



D3.2 – QualiExplore for Data Quality Factor Knowledge

WP3 – BUILD: Manufacturing
Data Quality



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	D3.2 – QualiExplore for Data Quality Factor Knowledge			
WORK PACKAGE	WP3 – BUILD: Manufacturing Data Quality			
DATE OF DELIVERY	CONTRACTUAL	June 2022	ACTUAL	June 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	<p>This document is the second result of T3.1. Its relationship to other documents is as follows:</p> <ul style="list-style-type: none">• D3.1 Data Quality Guidelines: Mentions this deliverable as a data quality management measure (M18)• D3.10 QualiExplore for Data Quality Factor Knowledge: follow-up deliverable (M24)			
EXTERNAL ANNEXES/SUPPORTING DOCUMENTS	None			
READING NOTES	None			
ABSTRACT	<p>This deliverable is one of the two results of Task 3.1, “Manufacturing Data Quality Strategy”. It presents 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. D3.1 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.</p>			



Document History

VERSION	ISSUE DATE	STAGE	DESCRIPTION	CONTRIBUTOR
0.1	02-May-2022	ToC	Create ToC and Section Description	BIBA
0.2	31-May-2022	1 st Draft	1 st Draft available for review	BIBA
0.3	10-Jun-2022	Internal Review	Internal Review Process	ENG, WHI
0.4	20-Jun-2022	2 nd Draft	Process comments from internal review	BIBA
1.0	30-Jun-2022	Final Draft	Final quality check and issue of final document	CERTH

Disclaimer

Any dissemination of results reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Copyright message

© i4Q Consortium, 2022

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged.



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.....	8
1.2.2 Editing environment.....	10
1.2.3 Knowledge base	13
1.2.4 Natural language interface	13
2. Technical Specifications	15
2.1 Overview	15
2.2 Architecture Diagram	15
3. Implementation Status	17
3.1 Current implementation.....	17
3.2 Next developments.....	18
3.3 History.....	18
4. Conclusions	19
5. References	20
Appendix I	21

LIST OF FIGURES

Figure 1. Tree structure to present data quality factors	9
Figure 2. Filter categories and filter statements.....	9
Figure 3. Login screen.....	10
Figure 4. Admin view of the filter page	10
Figure 5. Admin view of the tree page	11
Figure 6. Editing filters categories and statements	11
Figure 7. Editing the tree structure	12
Figure 8. Editing filter descriptions and filter statement assignments	12
Figure 9. Chatbot widget for natural language interactions.....	13
Figure 10. Example dialog to access quality factor knowledge.....	14
Figure 11. QualiExplore architecture in i4Q.....	16
Figure 12. Final version of QualiExplore in i4Q.....	18



LIST OF TABLES

Table 1. Feature overview and version	8
Table 2. QualiExplore requirements	17
Table 3. Release Dates	18
Table 4. History	18



ABBREVIATIONS/ACRONYMS

AD	Analytics Dashboard
API	Application Programming Interface
BDA	Big Data Analytics
DA	Data Analytics
GraphQL	Name of a query language for APIs
ISO	International Organisation for Standardisation
JSON	JavaScript Object Notation
MongoDB	Name of a document database
Neo4J	Name of a graph database
QE	QualiExplore



Executive summary

This deliverable is one of the two results of Task 3.1, “Manufacturing Data Quality Strategy”. It presents QualiExplore, 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. This i4Q tool is an example of a measure to raise awareness of data quality. D3.1 outlines these and other 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.

QualiExplore 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.

QualiExplore has nine requirements. Three are fulfilled, one is partially fulfilled, two are not fulfilled yet, and three must be revised because they might not be relevant or in the tool’s scope. Current i4Q features include an editing environment for filters and factor knowledge, a database to store the knowledge, and the first implementation of a chatbot widget. QualiExplore’s next version will integrate the chatbot’s backend with the Rasa Open Source framework, a graph database with factor knowledge in Neo4J, and content updates. A final version will include bug fixes and content updates.



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.

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

QualiExplore is 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. QualiExplore 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.

