



---

# Mapping between MagiCster prototypes and the DIPPER architecture

---

Johan Bos, Valeria Carofiglia, Berardina de Carolis,  
Stephen Crampton, Roberto Grassano, Jarmo Laaksolahti,  
Thomas Rist, Catherine Pelachaud, Isabella Poggi,  
Fiorella de Rosis, Markus Schmitt, Mark Steedman, Claus Zinn

Distribution: Unrestricted

---

**MAGICSTER**  
Embodied Believable Agents  
IST-1999-10982 Deliverable 4.2.1

February 2002

*The deliverable identification sheet is to be found on the reverse of this page.*

<b>Project rf. no.</b>	IST-1999-29078
<b>Project acronym</b>	MAGICSTER
<b>Project full title</b>	Embodied Believable Agents

<b>Security</b>	Unrestricted
<b>Contractual delivery date</b>	M15 = Feb 2002
<b>Actual date of delivery</b>	February 2002
<b>Deliverable number</b>	4.2.1
<b>Deliverable name</b>	Mapping between MagiCster prototypes and the DIPPER architecture
<b>Type</b>	Restricted
<b>Status &amp; version</b>	Final 1.0
<b>Number of pages</b>	6 (excluding front matter)
<b>Contributing WP</b>	4
<b>WP/Task responsible</b>	EDIN
<b>Other contributors</b>	All partners
<b>Author(s)</b>	Johan Bos, Valeria Carofiglia, Berardina de Carolis, Stephen Crampton, Roberto Grassano, Jarmo Laaksolahti, Thomas Rist, Catherine Pelachaud, Isabella Poggi, Fiorella de Rosis, Markus Schmitt, Mark Steedman, Claus Zinn
<b>EC Project Officer</b>	Jakub Wejchert
<b>Keywords</b>	believable agents, human-machine dialogue, natural language processing
<b>Abstract</b>	This is a complement to D4.2, as requested by the reviewers, covering the proposed relationship between the DIPPER architecture and the planned prototypes.

The partners in MAGICSTER are:

<b>University of Edinburgh ICCS</b>	EDIN
<b>Universit degli Studi di Roma "La Sapienza"</b>	ROMA
<b>Deutsches Forschungszentrum fur Kunstliche Intelligenz</b>	DFKI
<b>Swedish Institute of Computer Science</b>	SICS
<b>Universit degli Studi di Bari</b>	BARI
<b>AvatarMe</b>	AME

For copies of reports, updates on project activities and other MAGICSTER-related information, contact:

The MAGICSTER Project Administrator  
University of Edinburgh  
2 Buccleuch Place  
Edinburgh, Scotland EH8 9LW

Copies of reports and other material can also be accessed via the project's administration homepage, <http://www.ltg.ed.ac.uk/magicster/>

©2002, The Individual Authors

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

## **Mapping between MagiCster Prototypes and the DIPPER Architecture**

Johan Bos, Valeria Carofiglio, Berardina de Carolis, Stephen Crampton, Roberto Grassano, Jarmo Laaksolahti, Thomas Rist, Catherine Pelachaud, Isabella Poggi, Fiorella de Rosis, Markus Schmitt, Mark Steedman, Claus Zinn

---

## Table of contents

<b>1</b>	<b>INTRODUCTION</b> .....	<b>3</b>
<b>2</b>	<b>FACIAL MODEL, AVATAR PLAYER, AND SPEECH SYNTHESISER (ROMA, AME, EDIN)</b> .....	<b>4</b>
<b>3</b>	<b>THE MEDICAL DOMAIN (BARI, ROMA)</b> .....	<b>4</b>
3.1	System Architecture .....	4
3.2	Inputs .....	5
3.3	Discourse Planning and Dialog Management .....	5
3.4	Affective Model: the ‘MIND’ Module .....	5
3.5	Utterance Realiser and Markup: the ‘MIDAS’ Module.....	5
<b>4</b>	<b>THE MEETING NEGOTIATION DOMAIN (AVATAR ARENA: DFKI)</b> .....	<b>6</b>
4.1	Centralised versus Distributed Character Control .....	6
4.2	The Distributed Approach.....	6
4.3	The Centralised Approach.....	6
<b>5</b>	<b>THE INTERACTIVE NARRATIVE DOMAIN (SICS, DFKI)</b> .....	<b>7</b>
5.1	System Architecture .....	7
5.2	Inputs .....	7
5.3	Discourse planning and dialogue management.....	8
5.4	Sentence Realiser and Utterance Realiser.....	8

## 1 Introduction

The overall architecture of the MagiCster prototypes can be visualized as in Figure 1.1.

