3. February 2014: GGBN Data Standards and Policies and Practices Task Force Members will review the current set of GGBN data standard fields form

4. March 2014-September 2014: SI NMNH use cases for DNA and Tissue Data will be documented and the GGBN Data Portal functionality will be expanded to meet the needs of these use cases.

5. April 2014-October 2014: The DwC DNA and Tissue extension will be tested with SI NMNH tissue and DNA data using IPT.

6. June 2014: Mockup/draft version of new GGBN Data Portal for GGBN 2014 conference.