CHC NENC: Developing a Technical Infrastructure – Final Report

Connected Health Cities
North East and North Cumbria

Authors: Prof Paul Burton, Dr Olly Butters, Dr Stuart Wheater, Dr Nick Booth

Date: 30 April 2020

Status: FINAL (v1.0)

Contents

Background

In January 2016 the Connected Health Cities project began. It comprised of 4 sub-regional projects, in North East England and North Cumbria (CHC NENC), North West coast of England (including South Cumbria, Lancashire, Merseyside and Cheshire), Greater Manchester, and Yorkshire and Humberside. The whole project was funded by the Northern Health Sciences Alliance, and coordinated by a fifth organisational entity, known as “the hub”, based in Manchester. CHC NENC was based in Newcastle University, in the Institute of Health and Society.

The geography of the NENC included the conurbations around Newcastle and Sunderland in the north, and Middlesbrough, Stockton, Hartlepool and Darlington in the South. In addition, more rural areas around Carlisle and Whitehaven/Northern Lakes, Northumberland, parts of North Yorkshire and County Durham (Teesside facing populations as opposed to Leeds facing), were included. The overall population covered was about 3.6 million individuals. The Newcastle University project was late in starting and began on 28th March 2016.

The population of the North East of England is in transition. It was an industrial powerhouse in the 19th century until the post-world war two period, with coal mining, shipbuilding, the rail industry and engineering dominating. In Teesside the massive ICI chemical industry, the largest in the UK, grew from the early 20th century until the 1970s. But after the 1970s these industries went into decline. Now there is a residual inherited problem of unemployment and poverty, with associated higher mortality and morbidity than the south of England. Early signs of regeneration of high technology industry around the six universities are beginning to show. There is also a very high reliance on employment from the Public Sector. The imbalance with the relatively booming SE of England is stark, and not apparently improving. Current initiatives in rebalancing include the Northern Powerhouse initiative, which indeed funded this project.

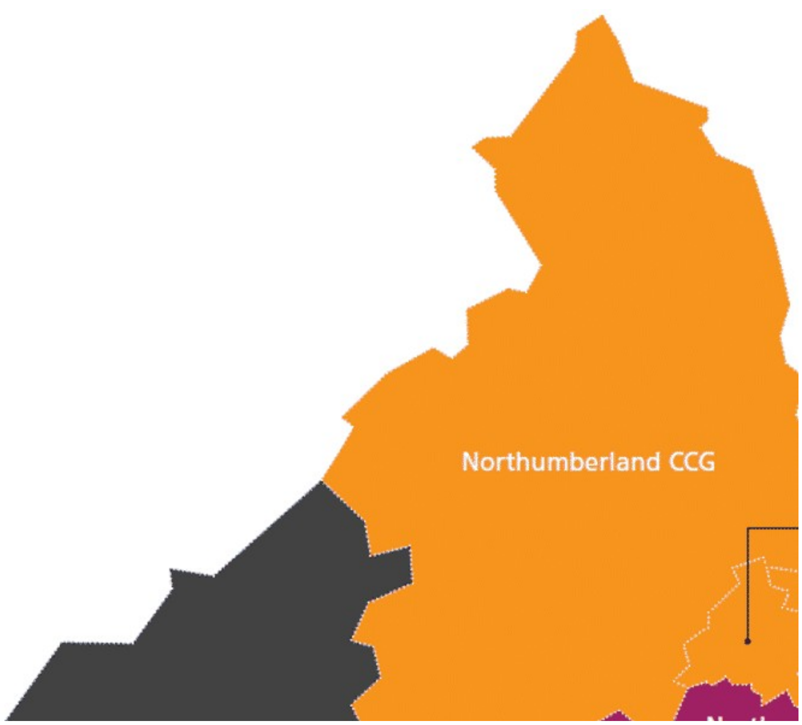


Figure 1: Clinical Commissioning Groups of the North East and North Cumbria (NENC)

