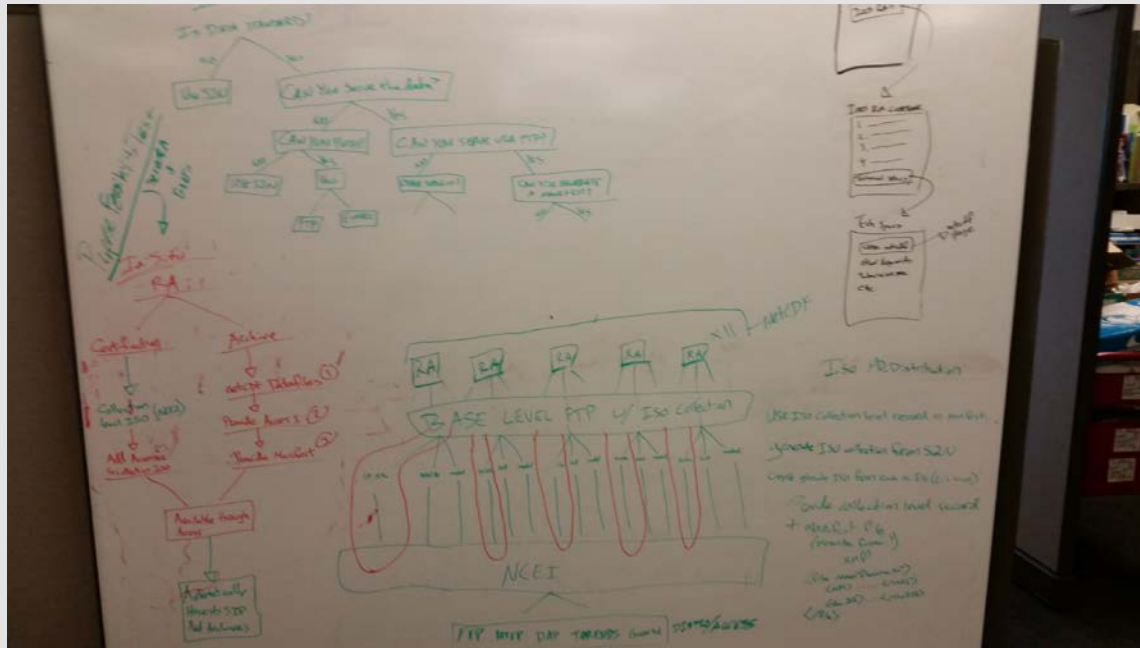




NCEI-IOOS Project Updates



Mathew Biddle

Mathew.Biddle@noaa.gov

May 28th, 2015

IOOS DMAC Meeting, IOOS Program Office
Silver Spring, MD





Outline

- Information about [NCEI](#)
- Certification
- Archiving
- Access
- Future



National Centers for Environmental Information

- Revealing the Past, Interpreting the Present, and Informing the Future
- NOAA's National Centers for Environmental Information (NCEI) is the merger of the National Climatic Data Center, National Geophysical Data Center, and National Oceanographic Data Center as approved in the Consolidated and Further Continuing Appropriations Act, 2015, Public Law 113-235.
- The newly merged organization under NESDIS is called the National Centers for Environmental Information (NCEI).
- NOAA requested the merger to increase integration across the three centers. By using consistent data stewardship tools and practices across all of our science disciplines and by forging an improved data management paradigm, we expect to provide users with improved access to environmental data and information archive products. (archive, IT, administration).
- The merger will allow the Data Centers to continue the successful tradition and mission of stewarding the Nation's environmental data and providing outstanding use-inspired products and services to the American public. It will provide much-needed information from and access to oceanographic, geophysical, and climatic data in a fully integrated way.
- A top priority during the merger will be to build on the full spectrum of climatic, oceanographic, coastal, and geophysical science and services the Data Centers currently deliver.

<http://www.ncei.noaa.gov/>



NCEI Tiers of Data Stewardship

6: National Services and International Leadership

- Lead, coordinate, or implement scientific stewardship activities for a community or across disciplines
- Establish highly specialized levels of data services and product assessments

5: Authoritative Records

- Combine multiple time series into a single, inter-calibrated product
- Establish authoritative quality, uncertainties, and provenance
- Ensure products are fully documented and reproducible

4: Derived Products

- Build upon archived data to create new products that are more broadly useful
- Distill, combine, or analyze products and data to create new or blended scientific data products

3: Scientific Improvements

- Improve data quality or accuracy with scientific quality assessments, controls, warning flags, and corrections
- Reprocess data sets to new, improved versions and distribute to users

2: Enhanced Access and Basic Quality Assurance

- Create complete metadata to enable automated quality assurance and statistics collection
- Provide enhanced data access through specialized software services for users and applications

1: Long Term Preservation and Basic Access

- Preserve original data with metadata for discovery and access
- Serve as expert advisors on standards for data providers.
- Archive only necessary data using appropriate retention schedules
- Safeguard data over its entire life-cycle
- Coordinate support agreements for sustainable data archiving
- Provide data citation services by minting DOIs



