GLOBAL BIODIVERSITY INFORMATION FACILITY

Mobilising Multimedia Resources in Biodiversity

2nd report of the

GBIF Multimedia Resources Task Group

March 2009

www.gbif.org

Suggested citation:

Morris, Robert; Annette Olson, Chris Freeland, Gregor Hagedorn, Greg Riccardi, Mihail-Constantin Carausu, Eamonn O' Tuama, and Vishwas Chavan (2009). Mobilising multimedia resources in biodiversity: 2nd report of the GBIF Multimedia Resources Task Group (MRTG), March 2009 Pp. 22.

L to R: Chris Freeland, Greg Riccardi, Eamonn O' Tuama, Annette Olson, Mihail-Constantin Carausu, Vishwas Chavan, Gregor Hagedorn, and Robert Morris

Cover Credits: Ciprian Marius Vizitiu, 2009.

CONTENT

Introduction:

This document contains the report of the 2nd meeting of the Global Biodiversity Information Facility's Multimedia Resources Task Group (MRTG)[1] held at its Secretariat in Copenhagen, Denmark during 23-26 February 2009. Meeting was chaired by Prof. Robert A. Morris, University of Massachusetts at Boston, USA.

Participants:

1. Prof. Robert A. Morris, University of Massachusetts at Boston, USA
2. Dr. Annette Olson, National Biological Information Infrastructure, US Geological Survey, Reston, VA, USA
3. Dr. Greg Riccardi, Florida State University, Tallahassee, USA
4. Dr. Gregor Hagedorn, Institut fur Pflanzenvirologie, Berlin, Germany
5. Mr. Chris Freeland, Missouri Botanical Garden, St. Louis, USA
6. Mr. Mihail-Constantin Carausu, Danish Biodiversity Information Facility (DanBIF), Copenhagen, Denmark
7. Dr. Eamonn O' Tuama, Global Biodiversity Information Facility, Copenhagen, Denmark
8. Dr. Vishwas Chavan, Global Biodiversity Information Facility, Copenhagen, Denmark

Dr. Nicholas King, Dr. David Remsen, Mr. Juan Bello, Mr. Samy Gaiji, and Mr. Timothy Robertson of the Global Biodiversity Information Facility attended various sessions of the meeting.

Background:

In March 2008, GBIF established the 'Multimedia Resources Task Group (MRTG)' under the chairmanship of Robert A. Morris, Professor, University of Massachusetts at Boston, USA (http://www.gbif.org/News/NEWS1205258694) to provide recommendations on how to mobilise multimedia resources in biodiversity through the GBIF network. The first meeting of MRTG was held in Copenhagen, Denmark during 19-21 June 2008. One of the major recommendations of MRTG was to develop a multimedia metadata schema to support the discovery of multimedia resources in biodiversity, as well to facilitate exchange/sharing of species occurrence data associated with these resources.

A meeting of experts in management of biodiversity multimedia resources, and metadata schemas for such resources, was held during 12-13 September 2008 at the Lilly Laboratory, Marine Biological Laboratory, Woods Hole, MA, USA. The major objective of the meeting was to draft the metadata schema to support the discovery of, and access to, multimedia resources in biodiversity.

Recommendations of the MRTG were discussed by the GBIF Science Committee in its meeting held at Arusha, Tanzania on 3rd November 2008. The Science Committee opined the need for further work on the recommendations as well the Multimedia Metadata Schema to turn it into a more mature product within next 6 months.

[1] MRTG: GBIF Multimedia Resources Task Group (MRTG) – http://wiki.gbif.org/gbif/wikka.php?wakka=MultimediaResourcesTaskGroup

In December 2008, MRTG submitted a proposal to the Biodiversity Information Standards (TDWG) for a 'Joint GBIF-TDWG Multimedia Resources Task Group'[2] with an objective to prepare a Multimedia Metadata schema for the TDWG ratification process.

The present MRTG meeting (23-26 February 2009) in Copenhagen was called to further refine the Multimedia Metadata schema and debate recommendations made earlier. In addition to refinement of Multimedia Metadata schema, MRTG discussed issues related to technical implementation of Multimedia Metadata Schema and training needs for mobilising multimedia resources in biodiversity.

Agenda:

1. Refinement of Multimedia Metadata Schema
2. Technical Implementation of related issues
    a. Multimedia Metadata schema within GBIF Data Architecture
    b. Integration of Authority Services
    c. Provenance
3. Training Needs for mobilising multimedia resources in biodiversity

Outcomes:

1. **Multimedia Metadata Schema:** The draft schema prepared at the Woods Hole meeting (September 2008) was re-visited. The current version 0.6 is an outcome of over 40 hours of intense debate on several aspects including cross-walks with existing schemas such as Darwin Core (DwC), Dublin Core, IPTC, Natural Collections Descriptions (NCD), etc. (http://www.keytonature.eu/wiki/MRTG_Schema_v0.6). The schema ver. 0.6 falls into five groups, viz., Management, Attribution, Content Coverage (Spatial and Temporal), Content Subject, Technical Metadata and Related Resources. A non-normative document was prepared during the meeting (Annexure 1). The schema will be submitted to TDWG for seeking initial feedback by April 2009 and for ratification in May/June 2009.

   Annexure I contains Non normative Multimedia Metadata schema document, which correspond to draft ver. 0.6 of Multimedia Metadata schema as found on http://www.keytonature.eu/wiki/MRTG_Schema_v0.6. Discussion on the Multimedia Metadata schema continues on http://www.keytonature.eu/wiki/MRTG_Schema_v0.7 where some minor additions will be made before soliciting public comment.

2. **Updates on Recommendations:** The Task Group felt that it would be appropriate to debate agenda item 2, and 3 in light of its September 2008, Recommendations[3]. Thus for reporting purpose, relevant recommendations (September 2008) are reproduced together with updates (indicated below as 'CPH2 Update'[4]) as debated during the 23-26 February 2009 meeting of the Task Group.

---

[2] Joint GBIF-TDWG Multimedia Resources Task Group: http://www.tdwg.org/activities/img/multimedia/charter/

[3] Morris, Robert; Annette Olson, Eamonn O' Tuama, Greg Riccardi, Greg Whitbread, Gregor Hagedorn, Ivan Teage, Mikko Heikkinen, Patrick Leary, Vijay Barve, and Vishwas Chavan (2008). Recommendations of the GBIF Multimedia Resources Task Group. September 2008 Pp. 18.

[4] CPH2 Update: Recommendations made by the GBIF Multimedia Resources Task Group at its Copenhagen meeting during 23-26 February 2009.

