STOP-IT

Deliverable 7.2: Validation Plans including the KPIs per demo

KWR
May, 2019
www.stop-it-project.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 740610.

The publication reflects only the authors' views and the European Union is not liable for any use that may be made of the information contained therein.

STOP-IT

D7.2 Validation Plans including the KPIs per demo

SUMMARY

Being part of WP7 that aims at building on-site integration, demonstration and validation activities for STOP-IT, D7.2 has the objective of defining the validation plan, i.e. the methodology that sets the parameters to test and validate the STOP-IT system, both as a single platform product and as an ensemble of tools. The validation methodology of D7.2 focuses on the end user (i.e. FR) experience, which is gained during STOP-IT demo activities on both an individual tool level (the use of a single tool, according to the use cases of D7.1) and a platform level (the experience obtained from the use of the end STOP-IT product as a whole). Key Performance Indicators (KPIs) are included as part of the validation plan in the form of tool and platform *traits*, i.e. indicators of excellence. These traits characterize tool and platform performance, as seen by the end user, and thus allow him/her to evaluate performance based on qualitative questions linked to these traits. Seven main trait categories are identified that are applicable to both tool and platform level, which formulate seven different chapters with questions for the questionnaires addressed to the end users. Moreover, D7.2 provides ways to quantify trait quality performance – in the form of a single metric per trait – and proceeds to link the presented validation methodology with the demonstration pilots, explaining how it is applied in WP7 demonstration activities. Finally, D7.2 links the developed validation methodology with user requirements, which are defined in a previous deliverable (D3.3).

DELIVERABLE NUMBER	WORK PACKAGE
D7.2	WP7
LEAD BENEFICIARY	**DELIVERABLE AUTHOR(S)**
KWR	Dimitrios Bouziotas (KWR)

QUALITY ASSURANCE

Reviewer 1: Ioannis Tsoukalas (ICCS)
Reviewer 2: Nikolaos Bakalos (ICCS)

PLANNED DELIVERY DATE	ACTUAL DELIVERY DATE
31/05/2019	13/06/2019

DISSEMINATION LEVEL	☑ PU = Public ☐ PP = Restricted to other programme participants ☐ RE = Restricted to a group specified by the consortium. Please specify: ________________________________ ☐ CI = Classified Information, RESTREINT UE (Commission Decision 2015/444/EC)

Table of contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

AB: Aigües de Barcelona (STOP-IT Partner)

ARC: Alarms Reception Center

ACSEL: Access Control System using Electronic Locks

API: Application Programming Interface

APT: Advanced Persistent Threats

AVAT: Asset Vulnerability Assessment Tool

BWB: Berliner Wasserbetriebe (STOP-IT Partner)

CP: Cyber-Physical

CPSTP: Cyber Physical Threats Stress Testing Platform

CRFT: Cyber Risk Fault Tree

CTSS: Cyber Threat Sharing Service/System

CVT: Computer Vision Tools

D: Deliverable

DoA: Description of Action

EC: European Commission

EPANET: Environmental Protection Agency's software for water distribution system modeling

ERS: Expert Reasoning System

EU: European Union

EUT: EURECAT (STOP-IT Partner)

EVI: Enhanced Visualization Interface

FCAC: Fine-grain Cyber Access Control

FR: Front Runner

FT: Fault Tree

FTCS: Fault tolerant Control Strategies for Physical Anomalies affecting SCADA system

EVI: Enhanced Visualisation Interface

HPD: Human Presence Detection

ICCS: Institute of Communication and Computer Systems (STOP-IT Partner)

ICT: Information and Communications Technology

IDS: Intrusion Detection System

InfraRiskCP: InfraRisk for Cyber Physical threats

I/O: Input/Output

IP: Internet Protocol

IPS: Intrusion Prevention System

IT: Information Technology

IWM: Interoperability Water Middleware

JDet: Jammer Detector

KPI: Key Performance Indicators

KWR: KWR, WaterCycle Research Institute, NL (STOP-IT Partner)

LCoP: Local Community of Practice

MEK: Mekorot (STOP-IT Partner)

MOT: Multi-objective Optimisation Tool

NTSA: Network Traffic Sensor and Analysers

OPWS: Optimised Public Warning System

PLC: Programmable Logic Controllers

PU: Public

QA: Quality Assurance

RAET: Risk Analysis and Evaluation Toolkit

REN: Reasoning Engine

RGIP: RISA GEN Integration Platform

RIDB: Risk Identification Data Base

RRMD: Risk Reduction Measures Database

RSDP: Real-time Sensor Data Protection

RTAD: Real-Time Anomaly Detector

RE: Restricted

SCADA: Supervisory Control and Data Acquisition

SP: Scenario Planner

STP: Stress-Testing Platform

TCP: Transmission Control Protocol

UDP: User Datagram Protocol

UHF: Ultra High Frequency

UI: User Interface

UX: User Experience

URI: Universal Resource Identifier

VAV: Oslo Kommune Vann-og avløpsetaten (VAV)

WP: Work Package

XACML: eXtensible Access Control Markup Language

XL-SIEM: Cross Layer Security Information and Event Management

Executive Summary

Being part of WP7 that aims at building on-site integration, demonstration and validation activities for STOP-IT, D7.2 has the main objective of defining the validation plan, i.e. the methodology that sets the parameters to test and validate the STOP-IT system, both as a single platform product and as an ensemble of tools. The validation plan that is designed herein will be used during STOP-IT demonstration activities (T7.4) by the Front Runners (FRs) in order to assess, evaluate and quantify the performance of STOP-IT products at platform and tool level, based on the user experience of the FRs.

The validation methodology of D7.2 focuses on the end user (i.e. FR) experience (UX), gained during STOP-IT demo activities on both an individual tool level (e.g. the use of a single tool, according to the use cases of D7.1) and a platform level (e.g. the experience obtained from the use of the end STOP-IT product as a whole). The methodology is based on identifying key parameters (named tool and platform *trait*s) that characterize tool and platform performance, as seen by the end user, and thus allows him/her to evaluate performance based on qualitative questions linked to these traits. Seven main trait categories are identified (Table 2) that are applicable to both tool and platform level, which formulate seven different chapters with questions for the questionnaires. The generated questionnaire templates are given in Annexes B and C. Having questionnaires at tool and platform level, the end users are able to provide validation scores (in a rank from 1.0 to 5.0) for every trait category. These scores can be then aggregated and used to assess the quality of an individual tool or of the platform as a whole, as explained further in Section 3.4. The scores can be further used for relevant reflection activities (e.g. T7.5) and to provide feedback to the developers in order to improve the design of specific tools or platform functions. The validation process is summarized in Figure 8 and its main attributes are outlined in Table 1.

The validation plan proposed in this deliverable is also directly linked with previous STOP-IT documents related to the demonstration activity planning and user requirements. This report is organized as follows: firstly, it expands upon and relates to concepts described in D7.1, where the demonstration activities and use cases for each tool are described in detail. Secondly, important end user requirements, as identified in D3.3 (WP3), are linked to the higher-level trait parameter system used in D7.2. Finally, links are drawn between specific platform traits and societal Key Performance Indicators (KPIs), as defined in the Description of Action (DoA), in order to provide insight on the impact STOP-IT has on the accuracy, preparedness and exposure of FR activities with respect to Cyber-Physical (CP) risk.

Table 1: Overview of the scope and main attributes of the validation plan.

Attribute	Description
Subject (what)	- Validation plan for the activities of WP7. - Questionnaire templates that stem from the validation plan. - Questionnaire templates on two levels: individual **tool** level and whole **platform** level.
Target group (who will fill the questionnaires)	- **WP7 scope**: Front Runners at the end of demonstration activities. - **General scope**: End users of STOP-IT tools and the STOP-IT platform.
Way to fill the questionnaires (how to fill)	- Print questionnaires before each demonstration activity starts. - Fill tool questionnaires only for the subset of tools that have been selected and demonstrated per FR. - Tool level questionnaire (Annex B): **once per demonstrated tool** (e.g. 8 demonstrated tools for a FR → 8 questionnaires filled for that FR). **Exception:** tools that are integral to the platform (see Chapter 5 and Section 5.1). - Platform level questionnaire (Annex C): **once** after the completion of a demonstration activity cycle, to reflect on the use of the whole platform. A demonstration activity cycle ends when all selected tools have been demonstrated and the FRs have completed their experience from the integrated STOP-IT product.
Processing group (who will process the questionnaires)	- Collection: T7.4 demo activity organizers, guided by T7.4 task leaders. - Processing/evaluation/impact assessment: T7.5 partners and task leaders.
Way to process the questionnaires (how to process)	- Filled questionnaires collected as part of T7.4. - Filled questionnaires sent to analysis/reflection tasks, e.g. T7.5. - Analysts choose a relevant metric approach, based on the information contained in Section 3.4, and calculate scores on tool/platform level. - Displaying and comparing scores across trait categories without unifying them to one quantity is suggested. - Results can be then shared with developers (WP6) and dissemination/exploitation tasks.

1 Introduction

The main objective of this document is to design the validation plan, i.e. the approach with which the individual tools as well as the whole platform of STOP-IT will be evaluated following the demonstration activities of WP7. A validation framework based on user experience is formulated that relies on questionnaires both at a tool and a platform level, serving as an integral part of the demonstration activities at the FRs organized in Tasks 7.3-7.4.

As part of the proposed design, the validation process relies on the end users (i.e. FRs), following the experience they obtained from the demonstration activities of Tasks 7.3-7.4. Since, as explained in D7.1, different modules and tools are selected by every FR, validation is based on two levels:

- On the lower level, every individual tool tested by a FR during a demonstration activity is validated by the end users, based on a number of identified parameters which are herein named *tool traits*. Validation on the tool level targets the performance of every one of the tools proposed and developed as part of the STOP-IT platform, based on the activities of WP4 and WP5. Since, as stated in D7.1, all tools are planned to be tested in at least one site, validation at the tool level will capture the entire array of STOP-IT options at strategic, tactical and operational levels.

- On the higher level, besides individual tool use, every FR gains experience in the use of STOP-IT platform as a whole and can thus evaluate it based on his/her expectations and requirements, described in more detail in D3.3. To capture this higher level of user experience, the end user also validates the whole platform based on a number of identified parameters which are, in a similar and methodologically consistent way to the tool level validation, named *platform traits*. Validation on the platform level targets the performance of the platform as a whole, rather than the performance of a specific function which occurs on a tool level.

The basis for validation is in any case the demonstration activities which are planned in the four (4) pilot sites of the FRs: 1) Aigües De Barcelona (AB), 2) Mekorot Water Company Limited (MEK), 3) Berliner Wasserbetriebe (BWB) and 4) Oslo Kommune Vann-og avløpsetaten (VAV). Following the completion of the demonstration activities, participating end users are requested to answer validation questionnaires on a tool and platform level (included in this deliverable as Annexes B and C), targeting the tools that they selected during the demo activities as well as the integral platform experience. The questionnaires are to be included as part of the material provided during T7.4.

Besides evaluating tool and platform performance against specific parameters (traits), the validation plan includes open feedback questions that aim at capturing any issues seen during the demonstration activities, so to provide information on the challenges the end users face while using the STOP-IT platform. The feedback collected can be then used for the reflection process initiated later at T7.5, where the lessons learnt from the validation process will be outlined.

Beyond presenting the validation framework design, this report creates conceptual links between the validation methodology and a number of important deliverables that describe user requirements as well as the expected outcome of STOP-IT. The user requirements, presented in D3.3, are linked to the tested parameters. Moreover, the way they are validated is presented in a similar way to the connection made between tool selection and user requirements seen in D7.1. In addition to this, this report provides the (societal) KPIs defined in the DoA and the parameters (traits) that are evaluated on the platform level, following the experience of the end users.

The remainder of this report is organized as follows: **Section 2** presents the objectives of D7.2, focusing on validation planning. **Section 3** analyses the methodology that is used as a basis for validation planning on both tool and platform level and presents relevant metrics. **Section 4** provides conceptual links between the validation plan and other important STOP-IT deliverables, such as D7.1 (piloting activities) and D3.3 (user requirements). **Section 5** provides templates for the application of the validation plan per FR, while **Section 6** provides the conclusions. With regards to the Annexes, **ANNEX A** provides a summary of the STOP-IT tools, upon which the validation will be based, in a consistent manner with the tool presentation seen in D7.1. Finally, **ANNEX B** and **ANNEX C** provide the questionnaire templates that are used to generate user feedback on a tool and platform level respectively.

Please note that since multiple aspects directly linked to the STOP-IT platform demonstration are not yet finalised at the time of writing, such as the final format of the platform and tools (i.e., part of WP4/WP5 output encapsulated through WP6), slight amendments to this report may occur following the pilot design and execution phase, without of course changing the core of the validation methodology, i.e. the approach based on end user experience that relies on selected model and platform traits. The proposed validation framework is also generic and able to adapt to changes in the tool selections for each FR, as explained in Chapter 5.

2 Objectives

The main objectives of this deliverable are the following:

- To define and elicitate the **validation plan** of the STOP-IT tools and the STOP-IT platform as a whole.
- To propose **validation parameters** that encompass validation planning and are used as the basis for evaluation.
- To **link** the proposed methodology with the requirements of the end users, as described in deliverable D.3.3.
- To include a number of societal KPIs seen in the DoA as part of the parameterization process.
- To **provide validation material** for the forthcoming STOP-IT demos, i.e. evaluation forms that will be used to generate end user feedback.

3 Methodology

3.1 A validation approach based on end user experience

The designed validation methodology for STOP-IT follows an end user perspective as the basis for evaluation, where the end users (i.e. FRs) themselves are allowed to validate, based on their experience, the individual tools and platform as part of the (reflection on) the demonstration of STOP-IT as an integrated platform product. The vehicle for validation thus becomes end user experience, gained from the demonstration activities themselves. This methodology focuses heavily on the concepts of User Experience (UX) and links validation with end user satisfaction, which has been found to be one of the key factors leading to the success of information systems (Kim et al., 2014; Bokhari, 2005; Al-Khaldi & Olusegun Wallace, 1999; Gelderman, 1998).

Without overanalysing semantics, it is important to clarify that, within the context of this report, UX is defined as the integrated process that comprises the end user acquiring, using and reflecting on the use of a product. According to this definition, UX encompasses all aspects of the end users’ interaction with the product (Norman & Nielsen, 2018), including the reflection process that follows product use, as well as the preparation (purchase, licensing and installation) process that precedes actual usage. Having this in mind, it is important to distinguish UX from some frequently-used terms in software testing and validation:

- UX is different from the user interface (UI), which is the front-end environment that the user interacts with. While UI is an important part of the design, it can be considered only one aspect of the UX, which stands for the integrated user experience with the product.

- UX is different from usability, which can be considered – as explained further in Section 3.2 – a quality attribute of the product, covering aspects such as ease and efficiency of use.

To measure user satisfaction, a qualitative approach is designed based on a set of questions targeting the end users. In a similar manner to past literature on information system validation (Lee, 2003; Zviran et al., 2005; Zviran & Erlich, 2003), questionnaires form the basis of validation, so that the end users of every demo activity (i.e. the FRs and generally users of the STOP-IT platform) answer a number of questions that target to assess the quality of the product itself on both tool and platform level. What is defined as *product* in this case is the integral modular STOP-IT platform where tools at both strategical/tactical and operational levels are encapsulated in the form of modules, as demonstrated during the activities of WP7. An overview of the tools and corresponding modules that are considered as products in this study is given in ANNEX A: Brief Description of the STOP-IT Tools.

3.2 Definition of quality parameters for the STOP-IT tools and platform

Having clarified what a *product* refers to for the STOP-IT platform in Section 3.1, a question that follows is what constitutes *product quality* and how an approach that targets quality can be developed. There is a multitude of ways with which stakeholders perceive product quality, all based on the presence – or absence – to a certain degree of a set of attributes (Laplante, 2007). Following the formal definition seen in ISO 8402:1994 *Quality Management and Quality Assurance – Vocabulary Standard* (ISO/TC 176, 1994) what is defined as quality in this document is:

[…] *the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs.*

The ending part of this definition – the ability to satisfy stated or implied needs – is guaranteed by following a UX-based approach, as stated in Section 3.1. What is also evident in this definition is the link of quality with a number of features and characteristics, inherent to the product, which constitute *behavioural attributes* of the products (Voas & Agresti, 2004) that can be experienced by users throughout the product usage. Defining this set of behavioural attributes – or, in other words, quality parameters - then becomes a crucial step in any quality assessment study and many studies are dedicated to outlining the most inclusive, representative attributes for their product types (Yahaya et al., 2008; Voas & Agresti, 2004; Jamwal, 2010; Dromey, 1995).

Following the same approach, this study identifies a number of quality parameters which can be seen as tool and platform *traits*, i.e. desirable properties that the product should have to satisfy end users. The definition of a trait is borrowed by the school of virtue ethics (Van de Poel & Royakkers, 2011), as a product analogue of virtues, i.e. indicators of person excellence. This approach facilitates the connection of these (functional and non-functional) high-level product attributes with the end user desirability (which is the evaluator of excellence in a UX-based approach) and allows this linked concept to be presented in a form of virtue qualities that the developers should strive for in their products. Drawing upon past literature (Yahaya et al., 2008; Voas & Agresti, 2004; Jamwal, 2010) as well as relevant deliverables (notably D3.3), seven (7) main trait categories are identified in this report, along with an array of partial characteristics per category, seen in Table 2. Having *traits* provides high-level attribute classification which is important to facilitate discussion on quality (Jamwal, 2010; Dromey, 1995), while *partial characteristics* provides a lower-level elaboration on the general trait concepts that facilitates the design of questions and metrics. These traits are:

- The **ease of installation (IN)** a product has, which includes partial characteristics such as:
    - the (in-)dependence of the product to third-party platforms and operating systems
    - the (in-)dependence of the product to third-party and their accessibility to the end user
    - the installation process (format of the installer/wizard)

  - the existence of documentation or guidance material for the installation process
- The **facilitation of user learning (L)**, which includes partial characteristics such as:
  - the availability of learning material (tutorials, documentation, featured examples), as seen by the end user during the demonstration
  - the learning curve and time investment needed by the end user to become acquainted with the product
- The **data requirements (D)** of a product, which includes partial characteristics such as:
  - the amount and form of data needed (i.e. common vs. more esoteric formats that are less available)
  - the source availability of these data, defined by how easy it is for the end users to obtain the data (e.g. from their own line of work or through open data/repos)
  - the degree of data preparation needed and the dependence of data to pre- and post- processing modules and third-party tools
- The **support (S)** of a product, which focuses on materials available to solve problems throughout the usage cycle of the product (rather than the initial learning phase, which is the focal point of the **facilitation of user learning** trait). Support includes partial characteristics such as:
  - access to support resources, i.e. troubleshooting guides or a wiki
  - access to live support, e.g. through a forum or live (e-mail or telephone) correspondence
- The **integrity (IG)** of a product, which includes structural aspects that ensure smooth tool runtime, such as:
  - the stability of the product (i.e. existence or absence of bugs, crashes, lag and generally unexpected behaviour)
  - the interoperability with third-party software/hardware, including how interoperable the product is with other STOP-IT platform components. Unlike product dependence during installation, interoperability aims to evaluate how smooth it is for the end user to connect different tools together through shared I/O protocols
  - the overall reliability of the product in persistently generating the results the user would like to have
  - the level of security a product has, i.e. the protocols and encryption technology used by the product, in case security is a core functional requirement of that particular product
- The **usability (UB)** of a product, which includes partial characteristics such as:
  - the conceptual clarity and simplicity (or, vice versa, complexity) of the product and the general intuitiveness it has
  - the design and functionality of the user interface (UI), which should be in line with the requirements of end users
  - the overall aesthetics and visualization of the product, as experienced by the product UI or frontend
- The **usefulness (UF)** of a product, which includes partial characteristics such as:
  - the importance of the product in answering STOP-IT research questions and the general requests of the end users regarding Cyber-Physical (CP) system assessment

- the ability of the product to be (re-)used and generally be useful in the operational context of the end user (FR)

What is noted is that these traits refer to both (functional and non-functional) intrinsic properties of the product itself, as well as attributes that are integral to the product life cycle (installation, usage and disposal) and important to the user experience. Moreover, they have been shaped to not only encompass software but also hardware technologies that are featured as STOP-IT products. What is also notable is that this study makes a distinction[1] between *usability*, i.e. the design elements that facilitate use (Dromey, 1995), and *usefulness*, i.e. the frequency of use and importance of the tool for the end users, as reflected by them. This distinction is important to identify cases that are rich in usability features but lack the applicability in the operational context of the FRs. Since usefulness focuses on how useful the tools are for the operations of the FRs, regardless of their design features, this trait is particularly important to assess if the product or platform fulfils the operational demands of the FRs in terms of CP risk assessment.

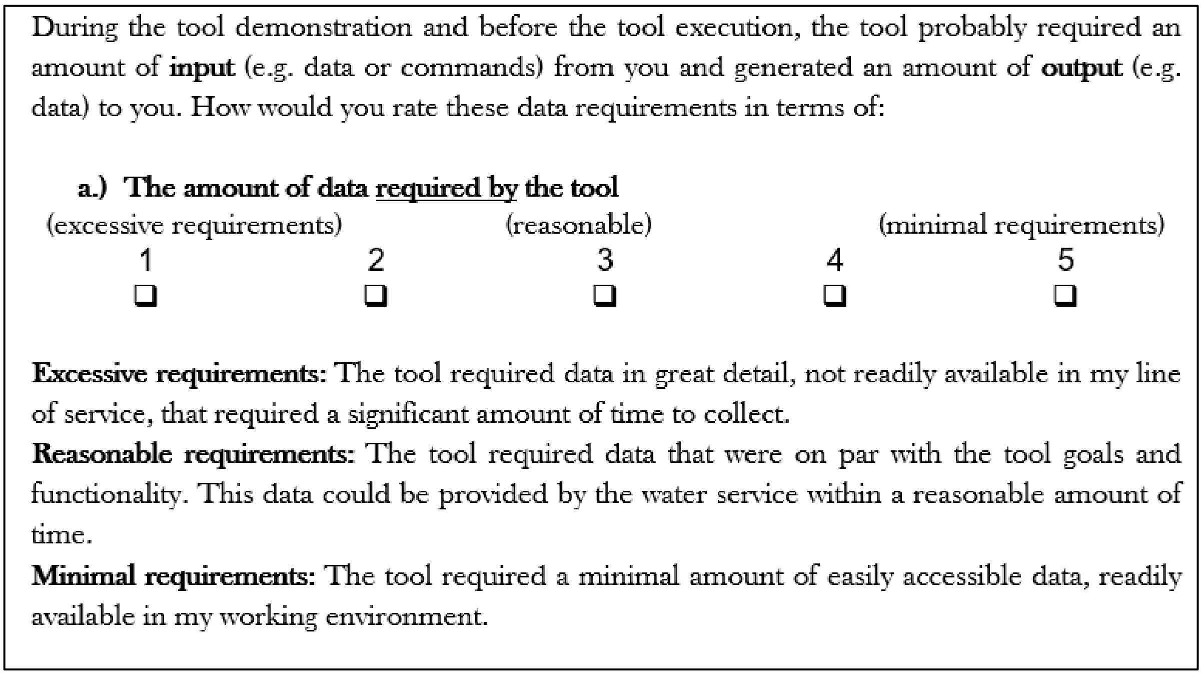
Table 2: Overview of the product traits, along with their partial characteristics.

