

GLOBAL  
BIODIVERSITY



INFORMATION  
FACILITY

# GBIF and Metadata - Current Status and Plans

Éamonn Ó Tuama

PIIB Metadata Workshop,  
Bogotá, 13 September, 2010



# Outline



- Why metadata?
- Current metadata handling in GBIF portal
- GBIF metadata plans – scoping the system
- GBIF Nodes Survey
- Promoting/incentivising metadata preparation
- Resources at GBIF
- GBIF work programme activities

# Why metadata?

”Fourth Science Paradigm”

”Data Intensive Science”

# ”Digital Data Deluge”

**Key requirement :**

high quality metadata for long term  
curation and use of datasets

e-Infrastructure Reflection Group  
Metadata Implementation Framework TG  
Data Publishing Framework TG  
Content Needs Assessment TG

*The Fourth Paradigm: Data-Intensive Scientific Discovery*  
<http://research.microsoft.com/en-us/collaboration/fourthparadigm/contents.aspx>

*e-Infrastructure Reflection Group (European Strategy Forum on Research Infrastructures). Report on Data Management, November 2009.*  
[http://www.e-irg.eu/images/stories/publ/task\\_force\\_reports/dmtfjointreport.pdf](http://www.e-irg.eu/images/stories/publ/task_force_reports/dmtfjointreport.pdf)

# Why metadata?

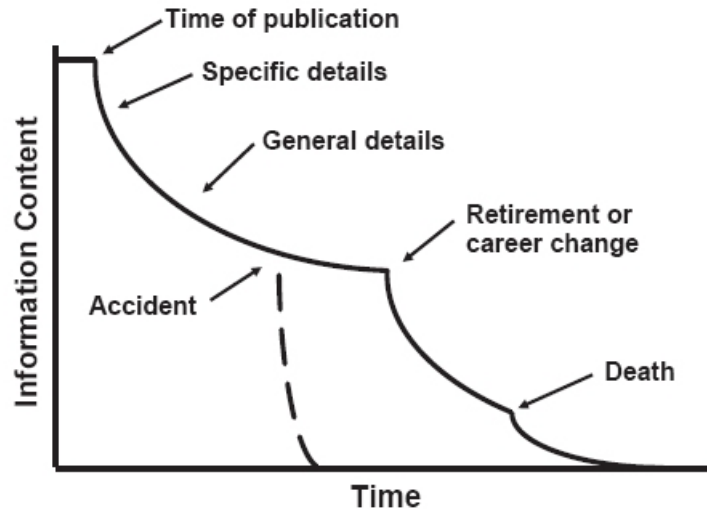


Fig. 1 – Illustration of the natural degradation in information content associated with data and metadata—information entropy (from Michener et al., 1987, by permission of the Ecological Society of America).

Metadata supports:

- Discovery
- Interpretation/Evaluation
  - Provenance
  - Quality
  - Fitness-for-use
- Analytical re-use

William K. Michener, Meta-information concepts for ecological data management, *Ecological Informatics*, Volume 1, Issue 1, January 2006, Pages 3-7, ISSN 1574-9541, DOI: 10.1016/j.ecoinf.2005.08.004.

(<http://www.sciencedirect.com/science/article/B7W63-4HJRS57-3/2/ea2e08412c6776456f540e66983546c0>)

# Historical situation at GBIF



Focus on museum specimens, later expanded to individual observation events

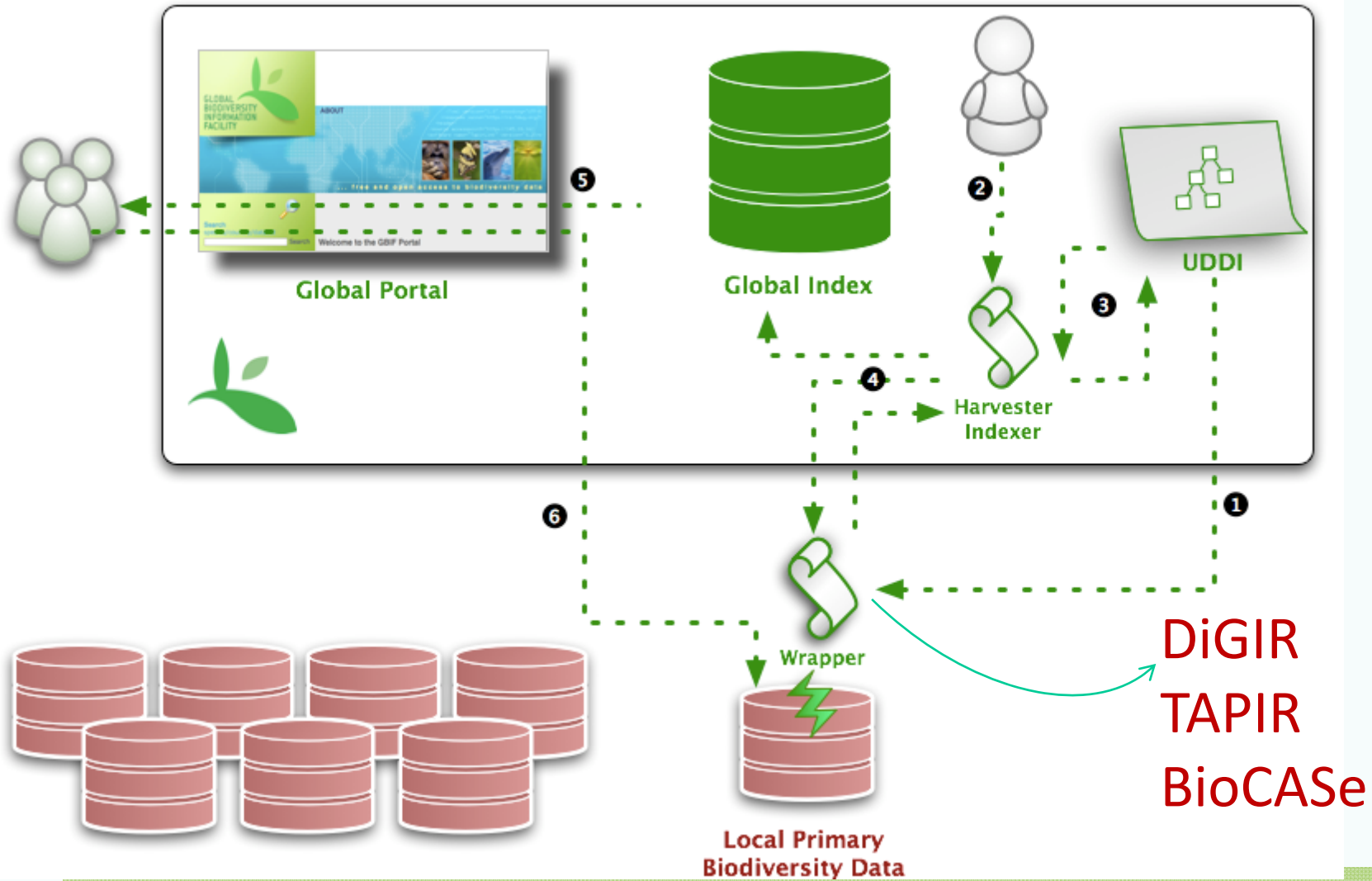
Focus on individual record vs dataset

Use of Darwin Core (includes some metadata descriptors, e.g., positional accuracy)

