Metadata in the European e-Infrastructure

Keith G Jeffery
Science and Technology Facilities Council
Rutherford Appleton Laboratory, OX11 0QX UK
e-mail: keith.jeffery@stfc.ac.uk

STRUCTURE

Introduction

e-Infrastructure: GRIDS

Research Challenges

Metadata

SOA

Information Systems Engineering Solution

Conclusion

Rutherford Appleton Laboratory

Alliance for Permanent Access

The Alliance for Permanent Access aim is to foster the development of an ecosystem of trusted digital repositories to enable Europe to fully exploit the potential of European scientific collaboration. The accessibility (permanent), quality assurance and preservation of key data collections are and will be one of the major driving forces for the advancement of science but also for applications thereof.

http://www.alliancepermanentaccess.eu/

Annual Conference 24th November 2009

Science & Technology
Facilities Council

ERCIM

European Research Consortium
for Informatics and Mathematics

www.ercim.org

20 European countries - major labs or consortia of universities
12000 ICT researchers
- Working groups
- Fellows programme
- Cor Baayen Award

Strategy documents for EC and national governments
R&D projects, networks of excellence etc
> 200 spin-out companies, strategic seminar, innovation magazine
European Office(s) of W3C and on W3C Steering Committee
ERCIM News

www.eurocris.org

Linking together systems in each country managing research information

- Funders of research
- Organisations performing research

For

- Strategic decision-making about what research to fund /do
- Finding research partners and competitors
- Finding innovative ideas for technology transfer / exploitation
- Informing the media / public

But first...Context

- Internet
  - 1.5 billion fixed connections
  - Estimated 4 billion mobile connections
- Digital Storage
  - Estimated 280 billion Gigabytes
- Users :
  - Asia 550 million 14% penetration
  - Europe 350 million 50% penetration
  - USA 250 million 70% penetration
- Expect technology to grow ~ 1 order of magnitude each 4 years
  - and accelerating

Scalability
Trust & security & privacy
Manageability
Accessability
Useability
Representativity

But first…Context

- The vast majority of computers – 98% - do not have keyboard, mouse, screen
  - They are in cars, planes, washing machines, mobile phones
- The most-used operating system is NOT Windows (or even Unix/Linux)
  - Symbian in mobile phones or specialised operating systems (e.g. Contiki) in embedded systems

Science & Technology Facilities Council

But first…Context

- The number of computers will vastly out-number humans on the planet very soon;
- Everything will be computerised;
  - Sensor networks
    - Home, healthcare, environment, industrial processes, transport systems….
  - Control systems
    - Industrial, transport, home (central heating)…

So…

- This is the ‘internet of things’ or ‘future internet’
- We need to :
  - Manage the huge numbers, sizes
  - Integrate the different kinds of systems
  - Into one environment leading to human decision-making
    - whether managing a business, shopping, media choice, social interaction
- But there is a problem…in last 20 years
  - Data storage density increased ~10**18
  - Processor power increased ~10**15
  - BUT broadband capacity increased ~10**4

- This has implications for Information Systems Engineering!
- In fact the requirement and limitations challenge the very basis of traditional computer science / ICT

So…

- This is the 'internet of things' or 'future internet'
- We need to :
  - Manage the huge numbers, sizes
  - Integrate the different kinds of systems
  - Into one environment leading to human decision-making
    - whether managing a business, shopping, media choice, social interaction
- But there is a problem…in last 20 years
  - Data storage density increased ~10**18
  - Processor power increased ~10**15
  - BUT broadband capacity increased ~10**4

IN FACT WE HAVE TO MAKE ICT **AUTONOMIC**
AND TO DO THIS REQUIRES **METADATA**

STRUCTURE

Introduction

e-Infrastructure: GRIDS

Research Challenges

Metadata

SOA

Information Systems Engineering Solution

Conclusion

The GRIDs Vision

- The end-user interacts with the GRIDs environment to clarify the request
  - using a ‘device’ or ‘appliance’
- The GRIDs environment proposes a ‘deal’ to satisfy the request
  - which may or may not involve money
- The user accepts or rejects the ‘deal’

