

AI, Data Governance & Ethics Committees

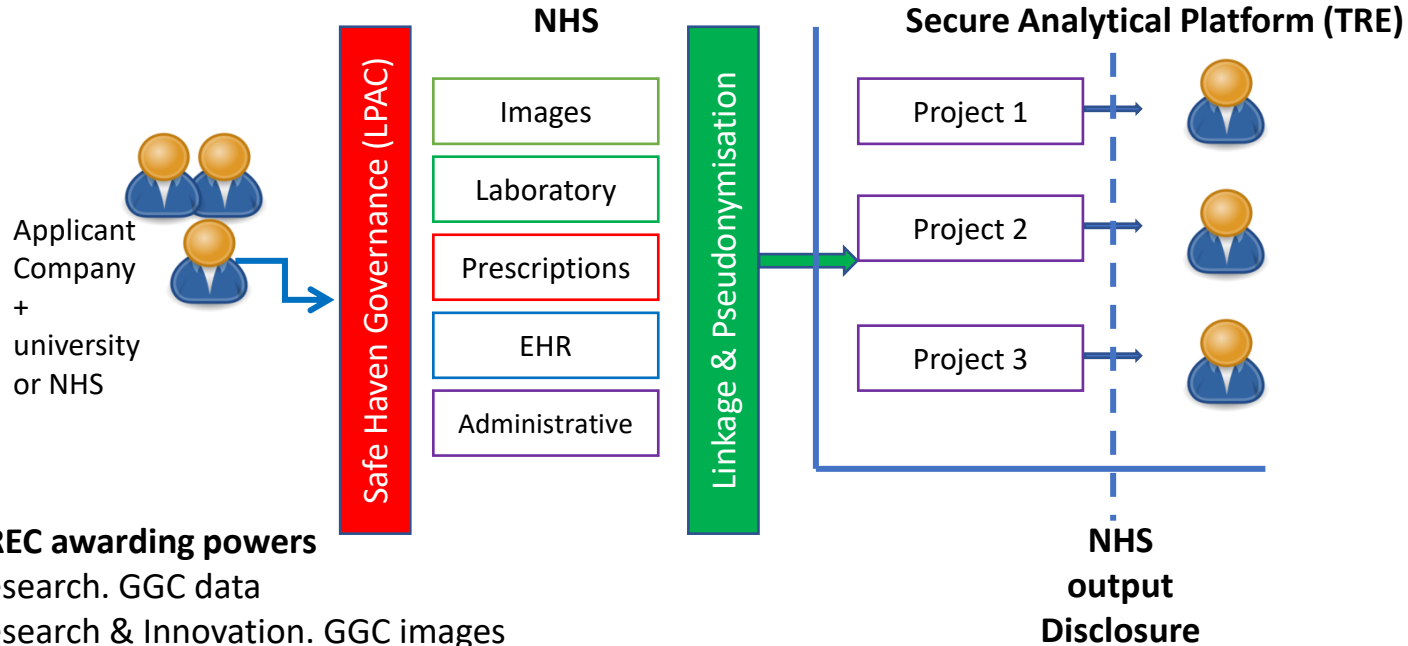
Reflections from our experience

Dr Roma Armstrong

NHSGGC R&I Innovation Lead



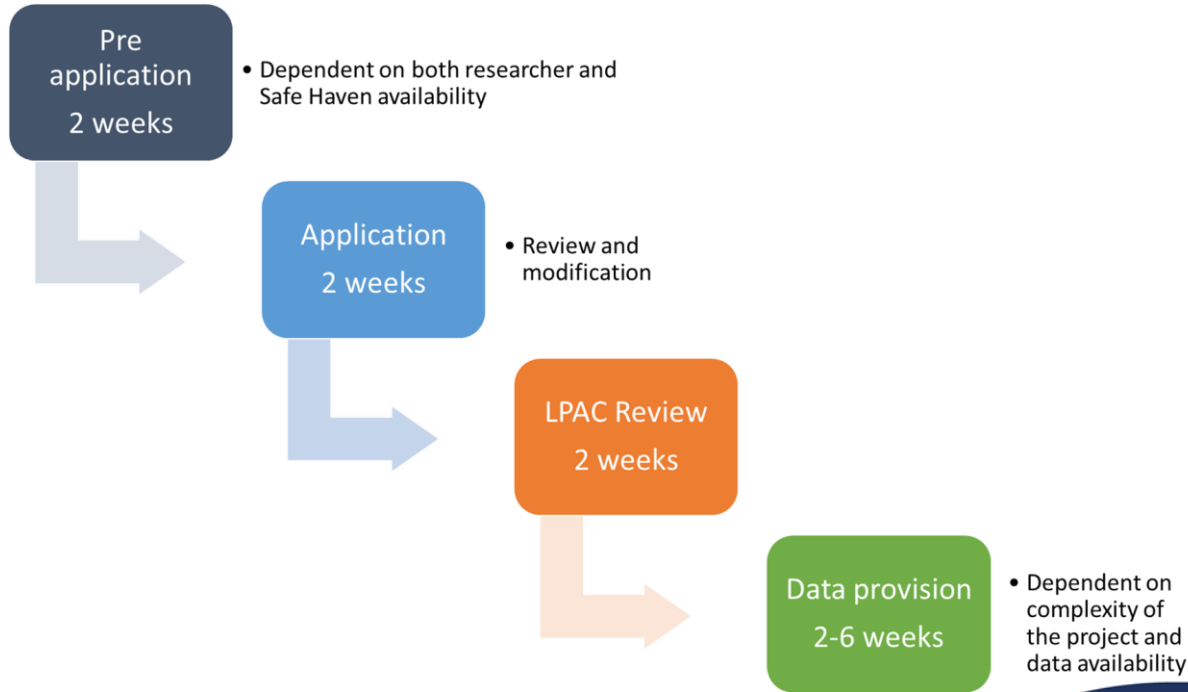
Underpinning Governance – use of un-consented data



Delegated REC awarding powers

- 2012 Research. GGC data
- 2016 Research & Innovation. GGC images
- 2017 Renewal. Research & Innovation Extended to WoS
- 2020 Amend to join images & data REC. Explicit re range of text & possible linkage to genomic, proteomic, transcriptomic data

