



Project Title International Cooperation for Digital Standardisation

Project Acronym INSTAR

Grant Agreement No. 101135877

Start Date of Project 01.01.2024

Duration of Project 30.06.2026

Project Website <https://www.instarstandards.org/>

D6.3 – Standards Tracer Portal

Work Package	WP6
Lead Author (Org)	Homer Papadopoulos (NCSRD)
Contributing Author(s) (Org)	George Balaskas (NCSRD), Teresa Papazoglou (NCSRD)
Due Date	31.12.2024
Date	29.01.2025
Version	V1.0

Dissemination level

(X) PU: Public

() PP: Restricted to other programme participants (including the Commission)

() RE: Restricted to a group specified by the consortium (including the Commission)

() CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Author	Notes
0.1	05.12.2024	Homer Papadopoulos (NCSRD)	Table of Contents
0.2	28.12.20205	Homer Papadopoulos (NCSRD). George Balaskas (NCSRD), Teresa Papazoglou (NCSRD)	First draft
0.5	22.01.2025	Luigi Colucci (TrustIT)	First review
0.6	23.01.2025	Homer Papadopoulos (NCSRD)	Implementation of comments
0.7	24.01.2025	Antonio Kung (Trialog)	Second review
1.0	29.01.2025	Homer Papadopoulos (NCSRD)	Implementation of comments and final release

Disclaimer

This document contains information which is proprietary to the INSTAR Consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by any means to a third party, in whole or parts, except with the prior consent of the INSTAR Consortium.

Table of contents

D6.3 – Standards Tracer Portal	1
1 Introduction.....	7
2 Standards Tracer Portal	9
3 Standards Repository and Rolling Plans.....	11
4 News Aggregation Through Web Scraping.....	13
5 Database Construction for Standards via Web Scraping.....	16
5.1 Scraping Process.....	16
6 Standards Chatting Tool.....	17
6.1 Search Infrastructure	17
6.2 Document Storage Architecture	18
6.3 RAG Implementation	19
6.4 Data Integration and Security.....	19
6.5 Access Control	21
6.6 User Interface.....	21
7 Conclusion and Future Directions	26

List of Figures

Figure 1 Portal functionality	9
Figure 2 Standards Repository	11
Figure 3: News Repository.....	13
Figure 4 : News Article	15
Figure 5 : Rag approaches	19
Figure 6 : Available document store and their types.....	20
Figure 7: User Control Settings	21
Figure 8 : User Interface of the chatting tool.....	22
Figure 9 : Retrieved Information with sources and relevancy score.....	23
Figure 10 : Retrieved information from a GraphRAG document store with the relevant knowledge graph and table of entities	24

Terminology

TERMINOLOGY/ACRONYM	DESCRIPTION
AI	Artificial Intelligence
BLU	BluSpecs
CSA	Coordination and Support Action
EC	European Commission
ICT	Information Technology
INSTAR	International cooperation for digital standardisation
IoT	Internet of Things
NCSR"D"	National Centre for Scientific Research "Demokritos"
Partners	Refers to the INSTAR Consortium partners
WP	Work Package
ISO	International Organization for Standardization
IEC	International Electrotechnical Commission
ETSI	European Telecommunications Standards Institute
NIST	National Institute of Standards and Technology
EU	European Union
MSP	Multi-stakeholder Platform
AUWP	Annual union work program
ASTM	American Society for Testing and Materials
AFNOR	Association Française de Normalisation
BSI	British Standards Institution
HTML	Hypertext Markup Language
IOT	Internet of Things
URL	Uniform Resource Locator
RAG	Retrieval Augmented Generation
BM25	Best Matching 25
LLM	Large Language Model

Executive Summary

Work Package 6 focuses on the creation of the **Standards Tracer Portal**, a software tool designed to automate the collection, analysis, and dissemination of standards-related information. By integrating APIs, web crawlers, and AI-driven analytics, the portal ensures that policymakers, industry stakeholders, and organizations have seamless access to up-to-date, comprehensive standards data related to INSTAR work streams. Key features include a dynamic repository of public standards, real-time updates on emerging technologies, and advanced search capabilities that allow users to efficiently locate, analyze, and share relevant information.

The portal also incorporates cutting-edge tools, such as AI-powered chat functionalities and a hybrid vector search system, to enable intelligent query resolution across a vast database of standards and technical documentation. With its ability to deliver real-time updates, actionable insights, and tailored information retrieval, the Standards Tracer Portal empowers stakeholders to make informed decisions.

The implementation of the **Standards Tracer Portal** aligns with global and EU-specific standardization objectives by fostering innovation, ensuring interoperability, and supporting critical domains like cybersecurity and emerging technologies. Through systematic methodologies and intelligent features, this tool aims to become a key resource for technical experts, policymakers, and industry leaders navigating the complexities of the standardization ecosystem.

1 Introduction

Work Package 6 is dedicated to the development of the **Standards Tracer Portal**, an advanced software solution designed to address the critical need for up-to-date and comprehensive standards-related information. This portal aims to empower policymakers, industry stakeholders, and organizations by automating the collection, analysis, and dissemination of standards from global, regional, and national SDOs and other relevant sources. By leveraging cutting-edge technologies such as APIs, web crawlers, and AI-powered analytics, the portal will offer a dynamic and user-friendly platform for accessing and engaging with standards data.

The Standards Tracer Portal aims to become a tool that ensures stakeholders have access to the latest standards and technical documentation in real time. Its functionality includes automated data collection from key SDOs, including international bodies like the **ISO** and **IEC**, regional organizations such as **ETSI**, and national entities like **NIST**. Using the web scraping approach, the portal has the ability to extend its coverage to additional websites that provide valuable, complementary information not found in its primary data sources, particularly when these sites do not require login credentials. For example the web scraping will be used to access and collect data from the Genorma (Genorma.com) platform which is a pan-European information point for standards and legislation in Europe. It covers EU and international standards from Iso and IEC. The platform provides more information on preliminary work items of ISO (which ISO does not publish publicly), such as ISO/IEC 27568 (<https://genorma.com/en/standards/iso-iec-pwi-27568>). In addition, the portal will integrate a dedicated repository of standards, including resources like the EU's ICT Rolling Plans, alongside advanced search capabilities, real-time news aggregation, and AI-driven analytical tools.

Prior to the development of the Standards Tracer Portal, policymakers and industry stakeholders faced the challenge of manually navigating multiple platforms to obtain comprehensive standards-related information. For instance, they would need to visit the EU portal to access ICT Rolling Plans, search ISO and ETSI websites for relevant standards, and consult IEC and NIST platforms for news and updates. This fragmented process was not only time-consuming but also inefficient in ensuring access to the latest developments. The Standards Tracer Portal addresses these challenges by centralizing all these resources into a single, integrated platform. In addition, the portal's AI-powered Standards Chatting Tool enables users to interact with the standards database to receive precise and timely answers to complex queries, streamlining workflows and empowering stakeholders to make well-informed decisions

The methodology that has been adopted for the portal concerns the following steps:

Identify key relevant web portals (open to public) within their respective industries to ensure comprehensive coverage of relevant standards across different sectors.

