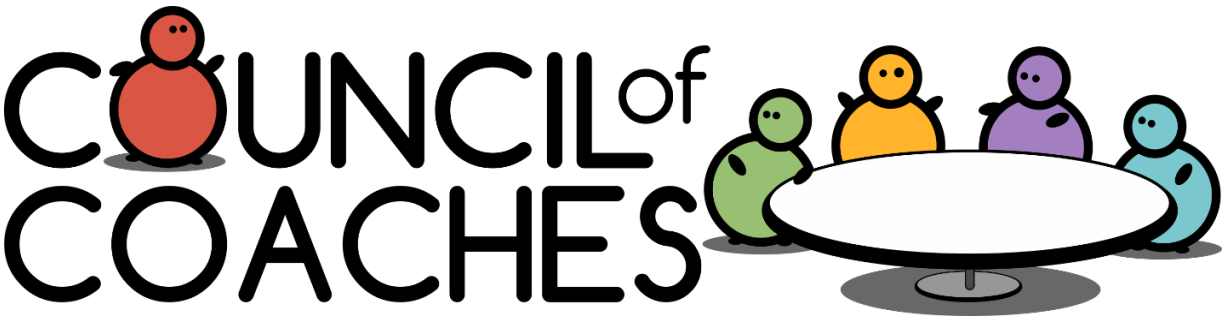


D2.7: Report on ethical, social and legal considerations and implications

Dissemination level: Public
Document type: Report
Version: 1.0.0
Date: July 30th, 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement #769553. This result only reflects the author's view and the EU is not responsible for any use that may be made of the information it contains.

Document Details

Project Number	769553
Project title	Council of Coaches
Title of deliverable	Report on ethical, social and legal considerations and implications
Due date of deliverable	June, 2019 (M22)
Work package	WP2: Responsible Research and Participatory Design
Author(s)	Sita Ramchandra Kotnis (DBT), Bjørn Bedsted (DBT)
Reviewer(s)	Harm op den Akker (RRD), Jorien van Loon (CMC)
Approved by	Coordinator
Dissemination level	PU - Public
Document type	Report
Total number of pages	49

Partners

- University of Twente – Centre for Monitoring and Coaching (CMC)
- Roessingh Research and Development (RRD)
- Danish Board of Technology Foundation (DBT)
- Sorbonne University (SU)
- University of Dundee (UDun)
- Universitat Politècnica de València, Grupa SABIEN (UPV)
- Innovation Sprint (iSPRINT)

Abstract

In this report we present an analysis of the responsibility issues pertinent to the Council of Coaches project along with an outline and assessment of the solutions that the project has sought to implement to manage these issues. We define a 'responsibility issue' (RRI-issue) as an issue where the ambitions of the project to do good at same time imply a risk of doing harm. An RRI-issue may be ethical, social, legal, and/or medical of nature and often implies several of these qualities. To manage such issues, the Council of Coaches project has implemented a work package for responsible innovation and participatory design coordinating and facilitating RRI-work and reflection as the project unfolded. In the context of this work package, the collaborating researchers, the research management, and the exploitation management of the project have been supported in reflecting on responsibility issues in the project while at the same time being confronted with inputs and assessments from potential end-users and stakeholders in the field of eHealth. This report describes the RRI-issues we have been working with and their potential implications.

Table of Contents

List of figures

List of tables

Symbols, abbreviations and acronyms

CMC	Centre for Monitoring and Coaching
COUCH	Council of Coaches
D	Deliverable
DBT	Danish Board of Technology Foundation
EC	European Commission
ELSI	Ethical, Legal, and Social Implications
EU	European Union
FP7	EU 7th Framework Programme for Research and Development
GDPR	General Data Protection Regulation
HGP	Human Genome Project
ISPRINT	Innovation Sprint
M	Month
MS	Milestone
R&I	Research & Innovation
RRD	Roessingh Research and Development
RRI	Responsible Research and Innovation
STIR	Socio-Technical Integration Research
SU	Sorbonne University
UDun	University of Dundee
UPV	Universitat Politècnica de València
UT	University of Twente
WP	Work Package

1 Introduction

In the Council of Coaches project, RRI-work on ethical, social, and legal considerations and implications has been developed, carried out and fed back into technical, product and process development of research and project work from the very outset. This report outlines the issues that have been at the core of this work throughout the project, how we have sought to anticipate, handle and take their implications into account and how we evaluate the results of this work.

1.1 Concern and scope of this report

The purpose of this report is to present the responsibility issues pertinent to the Council of Coaches project along with an analysis and assessment of the solutions that the project has sought to implement to manage these issues. We define a 'responsibility issue' (RRI-issue) as an issue where the ambitions of the project to do good at same time imply a risk of doing harm. An RRI-issue may be ethical, social, legal, and/or medical of nature and often implies several of these qualities. To manage such issues, at the outset of the project the Council of Coaches implemented a work package for responsible innovation and participatory design, coordinating and facilitating RRI-work and reflection as the project unfolded. In the context of this work package, the collaborating researchers, the research management, and the exploitation management of the project have been supported in reflecting on responsibility issues in the project while at the same time being confronted with inputs and assessments from potential end-users and stakeholders in the field of eHealth. This report describes the RRI-issues we have been working with and their potential implications.

1.2 Objectives of the Council of Coaches project

The goal of the Council of Coaches project has been to introduce a new concept of virtual coaching by advancing the state of the art in applying self-learning artificial intelligence for highly tailored interaction. The idea is basically to offer the user feedback on his/her personal and medical issues by way of a collection of multiple autonomous agents, each representing a specific type of knowledge, which are able to educate and motivate the user in interactive group discussions in a holistic, personal and inclusive manner. The physical and technical realization of the idea consists of a platform to which coaching applications from various sources may be attached. Along with the platform itself, the Council of Coaches project has also developed a number of virtual agents to populate the platform, each with their own expertise, personality and style of interacting with the user. The prototypical scripts (argumentative and cognitive frameworks) for the actual exchanges with the user has been co-designed with user groups modelling the target population of elderly users and users with diabetes or chronic pain to ensure intuitive, personal human-computer interaction. Throughout, there has been an explicit focus on providing a rich user experience through a high level of attention to visual and audio design, character development, interactive storytelling, and user interface design.

The project has demonstrated a fundamental potential of opening up radical new research directions in the field of virtual coaching and human-computer interaction in general and the multiple-agents approach itself also proves very promising for user involvement, engagement, and education in a number of fields beyond lifestyle and healthy life choice. In the long term, the Council of Coaches platform may grow to benefit users in areas far beyond health applications with councils being built by third-party providers, e.g. environmental debates and issues of sustainable choices and trade-offs.

The agenda of the Council of Coaches project has been technologically as well as commercially ambitious and has a huge potential to do good and address various social, medical, educational, organisational, and financial issues in modern society. However, when designing innovative technological solutions for a better society, all good intentions always also comes with a certain risk of blind spots. In the eagerness for meeting technological challenges, every attempt to build a solution or solve an issue also contains the prospect for causing unintended problems. Therefore, every EU-funded project must reflect on and come up with a plan for managing their research and innovation in a responsible manner. In the Council of Coaches project this need has been a core priority from the very outset.

Alongside the development of the Council of Coaches platform, virtual coaches and technical/argumentative knowledgebases and frameworks, it has been an objective of the project to experiment with and develop a protocol for responsible research and innovation work in accordance with the RRI-concept stemming from the EU's 7th Framework Programme for Research and Development (FP7) which has been gaining increased importance during the implementation of Horizon 2020. This protocol – the Council of Coaches RRI Vision – has served as a guideline throughout the project. Its function and purpose has been to serve as an institutional co-creative formal checklist which, together with the socio-technical integration process, has helped us keep our focus on the identified issues and held each other invested and responsible in fulfilling our mutual obligations towards responsible development.

It has been our wish to embody a responsible approach to virtual coaching and lay the groundwork for the Council of Coaches platform – Agents United – in a responsible manner that lives up to the values of society during the time of its existence, in general and in specific relation to the health prototypes developed by the project.

In the following a bit of historical, conceptual, and empirical background is reviewed before the main responsibility issues for the Council of Coaches project, their development and the solutions pursued are described and discussed.

1.3 What is an 'RRI-issue'?

Responsible Research and Innovation means involving society in science and innovation 'very upstream' in the processes of R&I to align its outcomes with the values of society (RRI Tools https://www.rri-tools.eu/about-rri)

RRI - the European Commission's umbrella concept for the five RRI 'keys' (public engagement, science education, gender equality, ethics, and open access) connects different aspects of the relationship between R&I and society. It serves as an overarching framework for the scope of responsibilities that the Commission foresee necessary to engage in order to meet the ethical and societal demands of contemporary research and innovation.

RRI is intended to anticipate and assess potential implications and societal expectations, with the aim to foster the design of inclusive and sustainable research and innovation and has been a 'cross-cutting issue' in Horizon 2020, and promoted throughout Horizon 2020 objectives (Responsible Research and Innovation/Horizon 2020 http: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation).

The European Commission has prioritised the development of the RRI concept and invested significant resources in making it tangible and operable as policy as well as at a practical level. However, how to translate and operationalise the declarations and principles in concrete project work puts individual research consortia in somewhat uncharted territory anyway with a great deal of sense-making to do on their own. To steer this process, the Council of Coaches dedicated a work package to explicitly focus on RRI and oversee the correctness of the innovation process and its implementation in terms of ethics and societal expectations. The ensuing RRI-approach has been inspired by *The RRI Practice project* (https://www.rri-practice.eu/) which aim has been to develop a 'practical handbook' for the actual implementation of RRI together with *The Responsible Industry framework* (Stahl, 2017) which suggests that the implementation of all R&I activities should be accompanied by user and stakeholder engagement and not least *The Socio-Technical Integration Research* (STIR) method (Fischer & Schuurbiers, 2013). Our methodological adaptation and ensuing construct have been described elsewhere (D2.1 and D2.8). However, for clarification here, central to the approach is the identification of and mutual agreement on a number of particularly pertinent issues that we, from the very beginning of the project, foresaw as in need of special attention and management in relation to the work the Council of Coaches project were to pursue, in order to steer clear of well-intended but double-edged

actions and dispositions that could simultaneously hurt someone or something in the process. These 'particularly pertinent issues' are what we call our *RRI-issues*.

