

From Data Discoverability to Data Navigability

Tito Castillo

Application Architect, University College London Hospitals NHS Foundation Trust

Founder & Managing Director, Xperimint Ltd



Enhancing Discoverability of Public Health and Epidemiology Research Data: Summary

July 2014



Arofan Gregory
Open Data Foundation

Samuel Moore and Brian Hole
Ubiquity Press, London

Tito Castillo, Christiana McMahon and Spiros Denaxas
Farr Institute at UCL Partners, University College London

Veerle Van Den Eynden, Hervé L'Hours, Lucy Bell, Jack Kneeshaw
and Matthew Woollard
UK Data Archive, University of Essex, Colchester

Chifundo Kanjala, Gareth Knight and Basia Zaba
London School of Hygiene and Tropical Medicine

Public Health Research Data Forum

Joint Statement, 2011

Immediate goals

- Data management standards support data sharing
- Data sharing is recognized as a professional achievement
- Secondary data users respect the rights of producers and add value to the data they use

Longer-term aspirations

- Well documented data sets are available for secondary analysis
- Capacity to manage and analyse data is strengthened
- Published work and data are linked and archived
- Data sharing is sustainably resourced for the long term

My own journey

Chelsea Harbour

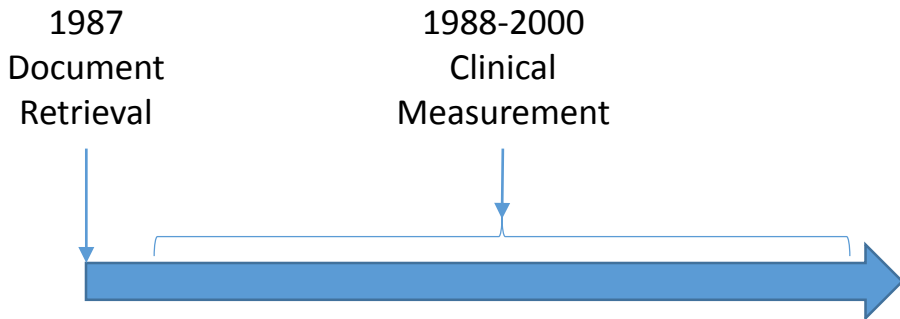
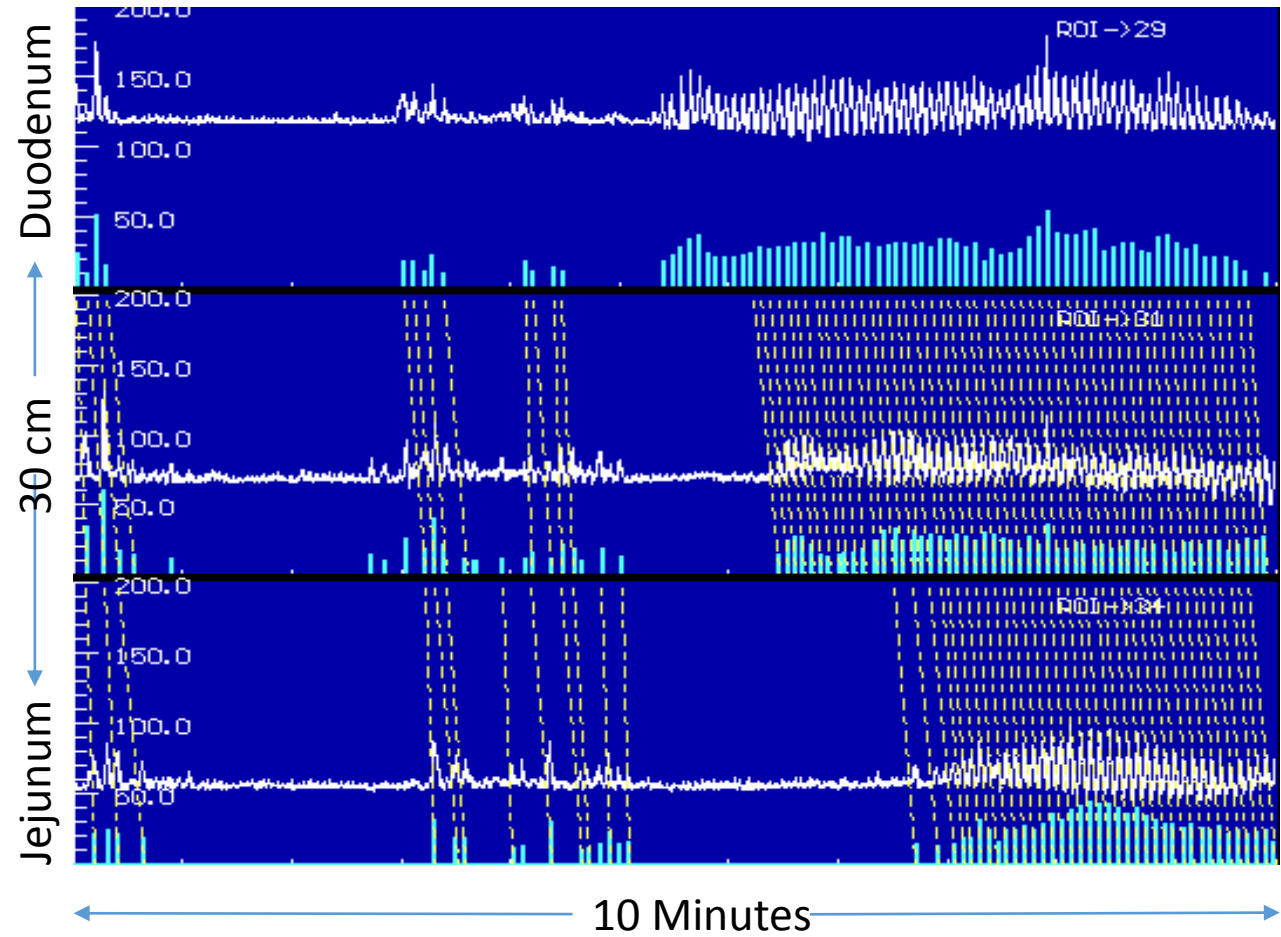


