RDA Interest Group Draft (renewed) Charter

(draft v0.7)

Name of Proposed Interest Group: Biodiversity Data Integration IG

Introduction

Disciplinary groups play a pivotal role in the RDA ecosystem. Not only do they provide a mechanism for engaging with the outputs of the RDA community, they are instrumental in providing input to the cross-cutting technical and socio-cultural groups. With its formation in 2013, the Biodiversity Data Integration IG has been one of the pioneering groups in RDA, focusing from the outset on linking the wider communities of practice in the biodiversity data domain with experts and other communities within RDA.

The COVID-19 outbreak was a game changer that will have a permanent effect on digital strategies. It has made it abundantly clear that the domain needs to expand its digital strategies beyond simply adding content online. This includes strengthening relationships with our audiences, supporting the development of people and their digital skills, evaluating joint policies and the ability to evaluate choices in the ever changing technology landscape. The BDI IG aims to be a platform for discussion and development of recommendations and guidance that can help institutions and researchers in the domain in their digital transformation.

This charter has been updated because there are many developments in the field of biodiversity data now that would benefit with alignment of developments in other scientific domains and vice versa. Examples are the global movement to FAIR data, implementing RDA recommendations towards a FAIR DO infrastructure for specimens with Natural Science Identifiers, an extended Catalogue of Life and the establishment of a Global Alliance for Biodiversity Knowledge.

Objectives

The group aims to steer discussion and coordinate efforts that create synergies among infrastructure developments around biodiversity data, and to connect these with RDA developments in the wider inter-disciplinary research infrastructure landscape. Infrastructures cannot operate in isolation but need to be part of the same value chain to deliver scientific outputs. This overall aim can be divided in three main objectives:

1. **Impact/effect/influence and disseminate some of the wider RDA discussion/work strands.**
   Many of the recommendations produced in RDA together are forming the foundations for how to design and develop systems, workflows and policies in such a way that data integration and interoperability problems can be overcome and enable interdisciplinary science. These recommendations also work toward establishing FAIRness of the data to enable Open Science. FAIRness means

compliance with the fifteen guiding principles of FAIR (Findable, Accessible, Interoperable, and Reusable).
A recent survey carried out by the BDI IG indicated however that two-thirds of the respondents from the Biodiversity data community are not yet aware of the RDA recommendations. The group aims to discuss the impact/effect/influence of some of the wider RDA discussion/work strands, such as data fabric, digital object architecture, etc. and their implications for the biodiversity/geodiversity domain. The BDI IG can then promote these recommendations further and write guidelines for application in biodiversity data. For FAIR principles implementation, the BDI IG aims to connect with the BiodiFAIRse Implementation Network in order to collectively build communities of practice.

2. **Crossbreed biodiversity standards and solutions with other domains.**
The biodiversity data community itself has made significant progress on issues like liberating biodiversity data from literature, semantic enhanced publishing, community standards development procedures (TDWG), industrial scale digitisation of objects, a common data exchange format (DarwinCore Archives) and community curation of data (using tools like WikiData, Catalogue Of Life Clearinghouse). The BDI IG aims to crossbreed biodiversity standards and technology and produce neutral solutions with other domains in order to achieve cross-domain interoperability. It may also disseminate the progress to other disciplines and ultimately transform these into RDA recommendations for cross-disciplinary use.

3. **Create recommendations and guidelines for specific biodiversity data integration issues.**
As a third objective, the BDI IG aims to create recommendations and guidelines specific for biodiversity data integration, e.g., the adoption of (TDWG) community standards or community policies across biodiversity data infrastructures, or guidelines for connecting biodiversity data through globally adopted solutions for taxon names and persistent identifiers for occurrence records and specimens.

User scenario(s) or use case(s) the IG wishes to address

The use cases the IG wants to address have changed over time. This triggered an update of this charter. The following areas were recently identified by the IG members:

- **Data integration for biodiversity data**
    - **Data integration through Natural Science Collections**: Create recommendations that enable linking data together using specimens (hard evidence) as the basis. Several specifications and standards are already in place,e.g., DarwinCore, ABCD, EML, or in development, like OpenDS, MIDS, natural sciences identifiers (NSId), and the digital specimen and extended specimen concepts. The BDI IG will be used as a platform to discuss and enhance these as part of the wider RDA landscape.

