



HDR UK Text Analytics Workshop 12/03/2021

Generation Scotland Datasets

Archie Campbell
Chief Technical Officer
Generation Scotland
University of Edinburgh

genscot@ed.ac.uk

www.generationscotland.org

**A population & family based study of
genetic influences on disease & well being
24,000 participants in 7,000 families**



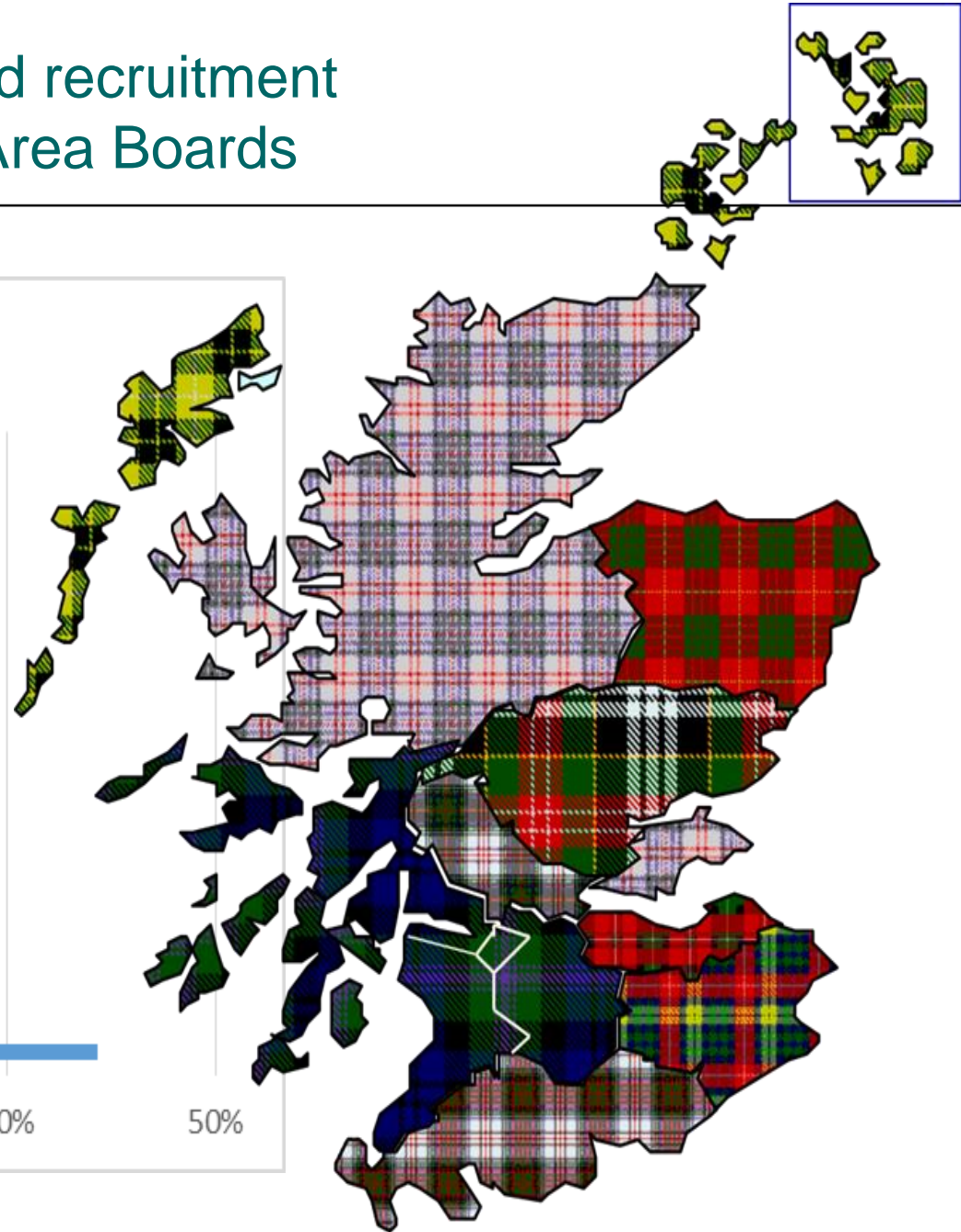
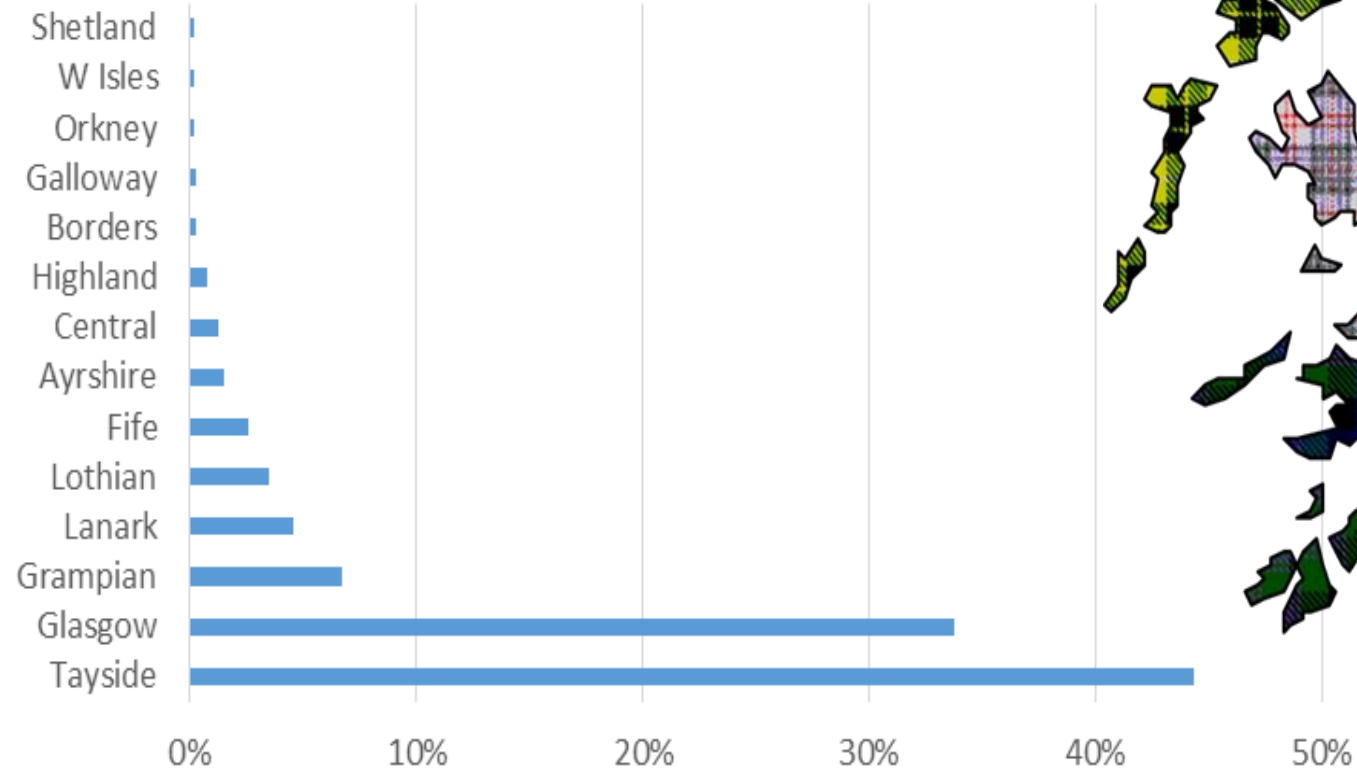
- ✓ Heart Disease
- ✓ Lung disease
- ✓ Diabetes
- ✓ Obesity



