

# Developing an Integrative Platform for Cancer Research: a Requirements Engineering Perspective

Anthony Finkelstein<sup>1</sup>, Leah Goldin<sup>1</sup>, Jeff Kramer<sup>2</sup>, Helen Parkinson<sup>3,4</sup>, Vito Perrone<sup>1</sup>,  
Fiona Reddington<sup>3</sup>

1 University College London, London UK

2 Imperial College, London UK

3 NCRI Informatics Coordination Unit, London UK

4 European Bioinformatics Institute, Cambridge UK

## Abstract

The NCRI Informatics Initiative has been established with the goal of using informatics to maximise the impact of cancer research. A clear foundation to achieving this goal is to enable the development of an informatics platform in the UK that facilitates access to, and movement of, data generated from research funded by NCRI Partner organisations, across the spectrum from genomics to clinical trials. To assure the success of such a system, an initial project has been defined to establish and document the requirements for the platform and to construct and validate the key information models around which the platform will be built. The platform will need to leverage many projects, tools and resources including those generated by many e-Science projects. It also required contributing to the development of a global platform through a close interaction with similar efforts being developed by the NCI in the USA. This paper recounts our experience in analysing the requirements for the platform, and explains the customised analysis approach and techniques utilised in the project.

## 1. Introduction

A critical factor in the advancement of research in any e-science domain is the ease with which data can be integrated, redistributed, and made accessible and analyzable. This is particularly true in the cancer research domain where tackling these challenges is key to integrating diverse scientific data to inform clinical decision-making and enabling a move towards personalised medicine. A number of factors have so far hampered the ability of researchers to achieve interoperability across the various data sets in the cancer domain and to provide effective access to integrated data and services. Data sets have been generated by different research groups around the world working across the cancer research spectrum, that is, from basic to clinical cancer research. These distributed and heterogeneous data sets have been recorded and often made accessible (typically via web sites) in non-standardized ways, that is, using different vocabularies, data structures, metadata standards and service interfaces. In recent years, the need for standards has been recognized and a number of projects and initiatives have been established to

define reference ontologies (within specific sub-domains or across sub-domains), common metadata elements, data representation formats, common data repositories, and so forth. Currently there are many informatics projects, resources and standards in use in the UK and internationally. Many of these are excellent research tools, but they have evolved separately and so present an incoherent, fragmented landscape. Furthermore, it is unclear whether these resources and standards have responded more to technical imperatives than user requirements.

### 1.1 The NCRI Informatics Platform

In this context, the goal of the NCRI Informatics Initiative (hereafter NCRIII) [1] is to increase the impact of UK cancer research and improve prevention and treatment of cancer by effective use of informatics to manage and exploit the vast amounts of diverse information currently generated. The envisaged support will be provided through an integrative platform [16] whose main aim is to enable the creation of an open community where the different informatics

tools and resources available in the UK and worldwide can interoperate with one another as a coherent whole. This will reduce duplication of effort and funding and leverage existing resources for maximum benefit.

## 1.2 The NCRIII Platform Requirements Analysis Project

As with any complex and innovative software system project, a fundamental activity is to reach a clear understanding of what the system's goals and requirements are. Lack of proper requirements analysis is recognized as the major source (almost 50%) of failure for software system projects [2]. E-science projects are, we would contend, not immune to this. A proper requirements analysis should: identify what is missing in the context where the system will operate, define stakeholder goals and expectations; eliminate ambiguities, inconsistencies and problems potentially leading to late faults; provide the information needed to support the system construction and the evaluation of competing technical options; enable system change and evolution.

