Design goals, technical implementation and pratical use of Another data entry system (Andes)

Daniel Ricard, David Fishman, Lindsay Beazley, Denis Bernier, Brian Boivin, Jamie Emberley, Ryan Martin, Jenni McDermid and Nicolas Rolland

Science Branch
Gulf Region
Fisheries and Oceans Canada
Moncton, New Brunswick, E1C 5K4, Canada

2022

Canadian Technical Report of Fisheries and Aquatic Sciences ####

Canadian Technical Report of Fisheries and Aquatic Sciences

Technical reports contain scientific and technical information that contributes to existing knowledge but which is not normally appropriate for primary literature. Technical reports are directed primarily toward a worldwide audience and have an international distribution. No restriction is placed on subject matter and the series reflects the broad interests and policies of Fisheries and Oceans Canada, namely, fisheries and aquatic sciences.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts*.

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Numbers 1-456 in this series were issued as Technical Reports of the Fisheries Research Board of Canada. Numbers 457-714 were issued as Department of the Environment, Fisheries and Marine Service, Research and Development Directorate Technical Reports. Numbers 715-924 were issued as Department of Fisheries and Environment, Fisheries and Marine Service Technical Reports. The current series name was changed with report number 925.

Rapport technique canadien des sciences halieutiques et aquatiques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui ne sont pas normalement appropriés pour la publication dans un journal scientifique. Les rapports techniques sont destinés essentiellement à un public international et ils sont distribués à cet échelon. II n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques de Pêches et Océans Canada, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports techniques peuvent être cités comme des publications à part entière. Le titre exact figure au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la base de données *Résumés des sciences aquatiques et halieutiques.*

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre.

Les numéros 1 à 456 de cette série ont été publiés à titre de Rapports techniques de l'Office des recherches sur les pêcheries du Canada. Les numéros 457 à 714 sont parus à titre de Rapports techniques de la Direction générale de la recherche et du développement, Service des pêches et de la mer, ministère de l'Environnement. Les numéros 715 à 924 ont été publiés à titre de Rapports techniques du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 925.

Canadian Technical Report of
Fisheries and Aquatic Sciences nnn

2022

DESIGN GOALS, TECHNICAL IMPLEMENTATION AND PRACTICAL USE OF ANOTHER DATA ENTRY SYSTEM (ANDES)

by

Daniel Ricard, David Fishman, Lindsay Beazley, Kevin Pauley, Denis Bernier, Brian Boivin, Jamie Emberley, Ryan Martin, Jenni McDermid and Nicolas Rolland

Science Branch
Gulf Region
Fisheries and Oceans Canada
Moncton, New Brunswick, E1C 5K4, Canada

Cat. No. Fs97-6/nnnE-PDF ISBN ISBN GOES HERE ISSN 1488-5379

Correct citation for this publication:

Ricard, D., Fishman, D., Beazley, L., Pauley, K., Bernier, D., Boivin, B., Emberley, J., Martin, R., McDermid, J. and Rolland, N. 2022. Design goals, technical implementation and practical use of Another data entry system (Andes). Can. Tech. Rep. Fish. Aquat. Sci. nnn: vii + 16 p.

CONTENTS

ABSTRACT

Ricard, D., Fishman, D., Beazley, L., Pauley, K., Bernier, D., Boivin, B., Emberley, J., Martin, R., McDermid, J. and Rolland, N. 2022. Design goals, technical implementation and practical use of Another data entry system (Andes). Can. Tech. Rep. Fish. Aquat. Sci. nnn: vii + 16 p.

A data entry system to facilitate the capture of information collected during scientific activities conducted by DFO Gulf Region was developed. Another data entry system (Andes) was implemented as a browser-based environment that facilitates its deployment in a variety of data capture scenarios. The design goals and implementation details of Andes are described, and a number of usage cases are presented. To promote endorsment of Andes by other scientific groups and to ensure the pereniality of the project, the application relies on Open Source software and uses a shared code development platform.

