Challenges & Opportunities in Medical Big Data

Prof Richard McClatchey UWE, Bristol UK

Experiences from EC Grid/Cloud Projects:

MammoGrid 2002-2005

Health-e-Child 2006-2010

NeuGRID & N4U 2008-2015

Richard McClatchey, CBMS 2015 Sao Carlos 23rd June 2015

1

- Generation of Big Data
- Its management
- The role of meta-data
- Past, & present challenges
- Future directions & concerns

Medicine is Data Driven

David Gotz
IBM T.J. Watson Research Center

BIG (medical) DATA : Scale

- Google processes 100 PB a day (2015) 10^{17} bytes
- Facebook has >100 PB of user data + 20-25 TB/day (12/2014)
- eBay has >100 PB of user data + 100 TB/day (12/2013)
- CERN's Large Hadron Collider (LHC) generates 20-25 PB a year
- **US Healthcare data in 2011 150 exabytes (1.5 * 10^{20} bytes) and by 2020 estimated to be of the order of yottabytes (10^{24} bytes)**
- Digitisation in medicine is producing a **tsunami of data**

640K ought to be enough for anybody.

Big data produced by…

- Development of continuous monitoring and remote sensors

- Massively increased imaging data (MRI, CAT, PET, US)

- Genomics, proteomics, NGS methods

- Data from heterogeneous sources in multiple places (electronic medical records, lab and imaging systems, physician notes, diagnoses / test results, correspondance)

- Unstructured data & ***meta-data***

- So, new approaches needed for accessing, manipulating, visualizing

- Requires entirely new perspective

So, What Can We Do With All This Data?

David Gotz
IBM T.J. Watson Research Center

We need good data management...

...helps research to be:

Replicated and verified

Preserved for future use

Hetereogeneously integrated

Linked with other research products

Shared and reused

...helps researchers:

Meet funding requirements

Increase visibility of research

Save time and effort (avoid data loss)

Deal with an ever-increasing amount of data

http://www.healthcare-informatics.com/article/guest-blog-data-management-challenge-unlocking-value-clinical-data-many-times-requires-enter

neuGRID for Users (N4U) : Services 4 Users

- EU Framework 7 Integrated Infrastructure Initiative, I3
- Started July 2011, 42 months, funded at €3.5M
- To provide: an e-Science environment by developing and deploying the neuGRID infrastructure to deliver a Virtual Laboratory framework to offer neuroscientists access to a wide range of datasets, algorithm applications, and access to computational resources, services, and support
- **Partners:**
  - **IRCCS Fatebenefratelli, Italy; , University of West of England, Bristol, UK;, Maat G Knowledge Spain;**
  - **Hospital University of Geneva, Swizerland. VUmc - Vrije Universitet Medical Center, Amsterdam, NL**
  - **Karolinska Institutet, Stockholm, Sweden;, CNRS, France, CEA, France;**
 - **CF consulting, Milano, Italy, MNI Montreal, Canada , UCLA, USA**

Past & Current example NeuGRID & N4U

neuGRID

neuGRID for Users, N4U

User services

- Cortical thickness pipeline
- Core databasing
- Web portal
- LONI WMS

- Wider multimodal software portfolio for researchers and diagnostic neuroscientific communities
- Advanced Data Base Management system
- More representative datasets
- Data protection extension
- Educational programs

GRID services

- Security Services
- Medical Quering Services
- Provenance Services (CRISTAL)
- Grid Gluing abstraction Services

- Knowledge management
- Analysis services
- Workflow authoring extension
- Advanced querying extension

