Protocolos y herramientas para integrar y publicar información de biodiversidad

Hojas electrónicas e IPT

Aurelio Sanabria
asanabria@inbio.ac.cr

http://www.inbio.ac.cr/

Agenda

- Contexto.
- Flujos de datos.
- Estándares para compartir datos
  - Darwin Core (DwC)
  - Plinian Core (PliC)
  - Ecological Metadata Language (EML)
- Publicar estos datos
  - Hojas de cálculo.
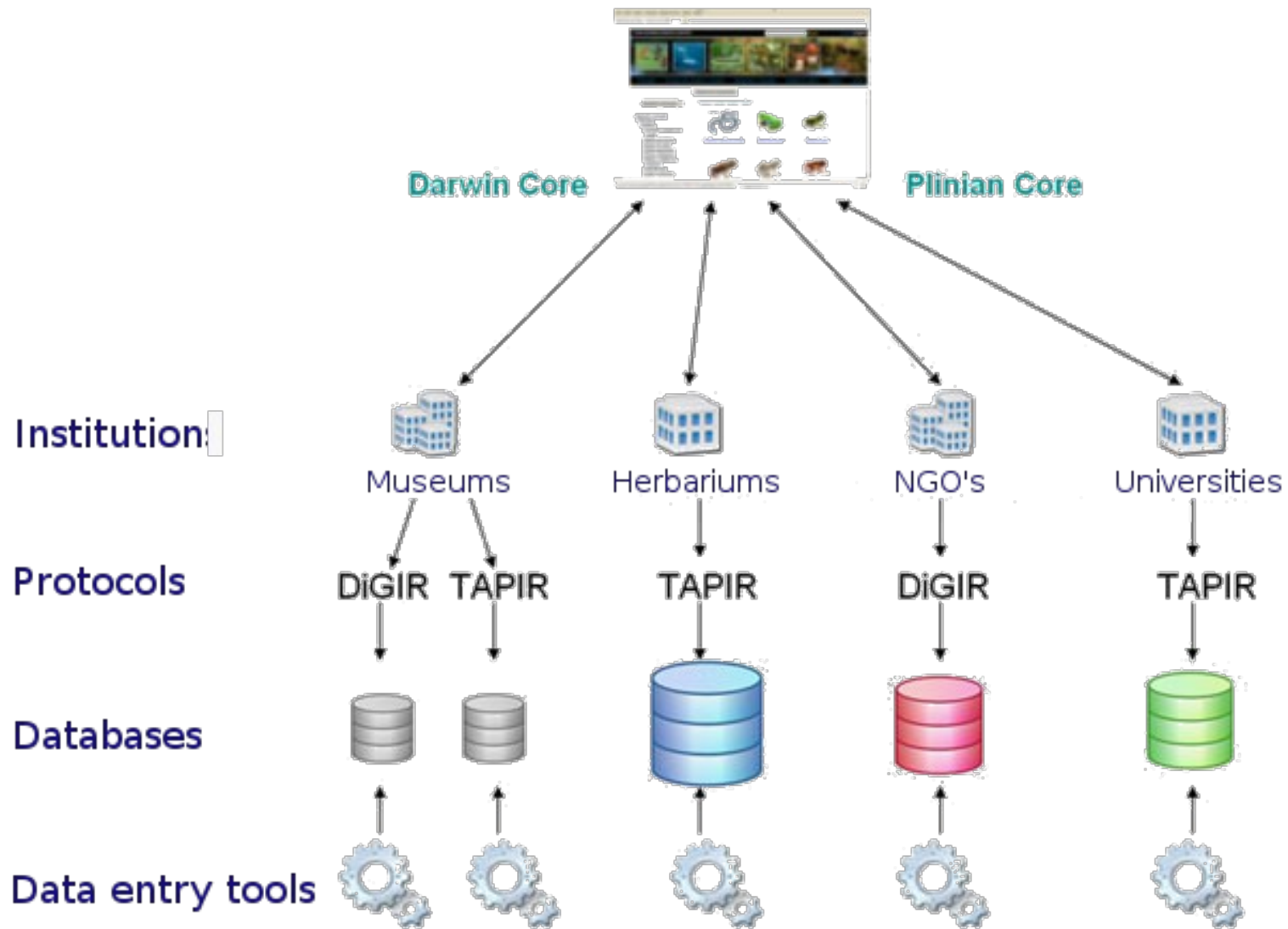 - Herramienta integrada de publicación (IPT).