Traits	Partial characteristics
ease of installation (IN)	- System (in)dependence (cross-platform) - Dependence on third-party software and their accessibility (open-source / commercial) - Dependence on third-party library and modules and their accessibility (open-source / commercial) - Installation process and help (wizard vs. manual, documentation/installation guides)
facilitation of user learning (L)	- availability of learning material (tutorials, documentation, examples) - time investment needed from end users - learning curve
data requirements (D)	- amount of data needed - form of data needed (common vs. esoteric formats) - source availability (open data vs. commercial repos) - data preparation / dependence of data to pre- and post-processing (other tools)
support (S)	- access to support resources (troubleshooting guides, wikis) - access to live support (forum, e-mail correspondence)
integrity (IG)	- stability (bugs/crashes) - interoperability with software/hardware, interoperability of components with each other - reliability - security protocols, encryption etc.

[1] More information on this distinction is available in these sources: [1] and [2].

usability / ease of use (UB)	- conceptual clarity/structural simplicity - interface design (less is more) - intuitiveness - tool aesthetics and visualization
usefulness (UF)	- importance of product in answering STOP-IT research questions - ability of the product to be (re)used in the context of the FR

The next step that follows the definition of the product traits is the evaluation method, i.e. finding a way to measure each trait. Similarly to past works on quality validation that focus on end user evaluation (Jamwal, 2010; Kim et al., 2014), the proposed validation approach focuses on surveys that target the end users and are answered at the end of each demonstration activity. End users are requested to answer a set of questions, which are formed based on the partial characteristics each trait has; through these questions, the end users give a qualitative ranking (on an ordinal scale from worst (1) to best (5)) for these characteristics. Since the ranking is qualitative it may differ from question to question, but always scales to the excellence of the specific partial characteristic and therefore trait. To facilitate questionnaire understanding, a brief text is also provided that explains the different ranks. Figure 1 shows this concept in action for the partial characteristic of *amount of data needed* for the product, which belongs to the parent category-trait of *data requirements*. Some partial characteristics also include conditional and open questions, in order to provide the necessary feedback to improve model design and identify issues in products that need to be addressed, as seen during end user experience. Through this categorization, questions are grouped together to provide insight on the parent trait categories, so the questionnaire has in total seven (7) sections with a group of questions within each section.



During the tool demonstration and before the tool execution, the tool probably required an amount of **input** (e.g. data or commands) from you and generated an amount of **output** (e.g. data) to you. How would you rate these data requirements in terms of:

a.) The amount of data <u>required by</u> the tool

(excessive requirements)		(reasonable)		(minimal requirements)
1	2	3	4	5
❑	❑	❑	❑	❑

Excessive requirements: The tool required data in great detail, not readily available in my line of service, that required a significant amount of time to collect.
Reasonable requirements: The tool required data that were on par with the tool goals and functionality. This data could be provided by the water service within a reasonable amount of time.
Minimal requirements: The tool required a minimal amount of easily accessible data, readily available in my working environment.

Figure 1: Example of a question targeting a specific partial characteristic in a trait category.

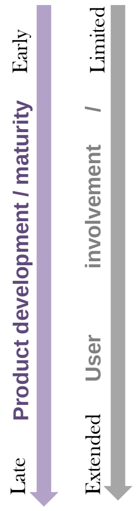
3.3 Linking validation with demonstration activities

Following the definition of product quality parameters, it is important to identify and analyse how these parameters are going to expose themselves to the end user through the demonstration activities. While an exhaustive presentation of each demo and tool use case would be outside the scope of this report (and is treated in D7.1), having a higher-level abstraction of the demonstration activities in the form of a typology would be useful to outline links between validation and demonstration activities, especially considering the user involvement and experience. Regardless of the use case design details for each product, one could deduce that there are generally three (3) types of demonstration activities for STOP-IT tools from a user involvement perspective:

- **Type I** activities, which offer short, top-down demonstrations of a tool to the FRs, where the FR does not gain hands-on experience. For instance, a Type I demo activity could be a short presentation or a webinar.
- **Type II** activities, which offer pilot applications of a tool to the FRs with a simple case (e.g. a toy or training model) and simple, premade data. For instance, a Type II demo activity is a workshop or tutorial that follows the installation of a new technology.
- **Type III** activities, which offer more extended applications of a tool to the FRs with real cases. In this case, the FR co-design a case based on their data and the tool has to be demonstrated against it; this type of activity provides a more extensive hands-on experience for the end user but comes at a higher learning intensity and cost of time.

The ranking of these activities is done in order of ascending user involvement and experience gain (with Type III providing the most in-depth, hands-on demonstration experience) and is summarized in Table 3. These demonstration types are also implicitly linked to the maturity of each product and the platform as a whole, as Type I activities tend to be used through earlier phases (e.g. to present a concept or preliminary design), while Type II/III activities require the product to at an advanced or, ideally, completed design phase.

Table 3: Typology of demonstration use cases.


Late Product development / maturity Early
Extended User involvement / Limited

Type	Demonstration use case - Description
Type I	- A display/top-down demonstration of a case to the FRs - No hands-on user experience and direct engagement from FRs **Examples**: webinar, presentation
Type II	- a pilot application in the FR, with a simple model (e.g. a toy model) and simple data - data is premade and fit-for-use for the model - FRs get limited hands-on experience **Examples:** workshop, tutorials
Type III	- an application of a real case with the model - the case/data is now fit for the FRs and the tool has to adapt to them - FRs get extensive hands-on experience - comes at higher learning intensity / cost of time **Example:** real case demo

Based on the information provided by the extended use cases description of D7.1 (Chapter 5), it can be inferred that demonstration activities in STOP-IT are active events where the end user is fully involved and gains extensive hands-on experience as the tool is being applied in his/her operational context; this means that activities of WP7 will gravitate towards Type II/III demonstrations, with the Type I demonstration types having been reserved for the earlier phases of product development seen in WP4 and WP5 in order to provide quick updates to the FRs. In that case of more active demonstrations, a number of distinct phases are commonly found (Figure 2) that include:

- The installation of the product (tool or platform) to the FR environment. This is the first phase where issues of software and hardware dependencies are solved and the user acquires a first view on how easy it is to install the tool in his/her operational context. This is apparently the domain where the *ease of installation* trait becomes dominant.
- The access and execution of a tool or product. This is the active phase where the tool is being demonstrated and the user gains hands-on experience, possibly with the aid of experts or developers throughout the demonstration activity. Multiple product traits play a major role in the experience of this phase, including the *data requirements*, *integrity* and *ease of use*.

- Finally, the user completes the demonstration activity, reflects on the demonstrated and taught material and proceeds to the validation of the product. This is the reflection phase, where the user can infer, based on his experience, if the demonstrated product is useful for his/her line of work, thus providing input on the *“usefulness”* trait.

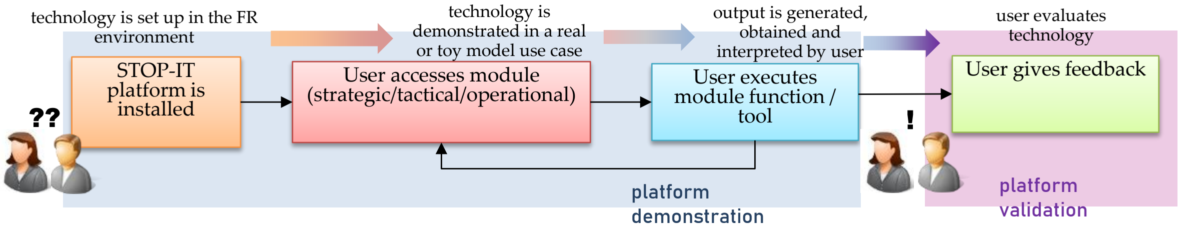

technology is set up in the FR environment
technology is demonstrated in a real or toy model use case
output is generated, obtained and interpreted by user
user evaluates technology
STOP-IT platform is installed
User accesses module (strategic/tactical/operational)
User executes module function / tool
User gives feedback
??
!
platform demonstration
platform validation

Figure 2: The demonstration phases in the case of Type II/III demonstration events.

Based on these phases from the end user perspective, it becomes evident that: (a) several of the traits are directly linked to one action or a group of actions within the demo activity, (b) other traits (e.g. “facilitation of user learning”) act as a catalyst throughout the user experience process, thus making the progression between actions smoother and quicker, (c) “usefulness” is a context-dependent attribute that comes from the reflection after the completion of the demonstration activity. These links between traits and the demonstration activities are depicted in the user action diagram of Figure 3.

While the link between product traits and the demonstration is straightforward on a tool level (Figure 3), the encapsulation in the form of a STOP-IT platform (WP6) adds a layer of complexity and creates a composite system (Voas & Agresti, 2004), as it encompasses WP4 and WP5 tools under a common modelling framework and UI, which is accompanied by a set of platform-specific functions such as the reasoning engine and the Enhanced Visualization Interface (EVI). An extra layer of understanding, learning and viewing results on a whole platform level is then added in terms of user actions. In any demonstration activity that features the STOP-IT platform, the end user has to access a specific tool (e.g. through the corresponding module), run the tool, interpret tool results, and possibly return these results to the platform level and use them for another tool (Figure 4). A platform-level demonstration activity can thus include a single tool, a series of tools (in the form of a modelling chain, with interconnected I/O) or a set of non-connected tools. Likewise, following the platform demonstration, the end user reflects on the usefulness of the platform as a whole, besides the individual tool reflection. To capture this added layer of complexity, two questionnaire templates are produced in this study: one that addresses validation at an individual tool level and one that addresses validation at a platform level. Notably, there is no difference in the attributes or partial characteristics (Table 2) between the levels however, as both levels constitute information systems and the general principles discussed in Section 3.2 hold for the platform as a whole as well as the individual tools. There is, however, difference in some of the questions asked, in order to highlight different aspects of the systems especially with regards to usefulness and also capture the interoperability of the tools with each other at a platform level.

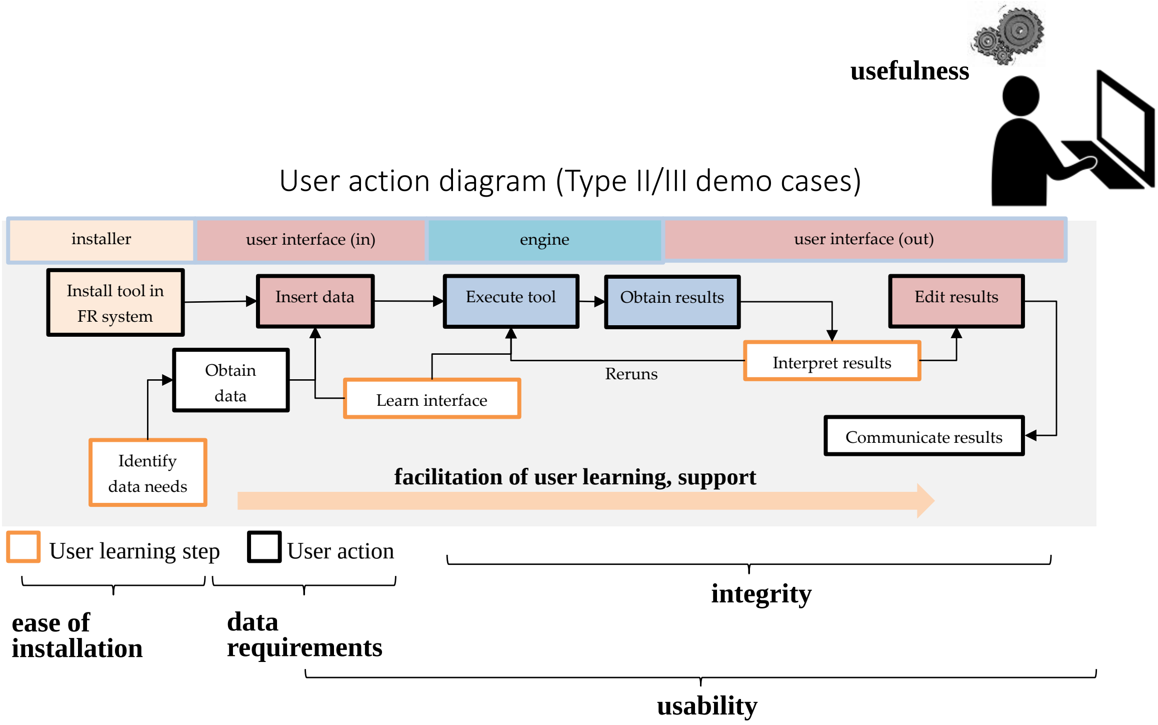


Figure 3: User action diagram in the case of Type II/III demonstration activities.

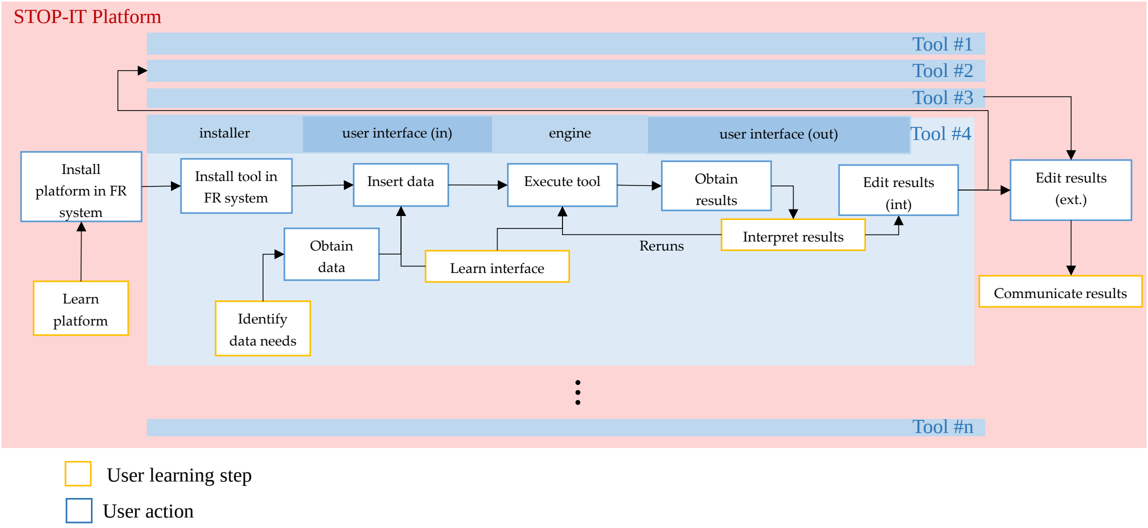


Figure 4: User action diagram in the encapsulated case (STOP-IT platform demonstration).

3.4 Quality assessment at multiple scales

3.4.1 Quality metrics

Quality is approached in this validation plan as a set of key attributes (i.e. traits) which the individual tools as well as the platform aim at having. Based on the ranking questions (which are in turn derived from the partial characteristics of Table 2), each trait can be given a specific score in the range of 1 (low) to 5 (highest). A tool (or the whole platform, for the platform level questionnaires) can thus have scores for each of the considered traits (Table 2), which can be then either presented separately or combined together to have a single pass/fail metric. While a linear combination of these scores yields a single metric for the tool or system, it is generally good practice to also display the performance of the tool and platform at an the level of individual attributes (Voas & Agresti, 2004), in order to reveal strong and weak aspects of the product character that provide useful feedback for design improvements.

The validation plan offers two ways to evaluate and demonstrate validation results, depending on the scope of the evaluation:

1.) At an individual attribute level for the tool and/or platform, by sharing the score of each one of the seven identified traits. The stakeholders can then evaluate the performance of each individual trait separately and compare them, in order to identify where particular tools performed well and where more improvements on the design could be made. Results of traits could be communicated concisely and efficiently with the right visualization, for instance with a web graph such as the one seen in Figure 5, where individual trait scores have been normalized (e.g. a score of 4/5 has been converted to 0.8).

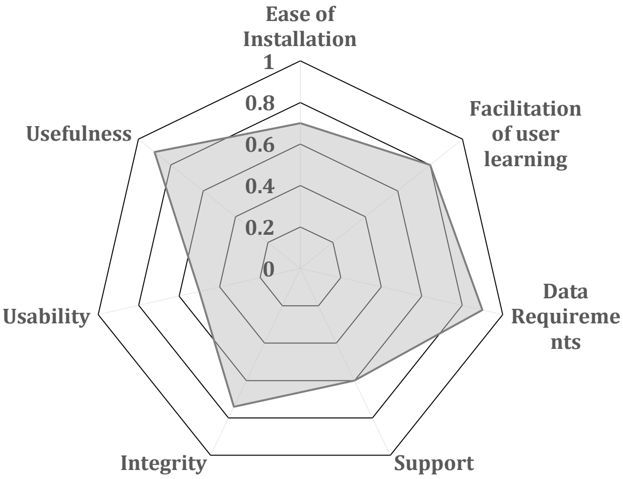


Figure 5: Validation results plotted at an individual attribute (i.e. trait) level.

2.) At an aggregate level, in order to provide a scalar quality metric for the whole tool or platform. This is typically done by calculating the weighted mean score for all attribute (trait) categories (Voas & Agresti, 2004):

$$S_t = \frac{1}{k}\sum_{i=1}^{k} w_i X_i^{(t)} \qquad (1)$$

where $S^{(t)}$ is the aggregate score value of the tool t , k is the number of traits, $X_i^{(t)}$ is the individual score of the trait *i* for the tool *t* and $\boldsymbol{w} = \{w_1, w_2, \dots, w_k\}$ is an appropriate weight vector with its elements having the sum of 1.0. Choosing an appropriate weight vector is empirical, with the choice being dependent on the scope of validation, the type of product and the relative priority stakeholders give to different attributes (Voas & Agresti, 2004), with some systems focusing more on aspects of security and performance (and thus integrity) and others focusing more on usability. This plan offers two options to choose the weight vector:

a.) Choose equal weights, in order not to undermine the contribution of a single attribute towards the final score. This gives a fair weighting scheme across all attributes and the aggregate score S_t is thus simplified to:

$$S_t = \frac{1}{k}\sum_{i=1}^{k} X_i^{(t)} \qquad (2)$$

b.) Choose to magnify functional and non-functional attributes that depend on the tool architecture (IN, D, IG, UB, UF) while at the same time reducing the impact of attributes that relate to the learning and support experience of the end user (L, S). This distribution is recommended if, through validation, one would like to obtain a view of the tool scores whose focal point is on issues of tool architecture and functionality, rather than tool support and documentation. In that case for instance, a weight distribution similar to the one depicted in Table 4 could be employed.

Table 4: Weight distributions for the aggregate scoring case.

Cases	IN	L	D	S	IG	UB	UF
Equal weights	1/7	1/7	1/7	1/7	1/7	1/7	1/7
Focus on tool architecture	0.18	0.05	0.18	0.05	0.18	0.18	0.18

In the case of the aggregate score on a tool or platform level, a pass/fail score threshold could be also set up to quantitatively evaluate total system performance (e.g. with 2.5/5.0 being a reasonable threshold choice). However, having a strict quantitative threshold is not recommended; instead, it is better practice to evaluate tool and platform performance by interpreting, discussing and reflecting on performance at an individual attribute (trait) level (e.g. through Figure 5), where strong and weak domains of the STOP-IT products can be outlined, rather than rely on binary aggregate metrics as sole indicators of quality.

3.4.2 Quality assessment at tool, module and platform level

The aforementioned metrics depend on input from the questionnaires, which are provided at an individual tool and platform level. However, results can be also aggregated to a module level if needed, in order to validate the performance of the different STOP-IT modules. Since modules in STOP-IT are containers of tools, a single quality estimate S_m can be extracted by calculating the mean score of the tools included in that module:

$$S_m = \frac{1}{n}\sum_{i=1}^{n} S_i \tag{3}$$

where[2] S_i is the (aggregate) score of the tool *i.* There is no need for weights in the aggregation from tools to modules, as all tools are considered to be equally contributing to the functionality of STOP-IT. Likewise, if the evaluation in T7.5 needs to focus in one specific trait of a module, this particular attribute performance can be calculated as the mean of the corresponding attribute score of the tools within that module. While the aforementioned aggregation works on most STOP-IT modules, there are two modules (module 8 and 9) which constitute basic functions in the integrated platform (WP6) and are not standalone CP risk assessment products; these modules feed into specific *platform traits* and can be thus evaluated by looking at the corresponding trait performance on the whole platform level. Table 8 provides an overview of the tools and modules of STOP-IT and the corresponding validation levels. ANNEX A, which is taken from the information provided in D7.1, offers more information on the functionality of each tool and can be used as a reference base, combined with Table 8. With regards to validating the whole platform, one may observe that there are two possible ways to do this through the proposed methodology:

a.) To rely on the results of the platform-level questionnaires only, and thus provide a metric similar to the one seen in Equation (1):

$$S_p = \frac{1}{k}\sum_{i=1}^{k} w_i X_i^{(p)} \tag{4}$$

where $X_i^{(p)}$is now the trait score of the whole platform, obtained from the platform-level questionnaires, or

b.) To combine these results with the ones obtained at the lower (tool) level, for instance by calculating the mean score of all tools and then pooling it with the score obtained at the platform level through equation (4):

$$S_{p,combined} = 0.5 \cdot S_p + 0.5\frac{1}{n}\sum_{t=1}^{n} S_t \tag{5}$$

[2] The difference in indexing here (*i* over *t*) wishes to underline that only a subset {i} of the total tools {t} counts towards the assessment of a specific module.

where n is the total number of tools and S_t the (aggregate) tool scores, obtained through equation (1). In this case equal weights are assumed, as there is no indication that either the platform-level or the tool-level questionnaires are more important to the outcome.

The validation plan allows flexibility in calculating these scores depending on the scope of the evaluation and reflection activities (which will be further decided in T7.5). For cases of rapid assessment, it is suggested to rely on the platform-level questionnaires only in order to present platform performance and either look at all attributes (Figure 5) or provide a scalar quality estimate through equation (4).

