

Regional Data Base

Summary

Recognising that Council has endorsed the further development of the Regional Database for Commercial catch sampling (RDB) both in 2014 (680 000 DKK from equity) and 2015 (300 000 DKK from equity). The future of the RDB now stands at a cross-roads. By redesigning the RDB now to anticipate and answer the needs of the ICES assessment groups, ICES will be in a strong position to:

- Reduce the workload for the countries in estimating and providing data (one data call for detail data that would also serve the raised data for stock assessments)
- Ensure quality assured standardised statistical methods (expert driven) are used for estimating the data for the stock assessment
- Provide a commercial catch data processing platform for all ICES countries (to avoid an EU and non-EU system for ICES stock assessments)
- Describe and document data quality by using common quality checks across all countries' data

Described in this paper is a plan of how this could be achieved. The effort amounts to 4.5 person years – 2.5 of these person years can be sourced from the existing pool of resources within the Secretariat and focussing almost entirely on this development. A further 2 person years of a technical resource would be needed to achieve the timeframe of achieving this development in 2 years. This would mean an investment of 1 million DKK from equity to support this activity.

1 Achievements

The RDB has been hosted and maintained by ICES under agreement with the European Commission (MoU) since 2012, in addition ICES have provided funds for the further development of this system. Based on this funding model the following has been achieved:

Maintenance:

- Operational system, actively used and maintained
- Regional standardization of codes and quality control of input data
- Helpdesk for data providers and users
- Delivery of regional data for the three Regional Coordination Meetings (RCM's) committed to its use
- Agreed data provision (data call) and data access (data policy)
- Main work platform of the three RCM's with all respective countries uploading data to the RDB

Development:

- Further standardisation of codes and quality control of input data, improvements of uploads, and report outputs
- Support for the new landing categories (i.e., landings above and below the minimum conservation size) resulting from the EU landings obligation

2 The main challenge for the RDB

In order to use the fisheries dependent data collected by member countries to provide a documented, quality assured and accurate description of the fisheries and their catches to be used in scientific advice on management of fisheries, there is a strong recommendation from ICES expert groups (PGDATA, WGCATCH, WKRDB), the EMFF funded fishPi project, and the wider ICES end user community to make it possible to raise the collected data to fisheries levels using statistically sound methods. To accomplish this the RDB must be further developed to ensure that:

- The input format support design based sampling and probability information (data need for statistical raising);
- The raising method support statistical raising.

In Figure 1 the current process for raising data is demonstrated. The first vertical arrow from the left hand side shows that data are exported from the RDB for use by the RCM's. The dotted blue line indicates that the system cannot raise/estimate biological data using statistically sound methods for the international stock assessment. This creates a system that falls short of its aim to allow multiple uses of the data, and where countries deliver data twice; Detailed data to the RDB for the RCMs and raised/estimated data to InterCatch for the stock assessment, and where the end user (ICES) is not able to make full use of the RDB as a tool to evaluate the quality of the data since there is no documentation of the national raising/estimation methods.