Only very basic metadata at dataset level

No standard catalogue/search interface available via GBIF portal

# Current GBIF Portal



# Current GBIF dataset metadata

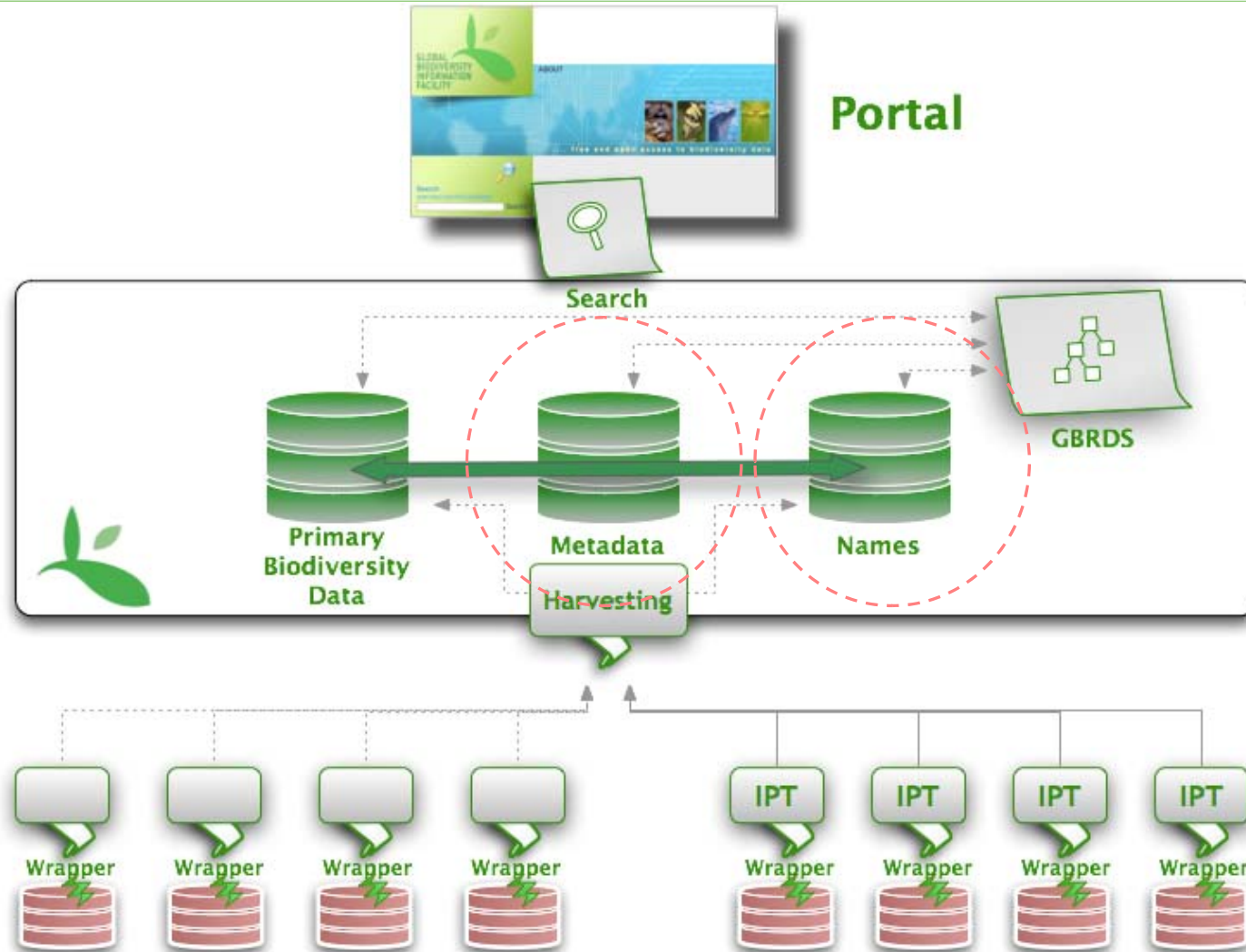
via the  
UDDI  
registry,  
DiGIR,  
TAPIR,  
BioCASE

- Data provider details (Name; Website; GBIF participant; Description; Country; Added to portal; Information updated)
- Provider (DiGIR, BioCASE, TAPIR) binding
- Name
- Website
- Description
- Citation
- How to cite this dataset
- Basis of record
- Access point URL
- Added to portal
- Information updated
- Contacts (Name, Role, Address, Email, Telephone)
- Data networks

and via the  
indexing  
process -

- Number of occurrence records indexed
- Number of records shared by provider
- Number of occurrences with coordinates
- Number of occurrences with no geospatial issues
- Number of species
- Number of taxa

# Expanding the GBIF Portal





# Service Oriented Architecture

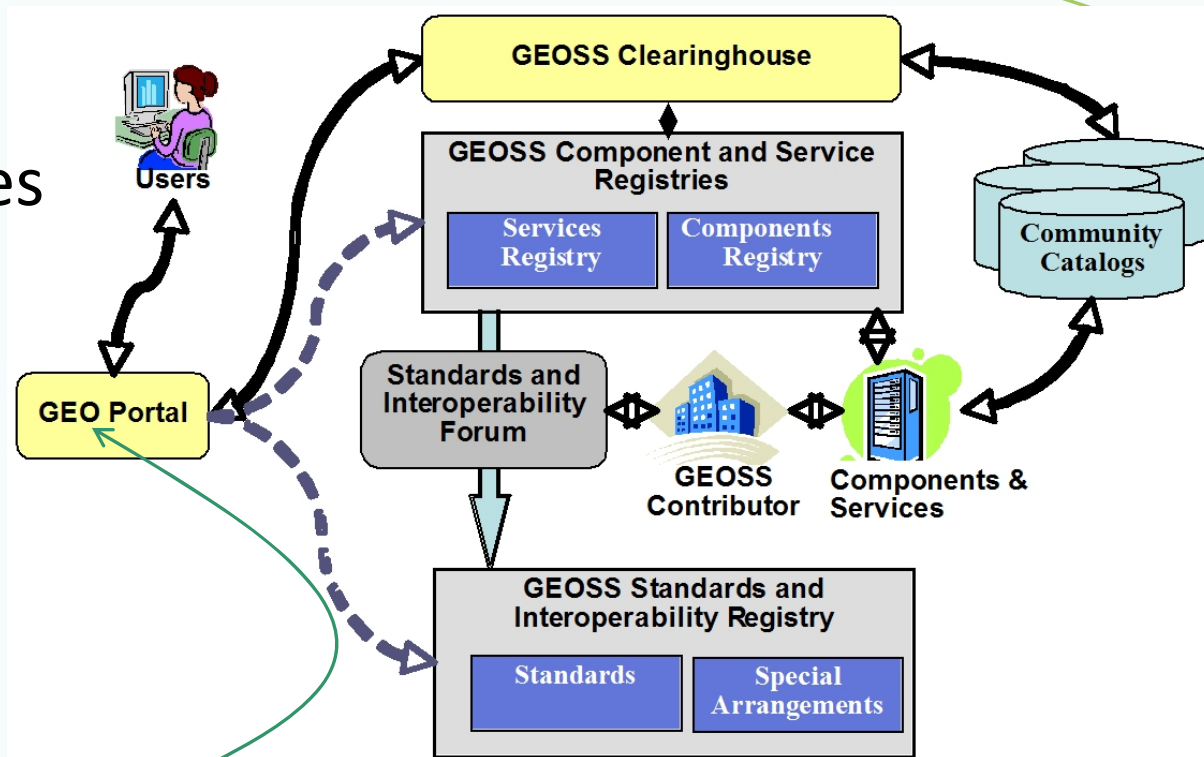


Key Features -

Loose coupling of services

Consumers discover services

Metadata and catalogue services are key components



Interactions of GEOSS Registries, Portal and Clearinghouse

**GBIF Role  
in GEO BON**

Extracted from: *GEOSS Core Architecture Implementation Report*  
([http://portal.opengeospatial.org/files/?artifact\\_id=24315](http://portal.opengeospatial.org/files/?artifact_id=24315))

