



Secure networking and AAAI

Farr Institute, Administrative Data Research Centres, Medical
Bioinformatics

22th May 2014

Overview

- Research Council funding for big data in health, medical and administrative data
- Need for secure data access and transfer to authenticated, authorised users
- Working with Janet
 - Secure networking
 - AAI

Big Data funding for health, medical and administrative data

- MRC £20M for the four Farr Institute nodes, for eInfrastructure and buildings, June 2013
- ESRC £34M for four Administrative Data Research Centres (ADRC) and Administrative Data Service, Nov 2013
- MRC £39M for six Medical Bioinformatics Initiative projects, Feb 2014

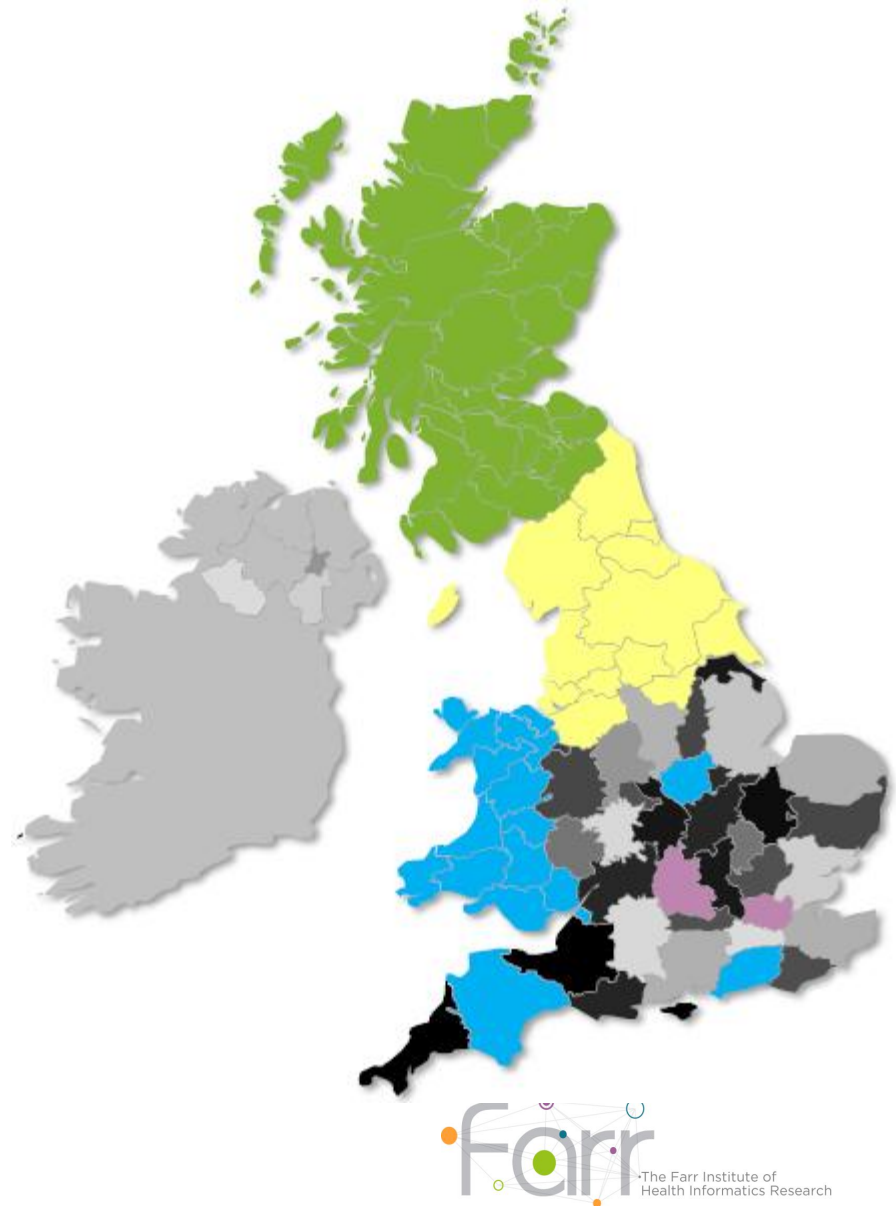
Funded Infrastructure

- ★ Farr Institute
 London
 Manchester N8
 Scotland
 Wales
- ★ Medical
 Bioinformatics
 UCLP-Crick-
 Sanger-EBI
 Leeds-N8
 Warwick-Swansea
 Oxford
 Uganda-Oxford-
 Sanger
 Imperial-Sanger-EBI
- ★ ADRC
 So'ton-UCL-IFS-IoE
 Swansea
 Edinburgh
 Belfast
 Essex (ADS)



