

## D5.6: Evaluation of the Dialogue and Argumentation Framework

**Dissemination level:** Public

**Document type:** Report

**Version:** 1.0.0

**Date:** August 31<sup>st</sup>, 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

<b>Project Number</b>	769553
<b>Project title</b>	Council of Coaches
<b>Title of deliverable</b>	D5.6: Evaluation of the Dialogue and Argumentation Framework
<b>Due date of deliverable</b>	August 31 <sup>st</sup> 2020
<b>Work package</b>	WP5
<b>Author(s)</b>	Mark Snaith (UDun), Jacky Visser (UDun), Tessa Beinema (RRD)
<b>Reviewer(s)</b>	Catherine Pelachaud (SU), Harm op den Akker (RRD)
<b>Approved by</b>	Coordinator
<b>Dissemination level</b>	Public
<b>Document type</b>	Report
<b>Total number of pages</b>	30

## 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

This deliverable presents the results of the evaluation of the Dialogue and Argumentation Framework (DAF), using the methods presented in Deliverable 5.5. The two aims of the evaluation were to determine: 1) whether or not scripted dialogues can be used to underpin the development of abstract dialogue protocols; and 2) how realistic DAF-driven dialogues are, based on several metrics, as determined by a panel of experts.

The results show that scripted dialogues can indeed be used to underpin development of abstract dialogue protocols, and that the DAF-driven dialogues are in some metrics comparable to real dialogues, but that further investigation is required.

## Table of Contents

1	Introduction.....	6
2	Objectives.....	7
3	Method 1: recreating concrete dialogues .....	8
3.1	Approach .....	8
3.2	Methodology .....	8
3.2.1	Changes to methodology.....	8
3.2.2	Input dialogue.....	9
3.3	Results.....	9
3.3.1	Recreated dialogue .....	9
3.3.2	New dialogues.....	10
3.4	Discussion.....	11
3.4.1	Recreated dialogue .....	11
3.4.2	New dialogues.....	11
4	Method 2: expert-based evaluation .....	12
4.1	Approach .....	12
4.2	Methodology .....	12
4.2.1	Changes to methodology.....	12
4.2.2	Dialogue transcripts .....	12
4.2.3	Evaluation process.....	13
4.3	Results.....	14
4.4	Discussion.....	16
4.4.1	Interruptions .....	16
4.4.2	Useful .....	16
4.4.3	Poor-scoring combinations .....	16
5	Summary and conclusions.....	17
6	Bibliography .....	18
	Appendix 1: WOOL script for goal setting dialogue.....	20
	Appendix 2: DGD specification for new protocol.....	27

## List of figures

Figure 1: Simplified updated DAF architecture including the WOOL platform.....	8
Figure 2: Evaluation matrix (from D5.5).....	14

## List of tables

Table 1: Average scores for the evaluation of real dialogue 1.....	14
Table 2: Average scores for the evaluation of real dialogue 2.....	15
Table 3: Average scores for the evaluation of DAF-driven dialogue 1. ....	15
Table 4: Average scores for the evaluation of DAF-driven dialogue 2. ....	15
Table 5: Average scores across all dialogues.....	16

## Symbols, abbreviations and acronyms

CMC	Centre for Monitoring and Coaching
COUCH	Council of Coaches
D	Deliverable
DBT	Danish Board of Technology Foundation
DGDL	Dialogue Game Description Language
DGEP	Dialogue Game Execution Platform
EC	European Commission
ISPRINT	Innovation Sprint
M	Month
MS	Milestone
RRD	Roessingh Research and Development
SU	Sorbonne University
UDun	University of Dundee
UPV	Universitat Politècnica de València
UT	University of Twente
WP	Work Package

# 1 Introduction

The Dialogue and Argumentation Framework (DAF) provides the core logic component to the Council of Coaches technical demonstrator, and the Agents United open agent platform. One of its key strengths is the ability to define abstract dialogue structures that govern the flow of a dialogue, but without stipulating the exact content. The content itself is derived from variables stored in the shared knowledge base (SKB), and models of argument that can make inferences from those variables. For this to be of practical use, however, the dialogues need to provide an engaging and realistic user experience. This is true for both structure and content.

The aim of this deliverable is to execute and evaluate the methods described in Deliverable 5.5 (Snaith, Beinema, op den Akker, & Pease, 2020): recreating concrete (scripted) dialogues, and expert-based evaluation related to aspects including flow, naturalness and usefulness.

For each of the two methods, we first summarise the approach and methodology described in D5.5 before reporting the results. In some cases, further investigation has required the methods reported in D5.5 to be adjusted. Where this has happened, we describe these adjustments along with appropriate justification for doing so.

## 2 Objectives

The primary objective of this deliverable is to report on the evaluation of the Dialogue and Argumentation Framework (DAF) on the basis of the methods described in D5.5. A secondary aim is to use the outcomes of the evaluation to set out directions for future work that will feed into both the Agents United open agent platform, and potential research projects.



## 3 Method 1: recreating concrete dialogues

### 3.1 Approach

This method involves the development of a dialogue game and associated content that, when executed within the Dialogue and Argumentation Framework, can both recreate the scripted dialogue upon which it was based, and lead to different dialogues based on changing user responses.

### 3.2 Methodology

#### 3.2.1 Changes to methodology

Following the design of the methods for evaluation, reported in D5.5, further technical discussions took place regarding how dialogues should be represented and executed. These discussions were prompted by the observation that certain dialogues contained only static content, and had very few paths that could be followed. Examples include the coaches introducing themselves to a user, and a coach explaining different types of goal. Representing such dialogues in Dialogue Game Description Language (DGDL), and executing them using the Dialogue Game Execution Platform (DGEP) introduces unnecessary complication. DGEP and DGDL are designed to underpin complex and dynamic dialogues where the exact flow cannot be predicted in advance. Where the exact flow is known in advance, it is much simpler to write a fully-scripted dialogue.

