Summary of Workshop: Data Automation & Data Management Options

April 1, 2020

Prepared by:
Marine Renewables Canada

Prepared for:
Offshore Energy Research Association (OERA)
Fundy Ocean Research Center for Energy (FORCE)

Summary of Workshop: Data Automation & Data Management Options

1.0 BACKGROUND

The Offshore Energy Research Association (OERA) has a mandate to enable sustainable development of Nova Scotia's energy resources by facilitating and funding collaborative research and development. It has supported numerous tidal energy R&D projects over the years and is now leading the Pathway Program in collaboration with the Fundy Ocean Research Centre for Energy (FORCE) with funding from Natural Resources Canada (NRCan) and Nova Scotia Department of Energy and Mines (NSDEM) – a coordinated R&D program that will define, test and validate Environmental Effects Monitoring (EEM) solutions for the instream tidal energy industry that can meet regulatory requirements. The program will increase the understanding of impacts from instream tidal energy projects in the Bay of Fundy and improve the understanding of fish and marine mammal interaction with instream tidal energy devices. The program will also improve data processing and analyses, so that results can be reported to regulators and disseminated to the public in a timely manner.

The main objectives of the Program are to i) define a DFO-approved solution for the tidal energy industry, ii) apply machine learning to data analysis to reduce reporting time and compliance costs, iii) minimize initial capital costs to developers, iv) develop regional capability to manage, process, analyze and report EEM data, and v) develop intellectual property that regional companies can exploit commercially in multiple marine industries, both regionally and globally.

To conduct this program successfully, OERA and FORCE are assessing different types of monitoring technology that can gather adequate data to inform regulatory requirements. To complete this assessment effectively, OERA and FORCE are consulting with experts through several workshops to gather information on how the technology may be able to work effectively in high-flow environments to gather required monitoring data.

The second workshop under the Pathway Program was focused on "Data Automation & Data Management Options." The Pathway Program contracted Marine Renewables Canada (MRC) to assist in leading workshop development, delivery and information-gathering. This summary report is the outcome of the workshop discussion and insights gathered during the workshop held on March 4, 2020.

2.0 WORKSHOP FOCUS

Instream tidal energy turbine projects, specifically in Nova Scotia but also elsewhere in the world, have an obligation to the local community and to regulatory bodies to provide environmental monitoring of their projects to ensure that there is no undue harm to marine animals or permanent alteration to habitat. This is critically important in ecologically sensitive and historically significant areas like the Bay of Fundy.

One of the greatest challenges facing monitoring of tidal energy devices is the tremendous amount of data generated from multiple types of monitoring instrumentation (i.e., active and passive acoustic devices). Much of the information gathered is 'data rich/information poor' (DRIP) and does not contribute to understanding effects of tidal turbines on the marine ecosystem. The tidal energy industry needs methods for dealing with this data more efficiently and effectively. The Pathway Program has identified echosounders, imaging sonars and passive acoustic monitoring instruments as technologies that are required to effectively monitor tidal energy turbines. However, each of these technologies require separate approaches for data automation and management. Moreover, these instruments are intended to be integrated on a cabled sub-sea platform and need some method (e.g., pre-processing) to filter and remove uninformative data prior to data storage (i.e., a control module). This is all in an effort to reduce costly 'data mortgages' for developers and improve the timely reporting of monitoring results to regulators.

The objectives of the workshop were to:

- Provide an overview of current data automation processes and existing algorithms for differing monitoring technologies (i.e. echosounders, imaging sonars, Passive Acoustic Monitoring (PAM) technologies),
- Discuss 'on-board' data management systems (e.g., control modules),
- Identify challenges and solutions to automation with multi-instrument platforms, and
- Discuss strategies to reduce 'data mortgages' and transition from 'data rich/information poor' monitoring to collecting meaningful data for understanding the effects of tidal turbines on marine animals.

3.0 WORKSHOP FORMAT

The Data Automation & Data Management Options workshop was held on March 4, 2020 and was facilitated by Kes Morton of Pisces Research Project Management Inc. Experts and industry representatives with knowledge/experience of data collection, automation and management and tidal energy development (Pathway Program partners) were invited to attend. Participants were provided with a backgrounder document and agenda in advance of the event that outlined the problem statement, objective, and desired outcomes:

Problem Statement: What data automation strategies and data management tools are available for dealing with enormous amount of data generated by integrated monitoring platforms.

Objective: Explore options for automating and pre-processing active and passive acoustic monitoring data and discuss options for data management.

Desired outcomes:

1. Identify solutions for automating the analysis of imaging sonars and PAM data
2. Identify solutions for pre-processing data from multiple integrated active and passive acoustic monitoring instruments

The workshop included brief presentations on key issues pertinent to the topic, followed by discussion from all participants. The workshop participants were:

Attended in-person:

- David Barclay – Dalhousie (Oceanography)
- Craig Chandler – SME
- Luiz Faria – OERA
- Meghan Flood – OERA
- Dan Hasselman, FORCE
- Ines Hessler – Meridian
- Jennifer LaPlante – DeepSense
- Scott Lowe – Dalhousie / DeepSense
- Bruce Martin - JASCO Applied Sciences
- Louise McGarry, FORCE
- John Moloney - JASCO Applied Sciences
- Jason Newport – DeepSense
- Mahtab Sarvmaili – Dalhousie / DeepSense
- Sarah Thomas - DP Energy
- Chris Whidden - DeepSense
- Benjamin Williamson - University of Aberdeen/ University of Highlands and Islands
- Mark Wood - Ocean Sonics
- Tony Wright – FORCE
- Lu Yang –DeepSense

Attended remotely:

- Ben Bifford, Ocean Networks Canada (User Engagement)
- James Joslin, University of Washington (AMP)
- Alex Lam
- Lilian Lieber, Queen's University Belfast
- Mike Morley, Ocean Networks Canada (Data Manager)
- Paul Murphy, University of Washington (AMP)
- Benoit Pirenne, Ocean Networks Canada (User Engagement)
- Ray Pieroway, SME (Lead Technologist – data acquisition)
- Damian Rohraff
- Adrian Round, Ocean Networks Canada (Director of Operations)
- Pal Schmidt, Queen's University Belfast
- Greg Trowse, Luna Ocean

The next section provides a summary of the input gathered and key points and/or questions raised.