International SDOs like:

- ISO (International Organization for Standardization) – Develops and publishes international standards across various industries. <https://www.iso.org/home.html>
- IEC (International Electrotechnical Commission) – Specializes in electrotechnology standards. <https://www.iec.ch/homepage>

Regional SDOs like:

- ETSI (European Telecommunications Standards Institute) – European standards for telecommunications. <https://www.etsi.org/>

Major National SDOs **like**:

- NIST: <https://www.nist.gov/>

This tool also offers a comprehensive database created through advanced web scraping techniques. The scraping process employed different approaches tailored to each website, utilizing tools such as BeautifulSoup for ETSI and Selenium for ISO. During this process, a total of **25,894 ISO standards** were downloaded by scraping **2,590 pages** of results, while **48,309 ETSI standards** were successfully collected. Only standards relevant to key categories such as Quantum Technologies, 5G+/6G, Data Technologies, Artificial Intelligence, IoT/Edge/Cloud/Internet, and Cybersecurity/eID were included in the database. Additionally, the tool features a powerful AI-based Standards Chatting Tool that utilizes hybrid vector search and Retrieval-Augmented Generation (RAG) approaches for intelligent query resolution.

The tool will be a significant asset for stakeholders requiring up to date standards information. The integration of user-friendly features—such as intuitive search filters, estimated reading times, and contextual metadata—will further enhance the user experience, enabling stakeholders to make well-informed decisions. For example, a non expert in standards engineer working in an SME could easily locate the latest specifications for implementing 5G+/6G technologies or cybersecurity protocols for their products.

2 Standards Tracer Portal

The Standards Tracer Portal incorporates advanced search capabilities that enable users to locate specific standards, documentation, and related content through keyword-based search functionality. Users can refine their searches using filters for document type, publication date, technical domain, and standardization body. The portal also features posting capabilities, allowing authorized users to contribute updates and annotations on standards and news documents, enabling the dissemination of knowledge within the standardization community and other interested users. To enhance user experience and productivity, the portal includes intelligent reading assistance features such as estimated reading time indicators for documents. Updates on Standards Tracer Portal are triggered on a monthly basis to ensure timely inclusion of new standards, revisions, or other relevant developments. This regular schedule allows for consistent monitoring and keeps stakeholders informed without overwhelming them with too frequent changes.

The Standards Tracer Portal incorporates advanced search capabilities that enable users to locate specific standards, documentation, and related content through keyword-based search functionality. Users can refine their searches using filters for document type, publication date, technical domain, and standardization body. The portal also features posting capabilities, allowing authorized users to contribute updates and annotations on standards and news documents, enabling the dissemination of knowledge within the standardization community and other interested users. To enhance user experience and productivity, the portal includes intelligent reading assistance features such as estimated reading time indicators for documents.

The Standards Tracer Portal implements an intuitive interface that prioritizes efficient access to standardisation information and technical documentation. As shown in Figure 1, the portal's main navigation structure provides direct access to Standards, News, and About sections, with an integrated search functionality and Chat assistance feature.

The Standards Tracer Portal implements an intuitive interface that prioritizes efficient access to standardization information and technical documentation.

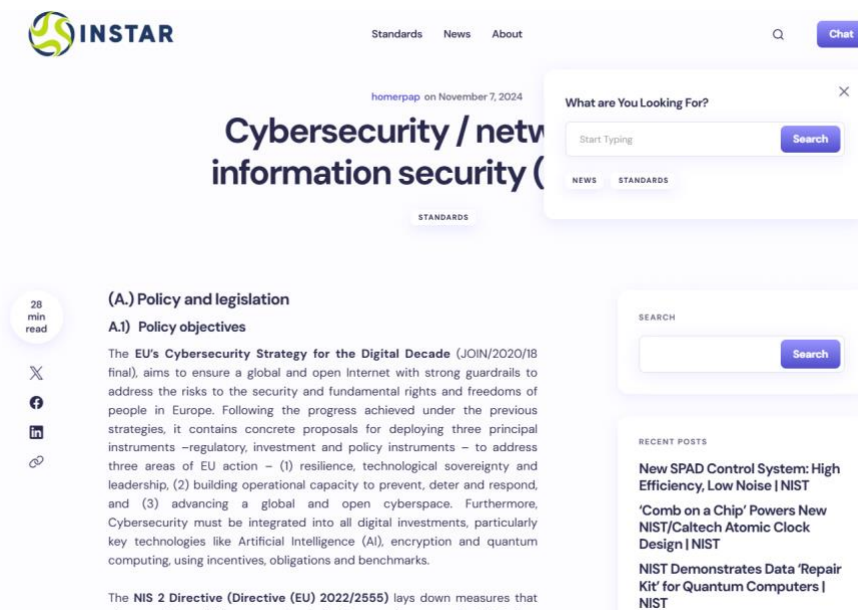


Figure 1 Portal functionality

The document presentation interface incorporates several key elements designed to enhance user productivity. Each document displays an estimated reading time (exemplified by the "28 min read" indicator) to help users effectively plan their review sessions.

Social sharing capabilities are integrated directly into the document interface, allowing users to disseminate important standards information through professional networks. The portal also maintains contextual information for each document, including publication dates and author attribution (as shown by "homerpap on November 7, 2024"), ensuring proper document provenance.

The search functionality is enhanced by category-specific filters (NEWS, STANDARDS) that enable users to quickly narrow their focus to relevant document types. This categorization system aligns with the EU's requirements for clear organization and accessibility of technical documentation, while supporting efficient information retrieval across the standardization landscape.

The Standards Tracer Portal extracts a set of metadata for each standard to ensure a detailed and accurate representation. Key fields include the title, short descriptions or abstracts, publication status, publication date, and the ICS (International Classification for Standards) number, all of which are crucial for understanding and categorizing standards.

