



Using Darwin Core
to make tracking data discoverable
via the
Atlas of Living Australia

Peggy Newman

Software Developer, ZoaTrack / Atlas of Living Australia

Biodiversity e-Infrastructure

- Australian region (broad)
- Aggregator - not curator or collector
- Open data for re-use
- Australian node of GBIF
- observations & specimens

Contributors:

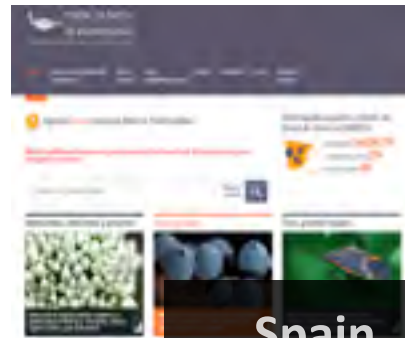
museums
herbaria
research institutions
government
citizen science

Users:

science research
ecology research
education
environmental assessment
conservation management/planning

Where does a species occur?
What species occur in an area?





Spain



France

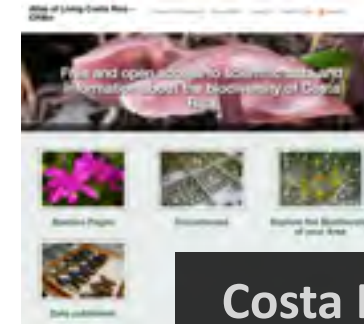


Scotland

ALA and GBIF facilitate other countries to develop platforms using ALA e-infrastructure.



