DataLoch

Governance for Non-Traditional TRE Use

DARE UK Mid-Sprint Exemplar, 21 April 2022

UK Research and Innovation

HDRUK Health Data Research UK

ADRUK

Challenge

How to make data access possible for non-traditional public-data users?

Traditional users: academic researchers / NHS clinicians doing quantitative, statistical research using standard software.

Examples of Non-Traditional users:

- Third sector/commercial researchers
- AI/Machine Learning tool developers

Issues in extending to non-traditional users:

- Demonstrating trustworthiness (in people and projects)
- New software security requirements
- Contractual arrangements & accountability
- Lack of understanding – new technique requirements vs data security

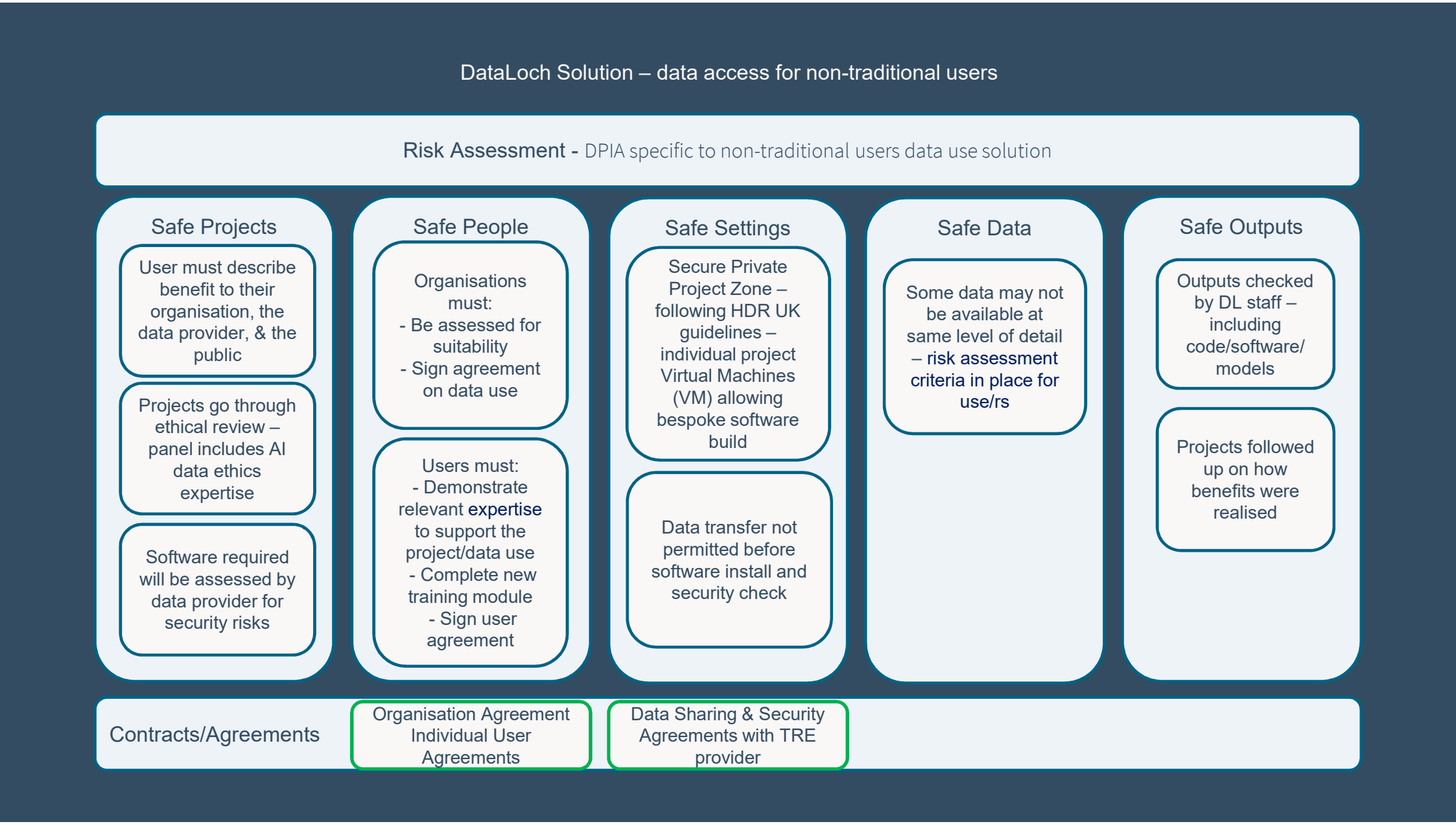
Additional requirements

From public engagement and stakeholder discussions:

