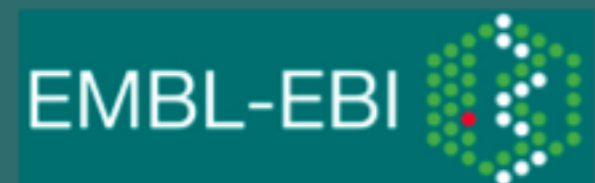


SIMBioMS - IT solution for data collection, harmonisation and analysis in large-scale biomedical genomics projects

11 June 2009

Maria Krestyaninova
Microarray Informatics group
European Bioinformatics Institute
EMBL-EBI (Hinxton, UK)
In collaboration with IMCS, (Riga,
Latvia) & FIMM (Helsinki, Finland)





MAIN MENU

- Home
- Software modules
- Downloads
- Documentation
- FAQs
- Publications
- Links
- Contacts

LOGIN FORM

Username

Password

Remember me

[Lost Password?](#)
 No account yet? [Register](#)

Home

SIMBiOMS

SIMBiOMS (System for Information Management in BioMedical Studies) is a multi-module solution for data management in biomedical studies.

Any research concerning human samples and/or utilizing high-throughput technologies yields such amount of information that conventional data storage solution might not be sufficient.

We offer here three software modules:

- Sample Information Management System (**SIMS**),
- Assay Information Management System (**AIMS**) and
- Sample Availability System (**SAIL**).

All three software modules are developed as a part of the integrated EU project **MoIPAGE** (Molecular Phenotyping to Accelerate Genomic Epidemiology) and the collaborative research project **ENGAGE** (European Network of Genomic and Genetic Epidemiology).

SIMS and AIMS can work either as united system or as two completely independent components. In turn, SAIL is independent web-based system for index of phenotypes availability in different cohorts and collections.

All systems are packaged in such a way that they can easily be installed either as local (e.g. on a laptop) or as centralized databases (to be used by a group of people).

SIMS and AIMS benefit from customizable interface, editable vocabularies and a choice of options for tackling data confidentiality issues. The systems provides a user with efficient means of control over data exchange process and at the same time helps to format the metadata in compliance with the standards accepted in functional

POWERED BY



FLASH TUTORIALS

- [SIMS tutorial](#)
- [AIMS tutorial](#)
- [SAIL tutorial](#)
- [GWA tutorial](#)

CONTACTS

E-mail:
support@simbioms.org

Started in 2005

as a collaboration between

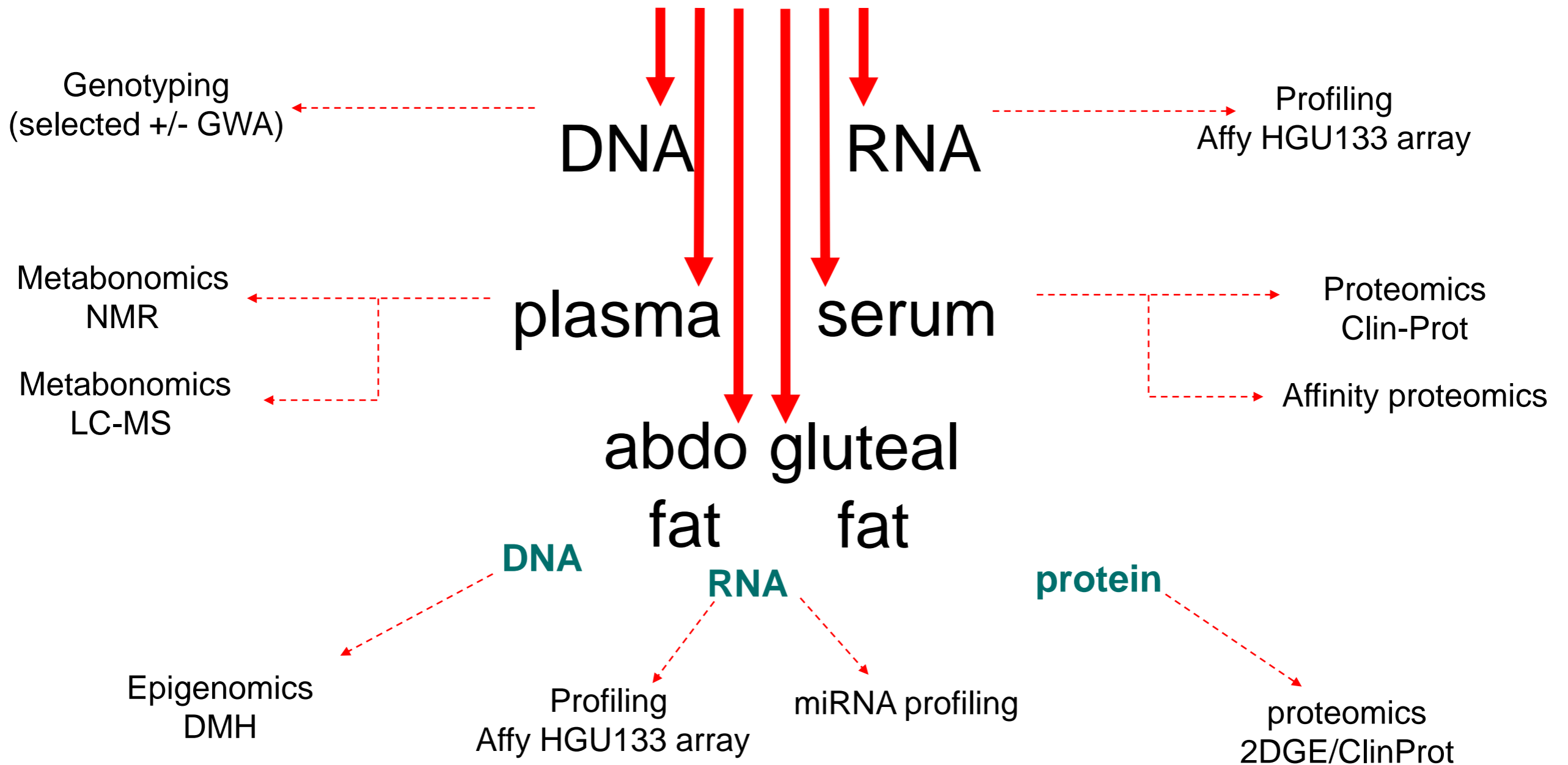
Institute of Mathematics and Computer Science (Riga
Latvia)

and European Bioinformatics Institute (Hinxton, UK)

since 2007:

The Institute for Molecular Medicine Finland (Helsinki)

subjects



~20 partners

multiple data sources

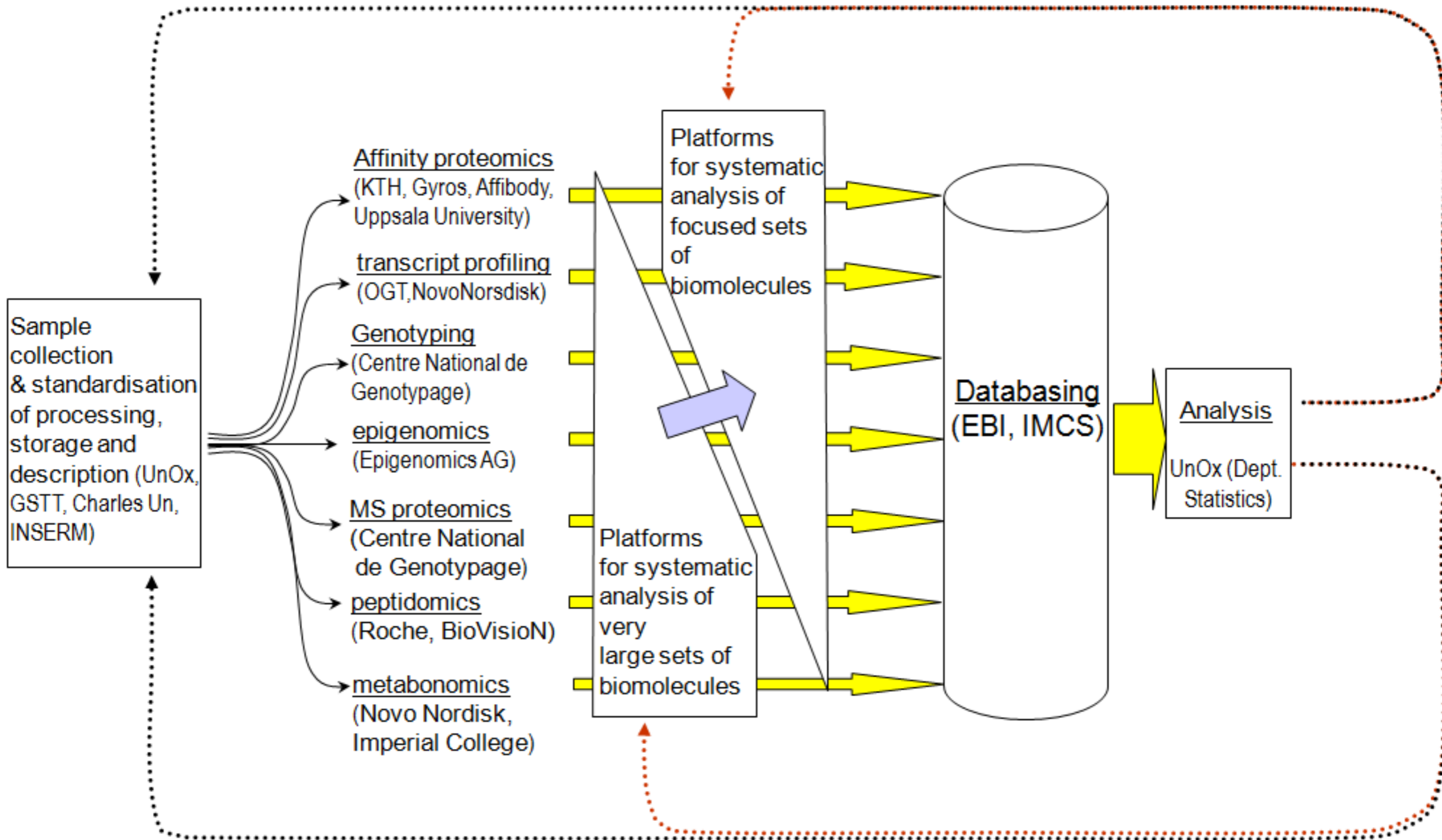
a variety of analysis strategies

10 000 – 100 000 samples

>10 molecular high throughput technologies

>100 000 assays

Informatics support within large integrated projects



ATLAS BETA

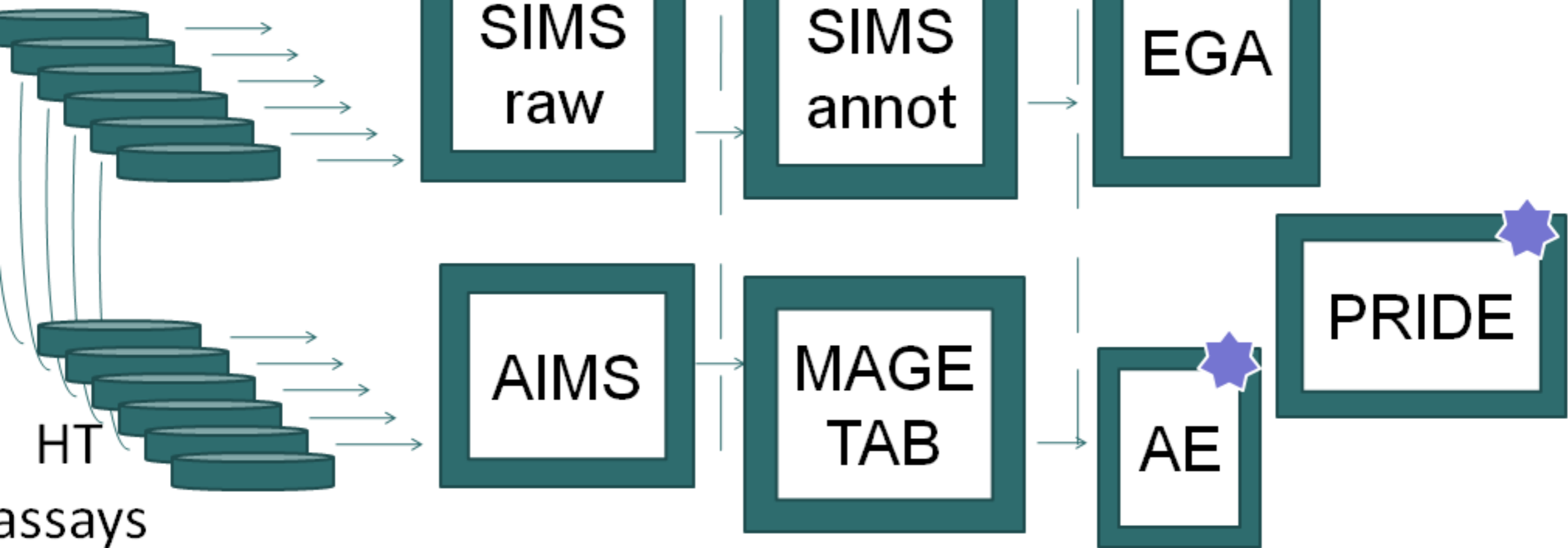
Genes: (all genes) Conditions: up/down in (aer_txt_expaccession:E- Organism: Any species

Search Atlas View results as: table heatmap expand conditions search with [EFO](#)

View as table

Factor Value	Number of studies	SORBS1		SPARCL1		YAP1		PTAFR		FFAR2		PTPN21		PLEKHC1		EMCN		NFIB		RASSF5		LILRA2		NPY1R		CAV1		PLIN		CLEC7A		PLSCR4	
		UP	DN	UP	DN	UP	DN	UP	DN	UP	DN	UP	DN	UP	DN	UP	DN	UP	DN	UP	DN	UP	DN	UP	DN	UP	DN	UP	DN	UP	DN	UP	DN
		Abdominal fat (organismpart)	1	1		1		1						1		1				1								1		1			
Blood (organismpart)	1							1		1										1		1								1			
Gluteal fat (organismpart)	1	1		1		1						1		1		1									1		1		1				
Female (sex)	1																																
Male (sex)	1																																
OX0904 (individual)	1																																
OX0694 (individual)	1																																
56 years (age)	1																																
52 years (age)	1																																

Biobanks*



PUBLIC

Disease -> genetic variant/molecular signature:

- relevant mutations, molecules and pathways*
- biobank/public biomolecule/trial data integration*
- samples available for a disease-study*



