The Knowledge Engineering Review, Vol. 00:0, 1–13. © 2008, Cambridge University Press
DOI: 10.1017/S000000000000000   Printed in the United Kingdom

A knowledge-rich distributed decision support framework: a case study for brain tumour diagnosis

DAVID DUPPLAW[1], MADALINA CROITORU[1†], SRINANDAN DASMAHAPATRA[1], ALEX GIBB[2], HORACIO GONZÁLEZ–VÉLEZ[3‡], MIGUEL LURGI[5], BO HU[1§], PAUL LEWIS[1], and ANDREW PEET[4]

[1] *Intelligence, Agents, Multimedia Group, School of Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK E-mail: dpd@ecs.soton.ac.uk, croitoru@lirmm.fr, sd@ecs.soton.ac.uk, bo01.hu@sap.com, phl@ecs.soton.ac.uk*
[2] *University of Birmingham, Birmingham B15 2TT, UK E-mail: a.j.gibb@bham.ac.uk*
[3] *Institute for Computing Systems Architecture, School of Informatics, University of Edinburgh, Informatics Forum, Edinburgh EH8 9AB, UK E-mail: h.gonzalez-velez@rgu.ac.uk*
[4] *University of Birmingham and Birmingham Children's Hospital, Birmingham B4 6NH, UK E-mail: acpeet@doctors.org.uk*
[5] *MicroArt, SL. Parc Cientfic de Barcelona C/Baldiri Reixac, 4-6 08028 Barcelona, E-mail: mlurgi@microart.cat*

Abstract

The HealthAgents project aims to provide a decision support system for brain tumour diagnosis using a collaborative network of distributed agents. The goal is that through aggregation of the small datasets available at individual hospitals much better decision support classifiers can be created and made available to the hospitals taking part. In this paper we describe the technicalities of the HealthAgents framework, in particular how the inter-operability of the various agents is managed using semantic web technologies. At the broad-scale the architecture is based around distributed *data-mart* agents that provide ontological access to hospitals' underlying data that has been anonymised and processed from proprietary formats into a canonical format. Classifier producers have agents that gather the global data from participating hospitals such that classifiers can be created and deployed as agents. The design at a micro-scale has each agent built upon a generic layered-framework that provides the common agent program code allowing rapid development of agents for the system. We believe our framework provides a well-engineered, agent-based approach to data-sharing in a medical context. It can provide a better basis on which to investigate the effectiveness of new classification techniques for brain tumour diagnosis.

1 Introduction

Brain tumours present a particular challenge to healthcare systems and medical research. Whilst they occur at all ages, they are a particularly common form of cancer in children and young adults. The young age of the patients, the comparatively high mortality and the burden of associated neuro-disability ensures that these tumours assume an importance which exceeds their overall incidence. Improved methods for diagnosis, treatment and understanding of brain tumours are an important goal of clinical and biomedical research.

Magnetic resonance imaging (MRI) is the standard investigation performed at presentation and allows the presence of the lesion to be confirmed and its location, size and structure determined.

[†] Madalina Croitoru is now at LIRMM, 161 rue ADA, F34392 Montpellier Cedex 5, Montpellier, France
[‡] Horacio González–Vélez is now at The Robert Gordon University, St Andrew Street, Aberdeen AB25 1HG, UK
[§] Bo Hu is now at SAP Research CEC Belfast, BT37 0QB UK

However, there are many questions which conventional MRI cannot reliably answer. In particular, the tumour type and grade is often poorly determined and the diagnosis is subsequently obtained by removing a sample of the tumour at operation and performing histopathology. However, surgery has an associated risk for the patient and improving the diagnostic and prognostic information available from non-invasive methods is an important objective.

Imaging research is currently focussing on techniques which give information on the biological properties of tumours. One of the most promising of these is magnetic resonance spectroscopy (MRS), a method which measures levels of various chemicals in tissue. Previous studies have shown MRS profiles to be a powerful characteristic of brain tumours aiding non-invasive diagnosis and providing novel biomarkers of prognosis. However, the method has not met with universal acceptance amongst clinicians. One of the barriers to the widespread clinical use of MRS lies in the difficulties of interpreting MRS data.

Radiologists are experts in interpreting images and they do this mainly by visual inspection. MRS data is difficult to interpret by visual inspection and the most accurate method of analysis lies in pattern recognition of the spectra. Making such techniques available in a clinical environment is challenging. Sophisticated software is required to present the results in a manner acceptable to clinicians working in a busy clinical environment. Furthermore, large numbers of cases in each diagnostic category are required to build robust pattern recognition classifiers and this necessitates the collection of data from multiple centres.

In HealthAgents (Arús et al., 2006; González-Vélez et al., 2009) we have created an agent-based architecture that allows the sharing of anonymised MRS patient data providing a basis on which to create classifiers that are able to produce better results than those that are limited to a hospital's local data. By providing better diagnosis of tumours based on non-invasive techniques, the hope is that prognosis and treatment can be better managed.

The nine partners of the HealthAgents project are MicroArt, Universitat de València, Universitat Autònoma de Barcelona (UAB), Institute for the Applications of Advanced Information and Communication Technologies (ITACA), Pharma Quality Europe (PQE), Katholieke Universiteit Leuven, University of Birmingham, University of Edinburgh, and University of Southampton. Being a multi-national, multi-centre project, the HealthAgents architecture is required to be platform-independent and scalable by design. It has therefore been devised so that any new functionality can be easily integrated into the general context by implementation of new agents.

There are many other systems designed for distributed storage and sharing of data, but most would not be able to be deployed in a medical context. Indeed, it is absolutely vital that patient data is kept secure at all times and the HealthAgents system has been designed such that all of the data-sets involved can be anonymised; that is, data that contains no direct link to the patient's identity. The architecture also provides various security enhancements to avoid the unauthorised access of data.

In the following sections we describe our contribution to this area of research. First, we motivate our research approach in section 2 complemented with the user requirements in section 3, and the layered, agent-based architecture—that is the basis of the framework—, which is delineated in section 4. Then, we properly describe the HealthAgents framework in section 5, and finalise by providing some evaluative comments and concluding remarks in sections 6 and 7 respectively. There is a section defining our nomenclature at the end of the paper.

2 Motivation

Widely conceived as parameterisable concurrent constructs, software agents possess four intrinsic characteristics (Brugali and Sycara, 2000):

- an intrinsic knowledge-based state that can evolve dynamically;
- changing reasoning capabilities that determine their internal behaviour through constraints or goals;

- a communication status that allows them to interact with other agents or human entities; and,
- a unique identity with roaming and service-advertising capacities.

In this vein, agent-based application frameworks follow a defined lifespan, where the architectural choices are first formulated with platform and domain-knowledge independence, and elemental ones—implementing concurrency and logic mechanisms—are then devised to determine implementation particulars (Brugali et al., 1997).

