Updated October 2021
DATA MANAGEMENT PLAN

I. Executive Summary

The Ocean Tracking Network (OTN) is a global infrastructure, monitoring, and data platform headquartered at Dalhousie University that aggregates animal telemetry data from multiple tagging systems to track the movements and survival of aquatic species worldwide. Information held by OTN's data system is funded by a broad spectrum of sources and represents a myriad of conservation interests. Researchers who depend on OTN for data quality and security assurance operate under a diverse range of data retention and security constraints. The objective of the data management plan is to satisfy these constraints while guaranteeing a pathway to public dissemination of all data held within the OTN system. It describes the obligations of the OTN Data Centre (OTNDC) to ensure the best quality information is shared with researchers and the public and the processes and workflows designed to guarantee the secure handling and timely dissemination of data stored and curated by the facility.

II. Purpose of the Plan

To provide a framework for managing the data generated by, shared with, and/or stored by OTN per internationally recognized best practices.

III. Goals and objectives

1. To maintain the confidence of funders, researchers, data users, and the public that the information is accurate, complete, reliable, and secure during and after the completion of the project. And to advance the principles of FAIR and TRUST in alignment with the Tri-Agency Statement of Principles on Digital Data Management.
2. To retrieve, quality control, and redistribute data to researchers in a timely manner.
3. To ensure data-related infrastructure is well maintained and secure from threats (physical and cyber). This includes performing routine maintenance checks, data backups, and software and infrastructure upgrades.
4. To build and maintain linkages to affiliate regional telemetry networks (nodes) and ensure interoperability and compatibility with OTN's data system. And to maintain a pipeline for public data to flow into global systems such as the Ocean Biodiversity Information System (OBIS) and the Global Biodiversity Information Facility (GBIF).
5. To optimize, refine, and re-invent the data infrastructure and processes to maintain alignment with globally recognized best practices in data systems and management. To maintain relevant certifications, e.g., OBIS Tier II node and adapt to new and emerging standards of data excellence and seek new accreditations that reflect OTN's continued commitment to high-quality standards, e.g., CoreTrustSeal and IODE/ADU.
6. To ensure OTN's data system is positioned to accommodate sovereign data sets and abides by relevant principles such as OCAP (Ownership, Control, Access, Possession).

IV. Data Acquisition

Responsible Parties: The intake of data from affiliated researchers is headed by the Data Acquisition Coordinator, with support from the Field Operations-Data Acquisition Coordinator (DAC) and other data team members as needed. The collecting investigator uploads most data via Plone—OTN's open-source Content Management System (CMS). Investigators are issued a

secure, password-protected repository folder that limits access to designated persons identified by the investigator and OTNDC. The identification of missing or delinquent data, support for amending or reformatting supplied data and metadata, and training for investigators on project-level data controls are provided by data acquisition personnel as needed. Specifics of how the OTN data system handles each data type are found in DAC checklists, data processing notebooks and SOPs.

Project creation and initial metadata harvest: Each project added to OTN's data system is assigned a unique collection code within its database Node and can be uniquely identified by adding a prefix denoting the node it was registered by, e.g., NEP.COHO.

OTN collects and stores metadata consisting of project metadata, deployment metadata, and tagging metadata. Metadata is collected on standardized templates to facilitate transmission into the OTN database. Detailed instructions on how to fill out templates are shared with new users to the data portal, as well as a list of frequently asked questions (FAQ) for network members to reference. Reference materials are also available on the members portal. Metadata include:

1. **Project***:* Contains essential project details, including PIs, collaborators, and contact details.
2. **Deployment***:* Contains all data pertaining to instrument deployment, including location data.
3. **Tagging:** Contains all details about tagged animals and is utilized to identify tagged animals.

Data sources collected by OTN include:

1. Electronic telemetry data*:* instrument-attached or implanted animal telemetry observation data loaded directly into OTN's data warehouse and/or aggregated via affiliate regional telemetry networks. Sources are:

  a. *Acoustic receivers*: The majority of data housed by OTN is detection data from acoustic receivers. Acoustic data is also obtained via fixed deployments that transmit acoustic data via satellites, e.g., live buoys and gliders with attached receiver units. Researchers provide data in the form of raw files offloaded from instruments or mirrored from affiliated regional networks and nodes. The data includes timestamped, georeferenced locations of stations that have observed the unique code emitted by each tagged animal.
  b. *Satellites*: OTN obtains satellite telemetry data directly from researchers who export track files to the data warehouse or via manufacturer file-drops with express researcher permission. Efforts are underway in conjunction with the US IOOS ATN to enable OTN to directly download instrument data from manufacturer web portals, provided that prior approval from the tagging researcher has been granted.