Within the NCRIII the need for a precise identification of the key goals and requirements has been considered of primary importance due to the complexity of the envisioned project and the apparent lack of solutions capable of filling the existing gaps. The platform development has been thus preceded by a project focusing on the requirements analysis. A multidisciplinary analysis group has been set up and the analysis has been driven by a set of use cases acquired by interviewing practitioners working in the field. Unlike use cases collected in other e-science project, which describe the main functionalities of the system being developed, our use cases needed to uncover interoperability needs. The main goals of our analysis have been to understand how the various initiatives operating in the cancer research field would effectively cooperate with one another, what the relative roles are, how they are actually used by practitioners, and what may make the community self-sustainable. This has required us to assume a higher level perspective in defining the use cases which are closer to user stories as defined in the Agile development [22] than to standard use cases as described in [3]. The core characteristic of our project along with the intrinsic multidisciplinary nature of the project team and the need to work side by side with domain experts, have required us to adopt an innovative approach and to customize the used analysis techniques. In this paper we describe the approach and the used techniques, motivating

them with respect to the project and domain characteristics.

Although our approach has been developed for supporting analysis of the cancer research domain, we believe that it has generic applicability to other projects in the broader context of e-science.

## 2. The Platform Context and Related Works

The context where the NCRI platform will operate is made up of a multitude of projects, resources and initiatives that aim to support cancer research in different ways. These range from local projects to global initiatives and address different issues ranging from data acquisition, storage, and mining to knowledge management. Furthermore, the field of cancer research is highly dynamic reflecting the emergence of new technologies, developing infrastructure and scientific discovery.

Our field analysis has led us to organize the different entities involved in the platform context into the following categories.

**Local Repositories:** belong to institutions or research groups and tend to host data about investigations performed by the local teams. They are important sources of data although the terminologies and data formats used to store the acquired data or specimens tend to be highly heterogeneous between repositories. Access to the data is usually limited to local researchers or exploiting direct contacts with them.

**Specialized Global Repositories:** aim to collect information about specific types of data e.g. ArrayExpress [4], and caArray [5], for micro-array data. Researchers need to be supported in both depositing and accessing the data they contain. Repositories may overlap with one another in the type of data they house and can use different data formats and metadata vocabularies to annotate the acquired data. Most of the repositories offer ad-hoc defined input forms for collecting data and a web site for accessing the stored data typically through a search engine. Other access services like API or Web Services may be offered.

**Knowledge Bases:** aim to collect domain knowledge in specific areas, organizing it so that deductive reasoning can be applied to them. Each repository focuses on specific data types and provides ad-hoc defined forms for input data and access functionalities which may include advanced user graphical inter interfaces. Examples include REACTOME (a curated KB of biological pathways) [6], PharmGKB (an integrated resource about how variation in human

genes leads to variation in our response to drugs) [7], UniProtKB/Swiss-Prot (a Protein Knowledgebase containing annotated protein sequences) [8].

**Scientific Libraries:** collect scientific publications containing research results in the field of interest. The basic information unit is a scientific paper. Examples include the Public Library of Science[9], and PubMed [10]

**Policies:** define the way information is collected and used. Examples include those provided by the Department of Health [12]

**Terminology Sources (or ontologies):** defined for offering terminology annotation services to the global cancer research community. They can vary in terms of addressed domains, coverage extension (domain specific vs. cross-domain), description language, adoption level, access service interfaces, etc. Examples include the NCI-thesaurus (cross-domain annotations in the cancer research domain)[5], the GO ontology (annotation of gene products), the MGED-ontology (descriptors for microarray experiments) [15], etc.

**Representation Formats:** defined to represent data for specific domains. Common representation formats are not available for all the cancer domains and often there are several proposals or “standards” in use. Examples include HL7 [13] and CDISC [14] for clinical information exchange, MAGE-ML [15] for microarray experiments, etc.

**Bioinformatics Services:** existing services used to elaborate raw data (e.g. microarray experiment data clustering), to search similarities between new biologic data and existing data stored in some repositories (e.g. BLAST), to translate data among different formats, etc.

Although not all existing projects, resources and initiatives can be easily categorized, this list provides a broad high-level view of the reference context showing its multiple facets.