    - **Data liberation from biodiversity literature, semantic enhanced publishing:** Establishment of the Biodiversity Heritage Library, TreatmentBank and integration of liberated taxonomic treatments. GBIF has provided much experience in semantically enhanced publishing, which can be disseminated further through the BDI IG and through support of RDA groups working on developing cross-disciplinary recommendations in this direction.
    - **Semantic initiatives for data integration**: to semantically describe taxon traits, phenomics, and interactions enabling the creation and visualization of linkages with and between e.g. species and specimens.
- **Increase usage of research data by biodiversity authorities**: Biodiversity research data are often not used by biodiversity authorities. The IG aims to promote adoption of RDA recommendations that aid the discovery, usage, open data sharing and trust in these data by these authorities. There is also a need to have a better understanding, from a policy standpoint, on how biodiversity data can contribute to e.g., the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) and the Intergovernmental Panel on Climate Change (IPCC).
- **Policies for referencing data**: Development of best practices and community policies to refer to biodiversity data using Persistent Identifiers (PIDs), which should influence requirements set by publishers and funders. Recommendations should push prospective publishing, making sure data are properly published from the scholarly and semantic point of view and to ensure that such data can directly flow into infrastructures.
- **Biodiversity data in Essential Biodiversity Variables (EBVs)**: The positioning of biodiversity data in "data cubes" is unclear; standards and pipelines need to be identified in collaboration with the monitoring and climate research domains and the Group on Earth Observations Biodiversity Observation Network (GEO BON). See also the example of how observation and specimen data can work together in one data cube specifically for creating models and indicators.
- **Develop guidelines to increase the synergy in achieving technology neutral solutions for enhanced data quality**: Data quality is a major concern in biodiversity informatics caused by the distributed nature of data acquisition and digitization, and also by specific problems in sub-domains, such as taxonomic data and geographic data. Initiatives like the Biodiversity Data Quality Framework developed in Biodiversity Information Standards (TDWG) will only work if synergy is achieved within the domain for implementing such solutions.
- **Promote synergy between existing initiatives to foster community adoption and experiences exchanges**: There are several initiatives and communities of practice that can contribute to foster community adoption of standards, as DataONE, LTER, iBOL, Map of Life, GBIF, NEON, ALA, OBIS, BioCASe, iNaturalist, GEO BON. In that sense, promoting connections will help to build a stronger network.

Participation
Participation in the group is open for anyone interested but is targeted to the biodiversity data domain. Where there is overlap between biodiversity and other domains, like earth science data in natural history collections, development of guidelines may need to be done in conjunction with other RDA groups like the Physical Samples and Collections IG.

The group has more than 150 members spread across the continents, including international experts in many of the biodiversity subdomains. Africa is currently underrepresented and should be targeted to balance the membership. The members also aim to expand participation to global and regional biodiversity infrastructures, to create better linkages with current developments in these infrastructures.

Outcomes
The BDI IG intends to be a platform for discussion and coordination of existing ongoing initiatives, however also aims to establish some outputs in the form of guidelines or recommendations. A number of examples have been recently mentioned that might lead to WGS to produce these outputs. These are:

1. Guidelines for linking biodiversity data to ecosystem services assessments;
2. Guidelines about how to markup biodiversity-related websites, in collaboration with Bioschemas' biodiversity group;
3. Recommendations for synthesis under constant taxonomic change;
4. Recommendations/standards for metadata versioning and reconciliation;
5. Closing the taxonomic gap between biodiversity and agrobiodiversity;
6. Recommendations for real-time transaction based data publishing and annotation;
7. Alignment on work for data standards to represent specimens as being carried out in ABCD community, IGSN, DOI, TDWG, DiSSCo etc.;
8. Persistent species identifiers;
9. Recommendations for biological data standards for ecological data;
10. Recommendations for how to handle species absence data;
11. Best practices for interoperability between physical data standards (i.e., Climate and Forecast) and biological data standards.