# GBIF Metadata Plans



Provide metadata for all datasets (inc names) offered through data portal

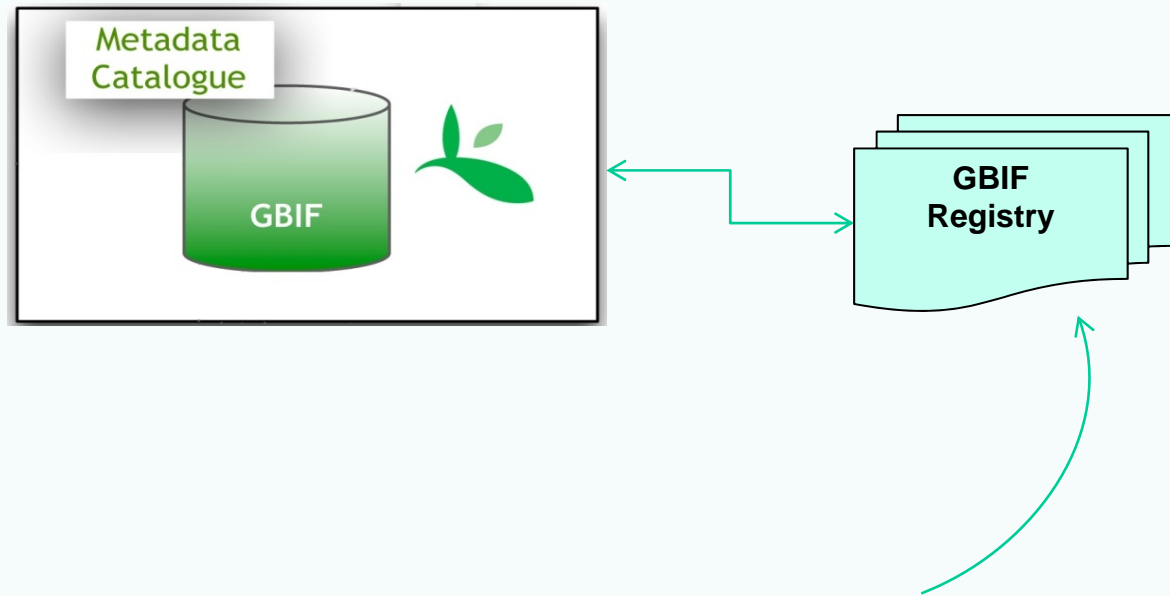
Provide metadata about datasets not yet available/digitised (inventory, priority setting)

Provide metadata catalogue search interface as component of GBIF portal

Share GBIF catalogue with other clearinghouse/catalogues, e.g., GEOSS

Share metadata with other networks, e.g., KNB (clearinghouse)

# GBIF Metadata Plans



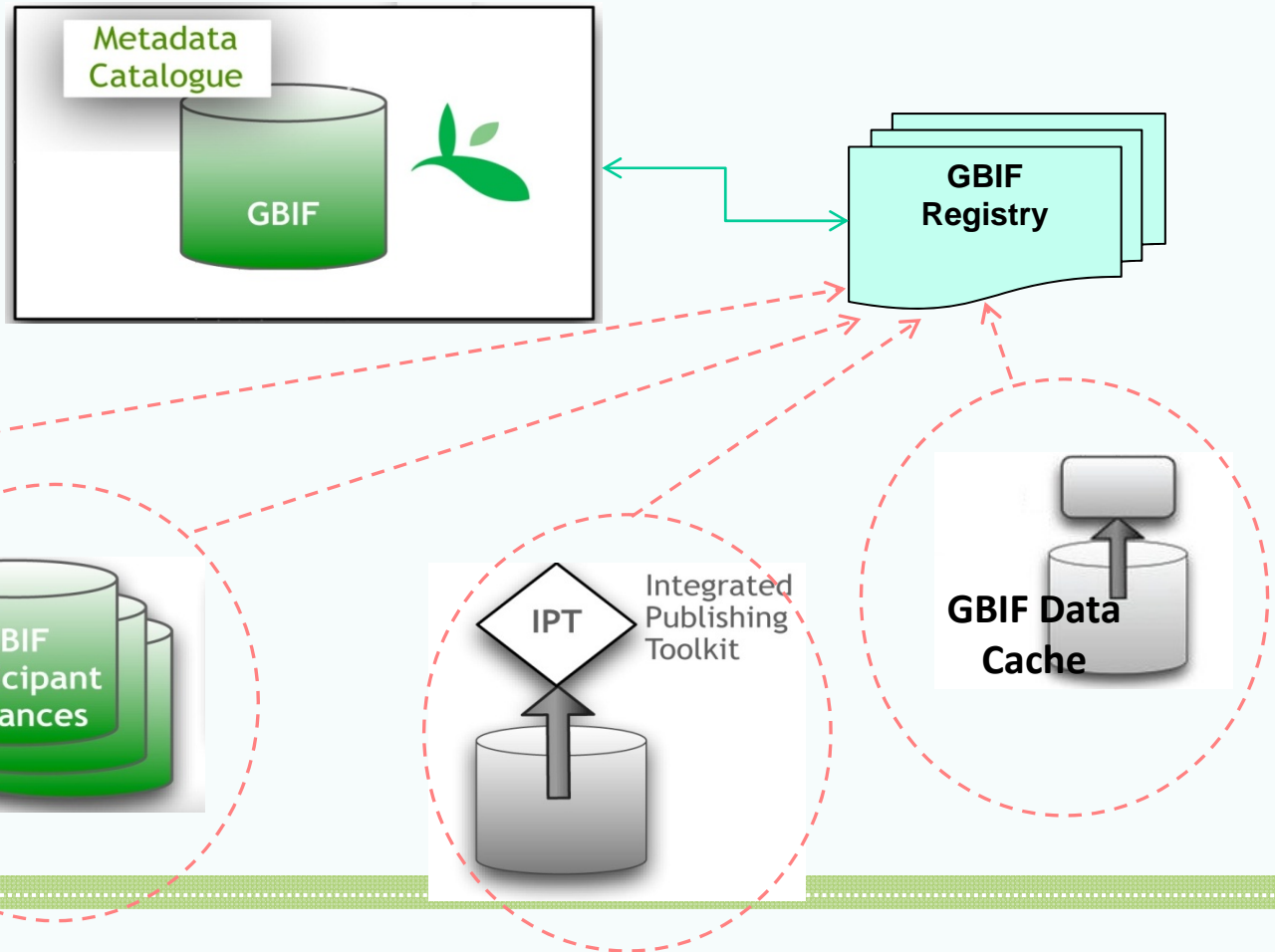
A registry is required to locate resources to put in the catalogue



