

STSM Report

Introduction

This is a report for a Short Term Scientific Mission (STSM) for the [COST Action ES1305: European Network for the Radar surveillance of Animal Movement \(ENRAM\)](#) under working group 3: [Visualizing spatiotemporal patterns of animal movement](#).

- **STSM title:** Data infrastructure proof of concept: implementation
- **Host institution:** Department of Bird Migration at the Swiss Ornithological Institute in Switzerland under the coordination of Dr. Silke Bauer and Dr. Felix Liechti.
- **Grantee:** Stijn Van Hoey from the LifeWatch team at the Research Institute for Nature and Forest (INBO) in Belgium.
- **Dates:** January 16 to 20, 2017

Purpose

The goals set out for this STSM were:

- To implement a data repository for bird migration altitude profile data (<https://github.com/enram/infrastructure>). This includes the development of a script to move data from the BALTRAD infrastructure to the data repository.
- To implement a database for bird migration altitude profile data and metadata. This includes the development of a script to periodically load data from the repository into the database.

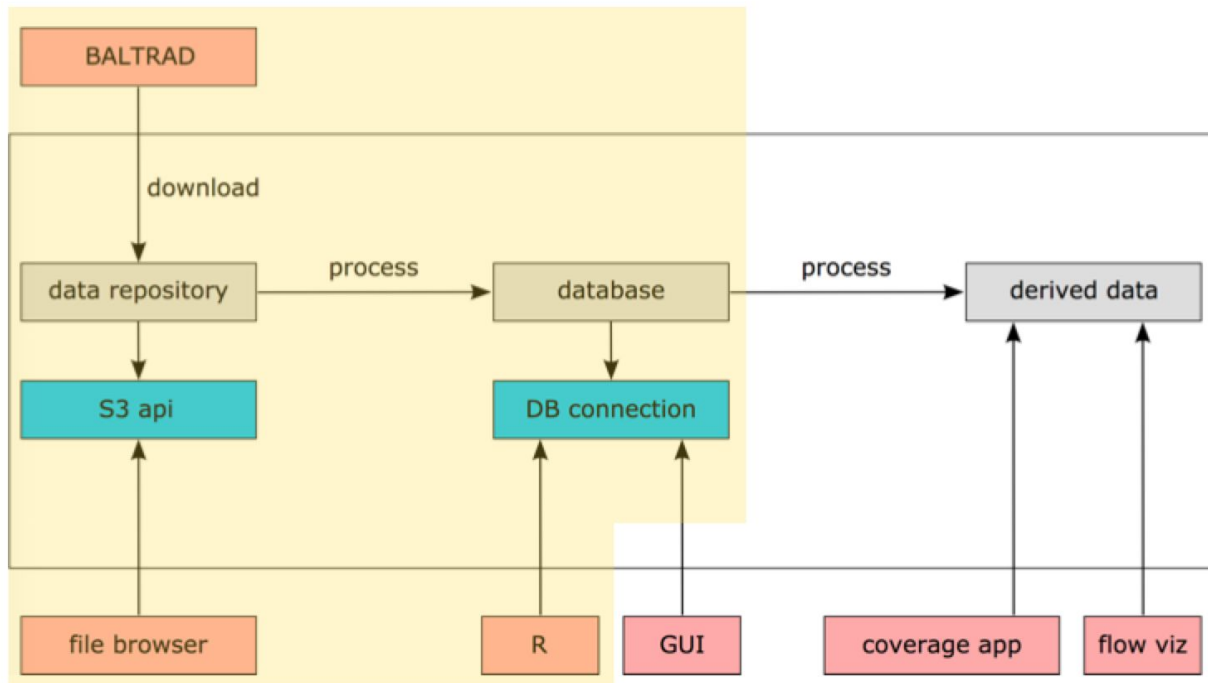
This STSM was done in close collaboration with Liesbeth Verlinden (University of Amsterdam, the Netherlands) and Peter Desmet (INBO, Belgium). The sections below describe the activities and results of all three STSMs.

Description of the work carried out

Monday 16 January 2017

Liesbeth, Peter, Stijn: Participate in the [Western European flyway workshop](#), which was held in parallel with the first three days of our STSMs. This provided an excellent opportunity to get direct feedback from the potential users of the ENRAM data infrastructure. At the workshop, we presented the initial goals of our STSMs (indicated in yellow in the schema below):

1. Develop a script to download bird profile data files daily from the BALTRAD FTP server, where these files are only kept for 3 days.
2. Develop an open data repository on Amazon S3 and a file browser website to access this repository.
3. Aggregate the bird profile data in a database for more convenient access and develop an R script to access the database programmatically.



At the workshop, we learned that:

- Polar volume and bird profile data on the BALTRAD FTP server are already being backed up on [Beehub](#), but these can only be accessed via a login.
- The [bioRad](#) package by Adriaan Dokter already includes functionality to read bird profile data files.

In consultation with the participants of the workshop, we decided to:

- Still setup a data repository for bird profile data on Amazon S3. This then clearly separates the workflow for private polar volume data - for which a data use agreement with the OPERA network is needed and which are only accessible via login on Beehub - from the workflow for public bird profile data - which will be accessible as open data on Amazon S3 and could be described in a data paper.
- Not to develop a data model and database for bird profile data, since there was no strong use case to do so, but rather extend the bioRad package to download and load bird profile data from the data repository on S3.