Farr Institute of Health Informatics Research

- The four nodes aim to harness health data for patient and public benefit by setting the international standard in trustworthy reuse of electronic patient records and related linkable data for large-scale research
- Distributed virtual research institute
 - Farr London
 - Farr HeRC N8
 - Farr Scotland
 - Farr Wales



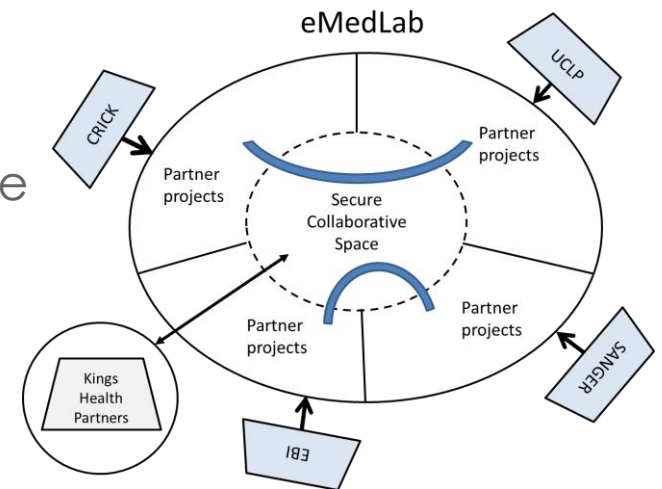
Administrative Data Research Network

- **The core aim of the ADRN is to facilitate access to and linkage of de-identified administrative data routinely collected by government departments and other public sector organisations**
- ADRC – England
- ADRC – Scotland
- ADRC – Wales
- ADRC – Northern Ireland
- Administrative Data Service (Essex)



Medical Bioinformatics Initiative

- **Building new ways of linking across complex biological data and health records to solve key medical challenges**
- UCL-QMUL-LSHTM-Crick-Sanger-EBI-KCL
 - Data-driven discovery for personalised medicine
- Oxford
 - Oxford Big Data Institute
- Leeds
 - Leeds MRC Medical Bioinformatics Centre
- MRC/UVRI Uganda Research Institute
 - Medical Informatics Centre
- Warwick and Swansea
 - Medical Microbial Bioinformatics
- Imperial
 - Medical Bioinformatics partnership



Many projects, one challenge

- All projects will use (and move) identifiable, sensitive data
 - Names, addresses, dates of birth
 - Medical records
 - Education and finance records
- All centres will have to authenticate, authorise and allocate resources to users
 - Bona fide researcher
 - Should only be able to see what they have ethical approval to see

Collaboration is key

- Each project or Centre has more than one university or institute
 - E.g. UCL-QMUL-LSHTM-Crick-Sanger-EBI-KCL
- Each project has a remit to collaborate with other projects
 - E.g. Farr nodes form one virtual centre
- Each project or Centre has a remit to make resources available to other academic groups
 - ADRC provides secure rooms and training to use admin data