# GBIF Metadata Plans



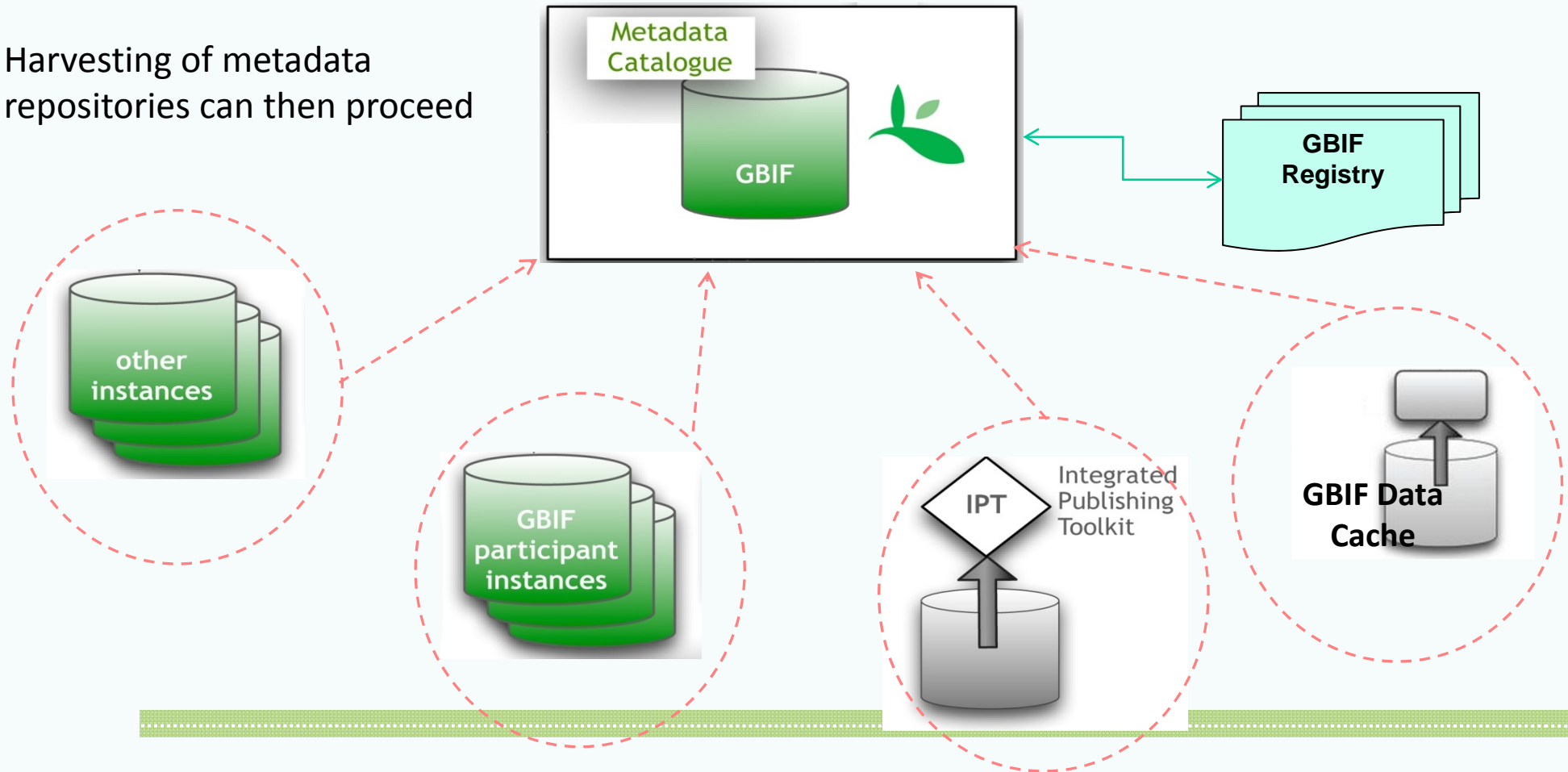
Metadata repositories are first registered



# GBIF Metadata Plans



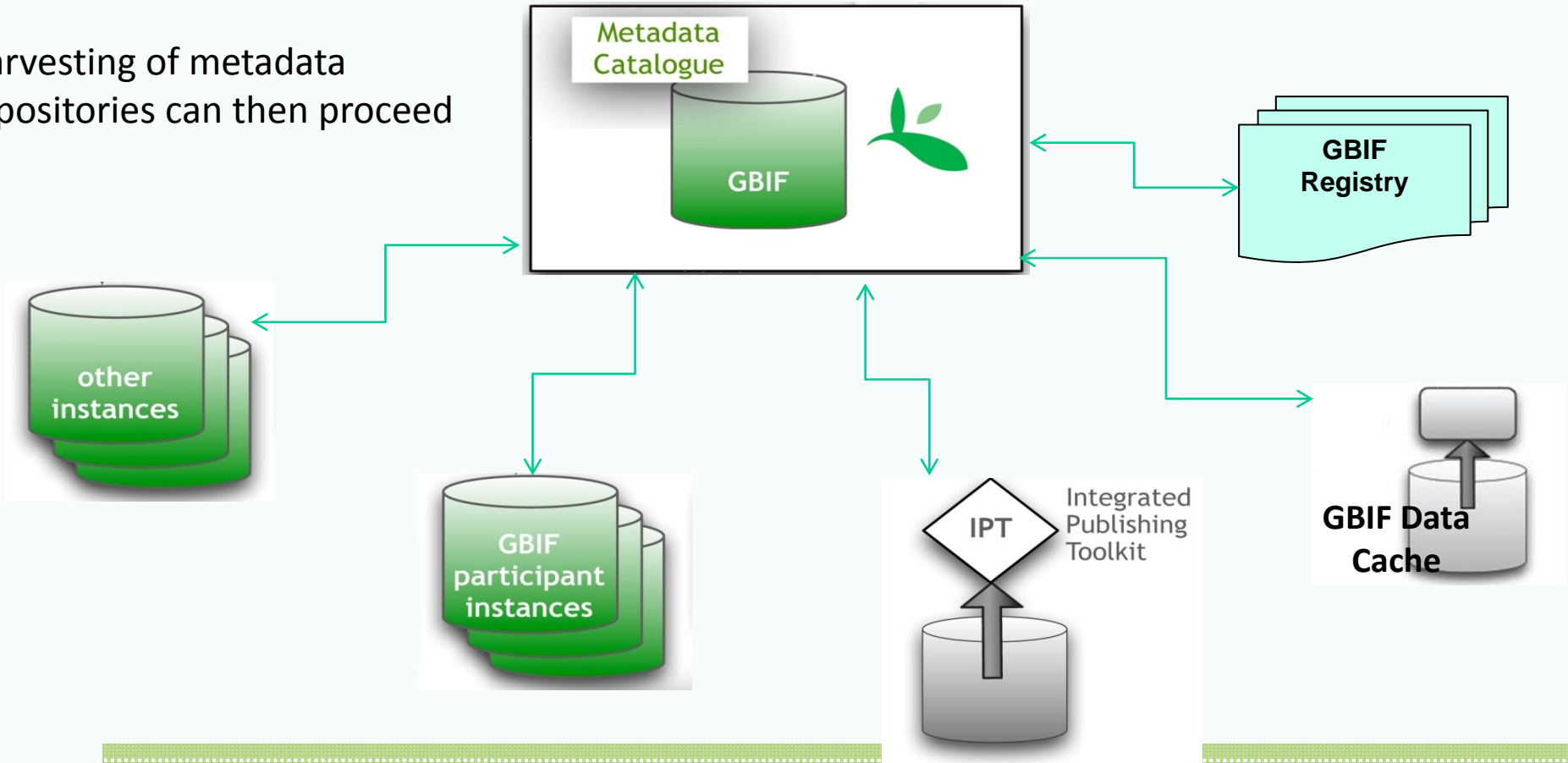
Harvesting of metadata repositories can then proceed



