SPECIAL

Scalable Policy-awarE Linked Data arChitecture for prIvacy, trAnsparency and compLiance

Deliverable D7.3

Quality, risk and data management plan

Document version: V1.0

SPECIAL DELIVERABLE

Name, title and organisation of the scientific representative of the project's coordinator:

Mr Philippe Rohou t: +33 4 97 15 53 06 f: +33 4 92 38 78 22 e: philippe.rohou@ercim.eu

GEIE ERCIM, 2004, route des Lucioles, Sophia Antipolis, 06410 Biot, France

Project website address: http://www.specialprivacy.eu/

Project	
Grant Agreement number	731601
Project acronym:	SPECIAL
Project title:	Scalable Policy-awarE Linked Data arChitecture for prIvacy, trAnsparency and compLiance
Funding Scheme:	Research & Innovation Action (RIA)
Date of latest version of DoW against which the assessment will be made:	17/10/2016
Document	
Period covered:	M1-M36
Deliverable number:	D7.3
Deliverable title	Quality, risk and data management plan
Contractual Date of Delivery:	30-06-2017
Actual Date of Delivery:	30-06-2017
Editor (s):	Philippe Rohou (ERCIM)
Author (s):	Sabrina Kirrane (WU), Philippe Rohou (ERCIM)
Reviewer (s):	Martin Kurze (DTAG), Rudy Jacob (PROX)
Participant(s):	Sabrina Kirrane (WU), Axel Polleres (WU) Philippe Rohou (ERCIM), Rigo Wenning (ERCIM)
Work package no.:	7
Work package title:	Project Management
Work package leader:	ERCIM
Distribution:	PU
Version/Revision:	V1.0
Draft/Final:	Final
Total number of pages (including cover):	35

Disclaimer

This document contains description of the SPECIAL project work and findings.

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

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

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 28 Member States of the Union. It is based on the European Communities and the Member States cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors (http://europa.eu/).

SPECIAL has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731601.

Table of Contents

Table of Figures

Table of Tables

1 Introduction

Processes related to quality, risk and data management within SPECIAL are summarised as follows in the Description of Work:

T7.3 Quality, risk and data management (Lead: ERCIM; Participants: WU; Duration: M1-M36) Quality Assurance and risk management is intended to ensure the production of concrete and high-quality results in line with the project work plan. To achieve this goal, a Quality Assurance Team is appointed to:

- Define and widely distribute the Quality Plan, to be a reference for all project participants; Encourage and verify that standards, procedures and metrics are defined, applied and evaluated;
- Adopt a procedure for identifying, estimating, treating and monitoring risks;
- Perform monthly Quality and Risk Reviews communicated to the General Assembly for appropriate action;
- Define a statement on the promotion of gender equality within SPECIAL practices and procedures; Produce a Data Management Plan (DMP) in accordance to [DM_H2020], as described in Section 3.2.5 below.

This deliverable reports on the processes put in place by the consortium to ensure that all contractual reports are delivered to the expected level of quality and timeliness.

It also re-visits and updates the risk table and mitigation measures proposed in the original Description of Work at proposal time, taking into account the experience gained during the first six months of operation.

Finally, the Data Management Plan (DMP), contractually due at M6, makes the third and final chapter of deliverable D7.3.

2 Quality

2.1 Overview

This chapter details the internal Quality Assurance processes put in place at the start of the project to ensure the production of quality deliverables throughout the lifetime of project SPECIAL.

In the Description of Work, the quality plan is one of the output of Task 7.3 'Quality, Risk and Data Management'. T7.3 is under the responsibility of the project coordinator (ERCIM), with the active participation of the scientific coordinator (WU).

Task T7.3 is active for the full duration of the project. Quality processes described here will apply to all SPECIAL deliverables listed in annex I of this document, as well as to possible additions to this list that may result from annual project reviews.

2.2 Quality Assurance process

2.2.1 Definition

As per the above definition of Task 7.3 'Quality, risk and data management', Quality Assurance is intended to guarantee the production of concrete and high-quality results in line with the project work plan. To achieve this goal, a Quality Assurance team is appointed to define and widely distribute the Quality Plan, to be a reference for all project participants, to encourage and verify that standards, procedures and metrics are defined, applied and evaluated.

With this in mind, the consortium has defined a Quality Assurance process based on a time line and a set of actions to be repeated for each project deliverable. In graphical terms, this is the sequence of events that will ensure proper internal review of the SPECIAL deliverables:

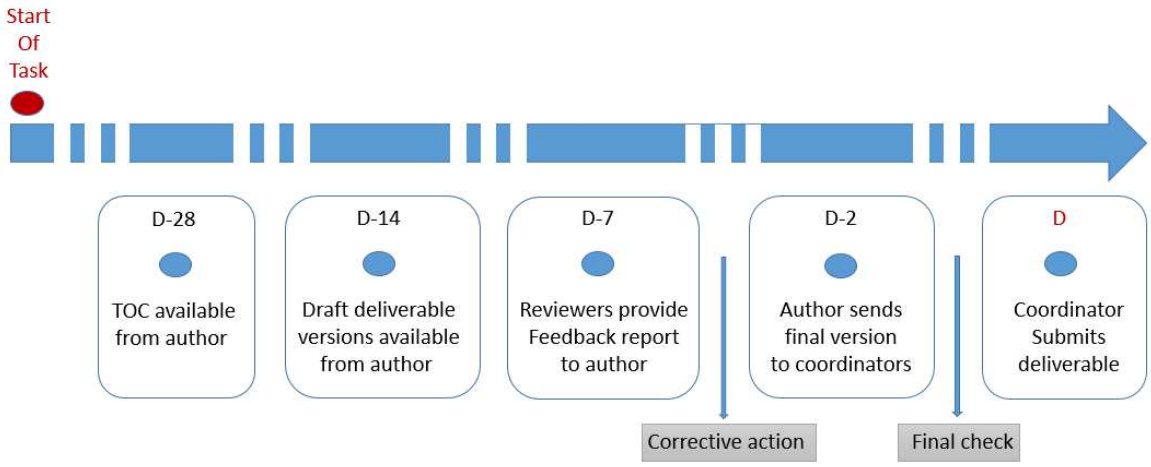


Figure 2.1 - The SPECIAL Quality Assurance process

Two internal reviewers for each project deliverable have been appointed by the Project Steering Committee (PSC) in the early stages of the project. Figure 3 below shows the detailed list of project deliverables together with their assigned internal reviewers. In a nutshell, the review process described in Figure 1 requires that:

- The author of the deliverable supplies a table of content for review at least four weeks before submission, and a first draft of the deliverable at the latest two weeks before submission;
- Internal reviewers write their review report using the internal review checklist described in annex 2 and send it to the author within a week;
- The author implements the changes and sends the final version back to the reviewers, to the WP leader and to the scientific coordinator, no later than two days before the deadline;
- Once last comments are resolved among all players and taken on board by the author, the deliverable is submitted to the EC by the project coordinator via the continuous reporting tool on the participant portal.

The end result of the implementation of this process is expected to be a set of quality deliverables delivered on time.