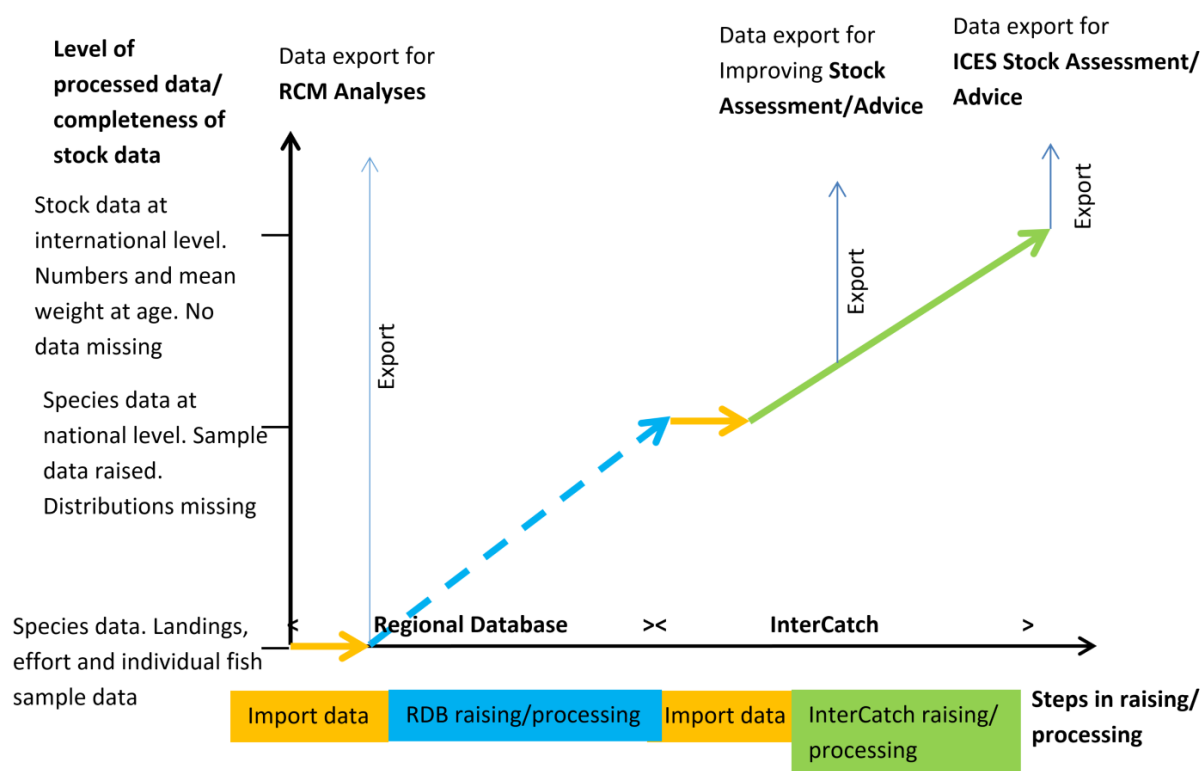


Figure 1 The current process for Stock Assessment

3 The shared vision for the RDB

- Reduce the workload for the countries in estimating and providing data, as the RDB would contain (or can utilise from R libraries) all needed methods
- Ensure quality assured standardised statistical methods (expert driven) are used for estimating the data for the stock assessment
- Provide a commercial catch data processing platform for all ICES countries (to avoid an EU and non-EU system for ICES stock assessments)
- Describe and document data quality by using common quality checks across all countries' data
- Reduce the workload for the countries submitting data, as the data raised in the RDB by a button click automatically will be made available to InterCatch for the ICES stock assessments
- Support the Regional Coordination Groups/Meetings with data and reports for their work
- Data are encapsulated within the RDB (the data is safeguarded in the RDB and the end user understands every change to the data)
- Leverage the body of work already existing in R code projects and developed further by the experts
- Links to other databases e.g. the VMS/Logbook database used by WGSFD, ByCatch regulation, Fisheries independent data (i.e. DATRAS)

The vision is illustrated in Figure 2. Not only would the RDB support the planning and reporting work of RCGs/RCM's, but it would be able to directly support the data needs for scientific advice on fisheries management including stock assessments and provide outputs at all levels in the process in a quality assured manner.

