

DARWIN CORE

Massata NDAO



OceanTeacher
GLOBAL ACADEMY



ISRA, Dakar

Darwin Core

DwC, C'EST QUOI ?

Darwin Core (abrégé en **DwC**) est une extension de **Dublin Core for Biodiversity Informatics**.

DwC est un ensemble de standards (normes) pour le partage des données de biodiversité.

Il a été développé par le Groupe International de Travail sur les Bases de Données Taxonomiques (**Taxonomic Databases Working Group : TDWG**) et permet aux propriétaires des données de publier des informations dans un format bien spécifique (compris et utilisé par tous !).



Darwin Core

DwC, C'EST QUOI ?

Le **DwC** comprend un glossaire de termes (champs, attributs, concepts, ...) destinés à faciliter le partage d'informations sur la diversité biologique.

Le **DwC** repose principalement sur des taxons, leur occurrence dans la nature, documentés par des observations, des échantillons, des spécimens, des informations connexes,

Darwin Core

QUI EN EST RESPONSABLE ?

TDWG : Taxonomic Databases Working Group

Le Groupe de travail sur les bases de données taxonomiques

Association scientifique et éducative sans but lucratif affiliée à l'Union internationale des sciences biologiques.



Darwin Core

MISSIONS du TDWG ?

TDWG : Taxonomic Databases Working Group

- ❖ Élaborer, d'adopter et promouvoir des normes et des lignes directrices pour l'enregistrement et l'interopérabilité des données sur les organismes.
- ❖ Promouvoir l'utilisation des normes par les moyens les plus appropriés et les plus efficaces.
- ❖ Agir en tant que forum de discussion (à travers des rencontres et des publications).



OceanTeacher
GLOBAL ACADEMY



Darwin Core

Principaux Projets utilisant le Darwin Core

- ❑ The Global Biodiversity Information Facility (GBIF)
- ❑ The Ocean Biogeographic Information System (OBIS)
- ❑ Mammal Networked Information System (MaNIS)
- ❑ Ornithological Information System (ORNIS)
- ❑ FishNet 2
- ❑ Sistema Nature 3.0
- ❑ Encyclopedia of Life

Darwin Core

AUDITOIRE ?

- Les éditeurs des données sur le biodiversité (organisations, institutions, chercheurs).
- Les utilisateurs d'informations sur la biodiversité
- Les systèmes de gestion des collections biologiques
- Les autres groupes de travail TDWG
- Les développeurs, les administrateurs de réseau et autres utilisateurs de la biodiversité



Darwin Core

QUE PEUT-ON PUBLIER ?

- Données d'occurrence
- Catalogue / Listes d'espèce
- Données d'échantillonnage
- Métadonnées



Darwin Core

Type de jeu de donnée : occurrence

Texte numérisé ou données multimédia détaillant des faits sur l'occurrence d'un organisme : le **quoi, où, quand, comment et par qui** de l'occurrence et de son enregistrement.

Darwin Core

Type de jeu de données : checklist / liste d'espèces



Checklists

“Un catalogue, ou une liste de taxons.”

Éventuellement aussi : noms vernaculaires, citations, ...

Classent généralement l'information par taxonomie, géographie et statut ou en combinant les trois.

Darwin Core

Type de jeu de données : données d'échantillonnage / Métadonnées

Données d'échantillonnage : Termes obligatoires (dans le fichier modèle)

- Code de l'Institut (ex : CRODT) : nom ou acronyme de l'institution ayant la garde de l'occurrence
- Numéro de Catalogue (ex : Crodt-CF-001) : identifiant (si possible unique) de chaque occurrence dans le jeu de données ou la collection
- Code collection (ex : Ichtyologie) : nom ou acronyme du jeu de données ou de la collection dont l'occurrence provient
- Nom scientifique (ex: *Arius sp.*) : nom scientifique complet avec paternité et date (si connus)
- Base d'enregistrement (ex : observation, spécimen, fossile, ...) : nature spécifique de l'enregistrement

Darwin Core

**Type de jeu de données : données d'échantillonnage /
Métadonnées**

Données d'échantillonnage : Termes importants

taxonomie, géographie (surtout latitude et longitude), conditions de prélèvement ou d'observation, contexte géologique, **date**, habitat, médias associés...

Possibilité d'ajouter des extensions (vocabulaire plus spécifique) :

- Taxonomie détaillée
- Ajout de noms vernaculaires
- Images
- ...