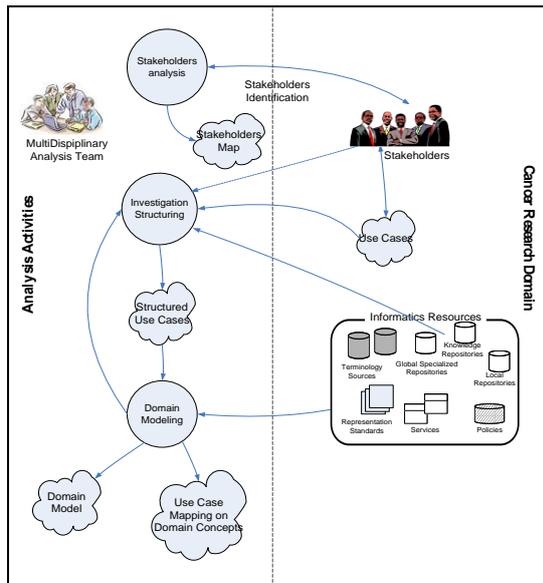
The issue of interoperability is currently being addressed via initiatives underway at the US National Cancer Institute Center for Bioinformatics (NCICB) [5] and the European Bioinformatics Institute (EBI) [4]. NCICB is coordinating an ongoing project called the cancer Biomedical Informatics Grid (caBIG) whose aim is to maximize interoperability and integration among initiatives funded by the NCI by providing open source infrastructure and tools plus vocabulary services (EVS – Enterprise Vocabulary Services [5] and a centralized metadata repository (caDSR – Data Standards Repositories [5]). The EBI provides a centralized access point to several bioinformatics databases and services through a unique web site.

Although the ultimate goal is similar, improving research efforts by enabling interoperability, a number of differences exist between the envisioned NRCRIII platform and the existing initiatives [16]. Essentially, while the existing initiatives are addressing the problem by undertaking an approach whereby tools, vocabularies, common repositories, etc are built and access is centrally provided by the lead organization, the NCRI initiative aims to provide a ‘glue’ layer enabling the different initiatives, resources and tools to communicate with one another in a more effective fashion. This requires close collaboration with aforementioned organisations and other existing projects including CancerGRID [17], CLEF [18], etc.

### 3. The Analysis Approach

A feature of e-science projects is the need to establish a multidisciplinary team including requirements engineers, software developers and domain experts with a broad vision of the specific fields, in this case the cancer biomedical domain. Such a team was composed including requirements experts from University College London and Imperial College and domain experts from the NCRI Informatics Coordination Unit and Task Force. In the initial phase, cancer related literature analysis and periodic team meetings have been intertwined to build up a common understanding of the platform’s high level goals and to define a common language. Outcomes of the preliminary activities have been a *context diagram* (whose main components have been described in outline in section 2), an initial set of *stakeholders* and a preliminary *domain model*. Subsequently, the analysis has been focused on understanding how the various elements operating in the platform context can interoperate one with another to underpin the needs of scientific researchers. In accordance with the NCRI vision of supporting community driven development, we have adopted a user centred approach for the analysis activities. To this end, a number of use cases, covering the whole spectrum of the cancer research as proposed by the NCRI matrix [1], have been acquired and used to drive the overall analysis process.

Figure 1 maps out the analysis flow focusing on the use case analysis.



**Figure 1: The use case analysis flow**

Information acquired in the initial phase has been used to begin the stakeholders analysis activity. The method used, known as the “onion model” [20], allows a progressive identification and classification of the key stakeholders by means of interviews and other elicitation techniques. The bidirectional line in Figure 1 points out that the information acquired in each interview with stakeholders may lead to the identification of further stakeholders. Besides progressing in building the *stakeholders map* (available on [19]), in each interview with potential users a number of *high level use cases*, akin to “user stories” in the Agile development [22], have been defined. They describe stories on cancer research investigations that can benefit from the introduction of the platform. A simple example is shown in Figure 2. Working side by side with our experts for defining such stories, we have encouraged them to avoid thinking how the platform could satisfy their research needs and to instead carefully state their goals and what they would ask of the platform to achieve their goals. Such an approach permits uncovering of real users’ needs and avoids potential biases, introduced by having in mind premature solutions [21].