On the one hand, agent frameworks for medical purposes have been extensively documented in the literature (Annicchiarico et al., 2008; Merelli et al., 2007), but its use has been traditionally confined to patient control and information (Huang et al., 1995).

Lanzola et al. (1999) deploy a prototype of hospital information system, mainly oriented towards the management of patients with leukemia, using an agents-based model. While the architecture falls short of showing multi-centre interaction or classification capabilities, its single framework approach approach is per se interesting.

The Artemis Project (Boniface et al., 2005) has previously dealt in a medical context with the transfer of critical patient data (mainly clinical data). It used a peer-to-peer-like architecture and used web services security protocols for the transfer of patient records. Like HealthAgents, the project has investigated ontological means of data representation.

The MIAKT (Medical Images and Advanced Knowledge Technologies) project designed and deployed a knowledge-based distributed decision support system to support the diagnosis and management of breast cancer cases using ontologies and web services. The framework developed there was capable of being used in different applications (Shadbolt et al., 2004).

The Intensive Care Agent Platform (ICAP) concentrates on the intensive care side on a single-node (intra-hospital) basis (Turck et al., 2007), providing little evidence of dynamic distributed multi-centre interaction.

On the other hand, from a neuro-oncological point of view, despite the advances in the application of MRI and MRS techniques for diagnosis (Rees, 2003), there exist a few research endeavours which report successful automated diagnosis of brain malignancies.

Having created a single-node decision support system using MRS classification supported on histopathological diagnosis, the Interpret project (Tate et al., 2006) has centred its efforts on the clinical aspects of the diagnosis and reported successful results for a handful of common malignancies.

Georgiadis et al. (2008) report the successful discrimination between metastatic and primary brain tumours using MRI. They employ distinct machine learning classifiers on a static patient record repository, but provide no evidence on the online classification via multiple repositories.

The eTumour project implements stringent quality control mechanisms for patient records (Wright et al., 2008) but has reported few advances from a generic software architecture standpoint.

That is to say, scant research has been devoted to deal with cancer and neuro-degenerative disease diagnosis using a generic distributed agent-based framework for the secure exchange and classification of MRS patient data. HealthAgents expands previous approaches to computer-aided brain tumour diagnosis with a distributed multi-centre agent architecture, an in-vivo classification method with negotiation, an additional number of cases located in different centres across Europe, and a web-based user interface. By working with hospitals and ensuring the security of the MRS data passing through the system, HealthAgents is being deployed in clinical environments.

3 User requirements

The diagnosis and management of brain cancer currently depends on histological examination of a brain biopsy and the optimum treatment varies with the tumour type and with its grade, i.e. its degree of malignancy. Both histology and grade have to be assessed using established but partly subjective criteria by a skilled neuropathologist.

Stereotactic brain biopsy has significant risks, and, moreover, histology is not accurate enough in some cases although still it is considered the gold standard methodology. Current radiological methods do not distinguish adequately between the large number (nearly one hundred) of recognised types of brain tumours and thus it is necessary to perform histopathological diagnosis on a biopsy.

Normally, most patients undergo surgery for cytoreduction and decompression and the resulting biopsy is available for analysis. However, there are a few cases where surgical resection or biopsy would not be considered, such as in very elderly and infirm patients with obviously malignant lesions, or patients with very slow-growing tumours in vital parts of the brain. In addition, there are certain pathologies (lymphomas and brain abscesses) in which pre-operative diagnostic certainty would avoid the open surgery step. An additional non-invasive method for accurately diagnosing and grading brain tumours would be a major advance in those cases.

There is a need to improve brain tumour classification, and to provide non-invasive methods for brain tumour diagnosis and prognosis, to aid patient management and treatment. As a consequence, HealthAgents employs two sources of biological information that allow improving the non-invasive typing and grading of brain tumours:

- *In Vivo* techniques: Magnetic Resonance Spectroscopy (MRS) is a non-invasive technique that provides biochemical information of tissue in vivo. MRS, coupled with conventional MRI, provides metabolite profiles of a single voxel or volume (SV) of tumour tissue. It also produces a molecular image of particular tumour metabolites in 10 minutes using multi-voxel (MV) techniques.
- *Ex Vivo* and *In Vitro* techniques: High-Resolution Magic Angle Spinning (HR-MAS) is a modality of MRS that is applied to biopsies in vitro in order to improve characterisation. DNA microarray analysis can be used to determine tumour phenotype from gene expression profiles.

The HealthAgents network is a set of sites interconnected to share resources. The main sites are located at Birmingham (UK), Barcelona (Spain) and Valencia (Spain). Birmingham is aggregating data from 50 different contributing centres while Barcelona and Valencia are providing data from 6 and 4 hospitals respectively. The Barcelona node is collecting brain tumour cases mainly from the validated database from the Interpret project. Globally, the different databases will comprise some 600 cases upon project completion. All these cases shall be quality checked and verified.

Two main decision-support functionalities are offered through the HealthAgents network. The groups carrying out pattern recognition, ITACA and The University of Leuven, are producing tumour classifiers. The databases of the contributing centres is updating all along the project and new cases are being stored into them. Then, potentially, with any new set of cases we could generate a new classifier or improve the available ones. On the other hand, the HealthAgents network offers the evidence-based Search Service (ebSS). The University of Edinburgh is the partner in charge of developing a data mining tool which complements the use of classifiers through a more general approach to seeking out and gathering information, especially textual evidence from the on-line literature and patient information.

In spite of the advantages inherent to a centralised system, a distributed system implementation is proposed, considering these following general aspects, which are sensibly better in a distributed system:

- Confidentiality. The Clinical Record Information of each centre is private and its confidentiality must be guaranteed. The centre should maintain the control of its information diffusion. In some cases they may decide to share this information to every HealthAgents sites (all or parts of it) but not in other cases and prefer to keep it private. In a centralised implementation, the entire Clinical Record of contributing cases should be made accessible.

- Robustness. Usually, in a distributed system, when a node is temporarily out of service, the global system would continue working and the other remaining nodes would be able to benefit from its use. Often, in a centralised implementation, if the server is down, all local sites must wait for the solution of the problem, being unable to benefit from the system.
- Speed. Since each contributing centre has its own copy of a classifier, the answer in the local predictions of d-DSS is sped up for each new case. This distributed architecture allows two main kinds of classifiers: those obtained from every public data case, and other bespoke, specific classifiers. A specific classifier is generated upon site demand and is trained not only with all the available public data, but also with the private data of the requesting site. The HealthAgents system has been built as a distributed network of nodes, where each contributing centre manages its own local site, with autonomous capacity to classify new tumour cases according to both, global classifiers and specific classifiers.