A. Updates on recommendations about Social Issues:

Recommendation 1: GBIF multimedia mobilization efforts should recognize the range of breadth and depth of IT resources available to publishers[5] of biodiversity media. Training and tools for suitable for organizations with sophisticated organizations will not be the same as those for smaller providers using, for example, personal image management tools.
CPH2 Update: MRTG reconfirmed that training and tools should be developed by GBIF, as described under other recommendations, but very small numbers of metadata fields are mandatory, which will ease the burden on some providers. This feature of the architecture may make it easier for GBIF to develop tools without complex user interfaces.

Recommendation 2: GBIF should commission a Training Manual for Mobilisation of Multimedia Resources, and Training courses for mobilizing multimedia resources related to biodiversity.
CPH2 Update: Training material needs must be met for contributors, content providers, aggregators. Participants will produce outlines of their own corresponding material to be placed on wiki as a guide to community developers of such documentation. Also, use cases, as suggested on the wiki, can provide supportive material for training.

Recommendation 3: GBIF should require that metadata about media resources is provided either without any restriction on its use or reproduction, or under a suitable open-content license (such as Creative Commons[6]). At the same time, GBIF should make it clear that this does *not* apply to the media resource itself. For resources, the metadata recommendations below are designed to insure that the copyright holder's resource usage terms or licenses are clearly available to users of GBIF.
CPH2 Update: The Multimedia Metadata schema ver. 0.6 provide for copyright attribution and terms of use. Recommended ways for protecting and managing copyrighted media resources are documented at http://www.keytonature.eu/wiki/MRTG_Best_Practices.

Recommendation 5: A GBIF data and metadata sharing agreement should provide that *if* a publisher's metadata supports thumbnail or other preview access (e.g. by a URL), then GBIF is granted the right to cache and display such a thumbnail.
CPH2 Update: To enable this recommendation there are specific fields within metadata schema ver. 0.6 that allow references to various versions of a multimedia resource, e.g. thumbnails, or previews.

Recommendation 6: Develop a comparison table of metadata support tools, together with attributes that ease the provision of metadata when the metadata architecture is complete. Consider selecting some for support and training.
CPH2 Update: The MRTG wiki provides for community participation and commentary about and existing tools (http://www.keytonature.eu/wiki/Image_Metadata_Management_Tools).

Recommendation 7: Develop and implement a strategy to create a cultural change toward routine geo-referencing of multimedia resources for which a location is meaningful with emphasis on geo-coding as close as possible to the time of acquisition.

---

5 Publishers: Throughout this report term "Publishers" or "Data Publishers" has been used instead of "data providers" as used in previous GBIF reports and communications. GBIF facilitate discovery, and access to data.

6 Creative Commons: http://creativecommons.org/

CPH2 Update: Metadata schema ver. 0.6 supports geo-referenced metadata, and will facilitate such cultural changes. Additionally metadata support tools that help create bounding boxes and provide coordinates, check records with gazetteers or other authorities, and more will help promote the use of geo-referencing.

Recommendation 9: GBIF Participants should establish one or more national, regional, and thematic multimedia repositories with the same level of service as those for other data types.
CPH2 Update: GBIF Integrated Publishers Toolkit (IPT)[7] developers will assist the MRTG to ensure that the standard is consistent with the requirements of IPT, and/or to identify any work resource requirements that may be needed to reconcile them. The meeting was attended by the representative of the DanBIF[8] which will establish a multimedia repository for Nordic region and be a reviewer and critic of the draft standards.

B. Updates on Recommendations about Technical Issues:

Recommendation 10: Multimedia metadata should be supported in an indexing or caching service with machine and human interfaces, at both the collection level and the object level (image, audio, and video).
CPH2 Update: The proposed metadata standard ver. 6.0 recognizes both the similarities and differences between multimedia objects and collections of such objects, and supports descriptions of each.

Recommendation 11: The proposed GBIF Global Biodiversity Resource Discovery System (GBRDS) and its Integrated Publishing Toolkit (IPT) should support discovery and mobilisation of multimedia metadata and resources.
CPH2 Update: See **CPH2 Update no. 9**. In addition, MRTG reconfirms that smaller metadata providers especially need assistance in 1) establishing their own multimedia repositories; 2) ingesting media metadata from their contributors; 3) providing tools to their contributors to assist in the ingestion process, and 4) responding to resource service requests based on queries containing metadata. Use cases, as suggested on the wiki, can provide supportive material for the development of needed tools.

Recommendation12: The design of required or recommended metadata should promote the ability of users of GBIF services to determine fitness for use without requiring the users to acquire underlying resources. At a minimum licensing or other access control terms should be available through GBIF services.
CPH2 Update: Substantial attention is paid in the Multimedia Metadata Standard ver. 0.6 standard to the ability to signal biologically relevant content metadata, including by existing standards such as the proposed standardized version of the Darwin Core. The committee has identified the need for but not provided for annotation and other feedback mechanisms to enhance the ability to discovery of fitness for use.

Recommendation 13: A single metadata schema should be developed that is able to treat resource collections and objects uniformly.
CPH2 Update: Multimedia Metadata schema ver. 0.6 ensures that both resource collections and objects are described through single schema.

Recommendation 14: Controlled vocabularies for metadata values should be encouraged and supported technically, but plain text should be supported for these as well.

---

7 Integrated Publishing Toolkit (IPT): http://ipt.gbif.org/.
8 DanBIF: http://www.danbif.dk/

CPH2 UPDATE: Multimedia Metadata schema ver. 0.6 facilitates the implementation of the recommendation above.

Recommendation 15: Make a provision in the metadata schema to specify that the copyright owner or available licenses are unknown.
CPH2 UPDATE: Multimedia Metadata schema ver. 0.6 facilitates such an implementation.

Recommendation 16: Support the identification of resources with publisher-defined GUID[9] schemes in resource or collection level metadata.
CPH2 UPDATE: Multimedia Metadata schema ver. 0.6 does this. However need better documentation.

Recommendation 17: Metadata standards should support the ability to express relations among described objects, e.g. that Image I isMemberOf collection C.
CPH2UPDATE: Multimedia Metadata schema ver. 0.6 supports this. Several such relations are already expressed in ver. 0.6.

Recommendation 18: Provide services for geomancy (geo-referencing) and scientific name recognition.
CPH2 UPDATE: This is accomplished in Multimedia Metadata schema ver. 0.6.

Recommendation 19: Metadata schema should allow support for the "documents" relation, which asserts that a multimedia object provides evidence for an assertion that something else (e.g. an observation) is a GBIF primary biodiversity datum in the sense of species occurrence, ecosystem occurrence, behavioural occurrence, etc.
CPH2 Update: See section "Related Resources" of Multimedia Metadata schema ver. 0.6.

