



d4SCIENCE

Project acronym	D4Science
Project full title	DI stributed coL aboratories I nfrastructure on G rid E nabled T echnology 4 S cience
Project No	212488

Deliverable No N/A

Fishery Country Profiles Production System (FCPPS) Vision

March 2008

SEVENTH FRAMEWORK PROGRAMME
Research Infrastructures

INFRA-2007-1.2.2: Deployment of
e-Infrastructures for scientific communities



DOCUMENT INFORMATION

Project	
Project acronym:	D4Science
Project full title:	DI istributed col L aboratories I nfrastructure on G rid EN abled T echnology 4 S cience
Project start:	1 st January 2008
Project duration:	24 months
Call:	INFRA-2007-1.2.2: Deployment of e-Infrastructures for scientific communities
Grant agreement no.:	212488
Document	
Deliverable number:	N/A
Deliverable title:	Fishery Country Profiles Production System Vision
Contractual Date of Delivery:	[N/A]
Actual Date of Delivery:	[03 2008]
Editor(s):	[Editor(s) Name]
Author(s):	[Yves Jaques]
Reviewer(s):	[Reviewer(s) Name]
Participant(s):	[FAO]
Work package no.:	[NA5]
Work package title:	[Title]
Work package leader:	[Partner short name]
Work package participants:	[Partner short name]
Est. Person-months:	1
Distribution:	[Public]
Nature:	Report
Version/Revision:	1.0
Draft/Final	[Draft]
Total number of pages: (including cover)	34
Keywords:	D4Science; grid; Country profile; FAO; Fisheries; VRE, Vision; requirements;

CHANGE LOG

Reason for change	Issue	Revision	Date
Comments from M. Taconet	1	0	06/04/08
FI review	2	1	16/04/08
FIES review	3	2	21/04/08
		3	

CHANGE RECORD

Issue: 1 Revision: 1

Reason for change	Page(s)	Paragraph(s)
Missing some stakeholders and resources.		

Issue: 1 Revision: 2

Reason for change	Page(s)	Paragraph(s)
Not adequate definition of current workflow.		
Better domain descriptions.		
Addition of flowchart appendix		

Issue: 1 Revision: 3

Reason for change	Page(s)	Paragraph(s)
Review by statistics department.		
Revision of users and stakeholders.		
Revision of functionalities		

DISCLAIMER

This document contains description of the D4Science project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium head for approval.

E-mail: info@d4science.research-infrastructures.eu

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of D4Science consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 27 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors.
(<http://europa.eu.int/>)



D4Science is a project partially funded by the European Union

LIST OF ABBREVIATIONS

D4Science	DIstributed colLaboratories Infrastructure on Grid Enabled Technology 4 Science
EC	European Commission
FCPPS	Fishery Country Profiles Production System
VRE	Virtual Research Environment
RFB	Regional Fishery Bodies
JDBC	Java Data Base Connector
API	Application Programming Interface
RSS	Really Simple Syndication
NEMS	News and Events Management System
XML	eXtensible Markup Language
HTML	Hypertext Markup Language
UN	United Nations
EGEE	Enabling Grids for E-scienceE
CSV	Comma-separated values
ESA	European Space Agency
FAO	Food and Agriculture Organisation of the United Nations

TABLE OF CONTENTS

1	INTRODUCTION	1
1.1	Purpose	1
1.2	Scope.....	1
1.3	Overview	1
2	POSITIONING	2
2.1	Business Opportunity.....	2
2.2	Problem Statement	2
2.3	Product Position Statement	3
3	STAKEHOLDER AND USER DESCRIPTIONS	4
3.1	Market Demographics.....	4
3.2	Stakeholder Summary	4
3.3	User Summary	5
3.4	User Environment	6
3.5	Workflow	7
3.6	Stakeholder Profiles	9
3.7	User Profiles.....	12
3.8	Key Stakeholder or User Needs	15
3.9	Alternatives and Competition.....	18
4	PRODUCT OVERVIEW.....	19
4.1	Product Perspective.....	19
4.2	Summary of Capabilities	21
4.3	Assumptions and Dependencies	22
4.4	Cost and Pricing.....	22
4.5	Licensing and Installation.....	22
5	PRODUCT FEATURES	23
5.1	User access definition	23
5.2	Data	23
5.3	Data access	24
5.4	View	24
5.5	Workflow definition	25
5.6	Reporting definition.....	25
5.7	Output	25
6	DOCUMENTATION REQUIREMENTS.....	26
6.1	User Manual.....	26
6.2	Online Help	26
APPENDIX A.	CURRENT WORKFLOW	27

1 INTRODUCTION

1.1 Purpose

The purpose of this document is to collect, analyze, and define high-level needs and features of the Fishery Country Profiles Production System (FCPPS).

1.2 Scope

It focuses on the capabilities needed by the stakeholders and the target users, and **why** these needs exist. The details of how the FCPPS fulfills these needs will be part of the larger requirements gathering process within D4Science, specifically the TRAC requirements system.

1.3 Overview

This document sets out the case for the FCPPS. It begins by defining the problem domain, including the opportunity the software represents. It goes on to cover the stakeholders, defining their roles and relationships to FCPPS. It also covers the key user needs which the software attempts to satisfy and possible other solutions to the solution here envisaged.

The vision then goes on to put the product in perspective: what are its capabilities, its assumptions, dependencies and licensing issues. The document then lists a set of product features before finishing with an overview of documentation requirements.