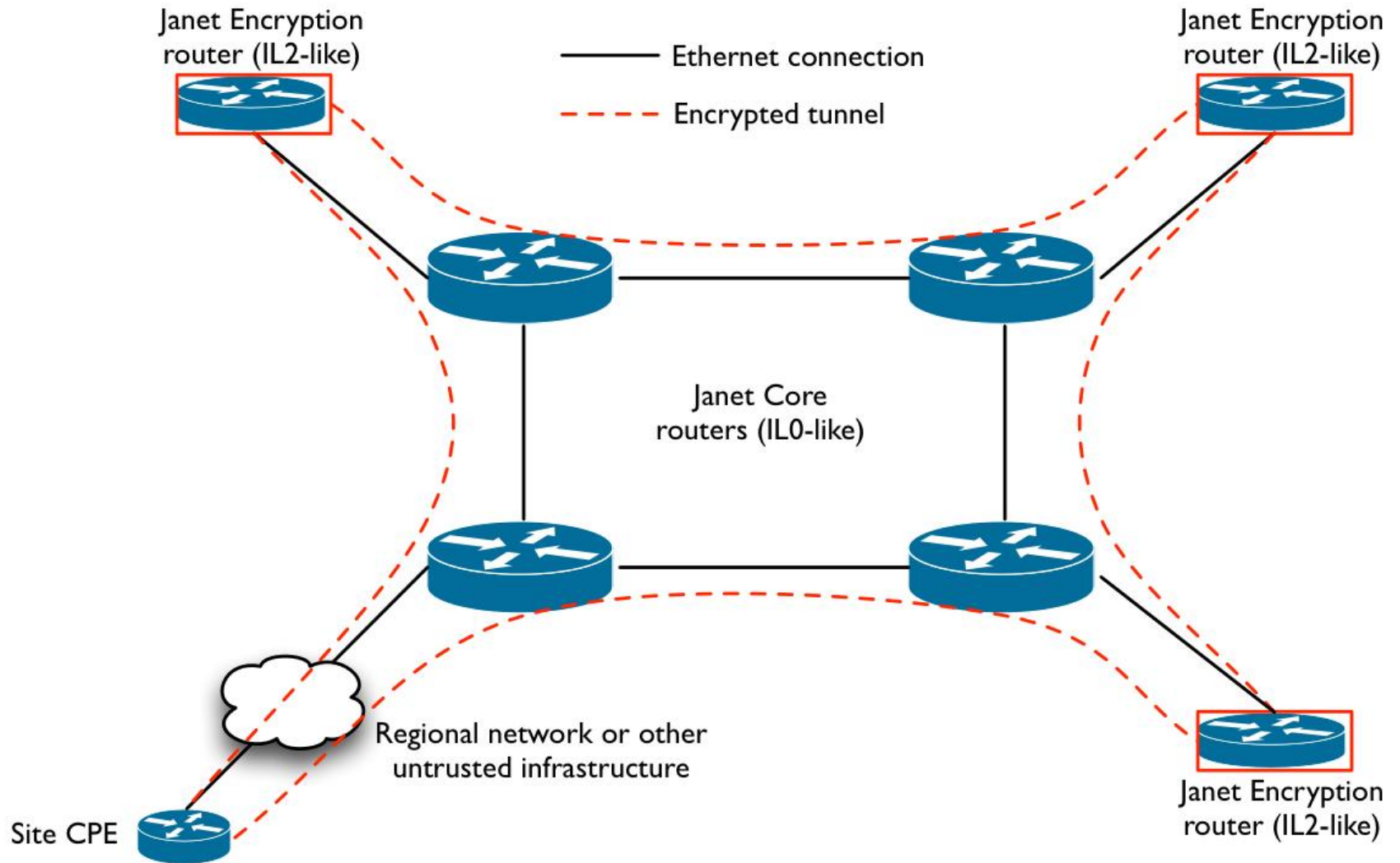
But

- There is no national system to allow user to log on to other host systems with own host ID
 - Except eduroam
- There is no national secure network layer

janet

- Farr eInfrastructure group and Janet met in autumn 2013 to discuss how to
 - Implement secure networking layer
 - Set up AAAI user management system
- With announcement of ADRC and Medical Bioinformatics funding we quickly realised these projects need same solutions
- Series of meetings and workshops to bring these groups together
 - UCL, Crick, Manchester, Leeds, Dundee, Swansea, Oxford, Southampton, Sanger,.....