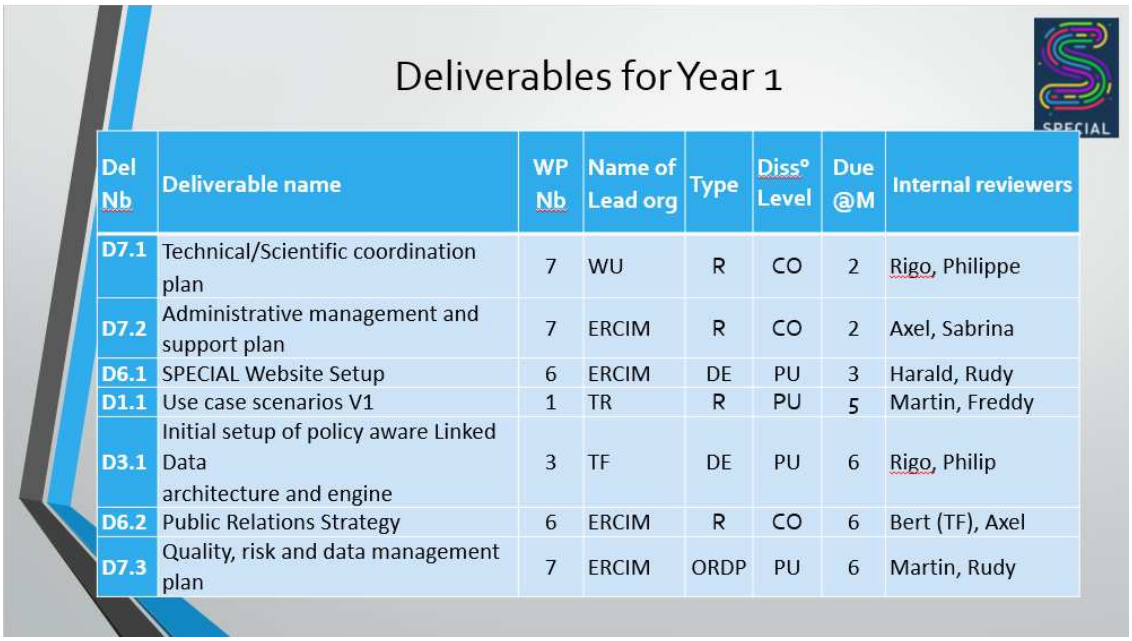
2.2.2 Implementation

2.2.2.1 Internal reviewers

At the kick-off meeting of the project, which took place at the European Commission in Luxembourg on 16-17 January 2017, the consortium discussed in further details and confirmed the Quality Assurance procedures described in the Grant Agreement.

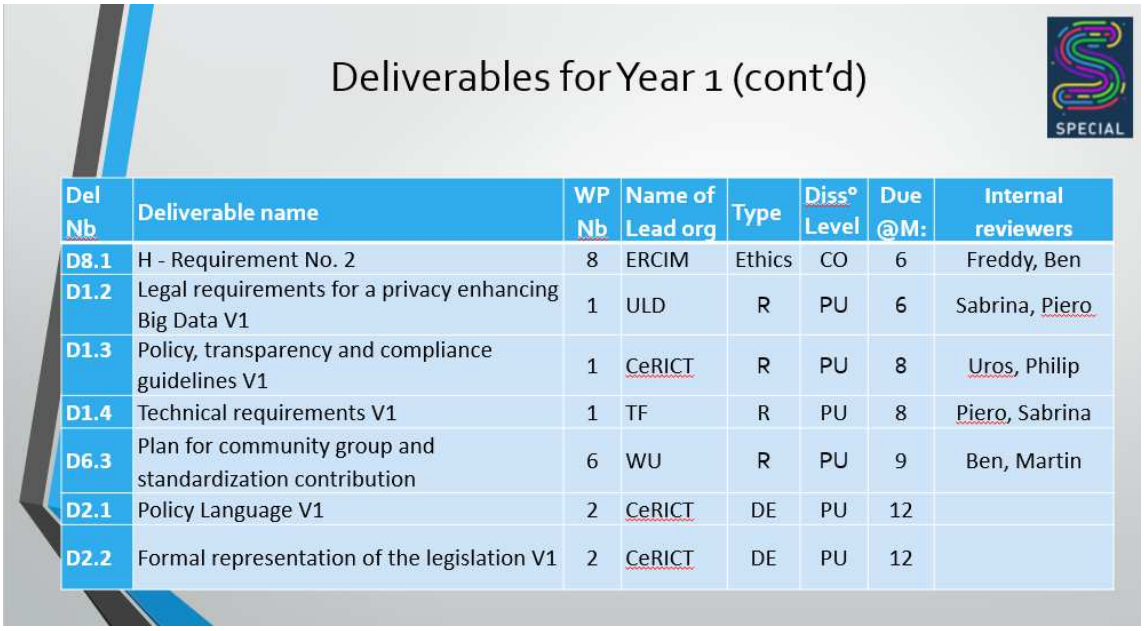
As a concrete result, a session at the kick-off meeting was dedicated to reviewing each single deliverable of Year 1, to reach common understanding of what exactly had to be delivered, and to identify the more suitable partners to review Year 1 deliverables.

This exercise yielded the following table of internal reviewers for Year 1 deliverables, as shown in this slide extracted from the WP7 presentation at the kick-off meeting:



Deliverables for Year 1

Del Nb	Deliverable name	WP Nb	Name of Lead org	Type	Diss° Level	Due @M	Internal reviewers
D7.1	Technical/Scientific coordination plan	7	WU	R	CO	2	Rigo, Philippe
D7.2	Administrative management and support plan	7	ERCIM	R	CO	2	Axel, Sabrina
D6.1	SPECIAL Website Setup	6	ERCIM	DE	PU	3	Harald, Rudy
D1.1	Use case scenarios V1	1	TR	R	PU	5	Martin, Freddy
D3.1	Initial setup of policy aware Linked Data architecture and engine	3	TF	DE	PU	6	Rigo, Philip
D6.2	Public Relations Strategy	6	ERCIM	R	CO	6	Bert (TF), Axel
D7.3	Quality, risk and data management plan	7	ERCIM	ORDP	PU	6	Martin, Rudy



Deliverables for Year 1 (cont'd)

Del Nb	Deliverable name	WP Nb	Name of Lead org	Type	Diss° Level	Due @M:	Internal reviewers
D8.1	H - Requirement No. 2	8	ERCIM	Ethics	CO	6	Freddy, Ben
D1.2	Legal requirements for a privacy enhancing Big Data V1	1	ULD	R	PU	6	Sabrina, Piero
D1.3	Policy, transparency and compliance guidelines V1	1	CeRICT	R	PU	8	Uros, Philip
D1.4	Technical requirements V1	1	TF	R	PU	8	Piero, Sabrina
D6.3	Plan for community group and standardization contribution	6	WU	R	PU	9	Ben, Martin
D2.1	Policy Language V1	2	CeRICT	DE	PU	12	
D2.2	Formal representation of the legislation V1	2	CeRICT	DE	PU	12	

Figure 2.2 - Initial nomination of internal reviewers at the kick-off meeting

All SPECIAL internal reviewers were made aware of this internal review process and agreed to comply with the timeline described in 2.2.1 above.

