

# GBIF Biodiversity Occurrence Data Publication in Africa: Cases from *icipe*

Kennedy Senagi  
(Data Manager and Postdoc Fellow)



International Centre of Insect Physiology and Ecology, *icipe*

[www.icipe.org](http://www.icipe.org)

Data Management, Modelling and Geo-Information Unit



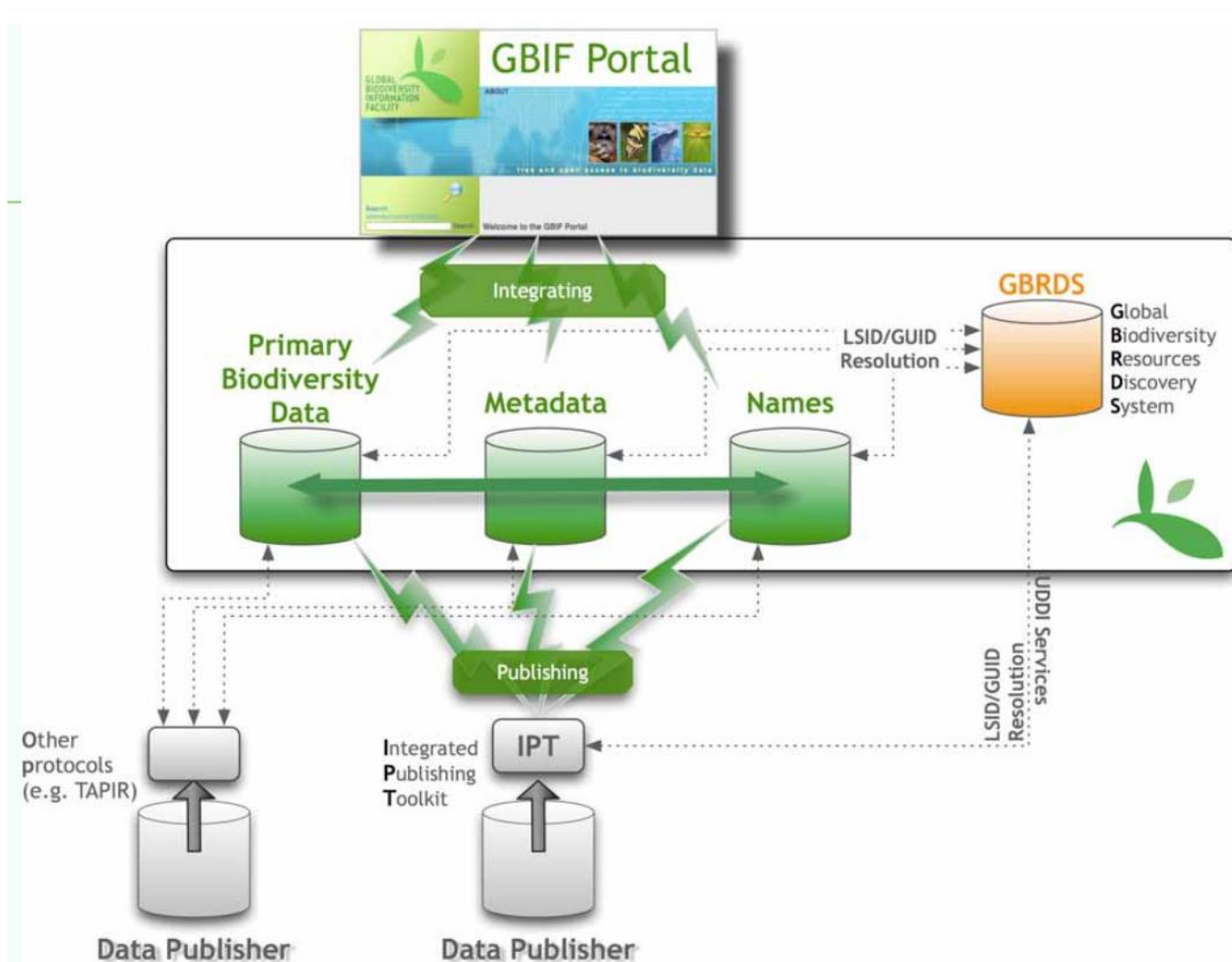
# The Stockholm Convention

- The Stockholm Convention is a global treaty that aims to protect human health and the environment from the effects of persistent organic pollutants.
- National government members (152+): Denmark, Kenya, Canada, USA, etc.
- The Stockholm Convention Regional Centre (SCRC) in Kenya is hosted by *icipe*.
- SCRC Kenya focuses on undertaking research and development for non-chemical alternatives to the use of hazardous pesticides including persistence organic pollutants (POPs) for management of pests and disease vectors.
- *icipe* leads significant reduction in the use of chemical-based pesticides using:
  - ‘Push–pull’ habitat management strategy.
  - Biopesticides.
  - Fruit-fly Integrated Pest Management (IPM).
  - Tsetse repellent collars; traps for control of tsetse flies
  - Habitat management for mosquito control; solar-powered
  - Honeybees; African Reference Laboratory for Bee Health