2 POSITIONING

2.1 Business Opportunity

The Country Profile VRE (FCPSS) is an opportunity for the Fisheries and Aquaculture department to both rethink how it produces and disseminates country-level data, while also creating an application environment that will offer a richer set of data than currently available today. FCPSS will provide a standardized platform more amenable to change in the number and type of data sources and in processing power and scalability.

The department has an opportunity to streamline its output and take advantage of new external data and internal data outputs such as Fishery fact sheets and the Supply and Utilization account database. The system can be configured such that thematic areas responding to different service or professional profiles such as resource and fishery inventories, management systems and statistics can be separated, allowing for a practical collaborative working environment.

2.2 Problem Statement

The problem of	difficult to manage country-level data workflows
affects	authors, researchers and managers producing country-level reports
the impact of which is	ad-hoc workflows so that reports are produced too slowly, lack important data and are often out-of-date. Traceability and quality assurance suffer as a result. Finally, the ad-hoc nature of the process means that it is hard to analyse or extract information and it is cumbersome to integrate this data into reports.
a successful solution would be	a system that supports the workflow needs for country-level reports, split into domains that express real user profiles, and that supports versioning and annotation of workflow objects. It would contain frequently updated data from a wide variety of quality data producers. It would allow users to create final reports by combining validated information into defined report templates that could be made available as PDF, HTML, XML and CSV.

2.3 Product Position Statement

For	Fisheries authors, managers and researchers
Who	produce reports containing country-level data
The FCPPS	is a Virtual Research Environment
That	manages access to multiple data sources divided by domain, including their annotation and versioning and permits production of structured text, tables, charts and graphs from these sources to be easily inserted into custom reporting templates that can be output in multiple formats.
Unlike	Microsoft Office
Our product	has a workflow and structured data that enables collaboration, annotation and versioning of the underlying data sources that make up the final product. It enforces a structured process and structured data while allowing users flexibility in actual content generation.

3 STAKEHOLDER AND USER DESCRIPTIONS

3.1 Market Demographics

The market for FCPPS covers several segments: international organisations such as FAO, regional fishery bodies such as the North East Atlantic Fisheries Commission (NEAFC), universities and centres of excellence, national ministries and scientists producing reports outside of their institutional structures as consultants.

FAO, as well as producing its own reports has long provided products that give its user market timely access to quality data that they can use in their own work. FAO has a strong reputation of producing data at global and regional level, with often comprehensive coverage not delivered by any other organisation. FCPPS can help insure FAO's continued reputation, providing a solid, extensible platform that meets the raw data access needs of its clients.

End users are expecting an ever richer set of data, from an ever wider array of sources. They expect not only to be able to access this data, but to be able to extract, analyse and output it in multiple ways. FCPPS will be an open-ended platform that delivers the rich set of functionality that our clients expect, in an environment that is robust and scaleable.

3.2 Stakeholder Summary

Name	Description	Responsibilities
Fisheries and Aquaculture Economics and Policy Division (FIE)	Umbrella division responsible for country profiles	Overall FCPPS product acceptance.
Development and Planning Service (FIEP)	Management systems	Definition, validation and exploitation of Management Systems domain and overall VRE output.
Fisheries and Aquaculture Information and Statistics Service (FIES)	Fishery and Resource Inventories and Statistics	User community leader. Definition, validation and exploitation of inventory and statistics domains.
Aquaculture Management and Conservation Service (FIMA)	Aquaculture	Definition, validation and exploitation of overall VRE output.
Partners	D4Science project partners	Ensure the system will meet requirements, disseminate project results, develop end-user training.
European Commission	D4Science project funder	Monitors the project's progress approves funding.

3.3 User Summary

Name	Description	Responsibilities	Stakeholder
Country-level scientific author	A national-level report-creating author using the FCPPS to create fishery and resource content.	Use system to download, upload, version and annotate fishery and resource inventories. Also may query and extract data.	FIES
Fishery and Resources Inventory manager	A FAO user managing the creation of country-level data by working with country-level authors to produce fishery and resources inventories.	Use system to download, upload, version and annotate fishery and resource inventories. Also may query and extract data and produce final reports.	FIES/FIEP
Management systems analyst	A FAO user generating management system content.	Use system to download, upload, version and annotate management systems. Also may query and extract data.	FIEP
Statistician	A FAO user creating statistical content.	Use system to generate and annotate statistics, and create tables, charts and graphs. Also may query and extract data.	FIES
Aquaculture manager	A FAO user contracting with country-level authors to create aquaculture-based content.	Use system to download, upload, version and annotate aquaculture content. Also may query and extract data and produce final reports.	FIMA
International organisation officer	An end-user using the FCPPS to extract needed data for support to national, regional and/or global level reports, e.g. the Investment Centre	Use system to create ad-hoc reports.	NA5 Community leader

Name	Description	Responsibilities	Stakeholder
	Division (TCI), an FAO group that undertakes agriculture and rural development project formulation.		
Non-governmental organisation	An end-user using the FCPPS to analyse country-level data in support of their organisations operational and/or policy goals.	Use system to create ad-hoc reports.	NA5 Community leader
FCPPS administrator	An administrator that adds/removes users and data sources, creates new reporting templates, and generally manages the system.	Adds/removes/maps data sources. Adds/removes users. Adds/removes report/process/query templates. Maintains system. Responds to user queries.	FIES system developer

3.4 User Environment

Country level reports involve a chain of users involved in the publication process. Managers contract with scientific authors at the national level who typically produce their reports as MS-Word documents and Excel spreadsheets. These documents are passed to managers (and back to authors several times) who send them in turn to editorial staff that streamline and standardize the content. Content is transformed from Word documents into structured formats such as XML. Other dynamic data may then be included from other analysts such as management systems and statistics. Edited content is then sent to translators to create multiple language versions. Finally the end-product is published.

