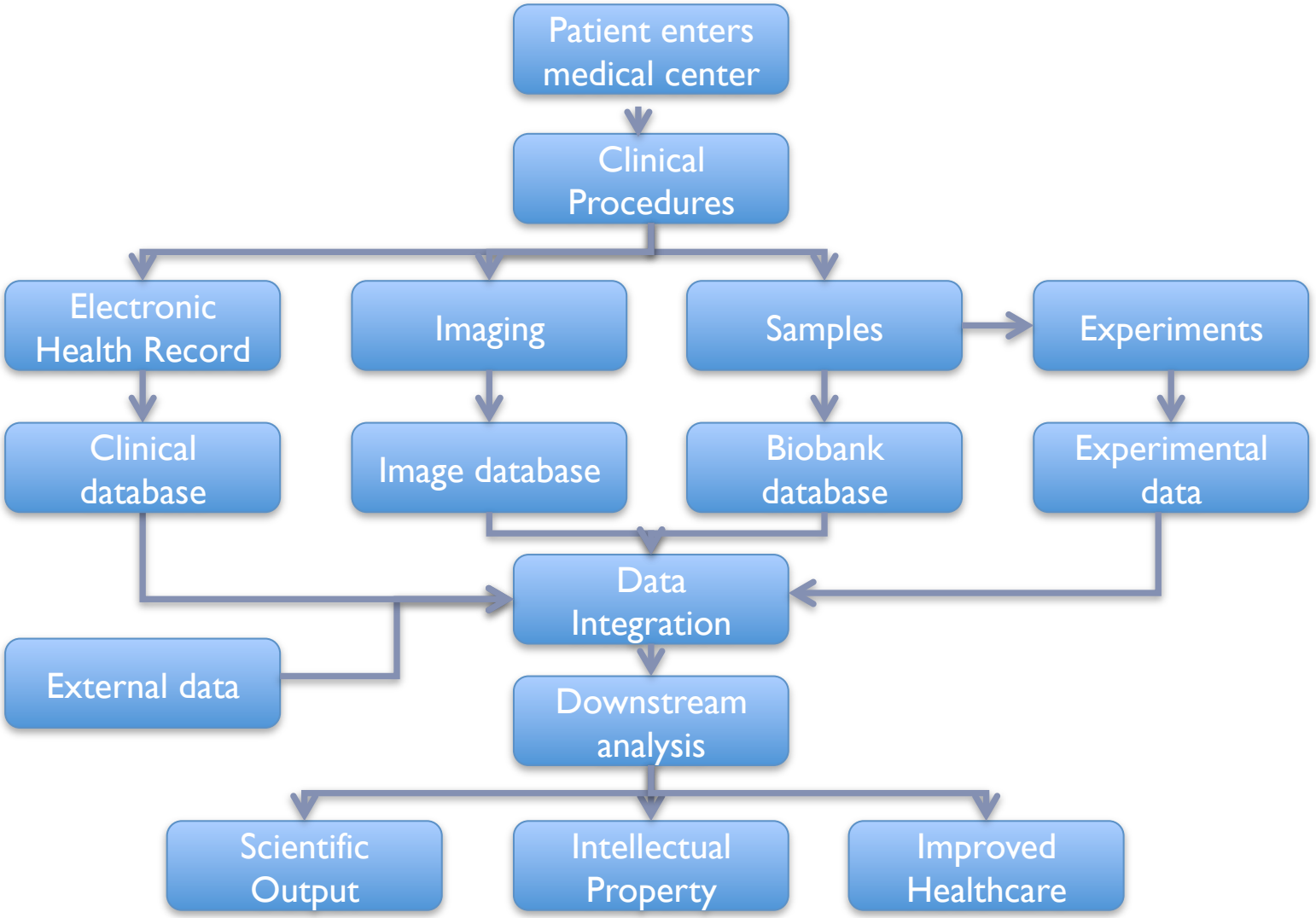
The tranSMART open source platform and the IMI-eTRIKS project

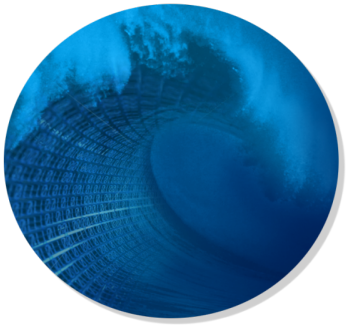
Ioannis Pandis, PhD \ Life Science Research Manager, Data Science Institute - Imperial College London

Translational Research Project Process



<i.pandis@imperial.ac.uk>

Data Challenges…: Not Just Analysis



Tsunami of genomic and clinical data

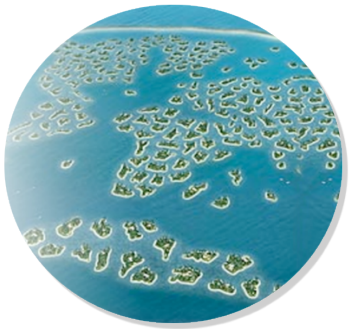
Consortia underestimate Data Management Effort

Multiple Partners; Fragmented Platforms

IT Research £ >> IT Service £

Weak standards Poor interoperability

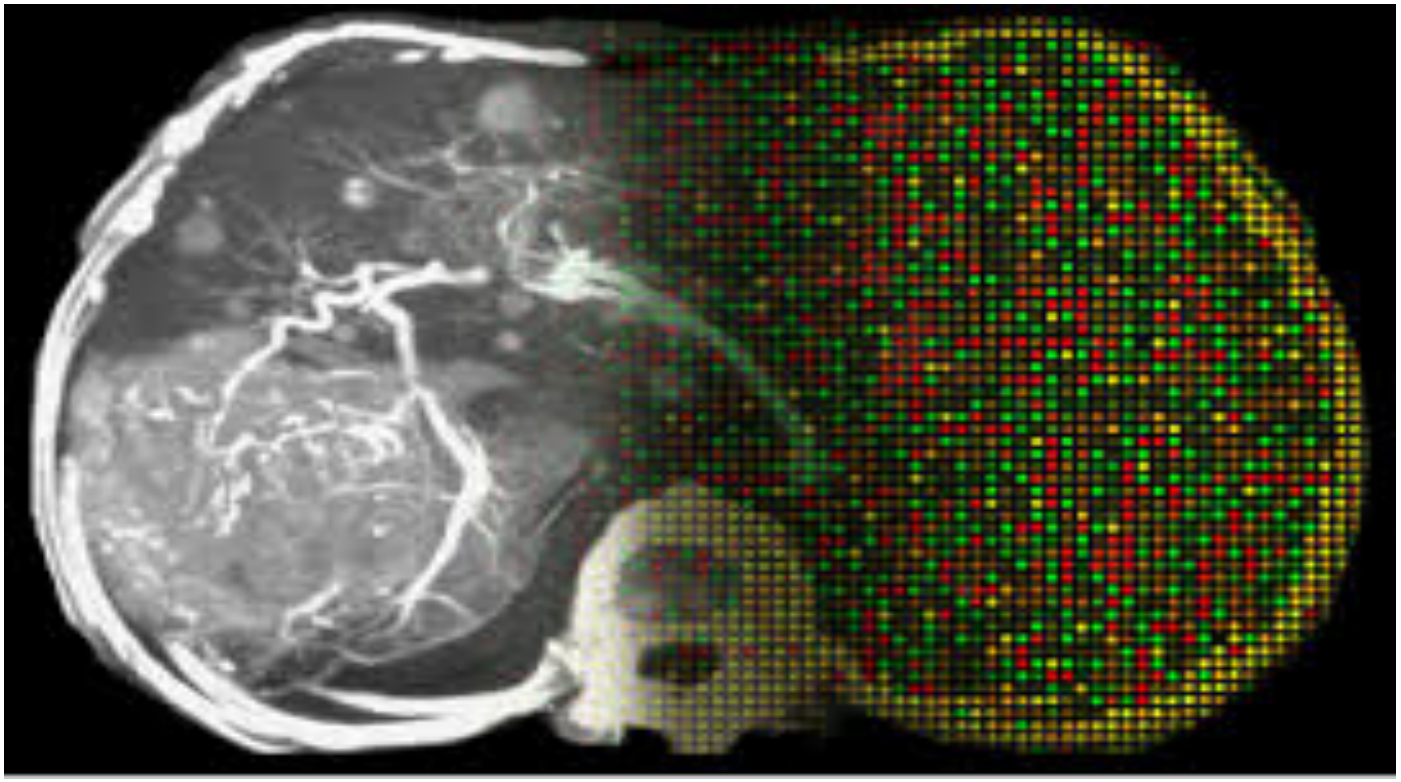
Patient Data Security



Islands of information

Sustainability

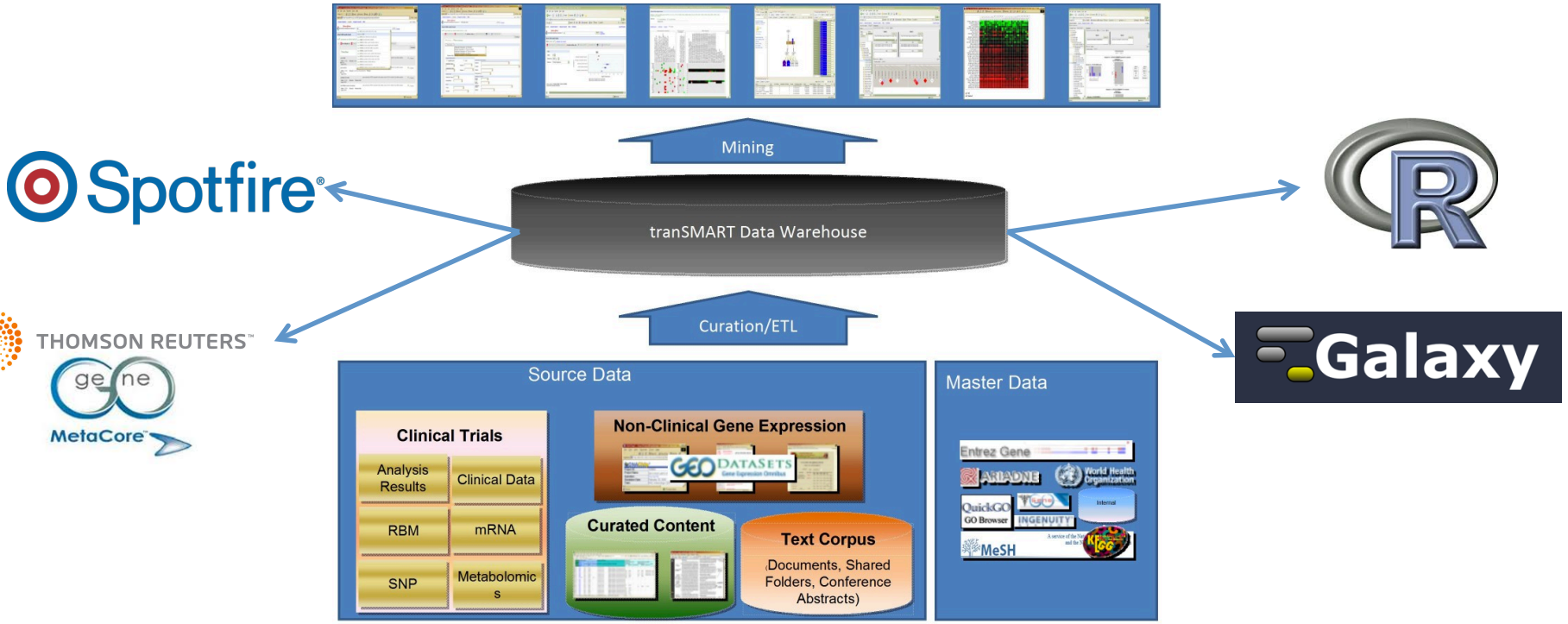
tranSMART : a fast growing platform



Objectives :