Australia



Costa Rica



Brazil

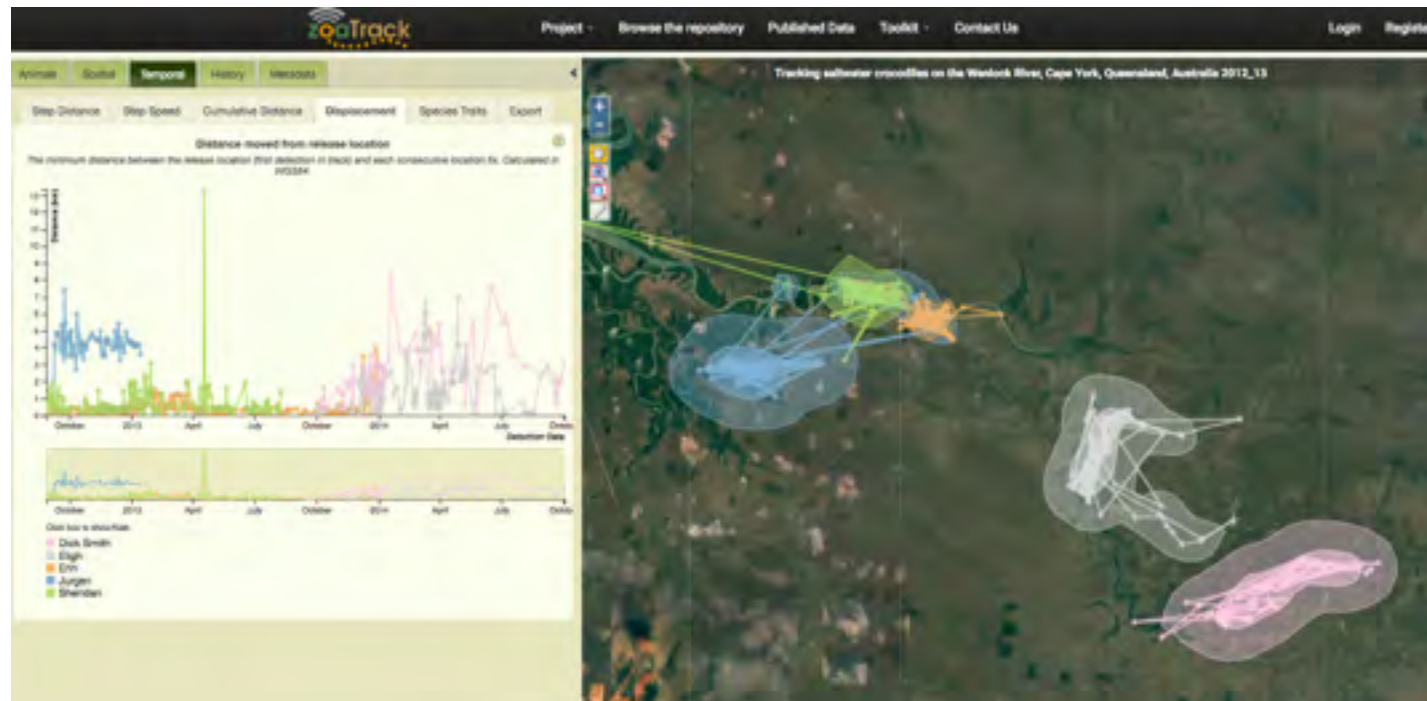


Argentina

More are being implemented or are under discussion.

ZoaTrack: separate application supported by ALA **specifically** for animal movement data

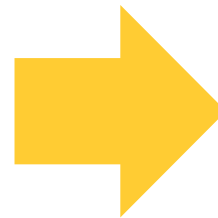
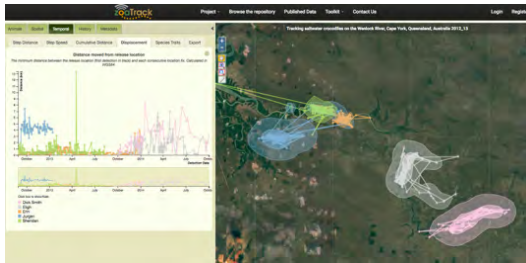
- analysis tools & visualisations (e.g., home range estimation)
- cleansing/filtering tools
- data embargoes to support a research data life cycle
- automated sensor feeds (e.g., Argos, Globestar)



Why share data between ZoaTrack and ALA?

- **Discoverability**
 - ALA is the go-to for Australia's biodiversity records
 - ALA is linked to GBIF and other Living Atlas platforms
- Tracking data is valuable, good quality, well described data that should be represented in the biodiversity record

How to share between such different databases?



- Multiple observations per individual
- Non standard data model
- Occurrence based datasets
- Simple Darwin Core

- ALA mainly uses Darwin Core - ingests flat files
- **Aggregate** each track into a single comprehensive record

Darwin Core Term	
basisOfRecord	MachineObservation
eventId	Identifier for the track (ie. sensor deployment on an animal). Foreign key to source system.
eventDate	Date range for first to last detection
eventRemarks	Sensor description
footprintWkt	Polygon representing track (home range, convex hull, alpha hull)
collectionId	URL link back to source system
decimalLatitude	Nominal occurrence (first detection)
decimalLongitude	Nominal occurrence (first detection)
measurement fields	Count of detections

Key Issues

- Discoverability – exposure - consistency
- General repository versus specialized
- Detailed Darwin Core record easily answers the two key questions
 - Where does this species/group occur?
 - What species occur in this area
- The difference between single occurrence records and tracks is made obvious to the users up front (no confounding)
- Ease of implementation

Next Steps

- TDWG involvement
- Upcoming Activities:
 - Darwin Core Hour discussion on Bio-logging and Camera Trap Data
 - TDWG 2018 Conference – workshop/symposium session on applying Darwin Core to Bio-logging data

WRAM & Swedish LifeWatch

Wireless Remote Animal Monitoring (WRAM)

The National Swedish Biotelemetry e-Infrastructure for sensor data from animals



Swedish LifeWatch (SLW)