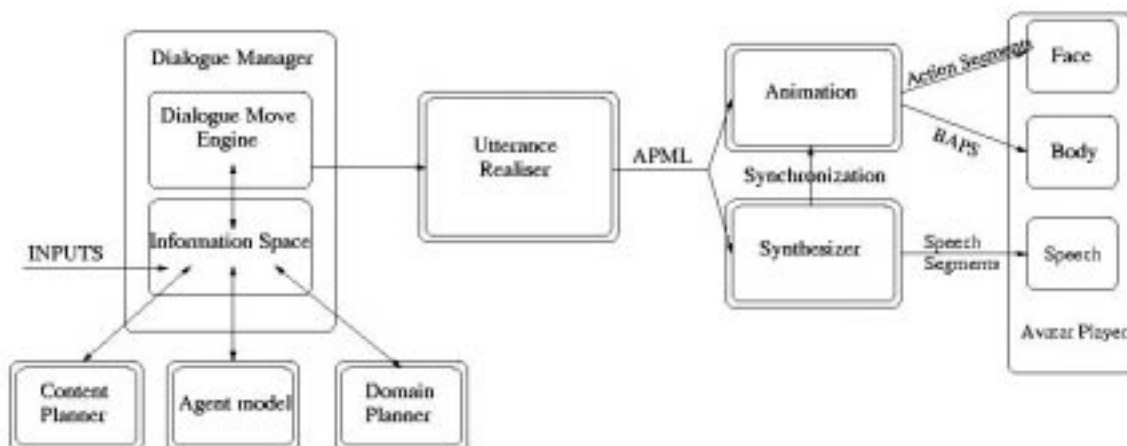


Fig. 1.1: Overall Architecture

This architecture is adapted from the TRINDI architecture developed under EU funding by a consortium including EDIN and is currently in use in other EU projects, notably SIRIDUS. TRINDI contributes the dialogue manager architecture, and the partition between a dialogue move engine, and a blackboard/hub-like Information Space in which the results of other inference engines can be posted. The agent model is composed of an affect component that relates the reaction of the agent to objects, events, persons in the world, a personality definition, and a social model that describes the relationship an agent has with other agents. The agent model is common to all scenarios and is domain-independent. The domain model however, is scenario specific and is different for each of the three scenarios. The dialogue manager takes inputs corresponding to interpreted utterances from other agents and generates logical forms that include information structure and correspond to individual utterances. The latter are passed to the utterance realiser, which generates XML marked-up text of a discourse move, including markers of propositional attitude, affect, and intonation. This marked up text is interpreted by both the face/avatar player and the speech synthesiser for animated spoken delivery.

In order to achieve reusability across MagiCster prototypes and ultimately across other projects, all of these modules are to be realised as OAA modules within the more comprehensive DIPPER system architecture for dialogue systems under development at EDIN (Deliverable 2.2).. Within MagiCster prototypes, modules can and do vary. This technical note is a summary of how the components of each of the three demonstrators map onto this common architecture and the extent to which it is intended that modules be shared.

The intention is that all demonstrators will use the same Animation/Speech-synthesis module developed by ROMA, AME, and EDIN: this is therefore described separately, in section 2.

Section 3 then relates work by BARI and ROMA towards an expressive emotional character in a medical domain to the architecture of Figure 1.1.

Section 4 introduces a framework under development by DFKI for the simulation of negotiation dialogues among avatars which are embedded in a social context. The framework will be illustrated by means of an appointment arrangement scenario.

Section 5 elaborates on interactive narrative as a further domain which under investigation by SICS and DFKI which might be used to demonstrate open multi-party dialogue settings.

---

## **2 Avatar Player and Speech Synthesiser (ROMA, AME, EDIN)**

---

To ensure portability of the body model, it follows the h-anim specification; while the facial model is compliant with MPEG-4 standards. An intermediate animation level representation is needed to ensure full synchrony with AvatarMe's facial model prototype: action segments. This level specifies each segment of action (inner raised right eyebrow, head turn left...) and their time of appearance and disappearance in the final animation. AvatarMe's player interprets the action segments and generates an animation. The action segments may also be translated into FAPs that will drive any MPEG-4 compliant facial player.