Recommendation 20: MRTG should propose a lightweight metadata schema by combining existing schemata of KeyToNature[10], the NBII Digital Image Library[11], and Morphbank[12].
CPH2 Update: This accomplished through Multimedia Metadata schema ver. 0.6.

Recommendation 21: Metadata should be able to specify media formats, including proprietary ones. The specification mechanism should be extensible. No particular format should be endorsed.
CPH2 UPDATE: This was accomplished through Multimedia Metadata schema ver. 0.6.

Recommendation 22: Develop metadata specifications that allow specification of media manipulation by the provider after acquisition.
CPH2 UPDATE: See "Resource Creation Technique" section of the Multimedia Metadata ver. 0.6.

Recommendation 23: Provide users of tagging systems (like Flickr[13], PicassaWeb[14], etc.) with facility for bulk assignment of metadata to media served by those systems.
CPH2 UPDATE: The Task Group recommends that GBIF should initiate development of tools that would facilitate such bulk assignment of metadata. MRTG believes that this is independent of the Multimedia Metadata schema.

---

9 GUID, Globally Unique Identifier:http://en.wikipedia.org/wiki/Globally_Unique_Identifier.
10 KeyToNature: http://www.keytonature.edu/.
11 NBII Digital Image Library: http://images.nbii.gov/.
12 Morphbank: http://www.morphbank.net/.
13 Flikr: http://www.flickr.com/.
14 PicasaWeb: http://picasaweb.google.com/.

Recommendation 24: Collaboratively with community and platform publishers, develop demonstration sites exhibiting practices which raise the utility of their media for science, environment protection, and education.
CPH2 UPDATE: GBIF should provide thought leadership in this area, as well implement exemplar applications, tools, etc.

C. Updates on Long Term Recommendations:

Recommendation 25: Develop demonstration sites for Flickr and similar other public folksonomy-based public multimedia repositories to demonstrate best practice for their tagging and machine API facilities.
CPH2 UPDATE: The Task Group recommends that GBIF initiate development of tools that will facilitate such bulk assignment of metadata. MRTG believes that this is independent of metadata schema.

Recommendation 26: Stimulate and encourage innovation around removing the human time intensive nature of metadata assignment, such as is provided by tools like BioGeomancer[15] and Herbis[16].
CPH2 UPDATE: The Task Group recommends that GBIF initiate development of tools that would facilitate such bulk assignment of metadata. MRTG believes that this is independent of metadata schema.

Recommendation 27: Organize collaboration with organizations, (e.g. NBII[17]) with experience in motivating disparate stakeholders towards developing strategies for mobilising them.
CPH2 UPDATE: The Task Group participants and others are invited to put their metadata ingestion documentation on wiki.

Recommendation 28: Develop mechanisms to allow providers to specify that appropriate metadata, particularly terms of use, be provided at the collection level, but served "by inheritance" as sub-collection or as record level metadata for objects in the collection.
CPH2Update: The Multimedia Metadata schema ver. 0.6 class mechanism need to be examined to check if the requirements for this recommendation are met.

---

15 BioGeomancer: http://www.biogeomancer.org/
16 Herbis: http://www.herbis.org/.
17 NBII, National Biological Information Infrastructure: http://www.nbii.gov/.

Annexure I

Non-normative 0.6 document for GBIF – TDWG Multimedia Resources Metadata schema ver. 0.6. (MRTG)

GBIF - TDWG Multimedia Resources Metadata Schema (MRTG)

A data standard for exchanging data describing biodiversity multimedia resources and collections

MRTG members who participated in drafting the proposed schema (in alphabetical order)

- Mr. Mihail-Constantin Carausu, Danish Biodiversity Information Facility (DanBIF), Copenhagen, Denmark
- Dr. Vishwas Chavan, Global Biodiversity Information Facility, Copenhagen, Denmark
- Mr. Chris Freeland, Missouri Botanical Garden, St. Louis, USA
- Dr. Gregor Hagedorn, , JKI, Federal Research Institute for Cultivated Plants, Berlin, Germany
- Prof. Robert A. Morris, University of Massachusetts at Boston, USA
- Dr. Dimitry Mozzherin, Encyclopedia of Life, Woods Hole, USA
- Dr. Annette Olson, National Biological Information Infrastructure, US Geological Survey, Reston, VA, USA
- Dr. Greg Riccardi, Florida State University, Tallahassee, USA
- Dr. Eamonn O' Tuama, Global Biodiversity Information Facility, Copenhagen, Denmark

Note: This is a non-normative document, which provides some background to the aims and uses of the proposed standard. The draft of the normative document has been separated and may be found on the MRTG Normative Discussion Website at http://www.keytonature.eu/wiki/MRTG.

Acronyms and named institutions and projects are listed in a Glossary in Appendix I.

The MRTG standard is the culmination of work on multimedia resource descriptions carried out by Key To Nature, the NBII Digital Image Library, MorphBank, and others, together with input from a number of other stakeholder communities including Encyclopedia of Life (EOL), the Biodiversity Heritage Library (BHL) and UMASS-Boston. The Global Biodiversity Information Facility (GBIF) commissioned the 'Multimedia Resources Task Group (MRTG)'

in March 2008 and it was approved in December 2009 by Biodiversity Information Standards (TDWG) as the 'Joint GBIF-TDWG Task Group on Multimedia Resources in Biodiversity'.

The standard was developed by the Joint GBIF- TDWG Multimedia Resources Task Group to fit with the suite of data standards being developed on behalf of the Global Biodiversity Information Facility (GBIF) by Biodiversity Information Standards (TDWG).

Funding has been provided by the Global Biodiversity Information Facility.

Grateful thanks go to Woods Hole Marine Biological Laboratory and the Encyclopedia of Life for hosting one of the meetings. This document, including some narrative is adapted from a corresponding document produced by the TDWG Natural Collections Descriptions (NCD) task group.

Summary

The Multimedia Resources Metadata schema ("MRTG schema") is a set of representation-neutral metadata vocabularies for describing biodiversity-related multimedia resources and collections.

Multimedia Resources are digital or physical artifacts which normally comprise more than text. These include pictures, artwork, drawings, photographs, sound, video, animations, presentation materials, interactive online media including, e. g., identification tool packages involving text and other media, etc. A multimedia collection is an assemblage of such objects whether curated or not and whether electronically accessible or not. For the purposes of this document we regard a collection of multimedia resources itself as a 'multimedia resource'. Wherever discussion or specification can apply only to a collection or only to a single media resource, we say so explicitly.