An overview of how quality can be estimated at multiple scales using this trait-based approach is given in Figure 6, with the steps being the following:

1.) Using the tool-level questionnaires, calculate each trait score X_i by averaging the score of the ranking questions inside that trait category.
2.) Present the tool level results by looking at all attributes (Figure 5) or by calculating a scalar estimate through equation (1).
3.) Module-level scoring occurs by averaging the score of the tools included inside that module. Exceptions are modules 8 and 9, which feed directly to relevant traits of the platform-level questionnaire (see also Table 8).
4.) Platform level scoring can be achieved by looking at the platform-level questionnaires only and repeating steps (1) and (2) for them, or by combining these results with the tool-level results through equation (5).

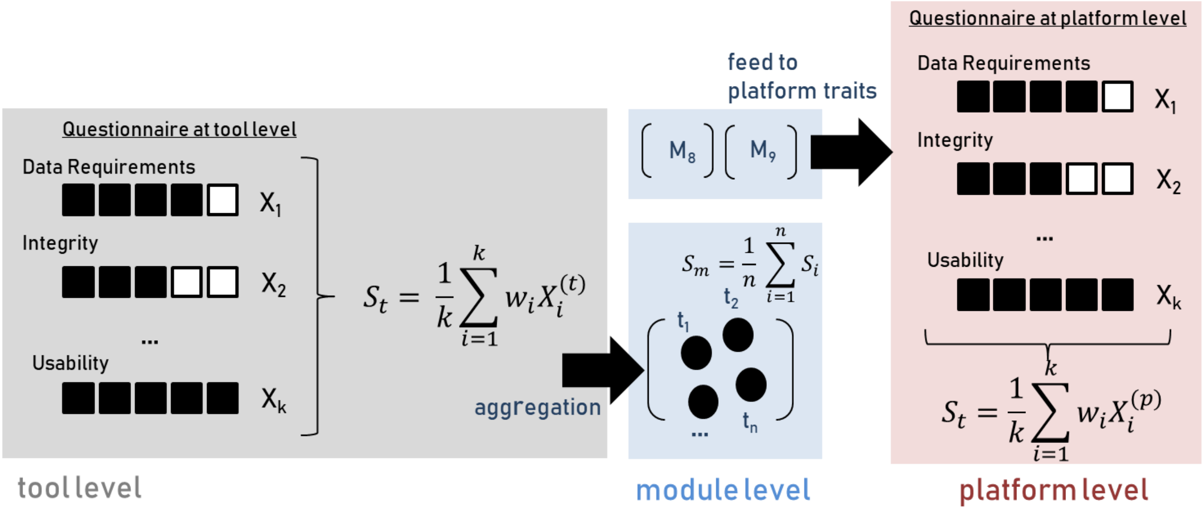


Figure 6: Ways to obtain quality metrics on a tool, module and platform level.

4 Integration of the validation plan with other deliverables

This section provides more information on how the concepts described in Chapter 3 are linked to the contents of other important deliverables. More specifically, the integration if the concepts with D7.1 – “Pilot plans and Report for the demo preparations” is discussed, since D7.1 provides information on the demonstration use case scenarios that form the basis for validation. Moreover, the important conceptual integration with the concepts of D3.3 - “Definition of end user requirements in Front Runner (FR) water utility” is analysed, as the contents of D3.3 examine what the end users demand from the functionality of the STOP-IT platform. Finally, a description of how the notions explained in Chapter 3 connect with the KPIs outlined in main documents such as the Description of Action (DoA) is provided.

4.1 Integration with demonstration use cases (D7.1)

D7.1 offers the definition of use case scenarios for each tool that will be used to design demonstration activities in all FRs, based on the tools they have selected to be demonstrated. Forty one (41) use cases in total are designed for the tools included in the STOP-IT platform. Despite the low level of analysis seen in the document, all of the afore-mentioned 41 use cases are active events that belong in Type II/III demonstration activities (Table 3) and follow the same high-level pattern that includes the steps described in Section 3.3. More specifically, these higher-level steps are the following (Figure 3):

1.) The technology demonstrated in the particular use case is installed in the system, and software/hardware dependence and licensing issues are solved.
2.) The user accesses the technology and gets acquainted with the front-end and UI.
3.) Necessary input datasets are collected from the user and fed into the technology. In case of the tool being a database, this step does not require third-party sources of data, but it is noted that any interaction with the users (including, for instance, generating database queries or retrieving subsets of data) requires – up to some degree – user input.
4.) The user accesses specific functions of the technology randomly or with a set pattern, as specified by the use case.
5.) Output from the technology is produced and presented to the user, which then evaluates it.
6.) In case of a use case that features multiple tools, this output can be fed (as input) to the next tool. The user then repeats steps 3-5 with the next tool that is part of the chain in order to further process the data and extract meaningful interpretable output.
7.) Following the end of the assessment or technology demonstration, the user is able to reflect on his experience and evaluate how useful the technology is in the context of his operations.

Since all use cases described in D7.1. follow the same pattern, the *trai*t and *partial characteristic* two-*level* evaluation scheme described in Section 3.2 holds for all cases and the questionnaire templates of this validation plan can be used in every designed use case.

It is noted that the planning of the demonstration activities should also carefully take into account the barriers relevant to the use case scenarios that are described for each case in D7.1, as they present potential bottlenecks that may shape how end users view that particular technology and thus its validation score.

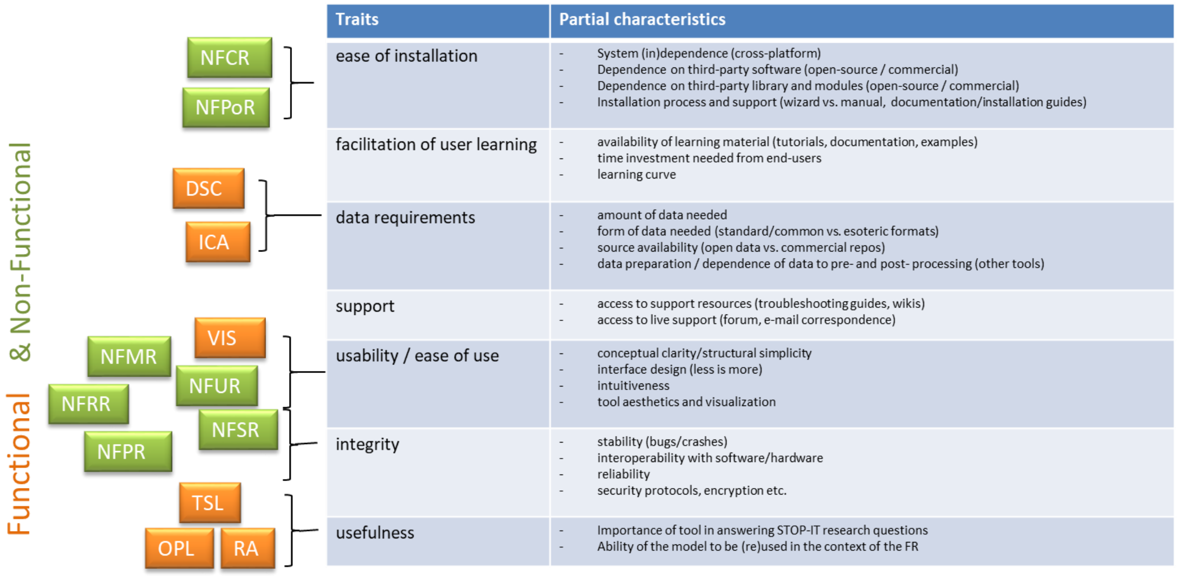
4.2 Integration with user requirements (D3.3)

D3.3 has the clear objective to collect all input information about end user requirements provided by the FRs and feeds its output directly to Task 6.1 (STOP-IT platform design), Task 7.1 (preliminary use cases) and Task 7.4 (demonstration activities of STOP-IT platform). Information on the FR user requirements is gathered through a round of two questionnaires, which also provide information on the platform design factors. The report then proceeds to the collection and analysis of the questionnaire results.

Similarly to D7.1, D3.3 examines user requirements at a very fine level, comprising a hundred and fifteen (115) identified requirements in thirteen (13) functional and non-functional categories. An overview of these categories is given in

Despite the large number of user requirements, D3.3 also provides an evaluation (ranking) of these requirements from each FR as mandatory, optional and not important for his/her perspective and line of work. D7.1 utilizes this deduction to study only the high priority requirements for the FRs and correlates these requirements with the tools and their use cases in order to describe how each requirement is covered by the STOP-IT platform. This report expands upon this concept to correlate how each requirement becomes validated by the developed validation plan. The results of this correlation can be seen in the sections that follow for each FR. Note that an initial selection of tools per FR for reference is provided in Section 5.2, based on the input from D7.1; this selection is indicative and might change.

Table 5 (functional requirements) and Table 6 (non-functional requirements). These requirements are exhaustive and, at times, conceptually overlapping with each other, but can be thematically mapped (with, in most cases, a many-to-one relationship) to parent trait categories; in most cases, they feed into particular partial characteristics. This thematic mapping between the functional and non-functional categories of D3.3 and the trait-partial characteristic system of D7.2 is given in Figure 7. Figure 7 indicates that most categories of D3.3, such as Data Sources and Collection (DSC) or Visualization (VIS), gravitate towards distinct traits (notably *Data Requirements* (DR) and *Usability* (UB)). The family of non-functional categories such as Portability (NFPoR), Security (NFSR) and Reliability (NFRR) balance their existence between the traits of *Usability* and *Integrity* and help shape partial characteristics in both of them. This mapping allows partial characteristics to be shaped by specific D3.3 categories and their requirements and was taken into account during the design of the questionnaires of the validation plan. A notable remark is this mapping identified two traits (*facilitation of user learning* and *support*) that were not found in any of the D3.3 categories. This mismatch is perhaps owing to the difference of perspective between D3.3, which focused on internal functional and non-functional requirements, and D7.2, which takes the user experience (UX) – and hence the learning process of the end user - as a departure point for the analysis.



Traits	Partial characteristics
ease of installation	- System (in)dependence (cross-platform) - Dependence on third-party software (open-source / commercial) - Dependence on third-party library and modules (open-source / commercial) - Installation process and support (wizard vs. manual, documentation/installation guides)
facilitation of user learning	- availability of learning material (tutorials, documentation, examples) - time investment needed from end-users - learning curve
data requirements	- amount of data needed - form of data needed (standard/common vs. esoteric formats) - source availability (open data vs. commercial repos) - data preparation / dependence of data to pre- and post- processing (other tools)
support	- access to support resources (troubleshooting guides, wikis) - access to live support (forum, e-mail correspondence)
usability / ease of use	- conceptual clarity/structural simplicity - interface design (less is more) - intuitiveness - tool aesthetics and visualization
integrity	- stability (bugs/crashes) - interoperability with software/hardware - reliability - security protocols, encryption etc.
usefulness	- Importance of tool in answering STOP-IT research questions - Ability of the model to be (re)used in the context of the FR

Figure 7: Thematic mapping between D3.3. categories and the trait methodology of this report.

Despite the large number of user requirements, D3.3 also provides an evaluation (ranking) of these requirements from each FR as mandatory, optional and not important for his/her perspective and line of work. D7.1 utilizes this deduction to study only the high priority requirements for the FRs and correlates these requirements with the tools and their use cases in order to describe how each requirement is covered by the STOP-IT platform. This report expands upon this concept to correlate how each requirement becomes validated by the developed validation plan. The results of this correlation can be seen in the sections that follow for each FR. Note that an initial selection of tools per FR for reference is provided in Section 5.2, based on the input from D7.1; this selection is indicative and might change.

Table 5: Functional requirement categories in D3.3.

Category Name		Requirements in category	Category Description
1	Data Sources and Collection (DSC)	DSC 01-14	collection of data
2	Information Correlation and Abstraction (ICA)	ICA 01-08	treatment of collected data
3	Tactical and Strategic Level (TSL)	TSL 01-09	requirements for tactical/strategic actions
4	Operational Level (OPL)	OPL 01-08	requirements for operational/levels
5	Risk Analysis (RA)	RA 01-15	requirements for risk assessment
6	Visualization (VIS)	VIS 01-12	aspects of visualization

Table 6: Non-functional requirement categories in D3.3.

Category Name		Requirements in category	Category Description
1	Performance (NFPR)	NFPR 01-08	resources and time behavior of platform
2	Compatibility (NFCR)	NFCR 01-10	compatibility with software/hardware
3	Usability (NFUR)	NFUR 01-05	Efficiency, effectiveness and satisfaction on user interaction
4	Reliability (NFRR)	NFRR 01-09	platform reliability
5	Security (NFSR)	NFSR 01-10	platform security
6	Maintainability (NFMR)	NFMR 01-04	platform modularity/reusability
7	Portability (NFPoR)	NFPoR 01-03	Transferability to multiple systems/VEs

4.2.1 AB

Requirements Addressed (only the highly prioritised have been included)	How the requirement is covered	How the requirement is validated
DSC-02: STOP-IT must be a flexible platform that collects/treats information from different sources and formats.	The selected set of tools collects information from various data modalities.	The questions of the Data Requirements (DR) traits on the individual tool, as well as the platform level provide scoring on the way data is treated in STOP-IT.
DSC-06: The data collection system must collect security policy information. For instance, permissions and prohibitions assigned to users in order to access data or any kind of resources from the system. e.g., User ABC is denied access to file XY.	In general, this requirement is applicable to any tool in which authorization access is required. Moreover, InfraRisk CP, RAET, and SP only work when the user has access to specific resources of the water network model. Finally, the CVT, while it does not act as an access control mechanism, it automatically detects suspicious actions, so it implicitly adds intelligence in the permission control procedures.	There are questions that target security on both tool and platform level, as part of the Integrity (IG) trait category.
DSC-10: The data collection system must collect network and system events (e.g., cyber security alerts from intrusion detection systems).	This requirement is covered by the selection of InfraRisk CP, RAET, STP, FTCS, CTSS, RTAD.	Validation (and pass/fail checks) on these tools (InfraRisk CP, RAET, STP, FTCS, CTSS, RTAD) provide answers on whether this requirement is covered or not.
ICA-01: The STOP-IT platform must make use of an information correlation engine.	RAET, STP, RRMD, FTCS, CVT, CTSS, RTAD REN are all tools that correlate information from relevant data.	Validation (and pass/fail checks) on these tools (RAET, STP, RRMD, FTCS, CVT, CTSS, RTAD) provide answers on whether this requirement is covered or not.

ICA-02: STOP-IT must translate multi-source information received by the data collection system into a semantic (logic-based) common information representation.	InfraRisk CP, RRMD, RTAD, REN are all tools that exploit data from multiple sources to produce semantically enriched data.	Validation (and pass/fail checks) on these tools (InfraRisk CP, RRMD, RTAD) provide answers on whether this requirement is covered or not.
ICA-05: The information correlation engine must create a unified view of the monitored system from the multi-source information received by the data collection system.	RTAD, REN correlates data from multiple data collection systems.	Validation (and pass/fail checks) on RTAD provides answers on whether this requirement is covered or not. The integrity (IG) trait questions on a platform level provide insight for REN.
TSL-02: The STOP-IT platform must be able to evaluate potential cyber and physical attacks based on security policy and vulnerability information	STP, SP, RRMD all evaluate potential cyber-physical threats by assessing the vulnerabilities of a water distribution network.	Validation (and pass/fail checks) on these tools (STP, SP, RRMD) provide answers on whether this requirement is covered or not.
TSL-04: STOP-IT must allow operators to activate any selected mitigation action to prevent potential cyber or physical threats from being realized.	All the selected tools provide useful information regarding increased risk of threat realization to the relevant administrators of a Water Utility. This allows the water utility to activate its procedures and mitigate the effects of perilous situations.	Validation on the whole platform level – and more specifically with regards to the *usefulness* trait – provides answers on whether this requirement is covered or not.
OPL-03: Actions considered to mitigate an attack must reduce the risk level down to an acceptable level	All the selected tools aim to successfully cover this requirement. The level of success, as well as possible calibration of the tools to enhance their outcomes, is part of the demonstration activities of WP7.	A reflection on the total validation outcome (on both tool and platform level) will provide insights on whether this requirement is covered or not.
OPL-04 Reactive actions against a malicious event must be implemented upon validation of the system operator.	All the STOP-IT tools do not impose actions to the water utility. Rather they identify risks, and propose	Validation on the whole platform level – and more specifically with regards to the *usefulness* trait – provides answers on whether this requirement is

OPL-06: The operational response system must enable the operator to select and deploy proposed mitigation actions.	mitigation strategies that the system operator can choose to implement.	covered or not, with regards to the proposed mitigation options and the identification of risk.
RA- 02: Risk analysis and evaluation must assess assets criticality by calculating the impacts of risk events under customized end user driven scenarios	SP allows the system operator to plan specialized scenarios of events and calculate the impact.	Validation (and a pass/fail check) on SP provides answers on whether this requirement is covered or not.
RA- 04: The STOP-IT platform must allow for the analysis and evaluation of physical and cyber risks on water CIs and their combinations.	InfraRisk CP, RAET, and SP provide services that cover this requirement.	Validation (and pass/fail checks) on these tools (InfraRisk CP, RAET, and SP) provide answers on whether this requirement is covered or not.
VIS-01 The STOP-IT visualization system must allow users to customize information displayed in the dashboard.	These requirements act as specifications that a visualization engine, EVI will offer functionalities that address all these points.	Validation of the usability (UB) trait on the platform level provides insights on how EVI performs.
VIS-03 The visualization system must provide information about the actual system Vulnerabilities.		
VIS-04: STOP-IT must evaluate and analyse exploitable vulnerabilities.		
VIS-06: The visualization system must display the mitigation actions proposed by the tactical and operational response systems.		
VIS-09: The visualization system must enable historical data analysis.		

NFPR-01: STOP-IT components must run in virtual machines with appropriate speed and storage capabilities.	All the software-based tools will have the ability to run in VMs.	Questions related to the ease of installation (EI), integrity (IG) and usability (UB) traits for the tool level are able to validate the ability to run in VMs.
NFCR -01: STOP-IT must privilege the use of simple and standard formats (e.g., TXT, MDPA, XML, JSON, ...) as the formats for data object exchanges.	All STOP-IT technical partners use widely accepted data formats as part for their development process. Deviations from this requirement will only be allowed for specific technical reasons that will be explained in the respective tool description.	There are questions on the data requirements (D) trait for the tool level that target specifically the capacity to handle common standard formats.
NFCR-06: The STOP-IT platform must be designed to ease management and optimization.	This is the prime specification of the RGIP.	Validation of the usefulness (UF) trait on the platform level provides insights on how EVI performs.
NFCR-07: The proposed integrated STOP-IT solution must be fully interoperable in order to accommodate collaboration with existing (legacy) systems, sensors, security management tools and infrastructure while being able to use market-available components to be deployed over the large area of the CIs.	STOP-IT demonstrations will illustrate the success of this requirement, as it plans of demonstrating the developed tools, without disturbing any day-to-day operations of the water utility.	Questions of interoperability and dependence on third-party software and operational systems are parts of the ease of installation (EI) trait and specifically target this requirement.
NFUR-01: Usability of the system must be evaluated verified and improved through regular validation by STOP-IT FRs.	WP7 will execute plan validation activities in all of STOP-IT FR consortium partners. This deliverable (D.7.1) illustrates the initial planning of this activities, while D.7.2 will describe in detail the validation framework of WP7's validation activities.	This requirement is covered by the existence of the Usability (UB) trait category on both a tool and a platform level.
NFUR-02: Users must be able to customize the data displayed in the STOP-IT graphical interface in order to adapt its content to their needs.	EVI will offer functionalities that address all these points.	This requirement is covered by the existence of the Usability (UB) trait category on both a tool and a platform level.

NFRR-01: Databases in the STOP-IT system must be regularly backed-up and automatically restored in case of problems.	This is covered by the individual technology/tool providers. All relevant partners have taken actions in ensuring data protection and fast return to normal operations after an unexpected problem occurs.	Validation of the integrity (IG) of the corresponding database tools (e.g. RRMD) is able to answer whether this requirement is covered or not.
NFRR-06: The STOP-IT platform must not impact the operational environment (i.e., the monitored or protected systems) when experiencing a failure. **In other words, if we have a failure and the STOP-IT system is down for some minutes or seconds, it should not impact negatively the end user environment.**	All the demonstration activities will not disrupt day-to-day operations in the FRs.	There is no need to validate that requirement, as the day-to-day operation will not be disrupted during the demonstration activities.
NFRR-07: Regular operations in the STOP-IT platform must not increase the number of alerts raised, which could lead to an infinite loop of reaction/detection.	This requirement has been taken into account by all relevant partners during the development process. The planned demonstration activities will help to calibrate the tools as well, in order to minimize false positive alerts.	Questions on the integrity (IG) of the tool and platform level are able to validate whether the tool or platform led to unexpected events.
NFSR-07: A ticket must be generated for all detected events compromising the system confidentiality, integrity or availability	This is the planned output of all the STOP-IT tools.	Validation (and pass/fail checks) on all tools provide answers on whether this requirement is covered or not.
NFMR-02: STOP-IT must be comprised of a fully integrated system of functions from detection to reaction.	RGIP offers that service.	Questions on the integrity (IG) of the platform level are able to validate whether this requirement is met or not.
NFMR-03: The STOP-IT data model must be designed to support integration with new data sources in the monitored system	All the tools are developed in such a way that they can use most relevant formats of required data input. This will be illustrated by the piloting of these tools in multiple FRs with varying data formats and operational procedures.	There are questions on the data requirements (D) trait for the tool and platform level that target how the tool or platform handles data management.

NFPoR-01: STOP-IT components must be developed in a language that is portable among most current operating systems.	The software tools will be executable in the most popular Operating Systems (MS Windows, Unix based systems).	Any issues of portability will be uncovered and evaluated during using the Ease of Installation (EI) trait on a tool and platform level.
NFPoR-02 The STOP-IT platform must be able to host its different components in virtual environments (e.g., VMs).	All the software-based tools will have the ability to be executed in a virtualized environment.	Questions on the integrity (IG) on the tool and platform level are able to validate whether this requirement is met or not.