Pour plus d'informations (définition et traduction des termes Darwin Core) :

http://terms.tdwg.org/wiki/Darwin_Core

Darwin Core

Standards d'échange



Métadonnées



Données primaires



Données taxonomiques

**Ecological Metadata
Language (EML)**

- *Description riche des jeux de données*

Darwin Core

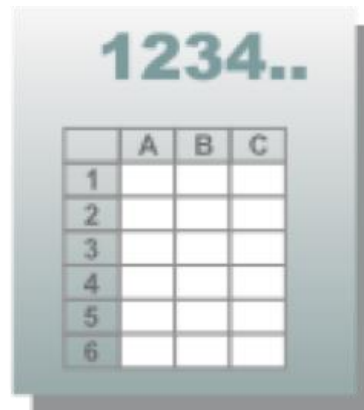
- *180 concepts*
- *Ratifié en 2009 (TDWG)*
- *Fichiers texte*
- *Extensible*

Darwin Core

Processus de publication des données

Simple Darwin Core :

Darwin Core exprimé sous forme de structure tabulaire simple

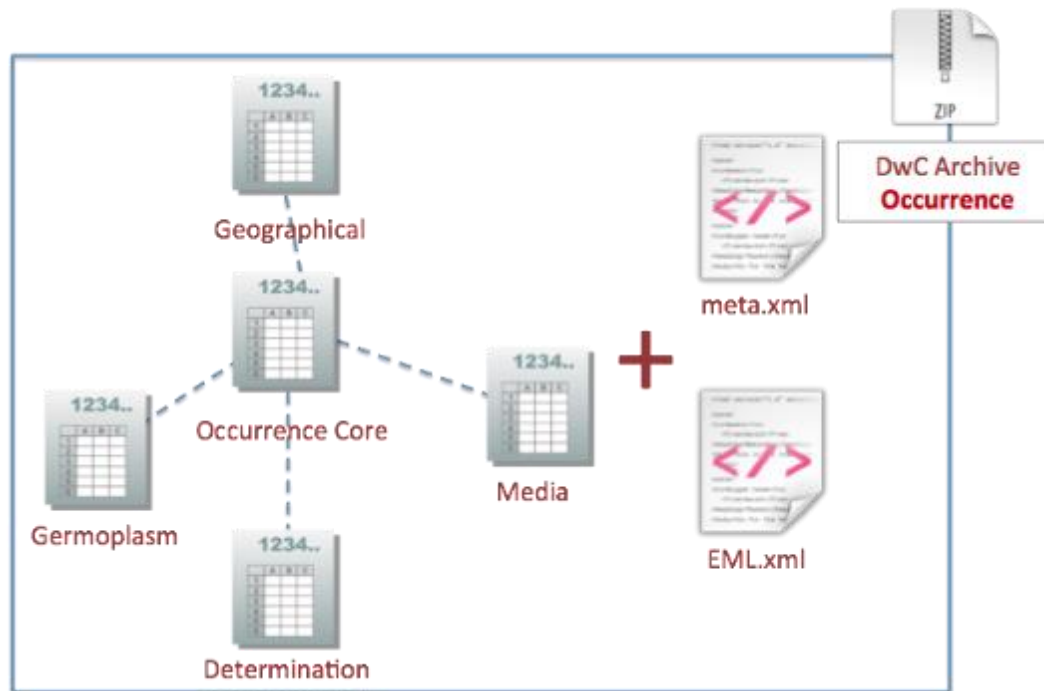


	A	B	C
1			
2			
3			
4			
5			
6			

Darwin Core

Processus de publication des données

Darwin Core Archive : un format plus avancé, qui permet l'usage d'extensions.