The GRIDs Vision

The GRIDs environment is such that

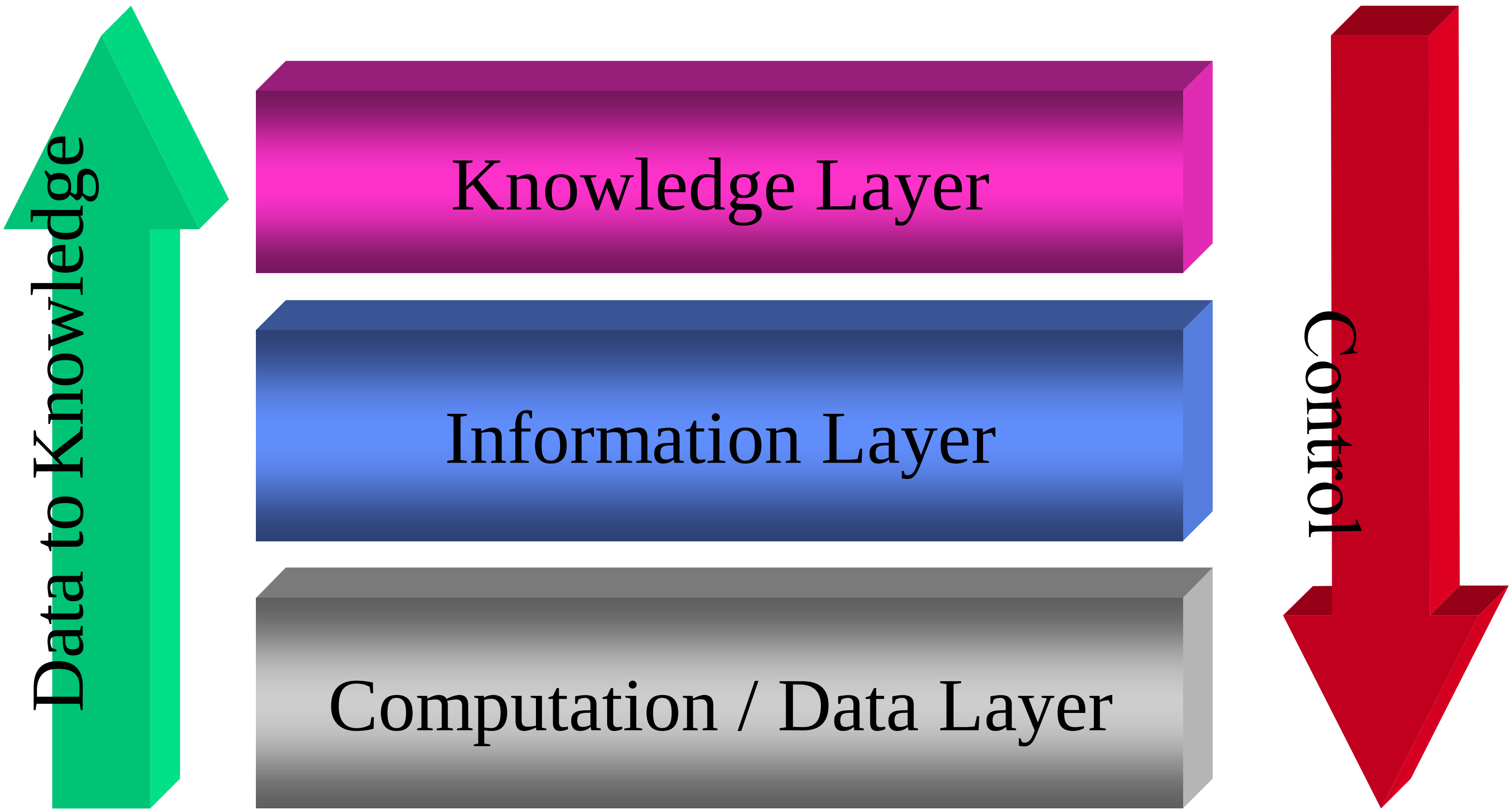
- A user can interact with it intelligently
- It provides transparent access to
  - data, information, knowledge
  - computation
  - instrumentation / detectors

Note.....

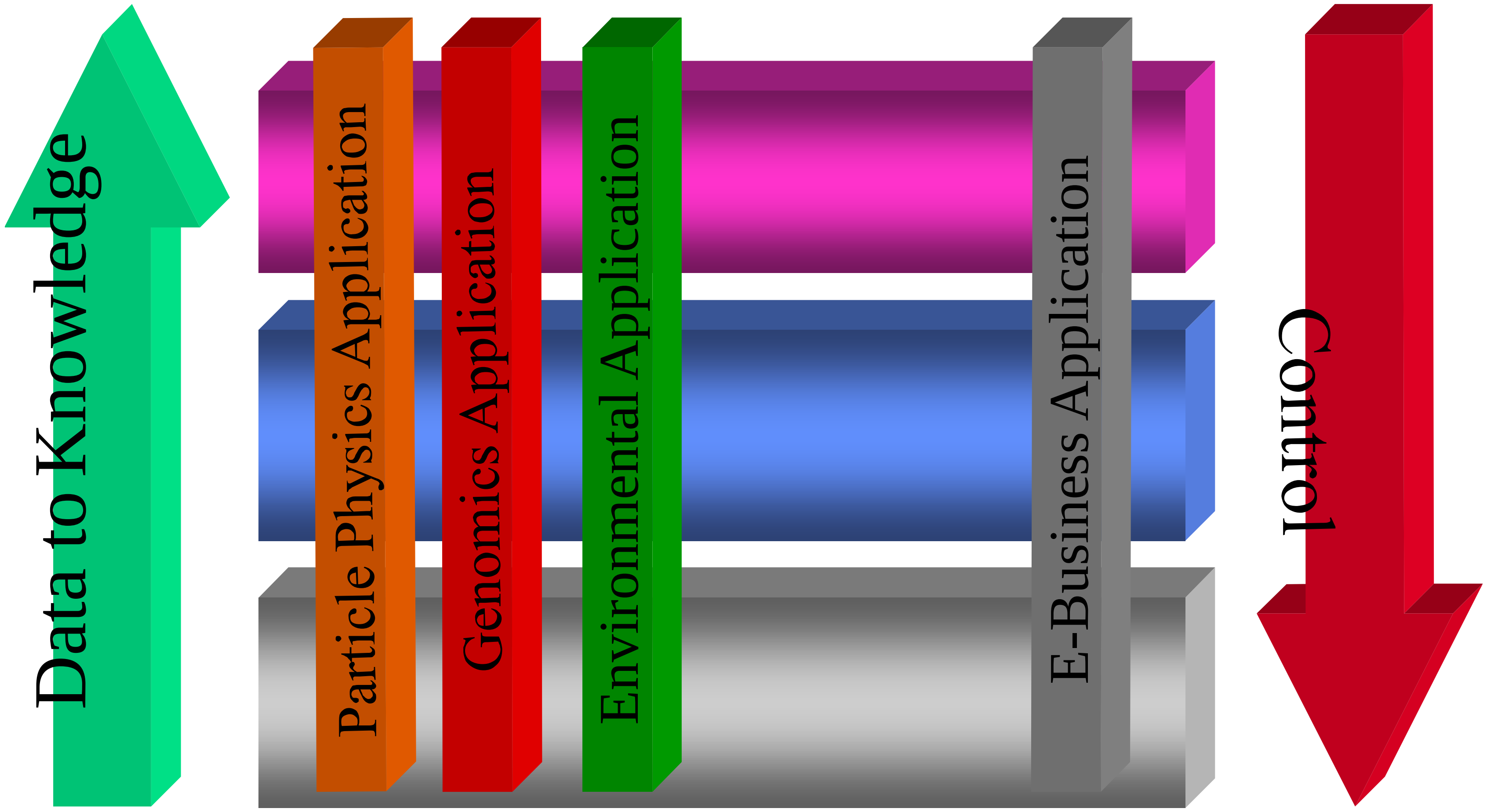
The requirements as originally defined for GRIDs imply existing e-infrastructure solutions are too low-level (programmer not end-user level)