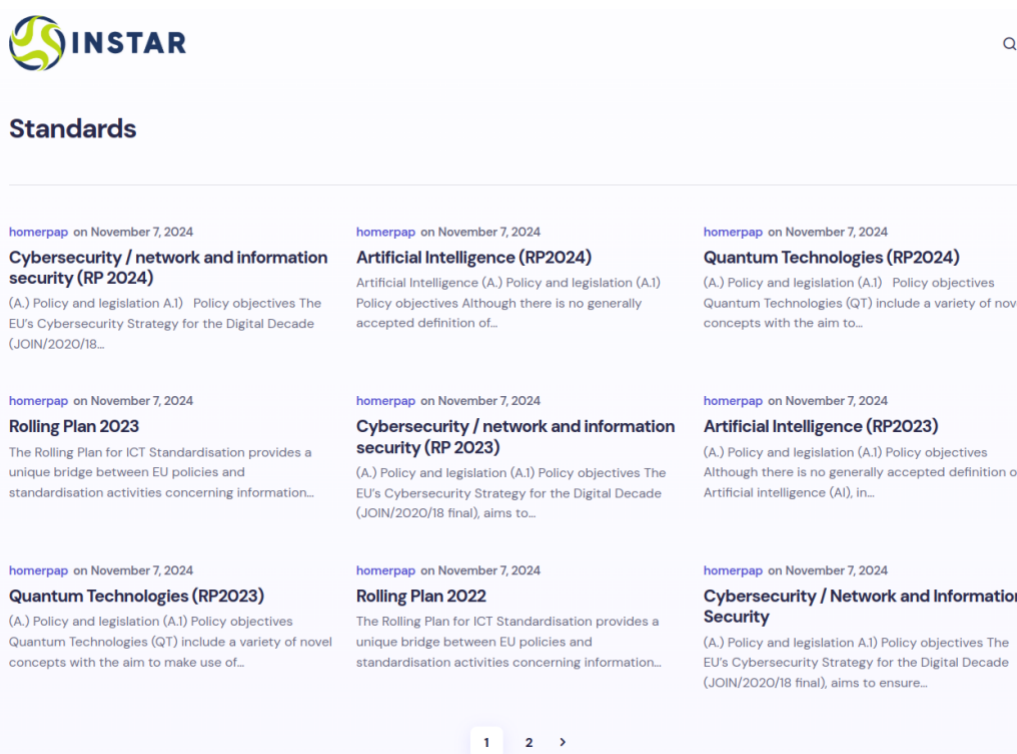
Additionally, the tool is able to capture the historical evolution of standards, leveraging accessible URLs to track multiple editions. For instance, it can retrieve the three editions of ISO 27001 (2005, 2013, 2022) based on available links. However, for more complex tracking, such as the progression of a standard through various development stages—e.g., ISO 27564 moving from PWI in 2022 to NP in 2023, AWI in 2024, and WD in 2024—web scraping should be customized and data storage mechanisms should be employed and the portal should provide such kind of information.

The portal also performs reference-based searches to uncover related updates or revisions that may not be immediately visible. All this information is systematically stored in the database supporting the portal's AI-based chat tool, allowing users to seamlessly explore standard histories, development stages, and related insights through interactive queries.

3 Standards Repository and Rolling Plans

The **Standards Tracer Portal** first section is dedicated to providing access to an organized repository of standards, including the Rolling Plans for ICT Standardisation for the years 2020-2024. Furthermore it focuses on key categories such as Data Economy, Cloud and Edge Computing, Cybersecurity / Network and Information Security, Artificial Intelligence, Quantum Technologies, 5G and Beyond, and Data Interoperability. These Rolling Plans serve as a bridge between European Union (EU) policies and standardisation activities in the domain of information and communication technologies. The repository is easily accessible via the Standards button in the navigation bar, enabling users to quickly find relevant information made available by the SDOs. In principle, the repository can display clusters of standards related to specific trade agreements or cooperation initiatives, such as an EU-Korea agreement focusing on particular standardization areas. It can also highlight the evolution of related standards over time. However, access to such information depends on the extent to which SDOs provide structured data on these relationships and agreements. At present, the repository does not have direct access to this type of data, which limits its ability to implement this feature fully. If this information becomes available, the repository can integrate it to enhance the tracking of standard clusters and their evolution in alignment with trade and cooperation frameworks.

The Rolling Plan for ICT Standardisation is the outcome of an annual dialogue involving a diverse group of stakeholders represented by the European multi-stakeholder platform on ICT standardisation (MSP). This dialogue ensures a comprehensive approach to identifying and prioritizing areas where standardisation can effectively support EU policies. The Rolling Plan does not aim to replace or replicate the work programs of various standardisation bodies (SDOs). Instead, it complements these programs by focusing on actions that directly align with EU policy objectives.



Standards

homerpap on November 7, 2024
Cybersecurity / network and information security (RP 2024)
 (A.) Policy and legislation (A.1) Policy objectives The EU's Cybersecurity Strategy for the Digital Decade (JOIN/2020/18...

homerpap on November 7, 2024
Artificial Intelligence (RP2024)
 Artificial Intelligence (A.) Policy and legislation (A.1) Policy objectives Although there is no generally accepted definition of...

homerpap on November 7, 2024
Quantum Technologies (RP2024)
 (A.) Policy and legislation (A.1) Policy objectives Quantum Technologies (QT) include a variety of novel concepts with the aim to...

homerpap on November 7, 2024
Rolling Plan 2023
 The Rolling Plan for ICT Standardisation provides a unique bridge between EU policies and standardisation activities concerning information...

homerpap on November 7, 2024
Cybersecurity / network and information security (RP 2023)
 (A.) Policy and legislation (A.1) Policy objectives The EU's Cybersecurity Strategy for the Digital Decade (JOIN/2020/18 final), aims to...

homerpap on November 7, 2024
Artificial Intelligence (RP2023)
 (A.) Policy and legislation (A.1) Policy objectives Although there is no generally accepted definition of Artificial Intelligence (AI), in...

homerpap on November 7, 2024
Quantum Technologies (RP2023)
 (A.) Policy and legislation (A.1) Policy objectives Quantum Technologies (QT) include a variety of novel concepts with the aim to make use of...

homerpap on November 7, 2024
Rolling Plan 2022
 The Rolling Plan for ICT Standardisation provides a unique bridge between EU policies and standardisation activities concerning information...

homerpap on November 7, 2024
Cybersecurity / Network and Information Security
 (A.) Policy and legislation (A.1) Policy objectives The EU's Cybersecurity Strategy for the Digital Decade (JOIN/2020/18 final), aims to ensure...

1 2 >

Figure 2 Standards Repository

Key aspects of the Rolling Plan include:

Support for EU Policies: The document identifies areas where ICT standardisation can contribute to achieving EU policy goals, such as fostering innovation, ensuring interoperability, and enhancing cybersecurity.

Defined Actions and Requirements: The Rolling Plan details the specific requirements for ICT standardization, translates these into actionable steps, and establishes a follow-up mechanism to ensure progress.

Complementarity with AUWP: The standardization actions outlined in the Rolling Plan are designed to complement other instruments, particularly the Annual Union Work Program (AUWP).

This section of the **Standards Tracer Portal** provides users with seamless access to these essential resources. By offering an easily navigable interface, professionals can quickly locate relevant information and gain insights into the standardization actions necessary to support EU policies. The repository acts as a reliable reference point, enabling informed decision-making and efficient knowledge retrieval.