Multimedia descriptions are digital records that document underlying multimedia resources or collections. MRTG is focused on biodiversity-related multimedia resources. It shares terminology and concerns with many well known and important standards for describing access to resources such as Dublin Core(DC), Darwin Core (DwC), the Adobe Extensible Metadata Platform (XMP), the International Press and Telecommunications Council (IPTC) the Metadata Working Group (MWG) schema, the Natural Collections Schema (NCD), and others.. Where there is an exact match to concerns of such standards, MRTG adopts their identifiers and definitions. Where this is unsuitable, recommended cross-walks are given. MRTG particularly intends to ease the burden of holders of descriptions specified either by DwC or DC to allow use of those existing descriptions where appropriate.

This document accompanies the normative part of the MRTG standard[1], which is published in draft form on the MRTG Wiki,[2] where we also invite readers to register and comment on the draft. The standard consists of the series of class and property definitions. Each is identified by unique Uniform Resource Identifiers (URI). Normative definitions are provided in the document. In addition MRTG will develop recommended representations for serialization of MRTG descriptions in several important forms including the TDWG RDF ontologies, XML Schema, and Comma Separated Values (CSV). These will be submitted to TDWG for comment and adoption.

It is expected that MRTG will develop further as experience is gained in the projects that are making use of it.

1 http://www.keytonature.eu/wiki/MRTG_Schema_v0.6

2 http://www.keytonature.eu/wiki/MRTG

Figure 1 below augments a portion of Figure 2 of the non-normative portion of the NCD document[3]., It shows a number of kinds of biodiversity data-centric resources and illustrates typical user communities, data and metadata standards, and network services that support the discovery, analysis and integration of data. We extracted from the NCD figure the resources and relations between them, which we augment with three types not in the main purview of NCD. These are: Observations, Ecological Models, and - the focus of this work - Multimedia Resources. Applications exploiting each kind of these resources find utility or sometimes require the use of multimedia resources to document them. For example, the Biological Heritage Library is a project that provides scanned images of legacy literature at a far greater rate than it can provide digitized versions based on optical character recognition, and these images remain available to document any subsequent derived products. Thus digitized legacy literature is documented by the page images. Most scientific literature of course is also illustrated by photographs, graphs, or other artifacts in the purview of MRTG. Even the providers of "Molecular DNA" resources sometimes will offer original data as digital images of microarray chips.

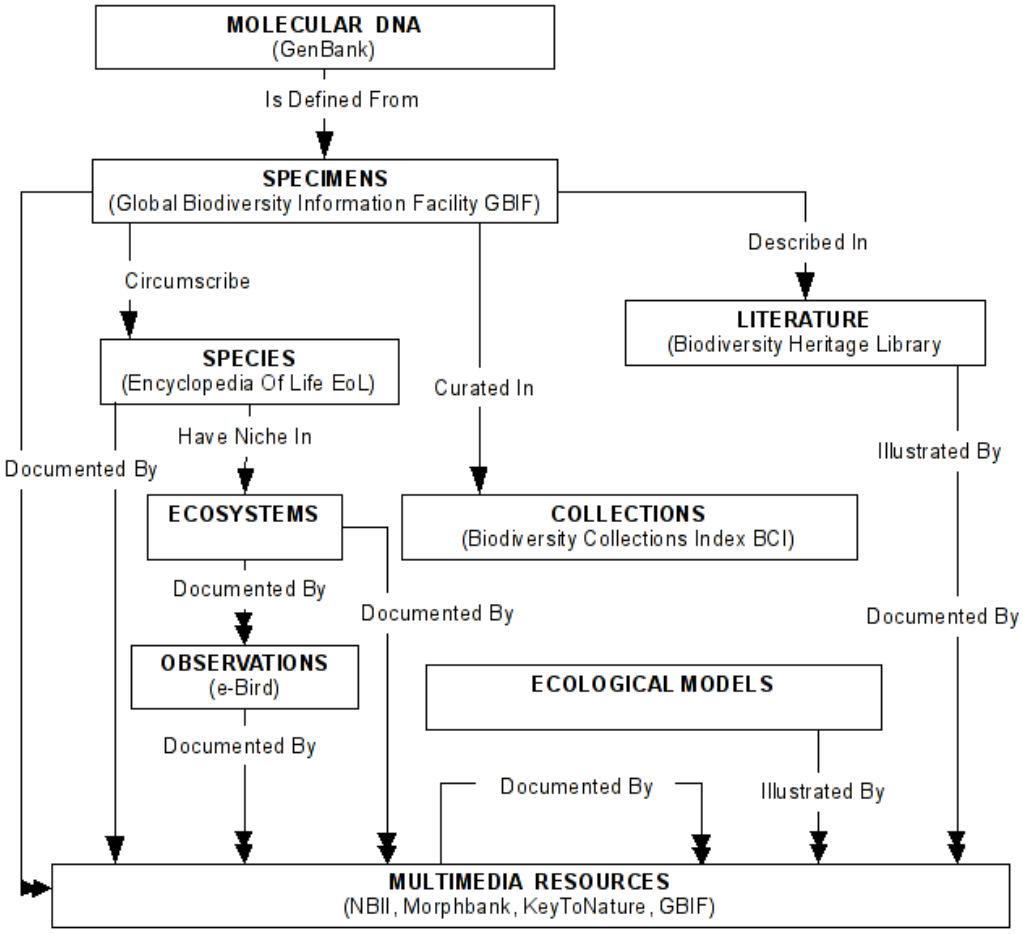

Figure 1. Relationships of Multimedia Resources to primary types of biodiversity resources

3 http://www.tdwg.org/fileadmin/subgroups/ncd/NCD_090.doc

Terminology in this document

There are many ways to organize metadata specifications, particularly as to the nomenclature of the constituents of the metadata. In this document and the normative documentation, we will closely follow a portion of the Dublin Core Metadata Initiative (DCMI) metadata nomenclature as described in Section 2.3 of the DCMI Abstract Model.[4].