Contexto

- Proveedores de datos
- Metadatos.
- Datos de observaciones y especímenes.
- Descripciones de especies.
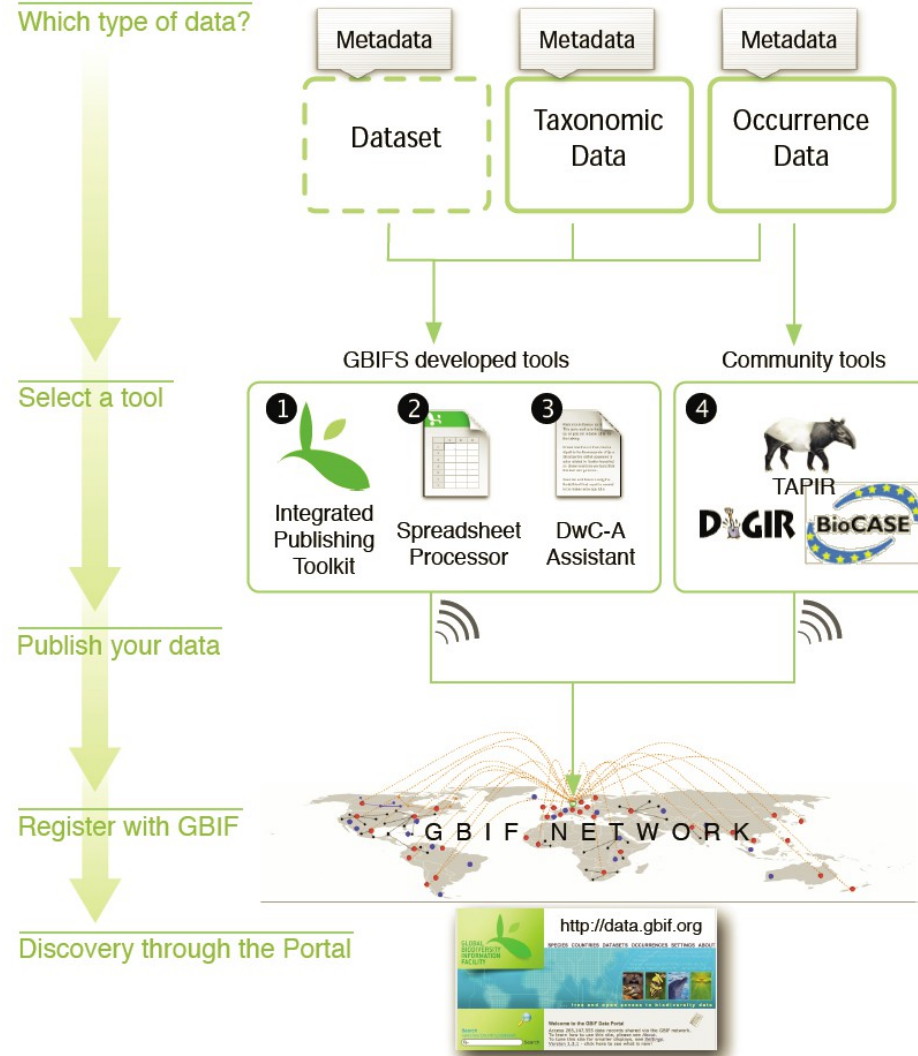- Datos taxonómicos.
- Registros de datos.

Flujos de datos

Flujos de datos

Estándares para compartir datos

- Darwin Core (DwC)
- Plinian Core (PliC)
- Ecological Metadata Language (EML)

Darwin Core

- Incluye un glosario de términos con la intención de facilitar el compartir información acerca de biodiversidad mediante la publicación de definiciones de referencia, ejemplos y comentarios para estos términos.

- El glosario promueve el uso consistente de términos a través de aplicaciones y disciplinas.