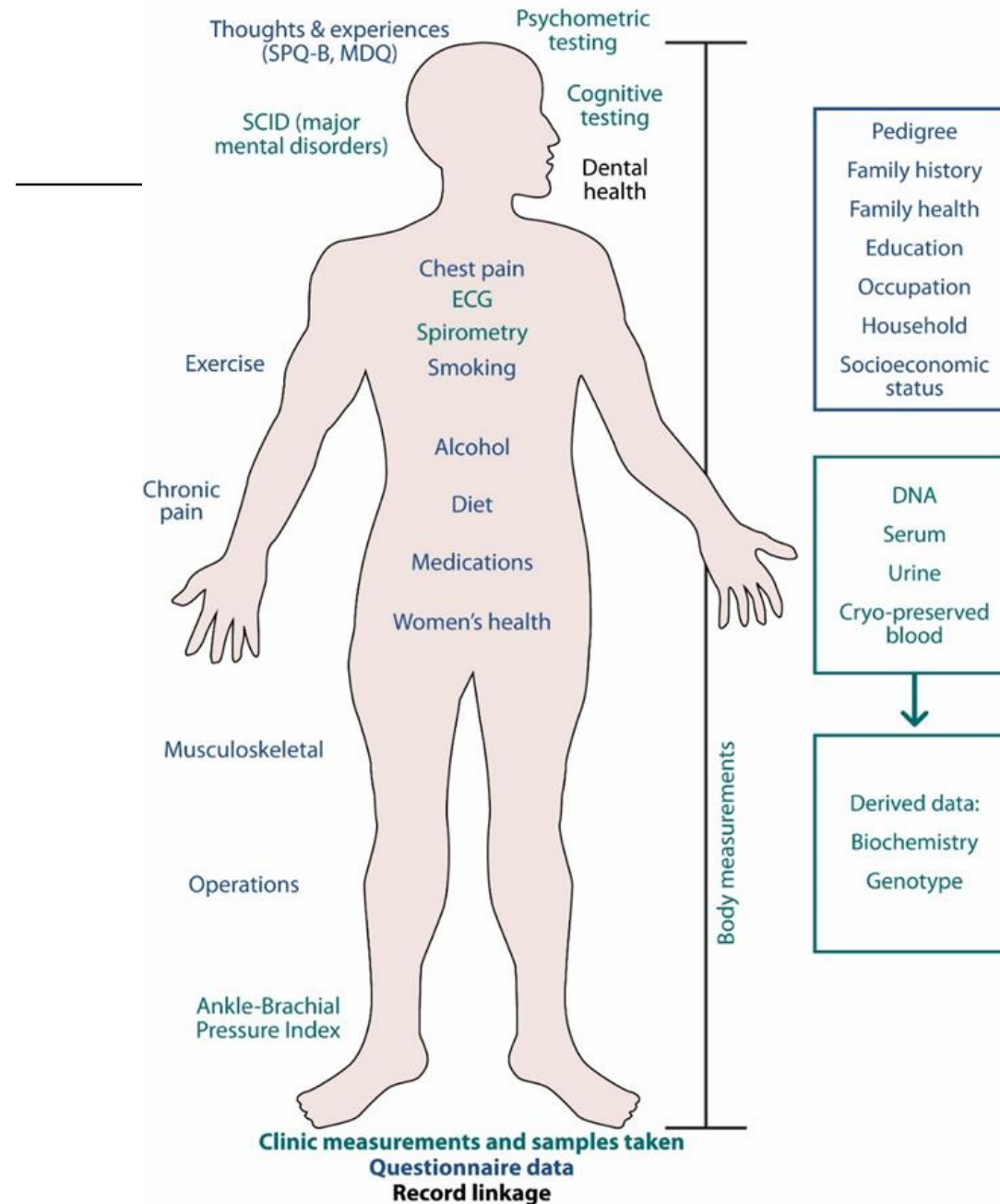
Depression
Cognition
Chronic pain

Generation Scotland recruitment NHS Scotland: 14 Area Boards

SFHS by Area



Phenotype and samples



2006-2011

**400 questions about
medical history, lifestyle &
habits**

150 clinical measures

Blood samples for DNA

Serum, plasma & urine

**Linked to NHS Scotland
health records**

Recontact

**Genetic and
epigenetic data**

Pedigree
Family history
Family health
Education
Occupation
Household
Socioeconomic
status