- A *term* is a *metadata item* that forms part of the description of a multimedia resource.
- A *term* has a type which is one of *'Property'*, or *'Class',* We refer to a term of type Property and Class as simply a 'Property' or 'Class', respectively.
- A "*value* is a *resource* - the physical, digital or conceptual entity or *literal* that is associated with a *property* when a property-value *pair* is used to describe a *resource*. Therefore, each *value* is either a *literal value* or a *non-literal value*:
- A *literal value* is a *value* which is a *literal*.
- A *non-literal value* is a value which is a physical, digital or conceptual entity.
- A *literal* is an entity which uses a Unicode string as a lexical form, together with an optional language tag or datatype, to denote a *resource* "[DCMI Abstract Model]. In MRTG, the language tag appears as a value assigned to the metadata record.
- A Property is a term that has a value. The datatypes of values are specified in the normative document. Typically, the values are either a member of a fixed set of literals, a URI, a numerical type, free text, or the datatype and values from an external controlled vocabulary referenced in the standard.
- A Class is a term that has a set of Properties. Thus, the values of the properties in this set define what it means for a resource (whether multimedia or not) to be a member of the class. Typically if M is a member of class C we say "M is a C". We attempt to minimize the number of classes, because we want to support simple and flat serializations in which structured representation is cumbersome or impossible.
- A Vocabulary is a set of terms.
- A Multimedia Resource is anything that a provider identifies as belonging to one of the possible values of the MRTG Type term - and at least one of the Subtype term - values. A mechanism is given by which providers can supply a privately defined subtype that will not collide with the MRTG-defined Subtype values.
- A MRTG record is a set of terms conforming to the normative document and which contain at least the six mandatory terms described below and which describe a single multimedia resource (possibly including a Collection). n identifier , which may have been assigned to the resource by an external authority or by the provider of the metadata record.

Every MRTG term has a plain text Name, a URI and a plain text normative Definition. URI's for terms conform to the http URI scheme. Informally, one may understand this thusly: an http URI has the syntax of an http URL, but there is no expectation that putting it in a web browser will result in any information being returned to the browser, and if it does, the return may have no relevance. At a future time, terms may resolve to RDF or other forms. This conformance requirement applies only to the URIs that identify MRTG terms. Any others, such as might arise if the values of MRTG properties are taken from another

4 http://www.dublincore.org/documents/abstract-model/

controlled vocabulary chosen by the user, as some MRTG properties permit. In this case, those values may involve URIs conforming to a scheme given by that external vocabulary.

Because http URIs are rather lengthy, MRTG documents follow a standard practice of introducing a short abbreviation comprising a "namespace qualifier" separated by a colon from a mnemonic name closely related to the term's Name. The result is known as a qualified name. For example the normative wiki documentation for the Identifier term renders its URI as " dcterms:identifier" but hovering over it will reveal that its actual URI is http://dublincore.org/documents/dcmi-terms/#title. In this document we will follow the qualified name convention that is established by the wiki rendering. In fact, most of the URIs for terms borrowed from external vocabularies (about half of them) do in fact resolve to something in relevant documentation for that external standard. Sometimes it is not precise because the documentation is a PDF document and several (different!) URIs might apparently resolve to the same place. MRTG solicits discussion on the wiki at points where contributors find our association of a MRTG term with that from another standard as misleading or otherwise inappropriate.

The principal namespace qualifiers for term URIs in this document are

- dcterms. The DCMI type vocabulary documented at http://dublincore.org/documents/dcmi-terms
- dwc: The Darwin Core vocabulary proposed at http://rs.tdwg.org/dwc/index.htm
- Iptc4ampExt Geographic extensions to IPTC with namespace http://iptc.org/std/Iptc4xmpExt/2008-02-29/ documented in http://www.iptc.org/std/photometadata/2008/specification/IPTC-PhotoMetadata-2008_2.pdf
- mrtg: terms defined in the normative documentation and not derived from other controlled vocabularies. A URI is yet to be assigned at this writing
- xmp: The Adobe XMP vocabularies with namespace http://ns.adobe.com/xap/1.0/
- xmpRights The Adobe XMP rights vocabulary documented at http://ns.adobe.com/xap/1.0/rights

Motivation and Rationale

Many valuable multimedia resources exist that have no information stored in databases. Some may have a web presence and others not. Even those available online may not be adequately discovered by search engines, or may be lost in the noise of images from unreliable sources. A brief descriptive record as defined by the MRTG standard can act as the “business card” for a multimedia resource, providing enough information to identify and locate media resources by researchers, aggregators, decision makers, educators, or the general public.

The standard enables the aggregation of multimedia resource descriptions from many sources and facilitates resource discovery, including establishing relationships among multimedia resources in several locations. MRTG records can also be used as an aid for multimedia resources management processes, allowing an institution to take a step back and see which collections are most in need of conservation or would benefit from a higher priority for item-level cataloguing.

Among important uses identified by the Task Group that are facilitated by the metadata are:

1. Discovery;
2. Evaluation of fitness-for-use prior to fetching a resource (especially relevant for

off-line resources);

3. Use of metadata records as potential taxon occurrence evidence, or other biological inferences such as evidence for species interactions, habitats, and phenotypic variation;
4. Identification aids.
5. Easing the burden of multimedia resource providers and producers to gather and serve resources contributed by a wide variety of producers and custodians, particularly those with little or no IT expertise or support.

To ensure that the barriers to use are as low as possible, only 7 properties of a MRTG record are considered to be mandatory:

1. Identifier (dcterms:identifier): An arbitrary code that is unique for the resource, with the resource being either a provider, collection, or media item. Whereas the identifier must be globally unique for providers and collections (e. g. a URI), identifiers for media items may be unique only within collection or provider.
2. Type (dcterms:type): Any dcmi type term from http://dublincore.org/documents/dcmi-type-vocabulary/ may be used. Recommended terms are Collection, StillImage, Sound, MovingImage, InteractiveResource, Text.
3. Title (dcterms:title): Concise title, name, or label of institution, resource collection, or individual resource.
4. Metadata Language(mrtg:MetadataLanguage): Language of description and other meta data (but not necessarily of the image itself)
5. Copyright Owner (xmpRights:Owner): The name of the owner of the copyright.
6. Copyright Statement (dcterms:rights): Information about rights held in and over the resource. A full-text, readable copyright statement, as required by the national legislation of the copyright holder. On collections, this applies to all contained objects, unless the object itself has a different statement.