Although rich in information, use cases acquired in this way need to be better structured to support further analysis. The **Investigation Structuring** activity, further described in section 4, aims thereafter to provide the use cases we acquired with a structure suitable for analysis. In summary, it enables the analysis team, and eventually the users involved in the use case definition, to systematically identify *what* the

platform should be able to provide the researchers with in order to fulfil their goals.

**Research Area:** Functional Genomics

**Actor:** Researcher in a microarray laboratory

**Description**

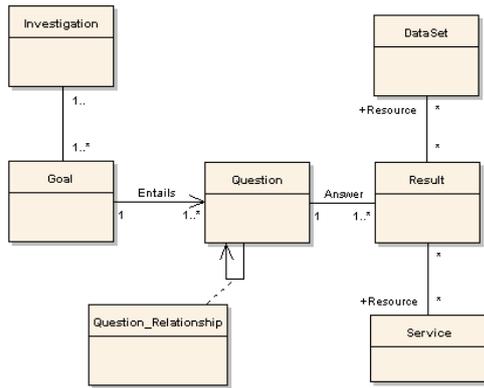
A scientist wishes to investigate genetic variation in tumour response to treatment with a specific class of chemotherapy. She would like to identify specimens of a specific tumour type, flash-frozen and prepared using a specific methodology, and for which there are associated medical records for treatment outcome. With sections of those specimens, the researcher would like to carry out microarray experiments for tumour cells and normal cells on the periphery of the tumour. She needs to store and analyze the data using conventional clustering methodologies. She would also like to compare the clusters to currently-known metabolic pathways, some of which are known to be chemotherapy targets. With the list of genes from the pathways of interest showing expression variation related to the chemotherapy treatment, the investigator can then identify common genetics variations in the public databases for subsequent follow-up. At the time of publication of her study she wants to maximize the impact of her achievements on the scientific community for follow-up studies.

**Figure 2: An example from the use case collection**

The *structured use cases* are then analyzed in the **Domain Modelling** activity with two main purposes. Firstly, the *domain model* is validated, or modified, to reflect the use cases. Secondly, the actual resources available in the cancer context, described in terms of domain model concepts, are mapped to the use cases enabling a precise study of the interoperability issues. Investigation Structuring and domain modelling derive from well-known techniques in the requirements engineering field but have been customized to address the specific needs of this project. The next sections describe how they have been used, highlighting the aspects of a typical e-science project that have influenced the customization of these techniques.

#### 4. Using an High Level Investigation Model in the Use Case Analysis

The use cases typically describe examples of investigations researchers perform in their daily work. The descriptions are completely unstructured and hard to understand for non-specialists like the computer scientists in charge of the analysis activities. An important step in their analysis is to structure them so that the investigation’s goals, flow and the informatics resources needed by researchers can be clearly identified. To this end, we have defined a high level model, shown in Figure 3 enabling simple but effective structuring of investigations in terms of three primary concepts – Goals, Questions and Results (GQR) – and two subsidiary concepts – Services and Data Sets. The model is grounded by goal oriented requirements engineering principles [23] and is inspired by the Goal/Question/Metric method [24] used to plan the measurement of success of a software system from the user point of view.



