D3.10 – QualiExplore for Data Quality Factor Knowledge v2

WP3 – BUILD: Manufacturing Data Quality

i4Q has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958205.

Document Information

GRANT AGREEMENT NUMBER	958205	ACRONYM	i4Q	
FULL TITLE	Industrial Data Services for Quality Control in Smart Manufacturing			
START DATE	01-01-2021	DURATION	36 months	
PROJECT URL	https://www.i4q-project.eu/			
DELIVERABLE	D3.10 – QualiExplore for Data Quality Factor Knowledge v2			
WORK PACKAGE	WP3 – BUILD: Manufacturing Data Quality			
DATE OF DELIVERY	CONTRACTUAL	31-Dec-2022	ACTUAL	30-Dec-2022
NATURE	Report	DISSEMINATION LEVEL	Public	
LEAD BENEFICIARY	BIBA			
RESPONSIBLE AUTHOR	Stefan Wellsandt			
CONTRIBUTIONS FROM	-			
TARGET AUDIENCE	1) i4Q Project partners; 2) industrial community; 3) other H2020 funded projects; 4) scientific community 5) manufacturing stakeholders			
DELIVERABLE CONTEXT/ DEPENDENCIES	This document is the second result of T3.1. Its relationship to other documents is as follows: • D3.9 Data Quality Guidelines: Mentions this deliverable as a data quality management measure (M24) • D3.1 QualiExplore for Data Quality Factor Knowledge: first version of this deliverable (M18)			
EXTERNAL ANNEXES/ SUPPORTING DOCUMENTS	None			
READING NOTES	None			
ABSTRACT	This deliverable is one of the two results of Task 3.1, "Manufacturing Data Quality Strategy". It presents i4Q QualiExplore, a web-based software tool for visualizing information quality characteristics and factors. Producers need to know the latter to assess their relevance for the specific use case and identify measures to manage them. This i4Q tool is an example of a measure to raise awareness of data quality. D3.9 outlines these measures in its proposed activity framework. QualiExplore is a standalone application using a permissive open source license (Apache 2.0), so stakeholders can easily exploit it.			

Document History

VERSION	ISSUE DATE	STAGE	DESCRIPTION	CONTRIBUTOR
0.1	05-Nov-2022	ToC	Create ToC and Section Description	BIBA
0.2	25-Dec-2022	1st Draft	1st Draft available for review	BIBA
0.3	06-Dec-2022	Internal Review	Internal Review Process	ENG, WHI
0.4	09-Dec-2022	2nd Draft	Process comments from internal review	BIBA
1.0	30-Dec-2022	Final Draft	Final quality check and issue of final document	CERTH

Disclaimer

Copyright message

TABLE OF CONTENTS

Executive summary **6**
Document structure **7**
1. General Description **8**
1.1 Overview 8
1.2 Features 8
1.2.1 Data quality factors and filter functions 9
1.2.2 Editing environment 11
1.2.3 Knowledge base 14
1.2.4 Natural language interface 14
2. Technical Specifications **18**
3. Implementation Status **20**
3.1 Current implementation 20
3.2 Next developments 21
3.3 History 21
4. Conclusions **22**
References **23**
Appendix I **24**

LIST OF TABLES

Table 1. Feature overview and version 8
Table 2. i4Q^{QE} requirements 21
Table 3. Release Dates 21
Table 4. History 21

LIST OF FIGURES

Figure 1. Tree structure to present data quality factors 9
Figure 2. Filter categories and filter statements 10
Figure 3. Login screen 11
Figure 4. Admin view of the filter page 11
Figure 5. Admin view of the tree page 12
Figure 6. Editing filters categories and statements 12
Figure 7. Editing the tree structure 13

Figure 8. Editing filter descriptions and filter statement assignments 13
Figure 9. User opens a conversation with chatbot 14
Figure 10. User asking for the assistant's capabilities 15
Figure 11. User asking for information about data privacy 16
Figure 12. Dialogue to access quality factor knowledge 17
Figure 13. i4Q^{QE} architecture 19

ABBREVIATIONS/ACRONYMS

AD	Analytics Dashboard
AI	Artificial Intelligence
API	Application Programming Interface
BDA	Big Data Analytics
DA	Data Analytics
DB	DataBase
GDPR	General Data Protection Regulation
GraphQL	Name of a query language for APIs
IP	Intellectual Property
ISO	International Organisation for Standardisation
JSON	JavaScript Object Notation
MongoDB	Name of a document database
Neo4J	Name of a graph database
PDF	Portable Document Format
QE	QualiExplore

Executive summary

This deliverable is one of the two results of Task 3.1, "Manufacturing Data Quality Strategy". It presents i4Q QualiExplore (i4Q^{QE}), a web-based software tool for visualizing information quality characteristics and factors. Producers need to know the latter to assess their relevance and identify measures to manage them. i4Q^{QE} is an example of a measure to raise awareness of data quality. D3.9 outlines these and other measures in its proposed activity framework. QualiExplore's main application is standalone using a permissive open-source license (Apache 2.0), so stakeholders can easily exploit it.

i4Q^{QE} has a *2-staged user interface* to grant users access to factor knowledge. The first stage serves as a filter because a high number of factors can cause information overload for users. Relevant *filter categories* include the user's goals, quality (information characteristics), and channels/sources. The goals include the perspective of the information user and the information creator/author. This approach is helpful because it emphasizes that many measures that avoid quality problems require both parties' involvement. Each category has several statements representing the user's interest in information quality problems and related factors. The indicated *factor categories* structure the factors and provide a link between statements and factors.

i4Q^{QE} has nine requirements. Three are fulfilled, one is partially fulfilled, two are not fulfilled yet, and three are not relevant or not in the tool's scope. Current i4Q^{QE} features include an editing environment for filters and factor knowledge, a database to store the knowledge, and the implementation of a chatbot widget. QualiExplore's second version integrates the chatbot's backend with the Rasa Open Source framework, a graph database with factor knowledge in Neo4J, and content updates. This bot is not open source and is not required to use i4Q^{QE}. A final version of the entire solution will include bug fixes and content updates. This document i4Q D3.10 v2 is an update of v1 of D3.2., for this reason it contains information of the 1st version together with the updates developed in this 2nd version.