The task cycle lasts for 3 to 6 months and involves at least three people directly (manager, author, editor) and several others optionally (statistician, management system analyst, translator, developer).

Major points of friction are the lack of a formal annotated workflow, easy access to versioned products and the mix of static and dynamic data creating partially out of date profiles. Much of the data needed is found in a variety of systems causing much wasted effort and a lack of process quality control. Conversion from unstructured to structured documentation is also time-consuming. Translation can be an additional problem as many translators also prefer using unstructured formats such as Microsoft Word.

3.5 Workflow

The following section describes both the current and the proposed workflows within the VRE. For a process-oriented view of the workflow as it is today please see Appendix A.

3.5.1 Current

In the current process for producing country-level reports FAO contracts with consultants in the country of interest to produce reports according to an agreed upon structure and length. The consultant uses available data and personal knowledge to create a document describing the country's fisheries sector. The consultant uses a FAO provided spreadsheet in which they inventory fisheries and aquatic resources existing or exploited by that country. Each row contains one item and each column is of a content type, such as "vessels" or "incidental species".

The returned spreadsheet is reviewed internally by several different experts before being returned to the author for corrections or clarifications. This process generally goes through several rounds until the content is accepted.

The validated spreadsheet is then converted into XML by a stand-alone configurable FAO-built conversion tool. The resultant XML is reviewed by an internal editor and usually requires some small modifications or additions.

There is additional content created by in-house experts covering the applicable management systems as well as a background overview for the country. This content is created as a Word document and typically manually pasted into the XML by an internal editor.

Finally economic and geographic statistical data is extracted from various UN and FAO systems and formatted. Queries to web services are also formulated and inserted in the XML for capture, production and fleets.

The final result is converted dynamically into HTML upon request, where additional FAO data such as catch and production time-series, legal regulations, publications and events are added to the result.

Following publication of the first version, translators are contracted and the reports translated and published in anywhere from 2 to 5 languages.

3.5.2 Proposed

In the proposed process for producing country-level reports, FAO would like to use the FCPPS as a way to centralise data services and maintain both workflow history and process annotation:

Fisheries and Resources inventory

1. Previous spreadsheets for the country are queried and extracted from VRE or if non-existent a new spreadsheet is created externally.
2. Fisheries and resources inventories are modified as needed, loaded into the VRE and automatically versioned.
3. Use existing or upload guidelines consisting of structured XML and attach to spreadsheets. These objects are also automatically versioned.
4. Annotations consisting of structured XML are attached.
5. If not already existing, a country-level user is created.

6. VRE alerts external consultant, who accesses and downloads inventories, guidelines and annotations.
7. External consultant completes inventories using information coming both from external paper and electronic sources, as well as digital library materials, time-series statistics and GIS data previously loaded in the VRE. Consultant may also refer to previous work histories and annotations.
8. External consultant uploads completed inventories and adds new annotations. Due to poor or expensive internet connectivity for some users, these annotations may be uploaded as part of a structured MS-Word document template (they would be indexed but the data would not be structured.)
9. VRE alerts FAO, who downloads completed inventories and annotations.
10. If validated go to step 12. If not validated (normally the case), FAO creates new annotations, possibly modifies the inventories and re-uploads.
11. Return to step 2.
12. Auto-convert spreadsheet to XML using FAO tool and edit result.
13. FAO adds general overview using information coming both from external paper and electronic sources, as well as digital library materials previously loaded in the VRE possibly including inventory annotations coming under this or other country profile thematic areas, e.g. statistics, management systems.
14. Go to XML integration section.

Management systems

1. Use existing or upload guidelines and a Word document template to VRE (automatically versioned).
2. VRE alerts internal expert to prepare a management system report for the specified country.
3. Internal expert prepares management system report, using information coming both from external paper and electronic sources, as well as digital library materials previously loaded in the VRE. Consultant may also refer to previous versioned work histories and annotations.
4. Internal expert uploads completed management systems document and adds new annotations.
5. VRE alerts FAO, who downloads completed document and annotations.
6. If validated go to step 8. If not validated, FAO creates new annotations, possibly modifies the document and re-uploads.
7. Return to step 2.
8. Auto-convert document to XML using FAO tool. XML stored, versioned and indexed.
9. Go to XML integration section.

Statistics

1. Use existing or upload new guidelines to VRE.
2. VRE alerts statistician to prepare statistical overview for the specified country.
3. Statistician queries time-series on catch, production, fleets, commodities, supply & utilization from FAO databases via the VRE. Also queries

economic and geographic data from UN Comtrade and UN-Data via the VRE. These queries may be saved queries from previous work, or modifications of saved queries.

4. Query output is prepared as needed using VRE outputting tools to create tables, charts and graphs.
5. Statistician attaches annotations to query output, singly and/or as a group.
6. VRE alerts FAO, who downloads completed statistics and annotations.
7. If validated go to step 9. If not validated, FAO creates new annotations.
8. Return to step 2.
9. Convert document to XML.

Integration

1. Combine Fisheries/Resources, Management System and Statistics XML using FAO XSLT.
2. Upload XML to VRE.
3. Using templating tools, assemble final report, adding general overview and other material as needed from VRE digital library.
4. Publish report (system defines a set of PURLs that can be used to access HTML, XML or PDF versions).