It was therefore decided to integrate WOOL into the DAF as an alternative dialogue execution engine, with the choice of engine being made depending on the topic chosen by the Topic Selection algorithm. Figure 1 shows a simplified version of the updated DAF architecture, incorporating the WOOL platform. The *Controller* component maps topics to dialogue execution engines; for instance, “goalsetting” is mapped to “DGEP/Utterance Generator”, while “introduction” is mapped to “WOOL”.

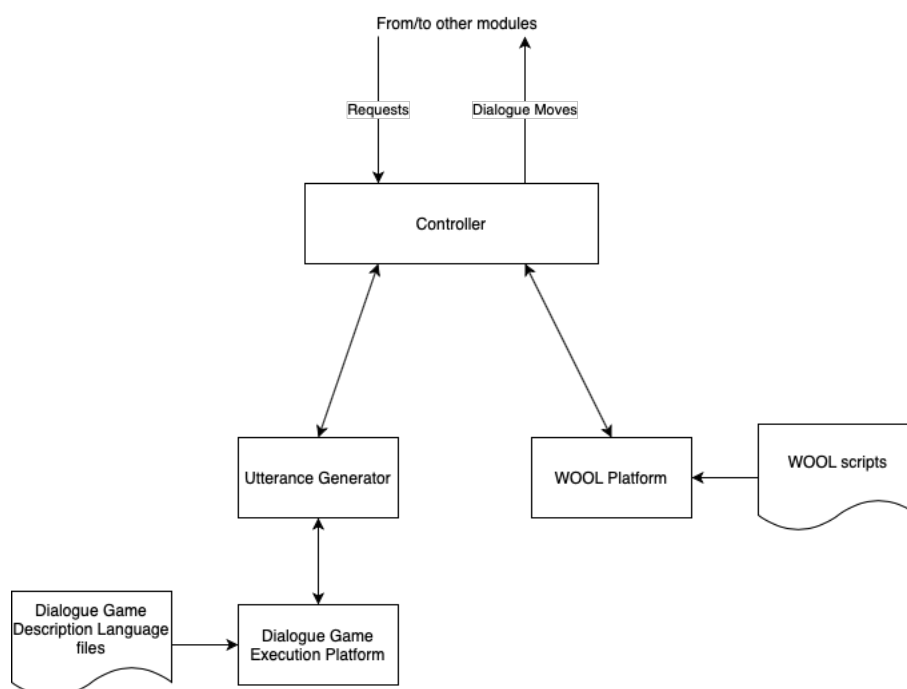


Figure 1: Simplified updated DAF architecture including the WOOL platform.

Since the purpose of this evaluation method is to determine whether or not abstract dialogue games represented in DGDL are capable of recreating scripted dialogues, only the actual goal-setting (i.e. dynamic) part of Example 2 in D5.5 (reproduced below) will be considered.

Furthermore, it was also decided that new dialogues (for the second part of this evaluation method) should be generated entirely from within the DAF itself rather than using the technical demonstrator. This was to ensure that the evaluation was entirely self-contained within the DAF and could not be influenced by other components in the technical demonstrator architecture.

### 3.2.2 Input dialogue

The following excerpt is taken from Example 2 in D5.5, and represents the actual goal-setting (i.e. dynamic) stage of that dialogue. The WOOL script underpinning this dialogue is provided in Appendix 1.

**Example 1: Physical activity coach Olivia wants to set a long-term goal with the user (Bob) (from D5.5).**

**Olivia:** The next step is to decide how many minutes would be a challenging but achievable goal for you in the long-term.

**Olivia:** The World Health Organization recommends that people aged above 65 years old participate in at least 2.5 hours of moderate to high intensity physical activity per week. Think of 3 sessions a bit shorter than 1 hour per week in which your heart rate is elevated. What do you think about this goal?

**Bob:** *It sounds good, let's do that!* | *That sounds too easy for me. Could you suggest a higher goal?* | *It might be a bit high for me. Could you please suggest a lower goal?*

**Olivia:** Ok, then I would like to suggest a long-term goal of 120 minutes of moderate to high physical activity intensity. Remember that we are discussing the long-term goal, not your goal for next week. Is that ok?

**Bob:** *I still think that is too high.* | *Oh, no, I think that is too low.* | *Yes, let's do that.*

**Olivia:** That is great! So just to make sure, we are working towards 120 minutes of moderate to high intensity physical activity per day.

**Emma:** Group activities are a great way to motivate you to reach your physical activity goal! And you might meet some interesting people as well!

**Olivia:** Yes Emma, that's true, and if you need any other tips on how to become more active you know where to find me.

## 3.3 Results

### 3.3.1 Recreated dialogue

The transcript of the dialogue recreated using DGEP is shown below. Note that in real use, the starting point of the dialogue would immediately follow a scripted (WOOL) dialogue where the user was asked to choose what type of goal (steps or active minutes) they wanted to set.

**Recreated Dialogue 1: Bob discusses a physical activity goal with Olivia and Emma.**

**Olivia:** The next step is to decide how many minutes would be a challenging but achievable goal for you in the long-term. (1)

**Olivia:** I would like to suggest a long-term goal of 150 minutes of moderate to high intensity physical activity per week. Is that OK? (2)

**Bob:** I'm not sure about that. Why do you suggest 150 minutes? (3)

**Olivia:** The World Health Organization recommends that people aged above 65 years old participate in at least 150 minutes of moderate to high intensity physical exercise per week. (4)

**Olivia:** So what do you say, can you manage 150 minutes? (5)

**Bob:** I'm still not sure. It might be a bit high for me. Could you suggest a lower goal? (6)

**Olivia:** OK, but remember that we are discussing the long-term goal, not your goal for next week. (7)

**Olivia:** I would like to suggest a long-term goal of 120 minutes of moderate to high intensity physical activity per week. Is that OK? (8)

**Bob:** Yes, let's do that. (9)

**Olivia:** That's great! So just to make sure, we are working towards 120 minutes of moderate to high intensity physical activity per day. (10)

**Emma:** Group activities are a great way to motivate you to reach your physical activity goal! And you might meet some interesting people as well! (11)

**Olivia:** Yes, that's true, and if you need any other tips on how to become more active, you know where to find me. (12)

### 3.3.2 New dialogues

The following dialogues were generated using the same dialogue game, but with different content and with different intentions. In the first, Bob is setting a long-term step goal, and in the second Debbie is setting a short-term exercise goal.