# Potential Benefits of Working with GBIF

- Mobilize and internationalize biodiversity data
- Promotes capacity building; grants, etc.
- Rich meta-data; Darwin core
- Networking
- Locals: enhanced tools built on the data
- Other resources: minting DOIs


# GBIF Portal Architecture



# GBIF Structure

- African head node;
- Country node; Kenya National Museum
- Institution node manager: Kennedy Senagi (icipe)

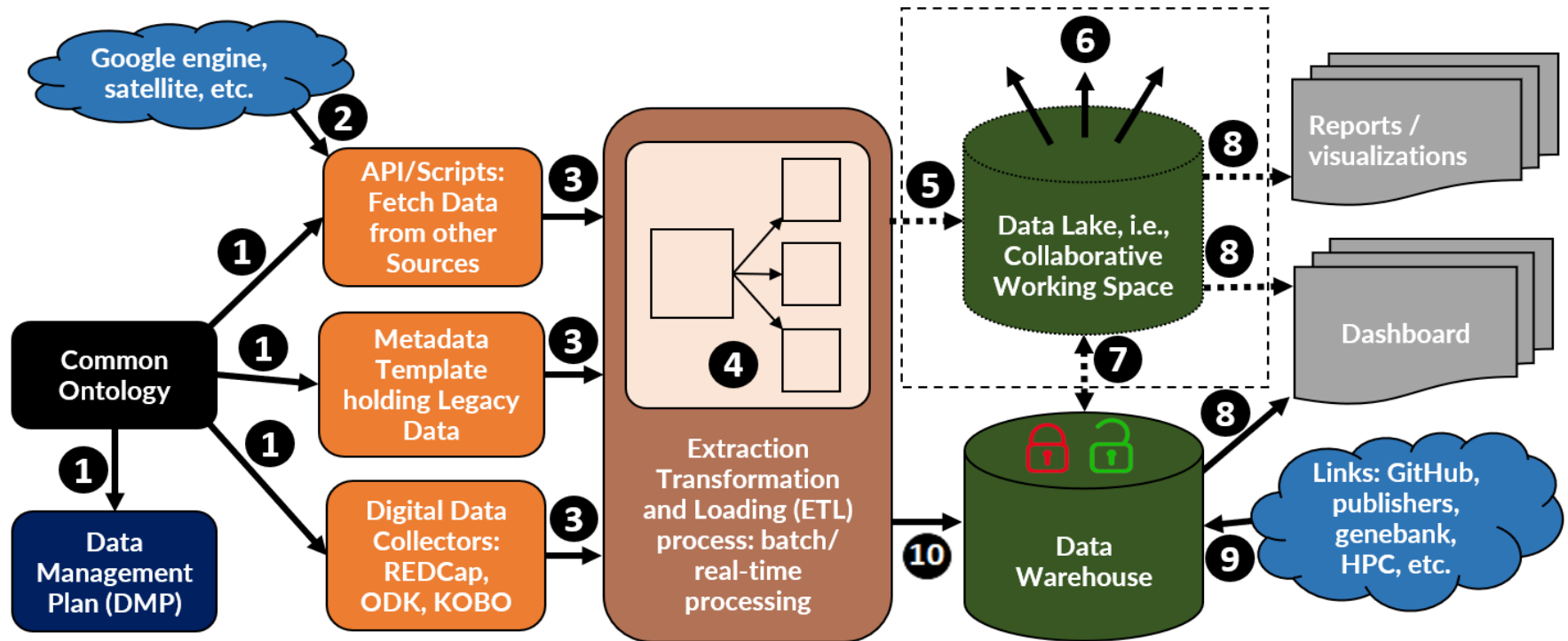
# Organization overview and mandate

- Founded in 1970. Prof T. R. Odhiambo. Currently, 500+ staff; both local & international
-  Kenya (Nairobi {HQ}, Mbita), Ethiopia (Addis Ababa), Uganda (Kampala) & Benin
- Activities: Insect science is supported by scientists from the related domain; research across Africa and beyond.
- Animal, Human, Environment, and Plant Health (4H) and supporting units → {1 Health}
- Mission/mandate: Developing and extending management tools and strategies for harmful and useful arthropods, while preserving the natural resource base through research and capacity building

# Data activities & Source of Funds

- Most, if not all, scientific activities are data-driven
- Activities:
  - Proposal writing; resource mobilization [Public Funded] – governments, well-wishers, etc.
  - Data collection: laboratory/experiment notes, genomic data, household data, field notes, photographs, geo-spatial, video and audiotapes, technical reports, simulation, etc.
  - Analysis – discover knowledge from data
  - Dissemination; publications, tools, etc. [Public Goods]
  - Archiving [Global/Digital Public Goods]; metadata, universal data/formats, license (CC BY 3.0)
- Pressure from donors:
  - Effective data management; the whole spectrum
  - Avail re-usable quality data to the public.
  - Data should meet universal standards
    - FAIR data management principles
    - Open Data
    - Open Science
    - Reproducible science
- Research Data Management and Archival (RDMA) Policy; complements:
  - The Kenya Data Protection Act 2019
  - international regulations