This initial list of internal reviewers was quickly extended after the kick-off meeting to all deliverables over the full duration of the project. The final list of deliverables and associated internal reviewers is captured in the table below (sorted by delivery date then by WP), which is available on the project repository to all project participants, ensuring the clear and transparent implementation of our Quality Assurance process:

Table 2.1 - List of Internal Reviewers for SPECIAL Deliverables

Deliverable	Deliverable Name	Delivery Date	Lead Participant	Responsible Person	Reviewer 1	Responsible Person	Reviewer2	Responsible Person
D7.1	Technical/Scientific co-ordination plan	M2	WU	Sabrina Kirrane	ERCIM	Rigo Wenning	ERCIM	Philippe Rohou
D7.2	Administrative management and support plan	M2	ERCIM	Philippe Rohou	WU	Axel Polleres	WU	Sabrina Kirrane
D6.1	SPECIAL Website Setup	M3	ERCIM	Bert Bos	WU	Sabrina Kirrane	ULD	Harald Zwingelberg
D1.1	Use case scenarios V1	M5	CeRICT	Ben Whittam Smith	TLABS	Martin Kurze	PROX	Freddy de Meersman
D1.2	Legal requirements for a privacy enhancing Big Data V1	M6	ULD	Eva	WU	Sabrina Kirrane	CeRICT	Piero Bonatti
D3.1	Initial setup of policy aware Linked Data architecture and eng	M6	TF	Uroš Milošević	ERCIM	Rigo Wenning	TUB	Philip Raschke
D6.2	Public Relations Strategy	M6	ULD	Eva	TF	Uroš Milošević	WU	Axel Polleres
D7.3	Quality, risk and data management plan	M6	ERCIM	Philippe Rohou	TLABS	Martin Kurze	PROX	Rudy Jacob
D7.4	Ethical guidelines and procedures	M6	ERCIM	Rigo Wenning	PROX	Freddy de Meersman	TR	Ben Whittam Smith
D8.1	H - Requirement No. 2	M6	ERCIM	Rigo Wenning	PROX	Freddy de Meersman	TR	Ben Whittam Smith
D1.3	Policy, transparency and compliance guidelines V1	M8	CeRICT	Piero Bonatti	TF	Uroš Milošević	TUB	Philip Raschke
D1.4	Technical requirements V1	M8	TF	Uroš Milošević	CeRICT	Piero Bonatti	WU	Sabrina Kirrane
D6.3	Plan for community group and standardisation contribution	M9	WU	Axel Polleres	TR	Ben Whittam Smith	TLABS	Martin Kurze
D2.1	Policy Language V1	M12	CeRICT	Piero Bonatti	TF	Uroš Milošević	TUB	Philip Raschke
D2.2	Formal representation of the legislation V1	M12	CeRICT	Piero Bonatti	ULD	Eva Schlehahn	ERCIM	Rigo Wenning
D1.5	Use case scenarios V2	M14	TR	Ben Whittam Smith	PROX	Freddy de Meersman	TLABS	Martin Kurze
D2.3	Transparency Framework V1	M14	WU	Sabrina Kirrane	ERCIM	Rigo Wenning	CeRICT	Piero Bonatti
D2.4	Transparency and Compliance Algorithms V1	M14	WU	Sabrina Kirrane	TF	Uroš Milošević	TUB	Philip Raschke
D1.6	Legal requirements for a privacy enhancing Big Data V2	M15	ULD	Harald Zwingelberg	CeRICT	Piero Bonatti	WU	Sabrina Kirrane
D3.2	Policy & events release	M16	TR	Ben Whittam Smith	TR	Ben Whittam Smith	TLABS	Martin Kurze
D4.1	Transparency dashboard and control panel release V1	M16	TUB	Philip Raschke	WU	Sabrina Kirrane	CeRICT	Piero Bonatti
D1.7	Policy, transparency and compliance guidelines V2	M17	CeRICT	Piero Bonatti	TUB	Philip Raschke	TF	Uroš Milošević
D1.8	Technical requirements V2	M17	TF	Uroš Milošević	ERCIM	Rigo Wenning	WU	Javier D. Fernández
D3.3	Backend Scalability and Robustness testing report V1	M18	WU	Javier D. Fernández	TF	Uroš Milošević	ERCIM	Rigo Wenning
D4.2	Frontend Scalability and Robustness testing report V1	M18	TF	Uroš Milošević	TR	Ben Whittam Smith	PROX	Freddy de Meersman
D6.4	Market Analysis and Plan for Exploitation	M18	ULD	Harald Zwingelberg	TF	Uroš Milošević	ERCIM	Rigo Wenning
D7.5	Periodic technical and financial report V1	M18	ERCIM	Philippe Rohou	ULD	Harald Zwingelberg	TF	Uroš Milošević
D8.2	POPD - Requirement No. 3	M18	ERCIM	Rigo Wenning	PROX	Freddy de Meersman	TR	Ben Whittam Smith
D5.1	Pilot implementations and testing plans V1	M19	TLABS	Martin Kurze	TUB	Philip Raschke	WU	Olha Drozd
D2.5	Policy Language V2	M21	CeRICT	Piero Bonatti	TF	Uroš Milošević	TUB	Philip Raschke
D2.6	Formal representation of the legislation V2	M21	CeRICT	Piero Bonatti	ULD	Eva Schlehahn	ERCIM	Rigo Wenning
D5.2	Public challenge report V1	M21	TF	Uroš Milošević	ERCIM	Rigo Wenning	ULD	Harald Zwingelberg
D2.7	Transparency Framework V2	M23	WU	Sabrina Kirrane	ERCIM	Rigo Wenning	ULD	Harald Zwingelberg
D2.8	Transparency and Compliance Algorithms V2	M23	WU	Sabrina Kirrane	TF	Uroš Milošević	TUB	Philip Raschke
D3.4	Transparency & compliance release	M25	TF	Uroš Milošević	WU	Sabrina Kirrane	CeRICT	Piero Bonatti
D4.3	Transparency dashboard and control panel release V2	M25	TUB	Philip Raschke	TR	Ben Whittam Smith	PROX	Rudy Jacob
D3.5	Scalability and Robustness testing report V2	M27	WU	Javier D. Fernández	ERCIM	Rigo Wenning	CeRICT	Piero Bonatti
D4.4	Usability testing report V2	M27	TF	Uroš Milošević	TUB	Philip Raschke	WU	Olha Drozd
D5.3	Pilot implementations and testing plans V2	M28	TLABS	Martin Kurze	CeRICT	Piero Bonatti	WU	Sabrina Kirrane
D5.4	Public challenge report V2	M30	TF	Uroš Milošević	ULD	Harald Zwingelberg	ERCIM	Rigo Wenning
D6.5	Final Report of the Community Group	M30	WU	Axel Polleres	TLABS	Martin Kurze	TR	Ben Whittam Smith
D3.6	Final release	M34	TF	Uroš Milošević	ULD	Harald Zwingelberg	ERCIM	Rigo Wenning
D4.5	Transparency dashboard and control panel release final relea	M34	TUB	Philip Raschke	TLABS	Martin Kurze	PROX	Rudy Jacob
D5.5	Pilot implementations and testing plans V3	M34	TLABS	Martin Kurze	WU	Sabrina Kirrane	CeRICT	Piero Bonatti
D5.6	Report on application guideline	M36	ERCIM	Rigo Wenning	TR	Ben Whittam Smith	TLABS	Martin Kurze
D7.6	Periodic technical and financial report V2	M36	ERCIM	Philippe Rohou	TF	Uroš Milošević	ULD	Harald Zwingelberg

2.2.2.2 Continuous monitoring

The SPECIAL consortium planned to hold a minimum of one plenary Telco or one quarterly face-to-face meeting every month of the project. Each one of these physical or virtual meetings will be the occasion to review upcoming deliverables and to ensure at management level that progress is on schedule and that potential issues are under control.