NCEI Tiers of Data Stewardship

6: National Services and International Leadership

- Lead, coordinate, or implement scientific stewardship activities for a community or across disciplines
- Establish highly specialized levels of data services and product assessments

5: Authoritative Records

- Combine multiple time series into a single, inter-calibrated product
- Establish authoritative quality, uncertainties, and provenance
- Ensure products are fully documented and reproducible

4: Derived Products

- Build upon archived data to create new products that are more broadly useful
- Distill, combine, or analyze products and data to create new or blended scientific data products

3: Scientific Improvements

- Improve data quality or accuracy with scientific quality assessments, controls, warning flags, and corrections
- Reprocess data sets to new, improved versions and distribute to users

2: Enhanced Access and Basic Quality Assurance

- Create complete metadata to enable automated quality assurance and statistics collection
- Provide enhanced data access through specialized software services for users and applications

1: Long Term Preservation and Basic Access

- Preserve original data with metadata for discovery and access
- Serve as expert advisors on standards for data providers.
- Archive only necessary data using appropriate retention schedules
- Safeguard data over its entire life-cycle
- Coordinate support agreements for sustainable data archiving
- Provide data citation services by minting DOIs



Archiving submission process

Two types of submissions:

1. One-off:

- a. One time or very infrequent submissions.
- b. Non-standard data sets.
- c. Now you can submit through the new Send2NCEI webtool!

2. Automation:

- a. Recurring submission.
- b. Well structured, consistent data sets.
- c. Develop the submission procedures with NCEI.
 - i. ATRAC (Advanced Tracking and Resource tool for Archive Collections) <https://www.ncdc.noaa.gov/atrac/index.html>
- d. NCEI Pipeline proposal.

Send2NCEI

The screenshot shows the NOAA National Centers for Environmental Information (NCEI) website. The header includes the NOAA logo and the text "NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION" and "NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION". Below the header is a navigation bar with buttons for "Home", "Access Data", "Submit Data", "Public Outreach", and "About". A search bar is located in the top right corner. The main content area is titled "My Submission Packages" and contains several sections:

- Create A New Submission Package**: A button to initiate a new submission.
- Submission Packages I Am Working On**: A table with columns for Reference ID, Status, Title, and Creation Date.
- Submission Packages NCEI Is Working On**: A table with columns for Reference ID, Status, Title, and Submission Date.
- Archived Submission Packages**: A table with columns for Reference ID, NODC Accession, Title, and Date Archived.
- Rejected Submission Packages**: A table with columns for Reference ID, Title, and Date Submitted.

At the bottom of the page, there is a footer with links for "Access Data", "Submit Data", "Intended Use of the Data?", "Online Store", and "Customer Service".

<https://www.nodc.noaa.gov/s2n/>





Archiving submission process

Two types of submissions:

1. One-off:

- a. One time or very infrequent submissions.
- b. Non-standard data sets.
- c. Now you can submit through the new Send2NCEI webtool!

2. Automation:

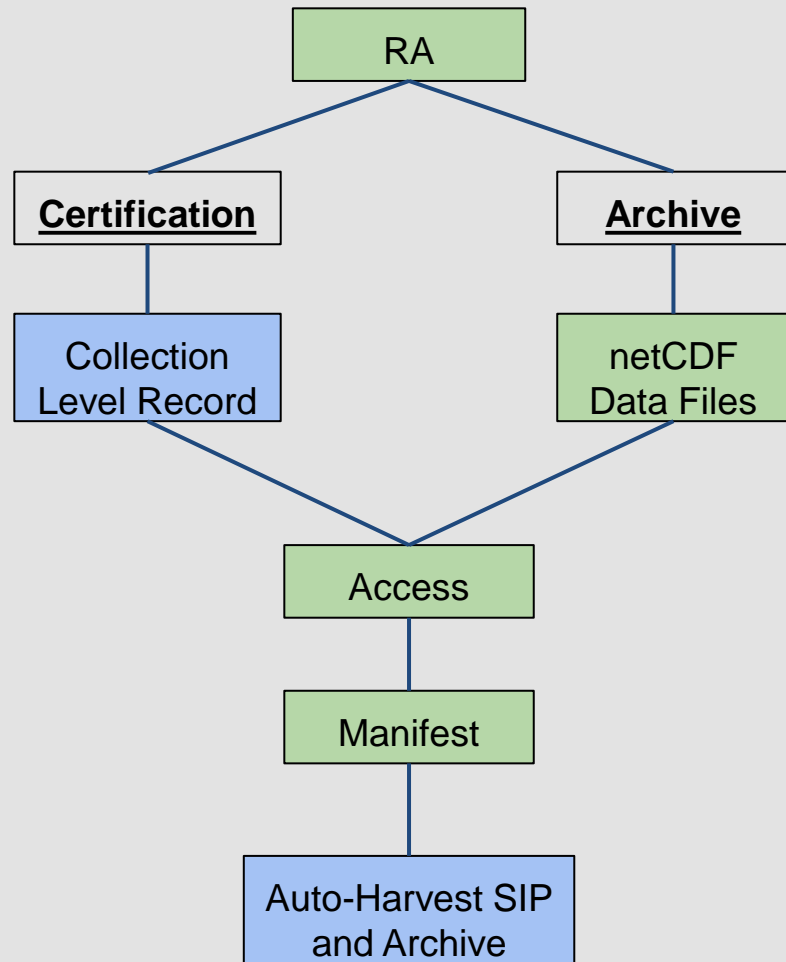
- a. Recurring submission.
- b. Well structured, consistent data sets.
- c. Develop the submission procedures with NCEI.
 - i. Submission Information Form (SIF).
 - ii. ATRAC (Advanced Tracking and Resource tool for Archive Collections) <https://www.ncdc.noaa.gov/atrac/index.html>.
- d. NCEI Pipeline proposal.



