Protecting Sensitive and Personal Research Data in the Cloud

Gary Leeming, University of Manchester

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?

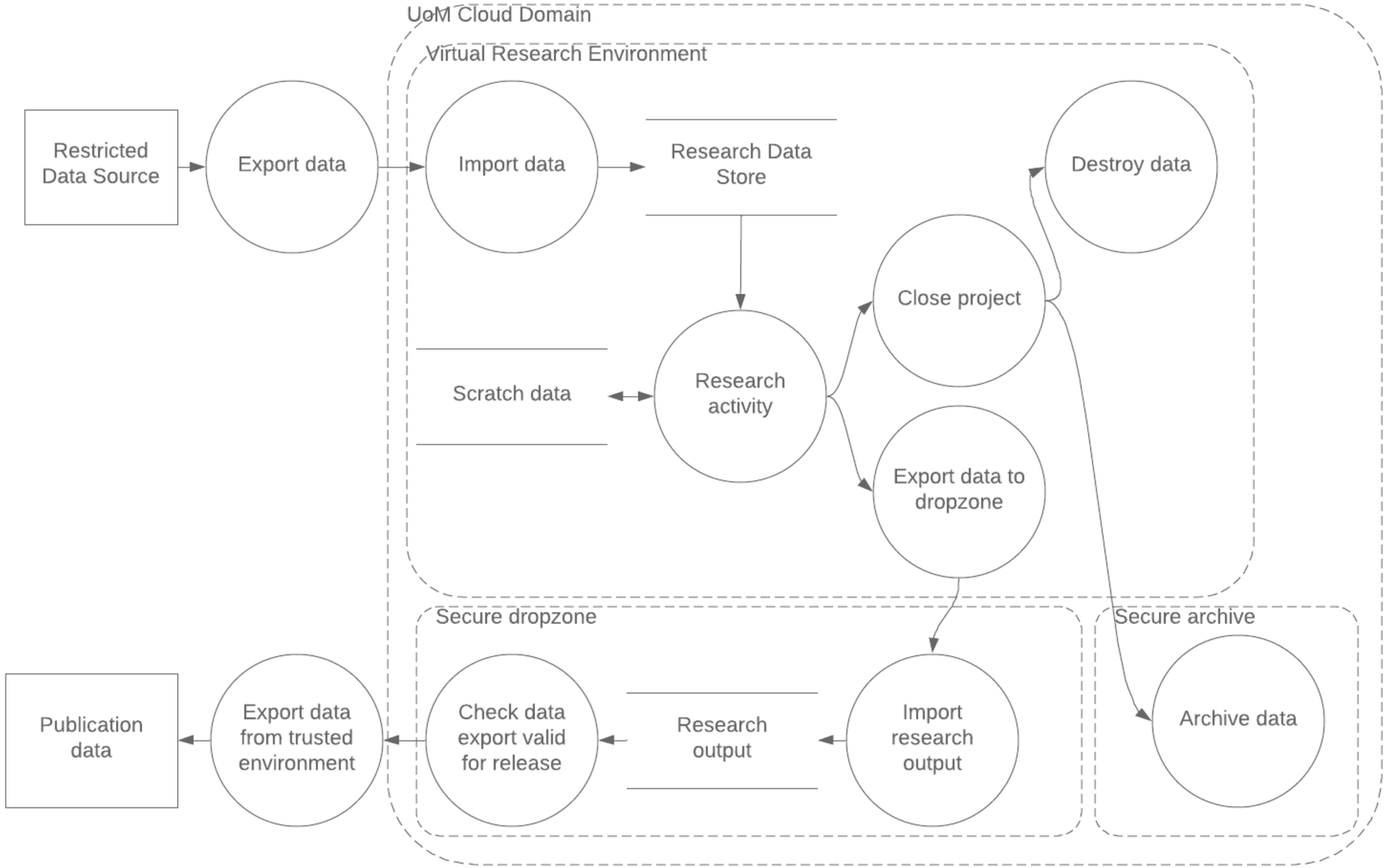
MANCHESTER 1824

What is Restricted Data Research?

