

Data Discovery and Sharing at UCLH

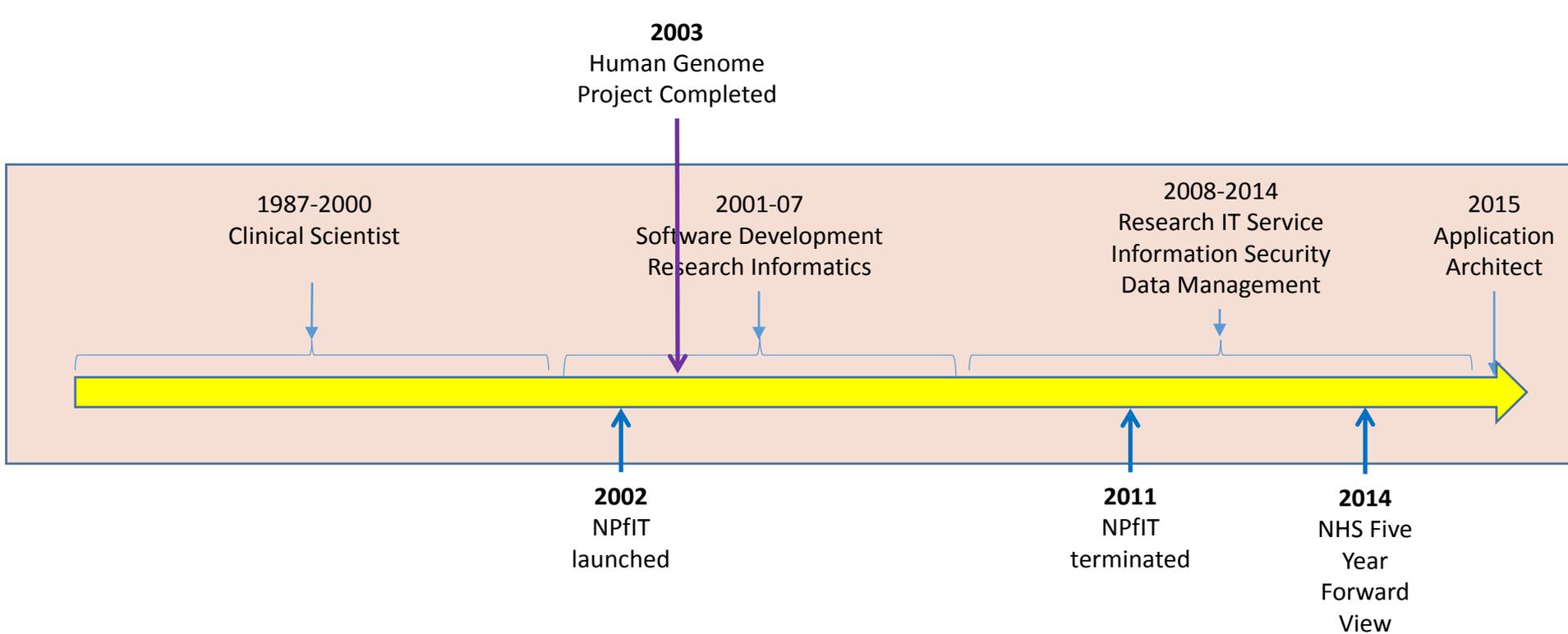
Dr Tito Castillo

Application Architect, UCLH

Senior Research Associate, Farr Institute

Associate Vice-Chair for Standards, BCS Health

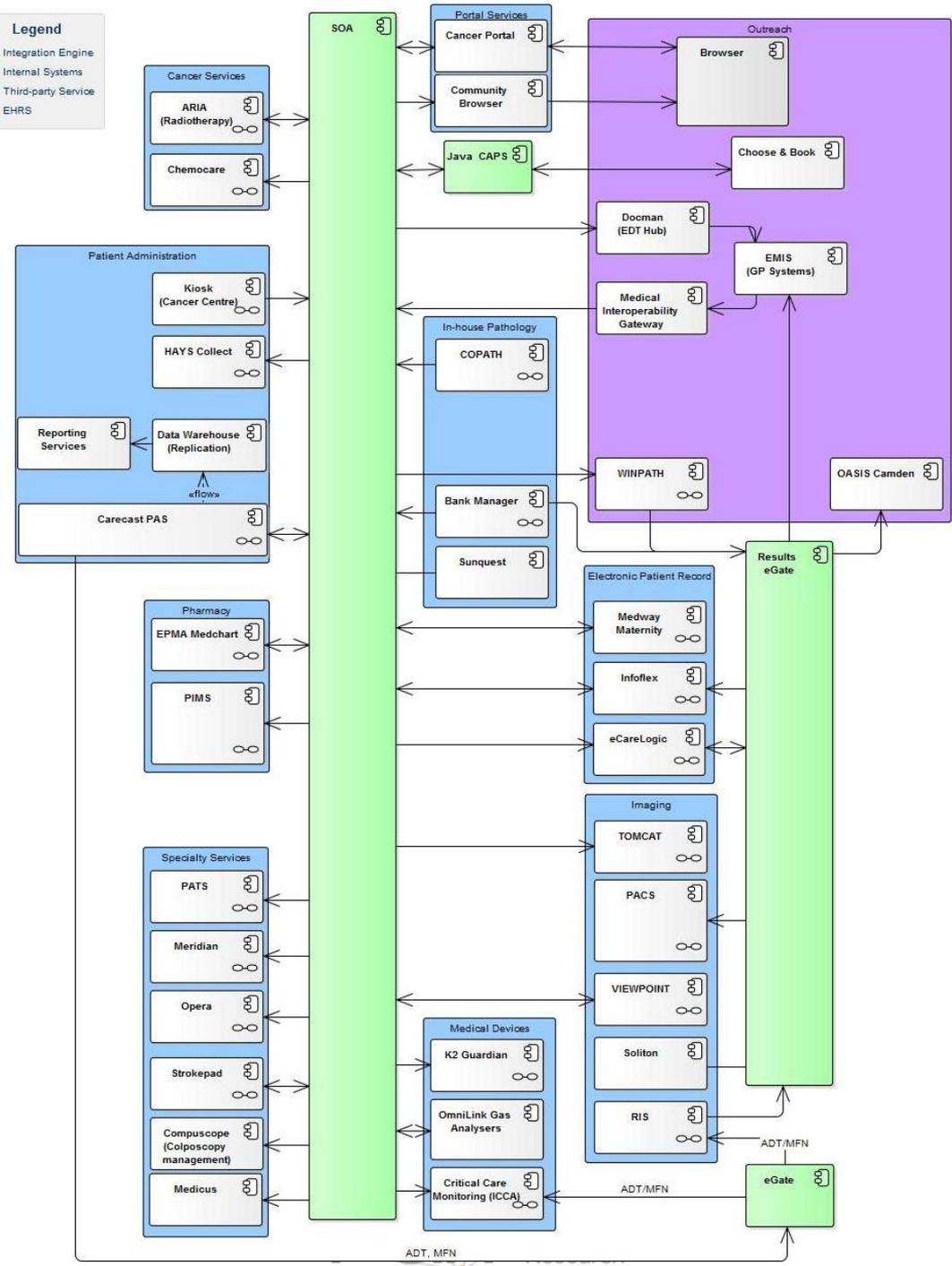
Health IT – personal journey



UCLH Context

UCLH Landscape

- PAS – GE Carecast (Tandem/NonStop)
 - Formerly IDX Corp
 - **Issues**
 - Difficult (expensive) to change
 - Difficult (expensive) to write-back
- eCareLogic (CGI)
 - Formerly Clinical Document Repository (CDR)
 - **Issues**
 - Originally developed within UCLH
 - Open API support
 - Black box solution
- Data warehouses
 - Multiple SQL Server instances
 - **Issues**
 - Historical evolution
 - Minimal documentation
 - Difficult to maintain



Strategic Objectives 2016/17

1. Provide the highest quality of care within our resources
2. Improve patient pathways through collaboration with partners
3. Support staff to deliver their potential
4. Achieve financial sustainability
5. Generate world-class clinical research

Challenges

- Multiple Data Access & Sharing Programmes
 - Health Informatics Collaborative
 - Genomic Medical Centre (Genomics England)
 - NIHR BioResource
 - Integrated Digital Care Records (Camden & Islington)
 - Digital Roadmap (Sustainability & Transformation Plans)
 - Exemplar Ward
 - Data Warehouse and BI Improvement

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

Data Governance

Data Governance

- Data governance encompass
 - People
 - Processes
 - Information technology
- Consistent handling of enterprise-wide data
- Healthcare data quality is extremely variable and often poorly described

Metadata

- Data is dynamic, passing through different stages of a life-cycle
- At each stage in this cycle it is possible add descriptive data (metadata)
- Sufficient metadata provide the necessary additional information for:
 - Discovery
 - re-use
 - Repurposing
 - comparison
- Metadata should be added as early as possible

ISO 11179 - Metadata Registry

- International standard for representing metadata for an organization in a metadata registry
- Provides consistent definitions of data across time, between databases, between organizations and between processes.
- Data elements are described by their semantics and their representations.
 - registered
 - uniquely identified within the register
 - named
 - defined
 - may be classified