P3G, SNOMED CT, ICD, TNM, ATC

EMBL-EBI



Informatomics challenges in genomic epidemiology projects

- Type of IT support varies greatly among partners (e.g LIMS, shared directory structure, folders in PC)
- Technologies develop faster than data exchange formats
- Data access control
 - Access decisions remain with the data generating body
- Differentiated permissions:
 - sample data
 - omics raw data
 - processed data
 - Results

CHALLENGE 1: multi-partner interaction

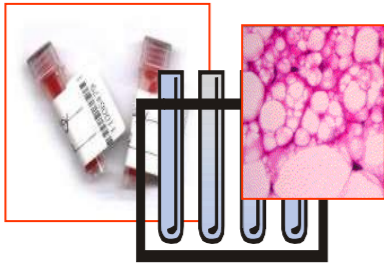
Architecture

SIMS



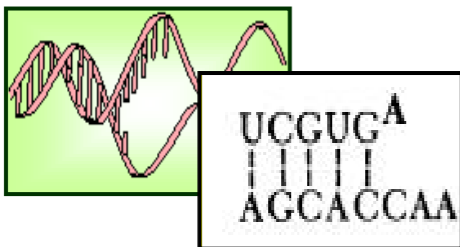
Sample Information Management System

AIMS

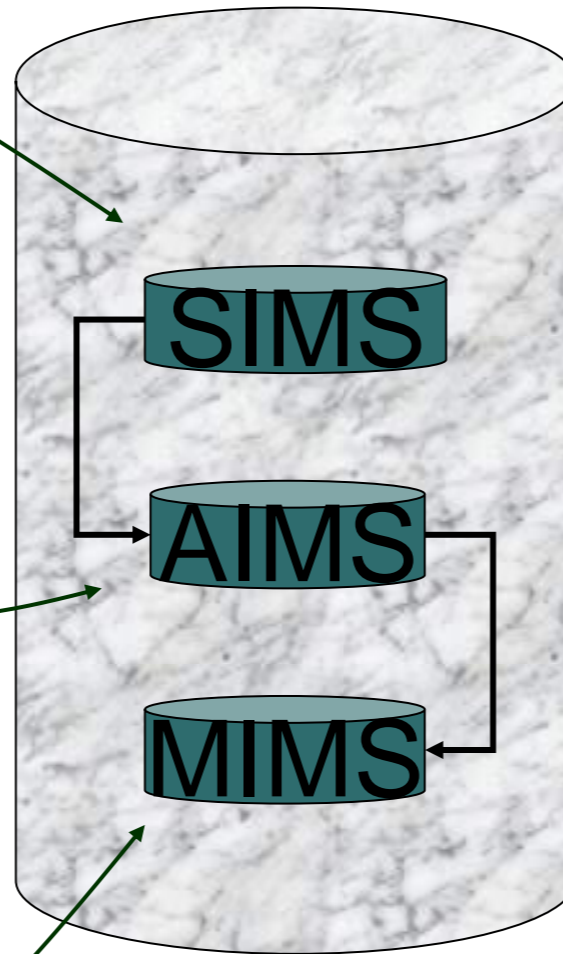


Assay Information Management System

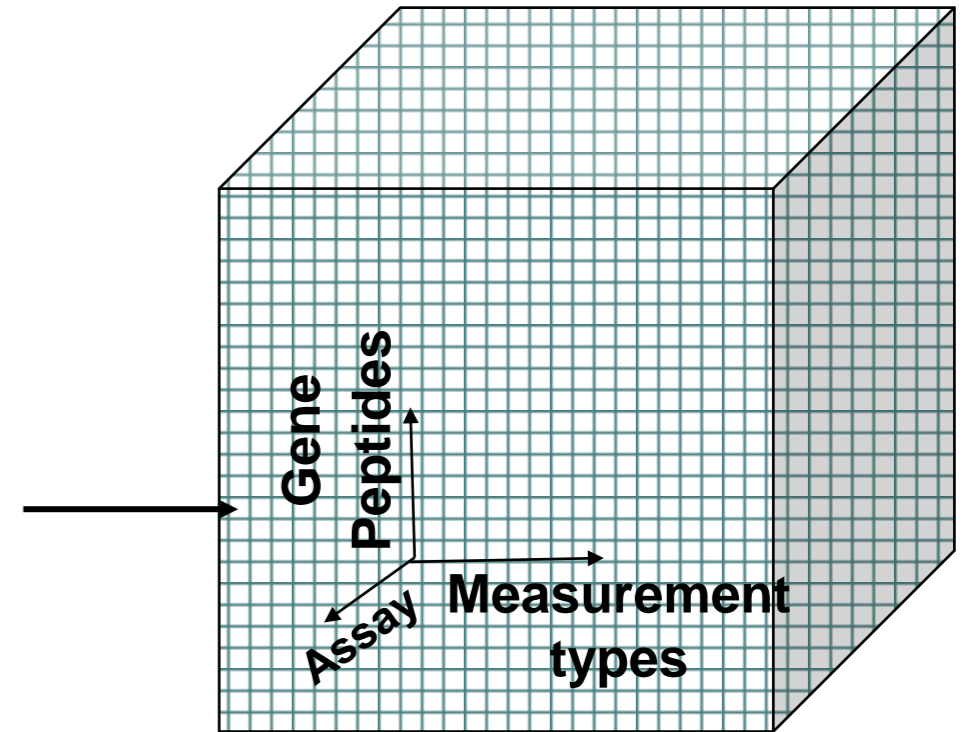
MIMS



Marker Information Management System



Repository



Data Warehouse

Dual functionality: core and customizable

- the core:
 - Sample tracking
 - Access rights
- Flexibility (configurable by the end-user)
 - experiment metadata
 - disease-specific annotation
 - Data types

Web interface for sample component (SIMS)



MOLECULAR PHENOTYPING TO ACCELERATE GENOMIC EPIDEMIOLOGY

Logout Persons Samples Aliquots Filter DisplaySettings Export

List of aliquots

(1-1000 of 5160 items [next](#))

Source: Person: Sample:

<u>Id</u>	<u>Date</u>	<u>Status</u>	<u>Location</u>	<u>Storage condition</u>	<u>Planned user</u>	<u>Fit for use</u>	<u>Processing</u>	<u>Transport condition</u>	<u>Received intact</u>	<u>Q</u>	<u>Creator</u>	<u>Create date</u>	<u>Edi</u>
1111-URN1-A	2005-Nov-23		LN2	room temperature	UOXF	yes	other	ambient	unknown as yet		admin	2005-Nov-23	edi
1111-URN1-B	2006-Apr-03		LN2	room temperature	UOXF	yes	other	ambient	unknown as yet	▼	admin	2006-Apr-03	edi
1111-URN1-C	2006-Apr-13		LN2	room temperature	UOXF	yes	other	ambient	no but usable		admin	2006-Apr-13	edi
1111-SFS1-A	2006-Jan-24		LN2	room temperature	UOXF	yes	other	ambient	unknown as yet		ocdem	2006-Jan-24	edi
1111-SFS1-B	2006-Jan-24		LN2	room temperature	UOXF	yes	other	ambient	unknown as yet		ocdem	2006-Jan-24	edi
1111-SFS1-C	2006-Jan-24		LN2	room temperature	UOXF	yes	other	ambient	unknown as yet		ocdem	2006-Jan-24	edi
1111-SFS1-D	2006-Jan-24		LN2	room temperature	UOXF	yes	other	ambient	unknown as yet		ocdem	2006-Jan-24	edi
OX0011-PLM1-A	2005-Jun-02		other	-80C	IMPERIAL	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0011-PLM1-B	2005-Jun-02		other	-80C	NOVO	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0011-PLM1-C	2005-Nov-04		LN2	room temperature	UOXF	yes	other	ambient	unknown as yet		admin	2005-Nov-04	edi
OX0012-PLM1-A	2005-Jun-02		other	-80C	IMPERIAL	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0012-PLM1-B	2005-Jun-02		other	-80C	NOVO	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0013-PLM1-A	2005-Jun-02		other	-80C	IMPERIAL	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0013-PLM1-B	2005-Jun-02		other	-80C	NOVO	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0014-PLM1-A	2005-Jun-02		other	-80C	IMPERIAL	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0014-PLM1-B	2005-Jun-02		other	-80C	NOVO	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0015-PLM1-A	2005-Jun-02		other	-80C	IMPERIAL	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0015-PLM1-B	2005-Jun-02		other	-80C	NOVO	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0016-PLM1-A	2005-Jun-02		other	-80C	IMPERIAL	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0016-PLM1-B	2005-Jun-02		other	other	NOVO	other	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0017-PLM1-A	2005-Jun-02		other	-80C	IMPERIAL	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi
OX0017-PLM1-B	2005-Jun-02		other	-80C	NOVO	yes	other	other	unknown as yet		ocdem	2005-Jun-02	edi

Assay Information Management System

Logout StudyGroups Assays Vocabularies Help

List of Assays

358 items

Change

Technology: CLINPROT MS

Change

Study group: CNG-MM-5

Filter

Edit

New

Column

Upload

Export

▼	Id	Name	Aliquot	D	Design		ExpAcquisition			ExpProcessing		SPreparation		SStorage	Edit	Clone
					Plate number	Well positions	Bead type	Plate numb	Matrix	FT	Smoothing	Barcode	Sample type	Storage after arrival		
▲	CNG-MM-5-1		OX0125-PPC1-A	📁	G1		C8	5	HCCA	none	none	P00002P		-80C	✎	🔄
▲	CNG-MM-5-2		OX0125-PPC1-A	📁	G2		C8	5	HCCA	none	none	P00002P		-80C	✎	🔄
▲	CNG-MM-5-3		OX0125-PPC1-A	📁	H1		C8	5	HCCA	none	none	P00002P		-80C	✎	🔄
▲	CNG-MM-5-4		OX0125-PPC1-A	📁	H2		C8	5	HCCA	none	none	P00002P		-80C	✎	🔄
▲	CNG-MM-5-5		OX0126-PPC1-A	📁	I13		C8	5	HCCA	none	none	P00002Q		-80C	✎	🔄
▲	CNG-MM-5-6		OX0126-PPC1-A	📁	I14		C8	5	HCCA	none	none	P00002Q		-80C	✎	🔄
▲	CNG-MM-5-7		OX0126-PPC1-A	📁	J13		C8	5	HCCA	none	none	P00002Q		-80C	✎	🔄
▲	CNG-MM-5-8		OX0126-PPC1-A	📁	J14		C8	5	HCCA	none	none	P00002Q		-80C	✎	🔄
▲	CNG-MM-5-9		OX0130-PPC1-A	📁	E21		C8	5	HCCA	none	none	P00002U		-80C	✎	🔄
▲	CNG-MM-5-10		OX0130-PPC1-A	📁	E22		C8	5	HCCA	none	none	P00002U		-80C	✎	🔄
▲	CNG-MM-5-11		OX0130-PPC1-A	📁	F21		C8	5	HCCA	none	none	P00002U		-80C	✎	🔄
▲	CNG-MM-5-12		OX0130-PPC1-A	📁	F22		C8	5	HCCA	none	none	P00002U		-80C	✎	🔄
▲	CNG-MM-5-13		OX0130-PPC1-A	📁	O13		C8	5	HCCA	none	none	P00002U		-80C	✎	🔄
▲	CNG-MM-5-14		OX0130-PPC1-A	📁	O14		C8	5	HCCA	none	none	P00002U		-80C	✎	🔄
▲	CNG-MM-5-15		OX0130-PPC1-A	📁	P13		C8	5	HCCA	none	none	P00002U		-80C	✎	🔄
▲	CNG-MM-5-16		OX0130-PPC1-A	📁	P14		C8	5	HCCA	none	none	P00002U		-80C	✎	🔄
▲	CNG-MM-5-17		OX0134-PPC1-A	📁	A5		C8	5	HCCA	none	none	P00002Y		-80C	✎	🔄
▲	CNG-MM-5-18		OX0134-PPC1-A	📁	A6		C8	5	HCCA	none	none	P00002Y		-80C	✎	🔄
▲	CNG-MM-5-19		OX0134-PPC1-A	📁	B5		C8	5	HCCA	none	none	P00002Y		-80C	✎	🔄
▲	CNG-MM-5-20		OX0134-PPC1-A	📁	B6		C8	5	HCCA	none	none	P00002Y		-80C	✎	🔄
▲	CNG-MM-5-21		OX0136-PPC1-A	📁	I17		C8	5	HCCA	none	none	P000030		-80C	✎	🔄
▲	CNG-MM-5-22		OX0136-PPC1-A	📁	I18		C8	5	HCCA	none	none	P000030		-80C	✎	🔄
▲	CNG-MM-5-23		OX0136-PPC1-A	📁	J17		C8	5	HCCA	none	none	P000030		-80C	✎	🔄
▲	CNG-MM-5-24		OX0136-PPC1-A	📁	J18		C8	5	HCCA	none	none	P000030		-80C	✎	🔄
▲	CNG-MM-5-25		OX0136-PPC1-A	📁	M11		C8	5	HCCA	none	none	P000030		-80C	✎	🔄
▲	CNG-MM-5-26		OX0136-PPC1-A	📁	M12		C8	5	HCCA	none	none	P000030		-80C	✎	🔄
▲	CNG-MM-5-27		OX0136-PPC1-A	📁	N11		C8	5	HCCA	none	none	P000030		-80C	✎	🔄
▲	CNG-MM-5-28		OX0136-PPC1-A	📁	N12		C8	5	HCCA	none	none	P000030		-80C	✎	🔄
▲	CNG-MM-5-29		OX0139-PPC1-A	📁	G3		C8	5	HCCA	none	none	P000033		-80C	✎	🔄

- tab delimited in, tab delimited out
- independent sample and assay data collection
- integrated export (subject<=sample<=>assay=>study)
- both LIMS and data exchange functionalities are important
- customisable interface AND compliance to public data exchange standards

CHALLENGE 2: using data exchange standards for data integration

Standards for systems biology

Alvis Brazma, Maria Krestyaninova and Ugis Sarkans

Abstract | High-throughput technologies are generating large amounts of complex data that have to be stored in databases, communicated to various data analysis tools and interpreted by scientists. Data representation and communication standards are needed to implement these steps efficiently. Here we give a classification of various standards related to systems biology and discuss various aspects of standardization in life sciences in general. Why are some standards more successful than others, what are the prerequisites for a standard to succeed and what are the possible pitfalls?