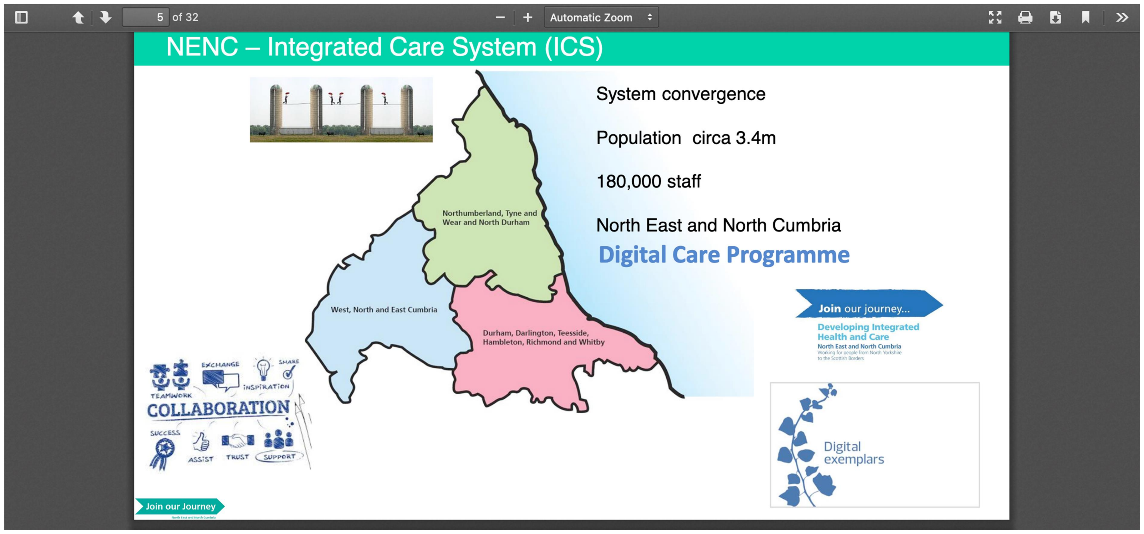


Figure 2: Integrated Care System (IC) of the North East and North Cumbria

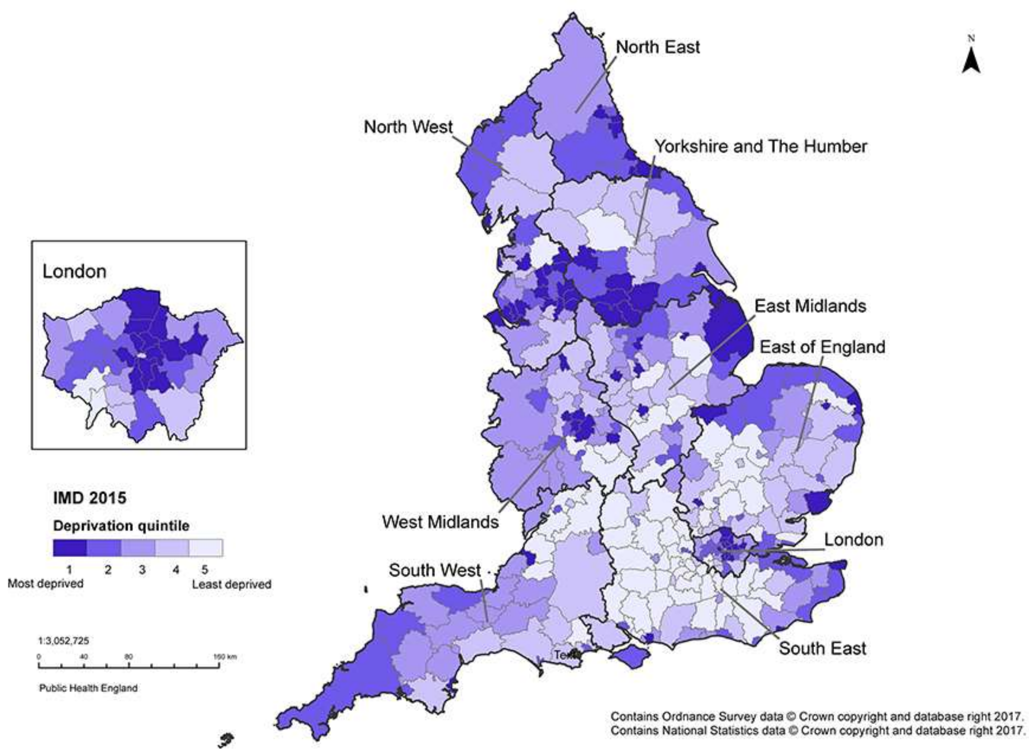


Figure 3: Levels of deprivation in England

Care Pathway Projects

Each of the four CHC sub-regions identified a number of sub-projects, loosely called Care Pathway Projects. The intention was to demonstrate regional data generation from clinical care. In NENC we chose the following projects:

- Durham University submitted a package around predictive modelling of (i) acute medical demand using data from Acute hospitals and the Ambulance service, and (ii) demand for GP services in "Healthy new towns" principally in the south of County Durham.

- The community of North Tyneside submitted a palliative care project designed around a perceived need for a single shared record for palliative care, between General Practice, community services (including district nurses, MacMillan and Marie Curie services, Hospice Care and Acute medical services in Northumbria Health Care Trust.
- Newcastle University submitted a novel project around the concept of record sharing of GP medical records with social care leads, for vulnerable families, with express record sharing permission from the families.

Final reports for each care pathway project are now available on the CHC website.

Overarching technology

A key goal of CHC NENC was to investigate and identify the required components – and options for those components – of an overarching technological framework that would facilitate the above (and future similar) projects without the need for the large investment that CHC necessarily delivered. Rather, from a strategic perspective it was recognised that we should build on that investment. Because of their tight timelines, the care pathway projects could not await the completion of this program of scoping work, so we discuss the details of this work separately from the CPPs as outlined below it was carried out in parallel.

Overarching technology principles

Work prior to the CHC project in the North East, at Newcastle University Business School, had indicated that, in a multivendor community[4], hub based front-line record sharing using commonly available document structures, such as pdf, HL7 CDA, and OpenEHR (using 'document' in the widest sense[5]), would offer a useful start. Indexing using IHE XDS profiles was already successfully undertaken in the LPRES project in North West Coast, using open-source based middleware supported by Tiani-Spirit[6]. In addition, an early piece of work under CHC-NENC was carried out by GE Healthcare Finnamore to collect stakeholder input as to what a shared record and analytic platform for the region might look like (bit.ly/2SiGjVZ). With these issues in mind, the intention was to create an open-standards based platform or hub for federated data analysis and information sharing across the NENC region.

Medical Information Gateway (MIG)

The first stage of the hub development was the decision to make the Medical Information Gateway (MIG) accessible to front line care at local hospitals. This allowed clinicians, with consent from patients, to look at their GP record in order to help their immediate care needs. This one-way real-time pull of individual's GP data is in constant use across the region. This was seen as the first step in building a truly connected NENC, and fed into the design of ARCHIE.

The ARCHIE subproject

With the early success of the MIG, and based on background research (above), a high-level picture of what the infrastructure needed to look like was then developed. It consisted of two major components joined together: A Trusted Research Environment (TRE) and a Health Information Exchange (HIE). A high level overview of this is shown in the figure below, and described in more detail later.

---

[4] 8 Acute trusts in the NE with most major EHR suppliers represented, 2MH trusts with different systems, and 2 GP suppliers predominate – "A perfect Storm of an interoperability problems" – Joe McDonald

[5] A circumscribed information object containing (formatted) text, sections, coded information, with metadata if available, signed or attributed to a particular individual clinican or set of clinicians.