docker

MANCHESTER 1824

Restricted Data VRE

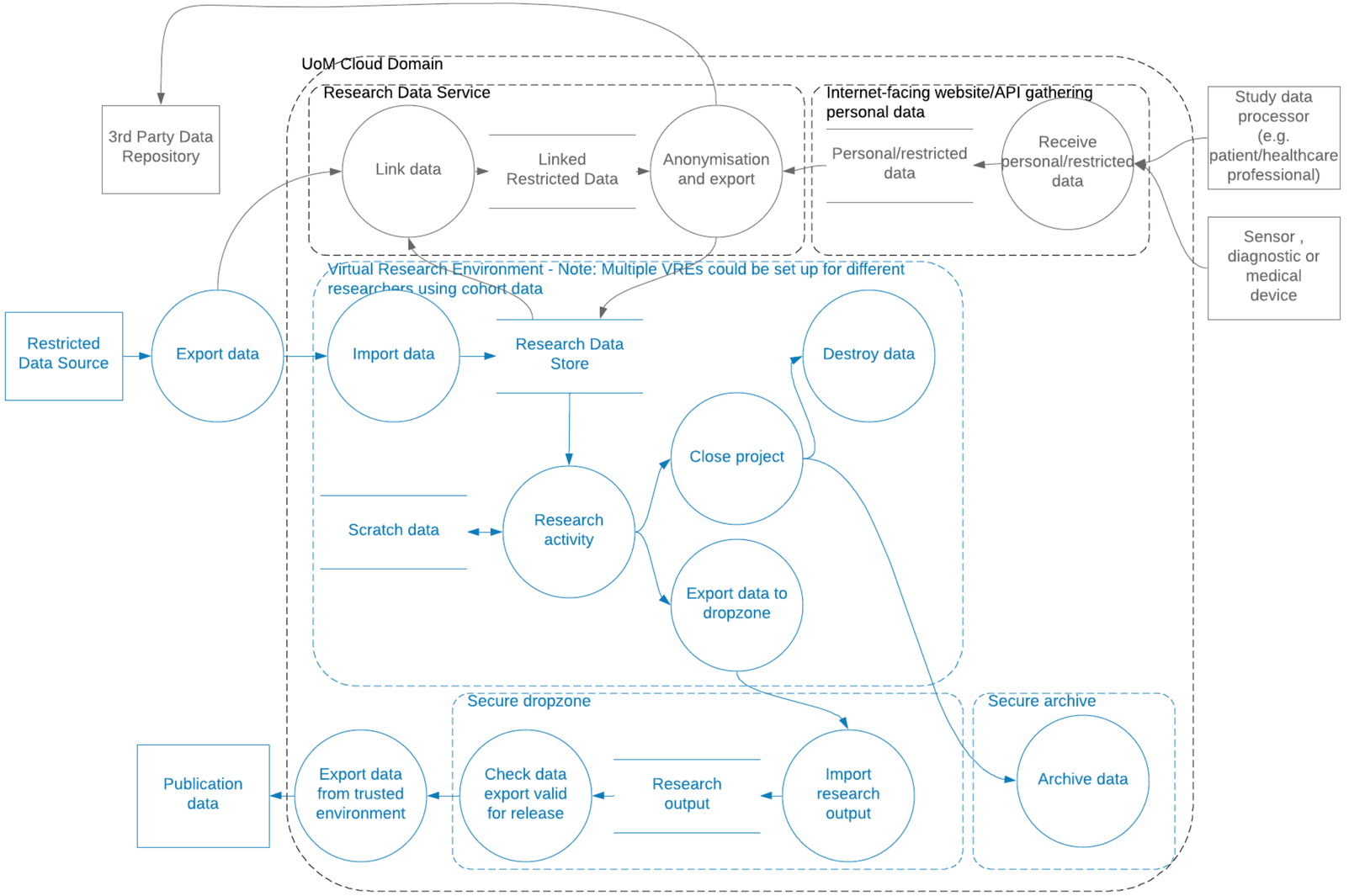


More Research Requirements

- Sensitive data from aircraft black boxes and airports received regularly for analysis
- Drug registry data collected every six months
- Geodata linked to medical records to be accessed by external researchers
- Good Clinical Practice