Data Documentation Initiative (DDI)

- Originally developed in Social Science to describe microdata collections
- Originally used for *post hoc* documentation
- Two main versions in use
 - v2.x (Codebook)
 - v3.x (Life-cycle)
- Widely adopted by biomedical cohort studies in public health and epidemiology
- Large user community and extensive tooling

DDI Data Life-cycle



Image from the Data Documentation Initiative Alliance

Statistical Data and Metadata Exchange (SDMX)

- Characterisation, exchange and reporting of statistical datasets
- Used routinely by most international statistical agencies
- Mature standard with rich toolsets available
- Model-driven approach to describe data and associated metadata
 - Information to be shared between processes or organisations
 - Relationships between the information objects
- Open source implementation www.sdmxsource.org

Bank for International Settlements (BIS)
European Central Bank (ECB)
Eurostat (the statistical office of the European Union)
International Monetary Fund (IMF)
Organisation for Economic Co-operation and Development (OECD)
United Nations Statistics Division (UNSD)
World Bank

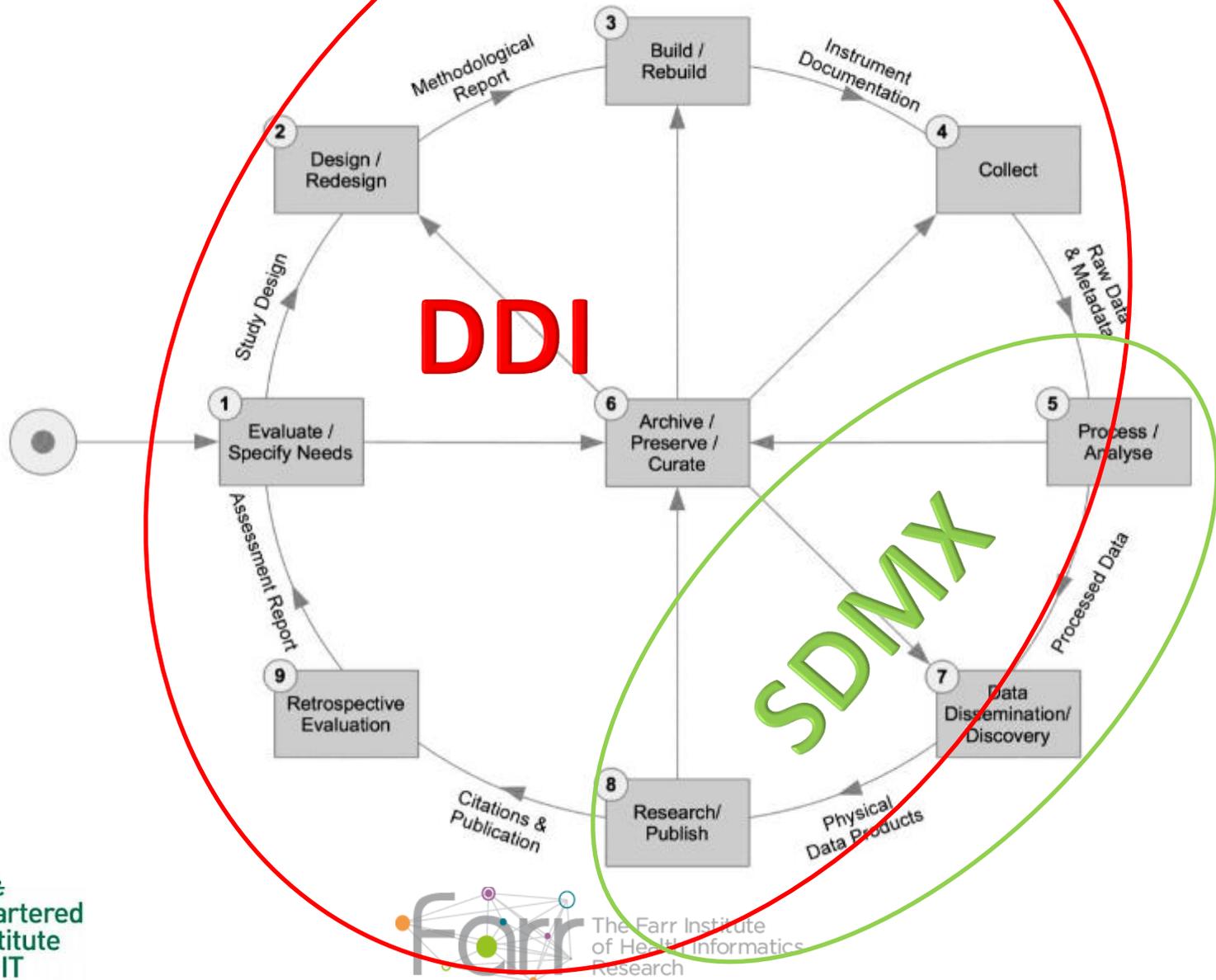
UCLH Approach

Data management

- Identify a suitable reference model against which data management processes can be mapped
- Generic Longitudinal Business Process Model (GLBPM)
 - Longitudinal and repeat cross-sectional data collection
 - Grounded in human science (research on people)
 - nine high level steps
 - three cross-cutting processes

Generic Longitudinal Business Process Model

I Barkow, W Block, J Greenfield, A Gregory, M Hebing, L Hoyle, W Zenk-Möltgen. DDI Working Paper Series (2013)



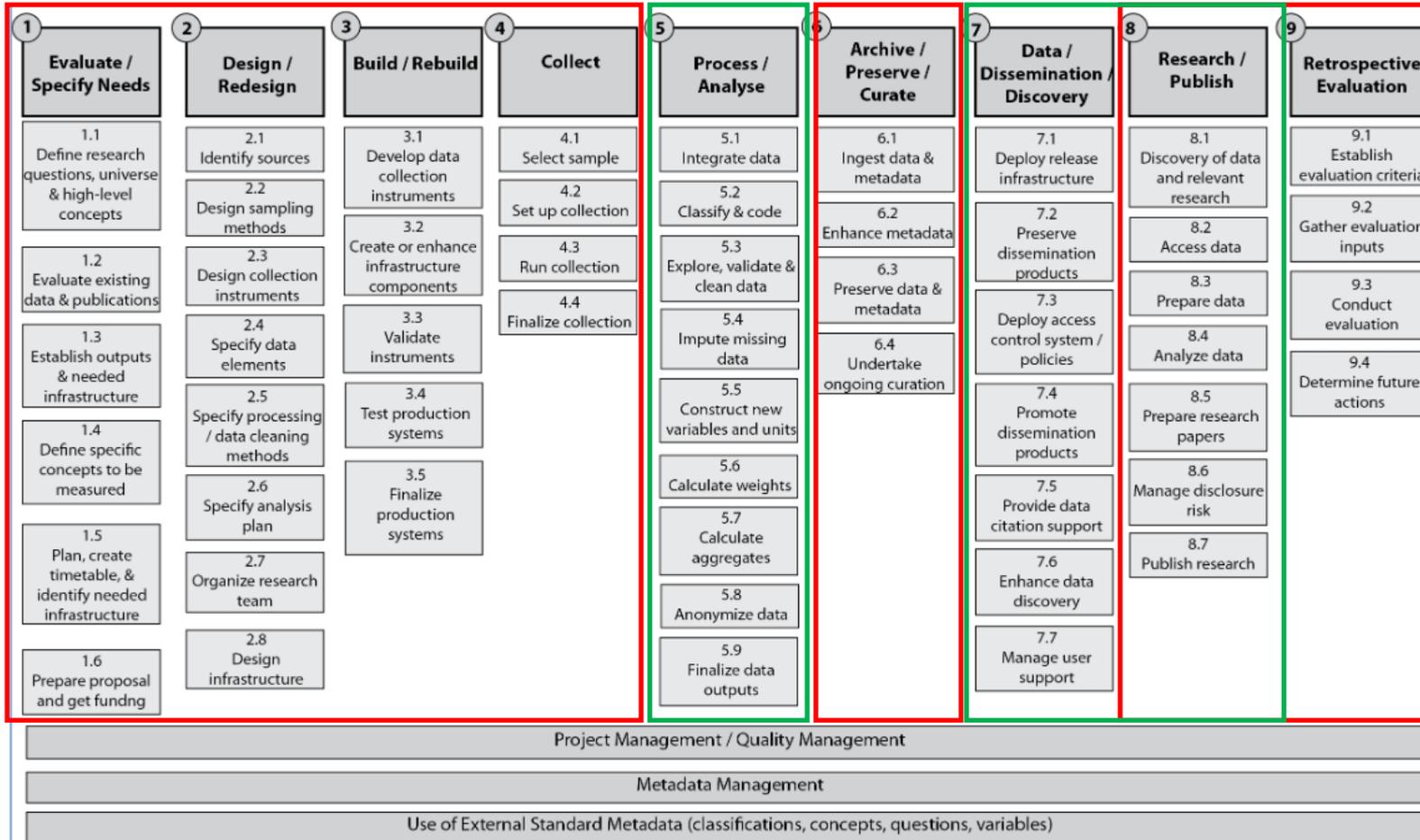
Generic Longitudinal Business Process Model

