Biodiversity Data Integration Interest Group

# Survey analysis

Report written by Ely Wallis, Atlas of Living Australia, CSIRO,

31 March 2020

# Summary of conclusions

The survey was conducted by the Research Data Alliance Biodiversity Data Integration Interest Group and was intended to be an input into a face-to-face RDA Plenary in Melbourne, Australia in March 2020. The meeting was ultimately cancelled due to the rapid escalation of the COVID19 pandemic.

Conclusions that can be drawn are:

* Approximately two thirds of a total of 94 respondents were familiar with RDA and its outputs but only approximately one third were familiar with the Biodiversity Data Integration Interest Group.
* 63% either agreed or strongly agreed that the BDI Interest Group descriptions and objectives need to be updated; 18% did not answer the question.

Respondents emphasised three areas they thought the group should focus on (each gathering over 40 votes), described as:

* Disseminate, Link and Find
* Data Management
* Data Description

The most important way to achieve the main objectives of the group, namely adoption of common tools and services to establish data interoperability within the biodiversity domain, was seen as being by:

* Implementing FAIR data principles in the biodiversity domain

Survey respondents most commonly described themselves as working in:

* Research (academic & non-academic, including citizen science) (31%)
* Technical support (ICT & Information Management) (19%)
* Policy (11%) and Collection management (10%)

In response to the question about whether respondents would be interested in participating in a task group:

* One third said they’d be keen to participate
* A different third said they’re interested but unable to participate.

The survey was completed by **94** people

|  |  |  |
| --- | --- | --- |
| **Question** | **% Yes** | **% No** |
| Are you already familiar with the Research Data Alliance and its outputs? | 62 | 38 |
| Are you currently a member of the Research Data Alliance? | 34 | 66 |
| Have you ever been actively involved in development of one or more of the RDA outputs? | 27 | 73 |
| Do you see RDA outputs and recommendations as relevant for your work? | 95 | 5 |
| Are you familiar with the Biodiversity Data Integration Interest Group (BDI IG)? | 38 | 62 |
| Are you a member of the Biodiversity Data Integration Interest Group? | 19 | 81 |

The most noteworthy of these initial questions was the response to ***Do you see RDA outputs and recommendations as relevant for your work?***Positive response was 95%.

Of the 36 people who responded **No** to the question**, *Are you already familiar with the Research Data Alliance and its outputs?***

* 3 of the 36 said that they’re a member of the RDA!
* 32 of the 36 said that they see RDA outputs and recommendations as relevant for their work, despite saying they’re not familiar with the RDA and its outputs.
* 4 of the 36 said they \*are\* familiar with the BDI Interest Group and 1 said they are already a member of the BDI Interest group

## Will you attend the BDI Interest Group session at RDA P15 in Melbourne?

Not relevant now but responses were: yes 6%, no 80%, not sure 14%

## Do the BDI IG group description and objectives need to be updated?

Raw data:

|  |  |
| --- | --- |
| I strongly agree | 34 |
| I somewhat agree | 25 |
| I have no opinion on this | 15 |
| I disagree | 3 |
| Left blank | 17 |

## The group title specifically mentions biodiversity, however especially in natural collections the scope is both biodiversity and geodiversity, do you see this as a problem?

Full text of answer options is:

* I think the group can work both on biodiversity and geodiversity related data integration issues, the title does not need to change.
* I think the group should work on both biodiversity and geodiversity related data integration issues, and the title needs to change.
* I think the group should focus on biodiversity only. Geocollections are out of scope for this group.
* The group should change its biodiversity domain and/or data integration focus to: (free text answer)

Raw data

|  |  |
| --- | --- |
| Work on both biodiversity and geodiversity issues, no title change | 41 |
| Work on both biodiversity and geodiversity issues, but change the title | 12 |
| Focus on biodiversity only | 22 |
| Change the focus of the group | 3 |

Three respondents wanted to change the focus of the group commenting that the group focus should be on:

* Natural sciences
* Modeling taxonomic evolution and persistent conflict. GNA is under powered, currently, to achieve this.
* I think that is important to include the agrobiodiversity which is a relevant and integrative topic today related to CBD and FAO and ODS and suposse another goals and problems especially in areas that are center of domestication and diversification on species important to food and agriculture at global level

## What should be the primary focus of the Biodiversity Data Integration Interest Group in terms of development of recommendations? (multiple options possible)

Raw data – note that respondents could pick several options

|  |  |
| --- | --- |
| Data Management | 46 |
| Data Collection | 18 |
| Data Description | 41 |
| Identity, Store, and Preserve | 33 |
| Disseminate, Link, and Find | 49 |
| Policy, Legal Compliance, and Capacity | 26 |
| No primary focus needed | 2 |

## Currently one of the main objectives of the biodiversity data interest group is adoption of common tools and services to establish data interoperability within the biodiversity domain. However there are more ways of achieving this, which one(s) do you see as the main [sic – most] important?