- Transparency of benefits to public, research organisation & data providers
- Trustworthiness criteria for organisational & individual users
- Accountability for users – to abide by security and fulfil project assurances
- Adequate security for data, software and research outputs

DataLoch solution – main features

- Application process amended to include: clearer public benefit question, organisational and project criteria
- Ethical review process for all projects – including AI data expertise
- Training module and user guide to support safe users.
- New secure TRE for non-traditional use/rs
- Contractual requirements – framework agreement with organisation (including IPR requirements), user agreements with individuals
- Follow up process to confirm data used fulfilled objectives


DataLoch Solution – data access for non-traditional users
Risk Assessment - DPIA specific to non-traditional users data use solution
Safe Projects
User must describe benefit to their organisation, the data provider, & the public
Projects go through ethical review – panel includes AI data ethics expertise
Software required will be assessed by data provider for security risks
Safe People
Organisations must:
- Be assessed for suitability
- Sign agreement on data use
Users must:
- Demonstrate relevant expertise to support the project/data use
- Complete new training module
- Sign user agreement
Safe Settings
Secure Private Project Zone – following HDR UK guidelines – individual project Virtual Machines (VM) allowing bespoke software build
Data transfer not permitted before software install and security check
Safe Data
Some data may not be available at same level of detail – risk assessment criteria in place for use/rs
Safe Outputs
Outputs checked by DL staff – including code/software/ models
Projects followed up on how benefits were realised
Contracts/Agreements
Organisation Agreement
Individual User Agreements
Data Sharing & Security Agreements with TRE provider

DataLoch

Considerations through project cycle – for non-traditional use

Project Request

- Demonstration of public, research org and public sector benefit
- Description of project hardware, software, data needs.
- Project sponsorship options discussed
- IPR of software and outputs described

Information & software governance

- Organisational and individual assessment
- Ethical, data & methods review
- Agreements with organisation and individuals
- Confirm data use limitations
- User training

Virtual machine setup in secure environment

- Software installation
- Software security assessment
- Data extract prepared and risk assessed

