



Imperial College
London



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

Vito Perrone, Anthony Finkelstein, Leah Goldin: University College London
Jeff Kramer: Imperial College, London UK
Helen Parkinson: European Bioinformatics Institute, Cambridge UK
Fiona Reddington: NCRI Informatics Coordination Unit, London UK

Vito Perrone

E-science All Hands Meeting
Nottingham, 19th September 2006



Structure of Talk

- Context
- Objectives and Challenges
- Analysis
- Final Remarks and Future Steps

Structure of Talk

- Context
- Objectives and Challenges
- Analysis
- Final Remarks and Future Steps

Context – Biomedicine and Informatics

- 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

Context – NCRI Informatics Initiative

- “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”



NCRI Informatics Initiative 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”

Our Project aim

Structure of Talk

- Context
- Objectives and Challenges
- Analysis
- Final Remarks and Future Steps

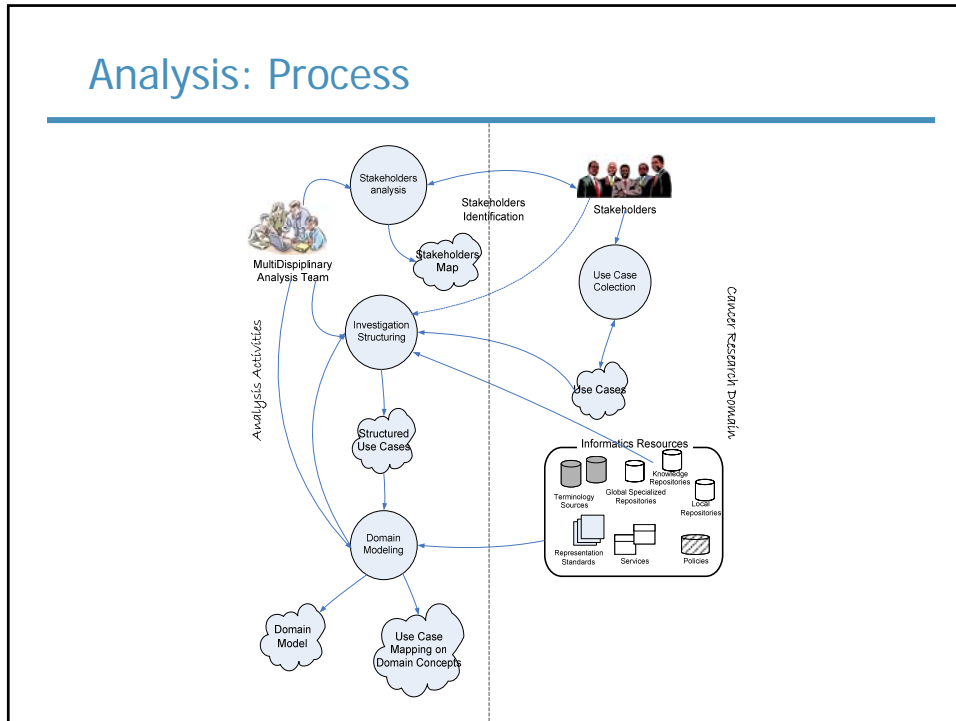
Objectives and Challenges

- 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
- Project Challenges
 - Heterogeneity
 - Innovative system
 - Multi-disciplinary team (req. eng. and NCRI domain exp.)
 - Multi-perspective (researcher, clinician, data base admin, etc.)
- Talk Objectives
 - Our experience in the user requirements elicitation and analysis
 - Describe motivations and use of our approach

Structure of Talk

- Context
- Objectives and Challenges
- Analysis
- Final Remarks and Future Steps

