Medical data in a Bioinformatics Core facility

Reinhard Schneider
Head of Bioinformatics Core facility

UNIVERSITÉ DU LUXEMBOURG

LCSB

DATA PROTECTION ACT 1998

These principles state in relation to healthcare that data should be:

- fairly and lawfully processed;
- processed for limited purposes, which include preventive medicine, medical diagnosis, medical research, provision of care and treatment and the management of healthcare services;
- adequate, relevant and not excessive, especially when obtaining, recording, holding, altering, retrieving, destroying or disclosing of data
- accurate;
- not kept longer than necessary;
- processed in accordance with the data subject's rights thus individuals are entitled to prevent processing
    - for direct marketing purposes
    - which will or likely to cause the data subject or another person unwarranted and substantial harm or distress
- secure thus any data subject who suffers damage due to unauthorised disclosure is entitled to compensation;
- **not transferred to countries without adequate protection.**

Data Confidentiality

The need for data confidentiality in medical applications has far reaching consequences, which are hard to grasp from the outset; for example, if the aim is to protect the information of a given individual at the highest possible level.

It implies that participating IT infrastructure must undergo **full inventory management**, including tracing equipment lifecycle even for broken components. An old backup, or a broken disk, cannot be just disposed off; personnel passwords have to be carefully protected; third-party systems and maintenance personnel have to be vetted for compliance with desired confidentiality level.

The physical environment of systems and their configuration have to be carefully controlled.

Techniques like anonymization or pseudo-anonymization (with full anomymization nobody can track back to the original data, pseudo-anonymization allows typically that the data-owner can trace back to the individual via an ID-number) might help in certain cases but, a DNA chain will always remain a unique identifier so, it is important to acknowledge early what the constraints are - and how to handle them.

Innovative Medicines Initiative (IMI) is a unique public/private partnership between the EU and the European Federation of Pharmaceutical Industries and Associations (EFPIA)

Key Required Features of eTRIKS KM Platform

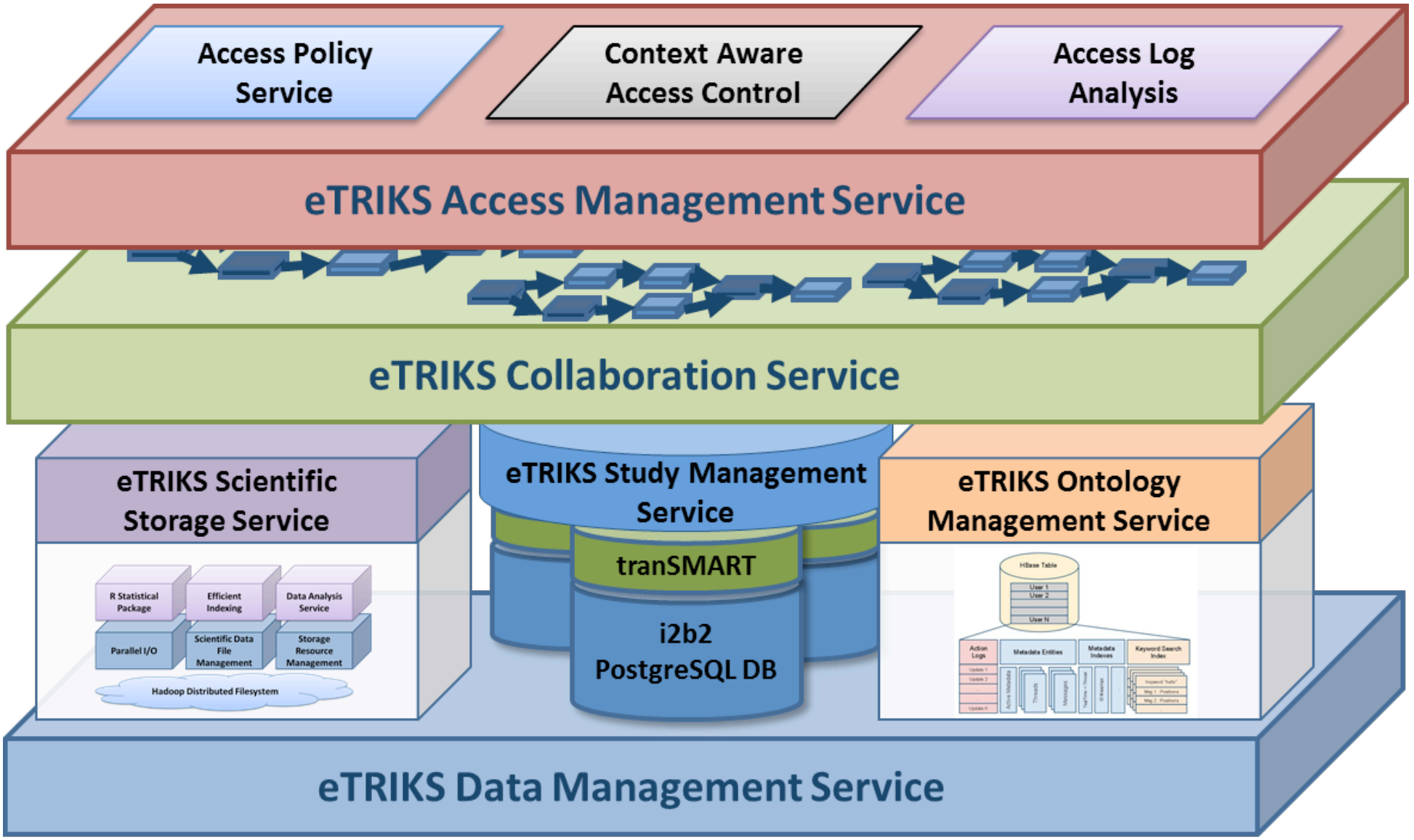
must go far beyond a classical data warehouse, to include the following key features:

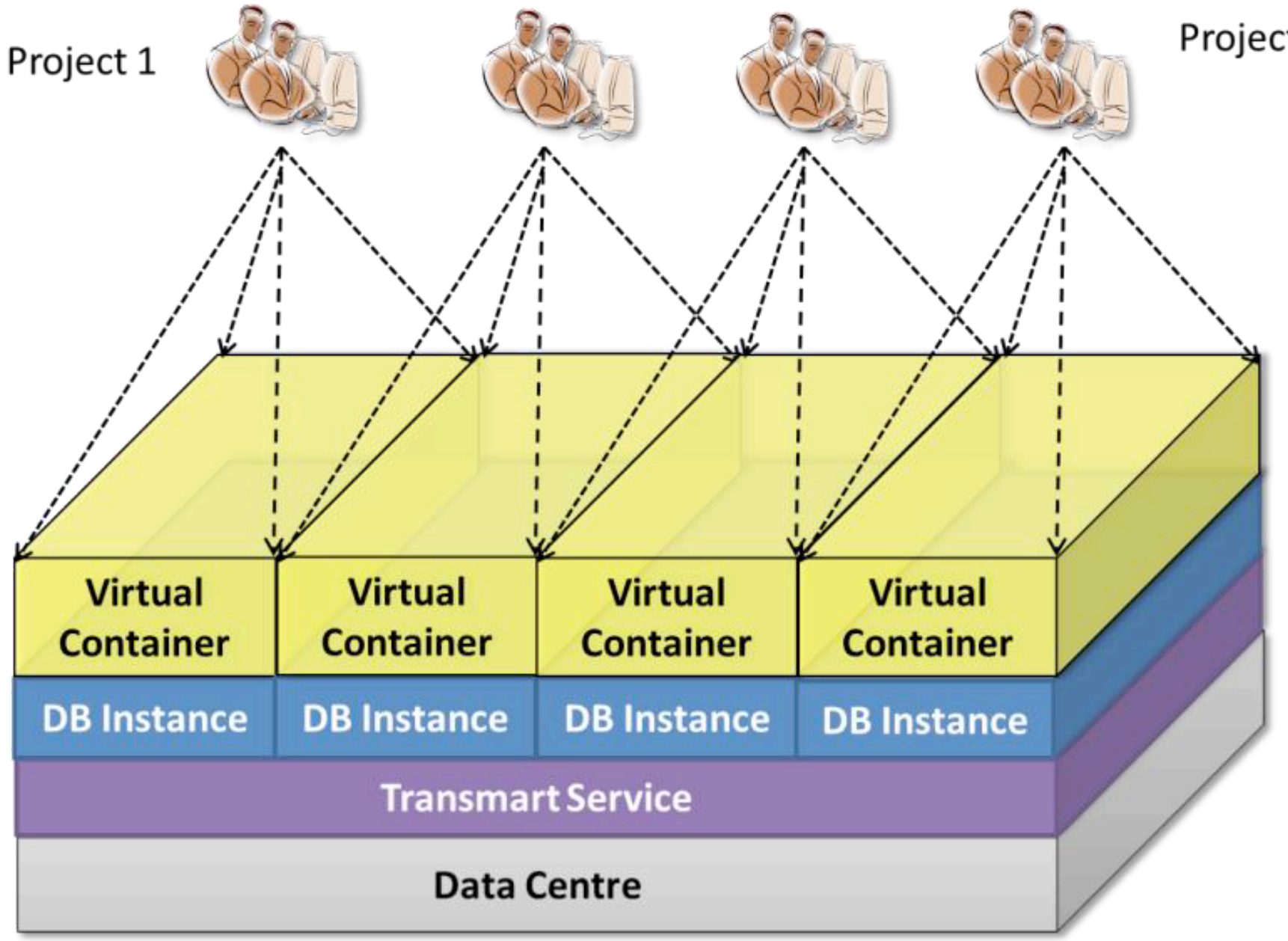
1) **A cloud-based biomarker discovery infrastructure**, which will be equipped with a rich set of analytical methods and tools for omics, imaging data and text, implemented and optimized on a high-performance computing platform, to support complex analytical tasks for biomarker discovery. Workflow technology will be used to support building of analytical applications.

