Sensitive Data Research in the Cloud

Gary Leeming

Chief Technology Officer, Connected Health Cities

Project S Solution Architect

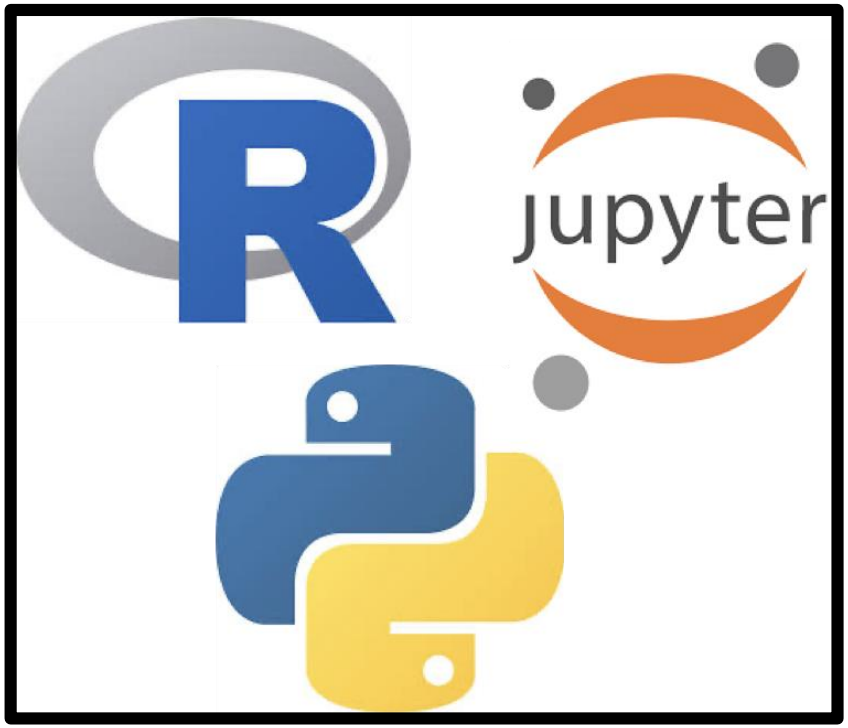
What is Restricted Data?

- Special category data under GDPR (Managing personally identifable data should be BAU?)
- Data defined as sensitive/secret/restricted by the data owner
- Definition usually risk-based
- UoM has 3 categories: Unrestricted, Restricted and Highly Restricted

MANCHESTER 1824

What is Research?

What is Restricted Data Research?



Research Data Management Lifecycle

Strictly host Claudia

Millions of workers to get extra holiday

Sex between 13-year-olds is

How it's cheaper to FLY to the US and

NHS lost track of 1.8m patient records in a year with sensitive information found in public bin and for sale on the internet

- The total is the equivalent of nearly 5,000 records going missing every day
- Error saw details of terminally ill patients were faxed to the wrong number
- Fines totalling £1million levied against NHS bodies in the last six months

By JACK DOYLE FOR THE DAILY MAIL

PUBLISHED: 00:21, 29 October 2012 | **UPDATED:** 08:45, 29 October 2012

NEWSPAPER OF THE YEAR

THE TIMES

Max 11C min 1C

Wednesday February 19 2014 | thetimes.co.uk | No 71124

Only 60p to subscribers £1.20

12 top trends from Times2 London Fashion Week

Ukraine erupts Nine people were killed in the capital, Kiev, yesterday when police fired on protesters as the President tried to hang on to power. World, pages 28, 29

NHS chiefs in climbdown over sharing patient data

Plan on hold after protests over confidentiality

Chris Smyth Health Correspondent

Plans to share the private medical data of NHS patients have been shelved for at least six months after health chiefs conceded that they had failed to explain the scheme properly and needed to restore public confidence.

In an embarrassing climbdown, NHS England has promised to ask doctors, patients and charities how best to explain the project, which will take information from GP records and link it with existing data from hospitals.

