

Bio-logging Data Standards & Interoperability

DATABASE TO DATABASE



Overview

- Who are we?
- Current Context
- Interoperability – Why and How?
- Biodiversity repository surveys – Key Findings
- Examples in practice
- Discussion

Database to Database Standards & Interoperability

Current Context

- Many, many databases → many, many purposes
- Systems are created to target functionality, species, geography, habitat, funding, etc
- Will they ever merge? Very unlikely.
- Do they need shared access to data?



seaturtle.org



Interoperability – Why?

A (re)statement of requirements

- **Users and Systems will always want access to data from different systems**
- We know from our own experience in this small group:
 - WRAM sensor records → Swedish LifeWatch
 - ZoaTrack animal movements → Atlas of Living Australia species occurrence
 - Movebank tracking records → GBIF
 - Many others in this room
- Base requirements for data access are:
 - Discoverability between systems
 - Provenance & attribution
 - Data are described using published vocabularies/dictionaries
 - Machine readability

Interoperability - How?

Current Solutions

1. Manually query multiple databases



2. Multiple databases have identical schema



3. An aggregator database harvests from all others



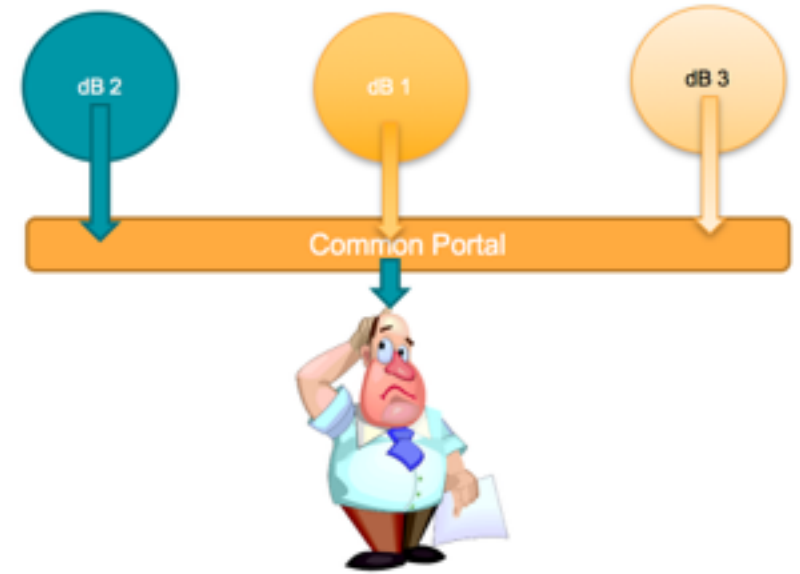
Challenges

Challenges	1	2	3
Same logins	😭	😐	😊
Same data formats	😭	😊	😊
Same data contents	😭	😊	😐
Ease of governance & implementation	😊	😭	😭
Uses standards	😭	😊	😊
Version control	😐	😐	😭

Interoperability – How?

One possibility

- **Establishing standards** could offer a common portal (hosted server / API / software?) that a user could query **once** and return results from **multiple databases**
- New databases can be included
- Standards allow ease of mappings between source system and users



Challenges	4
Same logins	😊
Same data formats	😊
Same data contents	😊
Ease of governance & implementation	😐
Uses standards	😊
Version control	😊

Biodiversity Repo Surveys – Key Findings

- 9 responses from biodiversity repositories
- Many use **Darwin Core** and **EML** (Ecological Metadata Language) – both are ubiquitous
- All seem to be gearing up to handle, or be able to handle, GB/TB volumes, and all mention that there are technical issues with performance, filtering and workflow.
- Diversity of opinion about how Bio-logging data should be represented – aggregated (summarised), or as individual records

Examples in practice

- OBIS-ENV-DATA
- Shared marine bio-logging database schema
- Oceanographic In-situ data Interoperability Project
- ZoaTrack / Living Atlases – Aggregating Movement data to Simple Darwin Core
- INBO / GBIF “Lossless” data
- WRAM / Swedish LifeWatch
- Others

Conclusion & Discussion

- Database to database standards & their place in our mindmap
- Issues for discussion?