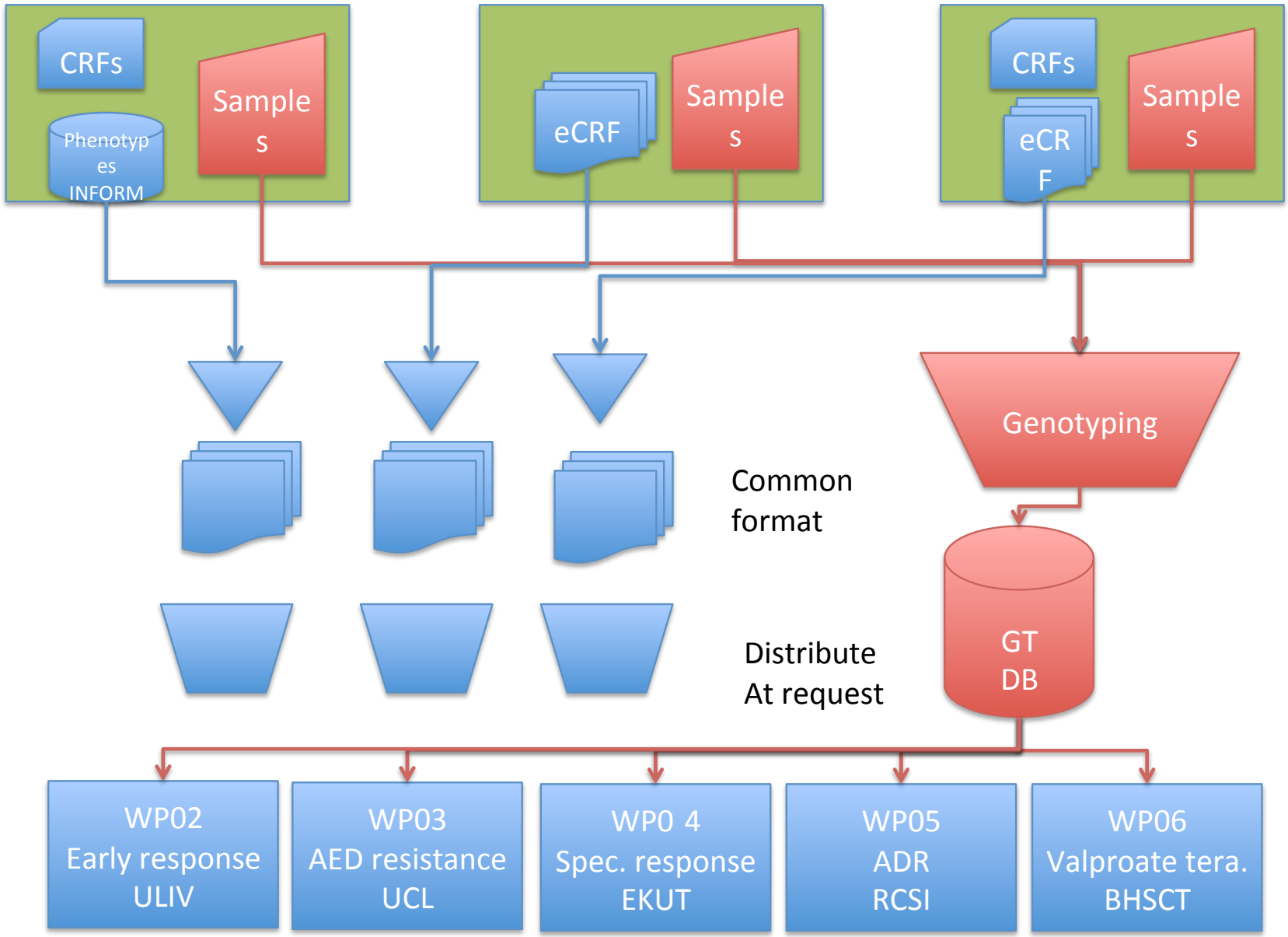
2) **A collaborative research management system** to support large-scale collaborations in IMI projects. The initial development will be based on the Analytical Register system developed by Imperial for the U-BIOPRED consortium based on Wiki technology.