- To customize the file browser website so it includes some information on the coverage of the data over time.

Tuesday 17 January 2017

- **Liesbeth:** Provide details about already existing data infrastructure and processing on Baltrad and production pipeline to Beehub; participate in the Western European flyway workshop.
- **Peter:** Document transfer procedure from BALTRAD FTP to Amazon S3 and test <https://github.com/rufuspollock/s3-bucket-listing> as file browser.
- **Stijn:** Initialize S3 bucket and develop script to connect to Baltrad FTP.

Wednesday 18 January 2017

- **Liesbeth:** Participate in the Western European flyway workshop.
- **Peter:** Customize <https://github.com/rufuspollock/s3-bucket-listing>: layout with Bootstrap, update and clean Javascript, use an HTML table for the file listing, and add breadcrumbs for navigation. Present progress of STSM at workshop.
- **Stijn:** Add functionality to parse file names (required for directory structure on file repository) and to initialize script. Add documentation and reporting.

Thursday 19 January 2017

- **Liesbeth:** Develop R script to search and read bird profile data in bioRad.
- **Peter:** Customize <https://github.com/DKirwan/calendar-heatmap> to visualize data coverage: allow to set start date, define min/max values and correctly position tooltips.
- **Stijn:** Add functionality to generate coverage data for coverage visualization and develop R script to search and read bird profile data in bioRad.

Friday 20 January 2017

- **Liesbeth:** Write example vignettes to search and read bird profile data and radar volume with bioRad and cleanup repository.
- **Peter:** Include s3-bucket-listing and calendar-heatmap functionality in the data repository. Submit new functionality as a pull request to the original calendar-heatmap library, so it can be used by anyone (<https://github.com/DKirwan/calendar-heatmap/pull/25>).
- **Stijn:** Add functionality to generate monthly zip files in the data repository.

Description of the main results

Data repository for bird profile data

ENRAM data repository for bird profiles detected by 🌤️ 📡

Served from <https://lw-enram.s3-eu-west-1.amazonaws.com>



You're here: [🏠](#) / [nl](#) / [dbi](#) / 2017

Name	Last modified	Size
↩		
nl/dbi/201701.zip	2017-01-20T12:19:04.000Z	1.3 MB

01/

A proof of concept of a data repository for ENRAM bird profile data has been deployed at <http://enram.github.io/data-repository>. It includes:

- An easy to navigate directory tree, structured as country > radar > year > month > day > hour. Each hour directory contains the original bird profile h5 files generated by the vol2bird algorithm. Each directory has its own URL so it can be referenced and shared (e.g. <http://enram.github.io/data-repository/?prefix=be/jab/2016/11/20/23/>). Breadcrumbs navigation is provided for reference and easier navigation up the directory tree.
- Zip files containing all data from one radar for one month for more convenient download. These files are generated each time new data is added and are available from the year directories (e.g. <http://enram.github.io/data-repository/?prefix=be/jab/2016/>).
- A calendar heatmap to visualize data coverage for one radar for one year. It is available from the year directories (e.g. <http://enram.github.io/data-repository/?prefix=nl/dbi/2017/>) and is based on the expected files vs the actual files in the data repository.

Data transfer procedure from BALTRAD FTP to the data repository

A daily data transfer procedure from the BALTRAD server to the data repository has been developed. The procedure provides the following functional steps:

- Transfer newly available data files from the BALTRAD server to the data repository, i.e. place them in the appropriate folder structure on Amazon S3.
- Recalculate the monthly data availability for each of the radars, which serves as the source data for the calendar heatmap.
- For each of the months for which new data has been added, the corresponding zip-files are regenerated and updated on the data repository.

To accommodate the above procedure together with other custom data transfers in between the different entities, a [Python module](#) has been developed. The module enables a high-level access and data exchange between the BALTRAD FTP server, the Amazon S3 data repository and a local file directory.

Functionality and documentation to read bird profile data with bioRad

A function (`retrieve_vp_paths`) has been added to the package bioRad that allows for easy searching and, in combination with the already existing functions, reading of bird profiles from the hierarchical directory structure of the data.

To illustrate the use of the new and some existing plotting functions, two vignettes ([Guidelines to load bird profiles into bioRad](#) and [Guidelines to create PPI plots from polar volume files in bioRad](#)) with example code were written.

Future work/foreseen publications

- Data currently not available in the data repository should be uploaded. This includes **2016 autumn migration data**, which need to be reprocessed, and data currently only available in Beehub.
- The vol2bird algorithm deployed on BALTRAD currently does not generate a bird profile data file if there was no bird movement detected in the radar volume data. That means that - based on the files alone - **no distinction can be made between no bird movement and no data because of an error** in the file processing chain. It was discussed with Adriaan Dokter to output a log file for each input file, so the distinction can be made between the two, both in the data repository and the coverage visualization. Once these log files are available, the data transfer procedure needs to be updated to interpret these correctly.

- The data transfer procedure currently runs in development modus. Once the data flow is stable, this should be **deployed in production**.
- It was suggested by the participants of the workshop that the ENRAM data repository and data flow could be described and published in an open access **data paper**, once these are fully functional, stable and documented.