4 The Architecture

Figure 1 shows the logical architecture of the system's data marts. All sites ('nodes' in HealthAgents parlance) are protected by firewalls and their HealthAgents connections are available in 'de-militarised zones' (DMZ) outside of these firewalls where agents are free to communicate. The databases behind the firewalls are not anonymised whereas databases in the DMZ are either link-anonymised or fully anonymised depending on local laws and norms. This is what necessitates the anonymisation process when porting data to the DMZ.

Agent technology provides access to the various functionalities a particular machine on the network can provide. For example, *data-mart agents* provide access to a database and *classifier agents* provide access to classification tools. By distributing the data and functionality of the system the system becomes more robust to failure of any one particular machine or node. To allow for agent discovery, a cooperative network of *yellow-pages agents* is used that allows agents to submit queries based on a description of a remote agent's functionality. Classification is the main functionality provided by the network and *classifier training agents* within the network are able to train classifiers from data sourced from the network matching specific requirements. These trained classifiers can then be shared on the network as *classifier agents* that provide classification for new data combinations.

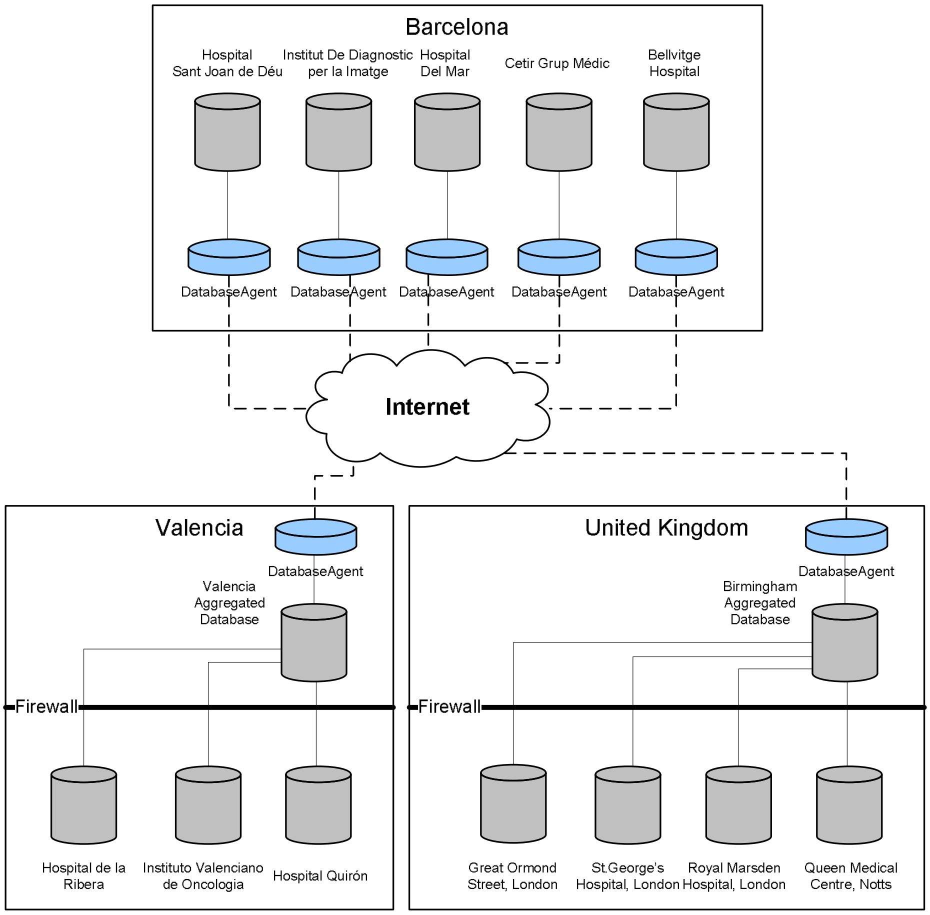
The agent network leverages Semantic Web technologies to provide inter-operability between the various agents by using ontologies. Here we refer to an ontology as "an explicit specification of a conceptualisation" (Gruber, 1993); that is, a well-defined vocabulary that explicitly defines what certain concepts mean and their relationship with other concepts. The agent network has an ontology for its own use and allows the use of any other ontologies for definition of data items.

The HealthAgents Language (or HAL) is the definition of the language that the agents use to communicate and it is defined in an ontology using the Resource Description Framework Schema (RDFS). The HAL ontology defines the messages and their properties that are used in agent communication and it is dedicated to this task. For example, it contains definitions of the message to submit a query to a yellow-pages agent and a definition of the expected result message. Figure 2 shows the top-level concepts in the HAL ontology.

Data-level constructs can be defined by importing ontologies into the HAL messages. To allow the communication of data relating specifically to brain-tumour diagnosis we have developed a domain ontology describing brain-tumour-related concepts. This nomenclature is called the HA-dom (HealthAgents Domain) ontology, and is an amalgamation of various sources, including the World Health Organisation's (WHO) brain tumour classification (Weiss, 1994).

Because data marts on the system are operated by hospitals and sites with relational database expertise, we have designed the data-marts to operate via an ontological mapping between HA-dom and a relational database using D2RQ (Bizer, 2006). This allows the agent system to fully

Figure 1 The logical architecture of the HealthAgents system, showing the various countries and hospital data marts involved. Data marts in the hospitals run entirely outside of the hospitals' own networks. The UK and Valencian data marts aggregate data from multiple hospitals into a single database, unlike the Barcelona hospitals who each provide their own data mart. The data marts are made available on the agent network through the database agents.

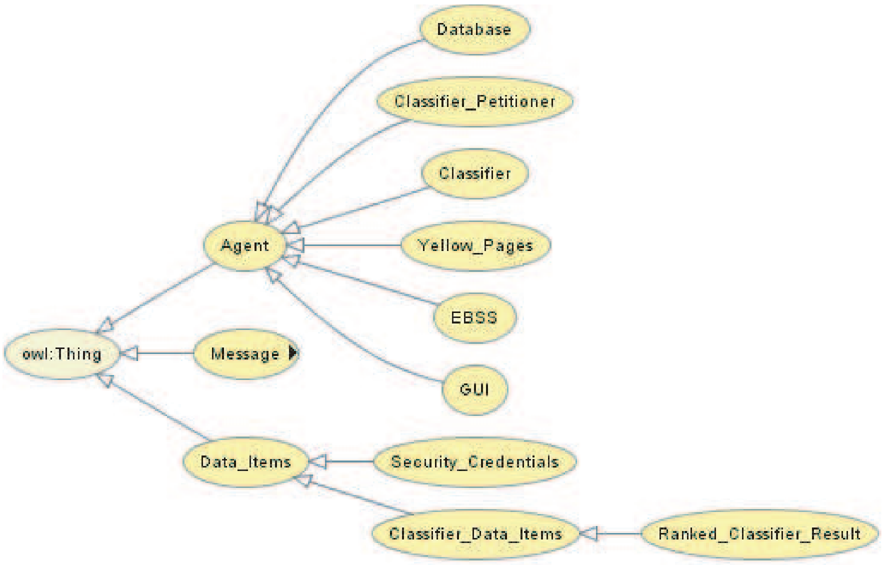


utilise the flexibility of semantic web technologies while also allowing the data sites to utilise their existing human expertise for relational database management. By designing the agent system to utilise semantic web querying mechanisms (such as Resource Description Query Language [RDQL] or Simple Protocol and RDF Query Language [SPARQL]) we build in maximum flexibility for integration of different functionality as the network grows as well as providing the ability to run more advanced reasoning over the data.