Thus, we define an RRI-issue as **an issue where the ambitions of the project to do good at same time imply a risk of doing harm**. An RRI-issue may be ethical, social, legal, and/or medical of nature and often implies several of these qualities.

2 Background

2.1 Ethical, social, and legal implications of research and innovation

The Ethical, Legal, and Social Implications (ELSI[1]) program was founded in 1990 as part of the Human Genome Project (HGP). The managers recognized that the information gained from mapping and sequencing the human genome would have profound implications for individuals, families, and society. While this information would have the potential to dramatically improve human health, they also realized that it would give rise to several complex ethical, legal, and social issues. How should this new genetic information be interpreted and used? Who should have access to it? How can people be protected from the harm that might result from improper disclosure or use? (https://ghr.nlm.nih.gov/primer/hgp/elsi)

The mission of ELSI was to identify and address issues raised by genomic research that would affect individuals, families, and society and became an integral part of the HGP. ELSI provided a new approach to scientific research by identifying, analysing, and addressing the ethical, social, and legal implications of human genetics research simultaneously with the basic science being carried out. In this way, problematic areas could be identified, and solutions developed before scientific information was integrated into health care practice (*ibid*).

ELSI was thus a pioneer way of thinking about the broad potential impacts of scientific research in terms of society, economy, politics, environment etc. In response to perceived shortcomings of the ELSI approach (i.e. criticism for being too oriented toward research outcomes, rather than processes of research and development), the RRI approach developed over the past decade or so succeeding ELSI, notwithstanding inheriting its focus on ethical implications in real time. Today, RRI is an integrated part of project requirements in the Horizon 2020 framework. It is now mandatory in all research projects – for proper research planning and for due diligence in the execution of project tasks and reporting – to keep an inventory of potential risks and sensitive areas to be able to anticipate and attend to research uncertainties and steer project developments in the right direction in good time, which may sometimes be before the fact. To be able to do this, however, implies a whole lot of reflective work and foresight.

A project like the Council of Coaches, with its focus on tech development in a healthcare context, an agenda to develop user-driven technology thus engaging with a large number of users belonging to vulnerable segments (chronically ill and elderly) in society, as well as projected artificial intelligence-mediated interaction with putative future users, must navigate potential ethical, social and legal situations and dilemmas on numerous levels. It is crucial to remain alert, aware, and self-reflective about the ways we choose to design research, tech, and interventions - from components to interactions with users. As noble and well-intentioned our decisions may be, often there is a flipside that must be considered. WP2 of the Council of Coaches project has had an explicit focus on accommodating this need throughout the project.

2.2 The Council of Coaches Responsibility Vision, an overarching ambition of responsibility

> *The Council of Coaches RRI Vision is "an agreement between the project partners about what responsibilities arise from the ambitions of the project, who needs to bear these responsibilities, and how the project is going to ensure that they do" (Council of Coaches RRI Vision D2.1)*

The starting point for the Council of Coaches' RRI approach has been the Council of Coaches RRI Vision, as laid out in D2.1. Throughout the project, the RRI Vision has served as the main document outlining the consortium's shared understanding and vision of what RRI means in concrete terms for the project

[1] In a European context we use the acronym ELSA (Ethical, Legal, and Social Aspects) for this approach.

and how this vision should be realised. The Vision came into being during a workshop in M5 where the consortium engaged in co-creative processes designed to create upstream reflection on research and innovation, and to facilitate related debate, negotiation and learning on how to implement RRI in the specific R&I processes.

The RRI Vision has also been the anchor and guideline of the work performed in T2.4, the Socio-Technical Integration Process, which signifies the concrete, empirical integration work performed by WP2 throughout the first two years of the project, with the aim of supporting the integration of ethical and societal perspectives in the research processes conducted by members of the consortium and furthering reflection on the RRI issues in real-time innovation to support anticipation and responsiveness.

The RRI vision has helped to identify risks and potentials arising out of the project's ambitions, which we have had to deal with, avoid or realise – and it has served as an example of how RRI can be operationalised with regard to the specific ambitions of a particular research project, thus demonstrating its potential as a genuine methodological hands-on RRI-tool.

3 Responsibility issues and the solutions pursued by the Council of Coaches

3.1 Process of identifying and working with the issues

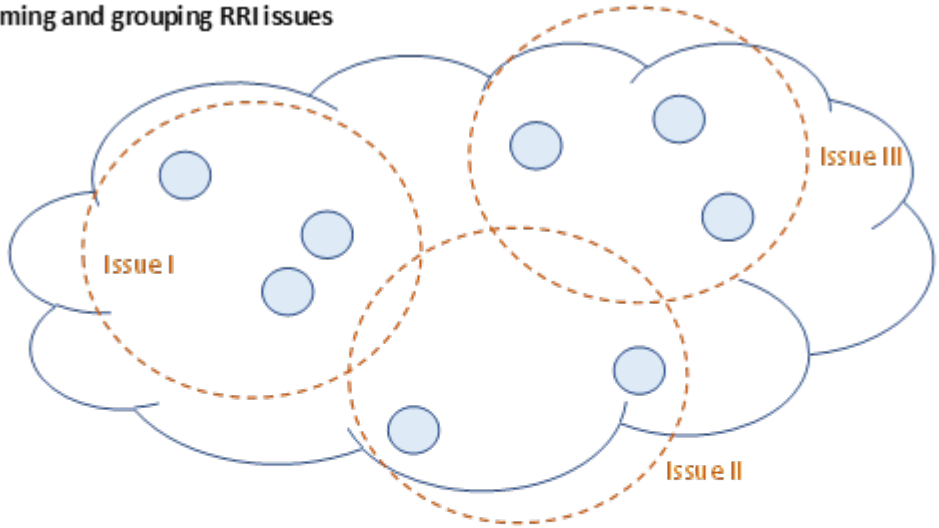
The RRI issues that we have pursued throughout the Council of Coaches project came into being by way of a strictly orchestrated, yet open and explorative co-creation process as described in D2.1.

The consortium embodied in 14 participants from all partners, met in Copenhagen in M5 for a two-days' workshop with the first day spent together with invited stakeholders and the next with the consortium only, working on synthesizing and streamlining the suggestions into concrete RRI issues.

In advance of the workshop all participants had been given a background brief (drafts of chapters 2, 3, and 4 of the ensuing RRI Vision D2.1) two weeks before to give them time to consider the RRI agenda, the approach to be taken, and the specific codes of conducts and ethics that might be relevant to the project. The actual workshop was divided into three sessions. Session one consisted of a brainstorm, informed by the background brief, of RRI issues that immediately occurred to the participants, and a common grouping exercise. The second session divided the participants into three groups, each of which developed a detailed description of one or several closely linked RRI issues, along with suggestions for how to handle them and who should be responsible. The third session was a group discussion of where in the project plan to place the responsibility to follow up on these suggestions (D2.1)

The three sessions of the RRI workshop format and their purposes are illustrated in Figure 1 below:

1. Brainstorming and grouping RRI issues



2. Explicating RRI issues

	Issue I	Issue II	Issue III
What is the issue?			
Why is it important?			
How can the issue be handled?			
Who is responsible?			

3. Concretizing responsibilities

RRI issues and solutions	WP1	WP2	WP3	WP5	WP6	…
Issue I						
• Solution I.A	X					
• Solution I.B	X		+	+		
Issue II						
• Solution II.A						
Issue III						
• Solution III.A		X		+	+	
• …						

Figure 1: Illustration of the three sessions of the RRI workshop and their purposes.

The output of the workshop was a prioritized list of issues that were to steer the project's RRI onward. There were four main issues and four additional issues on this original list. The issues are presented in the next section. In the last four chapters of this report, the main issues and our work operationalizing them and keeping them in check are described in further detail.

3.2 Main issues and 'sleeper issues' – a definition

At the workshop, four especially pertinent issues were identified by the consortium. Along with the identification, initial ideas for solutions and assignments of responsibilities for following up on these issues were also listed. These were the issues:

- **Issue #1: Privacy and informed consent**
- **Issue #2: Trust (not too little, not too much)**
- **Issue #3: Handling disagreement between coaches**
- **Issue #4: How to keep healthcare knowledge up to date?**

These issues are all reviewed, one by one, in further detail in the following chapters.

Also, during the workshop four additional issues surfaced. These were considered secondary to the main issues and were not explored in further detail – hence the naming as 'sleeper issues', indicating 'still there, but not in the centre of attention' – but were recorded and saved for further work along the line in T2.4 (Socio-Technical Integration Work). The thinking was that the 'sleeper' issues might prove important as the project progressed or further downstream in the phases leading towards product development. The sleeper issues will not be further described in this report but has been part of our horizon in the RRI work and facilitated debates carried out throughout the project. Table 1 below shows the sleeper issues and the reflections tied to them:

Table 1: List of sleeper issues and reflections.

Issue #5: Intrusion	
What is the issue?	There are several ways the coaches' behaviour can become intrusive. The coach may ask questions that are deemed too personal. The frequency of the information or questions provided on daily basis can be perceived as 'pushy'. Long talks/messages can be annoying.
How do we address it?	▪ Enable use-profiling (WP 2 and 3) ▪ Make questions more personal with time / build relationship ▪ Include a "do not disturb option" ▪ Enable "do not record data" for specific conversations ▪ Prompt the user to participate in the conversation ▪ Summary-option for long conversations with intentions of humour that might be annoying to some users
Issue #6: Personalisation	
What is the issue?	Who decides how the coaches are personalised and on what background? Are they individually shaped or even for all?
How do we address it?	We need insights into users' preferences and behaviours to know how to handle this issue.

Issue #7: Consistency/honesty	
What is the issue?	What if there is a discrepancy between sensory information and what the user says? If the user lies then the coaches may advice wrongly.
Issue #8: Privacy by design	
What is the issue?	How and to which extent can we ensure that organisations implementing Council of Coaches software downstream are not able, without proper legal authorisation, to mine personal data for other uses than those to which the user has given consent? This issue is especially important if the Council of Coaches were ever implemented in an ecosystem of other digital services, e.g. the national health care system.
Issue #9: Illegitimate exploitation	
What is the issue?	What can we do to prevent sharing / theft of the application?
Issue #10: Economic vs health incentives	
What is the issue?	How do we balance the need for the coaches to refer to GPs and other healthcare personnel in situations that go beyond the expertise of the coaches? This goes against the policy-level need to stimulate self-management and health to lessen the strain on public health resources.
Issue #11: Equality	
What is the issue?	What can we do to ensure that the Council is adopted and used by those who need it the most, not only those for whom use of the system is easiest?
Issue #12: Liability	
What is the issue?	Who is responsible in cases of damage due to wrong advice?