**Figure 3: Core concepts in the GQR method**

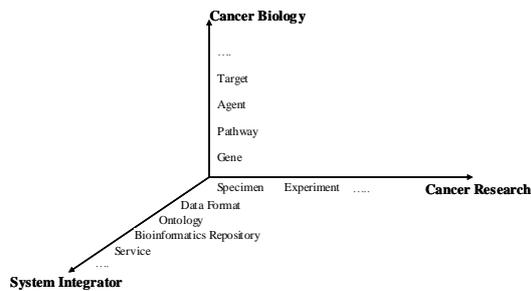
A **Goal** represents the investigation’s objective. Goals can be more or less specific like “the role of diet in cancer” or “investigate whether a disease responds to a drug”. A goal entails one or more **questions** that must be answered to achieve its fulfilment. Questions are answered by way of *data sets* or *services* (which in turn may consume data sets) building up the actual **results** the researcher was looking for. Data sets and services can be considered either inputs or outputs of an investigation. In the former case, one or more repositories or service providers that can offer the required data sets or services should be identified. In the latter case, repositories or service registries are used to publish data or services produced within the investigation. Results act as a bridge between the use case analysis and the domain modelling. They are used in combination with the domain model to identify what domain entities represent the information required in the result. Due to its simplicity, people without a software modelling background can easily understand the model. Due to the specialised nature of the various cancer research sub-fields, often only the experts interviewed for defining each use case could provide us with the needed information for structuring the use cases and identifying the sources for each result. This has been one of the main challenges we have had to cope with in defining it.

## 5. Using Domain Modelling to Master the Interoperability Problem

Defining the domain model is an important activity in any analysis process. This is particularly true in the context of e-science projects where it plays a central role in the whole development process. A domain model represents a rigorously organized and selective abstraction of the knowledge in the domain experts’ head. It addresses three main objectives: (1) acts as a bridge between the problem analysis and the

solution design so it is used to ensure that the analysis that went into it applies to the final product, the software system; 2) is the *backbone* of the language used by all the team members, including analysts, domain experts and developers; (3) is the teams’ agreed-upon way to structure the domain knowledge so that when new knowledge surfaces the domain model can be used to interpret it or to identify uncovered aspects. Defining the domain model is an iterative and incremental activity where rigorous analysis of meeting minutes, use cases, documentations, etc. by means of engineering techniques is intertwined with discussions with domain experts to reach a common vision. Moreover, the peculiarities of the e-science context require some properties to be considered in the model definition. Since the context is continuously evolving, it must be easy to accommodate inevitable changes. This requires the model to be extensible and flexible. Fine-grained and very specialized models are hardly extensible and it is often very difficult (if not impossible) to reach an agreement in heterogeneous teams involving domain experts with different specializations. In this light, our domain model has been defined to be sufficiently generic and to accommodate different points of view, while identifying the key entities and relationships. A second typical aspect of e-science projects is that their analysis requires a number of different perspectives to be considered [25]. As far as the perspective change, different typologies of entities and relationships may be required to describe the domain. For instance, in the above use case we can observe at least two perspectives. A *laboratory perspective* can be seen when the story talks about “specimens”, “experiments”, “protocol definition” (implicitly), “clustering services”, etc. A *biological perspective* comes out in the last part of the use case’s description when biological aspects are used to specify the searched information. On the other hand, since the platform will act as a system integrator allowing the interoperability among the different informatics resources, the *system integrator perspective* must be considered. The above mentioned data may be stored in “local repositories”, “specialized bioinformatics repositories” or “Knowledge bases”; data can be represented in different “representation formats” so that “translation services” published on “service registries” may be needed to integrate them; “terminology sources” are used to annotate data and services so that these can be retrieved by the platform; and so forth. These aspects are not directly described in the use case but can be defined analysis from the system integrator perspective with the

multidisciplinary team. An important issue we have identified in this project is thus that in order to clearly identify all the needed information, a multi-perspective analysis is needed. This entails a multi-dimensional domain model to be defined for supporting the analysis activities. Each dimension can be considered a domain model by itself used alone or combined with the other dimension in the analysis activities.



<b>Cancer Research</b>	Investigation, Experiment, Protocol, Individual Information, Research Sample, Clinical Record, Experiment Result, Policy, etc.
<b>Cancer Biology</b>	Gene, Protein, Pathway, Agent, disease, drug, target, tissue, SNP, etc.
<b>System Integrator</b>	Bioinformatics Repository, Ontology, Registry, Service, Data Format, Access Policy, Protocol, etc.