4.0 SUMMARY OF DISCUSSION

4.1 Introduction to workshop

The Pathway Program is a collaborative effort between industry, academia and regulators. Its focus is on defining, testing, and validating environmental effects monitoring solutions that will be acceptable for regulatory requirements of in-stream tidal turbines.

The Program includes three streams:

- Global Assessment – Examine and understand the global state of the science on environmental monitoring for instream tidal energy.

- Data Management – Establish an understanding and solution for how to manage the mass amount of data from instrumentation such as echosounders, imaging sonars, and PAM devices – and report results to DFO in timely fashion (quarterly basis)

- Technology validation/Field Trials – Test different technology that can be used with a platform including echosounders, imaging sonars, PAM, etc. Most testing will take place in the Bay of Fundy, with some activity in Grand Passage.

With three different types of technologies – imaging sonars, echosounders, & PAM – a large amount of data is generated. Once the data is collected it must be transferred onshore and processed. The development of algorithms is needed. The workshop discussed who will manage the data, process it and analyze it.

Workshop problem statement & objective
Responsible development of instream tidal energy in Nova Scotia and Canada requires us to monitor interactions between animals and tidal energy turbines. In reviewing and learning about the global experience around this issue, encounters of marine animals with tidal turbines occur infrequently. However, in order to detect those rare events,, a continuously operating monitoring program is required. The monitoring program will need to provide continuous data acquisition from multiple medium- and high-bandwidth instruments. The complexity of this issue is compounded when multiple instruments are included in an integrated monitoring package. Environmental monitoring solutions will need to ensure that data-rich information is transitioned into meaningful information that can be used by developers and regulators.

The Pathway program has identified a suite of environmental monitoring instruments that will be integrated onto a platform – echosounder, imaging sonar, and PAM. Each of these technologies requires separate approaches for data automation and management. Each provides different types of data and require their own approach.

Experience has been obtained from international integrated monitoring platform development programs that can inform the Pathway Program. Some solutions have been developed including: control modules to link activation of monitoring instruments; pre-processing to filter uninformative data prior to storage; and algorithm development to automate data post-processing.

The broad objectives of the workshop are to:

- Provide an overview of current data automation processes and algorithms for different monitoring technologies.
- Discuss 'on-board' data management systems.
- Identify challenges and solutions to data automation with integrated monitoring platforms.
- Discuss strategies to turn off the 'DRIP' and collect meaningful monitoring data for understanding effects of tidal turbines on marine animals.

Through presentations and discussions, the Pathway Program team aims to 1) identify solutions for automating analysis of imaging sonars and PAM data and 2) identify solutions for pre-processing data from multiple integrated active and passive acoustic monitoring instruments.

4.2 Strategies to reduce 'data mortgages' and transition from 'data rich/information poor' monitoring

- **Ocean Networks Canada**

ONC operates infrastructure for ocean science by building the infrastructure and data systems etc. to allow others to do research. ONC has familiarity with a number of similar instruments to what the Pathway Program has identified including: ADCP – RDI NORTEK, Kongsberg Mesotech Rotary Sonar 1171, Imagenex 881A Imaging Sonar, Imagenex 881A Profiling Sonar, COVIS Sonar (modified RESON multibeam Echosounder), ARIS/DIDSON imaging sonars (fish), ASL AZFP Echosounder, BioSonics DT-Xu Scientific Echosounder, and icListen hydrophones AF, LF,HF.

ONC's data collection philosophy is different than the Pathway Program as they do not limit the amount of data that initially come ashore. Its operations are not power or bandwidth limited, so ONC is able to collect high resolution raw data. All processing takes place onshore which allows for easier to upgrade equipment etc. This data processing includes:

- Scalar data processing (live):
  - Parsing for scalar measurements: temperature, salinity, etc.
  - Unit conversions, calibration formulae
  - QA/QA tests and flagging
  - Scalar data event detection – set detection or alert parameter

- Complex data processing (near-live) – stream the live data and post-process it ashore:
  - Production of various data products, particularly manufacturer's formats and other intermediate formats for speed and convenience.
  - Some file-based quality metrics (Video data QA/QC), Annotation generation.

ONC produces daily plots of most data but not all "complex" data – first step of visualizing data (*see slides for examples of daily plots*). Data products are primarily produced on-demand, on-the-fly in ~275 formats from the file archive and database (*see slides for visual samples).*

- ONC Data Collection Path includes the following:
    - Instruments on platform with power and communications path to shore
    - Instrument Driver at shore station which communicates with instrument (configuration, command and data capture). “Smarts” at the shore station – where they can be modified easily if needed.
    - Parser and QA/QC functions – all linked together with message bus
    - Message Bus to transmit data to Data Centre and commands to instruments/infrastructure.

- Automated detection capacity at ONC:
    - PAMGuard underway for automatic detections of Southern Resident Killer Whales in hydrophone data streams
    - Capacity and instrumentation for Earthquake Early Warning:
        - Accelerometer and GNSS data streams combined and processed in “mini shore station” at each site.
        - Detections sent to Event Correlation System at Data Centre

        See slides for example diagram.

ONC does not have many examples of directly processing close to the instrumentation. Its aim is to collect *all* data in case it is needed in the future.

- **University of Washington (Adaptable Monitoring Package)**

The Adaptable Monitoring Package (AMP) has been in development at University of Washington’s Pacific Marine Energy Center for the last nine years. The motivation for the AMP arose from **a** project aiming to deploy two tidal turbines in the Puget Sound in Washington. A plan was needed for the required environmental monitoring because at that time, very few options were available compared to now – a new system needed to be developed. The Applied Physics Lab at UW had been previously involved with the design and installation of the Cabled Array off the Washington/Oregon coast, which is similar to ONC, and this experience was leveraged in the AMP development. A key part of this work was to develop strategies to use the mass amounts of data in a meaningful way.

The AMP uses a suite of off-the-shelf instruments with an integration hub that allows them all to be synchronized and connected by a single cable to shore. Instruments have been used include two optical cameras, two imaging sonars, an echo-sounder, an ADCP, a hydrophone array, a water clarity sensor, and a fish tag receiver.