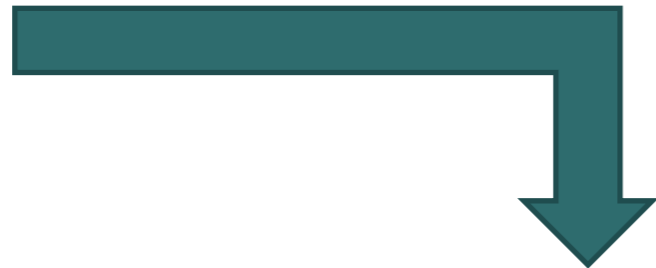
Barbarism is the absence of standards to which appeal can be made. José Ortega y Gasset

Historically, standards emerged from a need for a recommended practice in the manufacturing of products. The main purpose of standards is to help to fit together many pieces to make a useful whole. In information technology, standards are needed to exchange

What is the aim of these standardization initiatives? How many of these standards are used and how many have contributed to advances in biology? The goal of this review is to help systems biologists to navigate the rapidly changing maze of data standards in life sciences, and to encourage biologists to provide feedback on these developments. Such feedback is necessary to ensure that the developed standards are as close as possible to

Sample annotation harmonisation

Cohort 1
Cohort 2
Cohort 3
Cohort 4



Problems

- different languages (for headers)
- different assays, units, formats for same measure
- derived variables (eg HOMA)
- different component measures (eg cig smoking)
- different data missingness for related variables in same cohort
- repeated measures (which one to choose...?)
- different medical contexts, granularity...

Sample		Phenotypes					Genotypes			
		F Glucose	BMI	T2D Y/N	Age	Gender	GWA Y/N	SNP1	SNP2	SNP3
NFBC1		✓	✓	✓	✓	✓	x	✓	✓	✓
NFBC2		✓	✓	✓	✓	✓	✓	✓	✓	✓
NFBC3		x	✓	x	✓	✓	✓	✓	✓	✓
NFBC4		✓	✓	x	✓	✓	✓	✓	✓	x
NFBC5		✓	✓	✓	✓	✓	✓	✓	✓	x
NFBC6		✓	✓	✓	✓	✓	✓	x	✓	✓
NFBC7		✓	✓	✓	✓	✓	✓	✓	✓	✓
NFBC8		✓	✓	✓	✓	✓	x	✓	✓	✓
KORA1		x	✓	✓	x	✓	✓	✓	✓	✓
KORA2		x	✓	✓	✓	✓	✓	✓	✓	✓
KORA3		✓	✓	x	✓	✓	✓	✓	✓	✓
KORA4		✓	✓	✓	✓	✓	✓	✓	✓	✓
KORA5		✓	✓	✓	✓	✓	x	x	✓	✓
KORA6		✓	✓	✓	✓	✓	✓	✓	✓	✓
KORA7		✓	✓	✓	✓	✓	✓	✓	✓	✓
KORA8		✓	x	x	x	✓	✓	✓	✓	✓

This part easily expandable with standardised genetic and genomic data with controlled data formats, and ontological meta-data



Actual values could go in here (rather than availability) if deposit/access allowed



A few words about data access management (CHALLENGE 3)

Reasons for security

- Ethics
- Scientific competition
- Commercial interest

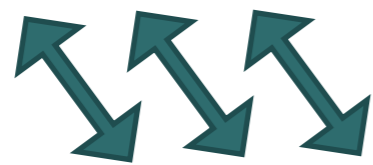
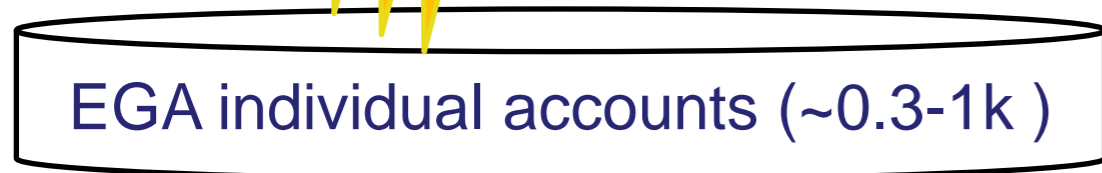
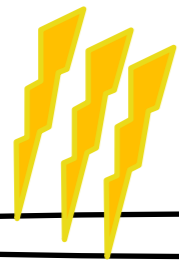
Data access show-case



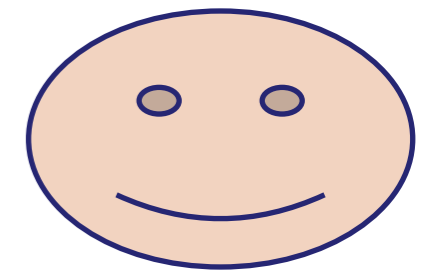
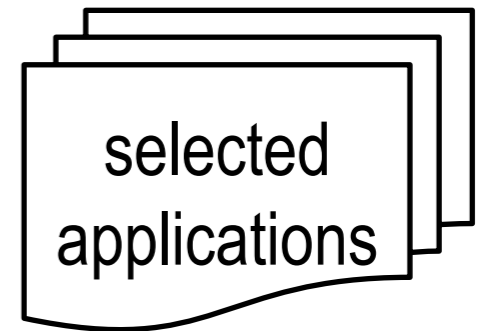
CDAC provides approval and amend in EGA the permissions for ind accounts

Application may cover broad areas of research for multiple projects, but may need to be updated or resubmitted as researchers, science or datasets available change

on-line DA mngnt tools

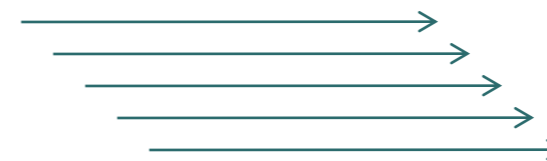


Researchers access approved data



“mediator” who prepares application to CDAC (individual ENGAGE researchers need to get institute to sign off on DTA)