Both of these dialogues were subsequently used in the expert-based evaluation described in Section 4.

#### New Dialogue 1: Alice helps Bob set a long-term steps goal.

**Alice:** Hey Bob, it's time to set your long-term steps goal. Let's have a think about what would make a good goal. How do you feel about 7,500 steps?

**Bob:** I'm not sure about that. Why do you suggest 7,500 steps?

**Alice:** General guidance says that 10,000 steps is normal, however for someone of your age 7,500 steps is recommended.

**Alice:** So what do you say; do you think you can manage 7, 500 steps?

**Bob:** I'm still not sure. 7,500 steps seems a bit high to me. Do you think it could be a bit lower?

**Alice:** No problem; let's see now...how do you feel about 6,00 steps?

**Bob:** That sounds a lot better! I can definitely manage that.

**Alice:** Fantastic! I'm sure you'll manage it and then some. So that's a long-term goal of 6,00 steps; next we'll discuss your short-term goal based on that.

#### New Dialogue 2: Charlie helps Debbie set a short-term minute-goal.

**Charlie:** Now that we've set your long-term minutes goal, we should look at a short-term goal to help you reach it. Now, your long-term goal is 60 minutes. Based on that, do you think you can manage 30 minutes of exercise a day?

**Debbie:** 30 minutes seems a bit low to me. Do you think it could be a bit higher?

**Charlie:** No problem; let's see now...do you think you can manage 40 minutes of exercise a day?

**Debbie:** 40 minutes now seems a bit high; do you think it can be a bit lower?

**Charlie:** No problem; let's see now...do you think you can manage 35 minutes of exercise a day?

**Debbie:** I'm not sure about that. Why do you suggest 35 minutes?

**Charlie:** Your long-term goal is 60 minutes, so this will get you off to a great start, however it's just a bit lower than 40 minutes but you should manage it. So what do you say; do you think you can manage 35 minutes?

**Debbie:** 35 minutes seems a bit low to me. Do you think it could be a bit higher?

**Charlie:** Well, it could but why don't we just set your goal at 35 minutes and see how you get on? So that's a short-term goal of 35 minutes.

## 3.4 Discussion

### 3.4.1 Recreated dialogue

There are some slight differences between the original dialogue script and the recreated version, in terms of both structure and content.

First, Olivia does not justify the goal at the point of proposing (statement 1); instead, the justification (4) comes following a query from Bob (3). This is because of the way that arguments are represented in the DAF: instead of presenting a full argument at once, the conclusion of the argument (in this case, suggesting 150 minutes) is first presented, with reasons subsequently being provided in response to a challenge.

Second, after Bob requests a lower goal (6) Olivia provides her reminder (7) before suggesting the new goal (8). This is a consequence of the abstract nature of dialogue games. When the goal is rejected, the dialogue game returns to the beginning to start the process again. Statements (1) and (7) represent the same dialogue move ("PrePropose"), but their content is different because it is not derived from the conclusion of an argument. Statements (2) and (8) are also the same dialogue move ("Propose") but because that content is derived from the conclusion of an argument, they have the same general structure, with the number of minutes being changed through a variable.

Finally, Olivia does not reference Emma by name (11). This is a minor technical issue whereby the DAF does not currently support agents referencing each other by name.

There are however key similarities. Both dialogues have the same outcome (Bob and his coaches agreeing a goal of 120 minutes), and contain the same broad content (Olivia justifying the original goal based on WHO advice; Olivia reminding Bob that it was a long-term goal he was setting; and Emma suggesting group activities).

It is therefore possible for the DAF to recreate scripted dialogues in terms of content and outcome, albeit with a slightly different structure.

### 3.4.2 New dialogues

The new dialogues used the same dialogue game as the recreated dialogue, but using different content and different coaching intentions. For the first new dialogue, the content and intention focused on setting a long-term steps goal, while for the second the content and intention focused on setting a short-term minutes goal.

This demonstrates that a scripted dialogue can be used to create an abstract dialogue game that is capable of underpinning completely new dialogues on a similar topic. In this case, the original scripted dialogue was based on setting a long-term goal for exercise minutes; the two new dialogues focused on setting a long-term goal for steps, and a short-term goal for exercise minutes.

## 4 Method 2: expert-based evaluation

### 4.1 Approach

This method involved generating two dialogues using the Dialogue and Argumentation Framework (DAF) and asking a panel of experts to evaluate the transcripts of these dialogues along several dimensions. To obtain a baseline measure, the experts were first asked to evaluate two excerpts of dialogue from the Patient Consultation Corpus (PCC) (Snaith, Pease, & Janier, 2018), the corpus of video-recorded and annotated interactions recorded earlier in the project. While these dialogues were created using actors representing patients, they were still “real” insofar as they happened between real people. This allows for an accurate comparison with dialogues obtained from the (DAF).

### 4.2 Methodology

#### 4.2.1 Changes to methodology

In D5.5, we presented an evaluation methodology that involved assembling a panel of experts to discuss the characteristics of the dialogue. Due to various factors, it was not possible to assemble this panel, either in-person or online. We therefore removed the plenary aspect of the evaluation and instead based the evaluation entirely on individual assessments.

#### 4.2.2 Dialogue transcripts

The following real dialogue excerpts were used to obtain the baseline score in the evaluation:

##### Real dialogue 1: Barbara persuades Linda to switch to wholemeal bread.

**Barbara:** How would you feel about switching to wholemeal bread?

**Linda:** I mean, will it really make a difference? Like, will it really make a difference switching?

**Barbara:** Not so much from a calorie point of view or a weight loss point of view, a piece of white bread and a piece of brown bread are, calorie content, roughly the same. However, brown bread has a lot more fibre in it, so actually what it will do is raise your blood sugar a little bit slower than the white bread and keep you feeling fuller for longer.

**Linda:** Ah, okay, I didn't know that.

**Barbara:** So that sounds like a small change and I just wonder how you feel about whether that is something you could switch to?