A number of strategies have been implemented to reduce Data Rich Information Poor (DRIP) Issues. The following directives have helped guide the development:

1. Monitoring Directives
    - Avoid biasing behavior of marine life, or you will only be monitoring yourself
    - Detect rare events / events of interest (if these events happen, it will be rarely)
    - Do not accrue un-manageable data / data mortgages – accruing data faster than you can process it

2. System/Software architecture
    - Computer that sits onshore
    - Data processing predominantly is conducted onshore
    - Connected by umbilical that leads to integration hub that connects to all instruments

3. Data Mortgages Issues - the instruments
    - ADP - Accruing a lot of data – but still manageable (not a high band width instrument)
    - Hydrophone array (PAM)– adds a significant amount of data – approx. 225 GB/day
    - Sonars – adds another 600 GB/day
    - Optical cameras – adds approx. 5 TB/day
    - If all instruments run continuously, it would be more than 5.8 TB/day – and this doesn't include other instruments on there (echo-sounder)

4. Data Acquisition methods – three options available, and the problems with each
    - Continuous acquisition – generates over 5 TB/day
        i. Leads to data mortgage, fails directive 3
        ii. Can cause behavioral response from lights, some acoustics, fails directive 1
    - Duty cycle acquisition
        i. Reduce data linearly by fraction of time acquiring
        ii. High likelihood of missing rare events, fails directive 2
    - Triggered – user lower band width instruments to trigger higher band width instruments
        i. Can reduce data archival
        ii. Provides initial processing for reviewers
        iii. Requires real-time processing modules to be happening in real time with the instruments and that all and instruments are well synced
        iv. This is the only solution that can satisfy all 3 directives

- **Key discussion points and take-aways:**

    - **Engagement with key stakeholders:** In order to ensure that the best solutions are pursued and developed in terms of regulatory requirements and acceptability, it is important for DFO and fishers to be an active participant in this work. Engagement and outreach with fishers with regards to this work and how data is being managed and analyzed will also be important.

    - **Strategies for data gathering:** There were two strategies discussed for data gathering – 1) collect all data possible (ex. ONC) and 2) Minimize the amount of data collected/data mortgage (AMP). ONC isn't limited by bandwidth which allows it to collect as much data as possible. At FORCE and in the Bay of Fundy, bandwidth will be an issue and therefore there is a limit on the data that can be collected. At the same time, the more data collected may satisfy the concerns and future needs of regulators and fishers.

While gathering data, it will be important to think of two sets of metrics and requirements:

1) Compliance and communication: Need to understand what is needed/required.
2) Science: Everything needs to be collected and recorded.

DFO/regulator role should be to communicate the metrics they require and assist in funding the science to meet those requirements. However, given the nascent nature of instream tidal energy, it is challenging for regulators to define exactly what is required. Currently, they are asking that developers provide advice/input and work with them to define requirements. A continuous dialogue and flexibility on these issues is needed.

- **Data management resources (human vs. technology):** Depending on the amount and type of data that is being collected there will be different needs and options with regards to the work being done by people vs. machines/technology. At ONC about 40% of staff work on the data side, but that is because they are collecting so much data. At AMP, the method for management varies depending on stage of deployment and logarithm development. Once the system is deployed and operational, human management is minimal (ex. AMP would use one person a day for a couple of hours). During the initial phase (tuning) more human resources would be required.

- **Triggered acquisition can reduce empty data**: By using lower band with instruments to trigger higher band with instruments, the amount of empty data transmitted is reduced. This strategy provides an initial processing for reviewers but the strategy also requires that all instruments are well synced and real-time processing modules.

4.3 Overview 'on-board' data management systems (control modules) and challenges for multi-instrument platforms

- **Data Management and Control for Multi-Instrument Platforms**

The characteristics of FLOWBEC (onshore post-processing) and AMP (onboard processing) were reviewed, illustrating the different methods for data management depending on onshore vs. onboard processing.

- **FLOWBEC – onshore post-processing**

FLOWBEC aims to monitor the entire water column (plankton, fish, seabirds, marine mammals) collecting both physical and ecological measurements. It operates continuously including across a spring-neap tidal cycle which allows for predictability and transferability of results between sites. To satisfy regulatory requirements while reducing the amount of monitoring required, FLOWBEC focuses on concurrent information on:

- Hydrodynamics - ADV & ADCP
- Animal distribution & identification - EK60 echosounder (multiple frequencies)
- Animal behavior - Multibeam sonar (predator-prey and animal-turbine interactions)
- Collect contextual data using a Fluorometer / turbidimeter
- Tried integrated a camera with limited results
- Integrated hydrophone

FLOWBEC targets co-registration because it allows for the most robust data set to be developed when co-registering between instruments. The control system is transferable to either an autonomous or cabled configuration:

- Autonomous configuration
  - 66+ kWh on board batteries for 2 week – 3 month deployment, depending on sensors
  - Everything recorded initially, delivering outputs to regulators and continuously develop robust algorithms in the background
  - Allows for deployment before/after or control/impact studies, floating or seabed turbines
  - Rechargeable in 24-h neap window
  - Recovery using small ROV to attach lift line

- Cabled configuration
  - Cabled to a structure (e.g. tidal turbines, wind turbine pilings)
  - Realtime data, longer endurance

FLOWBEC's onboard data management and control have several requirements. The central controller (Figure 1) provides:

- System health monitoring (temp, V/I, humidity, leak)
- Resiliency (fault tolerance), redundancy control
- Clock synchronization (drift)
- Ping scheduling (cross talk) – seeing same target on different instruments
- Modular expandable, Plug-n-Play using unified HW/SW interface – can swap in different instruments as needed
- Duty cycling / triggering (high power or audible instruments / visible lights)