The Swedish e-Infrastructure for Biodiversity & Ecosystem Research

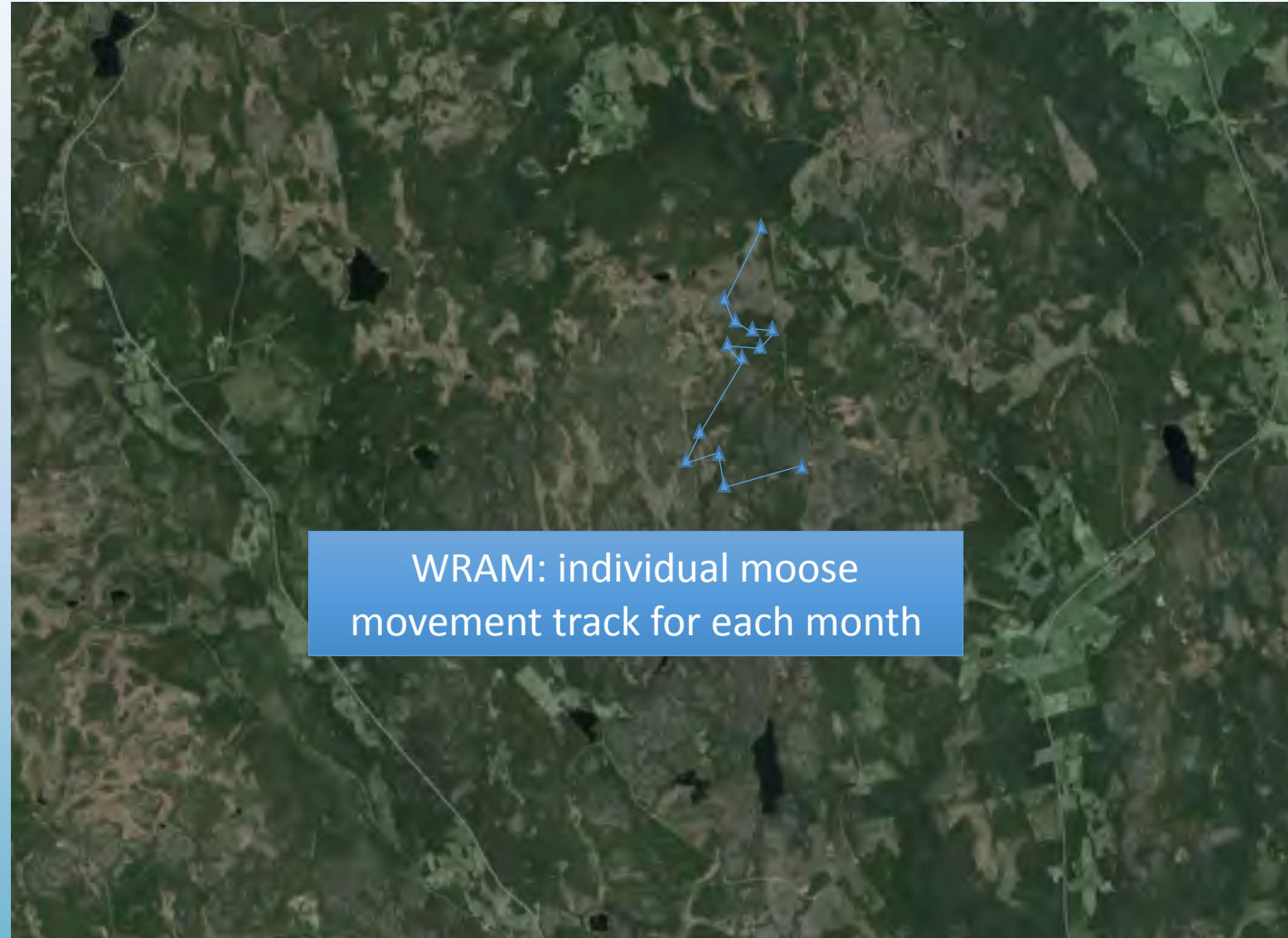


Task: WRAM delivers aggregated position data to SLW

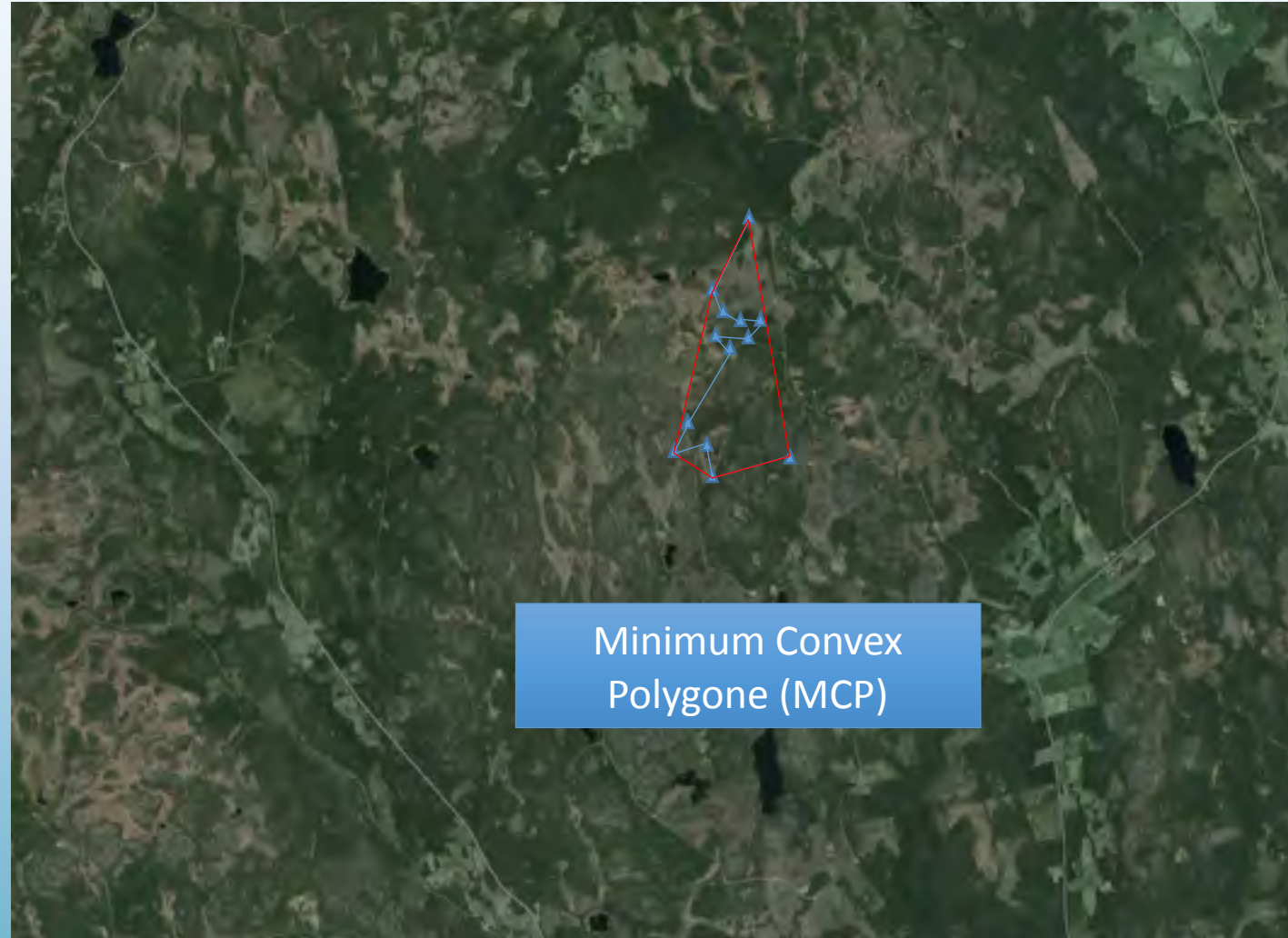
- WRAM currently contains e.g., 18.3M GPS records for 1,248 individual moose
- After aggregation 24,747 monthly 'occurrence records' with estimated position error can be delivered