The organisation of the network is such that each node has its own yellow-pages agent. This allows the agents on the local platform to find other local-platform agents with only a local platform search to the local yellow pages. A network of yellow-pages agents is created that share data they contain. After a while every yellow-pages agent knows about all other agents in the network, minimising the search delay. More about this organisation is described in section 5.1.

It is important to emphasise that a fully-functional prototype of the HealthAgents decision support system has been fully deployed, and has produced clinically-relevant classification results (González-Vélez et al., 2009).

Figure 2 The top-level concepts in the HAL ontology.



5 Framework

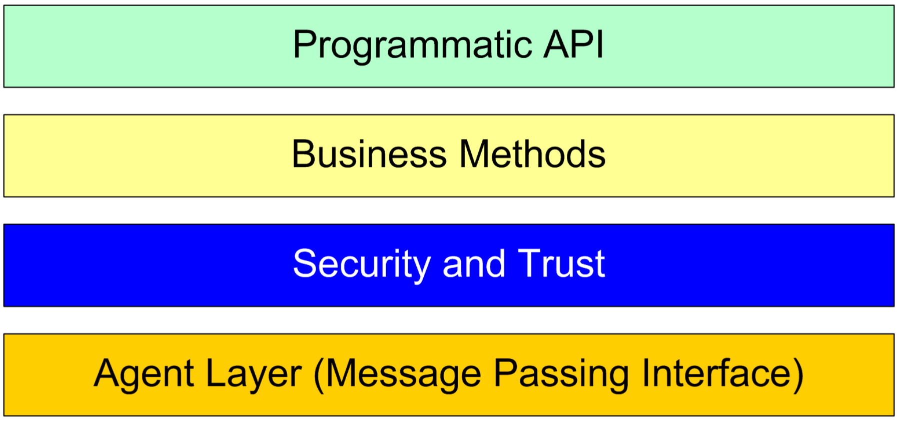
The HealthAgents framework provides the implementation of the agent framework on which the HealthAgents architecture is deployed. The whole design rationale behind the agent framework was to provide a means for integration of new functionalities into the network with the minimum of programming effort. This allows less-technical centres to provide agents for their specialised functionality, as well as minimising the development time of new functionalities. The agent framework has also been built to be agnostic to the specific application that runs upon it, so although it is designed and built with the application of brain-tumour decision support at the forefront, it is applicable in any domain, medical or other. In fact, the ease with which agents are created means it would be a good candidate for the basis of other projects requiring distributed networks.

In this section we describe the micro-level design of an agent in the network. All agents in the network use the same design that is based on a layered design pattern.

The general pattern is shown in Figure 3. The pattern enforces incoming messages to filter up through the various layers to get to the top where the agent's functionality is provided. So from the bottom to the top, the layers are:

- **Messaging Interface**: This is the network-level interface that is used to gather incoming messages and pass them on to the security layer. This layer is multi-threaded so that the agent can receive many incoming messages. This layer is also used to deliver messages onto the network. The messaging layer is also responsible for the parsing of incoming messages into a canonical format that the other layers in the agent can read.
- **Security Layer**: This layer provides a filter for incoming messages that ensures that they are suitable for processing at the agent. This layer can include certification checks, message typing checks and other simple filters.
- **Business Methods**: This layer is provided as part of the agent implementation and controls the flow of authenticated messages from the lower levels into the agent functionality. It deals with re-routing of messages that are identified as being intermediate messages in an ongoing conversation and handles framework-level messaging such as pings and pongs.

Figure 3 The basic layer structure of an agent.



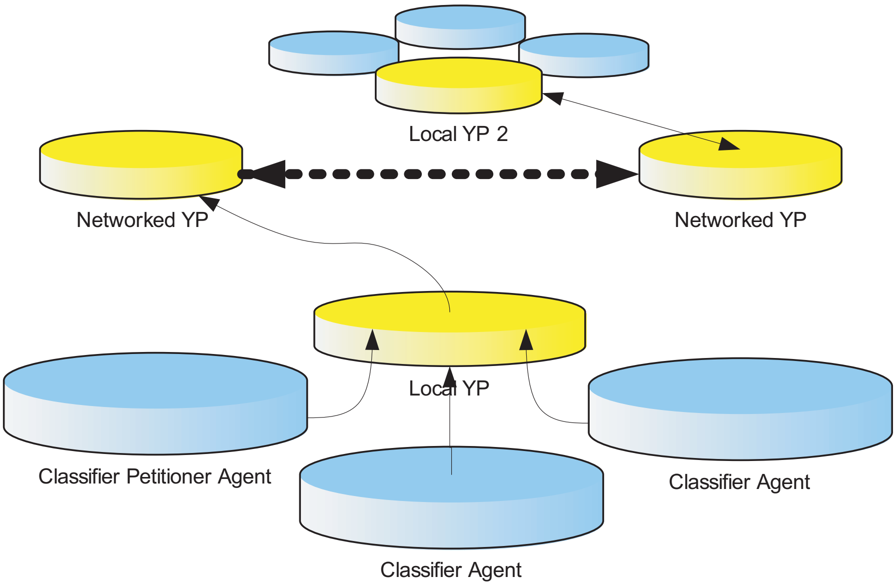
- **Programming Interface (API)**: This layer is provided as a simplified application programming interface (API) to which a developer can write their own agent functionality. This layer provides the functionality of the specific agents in the HealthAgents network - the data-marts, the classifiers, the yellow-pages, etc.

The messaging layer is made distinct from the general business methods of the framework, abstracted through a defined API, such that any message passing system can be employed. All application code is programmed on top of the HealthAgents framework allowing this to be migrated to any agent platform by re-implementing its agent (message passing) layer. For the HealthAgents project it was decided that the implementation of the agent layer would be provided by the Java Agent Development Environment (JADE), a ready-made, open-source, agent platform. The messaging layer API implementation is provided through a class that, in this case, uses the JADE agent platform for sending basic message that conform to the agent control language (ACL) defined by the Foundation for Intelligent Physical Agents (FIPA) (Bellifemine et al., 2001). The messaging layer is also responsible for ensuring secure passage of messages from one agent to another. The implementation we have provided uses a standard public-key methodology for encryption of messages that provides confidentiality (cannot be snooped upon), integrity (cannot be changed en route), identification (through digital signatures) and and non-repudiation (through digital signature verification).