This *i4Q* tool is an example of a measure to raise awareness of data quality. D3.1 outlines these measures in its proposed activity framework. QualiExplore 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 QualiExplore's 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.

1.2.1 Data quality factors and filter functions

QualiExplore 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. QualiExplore's content has a tree shape, as illustrated in Figure 1.

QualiExplore

Step-2
This step highlights the most relevant factors with a

[Selected Filters](#)

[Return to Step-1](#)

Discover all Quality Factors

[Edit Tree](#)

- ▼ Platform information quality
 - ▼ Collection quality
 - ▼ Accuracy
 - Semantic errors
 - Syntactic errors
 - Typographical errors
 - Bias
 - Measurement instrument information
 - Providing disinformation
 - ▼ Consistency
 - Standard application
 - ▼ Completeness

Quality Factor Information

Semantic errors

The semantic problem is a problem of linguistic processing. It relates to the issue of how spoken utterances are understood and, in particular, how we derive meaning from combinations of speech sounds (words).

Sources
[Link To Source](#)

[Edit](#) [Proceed](#)

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 QualiExplore.

QualiExplore

Step - 1
Select one or more items that fit to the task that you would like to do. QualiExplore will show you factors that influence the quality of the information that you can use in your task.

[Add More](#)

Goals

- I want to track other's products.
- I want that customers can track my products.
- I want to negotiate with partners.
- I want to upload products.
- I want customers to find my products.
- I want customers to trust my company.
- I want to understand cyber-attack risks.

Quality

- I am concerned my information is erroneous.
- I am concerned that my information is incomplete.
- I do not want my information to be contradicting.
- I am concerned that my information is outdated.
- My information should be credible.

Sources

- I want to connect sensors to the platform.
- I want to use platform forms.
- I want to work with maintenance reports.
- I want to upload files.
- I want to connect/use a third party tool.

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 QualiExplore 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 will simplify 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.

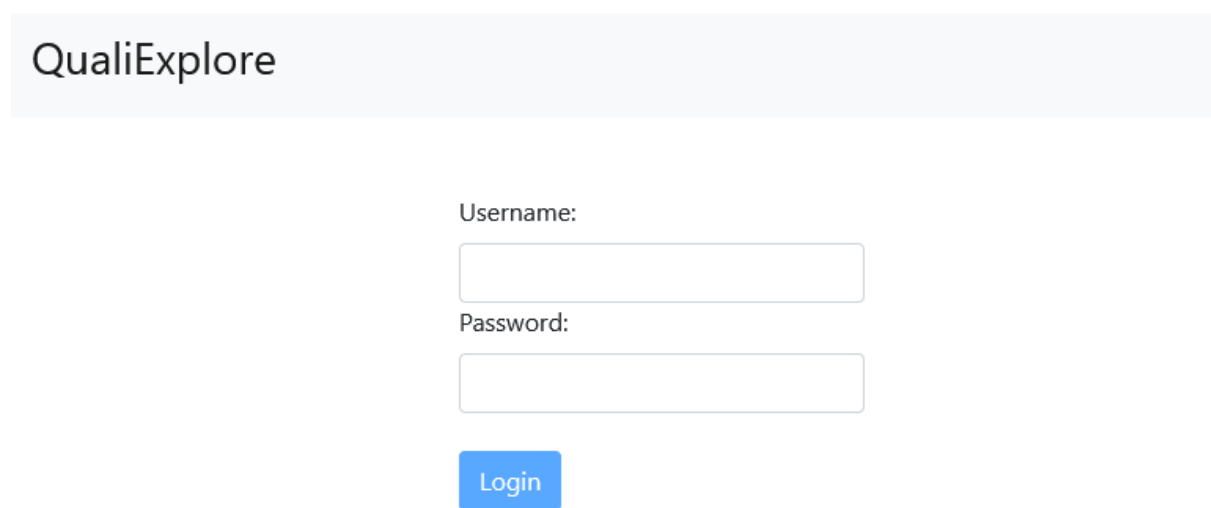


Figure 3. Login screen

If an admin user logs in, they will see editing buttons on QualiExplore's filter and factor sites. Figure 4 and Figure 5 present the admin views for both sites.