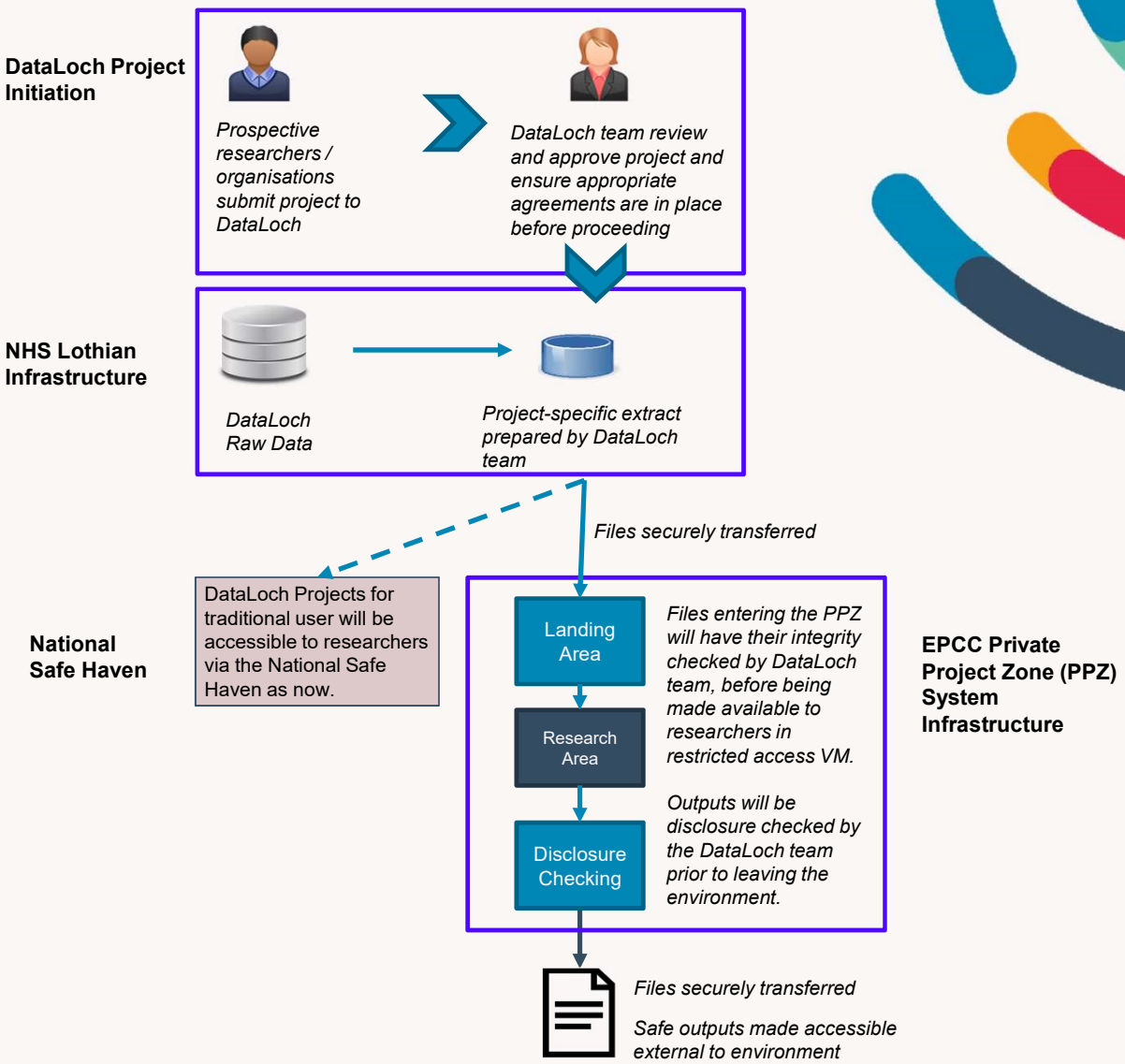
Project Activity

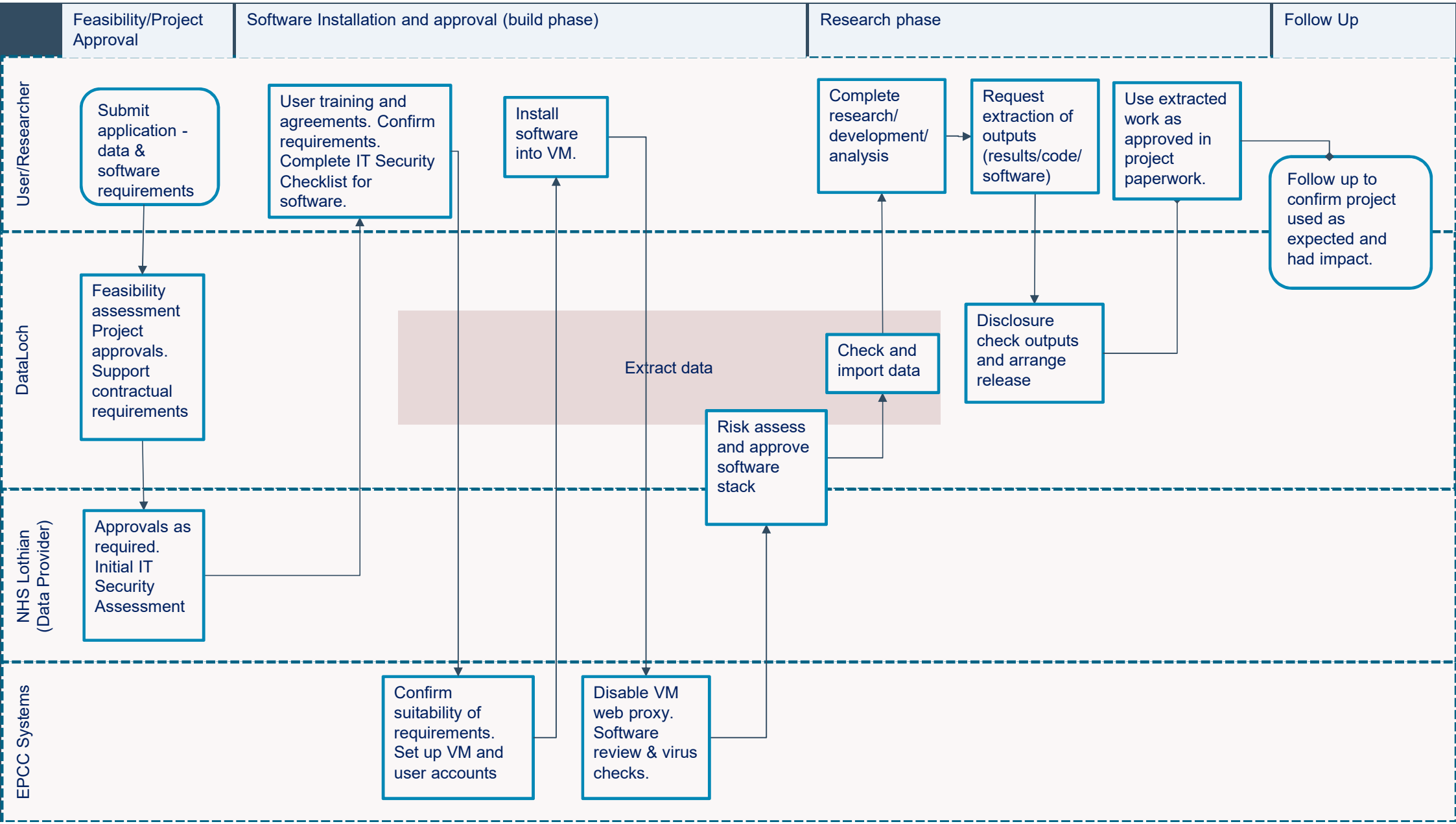
- Data transferred
- Outputs including code, stats results AND tools/models checked according to guidance for novel outputs