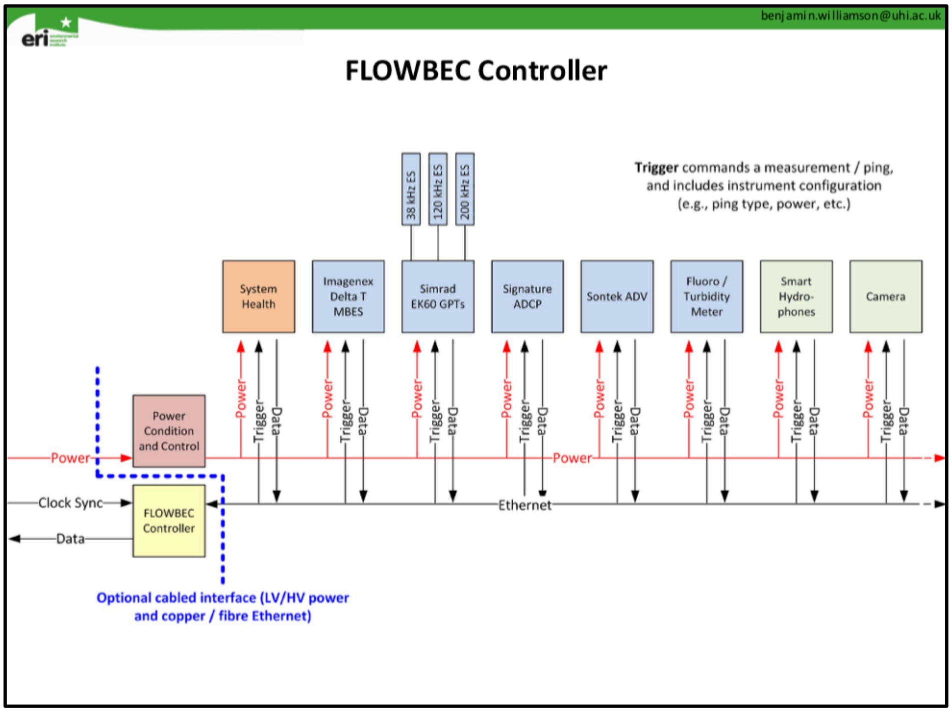

Figure 1: Schematic of the FLOWBEC central controller

The main differences between autonomous and cabled mode are that the autonomous method is powered with a battery and data parsing uses storage. The cabled method uses power conditioning and data is transmitted via cable.

Ping scheduling is done to mitigate cross-talk. FLOWBEC uses a flexible, modular and expandable interface that allows for real-time scheduling of different instruments and configurations. It can use lower power instruments for detection and switch to high power for identification. High power instruments or audible/visible instruments can be duty cycled.

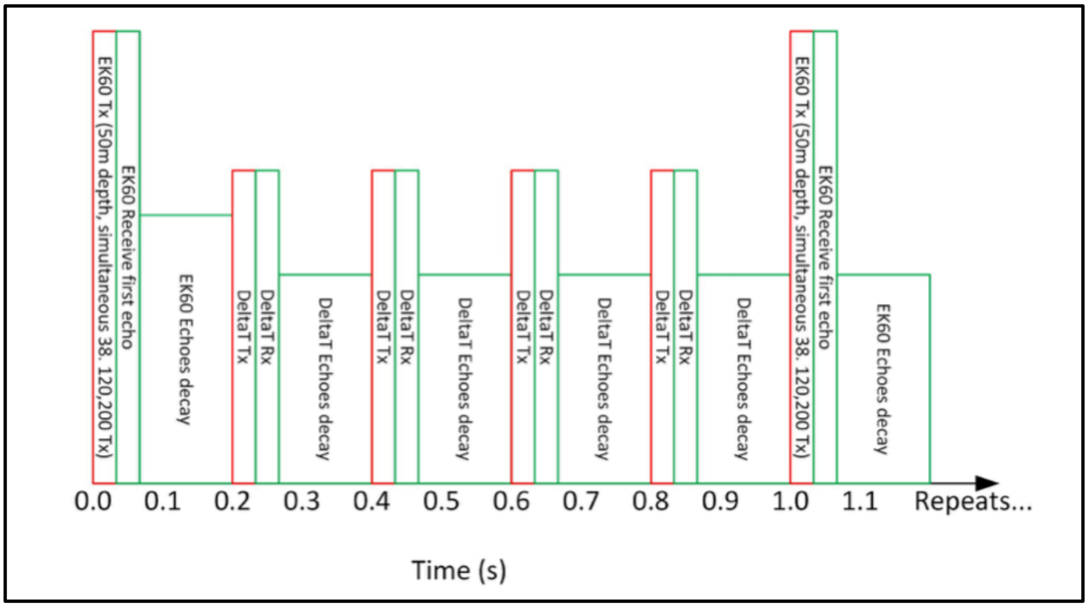

Figure 2: Schematic depicting an example of a duty cycling schedule implemented on FLOWBEC.

- **University of Washington (AMP/WAMP – on-board processing)**

The **Integration Hub (hardware backbone of AMP system; Figure 3)** is designed to be as modular as possible. The AMP main electronics bottle incorporates:

- Off-the-shelf Ethernet DAQs for adaptable control
- 10 output instrument ports configurable for 12, 24, or 48 VDC and serial or Ethernet communications
- Electrical isolation relays for each instrument port – isolation for ground fault control and power filtering to reduce the amount of noise
- Configurable trigger lines to/from each instrument port and DAQ
- High precision serial IMU
- Optional high voltage DC power input (200 to 400 VDC input)

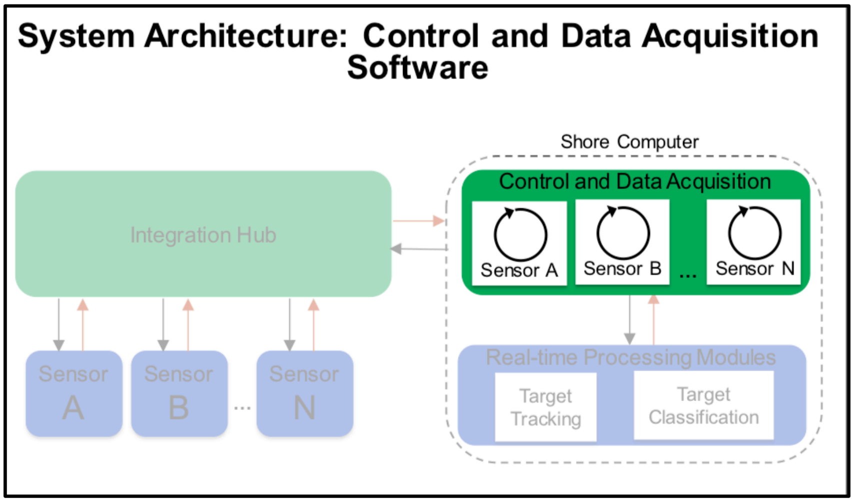

Figure 3: System architecture of the control and data acquisition software implemented in the AMP platform.