1987
Document
Retrieval



The Wingate Institute

Pressure waves recorded at three sites along the human small intestine, spaced 15 cm apart. The burst of contractions moving down the bowel is known as Phase III of the migrating motor complex.



National Programme for IT, UK NHS

THE  *DO NOT REMOVE*

Max 12C, min -6C Thursday December 8 2011 | thetimes.co.uk | No 70437

Exclusive NHS computer fiasco

● Patient database was axed in autumn but runs up £2bn bill ● Risk-taki

Dominic Kennedy, Chris Smyth
Laura Pittal

Taxpayers will foot the bill for a further £2 billion on a failed NHS IT project even though the Government has already pulled the plug on it.

The American technology company Computer Sciences Corporation (CSC) has boasted to Wall Street that it expects an extension of its contract to provide electronic patient records despite failing to deliver a fully functional version of its software.

Andrew Lansley, the Health Secretary, announced in September that he was ditching the scheme to create a national patient database



Times investigation
How Blair's dream became a nightmare
News, pages 6-8

because it had "let down" the health service.

An investigation by *The Times* into Britain's biggest IT procurement fiasco has unearthed a tale of risk-taking, mismanagement and broken promises for which taxpayers will pay the price.

In an indictment of Whitehall procedure that has saddled the NHS with an unfinished product:

- Health trusts were threatened with cuts unless they agreed to implement the system;
- Civil servants privately estimated that the software had a one-in-three chance of being delivered late;
- £250,000 was paid in bonuses by the Department of Health to 80 people as

Clinical Governance

UK National Health Service organisations became **accountable** for continually improving the quality of their services and safeguarding high **standards** of care by creating an environment in which excellence in **clinical care** will flourish