Follow up

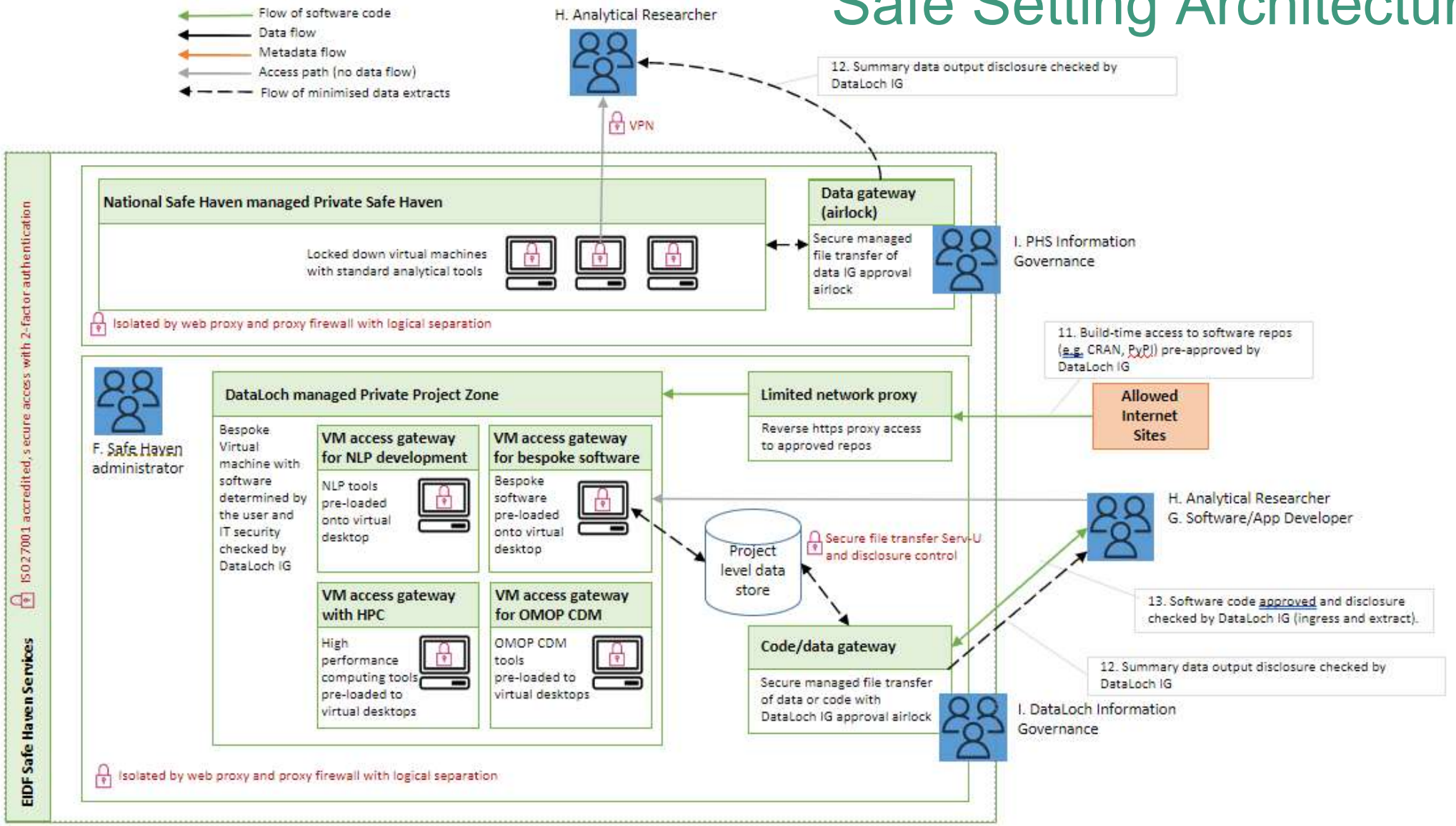
- Confirmation outputs used as expected and benefits realised for both organisation and public

Data Flow


DataLoch Project Initiation
Prospective researchers / organisations submit project to DataLoch
DataLoch team review and approve project and ensure appropriate agreements are in place before proceeding
NHS Lothian Infrastructure
DataLoch Raw Data
Project-specific extract prepared by DataLoch team
Files securely transferred
National Safe Haven
DataLoch Projects for traditional user will be accessible to researchers via the National Safe Haven as now.
Landing Area
Research Area
Disclosure Checking
Files entering the PPZ will have their integrity checked by DataLoch team, before being made available to researchers in restricted access VM.
Outputs will be disclosure checked by the DataLoch team prior to leaving the environment.
EPCC Private Project Zone (PPZ) System Infrastructure
Files securely transferred
Safe outputs made accessible external to environment


Feasibility/Project Approval
Software Installation and approval (build phase)
Research phase
Follow Up
User/Researcher
DataLoch
NHS Lothian (Data Provider)
EPCC Systems
Submit application - data & software requirements
User training and agreements. Confirm requirements. Complete IT Security Checklist for software.
Install software into VM.
Complete research/ development/ analysis
Request extraction of outputs (results/code/ software)
Use extracted work as approved in project paperwork.
Follow up to confirm project used as expected and had impact.
Feasibility assessment Project approvals. Support contractual requirements
Extract data
Check and import data
Disclosure check outputs and arrange release
Risk assess and approve software stack
Approvals as required. Initial IT Security Assessment
Confirm suitability of requirements. Set up VM and user accounts
Disable VM web proxy. Software review & virus checks.

Safe Setting Architecture



Informing the governance: views of the local population

Public consultation to explore conditions and recommendations around non-traditional use/r access to health care data.