Translation

1. XML translated versions of reports uploaded, versioned, associated and indexed.
2. Publish report (system defines a set of PURLs that can be used to access HTML, XML or PDF versions).

3.6 Stakeholder Profiles

3.6.1 Fisheries and Aquaculture Department (FI)

Representative	Marc Taconet
Description	Represents the overall interest of the department in the product as a solution to country-level dissemination needs.
Type	Manager.
Responsibilities	Responsible to see that the system is useful to the organization, and that value added knowledge products can be generated through smooth integration of information sources provided through various workflow processes.
Success Criteria	Success is a system that creates an environment in which end-users are accessing more data more easily and producing more powerful reports more efficiently. This stakeholder will be rewarded by presiding over a system that enables country-level reports of greater timeliness, quality and comprehensiveness.
Involvement	Requirements gathering. Product Validation.
Deliverables	Requirements generation. Validation testing.

Comments / Issues	
-------------------	--

3.6.2 Fisheries and Aquaculture Information and Statistics Service (FIES)

Representative	Sachiko Tsuji
Description	The service responsible for collecting, validating and disseminating time-series statistics on aquatic species capture and production, fishery fleets and fish commodities.
Type	Head statistician.
Responsibilities	Responsible to see that time-series are correctly deployed within the system.
Success Criteria	Success is a system that allows query access to time-series databases using a variety of keys such as species, country, water area, time, etc. and displays results in a tabular format. Ideally report generation should also allow the output of tabular data in chart and graph formats.
Involvement	Requirements gathering. Product Validation.
Deliverables	Requirements generation. Validation testing.
Comments / Issues	

3.6.3 Country Profiles Manager (FIPP/FIMA)

Representative	Jan Orzesko, Valerio Crespi
Description	A manager responsible for contracting country-level reports.
Type	Managers – FIPP, FIMA.
Responsibilities	Responsible to see that the system is usable by the scientific authors with which FAO contracts.
Success Criteria	Success is a system that creates a workflow that produces more powerful reports more efficiently. This stakeholder will be rewarded by managing over a set of country-level reports of increasing quantity, timeliness, quality and comprehensiveness.
Involvement	Product Validation.
Deliverables	Requirements generation. Validation testing.
Comments / Issues	

3.6.4 Management Systems Manager (FIPP)

Representative	Cassandra de Young
Description	A manager responsible for creating management systems.
Type	Managers – FIPP.
Responsibilities	Responsible to see that the system is usable for creating

	management systems.
Success Criteria	Success is a system that creates a workflow that allows easy streaming of management systems data into country profiles. This stakeholder will be rewarded by managing over a set of country-level management systems of increasing quantity, timeliness, quality and comprehensiveness.
Involvement	Product Validation.
Deliverables	Requirements generation. Validation testing.
Comments / Issues	

3.6.5 Knowledge and Communication Department (KCEW)

Representative	Giorgio Lanzarone
Description	The service responsible for disseminating publications, news and events.
Type	Information manager.
Responsibilities	Responsible to see that publication (EIMS) and news and events (NEMS) data is correctly deployed within the system.
Success Criteria	Success is a system that allows query access to the EIMS and NEMS databases.
Involvement	Product Validation.
Deliverables	Validation testing.
Comments / Issues	

3.6.6 FAO Legal Department (LEGN)

Representative	Barbara Mauro
Description	The service responsible for disseminating national, regional and international laws and regulations concerning fisheries and aquaculture.
Type	Information manager.
Responsibilities	Responsible to see that FAOLEX data is correctly deployed within the system.
Success Criteria	Success is a system that allows query access to the FAOLEX legislation database.
Involvement	Product Validation.
Deliverables	Validation testing.
Comments / Issues	

3.6.7 D4Science Project

Representative	Yves Jaques
Description	Responsible for organisation of day-to-day activities contributing to the development of the product.
Type	Manager
Responsibilities	Responsible to see that activities are organized and carried out according to project plan.
Success Criteria	Success is a system that meets user needs. This stakeholder will be rewarded by having delivered the system.
Involvement	Requirements generation. Product Exploitation. Product Validation.
Deliverables	Requirements. Testing.
Comments / Issues	

3.7 User Profiles**3.7.1 Fisheries & resources inventory compiler**

Representative	Elena Balestri
Description	An end-user that uses the system as an aid to managing the inventories created at national level and disseminated to national-level authors for completion.
Type	Scientific, but not necessarily a computer expert – possibly even a novice.
Responsibilities	Responsible to see that the system meets the needs they have for uploading and downloading inventory spreadsheets and their annotation, as well as versioning and querying.
Success Criteria	Success is being able to upload/download inventory spreadsheets, their various versions and their attached annotations. This stakeholder will be rewarded by a coherent workflow, and easy access to workflow history.
Involvement	Requirements generation. Product Validation.
Deliverables	Ongoing requirements definition. Validation testing.
Comments / Issues	Will require active support from user's service.

3.7.2 National-level author

Representative	Elena Balestri
Description	An end-user that uses the system as an aid to creating country-level content
Type	Scientific, but not necessarily a computer expert – possibly even a novice.

Responsibilities	Responsible to see that the system meets the needs they have for uploading and downloading inventory spreadsheets and their annotation, as well as versioning and querying.
Success Criteria	Success is being able to upload/download inventory spreadsheets, their various versions and their attached annotations. This stakeholder will be rewarded by a coherent workflow, and easy access to workflow history. They will also benefit from access to VRE resources that will help them in preparing their reports.
Involvement	Requirements generation. Product Validation.
Deliverables	Ongoing requirements definition. Validation testing.
Comments / Issues	Will require active support from user's service.

