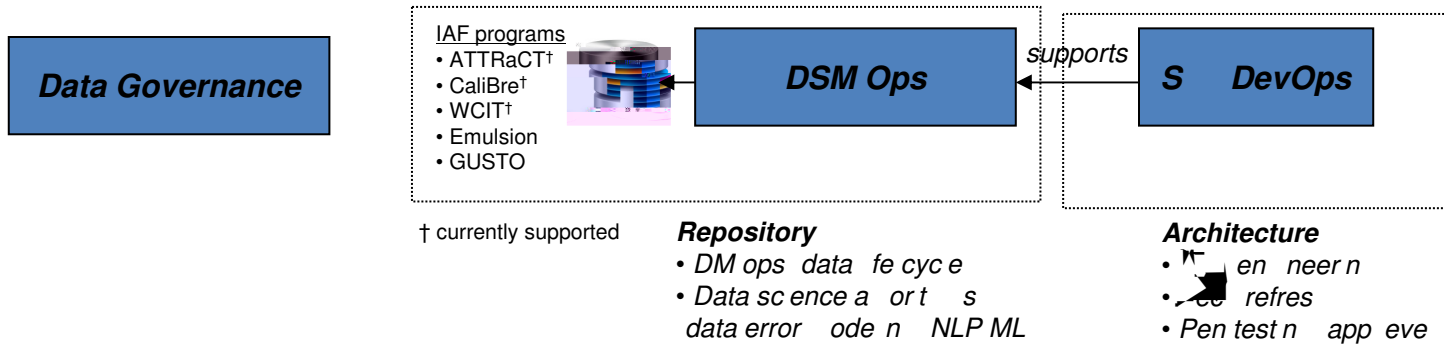


The
(Programme
for
BMRC clinical research
data management
Wong Wing-Cheong

BioMed DAR: Mission and Structure

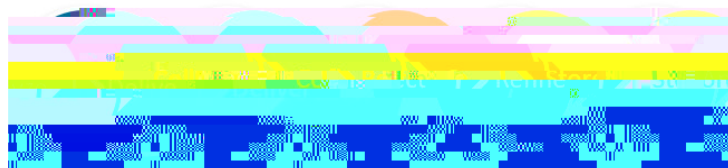
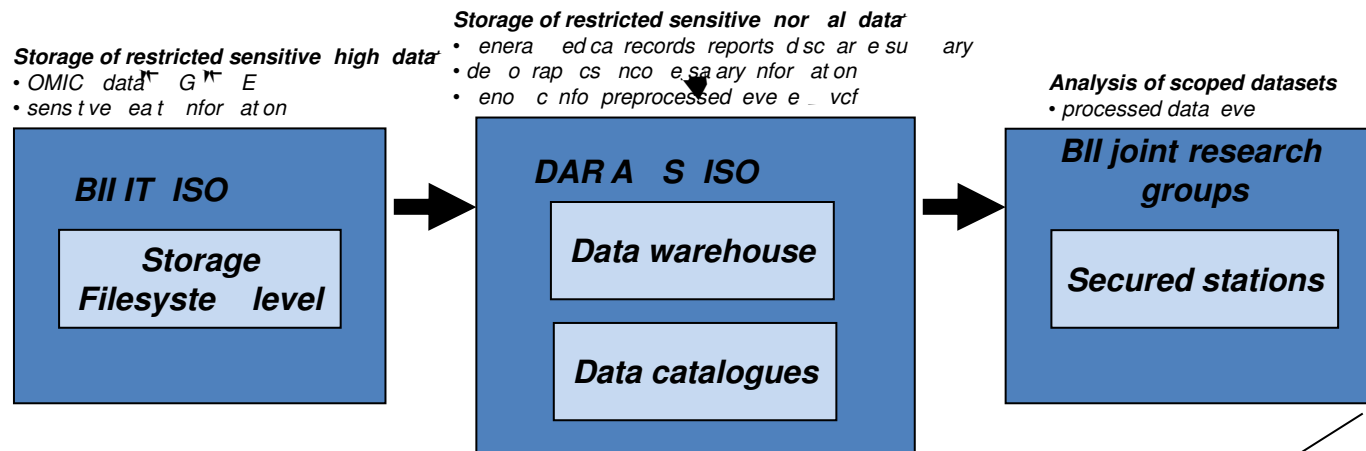
tru tur



