Interoperability framework for freshwater biodiversity data.
TFBIS Project 282 Scoping Report

Prepared for Department of Conservation

June 2012

3 September 2012 8.56 a.m.

Authors:
Andrew Watkins
Jochen Schmidt
Brent Wood

For any information regarding this report please contact:
Jochen Schmidt
Chief Scientist
Environmental Information
+64-3-343 8058
Jochen.Schmidt@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
10 Kyle Street
Riccarton
Christchurch 8011
PO Box 8602, Riccarton
Christchurch 8440
New Zealand

Phone +64-3-348 8987
Fax +64-3-348 5548

NIWA Client Report No: CHC2012-077
Report date: June 2012
NIWA Project: DOC12312

3 September 2012 8.56 a.m.

Contents

Reviewed by
NA (scoping report only)

……

Approved for release by
Jochen Schmidt

……

Formatting checked by

……

1 Project pathway

This report contains a brief documentation of the pathways, the major decisions, and the outcomes of the project. All discussions, minutes, etc. were publically announced and are globally visible through the internet (see below). A stakeholder group was formed to guide the project (see Acknowledgements), but the discussions were also widened to the user community using the dataversity blog. A detailed summary of the communications can be found in the Appendix.

This project was conducted in the following phases.

- A 'strawperson' for a fundamental Biodata Interoperability Framework was developed by NIWA and published on http://dataversity.org.nz/groups/interoperability
- The framework was publically discussed through email and post blogs on the dataversity website with relevant stakeholders. All emails were also posted on the dataversity website to ensure the whole discussion and process was documented.
- A phone conference was held in February 2012 with the stakeholder group to discuss the proposed framework. The minutes are available under https://teamwork.niwa.co.nz/display/NZEIIF/Biodata+Interoperability+Meeting%2C+Minutes%2C+02-02-2012
  The outcome and agreement of the meeting was:
  - Agreement in principle of several Regional Councils and other stakeholders to contribute to the project and happy with LCR/NIWA taking **technical** lead.
  - NIWA / Landare as the key New Zealand experts to develop a technical proposition for the interoperability framework.
  - Present Regional Council and DOC staff expressed their commitment to adopt a national interoperability framework for biodata.
- A workshop was held with NIWA and Landcare Research key scientists in March 2012 to discuss the development of a technical framework. The meeting minutes are available under https://teamwork.niwa.co.nz/display/NZEIIF/Biodata+Interoperability+Meeting%2C+NIWA+CHC%2C+26-03-2012
  The outcome and agreement of the meeting was:
  - All working further on collating information about existing global frameworks through and make available through the dataversity website.
  - Agreement that it would be useful to develop / test a biodiversity interoperability framework based on the Open Geospatial Consortium (OGC) Observation & Measurements Abstract Model. This is to complement the current suite of GBIF based interoperability services, which have their limitations in scope and for application in data federation.

  - NIWA to develop a technical recommendation for a biodiversity observation profile based on OGC standards as outcome of this project.

- In the following months the discussion on existing (global) standards did carry on and was documented on the dataversity website. The **NIWA interpretation** of this discussion is that while there are many technologies used internationally to enable biodiversity data communication between systems, there is a lack of standardisation of these approaches and there is not a dedicated global standard for satisfying relevant use cases for exchanging biodiversity observation information in a generic way (as defined by the OGC Observation and Measurement O&M Framework) (but multiple existing approaches). That means there is a need for synergizing and harmonizing the existing approaches.
- NIWA developed a technical recommendation for a biodiversity observation profile based on OGC standards as outcome of this project. This technical recommendation is attached as Appendix of this report and available as a living document under https://teamwork.niwa.co.nz/display/NZEIIF/Biodiversity+Interoperability+through+Open+Geospatial+Standards

Note:

- In the project proposal and contract we planned for three stakeholder workshops. This was not necessary as stakeholder agreement was reached very easily in the first workshop in terms of (1) NIWA and Landcare taking the technical lead, (2) and agreement of stakeholders on the straw man and general principles. We therefore decided to focus the project resources on producing an actual proposition for an OGC web-service based standard as per Appendix 2.