3.7.3 Management system analyst

Representative	Cassandra de Young
Description	An end-user that uses the system to upload/download, annotate and version management system reports.
Type	Scientific, but not necessarily a computer expert – possibly even a novice.
Responsibilities	Responsible to see that the system meets the needs for uploading/downloading, annotating and versioning of management system reports.
Success Criteria	Success is being able to easily access the workflow for management systems products, their various versions and the annotations attached to them. This stakeholder will be rewarded by having a coherent workflow and repository for the query and storage of management system reports and their process history.
Involvement	Requirements generation. Product Validation.
Deliverables	Ongoing requirements definition. Validation testing.
Comments / Issues	Will require active support from user's service.

3.7.4 Statistician

Representative	Stefania Vanuccini
Description	An end-user that uses the system in creating statistical charts, tables and graphs coming from various underlying data systems.
Type	Scientific, but not necessarily a computer expert – possibly even a novice.
Responsibilities	Responsible to see that the system meets the needs statisticians have for query, access, extrapolation, analysis and output of statistical data for country-level reports.
Success Criteria	Success is being able to utilize a wide variety of country-level statistical data and format for reports. This stakeholder will be rewarded by easier access to multiple data streams, and benefit from the ability to generate statistical data for presentation more

	easily.
Involvement	Requirements generation. Product Validation.
Deliverables	Ongoing requirements definition. Validation testing.
Comments / Issues	Will require active support from user's service.

3.7.5 Integration end-user

Representative	Elena Irde
Description	An end-user that uses the system to integrate the various data streams into a harmonious report.
Type	Scientific, but not necessarily a computer expert – possibly even a novice.
Responsibilities	Responsible to see that the system meets the needs for integration of various XML streams and for annotation and versioning of the integrated product.
Success Criteria	Success is being able to integrate and manage the various data streams and output a finalized country profile. This stakeholder will be rewarded by producing reports with less friction than is possible today.
Involvement	Requirements generation. Product Validation.
Deliverables	Ongoing requirements definition. Validation testing.
Comments / Issues	Will require active support from user's service.

3.7.6 System administrator

Representative	Francesco Calderini
Description	An administrator that sets up the system, adds/removes users and data sources, defines report types, fields end-user queries.
Type	An information system developer with average skills.
Responsibilities	Responsible to see that system administration is simple and not very time-consuming.
Success Criteria	Success is a system that can be administered to end-user satisfaction with a minimum of active involvement. It should be possible to set it up within a few days, and thereafter it should only require an hour here and there. This stakeholder will be rewarded by administering a system that users find indispensable to producing quality country-level reports.
Involvement	Product Validation.
Deliverables	Validation testing.
Comments / Issues	

3.8 Key Stakeholder or User Needs

- **Many data sources difficult to access**
 - *What are the reasons for this problem?*
 - Spread out amongst web-sites, spreadsheets, applications and databases
 - *How is it solved now?*
 - Users just hack around until they get more or less what they want
 - *What solutions does the stakeholder or user want?*
 - A one-stop shop with unified, integrated access in which they can easily find what they want and format it to their needs.
- **Ad-hoc workflow**
 - *What are the reasons for this problem?*
 - Spreadsheets and email are used, making it difficult to enforce a particular workflow and to review process history.
 - *How is it solved now?*
 - Users maintain emails in folders and generally a kind of compendium of annotations is produced and saved alongside the spreadsheet output.
 - *What solutions does the stakeholder or user want?*
 - Users want more formally defined processes together with query and versioning mechanisms that make it easy to find and trace workflow history and version history.
- **Ad-hoc annotation**
 - *What are the reasons for this problem?*
 - Email is used to annotate work. No management system is used.
 - *How is it solved now?*
 - Email.
 - *What solutions does the stakeholder or user want?*
 - A structured annotation format to create more coherence across the workflow with national-level authors.
- **Ad-hoc versioning**
 - *What are the reasons for this problem?*
 - Versions are merely held as separate files in a directory. No versioning system is used.
 - *How is it solved now?*
 - Store different versions on a network drive.
 - *What solutions does the stakeholder or user want?*
 - Formal versioning together with query functions so users can easily access and compare various versions.
- **Difficult to output data sources in desired formats**
 - *What are the reasons for this problem?*
 - Underlying data comes from many sources, so it's not so clear how to format it correctly.
 - *How is it solved now?*
 - Users do a lot of cutting and pasting and re-formatting with various products like excel.
 - *What solutions does the stakeholder or user want?*

