

# GloBAM Data Management Plan

*Peter Desmet*

*Cecilia Nilsson*

*Adriaan Dokter*

*2019-06-29*

## Contents

<b>Colophon</b>	<b>2</b>
<b>1 Introduction</b>	<b>2</b>
1.1 Research purpose . . . . .	2
1.2 Data manager . . . . .	3
1.3 How this DMP is maintained . . . . .	3
<b>Source data</b>	<b>3</b>
<b>2 European radar data (pvol)</b>	<b>3</b>
2.1 Data collecting . . . . .	3
2.2 Access . . . . .	3
2.3 License . . . . .	3
2.4 Geographical scope . . . . .	4
2.5 Temporal scope . . . . .	4
2.6 Format & file names . . . . .	5
2.7 Quality assurance . . . . .	5
2.8 Storage & backup . . . . .	5
<b>3 US radar data (pvol)</b>	<b>6</b>
3.1 Data collecting . . . . .	6
3.2 Access . . . . .	6
3.3 License . . . . .	6
3.4 Geographical scope . . . . .	6
3.5 Temporal scope . . . . .	6
3.6 Format & file names . . . . .	6
3.7 Quality assurance . . . . .	8
3.8 Storage & backup . . . . .	8
<b>4 Other external data</b>	<b>8</b>
4.1 Wind . . . . .	8
4.2 Land use . . . . .	8
4.3 Artificial light . . . . .	8
4.4 Wind energy installations . . . . .	9
<b>Produced data</b>	<b>9</b>
<b>5 Processing pipeline</b>	<b>9</b>
5.1 European pvol . . . . .	9
5.2 European pvol to vp . . . . .	9
5.3 vp archiving . . . . .	9

5.4	vp to vpts . . . . .	10
<b>6</b>	<b>Vertical profiles of aerial migrants data (vp)</b>	<b>10</b>
6.1	Source . . . . .	10
6.2	License . . . . .	10
6.3	Format . . . . .	10
6.4	Geographical scope . . . . .	10
6.5	Temporal scope . . . . .	13
<b>7</b>	<b>Time series of vertical profiles data (vpts)</b>	<b>13</b>
7.1	License . . . . .	13
7.2	Source . . . . .	13
7.3	Coverage . . . . .	13
7.4	Format . . . . .	13
	<b>Publishing &amp; preservation</b>	<b>13</b>
<b>8</b>	<b>Open data</b>	<b>13</b>

## Colophon

Distributions of this data management plan:

- Latest version (website): <https://enram.github.io/globam-dmp/>
- Latest version (pdf): <https://enram.github.io/globam-dmp/globam-dmp.pdf>
- Periodically archived version (pdf): *to be deposited on Zenodo.org*

Suggested citation for the latest version:

Desmet P, Nilsson C, Dokter A (2019) GloBAM Data Management Plan. Version 2019-06-29.  
<https://enram.github.io/globam-dmp/>

All versions are licensed under a Creative Commons Attribution License (CC BY 4.0).

## 1 Introduction

### 1.1 Research purpose

Migratory animals play significant roles in shaping ecosystems through a variety of transport and trophic effects that also represent services and disservices to human infrastructure, agriculture and welfare. Their aerial and terrestrial habitats have changed dramatically over the past decades and are expected to change further, particularly due to rapid climate change, increased urbanization, wind energy installations, and habitat fragmentation.

Within *GloBAM: Towards monitoring, understanding and forecasting global biomass flows of aerial migrants*, we aim to use weather radar data to quantify the biomass flows of aerial migrants (birds, insects and bats) from regional to continental scales across Europe and North America, over time-scales from days to years. We are particularly interested in identifying the drivers of migrant movements and abundances and will relate the timing and intensity of movements to a suite of atmospheric, climatic and landscape variables, exploring the implications for aerial migrants in a changing world.

For more information on GloBAM, see the project website.

## 1.2 Data manager

Peter Desmet is responsible for data management and DMP maintenance in GloBAM. He is also leading *WP1 - Data infrastructure*. Peter works as open data coordinator for the Open science lab for biodiversity at the Research Institute for Nature and Forest (INBO). He and his team have extensive expertise and experience with data management and open data publication meeting FAIR principles.

## 1.3 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 the website version of this DMP.
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 the DMP using the R package bookdown. The date of the build is used as the version number.

# Source data

## 2 European radar data (`pvol`)

This chapter describes polar volume (`pvol`) data from European weather radars: one the main data sources for GloBAM.