4.2.2 BWB

Requirements Addressed (high prioritization)	How the requirement is covered	How the requirement is validated
DSC-02: STOP-IT must be a flexible platform that collects/treats information from different sources and formats.	The selected set of tools collects information from various data modalities.	The questions of the Data Requirements (DR) traits on the individual tool, as well as the platform level provide scoring on the way data is treated in STOP-IT.
TSL -02: The STOP-IT platform must be able to evaluate potential cyber and physical attacks based on security policy and vulnerability information	STP, SP, RRMD all evaluate potential cyber-physic threats by assessing the vulnerabilities of a water distributio network.	Validation (and pass/fail checks) on these tools (STP, SP, RRMD) provide answers on whether this requirement is covered or not.
TSL -04: STOP-IT must allow operators to activate any selected mitigation action to prevent potential cyber or physical threats from being realized.	All the selected tools provide useful information regarding increased risk of threat realization to the relevant administrators of a Water Utility. This allows the water utility to activate its procedures and mitigate the effects of perilous situations.	Validation on the whole platform level – and more specifically with regards to the *usefulness* trait – provides answers on whether this requirement is covered or not.
OPL-03: Actions considered to mitigate an attack must reduce the risk level down to an acceptable level	All the selected tools aim to successfully cover this requirement. The level of success, as well as possible calibration of the tools to enhance their outcomes, is part of the demonstration activities of WP7.	A reflection on the total validation outcome (on both tool and platform level) will provide insights on whether this requirement is covered or not.
OPL-04 Reactive actions against a malicious event must be implemented upon validation of the system operator.	All the STOP-IT tools do not impose actions to the water utility. Rather they identify risks, and propose mitigation strategies that the system operator can choose to implement.	Validation on the whole platform level – and more specifically with regards to the *usefulness* trait – provides answers on whether this requirement is covered or not, with regards to the proposed mitigation options and the identification of risk.

RA- 02: Risk analysis and evaluation must assess assets criticality by calculating the impacts of risk events under customized end user driven scenarios	SP allows the system operator to plan specialized scenarios of events and calculate the impact.	Validation (and a pass/fail check) on SP provides answers on whether this requirement is covered or not.
RA- 04: The STOP-IT platform must allow for the analysis and evaluation of physical and cyber risks on water CIs and their combinations.	This is achieved through all the selected STOP-IT toolkits.	Validation (and pass/fail checks) on these tools (InfraRisk CP, RAET, and SP) provide answers on whether this requirement is covered or not.
VIS-04: STOP-IT must evaluate and analyse exploitable vulnerabilities.	This is achieved through the STOP-IT toolkits 1, 2, 7, 8.	Validation on the module level of the selected toolkits provides answer on whether this requirement is covered or not.
NFUR-01: Usability of the system must be evaluated verified and improved through regular **validation by STOP-IT FRs.**	WP7 will execute plan validation activities in all of STOP-IT FR consortium partners. This deliverable (D.7.1) illustrates the initial planning of this activities, while D.7.2 will describe in detail the validation framework of WP7's validation activities.	This requirement is covered by the existence of the Usability (UB) trait category on both a tool and a platform level.
NFUR-02: Users must be able to customize the data displayed in the STOP-IT graphical interface in order to adapt its content to their needs.	This is achieved through the visualization toolkits and particularly the toolkit 14.	This requirement is covered by the existence of the Usability (UB) trait category on a platform level.
NFRR-01: Databases in the STOP-IT system must be regularly backed-up and automatically restored in case of problems.	This is covered by the individual technology/tool providers. All relevant partners have taken actions in ensuring data protection and fast return to normal operations after an unexpected problem occurs.	Validation of the integrity (IG) of the corresponding database tools (e.g. RRMD) is able to answer whether this requirement is covered or not.

4.2.3 MEK

Requirements Addressed (only the highly prioritised have been included)	How the requirement is covered	How the requirement is validated
DSC-02: STOP-IT must be a flexible platform that collects/treats information from different sources and formats.	The selected set of tools collects information from various data modalities.	The questions of the Data Requirements (DR) traits on the individual tool, as well as the platform level provide scoring on the way data is treated in STOP-IT.
DSC-06: The data collection system must collect security policy information. For instance, permissions and prohibitions assigned to users in order to access data or any kind of resources from the system. e.g., User ABC is denied access to file XY.	Smart-Locks offers this exact feature for physical access.	There are questions that target security on both tool and platform level, as part of the Integrity (IG) trait category.
DSC-10: The data collection system must collect network and system events (e.g., cyber security alerts from intrusion detection systems).	The requirement is covered by FTCS, NTSA, RSDP, RTAD.	Validation (and pass/fail checks) on these tools (FTCS, NTSA, RSDP, RTAD) provide answers on whether this requirement is covered or not.
ICA-01: The STOP-IT platform must make use of an information correlation engine.	This is achieved through the selected toolkits, which are able to select information.	Validation (and pass/fail checks) on the selected tools provide answers on whether this requirement is covered or not.
ICA-02: STOP-IT must translate multi-source information received by the data collection system into a semantic (logic-based) common information representation.	This is achieved through the selected toolkits, where they collect multisource information, such as data from Wi-Fi signal, network traffic and anomaly detection.	Validation (and pass/fail checks) on the selected tools provide answers on whether this requirement is covered or not.
ICA- 05: The information correlation engine must create a unified view of the monitored	RTAD correlates data from multiple data collection systems.	Validation (and pass/fail checks) on RTAD provides answers on whether this requirement is covered or not.

system from the multi-source information received by the data collection system.		
TSL -02: The STOP-IT platform must be able to evaluate potential cyber and physical attacks based on security policy and vulnerability information	This is achieved through the selected toolkits which allow for a tactical and strategic response, particularly the FTCS, RTAD and RSDP, NTSA.	Validation (and pass/fail checks) on these tools (FTCS, RTAD, RSDP, NTSA) provide answers on whether this requirement is covered or not.
TSL -04: STOP-IT must allow operators to activate any selected mitigation action to prevent potential cyber or physical threats from being realized.	All the selected tools provide useful information regarding increased risk of threat realization to the relevant administrators of a Water Utility. This allows the water utility to activate its procedures and mitigate the effects of perilous situations.	Validation on the whole platform level – and more specifically with regards to the *usefulness* trait – provides answers on whether this requirement is covered or not.
OPL-04 Reactive actions against a malicious event must be implemented upon validation of the **system operator.**	All the STOP-IT tools do not impose actions to the water utility. Rather they identify risks, and propose mitigation strategies that the system operator can choose to implement.	Validation on the whole platform level – and more specifically with regards to the *usefulness* trait – provides answers on whether this requirement is covered or not, with regards to the proposed mitigation options and the identification of risk from the end user.
OPL-06: The operational response system must enable the operator to select and deploy proposed mitigation actions.	All the STOP-IT tools do not impose actions to the water utility. Rather they identify risks, and propose mitigation strategies that the system operator can choose to implement.	Validation on the whole platform level – and more specifically with regards to the *usefulness* trait – provides answers on whether this requirement is covered or not, with regards to the proposed mitigation options and the identification of risk.
RA- 02: Risk analysis and evaluation must assess assets criticality by calculating the impacts of risk events under customized end user driven scenarios	This is achieved through the selected toolkits, which provide information through multiple sources and therefore they are an assessment of the impact of the risks.	Validation (and a pass/fail check) on these tools provide answers on whether this requirement is covered or not.
RA- 04: The STOP-IT platform must allow for the analysis and evaluation of physical and	The selected toolkits allow analysis and evaluation of cyber-physical risks.	Validation (and pass/fail checks) on these tools provide answers on whether this requirement is covered or not.

cyber risks on water CIs and their combinations.		
VIS-02 The STOP-IT visualization system must allow users to customize information displayed in the dashboard.	EVI offers this feature.	Validation of the usability (UB) trait on the platform level provides insights on how EVI performs.
VIS-03 The visualization system must provide information about the actual system Vulnerabilities.		
VIS-06: The visualization system must display the mitigation actions proposed by the tactical and operational response systems.	EVI offers this feature.	Validation of the usability (UB) trait on the platform level provides insights on how EVI performs.
VIS-09: The visualization system must enable historical data analysis.		
NFPR -01: STOP-IT components must run in virtual machines with appropriate speed and storage capabilities.	All these toolkits will run on independent computer interfaces or virtual machines.	Questions related to the ease of installation (EI), integrity (IG) and usability (UB) traits for the tool level are able to validate the ability to run in VMs.
NFCR -01: STOP-IT must privilege the use of simple and standard formats (e.g., TXT, MDPA, XML, JSON, ...) as the formats for data object exchanges.	All the selected services will communicate with each other using standard formats for data object exchange.	There are questions on the data requirements (D) trait for the tool level that target specifically the capacity to handle common standard formats.
NFCR-06: The STOP-IT platform must be designed to ease management and optimization.	All the selected toolkits are configurable and easy to manage and optimize.	Validation of the usability (UB) trait on the platform level provides insights on this requirement.
NFCR-07: The proposed integrated STOP-IT solution must be fully interoperable in order	STOP-IT demonstrations will illustrate the success of this requirement, as it plans of demonstrating the	Questions of interoperability and dependence on third-party software and operational systems are parts of the

to accommodate collaboration with existing (legacy) systems, sensors, security management tools and infrastructure while being able to use market-available components to be deployed over the large area of the CIs.	developed tools, without disturbing any day-to-day operations of the water utility.	ease of installation (EI) trait and specifically target this requirement.
NFRR-01: Databases in the STOP-IT system must be regularly backed-up and automatically restored in case of problems.	This is covered by the individual technology/tool providers. All relevant partners have taken actions in ensuring data protection and fast return to normal operations after an unexpected problem occurs.	Validation of the integrity (IG) of the corresponding database tools (e.g. RRMD) is able to answer whether this requirement is covered or not.
NFRR-06: The STOP-IT platform must not impact the operational environment (i.e., the monitored or protected systems) when experiencing a failure.	All the demonstration activities will not disrupt day-to-day operations in the FRs.	There is no need to validate that requirement, as the day-to-day operation will not be disrupted during the demonstration activities.
NFRR-07: Regular operations in the STOP-IT platform must not increase the number of alerts raised, which could lead to an infinite loop of reaction/detection.	This requirement has been taken into account by all relevant partners during the development process. The planned demonstration activities will help to calibrate the tools as well, in order to minimize false positive alerts.	Questions on the integrity (IG) of the tool and platform level are able to validate whether the tool or platform led to unexpected events.
NFSR-07: A ticket must be generated for all detected events compromising the system confidentiality, integrity or availability	This is the planned output of all the STOP-IT tools.	Validation (and pass/fail checks) on all tools provide answers on whether this requirement is covered or not.
NFMR-03: The STOP-IT data model must be designed to support integration with new data sources in the monitored system	All the tools are developed in such a way that they can use most relevant formats of required data input. This will be illustrated by the piloting of these tools in multiple FRs with varying data formats and operational procedures.	There are questions on the data requirements (D) trait for the tool and platform level that target how the tool or platform handles data management.

NFPoR-01: STOP-IT components must be developed in a language that is portable among most current operating systems.	The software tools will be executable in the most popular Operating Systems (MS Windows, Unix based systems).	Any issues of portability will be uncovered and evaluated during using the Ease of Installation (EI) trait on a tool and platform level.
NFPoR-02 The STOP-IT platform must be able to host its different components in virtual environments (e.g., VMs).	All the selected components will run on independent computer interfaces or virtual machines.	Questions on the integrity (IG) on the tool and platform level are able to validate whether this requirement is met or not.

4.2.4 VAV

Requirements Addressed (only the highly prioritised have been included)	How the requirement is covered	How the requirement is validated
DSC-02: STOP-IT must be a flexible platform that collects/treats information from different sources and formats.	The selected set of tools collects information from various data modalities.	The questions of the Data Requirements (DR) traits on the individual tool, as well as the platform level provide scoring on the way data is treated in STOP-IT.
DSC-06: The data collection system must collect security policy information. For instance, permissions and prohibitions assigned to users in order to access data or any kind of resources from the system. e.g., User ABC is denied access to file XY.	In general, this requirement is applicable to any tool in which authorization access is required. InfraRisk CP, RAET, and SP only work when the user has access to specific resources of the water network model. Moreover, the Smart-Locks tools and XL-SIEM offers access control services in both physical and cyber resources. Finally, the CVT, while it does not act as an access control mechanism, it automatically detects suspicious actions, so it implicitly adds intelligence in the permission control procedures.	There are questions that target security on both tool and platform level, as part of the Integrity (IG) trait category.
DSC-10: The data collection system must collect network and system events (e.g., cyber security alerts from intrusion detection systems).	This requirement is covered by the selection of InfraRisk CP, RAET, STP, FTCS, CTSS, RTAD.	Validation (and pass/fail checks) on these tools (InfraRisk CP, RAET, STP, FTCS, CTSS, RTAD) provide answers on whether this requirement is covered or not.
ICA-01: The STOP-IT platform must make use of an information correlation engine.	RAET, STP, RRMD, FTCS, CVT, CTSS, RTAD, REN are all tools that correlate information from relevant data.	Validation (and pass/fail checks) on these tools (RAET, STP, RRMD, FTCS, CVT, CTSS, RTAD) provide answers on whether this requirement is covered or not.

ICA-02: STOP-IT must translate multi-source information received by the data collection system into a semantic (logic-based) common information representation.	InfraRisk CP, RRMD, RTAD, REN are all tools that exploit data from multiple sources to produce semantically enriched data.	Validation (and pass/fail checks) on these tools (InfraRisk CP, RRMD, RTAD) provide answers on whether this requirement is covered or not.
ICA- 05: The information correlation engine must create a unified view of the monitored system from the multi-source information received by the data collection system.	RTAD, REN correlates data from multiple data collection systems.	Validation (and pass/fail checks) on RTAD provides answers on whether this requirement is covered or not. The integrity (IG) trait questions on a platform level provide insight for REN.
OPL-04 Reactive actions against a malicious event must be implemented upon validation of the system operator.	All the STOP-IT tools do not impose actions to the water utility. Rather they identify risks, and propose mitigation strategies that the system operator can choose to implement.	A reflection on the total validation outcome (on both tool and platform level) will provide insights on whether this requirement is covered or not.
OPL-06: The operational response system must enable the operator to select and deploy proposed mitigation actions.	All the STOP-IT tools do not impose actions to the water utility. Rather they identify risks, and propose mitigation strategies that the system operator can choose to implement.	Validation on the whole platform level – and more specifically with regards to the *usefulness* trait – provides answers on whether this requirement is covered or not, with regards to the proposed mitigation options and the identification of risk.
RA- 02: Risk analysis and evaluation must assess assets criticality by calculating the impacts of risk events under customized end user driven scenarios	SPT allows the system operator to plan specialized scenarios of events and calculate the impact.	Validation (and a pass/fail check) on SPT provides answers on whether this requirement is covered or not.
RA- 04: The STOP-IT platform must allow for the analysis and evaluation of physical and cyber risks on water CIs and their combinations.	The selected toolkits allow analysis and evaluation of physical attacks, since the provide information from multiple sources.	Validation (and pass/fail checks) on the selected tools provide answers on whether this requirement is covered or not.
VIS-02 The STOP-IT visualization system must allow users to customize information displayed in the dashboard.	EVI offers this feature.	Validation of the usability (UB) trait on the platform level provides insights on how EVI performs.

VIS-03 The visualization system must provide information about the actual system Vulnerabilities.	EVI offers this feature.	
VIS-04: STOP-IT must evaluate and analyse exploitable vulnerabilities.	EVI offers this feature.	
VIS-06: The visualization system must display the mitigation actions proposed by the tactical and operational response systems.	EVI offers this feature.	Validation of the usability (UB) trait on the platform level provides insights on how EVI performs.
VIS-09: The visualization system must enable historical data analysis.	EVI offers this feature.	
NFPR-01: STOP-IT components must run in virtual machines with appropriate speed and storage capabilities.	All these toolkits can run on independent computer interfaces or virtual machines.	Questions related to the ease of installation (EI), integrity (IG) and usability (UB) traits for the tool level are able to validate the ability to run in VMs.
NFCR -01: STOP-IT must privilege the use of simple and standard formats (e.g., TXT, MDPA, XML, JSON, ...) as the formats for data object exchanges.	All the selected services will communicate with each other using standard formats for data object exchange.	There are questions on the data requirements (D) trait for the tool level that target specifically the capacity to handle common standard formats.
NFCR-06: The STOP-IT platform must be designed to ease management and optimization.	All the selected toolkits are configurable and easy to manage and optimize.	Validation of the usefulness (UF) trait on the platform level provides insights on how EVI performs.
NFCR-07: The proposed integrated STOP-IT solution must be fully interoperable in order to accommodate collaboration with existing (legacy) systems, sensors, security management tools and infrastructure while being able to use market-available	STOP-IT demonstrations will illustrate the success of this requirement, as it plans of demonstrating the developed tools, without disturbing any day-to-day operations of the water utility.	Questions of interoperability and dependence on third-party software and operational systems are parts of the ease of installation (EI) trait and specifically target this requirement.

components to be deployed over the large area of the CIs.		
NFUR-01: Usability of the system must be evaluated verified and improved through regular validation by STOP-IT FRs.	WP7 will execute plan validation activities in all of STOP-IT FR consortium partners.	This requirement is covered by the existence of the Usability (UB) trait category on both a tool and a platform level.
NFUR-02: Users must be able to customize the data displayed in the STOP-IT graphical interface in order to adapt its content to their needs.	All the selected components will have configuration parameters and support an ease of use interface, increasing the usability of the system.	This requirement is covered by the existence of the Usability (UB) trait category on both a tool and a platform level.
NFRR-01: Databases in the STOP-IT system must be regularly backed-up and automatically restored in case of problems.	This is covered by the individual technology/tool providers. All relevant partners have taken actions in ensuring data protection and fast return to normal operations after an unexpected problem occurs.	Validation of the integrity (IG) of the corresponding database tools (e.g. RRMD) is able to answer whether this requirement is covered or not.
NFRR-06: The STOP-IT platform must not impact the operational environment (i.e., the monitored or protected systems) when experiencing a failure.	All the demonstration activities will not disrupt day-to-day operations in the FRs.	There is no need to validate that requirement, as the day-to-day operation will not be disrupted during the demonstration activities.
NFRR-07: Regular operations in the STOP-IT platform must not increase the number of alerts raised, which could lead to an infinite loop of reaction/detection.	This requirement has been taken into account by all relevant partners during the development process. The planned demonstration activities will help to calibrate the tools as well, in order to minimize false positive alerts.	Questions on the integrity (IG) of the tool and platform level are able to validate whether the tool or platform led to unexpected events.

NFSR-07: A ticket must be generated for all detected events compromising the system confidentiality, integrity or availability	This is the planned output of all the STOP-IT tools.	Validation (and pass/fail checks) on all tools provide answers on whether this requirement is covered or not.
NFMR-03: The STOP-IT data model must be designed to support integration with new data sources in the monitored system	All the tools are developed in such a way that they can use most relevant formats of required data input. This will be illustrated by the piloting of these tools in multiple FRs with varying data formats and operational procedures.	The questions of the Data Requirements (DR) traits on the individual tool, as well as the platform level provide scoring on the way data is treated in STOP-IT.
NFPoR-01: STOP-IT components must be developed in a language that is portable among most current operating systems.	The software tools will be executable in the most popular Operating Systems (MS Windows, Unix based systems).	Any issues of portability will be uncovered and evaluated during using the Ease of Installation (EI) trait on a tool and platform level.
NFPoR-02 The STOP-IT platform must be able to host its different components in virtual environments (e.g., VMs).	All the selected components will run on independent computer interfaces or virtual machines.	Questions on the integrity (IG) on the tool and platform level are able to validate whether this requirement is met or not.

4.3 Exploring the impact of STOP-IT through the validation plan

The inclusion of the *usefulness* (UF) trait in the methodology (see Section 3.2 and Figure 3) creates a link between validation planning and the end user reflection process that follows each demonstration activity. By reflecting how useful the demonstrated STOP-IT products are in their operational contexts, the FRs – as tokens of expert judgment - are able to project and estimate the expected impact the STOP-IT platform will have to their line of work, as well as to the level of protection that can be achieved with the framework. This “reflection on projected impact” creates a surrogate validation scheme, based on expert end user judgment, to measure the impact of STOP-IT in the FR domains[3].

Table 7: Impact indicators (KPIs) as seen in the DoA.

Impact KPIs (as seen in the DoA)	Description	Treatment through the questionnaire templates (platform level, usefulness)
KPI_1	Accuracy improvement, higher detection of physical/cyber-attacks and incidents	YES
KPI_2	Latency reduction	YES
KPI_3	Response time	YES
KPI_4	Preparedness improvement	YES
KPI_5	Reduction of human exposure	YES
KPI_6	Cost effectiveness	YES
KPI_7	Business orientation	Not included, to be evaluated during exploitation activities (WP9).
KPI_8	Community involvement	Not included, to be evaluated through CoP activities (WP2).
KPI_9	Certification and knowledge transfer	Not included, to be evaluated through knowledge transfer activities at later phases (WP2, WP8).

To utilize this, specific questions on the expected impact of STOP-IT to the environment of each FR are included as additional indicators of the *usefulness* of the whole platform. These questions feed from the impact STOP-IT aims at having, as defined by relevant Key

[3] Without that surrogate scheme, the operational impact of STOP-IT would be tedious and even impossible to measure at the projected time horizon, as it is directly dependent on processes that last well beyond the development and demonstration horizon, such as the long-term frequent application of STOP-IT tools in operational FR environments or the repetitive use of STOP-IT tools against actual real (rare) events.

Performance Indicators (KPIs) in the STOP-IT Description of Action (DoA) document. These KPIs can be seen in Table 7. Since validation is based on the impact end users expect the STOP-IT platform to have on their operational domains, only KPIs 1-6 (i.e. KPIs which describe direct effects to the FR operations and level of protection and are relevant to the contents of WP3,4,5 and 6) are relevant in the case of D7.2. Besides them, there are two KPIs (8 and 9) which are relevant to the Communities of Practice (CoP) and are treated in the relevant deliverables of WP2 (D2.2 annual versions), while another KPI (7) is relevant to business orientation and is treated in the relevant deliverables of WP9.

5 Application

This section provides guidance to the demonstration activities by providing application templates for the validation plan per FR, based on the tools each FR has selected. These templates mark the tools of interest for each FR (based on the content of D7.1) and match it to the tool and platform level validation presented in this deliverable. The way validation is streamlined with the piloting activities (as delineated in D7.1) is also discussed.

It has to be noted that these templates are based on the **initial selection** of the tools from the FRs, seen through D7.1 and also updated with the most recent tool updates at the time of writing D7.2 (May 2019). As this selection is not finalized, the corresponding lists are indicative and might be updated in the period that follows the publication of this document, based on the actual implementations and developments of the tools. In case more tools are added to the list, every new addition should be accompanied by an assignment to at least one FR (see the tables in Section 5.2), so that all tools get demonstrated and validated at least once.