# GBIF Metadata Plans



Harvesting of metadata repositories can then proceed

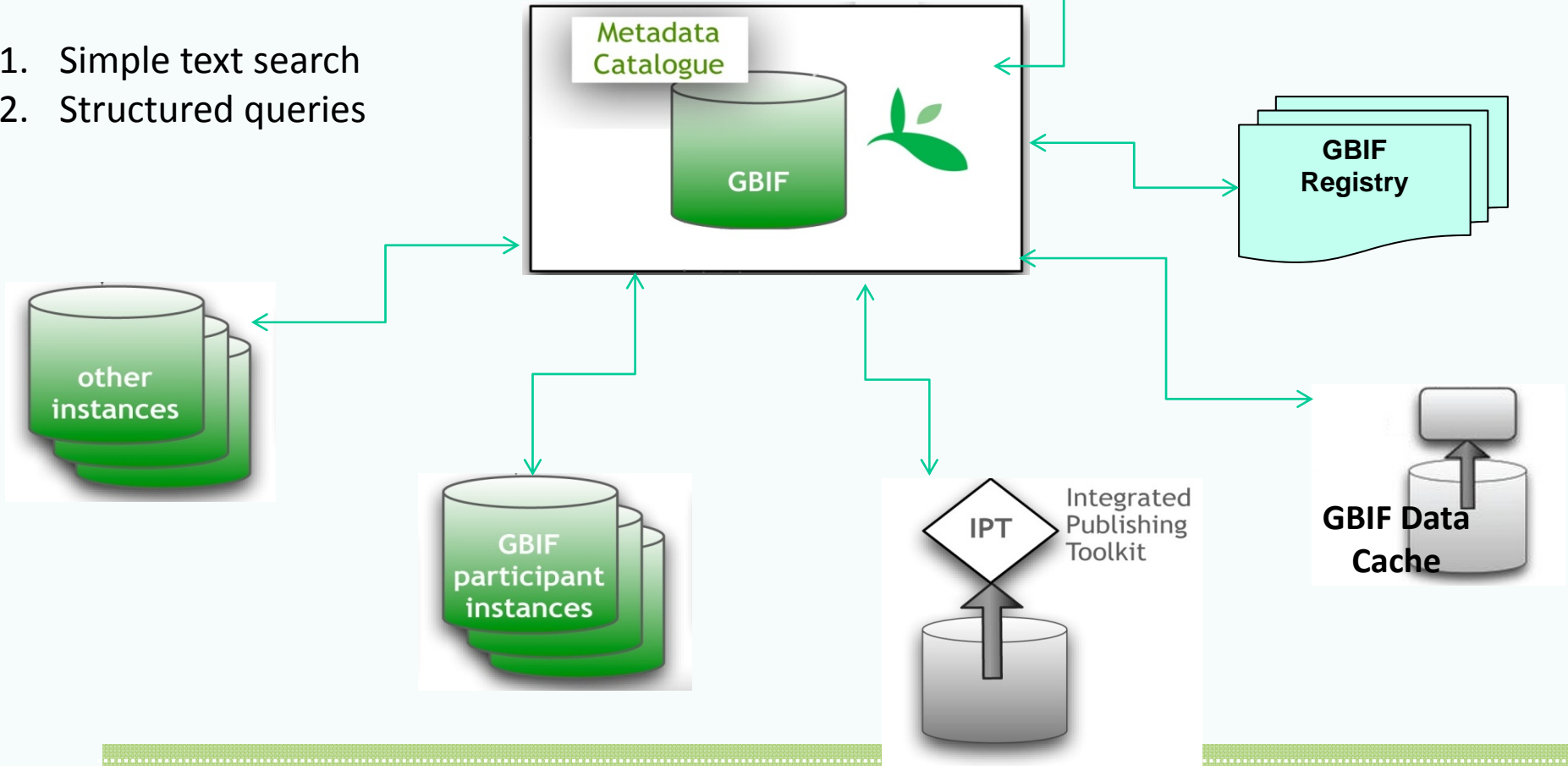


# GBIF Metadata Plans

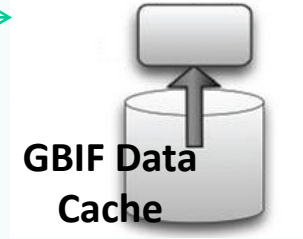
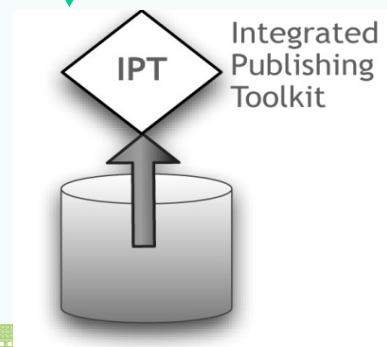
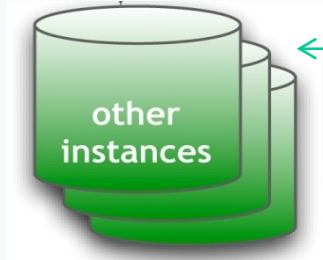
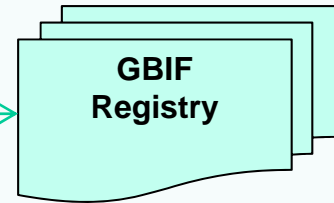
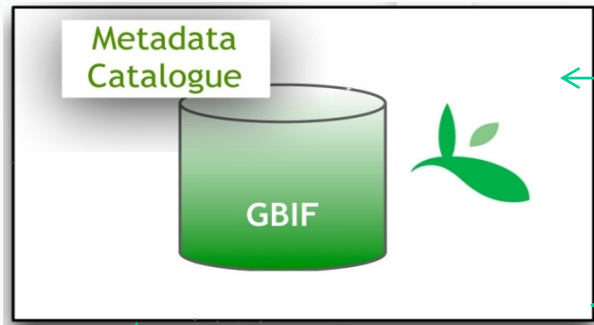
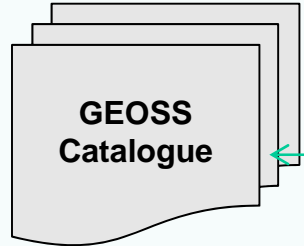


