NCEI 2017 Passive Acoustic Data Workshop

A two-day workshop was held at NCEI in Boulder, CO on September 14-15, 2017 for NMFS Fisheries Science Centers and Office of Science and Technology, NOAA National Marine Sanctuaries, BOEM, Navy, and San Diego State University scientists. The workshop brought together participants to discuss the progress, current efforts and future direction of the passive archive including an overview of the 2016-2017 Big Earth Data Initiative project. Below is a summary of the workshop's major topics.

Passive Acoustic Archive Accomplishments

Chuck has developed a dedicated data submission software package to facilitate the transfer of raw passive acoustic data files and associated metadata. It was based off of the water column sonar data submission software package and enhanced to support the critical components that describe passive acoustic data. The "Passive Packer" was first used to transfer ~2TB of Ocean Noise Reference Station Network data to NCEI.

The NRS data packaged and submitted to NCEI were archived and made available through the newly established passive acoustic data map viewer. The map viewer enables public access to the archived data. There users can filter the holdings based on parameters based on spatial-temporal limits, funding source, sample rate, duty cycle and number of channels, and request the data. So far we have received 6 requests. A passive acoustic archive project page was developed to facilitate sharing of information to stakeholders and the public.

A digital object identifier (DOI) has been created for the NRS dataset. It was decided to use one DOI per project such that all deployments associated with the NRS regardless of location or period of data collection would fall under the same DOI. The details of this permanent citation and dataset description information provided as metadata through the Passive Packer can be found here.

The ~3TB of passive acoustic data provided to NCEI by the NEFSC during the 2014-2015 pilot project will be moved from its current location on spinning disk to the archive as soon as we receive some additional details to complete the metadata record for ISO compliance. This process will allow the NEFSC data to be discoverable and available on the map viewer.

Feedback

The group provided great feedback on areas that our map viewer could be improved. **First and foremost was the size of the dots on the map. That known issue has already been corrected!**

Our dots may not be as sophisticated as the Navy passive acoustic Data Access page but at least they are bigger than they used to be.

All additional suggestions (e.g., filter by marine sanctuary status, color the dots by funding agency and source, and improvements to frequency range filtering) will be completed by the end of the June - though hopefully much sooner.

Project Overviews

Atlantic Deepwater Ecosystem Observatory Network (ADEON)

Jen Miksis-Olds and Terry Ridgway from UNH provided an overview on the ADEON project. This multi-year endeavor will collect ~250 TB of data. With their early planning and preparation (and project objective to develop a robust data management system), they have taken an active lead in ensuring NCEI will be able to archive this large volume as easily as possible.

NMFS Fisheries Science Centers

- NEFSC
  - Total volume estimated: 89-106 TB/year
  - Includes HARP, MARU, glider, towed arrays, and SoundTraps
- SEFSC
  - Total volume estimated: 35.5-36.5 TB/year
  - Includes towed arrays, LARPs, DIFAR sonobuoy, and acousonde; also joint work with NEFSC HARP and MARU and NRS
- SWFSC
  - Total volume estimated: 30-60 TB/year
  - Includes towed arrays, sonobuoys and DASBRs (drifters)
- PIFSC
  - Total volume estimated: 256 TB for 2017
  - Includes HARPS, towed array, sonobuoy, and DASBR
- NWFSC
  - Total volume estimated: 4 TB for 2017
  - Includes CPODS, vertical arrays, DTAGS, towed array, sonobuoys, PALs, EARs
- AFSC
  - Total volume estimated: ~8 TB (plus Manolo's data)
  - Includes bottom mounted, sonobuoys, towed arrays, acousonde, and CPODS and DTAGS

Navy-NOAA surtass settlement data

An upcoming project will see the deployment of acoustic recorders in 7 National Marine Sanctuaries and 1 National Monument to monitor the soundscape. A component of this project

is the archival of the data at NCEI. Volume estimates are expected within the next 6 months (~March 2018). Instruments include SoundTraps and, to a lesser extent, gliders.

Ocean Noise Reference Station Network

Acoustic recorders were deployed at 11 sites in 2016 and 12 sites are anticipated in 2017 with the addition of Buck Island. ~4 TB per year are expected. Since the first submission to the archive last year, more datasets have been recovered and are ready to be archived. Samara Haver will await Chuck's newest version of the Passive Packer, expected to be distributed by Jan 2018, before sending the data to NCEI. The possibility to expand their current effort to include Navy and BOEM assets would result in 30 additional sites and 10-130 TB/year.