TRE - fast turnaround for research project approvals



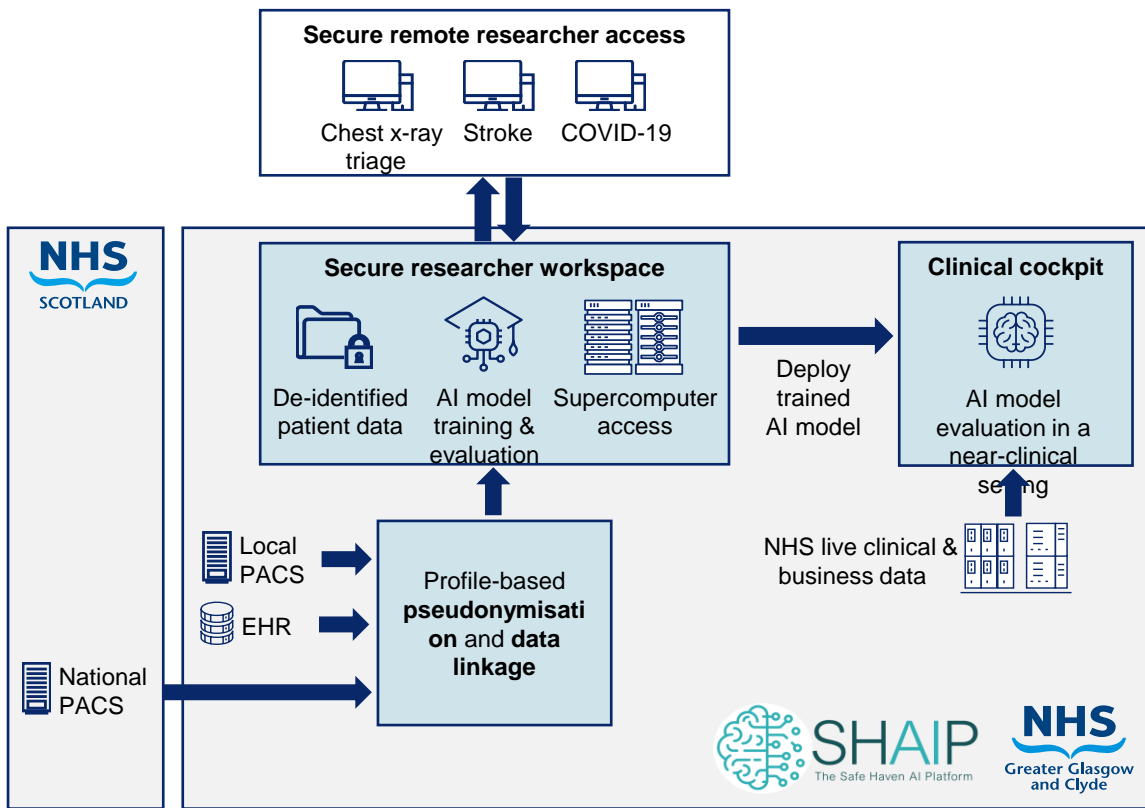
GENERIC IG APPROVALS

- Delegated REC approving powers
- TRE in University of Glasgow
 - ISO27001 accreditation
 - Scottish Government accreditation
- Signed DPIA and SSP
- Generic DPA

CONTRACTING

- Limited to:
 - approval of LPAC application
 - approval of researchers
 - funding in place
 - VPN access arrangements through UoG

Challenge 1 – governance before infrastructure is in place



No secure SHAIIP

No way to download at scale from National & Local PACS

No image pseudonymisation at scale

Need to develop anonymisation of text for NLP

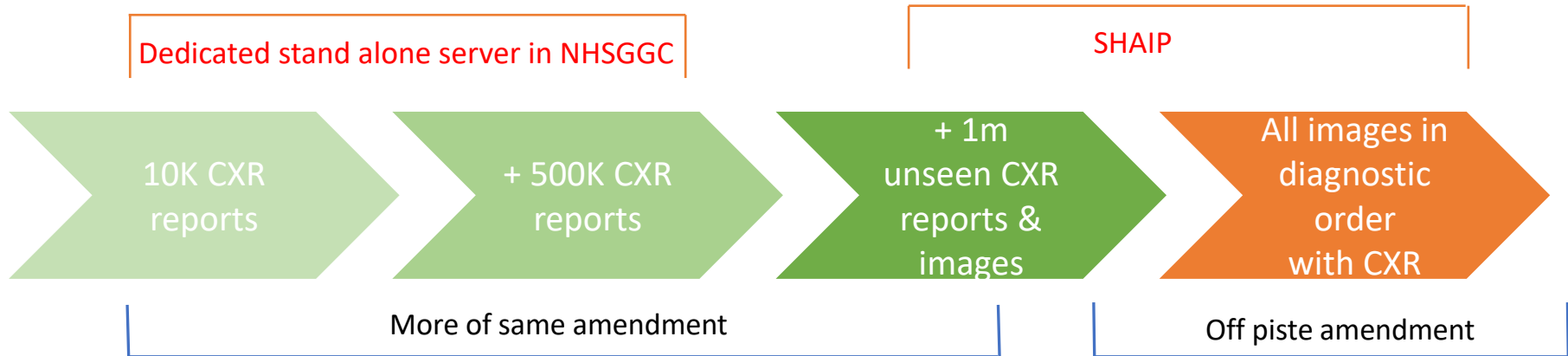
Audit & QA of all new processes by SafeHaven

Challenge 2 – additional contractual requirements

- Innovation Governance Group approval – strategic direction; NHS need; AI approach
- Delegated REC approving powers ✓
- Approval of LPAC application ✓
- Approval of researchers ✓
- Caldicott Guardian approval
- Project Agreement – indemnity, IP, benefit sharing
- Bespoke arrangements require DPIA and SSP (Cloud SSP)
- Sometimes DPA
- Third Party Access approval for network access
- New SOPs for supervised or unsupervised Letter of Access
 - person specific - signed off by individual, employer and GGC
 - clear articulation of data access rules
 - expansion of disclosure - all outputs, including but not limited to data, models, algorithms, software and graphs
- VPN access outwith UK – treat as equivalent to data transfer. ICO adequacy requirements/
country