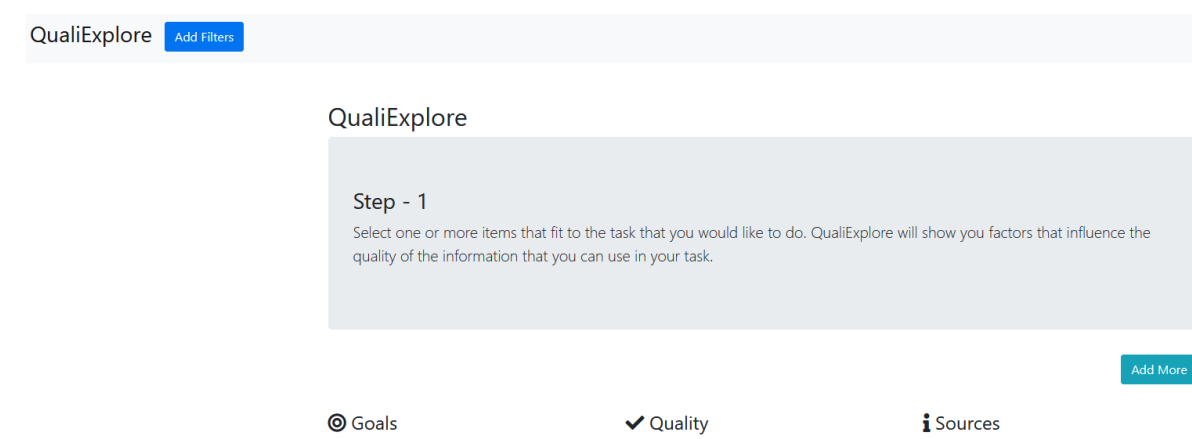


Figure 4. Admin view of the filter page

Discover all Quality Factors [Edit Tree](#)

- Platform information quality
 - Collection quality
 - Accuracy
 - Semantic errors
 - Syntactic errors
 - Typographical errors
 - Bias
 - Measurement instrument information
 - Providing disinformation
 - Consistency
 - Standard application
 - Completeness
 - Measurement frequency
 - Technical issue
 - Software bug
 - Standard application
 - Metadata

Quality Factor Information

Semantic errors

The semantic problem is a problem of linguistic processing. It relates to the issue of how spoken utterances are understood and, in particular, how we derive meaning from combinations of speech sounds (words).

Sources
[Link To Source](#)

[Edit](#) [Proceed](#)

Progress

0 of 4

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.

Edit your data ×

Filter category name

Your tasks are here

Statement 1

Statement 2

[Update](#) [Close](#)