5.1 Correlation between tools, modules and validation levels

The correlation between the different STOP-IT tools, their corresponding modules and the two levels of the validation planning (tool and platform level) are analysed in Table 8. Two main aspects are noted:

1. Most of the tools are a stand-alone encapsulation of a specific functionality offered by the STOP-IT toolkit and are thus able to be validated independently through the corresponding tool-level questionnaires. This applies for tools 1 (RIDB) to 22 (OPWS).
2. Some tools, notably tools 23 (REN) to 26 (IWM), constitute integral aspects of the STOP-IT platform and cannot be validated unless the STOP-IT platform is demonstrated as a whole. These tools include the ones belonging to modules VIII (Reasoning Engine) and IX (Enhanced Visualisation Interface for the water utilities) and are thematically linked to specific platform traits, such as platform usability and integrity. They thus form the basis of specific questions (belonging to distinct traits) in the platform-level questionnaire and the score of these traits reflects the performance of these particular tools.

Table 8: STOP-IT tools, corresponding modules and their validation levels.

A/A	Abbreviation	Tool Name	Module	Validation level	Notes
1	**RIDB**	Risk Reduction Measure Databse	I	stand-alone tool	
2	**InfraRisk CP**	InfraRisk for Cyber Physical threats	I	stand-alone tool	
3	**AVAT**	Asset Vulnerability Assessment Tool	I	stand-alone tool	

4	**SP**	Scenario Planner	I	stand-alone tool	
5	**RAET**	Risk Analysis and Evaluation Toolkit	I	stand-alone tool	
6	**RRMD**	Risk Reduction Measures Database	I	stand-alone tool	
7	**STP**	Stress Testing Platform	I	stand-alone tool	
8	**FTE**	Fault Tree Editor	I	stand-alone tool	Added 04/2019
9	**KPItool**	Key Performance Indicators tool	I	stand-alone tool	Added 04/2019 as part of D4.2
10	**Jdet**	Jammer Detector	II	stand-alone tool	
11	**NTSA**	Network Traffic Sensors and Analysers	III	stand-alone tool	
12	**RSDP**	Real-time sensor data protection	III	stand-alone tool	
13	**WQSP**	Optimisation Tool for Sensor Placement and Management		stand-alone tool	
14	**FTCS**	Fault-tolerant Control Strategies for Physical Anomalies affecting the SCADA system	I	stand-alone tool	
15	**CVT**	Computer Vision Tools	IV	stand-alone tool	
16	**FCAC**	Fine-grain Cyber Access Control	IV	stand-alone tool	
17	**Smart-Locks**	Access Control System using Electronic Locks	IV	stand-alone tool	
18	**HPD**	Human Presence Detection using WiFi signals	IV	stand-alone tool	
19	**CTsS**	Cyber Threat Sharing Service	V	stand-alone tool	
20	**RTAD**	Real-Time Anomaly Detector	VI	stand-alone tool	
21	**XL-SIEM**	Cross Layer Security Information and Event Management	VI	stand-alone tool	
22	**OPWS**	Optimised Public Warning System	VII	stand-alone tool	

23	**REN**	Reasoning Engine	VIII	platform, integral part	feeding to platform usability, integrity
24	**EVI**	Enhanced Visualisation Interface for the water utilities	IX	platform, integral part	feeding to platform usability
245	**RGIP**	RISA GEN Integration Platform		platform, integral part	feeding to platform integrity, usefulness
26	**IWM**	Interoperability Water Middleware		platform, integral part	feeding to platform integrity

5.2 Application overview

This section provides a brief overview of how validation is performed as part of a FR demonstration activity cycle. A demonstration activity cycle is defined as one or multiple demonstration events (organized as part of T7.4) that lead to all of the tools selected by each FR being demonstrated to him/her. Evidently, there can be multiple demonstration activities if many tools have been selected. In short, the end users have to fill:

- Questionnaires at the tool level **once for every tool demonstrated**, with the exception of tools that are integral parts of the STOP-IT platform and thus feed to platform traits (see Section 5.1). Tool questionnaires can be filled at the end of every tool demonstration activity.
- A questionnaire for the platform level **once** at the end of the demonstration activity cycle, i.e. once all tools have been demonstrated and the end users have completed their experience on the platform as a whole.

Table 1 summarizes these filling steps and also provides other aspects of the validation plan, such as the scope and target groups. A graphical depiction of the validation steps are given in Figure 8, which can be viewed as an extension of Figure 2. These steps are the following:

1. At the end of the demonstration activity, the end users (FRs) are requested to fill the questionnaires at tool level (one questionnaire needs to be filled per selected tool, with the exception of tools that are integral to the platform, according to Section). After filling the tool-level questionnaire, the platform-level questionnaire needs to be filled as well; this can be done only once at the end of the demonstration activity cycle, i.e. once all tools have been demonstrated and the FR has a clear picture on how the STOP-IT platform functions as a whole.
2. Once the questionnaires have been filled, they can be processed according to the instructions of Section 3.4 to calculate trait scores at the individual tool level. The scores can be displayed as a single metric per trait, or as a unified global metric per

tool (e.g. by assigning weights through Equations (1) and (2)); the former is suggested to be able to communicate strong and weak points for each tool in a more detailed fashion.

3. Trait scores at the whole platform level can be calculated as well. The platform scores can rely on the platform-level questionnaire only, or on a combination of the platform-level and tool-level questionnaires, as explained in Equation (5). Optionally and as explained in Section 3.4, the validation plan also allows aggregation at the module level, if this is desirable by the scope of T7.4/T7.5.
4. Once the results have been calculated and visualized, they can be further communicated to relevant stakeholders and used for reflection and dissemination activities. Moreover, they can provide direct feedback to tool developers in WP4, WP5 and WP6.

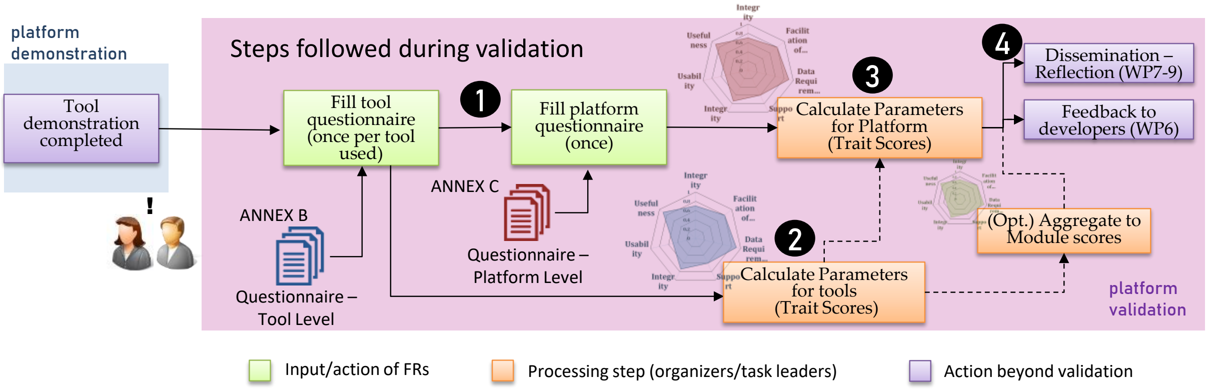

platform demonstration
Tool demonstration completed
!
Steps followed during validation
Fill tool questionnaire (once per tool used)
1
Fill platform questionnaire (once)
ANNEX B
Questionnaire – Tool Level
ANNEX C
Questionnaire – Platform Level
2
Calculate Parameters for tools (Trait Scores)
3
Calculate Parameters for Platform (Trait Scores)
4
Dissemination – Reflection (WP7-9)
Feedback to developers (WP6)
(Opt.) Aggregate to Module scores
Integrity
Facilitation of...
Data Requirem...
Support
Usability
Usefulness
platform validation
Input/action of FRs
Processing step (organizers/task leaders)
Action beyond validation

Figure 8: Overview of the steps that are followed during the validation process.

5.3 Application templates

5.3.1 AB

AB has initially selected a total of fourteen (14) tools for demonstration, including four (4) tools that are integral to the STOP-IT platform (i.e. belong to Modules VIII and IX). Two (2) recently added tools (FTE, KPItool) that are considered integral to RAET will be demonstrated as well, bringing the total number of demonstrated tools to sixteen (16). An overview of the selected tools is given in Table 9. The correlation of these tools to the validation plan is as follows:

- Tools 1 (InfraRisk CP) to 12 (Optimized Public Warning System OPWS) are validated through tool-level questionnaires, following the corresponding use cases described in D7.1. Based on the steps seen in Section 3.4.2, validation on these 12 tools is initially performed, following the distribution and completion of the relevant questionnaires by the FRs.
- Tools 13 to 16 constitute traits of the whole platform and are thus validated through a single platform-level questionnaire, following the corresponding use cases described in D7.1. In case multiple different members of AB join demo events for Tools 13-16, they have to collaborate to fill the platform-level questionnaire, according to their

experience. Alternatively, in case multiple members are fully exposed to the demo activities, multiple platform-level questionnaires can be filled (one per FR expert) and the final results can be averaged on the platform trait level.

- The filled questionnaires (12 in the tool level, 1 in the platform level) constitute the basis for validation for AB, according to the steps of 3.4.2.

Table 9: Selected STOP-IT tools for AB.

Selected STOP-IT tools AB
1. InfraRisk for Cyber Physical threats (InfraRisk CP)
2. Risk Analysis and Evaluation Toolkit (RAET)
3. Cyber-physical threats Stress-Testing Platform (STP)
4. Fault Tree Editor (FTE)
5. Key Performance Indicators tool (KPItool)
6. Scenario Planner tool (SP)
7. Risk Reduction Measures Database (RRMD)
8. Fault-tolerant Control Strategies for Physical Anomalies affecting SCADA systems (FTCS)
9. Computer Vision Tools (CVT) for automated surveying of the large-area of the water utility
10. Cyber Threat Sharing Service (CTSS)
11. Real-Time Anomaly Detector (RTAD)
12. Optimized Public Warning System (OPWS)
13. Reasoning Engine (REN)
14. Enhanced Visualization Interface for the water utilities (EVI)
15. Interoperable Water Middleware (IWM)
16. RISA GEN Integration Platform (RGIP)

5.3.2 BWB

BWB has selected a total of fourteen (14) tools for demonstration, including three (3) tools that are integral to the STOP-IT platform (i.e. belong to Modules VIII and IX). Two (2) recently added tools (FTE, KPItool) that are considered integral to RAET will be demonstrated as well, bringing the total number of demonstrated tools to sixteen (16). An overview of the selected tools is given in Table 10. The correlation of these tools to the validation plan is as follows:

- Tools 1 (RIDB) to 13 (Cross Layer Security Information and Event Management XL-SIEM) are validated through tool-level questionnaires, following the corresponding use cases described in D7.1. Based on the steps seen in Section 3.4.2, validation on these 13 tools is initially performed, following the distribution and completion of the relevant questionnaires by the FRs.
- Tools 14 to 16 (REN, EVI and RGIP) constitute traits of the whole platform and are thus validated through a single platform-level questionnaire, following the corresponding use cases described in D7.1. In case multiple different members of BWB join demo events for Tools 14-16, they have to collaborate to fill the platform-level questionnaire, according to their experience. Alternatively, in case multiple members are fully exposed to the demo activities, multiple platform-level questionnaires can be filled (one per FR expert) and the final results can be averaged on the platform trait level.

The filled questionnaires (13 in the tool level, 1 in the platform level) constitute the basis for validation for BWB, according to the steps of 3.4.2.

Table 10: Selected STOP-IT tools for BWB.

Selected STOP-IT tools BWB
1. Risk Identification Data Base (RIDB) 2. InfraRisk for Cyber Physical threats (InfraRisk CP) 3. Water Quality Sensor Placement Tool (WQSP) 4. Risk Analysis and Evaluation Toolkit (RAET) 5. Cyber-physical threats Stress-Testing Platform (STP) 6. Fault Tree Editor (FTE) 7. Key Performance Indicators tool (KPItool) 8. Scenario Planner tool (SP) 9. Risk Reduction Measures Database (RRMD) 10. Access Control System using Electronic Locks (Smart-Locks) 11. Fine-grain Cyber Access Control (FCAC) 12. Real-Time Anomaly Detector (RTAD) 13. Cross Layer Security Information and Event Management (XL-SIEM) 14. Reasoning Engine (REN) 15. Enhanced Visualization Interface for the water utilities (EVI) 16. RISA GEN Integration Platform (RGIP)

5.3.3 MEK

MEK has selected a total of eight (8) tools for demonstration, including one (1) tool that is integral to the STOP-IT platform (i.e. belong to Modules VIII and IX). An overview of the selected tools is given in Table 11. The correlation of these tools to the validation plan is as follows:

- Tools 1 (FTCS) to 7 (RTAD) are validated through tool-level questionnaires, following the corresponding use cases described in D7.1. Based on the steps seen in Section 3.4.2, validation on these 7 tools is initially performed, following the distribution and completion of the relevant questionnaires by the FRs.
- Tool 8 (EVI) refers to specific traits of the STOP-IT platform and has to be validated through a single platform-level questionnaire, following the corresponding use cases described in D7.1. The FR experts that will take part in the EVI demonstration will be required to fill the relevant traits of the platform-level questionnaire; the remaining pool of traits will be filled according to the general experience of the MEK members towards the STOP-IT platform. In case multiple members are fully exposed to the demo activities, multiple platform-level questionnaires can be filled (one per FR expert) and the final results can be averaged on the platform trait level.

The filled questionnaires (7 in the tool level, 1 in the platform level) constitute the basis for validation for MEK, according to the steps of 3.4.2.

Table 11: Selected STOP-IT tools for MEK.

Selected STOP-IT tools MEK
1. Fault-tolerant Control Strategies for Physical Anomalies affecting SCADA systems (FTCS)
2. Jammer Detector (JDet)
3. Network Traffic Sensors and Analysers (NTSA)
4. Real-Time Sensor Data Protection (RSDP)
5. Access Control System using Electronic Locks (Smart-Locks)
6. Human Presence Detection using WiFi signals (HPD)
7. Real-Time Anomaly Detector (RTAD)
8. Enhanced Visualization Interface for the water utilities (EVI)

5.3.4 VAV

VAV has selected a total of seventeen (17) tools for demonstration, including two (2) tools that are integral to the STOP-IT platform (i.e. belong to Modules VIII and IX). Two (2) recently added tools (FTE, KPItool) that are considered integral to RAET will be demonstrated as well, bringing the total number of demonstrated tools to nineteen (19). An overview of the selected tools is given in Table 12. The correlation of these tools to the validation plan is as follows:

- Tools 1 (RIDB) to 17 (Cross Layer Security Information and Event Management XL-SIEM) are validated through tool-level questionnaires, following the corresponding use cases described in D7.1. Based on the steps seen in Section 3.4.2, validation on these 17 tools is initially performed, following the distribution and completion of the relevant questionnaires by the FRs.
- Tools 18 and 19 (REN and EVI) constitute traits of the whole platform and are thus validated through a single platform-level questionnaire, following the corresponding use cases described in D7.1. In case multiple different members of AB join demo events for these two tools, they have to collaborate to fill the platform-level questionnaire, according to their experience. Alternatively, in case multiple members are fully exposed to the demo activities, multiple platform-level questionnaires can be filled (one per FR expert) and the final results can be averaged on the platform trait level.
- The filled questionnaires (17 in the tool level, 1 in the platform level) constitute the basis for validation for AB, according to the steps of 3.4.2.

Table 12: Selected STOP-IT tools for VAV.

Selected STOP-IT tools VAV
1. Risk Identification Data Base (RIDB)
2. InfraRisk for Cyber Physical threats (InfraRisk CP)
3. Water Quality Sensor Placement Tool (WQSP)
4. Risk Analysis and Evaluation Toolkit (RAET)
5. Cyber-physical threats Stress-Testing Platform (STP)
6. Fault Tree Editor (FTE)
7. Key Performance Indicators tool (KPItool)
8. Risk Reduction Measures Database (RRMD)

9. **Fault-tolerant Control Strategies for Physical Anomalies affecting SCADA systems (FTCS).**
10. **Network Traffic Sensors and Analysers (NTSA)**
11. **Real-Time Sensor Data Protection (RSDP)**
12. **Computer Vision Tools (CVT) for automated surveying of the large-area of the water utility**
13. **Access Control System using Electronic Locks (Smart-Locks)**
14. **Human Presence Detection using WiFi signals (HPD)**
15. **Cyber Threat Sharing Service (CTSS)**
16. **Real-Time Anomaly Detector (RTAD)**
17. **Cross Layer Security Information and Event Management (XL-SIEM)**
18. **Reasoning Engine (REN)**
19. **Enhanced Visualization Interface for the water utilities (EVI)**

5.4 Links with piloting activities

Besides delineating demonstration use cases per tool, D7.1. offers methodological information on the approach of the piloting activities. This information can be reviewed through the perspective of the proposed validation plan. The types of pilot activities (discussed in Section 3.2. of D7.1), as well as operational aspects of the demonstration activities (discussed in Section 3.3. of D7.1) are of interest to the concepts discussed in the validation report and are thus further analysed.

5.4.1 Links with piloting activity horizons

The pilots, as specified in D7.1, consist of a combination of three main activities: 1) short-term (simulation) piloting, 2) long-term piloting and 3) usability tests.

Short-term piloting aim at evaluating the feasibility and the technical performance of the STOP-IT system in order to validate a specific technology. In addition to this, as explained in D7.1, they offer a rapid assessment interface from FR representatives who gain first-hand experience of using the platform, its features and its implementation. This means that first impressions from participants can be also collected at the end of each activity. This definition enables both tool-level questionnaires and platform-level questionnaires to be used as part of the validation process for short-term pilots; tool-level questionnaires can validate whether the specific technology that was demonstrated was successful, while platform-level questionnaires will capture the ‘first impression’ the STOP-IT platform makes to the FRs as a whole. Besides validating specific technologies and capturing first impressions, the questionnaire could be useful to identify bottlenecks and technical issues related to the implementation of both the tools and the platform, since open questions are included as well; the STOP-IT developers can then capture this feedback and perform improvements that will be reflected in other piloting activities, such as the long-term pilots.

Long-term piloting, according to D7.1. has the aim of moving beyond an evaluation of the technical solution, in order to assess the conceptual aspects of the idea and how it impacts the users. Thus, the long-term pilots focus more on the operational aspects of the concept.

Given this definition, this type of piloting activities is leaning towards the validation of the *usefulness* trait of both specific tools as well as the platform as a whole. It is thus suggested to perform another validation cycle, using the same tool and platform-level questionnaire templates, in order to better capture how useful STOP-IT products were for the FRs on the long terms. Besides usefulness, reflecting on the performance of the other traits (especially the difference between short-term and long-term validation scores) is useful in order to check if any critical aspects of the 'first impressions' FR had have been taken into account through design improvements in later iterations of the STOP-IT platform.

The third type of piloting activities, usability testing, aims at discovering artefacts or unexpected behaviour while interfacing with the STOP-IT platform, in order to improve user experience. These artefacts can be captured and documented through the open questions of the questionnaires of the validation plans. If needed, shorter questionnaires (e.g. that only have open questions) that target usability testing can be deduced from the templates offered in this deliverable.

5.4.2 Links with piloting activity types

As described in D7.1, the pilots will involve a combination of surveys, focus groups and usability tests. According to D7.1, surveys are intended to allow the consortium to collect a broad range of data from users participating in both the long-term and short-term pilots, asking participants about their experience and thoughts on the demonstrated technologies, as well as suggestions on improvements. These intentions can be by definition captures through the questionnaire templates provided in D7.2.

A notable difference is that D7.1. suggests three layers of surveys for long-term piloting (pre, during and post), as well as no survey executions for the short-term pilots. However, the templates offered here are applicable and can be insightful to both short-term and long-term piloting horizons, as explained in Section 5.4.1. The decision on whether to include validation or not in all piloting phases has to be taken after piloted activities have been planned in detail, in order to better capture all aspects of the user experience; in case planning focus on the long-term and the FR experience from short—term piloting is very limited, the validation process in that phase can be omitted, with the long-term piloting activities being the focus of the validation. However, in case planning is evenly distributed over time and short-term piloting activities offer a complete experience with respect to the STOP-IT tools and platform, it is suggested to validate at the short-term scale as well, even if this validation is applicable to only a handful of tools.

Following the pilot testing activities, D7.1. also projects a number of focus groups, which will include the FR personnel and technical interested providers that took part in the trials, as well as additional participants from the LCoP. The proposed focus groups are intended to follow on after the long-term pilot trials in the four FRs and aim at eliciting qualitative insights into the experiences and views of participants who used the STOP-IT tools during the trial period. Evidently, the outcome of the validation questionnaires at the earlier piloting phases can provide a 'reflection basemap' for these focus groups. The participant-led discussions can be supported by data obtained through validation, in order to better capture, demonstrate and provide discussion points for the user experience.

5.5 Data collection and management during validation

The proposed validation plan, applied in the form of questionnaires handed to FR participants in the demonstration activities, will undoubtedly collect personal information on user preferences, views and suggestions of improvement. This section defines aspects of the data management that are important to ensure that the obtained data will be properly handled and stored during and after the completion of the pilot activities.

All information that has been generated or collected as part of the validation will be immediately anonymized and stored securely until immediately after the final project review meeting, **and in any case no longer than 6 months after that date**. Anonymization will happen at the individual user level (e.g. FR employee) who filled the questionnaire. The anonymized data will be kept for further analysis or in case information needs to be cross-validated. During this phase, the data will only be accessible by the consortium partners who have tasks directly relevant to the validation and reflection activities (T7.4 and T7.5). While the users who filled the questionnaires will be directly fully anonymized, the link between validation results and each FR will remain visible to the restricted pool of the consortium partners who have tasks directly relevant to the validation and reflection activities (T7.4 and T7.5), in order to better guide improvements of the STOP-IT platform and communicate results to each FR.

Likewise, external analysis and publication of the validation results is possible only after full anonymization in both users and FRs. Anonymized data will be used to drive comparative analyses and activities related to T7.4 and T7.5 across the three pilot sites, during the project. Each participant will be assigned a unique ID and this will be used to link their responses during analysis. Survey responses will not be attributable to participants. For cases where the participants have provided personalized details, the information will be filtered (extracted) after the collected data has been validated, and will therefore not be part of the data processing itself. The questionnaires themselves do not require extensive personal information and the personal details requested are always clearly stated at the beginning of the questionnaire in order to facilitate the filtering process.

Following the final project review meeting, the STOP-IT partners will retain anonymized data from all pilot sites **for a maximum period of 6 months**. All reflection activities, analyses and publications that depend on these datasets should be completed at the end of this period; once that period is reached, all data (incl. the anonymized datasets) will be promptly deleted.

6 Conclusions

The deliverable 7.2 describes the methodological approach and application of the validation plan, which provides a validation planning based on two levels (tool and platform level) based on a number of distinct traits. By employing both of these validation levels during the demonstration activities (WP7), STOP-IT technology providers are able to evaluate whether the end users of STOP-IT (FRs) are satisfied with different aspects of the demonstrated tools, including their functionality, usability, usefulness and data requirements, as well as their underlying supporting processes, such as installation and learning support. These aspects constitute the parameters that are important to the validation of the STOP-IT system, both at the level of individual tools (WP4-WP5) as well as at the level of the whole encapsulation (WP6).

