NZHS Data Workshop

Napier 2023

Group Discussion Results

Contents

Session One – Data Sovereignty/Ownership

How can we manage, store and use sensitive data?

- Work together (eg: what does each party need/want?).
- Possibility of setting up Iwi with their own database/data portal/data warehouse and then request data back from them.
- Put protection/security on sensitive data so only select users can view/access the data.
- Some ownership is well outside of the data collector's and Manager's realm.
- In most cases it will be determined on a case by case basis.
- Must record metadata to determine data quality.
- Need a data access agreement drawn up before any work/projects start.
- Very important to agree and decide on data access before the data collection begins.
- Treat the data collected as soft infrastructure.
- Could involve lawyers to draw up and manage data access agreements.
- Depends a lot on the purpose of use of the data.
- Question for data use = Can sensitive data be reported on to maintain government requirements and reporting consistency if the data is unable to be released normally?
- Some legal advice has been to not store any sensitive data within a council system due to LGOIMA's etc (eg: can not store it unless it can be released to public if requested).

Session One – Citizen Science

How can we collect citizen science data?

- Online portal with log in (eg: GeoNet with 'felt' earthquakes, coastal erosion photo points, 'smelt it' app for odour reporting)
- Annual rainfall data via paper forms
- Hills CSV format
- Publicly accessible electronic data collection forms (eg: Survey123, NIWA's Citizen Science app which includes NZwatercitizens SWQ data collection)
- NEMS for citizen science data

Things to consider:

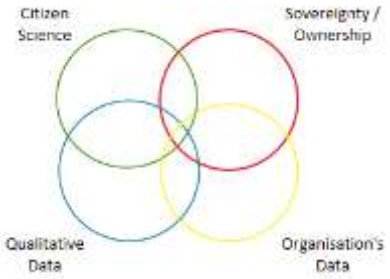
- What is citizen science (eg: paid or volunteer work?)
- Links to data sovereignty
- Need to collect methods and metadata to generate NEMS quality codes
- Regular training for consistency in community-based monitoring groups

How can we store collected citizen science data?

- With different quality codes so source is recorded as being collected externally
- Store in databases with appropriate metadata (eg: quality codes, notes/qualifiers, location notes, approval levels etc)
- Store data in a national database specifically for citizen science data

What can we use citizen science data for?

- Videos of monitoring sites (eg: flow data via videos and images and coastal erosion photos)
- Long term trends
- Compliance monitoring
- Can be used to fill gaps in monitoring network
- Can apply NEMS quality codes if methods and sensors are recorded
- Promoting public engagement and relationship building between communities and organisations



Session One – Qualitative Data

How can we collect qualitative data?

- Need fixed and limited categories to keep data collection consistent and make data analysable (eg: fixed weather and wind categories)
- Ideally collected using an electronic field app, so collection can be controlled (eg: limit the number of free text fields and keep it NEMS compliant)
- Photos to help with confirmation
- Need to set categories/scales and train people so that there is a baseline
- Ensure ongoing training
- Oral and written histories, National film archive
- Need testing to make sure field app or data collection forms are fit for purpose
- Can do training days with people collecting the data to ensure consistent results are collected

How can we store collected qualitative data?

- Store as comments/metadata alongside other collected data
- Use if a result is under question
- Potential for future systems to analyse qualitative data through machine learning
- HBRC store qualitative data in 'Puddle' system
- Categories could equal numbers (eg: for colour: blue = 1, green = 2, etc)
- Keep separate and apply different quality codes
- KISTERS product can do this well for ecology data
- Qualitative data is not easy to store in most databases
- Examples: GeoNet, Springwatch (UK), Smelt It app
- Use quantitative data to determine if qualitative data can be confirmed
- Can be stored as a parameter with categorical values
- Could be worth investigating if all councils are using the same categories
- Metadata needs to be recorded with all qualitative data

How can we display qualitative data?

- Photos
- Graphs and other charts
- Time-series with colour dots/legends
- Surveys (eg: stream surveys) can be displayed as a report

Session Two – Data Visualisation

What tools are we using to help visualise data?