Darwin Core

ABCD (Estándar TDWG)

- – > 1200 conceptos
- – XML
- – Utilizado en BioCASE y TapirLink

Darwin Core (pre-estándard v. 1.2, 47 versiones distintas)

- – 48 conceptos, especímenes
- – XML
- – Utilizado en DiGIR

Darwin Core (pre-estándard v. 1.4)

- – 46 conceptos (mas extensiones), especímenes
- – XML
- – Utilizado en TapirLink

Darwin Core (Estándar TDWG)

- – 172 conceptos (156 presentes en el "Darwin Core Simple"), información de biodiversidad.
- – CSV, XML, RDF, JSON, …
- – Utilizado en texto plano, TapirLink, Archivos Darwin Core…

Plinian Core

Es un conjunto de conceptos que define los atributos básicos necesarios para integrar y recuperar la información acerca de especies que usuarios especializados en biodiversidad o de otras áreas requieren.

Plinian Core

- Última versión: 2.3 Flat

- Nueva versión (en curso): 3.0

  - Esquema conceptual.

  - Diferentes implementaciones (más simples).

Ecological Metadata Language

- Es una especificación de metadataos desarrollado dentro del campo de la ecología.

- Mayormente diseñado para describir recursos digitales.

- Permite el uso de subconjuntos o "perfiles"

Ecological Metadata Language

Para describir recursos de datos
Incluso si no estan publicados.

EML.xml

- Título y resumen (Abstract)
- Cita y Atribución
- Contactos y Autores
- Alcance geográfico
- Métodos de muestreo.
- Bibliografía
- Y mucho más…

Hojas de cálculo

Hojas de cálculo

http://tools.gbif.org/spreadsheet-processor/

free and open access to biodiversity data
Darwin Core Archive Spreadsheet Processor

BETA

Language: English

GBIF

Directions:

1. Review the available spreadsheet templates
2. Download and complete the templates according to the user guidelines
3. Click the "Choose File" button on the right to upload the completed spreadsheet or
4. Email it to spreadsheets@tools.gbif.org
5. Receive a validated Darwin Core Archive file or EML metadata document.
6. To publish, put the Darwin Core Archive file and EML metadata document online, with a stable URL.
7. Follow the GBIF data registration guidelines to register the data files with GBIFS.

Note: The processor does NOT publish a datafile to GBIF. It provides a publication-ready file.

Available Spreadsheet Templates – The processor currently only supports a set of pre-configured spreadsheet 'template' files. Each is available in the older XLS format and the new XLSX format, which supports more than 64,000 rows.

Upload File:

Browse...

Process File

or

mail to:
spreadsheets@tools.gbif.org

The following templates are available:

Metadata - Use this template to describe a database or other data resource. Processor output is an Ecological Metadata Language Document that conforms to a GBIF metadata profile. [XLSX format]

Species Occurrence - Use this template to record or store basic species collections or observational data. [XLSX format]

Species Checklists – These templates provide different methods for recording and storing simple annotated species checklists. The different versions offer different methods for recording classification information.

Checklist 1 - Supports a 'normalised' classification where the classification is stored in two spreadsheet columns: one stores the ID of the taxon and another stores the ID of the parent taxon. [XLSX format]

Checklist 2 - Classification in columns where the user can define the specific higher taxon groups. Useful when using super- or sub- ranks or other ranks outside the core Linnaean ranks. [XLSX format]

Checklist 3 - Classification in columns is limited to the basic Linnaean ranks of Kingdom, Phylum, Class, Order, Family, Genus, etc. [XLSX format]

Project Website and additional documentation http://code.google.com/p/gbif-spreadsheet-processor/
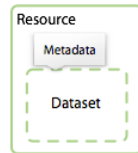
free and open access to biodiversity data

GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)

GBIF

Herramienta integrada de publicación (IPT)

- Servidor(es)
  - IPT(s)
    - Recursos(s)

Herramienta integrada de publicación (IPT)

Herramienta integrada de publicación (IPT)

Home	**Manage Resources**	Administration	About

MVZ Egg/Nests

This is the overview page for the *MVZ Egg/Nests* resource. Please start by filling in at least the mandatory metadata. Then you can upload and map source data to generate and publish a Darwin Core archive. Published resources can be registered with the GBIF Network. Registered resources can be removed, but they cannot be reverted to a private state.

Metadata

[Edit] ⬅

MVZ Egg and Nest specimens.