Following the validation plan, the proceeding steps of the WP7 include the detailed planning of demonstration activities, using both D7.1 and D7.2 as guidelines, as part of T7.4. This planning will design piloting activities and match the validation process, demonstrated in this report, to specific activities and timelines per FR. More specifically, the questionnaires provided as part of D7.2 will be filled from the FRs as part of the T7.4 activities. The questionnaire results can be in turn used for reflection and feedback to the developers, providing feedback for relevant tasks and deliverables (e.g. T7.5 or D6.7). Following the completion of the demonstration activities and the validation process, T7.5 will analyse the data from the filled questionnaires (using the metrics provided in Section 3.4 as a guideline) and reflect on the impact the demonstration activities had for the operations of the FRs; the validation plan will be evaluated as well as part of this process.

Since the STOP-IT tools and platform (WP4, WP5 and WP6 outcomes) are not yet finalised, the validation plan is designed in order to provide a flexible methodological base that holds even if more tools are added to the STOP-IT pool or if the selection of an individual FR changes. The validation methodology also holds, with minor changes, in case future WP7 activities identify end user experience areas (i.e. model or platform traits and partial characteristics) that have been omitted in the present validation plan; in that case, individual questions or sections can be added, subtracted or altered from the questionnaire templates. The validation plan is thus expected to be valid, with slight changes, in case the FRs select different tools, in case more tools are added to STOP-IT or in case the functionality of some tools is altered, provided that there is at least one demonstration of each tool in any FR. Likewise, the validation plan is valid, with slight changes to the provided questionnaires, in case different tool and platform attributes (traits) need to be outlined, explored and evaluated.

ANNEX A: Brief Description of the STOP-IT Tools

This section provides a brief description of the STOP-IT tools, based on information contained in D7.1. This information is provided here for continuity and consistency reasons. The reader is prompted to D7.1 and D6.1 in case more information on the STOP-IT functions, architecture and general framework is needed.

1.) Risk Identification Data Base (RIDB)

Risk Identification Database (RIDB)	
Brief Description of the Tool	RIDB is a list of risk events i.e. examples that assist the users in risk identification step. Risk events are related to the physical and/or cyber threats, which can occur in water distribution systems utilities. The RIDB identifies the type of threats, the sources of risk, the description of the events and the type of consequences produced. The RIDB is not considered as result of a comprehensive review, but as a list of events more relevant to the FRs.
Corresponding Module	Module I
Type of Threat Addressed	Cyber, physical and their combination
Required Input	Standalone MS-EXCEL file requesting manual input of risks/threats/events
Required Output	Repository of high-level descriptions related to risks/threats/events

2.) InfraRisk for Cyber Physical threats (InfraRisk CP)

InfraRisk for Cyber Physical threats (InfraRisk CP)	
Brief Description of the Tool	InfraRisk CP is a tool, based on the existing InfraRisk tool of SINTEF and upgraded to ensure compatibility with RIDB and RRMD, for assisting risk analysis of critical infrastructure and interdependencies with focus on cascading effects. It can be used to perform risk assessment at a generic level based on expert judgment.
Corresponding Module	Module I
Type of Threat Addressed	Cyber, physical and their combination
Required Input	• Event information • Physical System information (network/components) • Control and monitoring systems with connections • Reliability data of components (failure and repair)

	• Organizational and operational concerns
Required Output	• A "risk picture" in form of a Risk Matrix (based on the PHA) • Societal Critical Functions improvement potential

3.) Asset Vulnerability Assessment Tool (AVAT)

Asset Vulnerability Assessment Tool (AVAT)	
Brief Description of the Tool	AVAT is a tool (available as desktop or online application) acting as a procedural "step-by-step" guide for the assessment of vulnerability of water distribution system assets taking into consideration the specific characteristics of the assets, the importance of the components for water supply and their "attractiveness" to be attacked. AVAT calculates system wide and element-specific indexes requiring limited data from users and provides fast initial assessment of vulnerable areas in the network and the criticality of assets.
Corresponding Module	Module I
Type of Threat Addressed	Cyber, physical and their combination
Required Input	• A steady state hydraulic simulation EPANET (.inp file) • A data MS-Excel file, with a specific structure, containing default values of analysis, specific elements probabilities and system’s sources of the EPANET model
Required Output	• Todini’s Resilience Index • Connectivity Index • Node Reachability Index • Link Criticality Index

4.) Scenario Planner (SP) tool

Scenario Planner tool (SP)	
Brief Description of the Tool	Scenario Planner tool (SP) is a user-friendly graphical environment used for the development and design of threat scenarios and their effects (e.g. “triggered” interconnected, multiple occurring events, etc.) based on STOP-IT generic and predefined Fault Trees (FT). SP tool, which is a standalone application, enables users to design and configure scenarios to be examined in simulation platforms such as the STP. For the latter, SP automatically generates setup scenarios (through a wizard).
Corresponding Module	Module I

Type of Threat Addressed	Cyber, physical and their combination
Required Input	• Fault Trees • Identified Risks (RIDB) • Risk Reduction Measures (RRMD) • Tools (RAET)
Required Output	• Scenarios which may include: ○ Selected risks represented by activated paths of events in FTs ○ Selected risk reduction measures addressing the risks • Selected tools to simulate and analyze the risks

5.) Risk Analysis and Evaluation Toolkit (RAET)

Risk Analysis and Evaluation Toolkit (RAET)	
Brief Description of the Tool	Risk Analysis and Evaluation Toolkit (RAET) is a new web application that aims to support stakeholders in analysing and evaluating risks of cyber-physical threats, by providing access to and support the use of specific tools and models. Apart from encapsulating the tools developed under WP4 of STOP-IT (e.g. AVAT, SP, etc.), it assists the exploration of suitable models worth-considering for analysing and modelling water related problems. RAET provides also a set of KPIs for the analysis of the system performance under different scenarios.
Corresponding Module	Module I
Type of Threat Addressed	Cyber, physical and their combination
Required Input	• Selected infrastructure • Potential consequence of interest (water quantity, water quality, economic etc.)
Required Output	• Tools suitable to analyse and evaluate risks

6.) Risk Reduction Measures Database (RRMD)

Risk Reduction Measures Database (RRMD)	
Brief Description of the Tool	Risk Reduction Measures Database (RRMD) is a new web application that facilitates the identification, selection and prioritization of appropriate risk reduction measures as actions, activities or processes that can be applied in order to reduce the occurrence and minimize consequences of events. This RRMD contains qualitative indicators and application examples for selected certain types of events.

Corresponding Module	Module I
Type of Threat Addressed	Cyber, physical and their combination
Required Input	Risks documented in the RIDB
Required Output	Measure(s) matched with the given risks

7.) Cyber-physical threats Stress Testing Platform (STP)

Cyber Physical threats Stress Testing Platform (CPSTP)	
Brief Description of the Tool	Cyber-physical threats Stress-Testing Platform (STP) is an EPANET based platform which is currently being developed to provide a simulation environment for both physical and cyber sub-systems. The aim is to assess the behaviour of the cyber physical water system by deliberately stressing it under different attack scenarios, which can be developed through the Scenario Planner tool. STP will assess the system's response to a given attack and also allow the user to then simulate selected RRMs (from RRMD) and assess their performance against attack scenarios using a set of KPIs.
Corresponding Module	Module I
Type of Threat Addressed	Cyber, physical and their combination
Required Input	• Selected Tool, such as a Water Network Model based on EPANET • Selected scenario from Scenario Planner
Required Output	• Key Performance Indexes (KPI) • Points/results of interest on the Water Network based on the analysis executed

8.) Fault Tree Editor (FTE)

Fault Tree Editor (FTE)	
Brief Description of the Tool	Fault Tree Editor (FTE) is a tool for creating, editing and modifying fault trees. The FT Editor tool is a graphical fault tree user interface which has been initially developed for the needs of WP6. The latest version of the tool (v1.1.6) supports, among other options, the calculation of failure probabilities from the fault trees in case the probabilities of basic events have been defined.
Corresponding Module	Module I

Type of Threat Addressed	Cyber, physical and their combination
Required Input	• RIDB contents
Required Output	• Fault Tree designs

9.) Key Performance Indicators tool (KPItool)

Key Performance Indicators tool (KPItool)	
Brief Description of the Tool	The KPI tool is a MATLAB®- based standalone executable designed to assist in the evaluation process (by displaying suitable KPIs) of a threat scenario within the WP4 tactical and strategic planning of the water sector against CP attacks. The aim of the tool is to present a user friendly environment to deploy the STOP-IT KPI Framework in a structured way, based on the results of the RAET toolkit models.
Correspondin g Module	Module I
Type of Threat Addressed	Cyber, physical and their combination
Required Input	• A model of the water network based on EPANET • Selected scenario results (.csv file)
Required Output	• A graphical environment with Key Performance Indexes (KPI)

10.) Jammer Detector

Jammer Detector (JDet)	
Brief Description of the Tool	JDet provides security of Wireless Sensor Network (WSN) communications based on the Bit carrier product (owned by WS) which will be enriched with a Software Defined Radio module in order to analyse the wireless channel spectrum of several technologies with a unique solution.
Correspondin g Module	Module II
Type of Threat Addressed	Cyber
Required Input	Standalone Module
Required Output	Alert describing detected jamming models

11.) Network Traffic Sensors and Analysers

Network Traffic Sensors and Analysers (NTSA)	
Brief Description of the Tool	NTSA is a tool developed for the analysis of Netflow traffic data generated by routing and switching devices to detect anomalous behaviour in the traffic. By analysing the network traffic, it is possible to identify the normal behaviour of the system e.g., by defining the number of packets transferred during a given period of time, the volume of packets sent and received, the IP sources/destinations used in the communications, the port sources/destinations required for communications, the protocols used, etc., therefore, everything that falls outside this will be considered as suspicious, and the tool will alert the systems accordingly.
Corresponding Module	Module III
Type of Threat Addressed	The tool addresses network anomalies e.g., high volume of traffic during a given period of time; communications coming from unknown or malicious IP sources; communications going to unknown or malicious IP destinations; suspicious ports/protocols connections; and other actions that could lead to attacks such as brute force, DoS, and botnets.
Required Input	Netflow dataset about the network traffic to be used to train the model that will make predictions about the normal/abnormal behaviour of the system/network
Required Output	The model will provide data and images of the region considered to capture the normal/legitimate traffic points as well as the points that fall out of the region (which are considered to be anomalous)

12.) Real-time sensor data protection

Real-time sensor data protection (RSDP)	
Brief Description of the Tool	RSDP applies blockchain schemes to protect the integrity of all the data generated during a CI operation (logs, sensor data, etc.), both against intentional attacks or malfunction.
Corresponding Module	Module III
Type of Threat Addressed	Cyber
Required Input	Sensor data to be stored in the Cloud or in an alternative storage system and the identification of the device which generated that data.
Required Output	The requested data and the result of the data integrity test.

13.) Optimisation Tool for Sensor Placement and Management

Optimization Tool for Sensor Placement and Management	
Brief Description of the Tool	This tool is aimed at providing a sensor placement and sensor management framework, incorporating three interrelated modules: (1) Water Quality Sensor Placement Tool (WQSP), (2) Classification Model for Contamination Event Detection (CMCED), and (3) Contamination Source Intrusion Detection (CSID). An optimization methodology will be applied/developed for water quantity (hydraulic) and water quality sensor placement, a method for event detection of water quality intrusions, and a scheme for contamination source identification.
Corresponding Module	The tool will have three main modules: (1) a multi-objective optimization settings for the conjunctive placement of hydraulic and water quality sensors in water distribution systems, using evolutionary optimization such as genetic algorithms, (2) a module for event detection utilizing water quantity and water quality data collected by the sensors, and (3) an event detection module for estimating the most probable source intrusion locations, based on the water distribution system layout and information received from the water quantity and water quality sensors.
Type of Threat Addressed	The type of threat addressed is a contamination intrusion into a water distribution system. This intrusion can be a result of a terrorism action of deliberately injecting contaminants into the system or an occasional intrusion such as a pollutant entering a well.
Required Input	Water distribution system modelled in EPANET
Required Output	Locations of hydraulic and water quality sensors

14.) Fault-tolerant Control Strategies for Physical Anomalies affecting the SCADA system

Fault tolerant Control Strategies for Physical Anomalies affecting SCADA system (FTCS)[4]	
Brief Description of the Tool	FTCS focuses on the quantitative operation / management of the drinking water transport network through the SCADA system considering its integration with the distribution network. Mainly, for the case of the STOP-IT project, those anomalies / asset failures forcing the isolation of several parts of the network (i.e. for repairing the faulty assets) are considered. Once, a part of the network is isolated, the network has to be reconfigured to avoid the negative impact on supplying current and future demands. The tool will be used to assess the impact of isolating a given part of the network, and to support the decision-making process associated with the selection of the best reconfiguration that can be applied in order to avoid this

[4] Sub-component 1 of this tool includes the Methodology for physical anomaly detection affecting integrated SCADA assets while sub-component 2 relies on Fault-tolerant Control Strategies for Physical Anomalies affecting SCADA integrated Sensors/Actuators.

	impact. Mainly, this isolation process is triggered by physical attacks or malfunctions affecting the network assets (i.e. valves, pipes, pumps). From the functional point of view, the proposed tool relies on the fault-tolerant control (FTC) concept applied to the control strategies implemented in the SCADA system governing the transport network performance through the different existing actuators (i.e. pumps, automatic controlled valves) and using the information collected by the existing telemetry (i.e. pressure, flow and level sensors). Behind the FTC concept, there are two main sub-components: • Fault Diagnosis based on detecting the existence of an anomaly and isolating the faulty network component. Additionally, it may also give a rough estimation of the fault impact. • Controller reconfiguration: with the information provided by the Fault Diagnosis module, the control strategies implemented in the control module are re-design in order to minimize the impact of the faults in terms of the system performance. The STOP-IT project will focus mainly of the 2nd mechanism (Controller reconfiguration) since the network reconfiguration process is triggered by the network operator when the anomaly is detected/located using the SCADA system and the faulty assets are isolated for their reparation.
Corresponding Module	Module I
Type of Threat Addressed	The specific application of this technology corresponds to a problem raised by AB FR in which an unexpected (natural or intentional) failure affects the water transmission network which is monitored and controlled using a SCADA system. This network conveys drinking water from the treatment plants to the different consumptions points which are grouped in DMAs and when an event of this type occurs, water supply service to DMAs may be affected negatively in terms of volume and pressure. The intelligence implemented in the SCADA systems supports SCADA technicians to locate the faulty elements in this type of network and to set those interventions which allow to isolate the smallest part of the network containing the failure so that it can be repaired. In general, this isolation process may lead to consumers that completely lost the water service, consumers that have a drop in pressure and if the problem persists long enough to drain tanks, may lose service (Walski 2015) and unaffected consumers. The negative impacts of the isolation on the system are reduced in the second phase through recovery, which involves providing an optimal combination of operational interventions for network operators as temporary alternatives until pre-event failure conditions are restored. In large cities, a network of this type is segmented in PMZs using boundary valves. Inside every PMZ, DMAs are supplied using specific control points of the transmission network. However, for failure conditions, there are alternative paths enabled by opening/closing boundary valves that may bring water within the DMA from nearby DMAs inside the same PMZ or from a nearby PMZ, from other points of the transmission network (same PMZ or from another). Additionally, set-points of nearby PRVs and pumping stations can also be modified in order to avoid pressure drops and to keep tank levels inside safety values while the failure condition persists.

	In the STOP-IT project, the main focus is set on the recovery phase once the failure (i.e. pipe breakdown) is isolated. The proposed technology is adapted in order to optimize the process of selecting the needed interventions in order to minimize the negative impact of the failure isolation. This methodology deals with those alternative network configurations that in normal conditions are disabled.
Required Input	Configuration data: • List of tanks, pumps, valves, demands nodes • Hourly forecast of the electricity tariff for the next 24h hours SCADA data: • Hourly forecast of the demands for the next 24h • Initial state of the network elements (i.e. tank levels) (computed by the 1-step hydraulic model simulation) Hydraulic model (PICCOLO[5])
Required Output	Mitigation Strategies: Set-points for the network actuators (pumps and controlled valves) computed by the control module. Network performance – time evolution of the hydraulic variables

15.) Computer Vision Tools

Computer Vision Tools (CVT)	
Brief Description of the Tool	CVT combines computer vision and machine learning tools for automated surveying of water utilities' critical infrastructure, using a network of cameras to detect suspicious behavior.
Correspondin g Module	Module IV
Type of Threat Addressed	Physical
Required Input	OpenCV3 compatible video format
Required Output	Suspicious Behavior alert

16.) Fine-grain Cyber Access Control Tool

Fine-grain Cyber Access Control (FCAC)	
Brief Description of the Tool	The Fine-grained Cyber Access Control tool (FCAC), is an XACML-based authorization engine deployed in areas and resources system. Depending on the Policy Enforcement Point (PEP), the FCAC could perform access control for both cyber and physical entities. The process starts when an entity (e.g., a user, a machine) tries to access a system's resource (e.g., a file, a database, a printer, etc.), for which the system must evaluate if the permission is granted

[5] Piccolo is a hydraulic modelling tool for water networks developed by Suez group being used widely in Suez water utilities (http://www.safege.com/en/innovation/modelling-and-smart-solutions/)

	or denied. The Policy Enforcement Point (PEP) can be either a cyber-entity (e.g., software) or a physical entity (e.g., machine, equipment, tool) that communicates with the FCAC through the Policy Decision Point (PDP). Once this latter has the appropriate security policy in its possession, it will check if the entity has the right to access the resource, in such a case, an XACML response is generated to the PEP in order to grant the permission to the entity. Otherwise, a deny XACML response is generated. The PEP will enforce this response by authorizing or denying the requested access.
Corresponding Module	Module IV
Type of Threat Addressed	This tool is able to detect any user/entity trying to access unauthorized resources in the system. Please note that this tool does not address physical threats, it only addresses cyber threats in which access control violations may occur. The tool can be integrated with other tools/components that can act as physical enforcement points and can deal with physical threats.
Required Input	• Access request from a given entity • Security policies
Required Output	Access Response (denied/granted access) provided to the evaluated entity

17.) Access Control System using Electronic Locks

Access Control System using Electronic Locks	
Brief Description of the Tool	Smart-Locks are access control systems based on intelligent electronic locks, and dedicated applications to service employees and to central management system.
Corresponding Module	Module IV
Type of Threat Addressed	Physical
Required Input	Standalone module
Required Output	Access information logs

18.) Human Presence Detection using WiFi signals

Human Presence Detection using WiFi signals (HPD)	
Brief Description of the Tool	HPD is a technology which uses WiFi signals to analyze human body reflection to detect presence of persons in restricted areas.

Corresponding Module	Module IV
Type of Threat Addressed	Physical
Required Input	Standalone Module
Required Output	Intrusion Detection Alert

19.) Cyber Threat Sharing Service

Cyber Threat Sharing Service (CTsS)	
Brief Description of the Tool	This tool collects sources of existing threats from relevant feeds, structuring the information using standards to facilitate the exchange of the security identified threats.
Corresponding Module	Module V
Type of Threat Addressed	Cyber
Required Input	Cyber Threat Incidents (STIX™ v2 messages)
Required Output	Cyber Threat Incidents (STIX™ v2 messages)

20.) Real-Time Anomaly Detector

Real-Time Anomaly Detector (RTAD)	
Brief Description of the Tool	RTAD combines cyber, physical, behavioral and surrounding context information to detect unknown anomalies. This tool provides an additional layer of security by detecting potential threats from the logs of the system.
Corresponding Module	Module VI
Type of Threat Addressed	Situation Addressed (Core Modules)
Required Input	Input from various data modalities
Required Output	Alert based on threat classification

21.) Cross Layer Security Information and Event Management (XL-SIEM)

Cross Layer Security Information and Event Management (XL-SIEM)	
Brief Description of the Tool	XL-SIEM is a tool that works as an enhanced security data analytic platform, which receives events coming from different sources to generate correlated alarms that indicate the risk level, and detailed information about the event (description, IP source and destination, Port source and destination, Protocols).
Corresponding Module	Module VI
Type of Threat Addressed	The XL-SIEM is able to detect a wide variety of threats, as it receives feeds from intrusion detection systems (IDSs) and other types of sensors that evaluates and analyses the behaviour of the network/system. As such, the tool is able to detect brute-force attacks, network scans, policy violations, malware, web attacks, service attacks, and any kind of anomaly with a predefined pattern.
Required Input	Event Logs
Required Output	• Generated events based on malicious activities detected on the system • Risk level associated to each event, along with information about the IP source, destination, timestamp and sensor involved in the detection process • Alarms generated through the correlation of detected events • Risk Levels associated to each alarm, along with the number and description of correlated events, the duration of the anomalies, the IP source, and destination involved in the event.

22.) Optimised Public Warning System

Optimised Public Warning System (OPWS)	
Brief Description of the Tool	OPWS is a tool that has two main functions: the first is to allow the acquisition of incident data from external sources, and the second to notify relevant actors (including surrounding population) of an accident in order to protect citizens and decrease the impact of the event. The tool will receive alert events from external sources (to improve anomaly detection and include cascading effects) and will forward alarms and warnings produced by other STOP-IT tools to send Public Warning Messages to CI operators’ workforce and also to citizens reception device.
Corresponding Module	Module VII
Type of Threat Addressed	Cyber / Physical

Required Input	• External alerts [optional] (inputs as sensors, Social networks, Cascading events) • Confirmed Incident • Warning action treatment from other modules (as defined in D6.1, Module IX will send a list of communication actions)
Required Output	Warning message

23.) Reasoning Engine

Reasoning Engine (REN)	
Brief Description of the Tool	Reasoning Engine is a tool responsible of a continuous assessment of an organisation's risk exposure and response plan management by executing specific algorithms as a set of machine-readable model rules.
Corresponding Module	Module VIII
Type of Threat Addressed	Physical & Cyber
Required Input	• Detected risks (physical & cyber). • Scenarios assessment available from WP4. • The model of the water utilities CIs is necessary in the form of fault trees.
Required Output	Mitigation actions and risk exposure.