– Insufficient representativity
– Insufficient expressivity
– Insufficient resilience
– Insufficient dynamic flexibility

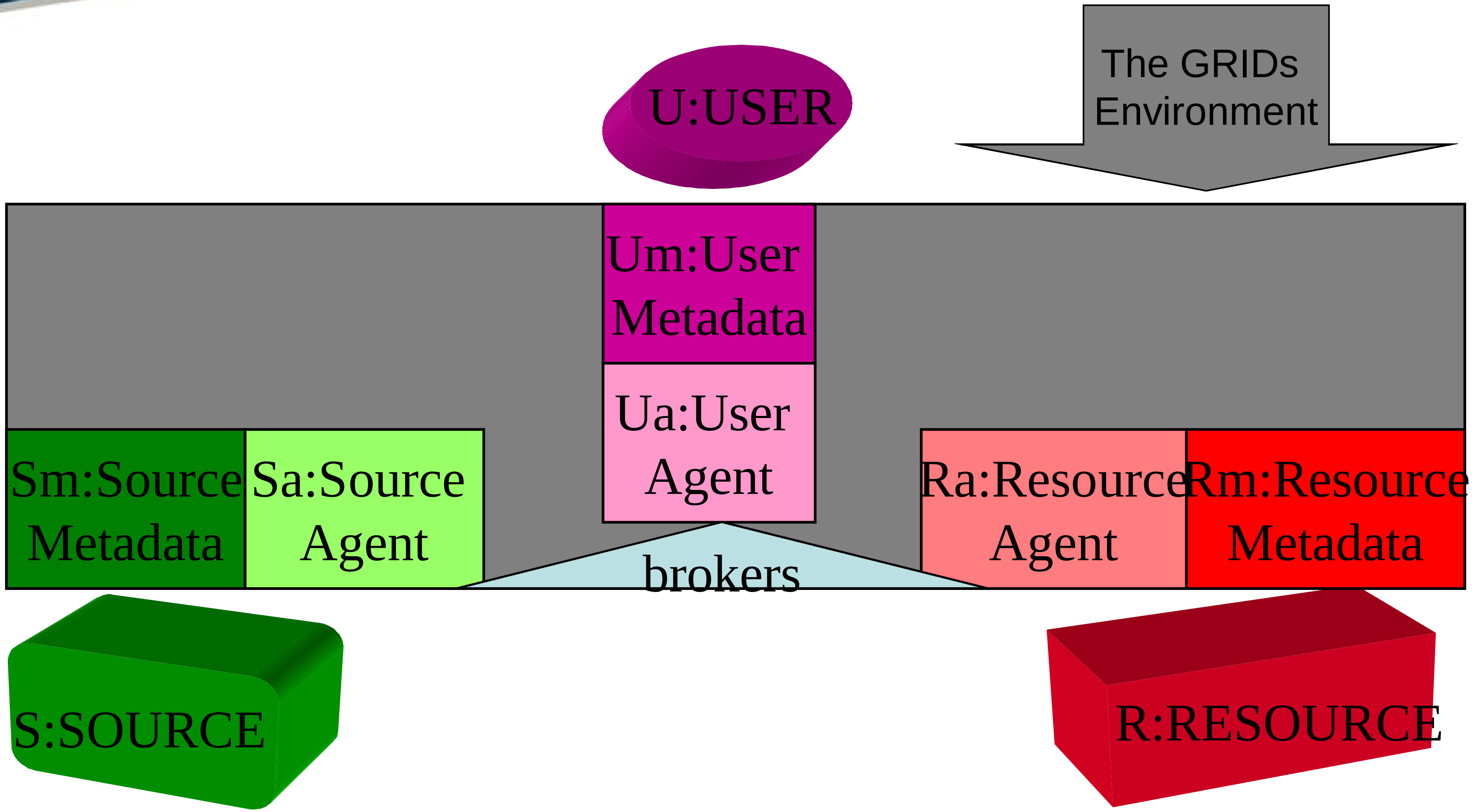
The GRIDs Architecture



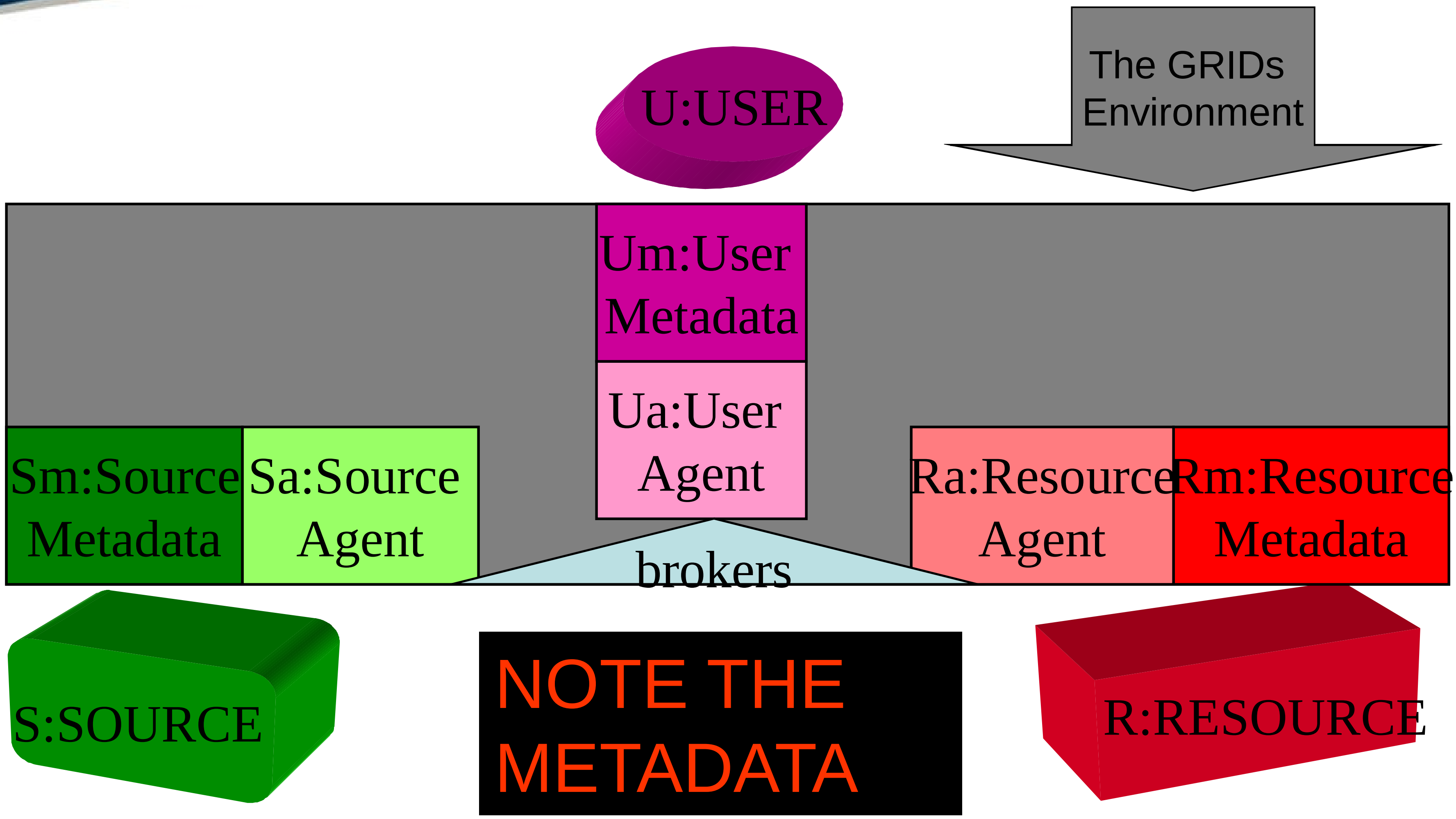
The GRIDs Architecture



A POSSIBLE ARCHITECTURE



A POSSIBLE ARCHITECTURE



STRUCTURE

Introduction

e-Infrastructure: GRIDS

Research Challenges

Metadata

SOA

Information Systems Engineering Solution

Conclusion

Research Challenges: 1

Metadata

- the need for metadata related to services, data/information/knowledge, agents;
- what is data, what is metadata?
- kinds of metadata and their use;
- representation and structure - syntax;
- semantics (meaning);