Research idea

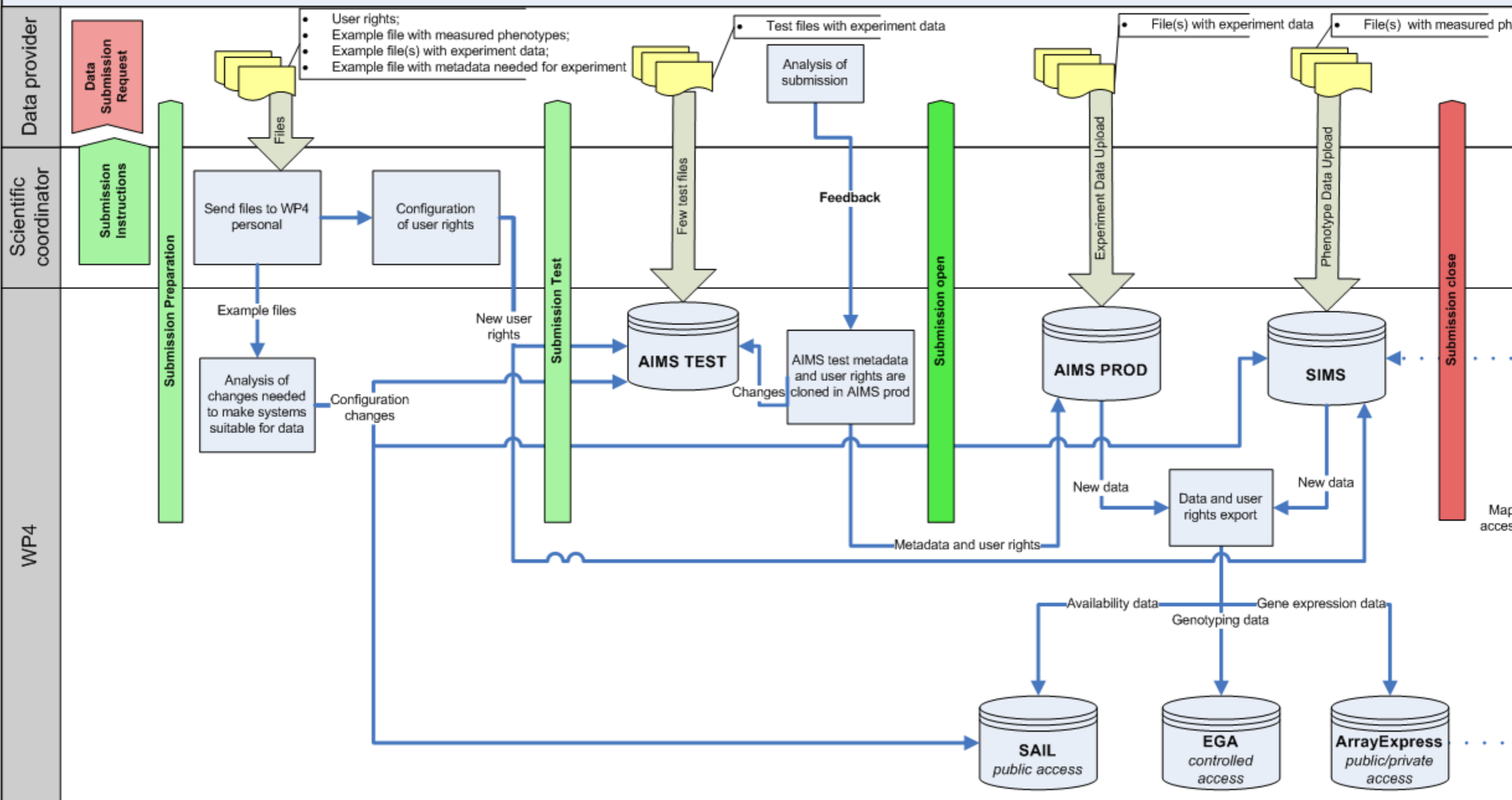


Data access management

- At each partner institution CDAC account holder (PI or PI's PA)
 - creates individual user account
 - grant access to studies available to the institution in accordance with policy documents signed by individuals
- Original collection-specific and newly developed policies can be referenced from the study
- In order to get data access to ENGAGE in EGA:
 - Apply for an account with ENGAGE PI at your institution
 - Sign necessary policy documents

So, operational model matters...

Data Submission Process in ENGAGE



Constructing a complex genomic epidemiology study with multiple data sources

light weight meta-data level integration at the stage of design

VS

deep data integration for downstream analysis

IT support for complex designs

Powering up and looking at complex traits:

1. enrich/control for phenotypes
2. maximise the sample size

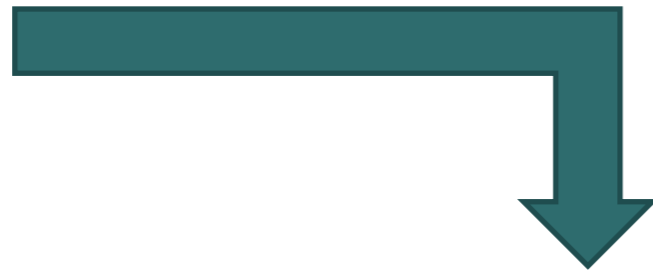
Technical/historical/ethical problems:

- inconsistent sample annotation
- diverse standards of medical records and procedures
- until study is designed, it's difficult to say whether access to data can be justified. Design of the study requires insights into data or metadata.

Problems

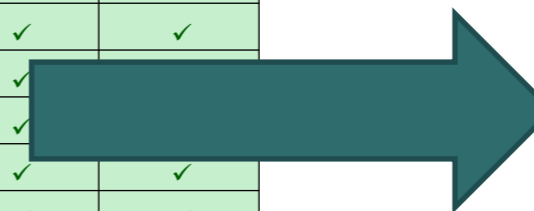
- different languages (for headers)
- different assays, units, formats for same measure
 - derived variables (eg HOMA)
- different component measures (eg cig smoking)
- different data missingness for related variables in same cohort
 - repeated measures (which one to choose...?)
 - different medical contexts, granularity...

Cohort 1
Cohort 2
Cohort 3
Cohort 4



		Phenotypes					Genotypes			
		F Glucose	BMI	T2D Y/N	Age	Gender	GWA Y/N	SNP1	SNP2	SNP3
Sample	NFBC1	✓	✓	✓	✓	✓	x	✓	✓	✓
	NFBC2	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NFBC3	x	✓	x	✓	✓	✓	✓	✓	✓
	NFBC4	✓	✓	x	✓	✓	✓	✓	✓	x
	NFBC5	✓	✓	✓	✓	✓	✓	✓	✓	x
	NFBC6	✓	✓	✓	✓	✓	✓	x	✓	✓
	NFBC7	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NFBC8	✓	✓	✓	✓	✓	x	✓	✓	✓
	KORA1	x	✓	✓	x	✓	✓	✓	✓	✓
	KORA2	x	✓	✓	✓	✓	✓	✓	✓	✓
	KORA3	✓	✓	x	✓	✓	✓	✓	✓	✓
	KORA4	✓	✓	✓	✓	✓	✓	✓	✓	✓
	KORA5	✓	✓	✓	✓	✓	x	x	✓	✓
	KORA6	✓	✓	✓	✓	✓	✓	✓	✓	✓
	KORA7	✓	✓	✓	✓	✓	✓	✓	✓	✓
	KORA8	✓	x	x	x	✓	✓	✓	✓	✓

This part easily expandable with standardised genetic and genomic data with controlled data formats, and ontological meta-data