[6] Tiani-spirit.com

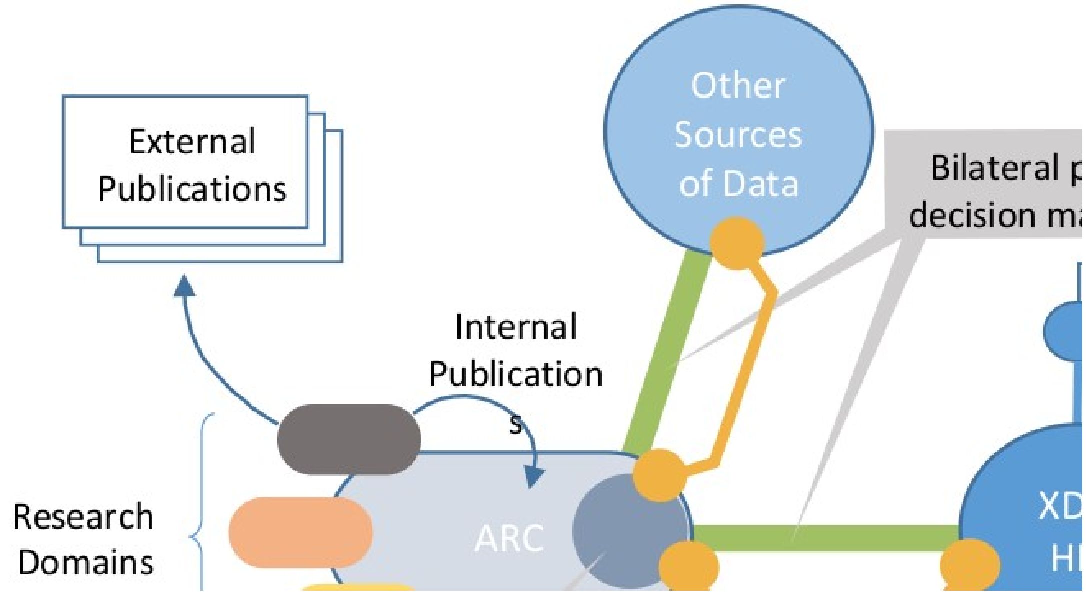


Figure 4: ARCHIE high-level overview

The naming of ARCHIE

There exist various synonyms for a TRE, one pertinent one being an 'Ark'. This was the term used in the original CHC grant, and was based on the impending 'flood of data'. We slightly modified this to 'ARC' and called the project to develop the combined components 'ARCHIE' (i.e. ARC + HIE). For the purpose of this report we will use TRE to describe the trusted research environment, but it is exactly equivalent to the 'ARC' component mentioned here. The focus under ARCHIE to specifically explore and prototype technology underpinning a TRE an HIE and enabling an effective interface between the two represents a distinctive feature of the approach of CHC-NENC in comparison to that in other CHC regions.

It should be noted that, entirely unpredictably, it is now important to avoid confusion between the term "ARC" to mean a TRE and the new NIHR- funded "ARC" project (Applied Research Collaboration) that has since been awarded to NENC led by Prof Eileen Kaner (Newcastle University). There is a particular risk of confusion because the NENC-ARC project is now applying and taking forward a number of the infrastructures developed under CHC-NENC and ARCHIE. In consequence we now try to consistently use the term TRE rather than ARC to refer to the Trusted Research Environment in CHC-NENC.

ARCHIE's early development

The diagram above shows a variety of data sources, data flows and usage domains with different users receiving different services depending on their role. In broad terms the HIE can be seen as the tool that connects all the data and allows it to flow, and the TRE is where, if required, the data may be analysed.

The HIE outlined here acts as a registry for what data are located where, and who is allowed to access them. In this sense it can act in a similar way to the MIG mentioned earlier by allowing front line care givers to access patient's GP data as part of a care session. Further to this it can facilitate the flow of data between other data sources (primary care, secondary care, third sector, etc) in an analogous way. A key point to this description is that the HIE doesn't hold any data itself, it merely facilitates the ability to view data available from third party data sources. This model is described as a publication and subscribe service: when new data about an individual are added to a source system it publishes the fact that the new data exist (but not the data themselves) to the HIE. If another system is subscribed to that information (e.g. that patient, or that kind of data) they get a notification and can act accordingly (e.g. automatically download the data and merge with local record, or automatically schedule an appointment with specialist etc).

The TRE in this scenario acts as just one potential data subscriber, its exact behaviour would depend on the role of the user using it. One scenario may be a clinician wanting to review or analyse key data about patients in their clinic. In this case it may be suitable for them to have fully identifiable data available to them. A different scenario may be a researcher (e.g. an academic or an internal hospital business analysist etc) wanting to analyse data from all the patients in a region, in which case the data would likely need to be pseudonymised or anonymised and appropriate governance permissions sought and obtained.