Infrastructure services

- Enactment Services
- Computing Services
- Storage Services

- Computational resources expansion
- Cloud compatibility development

Centre for Complex Cooperative Systems

UWE BRISTOL

It's a just a framework
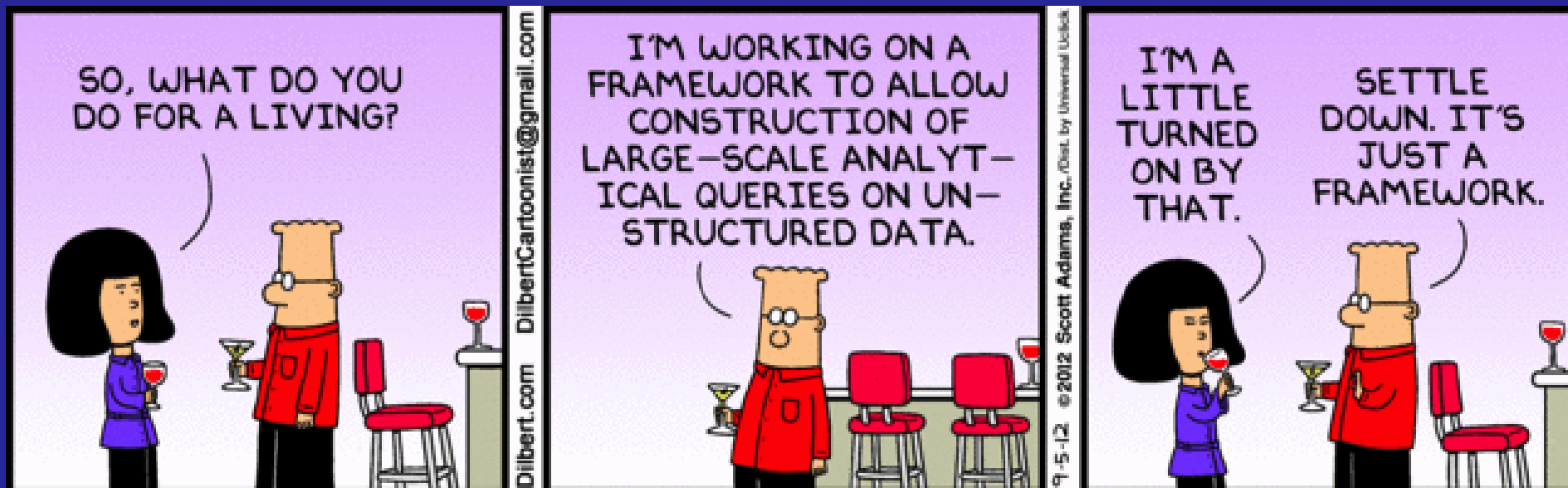
Dilbert Sept 5, 2012

Data Management Challenges in Medicine

Challenge 1 : Early 2000s **Can we collate & curate ?**
HealthGrids, Databases, Infrastructures

Challenge 2 : Late 2000s **Generate Information from data ?**
Data fusion, meta-data defintion, service-oriented architectures, querying & workflow systems.

Challenge 3 : Early 2010s **Making information useful ?**
Adding semantics to data, portals & visualisation, knowledge management, analysis/provenance services