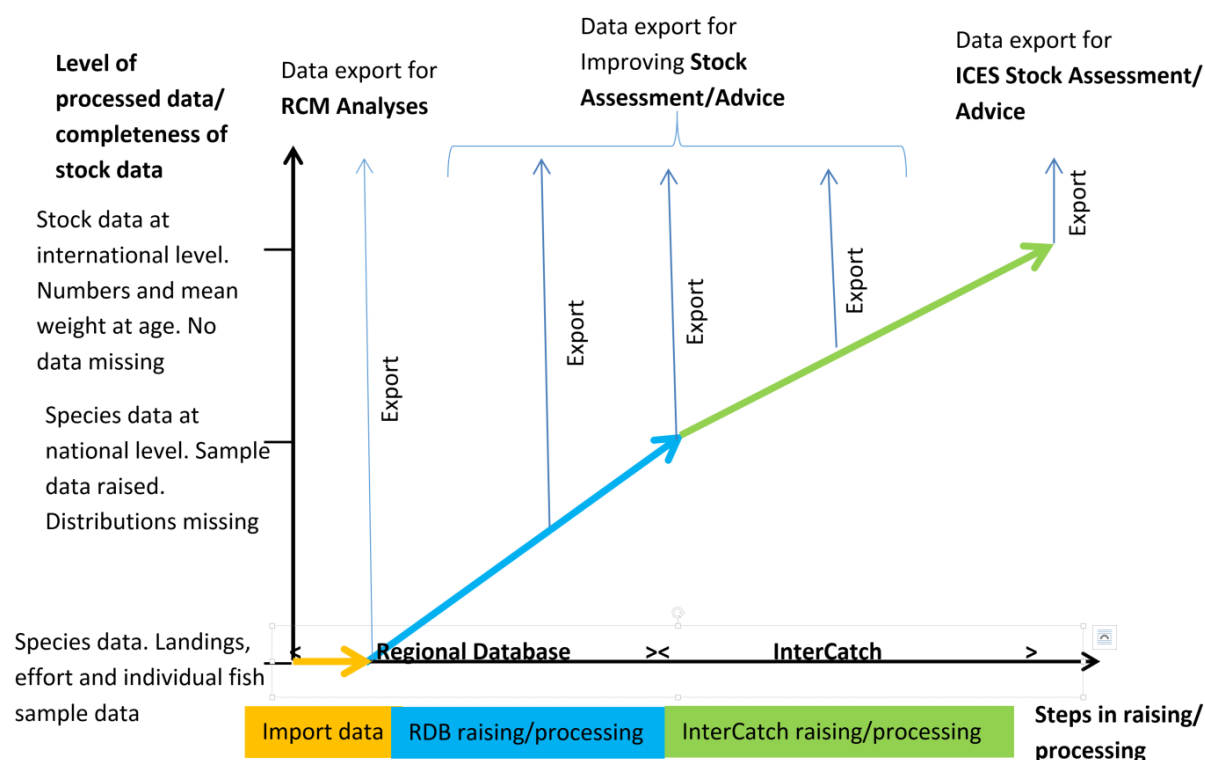


Figure 2 The vision for Stock Assessment

4 How to get to a statistically sound RDB

PGDATA, WGCATCH, WKRDB and the fishPi project recommend to update the data exchange input format with the necessary information that would enable statistically sound raising. The extra information concerns the sampling design and probabilities on all levels of the sampled imported data.

The raising should be based on statistical sound methods instead of the existing methods combining age-length-keys, etc. These statistical methods are documented and available already in *R*¹, so currently the existing RDB is implementing the raising methods 'behind the scenes', the new approach should be to call on the existing statistical methods available in *R*. This would make the raising more transparent, and easier for the experts to update the methods, if needed. It is important that the data providers and expert groups take an active part and are involved in a transparent process to ensure the RDB fulfils the needs for uploading their design based sampling information and raising data using statistical methods. Therefore, workshops involving all relevant groups should be considered to scope the further development of the RDB. Figure 3 below gives an overview of the flow of information and the interaction with the national experts to the two main end users, but data could also be extracted for other relevant end users e.g. STECF.

¹ <https://www.r-project.org/>

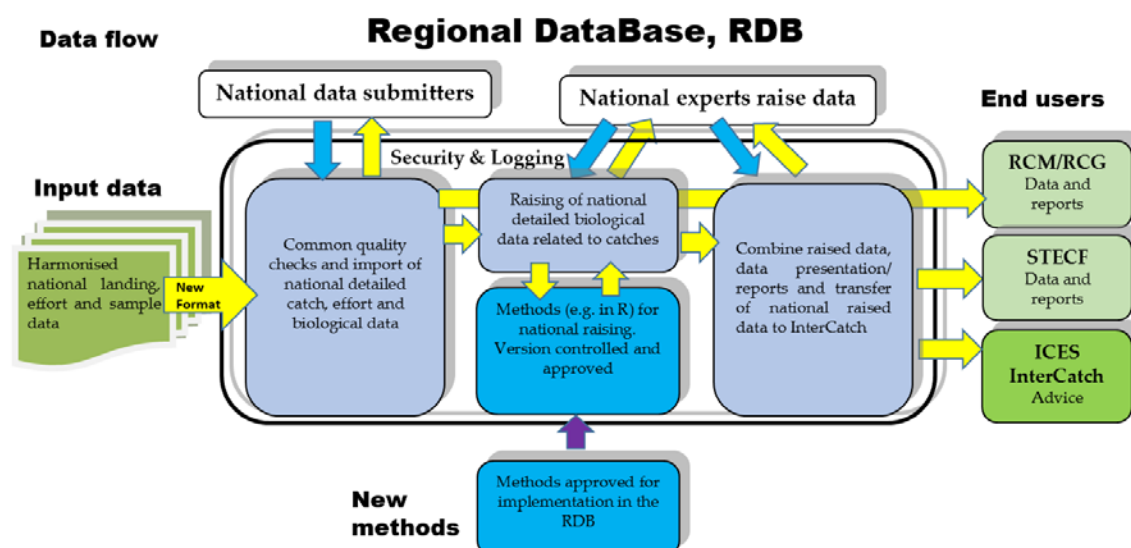


Figure 3 Architecture of the proposed RDB

5

This vision could be realised in 2 years. The project is estimated to require 4.5 person years in total effort.