### 2.1 Data collecting

European `pvol` data are collected by weather radars from **national weather services** in Europe.

### 2.2 Access

The national weather services have different approaches in processing, archiving and providing access to `pvol` data (from restricted access to open data). Even though some GloBAM partners (e.g. the University of Amsterdam) have archived subsets of these data for research use, the scattered scope and access to European `pvol` data is a serious barrier to GloBAM, which ideally requires access to all European `pvol` data using the same protocol.

The best source for consolidated `pvol` data is therefore **BALTRAD**, managed by SMHI (contact person: Günther Haase). `pvol` data are archived there, but access to that server is restricted to SMHI. See 5.1 for details on how `pvol` data are submitted to this archive.

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

### 2.3 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, which is what GloBAM intends to do. Storing, sharing or other use of these data is **restricted** and requires prior written consent of the Licensor.

## 2.4 Geographical scope

OPERA manages a list of radars, their unique codes (e.g. `odimcode`) and associated metadata. The list currently contains 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 5.1). Its precise geographical scope is currently **unknown**, but being gathered.

## 2.5 Temporal scope

The temporal scope of the BALTRAD archive is currently **unknown**, but being gathered. 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, see 2.7.

## 2.6 Format & file names

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()`.

File names are generally of the format `bejab_pvol_20190620T055500Z_0x9.h5`, where:

- `be`: two-letter code (ISO 3166-1 alpha-2)
- `jab`: three-letter code (last 3 letters from the ODIM code, see 2.4)
- `_pvol_`: indication that this is a `pvol` data file
- `20190620`: date in `yyyymmdd` format
- `T055500Z`: time in `hhmmss` format + timezone (Z for UTC time)
- `_0x9`: suffix (to be ignored)

## 2.7 Quality assurance

The scope and quality of the BALTRAD archive need to be assessed. 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)

## 2.8 Storage & backup

The BALTRAD archive is stored on a restricted FTP server. It is maintained by SMHI.

## 3 US radar data (pvol)

This chapter describes polar volume (pvol) data from weather radars in the United States: one the main data sources for GloBAM.

### 3.1 Data collecting

US pvol data are collected by the National Oceanic and Atmospheric Administration (NOAA), which operates a network of approximately 160 Next Generation Weather Radar (NEXRAD), WSR-88D sites in the US and associated territories.

### 3.2 Access

Archived polar volumes (pvols) (called “level II data” in the US) are stored on various services:

1. Amazon S3, see NEXRAD on AWS. Archived data are hosted in the noaa-nexrad-level2 Amazon S3 bucket in the `us-east-1` AWS region. These data are made available 10-20 minutes lag relative to real-time. There is also a real-time data stream, for which scans have not been merged yet into polar volumes.
2. Google Cloud.

### 3.3 License

There are no restrictions on the use of these data.

### 3.4 Geographical scope

The NEXRAD network covers continental US, Alaska, Hawaii, Guam, South Korea and Puerto Rico. Sites in use vary slightly over time, see [here](#) for more information, [maps here](#).

### 3.5 Temporal scope

The archive spans from June 1991 to present. However, during this time significant changes have been made. A resolution increase was implemented in 2008-2009 (so-called superresolution data). Dual polarization capabilities were added in 2010-2013.

### 3.6 Format & file names

See [this page](#) for available decoding tools of the radar format. US pvol data can be read in `bioRad` with `read_pvolfile()`.

From the AWS open data project documentation for archive data:

Each volume scan file of archival data is available as an object in Amazon S3. The basic data format is:

```
/<Year>/<Month>/<Day>/<NEXRAD Station>/<filename>
```

Where:

- `<Year>` is the year the data was collected

## NEXRAD Coverage Below 10,000 Feet AGL

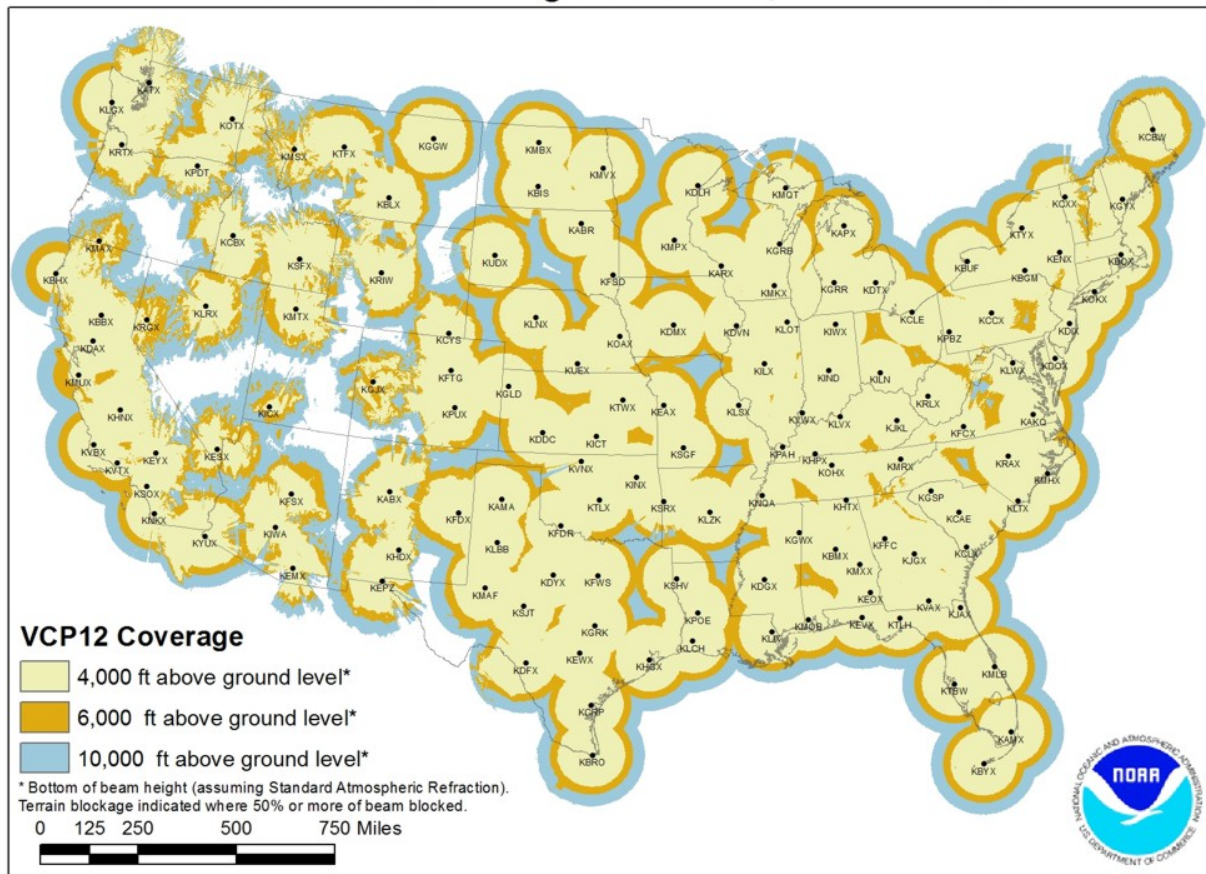


Figure 1: NEXRAD radar network in the lower 48 states

- <Month> is the month of the year the data was collected
- <Day> is the day of the month the data was collected
- <NEXRAD Station> is the NEXRAD ground station (map of ground stations)
- <filename> is the name of the file containing the data. These are compressed files (compressed with gzip). The file name has more precise timestamp information.

All files in the archive use the same compressed format (.gz). The data file names are, for example, KAKQ20010101\_080138.gz. The file naming convention is:

GGGGYYYYMMDD\_TTTTTT

Where:

- GGGG = Ground station ID (map of ground stations)
- YYYY = year
- MM = month
- DD = day
- TTTTTT = time when data started to be collected (GMT)

Note that the 2015 files have an additional field on the file name. It adds “\_V06” to the end of the file name. An example is KABX20150303\_001050\_V06.gz.

### 3.7 Quality assurance

Unknown.

### 3.8 Storage & backup

US pvo1 data are stored on AWS and other cloud services, see 3.2.

## 4 Other external data

This chapter describes the external data required by GloBAM.

### 4.1 Wind

### 4.2 Land use

### 4.3 Artificial light

Needed for WP3.

#### 4.3.1 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 may also use the new Black Marble product from NASA (not yet released).



## 4.4 Wind energy installations

Needed for WP4 (Judy Shamoun-Baranes).

### 4.4.1 Sources

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

# Produced data

## 5 Processing pipeline

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

### 5.1 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.

### 5.2 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).

### 5.3 `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.

## 5.4 vp to vpts

Not yet defined, but likely part of the ENRAM pipeline (see 5.3).