Genomic Revolution

Human Genome Project

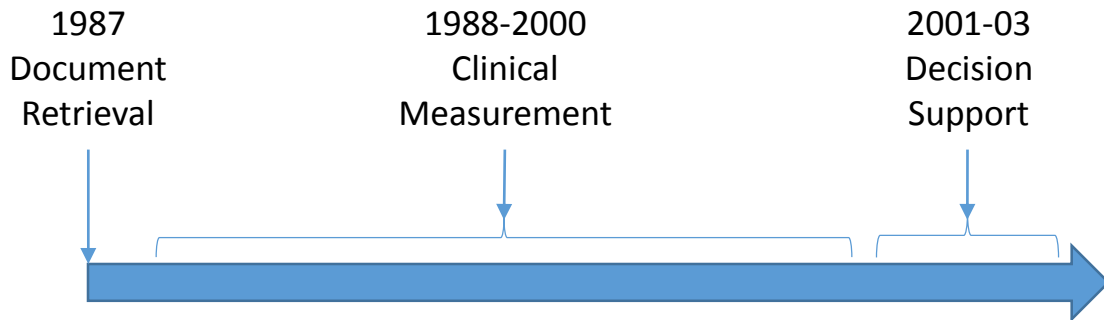
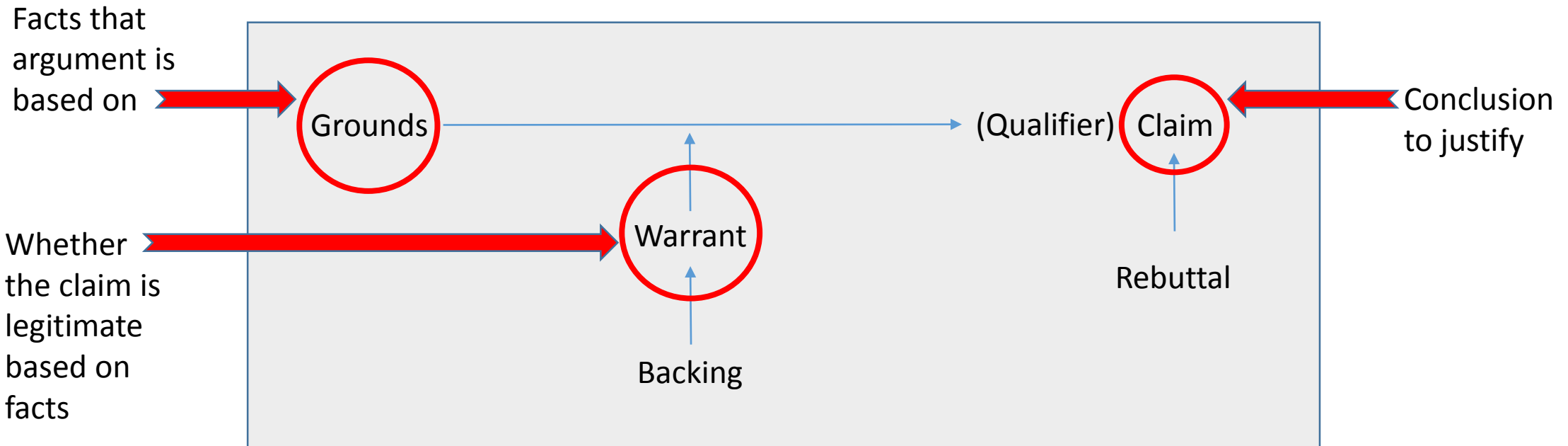
International, collaborative project to map the whole human genome.

Founded in 1990.

2001 draft human genome published in *Nature*.