Entry point in GBIF Portal for searching/  
browsing the metadata catalogue

1. Simple text search
2. Structured queries



# GBIF Metadata Plans



*Development status?*

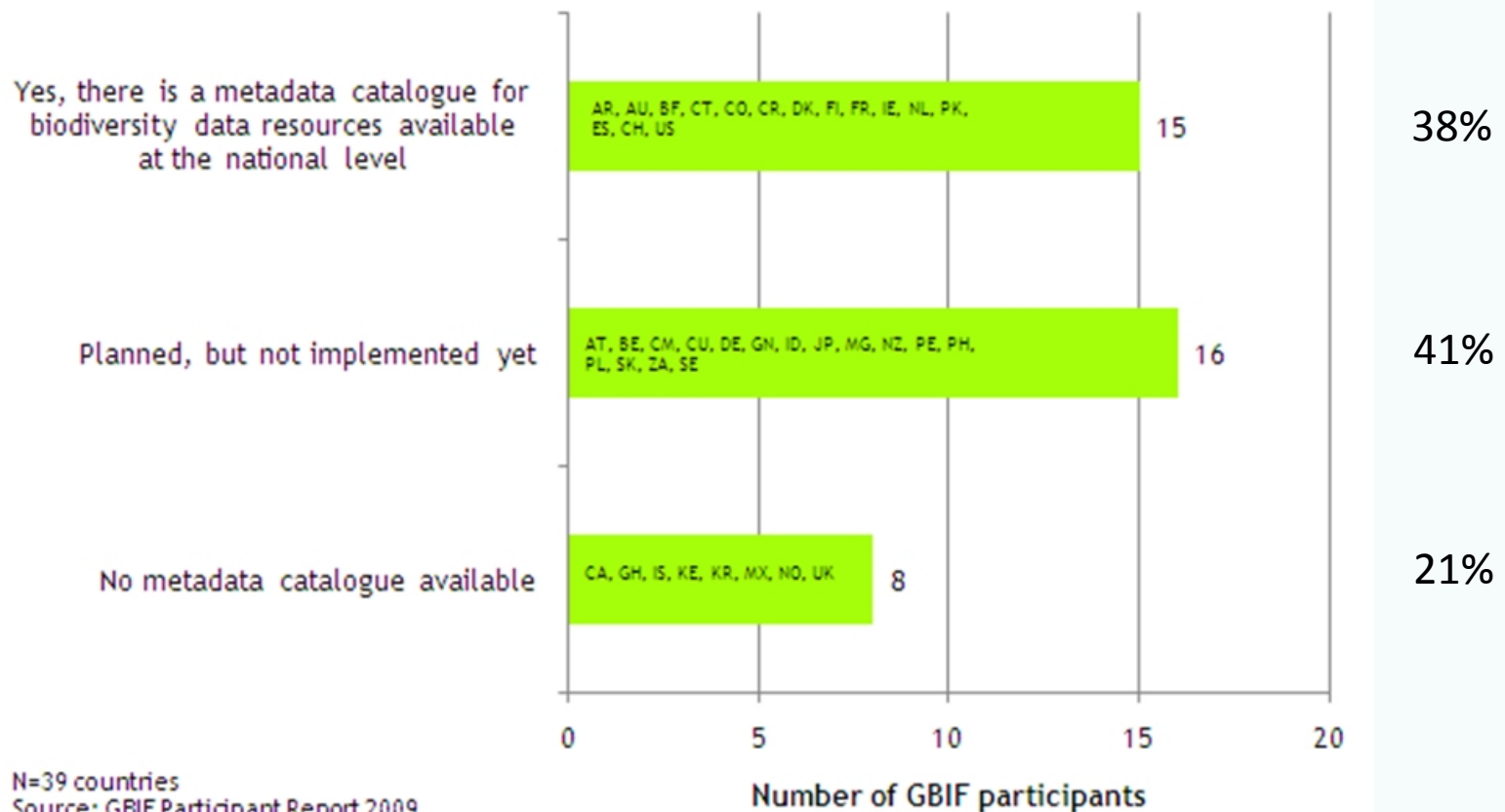




# GBIF Nodes Survey 2009



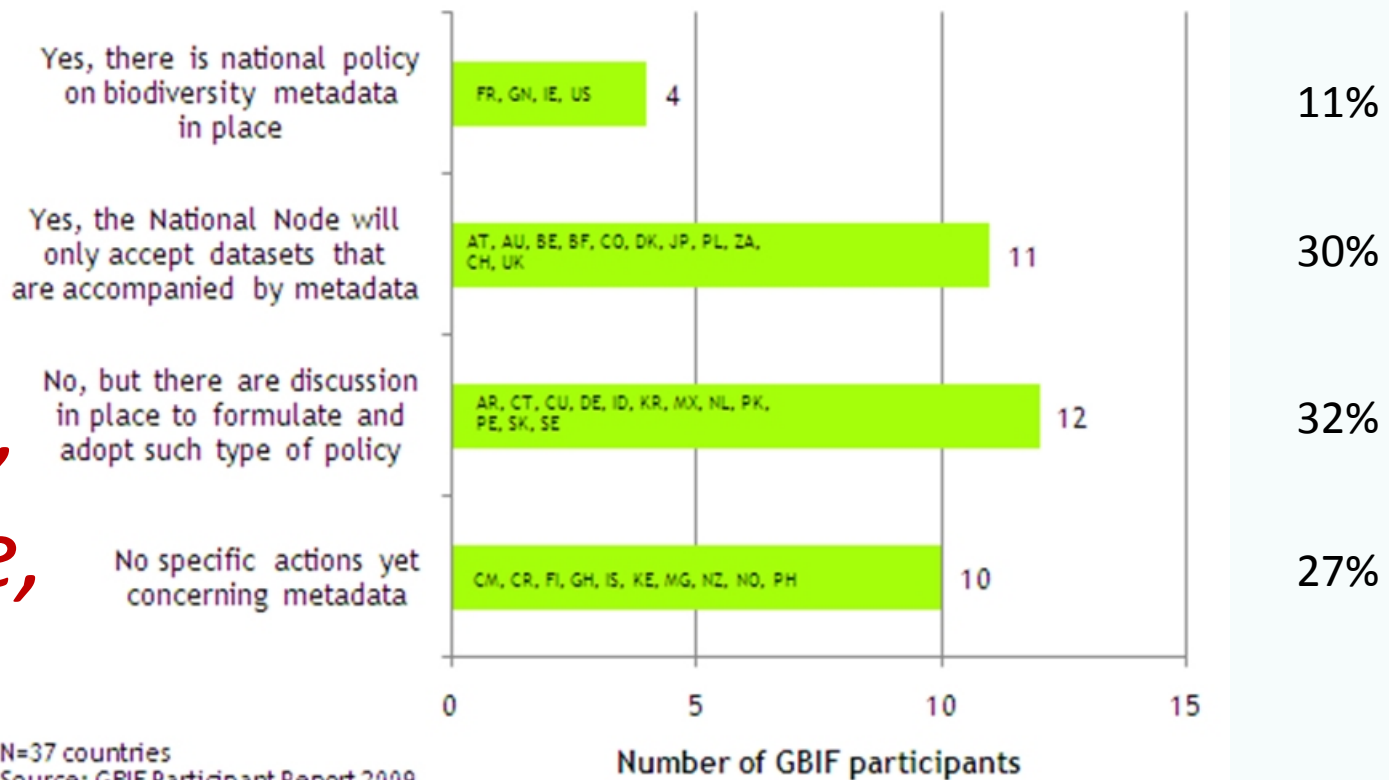
Does your National Node have a metadata catalogue for the biodiversity datasets and information resources within your country?



# GBIF Nodes Survey 2009

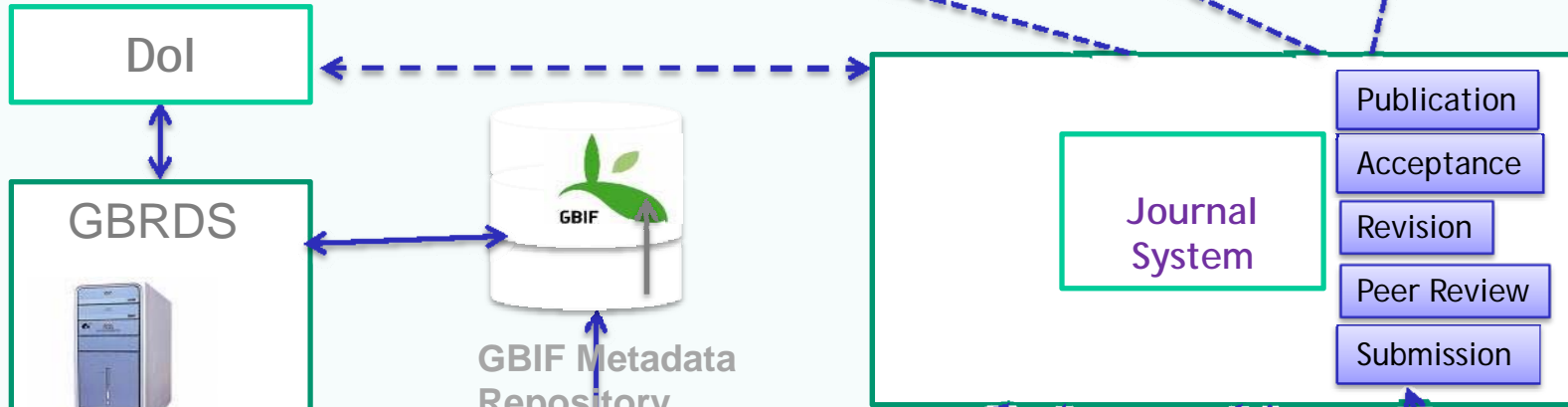
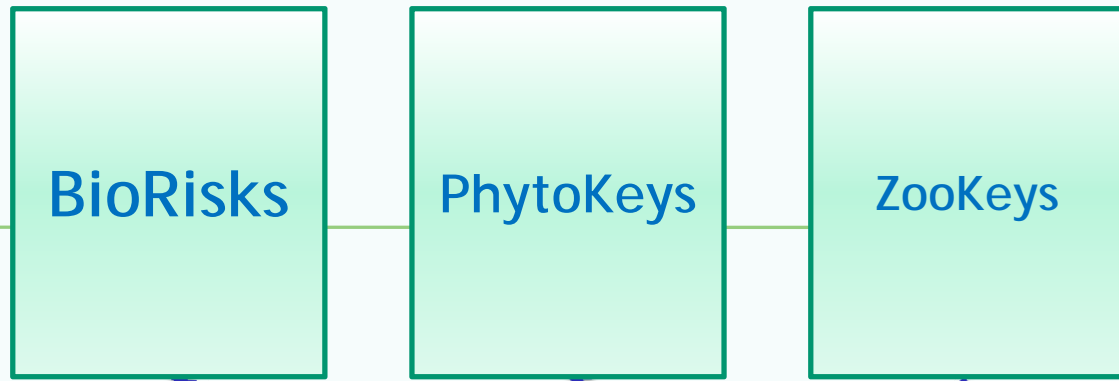