To ensure independence between the specification of the facial expressions and the facial models (that is we wish to be able to define facial expressions to be applied to any type of facial models) we define a set of tags using an XML format called APML (Affective Presentation Markup Language). The Facial and Avatar player interprets the APML tagged text spans, while the synthesizer interprets the words and intonational markers as a stream of time stamped segments and a pitch contour. The timings are then reintegrated with the APML markup, and the output of this interpretation is a file with the description of all action segments for the facial animation and a BAP file for the body animation.

More information on this function may be found for the animation player in the Deliverable 1.2 and for the speech synthesizer in Deliverable 2.2.

---

## **3 The Medical Domain (BARI, ROMA)**

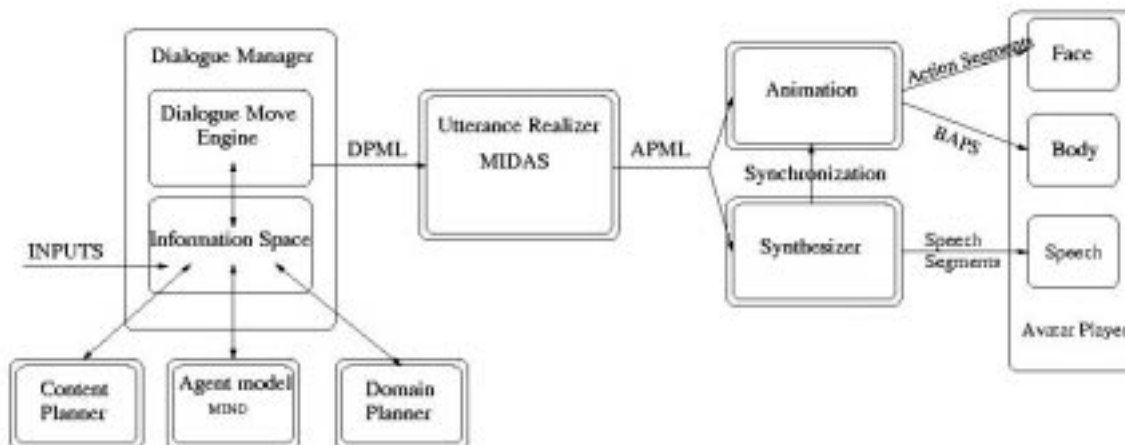
---

After the suggestions in the first Review Meeting, we shifted the focus in this domain towards *a system that provides information and suggestions to teenagers* about lifestyle problems (smoking or eating disorders). We hope that this particular application domain will appear to be more 'believable', because the target users are probably more open to discuss this kind of topics with an animated agent than a patient with a serious disease would be. Several hypermedia systems aimed at coping with this problem may be found on the web, and we are collecting them to build a corpus of data to analyse. In this scenario, we will mainly test 'monologs' (either one-shot or interactive) and persuasion dialogs between the avatar and a user.

### **3.1 System Architecture**

As described in Section 1. In particular, we develop the affective component MIND of the Agent Model (as described in 3.5) and an Utterance Realiser MIDAS (as described in 3.5) that will be used, as well, in other application domains.

---



### 3.2 Inputs

A *communicative goal* (that is translated into a ‘task’ for TRINDI), a *personality* for the Agent (as a combination of personality traits), a *social context* and a *domain model* (represented in TRINDI language)

### 3.3 Discourse Planning and Dialog Management

Discourse planning is done through a hierarchical decomposition of the main communicative goal into subgoals and through ‘recipes’ that are dynamically selected to achieve each subgoal. We are considering how to improve our planner to deal with ‘reactive planning’ (by possibly substituting it with Longbow).

### 3.4 Affective Model: the ‘MIND’ Module

As discussed in Deliverable 3.1, believability has several meanings and interpretations. In Magicster, it will be the result of the combination of several factors: the ability to endow the Agent with a personality and to adapt its behavior to this personality and to the social context is one of them. The MIND Module enables our Agent to think rationally but, at the same time, to ‘feel emotions’ when desirable or undesirable (past, present or prospective) events are evoked during the dialog. This aim is achieved thanks to a model of the affective component of the Agent’s mind, that is implemented through dynamic belief networks, as described in the Deliverable 4.2.