3) **An integrative knowledge production and management environment** that will enhance the current tranSMART system with improved knowledge production and management mechanisms such as curation automation and quality control support, terminology management, ontology-based knowledge annotation and assimilation mechanism, systems biology modelling and simulation support and semantic indexing and searching capabilities.

4) **A scientific knowledge service delivery platform** that goes beyond knowledge production and management for translational research to supporting the “knowledge-as-a-service” model, further developing and extending the “data-as-a-service” paradigm.


Access Policy Service
Context Aware Access Control
Access Log Analysis
eTRIKS Access Management Service
eTRIKS Collaboration Service
eTRIKS Scientific Storage Service
R Statistical Package
Efficient Indexing
Data Analysis Service
Parallel I/O
Scientific Data File Management
Storage Resource Management
Hadoop Distributed Filesystem
eTRIKS Study Management Service
tranSMART
i2b2
PostgreSQL DB
eTRIKS Ontology Management Service
HBase Table
User 1
User 2
User N
Action Logs
Metadata Entities
Metadata Indexes
Keyword Search Index
eTRIKS Data Management Service


Project 1
Project n
Virtual Container
Virtual Container
Virtual Container
Virtual Container
DB Instance
DB Instance
DB Instance
DB Instance
Transmart Service
Data Centre



UNIVERSITÉ DU LUXEMBOURG

LCSB

* Security for grid computing in the field of bioinformatics

* Information security development processes for sensitive (medical) data

* Securing medical data (Information, Data and System Integrity)

* Digital Rights Management for medical stored data

* Web and Wireless Security Bioinformatics and Medical Diagnosis

* Information Security Management