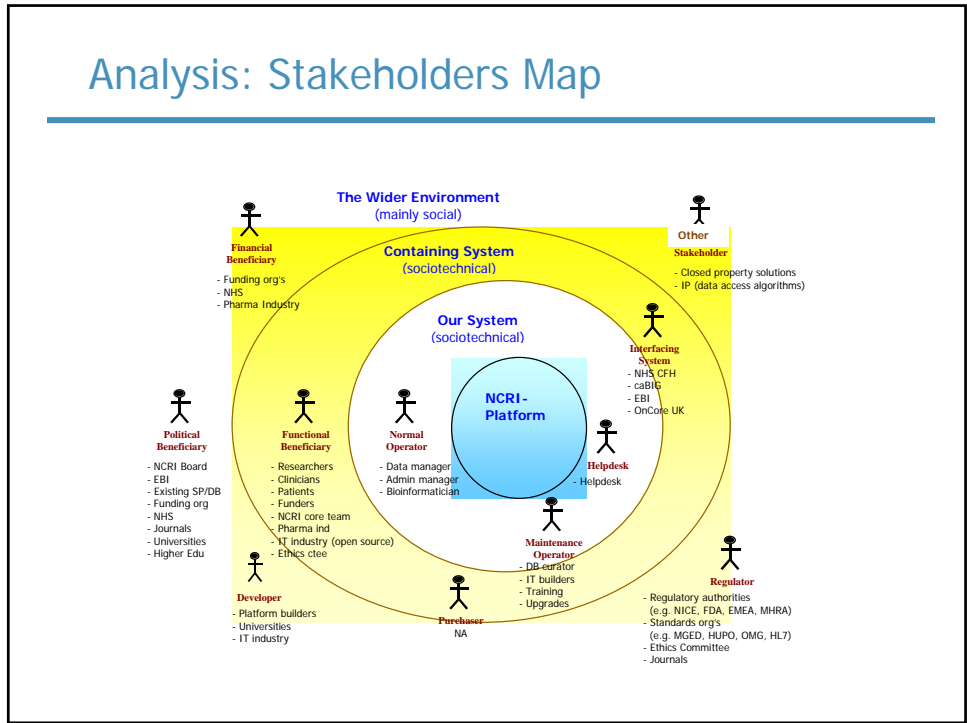
Analysis: Process



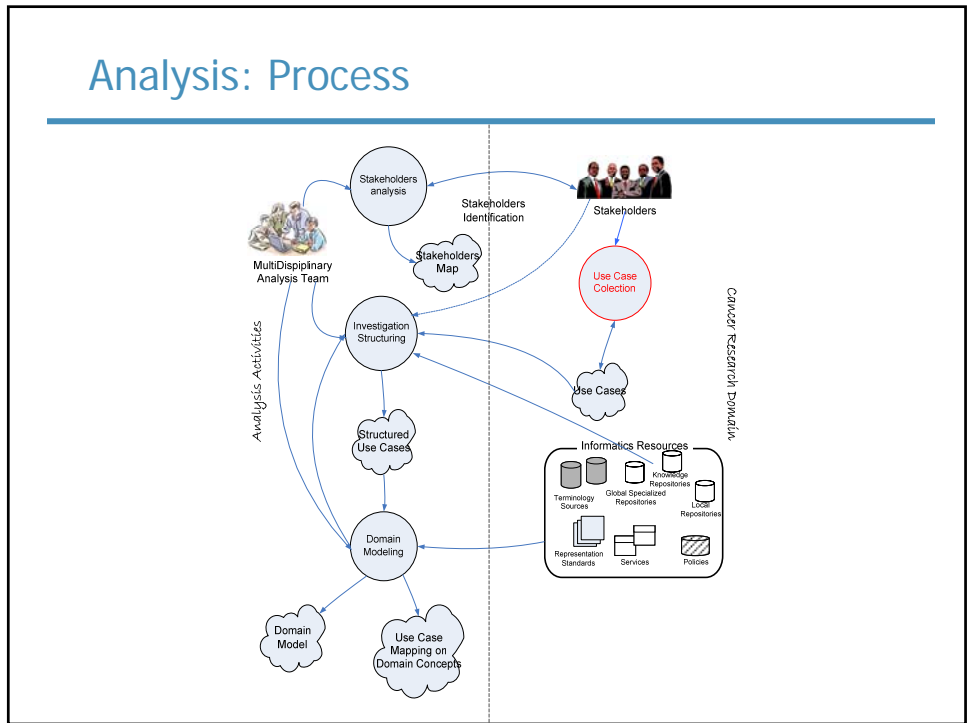
Analysis: Process





Analysis: Stakeholders Map



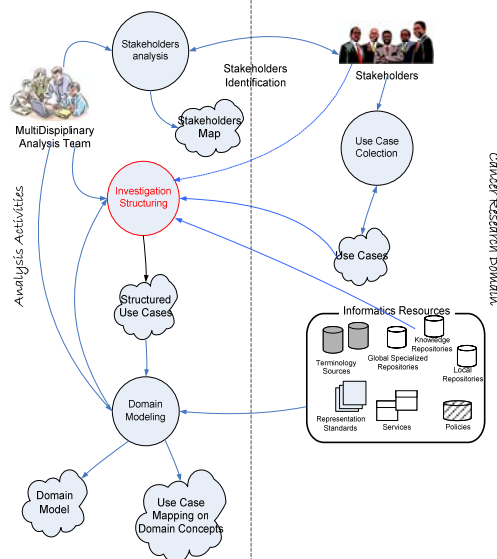
Analysis: Process




Analysis: Use case Collection

- Use case collection
 - “Stories” rather than cases of use for the system to be (Ex. )
 - Exploratory nature
 - Defined by domain experts
 - Avoid describing how (that is what researchers tend to) and focus on *goals* and *what*
 - Rich in information but unsuitable (as described by researchers) for our analysis (Ex. )