### 3.5 Utterance Realiser and Markup: the ‘MIDAS’ Module

Every dialog move is expressed according to the "Discourse Plan Markup Language" (DPML), an XML-based language that allows specifying RST-based discourse plans. This representation of what the Agent wants to communicate is not sufficient to insure that a consistent agent's behavior is produced, in the realisation phase. For this reason, an interface between the Dialogue Manager and the Speech, Face and Body Generator has been designed. This interface interprets the dialog move and translates it into a combination of verbal communicative acts, enriched with the ‘meanings’ that will have to be expressed through body, face and speech. These meanings will be interpreted by the Speech Synthesiser and the Facial and Avatar Player and will be translated into an appropriate combination of ‘signals’. Meanings are represented in a XML-based mark-up language that we called “Affective Presentation Markup Language” (APML). An algorithm translates the DPML-based tree-structure into APML, through a set of transformation rules that depend on the information attached to nodes in the discourse plan, as described in the Deliverable 2.2.

## 4 The Meeting Negotiation Domain (Avatar Arena: DFKI)

---

### 4.1 Centralised versus Distributed Character Control

*In contrast to the medical domain, the Avatar Arena used for the meeting negotiation domain is a multi-character application. Therefore, two fundamentally approaches to character control can be followed: a centralized approach or a distributed approach. In the following two subsections we will briefly describe and discuss the architectural consequences of each of the two approaches.*

### 4.2 The Distributed Approach

In the distributed approach each agent (avatar) that participates in a negotiation is conceived as an own entity that controls its behaviour and acting. Consequently, a scenario with  $n$  agents could be modeled as  $n$  instantiations of the overall architecture outlined in Fig. 1.1. That is, each agent would be equipped with its own Dialogue Manager, Utterance Realizer, Speech Synthesizer and Facial/Body Animation Engine. Negotiation and dialogue strategies would be defined from an agent-centric perspective. That is, an agent has to sense the world, recognize events and actions / utterances made by other agents, and decide on how to react or act appropriately. From the point of view of knowledge engineering, an agent's strategic negotiation knowledge and also its conversational skills need to be defined from an agent-centric perspective. For instance, each agent might have a strategy to prompt a greeting or to reject a proposal that does not meet its own preferences.

The attractiveness of this modeling approach lies in its generality as from the point of view of an agent, (a) it might make no difference whether a conversational partner is another synthetic agent or a real human, and (b) the number of negotiation partners can be kept variable and even may change during a negotiation. On the other hand, this generality comes at high computational costs and demands for a dialogue manager that can cope with multi-treated, multi-party conversations. Also, similar to un-chaired meetings, negotiation dialogues that emerge from self-controlled agents can easily become unstructured - if not incoherent - unless the agents are equipped with a broad repertoire of sophisticated conversational skills that enable them to "behave well" but at the same time safeguard their interests in a not necessarily cooperative group discussion.

While in the long run the distributed approach is more appealing and also technically more challenging, the first MagiCster demonstrator will be implemented following a centralized control approach.

### 4.3 The Centralised Approach

In contrast, a centralized approach to the control of a multi-agent negotiation scenario can be compared with writing a theater play or movie script involving multiple actors. That is, a single script author is in full control of all agents in the sense that she/he decides on what the single agents are going to utter, how they will react to utterances and actions by others, when to take the initiative etc. Consequently, knowledge about how agents negotiate and converse need to be specified from the point of view of a script-writer (rather than from an agent-centric perspective). An advantage of this approach is that it is computationally less complex. In fact, dialogue management can be kept much simpler because the script author determines all moves in the occurring dialogue games.

In the current implementation of the Avatar Arena we use one single planner for determining the dialogue moves of all agents. However, dynamically changing social relationships among the agents are taken into account when acting out the negotiation process in the form of a dialogue script.

Utterance realization is kept quite simple by using phrase templates in the leaf nodes of the emerging dialogue script. The phrase templates are enriched by APML tags in order to include information about the agents affective and cognitive states that hold while performing particular speech acts.