Tougher scrutiny of who is given access to the information has also been promised, after accusations that it would put patient confidentiality at risk. The health service is under pressure to undertake a national TV advertising campaign and to write to every patient individually.

NHS chiefs say that joining up patient data will allow doctors to track new diseases, assess new drugs and spot areas where the health service is failing. Even supporters of the "care.data" scheme say that the "cavalier" attempt to press ahead without building public support has damaged confidence in how the organisation looks after the material.

Katherine Murphy, chief executive of the Patients Association, added: "Patients need to be given confidence that information is in good hands in the NHS. Public trust and confidence has been tainted. This is what happens when you don't consult the right people."

The plan is backed by big health charities, but, in recent days, a procession of doctors have warned that it is confusing and has not been communicated properly. The NHS spent £1 million sending leaflets to every home in the country, but most patients said that they had not seen them.

Tim Kelsey, national director for patients and information at NHS England, said: "We have been told very clearly that patients need more time to learn about the benefits of sharing information and their right to object to their information being shared. That is why we are extending the public awareness campaign by an extra six months."

Data collection, which had been due to begin within weeks, will not start

Continued on page 2, col 3

How to tell if a tweet is telling the truth

Murad Ahmed Technology Reporter

It's the modern equivalent of *Call My Bluff*: what not to believe on Twitter. Now scientists are developing the ultimate "lie detector" for the digital age — a system that can instantly judge between truth and fiction in 140 characters or less.

Researchers across Europe are joining forces to analyse the veracity of statements that appear on social media in "real time". They hope to ensure that scurrilous rumours and false statements do not take hold.

The creators believe that the system, which has been named Pheme after the Greek mythological figure known for gossip and scandalous rumour, would have proved valuable during events such as the London riots in 2011. Panic was caused and police were diverted when Tweeters spread incorrect reports that animals had been released from London Zoo and that landmarks such as the London Eye and Selfridges had been set on fire.

The lead researcher on the project, Kalina Bontcheva, from the University of Sheffield's engineering department, said that the system would be able to test information quickly and track its provenance. This would enable governments, emergency services, health agencies, journalists and companies to respond to untruths.

"People do believe things they hear on the internet," she said. "In critical situations, you can instead show reliable information or alert the authorities before things get out of hand."

The makers said that the program could have warned Twitter users, such as Sally Bercow, the wife of the Commons Speaker, and the comedian Alan Davies, who were among those who spread rumours that linked Lord McAlpine to child sex abuse. Both were among a number of prominent people who agreed financial settlements with the late peer as a result of the incorrect claims.

Pheme will classify online rumours into four types: speculation, such as whether interest rates might rise; controversy — for example, the fierce debate over the MMR vaccine; misinformation, where something untrue is spread unwittingly; and

Continued on page 2, col 5

IN THE NEWS

Clegg faces rebellion
Infighting has broken out among the Liberal Democrats over Nick Clegg's plans to ensure that the party's 2015 election manifesto is coalition proof. News, page 2

Recovery takes hold
Britain is enjoying a recovery of slowly rising prices and strong growth after inflation fell below 2 per cent last month for the first time in more than four years. News, page 4

Butterflies soar
Butterflies prospered during last year's dry summer, with twice as many recorded on average at the same spots as during the wet summer of 2012. News, pages 18, 19

Arms plan for Syria
The US Secretary of State is secretly re-examining a plan to arm Syria's rebels that was put forward 15 months ago by the retired CIA chief General David Petraeus. World, page 29