Monitoring and control actions from the coordination team have been further described in earlier deliverables D7.1 'Technical/Scientific coordination plan' and D7.2 'Administrative management and support plan'.

Should any unplanned event arise that negatively impacts the submission deadline, the coordinator will contact the project officer at once to justify the delay and to request in writing an approval to postpone the deliverable by a reasonable amount of time.

2.3 Supporting Tools

With a consortium of nine beneficiaries, the tools required to manage the Quality Assurance process for deliverables do not require a high level of sophistication. Simple yet efficient methods will ensure the consistent quality of SPECIAL deliverables.

Deliverable templates

With the first deliverables due at M2, WP7 produced an initial deliverable template in the first month of the project. This MS Word template includes the logo of the project, which was produced during the kick-off meeting at M1. This template is available to partners on the BSCW project repository, and it is the one used here for this deliverable.

Approaching the more technical deliverables of M5 and M6, the scientific partners opted for collaborative writing based on the Latex document preparation system[1]. A Latex template is now available to partners who need or prefer the superior functionality offered by Latex for collaborative editing.

The outcome of the collaborative editing, whether done in MS Word or in Latex, is a SPECIAL deliverable in PDF format that looks the same regardless of the software that produced it.

E-Mail

In line with the internal review process described in 2.2.1, partners circulate via e-mail the successive versions of the draft and final deliverables.

BSCW

The final version of each deliverable is available to the consortium after submission to the EC, via a shared folder on the project repository (BSCW).

[1] https://www.latex-project.org/

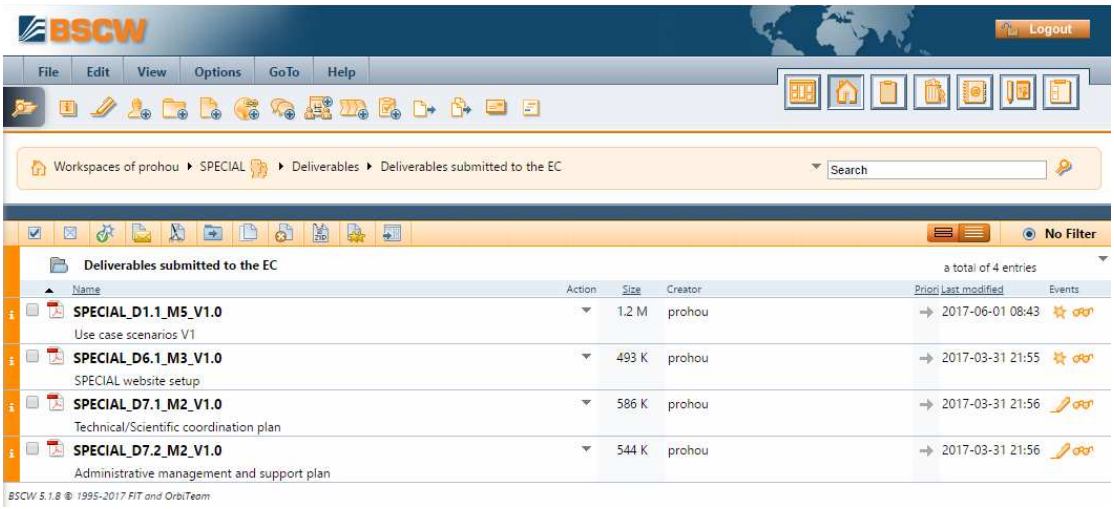


Figure 2.3 - SPECIAL deliverables on the project repository

Shortly before each project review, a special BSCW folder will be created for the reviewers and populated with the submitted versions of the deliverables for the period. Reviewers will be granted secure access to and invited to download from this location.

Web site

Once approved by the project officer at the project review, public deliverables for the reviewed period will be made available on the project web site[2].

2.4 Current State

At M6 of SPECIAL, the relevant procedures are well in place to ensure that the project will deliver quality deliverables, on time and in a controlled fashion. For the full duration of SPECIAL, each project deliverable has been assigned two internal reviewers, and all project participants know what their role is in supporting the successful implementation of the deliverable Q&A processes. Should there be any doubt during the course of the project, all necessary documentation is available on the project repository.

The consortium is committed to follow this procedure and to submit all SPECIAL deliverables on time and to the best possible level of quality.

[2] https://www.specialprivacy.eu/about/public-deliverables

3 Risks

3.1 Overview

This chapter details the internal framework for risk management put in place at the start of the project to ensure the smooth running of the SPECIAL project.

In the Description of Work, the risk management plan is one of the outputs of Task 7.3 'Quality, Risk and Data Management'. T7.3 is under the responsibility of the project coordinator (ERCIM), with the active participation of the scientific coordinator (WU).

In this chapter we: (i) outline the sound and practically applicable risk management methodology we will employ throughout the project; (ii) define procedures and role assignments to tackle risk identification, evaluation, monitoring and mitigation; and (iii) provide and overview of the risk register which will be used to support SPECIAL's workforce and governance structures to contribute to and follow the entire process.

The framework for Risk Management described here will apply to all risks identified to date, as well as to possible additions to this list that may be added as the project progresses.

3.2 Risk Management Methodology

The PRINCE2 risk management procedure[3] provides guidelines to project managers with respect to the identification and evaluation of potential risks and their ongoing monitoring. Although in the DOW we focused on negative risks, it is worth noting that in PRINCE2 risks can have either a negative and a positive impact on success of the project.

The risk management methodology described herein and depicted in *Figure 3.1* is closely aligned with PRINCE2's five steps to risk management:

1. Risk identification involves the identification of the source, casuse and effect of the risk should it materialise. Potential threats are recorded in a risk register (i.e. a list potential risks, including proposed mitigation measures).
2. Risk assessment involves estimating both the impact and the probability of the risk materialising (probability is ranked as low, medium or high).
3. Risk planning involves proposing risk mitigation measures and assessing said mitigation strategies for seconary risks.
4. Implementation is concerned with monitoring the risks, taking action if needs be and monitoring the effectiveness of any actions taken.