The Control and Data Acquisition Software has a number of aspects covered. For most instruments, data is recorded into a ring buffer and the last 15-60 seconds of data is kept in temporary memory. If an event of interest occurs and data is needed to be written to disc, a trigger is generated. Each instrument manufacturer has their own software packages for autonomously operating their instruments which makes it difficult to use every one of these packages independently as they all work differently. The AMP team developed its own software for managing this issue. There were a lot of benefits to doing this as it makes each instrument more modular. The software covers a number of different settings: Nexus System (Figure 4), Health Monitoring (Figure 5), Acquisition Setting (Figure 6).

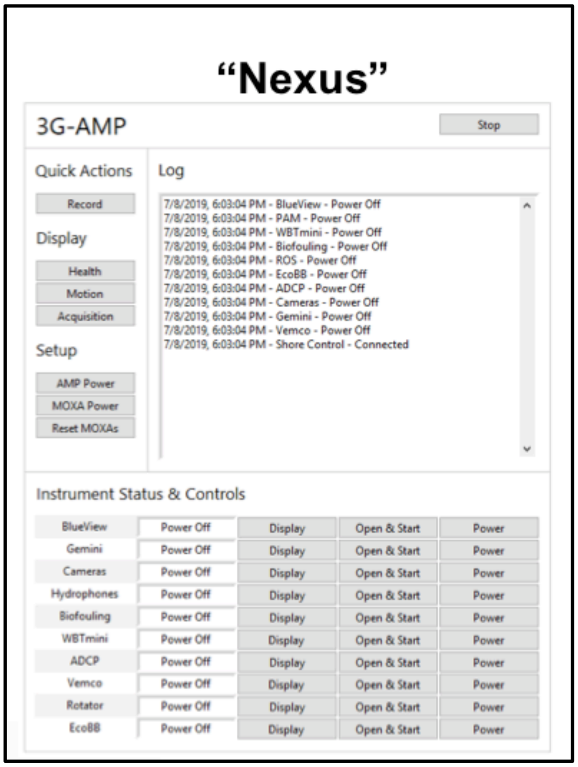

Figure 4: Nexus

Nexus features:

- Commonly used controls
- System log and status indicators
- System and instrument power controls
- Launch and display all other system and instrument control interfaces

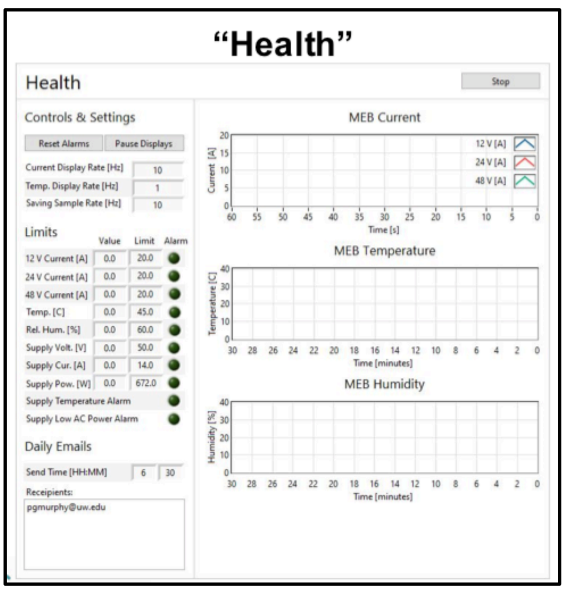

Figure 5: Health Monitoring Buoy

Health monitoring features:

- Power, temperature, and humidity monitoring. Automatically powers down the system when programmed limits are exceeded.
- Interfaces with power supply hardware to detect over-current and over-power conditions.
- Alarm emails
- Daily status emails

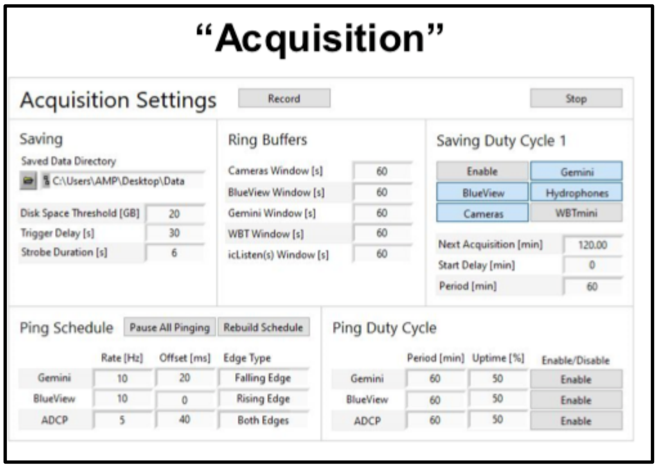

Figure 6: Acquisition Setting

Acquisition features:

- Generates "save triggers" in response to manual instruction (e.g., "Record" button), duty cycles, or messages from external software (e.g., detection software)
- Ring buffers
- Duty cycles
- Ping scheduling

A common data directory structure and file name convention was developed for all instruments. When data is saved, a trigger is generated that will automatically create a file structure:

- Level 1: "Date"
- Level 2: "Date Time"
  - a "buffer" folder
  - created when any "save trigger" is generated
- Level 3: "Instrument"
  - created when an instrument's "save trigger" is generated
- Level 4: "Files

There are a number of **challenges for multi-instrument platforms** around electrical noise and ground faults. The AMP work addressed these:

- **Challenge:** Electrical noise generated by electronics (e.g., switching noise from DC-DC transformers present in integration hub and in instruments) can produce artifacts in instrument data.
  - **Solution:** Power conditioning should be included at all electrical junctions.