2 Why an interoperability framework for freshwater biodiversity data?

The term “Bio data” is used synonymously with “Bio Observations” in this document and refers to any data about organisms captured in the field.

Bio observations are done by various organisations across the country and a number of agencies in New Zealand hold and / or curate archives and databases of freshwater biodiversity survey data. To date there is no easy and nationally agreed way to query or transfer freshwater biodiversity survey information across organisations and systems to assist use cases like

- NIWA freshwater scientist seeks all observation made in a particular year of a particular species in NZ. She wants to produce a species density map from that data.
- Yale PD wants to get observation information on a range of different freshwater fish species for New Zealand, time-tagged. He wants to study predator-prey relationships (real case;-)

- Regional Council Scientist wants to extract all available data for invertebrate species x for a particular catchment to create invertebrate abundance maps and derive the ecological health of different rivers.

Currently, users (scientists, managers, students) often have to go through painful processes of acquiring data manually through personal contacts and manual data manipulation to enable them to put the information into a common format. If all New Zealand biodiversity archives would be accessible (through a community standard) much energy and resources could be saved. That includes a (set of) community agreed protocols for freshwater biodiversity information exchange across systems and into user interfaces and explorative and analyses tools. What is missing is a useful and agreed interoperability infrastructure for sharing of freshwater biodiversity survey information. This infrastructure would allow much better and cost-efficient interpretation of primary data and make them accessible for ecological management.

The key paradigm of the biodata interoperability framework can be defined as:

Regardless how and where bio data is managed, we want stakeholders to be able to discover it in a taxa, spatial, and temporal context, provide consistent metadata, and be able to access the information in a consistent way.

3 Assumptions / use cases for a federated, interoperable bio observations infrastructure for New Zealand

The key user group we are focusing in this study are freshwater scientists or other end-users who are interested in acquiring all available bio data for a particular domain of interest and wanting to conduct some sort of analyses / investigation of that dataset. **Our understanding** of information use in the freshwater bio community is that there are many Scientists, Consultants, and other end-users who want to extract data and perform their own analyses in that format. Key problems will be consistency in taxa data, location data, and time data across different providers as well as methods. However, a detailed user needs analyses is outstanding and included as part of a proposed roadmap.

Stakeholders (Scientists, Managers, 'Citizen Scientists', etc.) want to be able to discover, access, understand and interpret bio observations held in various source systems by different agencies across the country. To be able to access, understand, and interpret bio observations in a federated information environment users need **consistent** information on the following core information types.

1. Where was the observation taken?
    a. In terms of coordinates of the observation
    b. In terms of the environmental feature observed (which river (reach), which lake (part))

2. When was the observation taken?
3. Other relevant meta information (who sampled, which organization, which project, purpose of sampling, etc.).
4. What taxa were observed?
5. Which observation (phenomena) was taken and what units were used (fish length, habitat, etc.)?
6. What method was used for the observation?

4 Requirements for a federated, interoperable bio observations infrastructure for New Zealand

The following lists the requirements to achieve a bio information infrastructure consisting of a federated network of source systems operated by different agencies across the country.

1. Taxa data are stored consistently in source systems.
2. Geospatial data (coordinates) are stored consistently in source systems.
3. Temporal data are stored consistently in source systems.
4. Phenomena descriptors ('measurements') are stored consistently in source systems.
5. Metadata are stored consistently in source systems.
6. Reference to observed features (river, lake, wetland) are stored consistently in source systems.
7. Descriptions and documentations of procedures / methods are stored consistently in source systems.
8. All the data is provided from the source systems through common standards into the federated environment.

"Consistently" means to agreed standards and /or complemented with understandable, documented, and traceable information.

5 Scoping a federated, interoperable bio observations infrastructure for New Zealand (Biodata Services Stack)

The findings of this scoping report propose that a New Zealand Interoperability Framework for bio observations consists of a **Biodata Services Stack** with the following components. Each of these systems / services needs to contain a (set of) national systems or protocols

and procedures / mechanisms to ensure these systems and / or protocols are used (in the source systems).

