

GloBAM Data Management Plan

Peter Desmet

2019-04-23

Contents

1	Intro	2
1.1	Distributions	2
1.2	How this DMP is maintained	2
1.3	Citation	2
2	European radar data (pvol)	2
2.1	Source	2
2.2	License	3
2.3	Format	3
2.4	Geographical scope	3
2.5	Temporal scope	4
2.6	Historical archive	4
3	US radar data (pvol)	4
3.1	Source	5
3.2	License	5
3.3	Format	5
3.4	Geographical scope	5
3.5	Temporal scope	5
4	Vertical profiles of aerial migrants data (vp)	5
4.1	Source	5
4.2	License	5
4.3	Format	5
4.4	Geographical scope	5
4.5	Temporal scope	8
5	Time series of vertical profiles data (vpts)	8
5.1	License	8
5.2	Source	8
5.3	Coverage	8
5.4	Format	8
6	Processing pipeline	8
6.1	European pvol	9
6.2	European pvol to vp	9
6.3	vp archiving	9
6.4	vp to vpts	9
7	External data	9
7.1	Climate	10
7.2	Wind	10
7.3	Land use	10
7.4	Artificial light	10
7.5	Wind energy installations	10

1 Intro

Welcome to the data management plan (DMP) for the project *GloBAM: Towards monitoring, understanding and forecasting global biomass flows of aerial migrants*.

1.1 Distributions

- Website: <https://enram.github.io/globam-dmp/>
- PDF: <https://enram.github.io/globam-dmp/globam-dmp.pdf>

1.2 How this DMP is maintained

1. This DMP is maintained and versioned on GitHub at <https://github.com/enram/globam-dmp/>.
2. Each chapter is an R Markdown file (`Rmd`) in the `src` directory of that GitHub repository. You can access it directly by clicking the pencil icon in the top navigation of this website.
3. Changes to the R Markdown files can be made by contributors to the GitHub repository or suggested by anyone as pull requests. Textual changes can be done directly on GitHub, code changes are better tested in RStudio first.
4. Accepted changes (i.e. changes to the `master` branch) will trigger an automatic build procedure that will generate a new version of this DMP using the R package bookdown. The date of the build is used as the version number.

1.3 Citation

Desmet P (2019) GloBAM Data Management Plan. Version 2019-04-23. <https://enram.github.io/globam-dmp/>

2 European radar data (`pvol`)

This chapter describes polar volume (`pvol`) data from European weather radars.

2.1 Source

`pvol` data are collected by **national weather services** in Europe. They have different approaches in processing, archiving and providing access to these data. Some **research labs** (e.g. UvA) also have archived subsets of these `pvol` data for research use.

The best source for consolidated `pvol` data however is **BALTRAD**, managed by SMHI (contact person: Günther Haase). `pvol` data are archived there, but access to that server is restricted to SMHI. See 6.1 for details on how `pvol` data are submitted to this archive.

The rest of this chapter discusses the specifics of the BALTRAD archive.

2.2 License

European radar data exchange is coordinated by the *European Operational Program for Exchange of Weather Radar Information (EUMETNET/OPERA)*. GloBAM has access to these data via the *EIG EUMETNET license agreement for Research use of OPERA data*, which grants a specific list of people (involved in ENRAM and GloBAM) access to radar data from selected countries under the following conditions (excerpt):

The Licensor grants the Licensees a non-Exclusive license to use intellectual property belonging to the EUMETNET members as defined below [...]

The OPERA Members Data will be provided for use within the ENRAM Project for the purpose of extracting animal migration information for scientific research.

The Grant of this License does not permit use of the OPERA Members Data licensed to be used for commercial purposes or exploitation for profit.

The license agreement allows us to access European `pvol` data for the sole purpose of extracting animal migration information for scientific research. Storing, sharing or other use of these data is **restricted** and requires prior written consent of the Licensor.

2.3 Format

The delivery of European `pvol` data is defined in the license agreement:

The Licensor will make the OPERA Members Data available to the Licensees in accordance to the following data description and technical specifications:

- a. Single-site polar volumes containing reflectivity, optionally also Doppler velocity, uncorrected reflectivity and dual-pol parameters. Data model ODIM as described in OPERA pages of the EUMETNET website (<http://www.eumetnet.eu>)
- b. Update frequency – 15 minutes
- c. Issue time – up to 120 minutes after data time
- d. Format: HDF5
- e. Delivery method: FTP via Internet using an ad hoc server. A username and password will be created for ENRAM.
- f. Availability of OPERA Members Data will be on the basis of reasonable endeavour
- g. Support: OPERA documents are available from the EUMETNET website's OPERA pages (<http://www.eumetnet.eu>). OPERA Program Manager will manage technical questions related to the OPERA Products but may address more complicated issues to a team member who may charge a fee.