Research Challenges: 2

Management of state

- detection of state across millions of individual nodes;
- maintenance of state across many nodes;
  - transactions and locking;
  - roll-back and compensation;

Research Challenges: 3 (1)

Data representativity

- data structures representing real-world inter-relationships;
  - data attribute value encoding (character set, media encoding), types, lengths;
  - data attribute value language;
  - fully connected graphs – the death of the hierarchy;
  - the time-machine: temporal duration of the inter-relationships;
  - certainty, probability of the inter-relationships
  - Incomplete and inconsistent information

Research Challenges: 3 (2)

Data representativity

- Interoperation
  - reconciliation of different data structures representing a similar real-world domain;
- data location / locality and replication
  - for business continuity;

Research Challenges: 4

Data quality, veracity and permanency

- detection of quality against metadata parameters e.g. precision, accuracy;
- provenance;
- temporal recording;
- data curation across media and policy evolution;

Research Challenges: 5

Trust, security and privacy

- policies declared, enforced and monitored through restrictive metadata;
- policy reconciliation for interoperation;

Research Challenges: 6

Management of service levels and quality of service

- policies declared, enforced and monitored through restrictive metadata;
- service level negotiation (e.g. lower price for lower performance);

Research Challenges: 7

Systems design, development, maintenance and decommissioning

- based on strong separation of:
  - services (processes),
  - data, information and knowledge
- assuming self-composition, self-managing and adjusting, self-maintaining properties ;
- assuming mobile code properties

STRUCTURE

Introduction

e-Infrastructure: GRIDS

Research Challenges

Metadata

SOA

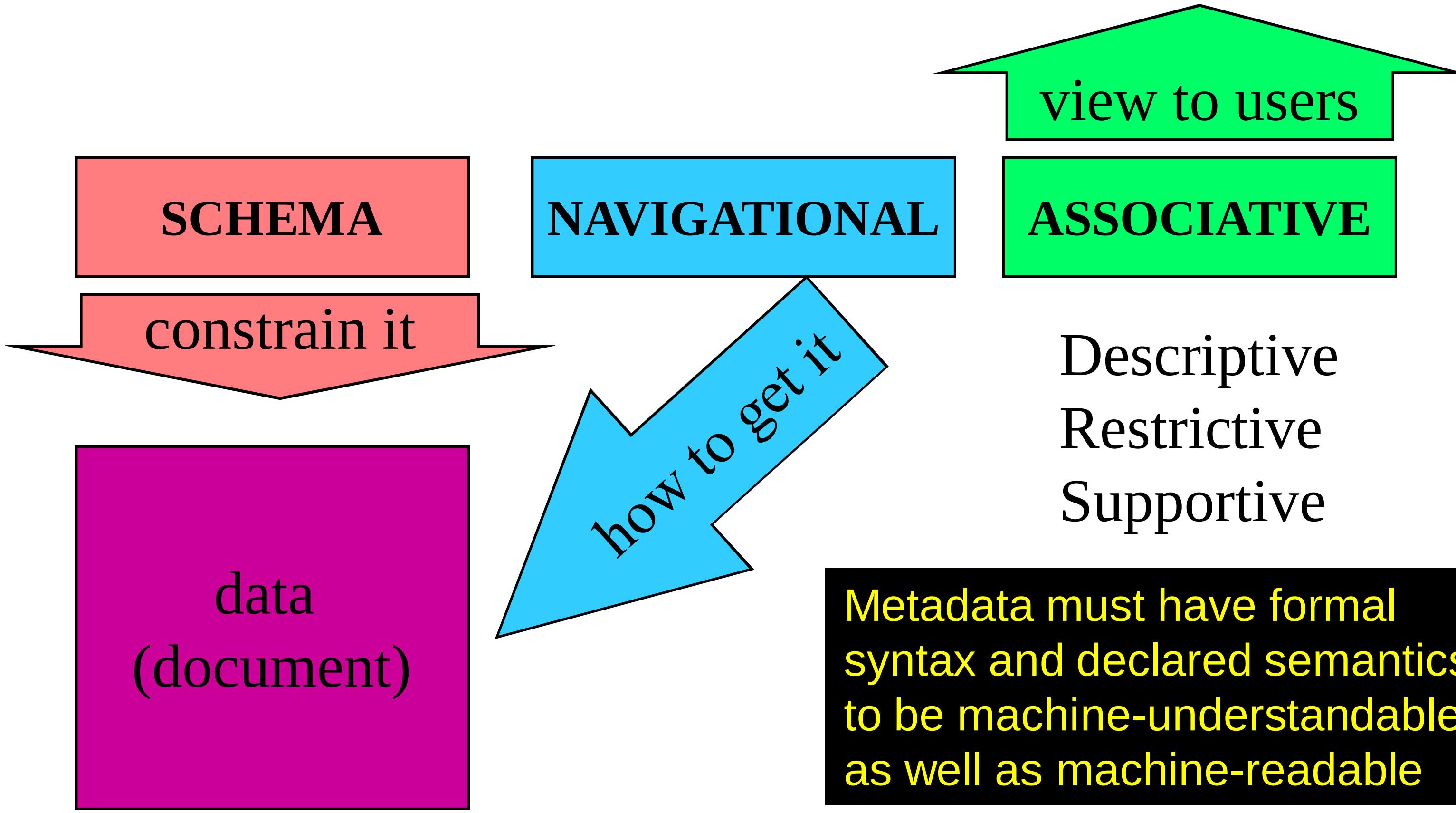
Information Systems Engineering Solution