- National Reference System for taxa data (Requirement 1 above). This exists through NZOR. However, this national infrastructure needs to be maintained and rolled out across the community / source systems through connectors and implementation guidelines.
- Biodiversity Phenomena Reference model (Requirement 4 above). We believe development of an ontology is required, including supporting systems which ensure the integration of the ontology into reference archives and / or institutional archives (source systems). This would ensure observations / measurements are stored consistently and therefore can be shared interoperably.
- National Reference System for Biodata Observations Methods / Procedures (Requirement 7 above). Development of an ontology or protocol is required, including supporting systems which ensure the integration of the ontology / protocol into reference archives and / or institutional archives (source systems). This would ensure methods used for observations / measurements are stored consistently and therefore can be shared interoperably.
- National Reference System for publishing biodata observational data (Requirement 8 above). This could be achieved through best practice implementation guidelines. More development needed to ensure the system meets end-user needs.

Minor issues also exist with respect to standardized recording of geospatial data, time data, meta data as well as observed feature data. Often we find in bio databases these entities recorded in inconsistent and / or not traceable way. Several guidelines already exist nationally and internationally and could be adopted.

- National Reference System for geospatial referencing (Requirement 2 above). This can be achieved through best practice guidelines which can be developed in cooperation with the New Zealand Geospatial Office.
- National Reference System for temporal referencing (Requirement 3 above). This can be achieved through best practice guidelines which can be developed in cooperation with the appropriate authorities / organisations.
- National Reference system for (other) metadata (Requirement 5 above). This can be achieved through best practice guidelines which can be developed in cooperation with the appropriate authorities / organisations.
- National Reference system for managing environmental features (‘observed feature’, for example the lake, river reach, beach, etc.) (Requirement 6 above). Ideally this would require a national catalogue of observed features. Diversity is great and this seems not feasible. Therefore we recommend can be achieved through best practice guidelines which can be developed in cooperation with the appropriate authorities / organisations. (“How to describe the observed feature”)

6 Recommendations

To fund a TFBIS project for:

- Conducting a full user needs analyses for a federated information infrastructure delivered through a Biodata Services Stack.
- Getting full agreement of the principles of a federated information infrastructure delivered through a Biodata Services Stack.
- Implement work streams for developing key areas / elements of a Biodata Services Stack as defined above. These include:
    - Biodiversity phenomena ontology and /or protocols and supporting systems.
    - Biodiversity methods register and/or recording protocols and supporting systems.
    - Protocols for bio observation publishing and web server and web client prototype implementation and / or supporting systems.
    - Implementation guidelines for the New Zealand Organisms Register NZOR.
    - Implementation guidelines for geospatial, temporal, and metadata references.
- Working with international organisations (ALA, GBIF, OGC) to include these developments into a global framework for biodiversity observation information interoperability. Given funding is available; New Zealand could take a lead role in that area internationally.
- Implementing the prototype for a number of databases involving key stakeholders and demonstrate interoperability, evaluate functionality.
- Modifying the prototype based on the outcomes of the evaluation.
- Support implementation of the finalized version for key biodiversity databases across New Zealand.

Roadmap / Implementation plan

As a guideline for this proposed project, the following draft roadmap was developed based on the findings from this scoping report.

What	Who	Timing	Output
User Needs analyses for a bio survey data interoperability framework	NIWA/Landcare/Dataversity (also evaluate existing material)	3 Month	Report on the needs that a bio survey data interoperability framework will address.

Stakeholder Workshop 1	Group of key stakeholders based on scoping report	3 Months	Direction confirmed Work groups established
Working group for biodiversity phenomena	Key experts	6 months	Recommend implementation for biodiversity phenomena ontology
Working group for biodiversity methods	Key experts	6 months	Recommend implementation for biodiversity methods catalogue
Working group NZOR implementation	Based on current NZOR infrastructure	6 months	Guidelines for implementation
Working group for geospatial, temporal and metadata	Key experts, led by NZGO	3 months	Standards and guidelines developed
Working group for biodiversity observation services	Key experts	6 months	Protocols for bio observation publishing
Stakeholder Workshop 2	Group of key stakeholders based on scoping report	3 months	Review developments and recommendations Agree on implementations
Implementation, biodiversity phenomena system	Key experts	6 months	biodiversity phenomena ontology
Implementation, biodiversity methods system	Key experts	6 months	biodiversity methods catalogue
Implementation, biodiversity observation services	Key experts	6 months	web server and web client prototype implementation and / or supporting systems
Stakeholder Workshop 3	Group of key stakeholders based on scoping report	3 months	Review implementations Confirm Biodiversity Services Stack

