

GLOBAL
BIODIVERSITY



INFORMATION FACILITY

Darwin Core

Francisco Pando, GBIF Spain



NATIONAL MUSEUMS OF KENYA

WHERE HERITAGE LIVES ON

KENBIF TRAINING WORKSHOP

Nairobi, 6th to 7th June 2011

Based upon



Darwin Core Ratified in the Year of Darwin

Gail E. Kampmeier

Illinois Natural History Survey

Institute of Natural Resource Sustainability

University of Illinois at Urbana-Champaign

Entomological Collections Network Meeting, Indianapolis, IN
13 December 2009

Informatics Infrastructure and Portal (IIP)

David Remsen

ECAT Program Officer

August 2010

What is Darwin Core?



- Data standard for sharing and publishing biodiversity occurrence data (i.e., “this species was in this place, at that time, according to someone”)
- It may pertain collections, individual observations, surveys,... almost anything
- It allows to discover and access very heterogeneous data in an uniform way

- Standard has evolved over 11 + years
 - Originally a product of the Species Analyst project out of the University of Kansas
 - Terms were first used in MaNIS (Mammal Networked Information System) as Darwin Core 2 with DiGIR protocol (Distributed Generic Information Retrieval)
 - The DwC is based on standards developed by the Dublin Core Metadata Initiative (DCMI) & is an extension of it (dc) for biodiversity information
 - DwC is part of a larger set of vocabularies & technical specifications under development or maintained by TDWG

Versions



- Historical versions shaped database structure
 - "Classic" (1.2), or "Draft" (1.4) or "Archive" (GBIF)
 - Extensions for curatorial, geospatial, paleontology, & specimen interaction information (pollination ecology)

Darwin Core Versions

Name	Namespace	Number of terms	XML Schema	Date Issued
Darwin Core 1.0	Not Applicable	24	(Z39.50 GRS-1)	1998
Darwin Core 1.2 (Classic)	http://digir.net/schema/conceptual/darwin/2003/1.0	46	[2]	2001-09-11
Darwin Core 1.21 (MaNIS/HerpNet/ORNIS/FishNet2)	http://digir.net/schema/conceptual/darwin/2003/1.0	63	[3]	2003-03-15
Darwin Core OBIS	http://www.iobis.org/obis	27	[4]	2005-07-10
Darwin Core 1.4 (Draft Standard)	http://rs.tdwg.org/dwc/dwcore/	45	[5]	2005-07-10
Darwin Core Terms (properties)	http://rs.tdwg.org/dwc/terms/	172	[6]	2009-10-09

Relationship to Dublin Core



- Darwin Core is based on the standards developed by the Dublin Core Metadata Initiative, and is an extension of Dublin Core for biodiversity information
- Simple set of terms for providing descriptive information about biological collections
 - Simple Darwin Core
- Like Dublin Core, emphasizes flexibility in order to encourage adoption and maximize re-use of data

Dublin Core in one slide

1. Title
2. Creator
3. Subject
4. Description
5. Publisher
6. Contributor
7. Date
8. Type
9. Format
10. Identifier
11. Source
12. Language
13. Relation
14. Coverage
15. Rights

Title: The Anatomy Lesson of Dr. Nicolaes Tulp
Creator : Rembrandt Harmenszoon van Rijn
Description : Dr. Nicolaes Tulp is pictured explaining the musculature of the arm to medical professionals. Some of the spectators are various doctors who paid commissions to be included in the painting. The painting is signed in the top-left hand corner Rembrandt f[ecit] 1632. It is the first known instance of Rembrandt signing a painting with his forename as opposed to the initials RHL (Rembrandt Harmenszoon of Leiden), and is thus a sign of his growing artistic confidence.

Type: oil painting
Coverage: 1632
Rights: Mauritshuis, La Haya
...