Source Data

[Choose File] No file chosen

[Clear] [Connect to DB]

Your data sources for generating a Darwin Core Archive. You can upload delimited text files (csv, tab, and files using any other delimiter) either directly or compressed (zip or gzip). To (re)upload a file, please select the local file then click "Add".

Alternatively, you can configure *SQL views* to databases in your local network. To create a new SQL source, please click "Connect to DB" without any file chosen.

mvzforvn [file]   2.4 MB, 5,255 rows, 39 columns. 16/06/11 [Edit]

Darwin Core Mappings

Darwin Core Occurrence ▾ [Add]

Your mapping between the source data and Darwin Core terms.

Darwin Core Occurrence   39 terms mapped to mvzforvn [Edit]

Visibility *Private*

[Public]

This resource is private to managers. To make this resource available to everyone, please make it public.

Published Release

[Publish]

When publishing a new release, a new EML version and a Darwin Core Archive (DwC-A) will be created. A DwC-A bundles all data sources with mappings and metadata in one zipped archive.

Last Publication   Version 1 from 16/06/11 Publication Log
Darwin Core Archive   download, 5,255 records
A new EML version has been published   download view
RTF   download

Basic Metadata: *MVZ Egg/Nests*

You must fill in at least these basic metadata before you can make this resource public.

For each contact you must supply at least a last name, a position, or an organisation. Title and Description are required.

Title

MVZ Egg/Nests

Description

MVZ Egg and Nest specimens.

Metadata Language

English

Resource Language

English

Subtype

checklist resource

Resource Contact

First Name

Carla

Last Name

Cicero

Position

Organisation

Address

City

Berkeley

Home	**Manage Resources**	Administration	About

MVZ Egg/Nests

This is the overview page for the *MVZ Egg/Nests* resource. Please start by filling in at least the mandatory metadata. Then you can upload and map source data to generate and publish a Darwin Core archive. Published resources can be registered with the GBIF Network. Registered resources can be removed, but they cannot be reverted to a private state.

Metadata

[Edit]

MVZ Egg and Nest specimens.

Source Data

[Choose File] No file chosen

[Clear] [Connect to DB]

Your data sources for generating a Darwin Core Archive. You can upload delimited text files (csv, tab, and files using any other delimiter) either directly or compressed (zip or gzip). To (re)upload a file, please select the local file then click "Add".

Alternatively, you can configure *SQL views* to databases in your local network. To create a new SQL source, please click "Connect to DB" without any file chosen.

mvzforvn [file]   2.4 MB, 5,255 rows, 39 columns. 16/06/11 [Edit] ⬅

Darwin Core Mappings

[Darwin Core Occurrence ▼] [Add]

Your mapping between the source data and Darwin Core terms.

Darwin Core Occurrence   39 terms mapped to mvzforvn [Edit]

Visibility *Private*

[Public]

This resource is private to managers. To make this resource available to everyone, please make it public.

Published Release

[Publish]

When publishing a new release, a new EML version and a Darwin Core Archive (DwC-A) will be created. A DwC-A bundles all data sources with mappings and metadata in one zipped archive.

Last Publication   Version 1 from 16/06/11 Publication Log
Darwin Core Archive   download, 5,255 records
A new EML version has been published download view
RTF   download

Home	**Manage Resources**	Administration	About

MVZ Egg/Nests

This is the overview page for the *MVZ Egg/Nests* resource. Please start by filling in at least the mandatory metadata. Then you can upload and map source data to generate and publish a Darwin Core archive. Published resources can be registered with the GBIF Network. Registered resources can be removed, but they cannot be reverted to a private state.

Metadata

[Edit]

MVZ Egg and Nest specimens.

Source Data

[Choose File] No file chosen

[Clear] [Connect to DB]

Your data sources for generating a Darwin Core Archive. You can upload delimited text files (csv, tab, and files using any other delimiter) either directly or compressed (zip or gzip). To (re)upload a file, please select the local file then click "Add".

Alternatively, you can configure *SQL views* to databases in your local network. To create a new SQL source, please click "Connect to DB" without any file chosen.

mvzforvn [file]   2.4 MB, 5,255 rows, 39 columns. 16/06/11   [Edit]

Darwin Core Mappings

Darwin Core Occurrence ▾   [Add]

Your mapping between the source data and Darwin Core terms.