4 Privacy and informed consent in a medical coaching application

During our StiRRIng process in T2.4 (see D2.8 for further details about this process) following the **RRI Vision Workshop**, we conducted a total of four independent RRI workshops along with smaller interventions at consortium meetings. In the following, the outputs from **RRI Workshop 1, held at University of Valencia during the consortiums first Technical Integration Week on March 7th, 2018,** are presented, together with outputs from **RRI Workshop 4, held at University of Twente during the consortium's fourth Technical Integration Week on October 11th, 2018**, where part of the exercise was to think up the perfect stakeholder representation possible to help the consortium manage their ongoing RRI-work further. We have specifically chosen to report most extensively on these two events here in order to show the development of the issues and the management work in a clear manner over a prolonged time span. However, developments debated during the intermediary RRI workshops (Workshop 2 and Workshop 3) has been added to the outlines presented.

The structure used in this chapter is continued in chapters 5, 6 and 7.

4.1 Description of the issue and its potential implications

During the RRI Vision Workshop, the issue **Privacy and informed consent** was envisioned in the following manner (these forms are the cleaned and edited version of the notes that were produced in the groups during the sessions, together with transcripts of recorded discussions):

Table 2: Outset: Issue #1: Privacy and informed consent.

Issue #1: Privacy and informed consent	
What is the issue?	In order to function optimally, the Council of Coaches must be able to collect and process multiple kinds of personal and medical data (including physical and mental health data) and share this data with multiple actors, e.g. general physicians. At the same time, users must be able to understand how data is used and retain control over the data
Why is it important?	**Law** With the GDPR, enabling users to stay in control of their data (cf. the right to be forgotten) is a legal requirement for anyone implementing the Council of Coaches in a real-life situation with European citizens. One thing is that the Council of Coaches project must be able to grant this ability in trial settings, but anyone wishing to license the outputs downstream will be faced with this challenge as well. Proactively engaging with the issue of privacy and informed consent will therefore help to make the Council of Coaches more attractive to potential licensees/investors/etc. **Moral obligation** More than a legal requirement, there is arguably a moral obligation to ensure transparency and user empowerment when designing ICT solutions in the health area. **Product quality enhancement** Solving the issue of privacy and consent in a proactive manner would free the consortium to pursue design solutions that make full use of the data gathered and thus enhance the quality of the coaching support provided by the product. **Enhanced capability of generalising**

	Solving the issue of privacy and consent upstream manner would further enhance the flexibility and scalability of the Council of Coaches platform. We would have to develop specific solutions to ensure privacy and informed consent in a real-life implementation of the specific products being prototyped in the project. But beyond these specific solutions, the Council of Coaches platform – which should ideally be usable by developers targeting entirely different actor-networks – would also benefit greatly from a more generic solution to handling this issue. It would thus be very useful to design a 'plug and play' system for consent and use of personal data, which could be taken up by any application developed on the Council of Coaches platform.
How do we address it?	The group discussing this issue suggests a modular system of data collection and data processing in which the user is able to consent to gradually more types of data being used, dependent on the kind of advice the user wants to make use of and which external actors the user wishes to give access to the data. This would involve a 'basic' consent to the minimal data needed for the Council to work, and additional modules being added on, e.g. corresponding to each new coach being added (physical health, mental health, social wellbeing, etc.). The system would have to be able to distinguish different types of data and different purposes of use: **Types of data** Basic data, mental health, physical health, social **Purposes of collection and processing** Internal to the council, 3rd party uses (GP, insurance, …)
Who does what?	Such a system would have to cut across content modules, data collection mechanisms, and data processing – it is thus a cross-cutting issue which all technical work packages must work together to solve.
Who takes the lead?	LEGO system of consent (proactive approach) (lead T2.4: working withT1.1, T3.3, T7.1, T8.3P) ▪ User interface (push messages on consent) ▪ Data collection and storage (checking w/ consent database) ▪ Data processing (what kind of advice will be given compared with data gathered)

As described in the previous, this conceptualization served as our starting point. During the ensuing StiRRIng process, we kept stirring the reflection process in the consortium as to how the handling of the issues proceeded. This is the topic of the next section.

4.2 Early developments of the issue and management work

The first formalized inputs for our StiRRIng process took place in the form of **RRI Workshop 1**, held in connection with the first Council of Coaches Technical Integration Hackathon at the Polytechnic University of Valencia. The aims of the Hackathon were:

- To establish what is required to connect various pre-existing components that will be integrated into the final COUCH system
- To develop a Technical Demonstrator that plays out a scripted scenario
- Provide all attending partners with an insight into other partners' tools, components, platforms etc.

With the RRI issues as formulated in the RRI Vision in hand, we designed a workshop format suited to integrate ethical, social, and legal reflection into what the researchers gathered were already mentally tuned into and reflecting about.

The workshop consisted of two hours of facilitated group work around the halfway-mark of the hackathon (i.e. the morning of the second day). The idea was to integrate the RRI-issues in the mindset and discussions of the technical partners going into the second day of the hackathon.

A small presentation was made by WP2 to sum up on the different RRI-issues, stakeholder concerns and 'sleeper issues' for the people who were not present when the issues were identified in T2.1. The hackathon group was divided into four smaller groups, all with at least one member who was there when the RRI-issues were originally identified. The groups were all given one of the major RRI-issues, a list of Sleeper-issues and a list of stakeholder concerns. They had one hour to discuss following questions:

1. How does your issue relate to the themes of the hackathon?
2. How relevant is the proposed solution for the issue to the tasks of the hackathon? (What needs changed or added?)
3. What is the group's proposal for taking this issue into account?
4. Should any sleeper issues or stakeholder concerns be prioritised?

Lastly, each group presented their issue(s) and views upon how to deal with the issue(s), to the rest of the technical experts and DBT. The presentations were discussed in plenum and the discussion was recorded. The following presents the input from the Hackathon intervention:

RRI issue 1: Privacy and informed consent:

The issue

- *The Council of Coaches needs to use multiple data sources*
- *At the same time, we must support full control by the data subject*

Proposed solution(s)

- *A modular system of data collection and processing*
- *Gradual extension of consent to new types of data and purposes*
- *A 'LEGO' system of informed consent*

Table 3: Early developments: Privacy and informed consent.

ISSUE #1	PRIVACY AND INFORMED CONSENT
How does the issue relate to the themes of the hackathon?	It relates in a number of ways. In terms of middleware choices, how data is sent around, dialogue management and management of 3rd party external coaches. **Middleware choices**: Regarding "internal" middleware: AMQ, password free, public access, this should obviously change. Regarding "external" middleware: putting, e.g., the "sensor box" exposed on the UniversAAL network will lift the consent issue up beyond COUCH and brings it also into the context of other "3rd party" systems (can 3rd party X access sensor data Y?). **Sending around data**: Should there be consent to use 3G connection for data upload? We need to encrypt what we send as it is a distributed system.

	Dialogue management: We must respect the privacy of users. What do we say out loud? Do we have consent to ask about certain things? Asking about something is also a measurement that might require consent. This raises a number of questions: ▪ If a user mentions something in dialogue that is not part of the given consent, what do we do? Do we store it? Do we infer implied consent? Or do we forget (which is not ideal since the user probably mentioned it for a reason and it could be important). ▪ Are the consents a part of the Knowledgebase / a potential topic for conversation? In other words, can the coaches ask verbally for consent to gather data, or should this be handled in parallel or beforehand? ▪ Do we do speech recognition? If so, should we be able to recognize the speaker who gave consent in order to avoid including other people (e.g. other users, people who happen to be nearby) in a private conversation? **Management of 3rd party external coaches**: "New external coaches" do not automatically share all formerly given consents even when they join on the middleware. So, consent must in the future at least potentially be specified on a per-coach level.
How relevant is the proposed solution to the tasks of the hackathon? (What needs to be changed or added?)	Incremental consent seems very relevant. Consent reminders should be added to review permissions, ex. "it is already half a year ago that we talked about consent do you want to review the status across all devices?" Also, they could be informative, ex. "this is what module A is actually storing" (like with cookies where you never get this reminder). There could be cross-device consent or per-device consent with a referral to the general consent, ex: "you consented that I could collect movement activity data, do you want me to include this device in that?" Some coach advice needs also to consider whether we actually have consent to act on certain data. The knowledge base should incorporate awareness of missing or incomplete data (no consent on a certain device or no consent of a certain category of data as opposed to no data / no connection to device). For AWARE: make sure we do not ask for "more than we need" from the mobile phone. What is the current standard in relation to Fiware/Universaal for this "informed consent for external partners"? Not all coaches will be having access to all knowledge from all other coaches. Perhaps especially the "new extra coach". Is this a problem? Extra: Do we need consent to use a non-version of data for generalized improvement?

What is your proposal for taking this issue into account in relation to the system integration? (Who should do what?)	Consent to: "we acquire data for the next 6 months and afterwards will review with you again; you can also review intermediate if you want". Consent should involve "dialog topics" that the coaches are allowed to ask about, re: potentially sensitive issues (this relates to "intrusive" sleeper topics). There should be potential per-coach consent; for sure if a new coach joins, a full review for that coach is necessary. Is there an official explicit "end of coaching track involvement moment"? If yes, at that point there should be a conversation with the user about discontinuation of the running consents. Is consent acquired in a coach dialog, or in separate GUI interaction? If the coach asks for consent inside the application, this is tricky as the discussion might become pushy, in reality making the user non-free to say "no" - so this is probably not a good idea...
Should any sleeper issues be prioritized?	Overdoing the consent questions might get intrusive in itself. ("Intrusive" dialog could also be by choice and the result of a preferred coaching style and strategy.) **Personalization**: Should the user be aware of, and consenting to, certain kinds of personalization (like "we think this is your preferred coaching strategy")? Or should the user be able to modify some of this? Is personalization also "which coaches are in your council" in the first place? Should the system proactively be able to add new coaches? Who/what decides which coaches are present? **Privacy by design / access to AWARE data:** Instead of giving away data to a 3rd party (with consent to use it only for X) we can turn it around, say: The 3rd party gives us algorithms to turn data into X, we run it on the AWARE server and only give X back to the 3rd party. **Issue #9:** Illegitimate exploitation in relation to integration meeting / do we do open source? **Issue #12 Liability**: Precede each dialog move by a disclaimer (at least do not phrase things too strongly when presenting advice). This also relates to the overreliance thing.