**Linda:** Yea, I certainly think that's something, as long as I'm still going to feel full.

##### Real dialogue 2: Colin tries to understand how Anne has been feeling.

**Colin:** So just tell us how things have been?

**Anne:** Not great, to be honest. I'm finding the news quite difficult. I'm in the red zone apparently, and I'm really worried about what that might entail for me in the future.

**Colin:** So, when you're saying you're in the red zone, tell me a bit about what you understand by that?

**Anne:** That basically, the worst possible things can happen, like having your foot amputated...

**Colin:** Oh, okay

**Anne:** ...becoming blind.

**Colin:** Okay. So you're aware of the long-term consequences of diabetes, where your blood sugars would be at that red level, and what could happen as a consequence of that? And it sounds as if that's really got you worried about what your future might look like?

The following DAF-driven dialogue transcripts were used in the evaluation.

**DAF-driven Dialogue 1: Alice helps Bob set a long-term steps goal.**

**Alice:** Hey Bob, it's time to set your long-term steps goal. Let's have a think about what would make a good goal. How do you feel about 7,500 steps?

**Bob:** I'm not sure about that. Why do you suggest 7,500 steps?

**Alice:** General guidance says that 10,000 steps is normal, however for someone of your age 7,500 steps is recommended.

**Alice:** So what do you say; do you think you can manage 7, 500 steps?

**Bob:** I'm still not sure. 7,500 steps seems a bit high to me. Do you think it could be a bit lower?

**Alice:** No problem; let's see now...how do you feel about 6,00 steps?

**Bob:** That sounds a lot better! I can definitely manage that.

**Alice:** Fantastic! I'm sure you'll manage it and then some. So that's a long-term goal of 6,00 steps; next we'll discuss your short-term goal based on that.

**DAF-driven Dialogue 2: Charlie helps Debbie set a short-term minutes goal.**

**Charlie:** Now that we've set your long-term minutes goal, we should look at a short-term goal to help you reach it. Now, your long-term goal is 60 minutes. Based on that, do you think you can manage 30 minutes of exercise a day?

**Debbie:** 30 minutes seems a bit low to me. Do you think it could be a bit higher?

**Charlie:** No problem; let's see now...do you think you can manage 40 minutes of exercise a day?

**Debbie:** 40 minutes now seems a bit high; do you think it can be a bit lower?

**Charlie:** No problem; let's see now...do you think you can manage 35 minutes of exercise a day?

**Debbie:** I'm not sure about that. Why do you suggest 35 minutes?

**Charlie:** Your long-term goal is 60 minutes, so this will get you off to a great start, however it's just a bit lower than 40 minutes but you should manage it. So what do you say; do you think you can manage 35 minutes?

**Debbie:** 35 minutes seems a bit low to me. Do you think it could be a bit higher?

**Charlie:** Well, it could but why don't we just set your goal at 35 minutes and see how you get on? So that's a short-term goal of 35 minutes.

### 4.2.3 Evaluation process

Each panel member was:

1. provided with the two real dialogues, the two DAF-driven dialogues, and the score matrix in D5.5, and shown again here in Figure 2. The dialogues were numbered 1-4, and no panel member was told which dialogues were real and which were DAF-driven.
2. asked to complete a score matrix for each dialogue, using the numbers 1 ("poor") to 5 ("excellent") inclusive, or "N/A" if they felt that a certain aspect did not occur, or a certain metric did not apply to a certain aspect.

		Aspects of a Dialogue				
		Opening	Coaching	Closing	Turn-taking	Interruptions
<b>Evaluation Metrics</b>	Flow					
	Natural					
	Effective					
	Useful					
	Sensible					
	Specific					

Figure 2: Evaluation matrix (from D5.5).

### 4.3 Results

The dialogues were due to be evaluated by three experts, however due to unforeseen circumstances results were only returned from two. With this representing a small sample size, we decide to present here only a simple average (mean) of their scores in a combined matrix. Where an expert used “N/A”, we represent this as 0.

Tables 1 through 4 below show the average scores assigned by two panel members.

Table 1: Average scores for the evaluation of real dialogue 1.

	Opening	Coaching	Closing	Turn-taking	Interruptions
<b>Flow</b>	3.5	4.5	4.5	5	0
<b>Natural</b>	4.5	5	4.5	5	0
<b>Effective</b>	4	5	3.5	2	0
<b>Useful</b>	2	4.5	3.5	1.5	0
<b>Sensible</b>	4.5	4.5	4.5	4.5	0
<b>Specific</b>	4	4	4	1.5	0

Table 2: Average scores for the evaluation of real dialogue 2.

	Opening	Coaching	Closing	Turn-taking	Interruptions
Flow	3	2.5	3.5	2.5	3
Natural	3.5	3	4	4.5	4.5
Effective	3	3	3	0.5	1
Useful	1	2	3	1	1
Sensible	2.5	3	3.5	4	2
Specific	1	2	3.5	1	2

Table 3: Average scores for the evaluation of DAF-driven dialogue 1.

	Opening	Coaching	Closing	Turn-taking	Interruptions
Flow	2.5	3	3	3	0
Natural	2	2	3	2	0
Effective	3.5	3.5	3.5	1.5	0
Useful	1.5	3	3	1.5	0
Sensible	3.5	3	3.5	3.5	0
Specific	3.5	4	3	1.5	0

Table 4: Average scores for the evaluation of DAF-driven dialogue 2.

	Opening	Coaching	Closing	Turn-taking	Interruptions
Flow	3.5	2.5	2.5	3	0
Natural	3	2	3	3.5	0
Effective	4	1.5	2	0.5	0
Useful	2	2	3	1	0
Sensible	3.5	3	2.5	3.5	0
Specific	4	3	3.5	1	0



Table 5 shows the average scores across all dialogues.

