Cawthron Archival and Data Delivery Information System (CADDIS); Interim Project Report

June 2008

Author
James Lambie
Environmental Scientist – Ecology

June 2008
ISBN: 978-1-877468-50-6
Report No: 2008/EXT/896

CONTACT	**24hr Freephone 0508 800 800**		**help@horizons.govt.nz**		**www.horizons.govt.nz**
SERVICE CENTRES	**Kairanga** Cnr Rongotea & Kairanga-Bunnythorpe Rds Palmerston North **Marton** Cnr Hammond & Hair Sts **Taumarunui** 34 Maata Street **Woodville** Cnr Vogel (SH2) & Tay Sts	**REGIONAL HOUSES**	**Palmerston North** 11-15 Victoria Avenue **Wanganui** 181 Guyton Street	**DEPOTS**	**Levin** 11 Bruce Road **Taihape** Torere Road Ohotu
POSTAL ADDRESS	Horizons Regional Council, Private Bag 11025, Manawatu Mail Centre, Palmerston North 4442 | **F** 06 9522 929				

EXECUTIVE SUMMARY

The Cawthron Archival and Data Delivery Information System (CADDIS) is an Access database built and adapted by Cawthron Institute with specific versions (CADDIS-Fly) for Horizons, West Coast and Hawke's Bay Regional Councils.

CADDIS is based on a biometric and chemical water quality database built by Cawthron. The Cawthron database was identified by the three regional councils as being a suitable design to fit data the councils held. The CADDIS-Fly versions are modifications of that database, transferred to the regional councils under a Medium Advice Envirolink grant.

The CADDIS-Fly versions and the data within them remain the property of the respective regional councils. The data is structured in such a way that data can be transferred to national databases if desired.

Horizons' CADDIS-Fly was delivered in April 2008 and is presently undergoing testing by Horizons' Science Team staff. Horizons' macro-invertebrate data from 1999 to 2006 was loaded as part of the testing process, already saving Horizons time and money. Further (recently collected) fish data is presently being added.

So far some minor data issues have been discovered. These are to do with the uploading of the test data and not to do with the database function itself. As staff progress with different types of data entry, it is expected that bugs will be encountered.

A second Envirolink bid is being put through (May 2008) for the second phase of the project. This phase is to resolve any bugs discovered during testing and to produce a manual for data entry and database maintenance.

CONTENTS

1. Introduction

Horizons' Land and Water Plan identifies water quality as one of the main resource management issues for the Region. This sentiment is reiterated in the Proposed One Plan, which includes water quality and natural heritage (including aquatic biodiversity) as two of the four big resource management issues for the Region. In order to support One Plan implementation and effectiveness monitoring, more attention is being given to improving the quality of Horizons' aquatic bio-monitoring - from sampling design right through to data storage.

Through Horizons' function in State of the Environment monitoring for water quality, large amounts of ecological data have already been amassed. These are currently stored in project files, but are not readily accessible or stored effectively in perpetuity. A suitable data repository and effective data retrieval mechanism is needed as part of assuring the ongoing quality and longevity of Horizons' scientific information.

Although there has been development of databases at a national level for the purposes of historic information storage, these do not provide the resolution required for day to day use by regional councils. In October 2007, Hawke's Bay, West Coast, and Horizons Regional Councils put together a joint Envirolink project (442-HBRC51) to adapt the Cawthron Institute's database for ecological and chemical information (EcoChem) for regional councils' use. Over the course of project development, Cawthron settled on the name "Cawthron Archival and Data Delivery Information System" with the convenient acronym, "CADDIS-Fly" for the versions specifically adapted for each regional council.

Independent development of databases is known to have occurred in other councils, but Horizons, Hawke's Bay and West Coast Regional Councils perceived that the transfer of pre-existing regional council databases for their own use would require the same process as proposed for CADDIS. The councils settled on the CADDIS option due to the database having been operational within Cawthron for a number of years, and the proven track record of that database meeting information storage and retrieval requirements similar to those of Horizons, Hawke's Bay and West Coast Regional Councils.

The Cawthron Institute was contracted to deliver the database. The deliverables for this project include versions of the database specifically suited to each Regional Council, and open source code so that the councils can run and manage the database in-house at the completion of the project. A revised version containing Horizons' macro-invertebrate records from 1999 to 2006 was received by the Science Team in April 2008 and is presently being tested.