**Figure 4: The 3-dimensional domain model**

In our project, The Platform domain model has been organized across three dimensions. The **Cancer Research** which includes all the concepts involved in the investigation/experiment execution, like *samples*, *patient data*, *protocols*, *publishable data* (results of the experiment executions), *etc.* The **Cancer Biology** including concepts like *tumour*, *drug*, *gene*, *pathway*, *etc.*

The **System Integrator** whose aim is to model the environment where the platform will operate, the different types of available resources and their relationships. Lack of space prevents us from showing the three models but

Figure 4 shows the three dimensions of our domain model above and some examples of the included concepts below. The complete specification can be found on [19].

### 5.1 Examples use case analysis through the domain model

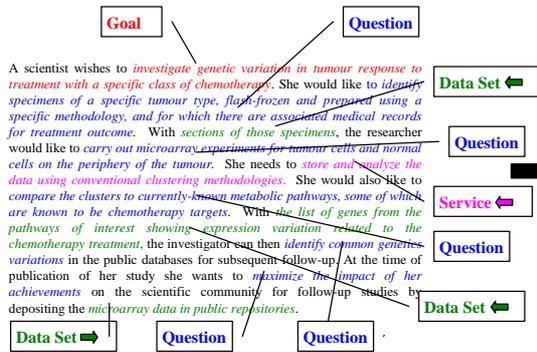
In this section we show some excerpts from the analysis of the use case shown in Figure2. The GQR analysis is usually conducted together with

the interviewed domain experts. Generally, the expert writes down the use case description and then, together with the analysis team, it is progressively structured whereby the method's concepts. The original description does not contain all the information that makes up the requirements. The GQR method has shown up to be effective in supporting the conversation between the analysts and the domain experts. For each question, experts are asked to identify what data sets and/or services may be used to produce the needed result, dragging in their knowledge of the context. Requirements are identified through this interaction and recorded as notes, at the beginning, and through semi-formal diagrams and tables afterwards.

Let us consider, as example, two questions among the five that follow from the goal.

- Q1.1 can be answered by identifying suitable specimens to be used in the experiment. This information can be found in tissue banks belonging to hospitals (e.g. Dundee tissue bank) or by means of global specialized repositories like those under construction in UK (e.g. OnCore [26]) or in US (e.g. caTissue [5]). The platform should thus be able to access to such repositories to query them and to produce the result R1, that is, a *list of specimens*. Using the cancer research perspective in the domain model, this result can be explained by way of the entities *CR\_ResearchSamples* and *CR\_MedicalRecords*. The platform will thus need to search all repositories which are known to include data about such kind of entities. From the *system integrator* perspective, these repositories are modelled as *SI\_InformaticsRepositories* (and in particular *SI\_GlobalSpecializedRepositories* for "OnCore" and *SI\_LocaRepositories* for "Dundee") and their data elements, representing the above mentioned entities, should be *semantically annotated* by metadata elements (*SI\_MetaDataElements*) whose domain is defined within a *IS\_TerminologySource* like the "NCI-Thesaurus" providing, for instance, concepts like *Organism*, *Disease*, *Disorder and Findings*, *Drugs and Chemical* to match the researcher's query attributes.

- Q1.4 can be answered by using the list of genes identified by the clustering service used to identify common genetic variations querying existing *knowledge bases* like REACTOME [6], PharmGKB [7], KEGG (Kyoto Encyclopedia of Genes and Genomes) [27], etc., or *scientific libraries* (e.g. PubMed) looking for papers which have reported about similar experiments. Currently, researchers use screen-scraping technologies to analyze and integrate data sets collected from different repositories.



