# Detection Extract Documentation

OTN and all of its partner nodes create detection extracts on a semi-regular basis, every ~3 months, following a cross-node coordinated detection matching event known as a Data Push. These detection extract files contain only the detections for the year corresponding to the suffix on the file name.

Detection extracts for previous years are often updated when new, but historical, detection data / tag records are loaded into the database and are matched to your project. To see when each record was last added or changed, please refer to the DateLastModified column.

The detection extract CSV files can end up being very large, and great care should be taken when opening them. The files are often too big to be opened and analyzed easily in spreadsheet programs (i.e. Excel) and should instead be handled using programming languages such as R or Python. OTN maintains acoustic telemetry analysis packages designed to work with this data format available in:

R ( <https://gitlab.oceantrack.org/GreatLakes/glatos> ).

Python ( <https://gitlab.oceantrack.org/otndc/resonate> ).

# For Taggers

## Matched to Animals YYYY

These detections have been mapped to animals reported to OTN as having been tagged by your project and can originate from any receivers in the OTN network, including your own (if applicable). This file also includes a data point marking the initial release of each tagged animal that was detected in this year’s file (the receiver column will contain the value RELEASE for these entries).

Please note that we are now matching your tags to all OTN-compatible databases worldwide. This may result in biologically improbable matches across large spatial scales. If any detections we have matched to your animals should not be attributed to your animals, please let us know. There are OTN developed tools (Python: [resonATe](https://gitlab.oceantrack.org/otndc/resonate), R: [glatos](https://gitlab.oceantrack.org/GreatLakes/glatos)) which will assist with identifying false detections via impossible velocities. OTN provides support to researchers for the use of these tools. Unmatching detections that you find to be in error will help us match those detections to the correct researcher. Please advise your local node-manager ASAP when you have determined a match to be incorrect.

# For Array Operators

## Detections Mapped to Other Trackers YYYY

Also called “qualified” or “matched” detections. For projects with array deployments, these detection extracts contain detections which have been mapped to tag releases by other projects. This does not contain detections matched to your own project’s tags – those are provided in the Matched to Animals YYYY extract type. You can see a total list of tracking projects detected by your project by clicking on your project’s collectioncode at: <https://members.oceantrack.org/data/discovery/bycollection.htm>, or by contacting your local node manager.

Some node members will receive Detections Mapped to Other Trackers YYYY - (Extended Form). This “extended format” contains extra information as agreed upon in your network’s Data Policy. Currently, FACT is the only node who receives these detection extracts. Please contact your local node manager or [OTNDC@dal.ca](mailto:OTNDC@dal.ca) if you have questions about your Data Policy or would like to learn more about these other networks.

## Unqualified Detections YYYY

These are detections made by your project’s receivers that have not been matched to any tag, in any cooperating data node. There may be several reasons for this. One: It is a test or sentinel tag that was not reported to the data node. Two: It is an ambiguous tag which means we have more than one set of tag metadata which the detection could belong to. Three: We have not received any release metadata for the tag. Four: It is an old-style sensor tag and we have not been able to determine the associated pinger ID.

## Sentinel Tag Detections YYYY

These are detections of known non-animal tags on your array. These include identified range test tags, sentinel tag deployments, transceiver detections and any detections of tags which fall within the transmitter ID range specified by vendors to belong to test tags. These are separated from the animal tags for ease of analysis and to reduce the apparent number of unknown animal tags.

# Archived Detection Extract Formats

In the past, there were several other detection extract formats provided. If your project existed prior to 2019, you may see some of these in your project folder. They are no longer being updated.

## Matched to Animals on FACT YYYY

These detections have been mapped to animals tagged by your project and detected by any receiver project in the FACT network. These detections are also included in the Matched to Animals YYYY extract for that year (it is the “master” copy, and still being supported).

## Matched to Animals on Other Deployments YYYY

This extract format was only for projects with both tags and receivers. These detections have been mapped to animals tagged by your project and detected by any receiver project EXCEPT your own. This was created because some researchers pre-analyze detections on their own array and were only interested in finding out non-self matches. If this is something you wish to do, you can filter your Matched to Animals YYYY detection extract to remove self-detections. These detections are also included in the Matched to Animals YYYY extract for that year (it is the “master” copy, and still being supported).