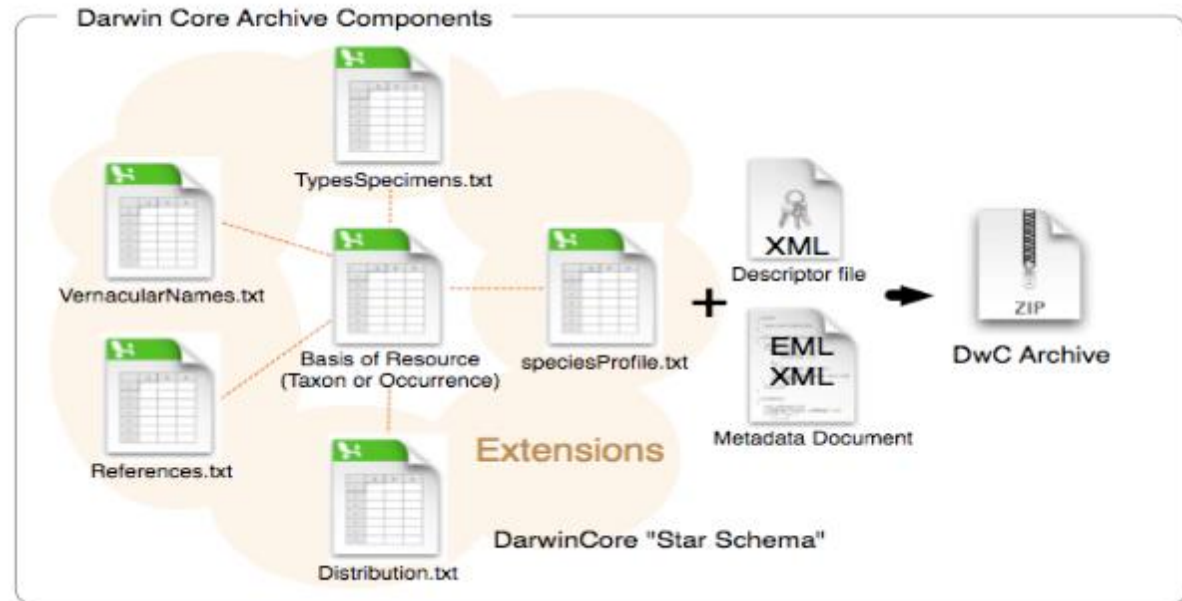
Darwin Core

Processus de publication des données

Darwin Core Archive

Contenu de l'archive

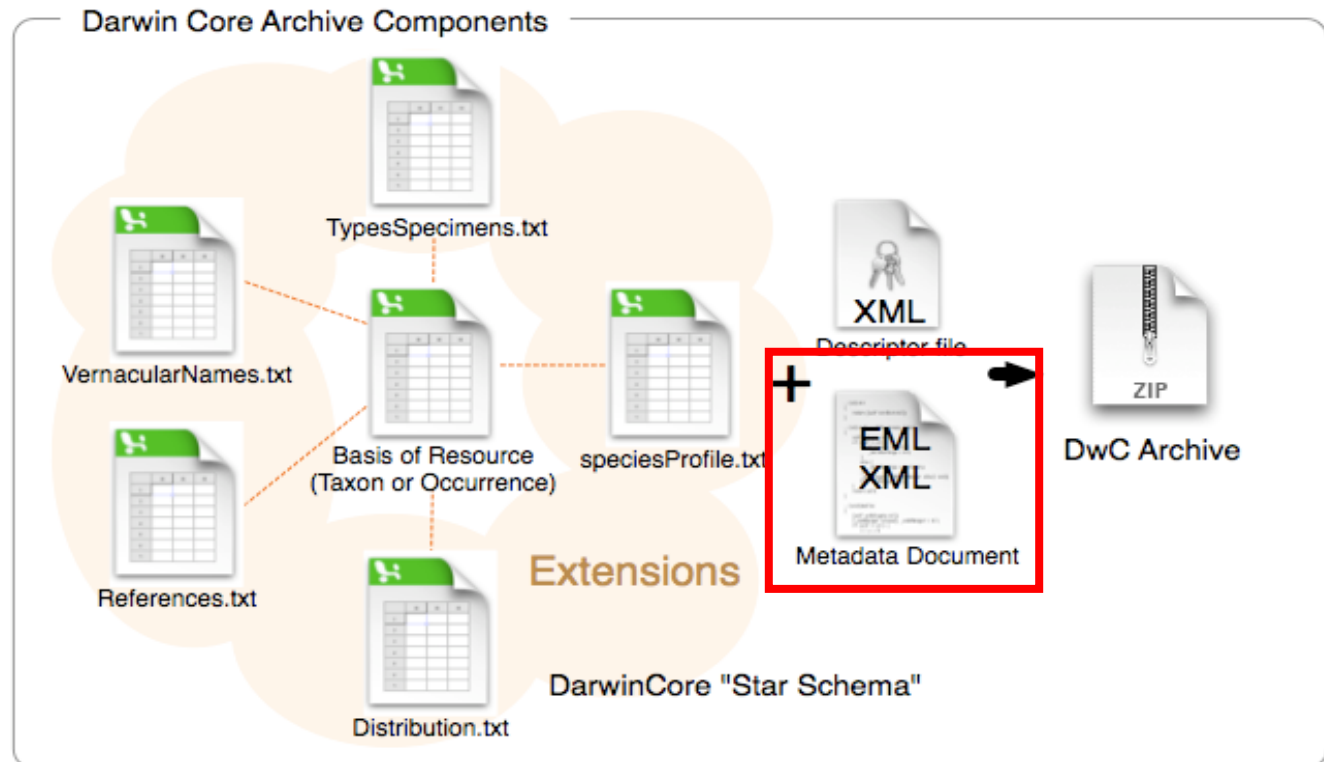
- Une archive correspond à un jeu de données
- Fichiers formatés par le standard Darwin Core
- Données d'occurrences ou taxonomiques