			Role-out plan
Role-out with key source systems	Key agencies	3 months	Implementation in key sources systems
Engagement, presentation of results		3 months	Presentations at major conferences etc.

Outputs

- NZOR supported as national system for ensuring taxa consistency through guidelines and procedures.
- National system for biodiversity phenomena developed and implemented.
- National system for biodiversity observation methods developed and implemented.
- Guidelines for geospatial, temporal, and metadata in contributing source systems developed and implemented.
- Guidelines / systems to support national consistency in publishing bio observations developed and implemented.
- Set of guidelines for implementing a bio observations archive / source system consistent with the national Biodiversity Services Stack as defined through the previous points.

7 Public resources and publication pathways

The project progress, meetings (minutes) and more discussion was all documented on the internet publically visible to anybody globally with internet access under the following resources.

- https://teamwork.niwa.co.nz/display/NZEIIF/New+Zealand+Environmental+Information+Interoperability+Framework
  This website documents all of NIWA's initiatives in interoperability.
- https://teamwork.niwa.co.nz/display/NZEIIF/Biodata+Observations+Interoperability
  This website contains relevant information related to this project, including the minutes of the meetings held sponsored by this project.
- https://teamwork.niwa.co.nz/display/NZEIIF/Biodiversity+Interoperability+through+Open+Geospatial+Standards
  This website contains the technical details / appendix of this final report.

- http://dataversity.org.nz/groups/interoperability
  This website was / is used as the communication platform for this project and contains valuable communication from involved parties.

8 Acknowledgements

Stakeholder group for this project is listed in the table below. We would like to thank all the stakeholders for assisting with the project, workshops, and discussions.

Key Stakeholder Group	
Jerry Cooper	Landcare Research
Lucy Baker	Ministry for Environment
Peter Hiemstra	Department of Conservation
Mark Peacy	Department of Conservation
Norm Thornley	Department of Conservation
James Lambie	Horizons, Dataversity
Jim Fretwell	Bay of Plenty Regional Council
Jim McLoed	Waikato Regional Council
Paul Barter	Cawthorn
Other involved stakeholders	
Dan Randow	Dataversity
Karl Majorhazi	Ministry for Environment
Dave West	Department of Conservation
Alistair Richie	Landcare Research
Aaron Wilton	Landcare Research
Kevin Richards	Landcare Research
Nick Spencer	Landcare Research

Appendix 1: Communication Records

The table below documents the responses on the draft final report.

Comments	Responses
Karl Majorhazi, MfE To my knowledge, this is the first attempt to develop a community-based set of standards based on an international framework. This would be an important point to mention as this work could spawn the development of standards in other areas.	Actually that is not true, in the geospatial domain there are many community-based set of standards based on an