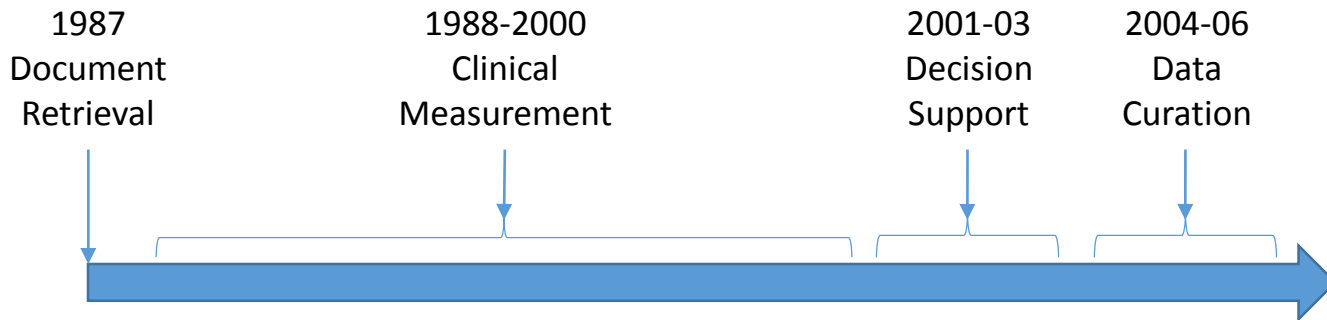


Decision support & Argumentation



Microarray Gene Expression

*... the complexity of the MAGE-ML format has made its use **impractical for laboratories** lacking dedicated bioinformatics support. We propose a simple tab-delimited, spreadsheet-based format, MAGE-TAB, which will become a part of the MAGE microarray data standard and can be used for annotating and communicating microarray data*



Data Re-use

Neurogastroenterol Motil (2006) **18**, 927–935

doi: 10.1111/j.1365-2982.2006.00824.x

The nocturnal jejunal migrating motor complex: defining normal ranges by study of 51 healthy adult volunteers and meta-analysis

S. M. SCOTT,* C. H. KNOWLES,* D. WANG, E. YAZAKI,† L. PICON,§ D. L. WINGATE‡ & G. LINDBERG¶

*Centre for Academic Surgery (GI Physiology Unit), Queen Mary's School of Medicine and Dentistry, London, UK
Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London, UK

‡The Wingate Institute, Barts & The London, Queen Mary's School of Medicine and Dentistry, London, UK

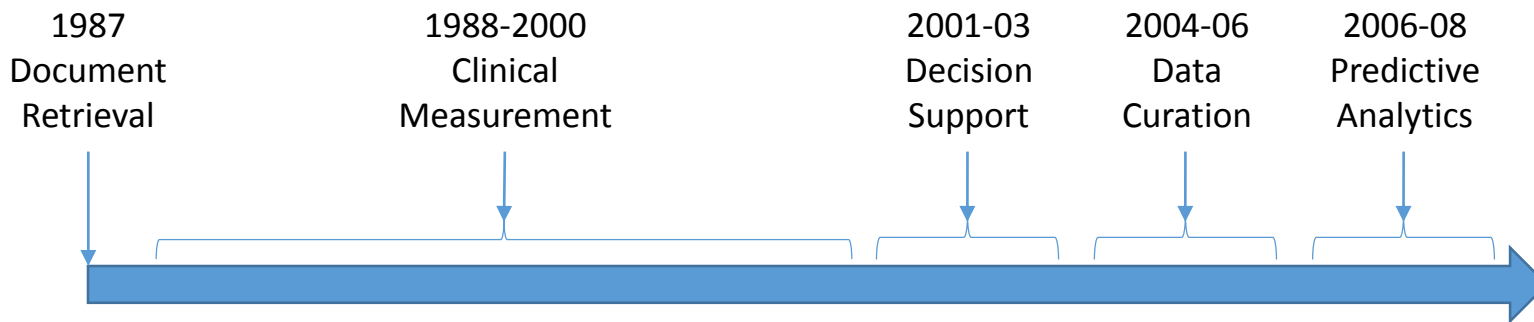
§Department of Public Health, Tours Medical University, Tours Cedex, France

¶Department of Medicine, Karolinska Institutet, Huddinge University Hospital, Stockholm, Sweden

Predictive Analytics (Health Dialog UK)