For both the HIE and TRE described above a key aspect of the infrastructure is ensuring the correct people have access to the correct data. While the HIE can enforce rules it cannot set them. This is where a patient portal comes in. By allowing patients the ability to define who can access their data, the HIE can then allow it to flow to the relevant users. The patient portal was scoped as part of CHC but is not outlined in this report.

With the overarching infrastructure split into discrete parts we were able to start developing, prototyping and applying the constituent parts in parallel.

Trusted Research Environment (TRE)

Goals and rationale

Our primary goals were two-fold:

(1) *To establish a Trusted Research Environment (TRE) to provide a platform for managing and analysing NHS-derived data for our CHC care pathway projects (CPPs) and other substantive analytic work under CHC*

Without access to a TRE that was HSCN compliant, ISO27001 certified and achieved NHS IG-toolkit level 2 (at least) it would have been impossible to realise, and secure timely governance clearance for, the analytic work based on the health or social care data underpinning CHC-NENC. This was particularly so for the CPP led by University of Durham - and the spin off projects, such as the NEWS2 scores project, that were subsequently based upon it . This work stream - which is described in more detail under the CPP reports - was led by Dr Camila Caiado and Dr Ian Briggs for Durham University - and resulted in the successful set up and application of a cloud-based TRE constructed by AIMES Technology.

(2) *To use the CHC TRE to explore, test and inform future strategic and operational plans to develop an NENC regional TRE providing a shared platform for analysis across the region, for Universities, Local Authorities and the NHS, jointly and severally*

To complement the Durham University TRE, we also procured the components of a second cloud-based TRE from AIMES which could interface with a sandpit Health Information Exchange (HIE) – see below – provided by Tiani-Spirit and based at Newcastle University. Jointly these provided the central infrastructure underpinning the "ARCHIE" subproject (see above) of CHC-NENC which was intended as a test-bed for exploring, prototyping and evaluating key components of a putative integrated NENC Health Informatics system which could later evolve into a fully-fledged regional Learning Health System. In effect, this represented the vision espoused by Prof Joe McDonald for the evolution of GNCR: at this time, Prof McDonald led both CHC-NENC and GNCR. The technology and governance work stream underpinning this vision was led and managed by a multidisciplinary group based at Newcastle University which included Prof Paul Burton, Dr Nick Booth, Dr Olly Butters, Dr Stuart Wheater, Prof Mike Martin, Prof Madeleine Murtagh and Mark Walsh. Key details of this workstream are outlined in the following sections.

Approach adopted and specific challenges

Requirements

To work as a trusted research environment the system was required to:

- Have a Health and Social Care Network (HSCN – previously known as N3) connection
- Be allowed to host NHS, and Social Care data, using appropriate IG safeguards

In addition to the above, the following business needs were also noted:

- Short set up time
- Ability to scale
- ISO 27001 certified
- IGToolkit level 2 or higher

Given these requirements the decision was made to procure a cloud-based solution which could offer a TRE as a service, instead of buying physical infrastructure and developing our own solution.

Procurement

With the above requirements in mind we used the G-cloud procurement service (www.digitalmarketplace.service.gov.uk) as this ensured an off the shelf solution that we could readily build upon. An additional constraint added to our procurement decision was the requirement to also have a Health Information Exchange included (see below). Given these requirements the only available company offering a solution via the G-cloud was AIMES (www.aimes.uk).

TRE v1 delivery

AIMES provided us with their standard Trusted Research Environment, which was a replica of environment supplied to Durham University for their care pathway project. In essence this was a Windows 10 remote desktop instances (a single Windows remote desktop being called a VDI) housed in a secure environment (the TRE). On top of their standard software we installed some additional tools (see software section below), and made DataSHIELD available (see below). We had two VDIs which were tied to two specific members of the team, with provision to have more VDIs made available to others in the TRE if requested.

Connecting to the TRE

The connection process was via a Virtual Private Network (VPN) connection and multi-factor authentication. The instructions for the VPN were written for windows, meaning the Linux user in the team spent a long time troubleshooting the settings. Moreover, the specific protocol used for connection is no longer available in the main Ubuntu distribution. The multi-factor authentication for the VPN was done with a username and password followed by a prompt from a mobile phone app called Duo Mobile (www.duo.com).