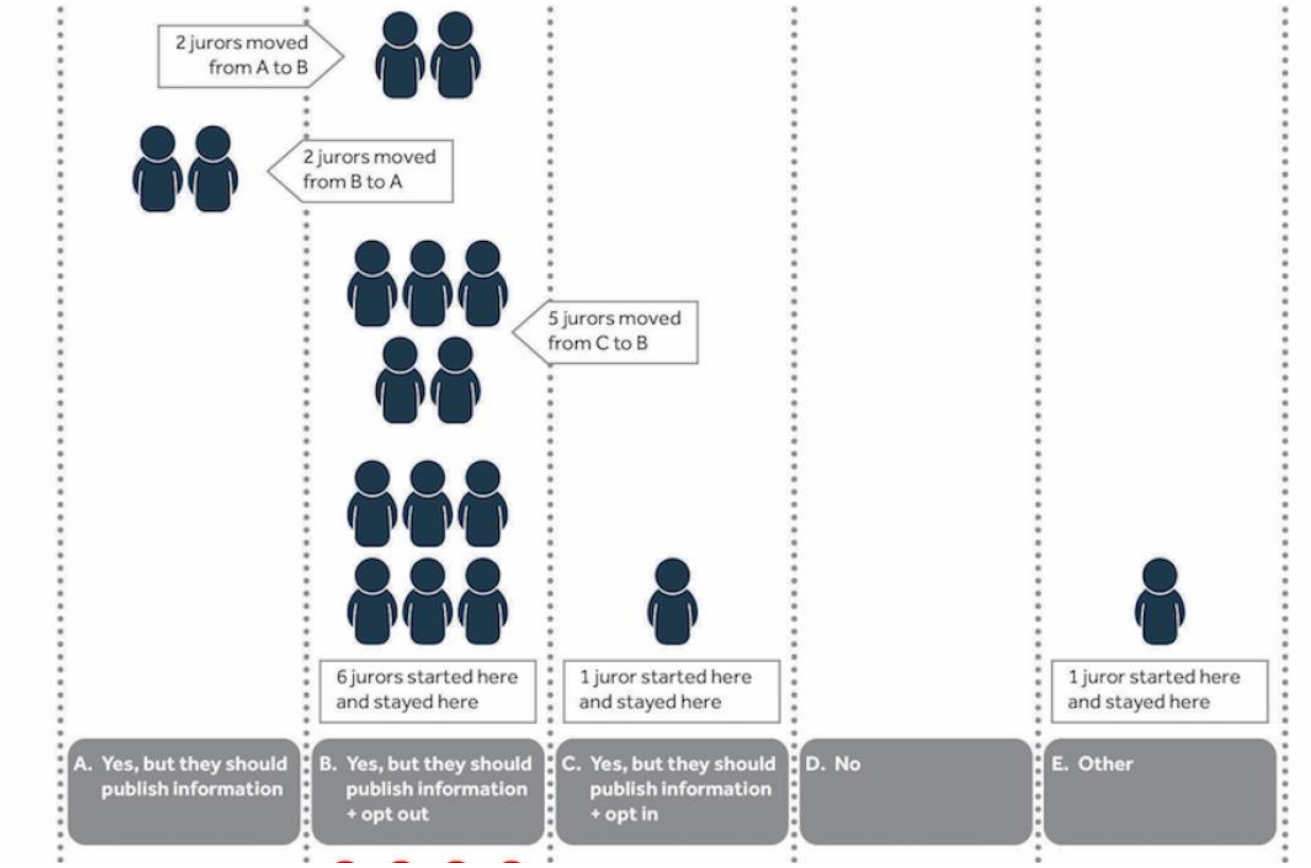
City on the brink
Manchester City lost 2-0 at home to Barcelona in the Champions League, leaving their hopes of reaching the quarter-finals hanging by a thread. Sport, pages 60–64