Survey – March 2022

- 595 respondents from City Region Deal area (Lothians, Fife, Borders)
- Report currently in development

Deliberative workshops – April 2022

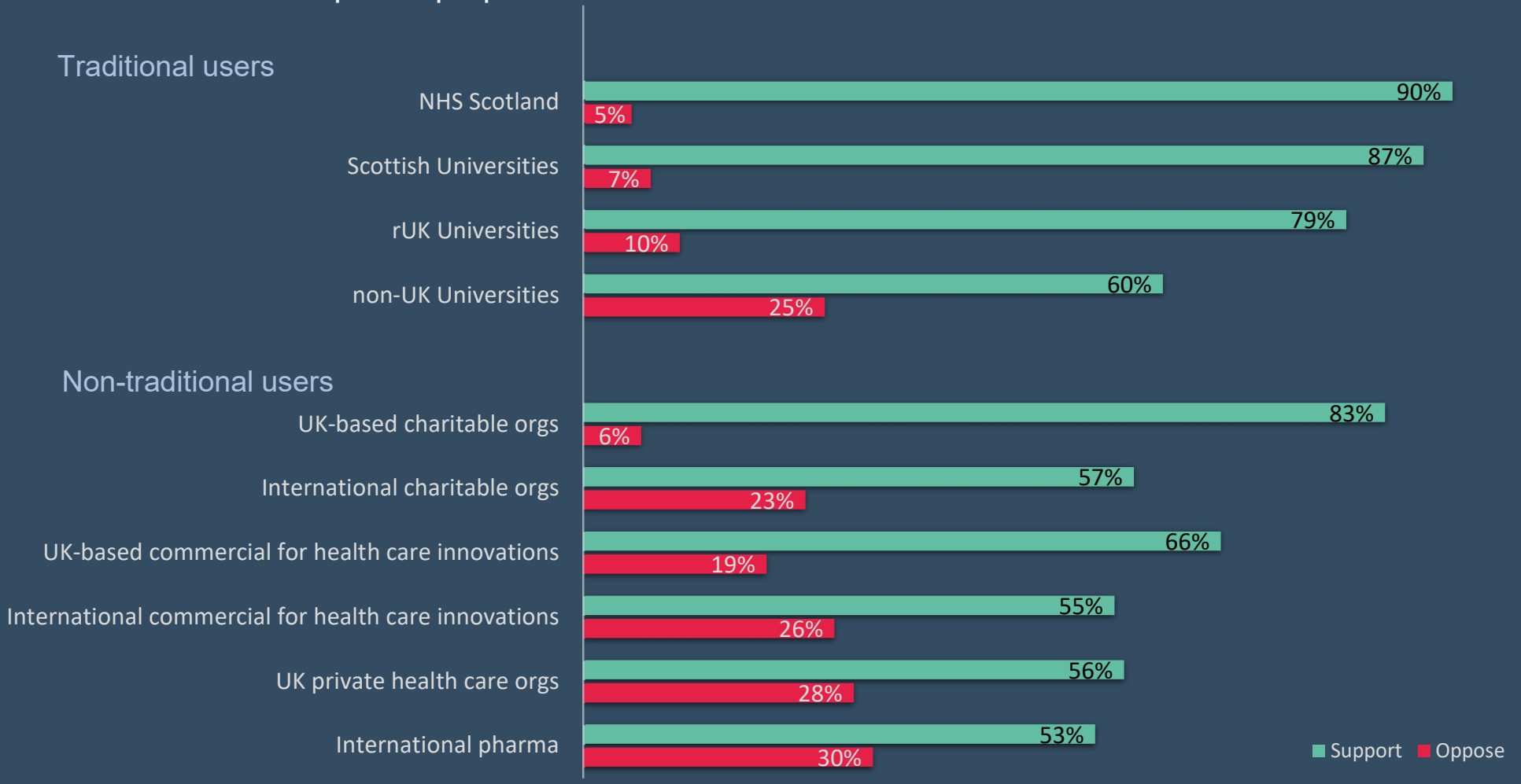
- 40 participants as broadly representative of local population
- Events in April with report on recommendations in June

Headline findings from the survey data follow on the next slides. These will be integrated with the workshop recommendations to inform governance decisions.