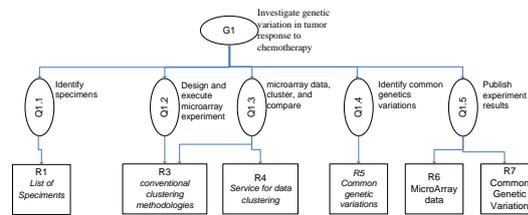
**Figure 5: Excerpts from the use case analysis**

R5 is the result of such integration that should be supported by the platform (as much as possible). Using the biological perspective in the domain model, we can explain the result in that saying that the platform should query repositories known to contain information about *CB\_Gene*, *CB\_Pathways*, *CB\_ExpressionVariation* and *CB\_Agent* (where a class of chemotherapy is considered an agent). From the *system integrator* perspective, these repositories are modelled as *SI\_KnowledgeBases* (e.g. “REACTOME” and “PharmGKB”) and *SI\_ScientificLibrary* (e.g. PubMed). As well as in the previous example, their information should be annotated with proper metadata. To build the required result, the platform should integrate data coming from these sources by exploiting the semantic relationships among concepts defined in the *IS\_TerminologySource* used as metadata source.

By these examples we have shown how the same use case is analyzed from different perspectives providing an explanation from both the user and system perspective on how the different resources should interoperate with one another to fulfil the high-level research goals.

## 6. Validation

The approach and models introduced in this paper are currently being validated in a companion demonstrator project (see [1] section Demo Projects, Imaging and Pathology). This project aims at developing services for 3D reconstruction of the macroscopic resected rectums using in vivo photographs and incorporating links to virtual histology images and MRI images. The project leverages technologies and services developed in the e.Diamond project [29] and is using data sets stored in repositories belonging to hospitals involved in a previous clinical trial, that is, the Mercury Trial [28]. Data (including images) is currently being collected and anonymized manually and transferred by coping it on CDs. Finally, the developed service has to be deployed so that it can be used in a distributed environment for more informed clinical decision making. Several user stories have been identified and are being analyzed in the context of this project to



evaluate the effectiveness of our approach in a real case. Moreover, the resources the platform would need to support the project are being collected. These range across the different kinds of resource identified in section 2. Analyzing how they interoperate with one another to fulfil the various goals identified in the user stories will provide a suitable test-bed for validation of the whole information architecture.

## 7. Discussion and Future Work

In this paper we have recounted our experience in the early analysis of the NCR1 Platform, a software system whose main aim is to enable the creation of a community supporting initiatives operating in the cancer research field in UK and world-wide. A multidisciplinary team composed by requirements engineers and domain experts was set up to carry out the analysis activities. Preliminary steps in the analysis process have been the identification of a core set of stakeholders and the main entities operating in the cancer research context, and the definition of a preliminary domain model to gain an initial common understanding of the problem within the multidisciplinary analysis team. Afterwards, the analysis has been driven by a set of use cases representing examples of research investigations that might benefit from the platform introduction. The main aim of our analysis has been to clearly understand how the various high heterogeneous resources making up the cancer research context might interoperate with one another to fulfil researchers’ goals. The high heterogeneity of the cancer research domain and the need of working in a multidisciplinary team have required us to develop a customised approach to analyze the acquired use cases. The approach involves using an ad-hoc defined high level investigation model and a domain model. The former allows the analysis team along with the interviewed domain experts structuring an investigation throughout its lifecycle pointing out what researchers would ask the platform with and what resources it can answer with. The latter, allows analyzing in details the multiple facets of the interoperability problem by facing them from three orthogonal dimensions representing the researcher, biology and system points of view. The approach has proven to be effective in analysing a variety of use cases taken from different heterogeneous sub-

domains of the cancer research and particularly suitable for working with multidisciplinary teams.

Although it has been defined and used in the context of cancer research, we believe that the basic ideas and principles can be of interest to other projects in the broader e-research field.