- **Challenge:** Ground faults in any subsystem can overload the system's power supply.
  - **Solution:** Thorough pre-deployment testing in salt-water. Subsystems electrically disconnected when powered down.

- **Challenge:** Acoustic emissions from an active acoustic instrument can manifest as noise in the data of an adjacent instrument (cross-talk).
  - **Solution**: Software "ping scheduling". Ping schedules are manually configured while reviewing live feeds from instruments.

- **Challenge**: Corrosion is an issue as a lot of the instruments are not meant for long tern deployment in this type of environment.
    - **Solution**: Dissimilar metal and andoic protection where isolation is not possible.

- **Challenge**: Biofouling can be hard to manage.
    - **Solution**: Have been successful in using a combination of wipers and UV lighting, antifouling paint, and copper.

- **Key discussion points and take-aways:**
    - **"Keep smarts onshore"**: Due to the inherent challenges of working in the marine environment, including corrosion and biofouling, computing and controllers are designed to be operated on short as much as possible.

    - **Robust training set is critical**: It is important to ensure that there is robust data that a model is learning from in order for consistency of results between sites and conditions.

4.4 Overview current data automation processes/algorithms for differing monitoring technologies

DeepSense, University of Aberdeen, and University of Washington provided overviews of data automation processes/algorithms for several different monitoring technologies – echosounders, imaging sonars, and PAM, respectively.

4.4.1 Echosounders (overview automation for echosounder data) – DeepSense

DeepSense is an organization focused on bringing together ocean organizations and helping support them to complete machine learning, AI and visualization projects. DeepSense has been working with FORCE and OERA to support automation for echosounder data. The aim is to analyse data to remove noise and to learn from it. This work is being conducted in three phases:

- Phase 1A & Phase 1B: Focus is on analyzing data from EK80 Echosounders.
  Phase 1A focused on mobile data collected over 24 hours. Eighty hours of manual processing as done to clean the data. DeepSense is developing a machine learning algorithm to reduce this time while maintaining the accuracy of manual post-processing. For Phase1B the focus is stationary data (sensor mounted to seabed) and months of data is collected and processed manually.

- Phase 2: Focus on developing an automated reporting template.
  The template reduces efforts for analysis preparation. In order to manage steps efficiently that occur during manual workflow, some steps need to be automated. Below are the key steps to be automated:

Manual Workflow	*Automated Workflow*
1. Collect Data 2. Calibrate Data – using salinity, temperature, etc. 3. Load Data into Echoview 4. Manually Clean Data 5. Export Binned Data 6. Manually Analyze Data 7. Manually Write Report	1. Collect Data 2. Calibrate Data 3. Load Data into Echoview **4. ML Model to Clean Data** 5. Export Binned Data **6. Automated Data Analysis** **7. Automated Draft Report** 8. Update Report

The ML Model process developed as part of the Automated Workflow essentially aims to replicate what a human does. Following are the ML steps:

- Collect annotated data
- Export out of echoview - raw & clean images; lines
- Split into partitions – training (70%)/validation (15%)/test sets
- Dataloader for loading pre-processed segments of data, in a random order
- Data augmentation
- Objective function design
- U-Net model in Pytorch (a Python Deep learning framework)
- Hyperparameter optimization

Once regions are excluded through the ML process, the data that is considered the 'valid=d' data, and is split into either columns or blocks (bin data). Reporting goes towards how much signal is seen in these column or blocks. You report how much signal you see in these columns or blocks. This supports the analysis to determine how often and where fish are seen (i.e. certain times, when turbines are running, where in the water column, etc.). The goal of an automated report is to 1) automate postprocessing and analysis of continuously collected echosounder data; 2) get standardized reporting, statistics, and figures; and 3) generate draft quarterly reports for manual testing. Following are the steps for developing the Automated Draft Report:

- Create a set of data standards – file names, variables etc.
- Eliminated manually having to analyze the code. This is automated smart processing, error handling, and smart resume
- Select statistics and figures of interest – what is the data that we need to show DFO
- Implement chosen analysis
- Draft reports
- Improve draft reports with stakeholder feedback

DeepSense will be making improvements to the model including automating processing of stationary surveys and analysis of results, as well as generation of quarterly reports.

- **Key discussion points and take-aways:**

  - **Time savings and efficiency**: To complete one training set using the automated workflow process that DeepSense has created takes about 20 seconds – a considerable time savings.

  - **Cross-validation:** Both FLOWBEC and AMP use cross-validation as tools – multiple instruments are used to determine what's in the water column. AMP uses the higher bandwidth imagery or acoustic data to classify targets. Lower bandwidth instruments are used to trigger the higher targets. FLOWBEC uses various instruments together. JASCO has been looking at flow noise at different heights through different instruments. It conducts a cross comparison between algorithms.

4.4.2 Imaging sonars (overview of algorithm development)

- **FLOWBEC Imaging Sonar – Overview of Algorithm Development**

FLOWBEC begins work with echosounders as that informs imaging sonar work. The objectives of using imaging sonar are to 1) detection, tracking, classification and kinematic metrics to monitor (ex. fish, bird and marine mammal distribution, animal behaviour, and interactions with marine energy devices) and 2) explanatory variables – investigating the predictability and transferability of results between sites.

FLOWBEC uses co-registration of multibeam and the EK60 for detection (Figure 7). The multibeam provides a 2D perspective – showing target velocity, behaviour, predator/prey interaction, turbine encounters, evasion, school morphology. However, the detection and discrimination of ecological targets from physical targets, is more challenging for multibeam because there is only one frequency. The EK60 (essentially an EK80 with narrow bandwidth) has high sensitivity and calibrated frequency response (ID) so absolute backscatter (abundance) can be inferred rather than just relative. The morphology of turbulence can be measured.