Effort (person months)	Tasks completed	Workshops
16 PMs	RDB system specification, 1 st phase design and development modules	1 st Workshop
15 PMs	Database design and development 2nd phase, upload and checks, version control of approved methods in R-script	2 nd Workshop
13 PMs	Raising methods process, overview and deletion of uploaded data, download of data including RCG reports	3 rd Workshop
11 PMs	Approval of raised national data and transfer to InterCatch, stock splitting	

Figure 4 person months and tasks

Financial aspects

The maintenance of the RDB is currently covered by the MoU with EU. The need for funds to cover the developments was discussed at the last Liaison Meeting. The Liaison Meeting is a meeting organized by the European Commission with the chairs of the various Regional Coordination Meetings and the main DCF data end users (ICES, JRC, and GFCM). All members endorsed the developments of the RDB and considered it the main tool for a regional coordination data collection programme. In relation to the financial aspects for the development, several possibilities were discussed and ICES was clear that including the developments in the MoU with the current budget ceiling is not an option.

From Commission (via the MoU with ICES)	Investment in Euros
ICES receives annually for hosting and maintenance	74 000 (annually)
From ICES Council	
Approved initial development in 2014	91 000
Approved further development in 2015	40 000
Development funding will be exhausted by March 2017	

Figure 5. Funding model for RDB

6 Linking to the Transparent Assessment Framework

Both ICES and the GFCM are working on assessment framework systems – the ICES version is under development and will start to be tested on a number of stocks in 2017. The goal of the transparent assessment framework is to have a fully encapsulated system – from the input data, to the stock assessment models, to the eventual stock assessment result outputs. Therefore, the proposed RDB development is a major building block in ensuring that there is a seamless link between the input data and the assessment modelling (See Figure 6).

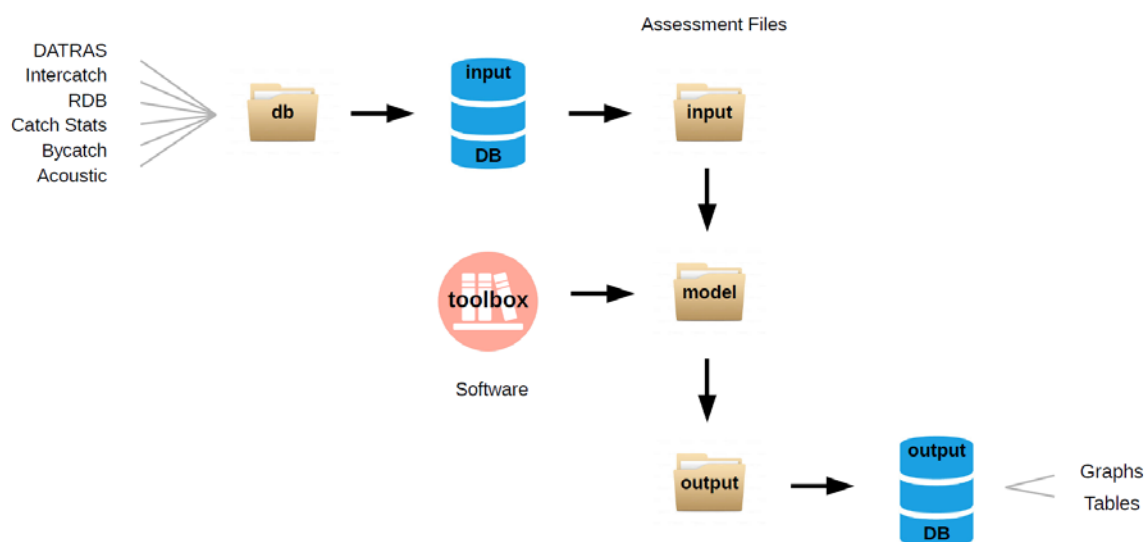


Figure 6. The Transparent Assessment Framework

7 Other considerations

7.1 The technical architecture of the RDB website

The current RDB web interface is built on outdated software architecture, in moving to a statistically sound RDB it would be logical to redevelop the interface at the same time using up to date technology, which will offer more functionality. It will be possible to reuse some of the existing code and logic, which will make the development faster, but it will be a good investment to spend the time developing the RDB using the latest software framework.