RA-NCEI Data Pipeline Proposal

- Build a pipeline for each CF Feature Type (Time Series, Profile, etc.).
- If the data has a DAC, send it to the DAC (HF Radar, Glider, CDIP, etc.).
- Start developing the pipeline with a simple feature type first.
- Want to focus on non-Federal assets initially.

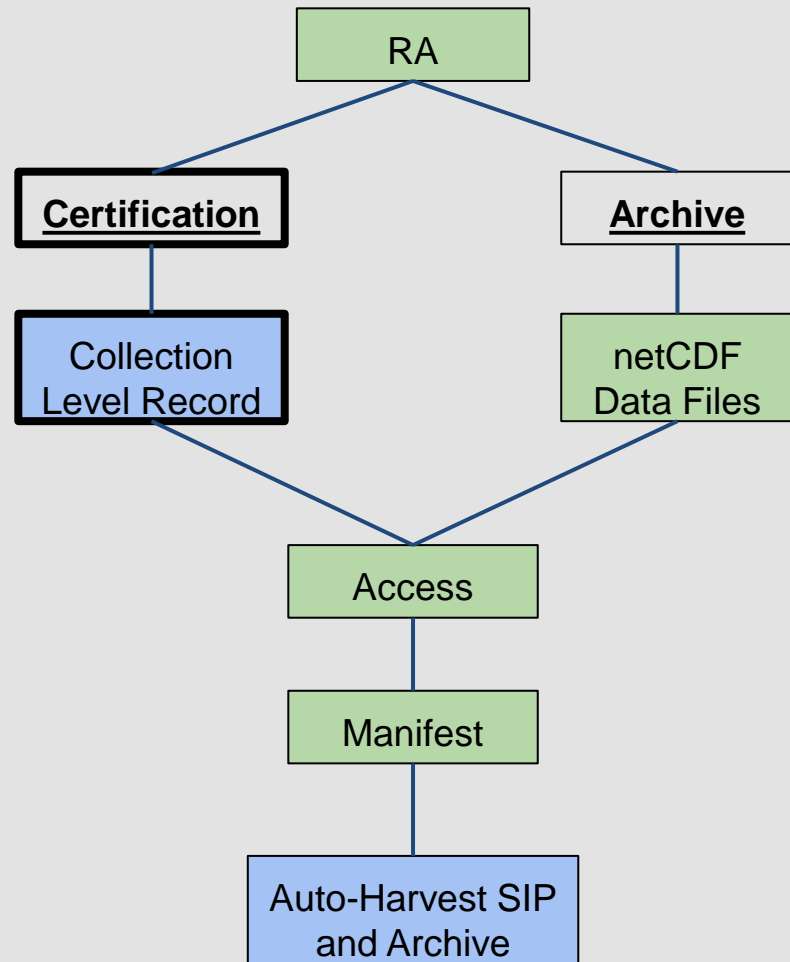
RA-NCEI Data Pipeline Proposal



NCEI

RA

RA-NCEI Data Pipeline Proposal



NCEI

RA



Certification

Requirements:

- Section 223.35 (f) of the certification form is complete.
- Any contracts with private industry must be cited.
- Any litigation holds are clearly stated and supporting documentation
- Any/all data you handle must be documented.
- Specific attention to:
 - data flow
 - data conversions
 - QA/QC

Checklist:

- In development.
- Fundamentally based on the guidance provided by the IOOS Program Office.

