Public Health e-Labs:  
A federated model for e-Epidemiology

1John Ainsworth, 2Patricia Baker, 3John New, 4Martin Gibson, 5David Pioli, 6Iain Buchan  
1Northwest Institute for BioHealth Informatics, University of Manchester, UK  
2Salford Royal Hospitals NHS Foundation & Primary Care Trusts, Salford, UK  
buchan@manchester.ac.uk

Problem:  The problem we address is the general lack of effective and efficient integration of health-related data, methods and expertise, across defined populations, for public benefit. This is a problem for most health systems throughout the world.

Population:  We focused on the Northwest of England, which has a rich diversity of social structures, health outcomes and research activity. The outputs of the pilot e-Lab projects were of high value to both healthcare and academic organisations. The Northwest Region of England is now preparing to form a confederation of e-Labs, known as ‘Northwest E-Health’. Internationally, in order to bring global benefit from sharing e-epidemiological data, methods and expertise, open standards are needed. We recommend that public health informatics groups come together to define global standards, or ‘meta-standards’, for inter-operable e-Health.

Each healthcare organisation that participates in the e-Lab is able to modify the data they expose and the security policy that controls data-release at any time. This ensures local control of the metadata of the data by the data controller. The metadata of the e-Lab is published to the central e-Lab where they are used for planning database queries.

Within the e-Lab federation there are three tiers of data sensitivity. The healthcare controllers maintain identifiable data and only release pseudonymised data to the NHS e-Lab. The NHS e-Lab is able to link at the individual record level using pseudonymous identifiers. Once the linking has been completed, the data are anonymised, before being delivered to the researcher using the research portal. This ensures that no identifiable data are released by the data controller and that no pseudonymised data leaves the safety of the NHS.

The e-Lab research portal gives the scientist access to data, workflows, computing resources and research objects. An aggregated metadata repository describes the data available across all the e-Lab nodes. This enable users to plan their data set request and determine the data that will be returned in accordance with their security privileges.

The process of creating a new linked-dataset begins with the user planning the data required using the metadata catalogue. A new Research Object is populated with the metadata of the required data and sent to the NHS e-Lab linking engine. The Link Controller then coordinates the linking of data from all the target sources (both clinical and non-clinical), gathering a pseudonymised data set from each. When the pseudonymous data are returned they are linked together, then anonymised before being released to the research portal. In exceptional circumstances, where secondary linkage of transformed data is required, the researcher works with pseudonymised data, within the healthcare organisation, under its information governance.

Conclusion and Recommendations:  The sustainability of an e-Lab federation relies on its value for service development for local care providers. The federated model permits a unique pseudonymisation scheme to be used for each data linkage request, thus reducing the feasibility of performing a brute force attack on identifiers such as NHS number. The sustainability of an e-Lab federation relies on its value for service development for local care providers. The federated model permits a unique pseudonymisation scheme to be used for each data linkage request, thus reducing the feasibility of performing a brute force attack on identifiers such as NHS number.

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

Each e-Lab Hub  
Expert analysis  
Software engineering  
Education and training  
Business development  
Coordination

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

We propose a federated model for e-Labs, using open standards-based technologies that are easy to deploy and maintain. Each local node of the federation will own datasets integrated datasets. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and use of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

The sustainability of an e-Lab federation relies on its value for service development for local care providers.

The federated model permits a unique pseudonymisation scheme to be used for each data linkage request, thus reducing the feasibility of performing a brute force attack on identifiers such as NHS number.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.

The developing functional map of e-Labs is naturally federal. Each local node of the federation will own datasets integrated across all nodes. Running a research protocol across multiple nodes will use a technology we call ‘research objects’. This approach promotes local ownership, enhancement and reuse of data; it matches existing NHS information governance arrangements and importantly avoids the need to build a centralised data warehouse.

The interface between the NHS network and the public Internet is unidirectional. We propose that the NHS e-Lab engine will call out to the e-Lab research portal to pull in new dataset requests. Once prepared, new datasets will be pushed back to the research portal and the user is notified.