Data is thus provided for every 15 minutes as `hdf5` in the OPERA ODIM_h5 format. These files can be read in `bioRad` with `read_pvolfile()`.

2.4 Geographical scope

OPERA manages a list of radar metadata (232 radars). This **OPERA database** is available as an interactive map derived from a `json` file containing the metadata. To keep track of changes, we archive this `json` on GitHub every time we update the DMP.

The license agreement allows access to data from 19 countries:

##	country	iso_code	radars	operational
##	Austria	AT	5	5
##	Belgium	BE	3	3
##	Croatia	HR	5	2
##	Czechia	CZ	2	2

##	Denmark	DK	5	5
##	Estonia	EE	2	2
##	Finland	FI	10	10
##	France	FR	31	25
##	Germany	DE	20	20
##	Netherlands	NL	3	2
##	Norway	NO	11	11
##	Poland	PL	8	8
##	Portugal	PT	4	3
##	Slovakia	SK	4	4
##	Slovenia	SI	2	2
##	Spain	ES	15	15
##	Sweden	SE	12	12
##	Switzerland	CH	5	5
##	United Kingdom	GB	16	16
##	Total	-	163	152

The BALTRAD archive does not contain `pvol` data for all of these radars/countries however (see 6.1). Its precise geographical scope is currently **unknown**, but being gathered in this spreadsheet.

2.5 Temporal scope

The temporal scope of the BALTRAD archive is currently **unknown**.

2.6 Historical archive

GloBAM needs a **2-3 year archive** of European `pvol` and derived `vp` data to tackle its research questions. By starting from a `pvol` archive, the generation of `vp` data can be reproduced if need be, e.g. to make sure adequate `vol2bird` settings/versions are used.

The `pvol` archive at BALTRAD is likely to be the best source for this archive, but its scope and quality need to be assessed before we can proceed. Tasks for this are listed here and include:

1. Get a file listing for `pvol` archive
2. Select subsets: first 2 days of data per radar/month for 2016, 2017, 2018 (72 days): either `pvol` or pre-merged scans
3. Transfer subsets to accessible FTP server
4. Merge to `pvol` (test 1)
5. Process with `vol2bird` to `vp` (test 2)
6. Store output `vp` files
7. Visual control of `vp` (test 3)

3 US radar data (`pvol`)

This chapter describes polar volume (`pvol`) data from weather radars in the United States.

3.1 Source

3.2 License

3.3 Format

3.4 Geographical scope

3.5 Temporal scope

4 Vertical profiles of aerial migrants data (vp)

This chapter describes vertical profiles of aerial migrants/biological signals (**vp**) data for Europe and the United States. Vertical profiles contain the speed, direction and density of aerial migrants at different altitudes for a specific radar location and time, and form the source data for GloBAM specifically and aeroecological research in general.

4.1 Source

European vp data are generated daily from **pvol** data on the BALTRAD server and then transferred to the public ENRAM data repository (see 6.3). **United States vp** data are being generated for the 25 year **pvol** archive, but not yet publicly archived.

vp files can be downloaded from the ENRAM data repository individually, as monthly zips, or automatically in **bioRad** with `download_vpfiles()`.

4.2 License

Data in the ENRAM data repository are available as **open data** under a Creative Commons Zero waiver.

For the **European data** it is recommended to acknowledge EUMETNET/OPERA in publications resulting from the use of these data as follows:

We acknowledge the European Operational Program for Exchange of Weather Radar Information (EUMETNET/OPERA) for providing access to European radar data, facilitated through a research-only license agreement between EUMETNET/OPERA members and ENRAM.

4.3 Format

A **vp** file is generated for each originating **pvol** file and thus has the same granularity. **vp** data are stored as **hdf5** files in the ODIM bird profile format specification. These files can be read in **bioRad** with `read_vpfiles()`.