- Users are generally following defined processes that they do over and over. They would like to be able to define these processes once and then have them ready to format the data they extract.
- **Data trapped in content, difficult to update**
 - *What are the reasons for this problem?*
 - Reports are often a mixture of authored content and extracted data.
 - *How is it solved now?*
 - It generally isn't, so the reports have a mixture of data that is obsolete and data that is not.
 - *What solutions does the stakeholder or user want?*
 - Defined reporting formats that could re-generate reports as needed could insure that dynamic data remains in a dynamic island amidst a sea of authored content.
- **Data variety not as wide as hoped**
 - *What are the reasons for this problem?*
 - too many systems are too hard to access.
 - query mechanisms vary greatly across underlying systems.
 - basketing/bookmarking systems vary greatly across systems.
 - there is no way to save standard processes, making the data variety somewhat ad-hoc.
 - *How is it solved now?*
 - it isn't.
 - *What solutions does the stakeholder or user want?*
 - a one-stop shop.
 - aggregations/extractions by regions.
- **Overall output is often out of date**
 - *What are the reasons for this problem?*
 - It's too expensive and time-consuming to regularly update reports for all countries.
 - *How is it solved now?*
 - It is not.
 - *What solutions does the stakeholder or user want?*
 - Systems that better separate authored content from dynamic data so that reports can be more easily kept up-to-date.
 - Systems which can help distinguish multiple authors and updates for the same country.
 - Systems that make it easier to produce and reuse data.
- **Content conversion too much effort**
 - *What are the reasons for this problem?*
 - Authored content is in unstructured Word documents.
 - *How is it solved now?*
 - Converted semi-automatically to XML using VB scripts.
 - *What solutions does the stakeholder or user want?*
 - Ideally, a system in which authored content is structured from the very beginning, including when translated.

- Standard templates that produce a common look-and-feel for all reports.

Need	Priority	Concerns	Current Solution	Proposed Solutions
access multiple data sources	High	data system API's	ad-hoc	web services, relational databases or CSV files access
annotation	High	should be possible to relate across annotations	not done	Diligent
control access to data	Medium	none	ad-hoc	User authorisationauthorization
create multiple output data types	Medium	none	ad-hoc	Use Diligent output tools
create reports	Medium	Current system seems far from perfect	ad-hoc	Use Diligent output tools
define process workflows	Medium	Will it be possible to make it user friendly?	not done	Create process templating mechanism
define report types	Medium	Will it be possible to make it user friendly?	Word templates and XML/XSL to HTML.	Extend Diligent templating tools and/or existing XML/XSL.
define various levels of user access to data	High	none	not done	User authorisationauthorization
extrapolate (infer) data	High	could be difficult to implement generically	ad-hoc	Extend current query tools and/or leverage existing analysis libraries
harmonize data	Medium	How to facilitate mapping.	ad-hoc	brought together under a common schema by local to global mapping

manipulate data	Medium	could be difficult to implement generically	ad-hoc	Leverage existing visualisationvisualization libraries
output data in various formats/visualisationsvisualizations for reporting	High	need to see if there are libraries that can be leveraged	ad-hoc	Leverage existing reporting tools such as BIRT.
query over heterogeneous data	High	meaningfulness of response	not done	Diligent indexing tools
versioning	High	Will it be possible to compare?	not done	Diligent storage
view heterogeneous data	High	usability tricky	time-series presented as tables, graphs and charts using bespoke services	Leverage existing visualisationvisualization libraries

3.9 Alternatives and Competition

Maintaining the status quo is one possibility, although upper management has requested the department look into alternatives.

Moving all country profiles production into a sophisticated content management / structured documentation system (most CMSs deal with content as formatting issue, not semantics) would be a partial solution to the problem. It would reduce friction but not solve fundamental problems of heterogeneous data access. Some of these CMSs could do a lot to answer the workflow issues (such as the [BSCW](#) system used by D4Science project) however.

Other systems that aggregate data sources and provide reporting functionalities could be considered such as Crystal reports, SAS or SAP. They are however typically quite expensive and subject to serious licensing restrictions.

4 PRODUCT OVERVIEW

4.1 Product Perspective

4.1.1 Output

There are currently three online country-level data products that are indexed at the following URL's:

- **Fisheries Country Profiles fact sheets**
(<http://www.fao.org/fishery/countryprofiles/search>)
 - *Thematic information on country level fisheries sector*
 - *Provided in XML and HTML format.*
- **National Aquaculture Sector Overview fact sheets**
(<http://www.fao.org/fishery/naso/search>)
 - *Thematic information on country level aquaculture sector*
 - *Provided in XML and HTML format.*
- **National Aquaculture Legislation Overview fact sheets**
(<http://www.fao.org/fishery/nalo/search>)
 - *Thematic information on country level aquaculture legislation*
 - *Provided in XML and HTML format.*

4.1.2 Data

Data storage will be handled by Diligent/EGEE grid services. The system will also interact with external data providers via web services, JDBC and/or CSV file upload. The list below includes sources currently in use together with several sources that need to be accessed for a smooth workflow.

Current data providers to Fisheries Country Profiles are:

Fisheries and Resources Inventories

- **Fisheries inventories**
 - *Produced as spreadsheets*
 - *Work is ongoing to produce them according to an agreed XML schema model.*
- **Resources fact sheets** (<http://firms.fao.org/firms/inventory/browse>)
 - *Thematic information on various fisheries domains*
 - *Provided in XML and HTML format.*
- **Introduced Species fact sheets** (<http://www.fao.org/fishery/introsp/search>)
 - *Thematic information on introduced species*
 - *Provided in XML and HTML format.*
- **Reference tables management system (RTMS)**
(<http://www.fao.org/fishery/rtms>)
 - *Classification systems, keyword lists, etc.*

Statistics

- **Online query panels** (<http://www.fao.org/fishery/topic/16140>)
 - *Provide species catch and production statistics, as well as statistics on commodities, fleets and supply & utilization in tabular format.*
 - *Produced in XML, HTML and CSV.*
- Other systems are consulted but not in a consistent manner, suggested:
 - **UN Comtrade** (<http://comtrade.un.org/>)
 - Trade data
 - **UN Data** (<http://data.un.org/>)
 - A cross-organisation initiative to provide UN data for multiple areas.
 - **Organisation for Economic Co-operation and Development (OECD)** (<http://stats.oecd.org/oecdstatws/datasetbrowser.asmx>)
 - Economic data

Management Systems

- **Fishery institutions** (Currently only available internally via JDBC)
 - Basic information on institutions, their membership, scope and mission.
- No other underlying provider, text contained in XML fact sheets listed in *output* section above.