- Ratified in 2009
- Significant additions/refinements
- Set of terms (172)
 - <http://rs.tdwg.org/dwc/terms/index.htm>
- Expressed via XML
- Simple Darwin Core (Subset)
- Express as Text
 - <http://rs.tdwg.org/dwc/terms/guides/text/index.htm>

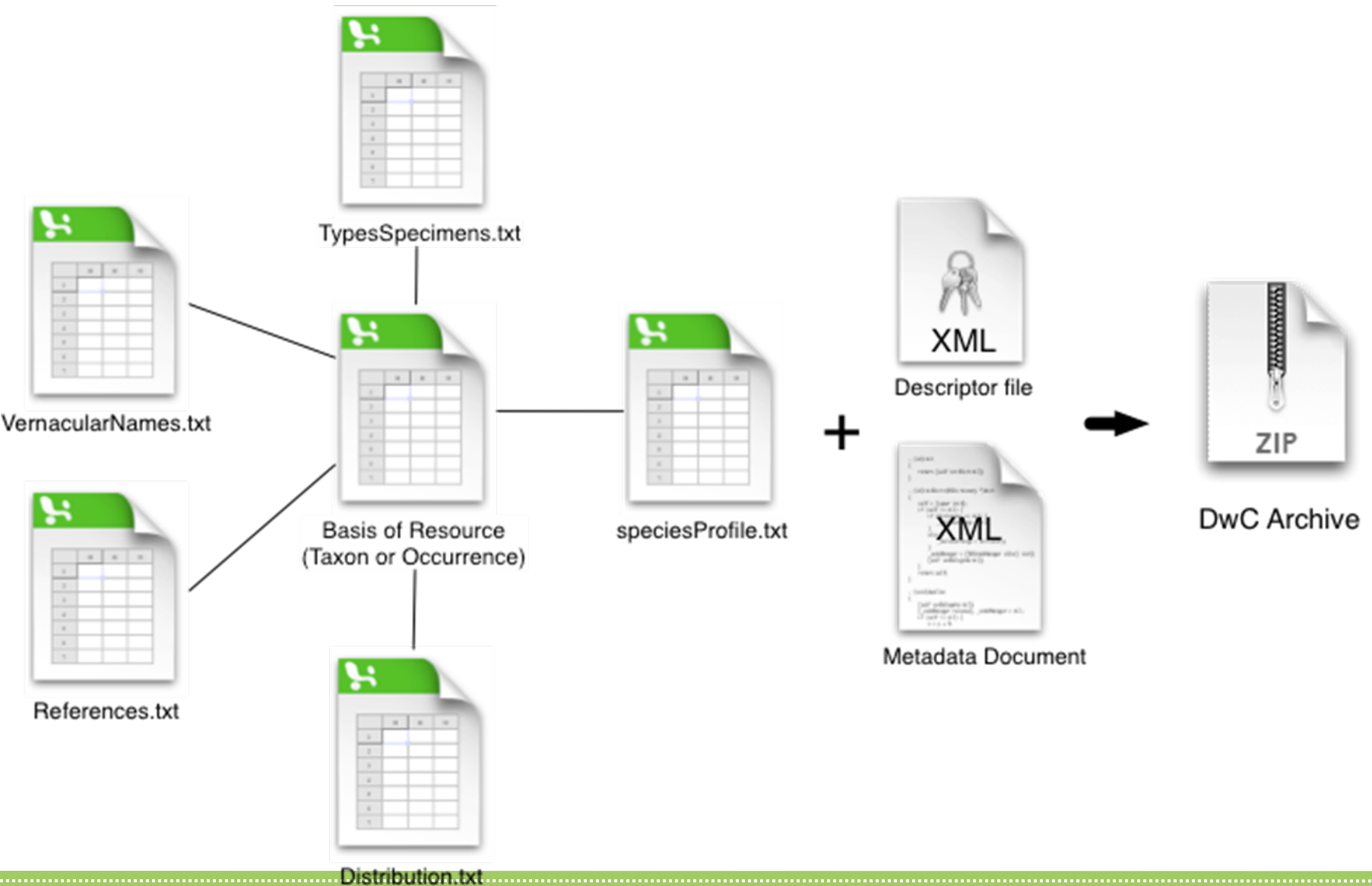
Simple Darwin Core



- Most prevalent structure
- A predefined subset of common Darwin Core terms that can be used in many biodiversity situations
- Easily shared in many ways including text files and XML documents
- Some “rules”:
 - No required fields
 - No field names can be repeated in a record
 - Use of controlled vocabulary for those fields that recommend the use of such vocabulary

[See a MS-Access database template of Simple Darwin Core >>](#)