Example : N4U's Virtual Laboratory
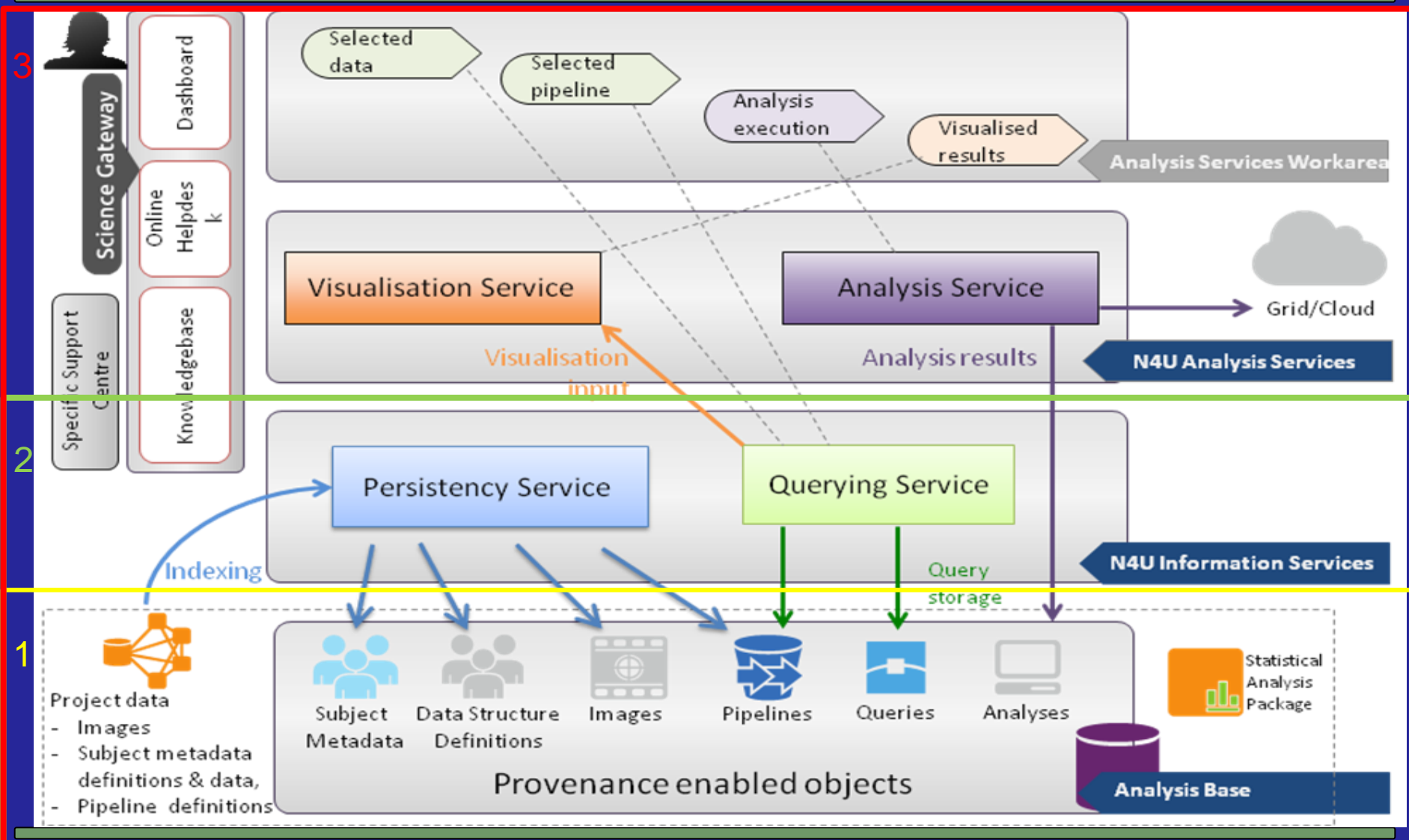
The rise of meta-data in analyses

- Data about data : **get information from data**

- Data (and algorithm) provenance - 7 W's:

  - **who** ran an analysis (username, role, identifier),

  - for **what** purpose, **what** the analysis aimed to achieve,

  - **what** were its outcomes/results

  - **when** was it run

  - **where** it was run

  - **which** datasets and algorithms were used

  - **how** it was executed

  - and lastly **why** the analysis was run

- Meta-data may soon take more space than raw data….

Data Management Challenges in Medicine

Challenge 1 : Early 2000s **Can we collate & curate ?**
HealthGrids, Databases, Infrastructures

Challenge 2 : Late 2000s **Generate Information from data ?**
Data fusion, meta-data defintion, service-oriented architectures, querying & workflow systems.

Challenge 3 : Early 2010s **Making information useful ?**
Adding semantics to data, portals & visualisation, knowledge management, analysis/provenance services

Where next ? : Late 2010s **Provide precision medicine**
Information -> Knowledge, Privacy & security frameworks, Standardised annotations & ontologies, Enhanced EHR, Transdisciplinary integration, Embedded decision support