- Power BI, R, Python, Snowflake, SQL, Puddle, Streamlit, Origin (graphing software), Jupiter notebooks, excel, Water-ride, API's, Tableau, Matlab
- LAWA, interactive GIS generated maps on council websites
- Hilltop Manager, Hydstra, Wiski, AQUARIUS Time-Series databases
- Graphs and excel tables/reports
- Interactive dashboards with useful legends and/or bands (eg: AQUARIUS WebPortal dashboards, R-Shiny dashboards)
- Websites and data portals (eg: NIWA's CliFlo database, council data portals)

What types of data are we wanting to visualise?

- Continuous and discrete environmental monitoring data (eg: Water level, flow, air quality, rainfall totals, water quality, water use data)
- Monitoring site locations
- Operational data for internal purposes (eg: scheduling work, system improvement, prioritisation of work)
- Matauranga Māori
- Missing data for data quality and processing purposes
- Real-time data for emergency management purposes
- Identifying outliers in data sets over time and space
- Help determine sites to target for gaps in flow ratings
- Discrepancies between systems
- Overdue verifications and site inspections
- NEMS/quality of data comparisons

Session Two – Data Reporting

What tools/software are we using to report data?

- LAWA
- IRIS, SQL, Microsoft Access, Time-series databases
- R and Python generated reports
- Moving away from written reports towards more interactive dashboards and visual reporting tools
- Move towards using live data for real-time decision making
- Online data portals/websites
- Similar methods to data visualisation methods

How can we maintain consistency when reporting on environmental data?

- Standards (eg: data sets, methods, algorithms, units, naming conventions, monitoring frequency)
- NEMS standards for processing and quality coding of data
- Collaboration on a national level for procedures and guidelines
- Testing, auditing and script checks
- Cross check nationally to ensure everyone is reporting on the same thing in the same way
- Training internally and externally
- Library of scripts for industry repository (eg: located on NEMS Tools and Resources website)
- Central data custodian in charge of these scripts (eg: MahiTahi role)
- Greater networking and visibility between sectors/LGOs

Session Two – Availability of Data

How are we making collected data available to end users?

- Hilltop API/webservice
- Snowflake, Power BI, Envirohub portal
- Data warehousing via APIs
- Data requests (eg: csv, excel outputs etc)
- LAWA
- LGOIMAs
- GIS maps for both internal and external access
- Dashboards that present data with context (eg: normal, low or high river level)
- Reports
- URLs and data portals/websites
- Metadata and comments usually provided if specifically requested only

What data are we sharing the most/least?

Most:

- Flow
- Water Level
- Rainfall
- Water Use/Abstraction
- Water Quality & Nutrient data
- Drinking Water
- Climate data (eg: wind data)
- Air Quality
- Webcam images (eg: flood cams)

Least:

- Methods
- Uncertainty
- Context behind data
- Quality Codes
- Terms of Use
- Qualitative data
- Compliance/Consent data (council dependant)

How do external agencies want to receive data?

File Formats:

- CSV, Excel, HTML, XML

Other:

- Reports, graphs, via codes/queries, statistics, access via API's if requested and ongoing.

Consistent formats make it easier to send data and for external agencies to work/use it (eg: Stats NZ prefer a specific format).

Session Three – Electronic Field Data Capture

What tools/apps are we using to collect data in the field?

- Survey123, GoCanvas, Fulcrum, Collector, Field Maps, Nintex forms, HAWQi, Microsoft Power Apps, hilltop forms, excel forms
- Custom in-house forms
- Paper forms
- Tablets, mobile phones and laptops

What types of data are we collecting using electronic methods?

- Mostly quantitative data, but also qualitative metadata
- Continuous data (water levels, rainfall, met data, soil data, consent water meter data, air quality data)
- Discrete data (mostly water quality and flow gauging data)
- Ecology & Biodiversity data (Wetlands, periphyton, WQ sampling, macroinverts)
- Site metadata (photos, videos)
- Survey data (bathymetry, BM level data, cross sections)
- Sensor calibration/validation data (eg: PT, TB3 cals)
- Asset management & sensor data
- Health & safety data (take 5's, risk assessments, traffic management etc)

How are we getting collected data into our systems/databases?