4 News Aggregation Through Web Scrapping

The second section of the **Standards Tracer Portal** focuses on delivering real-time updates and news related to Quantum Technologies standards and emerging technologies. The repository is accessible via the **News** button in the navigation bar. As a preliminary step, we have implemented the news aggregation and scrapping for the field of Quantum Technologies. This is achieved through an advanced web scrapping process, utilizing tools like [Selenium](https://www.selenium.dev/)¹ to aggregate and curate content from reputable sources worldwide. The following websites were scraped to ensure comprehensive and up-to-date coverage of Quantum Technologies standardization news:

- International Electrotechnical Commission (IEC)
- International Organization for Standardization (ISO)
- ASTM International
- Internet Engineering Task Force (IETF)
- European Telecommunications Standards Institute (ETSI)
- Association Française de Normalisation (AFNOR)
- British Standards Institution (BSI)
- National Institute of Standards and Technology (NIST)

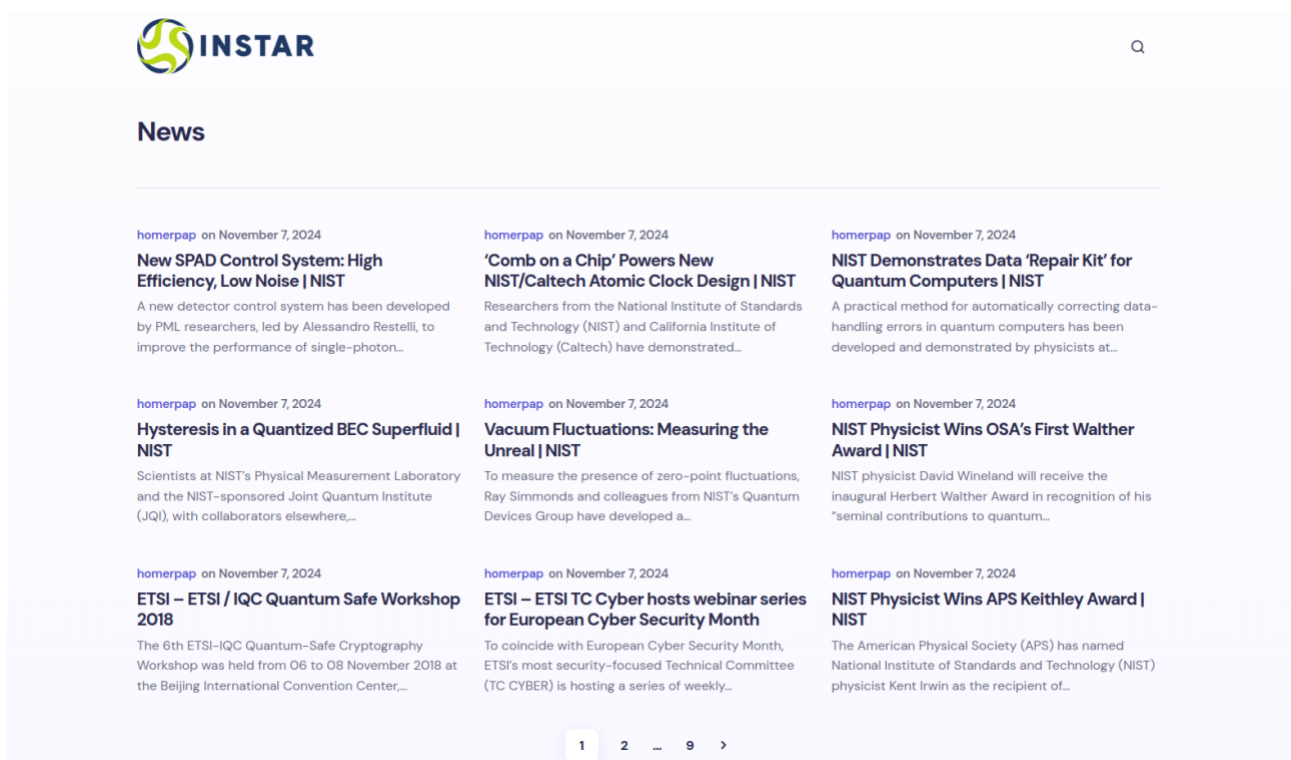


Figure 3: News Repository

¹ <https://www.selenium.dev/>

Web Scraping Process

Automation with Selenium: Selenium was employed to automate the navigation and interaction with the websites listed above. This includes accepting cookies, clicking through pagination, and extracting relevant HTML elements containing news articles, announcements, and updates.

Data Filtering: The scraped content was processed to filter out irrelevant information and ensure the inclusion of high-quality, credible news. Articles and reports related to Quantum Technologies standards, emerging technologies, and advancements in cybersecurity were prioritized.

- To achieve this, the following agents were employed:
- **Researcher:** Responsible for determining the relevance of the provided text to quantum standards or related technical developments. The agent ensured content alignment with the specified keywords and delivered a deterministic "yes" or "no" answer regarding relevance.
- **Reporting Analyst:** Tasked with cleaning and formatting the content by removing non-relevant elements (e.g., disclaimers, headers, footers) and ensuring the final output included only well-structured paragraphs focused on quantum standards and related updates.
- **Content Analyzer:** Focused on matching the cleaned content to 1-3 relevant categories within the domain of quantum standards and technology. Additionally, this agent extracted 3-5 keywords to facilitate improved tagging and classification.

A keyword list was utilized for this classification, encompassing core topics like:

- **Quantum Computing and Algorithms**
- **Quantum Cryptography and Cybersecurity**
- **Quantum Communication and Networks**
- **ICT Standardization and Standardization Bodies**
- **Quantum Sensing and Metrology**

Quantum Use Cases and Applications

These keywords ensured the accurate identification and prioritization of content related to quantum standards. Similar methodologies will also be applied to classify and filter content for other categories, including **Cybersecurity, Edge Computing and IoT, 5G etc.**



Figure 4 : News Article

By leveraging this automated process, the **Standards Tracer Portal** eliminates the need for users to manually browse multiple sources for information. Instead, they are presented with a continuously updated stream of news tailored to their professional needs. Examples of content include announcements about new standards, advancements in interoperability, and updates on cybersecurity policies. This section ensures that users remain informed about the latest trends and developments, empowering them to stay ahead in the rapidly evolving field of Quantum Technologies standardisation.

5 Database Construction for Standards via Web Scraping

The objective of this activity is to construct a comprehensive database of standards extracted from two major websites: ETSI (<https://portal.etsi.org>) and ISO (<https://www.iso.org/obp/ui>). This database acts as a knowledge base for an AI assistant, enabling users to query and interact with the content of the standards. The scraping process employed different techniques tailored to each website, leveraging tools such as [BeautifulSoup](https://www.crummy.com/software/BeautifulSoup/)² for ETSI and Selenium for ISO.