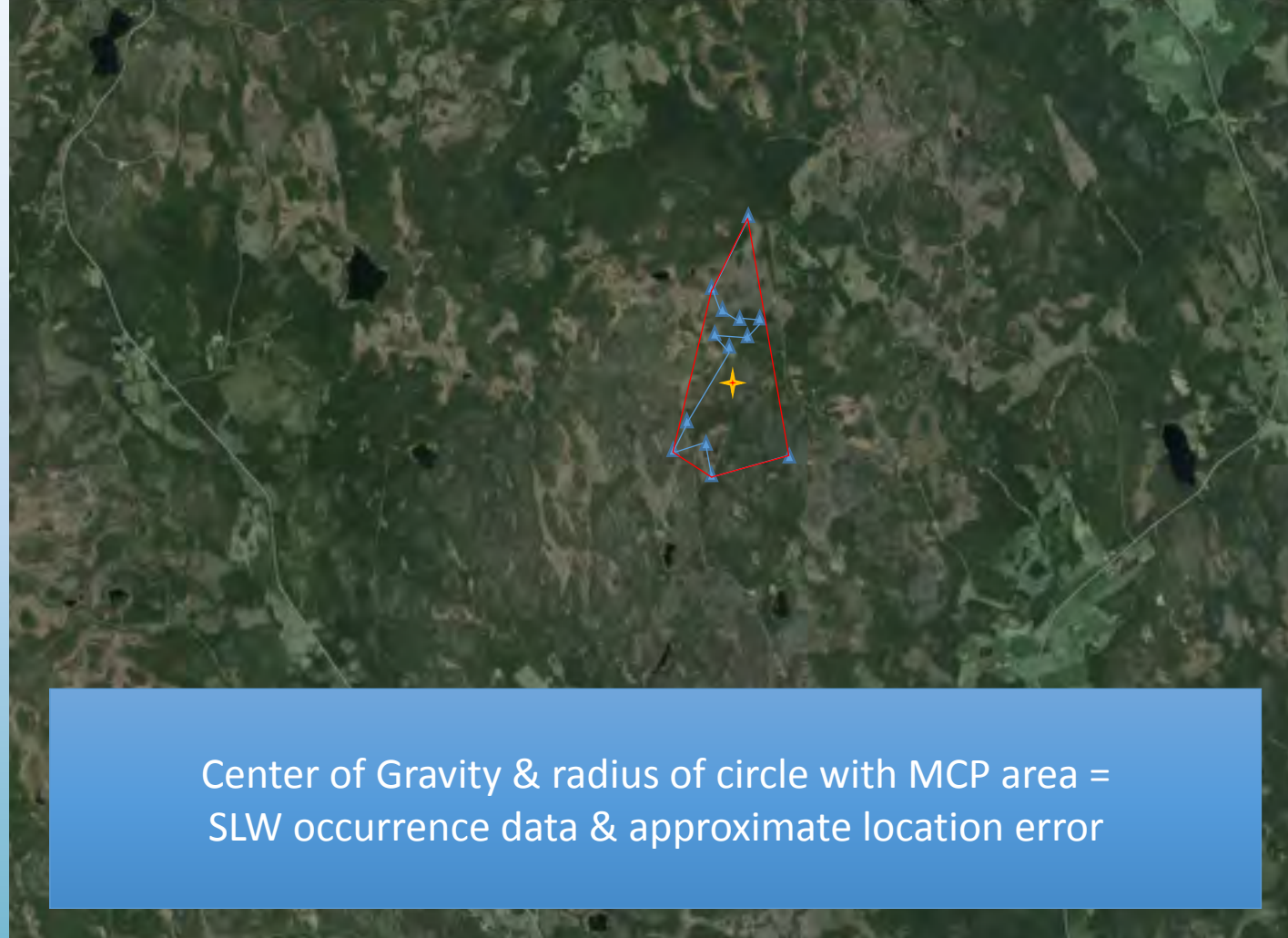
Aggregating data



Aggregating data



Aggregating data



Swedish LifeWatch - WRAM

The Analysis portal for biodiversity data

SWEDISH LIFEWATCH

Start Filter Calculation Format Result

Species observations provided by Swedish LifeWatch

Save settings Reset settings Download

<input checked="" type="checkbox"/>	data provider	Number of observations	Number of public observations
<input checked="" type="checkbox"/>	Species Observations System (Artportalen) (Swedish Species Information Centre (ArtDatabanken)) [?]	54,354,193	50,300,330
<input checked="" type="checkbox"/>	Observation database of Redlisted species (Swedish Species Information Centre (ArtDatabanken)) [?]	735,905	0
<input checked="" type="checkbox"/>	DNSA (The Swedish Museum of Natural History) [?]	20,487	20,486
<input checked="" type="checkbox"/>	MVM (Environmental data MVM, SLU) [?]	987,594	987,516
<input checked="" type="checkbox"/>	The National Register of Survey test-fishing (NORS) (Department of Aquatic Resources, SLU) [?]	2,675,474	2,675,474
<input checked="" type="checkbox"/>	The Database for electrofishing in streams (SERS) (Department of Aquatic Resources, SLU) [?]	397,719	394,470