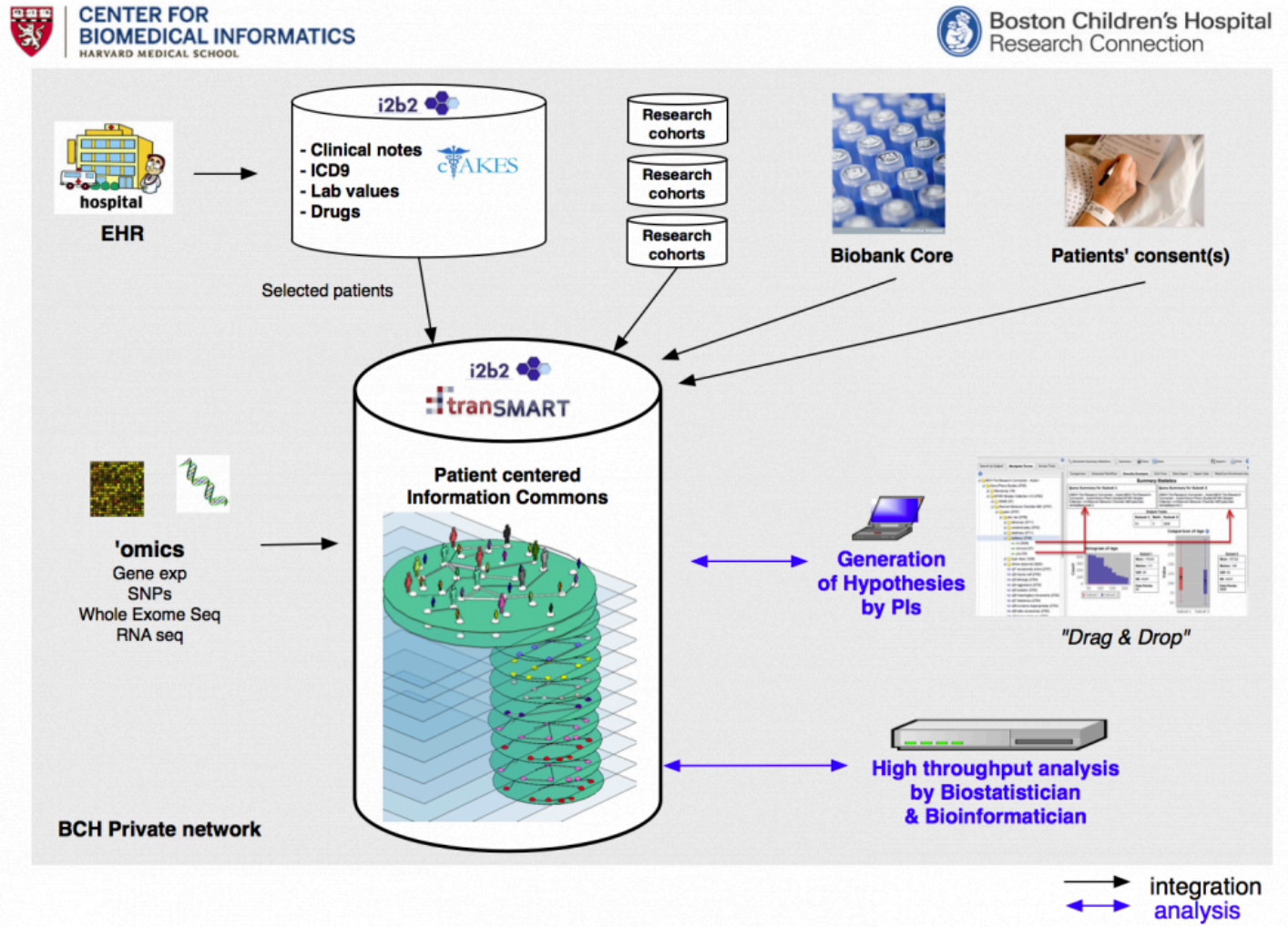
1. **Integration of clinical, biological and 'omics data in one place – hypothesis free –**
2. **Generation of hypothesis by Clinicians / Researchers**

tranSMART



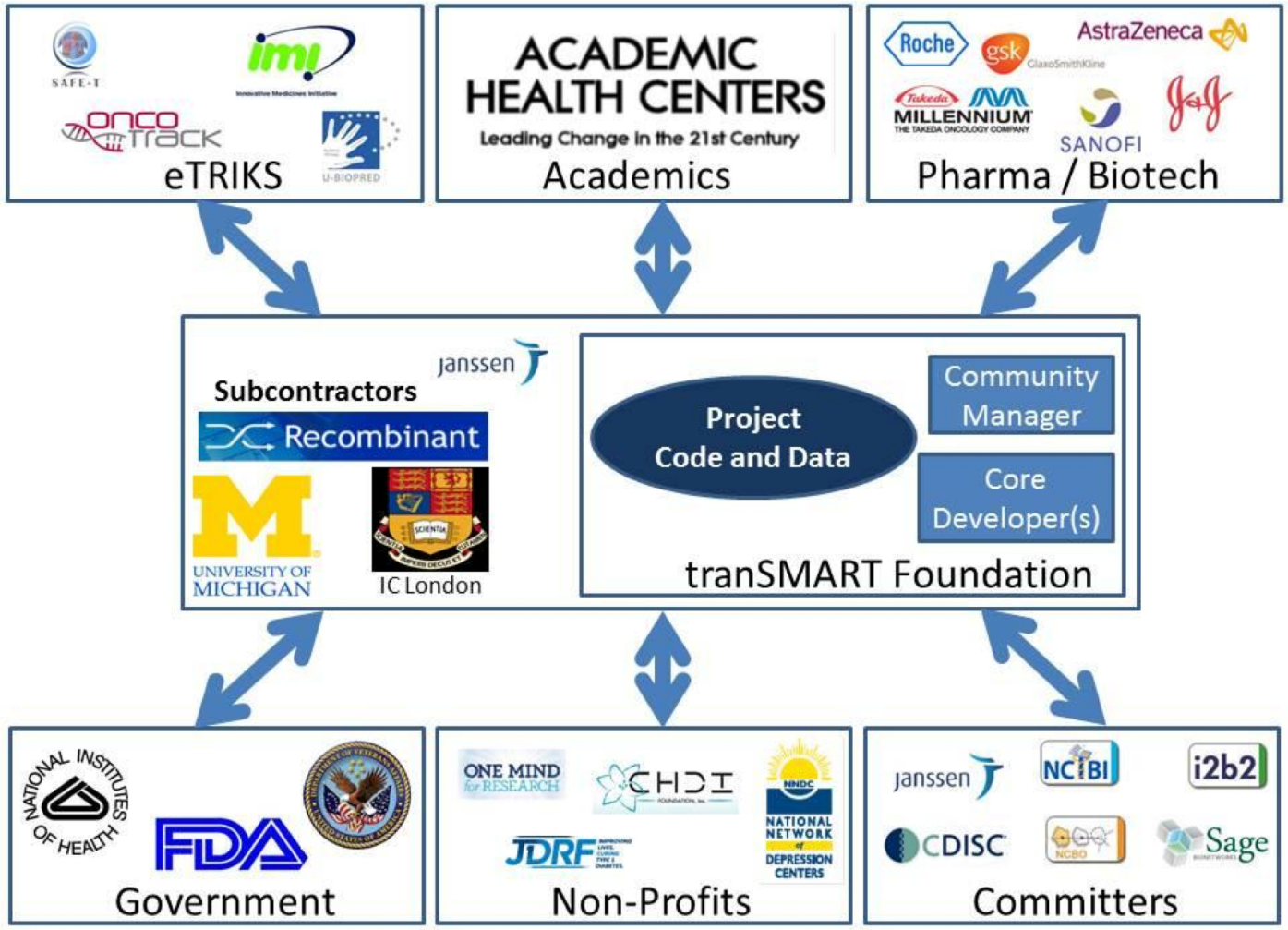
<i.pandis@imperial.ac.uk>

tranSMART @ Boston Children's Hospital



<i.pandis@imperial.ac.uk>

tranSMART Foundation



<i.pandis@imperial.ac.uk>

European Translational Research Information and Knowledge Management Services

Consortium of 16 Partners
Academic/Pharma/Coordination/Standards

Innovative Medicines Initiative

I Billion € — Public

I Billion € efpia — Private

Partnership

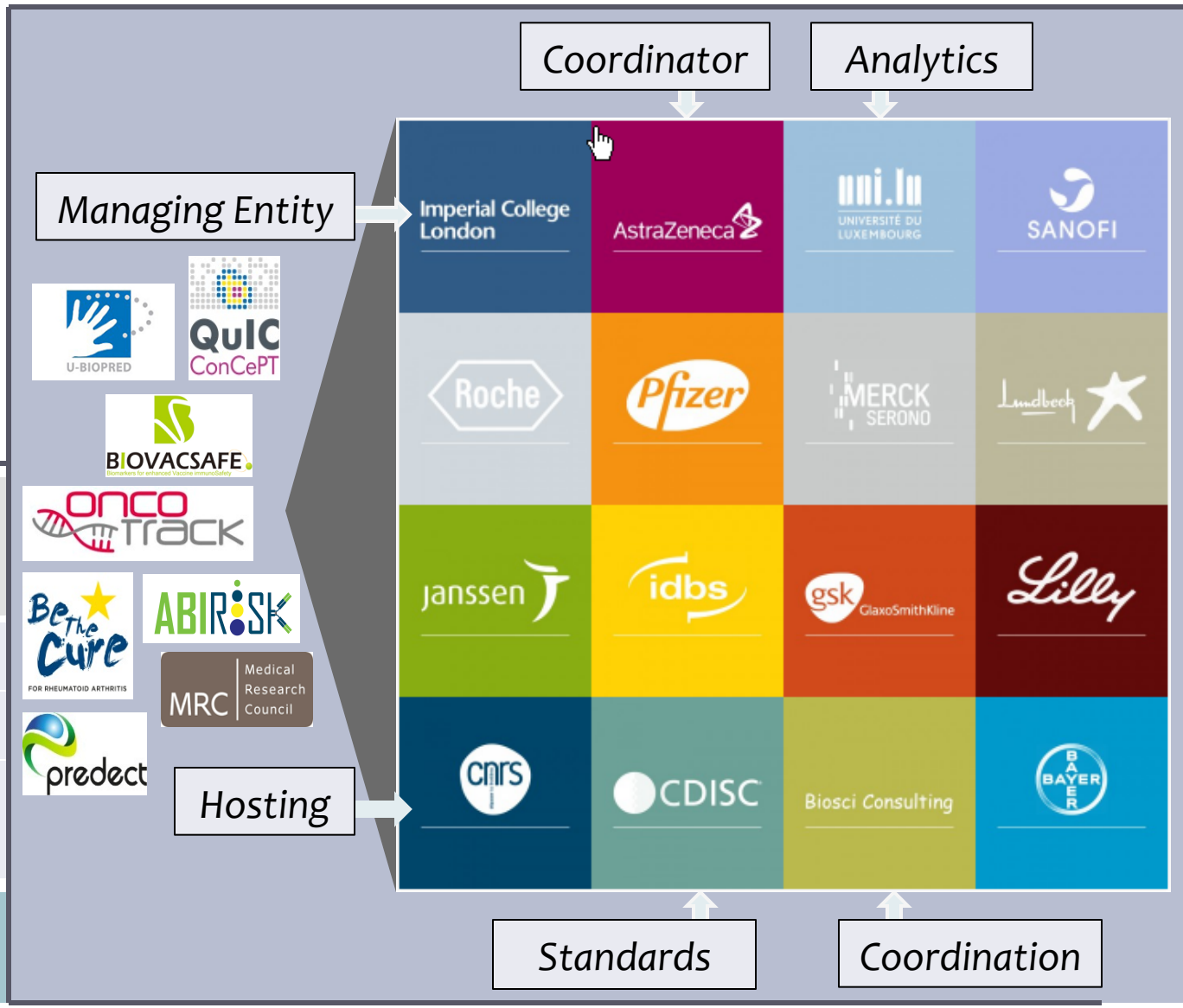
eTRIKS

~ $23M Euro

Oct-2012 – Sept-2017

Sustainable Open Platform

tranSMART v1.2



Goals from the *Full Project Proposal*

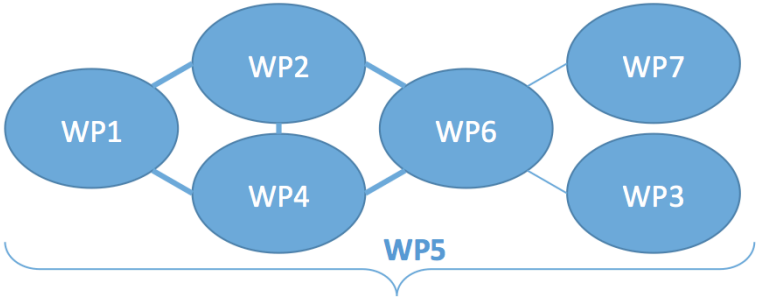
- **A cloud-based biomarker discovery infrastructure**
  - Support for OMICs, clinical data, cohorts
  - A rich set of analytical methods
- **A collaborative research management system**
  - Support large scale IMI programs
- **An integrative knowledge and management environment**
  - QC, ontology and terminology management, data provenance
- **A scientific knowledge service delivery platform**
  - Knowledge sharing, "knowledge as a service"
- **A federated knowledge sharing environment**
  - Support local instances of the eTRIKS platform (data privacy)
  - Share public data and data across projects where appropriate