Tethys

The database is growing with over 7.5 million detections. Discussions are underway on building a crosswalk between Tethys and OBIS. Marie is a member of ASA's passive acoustic metadata standards committee and the science committee of the International Quiet Ocean Experiment.

Navy (Living Marine Resources and SPAWAR)

The ongoing projects at LMR span several categories (monitoring and mitigation, standards and metric, and protected species) throughout the Atlantic and Pacific Ocean, and focus on the use of HARPs. The possibility exists to make 1-2 channels of Navy data available to the public. This would have to be cleared by PACFLEET. **The work to implement a sustainable storage system at SPAWAR is very impressive. Congrats!**

BOEM

Projects supported by BOEM include ADEON, AMAPP2, an initiative in the Gulf of Mexico (GOM) that is just getting started, North Atlantic Wind Energy, Maryland Wind Energy, RODEO, CA & HI PACMAC, and SPAM. Both ADEON and GOM will result in big datasets coming online in the next year. Both wind energy projects are expected to produce 45 TB of data. RODEO should amount to < 1TB. SPAM will comprise DTAGs and will also be smaller in size.

Future Focus

2017-2018 BEDI and Beyond

- **A new version of the Passive Packer will be distributed by January 2018.** This will allow Samara to package and submit the next batch of available NRS data.

- At this point, **we will also start archiving the backlogged NMFS data**. The volume we're able to take in right away (within FY18) may be limited (maybe < 75 TB) but we'll do as much as we can.
    - Following a recent conversation with Ryan Freedman and Lindsey Peavey from NOS we will **also take in the first preliminary dataset associated with the Navy-NOAA surtass settlement agreement.** This is a very small dataset collected in the Channel Islands and will serve to further test the robustness of the Passive Packer.
- We will expand our capability to handle towed array data
    - **I will meet with Shannon and Taiki on Oct 23** when I'm in La Jolla for the National Ocean Exploration Forum.
    - The Tethys schema will help greatly in handling the GPS and data file matching for mobile data
- We will continue to work with **Jen and Terry from ADEON** to ensure their data management needs are met. Since they are looking at ~250 TB by the end of the project, monetary support to steward these data will be crucial. **Can BOEM help with that?** Also of incredible importance is the planning and assurance of good metadata to accompany the numerous datasets and deployments throughout the project. We are grateful for Jen and Terry's active role in making this happen.
- Additional improvements to data discovery and access (i.e., **enhancements to the map viewer and a more automated system to deliver requested data**) will be completed. Your feedback will help drive what areas are improved.
- We will work through the **remaining NMFS backlog and current data as well as data from BOEM, NMS, PMEL, NPS, and possibly Navy**, and expanding our ability to handle all the components of mobile platforms are next on the list.

NCEI FY18 Plan

To begin the process of gaining NCEI leadership's approval to operationalize the passive acoustic archive (thus far our efforts have been really one year demos) with the goal of obtaining monetary support from NCEI for the basic stewardship activities that are core to NOAA and NCEI's mission, **I drafted a Project Plan and presented it to NCEI's Program Management Committee**. There are details that still need to be worked out before it heads up to NCEI's Executive Board for final approval this is due to the complication of cross-divisional resourcing within NCEI that needs to be established. The resourcing is being looked and I expect to have feedback by the end of October. The very good news is that **it is accepted that Tier 1 stewardship, our basic stewardship activities, (i.e., Chuck's salary) should be covered by NCEI Base funds.** Getting buy-in is a great first step. Getting the funds to back that up is another story. Other projects, tasks, and critical shortfalls throughout NCEI's four locations need to be

considered when distributing our limited base funds. However, I will continue to push hard on the value of this project.