Document structure

Section 1: Contains a general description of i4Q **QualiExplore**, providing an overview and a list of features. It is addressed to the final users of the i4Q Solution. This section provides content for the web documentation, which can be accessed online at: http://i4q.upv.es/2_i4Q_QE/index.html

Section 2: Contains the technical specifications of the i4Q **QualiExplore**, providing an overview and its architecture diagram. It is addressed to software developers.

Section 3: Details the implementation status of the i4Q **QualiExplore**, explaining the current status, next steps, and summarizing the implementation history.

Section 4: Provides the conclusions.

APPENDIX I: Provides the PDF version of the i4Q **QualiExplore** web documentation, which can be accessed online at: http://i4q.upv.es/2_i4Q_QE/index.html

1. General Description

1.1 Overview

i4Q QualiExplore is a web-based software tool for visualizing information quality characteristics and factors.[1] Producers need to know the latter to assess their relevance and identify measures to manage them. i4Q^{QE} has a *2-staged user interface* to grant users access to factor knowledge. The first stage serves as a filter because a high number of factors can cause information overload for users. Relevant *filter categories* include the user's goals, quality (information characteristics), and channels/sources. The goals include the perspective of the information user and the information creator/author. This approach is helpful because it emphasizes that many measures that avoid quality problems require both parties' involvement. Each category has several statements representing the user's interest in information quality problems and related factors. The indicated *factor categories* structure the factors and provide a link between statements and factors.

i4Q^{QE} is an example of a measure to raise awareness of data quality. D3.10 outlines these measures in its proposed activity framework. i4Q^{QE} is a standalone application and does not depend on other i4Q solutions. It uses a permissive open-source license (Apache 2.0), so stakeholders can easily exploit it.

1.2 Features

This section summarizes i4Q^{QE} features. It covers the implemented features of version one and the planned ones for version two, as summarized in **Table 1**.

Features	Version 1 (D3.2)	Version 2 (D3.10)
Data quality factors and filter functions	X	
Editing environment	X	
Knowledge base	X (MongoDB)	X (Neo4J)
Natural language interface		X

Table 1. Feature overview and version

i4Q^{QE} is a web-based software tool for the visualisation of information quality characteristics and quality factors using the Evolutional Data Quality Concept and a Data Life Cycle.

[1] The tool's first version resulted from the Horizon 2020 project NIMBLE (723810): https://www.nimble-project.org/wp-content/uploads/2019/11/NIMBLE-D6.4_Quality-Management-Handbook-1.1.pdf

1.2.1 Data quality factors and filter functions

i4Q[QE] uses the Evolutionary Data Quality concept outlined in D3.1 to present data quality characteristics and factors. Users can interact with this information to identify relevant quality factors for their use case. i4Q[QE] content has a tree shape, as illustrated in Figure 1.

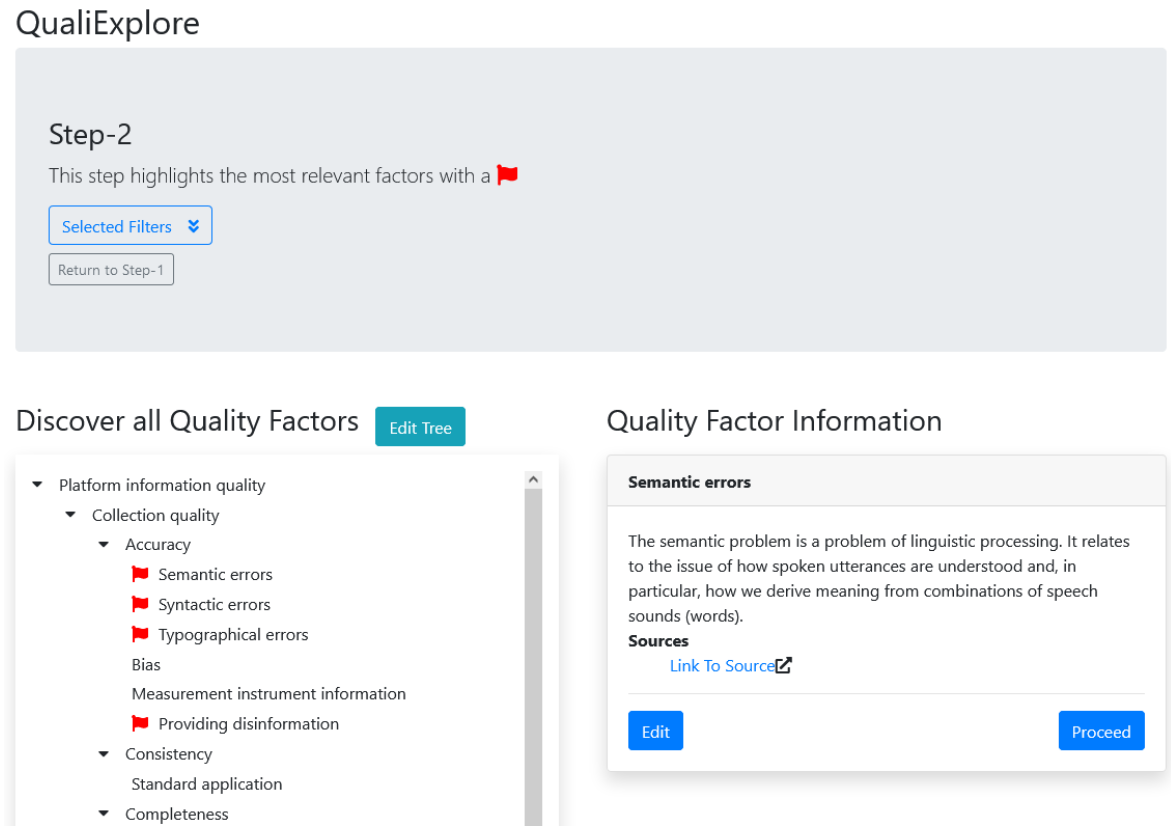

Figure 1. Tree structure to present data quality factors

The tree can become quite complex and hard to comprehend for users. Before users interact with the tree, they can describe their interest in production data quality. This approach helps highlight the most relevant branches and leaves in the tree.