The use of HAL for conversation follows the same philosophy: the abstraction of the communication language of the HealthAgents agents away from the agent platform used. Clearly applying abstractions on top of agent platforms minimises any extra reasoning we may automatically gain from the underlying agent platform, but economically, it makes good business sense to minimise dependency on external libraries. That said, for maximum flexibility the HealthAgents platform defines an API that allows any languages to be used for communication. An implementation of this API provides a parser for HAL utilising Sesame (Broekstra et al., 2002), allowing messages to be passed around the system using RDF (Beckett, 2004) (as synthesised in TURTLE (Beckett, 2007)). Using semantic web technologies in this way makes the easier to surmount the challenges of interoperability between the agents and inter-operability between the wide variety of data sources using different schemas. As the vocabulary that the agents understand gets larger the semantic web technologies provide the extensibility through reasoning over ontologies.

Security is a crucial part of the system and each message arriving at an agent has to be authenticated by the "security guard" that resides in the security layer. Again, this is realised as an API that can be implemented in different ways depending on the necessary level of security,

Figure 4 Network of Yellow-Pages Agents.



the type of the agent, or the location. This interface provides independence of the agent platform's secure passage functionality which may or may not be used. We are implementing a security layer that uses digital certificates to securely identify the route by which messages reached the agent and, using policy rules, allow or deny access to the incoming message.

Inter-operability is one of the challenges that needed to be addressed when constructing our distributed system. More specifically, communication can only be established if every agent can understand the languages used by others. Such a mutual understanding can be achieved by either forcing global consensus and regulating every participant to comply with it or recommending a reference vocabulary and requiring individuals to map their local vocabulary against the reference one. In HealthAgents we have developed HADOM for inter-operation at a data level, but hospitals are free to use their own schemas internally. A wide variety of databases exist within the HealthAgents consortium and having a shared conceptualisation of the domain in an ontology allows these databases' schemas to be mapped into the domain vocabulary to provide a consistent view into them from the agents' point of view.

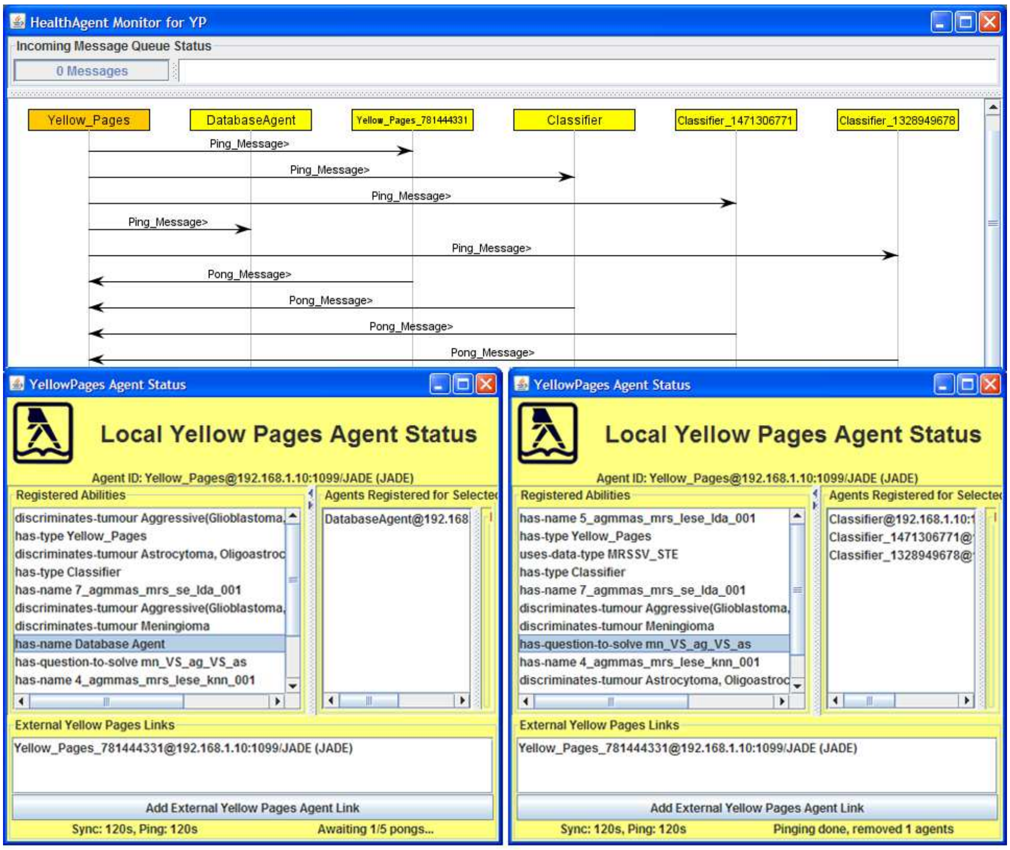
5.1 Directory Services

For agents to be able to communicate with each other they need to find the location of each other. Agents are identified using agent identifiers that are specific to the underlying message passing platform. The framework abstracts these into an API such that agent identifiers can be communicated without resort to understanding underlying agent identifiers.

The yellow pages agent is a directory of agent identifiers. These agent identifiers are indexed by the abilities that the agents have advertised themselves as having. The abilities are defined in the HAL ontology and also by the data-level ontology. When agents start up, they advertise their services to their local yellow pages which stores their record and flags it as authoritative. An authoritative record is the record created at the yellow-pages to which an agent originally registered.

The yellow-pages agents in the system then form a network that share data in a similar manner to the internet DNS system. The local yellow-pages agent is configured such that it has the agent identifier of one of the yellow-pages agents that are outside of the hospital nodes creating a

Figure 5 Message diagram displayed in real-time during the execution of the system showing the yellow-pages directory currency checking messages. The example shows 5 pings being sent and only 4 pongs being returned. The status screen of the yellow-pages agent is shown during and after the ping-pong mechanism to show that the database agent that did not respond has been removed from the directory (the message in the bottom right of the status screen also indicated the removal of one agent).



Agent ID: Yellow_Pages@192.168.1.10:1099/JADE (JADE)

discriminates-tumour Aggressive(Glioblastoma,
has-type Yellow_Pages
discriminates-tumour Astrocytoma, Oligoastroc
has-type Classifier
has-name 7_agmmas_mrs_se_lda_001
discriminates-tumour Aggressive(Glioblastoma,
discriminates-tumour Meningioma
has-name Database Agent
has-question-to-solve mn_VS_ag_VS_as
has-name 4_agmmas_mrs_lese_knn_001

DatabaseAgent@192.168

Yellow_Pages_781444331@192.168.1.10:1099/JADE (JADE)

Sync: 120s, Ping: 120s Awaiting 1/5 pongs...