5.1 Scraping Process

The ISO website employs dynamic content loading via JavaScript, necessitating the use of Selenium for automated interaction and scraping. To begin, the Selenium driver loads the ISO search page, and cookies are accepted to bypass the consent dialog. The search is initiated by selecting the "Standards" radio button and clicking on the search button. After that, the results are sorted by date using a designated sort button. Each standard's reference and details are accessed by clicking on the result entries.

For metadata extraction, key fields are gathered directly from the page. The URL is extracted from the current browser address, which provides a unique link to each standard's details. The reference number is retrieved from the title section of the page and serves as a unique identifier for each standard. The title and content are parsed from specific HTML elements on the page, which often include short descriptions or abstracts of the standards. In addition, more in-depth metadata such as publication status, publication date, and the ICS number are collected. These fields are crucial for understanding the status and categorization of each standard. Once the initial results are obtained, reference-based searches are performed to retrieve any additional metadata not immediately visible, such as more specific details about the standard's publication or revisions. By navigating through the website with Selenium, the scraper collects this comprehensive data, ensuring that the information is up-to-date and accurate. During this process, a total of **25,894 ISO standards** were downloaded by scraping **2,590 pages** of results.

The second scraping process for ETSI is more straightforward due to the simpler structure of the website. Unlike ISO, which requires handling dynamic content and JavaScript, ETSI's website primarily consists of static HTML, making it easier to parse. Using BeautifulSoup for this task, we extract the necessary metadata by parsing specific HTML elements, such as tables and links. We retrieve metadata such as the creation date, technical body, document number, and title directly from the static content, avoiding the need for complex interactions or waiting for JavaScript to load content.

Additionally, the extraction of PDF files is streamlined, as all necessary links are available in the HTML source. The PDFs are directly accessible through anchor tags with .pdf extensions, making the downloading process as simple as parsing the href attributes and saving the files. These files contain detailed information about each standard, including **Scope, References, Definition of Terms**, and the full analysis of the standard. As a result, **48,309 ETSI standards** were successfully downloaded during this process. This simplicity significantly reduces the complexity of the scraping process compared to ISO, where dynamic page elements must be handled with Selenium. By relying on BeautifulSoup and static content, the task becomes much easier and more efficient, leading to faster processing times for collecting the required standards metadata.

² <https://www.crummy.com/software/BeautifulSoup/>

6 Standards Chatting Tool

The Standards Chatting Tool combines vector search capabilities with multiple Retrieval-Augmented Generation (RAG) approaches to enable processing and retrieval of standards that we have scrapped in the previous steps. This system has access to 48,309 webpages from ETSI and 25,894 webpages from ISO through a dual-storage architecture that incorporates both traditional file-based document storage and graph-based knowledge representation. Users can use this tool to perform tasks such as retrieving specific standards or asking complex questions to gain deeper insights into certain topics. For example, an engineer might query, "What are the key 5G security standards published by ETSI? and relate with my antenna product" or "which is the standard I should apply to connect my Key Management System-KMS with the QKD system?". Similar a policy maker can query "List all the ISO standards apply to quantum networks?", "how many new standards ETSI released during the previous year for AI". The chatbot provides fast and detailed responses, making it an invaluable resource for navigating the extensive standards database.

The core architecture integrates hybrid vector search methodology that combines dense retrieval, sparse retrieval, and cross-encoder reranking to optimize search accuracy. This approach enables both semantic understanding and precise keyword matching across large document collections. The system's document storage capabilities are implemented through two distinct approaches: a File Document Store utilizing hybrid vector search, and a [GraphRAG](#)³ Document Store that leverages graph-based relationships for enhanced context awareness.

Access control mechanisms operate at both group and individual levels, enabling precise management of document access permissions. The system supports multiple RAG implementations, including Simple, [Rewoo](#), [ReACT](#), and Custom approaches, providing flexibility in handling various query types and complexity levels.

As a result, this tool can help focus on specific clusters within the Terms and Conditions of various standards and documents. By using advanced search and AI-driven capabilities, the portal can identify, extract, and categorize key sections of Terms and Conditions relevant to particular standards or domains, making it easier for users to analyse and understand the specific obligations, rights, and restrictions associated with each standard.

6.1 Search Infrastructure

The search functionality operates on a hybrid vector search architecture that integrates three search methodologies:

[Dense Retrieval](#)⁴ using vector search performs similarity search in high-dimensional spaces. This method captures semantic relationships between documents and queries, enabling understanding of contextual meaning beyond keyword matching.

[Sparse Retrieval with BM25](#)⁵ (Best Matching 25) algorithm provides keyword-based search capabilities. This method maintains precision for queries where exact keyword matching is necessary, particularly for technical documentation where specific terminology requires exact matches.

³ <https://arxiv.org/abs/2404.16130>

⁴ https://sbert.net/docs/cross_encoder/pretrained_models.html

⁵ <https://huggingface.co/blog/xhluca/bm25s>

Cross-encoder⁶ Reranking processes the final stage, performing detailed comparison between query and candidate documents. The implementation uses BERT⁷-based models to consider contextual relationships between queries and documents.

6.2 Document Storage Architecture

The system implements two document storage approaches:

File Document Store integrates with the hybrid vector search infrastructure. This storage solution handles traditional document retrieval scenarios where content matching and semantic similarity drive the search process. The implementation uses hybrid search capabilities for document retrieval.