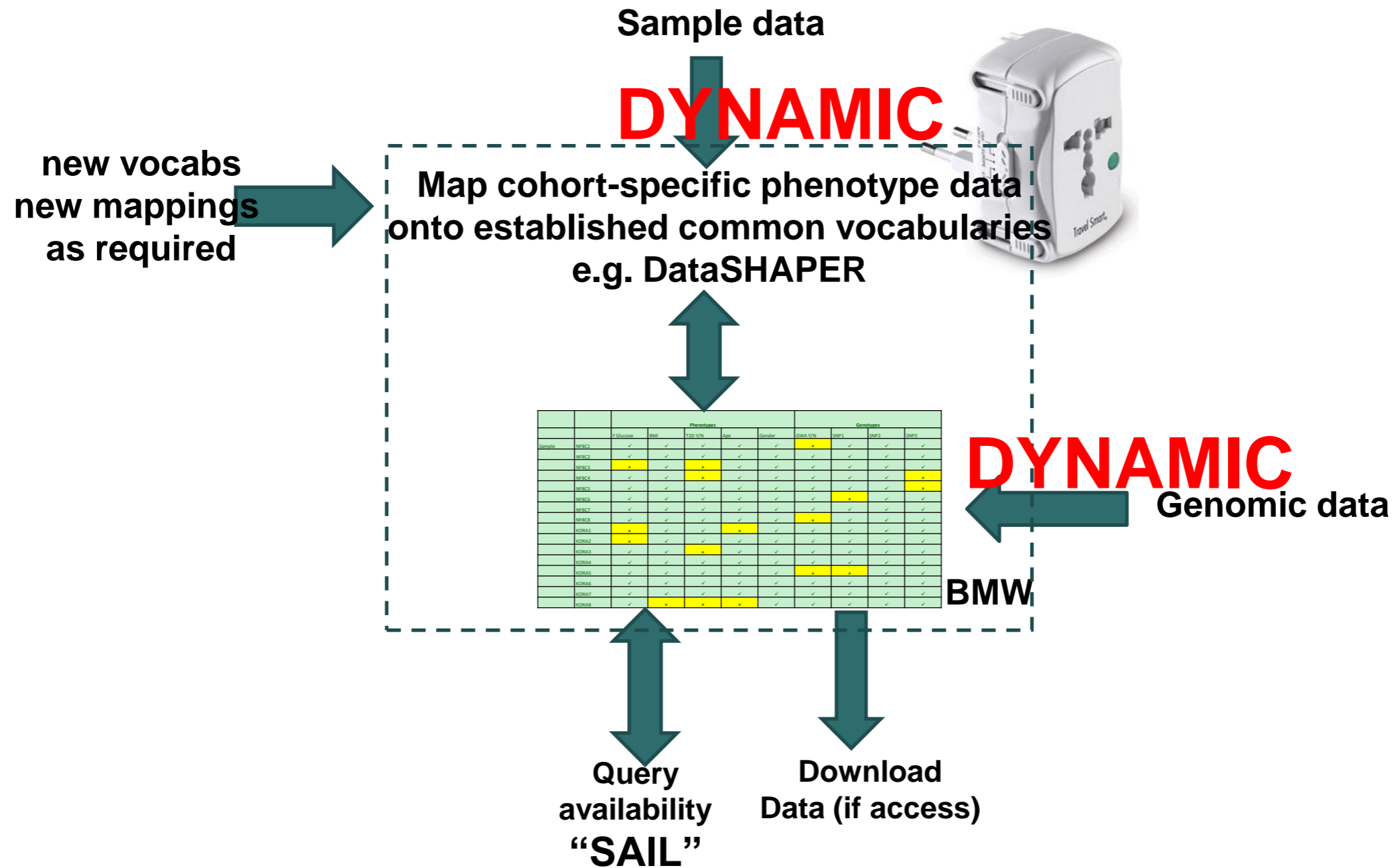


Actual values could go in here
(rather than availability) if deposit/access allowed

Harmonisation - translation of original biobank sample records into a common set of terms, to which samples can be re-mapped.

The ENGAGE (euengage.org) harmonisation is done within a context of a study for a limited number of variables. It provides added value annotation for samples on top of semantic links provided by the biobank standardisation initiatives and has a potential to provide meta-study view across multiple sample collections.

Ideal study design system

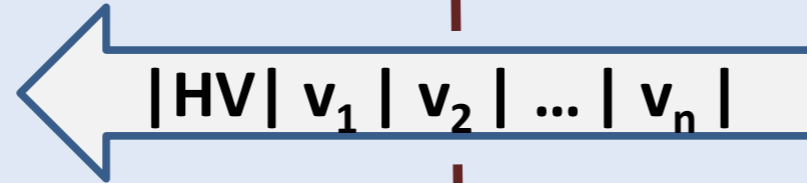
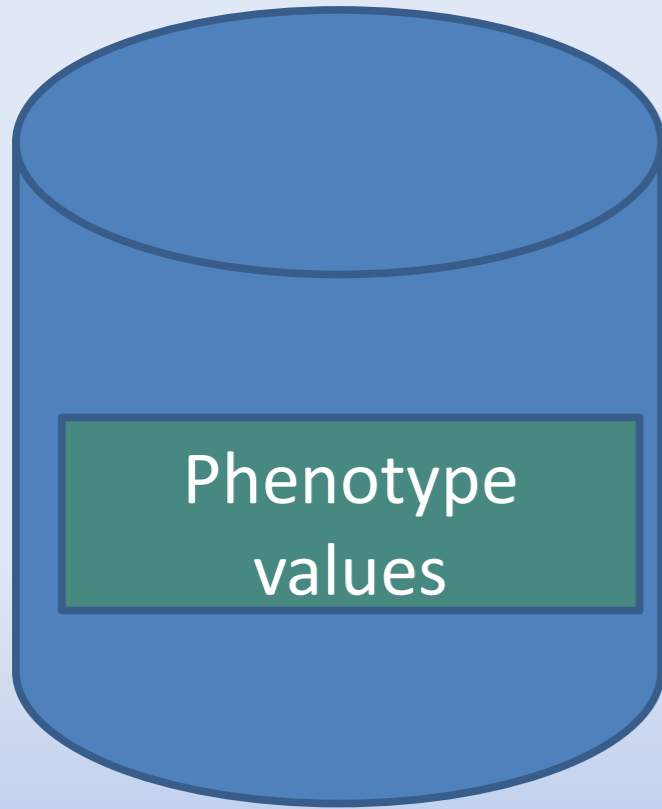


HV – harmonized vocabulary
 OV – original vocabulary

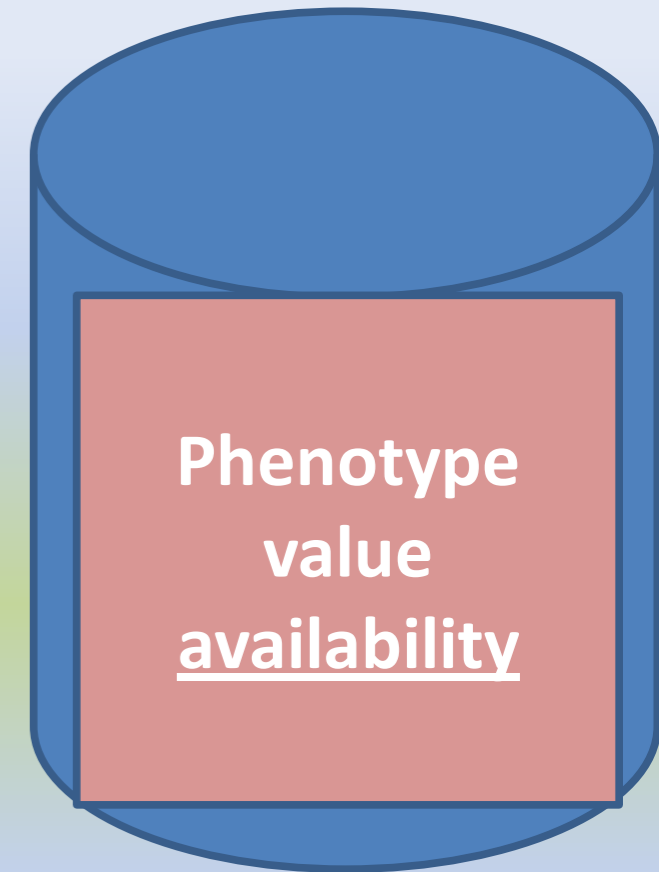
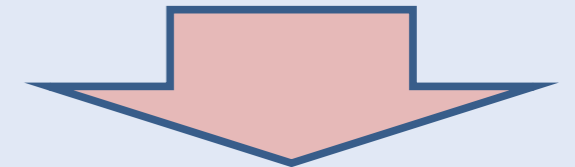
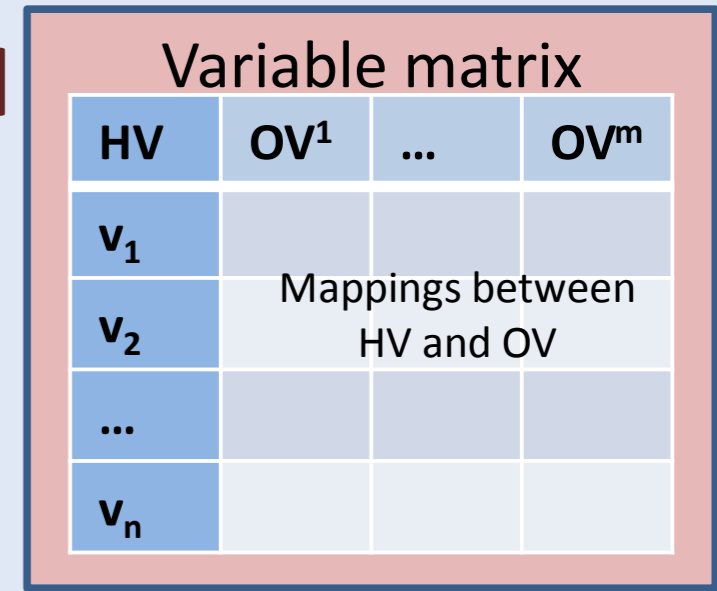
Framework