DNA
Serum
Urine
Cryo-preserved
blood

Derived data:
Biochemistry
Genotype

Linkage to other Datasets

CHI – Community Health Index

Administrative Data

- **NHS – SMR admissions**
- **NHS – PIS prescriptions**
- GROS – Mortality
- MIDAS Dental Data
- SCI Store lab results
- **GP records**
- A&E admissions
- NHS 24
- GP Out Of Hours
- ECOSS – virus tests
- PAX – scan images
- **Scan reports**

SMR - Scottish Morbidity Register

- SMR00 – Outpatient Attendance
- SMR01 – General/Acute Inpatient & Day Case
- SMR02 – Maternity Inpatient & Day Case
- SMR04 – Mental Health
- SMR06 – Cancer Registry
- SMR11 – Neonatal Inpatient
Scottish Birth Record (2002+)

Linkage Timeline

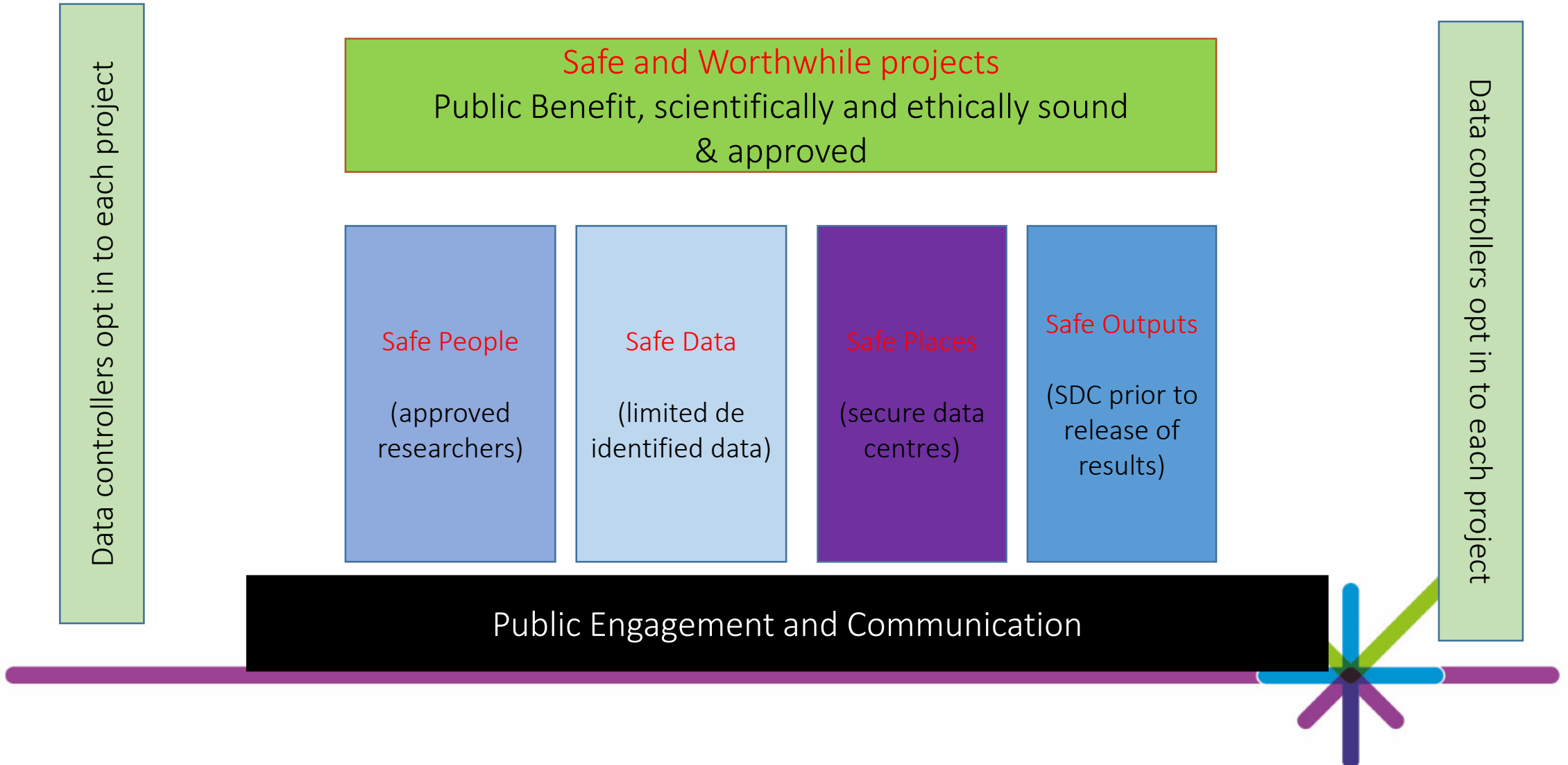
- 2012 – First SMR data release
- 2016 – Full SMR + **PIS**
- 2019 – PBPP program approval
- 2020 – **GP data**
Covid-19 data
Scan reports
- 202? – PAX