[3] PRINCE2 Risk Management, http://prince2.wiki/Risk

5. Communication is a continuous process that aims to ensure all stakeholders are kept up to date with respect to potential risks and risk management activities

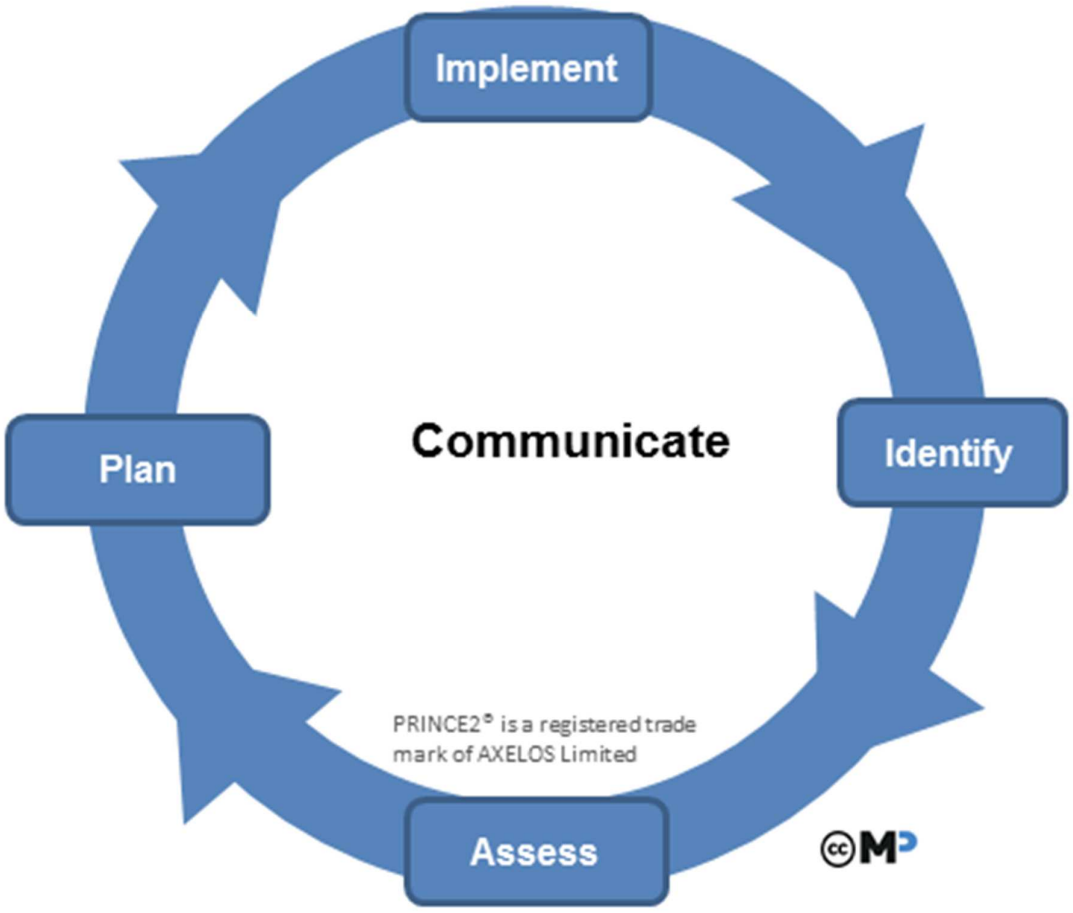


Figure 3.1 - The PRINCE2 RISK management Procedure

(Source http://prince2.wiki/File:Slide44.PNG)

3.3 Roles and Responsibilities

Task leader responsibilities

- Communicate potential risks to the work package leader
- Assist the work package leader with risk management activities, namely identification, assessment, planning and implementation
- Communicate risk status updates to the work package leader

Work package leader responsibilities

- Communicate potential risks to the project co-ordinators
- Perform risk management activities (i.e. identification, assessment, planning and implementation) for work package specific risks

- Assist the the project co-ordinators with risk management activities (i.e. identification, assessment, planning and implementation) across the various work packages
- Communicate risk status updates to the project co-ordinators

Project Co-ordinators (Scientific/Technical and Administrative)

- Provide feedback to the work package leaders on risk identification, assessment, planning and implementation
- Perform risk management activities (i.e. identification, assessment, planning and implementation) for risks that span multiple work package
- Communicate risk status updates to the General Assembly

3.4 Current State

At a minimum the risk register should include the following information:

- Unique risk number for each risk
- Author
- Date registered
- Risk Category (e.g. managerial, Implementation, impact)
- Detailed description of the risk
- Impact and probability
- Relevant work package(s)
- Proposed risk-mitigation measures
- Risk owner (e.g. task leader, work package leader, technical/scientific co-ordinator, administrative co-ordinator)
- Status

Table 3.1 provides a list of the initially identified risks from the DOW, along with the measures foreseen to mitigate those risks, that will be migrated into the risk register and expanded to include additional risks identified during the 1st six months of the project.

Table 3.1 - Critical Risks for Implementation

Risk num	Author	Date Registered	Category	Description of risk	Impact and probability	WPs involved	Proposed risk-mitigation measures	Owner	Status
1	All	Jan 2017	*Managerial*	IPR, licensing or other legal / ethics related issues arise among partners.	*Internal Risk; Low Probability*	WP7	The consortium foresees a number of measures to proactively face such issues, including a Consortium Agreement listing background/foreground. To name some measures: alternative licensing of specific S/W elements is allowed to withdraw related barriers, a clear IPR to inventor(s) policy is adopted, bodies that can handle stages of conflict before they escalate to a problem, and lastly, most partners have been cooperating in various contexts and already have well-established relationships.	ERCIM	Open
2	All	Jan 2017	*Managerial*	A partner / body underperforms, defaults or faces other severe operation issues.	*Internal Risk; Low Probability*	WP1, WP2, WP3, WP4, WP5, WP6, WP7, WP8	Governance structure and procedures (meetings, teleconference, collaboration tools, etc.) allow close monitoring of partner activities that allow any turbulence to be spotted	ERCIM	Open

							promptly. Moreover, measures are foreseen under the Grant and Consortium Agreement terms for the handling of any defaulting or underperforming partner.		
3	All	Jan 2017	*Managerial*	Lack of required competences for the completion of project's tasks.	*Internal Risk; Medium Probability*	WP1, WP2, WP3, WP4, WP5, WP6, WP7, WP8	The consortium members are selected specifically to fill in the pieces of the expertise required in the project. Any underperforming partner in technical or other activities related to lack of expertise will be identified early in the project and immediate rectification measures will be taken.	ERCIM	Open
4	All	Jan 2017	*Implement ation*	Policy framework and engine do not scale to the demanded volume and velocity	*Internal Risk; Low Probability*	WP1, WP2, WP3, WP4	The technology partners, led by the architects of the Big Data Europe project (TF) will from the start of the project onwards focus on solutions that scale. Scalability demands and requirements will be oriented by the use case partners and developed in an agile manner with the combined expertise of TUB, WU and CeRICT on	WU	Open