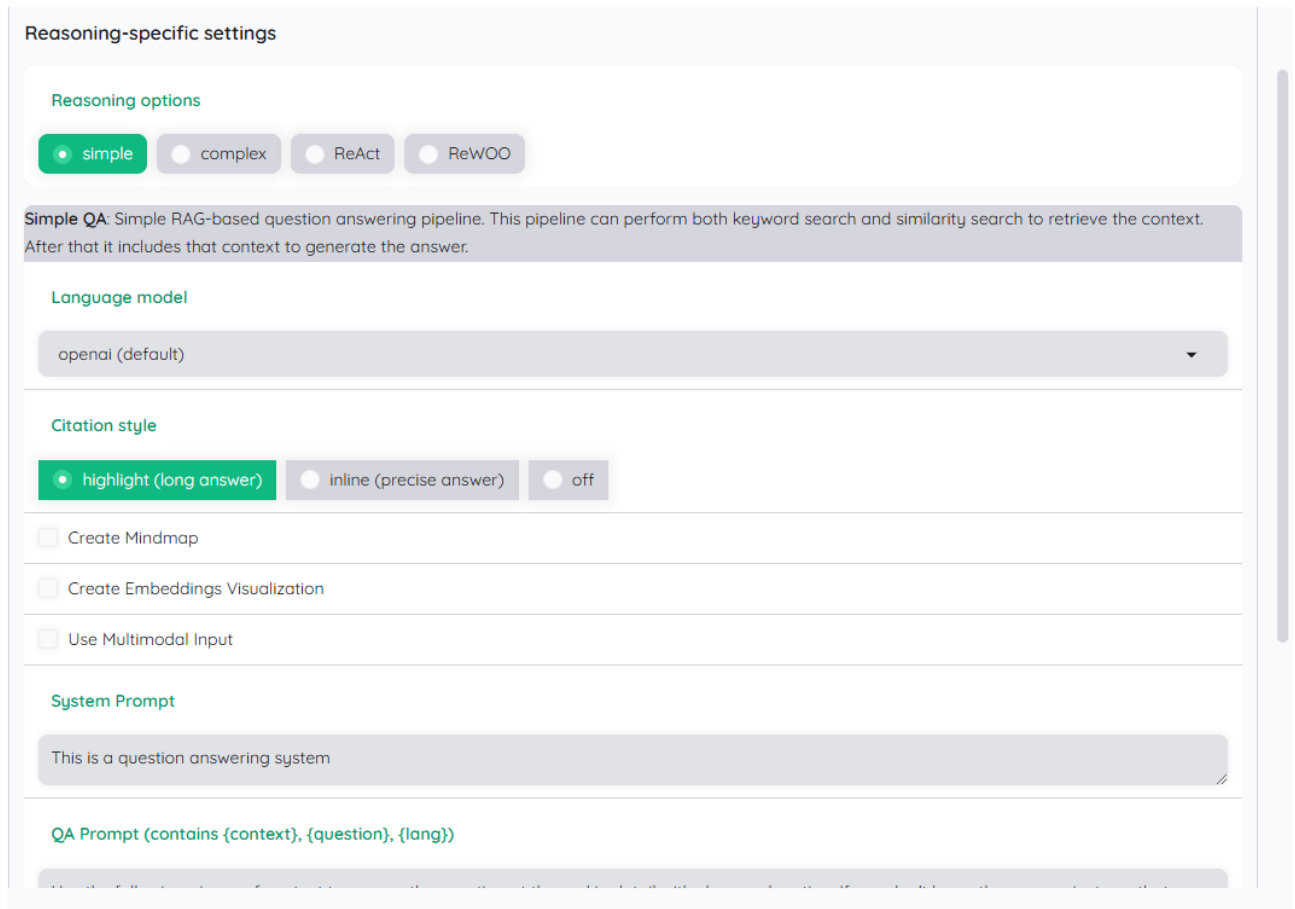
GraphRAG Document Store implements graph-based retrieval augmented generation. This approach creates interconnected knowledge graphs from documents, enabling query resolution through document relationships and contextual connections. GraphRAG traverses these relationships during query processing.

⁶ https://sbert.net/docs/package_reference/cross_encoder/index.html

⁷ <https://arxiv.org/abs/1810.04805>

6.3 RAG Implementation

The system includes multiple RAG approaches:



Reasoning-specific settings

Reasoning options

☒ simple
 ☐ complex
 ☐ ReAct
 ☐ ReWOO

Simple QA: Simple RAG-based question answering pipeline. This pipeline can perform both keyword search and similarity search to retrieve the context. After that it includes that context to generate the answer.

Language model

openai (default) ▼

Citation style

☒ highlight (long answer)
 ☐ inline (precise answer)
 ☐ off

☐ Create Mindmap
☐ Create Embeddings Visualization
☐ Use Multimodal Input

System Prompt

This is a question answering system

QA Prompt (contains {context}, {question}, {lang})

Figure 5 : Rag approaches

Simple RAG handles document retrieval and response generation for basic query-answer scenarios where direct information lookup meets the requirements.

Rewoo RAG implements a retrieval workflow that processes document selection and response generation through iterative refinement. This approach addresses complex queries requiring information synthesis from multiple sources.

ReACT RAG incorporates reasoning and action steps in the retrieval process. This method resolves queries through structured thinking steps, handling scenarios that require logical deduction or multi-step reasoning.

Custom RAG frameworks can be implemented for specific requirements, allowing tailored retrieval and generation strategies based on document collections or query patterns.

6.4 Data Integration and Security

Document Processing

The system processes 74,294 webpages from ETSI and ISO and News articles, creating a technical documentation database. This dataset is processed for both storage approaches:

Chat Files **Resources** Settings version: local

Index Collections LLMs Embeddings Rerankings Users

View Add

id	name	index type
1	Rolling Plan	FileIndex
2	Intelligent Analysis of File	GraphRAGIndex
3	News	FileIndex
4	Personal Files	FileIndex
5	File	FileIndex
6	GraphRAG	GraphRAGIndex

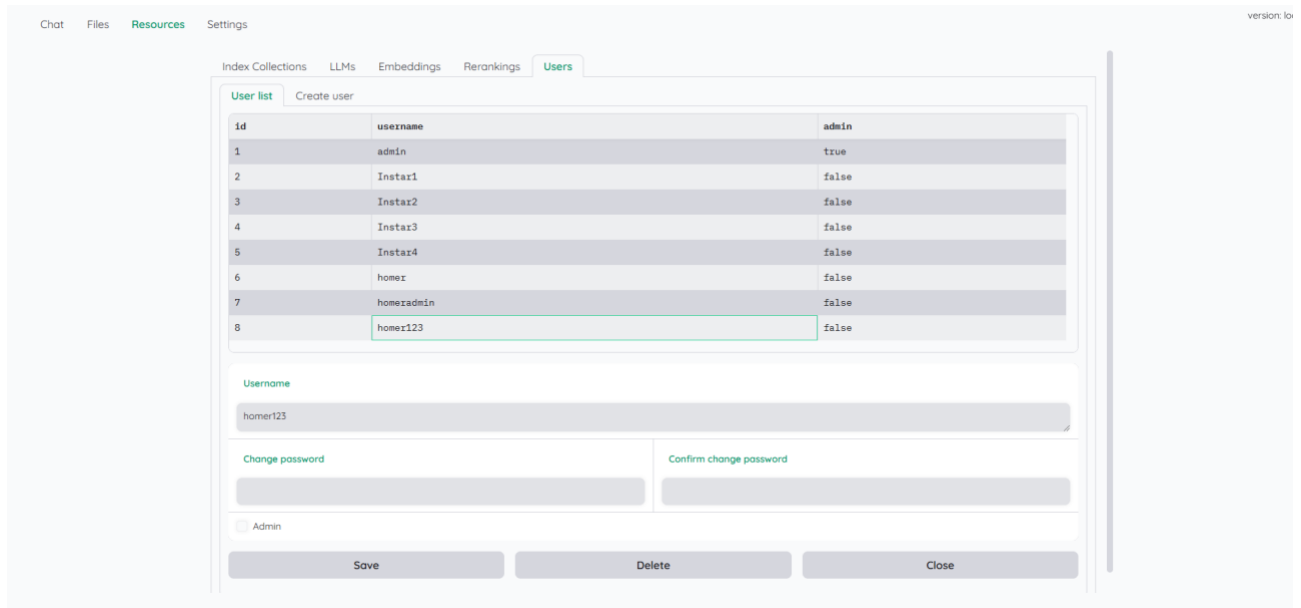
Figure 6 : Available document store and their types

The File Document Store generates dense and sparse representations for hybrid search functionality.