Secure network topology

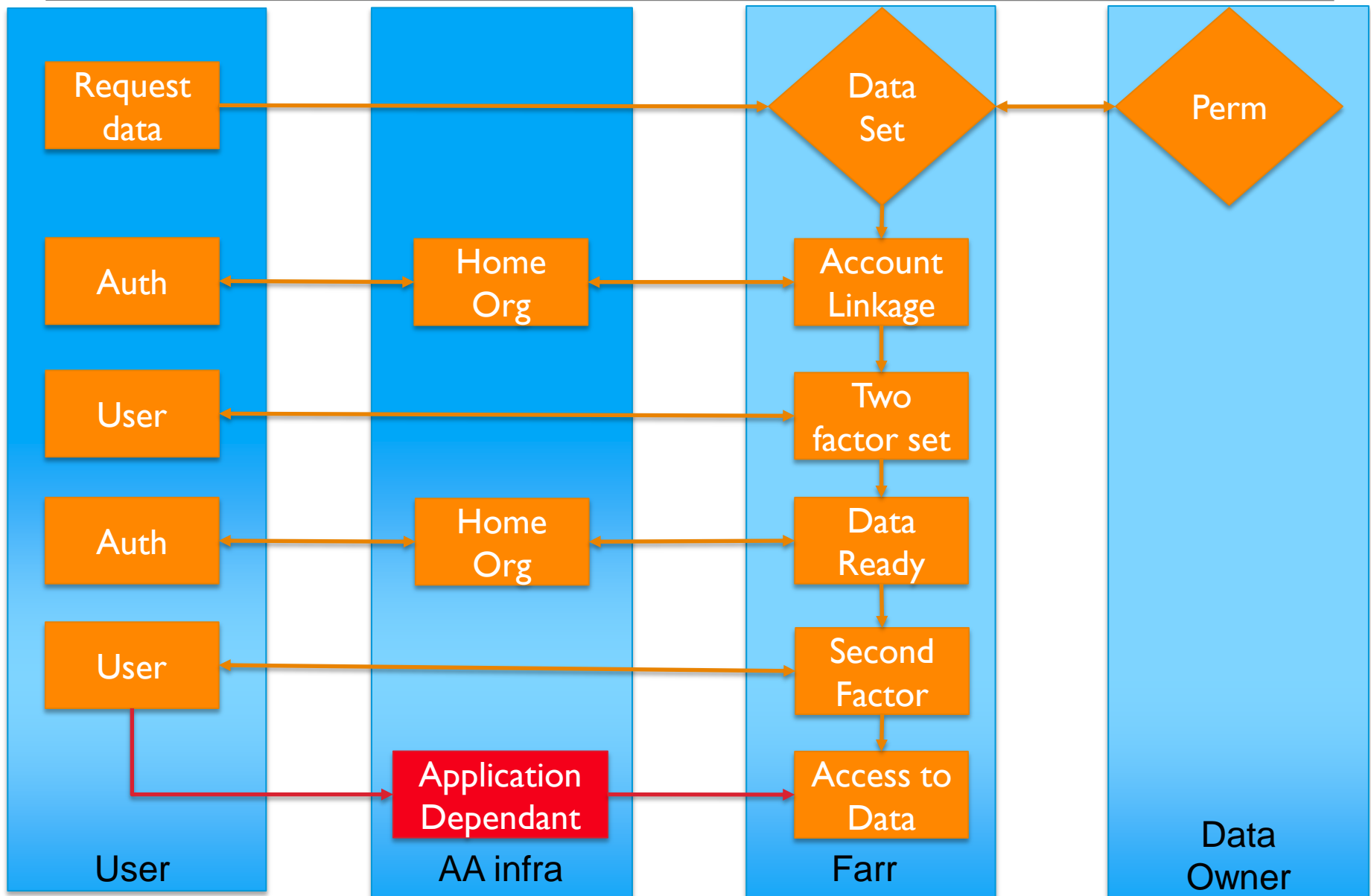


AAAI

- Authentication – It's me
 - Authorisation – Yes, you are you
 - Allocation – Yes, you can use this
 - Identity Management – system needed to know it is you
-
- thanks to Jeremy Yates and Dave Britton for this explanation

User Requesting Data (generic model)

janet



Meeting the Big Data challenge

- Being able to access data securely
- Being able to share data safely
- Being able to work together collaboratively
 - Solve the problem once for everyone
- Pilots identified
 - #1 is to share data-sets between the Manchester and Dundee data centres in a research project asking the question “Do diabetic drugs cause cancer?”
 - #2 researchers can use their home credentials to access the N8 HPC, HeRC and DiRAC facilities
 - build on existing work within the Janet Moonshot pilot
 - Subject to funding



WE SHOUTED OUT ABOUT
SOME NEW FUNDING
THEN THAT THING HIT HIM!



DeepFat '09