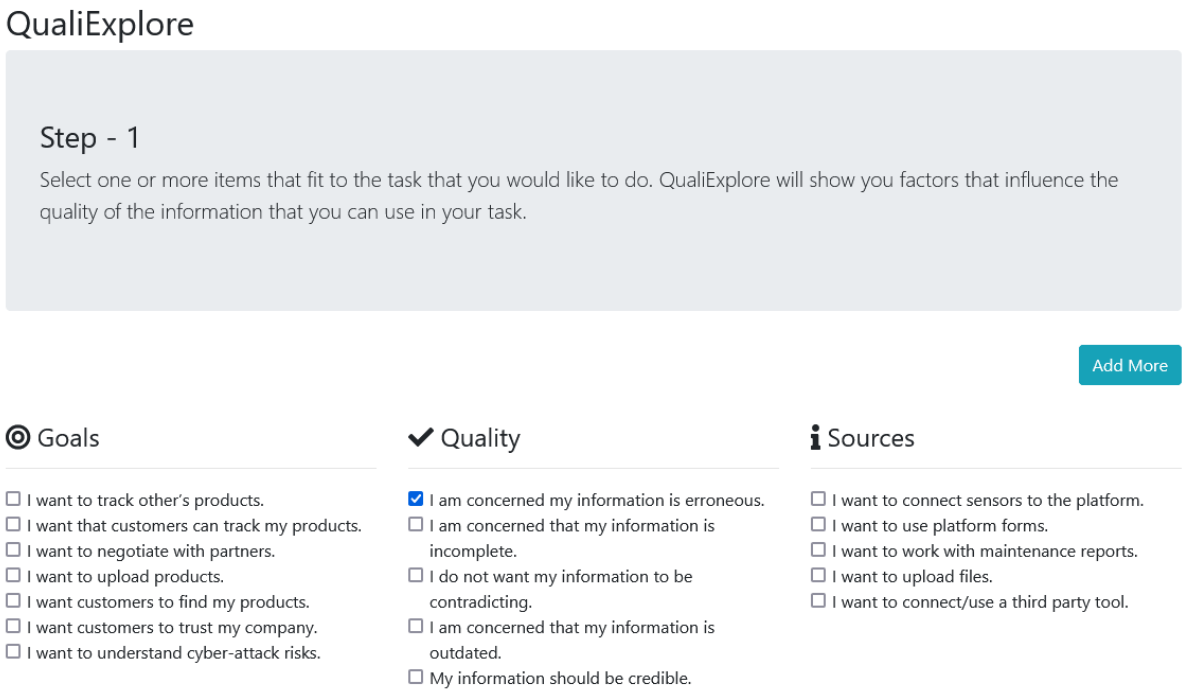

Figure 2 presents the filter website in i4Q[QE].

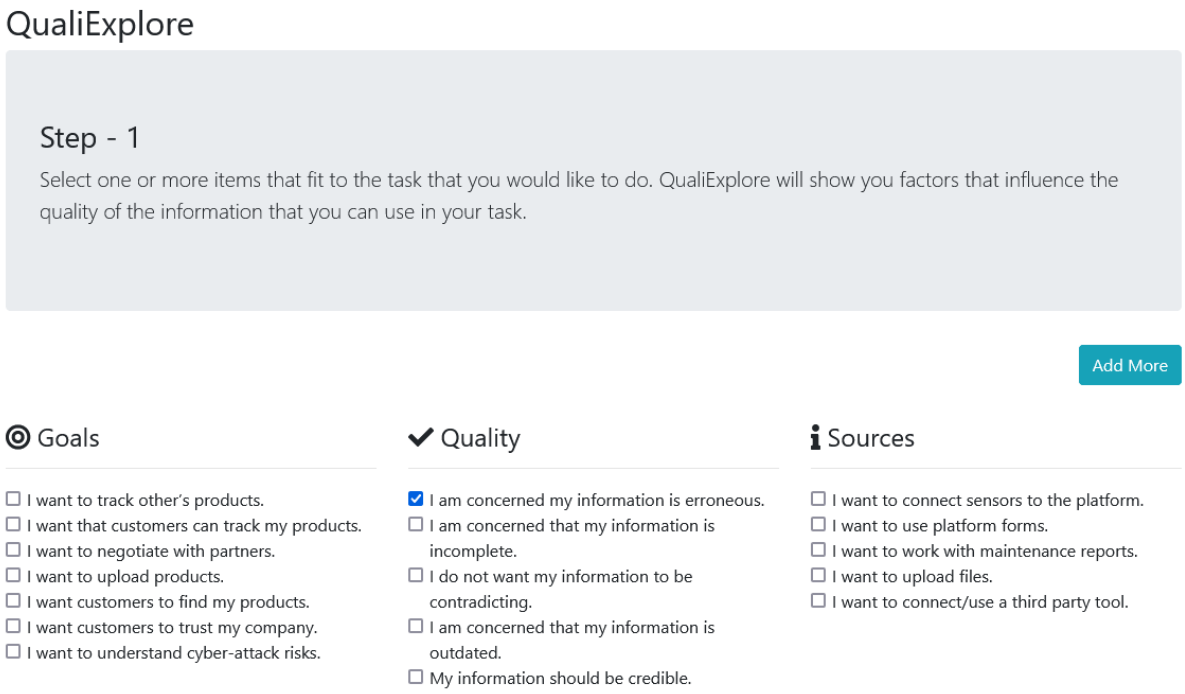

Figure 2. Filter categories and filter statements

The user's goals, quality (information characteristics), and channels/sources are relevant filter categories. Goals include the information user's and creator's perspectives. This approach is helpful because it emphasizes that many measures that avoid quality problems require the involvement of both parties.

A filter is a statement about production data and belongs to a category. Data quality factors can belong to one or more statements. The statements filter results through an "OR" logic, i.e., the more filters the user selects, the more factors i4Q[QE] will highlight.