4.4 Geographical scope

The coverage of the ENRAM data repository is recorded daily in `coverage.csv` and summarized here:

##	radar	2016	2017	2018	2019	Total
##	bejab	21	NA	194	111	326
##	bewid	19	12	205	107	343
##	bezav	21	NA	NA	NA	21
##	bgvar	21	NA	NA	NA	21

##	ctcdv	21	NA	NA	NA	21
##	ctpda	21	NA	NA	NA	21
##	czbrd	21	NA	286	111	418
##	czska	21	NA	286	112	419
##	deasb	NA	NA	221	93	314
##	deboo	31	17	305	93	446
##	dedrs	31	21	304	93	449
##	deeis	31	21	304	93	449
##	deemd	NA	20	49	NA	69
##	deess	31	21	304	93	449
##	defbg	31	NA	NA	NA	31
##	defld	31	21	303	92	447
##	deflg	NA	21	49	NA	70
##	dehnr	31	21	304	93	449
##	deisn	31	NA	NA	NA	31
##	demem	31	21	301	93	446
##	deneu	31	21	301	93	446
##	denhb	31	21	300	90	442
##	deoft	31	21	304	93	449
##	depro	31	21	302	93	447
##	deros	25	21	294	93	433
##	desna	NA	21	303	93	417
##	detur	29	21	303	93	446
##	deumd	31	21	304	93	449
##	dkbor	NA	34	85	112	231
##	dkrom	NA	38	85	112	235
##	dksin	NA	38	85	112	235
##	dkste	NA	38	85	112	235
##	dkvir	NA	27	85	112	224
##	eehar	NA	20	35	1	56
##	eesur	NA	20	35	NA	55
##	esalm	NA	38	287	112	437
##	esbad	NA	38	302	112	452
##	esbar	NA	38	303	112	453
##	escor	NA	38	287	112	437
##	eslid	NA	38	303	112	453
##	eslpa	NA	NA	14	NA	14
##	esmad	NA	38	301	112	451
##	esmal	NA	38	303	112	453
##	esmur	NA	38	302	112	452
##	espma	NA	38	303	112	453
##	essan	NA	38	302	112	452
##	essev	NA	36	300	112	448
##	esse	NA	38	303	112	453
##	esval	NA	38	294	112	444
##	eszar	NA	38	303	109	450
##	fianj	22	154	38	NA	214
##	fiika	22	153	39	NA	214
##	fikes	22	151	38	NA	211
##	fikor	22	146	38	NA	206
##	fikuo	22	156	38	NA	216
##	filuo	22	153	38	NA	213
##	fipet	22	99	NA	NA	121
##	fiuta	22	157	38	NA	217

##	fivan	22	157	39	NA	218
##	fivim	22	132	34	NA	188
##	frabb	21	38	320	102	481
##	frale	21	26	321	112	480
##	frave	21	NA	283	112	416
##	frbla	21	38	321	111	491
##	frbol	21	38	279	106	444
##	frbor	21	36	321	112	490
##	frbou	21	38	173	109	341
##	frcae	21	38	320	112	491
##	frche	21	38	317	112	488
##	frcol	21	38	312	111	482
##	frgre	21	38	320	107	486
##	frlep	21	38	320	107	486
##	frmcl	21	38	304	112	475
##	frmom	21	38	320	112	491
##	frmte	21	35	302	112	470
##	frnan	21	38	317	112	488
##	frnim	21	38	320	112	491
##	frniz	21	NA	281	112	414
##	fropo	21	38	241	59	359
##	frpla	21	38	315	112	486
##	frtou	21	37	305	112	475
##	frtra	21	38	320	112	491
##	frtre	6	38	316	112	472
##	frtro	21	NA	282	111	414
##	hrbil	NA	38	35	NA	73
##	hrosi	NA	38	35	NA	73
##	nldbl	22	4	NA	NA	26
##	nldhl	21	125	295	112	553
##	nlhrw	NA	21	294	112	427
##	plbrz	21	3	261	112	397
##	plgda	21	3	261	112	397
##	plleg	21	3	259	112	395
##	plpas	NA	3	259	112	374
##	plpoz	21	3	260	112	396
##	plram	21	3	260	112	396
##	plrze	21	3	222	112	358
##	plswi	21	3	260	112	396
##	ptfar	119	NA	NA	4	123
##	ptlis	NA	NA	NA	1	1
##	ptliz	122	NA	NA	NA	122
##	ptprt	113	NA	NA	NA	113
##	seang	21	137	310	100	568
##	searl	21	153	318	110	602
##	sease	21	16	NA	NA	37
##	sehem	NA	43	314	101	458
##	sehud	21	NA	NA	NA	21
##	sehuv	NA	36	295	103	434
##	sekaa	NA	NA	NA	6	6
##	sekir	21	144	318	112	595
##	sekk	21	157	171	NA	349
##	selek	21	118	208	99	446
##	sella	NA	NA	NA	6	6