Darwin Core Occurrence   39 terms mapped to mvzforvn   [Edit] ⬅

Visibility *Private*

[Public]

This resource is private to managers. To make this resource available to everyone, please make it public.

Published Release

[Publish]

When publishing a new release, a new EML version and a Darwin Core Archive (DwC-A) will be created. A DwC-A bundles all data sources with mappings and metadata in one zipped archive.

Last Publication   Version 1 from 16/06/11 Publication Log
Darwin Core Archive   download, 5,255 records
A new EML version has been published download view
RTF   download

free and open access to biodiversity data

GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)

Darwin Core Occurrence

The category of information pertaining to evidence of an occurrence in nature, in a collection, or in a dataset (specimen, observation, etc.).

Link: http://rs.tdwg.org/dwc/terms/index.htm#Occurrence

back to resource overview page

Mapping Source mvzforvn

For each property you can assign a dynamic value taken from the specified column of your data source mvzforvn or provide a fixed, static value that is used for all records.

Hide Unmapped Fields

occurrenceID	ⓘ	occurrenceID ▾	

The column in your source that holds a unique key for every record/row, also known as the "core id". If you want to map other sources, this identifier is required and will be used to link your source rows together.

Source Sample : MVZ:Egg:1219 | MVZ:Egg:1220 | MVZ:Egg:1222 | MVZ:Egg:1221 | MVZ:Egg:1223

Filter	ⓘ	▾	▾

You can define a filter to exclude records from the generated archive. Specify the criterion that matches the records to be included.

[Save] [Delete] [Cancel]

Record Level   Occurrence   Event   Location   GeologicalContext   Identification   Taxon

Record Level

type	ⓘ 📖	▾	Physical Object ▾
modified	ⓘ	▾	

GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)

GBIF

Darwin Core Occurrence

The category of information pertaining to evidence of an occurrence in nature, in a collection, or in a dataset (specimen, observation, etc.).

Link: http://rs.tdwg.org/dwc/terms/index.htm#Occurrence

back to resource overview page

Mapping Source mvzforvn

For each property you can assign a dynamic value taken from the specified column of your data source mvzforvn or provide a fixed, static value that is used for all records.

Hide Unmapped Fields

occurrenceID	ⓘ	occurrenceID ▾	

The column in your source that holds a unique key for every record/row, also known as the "core id". If you want to map other sources, this identifier is required and will be used to link your source rows together.

Source Sample : MVZ:Egg:1219 | MVZ:Egg:1220 | MVZ:Egg:1222 | MVZ:Egg:1221 | MVZ:Egg:1223

Filter	⟵⟶ ⓘ	▾	▾

You can define a filter to exclude records from the generated archive. Specify the criterion that matches the records to be included.

[Save] [Delete] [Cancel]

Record Level   Occurrence   Event   Location   GeologicalContext   Identification   Taxon

Record Level

type	ⓘ 📖	▾	Physical Object ▾
modified	ⓘ	▾	

Darwin Core Occurrence

The category of information pertaining to evidence of an occurrence in nature, in a collection, or in a dataset (specimen, observation, etc.).

Link: http://rs.tdwg.org/dwc/terms/index.htm#Occurrence

back to resource overview page

Mapping Source mvzforvn

For each property you can assign a dynamic value taken from the specified column of your data source mvzforvn or provide a fixed, static value that is used for all records.

Hide Unmapped Fields

occurrenceID	ⓘ	occurrenceID ▼	

The column in your source that holds a unique key for every record/row, also known as the "core id". If you want to map other sources, this identifier is required and will be used to link your source rows together.

Source Sample : MVZ:Egg:1219 | MVZ:Egg:1220 | MVZ:Egg:1222 | MVZ:Egg:1221 | MVZ:Egg:1223

Filter	ⓘ	▼	▼

You can define a filter to exclude records from the generated archive. Specify the criterion that matches the records to be included.

[Save] [Delete] [Cancel]

Record Level   Occurrence   Event   Location   GeologicalContext   Identification   Taxon

Record Level

type	ⓘ 📖	▼	Physical Object ▼

modified	ⓘ	▼	

GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)

GBIF

Home	Manage Resources	Administration	About

MVZ Egg/Nests

This is the overview page for the *MVZ Egg/Nests* resource. Please start by filling in at least the mandatory metadata. Then you can upload and map source data to generate and publish a Darwin Core archive. Published resources can be registered with the GBIF Network. Registered resources can be removed, but they cannot be reverted to a private state.