The GraphRAG Document Store analyzes documents to extract relationships and create interconnected knowledge graphs.

6.5 Access Control

The system uses individual level access control:



id	username	admin
1	admin	true
2	Instar1	false
3	Instar2	false
4	Instar3	false
5	Instar4	false
6	homer	false
7	homeradmin	false
8	homer123	false

Username: homer123

Change password:

Confirm change password:

☐ Admin

Save Delete Close

Figure 7: User Control Settings

Users receive access permissions managed at individual levels. The system administrators can create users and assign specific user permissions to each individual user. The users are created by the system administrators for specific individuals and institutions on a request basis, due to the high compute requirements by the system and the compute constraints that exist.

Document access controls operate at group and individual levels levels.

The access control system integrates with both document storage approaches.

6.6 User Interface

The Standards Chatting Tool will be accessed by users through the Instar website. We intend to interface the portal shown in Figure 8 that is made available in <https://infobutton.iwelli.com/> with the INSTAR portal. We will allow users by clicking on a **Chat** button in the navigation bar to be redirected to the Chat tool website and will be able to access the system. Due to compute constraints, selected users and institutions will be provided with login accounts.

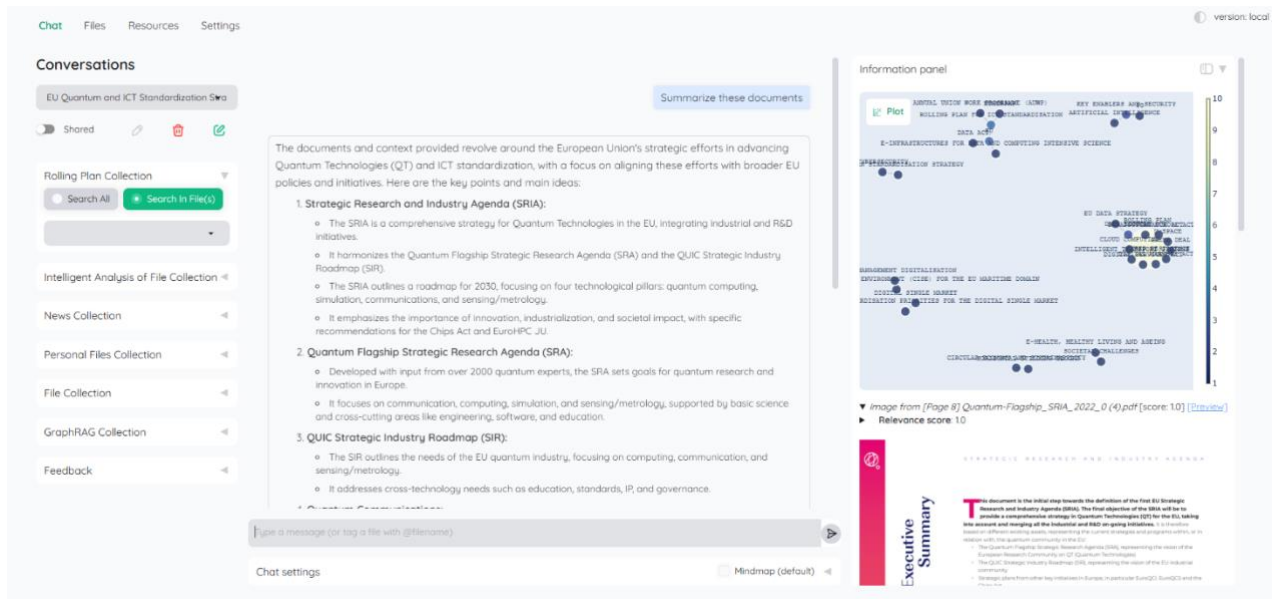


Figure 8 : User Interface of the chatting tool

The system implements an interactive chat interface that enables direct engagement with document collections. This interface serves as the primary point of interaction between users and the document storage systems, providing access to both the File Document Store and GraphRAG Document Store and allowing the user to choose which they prefer to use.

Users can access different document sets through dedicated collection spaces, including the Rolling Plan Collection, Intelligent Analysis of File Collection, News Collection, Personal Files Collection, standard File Collection, and GraphRAG Collection. This organization ensures efficient document management and retrieval while maintaining logical separation between different document categories.

Search functionality within the interface operates through two distinct mechanisms. The primary search capability allows users to query across entire collections, enabling broad exploration of available documents. Additionally, the system supports targeted searching within specific files or document subsets, allowing for precise information retrieval when users need to focus on particular documents or topics.

The information presentation layer consists of multiple integrated components working in concert. The main chat window serves as the primary interaction point, presenting document summaries and relevant excerpts in response to user queries. This is complemented by an information panel that provides additional context through visualizations and supplementary data. The system also maintains entity tables that display relevance scores and detailed descriptions, ensuring users can quickly assess the significance of retrieved information.

Information panel

▼ *Image from [Page 8] Quantum-Flagship_SRIA_2022_0 (4).pdf [score: 1.0] [E*

► **Relevance score: 1.0**

Executive Summary



STRATEGIC RESEARCH AND INDUSTRY AGEND

This document is the initial step towards the definition of the first EU Strategic Research and Industry Agenda (SRIA). The final objective of the SRIA will be to provide a comprehensive strategy in Quantum Technologies (QT) for the EU, taking into account and merging all the industrial and R&D on-going initiatives. It is therefore based on different existing assets, representing the current strategies and programs within, in relation with, the quantum community in the EU:

- The Quantum Flagship Strategic Research Agenda (SRA), representing the vision of the European Research Community on QT (Quantum Technologies)
- The QUIC Strategic Industry Roadmap (SIR), representing the vision of the EU industrial community
- Strategic plans from other key initiatives in Europe, in particular EuroQCI, EuroQCS and the Chips Act.

Starting from the analysis of these documents, the SRIA will propose an implementation path towards concrete actions that could be undertaken in relation with all QT initiatives within EU. The present document is a preliminary version of the SRIA, harmonising the scientific (SRA) and industrial (SIR) strategic agendas. It also provides more specific recommendations about the development of QT within the forthcoming work programmes of the Chips Act, in the area of semiconductors, and EuroHPC JU in the area of high performance computing. A final version of the SRIA, including all EU initiatives, will be published in 2023.