**Table 5: Average scores across all dialogues.**

	Opening	Coaching	Closing	Turn-taking	Interruptions
<b>Flow</b>	3.125	3.125	3.375	3.375	0.75
<b>Natural</b>	3.25	3	3.625	3.75	1.125
<b>Effective</b>	3.625	3.25	3	1.125	0.25
<b>Useful</b>	1.625	2.875	3.125	1.25	0.25
<b>Sensible</b>	3.5	3.375	3.5	3.875	0.5
<b>Specific</b>	3.125	3.25	3.5	1.25	0.5

## 4.4 Discussion

Overall, it can be seen that the DAF-driven dialogues scored below average on most of the metrics, however in some cases (e.g. Flow-Opening of DAF-driven dialogue 2) they performed above average. Here, we highlight some aspects that are of particular interest.

### 4.4.1 Interruptions

Both DAF-driven dialogues scored '0' for interruptions for all aspects. The explanation for this is straightforward: currently, the DAF does not natively support interruptions. When a dialogue move is received, it is assumed to be "complete". A means of supporting interruption has been implemented within the technical demonstrator, however because this does not take place within the DAF it would not have been appropriate to evaluate it as part of the present process.

### 4.4.2 Useful

Both DAF-driven dialogues scored poorly for "useful" across all aspects of dialogue. It should be noted, however, that the same was observed for the second real dialogue, and two out of the four observed metrics for the first real dialogue. As such, this suggests that further investigation is required into what exactly "useful" means in this context.

### 4.4.3 Poor-scoring combinations

Excluding interruptions, several combinations of aspects and metrics scored poorly across all or at least three dialogues (real and DAF-driven), such as *effective, useful and specific turn-taking* (all dialogues), and *effective and useful opening* (all but the first real dialogue). Similar to the issues with "Useful" as a metric, further investigation will be required to determine whether or not certain metric-aspect pairs are in fact measurable, and as such whether they are relevant in determining the quality of a dialogue.

## 5 Summary and conclusions

This deliverable has provided an evaluation of the Dialogue and Argumentation Framework (DAF), based on the methods described in Deliverable 5.5. Two different evaluations were performed: first, using an existing, scripted dialogue as the input for a (manual) development of an abstract dialogue game then determining whether or not that game can recreate the original dialogue, and also underpin new dialogues. Second, an expert-based evaluation to assess the quality of DAF-driven dialogues.

The first evaluation showed that this is indeed possible, with only relatively minor structural changes to the dialogue flow. A general dialogue game and associated content was able to replicate the outcomes and same broad content as the original scripted version upon which the game was based. The dialogue game, again when provided with relevant content, is also capable of underpinning completely new dialogues on the same broad topic (goal setting).

The second evaluation showed that in some aspects, DAF-driven dialogues are at least as good as real dialogues, however in other aspects they lagged behind. It was also however the case that even real dialogues scored poorly on certain aspects and combinations of aspects and metrics. As such, further investigation will be required along two strands: first, improvements to the DAF to make dialogues closer to real dialogues; second, determining whether or not the aspects and metrics used in the evaluation are all appropriate, including combinations thereof. These investigations will form the core of future work in the context of developing the Agents United platform.

## 6 Bibliography

Snaith, M., Beinema, T., op den Akker, H., & Pease, A. (2020). *D5.5: Methods for evaluating the Dialogue and Argumentation Framework*. The Council of Coaches Consortium.

Snaith, M., Pease, A., & Janier, M. (2018). *D5.2: Design of Dialogue Game*. Council of Coaches 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<sup>1</sup>.

Additional text in this document uses the Roboto font, designed by Christian Robertson and licensed under the Apache License, Version 2.0<sup>2</sup>.

The Council of Coaches logo and Blobmen graphics were *drawn freely* in Inkscape, licensed under the GNU General Public License<sup>3</sup>.

---

<sup>1</sup> Open Font License: [http://scripts.sil.org/cms/scripts/page.php?site\\_id=nrsi&id=OFL\\_web](http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=OFL_web)

<sup>2</sup> Apache License, Version 2.0: <http://www.apache.org/licenses/LICENSE-2.0>

<sup>3</sup> Inkscape License Information: <https://inkscape.org/about/license/>

## Appendix 1: WOOL script for goal setting dialogue

```

title: End
tags:
speaker: Olivia
colorID: 5
position: 14303,3450
---

===
title: Start
tags:
speaker: Olivia
colorID: 2
position: 12946,2825
---
Setting a goal for yourself is the first step towards becoming more active. Actually, did you know
that people who commit to a goal are more likely to become more active?

[[SetFinalGoal2]]
===
title: SetFinalGoal2
tags:
speaker: Olivia
colorID: 0
position: 12943,3108
---
You can set goals for the upcoming days or weeks. That's what we call "short-term" goals, or for
the upcoming months. That's what we call "long-term" goals.

[[SetFinalGoal3]]
===
title: SetFinalGoal4
tags:
speaker: Olivia
colorID: 0
position: 12946,3609
---
So let's talk about your long-term goal...

[[Do you have a suggestion?|SetFinalGoalSuggest]]
[[Why do I need a long term goal?|SetFinalGoalWhy]]
[[Can we talk about something else?|olivia-menu.Start]]
[[Goodbye.|End]]
===
title: SetFinalGoalSuggest
tags:
speaker: Olivia
colorID: 0
position: 13138,3958
---
First you should decide what kind of goal you want to set. We can look at the number of steps taken
or the minutes of higher activity per week or per day.

[[SetFinalGoalSuggest2]]
===
title: SetFinalGoalWhy
tags:
speaker: Olivia
colorID: 0
position: 12826,3966
---
Why? Well, there are two good reasons for setting a long-term goal for yourself.

[[SetFinalGoalWhy2]]
===
title: SetFinalGoalWhy2
tags:
speaker: Olivia
colorID: 0
position: 12551,3967
---
First, it can be a great motivation for yourself. It gives you something to work towards.

[[SetFinalGoalWhy3]]
===
title: SetFinalGoalWhy3
tags:
speaker: Olivia
colorID: 0