Figure 6. Editing filters categories and statements

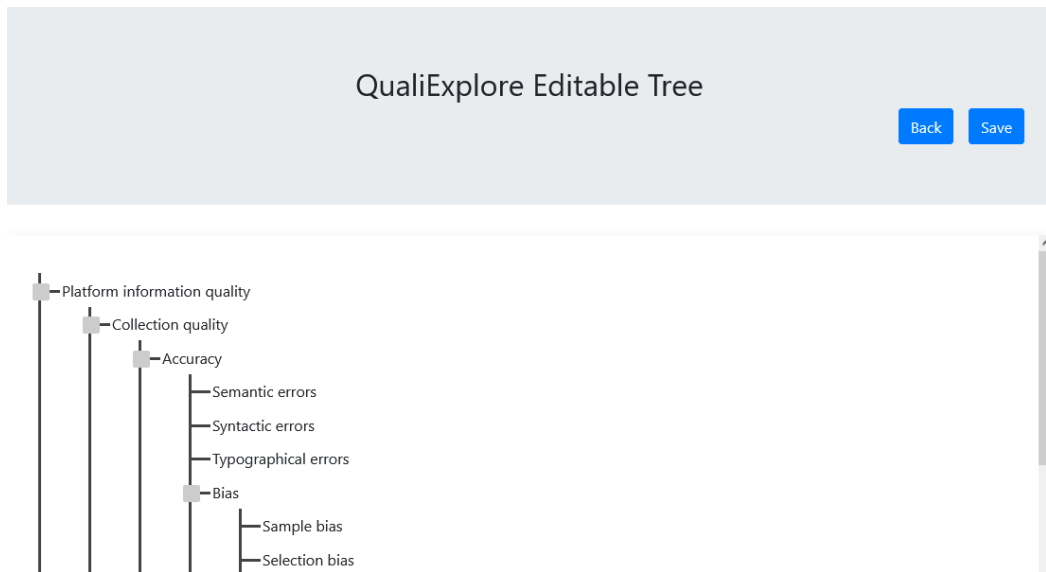


Figure 7. Editing the tree structure

The image shows a dialog box titled "Edit Sources and Links" with a close button (X) in the top right corner. The dialog is for editing the "Semantic errors" filter. It contains a text area with the following description: "The semantic problem is a problem of linguistic processing. It relates to the issue of how spoken utterances are understood and, in particular, how we derive meaning from combinations of speech sounds (words)." Below the text area is a URL input field containing "https://www.sltinfo.com/the-semantic-problem/". At the bottom of the dialog, there is a list of checkboxes for assigning statements to the filter: "Statement 1", "Statement 2", "Lorem ipsum dolor sit amet.", "Lorem ipsum dolor sit amet.", "Lorem ipsum dolor sit ame", "Lorem ipsum dolor sit am", and "test". At the bottom right of the dialog, there are two blue buttons labeled "Update" and "Close".

Figure 8. Editing filter descriptions and filter statement assignments

1.2.3 Knowledge base

QualiExplore 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 stores the knowledge as an intermediary solution in this deliverable. The second version will use 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 QualiExplore. It accesses the knowledge base, and users can ask about production data quality factors. **Figure 9** illustrates how users access the chatbot.

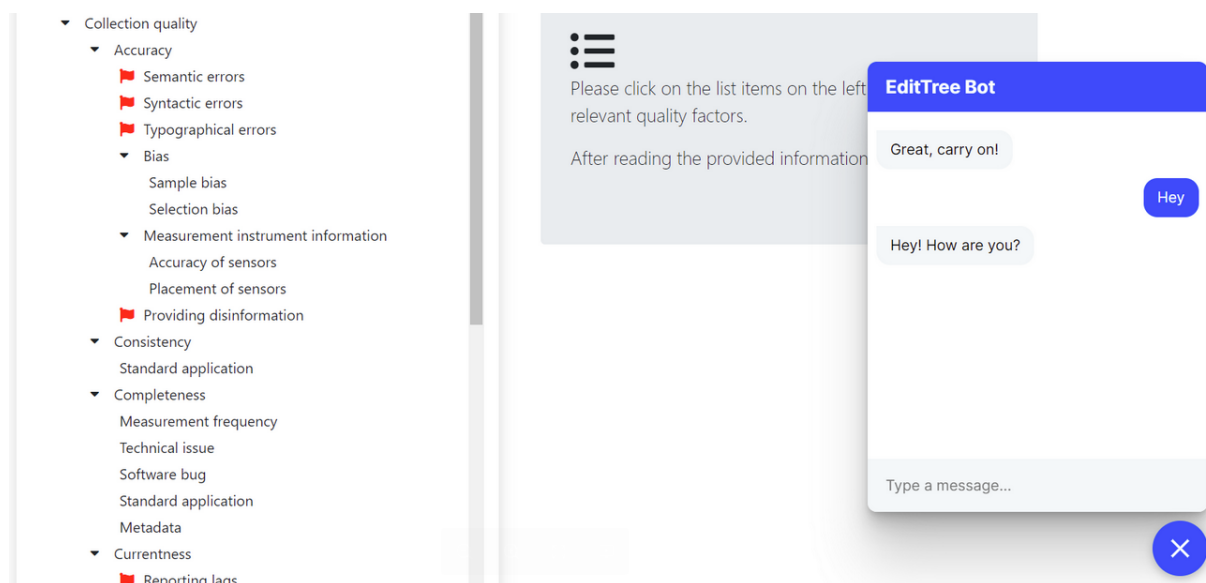


Figure 9. Chatbot widget for natural language interactions

This deliverable’s second version will detail how the chatbot works. It will contain example dialogs and the essential conversation patterns. **Figure 10** illustrates an example dialog demonstrating an interaction between a user and the bot.