Mechanism
From the group objectives it is evident that links to other relevant RDA groups would be key to the success of the BDI IG. These links need to be developed more. The BDI IG has already been working together with ELIXIR Bridging Force IG (Life Science domain) and organised joint sessions. Group members have been involved in biodiversity data integration related work in other groups, like the RDA/TDWG WG on Metadata Standards for attribution of physical and digital collections stewardship. But there are

many more opportunities. Ricarda Braukmann created a report[1] in which he summarised the current RDA Interest and Working Groups and analyzed them with respect to their relevance for social science researchers. The grouping in this report can also be used to identify groups of relevance for linking with the BDI IG. The report lists groups on general topics like policies and guidelines, data management and technical infrastructure aspects. When working on one of the BDI IG use cases, group members should use this overview to identify relevant groups and make a connection. Also, the group should establish connections with the Disciplinary Collaboration Framework IG which aims to bring together different research disciplines to discuss disciplinary-specific use cases of RDA outputs. Other relevant groups are the Social Dynamics of Data Interoperability IG, the Vocabulary Services IG and Sharing Rewards and Credit (SHARC) IG. The group should also establish links with GO FAIR Implementation Networks for biodiversity data like BiodiFAIRse, and, when endorsed, with GO FAIR IG. The mechanism for creating the linkages with these groups will be through organising joint sessions, group members participating in sessions of these groups, and ultimately through the establishment of joint WGs.

The group aims to meet during the plenaries (physical meetings with remote participants) but also to have virtual meetings in between the plenaries. The group members have indicated that having more virtual meetings might create more engagement and keep people close to the group between plenaries.

Where the group sessions at the plenaries may be more like dissemination events, the virtual meetings in between will have the aim of working sessions on specific topics. In general the group aims to develop activities that complement existing initiatives rather than developing new initiatives. This can, for example, include initiatives already under development in funded projects, infrastructures, standards development or other RDA groups. For working sessions the group aims to be specific on how people can tangibly contribute to the topic being worked on. Members are asked to accompany the organisation of working sessions with specific documents or ideas so that people can immediately interact.

Working groups may be established to work on larger tasks and produce tangible outputs in the form of recommendations or guidelines. A few ideas for new working groups have been proposed by the members:

- Minimum Information about a Digital Specimen recommendation (jointly with TDWG & Physical Samples & Collections IG);
- Open Digital Specimens (openDS) recommendation (jointly with TDWG);
- Guidelines for specimens citation in publications.

[1]Ricarda Braukman, 2019. RDA Overview for the Social Sciences - Report October 2019. DOI: 10.5281/zenodo.3580674

Activities may also be aligned with or add to work being carried out in already existing or past WGs. Examples are:

- I-ADOPT WG which aims at building an Interoperability Framework for describing observable properties;
- RDA/TDWG Attribution Metadata Working Group (attribution standard is further developed in TDWG).

To get a better understanding of the capacities of the group for development of activities, the group created during P15 a list of initiatives we are involved in or know of. Although such an overview rapidly gets outdated, it is helpful in identifying initial key areas of work. The list can be found in the session notes.

Timeline

The main goals for 2020 are to reactivate the group, inventory the current needs for a BDI interest group, update the charter and to identify key areas of work in current initiatives. In 2021 this will be used to discuss further work in the identified areas and establish working groups. Another goal for 2021 will be to create the connection with the identified RDA IGs and WGs of relevance, and to strengthen the connections with identified relevant initiatives in implementation networks and infrastructures.

Planned 2020 milestones:

- March: current interest group needs inventoried and discussed,
- June: draft updated Charter ready,
- September: group endorsed with updated charter,
- November (Plenary 16): first working group(s) established.

Planned 2021 milestones:

- First virtual working meeting in between plenaries organised,
- P17 and P18 joint sessions with one or two of the identified IGs of relevance,
- A P17 and/or P18 BDI IG session to disseminate WG results and/or establish new WGs.

Group Members

Chairs:

Wouter Addink, wouter.addink@naturalis.nl (Europe)

Sridhar Gutam, gutam2000@gmail.com (Asia)

Hamish Holewa, hamish.holewa@csiro.au (Australia)

Libby Ellwood, ellwoodlibby@gmail.com (USA)

Members:

see current members at:

https://www.rd-alliance.org/groups/biodiversity-data-integration-ig.html