Future : Ontologies & Semantics

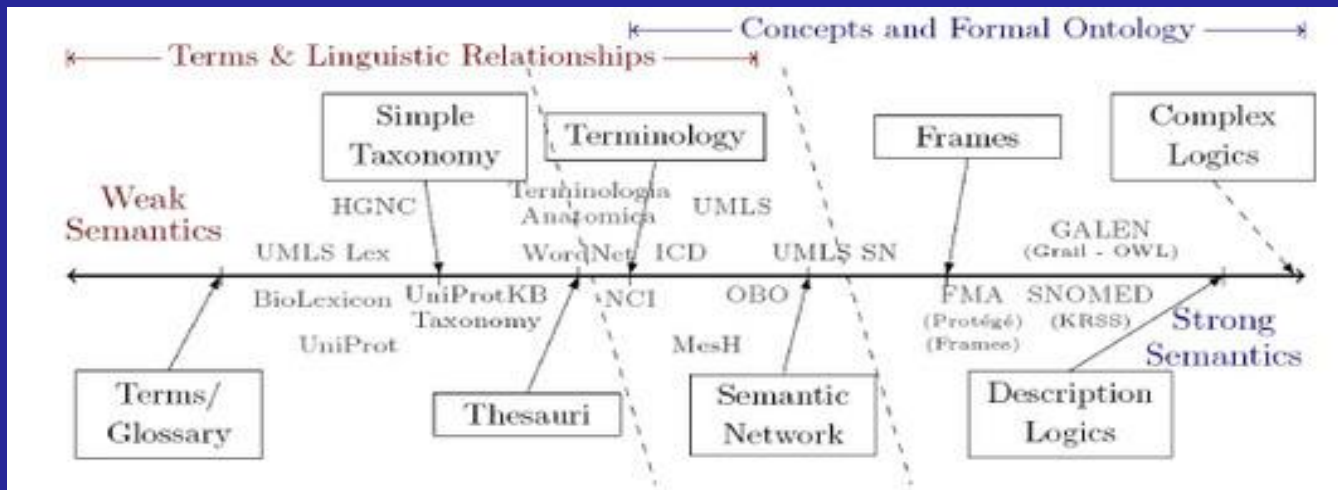- **Knowledge Management**
  - **Annotate data (such as genomes)**
  - **Access information (search, find, and retrieve)**
- **Data integration and exchange**
  - **Model dynamic cellular processes**
  - **Identify Drug Interactions**
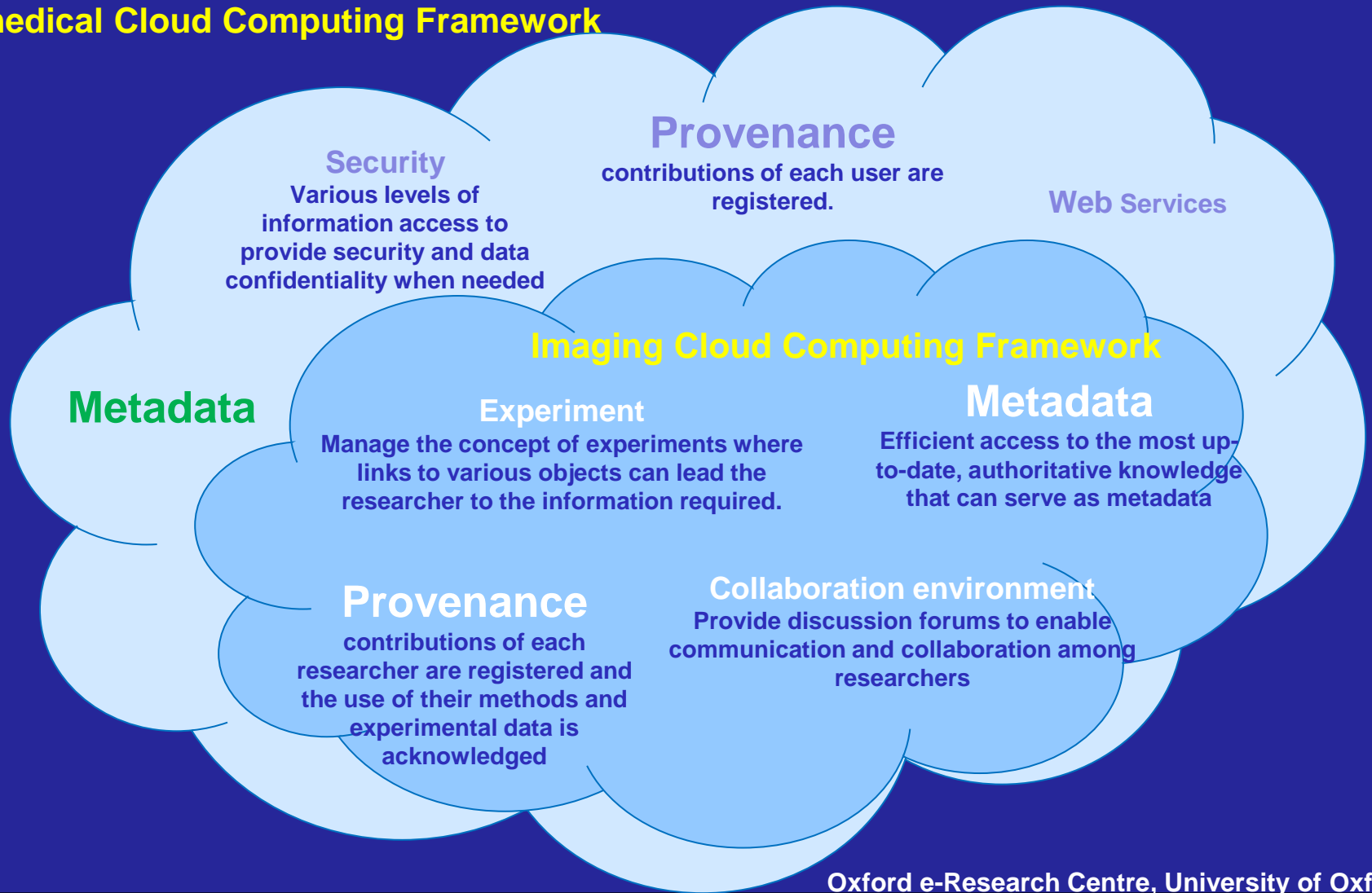
Many existing formalisms :