1.2.2 Editing environment

QualiExplore's original version allows developers to change filters, tree structure, and factors by modifying the related files. i4Q[QE] simplifies the process by introducing an editing environment for authorized users. It assumes that authorized users can log in and access the editing environment. Figure 3 illustrates the login screen. Visitors can freely browse the filters and factor knowledge.

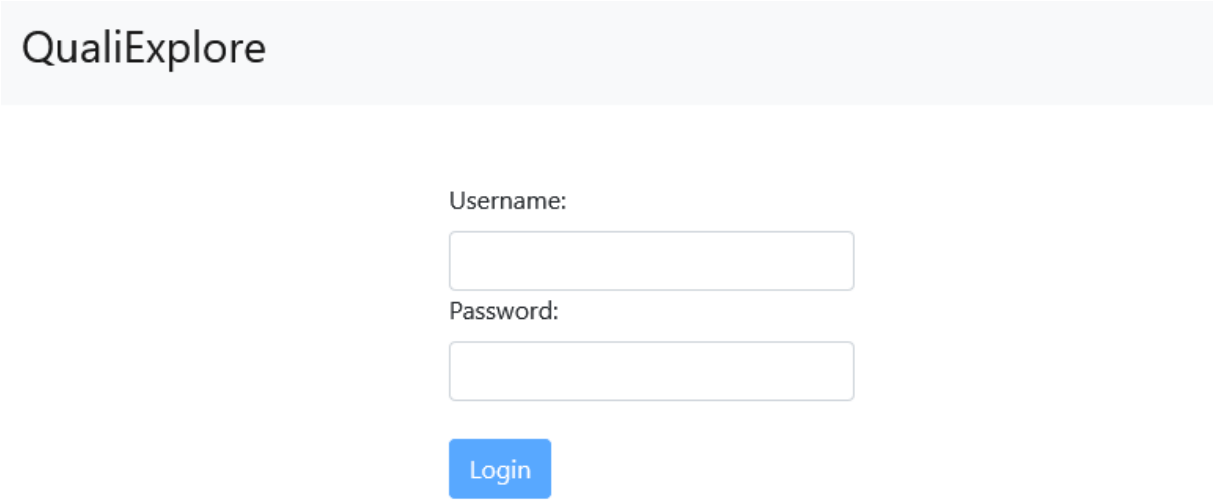

Figure 3. Login screen

If an admin user logs in, they will see editing buttons on i4Q[QE] filter and factor sites. Figure 4 and Figure 5 present the admin views for both sites.

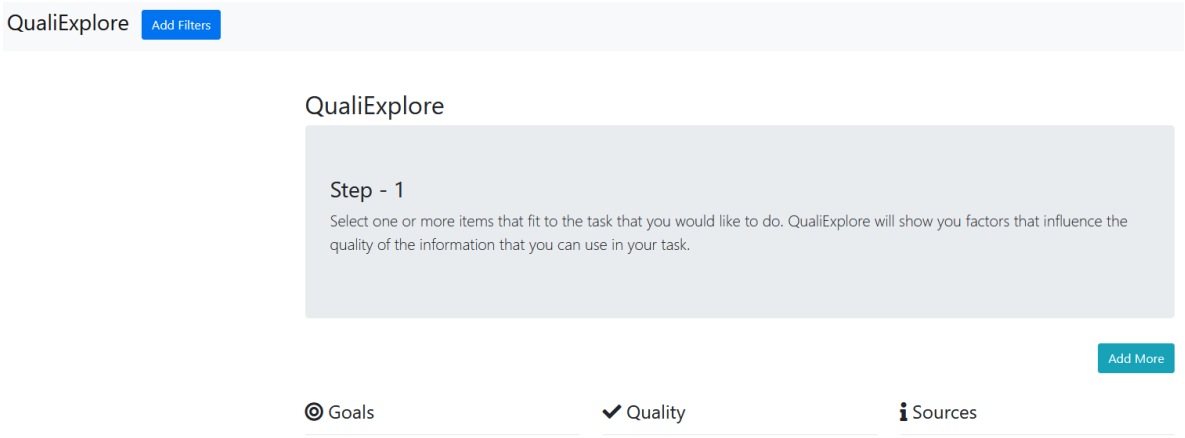

Figure 4. Admin view of the filter page

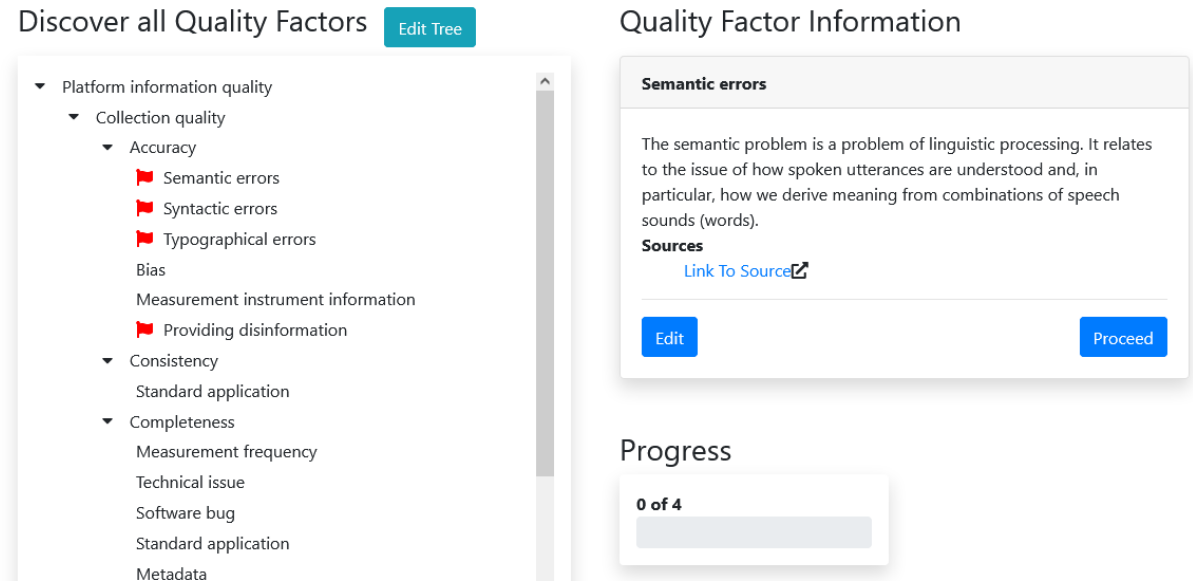

Figure 5. Admin view of the tree page

Figure 6, Figure 7, and Figure 8 present the actual editing environment to modify filter categories, filter statements, tree structure, and factor descriptions and their relations to filter statements.