24.) Enhanced Visualisation Interface for the Water Utilities

Enhanced Visualisation Interface for the water utilities (EVI)	
Brief Description of the Tool	EVI for the water utilities allows exploitation of the integrated data collected from the front-end and makes it available to the water utilities. In addition to specific applications, browser-based user interface will be made to control the workflow orchestration of the overall software system.
Corresponding Module	Module IX
Type of Threat Addressed	The EVI aims to help water utilities effectively respond to complex situations in disaster management by providing risk information available from the multiple modules involved in STOP-IT.
Required Input	• Detected events ○ Identified threats from Module II ○ RT fault diagnosis of sensors and assets operated by SCADA from Module III ○ Detected human presence & detected contamination event from Module IV

	○ Detected cyber threats from Module V ○ Detected threats from Module VI • Raw data ○ Water distribution model (also used in Module I) ○ Camera input (also used in Module IV) ○ Water quality sensors (also used in Module IV) ○ Fault trees (input from WP3 activities) ○ Risks, reduction measures and their connections (from RIDB and RRM of Module I) • Assessment results ○ Assessed vulnerability of assets from Module I (Asset Vulnerability Assessment Tool) ○ Assessed impact of potential incidents from Module I (Risk Analysis and Evaluation Toolkit) ○ Assessed risk exposure from Module VIII Response plans/suggestions for mitigating actions from Module VIII
Required Output	Various Visualisations on operators’ screen

25.) RISA GEN Integration Platform

RISA GEN Integration Platform (RGIP)	
Brief Description of the Tool	RGIP tool uses enhanced RISAGEN with parallel processing capabilities and dataflow schedulers, supporting modellers’ scaling, parallelisation and distributed execution.
Corresponding Module	Other components
Type of Threat Addressed	n/a
Required Input	Interfacing needs of modules
Required Output	Common data model

26.) Interoperability Water Middleware

Interoperability Water Middleware (IWM)	
Brief Description of the Tool	IWM is an interoperable and standard based middleware, that uses semantic capabilities to orchestrate platforms using mediation and matchmaking techniques.
Corresponding Module	Other components
Type of Threat Addressed	n/a
Required Input	Specific communication protocols defined in STOP-IT modules platform.

Required Output	JSON-LD with harmonized information.

ANNEX B: Questionnaire template – tool level

The following questions, directed at the end user (FR), aim at validating the performance of a tool within the STOP-IT platform, based on a set of **traits**, given as numbered sections, along with their partial characteristics, given as individual questions.

To fill this questionnaire, please provide:

- your ranking, in case of grading questions. If needed, an explanation of the different grades is provided below each question.
- your feedback, in case of open questions or conditional (Yes/No) answers.

Most questions are based on a grading/ranking evaluation that ranges from 1 (poor performance) to 5 (great performance). Open questions supplement some sections, allowing you to provide feedback back to STOP-IT tool developers.

1. Introduction

Please fill in the required information.

Demonstration Event Date: __ / __ / ____

Front Runner: ………………………………..

First Name: ……………………………………………………

Last Name: ……………………………………

Job Role in the FR: ………………………………………

Which is the tool that was demonstrated and that you are reviewing?

❑	RIDB	Risk Reduction Measure Databse	❑	CVT	Computer Vision Tools
❑	InfraRisk CP	InfraRisk for Cyber Physical threats	❑	FCAC	Fine-grain Cyber Access Control
❑	AVAT	Asset Vulnerability Assessment Tool	❑	Smart-Locks	Access Control System using Electronic Locks
❑	SP	Scenario Planner	❑	HPD	Human Presence Detection using WiFi signals
❑	RAET	Risk Analysis and Evaluation Toolkit	❑	CTsS	Cyber Threat Sharing Service
❑	RRMD	Risk Reduction Measures Database	❑	RTAD	Real-Time Anomaly Detector
❑	STP	Stress Testing Platform	❑	XL-SIEM	Cross Layer Security Information and Event Management
❑	FTE	Fault Tree Editor	❑	KPItool	Key Performance Indicators tool
❑	Jdct	Jammer Detector	❑	OPWS	Optimised Public Warning System
❑	NTSA	Network Traffic Sensors and Analysers	❑	WQSP	Optimisation Tool for Sensor Placement and Management
❑	RSDP	Real-time sensor data protection	❑	FTCS	Fault-tolerant Control Strategies for Physical Anomalies affecting the SCADA system

❑ Other (Please specify): ………………………………………………………………………………………..

Important Note: In case you are reviewing on of these tools: **REN** (Reasoning Engine), **EVI** (Enhanced Visualization Interface for the water utilities, **RGIP** (RISA GEN Integration Platform) or **IWM** (Interoperability Water Middleware), please use the platform questionnaire for validation instead.

2. Ease of installation

This set of questions gives insights on aspects of the data input the tool requires and works with.

Did you install this tool locally or is it a web/cloud service?

This was a local service, installed to my in—house hardware/software system (e.g. my work computer).	This was a web or cloud service, accessed online or through my intranet service.
❑	❑
(proceed to Sections 2.1 & 2.3)	(proceed to Sections 2.2 & 2.3)

2.1. Local installation

<u>If you installed the tool locally</u>, how would you rate the installation process in terms of:

a.) Installation time needed

(very long)		(reasonable)		(very fast)
1	2	3	4	5
❑	❑	❑	❑	❑

b.) Installation process simplicity/complexity:

(very complex)		(reasonable)		(simple and concise)
1	2	3	4	5
❑	❑	❑	❑	❑

Very complex: The installation process was lengthy and required special knowledge, e.g. installation of other tools first or technical hardware skills. As such, it had to be done by specialized personnel.

Simple and concise: The installation process was very simple and could be readily performed by me, without extra steps or pre-installation needs.

c.) Dependence on my current system:

(absolute dependence on current system)		(partial dependence)		(absolute independence)
1	2	3	4	5
❑	❑	❑	❑	❑

Absolute dependence on my current system: The demonstrated tool is ad-hoc and can only be run in the system it was installed. Any changes in the system will render the tool unable to run without substantial effort.

Partial dependence: The demonstrated tool is made to run on a specific standard of systems, e.g. specific versions of Windows.

Absolute independence: The demonstrated tool is cross-platform and able to run seamlessly in different systems.

d.) Integration with my current system:

(very limited integration)		(limited integration)		(seamless integration)
1	2	3	4	5
□	□	□	□	□

Very limited integration: The demonstrated tool was not able to run with my current system specs or required substantial effort to be installed in my current system.

Limited integration: The demonstrated tool was able to run with my current system, albeit with some effort and/or after installing some third-party software

Seamless integration: The demonstrated tool was installed easily and integrated fully with my current system, without the need from my side to change parts of my system.

e.) Dependence on third-party software/hardware:

(absolute dependence on commercial software/hardware)		(dependence on open-source software/hardware)		(stand-alone application)
1	2	3	4	5
□	□	□	□	□

Absolute dependence on commercial software: The demonstrated tool is fully dependent on software, hardware or libraries that are commercial and require licenses. An example is a tool that is distributed in the form of an MS Office or MATLAB add-on.

Partial dependence: The demonstrated tool is fully dependent on open-source software, which is openly accessible and is free. An example is a tool that is distributed in the form of a Python or R library.

Stand-alone application: The demonstrated tool is a stand-alone application, fully independent from third-party products. An example is a software product that is installed and runs as an executable, such as EPANET.

f.) Installation guidance and help

(no resources)		(limited resources)		(ample guidance)
1	2	3	4	5
□	□	□	□	□

No resources: The installer/installation process was manual and no means of aid were provided, such as a wizard, troubleshooting suggestions or guidance.

Limited resources: The installer/installation process offered help when needed in the form of simple documentation or very general steps/troubleshooting.

Ample guidance: There was rich supporting material to aid installation, such as troubleshooting guides, tips, clear instructions, a coherent installation manual, a special installer wizard etc.

2.2. Online service

<u>If this tool is a web or cloud service, accessible online</u>, how would you rate the access process in terms of:

a.) Loading time needed

(very long)		(reasonable)		(very fast)
1	2	3	4	5
❑	❑	❑	❑	❑

b.) Dependence on browsers/other online services

(absolute dependence)		(partial dependence)		(absolute independence)
1	2	3	4	5
❑	❑	❑	❑	❑

Absolute dependence: The demonstrated tool is able to run only in a specific browser version and/or is dependent on other web services (besides the STOP-IT platform) to be accessed. Example: a tool that is able to run only in Internet Explorer and requires the user to install Shockwave Player.

Partial dependence: The demonstrated tool is generally able to run in frequently used browsers and web technologies, with some exceptions (e.g. cannot be run in Chrome/Firefox).

Absolute independence: The demonstrated tool is able to run independent of the browser type and its version and is independent of any other web services, besides the STOP-IT platform.

2.3. Open Questions

Did you encounter any problems during the installation of the tool to your system (or your online access to it)?

Yes (major issues)	Yes (minor issues)	No
❑	❑	❑

In case you answered yes to the previous question, please explain the issues encountered:

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

3. Facilitation of user learning

This set of questions gives insights on whether the learning material provided along with the tool during its demonstration was satisfactory or not.

Was learning material (e.g. a tutorial, documentation, examples) provided to you during the demonstration phase, in order to facilitate your understanding and learning process involved with the tool?

Yes	No
□	□

In case you answered yes to the previous question, how would you rate this material in terms of:

a.) Its value in facilitating your understanding of the tool:

(not helpful)		(somewhat helpful)		(very helpful)
1	2	3	4	5
□	□	□	□	□

Not helpful: the material did not make it easier to understand the functionality of the tool and I am still confused on many aspects of the tool.
Somewhat helpful: the provided material simplified the learning process somewhat, but many aspects of the tool use remain challenging.
Very helpful: the provided material simplified the learning process significantly and helped me understand the tool functions considerably.

b.) The content of the provided material:

(too little and/or of bad quality)		(satisfactory)		(ample and/or of good quality)
1	2	3	4	5
□	□	□	□	□

Too little and/or of bad quality: the material that was provided was inadequate to facilitate the learning process or it was not very well explained.
Satisfactory: the provided material was of decent quantity and quality.
Ample and/or of good quality: the provided material was well-prepared and of good quality and helped me learn about the tool considerably.

In case you answered no to the previous question, what type of material do you think would be handy to facilitate user learning? (examples: tutorials, documentation, examples/toy models)

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

4. Data requirements

This set of questions gives insights on aspects of the data input the tool requires and works with.

During the tool demonstration and before the tool execution, the tool probably required an amount of **input** (e.g. data or commands) from you and generated an amount of **output** (e.g. data) to you. How would you rate these data requirements in terms of:

a.) The amount of data required by the tool

(excessive requirements)		(reasonable)		(minimal requirements)
1	2	3	4	5
□	□	□	□	□

Excessive requirements: The tool required data in great detail, not readily available in my line of service, that required a significant amount of time to collect.
Reasonable requirements: The tool required data that were on par with the tool goals and functionality. This data could be provided by the water service within a reasonable amount of time.
Minimal requirements: The tool required a minimal amount of easily accessible data, readily available in my working environment.

b.) The form/formatting of data required by the tool

(custom formats)				(common formats)
1	2	3	4	5
□	□	□	□	□

Custom formats: The tool required data in a specific form and I had to prepare/manually convert the data to that specific format.
Common formats: The tool required standard, common data formats, widely available in my working environment. Examples include: photos in .JPEG format, GIS data in .shp files, text data in plain .csv files.

c.) The form/formatting of data produced by the tool

(custom formats)				(common formats)
1	2	3	4	5
□	□	□	□	□

Custom formats: The tool produced data in a specific form and I had to prepare/manually convert the data to another format in order to use it further (e.g. for another function or tool).
Common formats: The tool produced standard, common data formats, widely available in my working environment that I could easily work further with. Examples include: photos in .JPEG format, GIS data in .shp files, text data in plain .csv files.

d.) The amount of preparation needed to load the data

(very large)				(very little)
1	2	3	4	5
❑	❑	❑	❑	❑

Very large: The tool accepts data only in specific formats and these have to be manually/externally converted by me. Substantial effort was needed to prepare the data. No automatic conversion process exists in the tool and I had to use a third-party converter to prepare the data.

Very little: The tool accepts data in multiple formats and/or features ‘smart’ converters internally, so that I didn’t need to spend a long time preparing my input data.

e.) The availability/accessibility of data requested by the tool

(very low)				(very high)
1	2	3	4	5
❑	❑	❑	❑	❑

Very low: The data requested from the model could not be found easily or were not openly accessible and required substantial effort to gather.

Very high: The data requested could be found easily in my working environment or were openly accessible.

5. Support

This set of questions gives insights on whether the tool offered support material (e.g. wiki, a forum, guides etc.) during its operation or not.

Was support (e.g. in the form of a help, a wiki, a forum etc.) provided to you along with the tool as part of the demonstration phase?

Yes	No
❑	❑

In case you answered yes to the previous question, how would you rate this support material?

(not helpful)		(satisfactory)		(very satisfactory)
1	2	3	4	5
❑	❑	❑	❑	❑

Not helpful: the material did not make it easier to understand the functionality of the tool and I am still confused on many aspects of the tool.
Satisfactory: the provided material was helpful when I ran into specific problems with the tool, but did not explain other problems I had.
Very satisfactory: the provided material was helpful when I ran into any types of problems with the tool and helped me find solutions.

6. Integrity

This set of questions gives insight on the integrity of the tool, i.e. the speed, stability and reliability of its structural functions.

6.1.1 Tool stability, reliability and security

Following the tool installation and preparation of its input data, you had a phase during the demonstration where the tool was executed (i.e. the model or internal tool processes were run and an analysis was made). In case your tool is a database, this runtime refers to the internal processing, e.g. changing data sheets or processing the data before they were being displayed. How would you rate this tool execution phase in terms of:

a.) The speed of tool execution

(very slow)				(very fast)
1	2	3	4	5
□	□	□	□	□

Very slow: The runtime/operational time of the tool was significant and the user had to wait a considerable amount of time before the results could be presented (e.g. the user had to wait a number of minutes).
Very fast: The runtime/operational time of the tool was very small and the model performed the calculations very fast (e.g. in a few seconds).

b.) The stability of tool execution

(very buggy)		(reliable)		(very reliable)
1	2	3	4	5
□	□	□	□	□

Very buggy: During the tool analysis, there were frequent problems and the results couldn't be displayed. These problems were structural, e.g. due to the tool crashing or freezing.
Reliable: Most of the analyses were run without errors. Some issues occurred at some more complex cases or when I did something that the tool did not expect.
Very reliable: The analysis of the tool was always able to run and the results were displayed with no problems. The tool performance was consistent, without any bugs or crashes.

c.) The security of tool execution

During the tool execution, the tool likely used a number of security protocols to ensure that the handled, processed and generated data cannot be seen by third users. Were you informed of or did you have any knowledge on the security protocols used for that particular tool?

Yes	No
□	□

In case you answered yes to the previous question, please rate your experience in the tool use in terms of how secure it was:

(not secure)		(secure)		(very secure)
1	2	3	4	5
❑	❑	❑	❑	❑

Very secure: The latest security protocols for that particular technology were used (e.g. encryption, user-restricted access etc.) and I was well informed of them, as part of the demonstration process.
Secure: A reasonable level of security was used and I had basic knowledge about it during the demonstration process.
Not secure: The tool, based on my experience, did not employ security protocols such as encryptions and I am concerned about its use as part of my regular water service.

Did you run this tool in your local disk drive or as part of a **Virtual Machine** environment?

I ran this tool locally, without any dependence on a VM environment.	I ran this tool through a VM environment.
❑	❑

What is a Virtual Machine? A Virtual Machine is an emulation environment for a computer system, where the tool runs on predefined hardware/software settings, as opposed to a local run where the tool has to be installed and run in your local system (e.g. a classic Windows installation). If you are unsure about this, ask your demonstration tool guides/experts.

If you ran this tool in a VM environment, how would you rate this experience in terms of:

d.) The stability of the Virtual Machine environment where the tool was executed?

(unstable)		(reliable)		(very stable)
1	2	3	4	5
❑	❑	❑	❑	❑

Unstable: During the tool analysis, many times the Virtual Environment itself would freeze or disconnect and I was not able to access the tool itself.
Very stable: The Virtual Environment itself ran smoothly and I was able to access the tool, as well as the tools connecting to it, on every occasion.

6.1.2 Open Questions

Did you encounter any problems during the **operation** of the tool to your system (or your online access to it)?

Yes (major issues, instabilities etc.)	Yes (minor issues, e.g. some bugs)	No
❑	❑	❑

In case you answered yes to the previous question, please explain the issues encountered:
………………………………………………………………………………………………………

………………………………………………………………………………………………………

7. Usability

This set of questions gives insight on how easy and pleasurable it is to use the tool, thus exploring its structural simplicity, aesthetic and functional aspects of its interface and intuitiveness.

Throughout the tool demonstration, you probably interacted with the tool or technology through a command line or user interface (UI) that included all buttons, commands, graphics etc. that enabled interaction with the tool. There could be also the case that this tool is only a protocol or (hardware) technology, so it was installed in your system and does not have a specific interface you can interact with.

Did the tool have an interface or was it just a protocol/technology?

The tool featured a user interface (e.g. graphics or command line)	This tool was a protocol or technology so there is no interface I can interact with.
❑	❑
(proceed to Section 7.1)	(proceed to Section 7.2)

7.1 Tool with a user interface

How would you rate this user interface in terms of:

a.) the time it took you to get acquainted with the interface:

(excessive)		(reasonable)		(minimal)
1	2	3	4	5
❑	❑	❑	❑	❑

excessive: It took me a long time to get used to the graphics and functions of the tool interface and I am still unsure about what many of the options do.
reasonable: The amount of time needed to get acquainted with the buttons and graphics was reasonable and in par with the tool goals. I now know what most options do.
minimal: I learned how to interact with the tool very quickly and got used to it very quickly as well.

b.) user interface functionality:

(cumbersome)		(functional)		(very functional)
1	2	3	4	5
❑	❑	❑	❑	❑

cumbersome: The user interface is complicated and a considerable amount of time is required to explore the options and functions of the tool.
functional: The user interface offers a decent level of functionality, even though some aspects could be improved (e.g. some options could be simplified).
very functional: The user interface is simple and functional, on par with the tool goals.

c.) the design of the user interface:

(basic)		(good)		(beautiful)
1	2	3	4	5
❑	❑	❑	❑	❑

basic: The user interface works with a very crude design, i.e. is a simple command-line, or is a primitive graphical user interface.
functional: The user interface is designed to serve the basic functions of the tool and facilitate the user experience.
very functional: The user interface is beautifully designed and offers a pleasurable user experience.

d.) the overall intuitiveness of the user-tool interaction:

(not intuitive)		(reasonable)		(very intuitive)
1	2	3	4	5
□	□	□	□	□

not intuitive: During the user-tool interaction, actions from my side frequently do not make sense or are not easy to deduce and I must spend a considerable amount of time to learn them. The sequence of actions needed from me is confusing.
reasonable: During the user-tool interaction, I occasionally have to look out where to find specific options and/or actions. However, the general experience is not cumbersome and I can interact with the tool without overall confusion.
very intuitive: The user-tool interaction works in a very intuitive way. I know or can easily guess where I can find the tool options without a lot of learning.

e.) the functionality of the tool in general:

(unnecessarily complex)		(functional)		(very functional)
1	2	3	4	5
□	□	□	□	□

unnecessarily complex: The tool in general looks very complex and offers a lot of options that I'm not going to or wouldn't like to use.
functional: The tool offers interesting options, even though some aspects could be improved.
very functional: The tool feels 'just right' and it has complexity and functionality in par with the tool goals. I find it very functional and would like to use it further.

7.2 Protocol or Technology

How would you rate your experience with the tool in terms of:

a.) the general way the tool runs in your systems so far:

(problematic)		(functional)		(very functional)
1	2	3	4	5
□	□	□	□	□

problematic: The protocol or technology is not able to run multiple times and I had to uninstall it/operate the system without it.
functional: The protocol or technology is operational during most times, with slight issues that do not bother me or cause downtime to other services.
very functional: The protocol or technology has a seamless operation to my working environment and is always working well.

8. Usefulness

This set of questions gives insight on how useful the particular tool is in the context of the FR service, both as a stand-alone product but also as part of the STOP-IT platform.

The end product of STOP-IT is to provide a framework for the identification, detection, assessment and mitigation of cyber-physical risks in your water system. Based on your experience, the role of **this particular tool** was in which part of the afore-mentioned general goal?

(Note: multiple answers are accepted)

Risk identification (e.g. RIDB)	Risk detection (e.g. NTSA)	Risk assessment and analysis (e.g. AVAT)	Risk mitigation (e.g. RRMD)
□	□	□	□

a.) Based on your experience from the demonstration, how well do you think the tool performs the <u>specific function</u> that it was designed/supposed to do?

(very limited success)		(partial success)		(success)
1	2	3	4	5
□	□	□	□	□

Very limited success: The tool's function deviates from what was promised and only a small part of the tools' objectives are currently fulfilled.
Partial success: The tool's function is exactly what was promised, albeit with a number of limited mishaps during the tool operation. Core functionalities are as promised, even though the tool could be improved to serve its functional requirements.
Success: The tool works exactly as it was envisioned and all of its requirements are covered.

b.) How do you view the specific tool as <u>part of the whole STOP-IT platform and its goals</u> (i.e. the provision of cyber, physical and cyber-physical risk assessment and treatment services at strategic, tactical and operational levels)?
Consider also the role of the tool in the identification-detection-analysis-mitigation chain that was analysed at the beginning of this section.

(a niche/optional part)		(a useful part)		(an integral part)
1	2	3	4	5
□	□	□	□	□

A niche/optional part: The tool's function has a niche part in the STOP-IT platform and/or does not directly aid/actively contribute to the goals of the platform as a whole.
A useful part: The tool's function offers useful functionality that helps/contributes to the general goals of the STOP-IT platform.
An integral part: The tool's function is important and can be considered an integral part of the STOP-IT platform services.

c.) Would you consider the specific tool as a useful addition to the needs and challenges of your water service?

(not that useful)		(useful)		(very useful)
1	2	3	4	5
□	□	□	□	□

Not that useful: The tool is not useful, in its present form, to the needs and challenges of the water services my company provides.

Useful: The tool is a useful addition to the needs and challenges of the water services my company provides.

Very useful: The tool is a highly desirable addition to the needs and challenges of the water services my company provides.

d.) Is the tool efficient at raising your awareness on cyber-physical risks on your system?

(inefficient)		(moderately efficient)		(very efficient)
1	2	3	4	5
□	□	□	□	□

Inefficient: The tool did not give any insights on the cyber-physical risk my company faces while operating their drinking water system.

Moderately efficient: The tool provided insights and made me aware of the cyber-physical risks involved in my line of business.

Very efficient: The tool provided me with good insight and made me well aware of the cyber-physical risks that are involved in my line of business.

ANNEX C: Questionnaire template – platform level

The following questions, directed at the end user (FR), aim at validating the performance of the whole platform, following the demonstration process. Validation is based on a set of **traits**, given as numbered sections, along with their partial characteristics, given as individual questions.