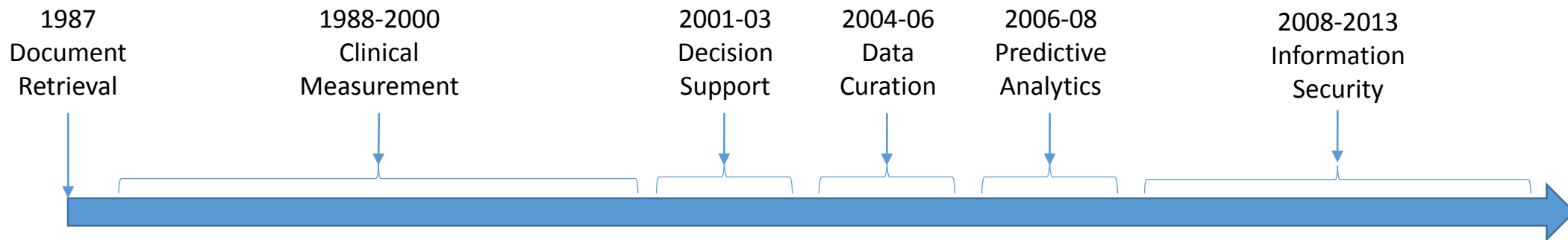
- Commissioned by the UK Department of Health
- Kings Fund, Health Dialog and New York University
- Patients at Risk of Rehospitalisation (PARR++)
- Predictive model based on UK inpatient data
- Freely distributed to all NHS organisations

PARR++

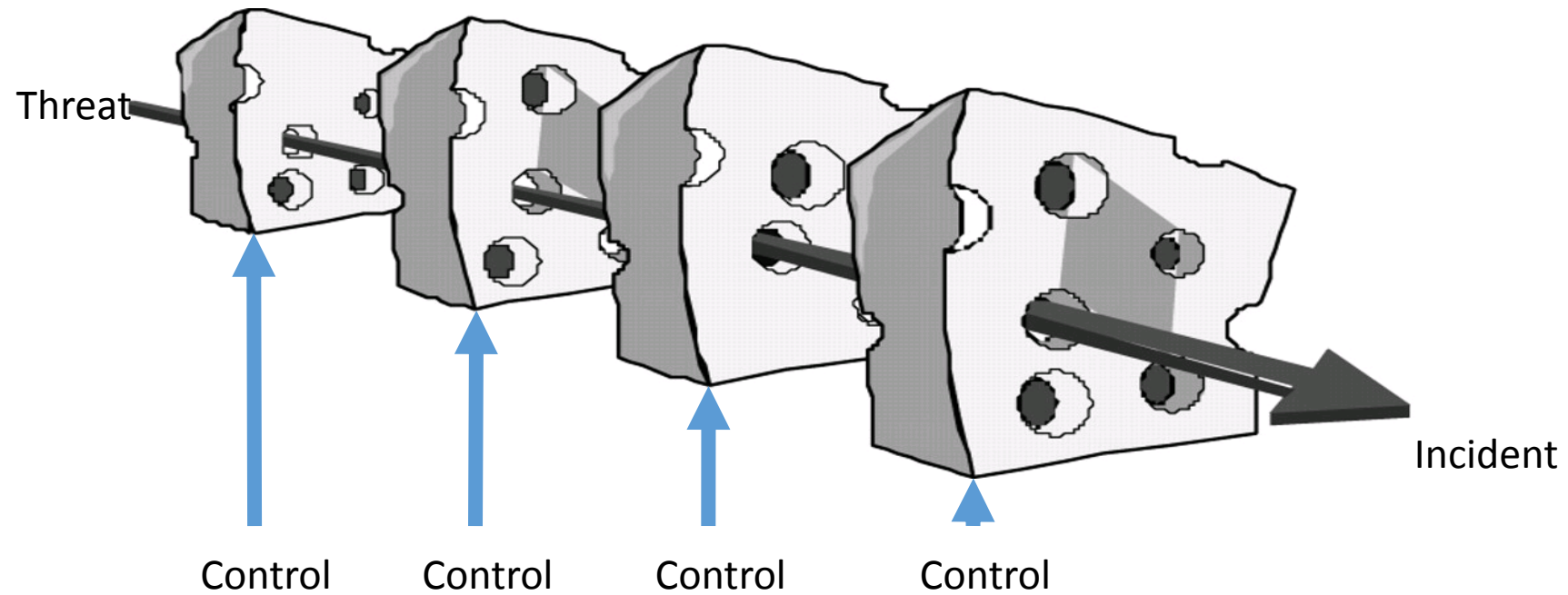


Information Security (UCL Inst of Child Health)