Collection Level Record

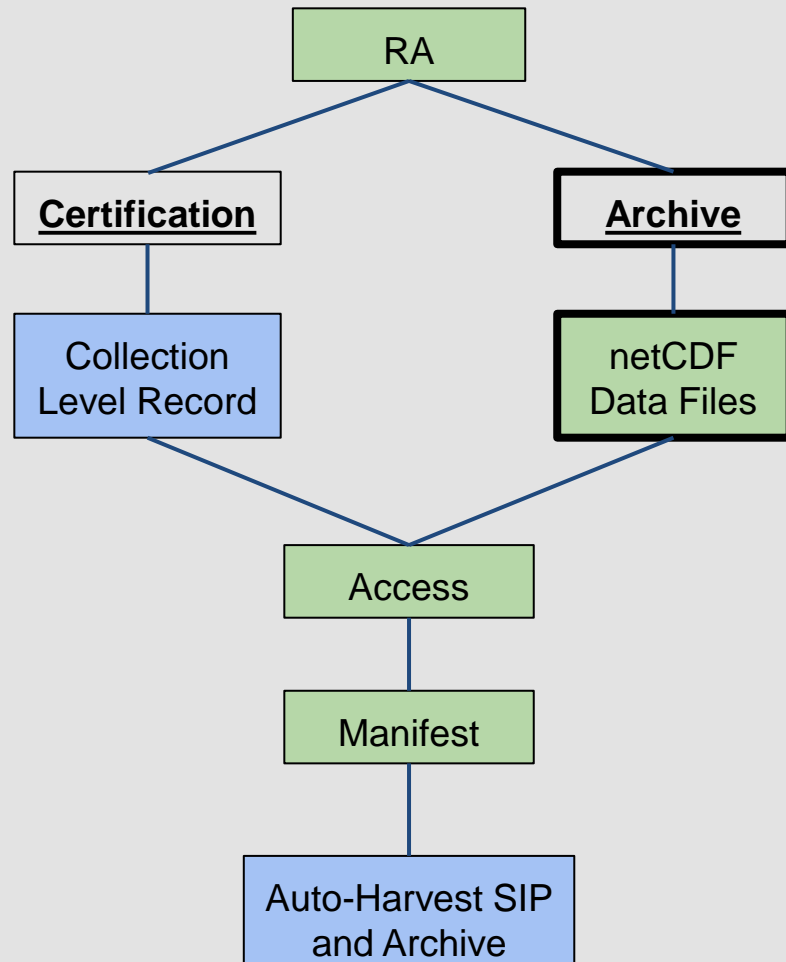
If a Certification exists:

- Developed from Certification documents.
- Some iteration between RA and NCEI will occur.

If a Certification does not exist:

- Iteration between RA and NCEI will develop the record.

RA-NCEI Data Pipeline Proposal



NCEI

RA





NODC NetCDF Templates

- Provides guidance for formatting your data in netCDF.
- Primarily follow ACDD 1.2 and CF 1.6 with a few added attributes/variables.
- Decision Tree.
- Data can be served to the public through NCEI's various web services (FTP, HTTP, DAP, THREDDS...).
- Tier 2 Stewardship.

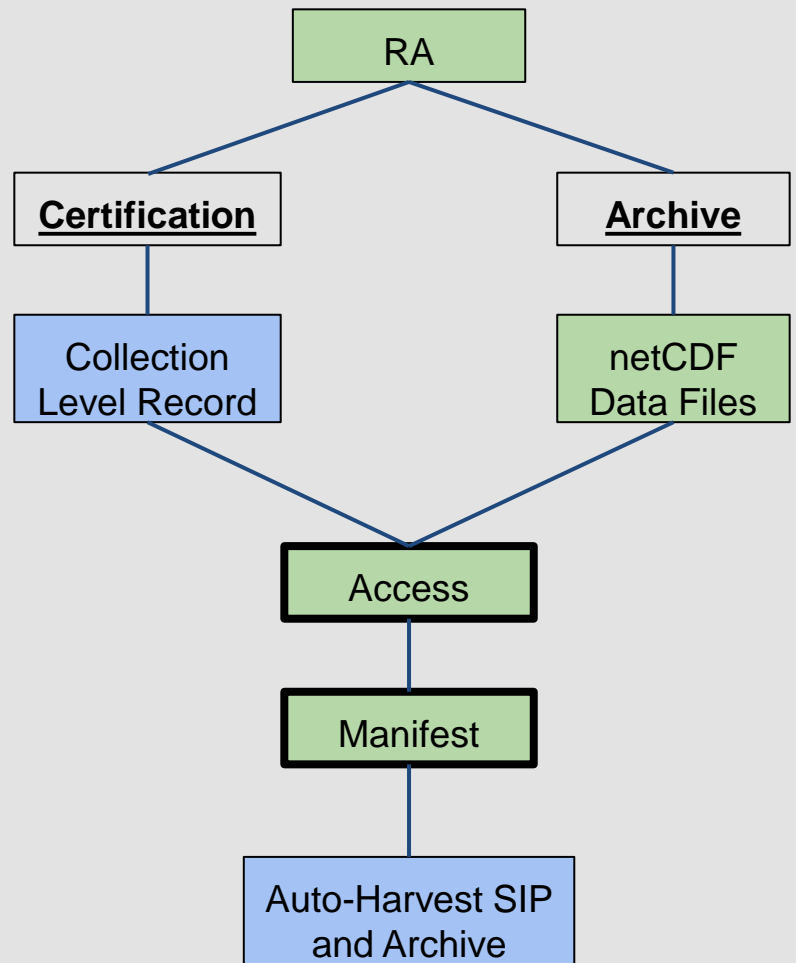
Updates:

- Working on updates for ACDD 1.3.
- Broadening the application of our templates to fit the NCEI scope.