Existing standards

MRTG intends to provide metadata that describe either media resources themselves or collections of them. There are several well known or newly emerging standards which address these concerns, so one may ask: why not simply use them? In fact, MRTG does exactly that in about half of its 80 elements, almost all of which are optional. Indeed, as shown above, most of the six mandatory terms come from external controlled vocabularies. However, all existing controlled vocabularies, most notably the widely used Dublin Core -present very little opportunity to provide media resource content metadata that is specifically biologically relevant. Use of the Dublin Core alone would make it difficult to do media resource discovery with high precision. Thus, one consequence of using Dublin Core alone would be that queries will not be selective enough. By contrast the Darwin Core, presently being finalized as a TDWG standard[5] has more support for some such concerns, but not much about important intellectual property rights issues, nor ways to express relations between alternate versions of media resources (e.g. different resolution versions). In turn, neither of these controlled vocabularies has mechanisms for capturing technical metadata, such as EXIF, that the imaging systems themselves, or metadata embedding tools, such as Adobe Photoshop(tm) and the GIMP open source image

5 http://rs.tdwg.org/dwc/index.htm

editor, can insert into media files and streams. To address this, and in furtherance of the above goals, MRTG should be regarded as a synthesis of DC, DWC, and, where those are inadequate, some forward looking metadata standards that the camera manufacturers are presently planning to support within the cameras themselves, much as they now use EXIF[6]. Where any of these standards suffice, MRTG metadata terms and definitions are those of such standards. In some instances, we find that none of these address concerns that our experience suggests are held by a wide variety of image contributors, especially those with limited access to sophisticated IT staff and Digital Librarians. The MRTG schema might be regarded as an extension to the union of small subsets of three accepted standards (together with a framework to insure that use of metadata from several standards can be understood by people and machines as referring to the same resource. Put another way, half of MRTG may be viewed as a wrapper around DWC, DC, XMP, and IPTC.[7]

Since the overwhelming portion of the MRTG metadata fields are optional, a resource provider that can already serve, Dublin Core metadata, could essentially serve little else but that, plus a suitable globally unique identifier to tie all the metadata to the same object. Similarly, a provider describing image content entirely with Darwin Core terms, might have little more to do. Both such providers would, however, find that value-added services such as metadata-indexers and caching aggregators. would be less likely to hold references to their media resources than if they had richer metadata. This gives a clear strategy for providers to increase the utility of their multimedia resources with little or no impact on their IT cyberinfrastructure services. They may need only to update mappings between their internal field names and the metadata terms specified by MRTG as personnel become available to do so.. As more resources become available for additional metadata provision, or and as community annotation mechanisms arise that providers can exploit, they can add the additional metadata at a pace determined by their own resources. If harvesters of the metadata monitor the (optional) MRTG:Metadata Modified property, the updated metadata can automatically be pulled by those value-added services, and more queries will return the provider's metadata and references to its media resources.

Common Concerns with other biodiversity information standards

MRTG regards Collections of Multimedia Resources themselves as a kind of Resource. Many types of Collections are describable in the pending TDWG Natural History Collections (NCD) proposed standard. If a provider wishes only to provide for discovery of a multimedia Collection without regard to discovery of and access to its contents (other than subCollections), it will often be immaterial whether NCD or MRTG metadata, or both, are served. This is all the more so if the NCD CollectionIdentifier and the MRTG Identifier have the same value. While MRTG Collection types are richer than NCD types, it is an open question whether MRTG's variety in this case is utilized.

There is substantial overlap with concerns of DarwinCore, notably with respect to

6 The Metadata Working Group (MWG, http://www.metadataworkinggroup.org/) is an industry consortium (Adobe, Apple, Canon, Microsoft, Nokia, and Sony) organized to specify how to exploit the Adobe Extensible Metadata Platform, XMP (http://en.wikipedia.org/wiki/Extensible_Metadata_Platform) for embedding into common image file formats metadata in several widely used controlled vocabularies. Although MWG's thrust is mainly toward consumer applications, over two dozen open source and commercial software products and platforms support XMP and Adobe has placed a Developers' Toolkit under an open source license. Along with proposals for standarized serializations of the representation-neutral MRTG metadata schema, MRTG intends to propose a TDWG standard way of embedding such serializations in multimedia files using XMP.

7 IPTC is mature standard from the International Press and Telecommunications Council (http://www.iptc.org). Its Intellectual Property Rights support finer grained controlled vocabularies than dc, providing better machine processing for discovery and fitness-for-use.

taxonomic, geographic, and temporal coverage of the data being described by the metadata record. We use DWC terms for most of those metadata, except in the case of geography, where it seems advisable to follow the standards for recording geographic locations that is set and supported by the producers of media management and manipulation software (Adobe, Microsoft, etc.) as well as the camera manufacturers themselves. Data in the XMP/IPTC standards will be embedded in images and are supported in media databases.

Concerns not emphasized in other biodiversity information standards

Some of the concerns mentioned here are also those of bibliographic metadata such as the Dublin Core,. These are, however, not explicitly of detailed concern in existing TDWG biodiversity standards, and some are not adequately addressed by DC. Some such concerns are below.

Size: Individual multimedia resources such as images, and especially video and sound are very large compared to specimen records, observation data, or species descriptions. The main consequence of this is that multimedia metadata must support use cases for which humans or software agents can, without fetching the resource, attempt to assess the fitness of the underlying media resource for the desired use, typically by use of a search based on a fine-grained controlled vocabulary. However, absent hit-and-miss natural language searches, it is not possible even using both DC and DWC for a metadata provider to answer a request of the form "Supply me with sizes and URL access points for still images of Dictyophora indusiata which has Spanish metadata available".

Intellectual Property Rights: DWC describes physical objects, whose ownership is generally governed by property laws not considered part of the Intellectual Property Rights corpus of law. Some impending standards about scientific literature address these, but rarely are publication reproduction permission issues as varied as for multimedia, which have a history of being treated as creative works of art, not necessarily as facts.

Provenance: For any scientific data, it is clearly important to know how and when the data may have been changed from its original gathering. This is particularly important for media, which are commonly edited for one or another purpose. If carelessly done, this may destroy some if the modified object's utility. No TDWG standards or proposed standards seem very strong about provenance, including MRTG, which provides only the Derived From property in order to provide a reference to another resource. This is somewhat akin to the NCD DerivedCollection term, which identifies a Collection record as having been produced by a query do another Collection. However, that apparently does not; identify the source collection or the query. A future version of MRTG will add more provenance terms.

Multimedia Resource Descriptions

The term Multimedia Resources encompasses a wide variety object of interest to biologists and the communities with whom they interact for research, education, and public service. Some instances of multimedia are familiar. These include:

- Still images from cameras, scanners, or medical and industrial imaging devices
- Movies with or without sound
- Audio recordings

In some of the above cases, these resources may exist in electronic or non-electronic form

or both. The electronic form may be analog or digital, the latter being more amenable to storage and exchange with computers. The digital form may have been born digital, i.e. originally captured as a digital object, or it may have been created from a non-digital object. As with biological specimen records, publications, field notes, experimental data and other artifacts of the practice of science, there is a large quantity of such material that has not yet been digitized. Yet that may be available, albeit with greater expense and inconvenience than digital resources. These analog (including paper) resources still require descriptive metadata to promote discovery and to ascertain fitness for use. At least as important, some of the metadata is itself of scientific and educational use even if the object is not conveniently accessible. Evidence for geo-referenced taxon occurrence is one such use.