RÉSUMÉ

Ricard, D., Fishman, D., Beazley, L., Pauley, K., Bernier, D., Boivin, B., Emberley, J., Martin, R., McDermid, J. and Rolland, N. 2022. Design goals, technical implementation and practical use of Another data entry system (Andes). Can. Tech. Rep. Fish. Aquat. Sci. nnn: vii + 16 p.

Un système d'entrée de données pour faciliter la saisie des informations recueillies lors des activités scientifiques menées par la Région du Golfe du MPO a été développé. Another data entry system (Andes) a été mis en œuvre en tant qu'environnement basé sur un navigateur qui facilite son déploiement dans une variété de scénarios de saisie de données. Les objectifs de conception et les détails de mise en œuvre d'Andes sont décrits, et un certain nombre de cas d'utilisation sont présentés. Pour promouvoir l'utilisation d'Andes par d'autres groupes scientifiques et assurer la pérennité du projet, l'application s'appuie sur des logiciels Open Source et utilise une plate-forme de développement de code partagé.

1 Introduction

Scientific activities conducted by Fisheries and Oceans Gulf Region include ecosystem surveys and port sampling activities. These monitoring programs require the collection of a variety of information about marine ecosystems and about commercial fishery activities.

To support the data collection during ecosystem surveys, a paper-based system was initially used on board research vessels. With the advent of computing capabilities, the paper data sheets were digitized and the information was stored for subsequent analyses. In the 1980s, a data entry system was created by DFO to allow for the direct entry of data in digital format. The Groundfish Survey Entry (GSE) system was used during research cruises from the mid 1980s until the deployment of its successor, the Ecosystem Survey Entry (ESE) system in the early 2000s. Both systems were Microsoft Access applications.

While the ESE is a highly functional system, it also poses some limitations:

- Despite being used over a local area network (LAN), the system is not designed to work in a networked configuration.
- The interface of the system currently does not support for multiple locales, meaning that it is a unilingual tool.
- Visual Basic for Applications (VBA), the main development framework used to build the application, is a deprecated programming language and is no longer supported by Microsoft (REF NEEDED). To adapt the software using modern libraries and Application Programming Interfaces (API) in VBA will become increasingly difficult over time.
- The use of a version control system (VCS) is incompatible with MS Access. The absence of a VCS presents serious challenges for managing, disseminating and troubleshooting versions of an application across multiple platforms and is a major impediment for a collaborative approach to development.

Furthermore, the lack of a true server-side application (e.g., web application) meant agile development was effectively impossible; especially in the context of a mission that was underway. While the status quo was meeting the immediate data capture needs of scientists, a proactive stance towards addressing the above spurred the development of **An**other **d**ata **e**ntry **s**ystem (Andes). This project began during the summer of 2018, the first field deployment on Andes took place in the fall of 2019 during the southern Gulf of St. Lawrence September survey, where it was piloted in parallel to the ESE on board *CCGS Teleost*.

This report documents the design principles that guided the development of Andes, provides technical details about its implementation and details the usage cases encountered thus far. Finally, we provide guidelines for assessing the suitability of Andes for noval applications (e.g., other field activities and sampling programs) and discuss the ways in which this system could gain further adoption within DFO.

2 Methods

The ESE was the starting point in the development of Andes. The data entry software had to first replicate the capabilities of the ESE. Early version of Andes achieved those goals and the system was further developed over its usage by scientists in the Gulf Region.

2.1 Usage cases

2.1.1 Ecosystem survey data entry

The Andes system is designed to be deployed as the main data entry platform during ecosystem surveys. The application's data entry interface for usage in a wet laboratory (Figure 3) allows the capture of all the information required by the sampling protocols used for the survey

2.1.2 Oceanographic metadata collection

Biological surveys often collect complementary oceanographic data. Coordinating these data with the associated biological data is a burdensome logistical challenge. Historically, the different types of data were captured and recorded using separate systems that operated in complete isolation from one another. The original intent of the Oceanographic module of Andes was to streamline the capture of oceanographic data and to simplify its association with the corresponding biological data. The Oceanographic module has since been evolving into a stand-alone component of the application that is capable of being deployed independently of the Ecosystem Survey module.