<http://www.nodc.noaa.gov/data/formats/netcdf/v1.1/index.html>



RA-NCEI Data Pipeline Proposal



NCEI

RA

Current IOOS RA Automations

- SECOORA
- GLOS
- AOOS CSESP

The screenshot shows the NOAA National Oceanographic Data Center (NODC) search results page. The search query is "Integrated Ocean Observing System Data Asse". The results list several records, all of which are "Oceanographic and surface meteorological data collected from station [station ID] by Everglades National Park (ENP) and assembled by Southeast Coastal Ocean Observing Regional Association (SECOORA) in the Coastal Waters of Florida and North Atlantic Ocean from 2014-02-13 to 2015-04-30 (NODC Accession [ID])". The records listed are for stations bdfv1, bnkf1, bobf1, bwsf1, and canf1. The page includes search filters for "WHEN" (Dates overlap range, Dates within range) and "WHERE" (Anywhere, Intersecting, Fully within). A map of the United States is shown with a red box highlighting the Southeastern region. The NOAA logo and "NATIONAL OCEANOGRAPHIC DATA CENTER (NODC)" are visible at the top.

Key: "Integrated Ocean Observing System Data Assembly Centers Data Stewardship Program"

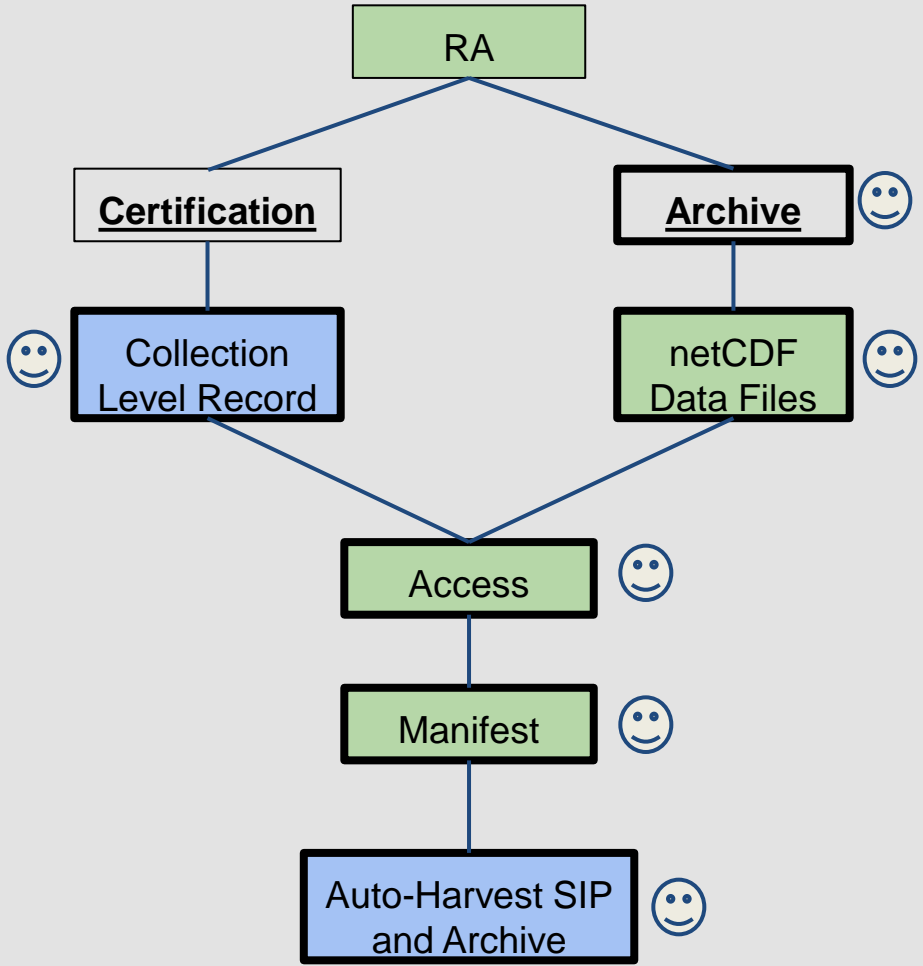


How do we manage SECOORA data?

- NetCDF data files are posted on SECOORA FTP for NCEI to harvest.
- Validation through checksums.
- Disseminate and generate Archival Information Packages (AIP).
(e.g. <http://accession.nodc.noaa.gov/0118736>)
- Each AIP is one station that gets updated monthly.
- Automation started on May 28, 2014.
- Current volumes (as of 5/20/2015):
 - min = 0.264 MB*
 - max = 6.380 MB*
 - average = 1.432 MB*
 - total = 88.784 MB (~12 months, 62 AIP)
- Data Access Statistics
 - <http://www.nodc.noaa.gov/webstats/secoora/ftp/2015/04/secoora-ftp.html>
 - <http://www.nodc.noaa.gov/webstats/secoora/www/2015/04/secoora-www.html>

*The sizes are per AIP.