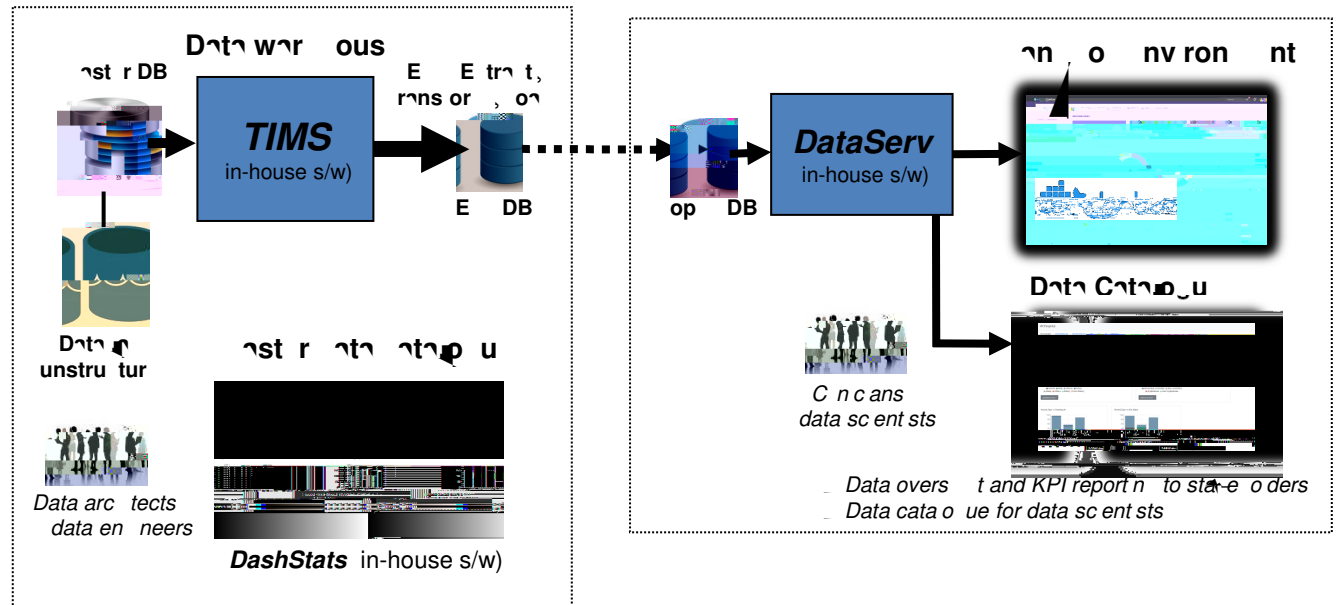
DAC : Data Access Committee
 DSM : Data Science Management
 ML : machine learning
 NLP : Natural Language Processing
 S/W DevOps: Software Development Operations

BioMed DAR: B integration

present report on BioMed DAR FY 2011



BioMed DAR: Technical Stack!



- **Data life cycle activities** maintenance, synthesis, catalog

BioMed DAR: Datasets " Data life-cycle

Datasets and their characteristics

Dataset	Content	Int. S.S.	Catalogue	Access	Archival	Purification
ATTRaCT	☒	☒	☒	☒		
CaliBre	☒	☒	☒			
WCIT	☒	☒	☒			

In addition: GUSTO, Emulsion

About Data Lifecycle

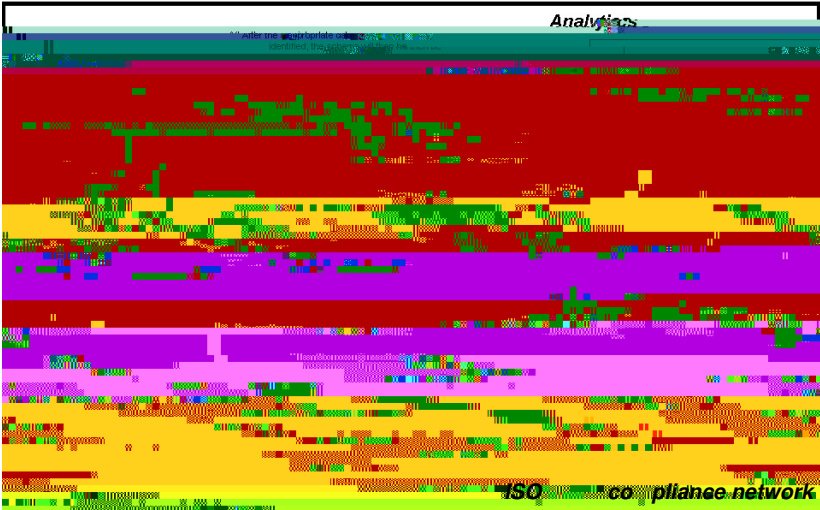
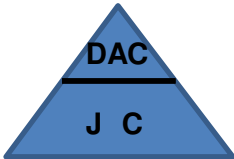
The sequence of stages that a particular unit of data goes through from its initial generation or capture to its eventual archival and/or deletion at the end of its useful life.

Core processes to data lifecycle ops

1. Maintenance (cleansing, integration, movement, governance)
2. Synthesis (data creation from existing data e.g. FASTQ to vcf)
3. Cataloguing (discovery, describe, organize)
4. Usage (data access enabling and sandboxing)

Core systems support to data lifecycle ops

1. s/w engineering (data warehouse, catalogue and request/approval systems)



1. Introduction

10% A & ' () \$