2.1.3 Commercial sort sampling

Another use of the software is to support data collection during sampling of commercial port sampling activities. (Benoît and Daigle 2007)

2.2 Technical implementation of design principles

The following section outlines the design principles that played a role in decision-making and provide a description of how they were implemented in the project.

2.2.1 Flexibility, scalability and reliability

Within the context of DFO Science, there is a high number of use-cases under which this application will be deployed. For example, the number of users might range from a single user to several dozen. Similarly, the application might sometimes be used in a networked setting while other times used in the field on a stand-alone tablet. Accordingly, the application has to accommodate a wide range of practical scenarios.

The Django Web Framework was selected for the backend of this application due to its modularize nature; virtually all aspects of the framework can be decoupled. Furthermore, the Django framework is written in pure Python; an open-sourced, generalized object-orientated programming language that is popular for use in data-heavy applications. While web applications are most often used over a network, the django library comes with a development web-server that permits users to serve and use the application locally. In this scenario, a single computer acts simultaneously as server and client. While there are important limitations to the use of the django development web-server in a full-scale production environment, this configuration is adequate for the stand-alone use-cases referred to above.

2.2.2 Version Control / Source Control

The utilization of a version control system (VCS) is an indispensable component of a modern day development workflow. Version management is especially important in the context of having concurrent instances of the application in production at a given time. Knowing the version of a production instance is necessary to resolve any issues that might arise. Similarly, databases and backup files are intimately linked to an application's version number. In order to successfully re-instantiate a backup file, the precise version under which it was produced must be known. The VCS will also provide an indispensable framework for the coordination, examination and integration of contributions from collaborators.

It was decided to use Git as the VCS for this project due to the fact that it is both widespread in use and open-sourced. The remote repository for this project is currently hosted on the Gulf Science organizational GitHub account. The project is additionally making use of GitHub infrastructure including, pull requests, issue management (e.g., bugs, feature requests and general enhancement requests) and security alerts.

2.2.3 Unit testing

The application also needs to be reliable especially considering its use in the field setting. The implementation of unit tests is a practical way to ensure the maintenance of functionality over time. On one extreme, the addition of any code can be preceded by the creation of unit tests (i.e., test-driven development). This approach will maximize the stability of an application but can hinder the momentum of a project, especially in its early stages. On the other extreme, application development in the complete absence of unit testing occurs at a relatively fast pace but will result in a project that is vulnerable to breaking in unexpected ways and one that is difficult to maintain and pass on to other developers.

Using the built-in Python / Django testing framework, this project will seek to employ a middle ground between the above two approached. While the goal is not to implement test-driven development, the use of unit tests are highly encourages, especially to back up the core functionality of the application.

2.2.4 User-defined protocols

The flexibility of project leads to design, modify and report on their sampling protocol without depending on developers is very important.

2.2.5 User Interface

By ensuring the application has a modern and intuitive interface, the barriers related to onboarding new users are significantly reduced. Furthermore, an intuitive interface will reduce the need for extensive help documentation. Wherever extra annotation is required, documentation will be inserted directly in the application in the form of tool tips and help bubbles. By appealing to the end users' intuitions and by providing in-situ help documentation, we reduce the likelihood fields and features being used incorrectly.

The above was achieved by building the frontend of the application in HTML 5, javascript and CSS. Most users will be familiar with the flow and functionality of a web browser and will be comfortable navigating and entering data into a website. Facilitated by the Django model and form classes, all controls (i.e., fields) on the website will contain verbose descriptions and help text. The Bootstap v5.0 css and javascript libraries were utilized in order to give the application a sleek, modern look and to ensure compatibility with different types of devices (e.g., personal computers vs. tablets vs. mobile devices). The Bootstrap libray also provides palatable styles for displaying help text in the form of popovers and tooltips.

2.2.6 Reactivity

In the context of being on a research vessel survey, data entry happens at a very fast rate; often with multiple transactions per second. Accordingly, it is imperative that the application does not create a bottleneck for data entry and is able to keep pace with experienced technicians. The usage of a reactive javascript library in conjunction with an API would allow data entry to occur without webpages having to constantly refresh.