SECOORA



NCEI

RA

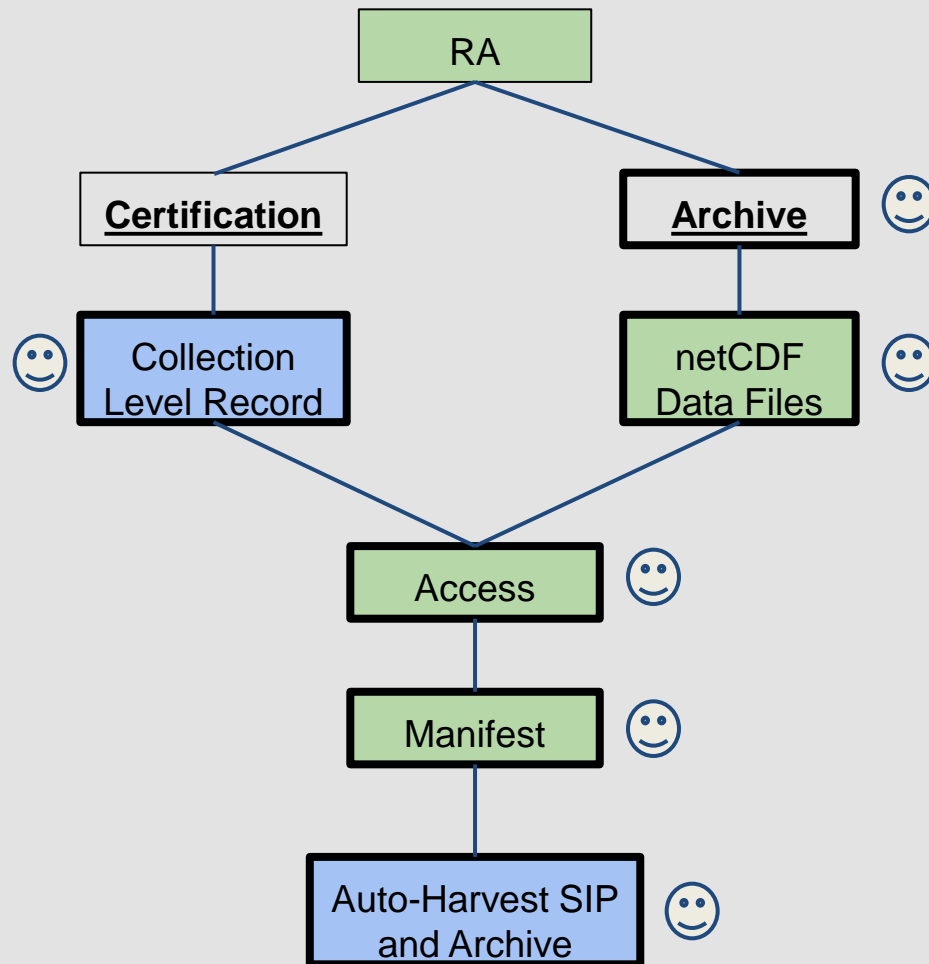


How do we manage GLOS data?

- NetCDF data files are posted on GLOS FTP for NCEI to harvest.
- Validation through checksums.
- Disseminate and generate Archival Information Packages (AIP).
(e.g. <http://accession.nodc.noaa.gov/0123647>)
- Each AIP is one station that gets updated monthly.
- Automation started on December 11, 2014.
- Current volumes (as of 5/20/2015):
 - min = 0.360 MB*
 - max = 3.236 MB*
 - average = 1.363 MB*
 - total = 29.976 MB (~6 months, 22 AIP)

*The sizes are per AIP.