	international framework.
There is a discussion of the benefits of the framework and problems that it solves. This could go further to mention the productivity gains and wider benefits that we could expect if we were to make use of the standards. For example, there is the NASA report that estimated a saving of around 25% on projects that used open standards in favour of proprietary ones. There may be other studies that could provide relevant benchmarks.	Cost/benefit analyses not in scope of this report.
It would also be good to mention the projects that are relying on the outputs of this project and the uses that the data is, or could, be put to.	Not in scope of this report.
I would also like to see some consideration given to the ongoing management, maintenance and funding of the standards in the scoping report. TFBIS are the funders for this report but I doubt they will be responsible for maintaining it.	This is why it would be good to get international buy-in via OGC as mentioned in report.
Other questions that come to mind are: Who will own the outputs of the project? How will the implementation be managed? How will the use of standards be communicated and promoted? Who will ratify the standards and future updates? How will compliance with the standards be assured? I'm sorry I don't have the answers for you, but I think that some consideration of these questions would give the project funders, future funders, users and participants more confidence that this will be a good investment.	Outside the scope of this report, but governance could possibly be managed through a national body like National Environmental Monitoring Standards (NEMS)
Mark Peacey, DOC	
I can't comment on much of the biodiversity\bio observations side of it but from from a GIS perspective I question why there is a need for OM XML. I haven't had much time to look at the various pages linked to on the NIWA website but I would see the need for having to transform data as a significant barrier to uptake. I see the real value in **this work coming from standardised look up tables, domains, names and common approaches** no matter what system an organisation uses. DOC has very few in-house developers (let alone any with spare time) and I'd hazard a guess that most councils and agencies are in a similar situation. Using existing OGC specifications such as WMS and WFS which are defined and integrated in a number of software packages, including many open source	WMS and WFS definitely have to be in the mix of provided services. It is hoped that an OM profile would allow development of client tools in the open source community which in turn will support end-users.

ones, would make life a lot easier and allow agencies to get their data out without significant extra effort and development work.	
Dave West, DOC As an "involved" stakeholder I too tender my apologies for not meeting 13 August deadline and endorse my less tardy colleagues comments. As an ecologist I also confess I found it hard to follow the full meaning of the multiple lines of technical dialogue around this project. I also have numerous FBIS and Freshwater Fish Database, "data ingestor" and national reporting (yes NEMaR & LAWF) dialogues banging around my head, that may mean I bring extra baggage to this discussion. So bear that in mind when you read my comments in attached and below ;) First off I too am a convert of the need for federated databases and NZ cannot afford to keep creating separate expensive databases. I also commend efforts to use open source and free software even though I myself are currently locked into the ESRI ecosystem. Which as much as we try and avoid is an important part of the business for government departments so needs to be considered. I liked the "Simplified abstract model" diagram in the primer document and would have been good to have similar picture in the final report (especially for the ecologists :)) If the final application/tool (as Jim says it is not clearly identified, is it the "Biodata Services Stack"?) can accommodate biodiversity monitoring data for range of purposes (SoE----> species conservation) then even better. Note there will be an equal (and dare I say it "greater") need for a physical Water Quality & Quantity database that should align (be map-able?) alongside biodiversity data???? I wonder when end users such as myself should get input to the process? Maybe it is that our needs are simply met by a clever front end after the real work of getting the data streams working? But I think ecologists know of and are nominal	We choose not to as the final model more complicated. The model could be extended to cover any type of sampling data. The developed protocols and guidelines should support local systems for how to integrate them in a federated infrastructure.

"custodians" of some important datasets that are not always consistent in terms of date or map reference format etc. Will it be my job to standardise these? as I don't think it is the job of our data infrastructure people. Yes I hear the QA/QC mantra....but I like many ecologists in these times have bigger fish to fry than cleaning up other people's data. Maybe it is buried in the scoping report but I would have thought a data ingestor would have been an integral part of this so important data is not excluded because it is not consistent with some NZ standard. I also like names for freshwater things so unless it is captured in "Biodiversity phenomena ontology" there is a job to establish an agreed list of spatially referenced names for NZ rivers (I know DOC has a partial one), lakes (DOC and NIWA have parts of one?) and wetlands. Could we not have a local mirror site of a international freshwater web service such as http://data.freshwaterbiodiversity.eu/ ? Thanks for chance to comment and note that DOC "Key Stakeholder" comment will come via our members on that group.	Will be interesting if something like this can be developed given the diversity of vocabularies around. Could be part of implementation.
James Lambie, Horizons I second Jim's analysis – I so far don't read anything in there that would stray from the path Horizons has itself for information management. I would like to add that I think it is time to include the Land and Water forum into the wider audience of this project. As you are aware, LAWNZ has aspirations for reporting regional council biodiversity statistics into one place and I would hate to see barriers to collaboration put up because of perceived competition in this space. To that end, if you agree, I would like to pass the scoping document onto the LAWNZ team for their perusal.	OK to pass on to LAWNZ architects.
Jim McLoed, Waikato Regional Council. I have discussed this with various Waikato Regional Council staff including the Project Manager of IRIS, Derek Postlewaight. (IRIS being the Integrated Regional Information System being	

