NZHS Data Workshop – Queenstown 2024

Group Discussion Results

Session One – Quality Assurance & Quality Control of Environmental Data

Is your organisation using automated QA/QC tools/processes?

If yes,

- *How is this being carried out? (eg: python, R, SQL, database tools…)*
- *What parameters/types of data are being automatically QA/QC'd?*
- *Is your organisation carrying out QA/QC of data in real-time?*

If not,

- *What is the average time lag between data being collected and data being processed?*
- *Are you currently working towards automated QA/QC of data?*
- *What are the major roadblocks/issues that are preventing or slowing down a move towards data QA/QC automation?*

- ECAN Groundwater = Quality assurance processes largely manual (eg: visual checks, double checking).
- ECAN Surface water quality = Automated quality assurance process that runs monthly to pull out max/mins and ranges for sites. Continuous data checks are mostly still manual.
- ECAN Overall = Currently looking at automated quality assurance of groundwater quality data via python. Manual quality assurance delayed and can be 3-6 months delay between collection and data processing (eg: LAWA displays unprocessed/raw real time data that is not even available on the ECAN website).
- BOPRC = Discrete water quality is fully automated using Labware database algorithms. Working on automating time series data QA/QC in real time using HydroCorrect tool in AQUARIUS database. Have multiple data checks using python and R reports. Time between collection and processing/approval is ~ 1 month/30 days in most cases. Will be flagged if outside this time.
- PDP = Hard to be automated when data points/collection is not ongoing.
- A few organisations not currently working towards automated QA/QC of data.
- Average time lag ~ 3 months.
- Matrix suggests a quality code, semi-automated.
- Important to communicate uncertainty to the public/end users of the data.
- Inter-lab comparisons in New Zealand are not done to the level it should be. Tried to implement with NEMS, but it just didn't happen.
- Automated > Telemetry > Decisions
- Manual > Time series download process by humans.

Session One – Data Audits / Data Reviews

Is your organisation carrying out regular data audits or in-depth data reviews?

If yes,

- *How is this being carried out? (eg: automated process vs manual process).*
- *How are reports or checks being generated? (eg: python, R, SQL, database generated reports etc).*
- *What parameters/types of data are being audited/reviewed?*
- *How are completed data audits/reviews recorded?*

If not,

- *Are you currently working towards being able to carry out data audits/reviews?*
- *What are the major roadblocks/issues that are preventing these from being carried out?*

- ORC = Not currently doing this. Paused when moving from Hilltop to AQUARIUS Time-series database. Have not set it back up yet. Lack of reports (working on this) and staff with enough knowledge/experience to carry out reviews. Aim to carry out annually.
- BOPRC = Yes, carrying out partially automated data audits every 2 years for rainfall, water level and flow data (will include groundwater level data soon). Have recently switched to 100% digital data audit, so no printing/scanning/hard copy of audit. All checks are carried out in the AQUARIUS Data Review tool or R/python reports. Audit certification reports are saved to location in database. Data reviews are carried out every 2-3 months depending on the parameter.
- NIWA = Yes, annual data reviews. Partially automated using C#, but could be improved as some are copy and paste.
- TRC = Five yearly reviews/audits completed externally.
- E3 Scientific = Not really, only use excel.
- ECAN Groundwater = If yes, it does not sit with technicians to complete. Duplicate samples are checked for.
- ECAN Surface water quality = Have internal collection audits run 'in-team'. Quality assurance process runs checks on discrete data.
- PDP = Duplicates and blanks etc. Changes project to project. Data checked internally at least once.
- KISTERS = Western Australia government has automated process from sensor to archive. If all tests passed, then data is sent to archive. Quality assurance checks are rule based, set up through mathematical analysis. Instrument calibration checks.
- Difference in approach to monitoring data and scientific research data. Monitoring data undergoes more review.
- Flood warning data is not reviewed prior to sending data out, as it is time critical data.
- Manual data reviews every 3-4 months.
- Manually data audit manually once a year.
- Automating this process is in progress, using python and/or R.
- All data reviewed manually.

Session Two – Citizen Science & Data Sovereignty

What data is your organisation currently collecting via Citizen Science methods?

- NIWA = Manual rain gauge data from farmers. Snow measurements. Fish passage data.
- NCC = Coast Snap app for coastal monitoring.
- ECAN = Farmer's groundwater levels and some consent data. Some MAR projects (eg: groundwater sampling). No large-scale citizen science yet, but we help community groups and assist with training, but don't store their data (with a few exceptions).
- BOPRC = Daily manual rain gauge readings.
- Daily rainfall measurements.
- Water quality sampling data from community-based groups.
- TRC = If there is community sourced data it's not coming through the Science field team.
- ORC = Has a team that works with community. Farmers collecting data.
- GWRC = Have some Wairarapa groups sending water quality samples to Hill labs, but these haven't been good with paper forms, so very keen to trial CBM tool (eg: Survey123 forms). EnviroSchool / SMACK, but not really getting or using the data.

