Modelling-in-the-Large

Anthony Finkelstein

Subtitles:

How Requirements Engineering can cure cancer …

How Requirements Engineering can eliminate diabetes …

How Requirements Engineering can solve the problems of climate change …

What is a Keynote Talk?

An excuse to be 'programmatic'. To ask questions without offering answers. To put forward outrageous generalisations from narrow foundations.

Three initial excursions

related to my major themes

But basically just things
I want to say!

1

As a consequence of the growth of CS as a discipline we largely abandoned direct teaching of software development to engineers and scientists

They learn to program
(mostly through experience)
But they do not learn
software engineering

2

We have focused on advancing research in software engineering and we have neglected our ‘service’ role

We can (and should) contribute as practitioners. What matters is that science advances as a whole and not just our little corner

3 We have a professional responsibility to contribute to the solution of big and meaningful problems

Disease, poverty, environment, crime, security, energy, democracy …

Two thematic threads

(a) Requirements Engineering used to 'understand science'
(b) Requirements Engineering used in 'supporting science'

a

Cancer Research:

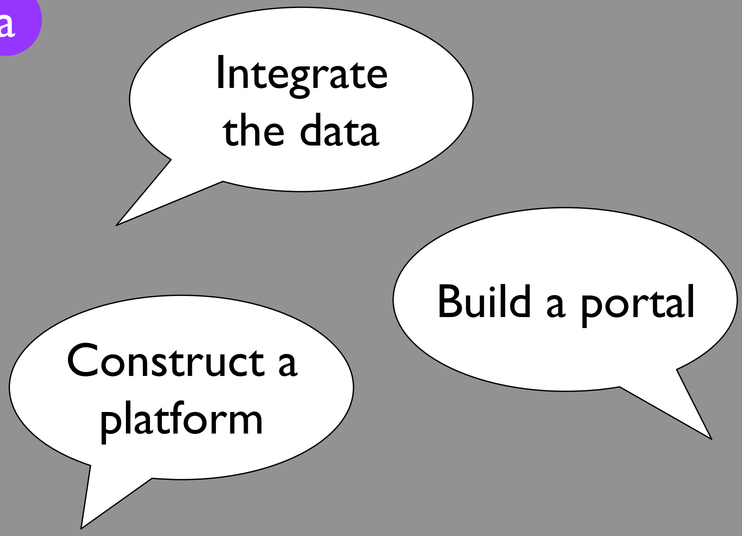
- improved 'high throughput' techniques from molecular biology
- large scale trials data
- growth in scientific literature
- new imaging technologies

a

Has led to a rapid growth of data resources … and some 'services'

Heterogeneous
Distributed
Locally managed
and of variable quality

a



a

But what exactly does this mean? What is data integration? How is this data to be accessed? How is it to be used?

– There is a general sense that there is ‘something’ of significant value in the fused data. But what and how to use it is unknown.

a

The science has created new possibilities that requires genuinely new types of infrastructure and ways of working (processes) in order to realise them

a

What is needed …

- A way of understanding how these new resources might be used
- A way of understanding how scientists work with complex data resources in sync with experimental and clinical work
- A plan for how to build the support necessary for future cancer research

a

In other words … Requirements Engineering is needed to help to build a new set of 'system metaphors' and 'concept of operation' for infrastructure to support cancer research

Scenario and use case analysis
Goal-oriented analysis
Modelling
Systematic elicitation
Observational methods

a

What we have been doing

NCRI Informatics Initiative

[in conjunction with NIH caBIG]

Funded by

CANCER RESEARCH UK

What we have ‘discovered’ …

- The centrality of the ‘investigation’
- The need for ‘lightweight’ data transformation support
- The fuzzy boundary between ‘external’ and ‘internal’ data
- The complexities of scientific ‘data sharing’
- The importance of the ‘community’
- The move away from a static understanding of integration
- Hidden data management problems (versioning)

a

a

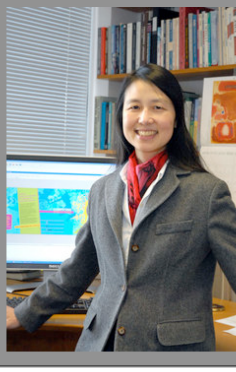
Vision …

Through Requirements Engineering we might effect a radical change in the way that cancer scientists work with their data and make a small contribution to finding a ‘cure for cancer’