Weak Semantics

Strong Semantics

http://www.mkbergman.com/wp-content/themes/ai3v2/images/2007Posts/070501d_SemanticSpectrum.png

Future: Cloud Computing & Meta-data

Biomedical Cloud Computing Framework

Security
Various levels of information access to provide security and data confidentiality when needed

Provenance
contributions of each user are registered.

Web Services

Metadata

Imaging Cloud Computing Framework

Experiment
Manage the concept of experiments where links to various objects can lead the researcher to the information required.

Metadata
Efficient access to the most up-to-date, authoritative knowledge that can serve as metadata

Provenance
contributions of each researcher are registered and the use of their methods and experimental data is acknowledged

Collaboration environment
Provide discussion forums to enable communication and collaboration among researchers

Oxford e-Research Centre, University of Oxford, UK

Centre for Complex Cooperative Systems

UWE BRISTOL

Future : Transdisciplinary integration

Materials, energy, IT

Ubiquitous computing

Ambient intelligence
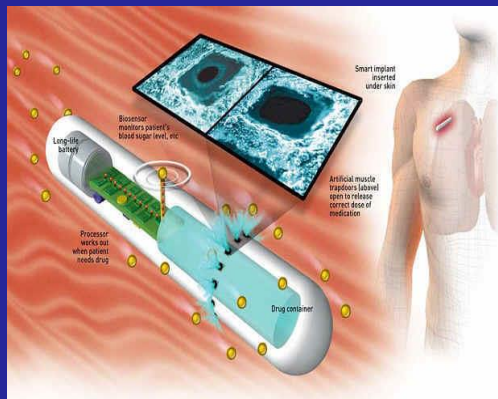
Embedded intelligence Smart pills