							distributed systems, scalable Linked Data query answering and reasoning, and reasoning about policies.		
5	All	Jan 2017	*Implement ation*	Dependency chains between related tasks	*Internal Risk; Low Probability*	WP1, WP2, WP3	WP2's tasks related to the development of a policy language depend on the requirements elicited from the use cases. In turn, WP 3 depends partly on WP2's inputs. So any delay or variation in the definition of the use cases may have a domino effect on some of the core technical work packages. If necessary, WP2 may start with a generic approach, by assessing the scalability of the main policy constructs introduced in the literature, studying scalable implementations whenever possible. The agile development process adopted by SPECIAL will help to compensate any delayed specific input coming from the use cases, allowing WP2 and WP3 to start their work in advance, and integrate use-case specific	WU	Open

							requirements when they become available.		
6	All	Jan 2017	*Implement ation*	Introduction of new standards, law and certifications affecting the success of the project activities	*External Risk; Low Probability*	WP1, WP2, WP3, WP4, WP5	As with technological changes, the consortium will monitor pertinent standards and the responsible partners will provide recommendations on how to address these in the project in case changes arise. The fact that there are two rounds of design and development in the project also allows for appropriate adjustments.	WU	Open
7	All	Jan 2017	*Implement ation*	Issues obtaining quality simulated data	*Internal Risk; Low Probability*	WP3, WP5	The use case partners will take all necessary actions to ensure that the simulated data is both realistic and representative of the real data used in the proposed usecase scenarios.	WU	Open
8	All	Jan 2017	Impact	Applicability of project results	*Internal Risk; Low Probability*	WP5	We have chosen use cases from diverse domains (telecom and financial sectors) that expose complementary critical aspects of the challenges we want to solve with respect to privacy-aware solutions that scale to Big Data. Moreover,	ERCIM	Open

							we will actively (by having allocated resources in our work plan) liaise with other ICT-14 projects and the CSA in the ICT-18 call to guarantee applicability of results and impact.		
9	All	Jan 2017	*Impact*	Insufficient end-user engagement in the pilot hacking challenges.	*Internal Risk; Low Probability*	WP5	The project team will use all available dissemination channels to promote awareness, e.g. through standardisation activities, existing networks, social media outlets. Rewarding services could be included to increase the level of user engagement.	ERCIM	Open

4 Data Management Plan

The data management plan described here-in relates to synthesised data ONLY. PROX is the only partner in the project that will collect data via user studies. This data will NOT be made available (even in an anonymized format). Specific guidance with respect to the assessment of ethical issues concerning the collection, processing and disclosure of personal data are described in deliverable D.8.1 SPECIAL Ethics Guidelines.

4.1 Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

- All three industry partners will provide their expertise and domain knowledge in order to enable the generation of simulated data, which we will use in our public challenges.
- Additionally, synthesised data (e.g. datasets, queries and policies) will be generated and used for benchmarking and stress testing purposes throughout the projects.

What types and formats of data will the project generate/collect?

The following types of data will be generated:

- Synthesised ledger entries containing processing and sharing events – who processed/shared, what data, for what purpose, with whom, under what usage conditions.
- Synthesised usage policies – stipulating what data can be used for what purpose by whom and under what usage conditions.
- Synthesised domain specific data (e.g. telecoms and financial data) – encrypted simulated telecoms and financial data and associated metadata.
- Ontologies used to describe ledge entries, usage policies, domain specific data and metadata (e.g. temporal, provenance, permissions, obligations).
- Certain aspects of the General Data Protection Regulation (GDPR) that are necessary for compliance checking within the remit of SPECIAL will be made available in a machine readable format.

All public data will be published according to the 5-star deployment scheme for Open Data[4] :

Tim Berners-Lee, the inventor of the Web and initiator of the Linked Data project, suggested a 5-star deployment scheme for Linked Data. The 5 Star Linked Data system is cumulative. Each additional star presumes the data meets the criteria of the previous step(s).

☆ Data is available on the Web, in whatever format.

☆☆ Available as machine-readable structured data, (i.e., not a scanned image).

☆☆☆ Available in a non-proprietary format, (i.e. CSV, not Microsoft Excel).

☆☆☆☆ Published using open standards from the W3C (RDF and SPARQL).

☆☆☆☆☆ All of the above and links to other Linked Open Data.

Figure 4.1 - 5 Star Linked Data

Will you re-use any existing data and how?

- Where possible existing ontologies will be used used to describe both data and metadata (e.g. temporal, provenance, permissions, obligations). PROV[5] and OWL-Time[6] ontologies can be used to represent provenance and temporal information respectively. Additionally there are a number of general event vocabularies such as the Event[7] ontology and the LODE[8] ontology that could potentially be adapted/extended in order to model our data processing events. Likewise, vocabularies for expressing policies such as the Open Digital Rights Language[9] which the W3C's Permissions and Obligations working group hopes to standardise in the near future, could be adapted/extended.

What is the origin of the data?

- Synthesised ledger entries, usage policies, domain specific data will be generated by the SPECIAL team.
- PROV[5], OWL-Time[6], Events[7], LODE[8] and ODRL[9] are ontologies that are openly available for reuse. Any adaptations/extensions developed by SPECIAL will likewise be given back to the community.

[4] 5 Star Linked Data, https://www.w3.org/2011/gld/wiki/5_Star_Linked_Data

[5] PROV, https://www.w3.org/TR/prov-overview/

[6] OWL-Time, https://www.w3.org/TR/owl-time/

[7] Events, http://motools.sourceforge.net/event/event.html

[8] LODE, http://linkedevents.org/ontology/

[9] ODRL, http://w3c.github.io/poe/model/

What is the expected size of the data?

At this stage of the project it is difficult to quantity. Size can have many meanings:

- number of data sets
- number of data fields
- size of data fields (an image is usually bigger that a binary field

To whom might it be useful ('data utility')?

- Computer Science, Semantic Web and Privacy researchers can make use of some/all of the data for reesearch and benchmarking purposes.
- Companies and researchers may leverage and extend the ontologies developed by SPECIAL.
- Legislators, companies, and researchers may benefit from the subset of the GDPR that will be made availble in a machine readable format.

4.2 Findable, accessible, interoperable and reusable (FAIR) data

This section is based on the guidelines for effective data management in in the course of a Horizon 2020 project, provided by the European Commission[10]. The primary objective is to ensure that research data is findable, accessible, interoperable and re-usable (FAIR).

4.2.1 Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

All synthesised data will be made available as Linked Data. Underpinning the Linked Data Web is a set of best practices for publishing and interlinking structured data, know as the Principles of Linked Data. These principles are defined by Tim Berners-Lee [11]as follows:

” 1. Use URIs as names of things.

2. Use HTTP URIs so that people can look up those names.

3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL).

4. Include links to other URIs. so that they can discover more things. ”

The term thing is used to refer to both real world entities and abstract concepts, commonly referred to as resources. The LDW builds on the existing web infrastructure, by using HyperText Transfer Protocol (HTTP) URIs to identify things, as well as documents. However, URIs only support a subset of the American Standard Code for Information Interchange (ASCII) character set. Later the W3C introduced Internationalised Resource Identifiers (IRIs), that provides support for the richer Unicode character set. Although the principles defined by Berners-Lee refer to URIs, as there is a mapping from IRIs to URIs, it is also possible to use IRIs.