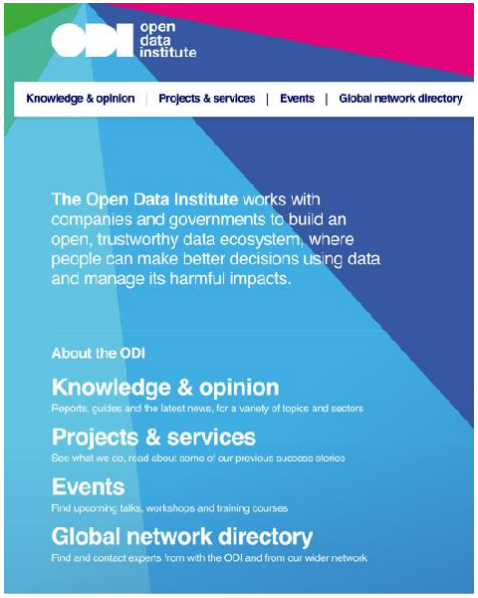
Our work with Citizens Juries

- **Citizens Jury**: Comprehensive public engagement process that allows decision makers to hear thoughtful input from an informed microcosm of the public
- Two juries (18 jurors each) met separately over 3 days
- **Jury charge**: "Should the NHS be allowed to create anonymised copies of patient records for secondary use?"

Tully et al., J Med Internet Res. 2018;20(3):e112.

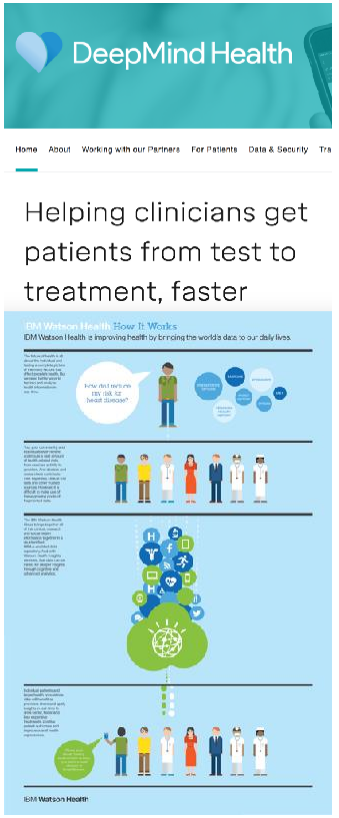
What the jurors said




open data institute
Knowledge & opinion | Projects & services | Events | Global network directory
The Open Data Institute works with companies and governments to build an open, trustworthy data ecosystem, where people can make better decisions using data and manage its harmful impacts.
About the ODI
Knowledge & opinion
Reports, guides and the latest news, for a variety of topics and sectors
Projects & services
See what we do, read about some of our previous success stories
Events
Find upcoming talks, workshops and training courses
Global network directory
Find and contact experts from with the ODI and from our wider network