RRI-integration going forward:

Table 4: RRI-integration going forward: Privacy and informed consent.

The Issue	Who solves the issue
Issue #1: Privacy and informed consent	On whose responsibility it is to solve issue #1, there was no direct answer given at the workshop. We therefore decided to sum up with the delegation of tasks made in D2.1: *LEGO system of consent (proactive approach) (lead T2.4: working withT1.1, T3.3, T7.1, T8.3)*

4.3 Later developments of the issue and management work

At the time of our **RRI Workshop 4,** which was held at the University of Twente during the consortium's fourth Technical Integration Week on October 11th, 2018, the Council of Coaches project was well under way. The platform was technically up and running, a Technical Demonstrator had been developed and we had launched our first Functional Demonstrator too. The virtual coaches-characters were coming into life and personality. And some ethical, social and legal concerns had shifted around and changed – some disappearing, others growing into an even harder nut to crack.

After two intermediary RRI workshops held at the University of Twente in April and the University of Dundee in June, this time, we tried out a slightly different format for stirring reflection.

The aim of this workshop, namely, was twofold. First, to foster self-reflection on ethical, social and legal aspects of actual project work and potentially contribute to change as usual. Second, to generate concrete input for the upcoming stakeholder workshop in February 2019 (T2.2). We wanted to facilitate a process where the researchers in the consortium would themselves identify the most relevant RRI issues to bring up for discussion at the stakeholder workshop and, in addition, identify the most relevant stakeholder to invite for the stakeholder workshop.

Reviewing our materials from the StiRRIng process until Workshop 4 had made us realize that the discussions in the previous RRI workshops had tended towards the legal and functional aspects of the different RRI issues, and less on societal and personal aspects. Thus, we wanted to broaden the debate about the RRI issues and force the participants to view the issues from different perspectives. To do this, we identified three general aspects that form part in all three RRI issues that we planned to address at the upcoming stakeholder workshop: 'Trust – not too little, not too much', 'Privacy and informed consent in a GDPR context' and 'Handling knowledge: conflicting knowledge sources in design and user interfaces'[2]. The three aspects we identified were: 'societal', 'personal' and 'functional'. These specific aspects were chosen because all the three overall RRI issues give rise to additional issues within the three aspects. At the workshop, these three aspects were used as thinking tools to help the participant approach each of the three RRI issues from different perspectives. The workshop was kicked off by a small presentation summarizing the RRI process so far and explaining the aim of the workshop. After the presentation, the participants were divided into three groups and guided through the following exercises:

1. In the first exercise, each member of the group had to identify as many relevant RRI issues as possible within all three aspects ('societal', 'personal' and 'functional') of one of the three overall RRI issues ('Trust – not too little, not too much', 'Privacy and informed consent in a GDPR context' and 'Handling knowledge: conflicting knowledge sources in design and user interfaces'). The identified issues were written down on sheets of paper. After the exercise, the sheets were rotated so that each group now had sheets with issues identified by another group within another overall RRI issues.
2. In the second exercise, each group now had to discuss the issues identified by another group and add additional issues if they could think of any missing issues.
3. The sheets were rotated a second time.
4. In the third exercise, the participants now had to identify and write down stakeholders they would like to invite into a discussion about the RRI issue they were working with.
5. In the fourth exercise, each participant, based on the discussions during the workshop, had to make a top 3 of relevant issues for the stakeholder workshop to cover within either 'Trust – not too little, not too much', 'Privacy and informed consent in a GDPR context' or 'Handling

[2] We had fabricated this issue by collapsing Issue #3: Handling disagreement between coaches and Issue #4: How to keep healthcare knowledge up to date? Which made good sense for our stakeholder discussion-framing.

knowledge: conflicting knowledge sources in design and user interfaces'. The workshop ended with a general discussion.

These were the outcomes from the work on" Privacy and informed consent in a GDPR context" during Workshop 4:

Outcome of the first and second exercise:

Table 5: Outputs from Workshop 4: Privacy and informed consent in a GDPR context.

Functional issues	Societal issues	Personal issues
▪ How often should the user be asked to consent to storage of personal data? ▪ Should location history be shared? And with whom? Should the users have an option to option it out? – If yes: what does the option out pattern tell about you? ▪ How do you protect the data with an open agent platform? ▪ How does the system deal with users that want to retract their data? ▪ Does a user understand what data is, and how can we explained it in a way that is easy to understand? ▪ Should we ensure full access to view the database or access to view a more readable data base? ▪ For how long can a system store data? ▪ Can we actually deal with users saying "no" to some data and "yes" to other? ▪ Should we group data in 3 sensible "trust groups" for privacy?	▪ Could health insurance become more expansive if you do not comply with the coach's advice? – If yes: this might be an issue ▪ Should it be possible for the user to share "goal achievement" with friends/their support group? - gamification ▪ What kind of data from a user should be shared with a doctor?	▪ If the user forgets to log out/turn off the system, should the visual feed then turn off automatically? And if yes: When? – We might let the user decide. ▪ How should we deal with users changing their mind about what data they want to share and have stored? – We might overestimate the rationality of users. ▪ How can we share other's data anonymously and should this even be possible? ▪ Should a user be allowed to view data from other users as a reference?

Outcome of the third exercise:

Within 'Privacy and informed consent in a GDPR context' the following stakeholders were identified:

- Experts in elderly peoples' understanding and use of technology
- Experts in data protection and data analysts
  - To clarify the legislative framework
- Health professionals
  - What is their practice regarding informed consent?
- End users:
  - To ask what they think is important regarding privacy
- Insurance companies:
  - Future scenario: Will the expenses of your insurance company raise if you don't comply with the advice of COUCH?
- Companies that store data

Outcome of the fourth exercise:

Several participants raised questions about how to deal with users that don't want to give full consent (only consent to share specific kind of data) and users that want to revoke data?

- How do we build a system that can deal with users consenting to some kinds of data and not to other kinds of data?
- Can the system provide reliable advices with only a partial picture?
- What if a user only wants to give consent to share data with specific people – e.g. physicians?
- Can we deal with users saying "no" to some data and "yes" to other? – can the system provide reliable advice with only a partial picture?

Another issue (that was also raised under 'Trust – not too much, not too little') was: Who is legally responsible?

- Who are legal responsible if the provided advices are not compliant with the newest knowledge within the field?
- Who is legally responsible if the system crashes?
- How can we make sure that the user accepts that the system is fallible? It can crash, make wrong decisions (which can happen with doctors too)

In addition, some of the participant identified issues regarding how to communicate the consent form:

- How do we make sure that the user understands what her/he provides consent for?

Besides, other relevant issues were identified:

- How do you protect data within an open agent platform?
- Should it be possible for the system to enhance the learning by stealing data from another patient?

People might use the system differently – some might use it regularly others might use it occasionally or like a game. This might have consequences for the big data.

After this description of the perspectives explored and the reflection carried out, the next section sums up the solutions pursued by the consortium to manage, solve or otherwise handle the issue.

4.4 Solutions pursued

The issue of Privacy and informed consent has in many ways been the most remarkable issue we have worked with in the StiRRIng process, in terms of extensiveness as well as urgency. In May 2018 GDPR was implemented throughout Europe, which made reflections and actions within ethical, social, and legal implications of research not only a timely, but also a mandatory concern. And the consortium as a whole also underwent a steep learning curve in relation to the implications of this issue in terms of patenting law and legislation and very different proceedings depending on whether a device is labelled as medical or rather as a life-style application. Some of the most prominent turns of this RRI issue within the scope of the Council of Coaches project – and concluding reflections on our handling of the issue by consortium members is outlined below.

Medical device or life-style application?

In the late autumn of 2018, Council of Coaches hosted a workshop in Brussels on: "Innovation uptake in eHealth with patient-centeredness and gamification. Regulatory challenges and opportunities ". The speakers covered a wide field, from the medical industry, regulation, policy making, research, academia, and SMEs. The scope of the workshop was to brainstorm on the status of the regulatory and ethical scenario in the context of eHealth, exposing the state of the art of several EU Horizon 2020 eHealth projects. During the talks and not least the discussion afterwards, patenting was raised as a theme and a long debate on the differences in legal demands and procedures applying to medical devices and life-style devices, respectively. In our second stakeholder workshop in Brussels, February 2019 (T2.2), the theme came up again. Here, it was stated that relevant legal frameworks go beyond GDPR. Other

relevant frameworks include the European medical devices directive and good medical practice standards. One participant recommended outright: talk to a lawyer. These issues are not going to be easy to map or settle, and they are going to be crucial for any attempt at going to market. Simple measures – such as defining the COUCH product as 'not a medical device' or providing a disclaimer saying that 'COUCH assumes no responsibility for what you do with this information' – is bound to be legally problematic at least in some cases. So, between the COUCH development process and the implementation of COUCH in any real context there still remains a legal hurdle to cross.

An easy-to-understand version of informed consent for user-friendly acceptance on the go using the Council of Coaches?

At some point during the scripting process for the Functional Demonstrators, an idea surfaced that it would be enormously handy and very user-friendly – not least considering our target group consisting diabetes patients, patients with chronic pain and elderly people – to have the opportunity for a short and easy-to-understand version of the full GDPR. When the user is using the actual application, it breaks with the illusion and universe of the interaction that every time the system wants to store new information, the user must consent to this in an external layer of interaction, which is moreover extremely voluminous and very hard to read. We discussed how it would be neat to actually script the actual meaning of the different types and levels of consent, as in 'what you are saying yes to', into a lingo that reflected the Council of Coaches Universe, which is (at this point) light, humorous and cartoony. However – at the stakeholder workshop in Brussels February 2019, when we asked the participants about their opinions of this, it soon became clear that this was not a feasible way to move forward. Any kind of abbreviation of the GDPR will deem this version not legally binding.