- Confidentiality
- Integrity
- Availability



Reason's model

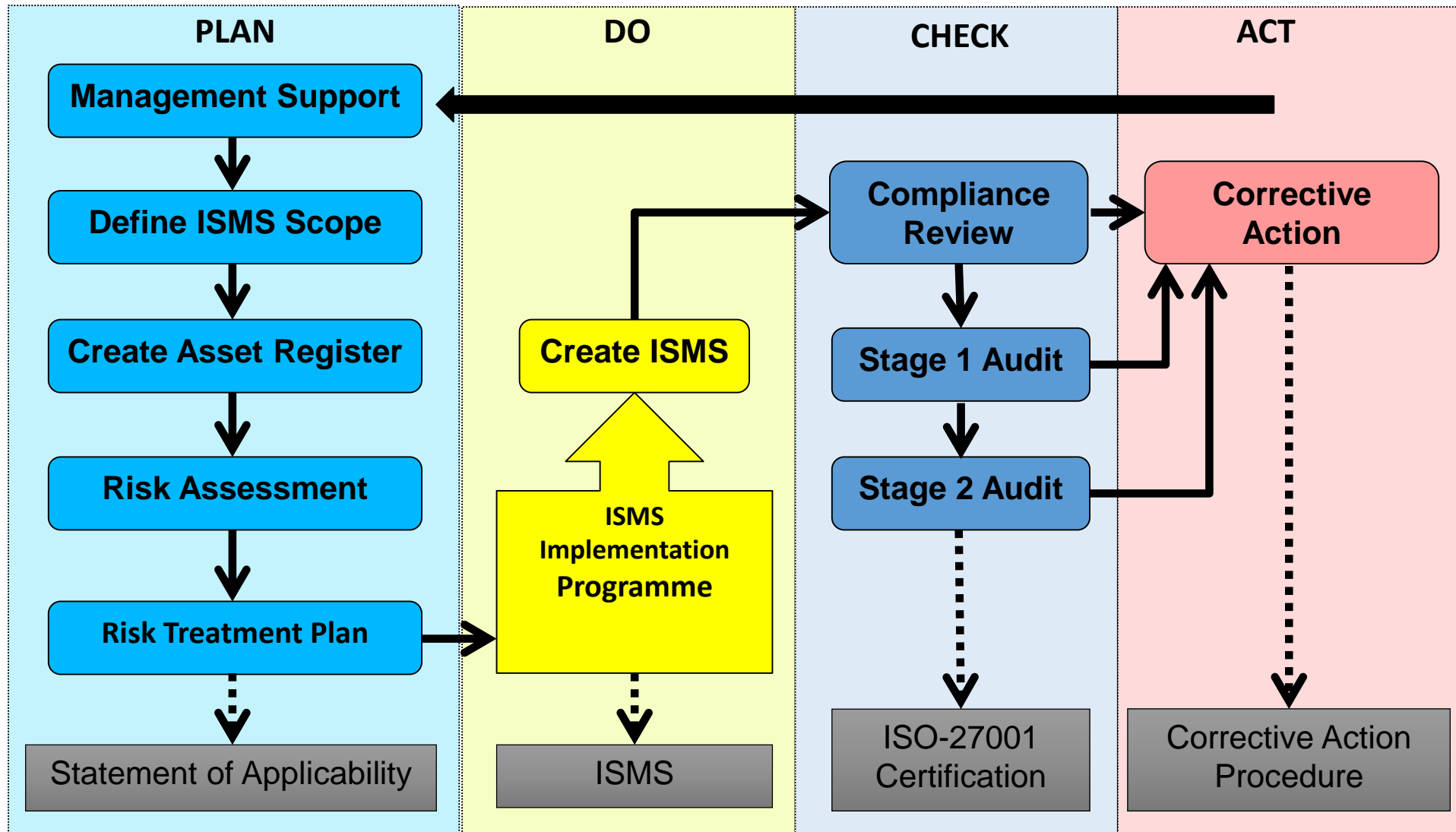


Information Security Management Systems

- International standard for information security
 - ISO-27001:2005
 - Describes requirements (i.e. what you 'shall' do)
 - Independently audited
 - Associated code of practice
 - ISO-27002:2005
 - Provides guidance (i.e. what you 'should' do)
- An ISMS is dynamic



