FAIR data for the ‘long tail of science’

Program

6 September - 10 September 2021

Venue: Lorentz Center@Online

All the times are provided in CET

Monday

This day will be centred around introducing participants, their goals in attending the workshop, and the expertise they bring. We will also introduce and discuss the aims of the workshop, and the main topics and activities planned for the next four days. As participants come from different (and currently not well-connected communities), we will organize a discussion in three break-out groups: each will contain at least one person from each of the three groups of expertise (ecologist, database managers, e-infrastructures/FAIR data). Workshop organizers will moderate the break-out groups, and we will use whiteboards to (a) discuss why it is important to create a FAIR data landscape and what are the obstacles in creating it; (b) formulate the main concepts important to agree on during the next days.

10:00-10:15 Intro by Lorentz about the set-up etc.
10:15-10:30 General introduction and objectives of the workshop
10:30-11:00 Brief introduction by each participant (1 min per person, prepare slides)
11:00-11:15 Break
11:15-12:00 Break-out groups (4) each moderated by one organizer (based on the two topics mentioned in the above text, (a) and (b)
12:00-12:30 Report from break-out groups
12:30-14:00 Lunch, off-screen time
14:00-14:20 Talk on Open data and European Open Data Landscape (Karel Luyben)
14:20-15:00 Discussion on the long-tail of data in the European landscape
15:00-16:00 Option for online socializing – in Wonder

Tuesday

Achieve goal 1: Standard data format and identifying problems in communication between databases

Please watch pre-recorded talks on Data formats (Mobebank, SPI-Birds, GBIF, BTO/EURING) and FAIR data. These are accessible to participants via Research Drive, under ‘Talks Day 2’ folder. Based on the talks, please add your questions and add your ideas to the shared Google docs.

The aims of the second day are to identify the similarities and differences in data formats, meta-data formats, and semantics used by different databases, and their link to FAIR data, EML and Darwin Core. In pre-recorded talks, one representative of each database that integrates the data for a research community (SPI-Birds, EURING, Movebank, GBIF) will present their meta-data
and data formats and argue for why these were chosen. This will be followed by a live Q&A. Then, based on pre-recorded talk by Barend Mons on FAIR data we will discuss what (shared) meta-data or data fields would be best to use to enable communication between the databases. In the afternoon we will try to implement these changes which will serve to also identify any additional problems that need to be addressed in the Wednesday session.

10:00-10:30 Plenary discussion on the Monday activities. Moderator Marcel Visser
10:30-11:00 Q&A on pre-recorded talks on the meta-data and data formats used by different databases (SPI-Birds, EURING, Movebank, GBIF)
11:00-11:15 Break
11:15-12:30 Discussion on interoperability between databases & their alignment with FAIR, EML, and Darwin core
12:30 -14:00 Lunch/option for socializing
14:00-15:00 Hands-on: connecting databases (based on data format and semantics)

Wednesday

Achieve goal 2: Semantics and data interoperability, data use

We will first present a summary of the issues that prevent creating a connection between databases based on meta-data and data fields, as identified in the previous day's session. We will then use this to develop a roadmap to best link the data contained within different databases (e.g. data on the same individuals that are present in different databases). This roadmap will be circulated among participants (after the workshop) for further comments.

The afternoon will be dedicated to a hackathon. Here, SPI-Birds code developers and other ecologists, joined by participants who will join online, will be developing codes to test the suitability of data hosted at SPI-Birds database for testing a specific scientific question: document spatial and temporal variation in adult survival in a number of small passerine bird species. The developed pipeline will need to produce a ready to be analysed dataset for capture/recapture analysis as a pipeline 'check'. Has the work done in the workshop led to an efficient pipeline? Here, junior researchers will be merged in groups with more senior researchers. The results of this session will be discussed on Thursday.

10.00-10:45 Plenary discussion on the Tuesday activities. Moderator Stefan Vriend
10:45-11:00 Intro to break-out groups
10:30-11:15 Break
11:15-12:00 Break-out groups - Creating a roadmap to connect databases (based on semantics and data/meta-data field)
12:00-12:30 Reporting back from break-out groups
12:30-14:00 Lunch/off-screen time
14:00-15:00 Hackathon - Case study: getting SPI-Birds standard data into a format for survival analyses

Thursday

Achieve goal 3: Technical architecture and support, case study evaluation
Please watch pre-recorded talks on Technical aspects of Mobebank, SPI-Birds, GBIF, and BTO/EURING. These are accessible to participants via Research Drive, under ‘Talks Day 4’ folder. Based on the talks, please add your questions and add your ideas to the shared Google docs.

Each database (EURING, SPI-Birds, Movebank, GBIF), will present, in a pre-recorded talk, what they find as weak, and what as strong aspects of their infrastructure. Based on the input from database managers, and e-infrastructure experts, we will together create a roadmap for technical improvements of databases. In the afternoon we continue the hackathon. We will end the day with a report by two junior researchers from the group that addressed the scientific question and evaluate the pipelines developed during the hackathon. Provide suggestions and agree on follow-up steps.

10.00-10:30 Plenary discussion on the Wednesday activities. Moderator AC
10:30-11:00 Discussion on talks on the technical infrastructure of the 4 databases
11:00-11:15 Break
11:15-12:30 Discuss the potential solutions to problems and created a roadmap
12:30-14:00 Lunch/option for online socializing
14:00-15:00 Hackathon (survival data) continues and wrap-up

Friday

Funding, conclusions, and planning the future

10.00-10:30 Wrap-up on the Thursday activities. Moderator Erik Matthysen.
10:30-11:00 Talk on funding landscape by Cees Hof followed by Q&A
11:00-11:15 Break
11:15-12:30 Roadmap to financial support: each database, and the overall integration of databases.
12:30-14:00 Lunch/socialising option
14:00-15:00 Wrap-up of the Workshop and outline follow-up actions after the Workshop