Challenge 3 – Project amendment – changes to cohort

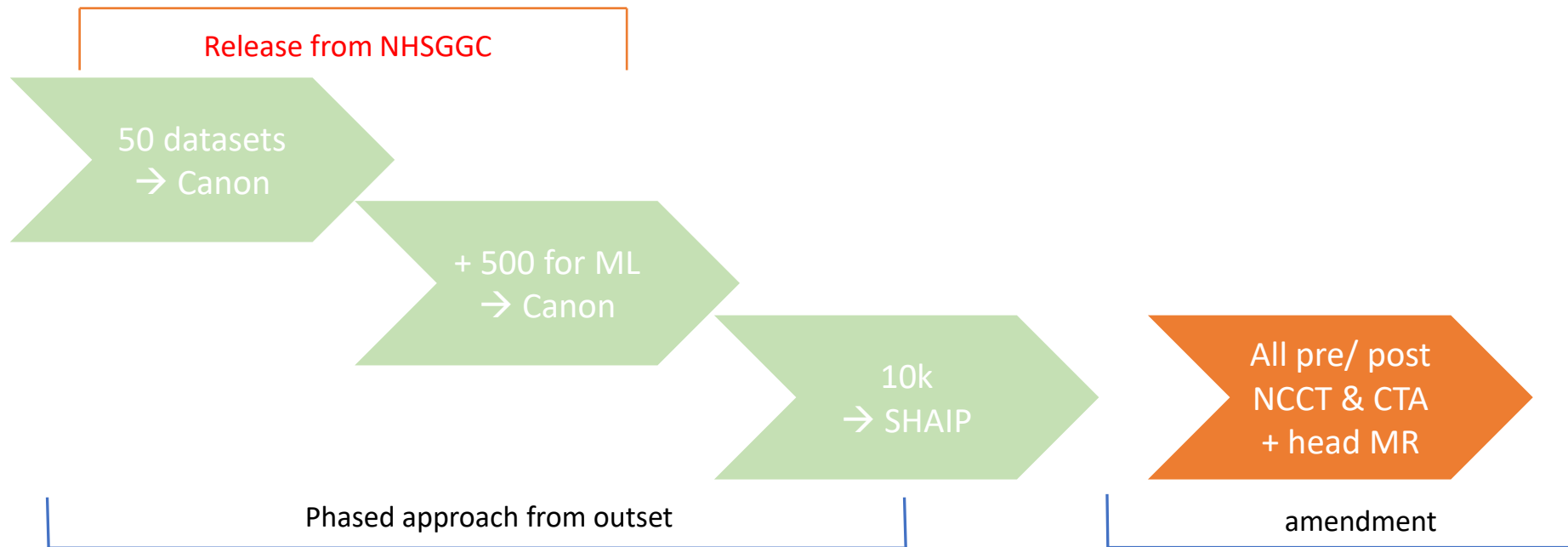
Bering: Developing Chest Radiograph Machine Learning: AiCXR - CI David Lowe



- Good justification given
 - unsupervised pipeline AUC 0.88 → labels reports normal/abnormal
 - want to achieve AUC of 0.98/0.99
- Art of possible only just being understood
 - 1.5m x-rays retrieved at ½ second each

Canon: Acute Stroke Head CTs and Free Text Reports - CI Keith Muir

Index event – NCCT of head on admission, report, IDL, FDL, basic demographic, ICD-10 codes



30,000 CTs retrieved at 4 second each

Canon: Acute Stroke Head CTs and Free Text Reports - CI Keith Muir

Text – endoscopy reports, US
carotid Doppler reports, ED
letters, theatre Op notes,
outpatient letters
Structured data – allergies, >
lab, > meds, diagnostic &
procedures