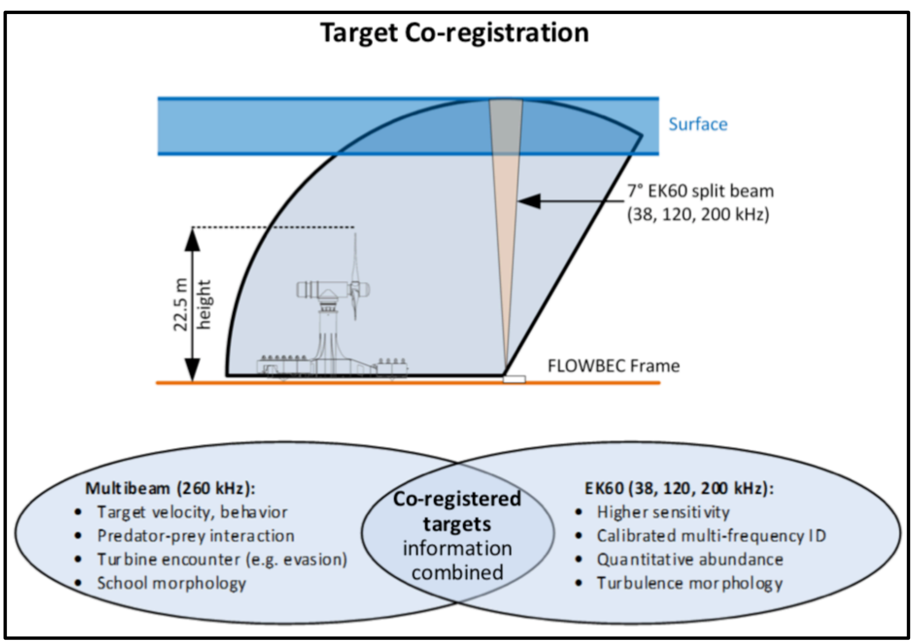

Figure 7: Example of target co-registration process on FLOWBEC system

The process for co-registration of multibeam and the EK60 includes the following steps:

- Step 1 – Detection – can be informed by echosounder detections
- Step 2 – Track – multibeam imaging sonar – images vertically through the water column to measure morphology over time, movement, speed, vertical distribution/evasion
- Step 3 – Classify - using frequency response

Backscatter from physical sources such as sediment and air bubbles does occur. 98.9% of back scatter is from physical sources and a very small percentage is actually an ecological target. FLOWBEC uses and adaptive processing approach (echosounder and multibeam) to better identify fish in the water column. This includes:

- Turbulence delineation
- Additional scale-dependent filtering
- Target morphology requirements
- Multi-frequency validation

Multibeam processing requires a few more steps. FLOWBEC uses a multibeam to allow it to operate in autonomous mode. The steps for processing include:

1. Noise detection and filtering
2. Adaptive suppression (3D median filter) – sites change (with wind, waves, tidal cycle etc.), so a window is set that moves through and adaptive filtering and tries to filter out this background noise.
3. Segmentation, object detection
4. Morphology and intensity requirements
5. Tracking algorithms (currently being revised):
    a. multi-hypothesis
    b. global nearest neighbor
    c. voting in/out (intermittent detection) – dealing with a noisy site with imperfect sensing.
6. Classification
    a. Size, movement, number of targets – fish schools & diving sea birds
7. Co-registration with EK60 – real strength of the multi-instrument platform.

When a target is detected on both instruments, the most information possible is gathered and leads to the most robust detection.

When co-registering targets, FLOWBEC sets thresholds to associate between the two different instruments. Different frequencies and different modalities of sensing are used, so the returns will differ, but have differing capabilities under changing conditions. Ultimately, the goal is to identify a target in the water column at approximately the same place, time and with the same characteristics. This multi-dimensional association approach works to say it's the same thing on two different instruments. To assist with this work, FLOWBEC filters by turbulence boundaries (running with an

EK60 it can mask out the turbulence at the surface) and uses multi-frequency validation (setting thresholds based on 30 years of hydroacoustic fisheries information from the literature).

Multibeam validation is performed using cameras and passive acoustics where possible. Instruments are also taken to other high flow tidal sites where there aren't turbines and put on vessels, using video for validation of targets to build up a training dataset.

- **AMP/University of Washington Imaging Sonar**

The AMP uses multibeam sonars with optical cameras for classification rather than echosounders. All instruments run continuously, saving data in 60 second ring buffers. Data goes through a pre-processing step that basically subtracts the rolling background to get the foreground images for each of the sonars. Data from this process is used to pull out and classify targets and send triggers for archival storage.

The University of Washington has been testing in Sequim Bay, Washington which is a moderate tidal site. During a 2017 deployment, researchers were able to conduct real-time processing using the BlueView multibeam sonar to detect different targets. During this deployment, a duty cycle data set was simultaneously collected to develop statistics on capabilities of the target detection and tracking method. A true negative rate of 99% was able to be demonstrated by looking at the duty cycle and any targets that might have been missed. The true positive rate was 58% -- essentially illustrating how many of the targets detected were actually targets. In comparative acquisition to the 15-minute duty cycle acquisition, the data mortgage was able to be reduced from approximately 170 GB (of close to empty data each day) to 28 GB which was a very significant reduction. The University of Washington's goal is to now reduce this data mortgage even further.

A classifier has been developed that distinguishes between different targets (ex. seals, diving birds, fish schools, small targets, non-biologics). The classifier uses a several different features of targets to establish the results (shape, intensity of the acoustic target, motion). ADCPs are also used to detail what currents are doing simultaneously with the targets.

In 2019, the University of Washington conducted additional research at a new tidal site to validate the results of the classification system. Despite the fact that the different sites were used in 2017 and 2019, where only 150 m apart, the model was negatively impacted as it received much higher false positive rates, complicating the triggering. It is, therefore, very important to re-train a data set between sites and different conditions. A key take-away is that while it is important to detect and classify these targets, the easiest and most accurate method is to classify biological and non-biological targets to determine whether to or not to archive the data -- this can reduce your data mortgage.

- **Key discussion points and take-aways:**
    - **Multi-frequency**: The echosounder surveys at the FORCE site have shown that encounters with fish are typically single fish, not schools and therefore it is harder to classify. Using multi-frequency is the best method for distinguishing between physical vs biological back

scatter. For single targets, these are often site specific and characteristic site-specific filtering (and even seasonal differences).

