

Project Assured Data Access

Henry Hughes

-
- Significant investment in medical research specifically in trying to unlock the value of the data collected by the government and NHS in order to further knowledge and research
 - Drivers
 - The NHS collects significant amounts of health data in treating us
 - This data could help deliver improvements in understanding the causes of illness and improve treatments and therefore outcomes for us all
 - In terms of health research it's a significant resources, it could also help give the UK a competitive advantage internationally
 - Challenges
 - Health data is very personal and sensitive
 - Significant number of ethical, consensual and practical hurdles to making use of the data for research

-
- More precisely ‘The FARR Institute of Health Informatics Research’ - £37.5M award lead funder MRC
 - Four nodes, located in Universities across the country, London, Manchester, Swansea and Dundee
 - Aim: “deliver high-quality, cutting-edge research linking electronic health data with other forms of research and routinely collected data, as well as build capacity in health informatics research”
 - “The Farr Institute aims to provide the physical and electronic infrastructure to facilitate collaboration across the four nodes, support their safe use of patient and research data for medical research, and enable partnerships by providing a physical structure to co-locate NHS organizations, industry, and other UK academic centers.”
 - Linkage with the Medical Bioinformatics Initiative

- Consortium lead by Southampton awarded £7.6M to help launch a project to give access to government data (and other agencies including tax, education and health) for academic research
- Funded by Economic and Social Research Council (ESRC)
- Vision is “to see this data transformed into knowledge and evidence which can be used to inform public and economic policy - helping to tackle some of the major issues facing society, in an innovative and efficient way.”
- "We will manage and maximise the use of new data linkages across Government departments and sectors to give safe, secure and strictly managed access to anonymised data for research purposes."

-
- Requirement for connectivity to move electronic health data securely and support research collaboration
 - Challenge in demonstrating that health data is appropriately protected – how do we give the public confidence?
 - Can a standardised approach to the connectivity element:
 - Provide economies of scale in secure connectivity?
 - Give the public greater confidence through external verification?
 - Support longer term higher bandwidth requirements (10Gbit/s +)?
 - This approach be extended and used more widely?
 - Project is in two parts:
 1. Secure connectivity
 2. Access and identity management

1. Inter-Farr – Transfer of data sets as part of the research project into diabetic drugs and cancer. Initial trial between Farr centres at Manchester and Dundee. Data being transferred would be anonymised – but with a non-zero risk of re-identification.
2. Intra-Farr – Transfer of data to support the ALSPAC project between Swansea and Bristol. Swansea is acting as the main data centre, but some data is also held at Bristol.
3. ADRC/Farr Pod to Data Centre – Connectivity between accredited secure rooms that can be connected to ADRC data centres for remote working. Initial trial connecting secure rooms to Southampton ADRC.

1. Dementia Study – The objective of this will be to enable a researcher to use their home credentials to authenticate in order to request the creation and delivery of collated datasets.
2. This project will enable researchers from HeRC and other N8 institutions to access a new large shared memory computing facility at Leeds using their home institution credentials. Building on existing work within the Janet Moonshot pilot the project will demonstrate access to the N8 HPC, HeRC and DiRAC facilities using the same credentials.
3. eMedLab will help the project partners to analyse human genome data and medical images, together with clinical and other physiological and social data. This pilot will demonstrate how a common AAI will allow researchers to access different datasets using a common credential.

-
1. Encrypted VPN infrastructure between organisations, providing enhanced confidentiality and integrity.
 2. Infrastructure to be operated using agreed processes, procedures and controls with external certification to ISO27001 information security standard.
 3. Range of policy and operation process development
 4. Routing hardware will be deployed to each organisation and scaled to support agreed traffic levels (1 Gbit/s initially).
 5. Public IPv6 address space will be used.
 6. Likely to require connection agreements with each connected organisation.

-
1. Identify if any changes or new policies are required to support this activity
 2. Explore using Moonshot to provide access to data sets and compute resources
 3. Explore using the UKfederation to provide data controllers with ‘member in good standing’.
 4. Demonstrate an integrated approach with a data controller’s second factor.
 5. Evaluate a range of second factor approaches that could be deployed in this context (Google Authenticator, SMS, Hardware tokens)
 6. Identify if any changes are required to existing services to support these activities (certificates, incident response, the Ukfederation or Moonshot)

-
1. Agreement on requirements and use cases – *complete June 14*
 2. Funding approval – *complete October 14*
 3. Detailed project planning – *in progress*
 4. Detailed design and architecture of infrastructure – *in progress*
 5. Operational standards development and controls – *Q1 15*
 6. Infrastructure deployment, installation and commissioning – *Q2 15*
 7. Initial operational and testing with customers – *Q3 15*
 8. Customer trials begin – *Q4 15*
 9. External certification ISO27001 process – *Q1 16*
 10. Recommendations – *Q2 16*



janet

Questions?