However, it is not enough to simply use URIs to refer to things. According to the Linked Data principles, it should be feasible to use the URI to return a description of the resource (commonly referred to as dereferencing). As URIs often represent real world entities, it is common practice to use different URIs

[10] Data Management, http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

[11] Principles of Linked Data, https://www.w3.org/DesignIssues/LinkedData.html

to represent the resource and the document that describes it. Two different strategies, that can be used to dereference URIs exist, namely 303 redirects and Hash URI's. In the case of 303 redirects, when a client attempts to dereference a resource, the server responds with a 303 See Other, and a URI for the document that describes this resource. The client subsequently uses this new URI to retrieve the description of the resource. Whereas, in the case of hash URIs a # separator is used to append an identifier, which identifies the resource, to the end of the URI. Prior to attempting to dereference the resource, the client strips off the # and the identifier, making it possible to distinguish between the physical resource and the document that describes it. The final principle refers to the linking of URI's. Just like the web of documents uses reference links to enable humans and machines to navigate web pages, the web of data is constructed in a similar fashion. By using RDF to describe resources, it is possible not only to link structured data, but also to describe complex relations between resources in a machine readable format.

What naming conventions do you follow?

All class, properties and instances will be provided with a unique IRI, that will be accessible via the SPECIAL web server. The prefix used for all public data items will be https://data.specialprivacy.eu/.

As the data management plan is a living document more specific naming conventions will be worked out at a later stage if needs be.

Will search keywords be provided that optimize possibilities for re-use?

SPECIAL will adopt the Comprehensive Knowledge Archive Network (CKAN) [12] web-based open source management system, which is developed by the Open Knowledge Foundation, for the storage and distribution of open data. CKAN is already the tool of choice for many national and local governments, research institutions, and other organisations which collect a lot of data. CKAN provides powerful search and faceting, browsing over distributed data sources.

Do you provide clear version numbers?

The CKAN ckanext-datasetversions[13] extension provides support for different versions of a dataset.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

CKAN provides a rich set of metadata for each dataset. The following information can be found on the CKAN website[14]: "

- *Title – allows intuitive labelling of the dataset for search, sharing and linking.*
- *Unique identifier – dataset has a unique URL which is customizable by the publisher.*
- *Groups – display of which groups the dataset belongs to if applicable. Groups (such as science data) allow easier data linking, finding and sharing amongst interested publishers and users.*

[12] CKAN, https://ckan.org/

[13] CKAN datasetversions, http://extensions.ckan.org/extension/datasetversions/

[14] CKAN metadata, https://ckan.org/portfolio/metadata/

- *Description – additional information describing or analysing the data. This can either be static or an editable wiki which anyone can contribute to instantly or via admin moderation.*
- *Data preview – preview .csv data quickly and easily in browser to see if this is the dataset you want.*
- *Revision history – CKAN allows you to display a revision history for datasets which are freely editable by users (as is thedatahub.org)*
- *Extra fields – these hold any additional information, such as location data (see geospatial feature) or types relevant to the publisher or dataset. How and where extra fields display is customizable.*
- *Licence – instant view of whether the data is available under an open licence or not. This makes it clear to users whether they have the rights to use, change and re-distribute the data.*
- *Tags – see what labels the dataset in question belongs to. Tags also allow for browsing between similarly tagged datasets in addition to enabling better discoverability through tag search and faceting by tags.*
- *Multiple formats (if provided) – see the different formats the data has been made available in quickly in a table, with any further information relating to specific files provided inline.*
- *API key – allows access every metadata field of the dataset and ability to change the data if you have the relevant permissions via API. "*

If needs be, CKAN allows for the specification of additional metadata items in the form of name value pairs.

4.2.2 Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

- All three industry partners will provide their expertise and domain knowledge in order to enable the generation of simulated data, which we will use in our public challenges. PROX is the only partner in the project that will collect data via user studies. This data will NOT be made available (even in an anonymized format).
- Additionally, synthesised data (e.g. datasets, queries and policies) will be generated and used for benchmarking and stress testing purposes throughout the projects.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

By default, commercially sensitive data belonging to SPECIAL usecase partners Prox, TLabs and TR will be closed

How will the data be made accessible (e.g. by deposition in a repository)?

As stated in section 4.2.1, synthesised data is will be made available as Linked Data.

What methods or software tools are needed to access the data?

Data will be accessible via HTTP and queryable using RDF (Resource Description Framework) query languages such as SPARQL

Is documentation about the software needed to access the data included?

Resource Description Framework (RDF) is a standard model for data interchange on the Web. Relevant documentation is provided by the W3C.

Is it possible to include the relevant software (e.g. in open source code)?

A pointer to documentation on the relevant standards can be include in open source code.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

As stated in section 4.2.1, synthesised data is will be accessible via the SPECIAL web server. The prefix used for all public data items will be https://data.specialprivacy.eu/.

Have you explored appropriate arrangements with the identified repository?

Not applicable.

If there are restrictions on use, how will access be provided?

All synthesised data will be publically available

Is there a need for a data access committee?

Not at the moment.

Are there well described conditions for access (i.e. a machine readable license)?

The default license for SPECIAL public data will be CC-BY.

How will the identity of the person accessing the data be ascertained?

Not applicable.

4.2.3 Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

Where possible data will be published as Resource Description Framework (RDF). RDF is a standard model for data interchange on the Web. Relevant documentation is provided by the W3C.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

By default, SPECIAL will adopt the RDF data model, and meta language such as RDFS and OWL. Like RDF both RDFS and OWL are W3C specifications

Additionally we will reuse existing RDF ontologies such as PROV[5], OWL-Time[6], Events[7], LODE[8] and ODRL[9].

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

Yes, where possible we will reuse standard vocabularies.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Should existing existing ontologies and vocabularies not meet our needs we will be sure to publish our extensions or new ontologies so that others can reuse.

4.2.4 Increase data re-use (through clarifying licences)

How will the data be licensed to permit the widest re-use possible?

The default license for SPECIAL public data will be CC-BY.

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Generally speaking, data will be made available when papers are published or releases of the SPECIAL software are deployed.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

All synthesised data will be publically available.

How long is it intended that the data remains re-usable?

The specialprivacy.eu domain name has been reserved by ERCIM/W3C – the project coordinator – for 5 years. It has been registered on 17/01/2017 and will expire on 17/01/2022, over-running the project life time by two years. A sustainability plan will be put in place during year 3 of the project.

Are data quality assurance processes described?

Data quality assurance processes will be described before the 1st public release.

4.3 Allocation of resources

What are the costs for making data FAIR in your project?

As we aim to use the existing web infrastructure at the current point in time no additional costs are foreseen.

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

Not applicable.

Who will be responsible for data management in your project?

The technical/scientific co-ordinator.

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

A sustainability plan will be put in place during year 3 of the project.

4.4 Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

For security reasons, six months after the end of the project and with the approval of the project coordinator, ERCIM will copy the whole website to a “static” version which will replace the online dynamic version. Content will still be available online but will stop being editable.

