

THEMATIC GROUP 1: ALA TECHNOLOGY

Members: David Martin, Veronika Johansson, Maxim Shashkov, Luka Katušić, Marina Samejima, Gautam Talukdar, Maofang Luo, Rui Figueira, Baba Ainina and Cristina Villaverde.

We found 8 topics but we couldn't discuss them all.

1. Infrastructure: cost and stability

- ALA is using Amazon EC2 and CSIRO's cloud infrastructure. The cost in Amazon depends on how much data, how many components, how update you want to keep your data. E.g. in UK 35.000 pounds/per year
- It is recommended that the servers are maintained by someone else, not by the node itself. Hence using a cloud infrastructure is recommended.
- In Europe the infrastructure can be for free thanks to EGI (European GRID Initiative). Fr, Pt and Sp are already using it. However in Spain this initiative has finished and they want money to continue supporting the infrastructure. Possibility to move to the Portuguese infrastructure
- Migrating from one infrastructure to another can take about two weeks
- Installation should be done with Ansible scripts but we there is no easy workflow about that

2. GBIF vs ALA

There is an overlap between both initiatives as they have same kind of data, but there are some differences. ALA and GBIF are planning to align data quality processing libraries and already make use of common software libraries (e..g darwin core archive, name parsing) ALA has a lot of components (citizen science, species information, etc). GBIF only focused on occurrence data; local installations of ALA might engage locals to mobilize data and use GBIF.

National portal also allows more analysis capabilities, with regionalizations for a country (e.g. political boundaries, land classification, national parks, areas of scientific interest), a country specific taxonomy and species lists including national legislative lists.

ALA can also be used to handle processing of sensitive records, which can help mobilize datasets which have sensitive records.

3. Feedback Data Enhancements/Corrections

We recommend that if data downloaded become better quality should be uploaded again to GBIF. There are two ways of doing this, if scientists improve data they should report to the data provider. ALA supports that users do that themselves but need authentication.

4. Sample based data in ALA

- ALA needs better support for sample based data
- It converts it to occurrences
- Not working on it really actively in ALA, but it is an area of interest
- Same with molecular data, although ALA tools have been used for this (e.g. <http://base.ala.org.au/datacheck/datasets>)
- Data currency is the main issue in Estonia

- Estonia has extended ALA to support sequence data

Other topic areas not fully discussed

- Funding portal developments
- Community driven data feedback
- Keeping with developments
- Ongoing funding for workshops