Conclusion

Metadata

- Note importance in the GRIDs (e-infrastructure) concept
  - UK e-Science programme
  - Further EC projects
- Note first of the major challenges to build the e-infrastructure
  - To meet the requirements defined back in 1999
- Note require metadata to provide a solution to all the other 6 challenges

Classification of Metadata



Metadata must have formal syntax and declared semantics to be machine-understandable as well as machine-readable

STRUCTURE

Introduction

e-Infrastructure: GRIDS

Research Challenges

Metadata

SOA

Information Systems Engineering Solution

Conclusion

SOA Concepts

- Code Modules (units, atomic units, components)
- {subroutines, coroutines, agents/daemons}
- Offered as service
- With description for
  - Discovery
  - Utilisation (composition)

SOA Experience

- Obvious benefits not realised – Why?
- Compare with other engineering
  - Standardisation of interfaces
  - Documentation of product and interfaces in detail
    - Description [WSDL]
    - Composition /orchestration [WS-BPEL]
    - Choreography [WS-CDL]
    - Execution <various>
    - Maintenance <?>
    - Conditions of use <various>

SOA - Solution

- Standardisation of interfaces
- Interface to :
  - – Describe for discovery
  - – Describe for composition
  - – Describe during execution

{ functional

Non-functional

SOA - Solution

- Standardisation of interfaces
- Interface to :
  - Describe for discovery
  - Describe for composition
  - Describe during execution

{ functional / Non-functional }

THIS REQUIRES METADATA

STRUCTURE

As a mature discipline...

We should learn from those more mature

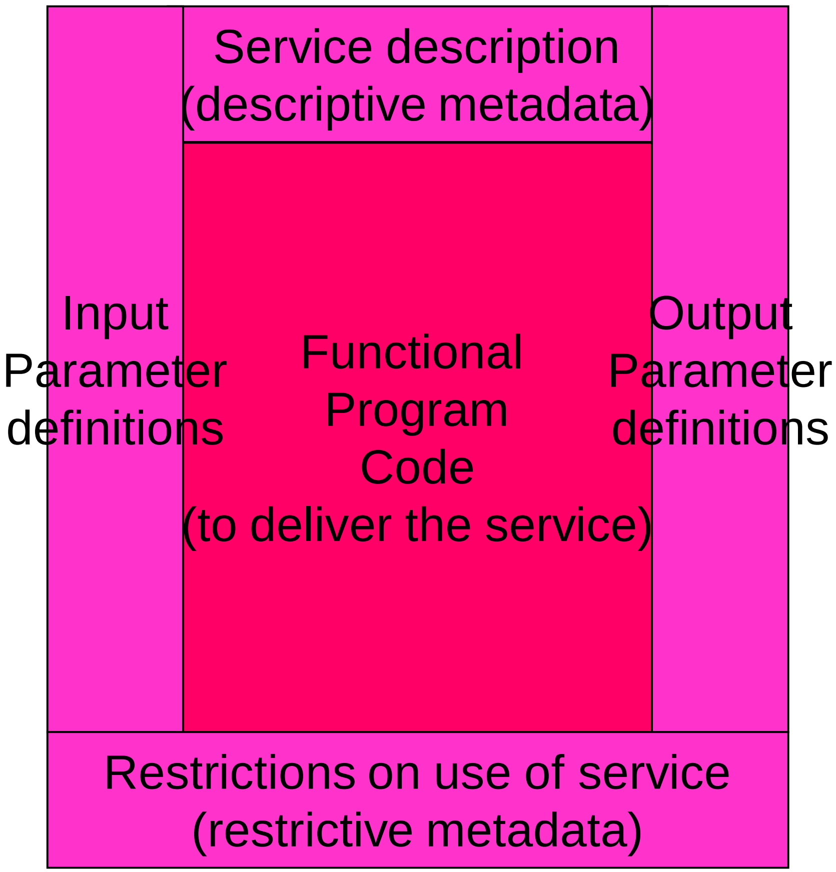
-Do not reinvent (the wheel)

-Novel solutions not always required

-From real engineering:

- Standard components
- Well-defined functional and non-functional properties (metadata)