# Software and Data Engineering Pipelines



## Key:


1. Common ontology annotated to variables in the DMP, new data from collectors, legacy data, and data from cloud systems
2. Fetch readily available data from other sources e.g. Google engine, satellite, etc.
3. Load raw data to memory for further cleaning
4. Extraction transformation and loading (ETL) processes: cleaning, profiling, and annotation; batch/real-time
5. Load clean and annotated raw data to the data lake for analysis and visualization [work in progress]
6. Build data analysis pipelines e.g. modeling and machine learning etc. [work in progress]
7. Execute APIs to archive meta-data, dictionary, protocols, experiment plan, raw data, and clean data [work in progress]
8. Execute APIs to summarize data and provide reports/visualization/dashboards [work in progress]
9. Data resources in the data warehouse updated with links to external data sources
10. Load clean and annotated raw data to the data warehouse for visualization and archiving





Article

# A Novel Tightly Coupled Information System for Research Data Management

Kennedy Senagi \*  and Henri E. Z. Tonnang

International Centre of Insect Physiology and Ecology, Nairobi 30772-00100, Kenya

\* Correspondence: ksenagi@icipe.org



Citation: Senagi, K.; Tonnang, H.E.Z. A Novel Tightly Coupled Information System for Research Data Management. *Electronics* 2022, 11, 3196. <https://doi.org/10.3390/electronics11193196>

Academic Editor: Yeliz Karaca

Received: 2 September 2022

Accepted: 29 September 2022

Published: 5 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Abstract:** Most research projects are data driven. However, many organizations lack proper information systems (IS) for managing data, that is, planning, collecting, analyzing, storing, archiving, and sharing for use and re-use. Many research institutions have disparate and fragmented data that make it difficult to uphold the FAIR (findable, accessible, interoperable, and reusable) data management principles. At the same time, there is minimal practice of open and reproducible science. To solve these challenges, we designed and implemented an IS architecture for research data management. Through it, we have a centralized platform for research data management. The IS has several software components that are configured and unified to communicate and share data. The software components are, namely, common ontology, data management plan, data collectors, and the data warehouse. Results show that the IS components have gained global traction, 56.3% of the total web hits came from news users, and 259 projects had metadata (and 17 of those also had data resources). Moreover, the IS aligned the institution's scientific data resources to universal standards such as the FAIR principles of data management and at the same time showcased open data, open science, and reproducible science. Ultimately, the architecture can be adopted by other organizations to manage research data.

**Keywords:** information systems; research data; data management; data engineering; software engineering; common data model

# Links to the Data Resources

*icipe* Main website - <http://www.icipe.org/>

DMMG data resources landing page- <http://dmmg.icipe.org/>

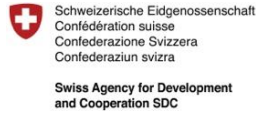
Links to Specific Data Assets/Resources Platform:

- ❑ Data Management Plan (DMP) - <http://dmmg-dmp.icipe.org>
- ❑ Common Ontology (CO) - <http://dmmg-co.icipe.org>
- ❑ Metadata Template (Legacy Data) - <https://github.com/icipe-official/Metadata-Template>
- ❑ Data Collector (REDCap) - <https://redcap.icipe.org/>
- ❑ Data Collector (ODK) - <https://odk-server.icipe.org/>
- ❑ Data Collector (KoBoToolbox) - <http://kf.dmmg-apps.icipe.org/>
- ❑ Project Manager Viewer (PMV) - <http://dmmg-pmv.icipe.org>
- ❑ Data Lake / Common Data / Working Area - work in progress
- ❑ Version Control (GitHub) - <https://github.com/icipe-official>
- ❑ Data Warehouse (CKAN) - <http://dmmg.icipe.org/dataportal>
- ❑ Services Request - [http://dmmg.icipe.org/service\\_request.html](http://dmmg.icipe.org/service_request.html)

# Biodiversity Occurrence Data

- *icipe* Data warehouse -  
<https://dmmg.icipe.org/dataportal/>
- GBIF Data warehouse -  
<https://cloud.gbif.org/icipe/manage/>

# Donor Acknowledgement



# Thank you



## International Centre of Insect Physiology and Ecology

P.O. Box 30772-00100, Nairobi, Kenya

Tel: +254 (20) 8632000

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

Website: [www.icipe.org](http://www.icipe.org)

Support *icipe*: [www.icipe.org/support-icipe](http://www.icipe.org/support-icipe)

 [facebook.com/icipe.insects/icipe](https://facebook.com/icipe.insects/icipe)

 [twitter.com/icipe](https://twitter.com/icipe)

 [linkedin.com/company/icipe](https://linkedin.com/company/icipe)