<input checked="" type="checkbox"/>	Wireless Remote Animal Monitoring (WRAM) (Umeå Center for Wireless Remote Animal Monitoring (UC-WRAM), SLU) [?]	7,906	7,906
<input checked="" type="checkbox"/>	Shark SMH (SMH) [?]	557,958	557,958
<input checked="" type="checkbox"/>	The database for coastal fish (KUL) (Department of Aquatic Resources, Institute of Coastal Research, SLU) [?]	238,257	238,256
<input checked="" type="checkbox"/>	Lund Botanical Museum (LD) via GBIF (Lund University) [?]	354,949	353,644
<input checked="" type="checkbox"/>	Bird ringing centre in Sweden via GBIF (Swedish Museum of Natural History) [?]	4,822,723	4,769,322
<input checked="" type="checkbox"/>	Porpoises (NPM) via GBIF (Swedish Museum of Natural History) [?]	2,098	2,098
<input checked="" type="checkbox"/>	Herbarium of Oskarshamn (OH) via GBIF (Biological Museum in Oskarshamn) [?]	125,370	124,857
<input checked="" type="checkbox"/>	Herbarium of Umeå University (UME) via GBIF (Umeå University) [?]	65,763	65,685
	Sum	65,356,374	60,504,980

At least one data provider must be selected

My Settings

Data
Data Providers (14 selected) [\[?\]](#)

Filter
 Occurrence [\[?\]](#)

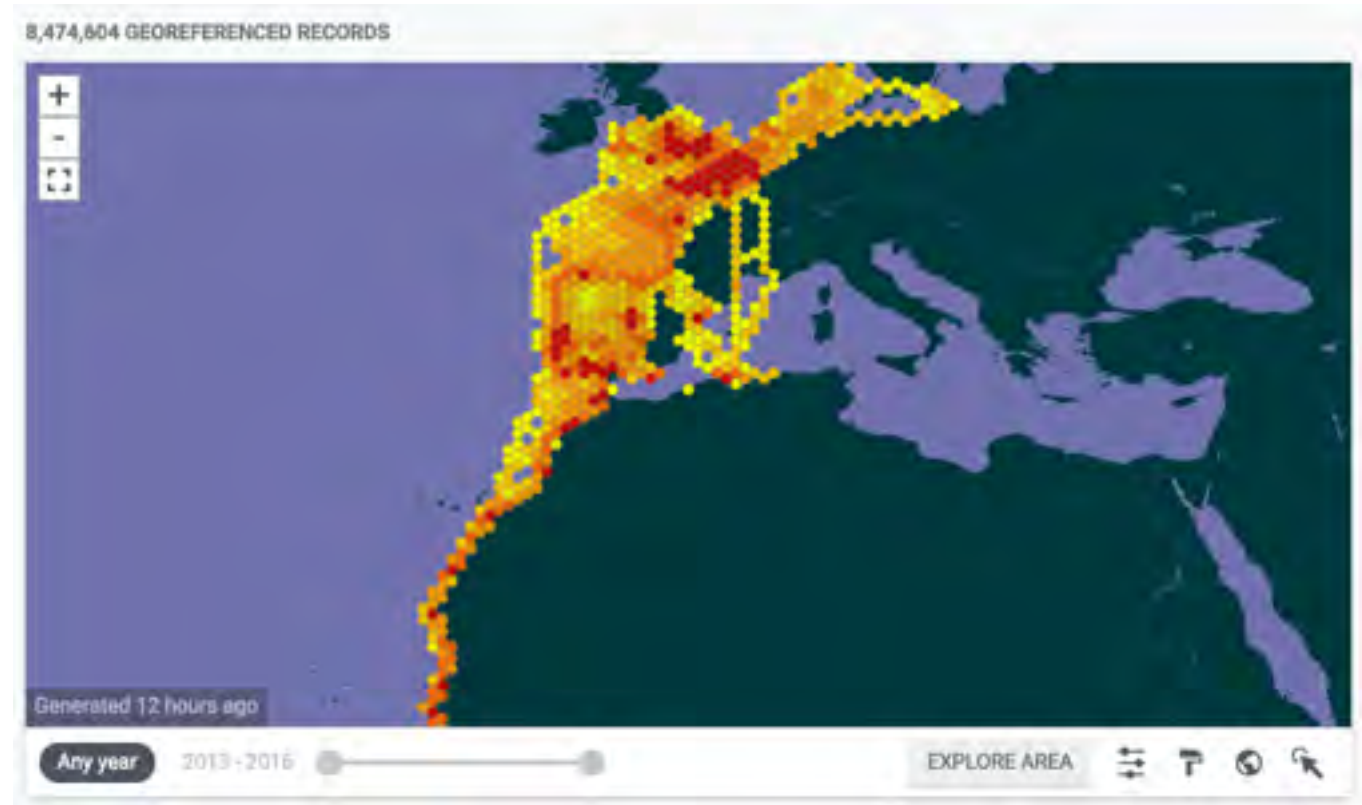
Calculation
 Grid statistics [\[?\]](#)
 Summary statistics [\[?\]](#)
 Time series [\[?\]](#)