The Django REST framework was used to construct the WebAPI component of the application. The project took a hybrid approach, combining the use of standard Django views and Vue.js frontend applications embedded in the templates. The latter were built to avoid the need for constantly reloading webpages and to keep optimize the flow of traffic across the network. Reactive javascript frontend applications also provide a better experience from the point of view of an end-user.

2.2.7 Multilingualism

Allowing users to utilize the application in the language of their choice is of considerable importance to this project. Previous tools that have been used were uni-lingual and this by itself would have limited the scope of their use in a national context. The Django framework has excellent support for internationalization and localization, including the translation of text and the formatting of dates, times and numbers. It achieves this using a system of “hooks” used by developers to indicate which parts of the code should be localized. See Django - Internationalization and localization for more details on this process. In our application, an end-user can toggle between English and French by simply clicking on a button. In this way, each client can view the application in the language of their choice.

2.2.8 Backup Regime and infrastructure

- Backup are automatically created when sets are closed
- All backup files reference the version of the application which produced them. Even old backups can be loaded by locally re-instantiating the version of the application under which it was produced

3 Results

3.1 User definitions and permissions

Andes provides access to its different components differently based on the credentials of users. While the chief scientist can change sampling requirements for the mission, other users won't.

3.2 Ecosystem survey app

This component of Andes replicates the capabilities of the ESE for capturing detailed information on length, weight, ageing material, maturity, etc. about fish and invertebrate specimens.

3.2.1 App used in the laboratory

The app running on computers in the wet laboratory of the vessel is used for all entry of data related to measurements and observations of marine organisms.

3.2.2 App used in the bridge

The console to be used by navigation officers is meant to run on a tablet in the bridge of the vessel conducting the survey.

This console contains two separate sections: 1) the Set Card section and 2) the Fishing section.

Additionally, a number of events that take place during fishing can be captured in the console.

Information coming from Scanmar sensors are captured and stored by ANDES.

3.2.3 Forecasting tool

Part of the task-specific suite of utilities is a forecasting tool that provides real-time prognostics of survey completion targets based on assumed transit speeds and time spent fishing and processing the trawl catches.

3.2.4 Progress map

Another task-specific utility is the completion map which shows what strata have been completed based on target and minimum number of sets per stratum.

3.2.5 Reporting capabilities

Two types of reports are available for Andes: 1) reports meant to be used during field activities and 2) reports meant to be used after field activites are completed.

3.3 Shrimp ecosystem survey app

Quebec: Denis and Brian could contribute here?

3.4 Oceanography app

The oceanography data collection is handled by a separate app that is used in the deployment of the rosette and the processing of water samples obtained at different depths.

3.5 Cruise dashboard

This utility is meant to obtain a summary of the cruise, it includes the current position and speed of the vessel, the list of the most recent catches in the wet laboratory, as well as running totals of specimens captured.

3.6 Port sampling app

Andes was adapted to support port sampling activities where technicians obtain length frequency samples from commercial fishing activities. To facilitate deployment, the port sampling app of Andes is deployed on field tablets with a custom interface that is suitable for use on boats, outdoor and in inclement weather conditions (Figure 4).

3.7 Tasks for administrative user

Prior to a survey, a number of configurations must be done for Andes to function as intended. At a minimum, the following information must be ingested by Andes in order for a mission to take place:

1. Strata polygons
2. List of stations
3. Species list
4. Sampling requirements

3.8 Backup regime and infrastructure

3.8.1 Unit Tests

A number of unit tests have been created for the application.

3.9 Data models

Simplified Entity Relationship Diagrams (ERD) of the following components:

3.9.1 Sampling Protocols

3.9.2 Ecosystem Survey

3.9.3 Port Sampling

3.9.4 Oceanography

3.9.5 Integration with navigation system and trawl sensors

Full ERD on Figure 1

3.9.6 Collaborative development

- translations were done in house
- they must be up-kept along with development
- at the time of writing this document, the project has a total of 6 contributor to the code.