Once the initial set up was done the VPN connection process worked well on both Linux and Windows. The major downside to the VPN was that once a connection had been made, all network traffic was funnelled through the VPN (and therefore AIMES). This had the effect of blocking all other network traffic from the connecting computer, i.e. it was not possible to browse the web/read email/etc once the VPN was connected. This had an impact on troubleshooting program code and reading documentation etc.

Connecting out of the TRE

Once a user had logged on to their VDI in the TRE they can only access websites which are on the TRE URL whitelist. For this exploration and development phase we set this whitelist to be completely open - i.e. any website was allowed to be visited. This would not be the case for a live system and is not the case for the next generation version of the TRE (see below).

Data storage and partitioning

Like most enterprise Windows environments the file storage was based on network shares connected to the VDIs. This allowed project based access control lists to be set up which meant different users only had access to their relevant files.

Software

We specified a list of software for AIMES to install on each of the VDIs:

- R
- RStudio
- Microsoft Office
- Git client
- Python3
- Atom
- Notepad++
- SSH client

SFTP

An essential functionality of the TRE is the ability to transfer data in and out of the system. This was achieved using Secure File Transfer Protocol (SFTP). To enable this a user had to log into the VPN (see above) then they could transfer data from their local computer, via an SFTP client, into the TRE. One important limitation to this is that by connecting to the VPN all local network connections are severed, this includes connections to network drives, so only data that are physically stored on the hard drive of the connecting computer could be transferred in. This is a limiting factor if large amounts of data needs to be transferred, or if the governance on the data is such that it cannot be stored locally.

v2 TRE delivery

The delivery of version two (next generation) of the AIMES TRE conferred additional functionality to that offered by the first version, a key component being the removal of the VPN. As noted earlier the connection to the VPN took complete control of the user's local network connection, inhibiting them from e.g. reading email or browsing the web. Version 2 of the TRE does away with the requirement for a VPN, and instead uses either the client software VMWare Horizon Client, or allows connection with a web browser via HTML5. In both cases the connection process still requires multifactor authentication to verify the identity of user, and the connection itself is still encrypted over the wire.

In conjunction with the removal of the VPN connection the SFTP process was modified. Instead of requiring a stand-alone client on the user's machine, data can now be transferred in and out with a web-based SFTP client. Access to this web-based client requires the multi factor authentication outlined above. By moving away from the VPN connection, it now makes it possible to upload/download data using network shares on the user's computer.

Other issues and challenges

Long timescale associated with procurement process

It took a long time for the procurement process to take place. This was due to a combination of a difficulty in describing what we required (this was a project to explore what was available after all), and the local organisational bureaucracy that had to be navigated.

Information security adverse event

On delivery of the TRE to Newcastle University it was discovered that access to another project's data was granted inadvertently. The data was anonymised, so the impact on the data subjects was minimal, however it did highlight a flaw in AIMES' internal processes. These have now been amended to mitigate the risk of this happening again. Newcastle University's information security team, AIMES and the other project (Durham University) were all notified of the event, and the external data controller followed all relevant procedures.

Care pathway projects alignment with the TRE

Of the three main care pathway projects in the region one was already using AIMES' TRE (Durham), while the other two (SILVER and palliative care) were focussed on building integrations with primary and social care systems. This focus meant that their need for the regional TRE was markedly less. However, in both cases alternative analytic and integrative platforms were developed – one using a third separate AIMES TRE. These other CPPs therefore provided a range of useful prototyping insights which fed into our broader understanding of how to set up an integrative health informatics infrastructure and were thus still able to feed into the learning and experience leading to the impact of CHC-NENC. (see Impact and sustainability of work on TRE and HIE, below).

TRE resource usage

One point of the TRE that was frustrating was its lack of flexibility with the VDIs not being able to be used by different people and being turned on all the time. This means that a VDI is turned on (and charged for) but not used for the vast majority of the time.

Issue notification and resolution

AIMES use a ticketing system called Autotask (autotask.net). We found this a little difficult to use at times, with some issues taking longer than we would hope to be resolved.