To fill this questionnaire, please provide:

- your ranking, in case of grading questions. If needed, an explanation of the different grades is provided below each question.
- your feedback, in case of open questions or conditional (Yes/No) answers.

Most questions are based on a grading/ranking evaluation that ranges from 1 (poor performance) to 5 (great performance). Open questions supplement some sections, allowing you to provide feedback back to STOP-IT tool developers.

1. Introduction

Please fill in the required information.

Demonstration Event Date: __ / __ / ____

Front Runner: …………………………………..

First Name: ……………………………………………………….

Last Name: ……………………………………….

Job Role in the FR: ……………………………………….

Did you also experience any of these tools as part of the platform demonstration?

□	REN	Reasoning Engine	□	RGIP	RISA GEN Integration Platform
□	EVI	Enhanced Visualization Interface	□	IWM	Interoperability Water Middleware

Note: These tools constitute parts of the platform that you might have experienced as part of the demonstration event.

2. Ease of installation

This set of questions gives insights on aspects of the data input the platform requires and works with.

Was the STOP-IT platform provided (as a whole service) locally or is it a web/cloud service?

STOP-IT was presented through local installations to my systems.	STOP-IT was presented as an online (web or cloud) service.	STOP-IT included both local and online demonstrations.
❑	❑	❑
(proceed to Sections 2.1 & 2.3)	(proceed to Sections 2.2 & 2.3)	(please answer all Sections)

2.1 Local installation

For the parts of the platform that were installed locally, how would you rate the installation process in terms of:

a.) Installation time needed

(very long)		(reasonable)		(very fast)
1	2	3	4	5
❑	❑	❑	❑	❑

b.) Installation process simplicity/complexity:

(very complex)		(reasonable)		(simple and concise)
1	2	3	4	5
❑	❑	❑	❑	❑

Very complex: The installation process was lengthy and required special knowledge, e.g. installation of other tools first or technical hardware skills. As such, it had to be done by specialized personnel.

Simple and concise: The installation process was very simple and could be readily performed by me, without extra steps or pre-installation needs.

c.) Integration with my current system:

(very limited integration)		(limited integration)		(seamless integration)
1	2	3	4	5
❑	❑	❑	❑	❑

Very limited integration: Many of the tools were not able to run with my current system specs or required substantial effort to be installed in my current system.

Limited integration: Most of the tools were able to run with my current system, albeit with some effort and/or after installing some third-party software

Seamless integration: All of the tools installed easily and integrated fully with my current system, without the need from my side to change parts of my system.

d.) Installation guidance and help

(no resources)		(limited resources)		(ample guidance)
1	2	3	4	5
❑	❑	❑	❑	❑

No resources: The installer/installation process was manual and no means of aid were provided, such as a wizard, troubleshooting suggestions or guidance.

Limited resources: The installer/installation process offered help when needed in the form of simple documentation or very general steps/troubleshooting.

Ample guidance: There was rich supporting material to aid installation of the platform, such as troubleshooting guides, tips, clear instructions, a coherent installation manual, a special installer wizard etc.

2.2. Online service

__For the parts of the platform that were a web or cloud service, accessible online__, how would you rate the access process in terms of:

a.) Loading time needed

(very long)		(reasonable)		(very fast)
1	2	3	4	5
❑	❑	❑	❑	❑

b.) Dependence on browsers/other online services

(absolute dependence)		(partial dependence)		(absolute independence)
1	2	3	4	5
❑	❑	❑	❑	❑

Absolute dependence: The demonstrated tools were able to run only in a specific browser version and/or is dependent on other web services (besides the STOP-IT platform) to be accessed. Example: a tool that is able to run only in Internet Explorer and requires the user to install Shockwave Player.

Partial dependence: The demonstrated tools were generally able to run in frequently used browsers and web technologies, with some exceptions (e.g. cannot be run in Chrome/Firefox).

Absolute independence: The platform functions were able to run independent of the browser type and its version and is independent of any other web services, besides the STOP-IT platform.

2.3 Open Questions

Did you encounter any problems during the installation of the platform as a whole to your system (or your online access to it)?

Yes (major issues)   Yes (minor issues)   No

❑	❑	❑

In case you answered yes to the previous question, please explain the issues encountered:
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………

3. Facilitation of user learning

This set of questions gives insights on whether the learning material provided along with the platform during its demonstration was satisfactory or not.

Was learning material (e.g. a tutorial, documentation, examples) provided to you during the platform demonstration phase, in order to facilitate your understanding and learning process involved with the platform?

Yes	No
□	□

In case you answered yes to the previous question, how would you rate this material in terms of:

c.) Its value in facilitating your understanding of the platform functions:

(not helpful)		(somewhat helpful)		(very helpful)
1	2	3	4	5
□	□	□	□	□

Not helpful: the material did not make it easier to understand what many tools did and I am still confused on many aspects of the STOP-IT functions.
Somewhat helpful: the provided material simplified the learning process somewhat, but many aspects of the platform remain challenging.
Very helpful: the provided material simplified the learning process significantly and helped me understand the platform functions considerably.

d.) The content of the provided material:

(too little and/or of bad quality)		(satisfactory)		(ample and/or of good quality)
1	2	3	4	5
□	□	□	□	□

Too little and/or of bad quality: the material that was provided was inadequate to facilitate the learning process or it was not very well explained.
Satisfactory: the provided material was of decent quantity and quality.
Ample and/or of good quality: the provided material was well-prepared and of good quality and helped me learn about the platform considerably.

In case you answered no to the previous question, what type of material do you think would be handy to facilitate user learning? (examples: tutorials, documentation, examples/toy models)

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

4. Data requirements

This set of questions gives insights on aspects of the data input the platform requires and works with.

During the platform demonstration, the tools probably required an amount of **input** (e.g. data or commands) from you and generated an amount of **output** (e.g. data) to you. How would you rate these data requirements of the tools and the platform as a whole in terms of:

f.) The amount of data required by the platform

(excessive requirements)		(reasonable)		(minimal requirements)
1	2	3	4	5
❑	❑	❑	❑	❑

Excessive requirements: The tools required data in great detail, not readily available in my line of service, that required a significant amount of time to collect.
Reasonable requirements: The tools required data that were on par with the tool goals and functionality. This data could be provided by the water service within a reasonable amount of time.
Minimal requirements: The tools required a minimal amount of easily accessible data, readily available in my working environment.

g.) The form/formatting of data used by the platform

(custom formats)				(common formats)
1	2	3	4	5
❑	❑	❑	❑	❑

Custom formats: The platform required data in multiple custom and specific forms and I had to prepare/manually convert the data to that specific format many times.
Common formats: The platform generally required standard, common data formats, widely available in my working environment. Examples include: photos in .JPEG format, GIS data in .shp files, text data in plain .csv files, using the JSON framework. The tools used these data types as input/output throughout the demonstration and did not have any problems communicating these common data formats.

h.) The amount of preparation needed to load and get the data

(very large)				(very little)
1	2	3	4	5
❑	❑	❑	❑	❑

Very large: The tools accepted data only in specific formats and these have to be manually/externally converted by me. Substantial effort was needed to prepare the data. No automatic conversion process exists in the platform and I had to use a third-party converter to prepare the data.
Very little: The platform accepted data in multiple formats and/or featured 'smart' converters internally, so that I didn't need to spend a long time preparing my input data. Tools many

times automatically recognized the data types that I inserted and used them without problems.

i.) During the demonstration, did the tools have to work together, in an Input/Output fashion, with other tools in the STOP-IT platform?

Yes, multiple tools were demonstrated in a serial fashion (tool chain). □
No, this demonstration a stand-alone demonstration of the tool. □

In case you answered yes to the previous question, please rate how well tools collaborated with other tools, in terms of input/output requirements:

(very limited integration)		(limited integration)		(seamless integration)
1	2	3	4	5
□	□	□	□	□

Very limited integration: The input to/output from each tool had to be manually and externally edited/converted to the format required by the tool. There was very limited interoperability between the two tools. It required a significant amount of time from me to do this conversion.

Seamless integration: All data conversion requirements to/from the tools were handled seamlessly by the STOP-IT platform or the tools themselves. There was no need to manually edit the data, as an extra step between the use of the different tools.

5. Support

This set of questions gives insights on whether the tool offered support material (e.g. wiki, a forum, guides etc.) during its operation or not.

Was support (e.g. in the form of a help, a wiki, a forum etc.) provided to you along with the STOP-IT platform as part of the demonstration phase?

Yes	No
❑	❑

In case you answered yes to the previous question, how would you rate this support material?

(not helpful)		(satisfactory)		(very satisfactory)
1	2	3	4	5
❑	❑	❑	❑	❑

Not helpful: the material did not make it easier to understand the functionality of the platform and I am still confused on many of its aspects.
Satisfactory: the provided material was helpful when I ran into specific problems with the platform, but did not explain other problems I had.
Very satisfactory: the provided material was helpful when I ran into any type of problems with the platform and helped me find solutions, as well as understand different functions of the STOP-IT platform.

6. Integrity

This set of questions gives insight on the integrity of the platform, i.e. the speed, stability and reliability of its structural functions.

6.1 Platform stability, reliability and security

Following the platform installation and preparation of its input data, you had a phase during the demonstration where one or multiple tools were executed (i.e. the models or internal tool processes were run and analyses were made). How would the experience of this execution phase in terms of:

e.) The speed of executing functions within the STOP-IT platform

(very slow)				(very fast)
1	2	3	4	5
❑	❑	❑	❑	❑

Very slow: The runtime/operational time of the platform as a whole was significant and the user had to wait a considerable amount of time before the results could be presented or before accessing different tools.
Very fast: The runtime/operational time of the platform was very small and the tools were executed quite fast, without waiting times in between.

f.) The stability of the platform

(very buggy)		(reliable)		(very reliable)
1	2	3	4	5
❑	❑	❑	❑	❑

Very buggy: During the analysis phase, there were frequent problems and the results couldn’t be displayed. These problems were structural, e.g. due to the tools crashing or freezing.
Reliable: Most of the analyses were run without errors. Some issues occurred at some more complex cases or when I did something that an individual tool (or the platform) did not expect.
Very reliable: The analysis was always able to run and the results were displayed with no problems at the platform level. No crashes were observed in between tool runs.

g.) The security of the platform

The platform employs a number of security protocols to ensure that the handled, processed and generated data cannot be seen by third users. Were you informed of or did you have any knowledge on the security protocols used as part of the STOP-IT platform?

Yes	No
❑	❑

In case you answered yes to the previous question, please rate your experience in the tool use in terms of how secure it was:

(not secure)	(secure)	(very secure)

1	2	3	4	5
❑	❑	❑	❑	❑

Very secure: The latest security protocols were used whenever needed (e.g. encryption, user-restricted access etc.) and I was well informed of them, as part of the demonstration process.

Secure: A reasonable level of security was used by the platform and I had basic knowledge about it during the demonstration process.

Not secure: The platform, based on my experience, did not employ security protocols such as encryptions to exchange data between tools or to display data to me and I am concerned about its use as part of my regular water service.

6.2 The higher-level (parallelisation, scheduling and scaling) applications of the platform

Did you experience any of the parallel processing, dataflow scheduling, scaling and distributed computing capabilities of the STOP-IT platform as part of the demonstration?

<u>**Note:**</u> These capabilities are demonstrated through the <u>**RISA GEN Integration Platform (RGIP) module**</u>.

Yes	No
❑	❑

In case you answered yes to the previous question, please rate your experience with these capabilities:

(very limited)		(satisfactory)		(very satisfactory)
1	2	3	4	5
❑	❑	❑	❑	❑

Very limited: The STOP-IT platform could not function properly in a parallel, scheduled or scaled manner and these capabilities were not sufficiently demonstrated.

Satisfactory: Based on my experience with these capabilities, STOP-IT platform is able to run in a parallel, scheduled or scaled mode.

Very satisfactory: The STOP-IT platform was fully able to perform parallel runs, to implement scheduled data flows and to scale applications to my need. This capability was properly demonstrated as part of the RGIP functionality of the platform and I am now aware of it.

6.3. Open Questions

Did you encounter any problems during the <u>**operation**</u> of the platform to your system (or your online access to it)?

Yes (major issues, instabilities etc.)	Yes (minor issues, e.g. some bugs)	No
❑	❑	❑

In case you answered yes to the previous question, please explain the issues encountered:

……………………………………………………………………………………………………..

……………………………..……………………………………………………………………………

7. Usability

This set of questions gives insight on how easy and pleasurable it is to use the platform, thus exploring its structural simplicity, aesthetic and functional aspects of its interface and intuitiveness.

Throughout the platform demonstration, you probably interacted with the platform in a higher-level user interface (UI) that guided you to find the right tools and visualized their results.

How would you rate this platform user interface in terms of:

f.) the time it took you to get acquainted with the interface:

(excessive)		(reasonable)		(minimal)
1	2	3	4	5
❏	❏	❏	❏	❏

excessive: It took me a long time to get used to the graphics and functions of the platform interface and I am still unsure about what many of the options do.
reasonable: The amount of time needed to get acquainted with the buttons and graphics was reasonable and in par with the platform goals. I now know what most options do.
minimal: I learned how to interact with the platform very quickly and got used to it very quickly as well.

g.) the user interface functionality:

(cumbersome)		(functional)		(very functional)
1	2	3	4	5
❏	❏	❏	❏	❏

cumbersome: The user interface is complicated and a considerable amount of time is required to explore the options and functions of the platform.
functional: The user interface offers a decent level of functionality, even though some aspects could be improved (e.g. some options could be simplified).
very functional: The user interface is simple and functional, on par with STOP-IT goals.

h.) the design of the user interface:

(basic)		(good)		(beautiful)
1	2	3	4	5
❏	❏	❏	❏	❏

basic: The user interface works with a very crude design, i.e. is a simple command-line, or is a primitive graphical user interface.
functional: The user interface is designed to serve the basic functions of the platform and facilitate the user experience.
very functional: The user interface is beautifully designed and offers a pleasurable user experience.

i.) the functionality of the platform user interface in general:

(unnecessarily complex)		(functional)		(very functional)
1	2	3	4	5
❏	❏	❏	❏	❏

unnecessarily complex: The platform UI design in general looks very complex and offers a lot of options that I’m not going to or wouldn’t like to use.
functional: The tool offers interesting options, even though some aspects could be improved.
very functional: The platform has complexity and functionality in par with the tool goals. I find it very functional and would like to use it further.

8. Usefulness

This set of questions gives insight on how useful the particular tool is in the context of the FR service, both as a stand-alone product but also as part of the STOP-IT platform.

8.1 STOP-IT usefulness value

The end product of STOP-IT is to provide a framework for the identification, detection, assessment and mitigation of cyber-physical risks in your water system. Based on your experience, how would you rate the STOP-IT platform as a whole in terms of:

e.) Risk identification (e.g. through the use of tools such as RIDB)

(very limited success)		(partial success)		(success)
1	2	3	4	5
❑	❑	❑	❑	❑

Very limited success: The platform function deviates from what was promised and only a small part of the platform objectives are currently fulfilled.

Partial success: The platform function is exactly what was promised, albeit with a number of limited mishaps during the tool operation. Core functionalities are as promised, even though some aspects could be improved.

Success: The platform works exactly as it was envisioned and all of its requirements are covered.

f.) Risk detection (e.g. through the use of operational tools such as NTSA)

(very limited success)		(partial success)		(success)
1	2	3	4	5
❑	❑	❑	❑	❑

Very limited success: The platform function deviates from what was promised and only a small part of the platform objectives are currently fulfilled.

Partial success: The platform function is exactly what was promised, albeit with a number of limited mishaps during the tool operation. Core functionalities are as promised, even though some aspects could be improved.

Success: The platform works exactly as it was envisioned and all of its requirements are covered.

g.) Risk assessment and analysis (e.g. through the use of tools such as RAET and AVAT)

(very limited success)		(partial success)		(success)
1	2	3	4	5
❑	❑	❑	❑	❑

Very limited success: The platform function deviates from what was promised and only a small part of the platform objectives are currently fulfilled.

Partial success: The platform function is exactly what was promised, albeit with a number of limited mishaps during the tool operation. Core functionalities are as promised, even though some aspects could be improved.
Success: The platform works exactly as it was envisioned and all of its requirements are covered.

h.) Risk mitigation (e.g. through the use of tools such as RRMD)

(very limited success)		(partial success)		(success)
1	2	3	4	5
❑	❑	❑	❑	❑

Very limited success: The platform function deviates from what was promised and only a small part of the platform objectives are currently fulfilled.
Partial success: The platform function is exactly what was promised, albeit with a number of limited mishaps during the tool operation. Core functionalities are as promised, even though some aspects could be improved.
Success: The platform works exactly as it was envisioned and all of its requirements are covered.

i.) Would you consider the STOP-IT tool as a useful addition to the needs and challenges of your water service?

(not useful)		(useful)		(very useful)
1	2	3	4	5
❑	❑	❑	❑	❑

Not that useful: The platform is not useful, in its present form, to the needs and challenges of the water services my company provides.
Useful: The tool is a useful addition to the needs and challenges of the water services my company provides and provides me with insight on cyber-physical risks.
Very useful: The platform is a highly desirable addition to the needs and challenges of the water services my company provides.

j.) Is the platform as a whole efficient at raising your awareness on cyber-physical risks on your system?

(inefficient)		(moderately efficient)		(very efficient)
1	2	3	4	5
❑	❑	❑	❑	❑

Inefficient: The platform did not give any insights on the cyber-physical risk my company faces while operating their drinking water system.
Moderately efficient: The platform provided insights and made me aware of the cyber-physical risks involved in my line of business.
Very efficient: The platform provided me with good insight and made me well aware of the cyber-physical risks that are involved in my line of business.

8.2 Reflection on the use of the STOP-IT platform for system operations

These questions request that you reflect on your experience with the STOP-IT platform as a whole, thus projecting it in your operational environment for the years to come. Project that you now can use elements from STOP-IT platform to assess, evaluate and mitigate cyber-physical risk in your system on all (operational, tactical and strategic) scales.

According to that experience and your projection using the STOP-IT platform:

how efficient do you consider STOP-IT to be in achieving a higher detection of physical/cyber attacks and incidents than the previous technology you were using before?

(No difference)		(50% better)		(100% better or more)
1	2	3	4	5
□	□	□	□	□

how efficient do you think STOP-IT is in achieving lower false positive rates, compared to the previous technology you were using before?

(No difference)		(50% better)		(100% better or more)
1	2	3	4	5
□	□	□	□	□

how many times faster do you estimate attacks and incidents are able to be spotted compared to the previous state of your operations (i.e. without the STOP-IT platform)?

(No difference)		(reasonably faster, e.g. 10 times)		(much faster, e.g. 20 or more times)
1	2	3	4	5
□	□	□	□	□

how much do you think the **average response time** of the STOP-IT platform is in **operational** scenarios and use cases? This only considers the operational tools offered by the platform and not the strategic or tactical tools.

(very large, e.g. 1 minute or more)		(reasonable, 5-10 seconds)		(very fast, e.g. 2 seconds or less)
1	2	3	4	5
□	□	□	□	□

how much do you estimate your **preparedness level** to be towards cyber-physical attacks, compared to the previous state?

(No improvement)		(Modest improvement – 50% better)		(Substantial improvement – 100% better)
1	2	3	4	5
□	□	□	□	□

how efficient do you estimate the STOP-IT platform to be in reducing your clients' exposure to cyber-physical attacks and their impacts?

(No improvement)		(Modestly efficient)		(Substantially efficient)
1	2	3	4	5
❑	❑	❑	❑	❑

how efficient do you estimate the STOP-IT platform to be in reducing the exposure of your personnel to cyber-physical attacks and their impacts?

(No improvement)		(Modestly efficient)		(Substantially efficient)
1	2	3	4	5
❑	❑	❑	❑	❑

8.3 Open Questions

Would you like to see some improvements in the services provided by the STOP-IT platform?

Yes (major issues)	Yes (minor issues)	No
❑	❑	❑

In case you answered yes to the previous question, could you please suggest some improvement areas?

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

References

Al-Khaldi, M.A. & Olusegun Wallace, R.S. (1999). The influence of attitudes on personal computer utilization among knowledge workers: the case of Saudi Arabia. *Information & Management*. [Online]. 36 (4). p.pp. 185–204. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0378720699000178.

Bokhari, R.H. (2005). The relationship between system usage and user satisfaction: a meta-analysis. *Journal of Enterprise Information Management*. 18 (2). p.pp. 211–234.

Dromey, R.G. (1995). A model for software product quality. *IEEE Transactions on software engineering*. 21 (2). p.pp. 146–162.

Gelderman, M. (1998). The relation between user satisfaction, usage of information systems and performance. *Information & management*. 34 (1). p.pp. 11–18.

ISO/TC 176 (1994). *ISO 8402:1994 - Quality management and quality assurance -- Vocabulary*. [Online]. Available from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=20115.

Jamwal, D. (2010). Analysis of software quality models for organizations. *International Journal of Latest Trends in Computing*. 1 (2). p.pp. 19–23.

Kim, S., Meng, J., Kalinowski, J. & Shin, D. (2014). The Development and Validation of an End-User Satisfaction Measure in a Student Laptop Environment. *American Journal of Business Education*. 7 (2). p.pp. 157–170.

Laplante, P.A. (2007). *What every engineer should know about software engineering*. CRC Press.

Lee, J. (2003). An end-user perspective on file-sharing systems. *Communications of the ACM*. 46 (2). p.pp. 49–53.

Norman, D. & Nielsen, J. (2018). *The Definition of User Experience (UX)*. [Online]. 2018. Available from: http://www.nngroup.com/articles/definition-user-experience/.

Van de Poel, I. & Royakkers, L. (2011). *Ethics, technology, and engineering: An introduction*. John Wiley & Sons.

Voas, J. & Agresti, W.W. (2004). Software quality from a behavioral perspective. *IT Professional*. [Online]. 6 (4). p.pp. 46–50. Available from: http://ieeexplore.ieee.org/document/1324574/.

Yahaya, J.H., Deraman, A. & Hamdan, A.R. (2008). Software quality from behavioural and human perspectives. *IJCSNS International Journal of Computer Science and Network Security*. [Online]. 8 (8). p.pp. 53–63. Available from: http://ieeexplore.ieee.org/document/1324574.

Zviran, M. & Erlich, Z. (2003). Measuring IS user satisfaction: Review and implications. *Communications of the Association for Information Systems*. 12 (1). p.p. 5.

Zviran, M., Pliskin, N. & Levin, R. (2005). Measuring user satisfaction and perceived usefulness in the ERP context. *Journal of computer information systems*. 45 (3). p.pp. 43–52.

STOP-IT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 740610.

The publication reflects only the authors' views and the European Union is not liable for any use that may be made of the information contained therein.