b

An influential paper …

Computational Thinking, Jeanette Wing, CACM, March 2006, 49, 3



b

"Computational thinking is using abstraction and decomposition when attacking a large complex task or designing a large complex system. It is separation of concerns. It is choosing an appropriate representation for a problem or modeling the relevant aspects of a problem to make it tractable. It is using invariants to describe a system's behavior succinctly and declaratively. It is having the confidence we can safely use, modify, and influence a large complex system without understanding its every detail. "

b

A quick switch of focus

A ‘reductionist’ scientific agenda has been highly successful … now we need to complement this with the ability to use the knowledge we have gained to understand ‘complex systems’

b

Example: Biomedicine

The Molecular Revolution

b

A revolution that has reshaped the life sciences

We now understand:

- the DNA sequence of many genes, up to whole genomes
- the mechanics of much of RNA synthesis
- the genetic code for specifying amino acids so that the backbone of a protein can be directly predicted
- the means by which one gene can generate many RNAs and therefore proteins

b

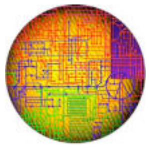
- how DNA sequences, called promoters, determine which genes are expressed
- how DNA binding proteins, called transcription factors, modify gene expression

Knocking-out and over-expressing genes and RNAs have revealed how particular genes contribute to certain biological processes.

b

The focus now is turning to whole biological systems: the heart, the cardiovascular system, the brain, the liver and to 'complex diseases' such as cancer …

systems biology



b

To succeed it is necessary to combine information from the many rich areas of biological information. Alongside the *genome*, our knowledge about genes, we place the *proteome*, *metabolome*, and *physiome,* our information about proteins, metabolic processes, and physiology

So as to build an *integrated* physiology of whole systems

b

The prize to be attained is immense!

- From 'in-silico' drug design and drug testing
- To individualised medicine that will take into account physiology and genetic profile

Systems biology has the potential to have a profound impact on healthcare and beyond.

b

Even if we had a catalogue of all the gene sequences, how they are translated to make proteins, which protein can interact with which, and the way in which the protein back bones fold, it is not possible to put them into a functionally meaningful framework

b

Cataloguing is not Understanding

- All proteins are post-translationally modified. These additions influence the actual shape of proteins
- Just because two proteins can interact, it does not mean that they do so in real cells
- Many functionally important, small molecules are synthesized by metabolism

b

Modelling

A bottom-up, 'data-driven' strategy, will not work — it is not possible to build an understanding of biological systems from an understanding of the components alone

What other approaches might be tried?

b

Use experimental information to build models at different biological scales, integrating these models to create an ‘orchestrated’ assemblage of models ranging from gross models of physiological function through to detailed models that build directly on molecular data

b

What does this mean …

- Matching modelling schemes to the systemic phenomena that are of interest (structural and behavioural, discrete and continuous)
- Integrating heterogeneous modelling schemes with different structuring schemes
- Synthesising the results of analysis
- Correlating the models with ongoing experimentation

b

What does this mean …

- Supporting the construction and evolution of complex models
- Supporting global collaboration around the models
- Supporting 'contested science' - conflict, inconsistency
- Supporting the model-experiment loop

I think this looks familiar!

b

But the scale is beyond anything familiar

- From gene to cell, from cell to organ, from organ to whole body physiology
  – And ultimately to individuals

Compare this with the current situation in medical informatics

b

This is a real Grand Challenge for software engineers and specifically for Requirements Engineering

And it applies also to climate modelling …
(and many others)

b

What I (actually we) have done:

A framework for selecting modelling schemes

A metamodel for systems biology

A model parameter repository

A prototype middleware for integrating heterogeneous modelling platforms

A unified ontology framework for systems biology models

A new versioning and impact analysis tool for systems biologists

(the first) modular, integrative, scale-crossing, hybrid model of liver glucose homeostasis

By way of a conclusion …

We have much to contribute

We should look outward rather than inward