developed collaboratively by Environment Southland, West Coast RC, Horizons, Taranaki RC, Waikato RC and Northland RC). WRC strongly supports the premise behind the Interoperability framework – agreeing data interchange standards, and standards on how info will be stored, identified, common standards on taxonomy and database cardinality. This is an excellent way to progress. Where WRC gets very cautious is committing to common applications for various organisations. The commitment to common applications across different organisations means for councils, duplication of common council reference data (places, people, requests for service etc,). It means that if you have officers that are involved in anything other than for example biodiversity, they will have two or more applications to use that contain parts of their work, and a host of other operational inefficiencies that will be introduced. In regard to such shared systems development WRC strongly agrees with and regularly cites Dan Randow on July 13 2011 "I think that the real shared system is the interconnected sprawl of heterogeneous systems connected with common standards. While there are of course benefits in collaborating to build shared systems, it is the standards and connections are steadily and sustainably evolving and that will long outlive its component systems". http://dataversity.org.nz/groups/dataversity_public_discuss/messages/topic/2H6xgp66FDKb17i9Ul0cgt#post-2H6xgp66FDKb17i9Ul0cgt In reading through the document and especially the recommendations, I haven't found any suggestion or indication of an intent to develop a common application. With this understanding, WRC is fully supportive of the Reports conclusions and recommendations and looks forward to participating in the next steps (as appropriate) and getting access to the outputs.	The envisaged system is a set of standards and protocols, so we are consistent here.
Jerry Cooper, Landcare Research We agreed it would be useful to build some OGC compliant services, which, in our case, could easily sit over existing services. This message implies there are no existing biodiversity	Agree and amended the text to reflect that comment.

interoperability services and there are, and both LCR and NIWA use them – OBIS/GBIF/NZOR. Don't forget this is a scoping report. In my view the outcome should be that we test the scenario outlined, before we go any further.	
"Biodata" needs to be defined.	Added a definition.
What kind of questions would a survey information interoperability framework solve? A user needs analyses is missing as part of the framework.	Not in scope. But good point. We need to do something like that as part of a new project. Added to work plan.
Comment on section 3: sounds simple. Our experience in decades of managing NVS data is that reality is more complicated, especially when it comes to defining and exchanging such data such that it maintains the information content intact, and yet is flexible and usable. If we can make the proposed framework work for NVS data then I'd be sold. Until then, I'm not sold. That is what we wish to test as the next phase in this project - not jump into something we haven' tested.	To be honest we do not agree that the simple use case described here will not work and we not see what is "more complicated". But we agree that a user needs analyses need to be done.
Comment on section 4: I think this is telling people what they should do. That might not be accepted, and often not necessary. I believe we should focus on making sure the data are exposed externally in a consistent manner, and there are many ways to achieve a coupling between internal systems and exposed services to achieve that.	We do not agree. Our basic premise is that information can only be shared (i.e. understood) if it is well managed. This requires data management standards on an national level.
On section 5: Why are ontologies needed, what do they solve?	Added comments in text.
On metadata systems: More detail work needed. There are many metadata standards. Who are we following, and why? That is a substantial piece of work in itself.	There are international standards (EML, ISO). We do believe this is not a major work.
Developing a national system is too big a leap. This test of a particular approach to a framework needs to be test in a small domain, results evaluated, issues identified for rolling it out, and	Hmmm. Up for discussion. Might be that we can

then a further project to do that.	achieve it by a step-by-step approach focusing on key components (like geospatial, taxa) first?

Overall the responses support the testing of a federated system through the development of national guidelines and standards through a well scoped study.

Appendix 2: Biodiversity Interoperability through Open Geospatial Standards

The technical details of this report are also available as a separate document and under the following website as a living document.

https://teamwork.niwa.co.nz/display/NZEIIF/Biodiversity+Interoperability+through+Open+Geospatial+Standards