When asked about where this, in terms of solutions to the issue, leaves us, a member of the consortium summed it up like this:

> "We were very careful with this issue from the outset and we though 'this is very important and we want to treat it right – and then we had that workshop in Brussels and it turned out that it was so much more complicated than we thought". We did everything right when we did all the patient testing in terms of our approach to informed consent there – but we have not 'solved' the issue in the Technical and Functional demonstrators and that's OK. We focused our efforts on other things and there are other projects out there working on solving these things. We have identified the problems and we are keeping our eyes on them and once a proper solution comes up – that is legal and easy to understand – then we will be the first project to implement, because it is not that difficult to implement. What you are going to be saying, how you are going to be saying it and what they are consenting to – that is the difficult thing."

For the Council of Coaches, the issue is part of the RRI package, and it is important in and off itself, but at the same time, it is not mainly what the project is about.

Still, while not a legally binding construction and not a route we pursued further, the necessity to keep the user in the driver's seat, also in terms of give out information had found its way into the structure in a slightly different manner:

> *"Some elements of this discussion and some elements of the idea of informed consent on details has found its way into the dialogue design where in many cases we have inbuild the option to state 'I don't want to say that! I don't want to share that. I don't want to answer those questions now. I don't want to answer those questions ever. At various points in the dialogue structure this comes back. And it is fair to say that this is influenced by the RRI discussions."* (Consortium member)

The LEGO approach

At the initial RRI Vision Workshop, the consortium's approach to privacy was to regard it as "a LEGO system of informed consent". This approach was in all essentials dismissed as the project unfolded. Instead, our understanding of privacy and informed consent within COUCH went from a technical issue in need of a technical solution to something much more all-encompassing:

> *"Another development that has been very obvious is how the constant focus and continued debate on privacy issues in the project has fostered a genuine awareness among the researchers in the consortium, so that the issue has most definitely become a shared responsibility throughout the consortium. The LEGO approach was very much a technical approach, but what we see now is that everyone is much more aware and has been incorporating this in their research."* (Consortium member)

Another reason for the LEGO approach to disqualify along the way has to do with the research context the project is embedded in:

> *"The LEGO approach is a really good approach if you are building a product, but since we are doing a project with a standard evaluation it simply is not relevant – because to do the evaluation, we need the consent of the user to use all of their data for research purposes – otherwise we cannot do the research. So, it is just less relevant in the current phase of product development. Later, when the product is launched, it becomes more relevant."* (Consortium member)

In the first phase of the project we had a distinct focus on 'where' in the project the RRI issues were located. Throughout the StiRRIng process, however, this has proved to be of very little importance. Rather, the RRI issue of privacy and informed consent has permeated the work carried out through and through:

> *"If you want to put it harshly: Privacy issues always complicate technical matters. Therefore, people try to take shortcuts and not implement it correctly. But what we have achieved here is that we have had a proper conversation where people thought about it and then saw all the intricate details and thought OK, better wait a little bit and do it correctly, than start implementing it wrongly, because building a database where you have the LEGO structure can be quite complicated – and that was actually something we discussed in the very first session we had (RRI Vision Workshop): what should be in the first LEGO brick, and what should be in the next? And how is the interaction going to be? Now all the researchers have been educated on this topic and they do not see it as a burden anymore, but they see it as a necessity. And therefore, they will think about it, but they will not take the short route. That might not be the optimal solution for a product, but that is also something that goes into the product development, where the product owner should take responsibility for implementing these things."* (Consortium member)

5 Balancing trust-building measures against risks of over-reliance

5.1 Description of the issue and its potential implications

During the RRI Vision Workshop, the issue **Trust (not too little, not too much)** was envisioned in the following manner:

Table 6: Outset: Issue #2: Trust (not too little, not too much).

Issue #2: Trust (not too little, not too much)	
What is the issue?	On the one hand, if users don't trust the advice given by the virtual coaches, they won't use it. On the other hand, if the users trust the advice too much, it might impede health or create social addiction to using the Council. The issue is thus one of handling trust-building in a reflected manner, where the design choices in the project and the exploitation choices after the project should all support a balanced approach to the relationship between the Council and its users.
Why is it important?	This issue is important because achieving a proper balance between trust and independence between the user and the Council will help to ensure that the user can get the most out of using the Council and avoid being subjected to the potential negative consequences of misinformation or overreliance on the Coaches. This issue is thus one that is closely related to the professional ethics of human coaches, where building trust while supporting the autonomy of the client is a crucial issue.
How do we address it?	The group working on this issue suggests a number of interrelated actions to handle this issue, some of which overlap to a certain extent: **Establishing trust** ▪ Quality assurance of information provided by the coaches ▪ Ensuring a good match between relevant characteristics of the user and the advice given to the user ▪ Establishing rapport with the user, e.g. through small talk where appropriate ▪ Enlisting advice from external actors when necessary ▪ Full transparency about what information will be shared – both at the point of registration and when data sharing occurs. This should apply to data sharing between coaches ("I'm going to tell Sue about your needle phobia, because…") as well as between the system and external actors ("I'm going to tell a professional about your suicidal thoughts, because…") **Avoiding overreliance** ▪ The Council should remind users to visit relevant health-care professionals and to prioritise their advice over the information provided by the Council ▪ While stakeholders suggested to ensure a physical likeness between the coaches and the users' real-world care professionals, going the opposite road of avoiding such likeness would help as a reminder to the user that this is 'just' a program ▪ Ensure that users have the Council of Coaches system explained to them in a manner that they understand and that they are warned about its limitations

	▪ In general, avoid too great a degree of realism in the appearance.
Who does what?	These issues cut across all the technical work packages and into implementation as well.
Who takes the lead?	**Building trust** ▪ WP2,3,4,5 Quality assurance of knowledge (lead T5.4, working with T4.4) ▪ WP3-6 Match between user characteristics and advice given (lead not yet identified) ▪ WP3,5 6 Establishing rapport, e.g. through small talk where necessary (lead not yet identified) ▪ Getting recommendations from doctors (how, who?) ▪ WP8 Virtual Agent Explaining the system (or other means of giving a clear explanation) (lead T8.1, working with T3.2) ▪ WP3, 5 Transparency of what information will be shared with e.g. GPs or other actors and with other coaches (internal sharing (explicit): 3.1) (internal sharing (implicit): in connection w/ informed consent) (external sharing, technical: WP7) (external sharing, contractual: 8.2) (to be taken up in T2.4 as well) **Avoiding overreliance** ▪ Quality assurance (see big paper) (lead 5.4, working with 4.4)) ▪ WP3, 5, 6 Remind users to visit human experts (modesty) (lead T3.1) ▪ WP6 Avoid likenesses with real doctors, too great realism in looks, etc. (lead T6.1, T6.2) ▪ WP8 Initial explanation of how the system works (lead T8.1, working with 3.2)

5.2 Early developments of the issue and management work

At the RRI Workshop 1 in Valencia, March 2018, the following developments were debated:

RRI issue 2: Trust (not too little, not too much)

The issue

- *The Council of Coaches must be trusted to have any effect*
- *But too much trust can create negative effects such as overreliance*

Proposed solution(s)

- *Various means of establishing trust*
- *Various means of avoiding overreliance*

Table 7: Early developments: Trust.

ISSUE #2	TRUST (NOT TOO LITTLE, NOT TOO MUCH)
How does the issue relate to the themes of the hackathon?	In fact, this issue is not that closely related to current talks in the hackathon. However, in time it certainly will be, so it would be reasonable to include something in our current scenarios.
How relevant is the proposed solution to the tasks of the hackathon? (What needs to be changed or added?)	Most of them are relevant. Here some suggestions that could be used in the hackathon scenarios: **Sensing**: The users need to be informed about the functionality and purpose of sensors and why we need these sensors. We must also let them know that these will not be used for commercial ends. **Dialogue, embodiment of characters:** There is a huge debate to be taken concerning the appearance of the characters, which probably should not be made to look too realistic. Maybe their looks and way of speaking should be aligned with the demographic they will serve, for example in the way they talk to you about your week etc. (the scripting of social talk.) **Duration of relationship**. We need to consider whether it should be possible to continue the interaction with the Council after one's problem is solved. (In cases where the original incentive to engage was based in a defined, solvable problem. Since in real-life coaching this would be the time to discontinue interaction with the coach)
What is your proposal for taking this issue into account in relation to the system integration? (Who should do what?)	Once we have a basic working system, we should start building in the proposed solutions above in the overall system architecture. Also, instead of giving away data to putative future 3rd party providers of coaches for their algorithms, we could ask for their algorithm, feed it our data and only give them the outcome (which most likely is what they say they need). Different decisions on this point will have very different outcomes/consequences.
Should any sleeper issues be prioritized?	**Issue #8: Privacy by design,** **Issue #9: Data security,** **Issue #11: Equality,** Should all be prioritized.

RRI-integration going forward:

Table 8: RRI-integration going forward: Trust.

The Issue	Who solves the issue
Issue #2: Trust (Not too little, not too much)	Issues #2 is an omnipresent matter, and must therefore be reflected upon throughout the whole project. D2.1 came up with following delegation: ***Building trust*** ▪ *WP2,3,4,5 Quality assurance of knowledge (lead T5.4, working with T4.4)* ▪ *WP3-6 Match between user characteristics and advice given (lead not yet identified)* ▪ *WP3,5 6 Establishing rapport, e.g. through small talk where necessary (lead not yet identified)* ▪ *Getting recommendations from doctors (how, who?)* ▪ *WP8 Virtual agent Explaining the system (or other means of giving a clear explanation) (lead T8.1, working with T3.2)* ▪ *WP3, 5 Transparency of what information will be shared with e.g. GPs or other actors and with other coaches (internal sharing (explicit): 3.1) (internal sharing (implicit): in connection w/ informed consent) (external sharing, technical: WP7) (external sharing, contractual: 8.2) (to be taken up in T2.4 as well)* **Avoiding overreliance** ▪ *Quality assurance (see big paper) (lead 5.4, working with 4.4))* ▪ *WP3, 5, 6 Remind users to visit human experts (modesty) (lead T3.1)* ▪ *WP6 Avoid likenesses with real doctors, too great realism in looks, etc. (lead T6.1, T6.2)* ▪ *WP8 Initial explanation of how the system works (lead T8.1, working with 3.2)* Once we have a basic working system, we should start building in the proposed solutions above in the overall system architecture (WP7)

5.3 Later developments of the issue and management work

At the RRI Workshop 4 in Twente in October, the theme of "Trust – Not too much not too little" yielded these outcomes:

Outcome of the first and second exercise:

Table 9: Outputs from Workshop 4: Trust (not too little, not too much).