Provisioning unstructured data – opportunities and challenges

Jackie Caldwell / 12 March 2021

SECURING TRUST

Data Controllers and Public



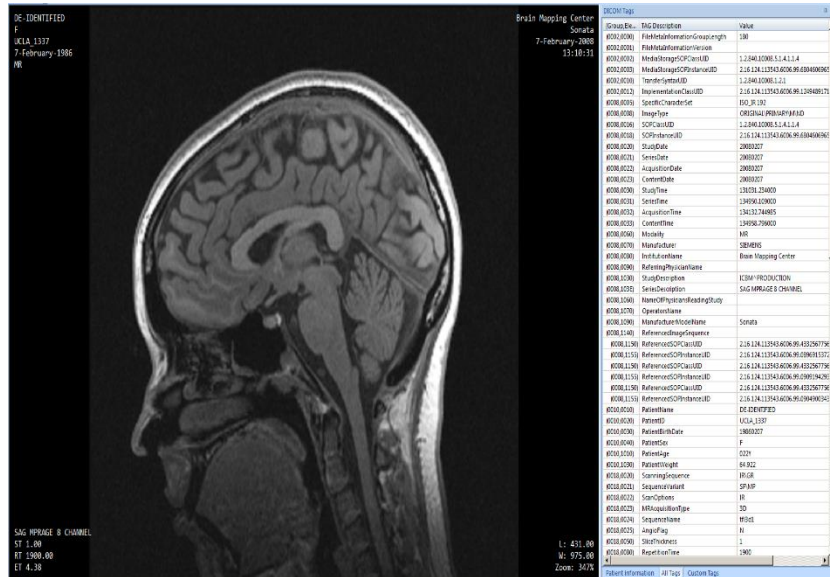
[illegible]

Exploiting the structure in unstructured data

Prescription Instruction	Dose Min	Dose Max	Dose Unit	Timing Min	Timing Max	Timing Unit	As Directed	As Required
ONE TABLET DAILY	1	1		1	1	day		
WEEKLY 1 TAB	1	1		1	1	week		
1 OR 2 CAPS 4 TIMES A DAY	1	2		4	4	day		
THREE 5ML SPOONFULS TWICE DAILY	15	15	ml	2	2	day		
1 SPRAY TO BE USED UP TO THREE TIMES A DAY	1	1	spray	0	3	day		
2 PUFFS PRN	2	2	puff					true
AS DIRECTED 2.5 ML	2.5	2.5	ml				true	



SCOTTISH MEDICAL IMAGING and PICTURES



- PICTURES - £4.5m grant from MRC
- Can be thought of as R&D for SMI
- One deliverable is improved cohort building capabilities



- Scotland has a national PACS system
- PBPP approval to create research ready images
- Currently deidentified MRI, CT and PET for 2010-2018
- Working on Xray deidentification, M/L VMs, GPUs etc etc
- Beta service launched



Challenges

- Handling the volume of data
 - – > 100 million prescriptions and rising each year
 - technical computing challenges of manipulating the > 2 PB of SMI data
- Maintenance
 - ensuring the deidentification and exploitation algorithms are maintained and remain accurate over time