After presenting the methodology in Part A, this document is divided into two main sections:

- In Part B a roadmap for 2030 is presented, merging the main elements of the SRA, SIR, and other relevant documents. Section B.1 is based on the usual four technological pillars (quantum computing, quantum simulation, quantum communications, quantum sensing metrology), and Section B.2 presents the associated transversal activities, including Quantum Resources, Innovation, Industrialization, and Societal Impact. At the end of each section there is a list of short and long term objectives, as well as more specific recommendations.
- In Part C the inputs from Part B are aligned with the specific framework of the Chips Act and EuroHPC JU, and again a list of specific recommendations is given.

Figure 9 : Retrieved Information with sources and relevancy score

Knowledge representation within the interface extends beyond text-based interactions through sophisticated visualization capabilities. The system generates interactive graphical representations that illuminate document relationships and concept hierarchies. These visualizations map connections between related documents and ideas, while also indicating the relative strength of these relationships through relevance scoring mechanisms.

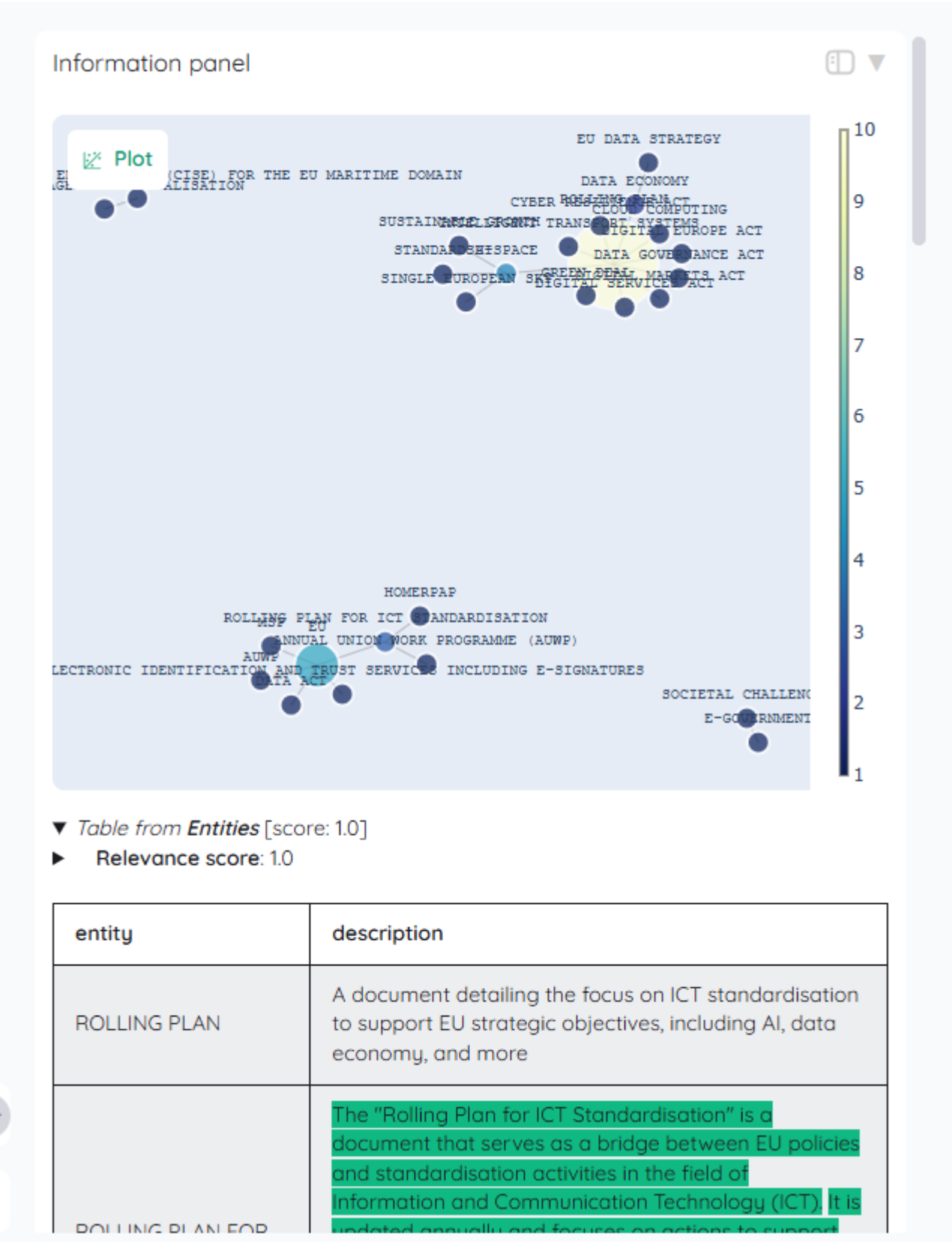


Figure 10 : Retrieved information from a GraphRAG document store with the relevant knowledge graph and table of entities

Query processing occurs through a natural language understanding system that interprets user intentions and retrieves relevant information accordingly. The system processes queries contextually, considering both explicit and implicit relationships within the document collections. Responses include direct information extraction from documents, contextual elaboration based on related content, and clear source attribution to maintain information provenance.

The interface integrates seamlessly with both the hybrid vector search capabilities and the GraphRAG infrastructure. When processing queries, it leverages the full spectrum of search methodologies, including dense retrieval through vector search, sparse retrieval via BM25, and cross-encoder reranking. This integration ensures that responses benefit from both semantic understanding and precise keyword matching.

7 Conclusion and Future Directions

The Standards Tracer Portal is designed to automate and centralize access to standards-related information. Advanced technologies like agentic web-scraping, LLMs, agent frameworks and knowledge graphs aim to deliver updates on relevant standards and technical documents, ensuring stakeholders have timely access to critical information.

A key focus is on developing AI-powered chat functionalities and intelligent reading assistance features, which can evolve into interactive learning modules. These tools will provide stakeholders like SMEs, professionals, students, and newcomers with an accessible and user-friendly way to understand standardization processes within SDOs.

By enhancing AI-driven analytics and LLM models, expanding data sources, and integrating user feedback, the portal will generate personalized newsletters to keep stakeholders updated on the latest trends and developments in standardization.

To ensure sustainability and maximize utility, the portal will introduce collaborative features like shared workspaces to enhance knowledge exchange.

Initially targeting SMEs, the portal will assist them in navigating regulatory and standardization requirements. It will inform SMEs about the technology stack that is more relevant for them to get involved in the standardization process, legal considerations such as IP issues, and the processes involved.

By lowering barriers such as providing access to simplified guidance, the platform aims to open opportunities for SMEs to participate in standardization processes while supporting compliance and risk management by offering automated alerts and proactive guidance on regulatory changes.

The long-term vision of the Standards Tracer Portal is to create a platform that addresses current needs while anticipating future challenges. Additionally, the portal will integrate multimodal LLMs with vision capabilities to assist SMEs in identifying suitable standards and regulations for their products, enabling them to effectively engage with complex technical documents.