2. Database description and data to be stored

The information management system is a pair of Microsoft Access (2003) databases consisting of a "front end" database called CADDIS-Fly linked to a "back end" database called CADDIS-Data

CADDIS-Fly is the day-to-day database for data entry, reporting, and retrieval. In this database, data entry forms direct the data entry operator to enter site name and location parameters, sampling method and replication parameters, and sample data.

Data that has already been inputted may be reviewed for quality control either by re-examining the data entry pages, or by harvesting data though specially designed report queries at the click of a button.

CADDIS-Data stores the data in data tables. The database has a very simple table structure consisting of 35 inter-related tables that hold site data, method data and sample data. A set of tables relates to species definition which ensures all biological records use consistent species names. This table can be updated in such a way that all records (including historic ones) use the most up-to-date species name.

At this point in time, it is intended that CADDIS-Fly be the repository for all of Horizons' aquatic biological monitoring data. The type of data sets that fall into this scope include;

- Macro-invertebrate data – surber and kick sampling of stream benthos for mayflies, caddisflies, stoneflies, other insects, crustaceans, snails, worms etc (collectively described as macro-invertebrates).
- Fish data – electro-fishing samples and other records of the presence of fish, including count data and presence data. It is envisaged CADDIS-Fly will replace Horizons' day-to-day use of the New Zealand Fresh Water Fish Database (NZFFD or FBIS), but Horizons will be able to bulk-deliver fish records from CADDIS-Fly to the NZFFD and FBIS as desired.
- Periphyton data – samples of algae and the other "slimy stuff" growing on the beds of streams.
- Habitat data – descriptions of sites including percentage cover of rock types, vegetation (in-stream and riparian), shading etc. that is collected as part of stream surveys but is presently not stored in Hilltop or Qualarc.
- Deposited sediment – description of the depth, ash-free dry mass weight, and any chemical analysis if done. The deposited sediment data is different from (though related to) suspended solids data which is archived elsewhere.

CADDIS-Fly is not intended to replace Hilltop or Qualarc for storing physico-chemical data collected from rivers. However data such as flow, temperature, conductivity, dissolved oxygen, and chemical content, could be stored in CADDIS-Fly if desired.

3. Database location and custody

While testing takes place, CADDIS-Fly resides on <\\hera\councildata\Research\Aquatic_Biodiversity\CADDIS_Fly>. The main database users – both in terms of data entry and retrieval – are the Science Team members, and therefore it is likely that the database will continue to reside here as the most appropriate place.

The present custodians are Carol Nicholson (Research Associate) and Jim Lambie (Environmental Scientist – Ecology). Carol has responsibility for data entry and data quality control, and Jim has responsibility for testing and maintenance.

Other potential users include:

- Catchment Information - particularly with respect to internal and external information requests; and
- Catchment Data - particularly with respect to water quality information (generation and archival).

Staff within these teams have been consulted over the last two years during the database project identification phase, and their input is likely to be required into the future. However, to date, they have not been privy to the testing of the database.

Under the standard contractual agreement attached to Envirolink funding, Horizons owns its own copy of CADDIS-Fly. All of the data in the database remains the property of Horizons.

4. Problems Encountered

The problems encountered so far have been very few and are not of a database "bug" nature. The version received from Cawthron in April 2008 was successfully loaded to Horizons' server and has been successfully run.

To allow Horizons' staff to gain insight to the database structure and function, the Cawthron Institute loaded historic macro-invertebrate data (average counts of insects and other in-stream critters) from 1999 to 2006. This saved Horizons an estimated 50 hours data entry time. Quality assurance checking has revealed some underlying data quality issues. These issues have nothing to do with the structure and function of the database itself.

A few recent electro-fishing records were added without problem.

5. Future Prospects

The Envirolink application is the first stage of a two-part application process. A second proposal (put forward in May 2008) focuses on documentation and support in the form of fixing any problems that are identified during testing, and developing a manual for data entry and database maintenance. It is anticipated this phase of the project will conclude before June 2009, subject to funding by the Foundation for Research, Science and Technology who administer the Envirolink fund.

CADDIS has the capacity to hold other water quality parameters such as physico-chemical data. Further advances could be made with Catchment Data to integrate other water quality data if that is deemed desirable. At present, there is an effort to integrate Horizons' bio-monitoring with physico-chemical monitoring by using the same sampling sites, sampling dates, and

ensuring all data are tagged with consistent site names. Having the separate kinds of data in two or more places, while inefficient, is not a barrier to integrating and analysing the data.

The CADDIS-Fly databases have been specifically designed to align with Cawthron Institute's CADDIS database. This may serve to assist feeding Horizons' data into national databases at any later stage if desired.