3.10 Deployment Checklist

Here is a deployment checklist for deploying an instance of the Andes application in a production environment:

- ☐ It is always a good practice to set the `DEBUG` setting to `False`. Using DEBUG in production can result in memory leakage. The debug setting is set via the project's .env file, located in the andes root directory (e.g. `/opt/andes_root/andes/.env`)
- ☐ Set Andes Leads and Chief Scientists to Admin users.
- ☐ Make sure GPS is working on site
- ☐ Make sure printer is working on site
- ☐ Make sure the backup works
- ☐ Make sure custom maps are working (if applicable)
- ☐ Check everything again after a reboot of the server
- ☐ Make sure you can run shell scripts from the front end (optional)
- ☐ Alderbarren setup
- ☐ Reset the Specimen table autoincrement (see basic_setup)
- ☐ If using cradlepoint, has web filtering been turned on?
- ☐ Check to make sure that the time/date on any clients (especially Bridge clients) are in sync with server time

4 Discussion

The current capabilities of Andes have evolved over the last three years where the system has been used to support field activities in the Gulf Region. The software has also been used by the Maritimes and Quebec regions of DFO.

Versioning details and why the backup process is so awesome.

Things will change.

4.1 Creation and maintenance of unit tests

4.2 Integration of Andes with existing data repositories

Discuss the steps required to go from data collected in Andes to transferring them to existing databases.

4.3 Future Directions

- AZMP / Oceanographic Surveys
- use for ecosystem surveys by multiple DFO regions
- use of barcodes and/or QR codes for identifying and tracking down samples (e.g. otoliths, tissue samples, . . .)