Functional issues	Societal issues	Personal issues
▪ How do you build trust if it's different for every person? ▪ What kind of trust do we want? ▪ How do we know that someone trust COUCH (or doesn't)? ▪ Should we show users why they should trust the virtual coaches? If yes: How? ▪ What if your app crashes or give fallible advices? – will this affect the trust of the user? ▪ System perception – measurement, instrument, social entity? ▪ The amount of trust influences the acceptance of advices – thus it is important to find a balance in how much the user trusts the system	▪ What if a virtual coach's advice disagrees with your doctor's advice? ▪ Should the user be allowed to use the system without professional guidance? ▪ Who is responsible if the system gives fallible advices? – Who owns the advice? ▪ How do we ensure that the user does not use COUCH like a game? ▪ When should the app send warning to a physician/real actor?	▪ Potential question from a future user: Can I trust the system better than the doctor? ▪ How do we avoid over-reliance? ▪ How do you convince someone who does not trust the system? – should you even try? ▪ Should the system adjust to the value of the end user to create trust? ▪ Do you trust the system if you do not know its training goals? – should the training goals be apparent? ▪ I will trust the system more if my doctor recommends it? ▪ Trust/distrust in the system might influence the trust in doctors – could this be an issue?

Outcome of the third exercise:

In the third exercise the participant had to identify relevant stakeholders within the three main issues: 'Trust – not too little, not too much', 'Privacy and informed consent in a GDPR context' and 'Handling knowledge: conflicting knowledge sources in design and user interfaces', based in the issues identified in exercise one and two.

Within 'Trust – not too little, not too much' the following stakeholders were identified:

- Potential users
- Health professional
- Health care app developers
- Philosophers – with expertise in how human trust in machines

Outcome of the fourth exercise:

The following issues were identified within the scope of 'Trust – not too little, not too much' in the final exercise:

Several participants raised question about how to make sure that the system does not become manipulation when gaining trust from the user. This concern related to the core of the trust RRI issues: the balance between too much trust and too little.

- How much trust is too much trust? Where do we draw the line?
- When does trust building become manipulative?
- How can we build trust without being manipulative?
- How do we develop trust in the system in a positive and not manipulative way?

Another general issue relates to the relationship between the virtual coach and real health professional.

- How do we make the system more trustworthy - and how might health professionals help to do this?
- Does health professional trust a system like COUCH?
- How do we ensure that the system does not reflect negatively on health professionals?
- In what way do health professionals want to use the system (if they do)?
- Should you be able to use the system without professional guidance?

In addition, a lot of the participants raised the question: Who is responsible if something goes wrong.

- Who is responsible if a virtual coach gives an incorrect advice? Who owns the advice?
- How can we make sure that the users accept that the system is fallible?

In addition, a single participant raised the question: How does the system even detect and model trust? How does it know that the user trust it?

5.4 Solutions pursued

In the Council of Coaches, our conceptualisation of the issue of trust has very much been a question of finding and nurturing the right amount and the right kind of trust. The users' baseline level of trust in the project has been extremely high, which is probably very much correlated to the history and identity of the host-research institution, RRD. The RRD has substantial experience with giving eHealth applications to end users, elderly users, and users with chronic conditions. Their experience is that when people are asked to participate in research as part of a science project, testing prototypes developed by scientists (even if these are not the white coat-wearing kind), a lot of authority is at play. So, when given an application – no matter whether it is a serious application developed in a serious project with lots of effort invested or a minor, less prepared test-style thing – the respondents tend to take the advice and the feedback given in the applications very serious. Therefore, a significant part of the work related to the issue of trust has revolved around lowering the level of trust that the project is born with due to its academic source.

Too much trust: trying to build trust-lowering mechanisms into the system

In designing and building the Council of Coaches, much effort has been invested in toning down the level of trust. One of the main purposes of the project has been to enable an application that will invite and encourage the users to consider different opinions and different sides of a story – and then make their own decision as to what they are going to do with their health. If one coach says you should stop eating and the other that you should start moving – then maybe you decide to start doing a little bit of both. We want the users to consider different types of advice, with a healthy, critical approach, which is why measures have been taken to counteract that too much authority is invested in the virtual coaches on behalf of the user. Two of the main trust-lowering mechanisms that for these reasons have been built into the system are the cartoony look of the coaches in the functional demonstrator and the amount of humour in the conversational style and exchanges. The humour is intended to create a light atmosphere between the user and the coaches, while their appearance has the purpose of installing a reflective distance, which also serve to lower the coaches' authority as they, after all, are 'just' cartoony-looking characters, dressed in casual clothes in a living room – not life-like personas dressed in white coats in a virtual doctor's office or clinic. Still, the aim is not 'just' to create critical space and distance, we still want the users to build relationships with the coaches over time, but we want the character of the relation to be 'just right'.

Trust 'at first glance' vs. building trust over time. Framing the relationship between the user and the system

Making up the back stories and the characters is a way to decide the level of trust. For example, we have chosen not to add a doctor to avoid a too authoritative path, and to avoid conflicts of interest with the user's real-life doctor. On the other hand, people who have tested the system have told us that they would like a peer coach, someone like themselves, in order to establish more trust. You may say that what we are trying to do is on the one hand to reduce the initial trust stemming unmediated from the source of the application, while on the other building up trust tied to actual use of the system. We have aimed for some kind of middle ground, reachable from both sides, and that is where we have ended up.

There is something like a trust at first glance which, as mentioned, in our case is very high because of the source of the application, the persons and organisations that provides it. We have taken a lot of steps to reduce this by way of things that are immediately apparent. The characters are casually dressed people in a living room, the whole thing being a bit cartoony. Right away in the conversation it is a bit funny and it is a bit light. But what if you do not know anything about the source or is neutral about the source? Trust also develops due to the things happening when you begin to use an application, you decide whether it is good for you or not good for you, and you decide whether to carry on or discontinue use. We distinguish between this trust at first glance the building of trust over time through use. So what we have done in the project has been to try to downplay trust at first glance, but build up trust in using the application over time, for example in building the relationship between users and coaches and just in general giving advice and feedback that are sound and useful and makes sense.

Too little trust – not so much an empirical problem

The opposite of too much trust is of course too little. If the users do not trust the system, they will stop using it. However, with the Council of Coaches' anchor in a renowned research institution and because of the development in a 'proof of concept' way, too little trust has not been an issue at all. At least in this phase of the Council of Coaches. However, since we have been taking all these steps to lower the amount of trust, we might actually face a problem when the project-part is over, and we want to put the product on the market. When by then the source or party that provides the application is no longer a well-known research institution, giving you the opportunity to contribute to science and knowledge and asking you to help to evaluate a product, but instead a company that wants to earn money, the situation concerning trust may be very different. By then the problem might switch around on itself and making sure to facilitate the right level of trust may become much more important. So, it may be the case that the steps we have taken in this phase of the project have in fact been harmful for the product development. However, we are aware of this possibility and have many discussions on the subject to enable reframing, should it become relevant.

Cultures of trust

We want to make sure that people use our system's advice as an outset to make their own decisions – not as information they need to follow strictly – and we want to enable and empower the user to make sound decisions about their health. Certainly, there is no 'one truth' about health coaching that we try to capture in a technical way, only a vast amount of quantitative data to explore and elicit knowledge from. Also, there exists no set agenda that will work for everybody, rather everybody has their own things that will and will not work for them. For these reasons, the system needs to be adaptable in its coaching and interaction with the individual user in order to establish results By triggering critical thinking, we aim to come up with a system that is best for 'you', something you can keep up for a longer time and thus more effective, hopefully.

The overall aim of the application was to stir the users' reflection about lifestyle choices to support their own health and empower them to make changes and form new healthy habits. We wanted to create something where the user is in the driver's seat and nudge them to take responsibility for the own health. Based on user feedback and broad evaluations there are strong indications that this goal has been obtained. However, how one responds to this kind of distributed authority on the area of health is also a cultural question. The Alma Mater of the Council of Coaches is the RRD, based in the Netherlands, and the user testing of the functional demonstrator has mainly been carried out in the Netherlands, Scotland and to a certain extent Denmark, which are all North-Western European nations allegedly with a strong cultural tradition for perceived equality between experts and clients and critical questioning of authority, also in fields that are traditionally characterised by the superiority of experts, such as the doctor's voice in the area of health. It is thus possible to wonder how willingly users in other parts of Europe will embrace the idea of being in the driver's seat concerning decisions related to their own health. Other cultural traditions of authority in e.g. medical questions may be perhaps less dialogic and rather tend towards bestowing the doctor full authority and expecting a final, restrictive answer rather than a proposal or an open debate, designating the patient a role of adhering to advice rather than questioning or discussing it. At the end of the day, the take-home massage is that different cultures might require different approaches to health coaching, whether virtual or the traditional 'old-fashioned' kind.

6 Multiple knowledge sources and potential conflicts between them

6.1 Description of the issue and its potential implications

During the RRI Vision Workshop, the issue **Handling disagreement between coaches** was envisioned in the following manner (see Table 10):

Table 10: Outset: Issue #3: Handling disagreement between coaches.

Issue #3: Handling disagreement between coaches	
What is the issue?	An interdisciplinary Council of Coaches will necessarily produce conflicting advice at some point. This demands a solution to solving conflicts arising between lines of argument modelled on different professions. For example, a patient with a cardiovascular disease and asthma profile is recommended to drink coffee by one coach but another couch may recommend the opposite.
Why is it important?	This issue is important because a user seeking medical advice from the coaches may either become confused by the conflicting advice coming from coaches drawing on different sources of information or may choose to make use of one piece of advice based on non-medical factors, e.g. how much the user likes one coach or the other. Ultimately, this may lead to misinformed and potentially dangerous health choices.
How do we address it?	Conflicts between the argumentation patterns of the different coaches can be identified in two ways: 1. We imagine that there would be something like COUCH "knowledge packages" for each of the coaches. Whenever we are about to update the system and add a new "package", we can define/identify those conflicts and 2. During execution, i.e. dialogue with the patient, it will be possible to identify conflict that were not apparent during the updating process. In both cases and based on the system definition, there will be agreed on a strategy, i.e. talk to a human expert to settle conflicts. Overall, it will be necessary to establish a hierarchy of conflicts and a modus operandi for solving them, including protocols for documentation of conflicts and probable solutions (guidelines, papers, etc.).
Who does what?	The issue of handling conflicts between the knowledge bases and argument patterns must ultimately be handled through organisational protocols to be followed by actors running implementations of the Council of Coaches.
Who takes the lead?	Developing a proper organisational response for handling these conflicts seems to fall in the vicinity of task 8.2 (exploitation and business planning) where criteria for implementation can be set out. However, exactly how to solve this issue remains unclear at the end of the RRI workshop and the STIRRING follow-up (task 2.4) will be charged with following up on this issue.