Darwin Core

Processus de publication des données

Les archives ont toujours un fichier de métadonnées EML



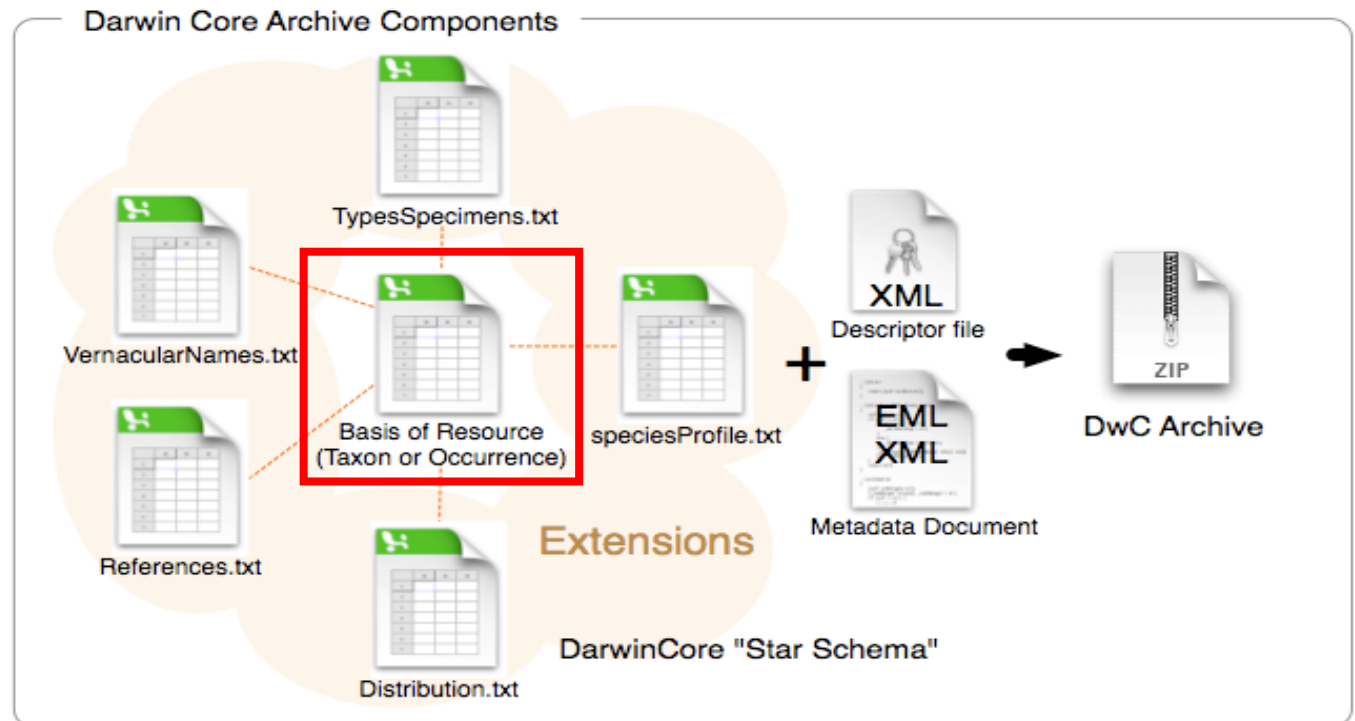
Darwin Core
Archive



Darwin Core

Processus de publication des données

Les archives ont toujours un fichier de données principal (Core Data File)

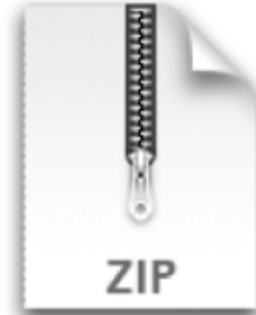
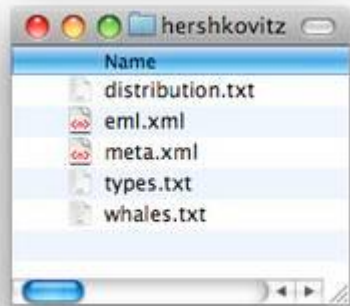


Darwin Core
Archive

Darwin Core

Processus de publication des données

Le dossier est zippé



Ceci est une archive
DarwinCore

- Fichiers de données (txt)
- Fichier de correspondance des colonnes (meta.xml)
- Fichier de métadonnées (eml.xml)

Darwin Core
Archive

Darwin Core

Processus de publication des données

Darwin Core Archive : avantages

- Format simple (Fichiers texte)
- Stockage efficace (compressé)
- Récolte efficace (un seul fichier)
- Simple d'accès (pas besoin de logiciel spécifique)
- Extensible

Darwin Core

Processus de publication des données

Méthode de publication : IPT (Outil de Publication intégrée)

- Logiciel serveur, nécessite une connexion stable
- Un IPT peut héberger plusieurs datasets, pour plusieurs institutions.
- Modes test et production
- Multilingue

Darwin Core

Pour installer l'IPT ?

- Version 2.3.2
- Serveur Tomcat
- Disponible sur www.gbif.org et Google Code