Neuron on chip

wireless

POTS

Cellular

WLAN

Rehabilitation engineering
Monitoring
Sensors: EEG, glucose,blood, DNA, …
Add-ons: vision, hearing, implants, …
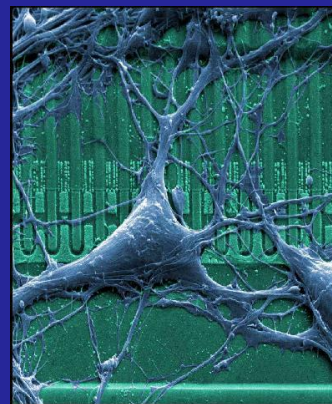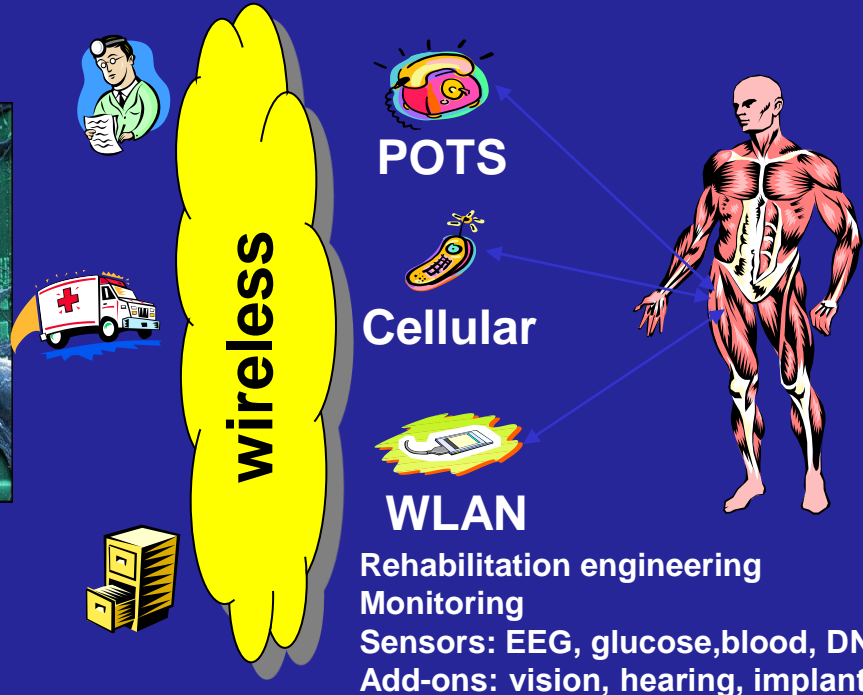
Prof. Dr. Ir. Bart De Moor ,ESAT-SCD K.U.Leuven / IBBT

Centre for Complex Cooperative Systems

UWE BRISTOL

Future:Embedded decision support systems

- Assistive health and wellness management systems

- Health telematics

- Intelligent environments, ambient intelligence, smart homes, home networks

- Home health monitoring and intervention

- Health vaults: personal medical data collection and processing

- Wearable sensor signal processing/wireless registration of physiological parameters

…and then there's ethics

"Make sure everything is done ethically. Within reason, of course."

Nicholas Anderson, PhD
NeuroDevNet, Toronto, Canada

Richard McClatchey, CBMS 2015 Sao Carlos 23rd June 2015

26

Future : Information security concerns

- Multilateral security for **community-centric healthcare IT** platforms

- System and software security of **critical community** (e-health) infrastructures

- Enabling technologies for **collaborative work in the e-health sector**

- Policy negotiation, enforcement and compliance

- **Privacy preserving** data-mining and statistical databases

- Body Area Networks (implanted devices, **wearable devices**,…) and Personal Area Networks

- E-government : **identity management**, delegation, controlled data exchange

Not everyone agrees ...

Journal of the American Medical Association (JAMA) June 22nd 2015, Michael Joyner MD and Nigel Paneth MD :

"the assumptions underpinning personalized medicine have largely escaped questioning. In this Viewpoint, we seek to stimulate a more balanced debate by posing 7 questions for the advocates of personalized medicine".

They conclude: "Even though personalized medicine will be useful to better understand rare diseases and identify novel therapeutic targets for some conditions, the promise of improved risk prediction,behavior change, lower costs, and gains in public health for common diseases seem unrealistic. Proponents of personalized medicine should consider tempering their narrative of transformative change and instead communicate a more realistic set of expectations to the public."