SOFTWARE ENGINEERING MEETS CANCER RESEARCH:
ENABLING INTEROPERABILITY OF DATA AND SERVICES

Anthony Finkelstein[1], Leah Goldin[1], Jeff Kramer[2], Helen Parkinson[3],
Vito Perrone[1], Fiona Reddington[3]

[1]Department of Computer science, UCL
[2]Department of Computer Science, Imperial College
[3]NCRI Informatics Coordination Unit

Context

Biomedical point of view
- Biomedicine has experienced an explosive growth
- Bottom-up growth
- Critical success factor: easing data access, redistribution, analysis and integration
- Multitude of data repositories, service interfaces, data formats, terminologies, etc.

Informatics point of view
- Under-investment in gathering requirements and conducting architectural analysis
- Little effort in understanding what the user needs are
- Poor requirements analysis causes failure or abandonment

Objectives and Challenges

NCRI Informatics Initiative Aim
"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"

Project Challenges
- Heterogeneity
- Innovative system
- Multi-disciplinary team (requirements eng. domain experts.)
- Multi-perspective (researcher, clinician, data base admin, etc.)

Project Aim
"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"

Project Objectives
- Define the system Scope
- Elicit *user requirements*
- Identify *integration needs* and understand the role different resources could play
- Define the high level architecture of the system to-be

Analysis Process
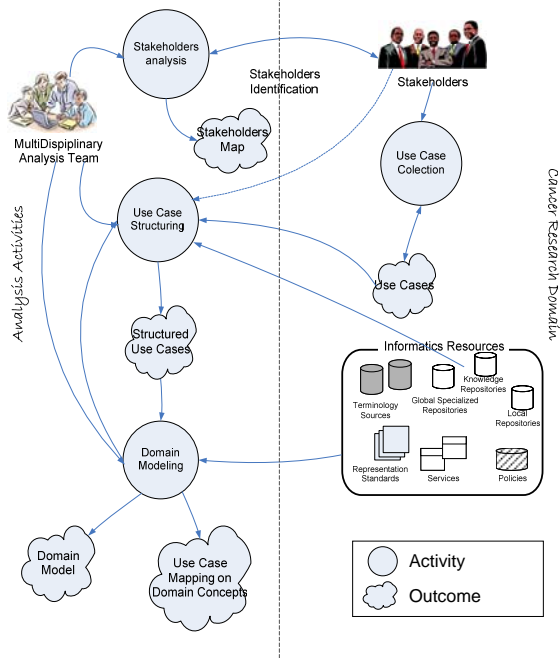
Use Case Collection
- Describe everyday *research investigations*
- "Stories" rather than cases of use for the system to be
- Exploratory nature
- Defined by domain experts avoiding describing *how* and focusing on *goals* and *what*
- Rich in information but unsuitable for our analysis

Use Case Structuring: Goal-Question-Results Method
- Ideas from the Goal-Questions -Metrics and Goal Oriented Analysis
- Driven by project needs: integration, multidisciplinary field heterogeneity, …
- Easy to understand and use for domain experts
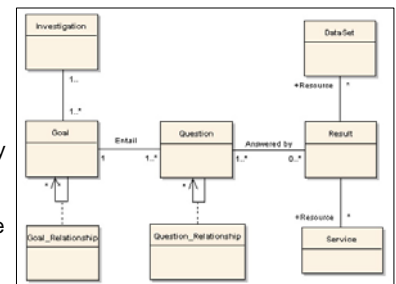
Domain Modelling
- High Level: Need for consensus across different sub-domains
- Multi-dimension: separate different types of concern to enable extensibility
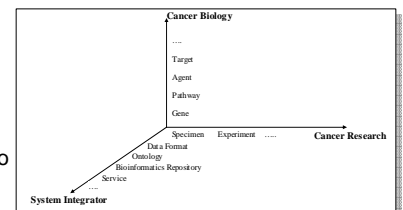
Conclusions and Next Steps

Key advantages
- Systematic structuring of research investigations
- Unveil user pains, integration needs, needed resources, resource roles
- Effective in supporting communication with experts
- Drive elicitation of extensive information
- Separation of concerns in the domain analysis

Cooperation
- On-going cooperation with the caBIG team (Harmonization workgroup) in defining an common domain model
- Future cooperation with related UK projects: CancerGrid, eDiamond, EBI, CLEF, myGrid, etc.

Next Steps
- Prototype User Interface (currently available, see references)
- Prototype Architecture: focus on 2-3 domains to validate the key requirements and estimate the whole development costs

References
- A. Dardenne, A. van Lamsweerde, S. Fickas: Goal-Dircted Requirements Acquisition. Science of Computer Programming, 20 (1993)
- R. Solingen, E. Berghout: The Goal/Question/ Metric Method McGraw-Hill, 1999
- 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
- Project Web site: www.cs.ucl.ac.uk/CancerInformatics