GLOS



NCEI

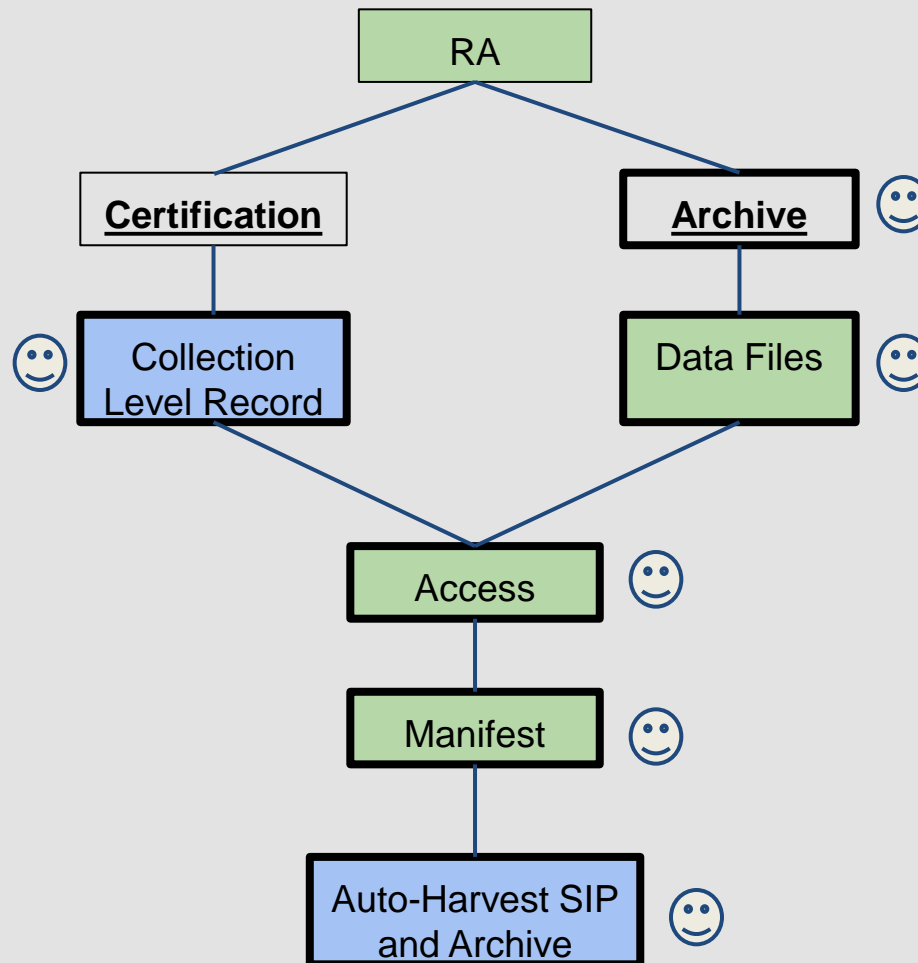
RA



How do we manage AOOS CSESP data?

- Six years of data
- Yearly updates.
- Data is posted on workspace... when a manifest appears, we pull the data.
- Disseminate and generate Archival Information Packages (AIP).
- Current volumes:
 - total = 540.9 MB (1 AIP)

AOOS CSESP



NCEI

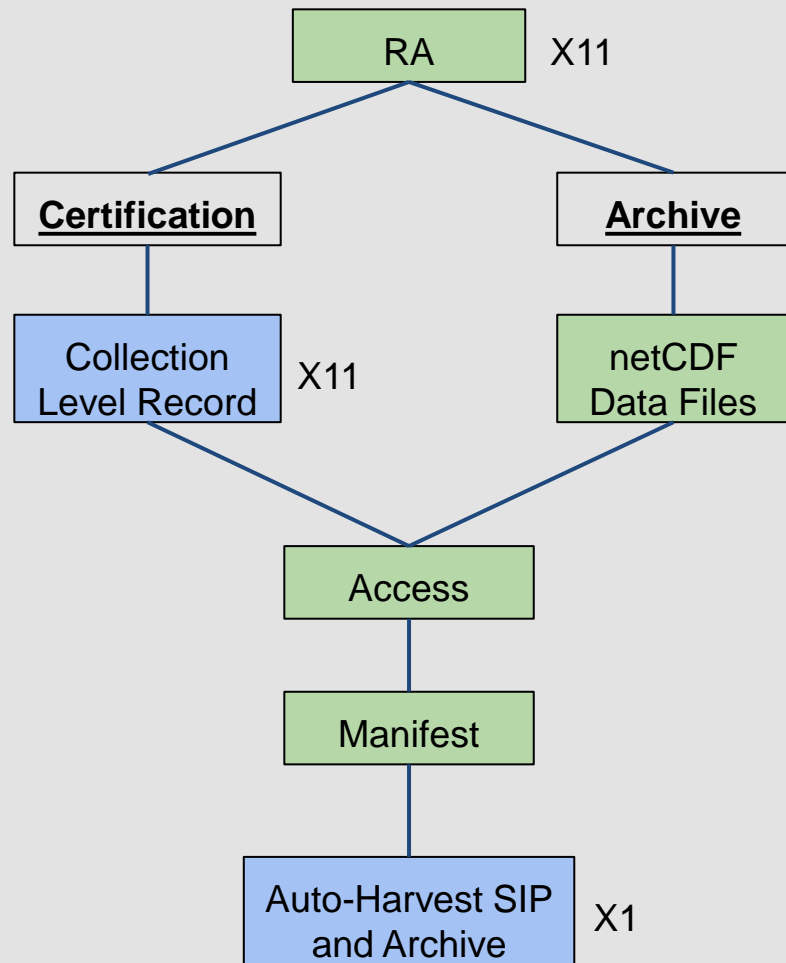
RA



Lessons Learned

- Consistency is key.
- Definitions for non CF keywords.
- Triage for data streams has been developed.
- Data not following CF and ACDD in netCDF requires more iteration.