- **Using multi-beam sonars with optical cameras and real-time classification works well for splitting the biological and non-biological data.** It is imperfect for target classification. However, splitting the data between biological and non-biological will reduce the amount of empty data.

4.4.3 PAM technologies (overview of data automation effort)

Passive acoustic technologies have far more standards and definitions in terms of what should be done, than some of the other instruments discussed (echosounders, multibeam).

PAM technology used around tidal turbines in Nova Scotia are aimed at detecting the presence of vocalizing marine life. PAM quantifies the existing soundscape and effects of the turbines on the soundscape. In Grand Passage, the focus is more on porpoises and the ambient soundscape.

When questions arise about the effect of the sound from a turbine on the soundscape, then different possible scales of effects within the effects of noise are looked at – essentially, a zonal approach. In terms of effects from noise, there can be physical damage from really loud sounds including temporary hearing or permanent hearing. This can lead to masking potential behavioral effects and that can warrant further investigation.

PAM data collected and managed for compliance can be challenging due to the large volumes of data collected. With large volumes of raw data, marine mammals are auto-detected and only snapshots of interest are stored. The soundscape data is compressed to 1 minute millidecades which is viewed as a useful metric of time for this use. Data QA must be done in real-time at a site by acquiring long term trends in spectra, known sources for daily calibration, and direct to web automated plotting.

Some of the **measurement challenges** experienced with PAM include:

- Flow noise over the hydrophones –the lower you are in the water column, the quieter it is.
- Integration of hydrophones into the system – power, clocks, storage
- Acoustic interference – different echo sounders interfere with each other and all interfere with the passive acoustic hydrophone because they are all close together; 60 meters away would be ideal.

To improve PAM for use in instream tidal sites, the following work and next steps are required:

- Research
    - Directional systems for tracking of porpoise around turbine
    - Audibility & TTS from new turbines
    - Automated fish detection in sonars – ML?
    - Interaction and / or avoidance by fish, porpoise – to the turbine, and also to the turbine with the echo sounders on
    - Scaling of effects from single to turbine farms

- Systems
  - What are requirements – for science, for compliance monitoring?
  - Prototype and acceptance of automated quarterly reporting
  - Sonar / PAM mounting and scheduling to achieve objectives
  - Systems & software integration, long term stability testing, redundancy

- **Key discussion points and take-aways:**
  - **Collaboration:** It would be ideal to explore how to continue testing data sets and associated findings that can be more transferable and easier to use across different sites. Agreement on how meta data is processed and reduced to a smaller set is needed. Developing standards for meta data is critical to collaborations and shared calibrations.
  - **Training and Re-training models:** Efficient and effective training of data models is necessary to ensure consistent results. Maintaining the models through long term deployments can create challenges and therefore must be addressed to ensure the highest quality data is obtained.
  - **Data QA must be done in real-time at a site** by acquiring long term trends in spectra, known sources for daily calibration, and direct to web automated plotting.

4.5 Summary of common themes and takeaways from the workshop

- Agreement by regulators on the type of reporting that is required is necessary as it will ensure that reporting is standardized and simplified and can be achieved within the time requirements. (Section 4.2)
- There were two strategies discussed for data gathering – 1) collect all data possible (ex. ONC) and 2) Minimize the amount of data collected/data mortgage (AMP). (Section 4.2)
- Depending on the amount and type of data that is being collected there will be different needs and options with regards to the work being done by people vs. machines/technology (Section 4.2)
- Triggered acquisition can reduce the amount of empty data. (Section 4.2)
- It is important to ensure that there is robust data that a model is learning from in order for consistency of results between sites and conditions. (Section 4.3)
- Due to the inherent challenges of working in the marine environment, including corrosion and biofouling, computing and controllers are designed to be operated on short as much as possible. (Section 4.3)
- Machine learning results can be as good as good as a human and much faster (ex. DeepSense echosounder automated reporting = 98%). (Section 4.4)
- Co-registration of algorithms is important and was done by both AMP and FLOWBEC to improve the accuracy of target detection (Section 4.4)
- Using multi-frquency is the best method for distinguishing between physical vs biological back scatter (Section 4.4)

- It would be ideal to explore how to continue testing data sets and associated findings that can be more transferable and easier to use across different sites. (Section 4.4)
- Efficient and effective training of data models is necessary to ensure consistent results. (Section 4.4)
- Data QA must be done in real-time at a site by acquiring long term trends in spectra, known sources for daily calibration, and direct to web automated plotting. (Section 4.4)
- Using multi-beam sonars with optical cameras and real-time classification is imperfect for target classification but does work well at splitting the biological and non-biological data, which also reduce the amount of empty data. (Section 4.4)
- Developing standards for meta data is critical to collaborations and shared calibrations. Agreement on how meta data is processed and reduced to a smaller set is needed. (Section 4.4)

5.0 Next Steps

The Data Automation & Data Management Options workshop generated a productive discussion with experts and industry, identifying key areas for consideration as OERA and FORCE move forward with the Pathway Program. Some of the discussion findings will assist in providing insight on standardizing data requirements with regulators.

As part of the Pathway Program, there will be additional workshops held that focus on other key aspects of the program.

APPENDIX A: Speaker Presentations (links)

Welcome & workshop purpose

Introductions

- Luiz Faria, OERA

Review of problem statement & objectives

- Dan Hasselman, FORCE

Strategies to reduce 'data mortgages' and transition from 'data rich/information poor' monitoring

- Adrian Round, Ocean Networks Canada
- James Joslin, University of Washington

Overview 'on-board' data management systems (control modules) and challenges for multi-instrument platforms

- Benjamin Williamson, University of Aberdeen/ University of Highlands and Islands (FLOWBEC - onshore post-processing)
- James Joslin & Paul Murphy, University of Washington (AMP/WAMP -on-board processing)

Overview current data automation processes/algorithms for differing monitoring technologies

- Echosounders (overview automation for echosounder data)
    - Jennifer LaPlante, Deep Sense

- Imaging sonars (overview of algorithm development)
    - Benjamin Williamson, University of Aberdeen/ University of Highlands and Islands
    - James Joslin & Paul Murphy, University of Washington

- PAM technologies (overview of data automation effort)
    - Bruce Martin, JASCO