MANCHESTER 1824

Restricted Data Virtual Service Environment (VSE)



More requirements

- Can I work from home?
- Tell me what to put in my data management plan to get the grant
- [3rd Party Organisations] over-classifies their data

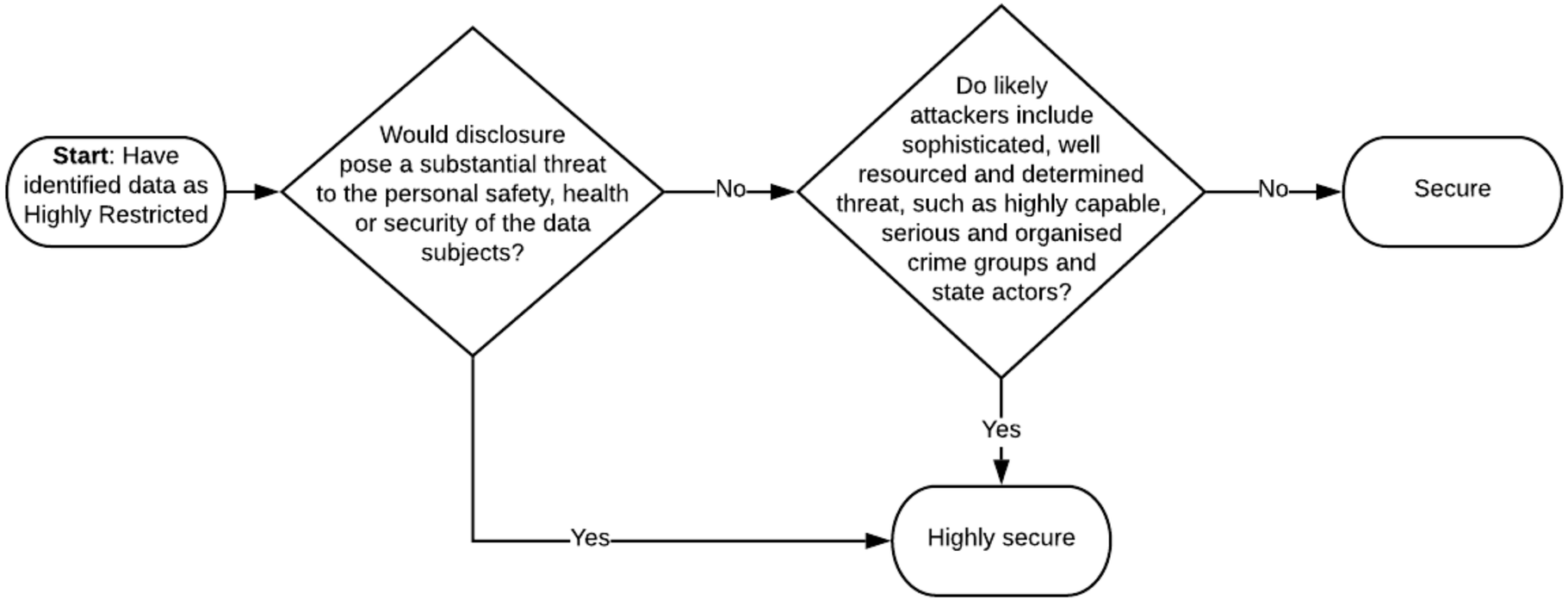
MANCHESTER 1824

Managing Data Flows

- Security is about process not technology
- Different parts of the process can be at different levels of risk
- Need to ensure that solution is appropriate to the risk at that point and can move between process requirements, e.g. Anonymisation

MANCHESTER 1824

Risk-based classification



MANCHESTER 1824

Turing 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.