6.2 Early developments of the issue and management work

At the RRI Workshop 1 in Valencia, March 2018, the following developments were debated:

RRI issue 3: Handling disagreement between coaches

The issue

- *The Council of Coaches will necessarily produce conflicting advices at some point (e.g. a coach might advise for a daily jog, while another might say it's bad for the knees)*
- *This might lead to misinformation and dangerous health choices*

Proposed solution(s)

- *With a knowledge-package for each of the coaches, it should be possible to identify conflicts when the system is to be updated with a new package*
- *During dialogue with patient, other issues that were not apparent during the update can be identified*
- *Hierarchy of conflicts and modus operandi for solving them*

Table 11: Early developments: Handling disagreement between coaches.

Issue #3	**Handling disagreement between coaches**
How does your issue relate to the themes of the hackathon?	This issue will not be solved at the hackathon but is ultimately something that needs to be taken care of and be on point in the final system. That being said the hackathon also works with scenarios that shows off some of the strengths in the system and can be illustrated at an early stage in the project. It would be good to integrate this issue and the proposed solutions to it, in the scenarios. The hackathon group proposed to illustrate following: ▪ Medical conflicts will appear and can be handled by argumentation framework. ▪ How interaction can be designed around medical conflicts
How relevant is the proposed solution for the issue to the tasks of the hackathon? (What needs changed or added?)	Identify where issue exists on architecture diagram – show where we need to account for the preferences in different parts of the architecture. Feed them into a strategy for resolving the conflict. Give the different medical advices to the user from different perspectives, and let him decide which one to follow.
What is the group's proposal for taking this issue into account?	We need to figure out what the medical principles are, that can then be used to generate preferences in the system, i.e. is there a medical principle that says; this concrete medical advice should be prioritized over a general health advice: Arrange dedicated discussion on principles (and preferences) in the consortium - what could they be; produce rough list. Take the list to stakeholder workshops - show a scripted demonstrator with extreme examples of conflicting principles. The stakeholder-workshop should help dealing with what sort of principles would solve the specific conflicts. With the technical demonstrator (completed in month9), we can in month 10 come up with a more concrete scenarios based on the stakeholder

	feedback. = A proper, fully functional, technical demonstrator of extreme example on how to deal with conflicting advices. Use feedback on the demonstrator to feed into next prototype – constant feedback-loop.
Who solves the issue and when?	In the original iteration of this issue (in D2.1) it was stated that this issue is ultimately going to be handled in the implementation of real-world applications and that D8.2 of the project would need to prepare the way for this, e.g. through development of organizational protocols (or templates for such protocols) for use by downstream actors. Nevertheless, the reiteration made here shows that there are important aspects of this issues that needs to be handled in the develop phase, specifically in the development of the knowledge base and argument structure (WP3). In this connection the tasks to be handled are: Develop a hierarchy of conflicts. Work out a modus operandi (a set of principles) for solving them; e.g. "medical knowledge trumps social knowledge", "fitness knowledge vs social knowledge must be solved by user" etc. Note: It is these principles that downstream actors implementing the Council of Coaches will have to revisit before implementation, since they have significant impact on the outputs from the system. The decision of which principles should be implemented in the prototypes (by WP3) should be made by the consortium as a whole and should be facilitated by the RRI group (WP2).
Should any sleeper issues or stakeholder concerns be prioritized?	The issue of Personalization came up: who decides how the coaches are personalized and on what background: We could by default prioritize concrete medical advices over all other things, but what if the specific user is into e.g. herbal medicine or remedies. If we only advocate normal medicine, we can eliminate those users. This sets up an example on a user type, that would disapprove certain principles in the system Transparency: Should the principles be apparent to the user – then the user will be able to manipulate the settings.

RRI-integration going forward:

Table 12: RRI-integration going forward: Handling disagreement between coaches.

The Issue	Who solves the issue
Issue #3: Handling disagreement between coaches	The issue should be dealt with in the development of the knowledge base and argument structure (WP3). The decision of which principles should be implemented in the prototypes (by WP3) should be made by the consortium as a whole and should be facilitated by the RRI group (WP2). D8.2 should bear the responsibility of ultimately preparing this issue to be handled in the implementation of real-world applications e.g. through

	development of organizational protocols (or templates for such protocols) for use by downstream actors.

6.3 Later developments of the issue and management work

At the RRI Workshop 4 in Twente in October, the theme of "Handling knowledge: conflicting knowledge sources in design and user interfaces" yielded these outcomes:

Outcome of the first and second exercise:

Table 13: Outputs from Workshop 4: Handling disagreement between coaches.

Functional issues	Societal issues	Personal issues
▪ Should we: 1) prevent conflicting knowledge? 2) Highlight conflicting knowledge? 3) Do nothing and let the user decide? ▪ How do we deal with people that want to revoke data if that data is already used in a machine learning model? ▪ Conflicting knowledge detection: Auto-detected with reasoning or manually analysis on advance? ▪ How do we ensure that users know they can trust the system?	▪ Who is legally responsible if advices in the system cause injury/death? ▪ Could a health professional be held accountable if they refer a patient to COUCH and the system gives bad advices? ▪ What will specialists think if a conflict is discovered in COUCH regarding their field? ▪ What will the general public think if a conflict is discovered in COUCH? ▪ How will the system probably scare the public? Nowadays, public are often scare of computers and computers are often blamed for everything. ▪ How can data sourced from jurisdiction with weaker data protection laws (e.g. much of Asia and Africa) be safely used by a European based organization? - An option could be to develop the system to comply the strictest law.	▪ What if a user doesn't give consent / only give limited consent? ▪ How transparent should we be regarding conflicting knowledge? Should users e.g. be able to see conflicting knowledge? ▪ How should the system process conflicting knowledge from the user? – If the user one day says something and the week after says the opposite? ▪ How can we convince public that their data is safe? ▪ How can we ensure that we only ask a bare minimum of information for the users? ▪ How to respond to a plausible accusation that the system unduly affects the user behaviour?

Outcome of the third exercise:

Within 'Handling knowledge: conflicting knowledge sources in design and user interfaces' the following stakeholders was identified:

- Multidisciplinary medical companies
- Someone who knows how to incorporate new medical knowledge into existing medical knowledge
- Someone who writes guidelines for health professional
- Someone who knows about cross-country advise/guideline resolution

Outcome of the fourth exercise:

Only a single participant chose to make a top 3 within this issue. The top 3 is shown below.

- When the coaches provide conflicting information in a discussion can you then make sure that the user adopts the correct information?
- How to make sure that the information is up to date?
- How to be safe, but useful of added value?

6.4 Solutions pursued

As the project progressed, it was very clear to the RRI team that we talked much less of this issue than we did of privacy and trust. However, this does not mean that the project has changes the intention of enabling situations where the user can choose between the coaches' perspectives and pic advice from one coach while another says something different. Rather, this scenario is closely tied to a future for the application where artificial intelligence in much more integrated than is the case at this point:

> "While building the demonstrators… this is still done manually, so we make sure that the content we put in isn't in conflict with contents available for other coaches or we make adjustments to make sure that everything works out. To be able to do this automatically… checking if content is conflicting, at this point would present some technical difficulties still. In the dialogues we do something like offering two different views on the same topic. That is something we do in some cases in the dialogues, so we make some remarks saying 'but also take this into account' or 'well this might not be for everybody', but we don't insert completely new coaches or completely new pieces of content without being sure that it works" *(Consortium member)*

But did we then 'solve' the problem by changing our outlook on the way it is scripted into the project?

> "Most of the scenarios we discussed in the beginning were mostly about medical information and since this is not going to be a medical device – it's going to be lifestyle device – also the advice you'll be giving might be less medically charged. So, the ethical issues might be… less" *(Consortium member)*

The original idea inscribed in the RRI issue here took into account that we might have a very extensive logic-based knowledge base-structure that could automatically come up with arguments or statements for the coaches and then you would have a higher risk for conflicts if you would put new information in. But at least for the Functional Demonstrator that risk is a lot lower since we use a lot of scripted dialogue so there is a lot more structure between statements for the coaches. For the Technical Demonstrator, the risk is a little higher but still OK, since also those dialogues have been carefully designed.

Hard and Soft Conflicts

The essential distinction to distinguish here is that of 'hard' and 'soft' conflicts.

> *A soft conflict is a situation, where different views are presented e.g. different ways to tackle a problem. This is a good thing, at least to try to think through this, so in our proof of concept study in the Functional Demonstrator we deliberately make content where those soft conflicts arise, so we can see if it is perceived to be a nice thing by users. On the other hand, yet the hard conflicts where – if you have a system that has fully automatic virtual coaches that decide by themselves what to say and you inject another coach, developed by another partner – these are all quite futuristic scenarios – then you can have a situation where one coach says A and the other coach says not-A, something that really does not match, and then you have a problem. So – to solve that technical problem is a situation that will not arise yet, since our system is not that automated yet. But in time to solve that problem will be an interesting challenge.* (Consortium member)

And yet – maybe we are in fact already on the move:

> *"In the dialogue and argumentation framework we do somehow automate what coaches say, so we have a system in place where coaches decide what to say by themselves, but it is not yet at a level where these issues arise. There is scripted dialogue and there is automated dialogue and we are somewhat in the middle, though mostly still on the scripted side so as we move towards more automation this will become a problem – and the good thing is: We are prepared for it, because we have been discussing it, we are aware of that."* (Consortium member)

7 Keeping knowledge up to date

7.1 Description of the issue and its potential implications

During the RRI Vision Workshop, the issue **How to keep healthcare knowledge up to date?** was envisioned in the following manner:

Table 14: Outset: Issue #4: Keeping knowledge up to date.