Constraints/Priorities

- **Must use *tranSMART***
  - Infrastructure must be a fully *open technology stack*
- **Priority: Support client projects**

Work Packages

	WP Number	WP Name	WP Leads
Biosci (Program Mngmt)	WP1	Platform Deployment	*CNRS*
	WP2	Platform Development	*Imperial/Sanofi/Pfizer*
	WP3	Data Standards	*Roche/IDBS/Lilly/CDISC*
	WP4	Curation and Analysis	*Luxembourg/Sanofi/ Merck Serono*
	WP5	Management/Sustainability	*AstraZeneca*
	WP6	Community and Outreach	*Janssen/Biosci*
	WP7	Ethics	*GSK/CNRS/Bayer*



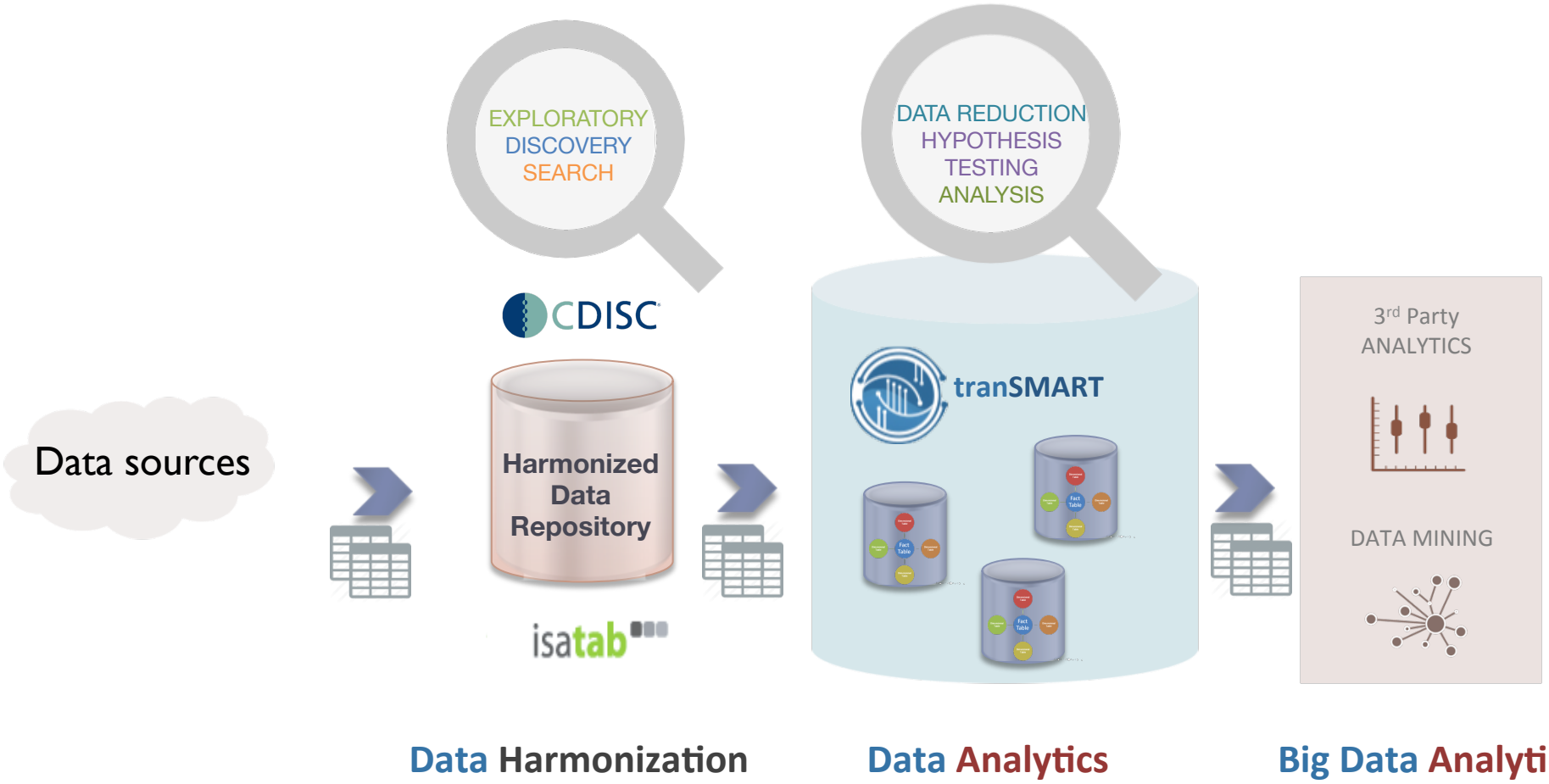
Horizontal Cooperation

Delivery Package: 2+ WPs. Client Delivery.
- eTRIKS/tranSMART 1.1, GUI
- ABI Risk, Oncotrack
- Public Server