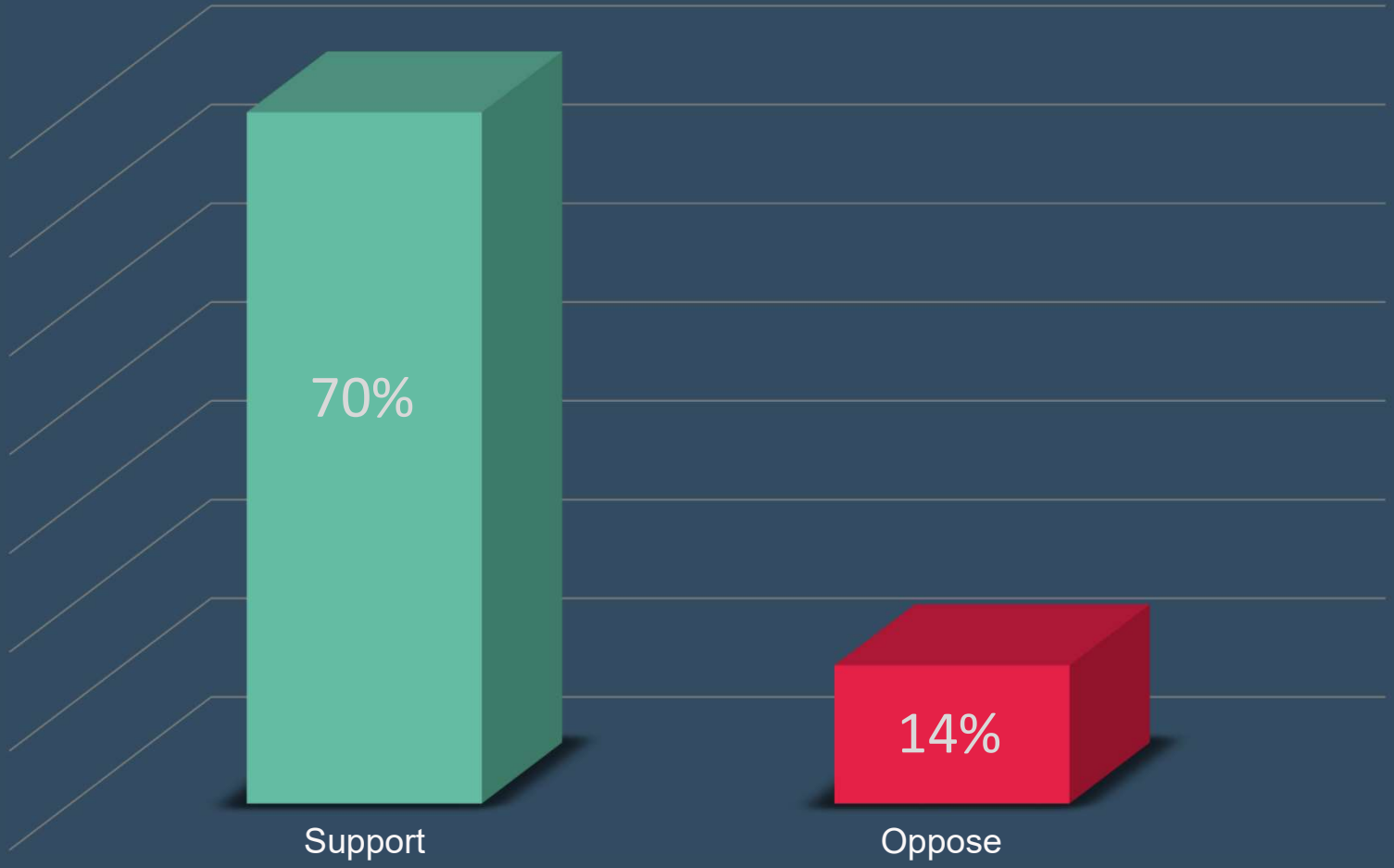
Q. Overall, which of the following statements is closest to your view?