Example Dialog (U: User, B: Bot)

U: Which factors should I consider when I *install new sensors* ← Reference to data life cycle
(extracted from text)

B: You should consider these factors:

- [Calibration](#)
- [Accuracy of sensors](#)

U: What about calibration?

B: Calibration means that you ...

B: It contributes to collection quality and affects the measurement's accuracy.

Reference to data life cycle

Quality characteristic

Figure 10. Example dialog to access quality factor knowledge



2. Technical Specifications

2.1 Overview

QualiExplore version 1 consists of three essential components (refer to **Figure 11**): 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. QualiExplore's second version will contain a natural language interface realized through a chatbot. This bot will use Rasa Open Source and a custom-built dialog model.

2.2 Architecture Diagram

Figure 11 presents the architecture for QualiExplore with the expected inputs and outputs.

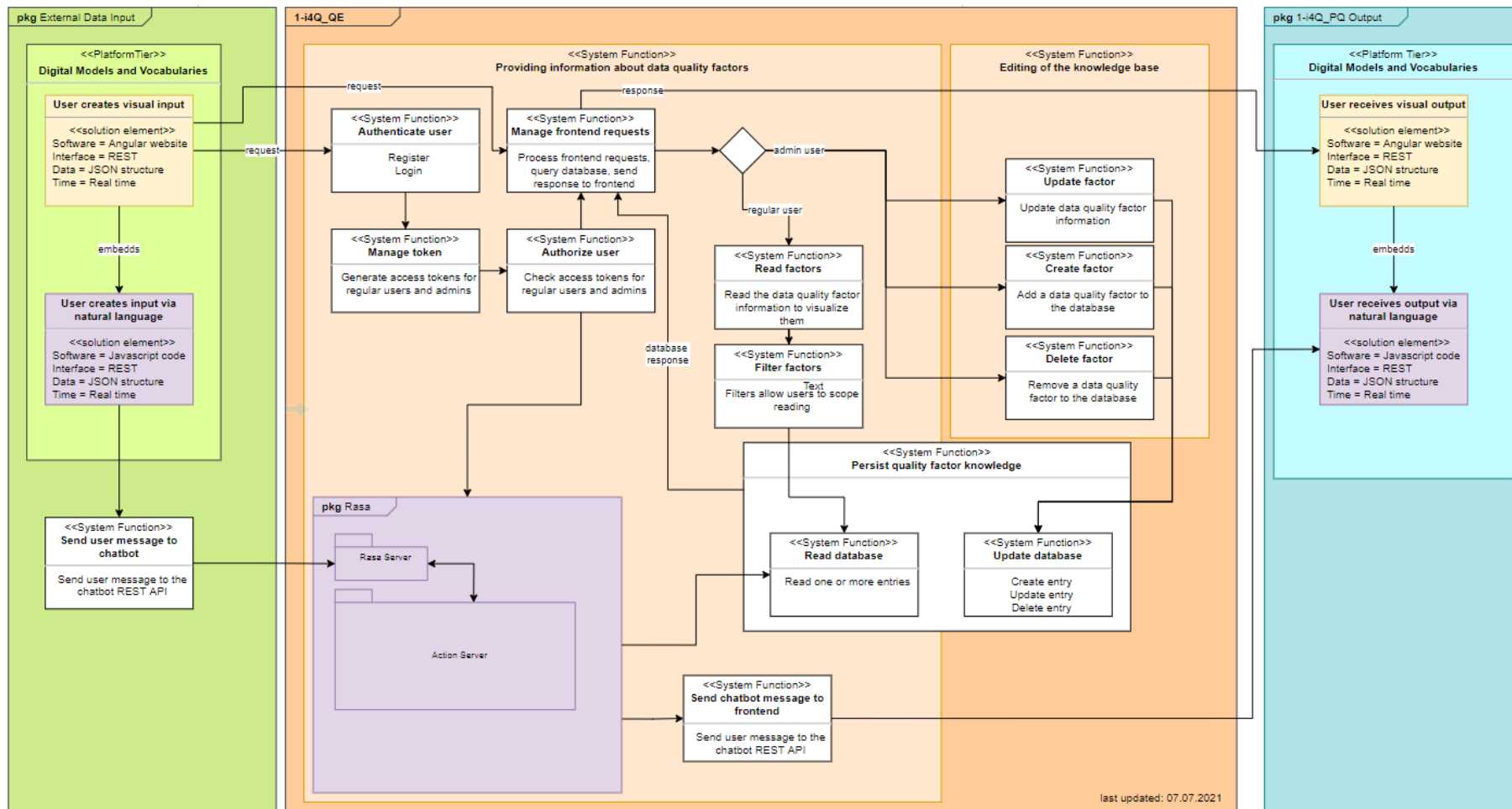


Figure 11. QualiExplore architecture in i4Q



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 will allow users to describe their interest in data quality factors via natural language. (partially 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. (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. (potentially 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. (potentially not relevant)
PC3r1	The system shall predict the product conformity.	QualiExplore does not predict data quality. (not feasible)

Table 2. QualiExplore requirements