Reset

INBO / GBIF Example: Lossless Data

Bird tracking - GPS tracking of Lesser Black-backed Gulls and Herring Gulls breeding at the southern North Sea coast

- 150 animals, ~8.5M records
- Each position is a Simple Darwin Core Occurrence record
 - Machine Observations
 - published to GBIF using IPT
 - Repetition, but volumes are not considered a problem



INBO / GBIF Example: Lossless Data

Darwin Core Mappings

<https://github.com/inbo/data-publication/blob/master/datasets/bird-tracking-gull-occurrences/mapping/dwc-occurrence.Rmd#transform-data>

Sample 100 Records

<https://github.com/inbo/data-publication/blob/master/datasets/bird-tracking-gull-occurrences/mapping/dwc-occurrence-100.csv>

```
# Map to Darwin Core
## Record terms
mutate[occurrenceID = paste(device_info_serial, format(date_time, "%Y%Md%MMNS"), sep = ":")] %>%
mutate[type = "Event"] %>%
mutate[language = "en"] %>%
mutate[license = "http://creativecommons.org/publicdomain/zero/1.0/"] %>%
mutate[rightHolder = "INBO"] %>%
mutate[accessRights = "http://www.inbo.be/en/norms-for-data-use"] %>%
mutate[datasetID = case_when(!$species_code %in% c("lbbg", "hg") ~ "https://doi.org/10.15468/82",
                             $species_code %in% c("lbbg", "hg") ~ "https://doi.org/10.15468/82") %>%
mutate[institutionCode = "INBO"] %>%
mutate[datasetName = case_when(!$species_code %in% c("lbbg", "hg") ~ "Bird tracking - GPS track",
                                $species_code %in% c("lbbg", "hg") ~ "Bird tracking - GPS track") %>%
mutate[basisOfRecord = "MachineObservation"] %>%
mutate[informationWithheld = "see metadata"] %>%
mutate[dynamicProperties = paste1("\device_info_serial:", device_info_serial, ", \catch_loc

## Occurrence terms
select[-sex, everything()] %>% # This field already exist, so just move it to correct place
mutate[lifeStage = "adult"] %>%

## Organism terms
mutate[organismID = ring_code] %>%
mutate[organismName = bird_name] %>%

## Event terms
mutate[samplingProtocol = "https://doi.org/10.1207/s10336-012-0908-1"] %>%
mutate[samplingEffort = paste1("{seconds_since_last_occurrence}", str_replace_na(calc_time_d
mutate[eventDate = format(date_time, "%Y-%m-%dT%H:%M:%S")] %>%

## Location terms
mutate[minimumElevationInMeters = as.integer(0)] %>%
mutate[minimumDistanceAboveSurfaceInMeters = as.integer(round(altitude, digits = 0))] %>%
mutate[decimalLatitude = sprintf("%.7f", round(latitude, digits = 7))] %>%
mutate[decimalLongitude = sprintf("%.7f", round(longitude, digits = 7))] %>%
mutate[geodeticDatum = "WGS84"] %>%
mutate[coordinateUncertaintyInMeters = as.integer(str_replace_na(round(h_accuracy, digits = 0),
mutate[georeferenceSources = "GPS"] %>%
mutate[georeferenceVerificationStatus = "unverified"] %>%

## Taxon terms
mutate[scientificName = scientific_name] %>%
mutate[countryCode = "Belgium"] %>%
```