Task Force: 3+ WPs

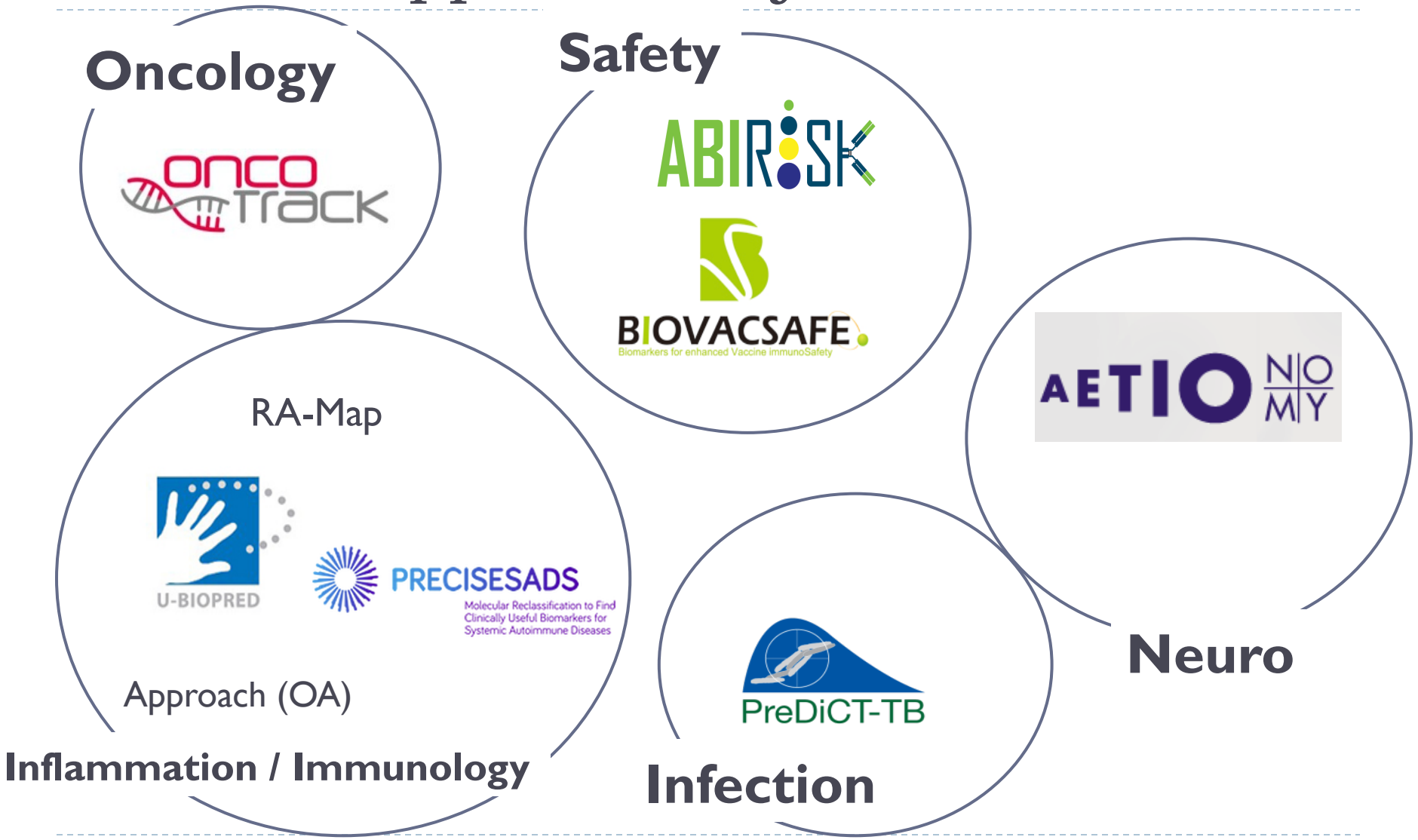
Functional Domain Teams: WP2/WP6

eTRIKS Platform Architecture



<i.pandis@imperial.ac.uk>

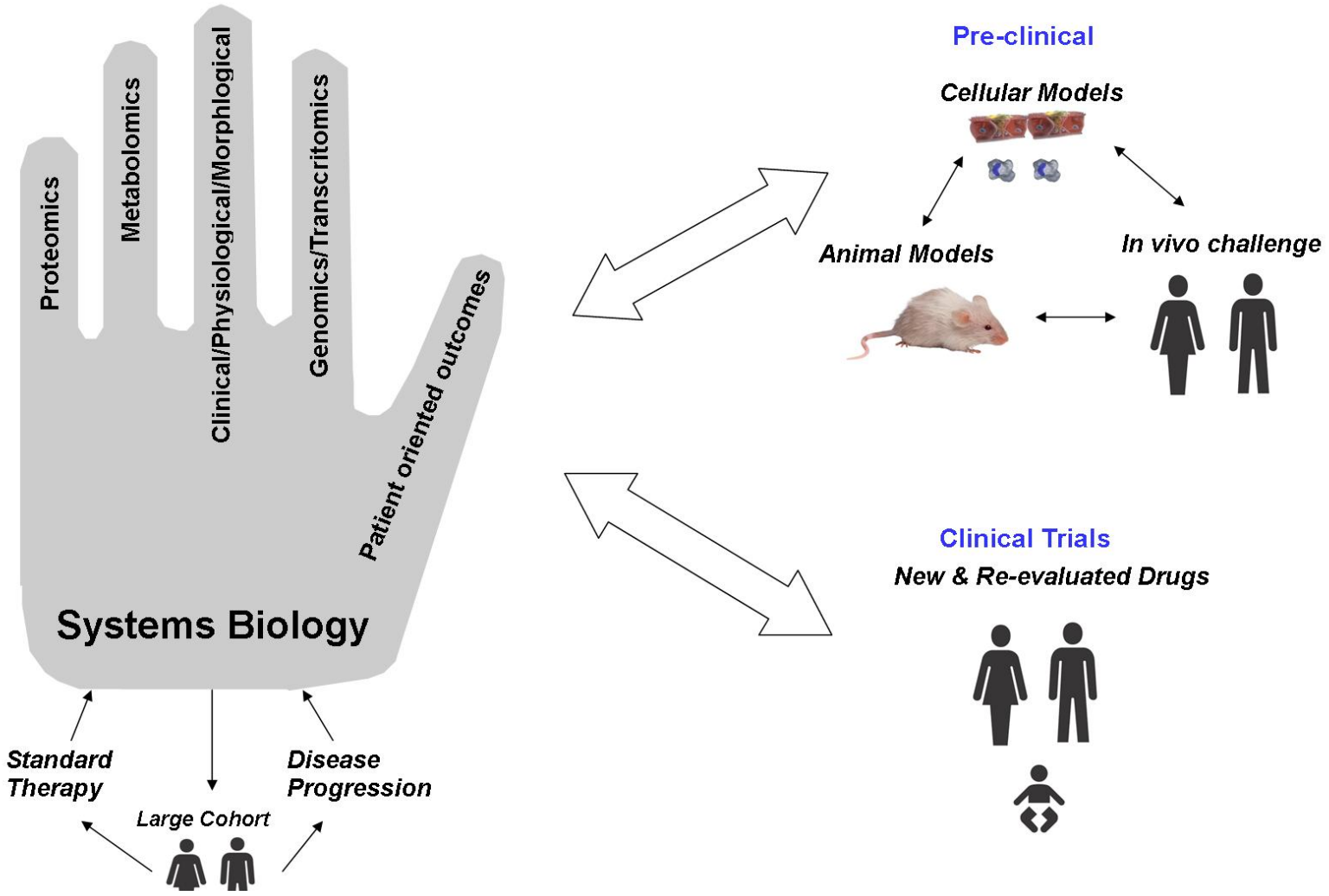
eTRIKS Supported Projects



<i.pandis@imperial.ac.uk>

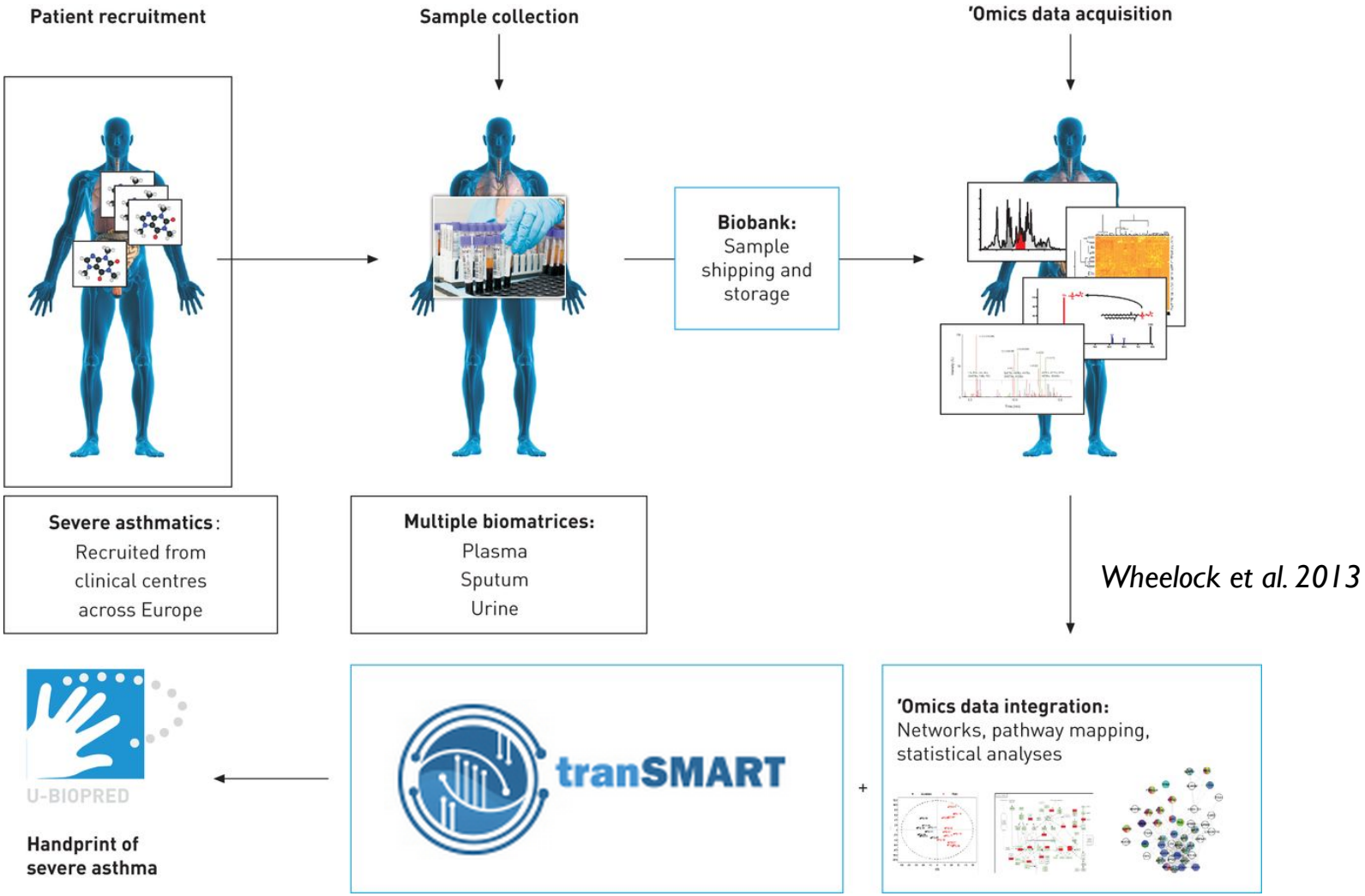
Unbiased BIOmarkers in PREDiction of respiratory disease outcomes

U-BIOPRED project



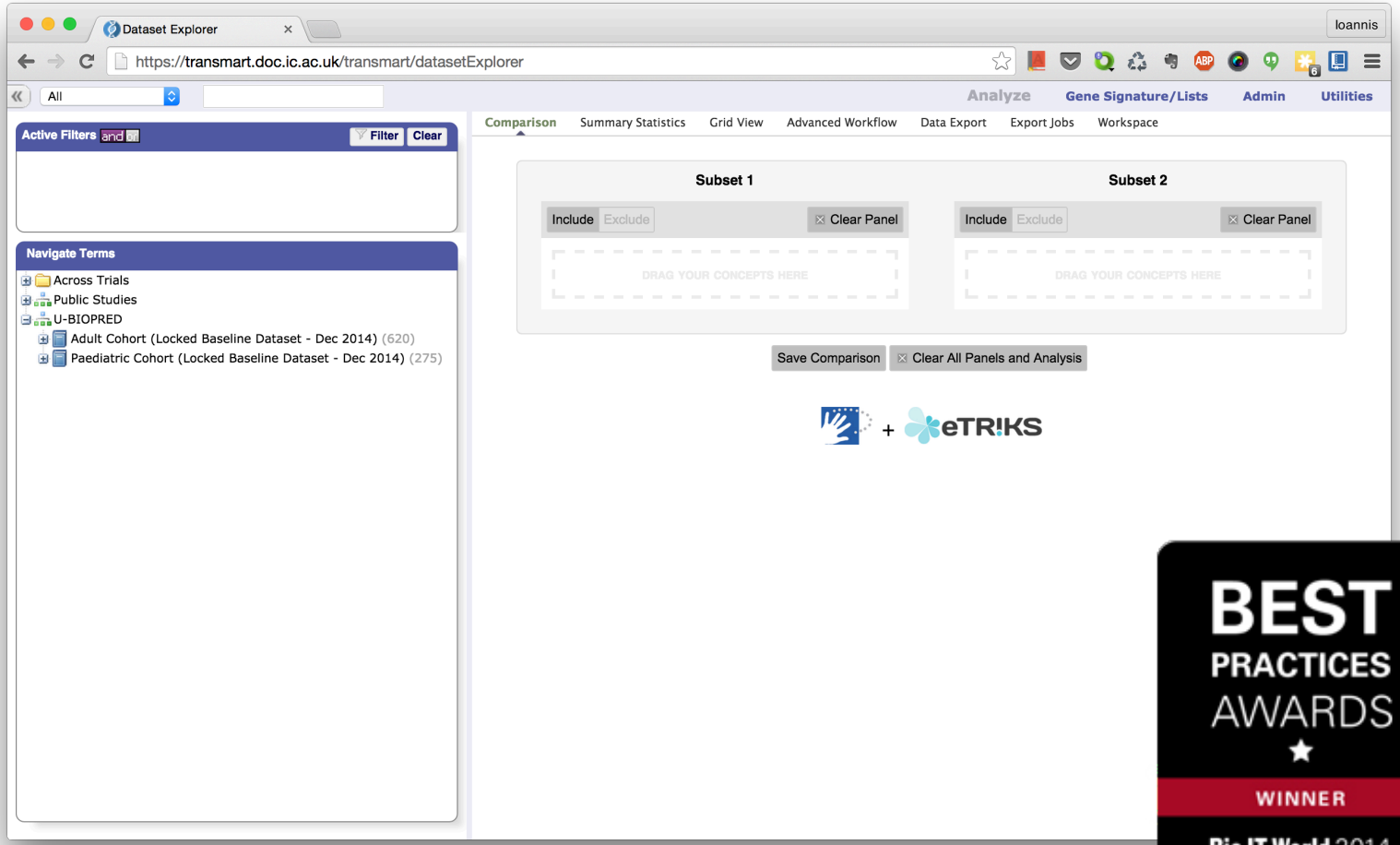
<i.pandis@imperial.ac.uk>

Sample and Data Flow



<i.pandis@imperial.ac.uk>

U-BIOPPRED eTRIKS-tranSMART



<i.pandis@imperial.ac.uk>

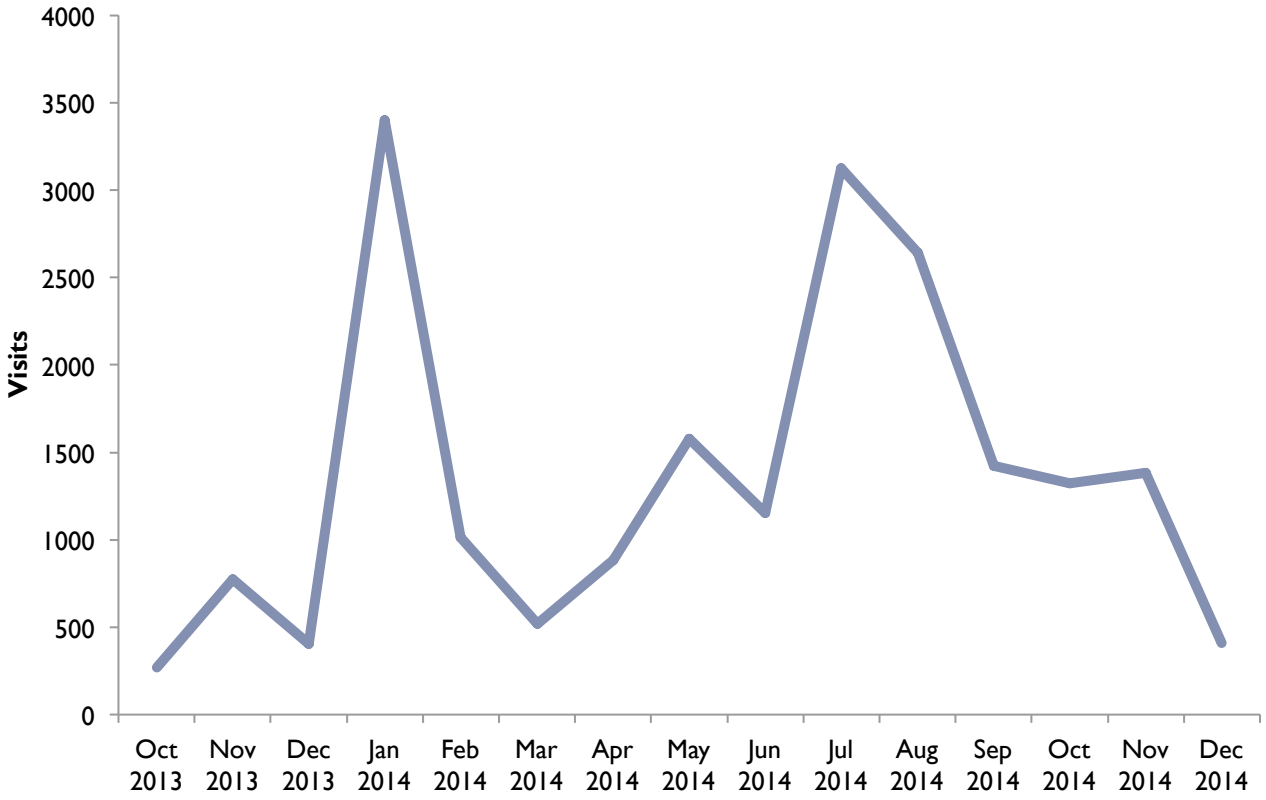
U-BIOPRED eTRIKS-tranSMART content to date

Cohort	Omics Type	Datasest	Number of Subjects/ Sample	Number of features	Total Features ($x10^3$)
Adult	Gene expression	Blood	312	54675	17058.6
		Bronchial Brushings	149	54675	8146.575
		Bronchial Biopsies	108	54675	5904.9
		Nasal Brushings	89	54675	4866.075
		Sputum	120	54675	6561
	Protein abundance	Serum Somalogic	579	1129	653.691
		Sputum Somalogic	121	1129	136.609
		Serum MS^e	131	130	17.03
		Sputum MS^e	100	3231	323.1
	Lipid abundance	Sputum MS^e	128	15622	1999.616
		Urine Eicosanoids	611	13	7.943
		Sputum Eicosanoids	321	13	4.173
	Breath metabolites	eNOSE	106	190	20.14
		GC/MS	62	7036	436.232
Paediatric	Lipid abundance	Urine Eicosanoids	35	13	
	Breath metabolites	Paediatric eNOSE	107	190	20.33
		Paediatric GC/MS	101	7036	710.636
Adult		Clinical / Phenotypic	620	2886	1789.32
Paediatric		Clinical / Phenotypic	275	1210	332.75
Total					**92656.3**

<i.pandis@imperial.ac.uk>

Visit Statistics

- Number of users: 227



<i.pandis@imperial.ac.uk>

Thank you!