Darwin Core Archives (DwC-A)



XML Descriptor file



```
<archive xmlns="http://rs.tdwg.org/dwc/text/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://rs.tdwg.org/dwc/text/ http://rs.tdwg.org/dwc/text/tdwg_dwc_text.xsd
  metadata="http://www.biodiv.org/docs/metadata/whale_catalogue.eml">

  <core encoding="UTF-8" fieldsTerminatedBy="\t" linesTerminatedBy="\n" fieldsEnclosedBy=" " ignoreHeaderLines="0"
  rowType="http://rs.tdwg.org/dwc/terms/Taxon">
    <files>
      <location>taxa.txt</location>
    </files>
    <id index="0" />
    <field index="1" term="http://rs.tdwg.org/dwc/terms/kingdom"/>
    <field index="2" term="http://rs.tdwg.org/dwc/terms/phylum"/>
    <field index="3" term="http://rs.tdwg.org/dwc/terms/class"/>
    <field index="4" term="http://rs.tdwg.org/dwc/terms/order"/>
    <field index="5" term="http://rs.tdwg.org/dwc/terms/family"/>
    <field index="6" term="http://rs.tdwg.org/dwc/terms/genus"/>
    <field index="7" term="http://rs.tdwg.org/dwc/terms/species"/>
    <field index="8" term="http://rs.tdwg.org/dwc/terms/infraspecies"/>
    <field index="9" term="http://rs.tdwg.org/dwc/terms/infraspeciesRank"/>
    <field index="10" term="http://rs.tdwg.org/dwc/terms/scientificNameAuthorship"/>
    <field default="ICZN" term="http://rs.tdwg.org/dwc/terms/nomenclaturalCode"/>
  </core>

  <extension encoding="UTF-8" fieldsTerminatedBy="\t" linesTerminatedBy="\n" fieldsEnclosedBy=" " ignoreHeaderLines="0"
  rowType="http://rs.gbif.org/terms/1.0/VernacularName">
    <files>
      <location>vernacular.txt</location>
    </files>
    <coreid index="0" />
    <field index="1" term="http://rs.tdwg.org/dwc/terms/vernacularName"/>
    <field index="2" term="http://purl.org/dc/terms/language"/>
    <field index="3" term="http://rs.tdwg.org/dwc/terms/countryCode"/>
  </extension>
</archive>
```

Authoring meta.xml



GBIF Metafile Maker v1.0

Core data file: GNA Vernacular Names meta.xml

Extension List

Select	Index	Term	Required	Static/Variable Mappings
<input checked="" type="checkbox"/>	1	dwc:vernacularName	Yes	
<input checked="" type="checkbox"/>	2	dc:source		
<input checked="" type="checkbox"/>	3	dc:language	Yes	EN
<input checked="" type="checkbox"/>	4	dc:temporal		
<input checked="" type="checkbox"/>	5	dwc:locationID		
<input checked="" type="checkbox"/>	6	dwc:locality		
<input checked="" type="checkbox"/>	7	dwc:countryCode		US
<input checked="" type="checkbox"/>	8	dwc:sex		
<input checked="" type="checkbox"/>	9	dwc:lifeStage		
<input checked="" type="checkbox"/>	10	gbif:isPlural		
<input checked="" type="checkbox"/>	11	gbif:isPreferredName		
<input checked="" type="checkbox"/>	12	gbif:organismPart		
<input checked="" type="checkbox"/>	13	Unmapped data		
<input checked="" type="checkbox"/>	14	dwc:taxonRemarks		

File Settings

File name/location: commonNam...

CSV file
 Tab separated
 Other

File encoding: UTF-8
Field delimiter: \t
Fields enclosed by: ""
Line ending: \r
Ignore header row: True

This extension supports the publication of vernacular name data to the Global Names Architecture.

- See a “real” Darwin Core Archive>>

At your command



Francisco [Paco] Pando

GBIF Spain Node Manager

GBIF NODES Committee Chair

GBIF Spain

Real Jardín Botánico - CSIC

Claudio Moyano 1

28014 Madrid, Spain

pando@gbif.es

www.gbif.es

Phone: + 34 91 420 3017

Fax: + 34 91 420 0157