Agent ID: Yellow_Pages@192.168.1.10:1099/JADE (JADE)

has-name 5_agmmas_mrs_lese_lda_001
has-type Yellow_Pages
uses-data-type MRSSV_STE
has-type Classifier
has-name 7_agmmas_mrs_se_lda_001
discriminates-tumour Aggressive(Glioblastoma,
discriminates-tumour Meningioma
has-question-to-solve mn_VS_ag_VS_as
has-name 4_agmmas_mrs_lese_knn_001
discriminates-tumour Astrocytoma, Oligoastroc

Classifier@192.168.1.10:1
Classifier_1471306771@
Classifier_1328949678@

Yellow_Pages_781444331@192.168.1.10:1099/JADE (JADE)

Sync: 120s, Ping: 120s Pinging done, removed 1 agents

persistent network of directory services, as shown in Figure 4. The local yellow-pages regularly send its directory to the other yellow-pages it knows of. These receive the data and store the entries as non-authoritative. That way it is always known that non-authoritative data may be out of date, as the currency of the directory depends on the period of synchronisation configured in the agents. The networked yellow pages agents then pass on their aggregated data to other yellow-pages agents it knows of until the network of yellow-pages are all synchronised.

The currency of any yellow-pages directory is tested regularly by pinging all of the agents in the directory. Agents which are still available respond (with 'pong' messages) and those that do not respond within a given timeout period are removed from the yellow-pages directory (see Figure 5). The response to pings is handled by the business methods layer of the agent design, and is thus handled automatically, without any intervention required of an agent developer.

In the event that a local yellow-pages agent has recently been started and has not yet received a data synchronisation message from a networked yellow-pages agent, it may be unable to answer queries from its local agents. In this case, the local yellow-pages agent forwards the search query on to one or more networked yellow-pages agents then aggregate and cache the results.

The yellow pages is an integral part of the agent framework, i.e., it does not run without it. However, this dependency meant that it was possible to make the system easier to program. The abstract classes of the framework have built-in methods for searching the yellow pages. The code snippet below shows the Java for searching for a classifier petitioner agent. A set is created containing the abilities of the agent for which we are searching (the ability is that it has the type ClassifierPetitionerAgent, as defined by HAL) and the *searchYellowPages* method does all the work of creating the appropriate messages and dealing with the waiting for the asynchronous result from the network.

```
HashSet<AgentAbility> s2 = new HashSet<AgentAbility>();
s2.add( new SpecificAgentAbility(
      new Concept(Ontology.AA_HAS_TYPE),
      new Concept(Ontology.CLASSIFIER_PETITIONER_AGENT) ) );
Set<AgentIdentifier> petitioners =
      getMessagingService().searchYellowPages( s2 );
```

5.2 Database communication

The HealthAgents Domain Ontology (HADOM) is used as the common referencing point among different hospitals maintaining their own vocabularies and database schemas. The integrity and independence of legacy databases are fully respected. The discrepancy between such schemas is resolved by dedicated mappings between each individual schema and the HADOM ontology. By referring to (e.g. with rdf:seeAlso) or making HADOM concepts equivalent to (e.g. with owl:sameAs) those from the external resources, any new advances and updates in the external resources can be reflected in HADOM. Thus far, HADOM is based on publicly available documents and information gathered from various sources such as the WHO classification of tumours affecting the central nervous system (Tatter, 1996; Weiss, 1994) and other relevant sources (DeAngelis, 2001).

These documents are based on universal knowledge of the domain and may not be strictly applicable in HealthAgents or not sufficient enough to reflect all the necessary project specific domain restrictions. Its use as a mediator in the interoperability of agents allows integration of functionalities into the system more easily than might otherwise have been in a purely relational database system. There is a remote agent in the prototype that acts as a proxy to the database.

All data-related information in the messages in the HealthAgents network is referenced to the HADOM ontology. A database agent transforms these messages into Structured Query Language (SQL) queries executed against a relational database.

6 Evaluation

The deployment of the agent framework to create a European-wide decision support system for brain tumour diagnosis is still underway, but the core functionality is now being tested with the data assembled by the data-collecting sites during the development of the agent framework.

USOU and MicroArt are acting as a small 2-node network of yellow-pages agents such that all joining nodes may eventually determine the location of all agents on the network.

The graphical user interface that has been developed at MicroArt for providing access to the agent framework, the data and all the classifiers, has been deployed at hospitals in the UK and Spain.

So far, all-but-one of the hospitals are running data-marts and the classifier producers at ITACA have deployed classifiers and an agent specifically for gathering data for producing new classifiers.

Tests are in the early stages, but it has been shown that data from one hospital can be classified by a classifier in a distant node, when that hospital has no previous knowledge of the classifier node, due to the network of yellow-pages. We have also shown that, with appropriate permissions, the classifier producers can gather data from data-marts as and when they require it for creating new classifiers.

6.1 Implementation

At the time of writing, the HealthAgents system is being deployed in various sites around Europe. Hospitals in Birmingham, United Kingdom, Barcelona, Spain and Valencia, Spain have installed the data-mart node and are gathering MRS data to be used for creating classifiers. Classifier nodes in ITACA and KULeuven are online providing pre-built classifiers that were trained on the data from the Interpret project. Such large-scale aggregation of data is not only novel, but somewhat unusual in the field of medicine.

However, the distributed nature of the aggregation means that the entry-costs for joining a system are very low and the amount of 'buy-in' required is very small; that is, hospitals, classification producers and other tool providers can leave and join as they see fit (within some specified bounds). Indeed, the consortium has had much interest in the system from parties external to the project, and the Province Hospital of Jiangsu has already agreed to join the network and provide MRS data to the distributed data-mart. As previously described, other decision support systems, based on MRS data do not have the flexibility of the HealthAgents system.

The agent-based distribution of the network provides low-entry requirements while also providing the flexibility to reconfigure networks as necessary, improving network robustness. The Interpret system was not based on semantic web technologies and the use of a domain ontology in HealthAgents, for formally specifying the data, provides a well-defined description of the data that eases inter-operability issues, particularly when hospitals who already have their own data schema join the network. The use of the ontology to define the data also provides future-proofing for the data schema. During the project, the World Health Organisation released a new version of their vocabulary and it was because the technology of ontologies provides great extensibility that we were able to extend the vocabulary to cover the new terms.