Information Security Management System (ISMS) Development



Data Management Planning

Typical content

- Description of the data to be collected / created
- Standards / methodologies for data collection and management
- Ethics and Intellectual Property concerns or restrictions
- Plans for data sharing and access
- Strategy for long-term preservation

DMP Tool

http://www.openmetadata.org/site/?page_id=373



HOME PRODUCTS SERVICES DEVELOPER BLOG ABOUT

Data Management Plan (DMP) Editor



- Downloadable application
- User documentation
- Open source code repository

Benefits of certification



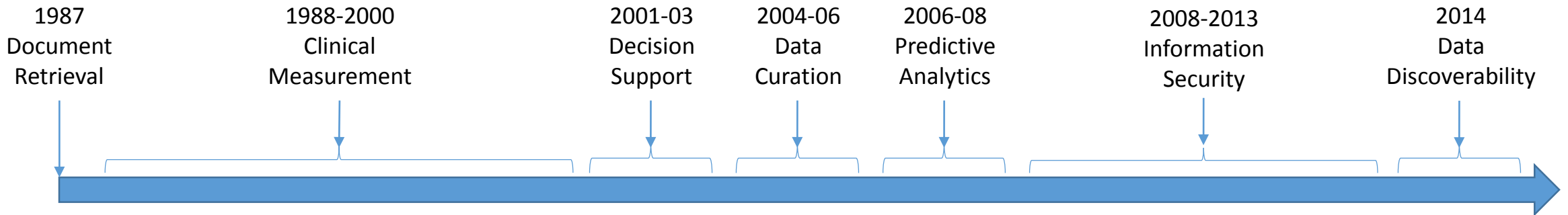
- 2012 formal certification of epiLab-SS service
- 2013 accreditation with NHS standards
- UCL among a group of universities funding to establish
 - Farr Institute of Health Informatics Research 2013
 - Administrative Data Research Centre 2014
 - Total of more than **£74 million**
- 2014 UCL receives ISO 27001 and NHS for large-scales secure service

Data Discoverability Report (Wellcome Trust)

Three Complementary Models

1. Centralised Portal Model
2. Data Journal Model
3. Linked Data Model

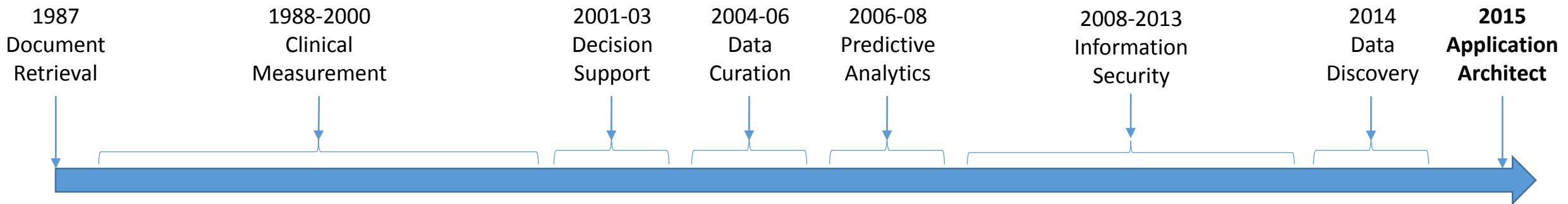
Standardised metadata and documentation is essential



Application Architect, UCLH

*I may not have gone
where I intended to go,
but I think I have ended up
where I needed to be.*

Douglas Adams, The Long Dark Tea-Time of the Soul





Service information

Date MONDAY 16TH MARCH 2015

Time QUOTE OF THE DAY

“HAPPINESS IS A
JOURNEY AND NOT
A DESTINATION.”

— BEN SWEETLAND