# 6 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.

## 6.1 Source

**European vp** data are generated daily from **pvo1** data on the BALTRAD server and then transferred to the public ENRAM data repository (see 5.3). **United States vp** data are being generated for the 25 year **pvo1** 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()`.

## 6.2 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.

## 6.3 Format

A **vp** file is generated for each originating **pvo1** 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()`.

## 6.4 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	173	388
##	bewid	19	12	205	171	407
##	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	172	479
##	czska	21	NA	286	174	481
##	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	176	295
##	dkrom	NA	38	85	175	298
##	dksin	NA	38	85	176	299
##	dkste	NA	38	85	175	298
##	dkvir	NA	27	85	176	288
##	eehar	NA	20	35	1	56
##	eesur	NA	20	35	NA	55
##	esalm	NA	38	287	167	492
##	esbad	NA	38	302	167	507
##	esbar	NA	38	303	168	509
##	escor	NA	38	287	167	492
##	eslid	NA	38	303	167	508
##	eslpa	NA	NA	14	NA	14
##	esmad	NA	38	301	167	506
##	esmal	NA	38	303	168	509
##	esmur	NA	38	302	167	507
##	espma	NA	38	303	166	507
##	essan	NA	38	302	167	507
##	essev	NA	36	300	164	500
##	essse	NA	38	303	161	502
##	esval	NA	38	294	167	499
##	eszar	NA	38	303	163	504
##	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	166	545
##	frale	21	26	321	175	543
##	frave	21	NA	283	176	480
##	frbla	21	38	321	167	547

##	frbol	21	38	279	140	478
##	frbor	21	36	321	176	554
##	frbou	21	38	173	173	405
##	frcae	21	38	320	175	554
##	frche	21	38	317	166	542
##	frcol	21	38	312	175	546
##	frgre	21	38	320	171	550
##	frlep	21	38	320	166	545
##	frmcl	21	38	304	175	538
##	frmom	21	38	320	176	555
##	frmte	21	35	302	161	519
##	frnan	21	38	317	175	551
##	frnim	21	38	320	176	555
##	frniz	21	NA	281	176	478
##	fropo	21	38	241	59	359
##	frpla	21	38	315	176	550
##	frtou	21	37	305	176	539
##	frtra	21	38	320	176	555
##	frtre	6	38	316	176	536
##	frtro	21	NA	282	175	478
##	hrbil	NA	38	35	NA	73
##	hrosi	NA	38	35	NA	73
##	nldbl	22	4	NA	NA	26
##	nldhl	21	125	295	175	616
##	nlhrw	NA	21	294	175	490
##	plbrz	21	3	261	174	459
##	plgda	21	3	261	173	458
##	plleg	21	3	259	174	457
##	plpas	NA	3	259	174	436
##	plpoz	21	3	260	174	458
##	plram	21	3	260	174	458
##	plrze	21	3	222	175	421
##	plswi	21	3	260	174	458
##	ptfar	119	NA	NA	21	140
##	ptlis	NA	NA	NA	5	5
##	ptliz	122	NA	NA	NA	122
##	ptprt	113	NA	NA	59	172
##	seang	21	137	310	136	604
##	searl	21	153	318	123	615
##	sease	21	16	NA	NA	37
##	seatv	NA	NA	NA	11	11
##	sehem	NA	43	314	165	522
##	sehud	21	NA	NA	NA	21
##	sehuv	NA	36	295	166	497
##	sekaa	NA	NA	NA	70	70
##	sekir	21	144	318	176	659
##	sekkrr	21	157	171	NA	349
##	selek	21	118	208	160	507
##	sella	NA	NA	NA	69	69
##	selul	21	148	183	NA	352
##	seoer	NA	36	307	167	510
##	seosd	NA	142	311	163	616
##	seosu	13	NA	NA	NA	13
##	seovi	21	59	NA	NA	80

##	sevar	21	NA	NA	NA	21
##	sevax	NA	133	312	161	606
##	sevil	21	153	318	29	521
##	silis	21	35	304	170	530
##	sipas	21	35	311	139	506
##	skjav	NA	38	301	175	514
##	skkoj	NA	38	301	168	507
##	skkub	NA	NA	280	174	454
##	sklaz	NA	NA	280	176	456

## 6.5 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.

## 7 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.

### 7.1 License

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

### 7.2 Source

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

### 7.3 Coverage

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

### 7.4 Format

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

# Publishing & preservation

## 8 Open data

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.