##	selul	21	148	183	NA	352
##	seoer	NA	36	307	103	446
##	seosd	NA	142	311	101	554
##	seosu	13	NA	NA	NA	13
##	seovi	21	59	NA	NA	80
##	sevar	21	NA	NA	NA	21
##	sevox	NA	133	312	100	545
##	sevil	21	153	318	29	521
##	silis	21	35	304	106	466
##	sipas	21	35	311	76	443
##	skjav	NA	38	301	112	451
##	skkoj	NA	38	301	105	444
##	skkub	NA	NA	280	112	392
##	sklaz	NA	NA	280	112	392

4.5 Temporal scope

Data transfer to the ENRAM data repository become more or less operational in March 2018 (with a gap in July 2018). Data from 2016 were uploaded for the European flyway study Nilsson et al. 2018.

5 Time series of vertical profiles data (vpts)

This chapter describes time series of vertical profiles (**vpts**). **vpts** are **vp** data bundled in time series (without data loss) and are a more convenient way for downloading and reading this type of data.

5.1 License

Not yet defined, but likely as **open data** under a Creative Commons Zero waiver.

5.2 Source

These files are not yet generated, but will be stored in the ENRAM data repository.

5.3 Coverage

Not yet defined, but likely the same as the **vp** data.

5.4 Format

Not yet defined, but likely text format (**txt**, **json**, **csv**) and tabular. See this issue.

6 Processing pipeline

This chapter describes the processing pipeline from **pvol** over **vp** to **vpts** data.

6.1 European pvol

BALTRAD collects `pvol` data for Europe. For `pvol` data to be available in BALTRAD, several conditions need to be met:

1. A country should send both reflectivity and radial velocity data to the OPERA data centre, which is called ODYSSEY. While many countries are sending reflectivity data, radial velocities are still unavailable for many countries.
2. ODYSSEY should forward these data to the BALTRAD datahub. It is standard policy to do so, but in practice data is not yet forwarded correctly for some radars/countries.
3. Currently BALTRAD and ODYSSEY store data at a 15 minute interval, higher resolution data is not yet available.

6.2 European pvol to vp

A server at BALTRAD processes any incoming `pvol` data with `vol2bird`. This pipeline is maintained by Günther Haase (SMHI).

The resulting `vp` files are stored for 2 days on a private FTP server to which we have access. The used `vol2bird` version is stored in the metadata of the `vp` files and can differ over time. The latest files are created with `vol2bird 0.3.20` (last tested April 11, 2019).

6.3 vp archiving

European `vp` data are archived daily by a pipeline running on Amazon Web Services (AWS) which transfers `vp` files from the BALTRAD FTP server to a public S3 bucket. This pipeline is maintained by Stijn Van Hoey (Open science lab for biodiversity) and its code and documentation are available in this repository. Note that any issues in the pipeline that are not resolved within 2 days (the time `vp` data are kept on the FTP server) **can result in data loss**.

The pipeline also updates a coverage file and bundles `vp` files in monthly zips per radar (e.g. `bejab201904.zip`). A public website allows easier file navigation of the S3 bucket.

United States `vp` data are currently not archived. See issue.

6.4 vp to vpts

Not yet defined, but likely part of the ENRAM pipeline (see 6.3).

7 External data

This chapter describes the external data required by GloBAM.

7.1 Climate

7.2 Wind

7.3 Land use

7.4 Artificial light

Needed by WP3.

7.4.1 Sources

Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band provided by NOAA. See that page for a description, filenaming conventions, data types/formats and providing credit.

We *might* also use the new Black Marble product from NASA once it is released.

7.5 Wind energy installations

Needed by WP4 (Judy Shamoun-Baranes).

7.5.1 Sources

Depends on the exact research questions and what is available. Will be updated in DMP once known.

8 Open data publication

This chapter describes the data of long term value that GloBAM will generate.

Not yet defined, but from the proposal:

The produced data will also be archived yearly as open data in an online research repository such as Zenodo (operated by CERN) where they will be assigned a Digital Object Identifier (DOI). These open data will be released under a Creative Commons Zero (CC0) waiver; formatted following open domain standards such as the ODIM bird profile specification; and documented with metadata describing their contributors, provenance, resolution, temporal and spatial coverage, and how to use these with open source software developed or contributed to by this project. To help users discover the data generated through this project, we will also develop an easy-to-use interface for the exploration of data at different spatial and temporal scales and offer derived data products that are ready to use in analyses. Materials supporting publications, such as software scripts and derived data products, will be deposited in open research repositories as well.