# Data Dictionary

OTN and all partner nodes follow Darwin Core standards for reporting biological data. This includes using terms found in the accepted Darwin Core vocabularies, with the intention to facilitate the sharing of biodiversity data. See the glossary of terms here https://dwc.tdwg.org/terms/ . You will see many of these terms in our detection extracts.

## Matched to Animals YYYY

**collectioncode** - the unique code which your node uses to identify your project.

**catalognumber** - the unique identifier assigned to each tagged animal. This is provided in the animal\_id value from the tagger’s metadata, or if blank, the data system will automatically assign this value. Transmitters are often reused in multiple animals, so this value remains unique to each tag deployment when the tagname (see below for definition) would not.

**scientificname** - the scientific name as provided in the tagging metadata.

**commonname** - the common name as provided in the tagging metadata.

**datelastmodified** - the date this record was last changed. In the case of 'release' records this will be the date OTN loaded your metadata into the database. In the case of detection records this will be the date the detection was assigned to your animal.

**detectedby** - the project code for the receiver deployment which detected the animal.

**receiver\_group** - not used at this time but will have same value as column detectedby. Intended for use with mark/recapture studies but has not been implemented.

**station** - for 'release' records this will be blank. For detection records on fixed receivers it will be the name of the station or mooring where the receiver was located. This column is used in many OTN analysis tools for grouping detections, which occurred in the same place, over multiple deployments.

Detections from gliders, active tracking and satellite data will have a unique station name per detection location.

**receiver** - if the value here is 'release' this represents the record for the release location and will be the first record for a particular animal. If it is a detection on a receiver it will be the manufacturer’s serial number for the receiver. If it is a satellite detection it will be the manufacturer’s serial number for the transmitter.

**bottom\_depth** - the bottom depth where the receiver was deployed if provided in the receiver operator’s submitted receiver metadata.

**receiver\_depth** - the depth of the deployed receiver if provided in the receiver operator’s submitted receiver metadata.

**tagname** - the complete transmitter ID; the code space plus tag ID for acoustic tags.

For satellite detections this value will be the PTT of the satellite tag.

In the case of sensor tags which have a pinger and sensor ids with a different code space this will be the tag code for the pinger on that tag. The sensor name will be in another column.

**codespace** - for release records and satellite detections this will be NULL. For acoustic tags it will be the code space (ex. A69-1601).

**sensorname** - if this is a sensor detection it will be the tagname (full transmitter value: code space plus id) for the sensor transmitter.

**sensorraw** - if this is a sensor detection this will be the raw value (ex: ADC value for InnovaSea equipment) that has been exported from the manufacturer’s software after processing raw detection files.

**sensortype** - if this is a sensor detection AND OTN has the vendor specifications for this sensor tag, this will be the sensor type as it appears in the specifications, based on the serial number and tag ID.

**sensorvalue** - if this is a sensor detection AND OTN has the vendor specifications for this sensor tag, this will be the calculated sensor value using the slope and intercept provided in the specifications, based on the serial number and tag ID.

**sensorunit** - if this is a sensor detection AND OTN has the vendor specifications this will be the units value provided in the specifications, based on the serial number and tag ID.

**datecollected** - date and time of release or detection.

**timezone** - in OTN-style detection extracts, this value will always be 'UTC'.

**longitude** - release location or receiver location at time of detection in decimal degrees.

**latitude** - release location or receiver location at time of detection in decimal degrees.

**the\_geom** - previously called “st\_setsrid\_4326”. This is geometry – Open Geospatial Consortium’s WKB (well-known binary) formatted values for the point of detection. These are calculated using a PostGIS function; st\_makepoint(longitude, latitude). This geospatial information can be used in mapping programs.

**yearcollected** - year extracted from datecollected.

**monthcollected** - month extracted from datecollected.

**daycollected** - day extracted from datecollected.

**julianday** - Julian day number extracted from datecollected.

**timeofday** - time extracted from datecollected.

**datereleasedtagger** - for future use. Currently NULL.

**datereleasedpublic** - for future use. Currently NULL.

**local\_area** - description of release or receiver location as provided in submitted metadata.

**notes** - any comments which appear on the detection record or on the animal metadata record stored in the database.

**citation** - the project citation associated with the project in column detectedby. For release records this will be the tracker citation. For animal detections it will be the citation for the project which operates the detection-array.