|  |  |
| --- | --- |
| Option | Count |
| Implementing FAIR data principles in the biodiversity domain | 20 |
| Achieving common protocols and best practices for data sharing | 17 |
| Wider adoption of biodiversity data standards | 14 |
| Convergence on Data Infrastructure through Digital Objects with persistent identifiers | 12 |
| Adoption of common biodiversity tools and services | 7 |
| Other (free text answer) | 7 |
| Left blank | 17 |
|  | 94 |

### Free text of the seven “Other” answers

**All of the above / the options are all closely related / several are important**

* All of the above, if possible. There is considerable overlap and commonality in the options offered.
* These seem closely related, and perhaps aspiring to FAIR biodiversity data practices may hit upon others in this list. Tools and services seems vague- are these for data generators, managers, or consumers?
* I see adoption of tools and services as most important, but all of the above are important.
* Achieving common protocols and best practices for data sharing, Wider adoption of biodiversity data standards, and Convergence on Data Infrastructure through Digital Objects with persistent identifiers
* Asks for "one(s)", but only allows one, so I would say Convergence of Data Infrastructure through digital objects/persistent identifiers, Adoption of common tools/services, achieving common protocols/practices for sharing (in that order).

**Biological systematics / incorporating disparate initiatives**

* I believe we need to make the next step for several biodiversity information initiatives to converge all around the same foundations of facts (nomenclature, taxonomy, name usage, literature references, occurrence) in such a way the same data does not have to be maintained in various places. The approach should be independent of technical solutions.
* Accommodate biological systematics better as a continuous, integration disrupting component; align evolving taxonomies and also phylogenetic insights. Foster bold new research into the theory of biodiversity informatics synthesis in light of conflicting taxonomic perspectives. Model taxonomic concept relationships.

## Do you have suggestions for experts that should become member of the interest group? (free text answers)

### Individuals mentioned by name

* Geoffrey Bowker (UC Irvine)
* Robert Montoya (Indiana University)
* Nico Franz (Arizona State University)
* Edward E Gilbert (Arizona State University)
* Carl Boettiger (UC Berkeley)
* David Bloom
* John Wieczorek
* Steve Baskauf
* Peter Desmet

### People with certain skills

**Taxonomists / managers of names checklists / national and regional focus**

* Data aggregation and the compilation and management of global, national and regional species checklists are not the same thing - aggregators should not attempt to restrict the science of taxonomy to meet the practical needs for informatics, and national needs should bot be ignored to meet the ideal for a global checklist. Hence some national/regional taxonomy checklists/managers should be added to the group
* Biodiversity data initiatives generally suffer from poor representation of the interests and knowledge of the data generators (i.e., taxonomists). Standards for data quality and support for the taxonomic work and expertise needed to ensure the reliability of biodiversity data lag severely behind the development of tools to analyze large compilations of often unreliable data.
* I've been involved in the Australian Faunal Directory and see that standardising taxon names worldwide as sometimes problematic. Experts may differ in what they consider to be the valid name of a particular species, especially for widespread taxa. And then - higher classification is a work in progress in many instances.

**Representatives of biodiversity infrastructures / aggregators**

* I recommend more focus on bringing in representatives of running infrastructures (DataONE, LTER, iBOL, Map of Life, GBIF, NEON, ALA, BioCASe, Zenodo, Amazon AWS, Google Earth Engine, rOpenSci, iNaturalist etc). There is valuable operational experience (infrastructure challenges, community adoption etc) and lessons learned that could help steer RDA outputs to be more readily useful. The research communities (e.g. GeoBON, environmental modeling) and the library community may be avenues to explore too.
* Just include the experts of LifeWatch, EU BON, GBIF, Encyclopedia of Life and Atlas of Living Australia and others
* Representatives of all the main biodiv/geodiv initiatives, compatible with a suggested new focus on convergence on data infrastructure through DOs with PIDs.

**General comment**

* Biodiversity data extends beyond natural history collections which should be taken into account. Try to make use or improve existing biodiversity data standards, e.g. Darwin core and mainstream the use across domains.

## In what user category do you fit the best?

|  |  |  |  |
| --- | --- | --- | --- |
| **User category** | **Number of respondents (total n=94)** | **User category percentage of respondents** | **Free text description of “other”** |
|
| Research (academic & non-academic, including citizen science) | 31 | 32% |  |
| Technical support (ICT & Information Management) | 18 | 19% |  |
| Collection management | 9 | 10% |  |
| Policy (institutional, national & international) | 10 | 11% |  |
| Other | 4 | 4% | 2 x Government (not policy), 1 x Research plus Education, 1 x Research plus Technical |
| Education (academic & non-academic) | 1 | 1% |  |
| Industry | 1 | 1% |  |
| No answer | 20 | 21% |  |

## There are some ideas for new task groups under the Biodiversity Data integration Interest Group (task groups need to deliver a tangible output and have 18 months to achieve that). Examples: 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 using Natural Science Identifiers (NSId).

Raw data (n=94)

|  |  |
| --- | --- |
| These ideas sound awesome, I would like to participate | 30 |
| I would be interested in using the outcomes but I cannot participate in the development | 28 |
| I am not interested in these outcomes | 5 |
| Great ideas, but RDA is not the right place to develop these outputs | 1 |
| I have some ideas for other taskgroups and outcomes | 10 |
| No answer | 20 |