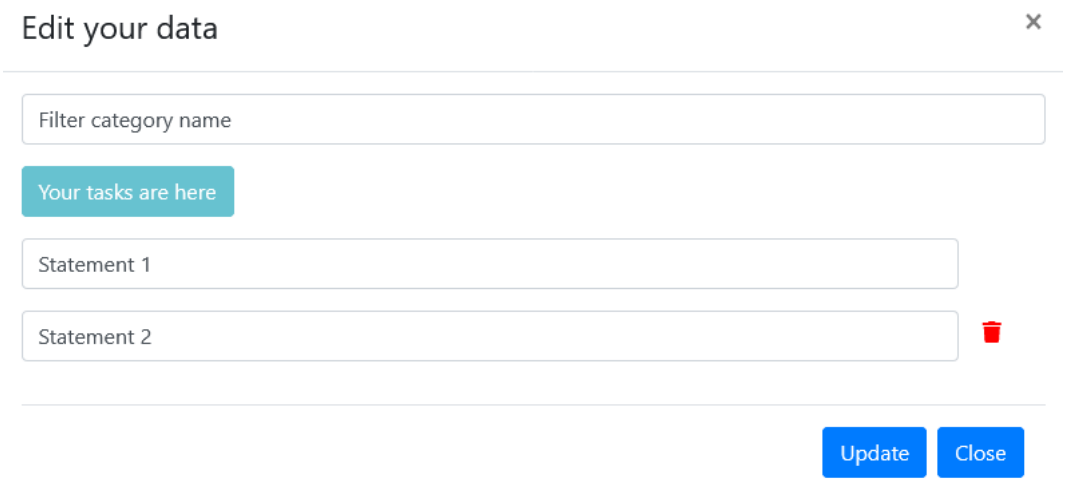

Figure 6. Editing filters categories and statements

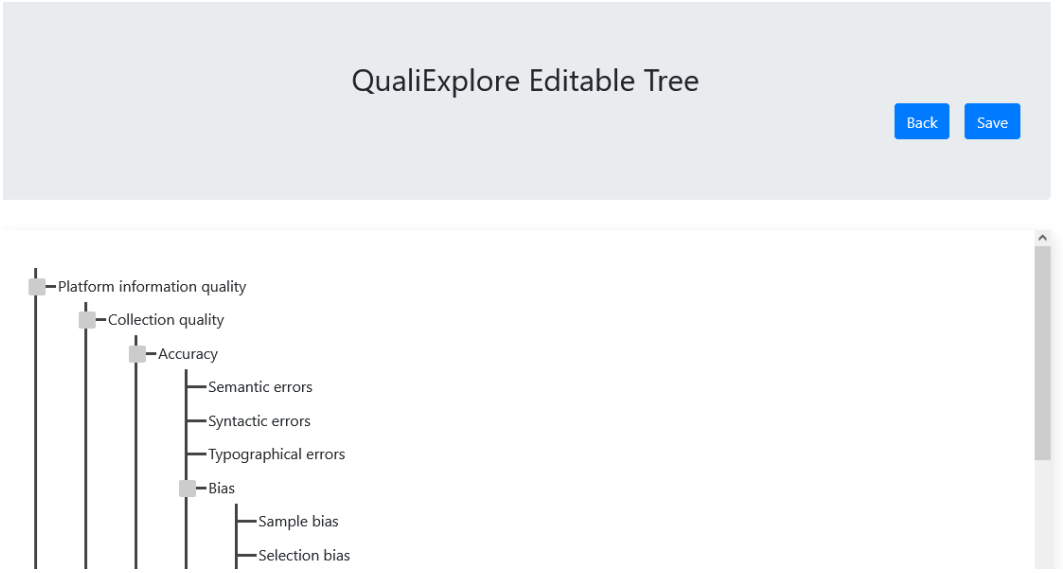

Figure 7. Editing the tree structure

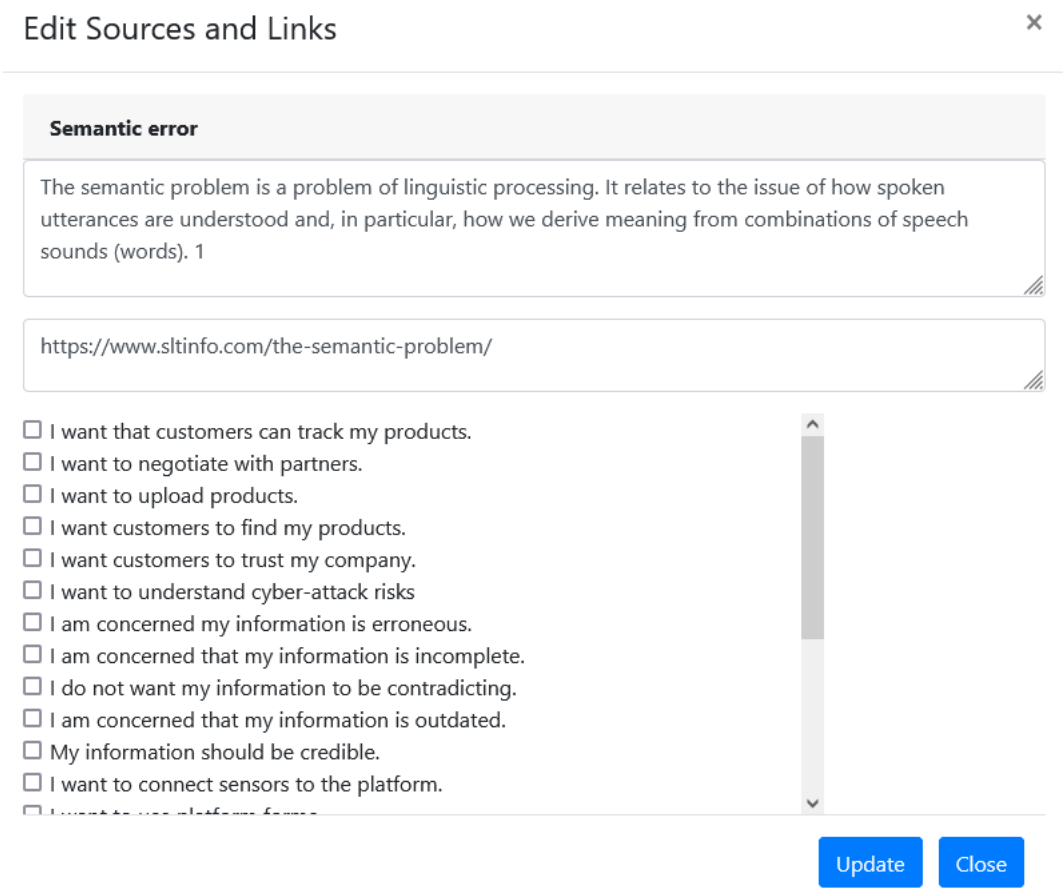

Figure 8. Editing filter descriptions and filter statement assignments

1.2.3 Knowledge base

i4Q^QE covers knowledge about data quality factors. For the user, a factor has a description, a source/reference for it, and a position in the tree-structure. Besides, factors belong to specific filter statements.

QualiExplore's original version stores the factor knowledge in files (JSON format). i4Q changes this structure to a database making it more flexible to extend and modify. A document database stored the knowledge as an intermediary solution in this deliverable. This version uses a graph to represent factors and filters. This change simplifies further improvements, such as presenting relations between factors and filter statements. Users will not see the difference in the front-end.

1.2.4 Natural language interface

The most innovative part of the QualiExplore improvements in i4Q is a natural language interface to interact with quality factor knowledge. This feature grounds on a chatbot that users can access while using i4Q^QE. It accesses the knowledge base, and users can ask about production data quality factors. Figure 9 illustrates the first step of the user interaction.