Services: Model 1



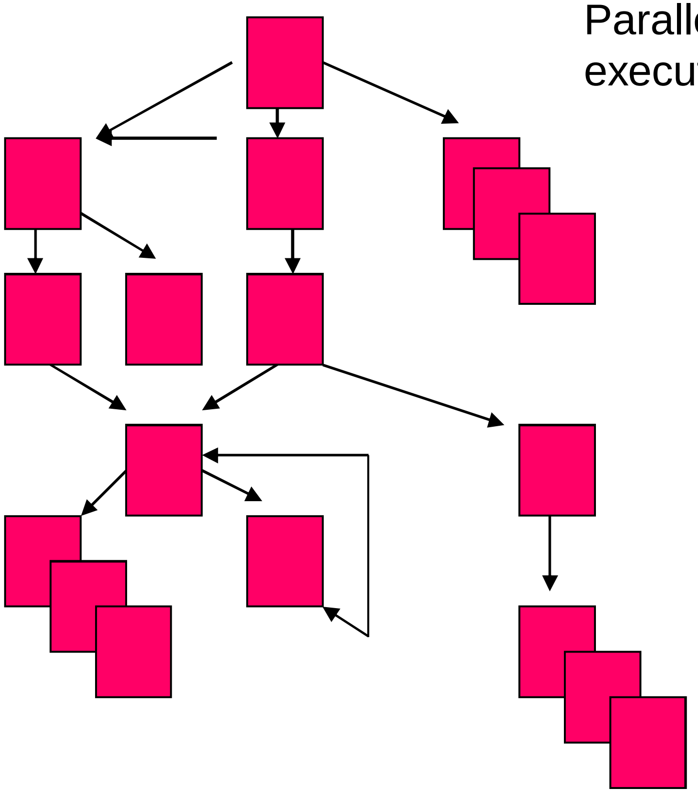
Description

- Location
- Requirements matching
- Composing
- Utilising

➔ metadata

Services: Model 2

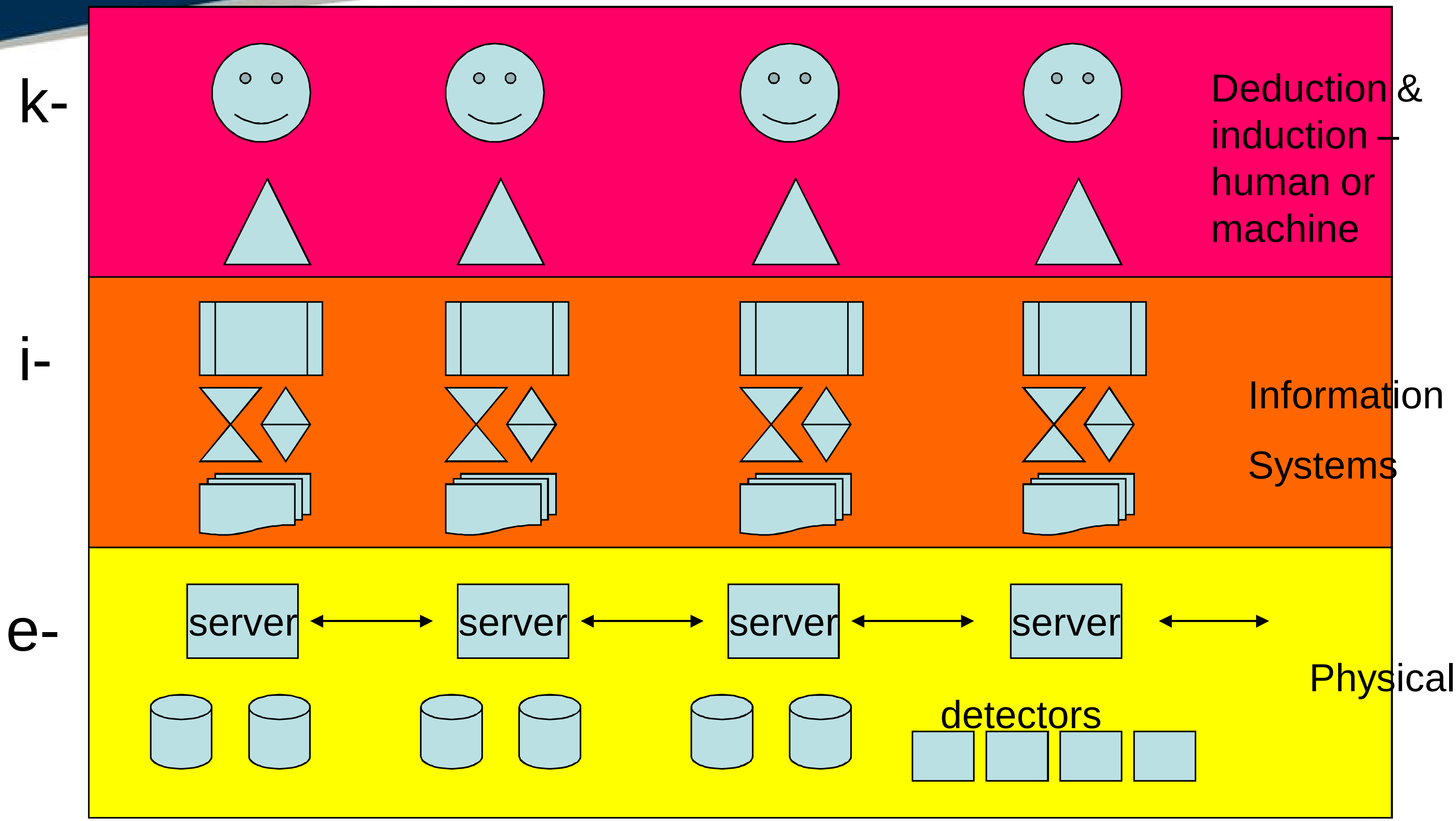
Multiple Instances Parallel execution



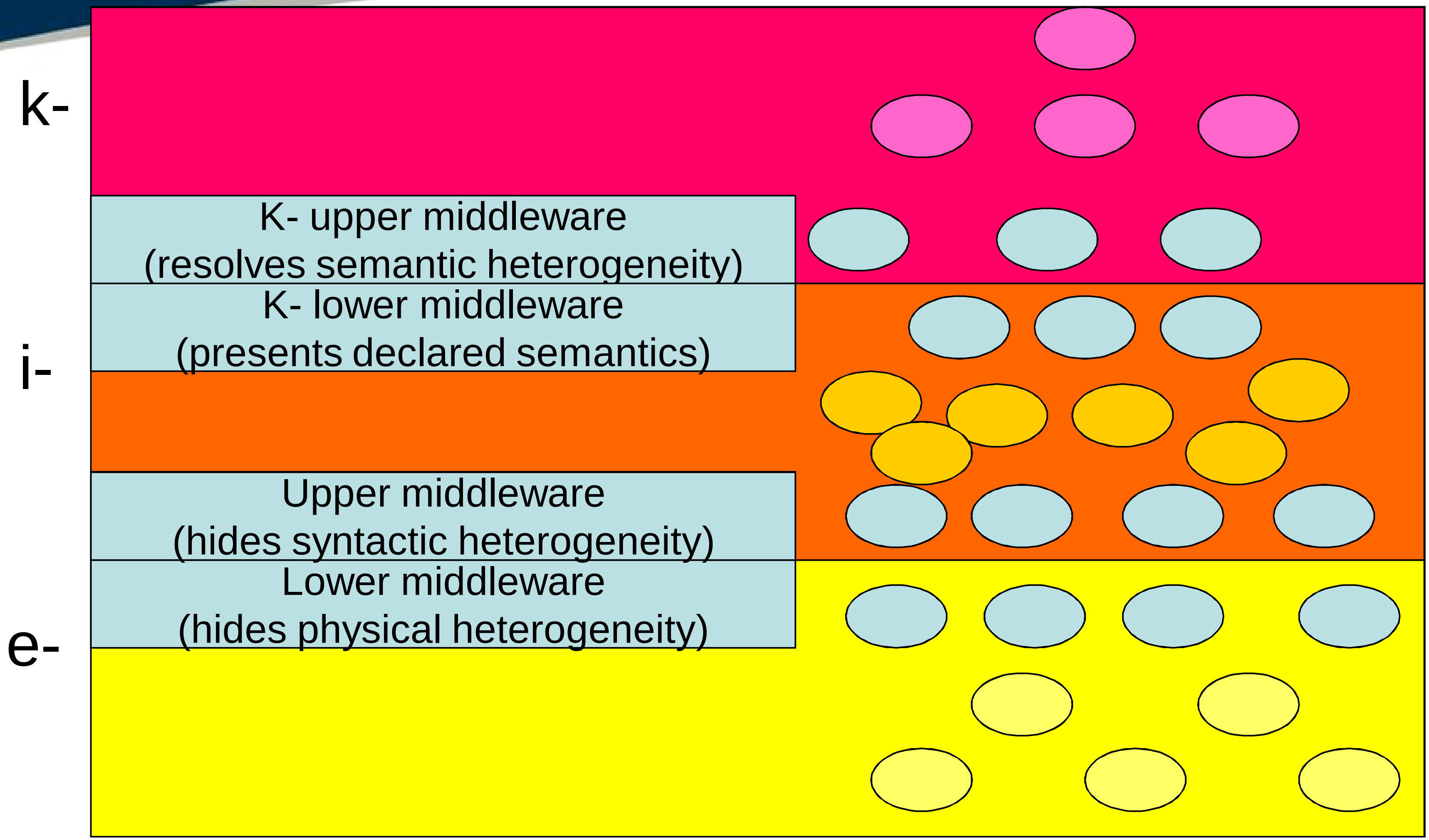
Composition

- End-to-end FR satisfaction
- End-to-end NFRs satisfaction
- Avoiding emergent properties
- Conditions of use of services
  - Processes
  - wrapped with data
  - wrapped with processing, storage etc
  - wrapped with real estate
  - wrapped with staff
- → metadata

e-,i-,k- infrastructure



Middleware – and as SOKUs



STRUCTURE

Introduction

e-Infrastructure: GRIDS

Research Challenges

Metadata

SOA

Information Systems Engineering Solution

Conclusion

Conclusion

There is a way forward for the e-infrastructure

It is based on bringing together all the relevant strands of R&D

- GRIDs
- Clouds
- SOA
- Web 2.0

Supported by:

Formal software engineering

Formal data/information/knowledge representation

Advanced standardised metadata

New systems development methods

Prof. Keith G Jeffery
CEng, CITP, FBCS, FGS, HFICS
Director, IT & International Strategy

STFC Rutherford Appleton Laboratory
keith.jeffery@stfc.ac.uk