```

```
position: 12287,3964
---
So, your long-term goal should be ambitious! Something that will make you feel really proud once
you achieve it.

[[SetFinalGoalWhy4]]
===
title: SetFinalGoalWhy4
tags:
speaker: Olivia
colorID: 0
position: 12021,3960
---
And the second reason is that it gives us something to cut into little pieces, something to tackle
day-by-day.

[[SetFinalGoalWhy5]]
===
title: SetFinalGoalWhy5
tags:
speaker: Olivia
colorID: 0
position: 11738,3965
---
So having a long-term goal makes it possible to set relevant and personal daily goals. And those
daily goals give us a concrete objective for every day!

[[Okay, can you suggest a goal?|SetFinalGoalSuggest]]
[[Can we talk about something else?|olivia-menu.Start]]
[[Goodbye.|End]]
===
title: SuggestStepsTaken
tags:
speaker: Olivia
colorID: 0
position: 13001,4529
---
Your activity tracker constantly measures how much you move, and tries to estimate this as a number
of steps.

[[SuggestStepsTaken2]]
===
title: SuggestStepsTaken2
tags:
speaker: Olivia
colorID: 0
position: 13001,4779
---
So when we're talking about "steps" it's not just about walking, it's about every little thing you
do throughout the day.

[[SuggestStepsTaken3]]
===
title: SuggestStepsTaken3
tags:
speaker: Olivia
colorID: 0
position: 13001,5029
---
So, should we set a step goal? Or should I tell you about active minutes?

[[I am ready to decide.|DecidePreferredUnit]]
[[What about active minutes?|SuggestActiveMinutes]]
[[Goodbye.|End]]
===
title: SuggestActiveMinutes
tags:
speaker: Olivia
colorID: 0
position: 13251,4529
---
Although every little bit of activity helps, it is also important to do things that require a little
more effort. Basically this is when you feel your heart beat going faster, maybe you become sweaty,
and you feel tired at the end of the activity.

[[SuggestActiveMinutes2]]
===
title: SetFinalGoalSuggest2
tags:
speaker: Olivia
colorID: 0
position: 13137,4219
```

```

---
I can tell you about each of the options, or let me know if you're ready to decide.

[[Steps taken?|SuggestStepsTaken]]
[[Active minutes?|SuggestActiveMinutes]]
[[I am ready to decide.|DecidePreferredUnit]]
[[Goodbye.|End]]
===
title: DecidePreferredUnit
tags:
speaker: Olivia
colorID: 6
position: 13639,4692
---
That's good to hear. What will it be?

[[Steps.|ConfirmPreferredUnitSteps|<<set $paPreferredUnit = "steps">>]]
[[Active minutes.|ConfirmPreferredUnitMinutes|<<set $paPreferredUnit = "minutes">>]]
[[Actually, could you explain the options again?|RepeatOptions]]
===
title: SuggestActiveMinutes2
tags:
speaker: Olivia
colorID: 0
position: 13251,4779
---
Your activity tracker can see how many time you spent doing a little extra, which we call "active
minutes".

[[SuggestActiveMinutes3]]
===
title: SuggestActiveMinutes3
tags:
speaker: Olivia
colorID: 0
position: 13251,5029
---
So would you like to set a goal for active minutes?

[[I'm ready to decide.|DecidePreferredUnit]]
[[What about steps?|SuggestStepsTaken]]
[[Goodbye.|End]]
===
title: ConfirmPreferredUnitSteps
tags:
speaker: Olivia
colorID: 0
position: 13936,4558
---
Good! I will help you work towards a long-term step goal then!

[[SetFinalGoalValue]]
===
title: ConfirmPreferredUnitMinutes
tags:
speaker: Olivia
colorID: 0
position: 13938,4865
---
Good! I will help you work towards a long-term goal for active minutes then!

[[SetFinalGoalValue]]
===
title: RepeatOptions
tags:
speaker: Olivia
colorID: 0
position: 13634,4152
---
Of course, which option would you like me to explain again?

[[Steps taken?|SuggestStepsTaken]]
[[Active minutes?|SuggestActiveMinutes]]
===
title: SetFinalGoalValue
tags:
speaker: Olivia
colorID: 2
position: 14239,4699
---
<<if $paPreferredUnit == "steps">>

```

Next we need to decide how many steps would be a challenging but achievable goal for you in the long-term.

```
[[SetLongTermGoalValueSteps]]
<<elseif $paPreferredUnit == "minutes">>
```

The next step is to decide how many minutes would be a challenging but achievable goal for you in the long-term.

```
[[SetLongTermGoalValueMinutes]]
<<endif>>
```

```
===
title: SetFinalGoal3
tags:
speaker: Olivia
colorID: 0
position: 12943,3353
---
```

You can think of it as setting aside some money every week (short-term goal) in order to save for a weekend trip (long-term goal). Similar with activity... you commit to become a little more active now, to have greater benefits later.

```
[[SetFinalGoal4]]
===
title: SetLongTermGoalValueMinutes
tags:
speaker: Olivia
colorID: 0
position: 14543,4855
---
```

```
<<if $userAge >= 65 >>
    The World Health Organization recommends that people aged above 65 years old participate in at least 2.5 hours of moderate to high intensity physical activity per week. Think of 3 sessions a bit shorter than 1 hour per week in which your heart rate is elevated. What do you think about this goal?
```

```
    [[It sounds good, let's do that!|LongTermGoalAccepted|<<set $paLongTermMinutesGoal = 150>>]]
```

```
<<elseif $userAge >= 18 >>
    The World Health Organization recommends that adults participate in at least 150 minutes of moderate to high intensity physical activity per week. Think of 3 sessions a bit shorter than 1 hour per week in which your heart rate is elevated. What do you think about this goal?
```

```
    [[It sounds good, let's do that!|LongTermGoalAccepted|<<set $paLongTermMinutesGoal = 150>>]]
```

```
<<elseif $userAge < 18 >>
    The World Health Organization recommends that those aged between 5 and 17 years old participate in at least 60 minutes of moderate to high intensity physical activity per day. What do you think about this goal?
```

```
    [[It sounds good, let's do that!|LongTermGoalAccepted|<<set $paLongTermMinutesGoal = 60>>]]
```

```
<<endif>>

[[That sounds too easy for me. Could you suggest a higher goal?|SuggestHigherLongTermGoal]]
[[It might be a bit high for me. Could you please suggest a lower goal?|SuggestLowerLongTermGoal]]
===
```

```
title: LongTermGoalAccepted
tags:
speaker: Olivia
colorID: 0
position: 15542,4723
---
```

```
<<set $paUserLongTermGoalSet = true>>
<<set $paUserLongTermGoalSetDate = $today>>
```

```
<<if $paPreferredUnit == "steps">>
```