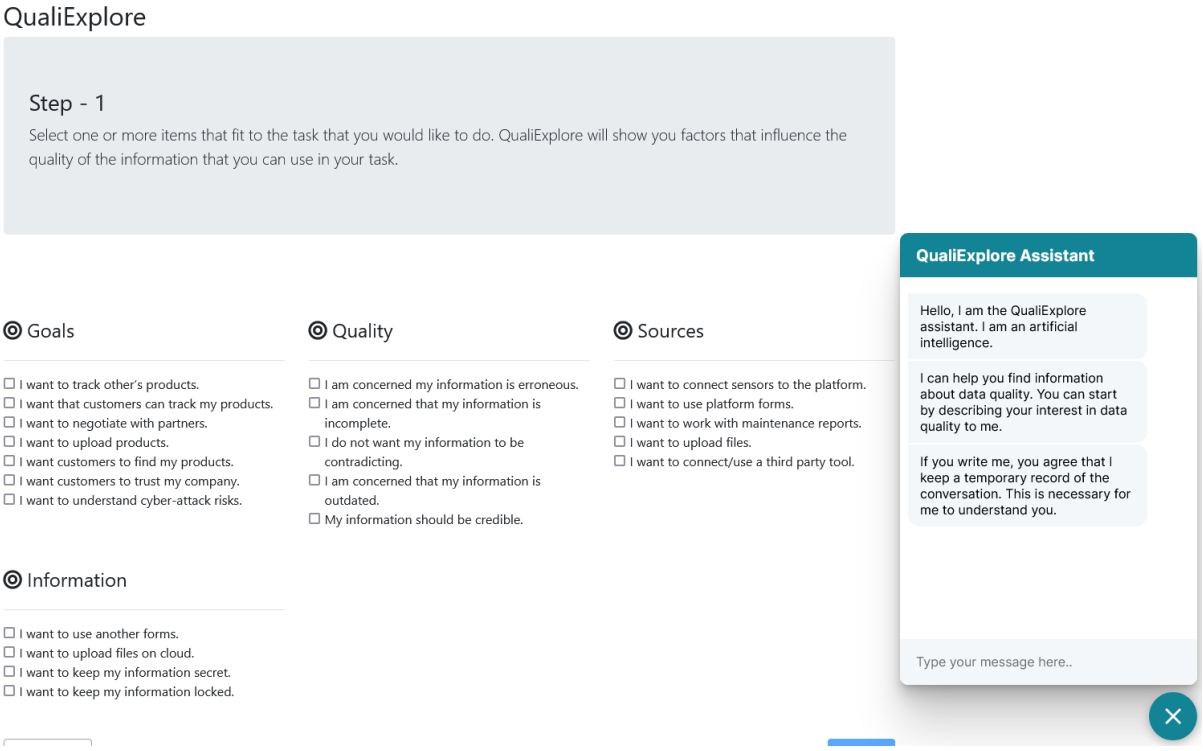

Figure 9. User opens a conversation with chatbot

The following interactions result from legal obligations (e.g., emerging AI Law) and best practices of building chatbots. The entire conversation is a series of turns where the user and the bot exchange their messages. The bot will initiate the conversation after clicking the chat icon.

The bot revealing itself is an obligation resulting from the emerging EU AI Act (Section 5.2.4.). Typically, chatbots also inform the user about their purpose in the first conversation turn.

Chatbots can create and record personal information about users, e.g., log their IP address, maintain a conversation history, or store their name to personalize the interaction. Therefore,

chatbots are subject to the General Data Protection Regulation (GDPR) and its legal obligations. It informs the user in the first turn about the fact that it may collect personal information.

Any further turns depend on the user – users familiar with the bot will likely ask more specific and directed questions while novices will likely be more explorative.

A typical interaction is asking for the bot's capabilities to understand the features. Capabilities for the assistant are illustrated in Figure 10.