7.2 Consequences of not developing the RDB

If the RDB is not developed to support statistical raising, then the countries will have to use other tools to raise their data according to statistically sound principles. This could result in an uncoordinated and undocumented approach within each region. The data, which are used for the stock assessment and subsequent catch advice, will not have been prepared or checked to the same uniform standard across countries. It will not be possible to fully document how the raising has been done, or which methods and data have been used. This would result in poorly documented data quality, which will affect the resulting assessment and advice.

8 Development tasks of the statistical RDB

The following is a more detailed description of the tasks for developing the RDB, so it fulfils all the requests of the RDB and support the countries in statistical raising of the data.

RDB system specification

Overall system specification of what functionalities the RDB should have. Coordination of approval and dialogues with a group of experts, who can approve/help with the specification of the functionalities and later perform the testing of the developed functionalities.

Specification	Interact with ICES RDB statistical expert group
	Identify the latest version of the updated exchange format
	Identify the outputs from raising methods in R
	Specification document on upper level of the overall RDB

Database design and development

Specification	Write detailed specification according to the latest exchange format and user security in the first round, the other modules will follow.
Development	Develop and implement the database design for the uploaded data and the user security in the first round, the other modules will follow. User id and creation date and time added to all import tables
Test	All of the above including unit and integration test

User security

The user security will be based on country, region, and a few needed roles

Specification	Write detailed specification of the user security and maintenance Write detailed specification of the menu structure
Development	Develop and implement: <ul style="list-style-type: none">• Setup the RDB in the TFS• User security maintenance• Login• The user security on pages

Test	All of the above including unit and integration test
------	--

Note: Maybe a role for updating the methods, but I think it should be in the hands of ICES, since we have the responsibility for making sure the R-script is working.

Code maintenance from ICES internal code system RECO

Specification	Write detailed specification of the code maintenance from RECO to lookup tables in the RDB
Development	Develop and implement the code maintenance from RECO to lookup tables in the RDB. Use procedures from the acoustic db.
Test	All of the above including unit and integration test

Upload and checks

The checks are the existing using XSD and with extra field dependency range checks (e.g. WECA) written in C# in the RDB. Maintain a list of checks, which are implemented in C#, so the users know which checks are performed.

Specification	Write detailed specification of the upload and checks Write detailed specification of the menu structure
Development	Develop and implement the data upload from the file to the database, data will not be deleted, data will be added and the latest version will be used for further raising Develop and implement the checking of the data: XSD, convert XQuery checks to C# Develop and implement an overview of checks implemented in C#
Test	All of the above including unit and integration test

Issues to look into:

- Should the overwrite rules include institute? This will open up for the possibility of having duplication of data, which should be avoided.

Note: Based as much as possible on web services so we in the future easily can let countries upload data automatically using a web service. Design the upload so it accommodate for adding extra information without huge changes to the RDB. The

upload have to be optimised and using the fastest technics. The existing code is not optimal and it can takes more than an hour to upload a file.

Checks in R for data upload

Interface with version controlled checks programmed in R.

Specification	Write detailed specification of the version control and interface to checks written in R
Development	Develop and implement a direct import of the uploaded data into temporary import tables for R checking Develop and implement of the version control and interface to checks written in R, so the checks are stored, can be viewed and executed
Test	All of the above including unit and integration test

Issues:

- Ensured that the checks written in are harmonised way or are using a template and can just be plugged in
- In case of errors how detailed are each check in feedback of the error to the user

Overview and deletion of uploaded data

Overview for countries and RCG

Specification	Write detailed specification of the filter and overview of the uploaded data with a functionality to delete the selected data. The deleted data will be flagged as deleted Write detailed specification of the RCG specific page with an overview of aggregated uploads
Development	Develop and implement the filter and overview of the uploaded data with an option to delete/set a bit in the deleted field of the selected data Develop and implement the RCG specific page with an overview of aggregated uploads
Test	All of the above including unit and integration test

Download of data

Country and RCG

Specification	Write detailed specification of the filter for the data which should be downloaded
Development	Develop and implement of the filter for the data which should be downloaded and the download functionality
Test	All of the above including unit and integration test

Note: Should be based on a service, so countries in the future can download data using a service.