We have not provided any quantitative evaluation of the agent framework's speed; this is a considered decision. Although we could measure the performance of the agent network for answering queries across the system, speed variances caused by the network reliability when using a European-wide distributed network would not show anything indicative of the performance of the framework itself. The main speed bottleneck of any particular agent is the parsing of an RDF message into objects. As we are using Sesame for this task, the performance of an agent's parsing routine depends on the performance of Sesame.

Nonetheless, to give an idea of the performance of the network:

> *classifications are usually returned within 10 seconds.* Most of that time is network-related. On a single desktop-specification machine, a classification request and reply (via the Classifier Petitioner Agent) happens in under 2 seconds. We have also found that the message encryption routines impart a large overhead (around 100%) on communication performance.

For a qualitative clinical evaluation, the HealthAgents consortium are conducting a Technological Acceptance Model (TAM) study, which strives to gauge the acceptance of the system from the point of view of the clinicians that use the system.

6.2 Discussion

The design of the agent architecture was primarily based around the need to provide aggregation of data from many different disparate locations and institutions. The technological know-how of these institutions may be limited, so the architecture was designed in such a way as to minimise the necessary familiarity with agent-based and distributed systems. The partners in our project that were building agents based on the framework have commented that the framework is easy to use and this is borne out by the large number of distinct agents that have been created during the project by non-agent experts. Debugging of truly distributed systems such as HealthAgents is very difficult and we encountered some challenges when deploying the system Europe-wide. We built some debugging tools (such as the message diagram and the yellow pages agent status screen shown in Figure 5) that really helped to diagnose some problems and generally we found agents performed as expected.

However, there are still a few small areas in the framework where a little tittivating could improve the overall simplicity of its use. In particular, the creation of the HAL ontology was effectively a distributed task, as programmers in the project found they required new predicates by which they could refer to values. In general, HAL was built in a relatively ad-hoc manner. That's not to say it is unworthy, but to say that should further programmers require new ways to refer to data structures, they will be required to understand at least the basics of semantic web technologies. However, there doesn't seem to be an easy way around requiring at least a basic understanding of RDF when all the communication is achieved using it. RDF is also the main cause for concern in the performance of the communications. The parsing of the messages into the Sesame RDF store (which also calculates all the inference) is one of the main bottlenecks of the framework. Of course, its performance is proportional to the speed of the machine on which it executes, so throwing computing power at the problem can ease the problem. We have also designed the language centre of the framework to be a plugin so should new ways of dealing with graphical data become available or industrialisation of the framework require that the communication language be made non-graphical, either can be easily achieved through implementation of one API. Of course, using RDF plays two-ways – its extensibility means that at least it is possible for new programmers to easily extend the language beyond its original scope; something which is much more difficult to achieve with relational and functional means of creating language schemas.

Another way we hope to improve the agent framework is through agent workflows. Currently, the workflow of an agent must be handled explicitly by the programmer of the agent by arranging their code such that it performs the appropriate workflow. This is not a trivial task for some agents, for example where certain messages must be handled prior to other messages, and it makes maintenance of agent code much more difficult. We researched agent workflow systems and built a module for controlling an agent's messaging agenda using the agent workflow language called Lightweight Coordination Calculus (LCC) (Robertson, 2004). Because LCC is a very simple-to-understand language the building of agents would become much easier for the programmers. Indeed, it would be possible for simple agents to be created by non-programmers. Conversion programs exist for converting other agent workflow languages to LCC so in that way we encompass a wider gamut of the workflow community. Agents following LCC workflows were tested in controlled situations and correctly followed the appropriate workflow. In the test scenario, the framework was extended to make these workflows as simple to use as possible. Unfortunately, due to time constraints the workflow language was not used in the final HealthAgents system, however we still hope to distribute this as part of the framework codebase and perhaps in the future the HealthAgents communication system will move to a workflow-based execution environment.

7 Conclusions

The fully-distributed agent-based design for the data-marts is novel in the field of computer-based decision support. It is important to realise that previous brain tumour decision support systems have been centralised making them limited in both the range of data available to them and their extensibility. So, because the HealthAgents system is distributed, new data that becomes available on the network will help to increase the accuracy of the decision support, thereby making the system more valuable as the community builds. Also, the layered approach of the agent framework makes the development of new agents very easy and therefore the possibilities for extension of the system become greater than other decision support systems. We have shown that the agent design is highly extensible which means that agents beyond the remit of the project can be deployed within the system easily and quickly which could lead to a collaborative network of agents for many different medical decision support scenarios.

The use of semantic web technologies has been an important factor in the design of the system such that the messages and the data payload of the messages is formally represented. With the increasing uptake of RDF, the system stands in good stead for interoperability with new and future systems. HADOM provides an important role in linking the nomenclature of hospitals across Europe into a formalised vocabulary that can be used to map schemas of existing patient record systems in hospitals into the system. HAL provides an extensible, hierarchical language by which agents can communicate with each other. Together HADOM and HAL provide a novel way in which decision support data can be viewed and show that the extensibility of semantic web technologies makes them ideal for distributed systems and distributed development environments.

The aim of the HealthAgents project is to build a distributed decision support system for brain tumour diagnosis and in this paper we have presented an extensible agent architecture that uses research-led methods to provide this novel application.

Acknowledgements

This research has been carried out under the HealthAgents research grant, funded by the Information Society Technologies priority of the European Union Sixth Framework Programme as an Specific Targeted Research Project with contract no.: IST-2004-27214 (2006–2008).

Nomenclature

ACL	Agent Control/Communication Language
API	Application Programming Interface
D2RQ	System for mapping relational database schemas to ontologies
DMZ	De-Militarised Zone (A fully-firewalled network)
FIPA	Foundation for Intelligent Physical Agents
HADOM	HealthAgent Domain Ontology
HAL	HealthAgents Language (The communication language of the agents)
HR-MAS	High-Resolution Magic Angle Spinning (high resolution ex-vivo form of MRS)
ICAP	Intensive Care Agent Platform
ITACA	Institute for the Applications of Advanced Information and Communication Technologies
JADE	Java Agent Development Environment
LCC	Lightweight Coordination Calculus agent workflow language
MIAKT	Medical Images and Advanced Knowledge Technologies (project)
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
MV	Multi-Voxel (form of MRS data-capture)
PQE	Pharma Quality Europe
RDF	Resource Description Framework

RDFS	Resource Description Framework Schema (vocabulary of describing vocabularies)
RDQL	Resource Description Query Language
SPARQL	Simple Protocol and RDF Query Language
SQL	Structured Query Language (for relations databases)
SV	Single Voxel or Volume (form of MRS data-capture)
TURTLE	Terse RDF Triple Language
UAB	Universitat Autònoma de Barcelona
WHO	World Health Organisation

References

R. Annicchiarico, U. Cortés, and C. Urdiales, editors. *Agent Technology and e-Health.* Whitestein Series in Software Agent Technologies and Autonomic Computing. Birkhäuser Verlag, Basel, 2008. ISBN 978-3-7643-8546-0.