The servers use RAID hardware and redundant power-supplies to ensure high efficiency and availability of our services. Those machines are hosted in a secured machine room in Sophia Antipolis, France. It is a limited-access facility, with air conditioning and uninterrupted power supplies. The systems are backed-up daily and are up and running 24/7.

Is the data safely stored in certified repositories for long term preservation and curation?

A sustainability plan will be put in place during year 3 of the project.

4.5 Ethical aspects

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

Given that we will only share synthesised data there are no foreseen ethical or legal implication.

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

Given that we will only share synthesised data, informed consent is not a consideration.

4.6 Other issues

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

No

5 Conclusion

As early as Project Month 1, the SPECIAL consortium has put procedures in place to ensure that the project will deliver quality deliverables, on time and in a controlled fashion. At Project Month 6, those procedures are completely documented, well publicised within the consortium, and adhered to by all project participants.

The good team work on each deliverable between its owners, its participants, its internal reviewers and the coordination team reinforces the good spirit within the consortium, keeps everyone aware of issues and progress, and ensures that all partners remain committed to the global success of the project.

For a project and a consortium the size of SPECIAL, we feel that the processes and tools put in place to limit risk and ensure quality are adequate and will prove sufficient and successful over the project life time.

6 Annexes

6.1 List of SPECIAL deliverables

Sorted by due date, then by WP.

Del Nb	Deliverable name	Leader	Type	Diss°	Due at M
D7.1	Technical/Scientific coordination plan	2 - WU	R	CO	2
D7.2	Administrative management and support plan	1 - ERCIM	R	CO	2
D6.1	SPECIAL Website Setup	1 - ERCIM	O	PU	3
D1.1	Use case scenarios V1	8 - TR	R	PU	5
D1.2	Legal requirements for a privacy enhancing Big Data V1	5 - ULD	R	PU	6
D3.1	Initial setup of policy aware Linked Data architecture and engine	6 - TF	P	PU	6
D6.2	Public Relations Strategy	1 - ERCIM (ULD)	R	CO	6
D7.3	Quality, risk and data management plan	1 - ERCIM	R	PU	6
D7.4	Ethical guidelines and procedures	1 - ERCIM	R	CO	6
D8.1	H - (Ethics) Requirement No. 2	1 - ERCIM	R	CO	6
D1.3	Policy, transparency and compliance guidelines V1	3 - CeRICT	R	PU	8
D1.4	Technical requirements V1	6 - TF	R	PU	8
D6.3	Plan for community group and standardization contribution	2 - WU	R	PU	9
D2.1	Policy Language V1	3 - CeRICT	P	PU	12
D2.2	Formal representation of the legislation V1	3 - CeRICT	P	PU	12
D1.5	Use case scenarios V2	8 - TR	R	PU	14
D2.3	Transparency Framework V1	2 - WU	P	PU	14
D2.4	Transparency and Compliance Algorithms V1	2 - WU	P	PU	14
D1.6	Legal requirements for a privacy enhancing Big Data V2	5 - ULD	R	PU	15
D3.2	Policy & events release	8 - TR	P	PU	16
D4.1	Transparency dashboard and control panel release V1	4 - TUB	P	PU	16
D1.7	Policy, transparency and compliance guidelines V2	3 - CeRICT	R	PU	17
D1.8	Technical requirements V2	6 - TF	R	PU	17
D3.3	Backend Scalability and Robustness testing report V1	2 - WU	R	PU	18
D4.2	Frontend Scalability and Robustness testing report V1	6 - TF	R	PU	18
D6.4	Market Analysis and Plan for Exploitation	5 - ULD	R	CO	18
D8.2	POPD – (Ethics) Requirement No. 3	1 - ERCIM	R	CO	18
D5.1	Pilot implementations and testing plans V1	7 - DTAG	P	PU	19
D2.5	Policy Language V2	3 - CeRICT	P	PU	21
D2.6	Formal representation of the legislation V2	3 - CeRICT	P	PU	21
D5.2	Public challenge report V1	6 - TF	R	PU	21
D2.7	Transparency Framework V2	2 - WU	P	PU	23
D2.8	Transparency and Compliance Algorithms V2	2 - WU	P	PU	23
D3.4	Transparency & compliance release	6 - TF	P	PU	25
D4.3	Transparency dashboard and control panel release V2	4 - TUB	P	PU	25
D3.5	Scalability and Robustness testing report V2	2 - WU	R	PU	27
D4.4	Usability testing report V2	6 - TF	R	PU	27
D5.3	Pilot implementations and testing plans V2	7 - DTAG	R	PU	28
D5.4	Public challenge report V2	6 - TF	R	PU	30
D6.5	Final Report of the Community Group	2 - WU	R	PU	30
D3.6	Final release	6 - TF	P	PU	34
D4.5	Transparency dashboard and control panel release final release	4 - TUB	P	PU	34
D5.5	Pilot implementations and testing plans V3	7 - DTAG	R	PU	34
D5.6	Report on application guideline	1 - ERCIM	R	PU	36

6.2 Internal review checklist

Author(s) responsible for the Deliverable:

WP leader:

Internal reviewer:

Assurance Point	Issues to be addressed	Assessment	Comments	Recommendations
Compliance with the objectives of SPECIAL	Does the deliverable comply with the overall objectives of the project?	YES NO PARTIALLY		
Compliance with the specific objectives of the workpackage	Does the deliverable comply with the WP Objectives as specified in the WP description?	YES NO PARTIALLY		
Correspondence with the description of work of the relevant activity	Does the deliverable correspond with the activity description as specified in the Application Form?	YES NO PARTIALLY		
Compliance with the deliverables format	Is the deliverable presented using the Project's deliverable format?	YES NO		
Adequacy of written language	Level of written English	EXCELLENT ADEQUATE POOR		
Overall assessment and suggestions for improvement:				
Date of Quality Assurance performed by the internal reviewer:				

6.3 DMP summary and follow-up tables

SUMMARY TABLE 1

FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined above.

DMP component	Issues to be addressed
1. Data summary	• State the purpose of the data collection/generation • Explain the relation to the objectives of the project • Specify the types and formats of data generated/collected • Specify if existing data is being re-used (if any) • Specify the origin of the data • State the expected size of the data (if known) • Outline the data utility: to whom will it be useful
2. FAIR Data 2.1. Making data findable, including provisions for metadata	• Outline the discoverability of data (metadata provision) • Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? • Outline naming conventions used • Outline the approach towards search keyword • Outline the approach for clear versioning • Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how

2.2 Making data openly accessible	• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so • Specify how the data will be made available • Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? • Specify where the data and associated metadata, documentation and code are deposited • Specify how access will be provided in case there are any restrictions
2.3. Making data interoperable	• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. • Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	• Specify how the data will be licenced to permit the widest reuse possible • Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed • Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why • Describe data quality assurance processes • Specify the length of time for which the data will remain re-usable
3. Allocation of resources	• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs • Clearly identify responsibilities for data management in your project • Describe costs and potential value of long term preservation
4. Data security	• Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	• Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

HISTORY OF CHANGES		
Version	**Publication date**	**Change**
1.0	13.10.2016	▪ Initial version