- ECAN = Survey123 / Hills Lab Data > LabMail / GIS SQL > Hilltop Sampler > Hilltop Manager
- TDC = In-house site visit app for field data collection > Hilltop
- MDC = Survey123 > Hilltop Manager
- BOPRC = Survey123 > AQUARIUS Time Series & Labware databases
- WRC = Fulcrum > WISKI
- AC = GoCanvas > Hydstra
- GWRC = Survey123 > python > SQL DB > Hilltop
- Automated process to import Nintex forms into Puddle
- Env Southland = Not on electronic data capture yet / manual processes
- CSV, HTML, XML, Excel imports
- Manual data entry
- HydroTel (continuous data and measurements) integrated into databases.
- Files > Some documents just get saved to archive databases

What data collection form templates could we add to the NEMS website?
- Survey123 and Nintex templates for data collection to set standards - Forms must be standardised, but doesn't matter what app/tool is being used - Need to agree on best practise according to NEMS standards - Pre-work checklist (H&S, gear, training requirements etc) - Field work checklists (gear and training etc) - Sensor calibration/validation forms - Forms including NEMS quality code matrices (eg: flow gauging QC matrix) - Consent flow meter install and verification standardised forms (blue tick + IrrNZ regulations + NEMS standards) - Forms for all NEMS standards that currently exist with minimum required fields and correct terminology - Future = Forms for all new NEMS standards published with minimum required fields and correct terminology

Session Four – Data Quality Assurance and Quality Control

What NEMS standards have we adopted / implemented?

- Rainfall Recording
- Water Level
- Open Channel Flow Measurement (gaugings)
- Water Temperature Recording
- Discrete Water Quality (parts 1 to 4)
- Dissolved Oxygen
- Water Meter Data
- Rating Curves
- Turbidity Recording
- Processing of Environmental Time-series Data
- Periphyton

Why have we not implemented an available NEMS standard?

- Some organisations are following guidance in NEMS, but not coding data for some standards
- Don't meet documentation requirements (eg: calibration requirements & recording etc)
- Time and resource requirements (hard to keep up with changes, training etc)
- Proliferation of data to be stored is a burden in some cases
- Some standards are too vague / hard to understand / confusing / too long
- No real push to adopt them (not compulsory)
- Inconsistencies in level of detail
- External users do not request quality codes assigned to data, so limited demand
- Waiting for databases and data to be 'cleaned up' or reviewed before adopting and applying quality codes
- Too expensive to get QC600 for some standards, so no drive for adoption
- Waiting for other organisations to adopt the standards and iron them out before adopting
- Competing work priorities will not allow for adoption at this stage
- Changing databases, which result in large scale changes and so adoption of new standards is not feasible at this stage
- No official NEMS training workshops for data processing, so hard to train staff correctly

What tools are we using or developing to automate data QA/QC?

- Python scripts
- R scripts
- SQL reporting
- PowerBI (eg: flags data out of range, then a person takes over, dashboards)
- System performance (internal benchmarking / KPIs)
- Virtual measurements
- Time series management databases
- Data entry tools
- Chain of custody tracing
- Most organisations do not have a heap of automation set up yet

What tools are we using to check for operational (or system) performance?

- R scripts (dashboards and emails)
- Azure
- Python scripts
- PowerBI (eg: gauging priorities)
- HydroTel / Hilltop alarms and notifications
- Client notifications (eg: LAWA)
- SQL reporting
- Daily checks / manual checks in databases for issues in data and outliers
- Manual checking of field inspection data
- Alerts for data exceedances and missing data
- Data capture rates in KPI reporting

How are we reviewing, auditing, or archiving data?

- Ad hoc, maybe not to NEMS standards
- Manual and automated checks
- Annual external data audits (eg: Auckland Council)
- Regular data audits (eg: BOPRC)
- Process > someone processes the data, then someone checks the data and archives it
- Using time series manager for archiving data
- Automated reporting / KPI reporting

Session Five – Open Session

Topic = What would you like to see on the NEMS Tools and Resources webpage?

- Videos, audio and written documents to cater for all learning styles (eg: training videos, sampling procedures etc with references to NEMS standards)
- Break up NEMS documents into manageable sizes with examples and videos
- Data processing examples in each database (eg: Hilltop, AQUARIUS Time-Series, Hydstra, Wiski) with references to NEMS standards
- NEMS quality code applications examples in each database
- Online training sessions on NEMS related topics (1-2 hrs)
- Links to scripts for consistent usage
- User forum for ideas and feedback
- Computer readable versions of NEMS standards
- Templates for electronic field data capture including minimum requirements to meet NEMS standards
- SOPs for standard and universal procedures (eg: working at heights, confined space, slackline work etc)
- Gear lists for specific environmental monitoring related tasks