The last two requirements above do not specifically relate to QualiExplore. PC5r7 is most likely not specific enough to be relevant (i.e., no specific feature will fulfill the requirement). PC3r1 requests a feature that is out of QualiExplore’s scope. We will monitor and reject or revise both requirements in this deliverable’s second version.

3.2 Next developments

The next critical step is deploying the chatbot and designing the dialog model with Rasa. This step involves designing example dialogs and collecting training data for Rasa’s natural language understanding functions. In parallel, we will migrate the JSON structures in the MongoDB into a graph structure to increase the knowledge’s maintainability and simplify future changes (e.g. relating different factors). We will store this graph in a Neo4J database and make it accessible via GraphQL. **Figure 12** illustrates QualiExplore’s final version and the two ways to interact with the database.

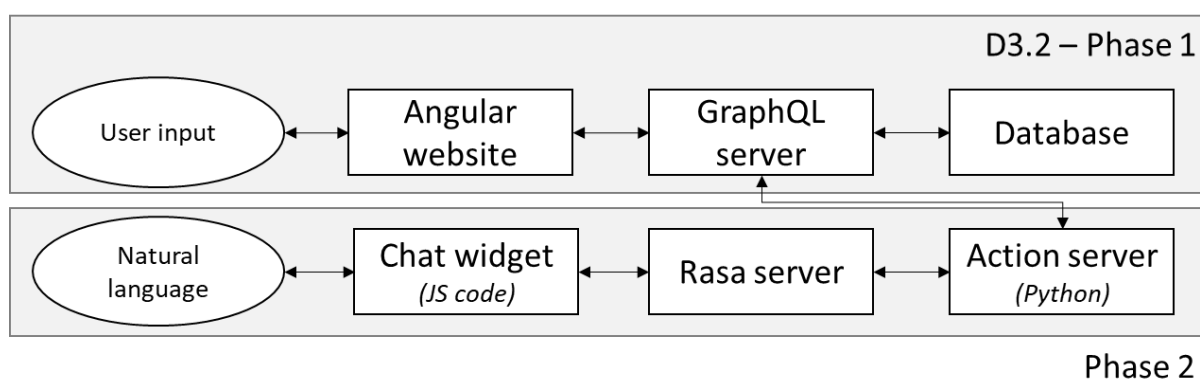


Figure 12. Final version of QualiExplore in i4Q

The planned releases are as follows:

Version	Release date	New features
2.0	December 2022	Chatbot integration, Neo4J graph database integration, content updates
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)

Table 4. History



4. Conclusions

This deliverable summarized the features, technical architecture, and implementation status for i4Q's QualiExplore tool. 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 D3.1 “i4Q Data Quality Guidelines” because QualiExplore is a complementary tool to the proposed guideline.