PHASE I

SIMS – Sample Information Management System (access control)

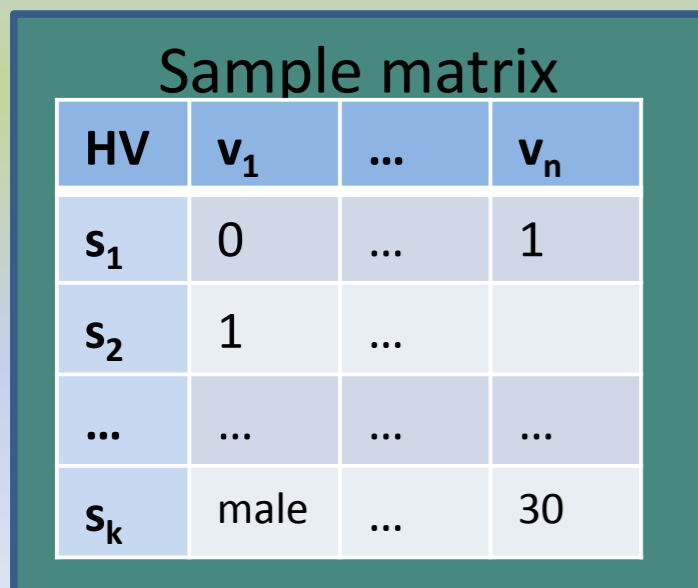


only HV



only availability data

PHASE II



availability data or real values

SAIL – Sample Availability Information System (public).
 Search and sample counts.

SAIL - defining the scope

- Aimed at large and complex content resources
- Summarizing, indexing and searching content

made possible through

- Plug-in vocabularies
- Vocabulary editing and harmonisation
- Sharing availability and meta data

two main functions of SAIL

- Report Constructor and Report Viewer - allow the user to build a complex filter and get the counting results
- Parameter manager: import of new vocabs, variables, mappings between, evidence tags to mappings, qualifiers for variables (e.g. Time)

How the tool looks like

Report constructor Repository Report 1

Parameters list Parameters tree

[NO FILTER] Select tag Search

Code	Name	Description	Records	V	E
MetS:DCHD	Date first CHD	Date first CHD	7	1	0
MetS:DMI	Date first MI	Date first MI	0	1	0
MetS:DS	Date first stroke	Date first stroke	0	1	0
MetS:EDU	Education	Education	2877	1	0
MetS:EOSIN	Eosinophils (0.02-0.05)	Blood Eosinophils	69	1	0
MetS:EXYR	Year examination	Year examination	8734	1	0
MetS:FAMH	Family History	Family History	69	1	0
MetS:FMHRT	Family history heart disease	Family history heart disease	998	1	1
MetS:FMSTRK	Family history stroke	Family history stroke	998	1	1
MetS:FMT2D	Family type 2 diabetes	Family type 2 diabetes	998	1	1
MetS:GENDT	Genotyping data	Availability of genotyping data	5913	1	1
MetS:GLU	Glucose	Glucose, mMol/L	16126	1	0
MetS:GLUM	Glucose medication	Glucose medication	0	1	1
MetS:GLUTM	Glucose w timing	Glucose w timing	5700	1	1
MetS:HB	Hb (13-17)	Blood Hb	69	1	0
MetS:HBA1C	Hba1C	Hba1C, %	10287	1	0
MetS:HDL	HDL	HDL, mMol/L	16169	1	0
MetS:HIP	Hip	Hip, cm	7525	1	0
MetS:HT	Height	Height, m	6811	1	0
MetS:IDF	ID Father	ID Father	0	1	0
MetS:IDM	ID Mother	ID Mother	0	1	0
MetS:IDP	ID Person	ID Person	5844	1	0
MetS:IMT	IMT	IMT	0	1	0
MetS:INS	Insulin	Insulin, mUM	660	1	0

Select Add Enumerations Relations View

Report request

Name	Enumeration	F
BMI		
Sex	Sex	
Transcriptomics data	Available	
Smoking status	Status	
Smoking quantity 1		
Genotyping data	Available	

Use split by repository
 Split by all repositories
 Split by selected repositories

Remove Clear Up Down

Make report

Search results:

Report constructor										Repository										Report 1									
Total records: 17271																													
Repository	NFBC66 5844												Genmets Case 946	Genmets Control 965	SW TG 8449	EGP 998		Oxford biobank 69											
MetS:BMI	5727												946	964	8305	996		69											
MetS:SEX.Sex	Man 2770			Woman 2957			@ 946	@ 964	@ 8305	Man 504	Woman 492	Man 39	Woman 30																
MetS:TRCPDT.Available	Not available 2770			Not available 2957			0					Available 39	Available 30																
MetS:SMK.Status	never 896	passed 608	current 1092	never 1174	passed 650	current 862	0																						
MetS:SMKQ1	59	353	1061	89	359	841	0																						
MetS:GENDT.Available	Available 59	Available 353	Available 1061	Available 89	Available 359	Available 841	0																						

How in practice?

- Fix the format for collecting the mapping data
- Provide annotation tools
- Redundancy control (unique ids for samples, merge/linking for variables)
- Re-annotate samples upon arrival in as much terms as possible
- Keep manual and automatic annotation separate

Conclusions

- Cooperative discovery IT follows the consensus logic of scientific community rather than imposing a standard data structure
- Fast customization as opposed to a generic, fit-it-all, solution
- Short term data interpretability is as important as compliance universal standards and domain-specific ontologies
- IT support for design is needed in addition to execution of analysis
- Cost-efficient access control for the sake of cooperation rather than competition
- number of data sharing transactions as a measure of service efficiency