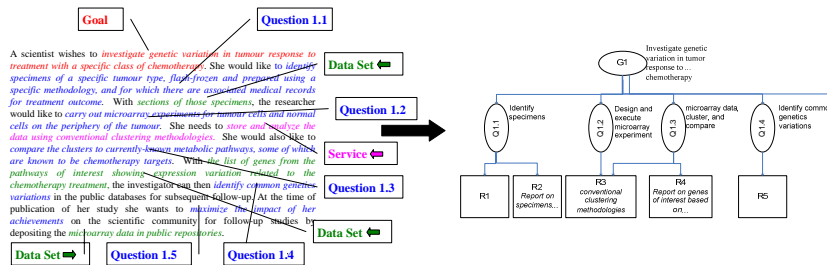
Analysis: Process



Analysis: Use Case Structuring

- GQR (Goal-Questions-Results) Method
 - Ideas from the Goal-Questions-Metrics and Goal Oriented Analysis
 - Driven by project needs: integration, multidisciplinary,...
 - Core meta-model 

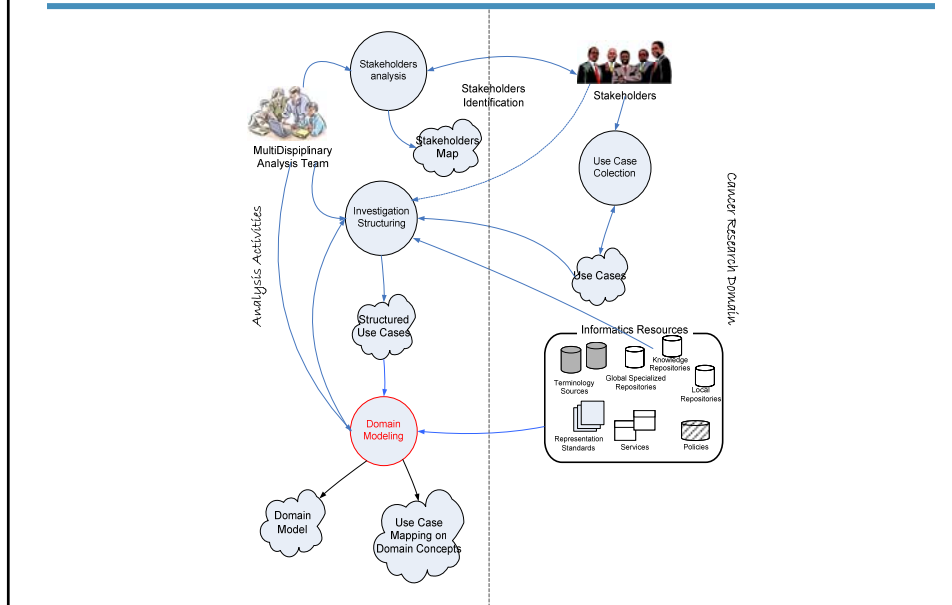
- Example



Analysis: Use Case Structuring

- Question Example
 - Q1.1: "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"
- What Analysis
 - Desiderata
 - Obstacles (users' pains)
 - Examples of Repositories, Records
 - Examples of data formats, keywords, ontologies, etc.
 - Level of adoption of informatics resources
 - ...

Analysis: Process

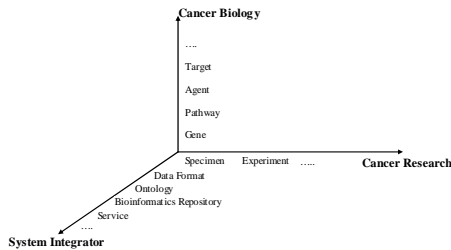


Analysis: Domain Modelling (1)

- Defining a domain model is a key task
 - Bridge between problem analysis and solution design
 - Language backbone
 - Help to structure the domain knowledge
- From existing domain models
 - Different types of concepts mixed together
 - Often specialized for a specific domain
- The Platform requirements
 - Need for consensus across different sub-domains
 - Unveil integration needs
 - Different perspectives

Analysis: Domain Modelling (2)