Version control of approved methods in R-script

Create a version of the new method and wrap it in to a stored procedure. Link to the method to the previous version of the same method or create it as a new method. Testing of the method inside the RDB would be needed. A person from an ICES statistical methods group (E.g. WGCATCH), who is authorised to approve methods should acceptance test and approve the method. The script will be save both in a SP and also in TFS. The method written in R would be in the ICES GitHub and from there the script would be evaluated, tested, outputs compared, and when approved it should be send to ICES. The final approved version would then be included in the RDB.

Specification	Write detailed specification of the version control of raising methods written in R
Development	Develop and implement the version control of raising methods written in R
Test	All of the above including unit and integration test

Raising process

A snapshot of uploaded data used for raising is taken, which makes it possible to reproduce the raising in exactly the same way, independent of later uploaded updates of data. It will also be possible to use other methods on exactly the same data and compare the output.

Specification	Write detailed specification of the execution of the version controlled methods in R, the snapshot of the sample data used for the raising, a page for settings for the raising
---------------	---

method, the output data from the raising methods and logging of everything. have to be specified

Development	Develop and implement the execution of the version controlled methods in R, the snapshot of the sample data used for the raising, a page for settings for the raising method, the output data from the raising methods and logging of everything
Test	All of the above including unit and integration test

Note: Differences in sample data can limit the statistical methods used, this should be incorporated in the selection of available methods.

Stock splitting

Specification	Write detailed specification of the stock splitting functionality
Development	Develop and implement of the stock splitting functionality
Test	All of the above including unit and integration test

Approval of raised national data and transfer to InterCatch

Specification	Write detailed specification of the approval of raised national data and transfer to InterCatch
Development	Develop and implement of the approval of raised national data and transfer to InterCatch
Test	All of the above including unit and integration test

RCC reports

Specification	Write detailed specification of the existing reports, selection and execution of a report
Development	Develop and implement the existing reports and the selection and execution of a report

Test	All of the above including unit and integration test
------	--

Project management

Specification	Guidance, decisions, coordination internal and external, meetings, status.
---------------	--

Reports from R scripts

Specification	Write detailed specification of the version control of approved reports from R scripts and the execution of the reports
---------------	---

Development	Develop and implement of the version control of approved reports from R scripts and the execution of the reports
-------------	--

Test	All of the above including unit and integration test
------	--

Workshops

There should be three workshops during the development of the RDB, the aim is both to get all countries involved, but also to focus on the latest improvement of the exchange format and raising methods.

1st workshop

The first workshop should focus on the latest exchange format for statistical raising, at one point it has to be determined that the format can fulfil the needs, and that format will then be frozen until a significant change has to be included. The workshop should also work through the RDB system specification, which include the overall functionality of the RDB. If functionality is missing or not clear the workshop can specify needs. Finally the workshop should approve the specifications.

2nd workshop

The second workshop should focus on checks, the version control of methods in the RDB and the use of the methods for raising.

3rd workshop

The second workshop should focus on the development of the RDB at the stage it is, and identify issues, which need to be dealt and suggest solutions, with special focus on the data raising process. Standard reports should also be specified.

9 The project timeline

The project could be achieved in 2 years. The project is estimated to require 4.5 person years in total effort.

Effort (person months)	Tasks completed	Workshops
16 PMs	RDB system Specification Database design and development 1st part User security Code maintenance from RECO Checks in R for data upload	1 st Workshop
15 PMs	Database design and development 2nd part Upload and checks Version control of approved methods in R-script Project management	2 nd Workshop
13 PMs	Raising process Overview and deletion of uploaded data Download of data RCG reports	3 rd Workshop
11 PMs	Approval of raised national data and transfer to InterCatch Stock splitting Reports from R scripts	

10 Architecture

Development methodology

Web application: MS Visual Studio 2015 using ASP.NET Core web application with the .NET framework

Database: MS SQL Server 2016 with R services

Project steering methodology

Agile project management