





A Semantic Big Biodiversity Data Integration Tool

Taysir Soliman, Alsayed Algergawy, Birgitta König-Ries, Majid Askar and Marwa Abdelreheim



Outline

- Big Data
- Big biodiversity Data
- Current Semantic Big Biodiversity Data integration tools
- Proposed tool
- Conclusion

What is Big Data?

Who's Generating Big Data?





Scientific instruments (all of us are generating data) (collecting all sorts of data)



Mobile devices (tracking all objects all the time)



Social media and networks

Transactions



Sensor technology and networks (measuring all kinds of data)

The progress and innovation is no longer hindered by the ability to collect data but, by the ability to manage, analyze, summarize, visualize, and discover knowledge from the collected data in a timely manner and in a scalable fashion



Ben Chams - Entolia

Characteristics of big data



Big Biodiversity Data







United Nations Decade on Biodiversity

PRESS RELEASE





Big data for biodiversity: Global Biodiversity Information Facility surpasses one billion records

- Global Biodiversity Information Facility provides open access to data about all life on Earth
- Records provide researchers and policy makers with unrivalled information resource, bringing together evidence on where and when species have been observed or collected
- Information result of the collective efforts of more than 1,200 institutions in 123 countries.

9 JULY 2018 - A global platform for sharing information about the world's biodiversity has passed a major milestone, with the publication of the one-billionth record of where species live, through the Global Biodiversity Information Facility (GBIF).

Current Semantic Big Data Integration Tools

- As Semantic Web technologies have been elaborated, one of the main research areas that have become widened is semantic data integration.
- Semantic data integration is a mechanism to associate heterogeneous data resources based on the meaning of the content.

What exactly we need ?

 In particular, we need data systems that enable us to move from 'what is where' questions to 'why is it there', 'what does it do' and 'what can we do about it'.

Current SBDI Tools

OntoText

Graph mining, Text Mining, IR, Semantic tagging, Semantic Search

Karma

Produces XML, KML, JSON files into a shared ontology

SBDI Tools

- Prepare datasets for analytics and visualization tools
- Handle uncertainty in biodiversity & monitoring datasets
- Semantically integrate heterogeneous data
- Provide knowledge graphs to express links between data

Proposed Tool: SBDI Architecture



Proposed Tool: SBDI Architecture



Conclusions

- Semantic Big Data Integration has become a necessity
- SBDI tool Architecture
- A prototype of SBDI will be developed and tested on a number of species.



THANK YOU