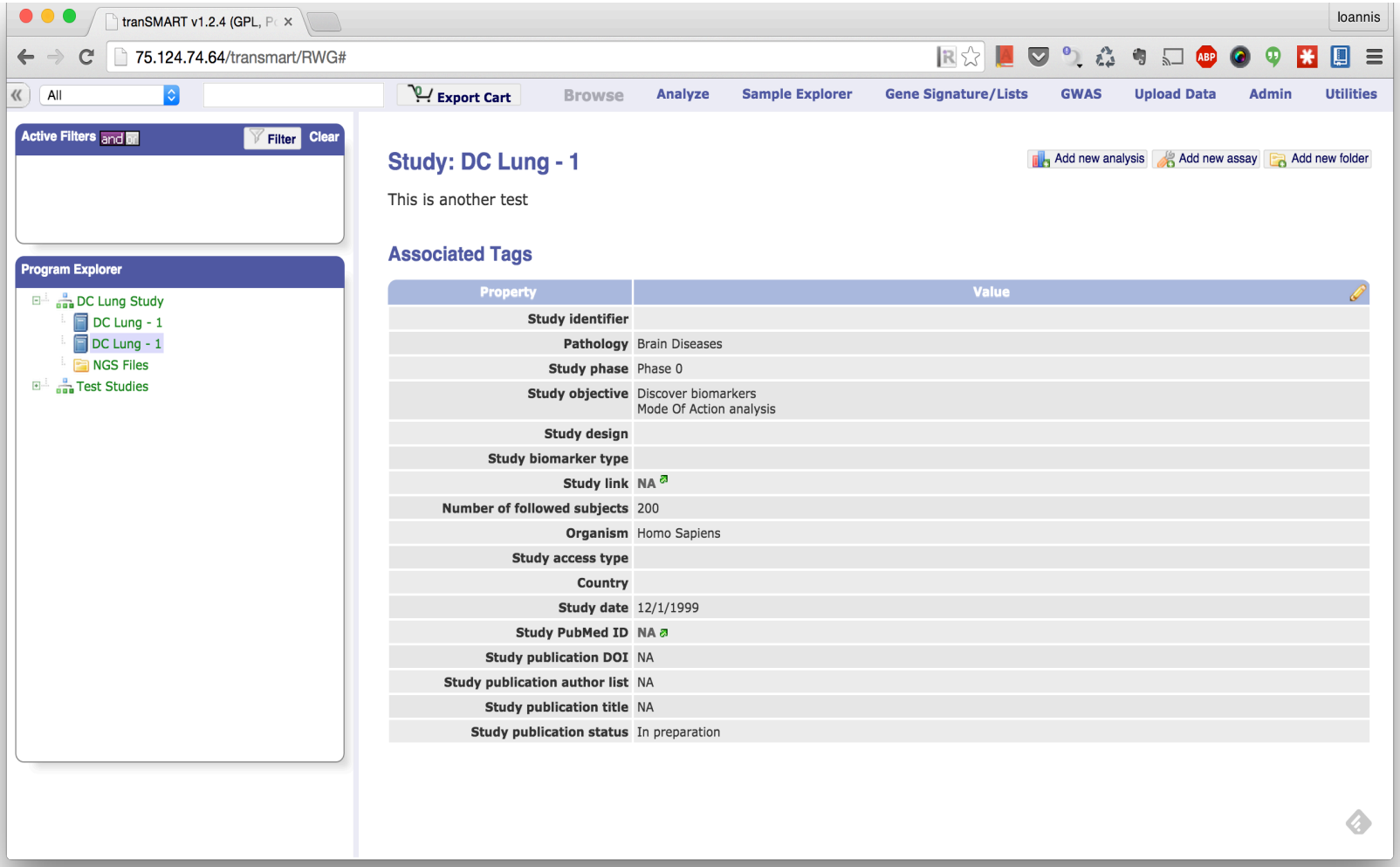
Imperial College London Data Science Institute

www.etriks.org

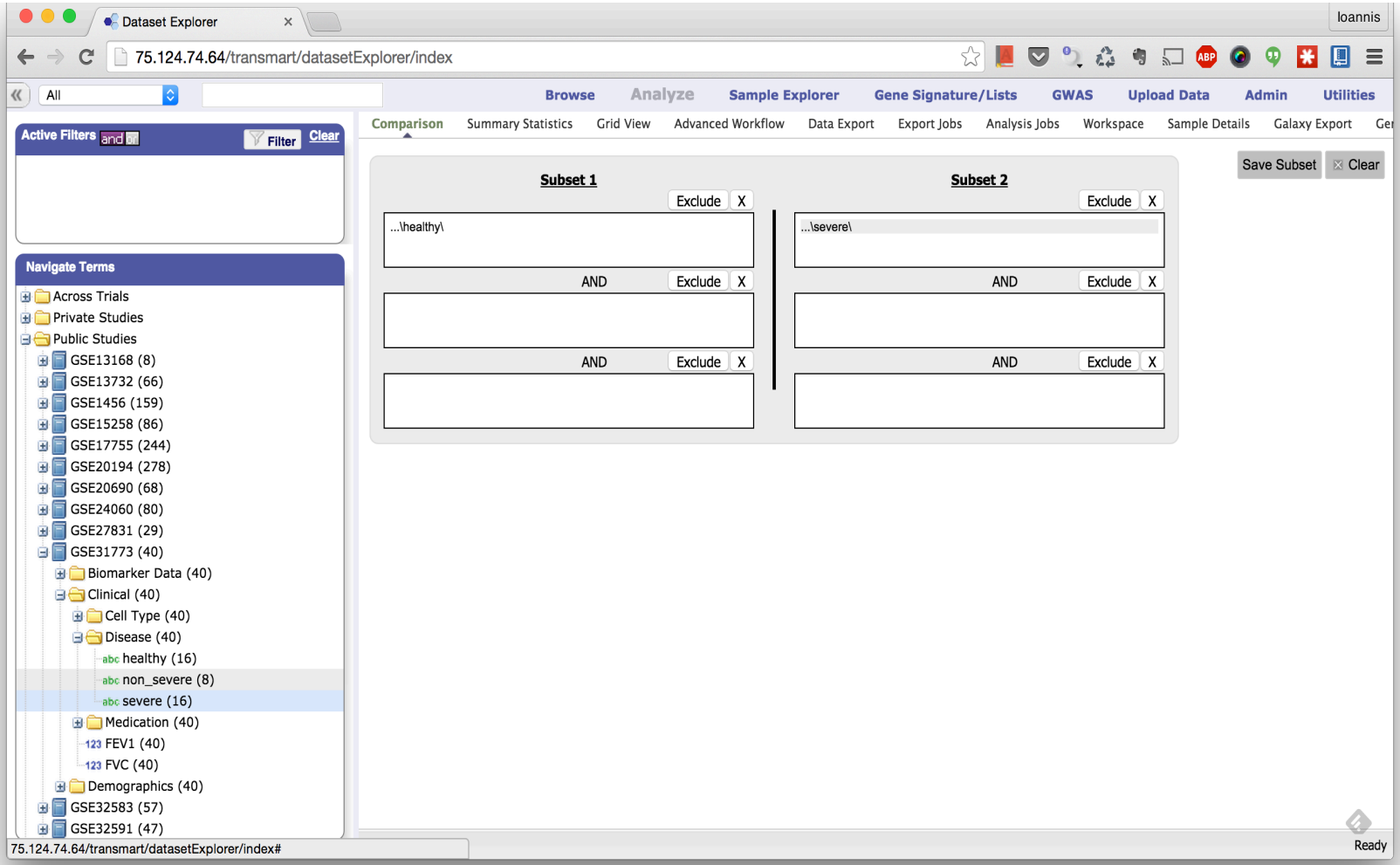
transmartfoundation.org

efpia

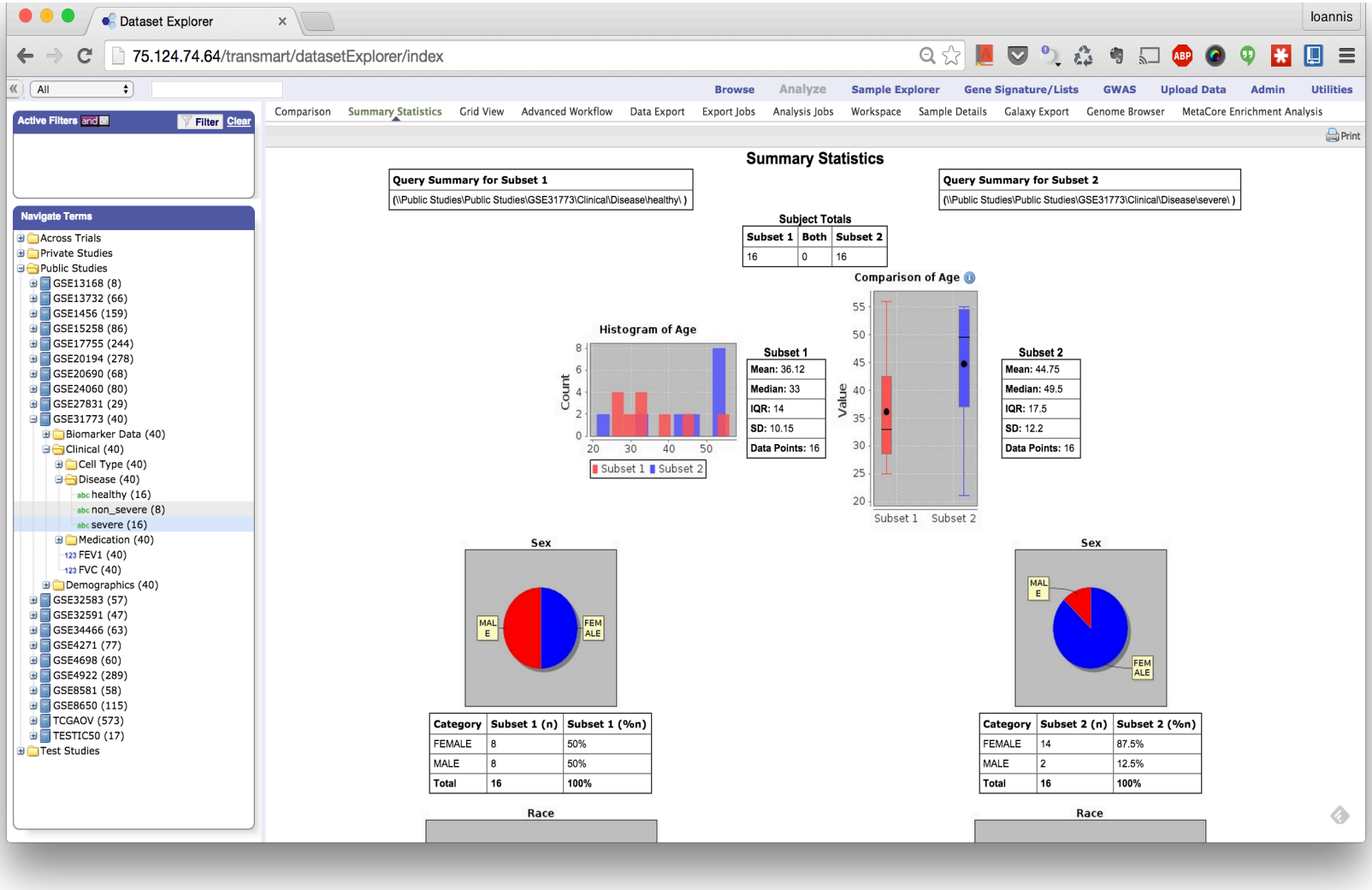
Browse and Search – Storing study information, metadata and associated raw files