What types of sensitive/restricted/confidential data is your organisation currently collecting?

- Commercial data.
- Consent data.
- Landowners' information.
- Water quality data from sensitive locations.
- PDP = Data collected for clients is private.
- NIWA = Energy producing clients' data.
- Sensitivity based on location and parameter in most cases (eg: rainfall data is less sensitive vs nitrate measurement have higher sensitivity).

How are groups collecting and supplying environmental data?

- Manual/paper-based forms emailed in.
- Pictures sent in via apps or emails.
- Emails or phone calls.
- CSV/excel files with timestamps and values.
- Snow data collected and supplied via app.
- Digitised recording for manual rain gauges has proved problematic for some older volunteers.
- Fish passage data has a dedicated citizen science app.
- GWRC = Data via Hill labs or scanned paper forms.
- Some community groups enter data into the NZ Freshwater Fish Database.

How are you storing this data?

- *Metadata, NEMS QC, stored in database, stored in file system etc.*

- Daily manual rainfall data entered into Hilltop database and NEMS QC 200 applied.
- Most water quality sampling data from community-based groups do not contain enough metadata to apply a NEMS quality code, but are entered into Hilltop.
- Stored in database, but with highest permissions/security needed to view & access data.
- Stored in database, but with restrictions on who can view/access the data.

Who has access to this data?

- *Internal staff only + landowner(s), Iwi only, specific internal staff only etc.*

- Public sector = Most available freely and some on request.
- Private sector = Some is requestable, as it was only collected for internal use, but is still not sensitive.
- Silos depending on governance.
- Open access if it's public data – public sector.
- Some separate data pages for private info containing data – secure.
- All data open to internal users.
- Some data is internal only, if sensitive or confidential, but mostly all available on data portals.
- Data is restricted to specific users, but can sometimes be shared accidentally due to confusion around who can have access.

Session Two – Electronic Field Data Capture & Qualitative Data

What type(s) of environmental data are currently <u>not</u> being collected electronically by your organisation?

- *What is preventing this data from being collected electronically?*

- Gauging cards not being collected digitally for some organisations.
- ORC = All paper-based forms currently. Trying to use Survey123, but limited by GIS team not creating Survey123 forms and struggling with IT.
- PDP = No electronic forms yet.
- MDC = 100% digitised forms. Scientists have access to create Survey123 forms too.
- TRC = Using manual paper-based book for groundwater data collection. Electronic forms are being used, but are being imported manually.
- BOPRC = Using electronic Survey123 forms, with only a few ecological forms not digitised.

- GWRC = Terrestrial monitoring still on paper, mostly because it feels easier at the time. However, then it is more difficult when back in the office.
- NIWA = Mixed paper and electronic forms, but mostly paper based. Some people choosing to not use electronic forms and instead stick with paper forms. These paper forms have been the same for a very long time, so potential for people to want to use familiar format, but also might be missing key information needed now. Sometimes the electronic forms are too complex and creates more work than just writing results on paper. Can face cultural issues getting staff on board with electronic forms.
- ECAN = Mixed paper and electronic forms, but mostly electronic with water quality data. Limited by GIS teams not creating Survey123 forms.

Are your electronic field data collection tools integrated into a data management system?

- *If not, what is preventing integration?*

- Yes, field teams.
- Not for telemetry or water meter forms.
- ECAN = Field form data imports automatically.
- ORC & ECAN = Both having difficulties with their GIS & IT teams with support in creating and integrating Survey123 field forms.
- BOPRC = Either all currently integrated or waiting for IT to complete integration of forms.

How are you collecting, storing and reporting on qualitative environmental data?

- *eg: fixed options in drop down menus vs free text.*
- *eg: assigning quantitative values to qualitative data (eg: yellow = 2).*

- Still stored quantitatively.
- Yes, do have drop down boxes for most fields.
- Free text for any comments about the site.
- ECAN = Mostly removed free text fields from forms.
- GWRC = Using quantitative values to remove multiples in naming conventions.
- BOPRC = Assigning quantitative values to qualitative data, then using legends or notes in time series to show true meaning/value of numbers (eg: 1.1 = Mild odour, 1.2 = Medium odour etc). Can view trends in qualitative data over time in a time series if done this way. Limiting number of free text fields wherever possible.

Session Three – Operational Checks & KPI Reporting

How does your organisation currently report on:

- *Individual staff KPI's*
- *Team/Organisational KPI's*

- MDC = Power BI reports, including staff target timeframes for data processing. Annual basis for team.
- GWRC = Used to have individual KPIs for data processing timeframes.
- TDC = Triennial data reporting.
- HBRC = Have scheduled tasks (eg: what we have to deliver). No set KPIs, just things like getting LAWA data on time.