5 Acknowledgments

6 Figures

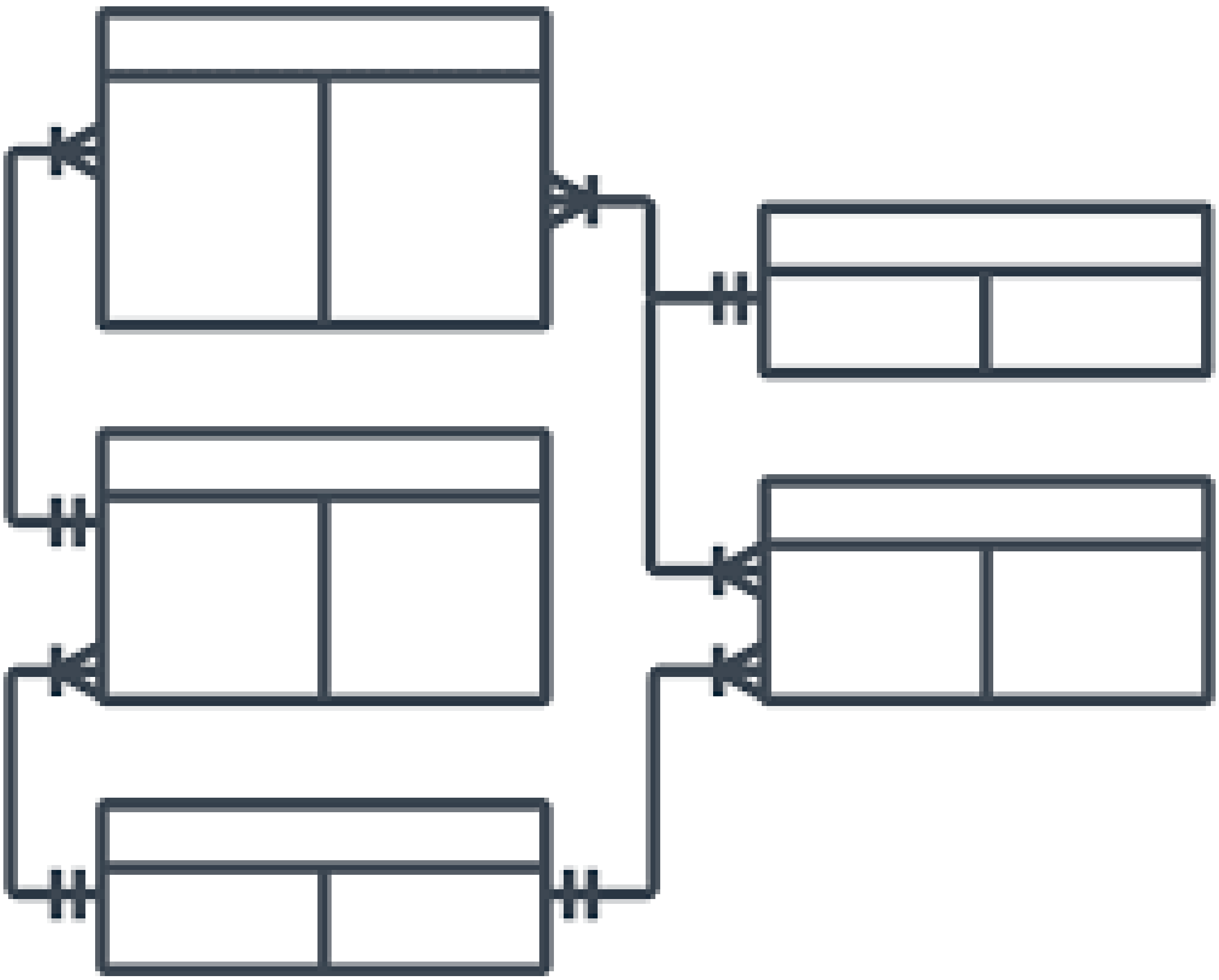

Figure 1. Entity relationship diagram of the data model used by ANDES.