Metadata

[Edit]

MVZ Egg and Nest specimens.

Source Data

[Choose File] No file chosen

[Clear] [Connect to DB]

Your data sources for generating a Darwin Core Archive. You can upload delimited text files (csv, tab, and files using any other delimiter) either directly or compressed (zip or gzip). To (re)upload a file, please select the local file then click "Add".

Alternatively, you can configure *SQL views* to databases in your local network. To create a new SQL source, please click "Connect to DB" without any file chosen.

mvzforvn [file]   2.4 MB, 5,255 rows, 39 columns. 16/06/11 [Edit]

Darwin Core Mappings

Darwin Core Occurrence ▼ [Add]

Your mapping between the source data and Darwin Core terms.

Darwin Core Occurrence   39 terms mapped to mvzforvn [Edit]

Visibility *Private*

[Public] ⬅

This resource is private to managers. To make this resource available to everyone, please make it public.

Published Release

[Publish]

When publishing a new release, a new EML version and a Darwin Core Archive (DwC-A) will be created. A DwC-A bundles all data sources with mappings and metadata in one zipped archive.

Last Publication   Version 1 from 16/06/11 Publication Log
Darwin Core Archive   download, 5,255 records
A new EML version has been published   download view
RTF   download

GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)

| Home | **Manage Resources** | Administration | About |

MVZ Egg/Nests

This is the overview page for the *MVZ Egg/Nests* resource. Please start by filling in at least the mandatory metadata. Then you can upload and map source data to generate and publish a Darwin Core archive. Published resources can be registered with the GBIF Network. Registered resources can be removed, but they cannot be reverted to a private state.

Metadata

[Edit]

MVZ Egg and Nest specimens.

Source Data

[Choose File] No file chosen

[Clear] [Connect to DB]

Your data sources for generating a Darwin Core Archive. You can upload delimited text files (csv, tab, and files using any other delimiter) either directly or compressed (zip or gzip). To (re)upload a file, please select the local file then click "Add".

Alternatively, you can configure *SQL views* to databases in your local network. To create a new SQL source, please click "Connect to DB" without any file chosen.

mvzforvn [file]   2.4 MB, 5,255 rows, 39 columns. 16/06/11 [Edit]

Darwin Core Mappings

[Darwin Core Occurrence ▼] [Add]

Your mapping between the source data and Darwin Core terms.

Darwin Core Occurrence   39 terms mapped to mvzforvn [Edit]

Visibility *Private*

[Public]

This resource is private to managers. To make this resource available to everyone, please make it public.

Published Release

[Publish] ⬅

When publishing a new release, a new EML version and a Darwin Core Archive (DwC-A) will be created. A DwC-A bundles all data sources with mappings and metadata in one zipped archive.

Last Publication   Version 1 from 16/06/11 Publication Log
Darwin Core Archive   download, 5,255 records
A new EML version has been published   download view
RTF   download

Recurso de datos publicado

✅ RTF published successfully!
Publishing resource version 1.

My Occurrence Resource

Publishing Status

May 28, 2011 5:49:50 PM

✅ Archive generated!

Continue to resource overview.