General data

- **FAO Corporate document repository** (<http://www.fao.org/documents/>)
 - *FAO official publications*
 - *Abstracts in Dublin Core / AgMES XML*
- **FAOLex** (<http://faolex.fao.org/faolex/>)
 - *Provides texts of legal documents relating to fisheries*
 - *Is in a UN-format ISIS* (<http://unesco.org/isis>) *database*
- **News and Events Management System (NEMS)** (<http://www.fao.org/nems/>)
 - *Provides fisheries news and events information*
 - *Is an RSS format*
 - *Provide species catch and production statistics, as well as statistics on commodities and fleets in tabular format.*
 - *Produced in XML, HTML and CSV.*
- Other systems are consulted but not in a consistent manner, suggested:
 - **Back-to-office reports**
 - Contain results of country-level missions in Word format according to a semi-standard template.

GIS data

- Currently not exploited in the fact sheets but would be useful, suggested:
 - **European Space Agency** (<http://www.esa.int/esaEO/index.html>)

4.1.3 Services

The FCPPS will exist within the grid services provided by the Diligent/EGEE network. Some basic tools that would greatly assist in the creation of country level reports are:

- **The R Project for Statistical Computing** (<http://www.r-project.org/>)
 - R is a language and environment for statistical computing and graphics. It is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.
- **BIRT** (<http://www.eclipse.org/birt/phoenix/>)
 - BIRT is an open source configurable Eclipse-based (for configuration) reporting and charting system.

4.1.4 Interface

Users will experience it as a browser-based application, and all its services will be available from within a standard web browser.

4.2 Summary of Capabilities

Customer Benefit	Supporting Features
Analysis of heterogeneous data	Process definition template mechanisms allow analysis workflows to be created.
Annotation of collections.	Objects in the VRE can have annotations attached.
Better traceability of process/workflow.	Actions performed on data can be stored and viewed in the future.
Better traceability of processReport definition.	Actions performed on data can be stored and viewed in the futureStandard report templates.
Data output in multiple formats	Data can be output as charts, graphs, tables and text.
Distributed workflow.	Multiple users can collaborate on the same products wherever they are connected to the web.
Easy to get data for many data sources	Grid services support access, definition and aggregation of multiple data sources.
Harmonization of data	Mapping features allow local to global schema mapping.
Manipulation of data	Advanced queries allow data subsets to be re-shaped.
Process definition.	Routinely performed actions can be automated/chained.

Query over heterogeneous data	Queries are built over multiple data sources allowing one-stop queries.
Table, chart and graph definition from statistics	Tools can take statistical data and produce user-friendly outputs.
Versioning.	Objects can contain multiple versions.

4.3 Assumptions and Dependencies

- Access to data sources is a strong dependency. Extension and/or modification to existing services may be necessary in order for the FCPPS to fully benefit from a data source. The state of a system's data services and the ability to upgrade them could impact the number of data sources added to the system.
- Tool integration may be difficult depending on the underlying operating system it is meant to run on.
- Internet connectivity is an assumption, something that for some countries is not so reliable, fast and/or is expensive. For these users some activities may be best performed offline such as preparing an annotation that can then be uploaded.

4.4 Cost and Pricing

Costs must be contained within the constraints of the D4Science budget.

4.5 Licensing and Installation

The system should be licensed as GPL or similar to not place constraints on its use by the community.

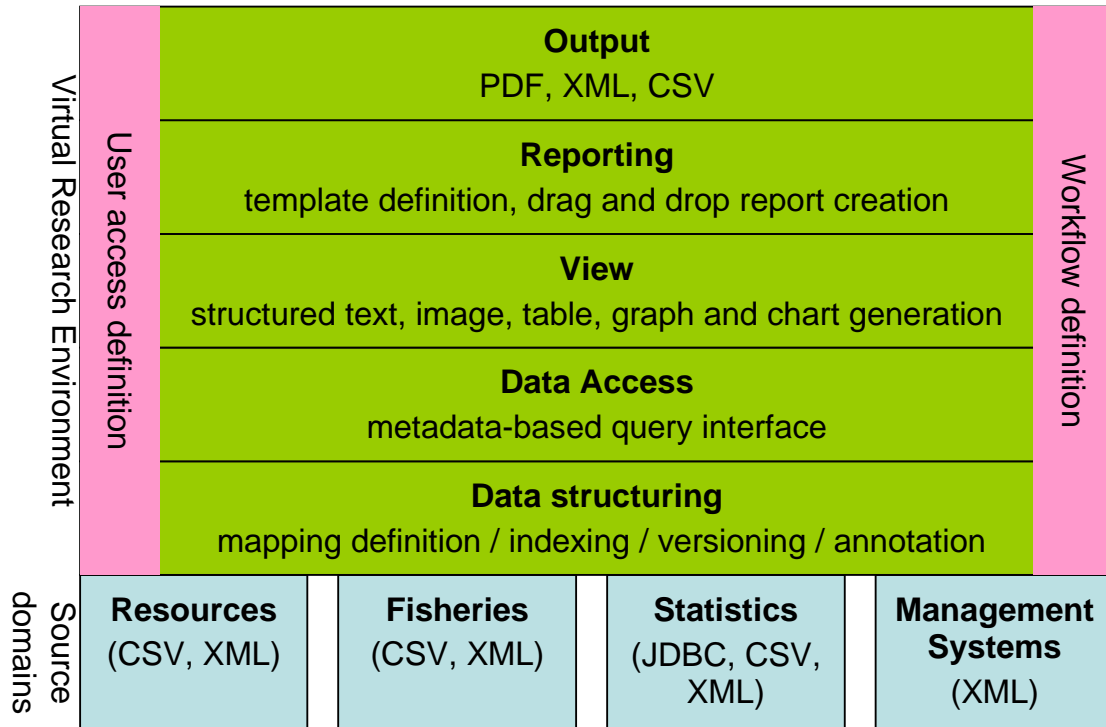
Users should not have to have any special installation on their computers to access the FCPPS. It should function within a browser.