Health Information Exchange (HIE)

Goal

To set up a third-party Health Information Exchange as a proof of concept to help the direction of any local/regional procurement of future HIEs. There was also specific interest in using the HIE as a vehicle for integrating FHIR with DataSHIELD.

Approach adopted and specific challenges

Procurement

As noted above, the only TRE-HIE pairing available through the G-cloud was with AIMES, who provide a Tiani-Spirit HIE via a subcontracting service.

Exploration/learning

From the very beginning it was difficult to explore and understand the HIE, we were not clear which components were actually installed in our instance. It took a long time to receive documentation, which was not very helpful once we had it.

Integrating with data sources

The whole point of a third-party HIE is that it connects other data sources together. We wanted to demonstrate this connection by setting up an external data source to the HIE. This was not achieved. A significant cause of this was the difference in working practices between commercial companies and us as a university project. Where we would ask 'can you suggest a simple data source software we can connect to the HIE?', or 'would this standards-based software be easy to connect to the HIE?', they would say 'tell us exactly the specification of the software you have installed and we will provide a quote to integrate with it'. This made demonstration of the HIE connected to other data sources impossible, this includes a demonstration of FHIR integration. This was important learning that fed into the later design and specification of a regional HIE – initially as part of the NENC LHCRE bid.

HIE API

We managed to connect to the HIE REST API using their bespoke functions. Using this we demonstrated adding patients and querying the HIE for the patients. Note that this was utilising the HIE's internal data store – there was no external connection to any other services.

Impact and sustainability of work on TRE and HIE

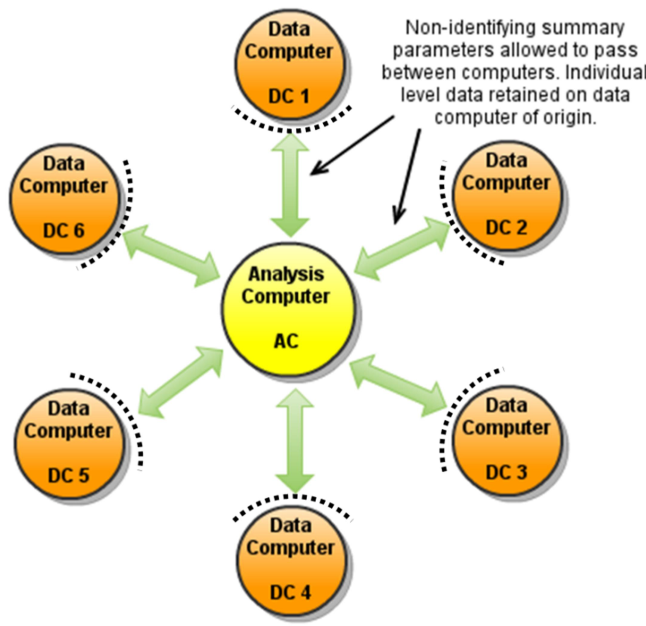
As a direct complement to the practical experience of setting up, obtaining governance permissions and actually using a cloud-based TRE presented by the CPP work led by Durham University, the extensive technical and methodological prototyping work based on the ARCHIE subproject and Newcastle University's sandpit TRE/HIE made an invaluable contribution to the ongoing strategic development of health care informatics in the NENC region. This therefore reflects a major impact of CHC-NENC and its ARCHIE subproject: (1) It informed and guided ongoing development of the Great North Care Record (GNCR) under the leadership of Joe McDonald who jointly headed up both the GNCR project and CHC NENC. (2) It directly contributed to strategic planning and design leading up to development of the NENC regional bid to the NHS England LHCRE program. This bid was successful in wave two, but because of concerns based on regional/local healthcare politics, the funding contract was unfortunately not signed before the LHCRE program was cut back when Boris Johnson became Prime Minister. In practice this meant that development of the TRE, HIE and patient portal were ultimately funded by routes other than LHCRE. Nevertheless, the CHC-driven experiential learning that fed into the LHCRE bid was still of value in informing the separate projects that funded the individual components. (3) The Durham University TRE has now evolved into an NENC regional TRE with additional components and capabilities based on work undertaken under ARCHIE. (4) The existence of this regional TRE was pivotal to the successful bid for the NIHR NENC ARC project and it now services a number of the analytic needs of that project. (5) It was also central to the recent success of the Northern Better Care Partnership HDRUK project led across six Northern Universities by Prof Munir Pirmohamed and Prof Paula Williamson (Liverpool University & CHC North West Coast). One of three primary work packages under this broader HDRUK program was the Digital Care Homes project which focuses on enhancing information utilisation in NENC care homes and is also critically dependent on the regional TRE. This care homes project is led by Mr Graham King (CIO, Newcastle upon Tyne Hospitals Trust) and Prof Suzanne Mason (Sheffield University & CHC Yorkshire), with Dr Camila Caiado, Prof Paul Burton and Prof Jo Knight (Durham University, Newcastle University, Lancaster University & CHC NENC and North West Coast) playing key roles in the development of methodology and technology underpinning the Care Homes project. (6) The data underpinning a number of urgent covid-related projects which are currently being set up in the North East will all sit on the regional TRE.