2. Oceanographic data: OTN deploys oceanographic instruments alongside their acoustic listening stations and utilizes a number of autonomous and remotely operated machines, including gliders, remotely operated vehicles (ROVs), and side-scan sonar to characterize the environment through which tagged species are tracked. Sources are:

  a. *Gliders*: OTN's Glider Group deploys surface and subsea platforms that collect oceanographic data and transmit subsets via satellite in near-real-time. This data is contributed to the Global Telecommunication System via a partnership with DFO-MEDS.
  b. *ROVs / Side-scan sonar*: OTN deploys ROVs and side-scan sonar instruments to characterize the composition/topology of the ocean floor within a study area to better plan the deployment of instruments or service/recover lost, unresponsive, or stuck moorings.

Information collected by these instruments is stored and archived at Dalhousie upon the conclusion of the mission.

c. *Stationary Data Loggers*: Oceanographic data collected by co-deployed stationary data loggers are shared where possible with the relevant regional association of CIOOS for optimal quality control and data discovery.

3. Audiovisual data: Some OTN platforms collect audiovisual data. This media is stored and archived at Dalhousie and is shared with partners on request. If a biological presence data is captured, an archival copy will be published with guidance from the camera-trap and machine observation working groups of TDWG, either through OTN or OBIS-Canada.

V. Data Products and Information Created/Managed by OTNDC

- **PostgreSQL/GIS database:** OTN maintains a database structure suitable for managing and connecting hundreds of distinct but co-related telemetry projects while maintaining autonomy within any one project's submitted data. This structure is defined as a set of SQL commands version-controlled and managed via a database migration engine (nomad) to allow a concerted upgrade regime across all instances of the data system. The Database Manager is responsible for database maintenance, authorship of new migrations, delivering new features, and fixing bugs.

- **Project-level reports:** OTNDC builds a report for each registered project that has deployed acoustic tags and every detection of deployed tags across any listening station reporting data into the network. This tag-focused report is called a detection extract and provides researchers with the fullest possible description of where tagged animals have been detected.

- **Key Performance Indicator (KPI) dashboard:** OTN tracks statistics that indicate its progress towards various goals via a KPI Dashboard application. The application harvests information daily from OTN's Database, OTN's knowledge management platform (Podio), and inventory management system (Snipe-IT) and annually via the annual survey of OTN-affiliated investigators. It produces on-demand reports and is shared with staff via a secure web portal.

- **Data processing JuPyTeR notebooks (Nodebooks):** OTN's data processing pipeline for the most common data sources is written in Python and maintained by OTN's programming team. Data workflows are defined in a collection of code notebooks called the "Nodebooks," which lower the barrier to data assimilation for non-technical regional node managers. Nodebooks ensure that data managers at OTN and its affiliates follow the same processes and standards for data ingestion, thus guaranteeing the validity of data aggregated across nodes.

- **Data Portals:** OTN's Global Map hosted on the OTN Data Portal offers high-level taxonomic, geospatial and temporal summaries of OTN's data holdings. These, along with the Network-level Statistics page, are designed to call attention to the work of affiliated OTN researchers and the global scope of OTN. Individual Project Metadata pages are generated using the supplied project metadata, identifying points of contact and species of interest as well as scientific goals for every project affiliated with OTN. Data used for discovery is shared via OTN's public data portals, GeoServer and ERDDAP, to optimize its utility for end-users building GIS applications or analysis pipelines.

- **End-user Analysis Tools:** OTN data personnel co-author and co-maintain a suite of community-developed tools for data analysis and visualization in R and Python.

VI: Documentation and Workflow

Received data and metadata (template) files are run through QA/QC processes, revised with the advice of the submitter, and finally loaded into the OTN data system via the Nodebooks. To maintain proper project scope and power of revision, each project writes data into its own separate schema. At scheduled intervals coinciding with the conclusion of seasonal fieldwork and data submission, the Nodebooks are also used to process the matching of tagged animals to detection data and create detection extracts for tagging projects (format described here). These detection extracts are shared securely with each project via private document folders hosted on OTN's Plone CMS. Summaries of all vetted data are generated into a separate schema to inform the Data Portal and public data endpoints. This process is referred to as a data push, whose workflow is global in scope and catalogued separately from Nodebooks. OTNDC maintains a FAQ page describing data flow activities, and additions are currently in development in OTNDC GitLab.