Considerable expansion of data per patient

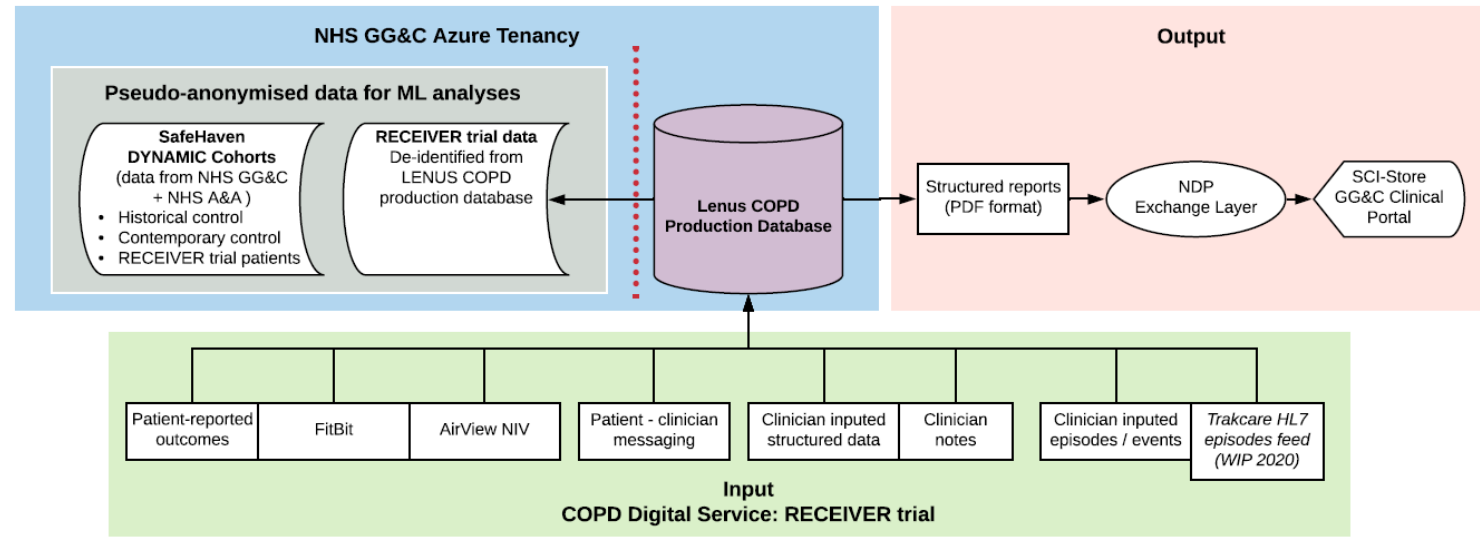
Neuro-oncology
tumour cohort

Amendment **X**
NEW PROJECT
Same container?

- Art of possible only just being understood
 - De-identification of text and NLP by DeepCognito
 - **audit & QA by SafeHaven of each new text source**

Challenge 4 – projects using the Health Data Exchange

DYNAMIC Connectivity and Data Flows



StormID - NHS GG&C Data Processing Agreement

Schedule 1 = Access to Lenus COPD Production Database and associated architecture for technical support and dataset preparation

Schedule 2 = Access to pseudo-anonymised data for ML analyses

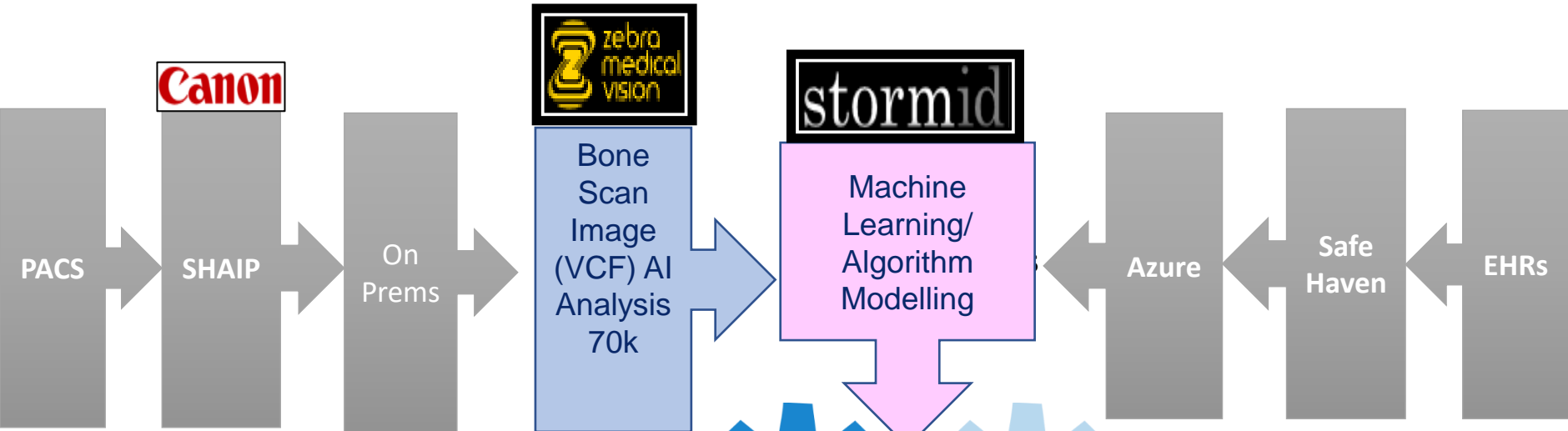
SafeHaven

Validation of pseudonymisation, audit of processes, disclosure



Challenge 5 – Agreeing governance route

OPTIMAL: OsteoPorosis Treatment Identification using Machine Learning



- 1) Docker – load algorithm on SHAIP
- 2) Run algorithm on Preams server
- 3) CE marked – DICOM forwarding from PACS using accession numbers