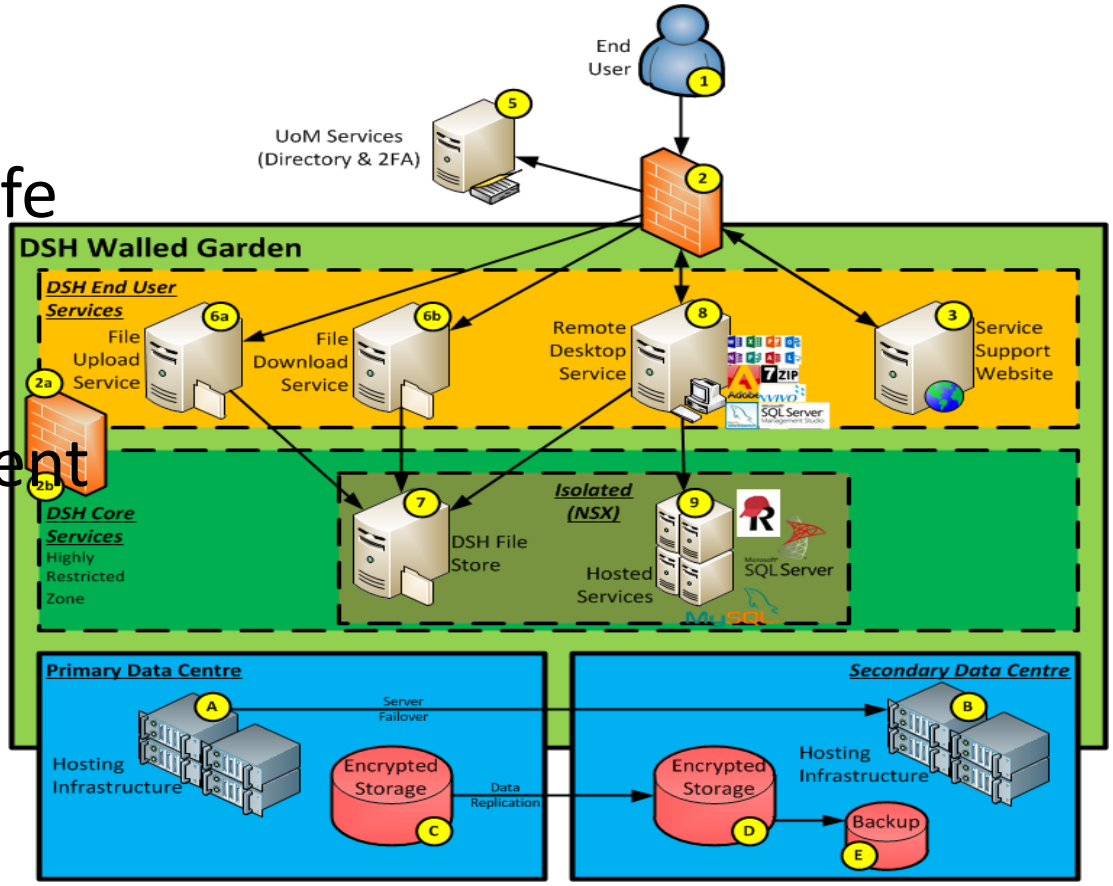
NHS

?


DeepMind Health
Home
About
Working with our Partners
For Patients
Data & Security
Helping clinicians get patients from test to treatment, faster
IBM Watson Health How It Works
IBM Watson Health is improving health by bringing the world's data to our daily lives.
IBM Watson Health

Current UoM Solutions

- Research IT Data Safe Haven
- CHI Trustworthy Research Environment



Project S

RESEARCH LIFECYCLE PROJECT

DEVELOP NEW SERVICE TO EXTEND/ENHANCE EXISTING CAPABILITY

PROVIDE SIMPLIFIED ACCESS AND INFORMATION FOR RESEARCHERS

Standards and Legislation

- ISO27001 (ISO27017, ISO27018?)
- NCSC - CyberEssentials (+)
- UK Gov 2016 14 Principles of Cloud
- GDPR
- Research ethics
- NHS Digital - IG Toolkit, Cloud guidance
- Data classification
- HIPAA