That is great! So just to make sure, we are working towards \$paLongTermStepsGoal steps per day on the long run.

```
<<elseif $paPreferredUnit == "minutes">>
```

```
    <<if $userAge >= 18 >>
```

That is great! So just to make sure, we are working towards \$paLongTermMinutesGoal minutes of moderate to high intensity physical activity per week.

```
    <<else>>
```

That is great! So just to make sure, we are working towards \$paLongTermMinutesGoal minutes of moderate to high intensity physical activity per day.



```

    <<endif>>
<<endif>>
<<if $coachEmmaEnabled>>
    [[SocialActivitiesPhysicalActivity]]
<<else>>
    [[SetLongTermGoalEnd]]
<<endif>>
===
title: SuggestLowerLongTermGoal
tags:
speaker: Olivia
colorID: 0
position: 14870,5000
---
<<if $paPreferredUnit == "steps">>
    <<if $userAge >= 65>>
        Ok, then I would like to suggest a long-term goal of 5.000 steps per day. Is that ok?

        [[Yes, let's do that.|LongTermGoalAccepted|<<set $paLongTermStepsGoal = 5000>>]]
    <<elseif $userAge >= 18 >>
        Ok, then I would like to suggest a long-term goal of 7.500 steps per day. Is that ok?

        [[Yes, let's do that.|LongTermGoalAccepted|<<set $paLongTermStepsGoal = 7500>>]]
    <<else>>
        Ok, then I would like to suggest a long-term goal of 10.000 steps per day. Is that ok?

        [[Yes, let's do that.|LongTermGoalAccepted|<<set $paLongTermStepsGoal = 10000>>]]
    <<endif>>

    [[I still think that is too high.|SuggestLowestLongTermGoal]]
    [[Oh, no, I think that is too low.|SetLongTermGoalValueSteps]]

<<elseif $paPreferredUnit == "minutes">>
    <<if $userAge >= 18>>
        Ok, then I would like to suggest a long-term goal of 120 minutes of moderate to high physical
        activity intensity per week. Remember that we are discussing the long-term goal, not your goal for
        next week. Is that ok?

        [[Yes, let's do that.|LongTermGoalAccepted|<<set $paLongTermMinutesGoal = 120>>]]
    <<else >>
        Ok, then I would like to suggest a long-term goal of 40 minutes of moderate to high physical
        activity intensity per day. Remember that we are discussing the long-term goal, not your goal for
        next week. Is that ok?

        [[Yes, let's do that.|LongTermGoalAccepted|<<set $paLongTermMinutesGoal = 40>>]]
    <<endif>>

    [[I still think that is too high.|SuggestLowestLongTermGoal]]
    [[Oh, no, I think that is too low.|SetLongTermGoalValueMinutes]]

<<endif>>
===
title: SuggestLowestLongTermGoal
tags:
speaker: Olivia
colorID: 0
position: 15185,5002
---
<<if $paPreferredUnit == "steps">>
    <<if $userAge >= 65 >>
        Let's just start with 4.500 steps per day, and we see how it goes. If it does not work for
        you, we can adjust it next week.

        [[Okay, let's do that.|LongTermGoalAccepted|<<set $paLongTermStepsGoal = 4500>>]]
    <<elseif $userAge >= 18 >>
        Let's just start with 7.000 steps per day, and we see how it goes. If it does not work for
        you, we can adjust it next week.

        [[Okay, let's do that.|LongTermGoalAccepted|<<set $paLongTermStepsGoal = 7000>>]]
    <<else>>
        Let's just start with 9.000 steps per day, and we see how it goes. If it does not work for
        you, we can adjust it next week.

```

```

    [[Okay, let's do that.|LongTermGoalAccepted|<<set $paLongTermStepsGoal = 9000>>]]
<<endif>>

<<elseif $paPreferredUnit == "minutes">>

    <<if $userAge >= 18>>
        Let's just start with 60 minutes per week, and we see how it goes. If it does not work for
        you, we can adjust it next week.

        [[Okay, let's do that.|LongTermGoalAccepted|<<set $paLongTermMinutesGoal = 60>>]]
    <<else>>
        Let's just start with 30 minutes per day, and we see how it goes. If it does not work for
        you, we can adjust it next week.

        [[Okay, let's do that.|LongTermGoalAccepted|<<set $paLongTermMinutesGoal = 30>>]]
    <<endif>>

<<endif>>

[[Maybe I can do a bit more...|SuggestLowerLongTermGoal]]
===
title: SuggestHigherLongTermGoal
tags:
speaker: Olivia
colorID: 0
position: 14889,4411
---
<<if $paPreferredUnit == "steps">>

    <<if $userAge >= 65 >>
        Wonderful to hear that you are willing to try a bit more! So, what about 10.000 steps per
        day? Remember, we are talking about long-term goals, not your goal for tomorrow.

        [[Yes, let's do that!|LongTermGoalAccepted|<<set $paLongTermStepsGoal = 10000>>]]
    <<elseif $userAge >= 18 >>
        Great to hear you're willing to step it up! Okay, what about 12.000 steps per day? Remember,
        we are talking about long-term goals, not your goal for tomorrow.

        [[Yes, let's do that!|LongTermGoalAccepted|<<set $paLongTermStepsGoal = 12000>>]]
    <<else>>
        Nice! Let's step it up some more then. What about 15.000 steps per day? Remember, we are
        talking about long-term goals, not your goal for tomorrow.

        [[Yes, let's do that!|LongTermGoalAccepted|<<set $paLongTermStepsGoal = 15000>>]]
    <<endif>>

    [[That might be too high, can we go lower?|SetLongTermGoalValueSteps]]

<<elseif $paPreferredUnit == "minutes">>

    <<if $userAge >= 18>>
        Okay, well... the World Health Organization suggests that for extra health benefits you can
        double the time on moderate to high intensity physical activity. What about aiming for 300 minutes
        per week?

        [[Yes, let's do that!|LongTermGoalAccepted|<<set $paLongTermMinutesGoal = 300>>]]
    <<else>>
        Okay, what about 120 minutes of moderate to high intensity physical activity per day?

        [[Yes, let's do that!|LongTermGoalAccepted|<<set $paLongTermMinutesGoal = 120>>]]
    <<endif>>

    [[That might be too high, can we go lower?|SetLongTermGoalValueMinutes]]

<<endif>>
===
title: SetLongTermGoalValueSteps
tags:
speaker: Olivia
colorID: 0
position: 14544,4554
---
<<if $userAge >= 65 >>
    There is a general trend to recommend 10.000 steps as a daily goal. However, for people aged 65
    or above, this value is often reduced to 7.500 steps. Should we start with a long-term goal of 7.500
    steps?

    [[Okay, let's do that!|LongTermGoalAccepted|<<set $paLongTermStepsGoal = 7500>>]]

<<elseif $userAge >= 18 >>