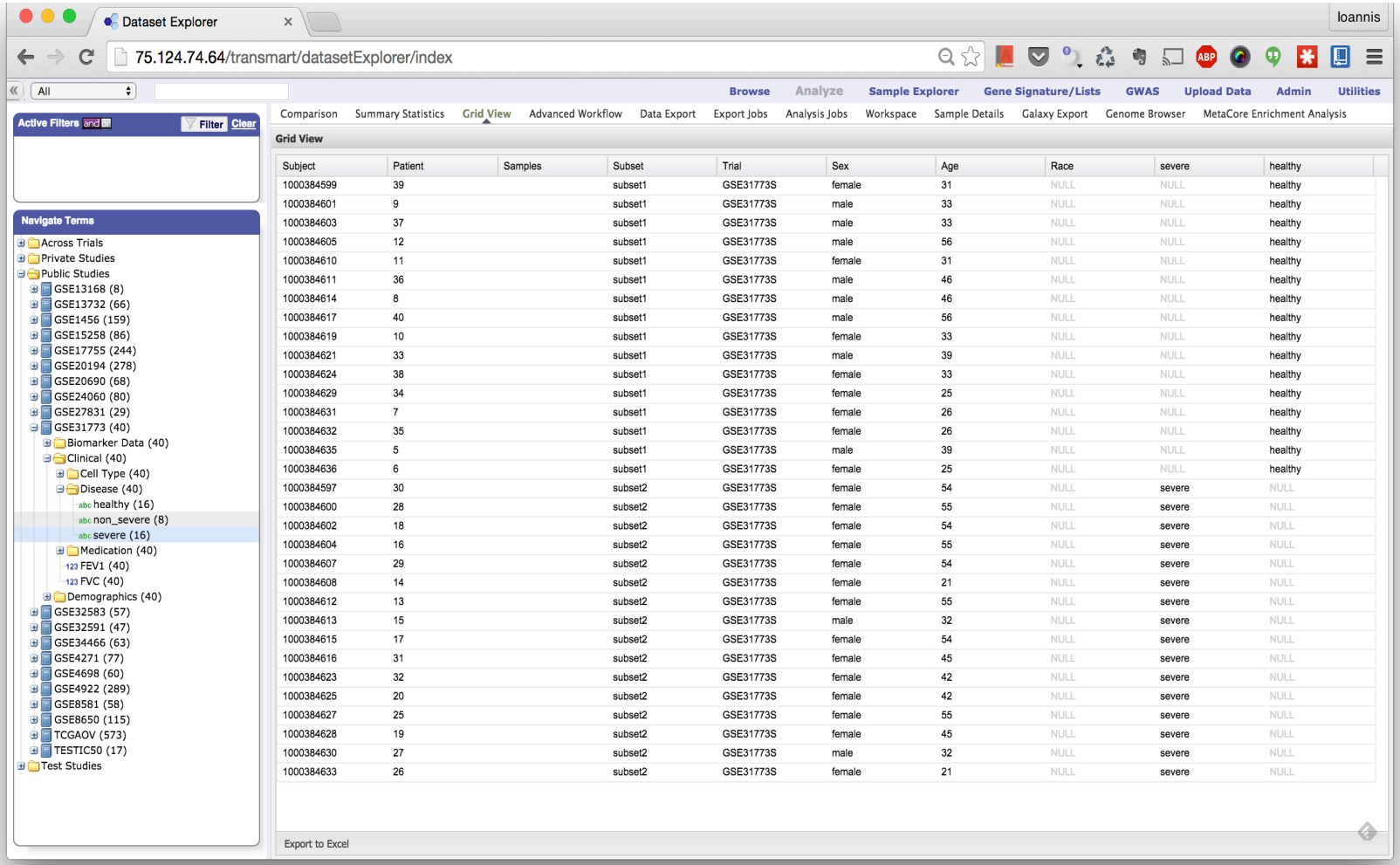
Analyze - Cohort Selection Tool



Analyze - Cohort Summary Statistics & clinical variable statistical analysis

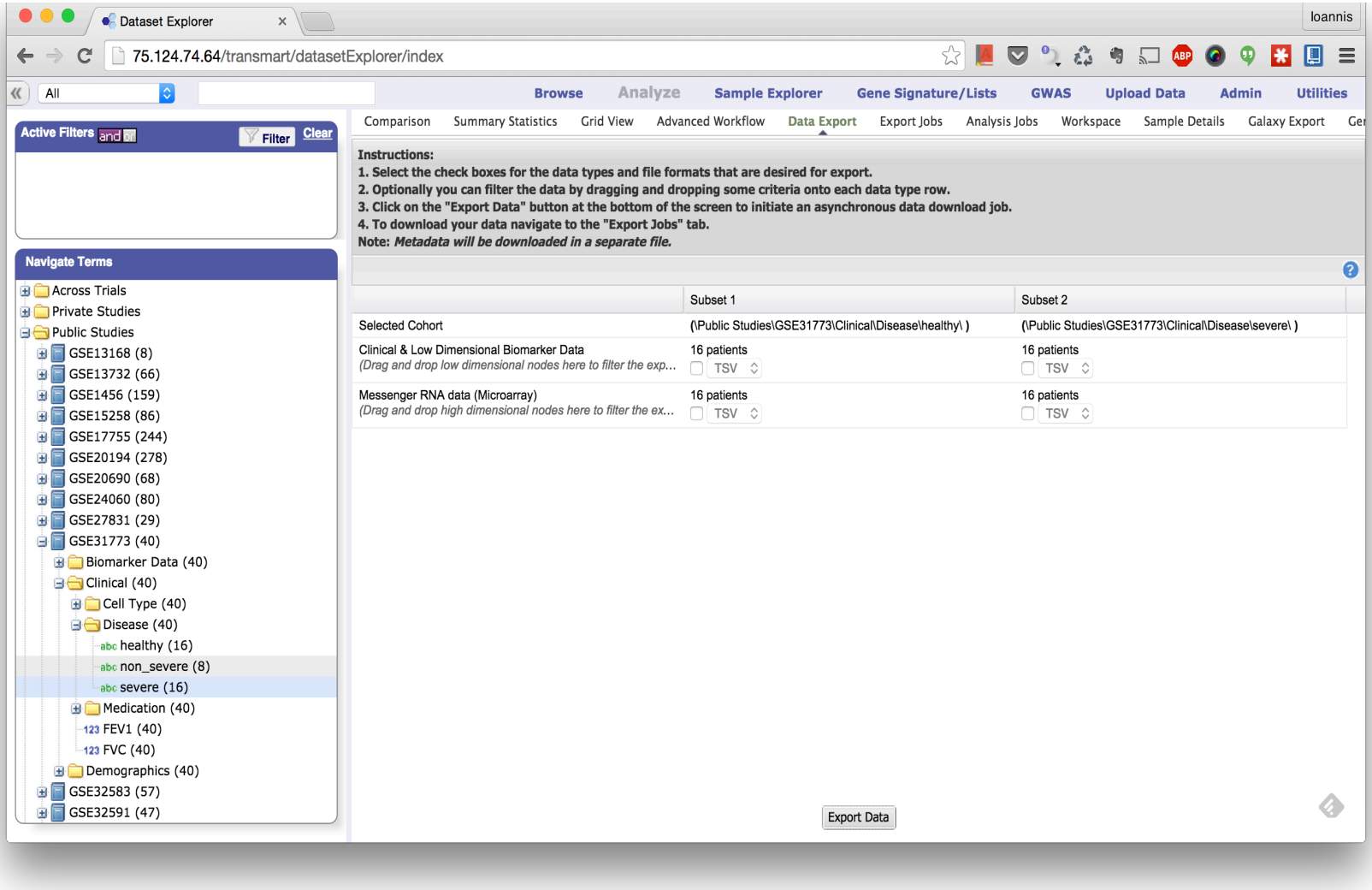


Analyze – Grid View for visual inspection and export of selected data

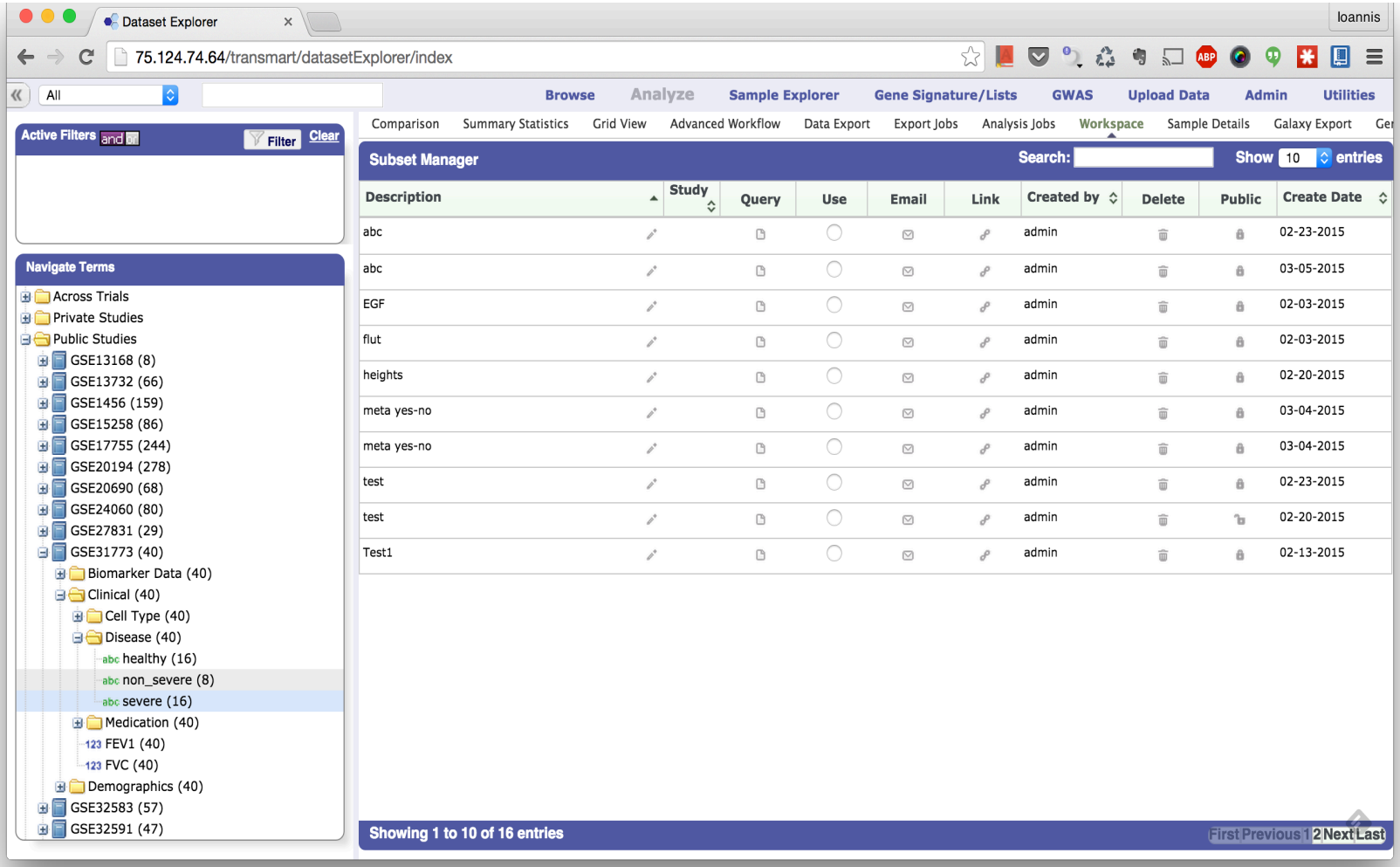


Subject	Patient	Samples	Subset	Trial	Sex	Age	Race	severe	healthy
1000384599	39		subset1	GSE31773S	female	31	NULL	NULL	healthy
1000384601	9		subset1	GSE31773S	male	33	NULL	NULL	healthy
1000384603	37		subset1	GSE31773S	male	33	NULL	NULL	healthy
1000384605	12		subset1	GSE31773S	male	56	NULL	NULL	healthy
1000384610	11		subset1	GSE31773S	female	31	NULL	NULL	healthy
1000384611	36		subset1	GSE31773S	male	46	NULL	NULL	healthy
1000384614	8		subset1	GSE31773S	male	46	NULL	NULL	healthy
1000384617	40		subset1	GSE31773S	male	56	NULL	NULL	healthy
1000384619	10		subset1	GSE31773S	female	33	NULL	NULL	healthy
1000384621	33		subset1	GSE31773S	male	39	NULL	NULL	healthy
1000384624	38		subset1	GSE31773S	female	33	NULL	NULL	healthy
1000384629	34		subset1	GSE31773S	female	25	NULL	NULL	healthy
1000384631	7		subset1	GSE31773S	female	26	NULL	NULL	healthy
1000384632	35		subset1	GSE31773S	female	26	NULL	NULL	healthy
1000384635	5		subset1	GSE31773S	male	39	NULL	NULL	healthy
1000384636	6		subset1	GSE31773S	female	25	NULL	NULL	healthy
1000384597	30		subset2	GSE31773S	female	54	NULL	severe	NULL
1000384600	28		subset2	GSE31773S	female	55	NULL	severe	NULL
1000384602	18		subset2	GSE31773S	female	54	NULL	severe	NULL
1000384604	16		subset2	GSE31773S	female	55	NULL	severe	NULL
1000384607	29		subset2	GSE31773S	female	54	NULL	severe	NULL
1000384608	14		subset2	GSE31773S	female	21	NULL	severe	NULL
1000384612	13		subset2	GSE31773S	female	55	NULL	severe	NULL
1000384613	15		subset2	GSE31773S	male	32	NULL	severe	NULL
1000384615	17		subset2	GSE31773S	female	54	NULL	severe	NULL
1000384616	31		subset2	GSE31773S	female	45	NULL	severe	NULL
1000384623	32		subset2	GSE31773S	female	42	NULL	severe	NULL
1000384625	20		subset2	GSE31773S	female	42	NULL	severe	NULL
1000384627	25		subset2	GSE31773S	female	55	NULL	severe	NULL
1000384628	19		subset2	GSE31773S	female	45	NULL	severe	NULL
1000384630	27		subset2	GSE31773S	male	32	NULL	severe	NULL
1000384633	26		subset2	GSE31773S	female	21	NULL	severe	NULL