I finally got the cost estimate numbers. Here is a detailed breakdown:

Passive Acoustic Data Cost Estimates	Tier 1 Costs		Typical Tier 1 Costs	
	1st Year	Annual	1st Year	Annual
Scenario: NCEI-NC CS (standard cost model)	$72,400	$16,200	$109,400	$16,200
Archival Stewardship & setup - NCEI Data Stewardship Division labor (1 product)	**$8,200**		**$9,800**	
Tier 1 - Hardware & Media - operational, 285 TB / yr	**$38,300**	**$16,200**	**$38,300**	**$16,200**
Tape drives / disks / licenses	$22,100		$22,100	
Tape media	$16,200	$16,200	**$16,200**	**$16,200**
Tier 1 - Hardware & Media - Historical, 457 TB	**$25,900**		**$61,300**	
Tape drives / disks			$35,400	
Tape media	$25,900		$25,900	

The “Tier 1 Costs” column is relevant to this project. The “Typical Tier 1 Costs” column notes what typically occurs. The “1st Year” column under this heading is for the first year of the project. These costs cover getting the archive fully set up. The “Annual” column is an estimate for subsequent years. The $16,200 under “Tier 1 Hardware & Media - operational, 285/yr” “Annual” assumes that another 285 TB of data will be archived at NCEI that near. Note the huge savings for “Tier 1 - Hardware & Media - Historical, 457 TB”. By leveraging existing capacity to accommodate the backlog data, we are able to save about $35k. Hooray!

Assuming a linear scale, **the cost per TB for the first year of current data is $134. Years following it is $56. For the backlogged data, the cost is also $56**.

You may be wondering, why are there costs to set up the archive when Chuck and I just presented the one we developed? Due to NCEI’s recent initiative (as decreed by NESDIS), we need to set up ALL our archives in NCEI-North Carolina (NCEI-NC). There are staff there to help and these costs reflect that. We will make use of previous work to keep the additional set up costs minimal. **Shown are estimates and additional technical conversations need to follow to figure out those details. The major benefit of using the NCEI-NC system is the greatly reduced**

media storage cost. If you recall, we first quoted $33,000 for just 75 TB of data (equivalent to $440 per TB). The long term gain will be well worth it.

I want to point out that these values, especially the "annual" cost are for the NCEI-NC component and do not cover CIRES staff (i.e., Chuck, myself, and a percentage of time from our fantastic GIS guy). I am working on having Chuck's salary covered by NCEI but just wanted you to be aware that these estimates show *most but not all* of the expected final cost.

The assumption of the media storage costs is that 285 TB can be archived at NCEI each year. Chuck and Veronica, our student intern, ran some numbers on our previous archive rates and identified that **215 TB per year is at the high end of our maximum intake.** So it's definitely not feasible for us to handle 285 TB in a year. We'll be aiming closer to a max of 200 TB each year. **The good news is that these are the compressed values - after FLAC - so the uncompressed volume of data sent to us for archive will be ~50% larger.**

Big Data Project

We discussed NOAA's Big Data Project (BDP) as a possible avenue to assist with accessing and analyzing large volumes of passive acoustic data for broadscale soundscape statistics. **I will present this concept to the BDP cloud source providers (e.g., Amazon, IBM, Google, Microsoft) on Friday November 3.** This follows Jonathan Blythe's presentation on October 6 that included the application of using cloud services to analyze passive acoustic data from the perspective of BOEM. Jason Gedamke and Kurt Fristrup will be joining me on the BDP call in November. Please let me know if you'd also like to join. I will circulate my slides prior to the call and will let you know how the presentation was received.

Data Projects

Another avenue to assist with data analysis is the Hollings Scholar program. This scholarship brings talented undergraduate students to NOAA facilities to conduct research over a 9 week period in the summer. NCEI has had a lot of success working with Hollings Scholars on various projects in the past. Jason and I submitted a project last year but did not receive interest from the perfect candidate. **We have submitted a proposal again this year in collaboration with Kurt to implement some of his scripts, develop visualization products of the archived data, incorporate the products into the map viewer, and create ESRI story maps.**

DTAGs and Animal Telemetry Network

The importance of ensuring the passive acoustic data remain connected to the telemetry component of the DTAG data was raised during the workshop. Through IOOS, the Animal Telemetry Network (ATN) has been developed. There is nominally an ATN Data Assembly Center (DAC), and the expectation is that NCEI will get data from them. From what I understand the effort is in its infancy and working with individual providers while the DAC gets sorted out shouldn't be a problem. If you'd like more information on the current status, please check out the report from the ATN Steering Group Meeting in June. There is also cross-pollination with NMFS PARR from ERDDAP to NCEI. ERDDAPs animal telemetry holdings are available here. There are several ongoing conversations but **since NCEI will ultimately house all the ATN archived data we will have some leeway in how it is presented. This will allow us to ensure the two datasets are closely linked and easily accessible.** I'll stay apprised of this topic as it develops.

Group dinner at the all too noisy The Med restaurant Thursday night.