- BOPRC = Python generated reports for data processing timeframes, sample data checks (eg: groundwater quality sample results vs historical results/ranges), python generated data approval reports for time series, field visits and ratings, team KPIs and individual staff KPIs in reports.
- NCC = Currently little to no KPI reporting.
- RDMW (Queensland) = Quality code analysis, missing records, rolled up by network.

Does your team carry out regular operational checks?

- *How often are these carried out?*
- *How are the results reported?*
- *What actions are taken if, for example, a site/sensor are continuously being flagged as failing this check(s)?*

- BOPRC = Yes, manually twice a week using HydroTel, reported via a team email. If sites/time series are being continually listed in results, then team leaders will investigate and ensure action is taken.
- NIWA = Yes, carry out regular operational checks using WebPortal and python scripts.
- NCC = Monthly site inspections.
- Telemetry checks, estimations for service or repairs or quarterly services by default.

Session Three – Data Visualisation

Is it enough to simply provide data to end users, or is it your organisation's responsibility to ensure end users understand data?

- Yes, if you provide data then you have a responsibility to provide relevant metadata.
- Metadata is key/crucial.
- Depends on who the end user is. It's not their job to understand our job.
- If going to environment court, data needs to be fully interpreted/understood correctly.
- Consultancies are paid for a service/product clients access and understand, so yes absolutely.
- Councils need to provide data to end users, but there is room for more in most cases.
- To an extent. Can provide context with the data to make it more understandable, but don't want to hand hold.
- Point people in the direction of LAWA fact sheets etc.
- No, usually clarify request or ask what is intended use of the data, so more ensuring right data is provided but not interpreting it.

In your organisation, who creates environmental data visualisation tools?

- *Digital/IT team, geospatial analysts, environmental techs with specialist skills, environmental data analysts?*

- ECAN = Data visualisation tools made by science team or analysts, some techs also.
- NIWA = Data scientists, field techs with specialist skills.
- GWRC = Data analysts create tools.
- TDC = Data analysts.
- MDC = Data analysists mostly.
- PDP = Uses their GIS team.
- BOPRC = Depends on the type of data/information. IT create dashboards, GIS maps/layers, science create Rshiny interactive dashboards, Environmental data team configure data portal and dashboards on the data portal.
- IT mostly, but lots of different teams for different purposes (eg: environmental monitoring data team configure the data portal, GIS team does mapping, science team do their own visualisations for reporting.

What data visualisation tools do you use? (eg: Power BI, Python reports, excel reports, dashboards, data portals etc).

- NIWA = Python and AQUARIUS WebPortal.
- ECAN = Power BI, website, python reports, data portals, data requests (csv files).
- BOPRC = Python reports, AQUARIUS WebPortal, R reports, relevant layers on GeoView/ArcGIS. Currently looking into Power BI applications for this.
- Need to have interactive data to be properly understood.
- Power BI - internally for environmental monitoring teams for reporting.
- R scripts feeding into Power BI dashboards in data portals.
- Data portals, Hilltop Hydro, website graphs.
- Microsoft Access – internal operational dashboards at MDC.

Session Three – NEMS Quality Coding of Environmental Data

What NEMS standards has your organisation adopted?

- PDP = Follows methodology but does not code to NEMS.
- ECAN = Discrete Water Quality NEMS (groundwater, rivers, lakes, coastal).
- NCC = Trying to follow NEMS as much as possible (eg: quality coding, calibrations, data processing etc).
- BOPRC = Have adopted most published and available standards (except air quality parameters) and are in the process of adopting water meter data NEMS.
- Dissolved oxygen and water temp.
- Hydrological parameters (eg: stage, rainfall, groundwater).
- TDC = Periphyton, rainfall, water temp, dissolved oxygen, open channel flow, turbidity, water level.
- MDC = Rainfall, discrete water quality (all four types), water level.

Is the application of any NEMS standards proving to be problematic with automated QA/QC processes?

- NCC = No automated processes yet.
- ES = Not aware of ay causing issues.
- ORC = Doesn't use any automated QA/QC processes.
- BOPRC = Have automated all discrete water quality sample results. Will use the HydroCorrect tool (AQUARIUS Time-series database) to automate continuous data in the future.
- Quality coding matrices need to support automation (eg: GWRC holding back on water level until matrix available).

Why has your organisation chosen to not adopt certain NEMS standards?

- ECAN = Have not adopted turbidity, water temp, dissolved oxygen, periphyton or macroinvertebrates.
- NCC = Lack of resourcing.
- Some situations where there is disagreement about specifics and some development needed.
- Soils – too expensive to validate to NEMS.
- Some variations applied to NEMS standards to make it work.
- List of standards keeps growing, so simply haven't got to them all yet.
- Resourcing – don't always have the people, processes or equipment that NEMS requires (eg: air quality).
- Don't agree with some requirements (eg: rainfall NEMS).
- Waiting for new versions to be published. No point starting if criteria is going to change with a newer version.