Osteoporosis Composite risk and machine learning model

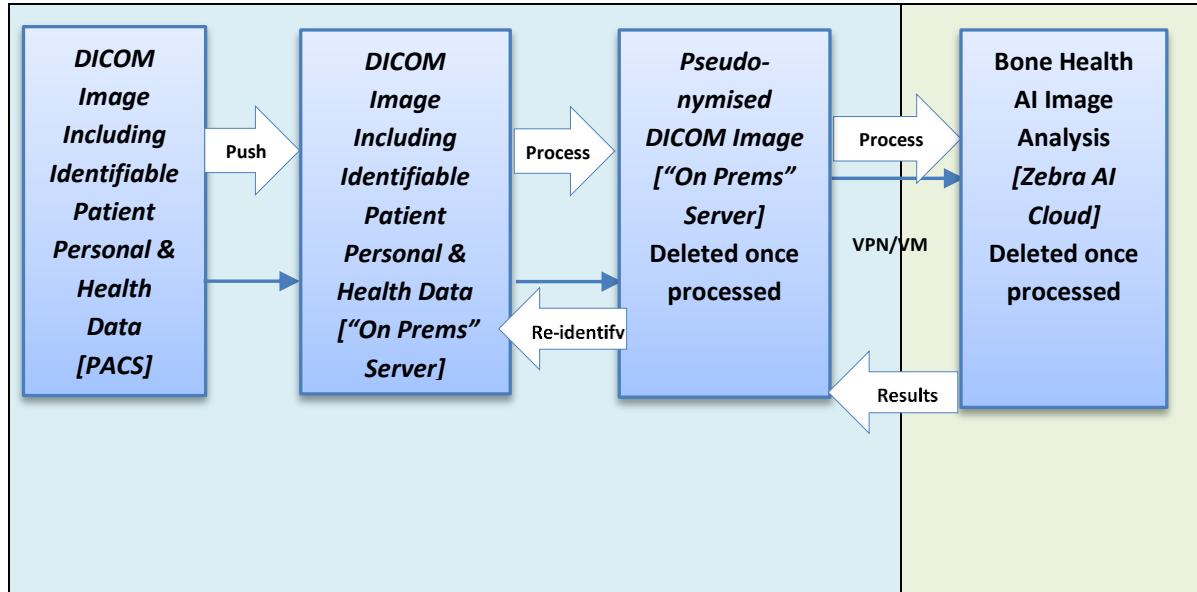
Companies offering Cloud AI solutions - problems

- Many companies want to offer a Cloud-based AI solution
 - involves transfer of identifiable images out of the NHS network
 - **Not acceptable to our IG**
 - offer their own pseudonymisation on Cloud outwith our network
 - **Requires validation by SafeHaven**
- Ensemble project (discovery phase)
Zebra product is CE marked – fully developed AI algorithm (not discovery phase)
 - **falls foul of need to inform patients (ICO rules)?**



More difficult in terms of governance than bringing model to the data

4)



Major effort – sending 70K in 5K batches
 SafeHaven validation of pseudonymisation
 Agreed in DPIA



Data Discovery -governance issues

- Substantial amendments to cohort builds will be needed:
 - we need to plan seen/unseen data from the outset
 - how will this work for fully anonymised cohorts?
- For algorithm development SafeHaven data is very useful but for deployment lower latency data will often be required - this needs to be validated too
- We need exemplars of moving models to other healthcare systems (within and outwith UK) to validate models and minimise bias
- Projects often need bespoke arrangements with complex governance requirements



Deployment testing -governance issues

- Much easier IG when the models are deployed within our network
 - may not suit company business model
 - easier with our electronic systems that engage directly with patients to inform, consent and remove from AI if required
- Often inconsistent advice from MHRA re requirement for a CIMD
 - software evaluation v CIMD
 - IRAS filters – feasibility or pivotal trial (nothing in between)
- Differences in IG requirements between AI automated processing and AI informed decision support need to be clearer
 - may be staged approach, starting with clinician review