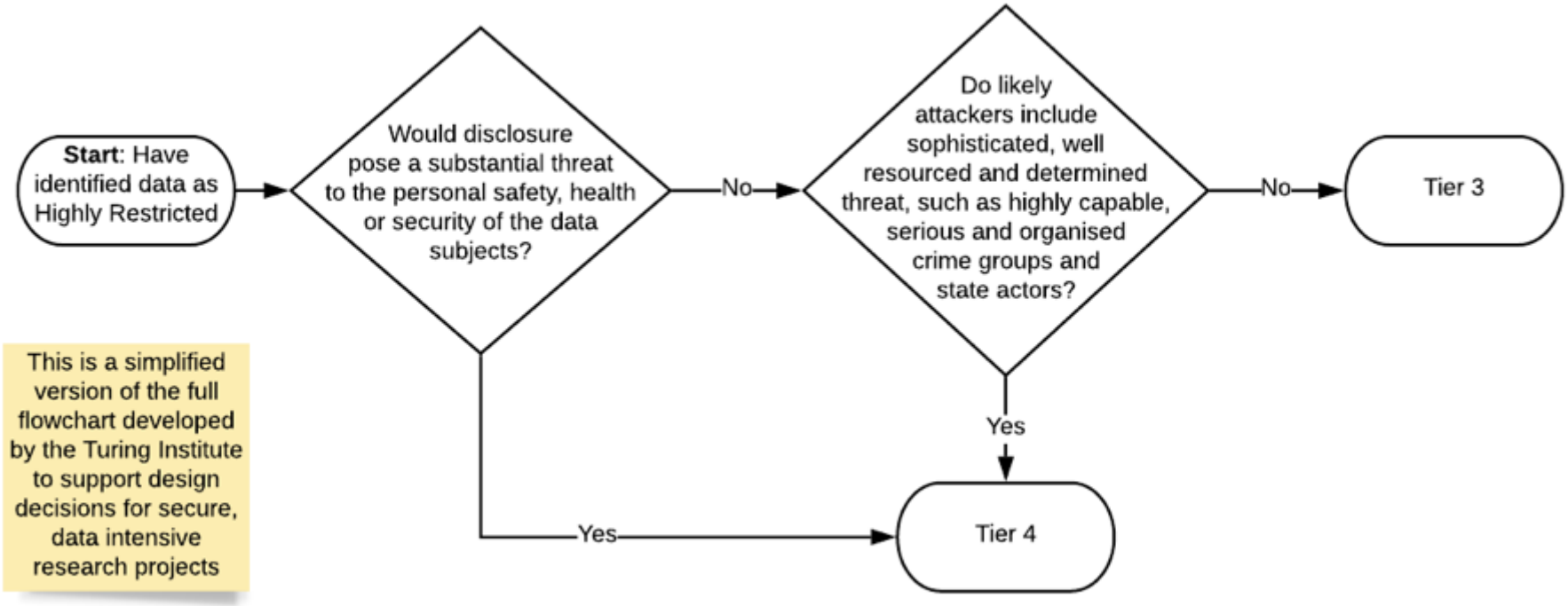
Data Challenges

- Classification of Types - Challenge of selecting appropriate classification (and avoiding over-/under-classification
- Define limitations of use - VM, physically secure, derived data, research object extraction
- Risks - Balance of risk vs business opportunity & need, value for researchers, protection University reputation, maintaining competitiveness (HPC access/data access)

Data Re-use and Discovery

- FAIR
  - Findable
  - Accessible
  - Interoperable
  - Reusable

Risk Management vs Classification



Arenas, Diego, et al. "Design choices for productive, secure, data-intensive research at scale in the cloud." *arXiv preprint arXiv:1908.08737* (2019).

Turing Institute Proposal

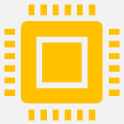
Turing Classification	University Classification	Risk (Reputation, legal, commercial, political)	Examples
Tier 0	Unrestricted	No risk if accessed by non-authorised actor	Public dataset, published paper
Tier 1	Unrestricted	Low risk if accessed by non-authorised actor	Research output intended for publication, non-personal research data
Tier 2	Restricted	Medium risk if accessed by non-authorised actor.	CPRD data extract, low-risk commercial in confidence data, low-risk IP
Tier 3	Highly Restricted	High risk if accessed by non-authorised actor, low-medium risk of attack	Detailed but anonymised hospital data, politically sensitive data, personal data where low risk of harm to the data subject
Tier 4	Highly Restricted	High risk if accessed by non-authorised actor, high risk of attack	Highly sensitive data, e.g. nuclear or pharmaceutical industry, personal data where high risk of harm to the data subject, e.g. refugee data.