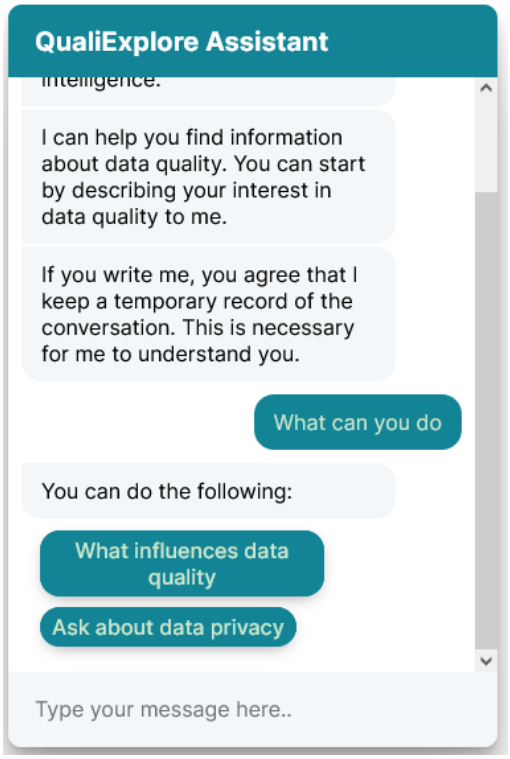

Figure 10. User asking for the assistant's capabilities

Besides the initial GDPR-related information, the chatbot has features to answer questions related to the GDPR. **Figure 11** summarizes some questions.

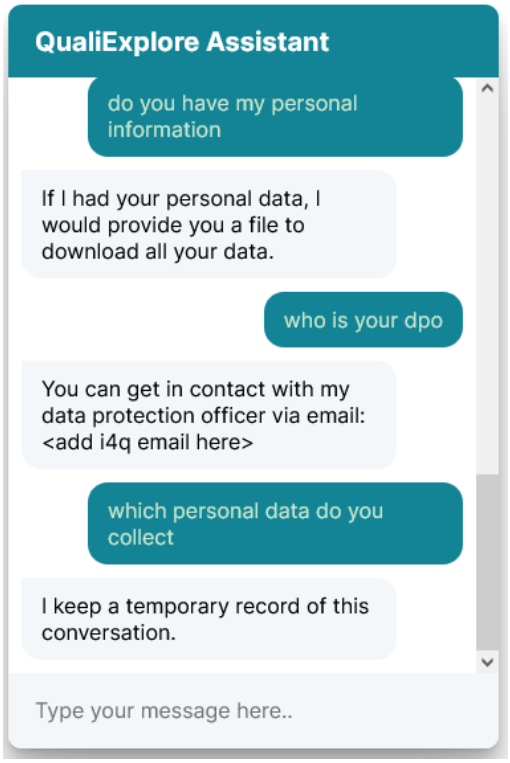

Figure 11. User asking for information about data privacy

The remaining interactions allow illustrating how users access the factor knowledge.

Figure 12 illustrates an example dialogue demonstrating an interaction between a user and the bot. The user asks for an overview (redirect to tree view page) and then about a specific interest where the bot responds with related factors.

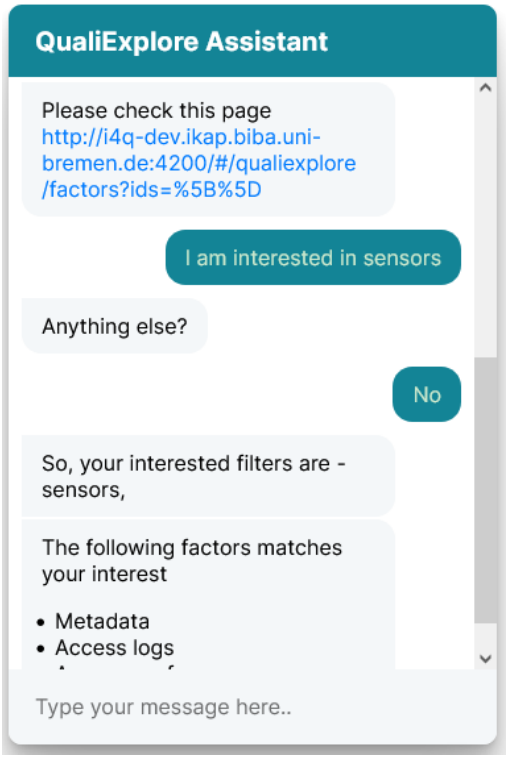

Figure 12. Dialogue to access quality factor knowledge

2. Technical Specifications

i4Q^{QE} version 1 consists of three essential components (refer to Figure 13): a website (front-end) that uses the Angular framework, the API manager GraphQL to simplify working with multiple services, and a MongoDB database to store and organize the data quality factor knowledge. i4Q^{QE} version 2 replaces the MongoDB with a Neo4J graph database and introduces a natural language interface realized through a chatbot. This bot uses Rasa Open Source and a custom-built dialogue model. Figure 13 presents the architecture for i4Q^{QE} with the expected inputs and outputs.