**unqdetecid** - unique detection ID: an auto-generated unique value assigned to each record in the extract file. This column is used in OTN analysis tools to exactly refer to the record.

**contact\_poc** -point of contact for the project you have matched to, as provided in the project metadata. For taggers: the point of contact of the receiver operator.

**contact\_pi** -principal investigator for the project you have matched to, as provided in the project metadata. For taggers: the principal investigator of the receiver operator.

## Detections Mapped to Other Trackers YYYY, Unqualified Detections YYYY, and Sentinel Tag Detections YYYY

**basisofrecord** -type of detection record. Anything logged by a receiver will be MachineObservation.

**insitutioncode** -a shortform of the institution associated with your project, as determined by project metadata.

**collectioncode** - the unique code which your node uses to identify your project.

**datelastmodified** - the date this record was last changed. In the case of detection records this will be the date the detection was matched to an animal or when it was marked as a Sentinel detection.

**datecollected** - date and time of release or detection.

**timezone** - in OTN-style detection extracts, this value will always be 'UTC'.

**trackercode** - the project code for the tagged animal that has been matched to this detection. Only available for Qualified detections.

**fieldnumber -** fulltransmitter value (ie: tagname; code space plus ID) that was detected.

**catalognumber** - the unique identifier assigned to each tagged animal. This is provided in the animal\_id value from the tagger’s metadata, or if blank, the data system will automatically assign this value. Transmitters are often reused in multiple animals, so this value remains unique to each tag deployment when the fieldnumber would not. For Sentinel detections: this is the unique identifier assigned to this detection. Catalognumber is not available for Unqualified extracts.

**yearcollected** - year extracted from datecollected.

**monthcollected** - month extracted from datecollected.

**daycollected** - day extracted from datecollected.

**julianday** - Julian day number extracted from datecollected.

**timeofday** - time extracted from datecollected.

**station** - for detection records on fixed receivers it will be the name of the station or mooring where the receiver was located. This column is used in many OTN analysis tools for grouping detections, which occurred in the same place, over multiple deployments.

Detections from gliders, active tracking and satellite data will have a unique station name per detection location.

**collectornumber -** the manufacturer’s serial number for the receiver.

**latitude** - receiver location at time of detection in decimal degrees.

**longitude** - receiver location at time of detection in decimal degrees.

**rcvrcatnumber** -receiver catalog number: the unique identifier assigned to this receiver deployment. Often a combination of station name, instrument model and serial number, as well as deployment date and time.

**sensorname** - if this is a sensor detection it will be the tagname (full transmitter value: code space plus id) for the sensor transmitter. Only available for Qualified detections.

**sensorraw** - if this is a sensor detection this will be the raw ADC value that has been exported from VUE after processing VRL files. Only available for Qualified detections.

**the\_geom** - geometry – Open Geospatial Consortium’s WKB (well-known binary) formatted values for the point of detection. These are calculated using a PostGIS function; st\_makepoint(longitude, latitude). This geospatial information can be used in mapping programs.

**notes** - any comments which appear on the detection record or on the animal metadata record stored in the database.

**scientific\_name** - the scientific name of the animal detected, only provided for animals whose data has been made public by the tag-owner. This column will not appear if there are no animals detected which have public species information. Only available in Qualified detections.

**tag\_contact\_pi** - principal investigator for the project you have matched to, as provided in the project metadata. For receiver operators: the principal investigator of the tag owner. Only available for Qualified detections.

**tag\_contact\_poc** - point of contact for the project you have matched to, as provided in the project metadata. For receiver operators: the point of contact of the tag owner. Only available for Qualified detections.

**species** -column provided in *Detections Mapped to Other Trackers YYYY - (Extended Form)* containing information about the animal detected. This will only be populated if the project is also a FACT network member and has therefore agreed to share this information. Only available for Qualified detections.