Current activity

- Safepod!
- Business Analysts completed interviews across all faculties - 50+ interviews
- Questionnaire completed online - 83 respondents
- Threat modelling activity
- Ongoing meetings with potential suppliers
- Turing framework
- Board scope meeting - SWOT
- Proof of Concepts
- Business Case for December 2019 Board in development

Project S Key Features

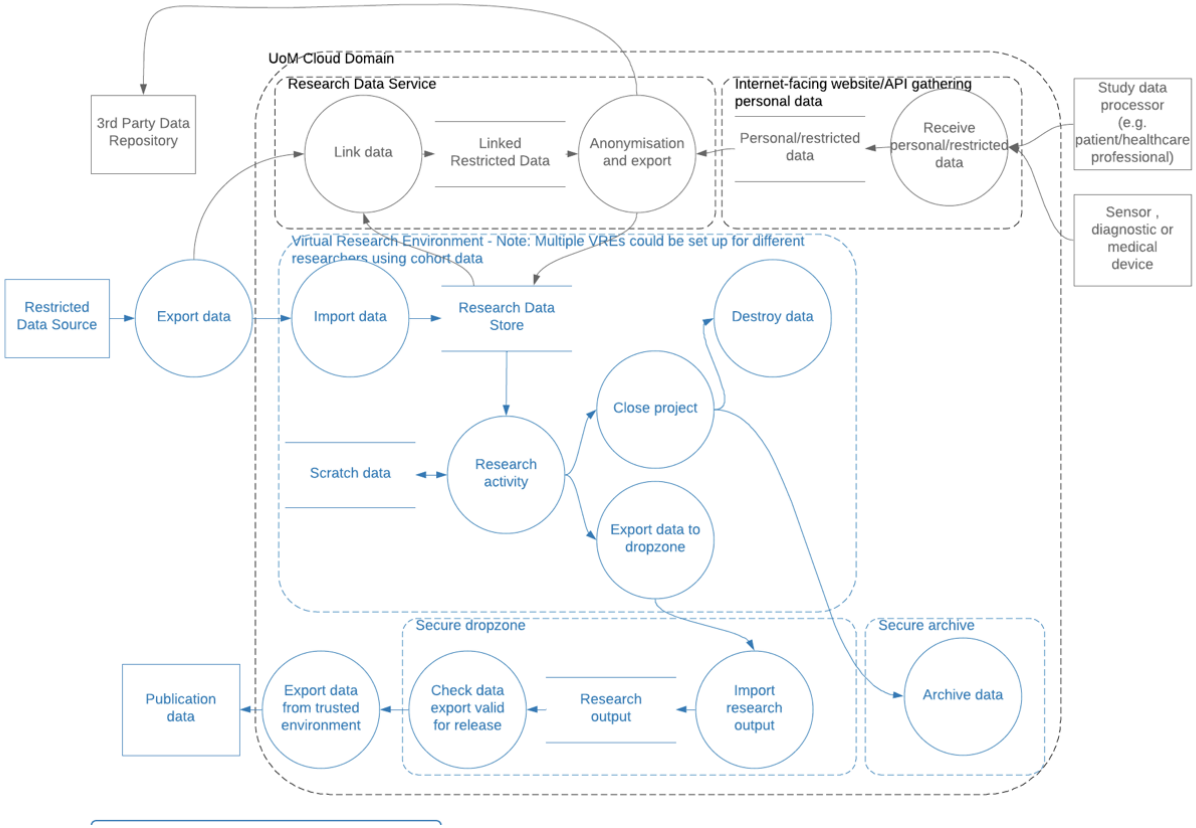
A management platform to enable Research IT to easily manage, audit and develop the platform for research projects



Secure Virtual Research Environments using template scripts to enable solutions to be quickly set up with appropriate controls and monitoring. These environments can be made available to researchers who are working on data with a variety of tools and options, from a basic Jupyter notebook to a virtual machine with access to enhanced compute, such as GPUs.

Secure Virtual Service Environments, also predominately scripted, to support the flow of data from external sources, such as public facing websites and devices, and manage the cohort dataset through the provisioning of secure VREs.

Highly Restricted Data Management Platform



Benefits

- Transparency of costs
- More consistent controls
- Better compliance and visibility of risk
- Updates and management of software
- Access to variety of compute and storage
- Collaboration opportunities