C. Arús, B. Celda, S. Dasmahapatra, D. Dupplaw, H. González-Vélez, S. van Huffel, P. Lewis, M. Lluch i Ariet, M. Mier, A. Peet, and M. Robles. On the design of a web-based decision support system for brain tumour diagnosis using distributed agents. In *2006 IEEE/WIC/ACM Int Conf on Web Intelligence & Intelligent Agent Technology (WI-IAT 2006 Wksp)*, pages 208–211, Hong Kong, Dec. 2006. IEEE.

D. Beckett. Modernising semantic web markup. In E. Dumbill, editor, *XML Europe 2004*. IDEA Alliance, 2004. URL `http://www.idealliance.org/papers/dx_xmle04/papers/03-08-03/03-08-03.html`.

D. Beckett. Turtle–Terse RDF Triple Language. Manual (version 20/Nov/2007), ILRT University of Bristol, `http://www.dajobe.org/2004/01/turtle/`, 2007.

F. Bellifemine, A. Poggi, and G. Rimassa. JADE: a FIPA2000 compliant agent development environment. In *AGENTS'01: Fifth Int Conf on Autonomous agents*, pages 216–217, Montreal, Canada, 2001. ACM Press.

C. Bizer. D2RQ. Freie Universität Berlin, 2006. `http://sites.wiwiss.fu-berlin.de/suhl/bizer/D2RQ/`(Last accessed: 13 Feb 2007).

M. Boniface, T. Leonard, M. Surridge, S. Taylor, L. Finlay, and D. Mccorry. Accessing patient records within virtual healthcare organisations. In P. Cunningham and M. Cunningham, editors, *eChallenges 2005: Innovation and the Knowledge Economy: Issues, Applications, Case Studies*, Ljubljana, Slovenia, Oct. 2005. IOS Press. ISBN 1-58603-563-0.

J. Broekstra, A. Kampman, and F. van Harmelen. Sesame: A generic architecture for storing and querying RDF and RDF schema. In *First Int Semantic Web Conf*, number 2342 in Lect. Notes Comput. Sc., pages 54–68. Springer Verlag, July 2002.

D. Brugali and K. Sycara. Towards agent oriented application frameworks. *ACM Computing Surv.*, 32(1):21–27, 2000.

D. Brugali, G. Menga, and A. Aarsten. The framework life span. *Commun. ACM*, 40(10):65–68, 1997.

L. M. DeAngelis. Brain tumors. *N. Engl. J. Med.*, 344(2):114–123, 2001.

P. Georgiadis, D. Cavouras, I. Kalatzis, A. Daskalakis, G. C. Kagadis, K. Sifaki, M. Malamas, G. Nikiforidis, and E. Solomou. Improving brain tumor characterization on MRI by probabilistic neural networks and non-linear transformation of textural features. *Comput. Meth. Programs Biomed.*, 89(1):24–32, 2008.

H. González-Vélez, M. Mier, M. Julià-Sapé, T. N. Arvanitis, J. M. García-Gómez, M. Robles, P. H. Lewis, S. Dasmahapatra, D. Dupplaw, A. Peet, C. Arús, B. Celda, S. Van Huffel, and M. Lluch-Ariet. HealthAgents: Distributed multi-agent brain tumor diagnosis and prognosis. *Appl. Intell.*, 30(3):191–202, June 2009. ISSN 0924-669X. doi: 10.1007/s10489-007-0085-8.

T. R. Gruber. A translation approach to portable ontology specifications. *Knowl. Acquis.*, 5(2): 199–220, 1993. ISSN 1042-8143. doi: 10.1006/knac.1993.1008.

J. Huang, N. Jennings, and J. Fox. Agent-based approach to health care management. *Appl. Artif. Intell.*, 9(4):401–420, 1995.

G. Lanzola, L. Gatti, S. Falasconi, and M. Stefanelli. A framework for building cooperative software agents in medical applications. *Artif. Intell. Med.*, 16(3):223–249, 1999.

E. Merelli, G. Armano, N. Cannata, F. Corradini, M. d´Inverno, A. Doms, P. Lord, A. Martin, L. Milanesi, S. Möller, M. Schroeder, and M. Luck. Agents in bioinformatics, computational and systems biology. *Brief. Bioinform.*, 8(1):45–59, 2007.

J. Rees. Advances in magnetic resonance imaging of brain tumours. *Curr. Opin. Neurol.*, 16(6): 643–650, 2003.

D. Robertson. Multi-agent coordination as distributed logic programming. In *ICLP*, volume 3132 of *Lect. Notes Comput. Sc.*, pages 416–430. Springer, 2004.

N. Shadbolt, P. Lewis, S. Dasmahapatra, D. Dupplaw, B. Hu, and H. Lewis. MIAKT: Combining grid and web services for collaborative medical decision making. In S. Cox, editor, *UK e-Science All Hands Meeting*, pages 784–791, Nottingham, UK, Sept. 2004. EPSRC. ISBN 1-904425-21-6. URL `http://www.allhands.org.uk/2004/proceedings/`.

A. R. Tate, J. Underwood, D. M. Acosta, M. Julià-Sapé, C. Majós, A. Moreno-Torres, F. A. Howe, M. van der Graaf, V. Lefournier, M. M. Murphy, A. Loosemore, C. Ladroue, P. Wesseling, J. L. Bosson, M. E. Cabañas, A. W. Simonetti, W. Gajewicz, J. Calvar, A. Capdevila, P. R. Wilkins, B. A. Bell, C. Rémy, A. Heerschap, D. Watson, J. R. Griffiths, and C. Arús. Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. *NMR Biomed.*, 19(4):411–434, 2006.

S. B. Tatter. *The new WHO Classification of Tumors affecting the Central Nervous System.* PhD thesis, Cornell University Medical College/The Rockefeller University, 1996. URL `http://neurosurgery.mgh.harvard.edu/newwhobt.htm`.

F. D. Turck, J. Decruyenaere, P. Thysebaert, S. V. Hoecke, B. Volckaert, C. Danneels, K. Colpaert, and G. D. Moor. Design of a flexible platform for execution of medical decision support agents in the intensive care unit. *Comput. Biol. Med.*, 37(1):97–112, 2007.

S. Weiss. *Histological Typing of Soft Tissue Tumours.* World Health Organization. International Histological Classification of Tumours. Springer, Berlin, second edition, 1994. ISBN 978-3-540-56794-3.

A. J. Wright, C. Arús, J. P. Wijnen, A. Moreno-Torres, J. R. Griffiths, B. Celda, and F. A. Howe. Automated quality control protocol for MR spectra of brain tumors. *Magn. Reson. Med.*, 59 (6):1274–1281, 2008.