Download the complete Publication Log

Log Message
Archive generation started for resource myoccurrenceresource 5:49:49 PM
Start writing data file for Darwin Core Occurrence 5:49:49 PM
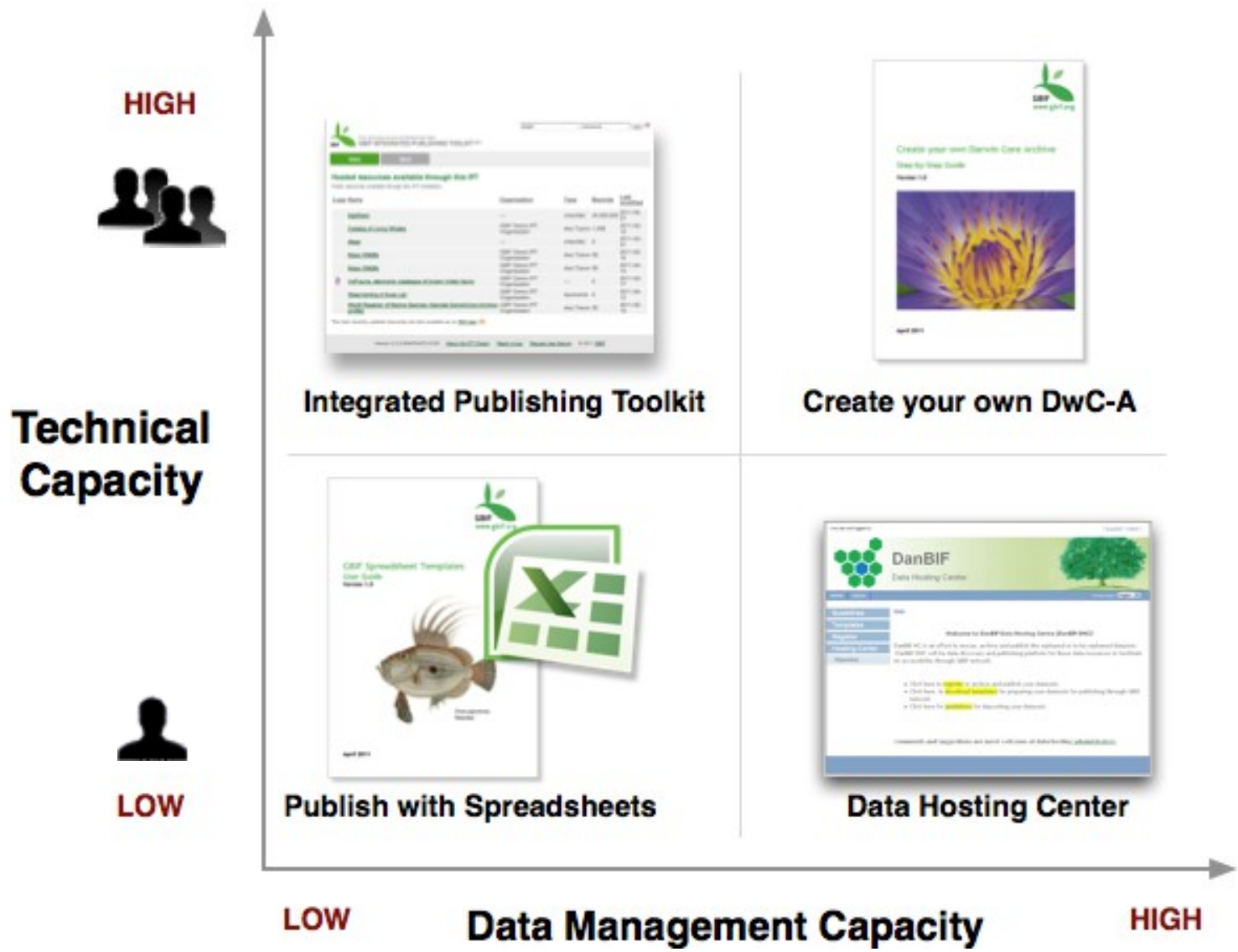Data file written for Darwin Core Occurrence with 12 records and 23 columns 5:49:49 PM
All data files completed 5:49:49 PM
EML file added 5:49:49 PM
meta.xml archive descriptor written 5:49:49 PM
Archive compressed 5:49:49 PM
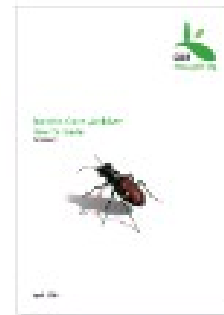Archive generated successfully! 5:49:49 PM

Technical Capacity — HIGH / LOW (vertical axis)

Data Management Capacity — LOW / HIGH (horizontal axis)

Integrated Publishing Toolkit

Create your own DwC-A

Publish with Spreadsheets

Data Hosting Center

Documentación

GBIF Metadata
Profile, How-to guide

- http://www.pliniancore.org

- http://www.tdwg.org/activities/darwincore/

- https://code.google.com/p/pliniancore/

- http://knb.ecoinformatics.org/software/eml/

GBIF Darwin Core
Archive, How-to
Guide