I Barkow, W Block, J Greenfield, A Gregory, M Hebing, L Hoyle, W Zenk-Möltgen. DDI Working Paper Series (2013)

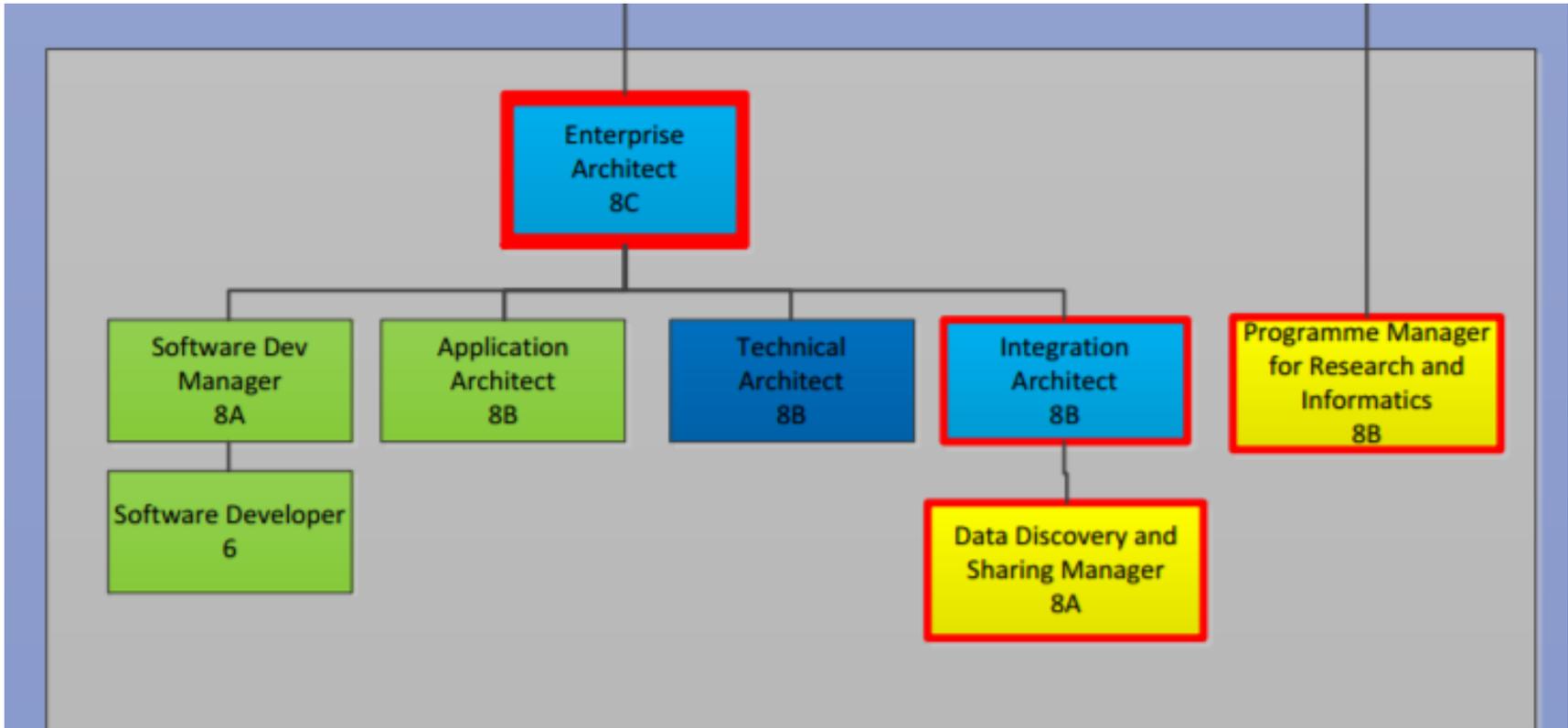
DDI

SDMX DDI

SDMX DDI



Business Solutions



Summary

- Despite the use of the NHS IG Toolkit data governance in the NHS is often poorly supported
- Multiple proprietary systems exist with undocumented/poorly documented data models
- New models of care combined with a drive to support biomedical research highlight the need to share, re-use and repurpose data
- Gap in existing healthcare standards

Personal perspective

NHS & Universities

Comparison

- Change vs Transformation
- Resilience & Business Continuity
- Information Governance vs Data Governance
- Relationship with industry

Barriers

- Pensions
- State registration
- Culture