Arenas, D., Atkins, J., Austin, C., Beavan, D., Egea, A. C., Carlysle-Davies, S., ... & Forrest, O. (2019). Design choices for productive, secure, data-intensive research at scale in the cloud. *arXiv preprint arXiv:1908.08737*.

MANCHESTER 1824

Threat Actors

Error by user	Malicious accounts person	External unskilled hacker
Malicious ex-user	Error by accounts person	Student hacker
Prof or senior staff member who "just wants to get it done"	Malicious hosting company worker	Knowledgeable external attacker
User who has been told by prof to "just get it done"	Error by hosting company worker	Malicious software developer
User who wants to try new plug in/ software to "just get it done"	Organised crime	Error by software developer
Malicious network team member	Nation state (UK, US, China, Russia, Iran)	Malicious package manager
Error by network team member	Competitor (for research or for students)	Error by package manager
Malicious first-line support	Commercial espionage	Error by retention policy definition
Error by first-line support	Malicious ex-network team member	Malicious time service maintainer
Social engineer	Malicious IT manager	Malicious DNS maintainer

Australian Threat Analysis

- **Basic premise: researchers are honest-but-sloppy**
  - ignorant of IT security
  - reliant on institutional IT security
  - driven by convenience
- **Secure remote analysis facilities for research are designed to protect against**
  - innocent acts-of-omission by researchers
  - acts-of-carelessness by researchers
  - malicious acts by non-users (i.e. external hackers)
- **But not necessarily malicious acts-of-commission by researchers**
  - e.g. filming the screen as they scroll through data

UNSW Sidney https://ipdln.org/sites/default/files/2018ConcurrentSessions/210/Sept-12-1515-C12-Room210-Jorm.pdf

MANCHESTER 1824

Data Management and Risk as a Lifecycle

- Development
- Testing
- Audit
- Risk identification
- Vulnerability Management

MANCHESTER 1824

Information and Security Standards and Guidance

- ISO27001
- NHS Digital Data Security and Protection Toolkit
- CyberEssentials (plus)
- NCSC 14 Principles
- Five Safes (UK Data Service)
- STRIDE
- CVSS
- Mitre Att&ck

MANCHESTER 1824

Cyber Essentials

- Firewall
- Secure Configuration
- Access Control
- Malware Protection
- Patch Management

NCSC 14 principles

1. Data in transit protection
2. Asset protection and resilience
3. Separation between users
4. Governance framework
5. Operational security
6. Personnel security
7. Secure development
8. Supply chain security
9. Secure user management
10. Identity and authentication
11. External interface protection
12. Secure service administration
13. Audit information for users
14. Secure use of the service

Key Features

- Management platform
- Secure VRE templates
- Secure Virtual Service Environment (VSE) templates

MANCHESTER 1824

Services

- Configuration of environments
- Deployment and testing
- Key management - Encryption keys, API keys, data identifiers
- IDAM - User and role management
- Software/VM Repo for approved images
- System health - patching, load
- Security Monitoring & Vulnerability Management
- Network configuration - Firewalls and subnets, no public internet, encryption in transit
- Disaster Recovery

MANCHESTER 1824

Challenges

- Usual cloud challenges - Supplier management, 3rd party resellers, cost/contract management, etc.
- Identity and roles
- DevOps model of services
- Ingress of software & scripts
- Research governance and finance process integration
- Fixed regions
- Serverless
- Making it easy to use….

MANCHESTER 1824

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

MANCHESTER 1824

“Think I have finally worked out what [the Highly Restricted Data platform] is. It's the Ginger Rogers to Research IT's Fred Astaire --- it has to do everything that the standard systems and staff can do, but backwards and in heels…”

–Anonymous