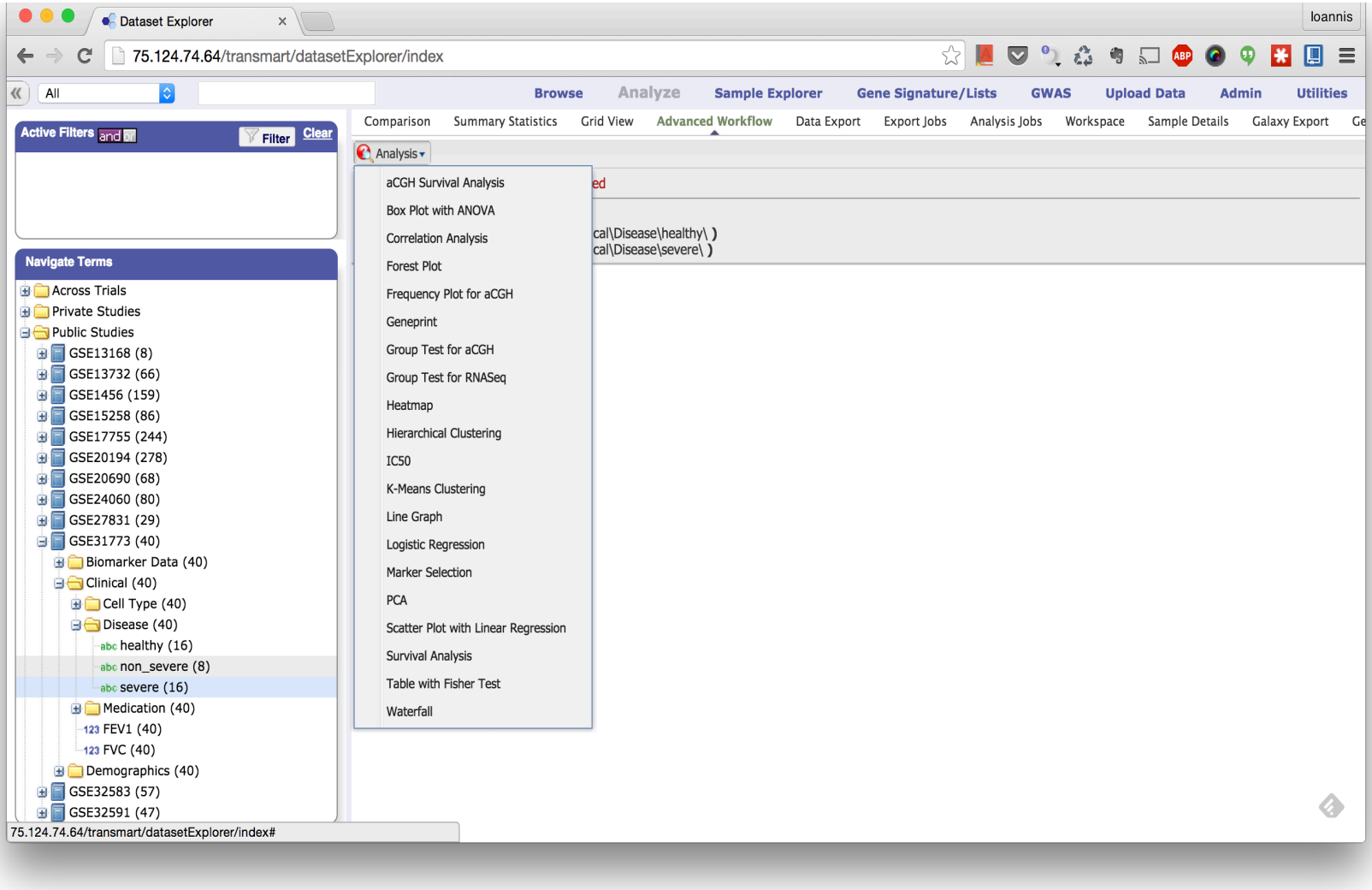
Analyze – Data Export for mass data (clinical and omics) export



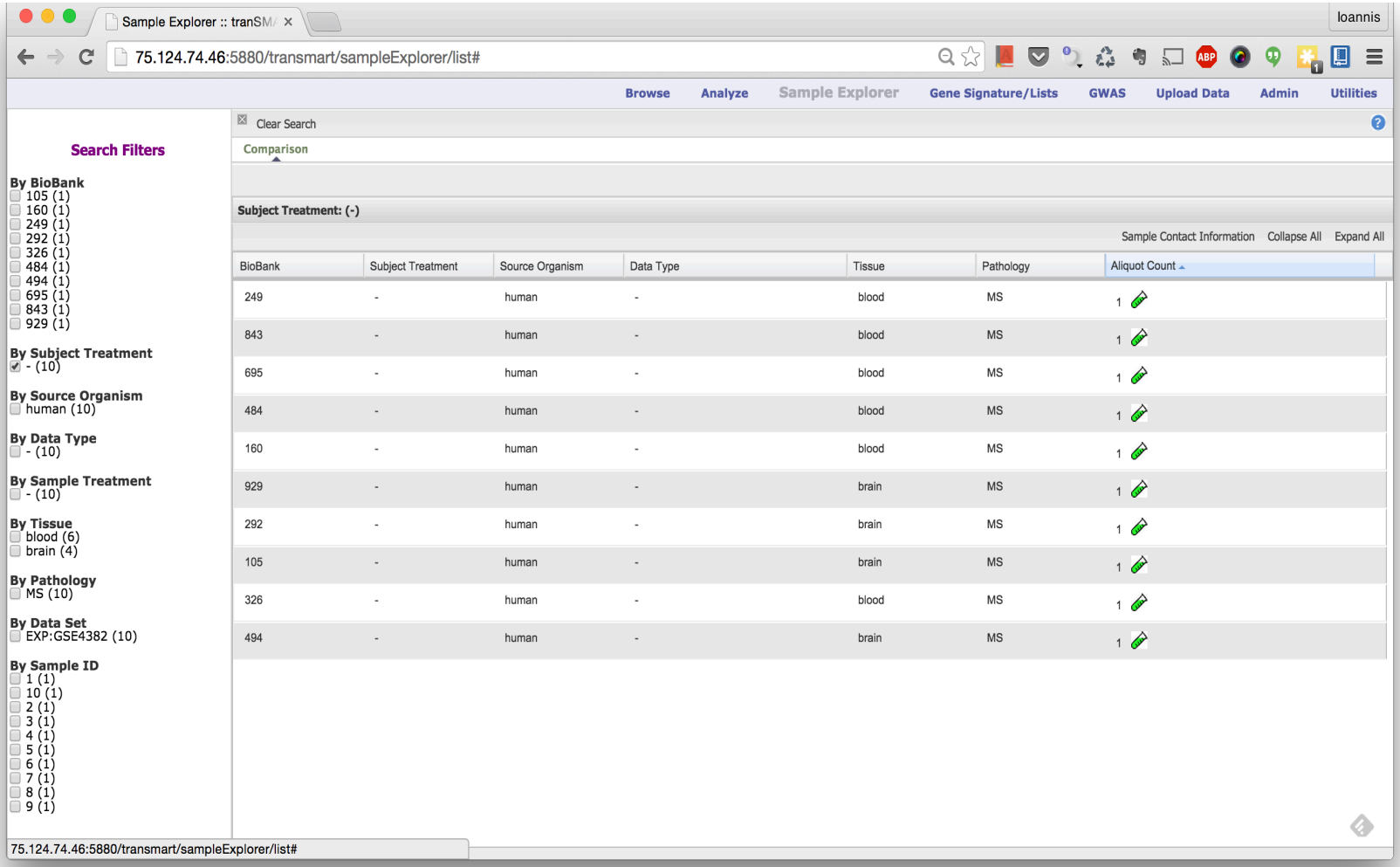
Analyze – Workspace for query saving and reuse



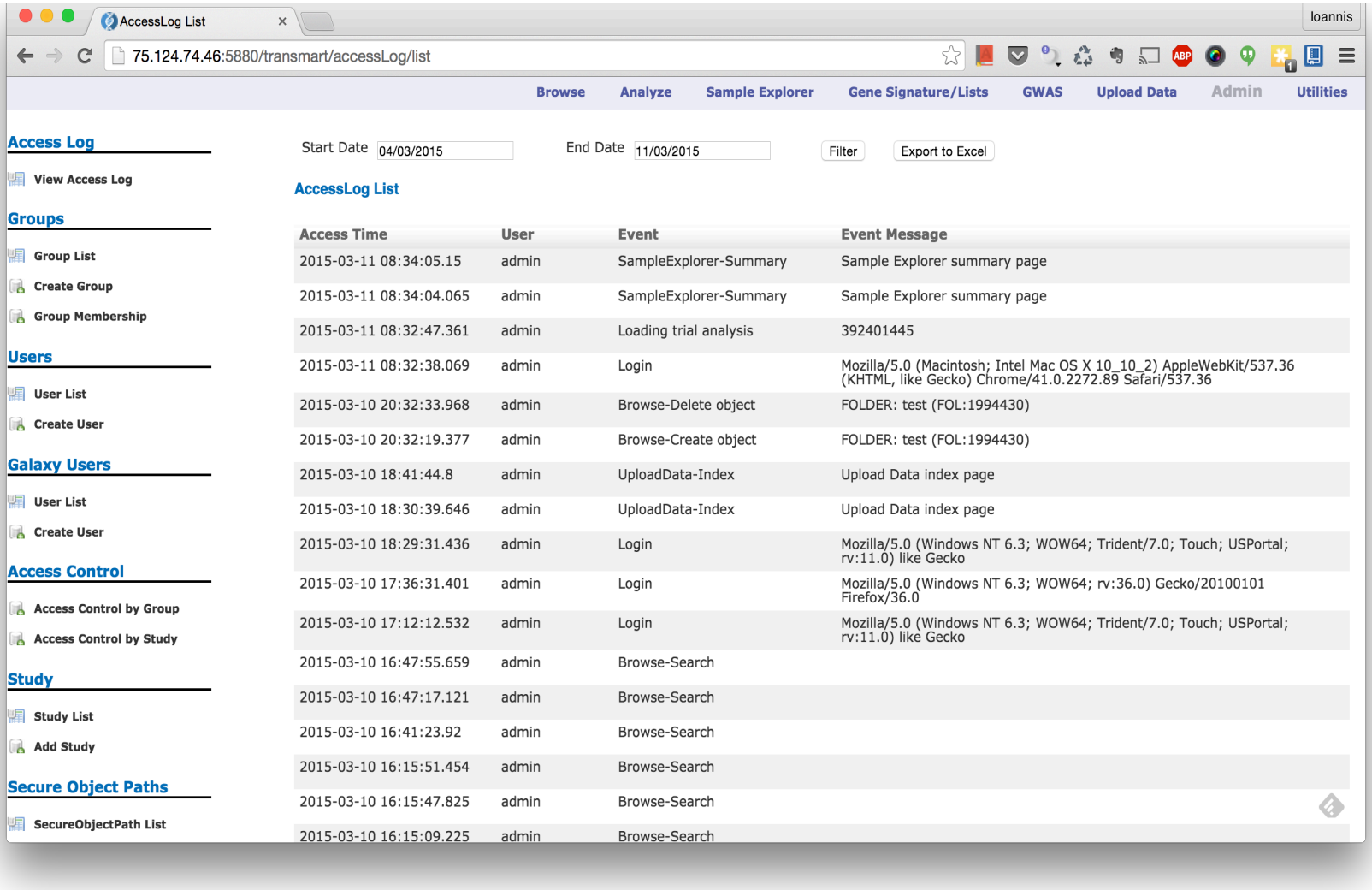
Analyze – Advanced workflow for omics and advanced analysis



