

Data Management Plan: Automatic Identification System (AIS)

NOAA Data Sharing Template

I. Type of data and information created

What data will you collect or create in the research?

Contextual statement describing what data are collected and relevant URL (IOOS Certification, f 1. ii)

AIS, real-time logging of data received from AIS antennas, with times added for preservation.

What data types will you be creating or capturing?

AIS is ASCII data in binary encoding. These data would need to be timestamped raw storage, as well as a local database/cache of MMSI metadata. The data might also produce a timeseries of lat/lon/time for aiding visual derivative products.

How will you capture or create the data?

Describe how the data are ingested (IOOS Certification, f 2.)

TCP socket application that writes received contents to disk with timestamp prepended. Async, post-processing will extract and transform the encoded information into a reasonable metadata storage system, such as a MySQL database and unencoded ASCII track history.

Describe how data are managed (IOOS Certification, f 2.)

Describe the data quality control procedures that have been applied to the data. (IOOS Certification, f 3.)

No QC. Minimal QA to ensure data acquired.

If you will be using existing data, state that fact and include where you got it.

What is the relationship between the data you are collecting and the existing data?

The data are provided by Marine Exchange. We maintain a few antennas and, in exchange for our piping those antenna outputs directly to them, they open up their entire data collection to us.

Data Management Plan: Automatic Identification System (AIS)

II. Expected schedule for data sharing

Adheres to the NOAA Data Sharing Procedural Directive. The System is an operational system; therefore the RICE should strive to provide as much data as possible, in real-time or near real-time, to support the operation of the System. (IOOS Certification, f. 4.)

How long will the original data collector/creator/principal investigator retain the right to use the data before opening it up to wider use?

How long do you expect to keep the data private before making it available? Explain if different data products will become available on different schedules (Ex: raw data vs processed data, observations vs models, etc.)

Real-time product: sent directly to browser as it comes in off the antenna.

Archival information: Unknown.

Explain details of any embargo periods for political/commercial/patent reasons?

When will you make the data available?

Unknown

Data Management Plan: Automatic Identification System (AIS)

III. Standards for format and content

Which file formats will you use for your data, and why?

How can the information be accessed? (IOOS Certification, f 1. ii)

ASCII. The native data format is already highly optimized for telemetry over radio. It is an incredibly dense data set, generating 3-4 GB of data per month, with up to 100 thousand samples per hour, or roughly 30 samples per second.

What file formats will be used for data sharing?

Zipped files of the raw data are internally used and JSON is used for real-time visualization.

What metadata/ documentation will be submitted alongside the data or created on deposit/ transformation in order to make the data reusable?

Unknown.

What contextual details (metadata) are needed to make the data you capture or collect meaningful?

The AIS encoding system, and various weather package add-on products which produce binary data that is sometimes encoded into the AIS messages. These add-on packages have proprietary data formats and may not be decode-able.

How will you create or capture these details?

The CORDC team has reverse engineered some of the fields, and we have found AIS parser implementations which have been adapted for the web and for local post-processing.

What form will the metadata describing/documenting your data take?

Unknown, flexible.

Which metadata standards will you use and why have you chosen them? (e.g. accepted domain-local standards, widespread usage)

Unknown. Whatever we're told to adapt. In fact, since data isn't allowed to be distributed, we may not even produce a formal metadata document.

Data Management Plan: Automatic Identification System (AIS)

IV. Policies for stewardship and preservation

What is the long-term strategy for maintaining, curating and archiving the data?

Points of contact- Individuals responsible for the data management and coordination across the region (CV's attached); (IOOS Certification f 1. i)

Maintain our good standing with Marine Exchange, which consists of providing ongoing maintenance of the AIS antennas installed at various HF RADAR stations. We shall notify Marine Exchange whenever our archival system's IP changes, so as to be allowed to maintain access to their data feed.

Identify the procedures used to evaluate the capability of the individual (s) identified in subsection 997.23(f)(1) to conduct the assigned duties responsibly. (IOOS Certification, f 1. iii)

Which archive/repository/database have you identified as a place to deposit data?

Documents of the RICE's data archiving process or describes how the RICE intends to archive data at the national archive center (e.g., NODC, NGDC, NCDC) in a manner that follows guidelines outlined by that center. Documentation shall be in the form of a Submission Agreement, Submission Information Form (SIF) or other, similar data producer-archive agreement (IOOS Certification, f 6.).

Time-based directory structure is used for the current archives. We could maintain a simple relational database such as MySQL for maintaining semi-static ship metadata and recent messaging information, kdb+ is good for high density time series:<http://kx.com/software.php> , and I've also heard good things of Postgres+PostGIS for geo-spatial data. None of these solutions have been implemented, and none are scheduled for implementation.

What procedures does your intended long-term data storage facility have in place for preservation and backup?

Data are backed up periodically with the CommVault suite.

How long will/should data be kept beyond the life of the project?

Indefinitely. Shipping records from the 1800's have been of great use for socio-economic research.

What data will be preserved for the long-term?

All information transmitted by the AIS transceiver, along with a timestamp of when the message was received by our acquiring server.

What transformations will be necessary to prepare data for preservation / data sharing?

Conversion from encoded formats to a more readily understood formats would be required. Separation of proprietary non-decodable information would be required since binary data is unintelligible by ASCII text readers, and can cause formatting/parsing problems. Aggregation of multiple files into a convenient transport format, such as .zip format, would be helpful, but not required.

What metadata/ documentation will be submitted alongside the data or created on deposit/transformation in order to make the data reusable?

Data Management Plan: Automatic Identification System (AIS)

Unknown. All, or none. Most certainly legal disclaimers and statement of ownership/copyright.

What related information will be deposited?

Location of AIS antennas owned/maintained by Scripps.

* Ideally, location of non-Scripps antennas as well, but we don't have that information.

Data Management Plan: Automatic Identification System (AIS)

V. Procedures for providing access

What are your plans for providing access to your data? (on your website, available via ftp download, via e-mail, or another way)

Describe how data are distributed including a description of the flow of data through the RICE data assembly center from the source to the public dissemination/access mechanism. (IOOS Certification, f. 2.)

Current state of AIS traffic hosted on the web: <http://cordc.ucsd.edu/projects/ais/>

We provide an internal TCP relay in support of local projects. The details of which I cannot disclose.

Will any permission restrictions need to be placed on the data?

Yes. We don't own the data, so permissions must be obtained from Marine Exchange.

With whom will you share the data, and under what conditions?

Unknown. Any UCSD researchers, under non-redistribution clauses.

Will a data sharing agreement be required?

Unknown. Probably.

Are there ethical and privacy issues? If so, how will these be resolved?

Yes. Since AIS contains real-time location information for public and private vessels, one could identify a prominent individual by their vessel name, cross reference that to an MMSI number, and watch for real-time position updates from their vessels... safety at sea, is a delicate issue. These issues aren't addressed, as anyone with an AIS receiver could obtain this information. It is public.

Who will hold the intellectual property rights to the data and how might this affect data access?

Unknown.

Data Management Plan: Automatic Identification System (AIS)

VI. Previous published data