Besides supporting the analysis activities, the investigation and domain models, mapped out in this paper, constitute the key information model around which the platform will be built. To achieve the platform's objectives, a semantic web service architecture is being designed following the Semantic Web Service Initiative proposed framework [30]. The platform will offer an environment where services requests (issued by the platform users) and offered services (issued by *service providers*) can be published. Adopting a model similar to the introduced Investigation Model, the platform will permit a simple workflow-like description of investigations where questions will correspond to service requests while results to offered services or to a set of available services integrated by *matchmaker services* provided by the platform environment. These services will be made accessible by providing semantic descriptions of their capabilities whose core concepts root in the Domain Model entities. These provide a high level conceptualization of the platform domain from three fundamental perspectives. This conceptualization can be used as a high-level ontology for semantically describing the web services offered by our platform. Such high level conceptualization will then be combined with a number of specialized ontologies (most of which already exist in the field) through ontology composition as described in [31].

Finally, we are currently analyzing and working in cooperation with several organizations which have already developed some of the informatics components the platform will need to use. Among others, we can mention the NCICB in US; the EBI; CancerGrid, CLEF and others.

## Acknowledgements

The project team wish to acknowledge the funding for this work provided by Cancer Research UK and to thank all the researchers who have cooperated in the analysis activities.

## References

- [1] NCRI Informatics Initiative  
[www.cancerinformatics.org.uk](http://www.cancerinformatics.org.uk)
- [2] Standish Group, CHAOS report,  
[www.standishgroup.com/chaos.htm](http://www.standishgroup.com/chaos.htm)
- [3] Alistair Cockburn: Writing Effective Use Cases. Addison Wesley, 2001.
- [4] European Bioinformatics Institute,  
<http://www.ebi.ac.uk>
- [5] National Cancer Institute – Center for Bioinformatics, <http://ncicb.nci.nih.gov/>
- [6] <http://www.reactome.org/>
- [7] <http://www.pharmgkb.org/>
- [8] <http://www.ebi.ac.uk/swissprot/>
- [9] <http://www.plos.org/>
- [10] <http://www.ncbi.nlm.nih.gov>
- [11] <http://www.geneontology.org/>
- [12] <http://www.dh.gov.uk/PolicyAndGuidance>
- [13] [www.hl7.org.uk/](http://www.hl7.org.uk/)
- [14] <http://www.cdisc.org/>
- [15] <http://www.mged.org/>
- [16] R. Begent, et al.: Challenges of Ultra Large Scale Integration of Biomedical Computing Systems", 18th IEEE International Symposium on Computer-Based Medical Systems, Dublin, Ireland, 2005.
- [17] <http://www.cancergrid.org/>
- [18] <http://www.clinical-escience.org/>
- [19] [www.cs.ucl.ac.uk/CancerInformatics](http://www.cs.ucl.ac.uk/CancerInformatics)
- [20] I. Alexander, S. Robertson, S.: Understanding project sociology by modelling stakeholders. IEEE Software, vol. 21(1), Jan. 2004.
- [21] Kuniavsky, M.: Observing the user experience: A Practitioner's Guide to User Research. Morgan Kaufmann, 2003
- [22] M. Cohn: User Stories Applied: For Agile Software Development. Addison-Wesley, 2004
- [23] A. Dardenne, A. van Lamsweerde, S. Fickas: Goal-Dircted Requirements Acquisition. Science of Computer Programming, 20 (1993)
- [24] R. Solingen, E. Berghout: The Goal/ Question/ Metric Method McGraw-Hill, 1999
- [25] A. Finkelstein, et al.: Viewpoints: a framework for integrating multiple perspectives in system development" Int. Journal of Software Engineering and Knowledge Engineering, vol. 2, 1992
- [26] <http://www.ntrac.org.uk/>
- [27] [www.genome.jp/kegg/](http://www.genome.jp/kegg/)
- [28] <http://www.pelicanancer.org/researchproject/s/mercury.html>
- [29] <http://www.ediamond.ox.ac.uk/>
- [30] M. Burstein et al.: A Semantic Web Services Architecture. Internet Computing, Sept. 2005
- [31] J. Jannink, P. Srinivasan, D. Verheijen G. Wiederhold: Encapsulation and composition of ontologies, Proc. of AAI Summer Conference, AAI (1998)