Is there a policy in place regarding the provision of metadata for biodiversity data?

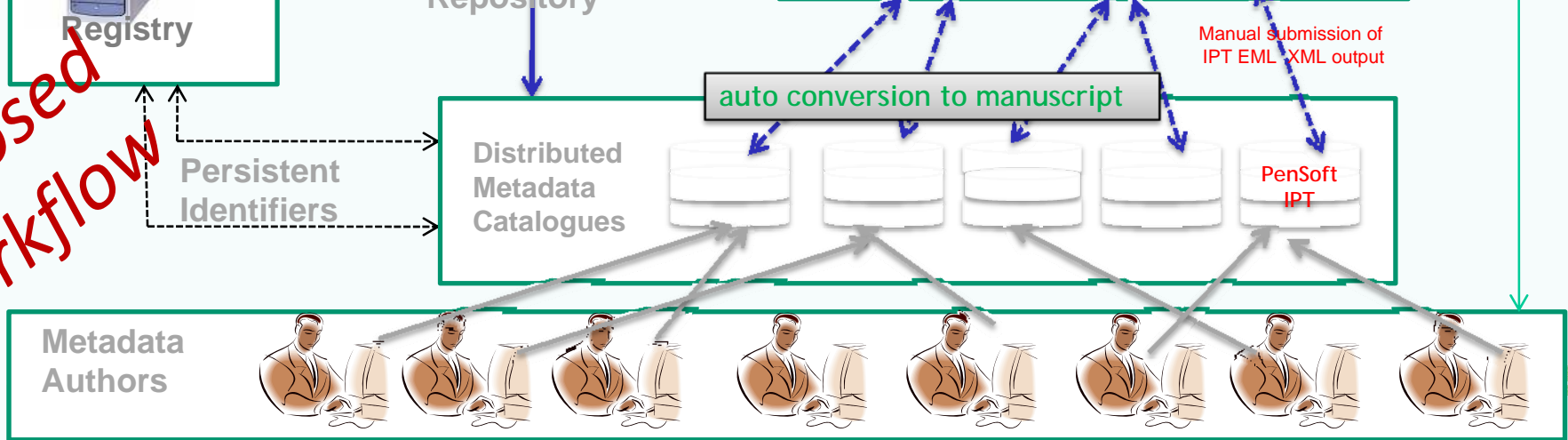


*Need to educate, promote, train !*

Data Paper as  
incentive to  
preparation of  
metadata



**Proposed  
workflow**



# Metadata Resources at GBIF



Metadata Strategy Document (2008)

*Metadata Requirements for Datasets delivered via the Global Biodiversity Information Facility (GBIF) Network*

[http://wiki.gbif.org/dadiwiki/wikka.php?wakka=FilesUpload/files.xml&action=download&file=GBIF-metadata-strategy\\_v.06.pdf](http://wiki.gbif.org/dadiwiki/wikka.php?wakka=FilesUpload/files.xml&action=download&file=GBIF-metadata-strategy_v.06.pdf)

Metadata Implementation Framework Task Group (MIFTG) Report

<http://www2.gbif.org/GBIF-MIFTG-Report.pdf>

The original wiki - used to collate documents and discussions relating to metadata standards:

<http://wiki.gbif.org/dadiwiki/wikka.php?wakka=CategoryMetadata>

The GBIF Google Code site - used to manage code and documents relating to the GBIF metadata system:

<http://code.google.com/p/gbif-metadata/> ; <http://rs.gbif.org/schema/eml/>

The GBIF Metadata Network community site - a forum for discussions

<http://community.gbif.org/pg/groups/5258/gbif-metadatanetwork/>

# Metadata Task Group (MIFTG) Report



Provides recommendations and guidelines on implementation of a metadata framework for the GBIF network, including:

- metadata specifications,
- controlled vocabularies,
- catalogue systems,
- network architecture

*“The Global Biodiversity Information Facility (GBIF) aspires to ... become a major provider of **discovery and access services** for a wide variety of biodiversity data types. A **distributed metadata catalog system** that describes and makes accessible general information on datasets of primary biodiversity data is recognised as an **essential component** of GBIF to achieve this objective.”*

# Metadata Framework Progress (2010)



- Secured EuroGEOSS funds (150K)
- Completed review of metadata catalogue systems ([GeoNetwork](#), [Metacat](#), [Mercury](#))
- Catalogue implemented Q4 2010 ([Metacat](#); [Solr/Lucene?](#))
- Metadata editing tools developed ([IPT metadata editor](#); [spreadsheet templates](#))
- GBIF EML metadata profile developed ([datasets](#), [checklists](#), [collections](#))
- 9 *GBIF Small Grants for Metadata Catalogues* issued ([Burkina Faso](#), [Columbia](#), [Denmark](#), [Finland](#), [Ireland](#), [Netherlands](#), [South Africa](#), [OBIS](#), [GMBA](#))

# GBIF Metadata Grants



Source	OAI-PMH Service	Metadata Standard
Burkina Faso	<a href="http://gbif.spconedd.org/oai2.php">http://gbif.spconedd.org/oai2.php</a>	Dublin Core
Colombia	<a href="http://190.254.17.3/metadatos/OAI-PMH">http://190.254.17.3/metadatos/OAI-PMH</a>	Dublin Core
DanBIF	-	EML
GBIF Finland	<a href="http://digi.luomus.fi/oai2">http://digi.luomus.fi/oai2</a>	Dublin Core
GBIF Ireland	<a href="http://w08-geonetwork.compass.ie:8090/geonetwork/srv/en/oaipmh">http://w08-geonetwork.compass.ie:8090/geonetwork/srv/en/oaipmh</a>	ISO19139; ISO19115; Dublin Core; FGDC; CSW-Record
NLBIF	<a href="http://145.18.162.102/d6_test/oai2">http://145.18.162.102/d6_test/oai2</a>	Dublin Core
OBIS	<a href="http://rudev.iobis.org/oai_pmh_dataprovider/GetData">http://rudev.iobis.org/oai_pmh_dataprovider/GetData</a>	Directory Interchange Format
Endangered Wildlife Trust (SA)	<a href="http://www.kmsystem.ewt.org.za/oai/oai.aspx">http://www.kmsystem.ewt.org.za/oai/oai.aspx</a>	Dublin Core



# Related Activities

*Forest ?#!*

*Swamp !\*\$?*

Annotations of data/metadata records

Descriptors for species distribution data related to transects and areas

Integration of species occurrence and genomic data (e.g. habitat info)