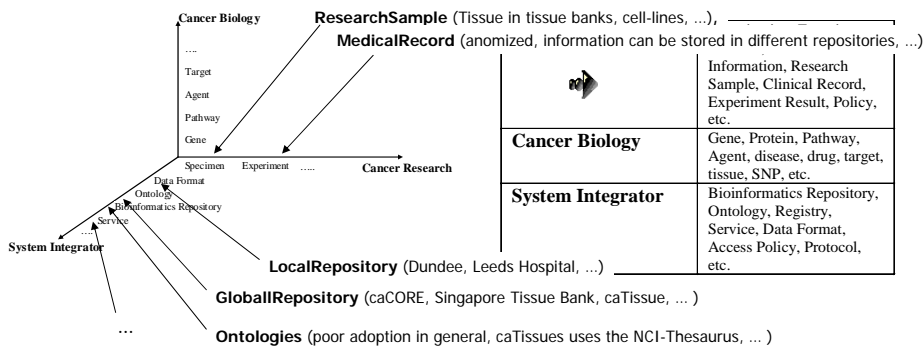
- Characteristics of our domain model
 - High Level
 - Extensible
 - Multi-Dimension to separate different types of concern



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

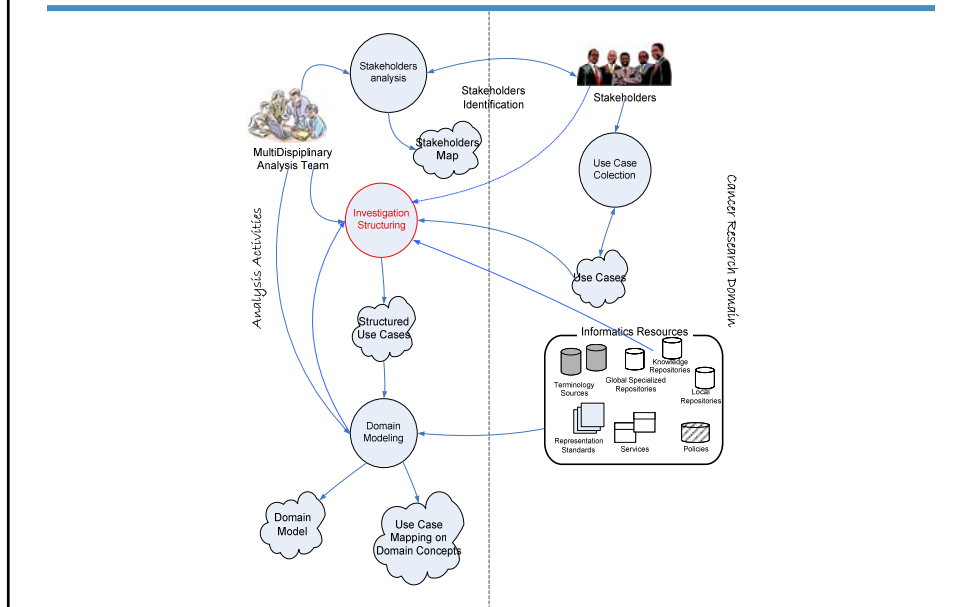
Analysis: Domain Modelling (2)

- Characteristics of our domain model
 - High Level
 - Extensible
 - Multi-Dimension to separate different types of concern



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

Analysis: Process



Structure of Talk

- Context
- Objectives and Challenges
- Analysis
- Final Remarks and Future Steps

Final Remarks

- NCRI Informatics Initiative: primary importance understanding user requirements and early architectural analysis
- Our approach
 - Motivations: exploratory goal, multidisciplinary team, improve communication with stakeholders, field heterogeneity
 - Key aspects
 - Systematic structuring of 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

Future Steps

- Links with key initiatives
 - On-going cooperation with the Imaging&Pathology demonstrator project in Oxford
 - On-going cooperation with the caBIG team (Harmonization workgroup) in defining a common domain model
 - Future cooperation with related UK projects: CancerGrid, eDiamond, EBI, CLEF, myGrid, etc.
- Prototype user interface
 - Used to validate user requirements
- System Prototype (follow-up)
 - Focus on 2-3 sub-domains
 - Validate requirements and integration approach
 - Detailed design of architectural components
 - Evaluate appropriateness, effectiveness and sustainability
 - Estimate development and maintenance costs

Contacts

- For questions about the NCRI Informatics Initiative
Dr Fiona Reddington - Section Head - Community Alliances
Tel: +44 (0)20 7061 8574
Fax: +44 (0)20 7061 8580
email: fiona.reddington@ncri.org.uk

- For questions about the PRM Project
Dr Vito Perrone – Computer Science Dep. - UCL
Tel: +44 (0)20 7679 0375 (Direct Dial)
Internal: 30375
Fax: +44 (0)20 7387 1397
email: V.Perrone@cs.ucl.ac.uk