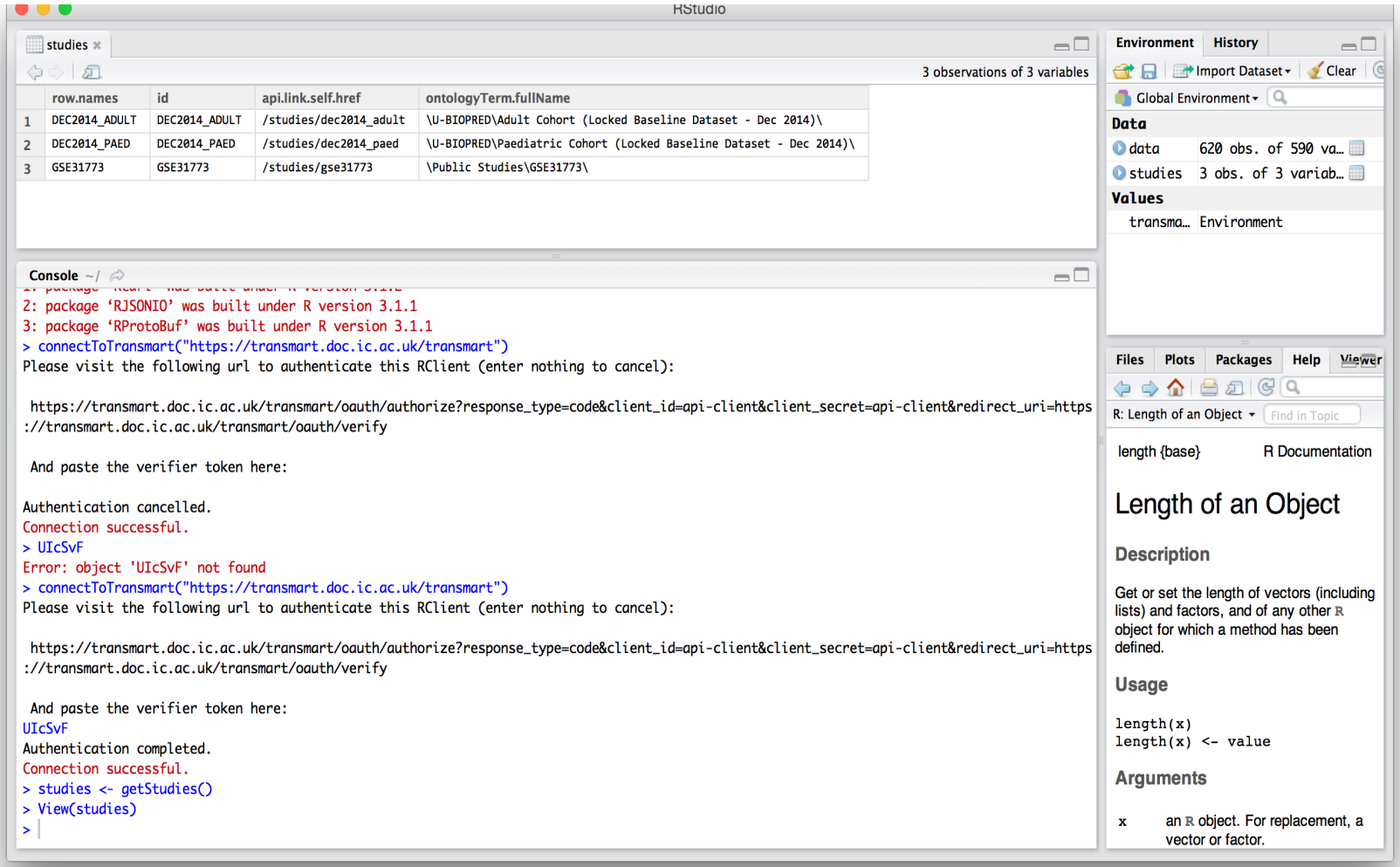
Sample Explorer – Biobank data visualiser



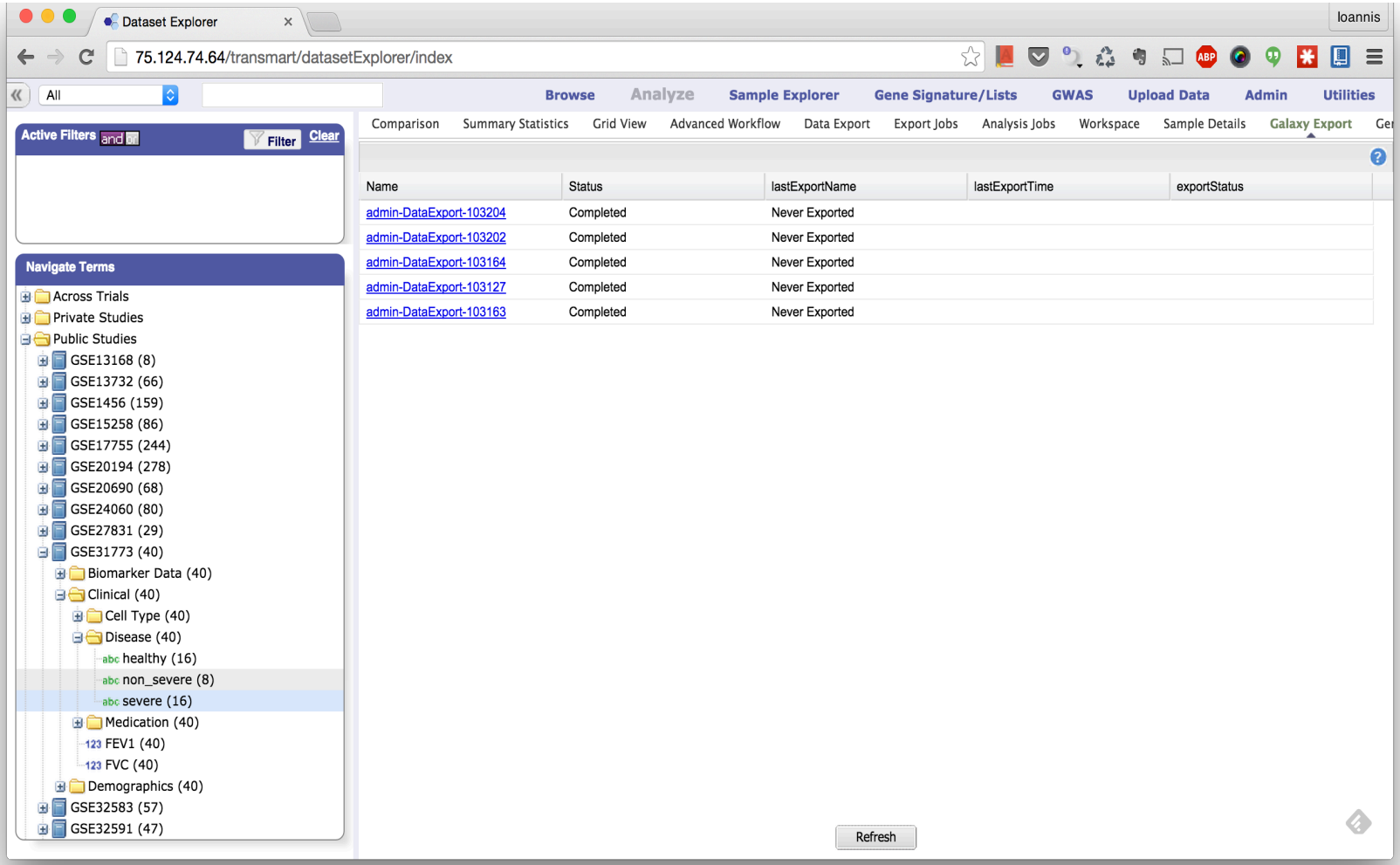
Admin Panel – Allowing granular access policy definition



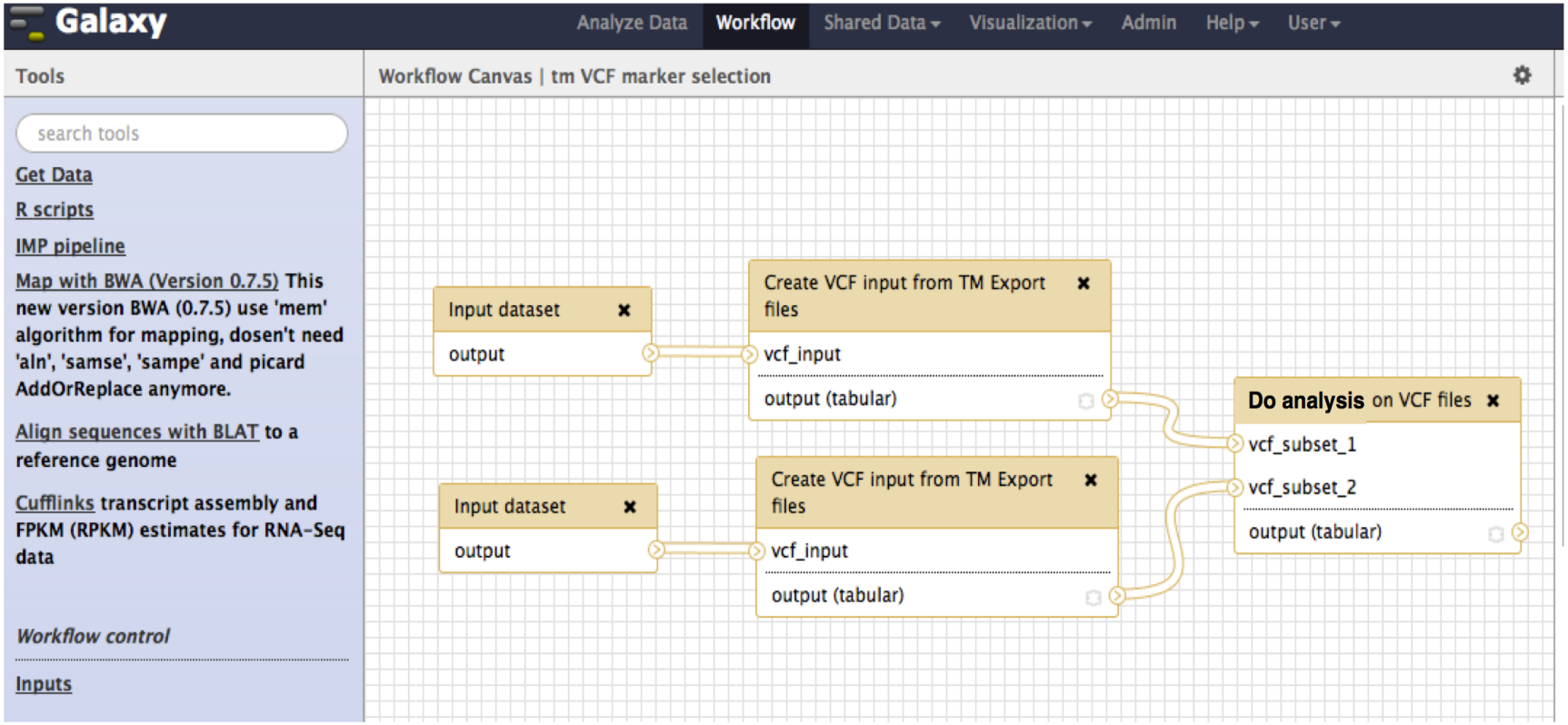
Rclient – Programmatic database querying



Galaxy Plugin – Connection to Galaxy and Advanced workflow development



Galaxy Plugin – Connection to Galaxy and Advanced workflow development



Genome Browser (Dalliance – HTML 5) – for omics data visualisatio