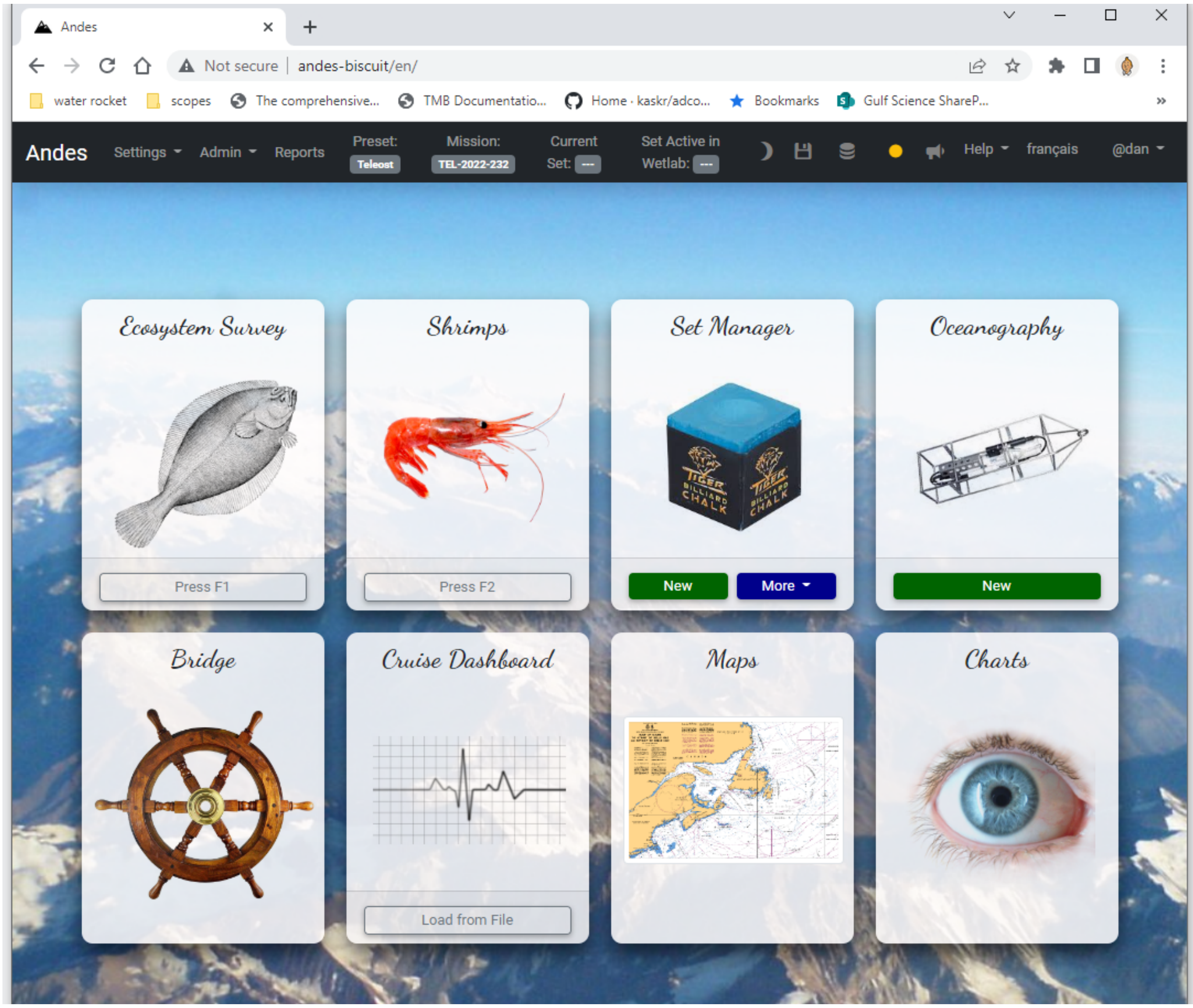

Figure 2. User interface of Another data entry system (Andes) showing the eight principal components of the application.

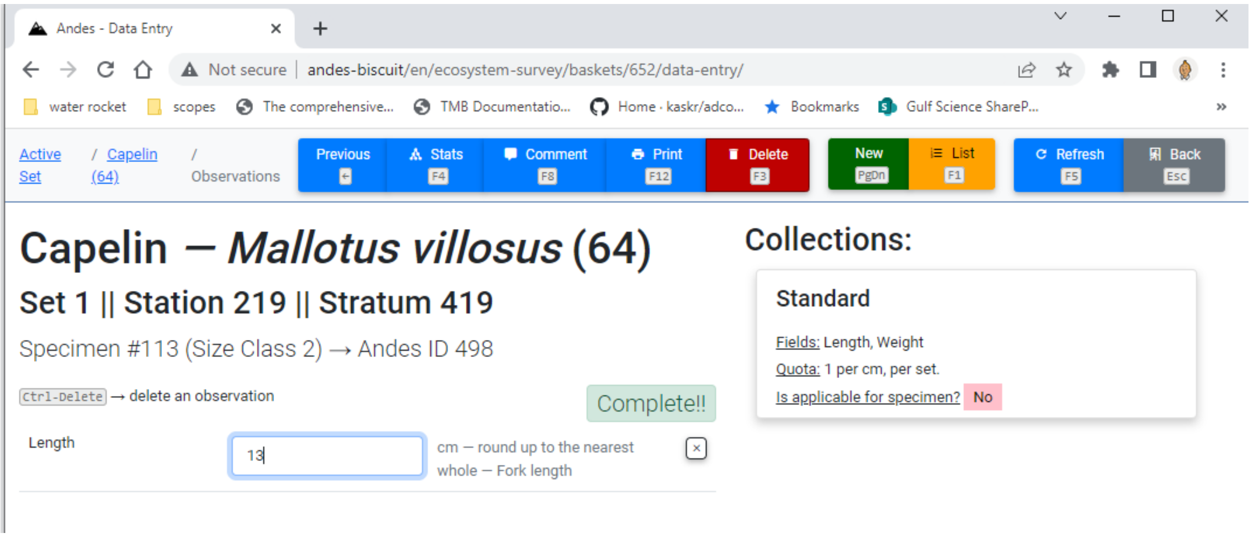

Figure 3. User interface of the Ecosystem Survey component of Andes when in use for processing samples in the wet laboratory. The screen capture presented here shows sampling of capelin lengths on an active set being processed in the lab.

Figure 4. User interface of the Port sampling component of Andes when in use for processing samples in the field.

APPENDIX A Placeholder for materials that would appear in an Appendix

7 References

Benoît, H.P., and Daigle, D. 2007. Procedures for commercial catch sampling of finfish and shrimp in the southern Gulf of St. Lawrence. Can. Manuscr. Rep. Fish. Aquat. Sci. 2833: iv + 63 p.