## Standardised vocabularies & ontologies:

- Vocabularies for Biodiversity Informatics - Recommended activities for the GBIF Work Programme [http://imgbif.gbif.org/CMS/DMS\\_.php?ID=1057](http://imgbif.gbif.org/CMS/DMS_.php?ID=1057)
- Position paper on vocabularies and ontologies (2010)

## Persistent Identifiers:

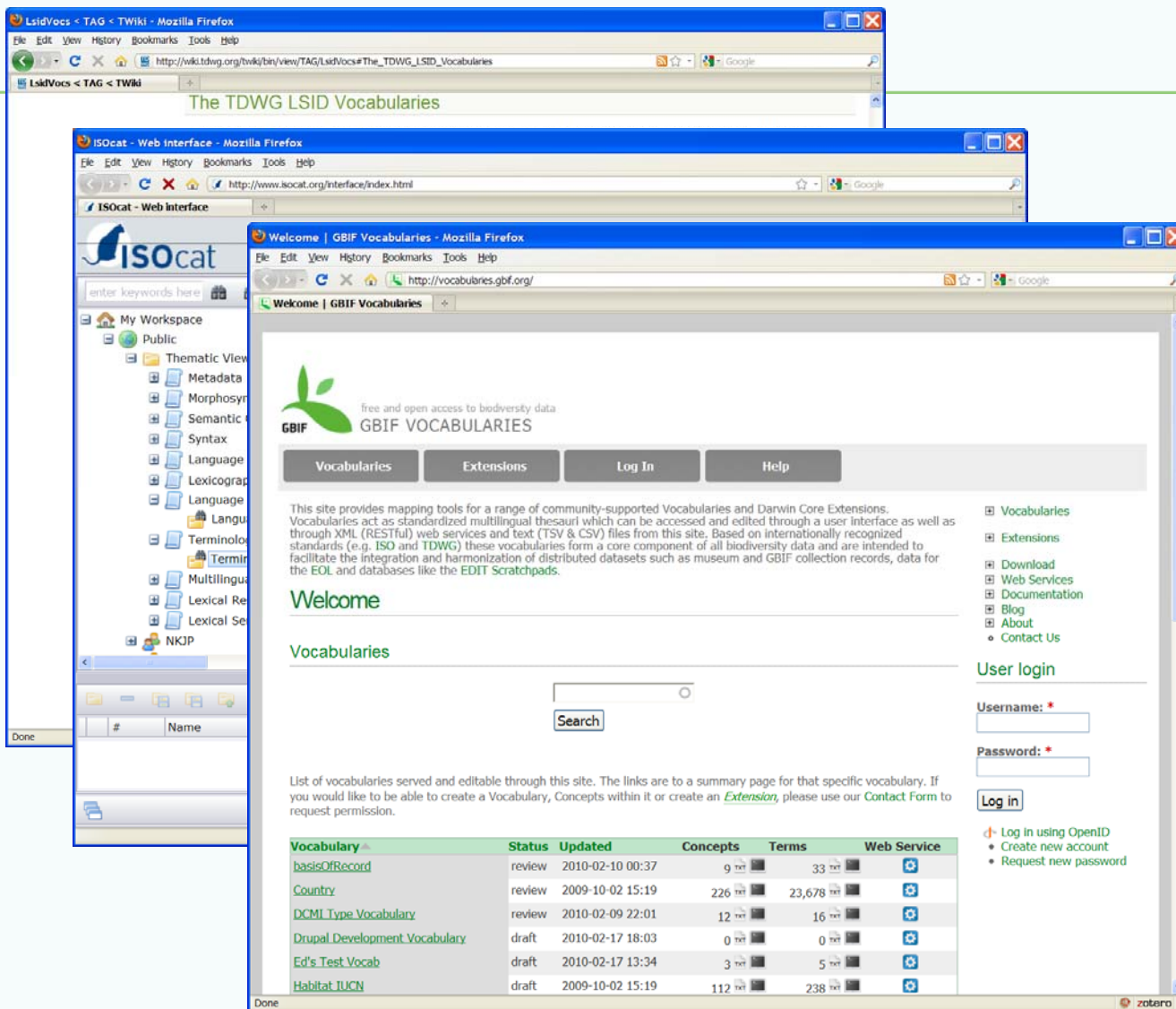
- Beginner's guide to issuing persistent identifiers (2010)
- Training event on persistent identifiers (2011)
- DOIs for datasets and data papers (2011?)

*Need for ...*

*Need for ...*







The image shows three overlapping browser windows. The top window is titled 'LsIDVocs - TAG < TWiki - Mozilla Firefox' and displays 'The TDWG LSID Vocabularies'. The middle window is titled 'ISOcat - Web interface - Mozilla Firefox' and shows the ISOcat web interface with a sidebar menu. The bottom window is titled 'Welcome | GBIF Vocabularies - Mozilla Firefox' and displays the GBIF Vocabularies homepage. The homepage includes a search bar, a list of vocabularies, and a user login section.

**GBIF VOCABULARIES**

free and open access to biodiversity data

Vocabularies Extensions Log In Help

This site provides mapping tools for a range of community-supported Vocabularies and Darwin Core Extensions. Vocabularies act as standardized multilingual thesauri which can be accessed and edited through a user interface as well as through XML (RESTful) web services and text (TSV & CSV) files from this site. Based on internationally recognized standards (e.g. ISO and TDWG) these vocabularies form a core component of all biodiversity data and are intended to facilitate the integration and harmonization of distributed datasets such as museum and GBIF collection records, data for the EOL and databases like the EDIT Scratchpads.

**Welcome**

**Vocabularies**

Search

List of vocabularies served and editable through this site. The links are to a summary page for that specific vocabulary. If you would like to be able to create a Vocabulary, Concepts within it or create an [Extension](#), please use our [Contact Form](#) to request permission.

Vocabulary	Status	Updated	Concepts	Terms	Web Service
<a href="#">basisOfRecord</a>	review	2010-02-10 00:37	9	33	
<a href="#">Country</a>	review	2009-10-02 15:19	226	23,678	
<a href="#">DCMI Type Vocabulary</a>	review	2010-02-09 22:01	12	16	
<a href="#">Drupal Development Vocabulary</a>	draft	2010-02-17 18:03	0	0	
<a href="#">Ed's Test Vocab</a>	draft	2010-02-17 13:34	3	5	
<a href="#">Habitat IUCN</a>	draft	2009-10-02 15:19	112	238	

User login

Username: \*

Password: \*

Log in

- Log in using OpenID
- Create new account
- Request new password

GBIF is driving the process to provide the social and technical mechanisms for a community to develop, publish and maintain vocabularies for their domain of practice.

# Metadata in 2011 GBIF Work Programme



**Goal 1:** Develop the GBIF Metadata Catalogue System  
consolidating and refining the GBIF metadata catalogue system  
and associated services established in 2010

**Activity 1:** Develop the GBIF EML metadata profile and IPT  
metadata editor

**Activity 2:** Develop best practice guide for writing GBIF EML  
metadata

**Activity 3:** Implement and deploy GBIF metadata catalogue

**Activity 4:** Develop the GBIF metadata catalogue network

# How to contact GBIF:



Web site: [www.gbif.org](http://www.gbif.org)

Data portal: [www.gbif.net](http://www.gbif.net)

GBIF Secretariat  
Universitetsparken 15  
DK-2100 Copenhagen Ø  
Denmark

E-mail: [info@gbif.org](mailto:info@gbif.org)

Phone: +45 3532 1470

Fax: +45 3532 1480

GBIF Secretariat building, supported by a grant from the Aage V. Jensens Fonde