RA-NCEI Data Pipeline Proposal



NCEI

RA





RA-NCEI Data Pipeline Proposal cont.

- RA requirements:
 - a. Data is formatted in *ioos compliance checker* validated netCDF.
 - b. Certification contains all relevant documentation about the data set.
 - c. RA will host the data on FTP/HTTP/DAP/THREDDS.
 - d. Manifest to document Submission Information Package (SIP).
- NCEI requirements:
 - a. Develop a collection record, with limited feedback from RA (provided the information is available).
 - b. Develop an acquisition procedure to pull the data, generate metadata, archive, and publish.
 - c. Provide various access mechanisms for the AIP's.

Access

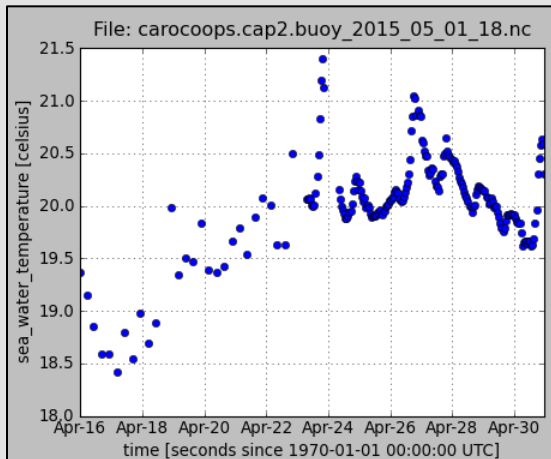
- <http://data.nodc.noaa.gov/geoportal/rest/find/document?searchText=%22Integrated%20Ocean%20Observing%20System%20Data%20Assembly%20Centers%20Data%20Stewardship%20Program%22&start=1&max=2500&contentOption=intersecting&f=searchPage>
- <http://data.nodc.noaa.gov/opensdap/ioos/contents.html>
- <ftp://ftp.nodc.noaa.gov/pub/data.nodc/ioos/>
- <http://data.nodc.noaa.gov/thredds/catalog/ioos/glos/catalog.html>
- <http://data.nodc.noaa.gov/thredds/catalog/ioos/secoora/catalog.html>
- <http://accession.nodc.noaa.gov/0124308/>



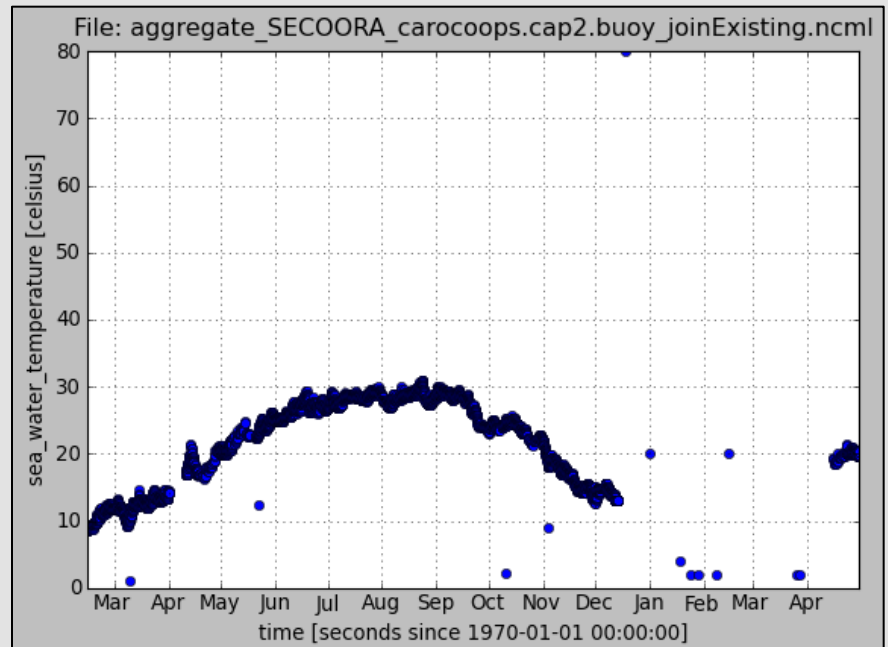
Future

- Feasibility test for the pipeline:
 - Develop the pipeline for the already automated SECOORA and GLOS process.
- NCEI Certification Checklist.
- Cookbook to submit data to NCEI.

Fun Stuff



http://data.nodc.noaa.gov/thredds/catalog/ioos/secoora/carocoops.cap2.buoy/catalog.html?dataset=ioos/secoora/carocoops.cap2.buoy/carocoops.cap2.buoy_2015_05_01_18.nc



http://data.nodc.noaa.gov/thredds/catalog/testdata/mbiddle/catalog.html?dataset=testdata/mbiddle/aggregate_SECOORA_carocoops.cap2.buoy_joinExisting.ncml



Thanks!

Questions?