Note: “Agree much more” was 50%

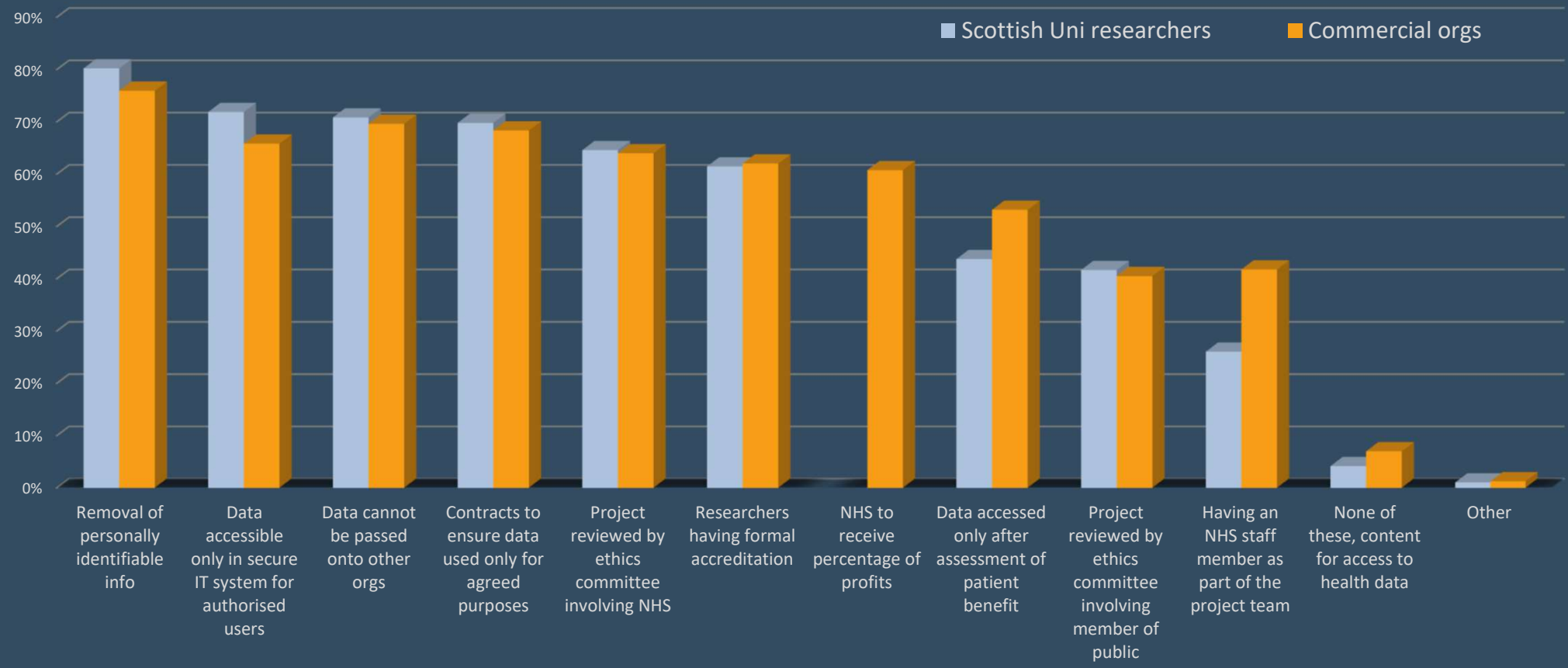
Q. Would you support or oppose each of the following types of organisations using health data about you for research and development purposes?



Q. Artificial Intelligence – or AI for short – involves the use of computers to analyse and act on health data. For example, AI can be used to detect diseases more accurately and in their early stages, or to identify patients who are more at risk of developing a health condition. Would you support or oppose AI developers using health data about you for these purposes?



Q. Which of the following conditions, if any, would you want to have in place before [these people] could access health data about you for research or R&D purposes?



Note: % above given as proportion of respondents who are content for each organisation to access the health data. Removed from the data are 4% of respondents who did not want Scottish Uni researcher to have access; and 21% did not want commercial organisations to have access.

DataLoch

To discuss your priorities and map out how DataLoch can support you, please contact:

DataLoch@ed.ac.uk

www.dataloch.org