As to animation and speech synthesis, two different player technologies are used. For the sake of an early test-bed, we simply relied on the MS Agent toolkit, a package which provides a high-level scripting interface for the animation of cartoon-style agents with simple text-to-speech output. However, an interface to the Avatar player has also been implemented and this implementation will be successively refined in accordance with the refinement of the avatar player itself.

## 5 The Interactive Narrative Domain (SICS, DFKI)

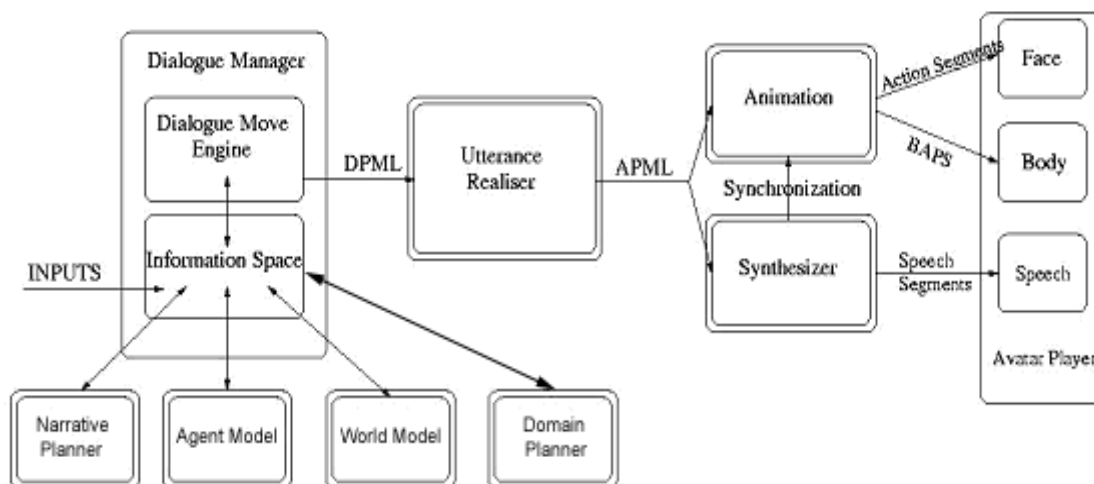


Figure 5.1 Components of the Interactive Narrative scenario

### 5.2 Inputs

How a dialogue is carried out depends on several factors. The *agent model* guides how an individual agent acts and reacts to events in the world. In the narrative scenario this model consists of the following components:

- A model of the agent's emotional state, beliefs, intentions and desires. This corresponds to the MIND module developed at BARI.
- A social model describing interpersonal relationships between agents. This builds on work done at DFKI and SICS.
- A model of attraction between agents which is developed at SICS.

The scenario also requires a *domain model* detailing concepts regarding the above components, e.g., concepts related to parties and/or attraction, and a *domain planner* making sure the dialogue makes sense from a domain perspective. In the above figure these two parts collapse into the domain planner component. We propose to use planners developed at Edinburgh under the DIPPER OAA architecture, such as Longbow, Beetle, and Beer for domain planning. In addition the scenario requires a *world model* holding the contents (and state) of the world (e.g., other agents that are part of the world or artefacts within the world). Finally, from a narrative point of view it is important that the dialogue strives to fulfil some narrative arch. This will be handled by the *narrative planner*, which is being developed at SICS. A brief description can be found in the next section.

### **5.3 Discourse planning and dialogue management**

The third application scenario introduces the concept of narrative as a force guiding the interaction between user and agent(s). Narrative guidance is concerned with the content of interactions rather than their surface realisation. For reasons of simplicity, the narrative guidance module, will explore a *centralised* approach, where one narrative planner guides several agents, although distributed solutions are also possible. To ensure that all agents are working towards some higher-level narrative goal the *narrative planner* monitors all agents' minds and can prioritise or alter goals that are currently pursued by agents. In addition, or alternatively, the planner can also decide to change the world e.g., by introducing a new item or character into the story in response to user actions. We are currently investigating an approach where guidance is provided by an *anticipatory agent system* where a simulation of the narrative world that runs faster than real-time is used to predict the effect of actions within the world.

### **5.4 Sentence Realiser and Utterance Realiser**

The third scenario will use the same mechanisms as the meeting arrangement scenario for aspects of dialogue management such as sentence realisation and utterance realisation. More information on this will follow when the prototype is closer to completion in deliverables D1.5, D2.5 and D3.4.