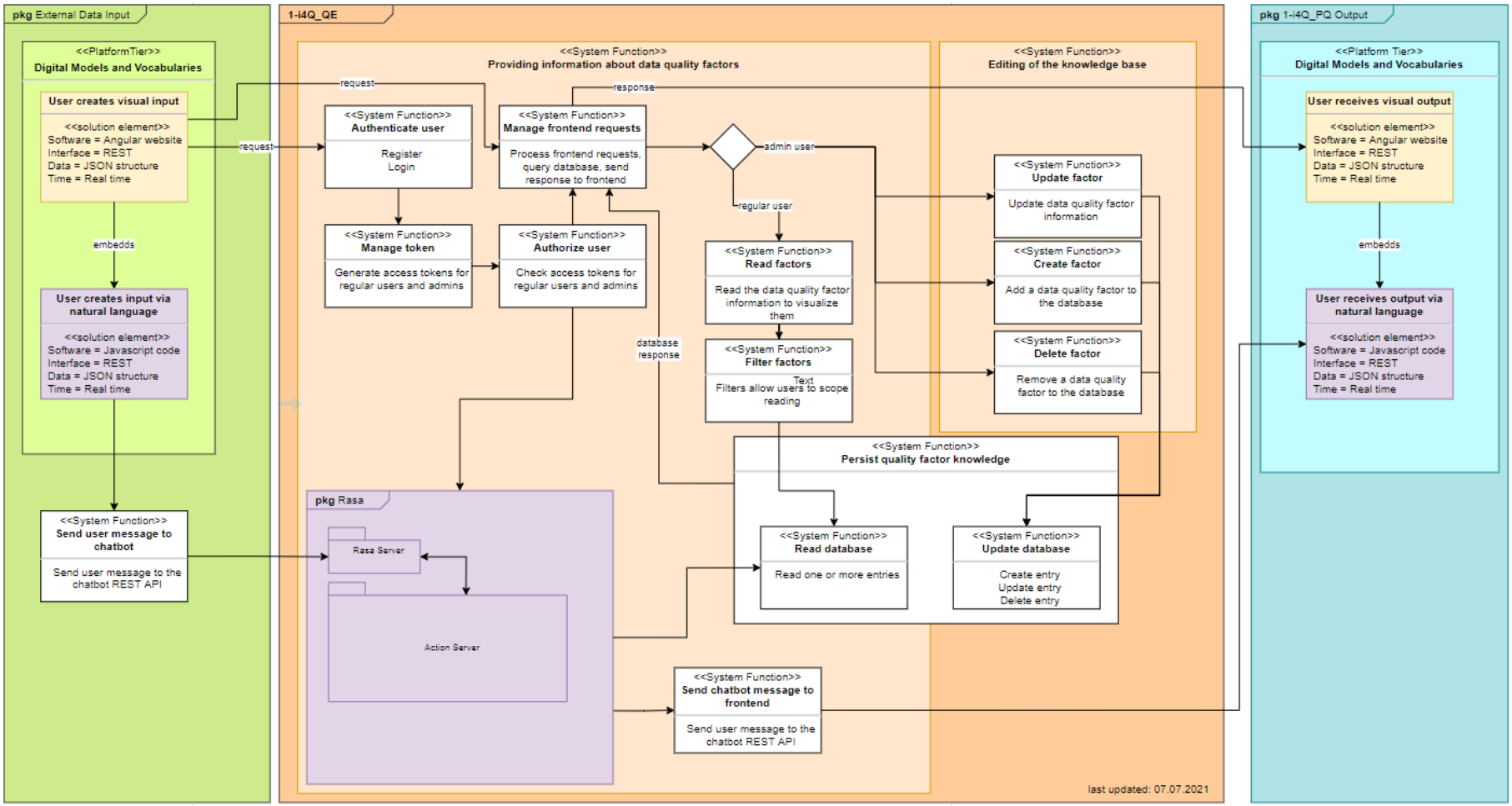

Figure 13. i4Q^{QE} architecture

3. Implementation Status

3.1 Current implementation

ID	Requirements	Progress descriptions
BIBA2r1	QualiExplore shall be accessible via a website (it is a web-based solution).	QualiExplore has an Angular-based front-end accessible via a web browser. (**fulfilled**)
BIBA2r2	QualiExplore users shall be able to create, edit, and delete factor descriptions.	QualiExplore has a (new) editing environment for filters and factor knowledge. (**fulfilled**)
BIBA2r3	QualiExplore factor descriptions shall be intelligible and relevant.	We will extend and improve factor knowledge during the i4Q project. Current descriptions are the first versions not tested and refined yet. (**not fulfilled**)
BIBA2r4	QualiExplore users shall be able to filter factors (to reduce cognitive load).	QualiExplore uses filter statements to highlight the most relevant factors. Version two allows users to describe their interest in data quality factors via natural language. (**fulfilled**)
BIBA2r5	QualiExplore contents should adopt acknowledged terminology from standards and literature (credibility).	QualiExplore uses the conceptual grounding of D3.1 "i4Q Data Quality Guidelines" and terms from ISO standards. (**fulfilled**)
BIBA2r6	QualiExplore shall use terminology used in other i4Q solutions (coherence).	QualiExplore's factor knowledge does not yet contain specific terms from other i4Q solutions. We will align during the evaluation. (**not fulfilled**)
BIBA2r7	Other i4Q solutions might integrate QualiExplore as a HTML/JavaScript widget (usability/ease of access).	Other solutions made no specific integration requests yet. (**not relevant**)
PC5r7	Quality issues/defects root cause analysis shall be realised using integrated data from production processes by the application and integration of several i4Q solutions, such as the i4Q_DA, i4Q_BDA and i4Q_AD	QualiExplore could support the process as a measure to raise awareness for production data quality. (**not relevant**)
PC3r1	The system shall predict the product conformity.	QualiExplore does not predict data quality. (**not feasible / out of scope**)

Table 2. i4Q[QE] requirements

3.2 Next developments

We will adjust the implementation and the knowledge base during the evaluation and fix bugs.

The planned major releases are as follows:

Version	Release date	New features
3.0	December 2023	Final version with content updates and bug fixes

Table 3. Release Dates

3.3 History

Version	Release date	New features
1.0	June 2022	Editing environment, MongoDB integration, Chatbot widget implementation (no backend)
2.0	December 2022	Chatbot integration, Neo4J graph database integration, content updates

Table 4. History

4. Conclusions

This deliverable summarized the features, technical architecture, and implementation status for i4Q^{QE}. The current implementation mainly allows first tests with end-users to improve the contents in the knowledge base. We will perform these tests in parallel with the evaluation of D3.9 "i4Q Data Quality Guidelines" because i4Q^{QE} is complementary to the proposed guideline.

Critical challenges are a) identifying the relevant quality factors and understandable descriptions and b) designing the dialogue model for the natural language interface (chatbot). The latter is a new application domain for chatbots, which means the reliability of the chatbot may require substantial improvement over time.

References

No references

Appendix I

i4Q **QualiExplore** ***(i4Q^{QE})*** *web documentation can be accessed online at:* http://i4q.upv.es/2_i4Q_QE/index.html *without putting the information in the deliverable as it is already in the web page.*