Critical challenges are a) identifying the relevant quality factors and understandable descriptions and b) designing the dialog 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.



5. References

No references



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

i4Q QualiExplore for Data Quality Factor Knowledge

General Description

QualiExplore is a web-based software tool for visualizing information quality characteristics and factors. Quality factors affect information characteristics (e.g. accuracy, timeliness, completeness, and credibility) and, therefore, influence information quality. Producers need to know these factors to assess their relevance and identify measures to manage them.

QualiExplore 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 may confuse 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.

This **i4Q** tool is an example of a measure to raise awareness of data quality. D3.1 outlines these and other measures in an activity framework. QualiExplore is a standalone application and does not depend on other **i4Q** solutions. It uses a permissive open source license (MIT), so stakeholders can easily exploit it.

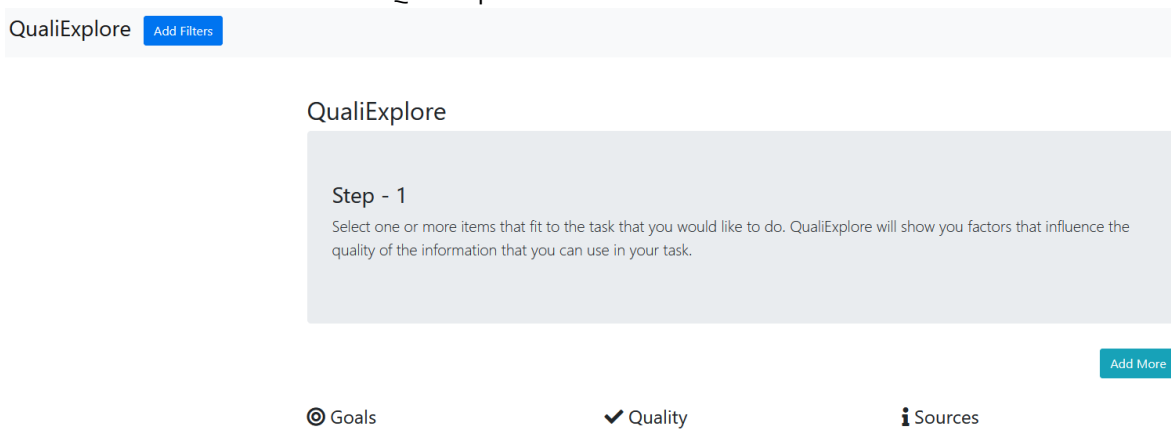
Features

1. **Data quality factors and filter functions:** QualiExplore 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. QualiExplore's content has a tree shape, as illustrated in **Figure 1**. 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 QualiExplore. 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 QualiExplore will highlight.

Editing environment: QualiExplore's original version allows developers to change filters, tree structure, and factors by modifying the related files. **i4Q** will simplify 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. If an admin user logs in, they will see editing buttons on QualiExplore's filter and factor sites.



2. **Figure 4** and **Figure 5** present the admin views for both sites. **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.
3. **Knowledge base:** 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 stores the knowledge as an intermediary solution in this deliverable. The second version will use 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.
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 QualiExplore. It accesses the knowledge base, and users can ask about production data quality factors.

ScreenShots

Screenshots are available in the core deliverable (figures).

Commercial Information

Authors

Company	Website	Logo
BIBA - Bremer Institut für Produktion und Logistik GmbH	http://www.biba.uni-bremen.de/	

License

The recent license is available here: <https://github.com/s-wel/qualiexplore>



Pricing

This solution uses a permissive open source license. Third parties can exploit it commercially.

Associated i4Q Solutions

This solution does not depend on other i4Q solutions. Its requirements are detailed in the open repository: <https://github.com/s-wel/qualiexplore>. The i4Q-related improvements are stored in separate branches and will be merged with the master once they are stable.

Installation Guidelines

The latest installation guideline is accessible in the open repository: <https://github.com/s-wel/qualiexplore>

User Manual

The open repository contains a brief user manual: <https://github.com/s-wel/qualiexplore>