```

```

    There is a general trend to recommend 10.000 steps as a daily goal. Should we strive to achieve
    10.000 steps in the long-run?

    [[Okay, let's do that!|LongTermGoalAccepted|<<set $paLongTermStepsGoal = 10000>>]]

<<else>>
    It is often suggested that people as young as yourself should reach 13.000 steps a day. Should
    we try 13.000 steps in the long-run?

    [[Okay, let's do that!|LongTermGoalAccepted|<<set $paLongTermStepsGoal = 13000>>]]

<<endif>>

[[That sounds too easy, can we go a bit higher?|SuggestHigherLongTermGoal]]
[[That sounds difficult, could we try lower?|SuggestLowerLongTermGoal]]
===
title: SocialActivitiesPhysicalActivity
tags:
speaker: Emma
colorID: 0
position: 15885,4724
---
Group activities are a great way to motivate you to reach your physical activity goal! And you might
meet some interesting people as well!

[[CommentSocialActivitiesPhysicalActivity]]
===
title: CommentSocialActivitiesPhysicalActivity
tags:
speaker: Olivia
colorID: 0
position: 16217,4727
---
Yes Emma that's true, and if you need any other tips on how to become more active you know where to
find me.

[[SetLongTermGoalEnd]]
===
title: SetLongTermGoalEnd
tags:
speaker: Olivia
colorID: 0
position: 15883,4352
---
So, is there anything else I can help you with?

[[I want to know more about goals.|olivia-coaching-goals-menu.Start]]
[[I want to discuss something else.|olivia-menu.Redirect]]
[[Goodbye.|End]]
===

```

## Appendix 2: DGD L specification for new protocol

```

System{GoalSetting{
    /* Description: A game to discuss a physical activity goal */
    turns{magnitude:multiple, ordering:liberal}
    roles{Person,Agent}
    players{min:2, max:3}
    player{id:AgentOne, roles{Person,Agent}, max:1, min:1}
    player{id:AgentTwo, roles{Person, Agent}, max:1, min:0}
    player{id:User, roles{Person,Agent}, max:1, min:1}
    backtrack{on}

    rule{id:StartingRule, scope:initial,
        {
            assign(AgentOne, speaker)
            & move(add, next, PrePropose, $User, AgentOne)
        }
    }

    interaction{PrePropose, $User, "",
        {
            move(add, next, Propose, $User, {p}, AgentOne)
        }
    }

    interaction{Propose, $User, {p}, "argument([set_goal(?)])",
        {
            assign(User, speaker)
            & move(add, next, AcceptGoal, $AgentOne, {p}, User)
            & move(add, next, ChallengeGoal, $AgentOne, {p}, User)
            & move(add, next, RejectGoalTooHigh, $AgentOne, {p}, User)
            & move(add, next, RejectGoalTooLow, $AgentOne, {p}, User)
            & save({p}, $AgreedGoal$)
        }
    }

    interaction{AcceptGoal, $AgentOne, {p}, "$p",
        {
            assign(AgentOne, speaker)
            & move(add, next, GoalAccepted, $User, {p}, AgentOne)
        }
    }

    interaction{ChallengeGoal, $AgentOne, {p}, "$p",
        {
            assign(AgentOne, speaker)
            & move(add, next, JustifyGoal, $User, {p,q}, AgentOne)
        }
    }
}

```

```

    }

    interaction{JustifyGoal, $User, {p,q}, "argument([?]>=$q)",
        {
            move(add, next, Propose, $User, {p}, AgentOne)
            & move(add, next, AcceptGoal, $Agent, {p}, User)
            & move(add, next, RejectGoalTooHigh, $Agent, {p}, User)
            & move(add, next, RejectGoalTooLow, $Agent, {p}, User)*/
        }
    }

    interaction{RejectGoalTooHigh, $AgentOne, {p}, "$p",
        {
            assign(AgentOne, speaker)
            & move(add, next, PrePropose, $User, AgentOne)
            & move(add, next, Breakout, $User, {p}, AgentOne)
        }
    }

    interaction{RejectGoalTooLow, $AgentOne, {p}, "$p",
        {
            assign(Agent, speaker)
            & move(add, next, PrePropose, $User, AgentOne)
            & move(add, next, Breakout, $User, {p}, AgentOne)
        }
    }

    interaction{Breakout, $User, {p}, "$p",
        {
            move(add, next, End, $User, {$AgreedGoal$}, AgentOne)
        }
    }

    interaction{GoalAccepted, $User, {p}, "$p",
        {
            move(add, next, End, $User, {p}, AgentOne)
            & move(add, next, Interject, $User, AgentTwo)
        }
    }

    interaction{Interject, $User, "$p",
        {
            move(add, next, End, $User, {$AgreedGoal$}, AgentOne)
        }
    }

```

```
        interaction{End, $User, {p}, "$p",  
            {  
                status(terminate, GoalSettingMain)  
            }  
        }  
    }}  

```