The system will need to be installed on a grid-hosted virtual organisation (VO) but it should not be necessary for a community to hose a gCube node in order to define and create an FCPPS.

Security is an issue that should be considered as an overhead. An administrator will need to define user rights and roles.

5 PRODUCT FEATURES

Product features are divided into a set of functional areas. Viewing the system as a set of functional layers:



Conceptual view of VRE functional layers

5.1 User access definition

5.1.1 Define user groups

The administrator shall be able to create ad-hoc group types with defined access to system data, services and outputs. Ideally, should be able to use the token-based FAO authentication system for FAO users.

5.1.2 Assign users to groups

The administrator shall be able to add users to groups.

5.1.3 Assign data to groups

The administrator shall be able to assign data sources to groups.

5.2 Data

5.2.1 Access web service heterogeneous data sources

The system shall be able to connect to multiple web service API's.

5.2.2 Access JDBC heterogeneous data sources

The system shall be able to connect via JDBC to multiple databases.

5.2.3 Harmonize multiple data sources

The system shall be able to aggregate multiple data sources. Users shall be able to define and/or import global schemas to which multiple local schemas may be mapped.

5.2.4 Define metadata for multiple heterogeneous data sources

The system shall allow users to define metadata for multiple data sources, preferably using OWL, RDF or XSD.

5.2.5 Index multiple heterogeneous data sources

The system shall be able to index multiple heterogeneous data sets. These indexes shall be periodically updated.

Annotations shall also be indexed

5.2.6 Annotate data sources

The system shall be able to annotate data sources. It shall be possible to relate annotations to one another.

5.2.7 Version resources

The system shall be able to maintain multiple associated versions of VRE resources.

5.3 Data access

5.3.1 Query over multiple heterogeneous data sources

The system shall be able to query over multiple heterogeneous data sets. Query interfaces and query results should be displayed using the metadata of source data.

5.3.2 Define custom query templates

Users shall be able to create templates defining new query combinations and custom fields based on underlying metadata of data sources. *Lower priority.*

5.3.3 Manipulate data source results

Users shall be able to manipulate data results by re-querying, restricting and filtering results. *Lower priority.*

5.4 View

5.4.1 View data source metadata

Users shall be able to view the source metadata extracted from data, and also compare metadata across data sources and data sets.

5.4.2 View data source text

Users shall be able to view the text content of data sources.

5.4.3 View data source images

Users shall be able to view images from data sources.

5.4.4 View data source tables

Users shall be able to view tabular data such as time-series.

5.4.5 View process history

Users shall be able to view the history of the selections and choices that have been made in arriving at a result set. *Lower priority.*

5.5 Workflow definition

5.5.1 Create process templates

Users shall be able create templates that define and execute common processes performed within the VRE. *Lower priority.*

5.5.2 Workflow templates

Users shall be able create workflow templates that define the methods by which common multiple user interactions take place (and hence the resources/choices that are seen). *Lower priority.*

5.6 Reporting definition

5.6.1 Create report templates

Users shall be able to create standard report templates that can be used to semi-automatically create reports with pre-made structures and pre-defined data source extractions. *Lower priority.*

5.6.2 Create multiple report formats

Users shall be able to create multiple format reports, such that one report template can output simultaneously a PDF, HTML and XML version. *Lower priority.*

5.7 Output

For all output it should be possible to output following the structured data defined in the Query section.

5.7.1 Output data source text

Users shall be able to output text extractions from data sources that can be easily put into report templates or exported.

5.7.2 Output data source images

Users shall be able to output images extracted from data sources or processed from data sources that can be easily put into report templates or exported.

5.7.3 Output data source tables

Users shall be able to output tabular data extracted or processed from data sources that can be easily put into report templates or exported.

5.7.4 Output data source charts

Users shall be able to output charts processed from data sources such as tabular data that can be easily put into report templates or exported.

5.7.5 Output data source graphs

Users shall be able to output graphs processed from data sources such as tabular data that can be easily put into report templates or exported.

5.7.6 Author reports

Users shall be able to author content directly into report templates.

6 DOCUMENTATION REQUIREMENTS

6.1 User Manual

Two types of user manual should be created; one for system administrators and a second for end-users, i.e. report creators. The end-user manual would be preferably tutorial-based as the application type is new to many users, and a tutorial helps users understand the actual limits of the system.

A system administrator manual should illustrate how to add nodes to the VRE, add/remove user groups and users, add/remove data sources, define output document templates, and define custom process templates.

6.2 Online Help

The system should provide at least on-line help for the end-user. This could in an initial stage as simple as point users to the correct portion of the user manual by hyper-linking to the relevant section.