OTN's Data Portal improves the discoverability of its affiliated projects via its Global Map and Statistics pages, which are hosted on Plone. Programmatically generated individual project metadata pages are built using supplied metadata and geospatial details of sampling effort, deployed stations, and tags released. Publishable projects are expressed onto the OTN IPT in Darwin Core Archives, harvested in turn by OBIS, and added to the broader global record of marine biodiversity. To maximize interoperability and reuse of metadata served by the Data Portal, OTN creates GeoServer geospatial data layers and filtered per-project data hyperlinks alongside each Project Metadata page. An "All Public Data listing" is maintained after every data push, and details of deployments, tagging, and detection events are shared via OTN's ERDDAP portal.

VII. Data Storage and Backup

Data is stored on a physical server maintained by Dalhousie's Information Technology Services (ITS). Backups are performed on a daily/weekly/monthly regimen by ITS to ensure optimal hardware operation. The physical server is maintained for the life of its warranty, and hardware upgrades are budgeted accordingly. Storage requirements are estimated on a five-year basis, and service-level agreements between OTN and ITS are reviewed and re-negotiated annually. Additional details on security are available in the Cybersecurity plan.

VIII. Publication and Data Sharing

OTN upholds FAIR and TRUST principles and is committed to the timely release of data to be publicly available per the Tri-Agency Statement of Principles on Digital Data Management. Data and metadata are published according to the OTN Data Policy, which collaborating investigators agree to upon registering their data with OTN. Data embargo requests are reviewed on a case-by-case basis by the International Scientific Advisory Committee (ISAC), and extensions to data embargoes will not be unreasonably withheld. Investigators are welcome to waive the embargo period and make data publicly available at any time before its expiry.Data uploaded primarily into an affiliated OTN node will abide first and foremost by the reporting or parent node's data policy, which can include rules governing the publishing of data and metadata. Matches made to detection events from affiliated nodes are governed by the data policy of the node to which the tag and animal morphology have been registered.

In preparation for data release, OTN uses internationally recognized standards, vocabularies, data formats, and open-source tools and software. Species names are verified using the World Register of Marine Species (WoRMS). The OTN database design was heavily informed by Darwin Core, a collection of terms and definitions that facilitate the sharing of information about biological diversity. OTN publishes unembargoed animal occurrence and sampling event data to OBIS in Darwin Core format via the Integrated Publishing Toolkit (IPT), under a Creative Commons Attribution 4.0 International (CC BY 4.0). OTN also employs two open-source data sharing portals popular in the oceanographic community, ERDDAP and GeoServer, for public data distribution of OTN animal presence data across various formats.

Publicly available data housed by OTN are shared through a number of endpoints and are updated following each data push. A complete list is available via the IODE ODISC as well as through OTN's members portal. OTN also maintains a publication data repository containing a list of datasets that directly support the publication of telemetry research.

IX. Ethics and Legal Compliance

OTN interfaces with two types of sensitive data:

1. **Indigenous-led data:** OTN defines Indigenous-led data as data informed by, pertains to, or collected by Indigenous peoples. As part of its commitment to EDIA (equity, diversity, inclusion, and accessibility), OTN abides by principles of OCAP for data governed by First Nation's data sovereignty. OTNDC staff have completed training in OCAP via the First Nation's Information Governance Centre (FNIGC). At the global level, OTN aligns its data system with CARE Principles for Indigenous Data Governance.

2. **Species-at-risk data:** Data pertaining to species at risk is often subject to extended embargos. OTN works closely with researchers to ensure data is not released that may cause undue harm. Additionally, data that may be made public, e.g., receiver locations, can be concealed to ensure exact locations are not easily available ("fuzzy data").

Researchers must agree in advance to have their names, project titles, and contact information published for all projects. Personal or other sensitive data, including fishing and vessel details, are not disclosed or stored. All researchers in possession of an OTN equipment loan are further required to sign a copy of the OTN Data Policy. This policy was drafted in conjunction with Dalhousie's Office of Research Services and is referenced in the legally binding equipment loan agreement between Dalhousie and the borrowing institution. Under the data policy, ownership of tagging data and mapped detections rest with the tagging project's PI(s). The PI on the receiving deployment retains ownership over the unmapped detections.

X. Selection and Preservation

Data held and quality controlled by OTNDC are critically important to the investigators who provide the observations and the broader and future research community. They form a global dataset of machine-observed animal presence spanning hundreds of species, are of immeasurable long-term value, and will be retained, shared, and preserved via OTN or per the archival strategy laid out in the data decommissioning plan.

XI. Review of the Plan

This Data Management Plan is updated on an annual basis, first under the guidance of the Director of Data Operations and second by the International Data Management Committee. Substantive changes are brought to the OTN Council for approval.