MRTG metadata also can describe resources less often thought of as multimedia objects. These include:

- Interactive software applications, either on the web or available for standalone use.
- Taxonomic identification keys
- Collections of multimedia resources.
- Web sites not otherwise falling into one of the above categories.

MRTG records

The normative MRTG metadata record specification is independent of the way in which those records are rendered into electronic form. MRTG intends to publish specifications for such rendering represented in RDF conforming to the TDWG LSID vocabularies, represented in XML constrained by an XML-Schema, and represented in plain text as comma separated values (CSV).

The language of the normative MRTG specification is English, but this in no way constrains applications from using labels or content of the metadata in local languages. Because its language is English, each metadata item in the normative document has an English label (which might, for example be part of a user interface), but these, too, are not required to be used by applications, although their use is strongly encouraged, at least in documentation.

As mentioned earlier, MRTG metatada record is a set of terms describing the underlying multimedia resource that the record describes . Each term is identified by a Uniform Resource Identifier (URI). These are URIs of the attribute, not of the underlying resource, and they simply specify which term is being provided. There are many URI schemes some of which have been registered with the Internet Assigned Names Authority (IANA). All MRTG term URIs, conform to the http URI Scheme . This is chosen because this widely used URI scheme uses the familiar internet URL syntax as its URI syntax. But this familiarity gives rise to a common misconception, namely that pasting the URI into a browser URL line, or providing it to some other application that respects the http protocol, should result in the application returning some information about the object identified by the URI. Such behavior is usually called resolution of the URI and is in no way guaranteed for MRTG term URIs. Where possible, we in fact try to make http URIs be resolvable, with the information returned being documentation for how the metadata attribute identified by that URI is defined or use. To reiterate: for MRTG term URIs, any such resolution will never contain information about the underlying multimedia resource being described. For this reason, few human-centric MRTG applications should ever present the URIs to users, nor use them as linking mechanisms. (One possible exception is an application for assigning metadata to multimedia resources, where such a use may provide a thesaurus

entry aiding the user in the semantics of the metadata property. However, the incidental nature of the resolution, and its lack of guaranteed long term persistence, makes even this an approach that should be considered with extreme caution.) Finally, note that some external controlled vocabularies are defined in PDF or other documents that do not have URL links directly to each defined term. In these cases, any resolution available from the normative document may only link to the beginning of the document, leaving it necessary to search in the document for the referenced definition.

Associated to each MRTG property is its value. The datatype of this value is also specified in the normative document. Datatypes can include free text, specific literals taken from a controlled vocabulary specified in the normative document, or a number of other datatypes specified and described in the normative document. In the case of a controlled vocabulary, it is important to note that whatever an application may present in a user interface, any MRTG metadata interchange MUST use the literals from a specified controlled vocabulary when one is specified, even if the record is declared to be a record in a different language than that of the controlled term. An important example is the Type metadata field, which is required to come from the corresponding vocabulary from Dublin Core. (We also add to that a mandatory field Subtype.. Similarly, agents answering MRTG metadata queries MUST be able to consume and respond to queries framed with the controlled vocabulary. Nothing in the normative document prevents a MRTG data provider from asserting it has no records with a given controlled term, nor from internally mapping between a controlled vocabulary and its internal attributes, whose names may well be in a language other than English. Only a small number of MRTG properties take values in a specific, English-based controlled vocabulary. This will become relevant only for metadata interchange. Of the six mandatory terms, only Type hasany such requirements.

A MRTG record consists minimally of the six mandatory fields (Identifier, Type, Subtype, Title, Metadata Language, Copyright Owner, and Copyright Statement). Later in this document we describe recommended serializations for the actual representation of the metadata, including one based on the TDWG LSID Vocabularies (an RDF/XML form), one based on an XML-Schema, and one based on Comma Separated Values (CSV). Each of these will be submitted to TDWG as standards in addition to the Normative, representation-free standard that accompanies this document.

In some cases, some metadata terms are necessarily related to others (e.g. various versions of an image must be associated the "main" version). However, spreadsheets and other flat sources of contributer metadata are regarded as particularly important, and in many of these it is difficult to represent such structural relationships. Consequently a MRTG metadatum is itself mainly flat, the exception being a few structures designated as Classes in the normative document. One consequence of this is that, for some purposes, a metadata Provider might have to make several metadata records available about the same underlying resource, because the representation neutral MRTG specification does not provide for "subproperties" on its properties, nor for relations in most cases. An important case surrounds multilingual metadata. Because each metadata record is in a fixed language specified by the Metadata Language property (this is the language of the record,, not the multimedia resource, in case it should have one), a Provider might have to offer several metadata records about the same multimedia resource. The six required terms must be provided in every metadata record, even if repeated in other metadata records. At the date of this writing, the normative document does not provide a mechanism for identifying a metadata record that might be overarching, in the sense that its optional terms may be regarded as defaults for any not specified in other records about the same resource. This point is under discussion on the MRTG Wiki.

Many items may be repeated in a MRTG record, but some may not, as indicated in the normative document. For example the Modified item corresponds to a date at which the media resource was modified and may be repeated to reflect the history of the resource.

By contrast, Date Available is a single date or a single range of dates at which the underlying resource became available

Implementation and Compliance

MRTG is defined in as representation-neutral a way as possible. It provides natural language definitions of classes, properties and instances that are identified by URIs and it makes recommendations on the use and content of properties from other vocabularies (Dublin Core , Darwin Core, and some parts of XMP.).

The URIs defined here may be used across a number of technologies, such as namespaces in XML Schema-validatable documents, RDF, and column headings in comma delimited text files.

This approach facilitates:

- Embedding of MRTG data within other standards such as descriptions of specimens or literature.
- The extension of MRTG records with other data types such as the extensive geographic controlled vocabularies of the Open Geospatial Consortium (OGC)
- Cross walking between technologies such as a Comma Separated Value file, an RDF graph, an XML document and a JSON object.

The MRTG representation-neutral normative standard itself does not provide an off-the-shelf, self validating exchange format. Multiple such exchange formats meeting different requirements can be defined and this standard allows mapping between them.