Issue #4: How to keep healthcare knowledge up to date?	
What is the issue?	Healthcare knowledge is fluid. System should give advice based on the latest medical knowledge. Adding a new coach or new "medical" insight results in a change in knowledge base. Such a change can cause unforeseen issues in the interaction between different domain knowledge bases. Bottom line: This can lead to the wrong advice.
Why is it important?	A lot of the credibility of the Council of Coaches as a potential tool for healthcare will rest on its ability to give access to reliable medical information, even if this information will be supplemented by other types of knowledge. As such, the product would need to have a reliable protocol for knowledge updates to attract investors/licensees/buyers downstream.
How do we address it?	The group that worked on this issue suggests to develop a mechanism or protocol for model-checking to be run after every change made to the knowledge-base. - Identifies potential risks that may be manually checked. In other terms, this approach to systematic updating of medical knowledge could perhaps be described as "verifiable by design". ▪ Manual checking of conflicts, what is the problem, what is the solution (e.g. by external expert) ▪ Knowledge verifiable by design (linked to quality assurance)
Who does what?	WP3 takes care of the Knowledge Design. WP5 looks at the Early design of the model checking tool. To be sure, designing a working version of such a checking tool would be an entire development project in itself, and is not part of the obligations of the consortium. The WP, however, can outline the initial requirements for the benefit of potential further development efforts after the end of the project.
Who takes the lead?	Knowledge base checking tool identifying conflicts between the "knowledge packages" of each coach. This should be closely linked to quality assurance (lead 5.4, along with WP3).

7.2 Early developments of the issue and management work

At the RRI Workshop 1 in Valencia, March 2018, the following developments were debated:

RRI issue 4: How to keep Healthcare Knowledge up to date?

The issue

- *A change in the knowledgebase when adding new medical knowledge can cause unforeseen issues in the interaction between different domain knowledge bases.*
- *This can lead to wrong advice.*

Proposed solution(s)

- *A mechanism or protocol for model-checking that can be run after every change made to the knowledge base. This can identify potential risks that may be manually checked*
- *Manual conflict-checking (what is the problem, what is the solution). This could be done by external expert*

Table 15: Early developments: Keeping knowledge up to date.

Issue #4	**How to keep healthcare knowledge up to date**
How does your issue relate to the themes of the hackathon?	The task of solving issue 4 is relevant to the hackathon as we need to create an architectural system for the interaction between the knowledgebase and user data that's intelligent enough and with enough variables to make sure that the advices given will always be the right advice. The advice shall take the newest healthcare knowledge into account, but not necessarily be based upon it alone.
How relevant is the proposed solution for the issue to the tasks of the hackathon? (What needs changed or added?)	It is relevant. The re-run solution that this group proposed as a solution, draws on the originally proposed solution in Copenhagen, but adds a mechanism that can automate a bigger part of the risk-checking system. This group also added a proposal of a "new knowledge coach".
What is the group's proposal for taking this issue into account?	The hackathon group identifies two specific ways to deal with the issue: **The first solution** is to make a system that can somehow compare the old knowledgebase and the advices given to the user based on that knowledgebase, to the new knowledgebase. That solution in steps would look like this: 1) Re-run all interactions that the system has had (this requires history of interaction). 2) Check if there is now relevant knowledge related to that specific user, which is different from the old knowledgebase. 3) Check if the knowledge differences are conflicting: a. If Solution1 (solution/advice based on old knowledge) is the same as Solution2 (solution/advice based on new knowledge) then there is no problem![3] b. If Solution1 is different from Solution2, then there might be a problem: 1. Check if it's an issue that the solutions are different. 2. Figure out how to resolve the problem → might mean that Knowledgebase needs changes. Here it is important to notice that the new solution might not be favourable to the old one. Maybe, the old solution fits better into the patients' lifestyle and is making great changes in his health-journey, while the new solution might be based on more accurate knowledge, but might have to start from scratch and will to some degree annoy the patient as s/he will have to adapt to a new health strategy.

[3] Supposedly Solution1 was checked when it was first created.

	This first solution creates a new potential issue as everything, both new and old data, is stored in the system, which might create issues with the data management policies. **The second solution** to the problem is to add a new coach that can articulate the new knowledge that is now in the knowledgebase and tell the user that he/she should be aware of that knowledge. The new coach can be very specific and tell the user, that the other coaches do not know about this knowledge. The only problem with this solution is that the other old coaches might not be able to understand advices and arguments from new knowledge-coach. That being said the new coach can still tap into the old coaches' understandings, and the old coaches also have the ability to say things like: "that sounds really interesting". Even though they don't know what is talked about. It is also possible to build the system in a way, so that a specific way of saying something from the new coach, will make the old coaches aware that what he is saying is important and that they cannot argue with it. Then they will stick to their confirming quotes. The second solution is, from a technical point of view, probably easier to implement.
Who solves the issue and when?	The hackathon group4 concluded that this issue should be dealt with (primarily by University of Dundee) in Task 5.4 and T3.3. T3.1 and T3.2, even though this issue is not one of the primary concerns, also is of relevance on this matter. A lot of this work will be done after the project when a different partner turns Council of Coaches into a continuously used application. It is not a job that should be done by the consortium. But of course, the framework should be done now.
Should any sleeper issues or stakeholder concerns be prioritized?	Relevant sleeper issues and stakeholder issues that could show up when addressing issue 4: Stakeholder issue #7: Internal ranking between the advices made by the coaches – what is the right advice to give, the new or the old one? This can be decided manually if there is a conflict detected when the old and new knowledgebase are compared. This is also done in the future and is not a job for the consortium. Sleeper issue #6 - Personalization: Group4 suggested that the new coach' personality should be designed as honest and with clear arguments, e.g.: There is new knowledge within the field of diabetes that says that running can be overdone, and should be limited to 10 km per day. The diabetes coach doesn't know about this yet". Also, Group 4 meant that the new coach should be updated, so s/he knows about the behaviour and preferences of the user, to adapt this into his coaching strategy. He should be clear about that as well: "The other coaches told me that you like to jog Bob. That's good!" Sleeper issue #12 - liability: The question of "who is responsible for damage if the coaches give wrong advices" is relevant for everybody and everywhere, but this issue will often be detected in the knowledgebase when it changes. Therefor a way of dealing with the issue of "who is responsible" is necessary before committing to a solution to issue 4.

RRI-integration going forward:

Table 16: RRI-integration going forward: Keeping knowledge up to date.

The Issue	Who solves the issue
Issue #4: How to keep healthcare knowledge up to date	Primarily by University of Dundee in T5.4 and T3.3. T3.1 and T3.2 are also of relevance on this matter, even though this issue is not one of the primary concerns here.

7.3 Later developments of the issue and management work

Note: *see 6.3, since at this point the debates of Issue #3 and Issue #4 ran in parallel.*

7.4 Solutions pursued

This issue also quieted a lot down during the first couple of years of project work. Part of the reason was that, for the Functional Demonstrator, National Guidelines were implemented. These guidelines have already been discussed by a multidisciplinary team and have been implemented into regular healthcare. This development trumps the multiple sources approach, which went into the original framing of the issue in the RRI Vision Workshop. (This is true at least until something becomes so complicated that guidelines clash.) Still, there is also another good reason that this issue has not caused a lot of stir during our stirring:

> *"When you have a system in place where coaches decide for themselves what to say, they will decide what to say based on some kind of formalized knowledge. E.g. "It's unhealthy to eat pizza!" and this statement is somehow stored as knowledge that artificial intelligence can use to say: "Stop eating pizza!" If suddenly new research shows that pizza is actually very healthy for you, we need to make sure that we update these knowledge rules, so from that point on, the coaches can use this information and say: "Eat more pizza!" But that is something you really have to solve technically; you have to have a structure in place, so those rules get updated if new knowledge arises. If you have an automated artificially intelligent system – which we don't, really. So if we get further and further into automating our dialogue systems, having the coaches make more and more decisions on what to say and when to say it, then it becomes relevant to keep the knowledge, on which they base their decisions, up to date. But it is really not that relevant at this point in time."*
> *(Consortium member)*

At this point in time and at this level of automation, the scenario of the coaches automatically updating medical domain knowledge is not yet relevant, but when we advance and get there, our systems are in principle designed to be able to handle that.

8 Conclusion

Having reached the end of this document, the overall question remains as to how well we managed to keep our focus on the issues. Did our RRI Vision make a difference? Did we solve the issues?

To begin with the last: None of these issues will ever be solved completely. The issues we are concerned with in this line of work are by nature always in the making, under development and under construction. Society is changing, stakeholder-interests are changing, how people interact with technology is changing, this is all ongoing. It is our unequivocal understanding that we implemented a very good case, but it needs to grow, and it needs to evolve with the technology, and the issues we pinpointed at the very beginning of this work will need to grow with it. The issues will also evolve as the product moves from research into a commercial product domain. Within the limitation offered by a research project, we believed we developed a very responsible prototype. Developing a responsible *product* will be next step and next phase. But due to our RRI Vision, a great number of distinct decisions from design to implementation were made thinking about these issues. That is the whole point. As we now approach evolving from proof of concept to research to perhaps clinical validation and then to a product, the issues will need to be framed and approached somewhat differently over time in this process. But we have the baseline settled with the discussion we have worked through.

9 Bibliography

Fischer, E., & Schuurbiers, D. (2013). Socio-technical Integration Research: Collaborative Inquiry at the Midstream of Research and Development. *Early engagement and new technologies: Opening up the laboratory. Philosophy of Engineering and Technology.*

Stahl, B. (2017). *Guide for the implementation of Responsible Research and Innovation (RRI) in the industrial context.* the Responsible-Industry Project Consortium.

Acknowledgements

The Council of Coaches project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement #769553. This result only reflects the author's view and the EU is not responsible for any use that may be made of the information it contains.

Headings and titles in this document, as well as the Council of Coaches logo use the Comfortaa font, designed by Johan Aakerlund and Cyreal and licensed under the Open Font License[4].

Additional text in this document uses the Roboto font, designed by Christian Robertson and licensed under the Apache License, Version 2.0[5].

The Council of Coaches logo and Blobmen graphics were *drawn freely* in Inkscape, licensed under the GNU General Public License[6].

[4] Open Font License: http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=OFL_web
[5] Apache License, Version 2.0: http://www.apache.org/licenses/LICENSE-2.0
[6] Inkscape License Information: https://inkscape.org/about/license/