Further Information

- GBIF Multimedia Resources Task Group wiki
  http://wiki.gbif.org/gbif/wikka.php?wakka=MultimediaResourcesTaskGroup
- Joint TDWG-GBIF MRTG Charter
  http://www.tdwg.org/charters/article/view/448/36
- MRTG Discussion Wiki http://www.keytonature.eu/wiki/MRTG This will also hold discussion to proposed serializations as they become available.
- MRTG Vocabularies v0.6 normative wiki pages for discussion:
  http://www.keytonature.eu/wiki/MRTG_Schema_v0.6.
- To register for the discussion wiki:
  http://www.keytonature.eu/wiki/Special:RequestAccount Registration not required except to contribute or to receive email notice of changes to the wiki.
- Register for the mailing list tdwg-img@tdwg.org at
  http://lists.tdwg.org/mailman/listinfo/tdwg-img. This email list tracks more general activity of the TDWG- Image Interest Group (IIG)
- Follow the TDWG IG news announcements at http://www.tdwg.org/activities/img/ This will have major announcements of MRTG developments, but not follow the day-to-day discussion on the normative wiki.

Appendix I: Glossary	
BCI	Biodiversity Collections Project. A central index of biodiversity collections around the world, based on NCD. http://www.biodiversitycollectionsindex.org/
BHL	Biodiversity Heritage Library. A consortium of institutions, dedicated to digitizing legacy biodiversity literature, mainly that which is out of copyright. http://www.biodiversitylibrary.org/
DC	Dublin Core. Metadata element set that is a standard for cross-domain information resource discovery. http://dublincore.org/documents/1999/07/02/dces/
DCMI	Dublin Core Metadata Initiative. The organization engaged in developing Dublin Core metadata standard. http://dublincore.org/
DWC	The Darwin Core is a proposed standard for representation of specimen records. It has been in wide use for several years in a number of unstandardized, sometimes inconsistent, versions. The proposed standard is at http://rs.tdwg.org/dwc/index.htm. The task group proposing it has a web page at http://rs.tdwg.org/dwc/index.htm intended to supplementhttp://wiki.tdwg.org/twiki/bin/view/DarwinCore/WebHome
EDIT	European Distributed Institute of Taxonomy. Consortium to integrate taxonomic research. http://www.e-taxonomy.eu/
EOL	Encyclopedia of Life. Information about many species. http://eol.org.
EXIF	A widely used tagging format for digital image metadata that is often embedded in the image files, particularly by modern digital cameras. Many image rendering applications can read and display EXIF data. See http://en.wikipedia.org/wiki/Exchangeable_image_file_format for a history and description/

GBIF	Global Biodiversity Information Facility. Interoperable network of biodiversity databases and information technology tools. http://www.gbif.org/
IANA	Internet Assigned Names Authority. Specifies the forms of, and registers instances of, names of various protocols in use on the internet. http://www.iana.org/. See especially information on the *IANA http URI scheme* at http://en.wikipedia.org/wiki/URI_scheme and
IPTC	IPTC is mature standard from the International Press and Telecommunications Council Its Intellectual Property Rights support finer grained controlled vocabularies than dc, providing better machine processing for discovery and fitness-for-use. The current version is a vocabulary for XMP. http://www.iptc.org
JSON	JavaScript Object Notation. Lightweight data-interchange format. http://www.json.org/
K2N	KeyToNature An EU project for the provision of online identification tools with a special focus on formal education. http://www.keytonature.eu
Morphbank	A specimen image repository http://www.morphbank.net/
MWG	The Metadata Working Group is an industry consortium (Adobe, Apple, Canon, Microsoft, Nokia, and Sony) organized to specify how to exploit the Adobe Extensible Metadata Platform, XMP (http://en.wikipedia.org/wiki/Extensible_Metadata_Platform) for embedding into common image file formats metadata in several widely used controlled vocabularies. Although MWG's thrust is mainly toward consumer applications, over two dozen open source and commercial software products and platforms support XMP and Adobe has placed a Developers' Toolkit under an open source license. Along with proposals for standarized serializations of the representation-neutral MRTG metadata schema, MRTG intends to propose a TDWG standard way of embedding such serializations in multimedia files using XMP. http://www.metadataworkinggroup.org/)
NBII	U.S. National Biological Information Infrastructure. Its image library (in May 2009 it will be officially names as the Library of Images From the Environment, LIFE) is at http://images.nbii.gov/ or http://life.nbii.gov/.

NCD	Natural Collections Description is a data standard designed for describing collections of physical objects such as specimens. It can accommodate collections of media objects, but cannot relate them to descriptios of the objects themselves. http://www.tdwg.org/activities/ncd/
OGC	Open Geospatial Consortium. Provides standards for geospatial data representation and exchange. http://www.opengeospatial.org/
RDF	Resource Description Framework. Lightweight ontology system to support knowledge exchange online. http://en.wikipedia.org/wiki/Resource_Description_Framework
TDWG	Taxonomic Databases Working Group. Now known as the Biodiversity Information Standards (TDWG), it is a group that develops standards and protocols for sharing biodiversity data. http://www.tdwg.org/
URI	Unique Resource Identifier. Generic term for linking web resources, includes URLs. http://en.wikipedia.org/wiki/Uniform_Resource_Identifier
XML	Extensible Markup Language. A simple flexible text format playing an increasingly important role in the exchange of a wide variety of data on the Web. http://www.w3.org/XML/
XMP	Adobe Extensible Metadata Platform (XMP) is a framework for embedding metadata into media files. Adobe provides a BSD-licensed open-source XMP developers toolkit which includes documentation about how to represent metadata in XMP. The XMP specification itself is licensed by Adobe under a "Public Patent License" (http://www.adobe.com/devnet/xmp/pdfs/xmp_public_patent_license.pdf), by which Adobe grants everyone the right to make XMP-compliant components of their applications, but it reserves the right to withdraw the license in case such a compliant component infringes "Essential Claims" of any patent. See http://www.adobe.com/devnet/xmp/ for download information. See Also MWG in this table.

Appendix II: MRTG Development History

2006, November	TDWG Image Interest Group initiated
2008, March	GBIF commissions Multimedia Resources Task Group
2008, June	GBIF Multimedia Resources Task Group meeting in Copenhagen, Denmark
2008, August	GBIF Multimedia Resources Task Group meeting in Woods Hole, USA
2008, October	TDWG Image Interest Group met in Freemantle, Australia at the 'TDWG Annual Conference 2008'
2008, December	Joint GBIF-TDWG Task Group on Multimedia Resources in Biodiversity commissioned
2009, February	GBIF Multimedia Resources Task Group met in Copenhagen, Denmark to refine the metadata schema
2009, March	GBIF – TDWG Multimedia Resources Metadata Schema (MRTG) ver. 0.6 drafted