Recognising the infrastructural value of the regional TRE created under CHC-NENC all six NENC universities have collectively agreed to provide ongoing baseline support funding (£20,000 per university per annum) that permits them to "buy into" the TRE facilities and they will now seek additional funding to support individual projects from the grant monies they apply for to support those projects. This provides what we believe will be a workable scalable model for sustainability going forwards. This would have been impossible without the investment in CHC-NENC.

DataSHIELD

Background

DataSHIELD (www.datashield.ac.uk, www.github.com/datashield) is an innovative software tool enabling secure remote analysis (or joint, parallelized analysis) of individual-person-data (microdata) from one or several data-sources simultaneously. Security is underpinned by preventing access to, or even visualisation of, the individual-level data themselves and by proactively blocking potentially disclosive analytic output. By avoiding the physical sharing of microdata, DataSHIELD can mitigate governance and intellectual property (IP) concerns that otherwise constrain data-sharing. It also circumvents the risk that when a data-set is *physically* shared with a third-party, its original custodian(s) will lose control over its ultimate fate and it could end up being copied to a jurisdiction with imperfect governance.



Commands are sent from a central analysis computer (AC) to several data computers (DCs) that store the data to be co-analysed. Each DC is located at one of the studies/institutions contributing data to the analysis. The data sets are analysed simultaneously but in parallel. The separate parallelized analyses are linked by non-disclosive summary statistics and commands that are transmitted back and forth between the DCs and the AC. Technical implementation of DataSHIELD is based entirely on open-source freeware. It employs a specially modified R statistical environment linked to an Opal database deployed behind the computer firewall of each DC. Analysis is then controlled through a standard R environment at the AC. DataSHIELD is most often configured to carry out a – typically fully-efficient – analysis that is mathematically equivalent to placing all data from all studies in one central database and analysing them all together (with centre-effects where required). Alternatively, it can be set up for study-level meta-analysis: estimates and standard errors are derived independently from each study and are subject to centralized random effects meta-analysis at the AC.

The international team leading DataSHIELD development is the same academic team that undertook the overarching program of technical and infrastructural development under CHC-NENC/ARCHIE. This team is headed up by Paul Burton as PI of the DataSHIELD project and lead architectural and informatics roles are played by Olly Butters and Stuart Wheater.

Prior to CHC the DataSHIELD project had focussed almost on the development of secure and federated analysis methods for academic research projects, particularly those working with large cohort studies. However, based on initial work in CHC, it rapidly became clear that there were potentially important uses of DataSHIELD in the analysis of health service data.

The relevance of these uses was later emphasised by the introduction of GDPR in May 2018: specifically, the historic reliance on using pseudonymisation as a sufficient way to make potentially sensitive health service data 'safe' for external analysis was formally recognised as often being inadequate. Whether clinically or personally sensitive information can be inferred from a particular data set containing healthcare information is determined by many different factors and conventional pseudonymisation on its own provides little reassurance that disclosure of identity or of sensitive information cannot, in principle, be possible. This means that if we are to provide transparency to data subjects under the GDPR, we ideally need to be able to add an additional layer of inferential security to sit on top of the data, particularly linked data, whenever those data are viewed as being particularly sensitive or at particular risk of information disclosure.

The need to consider implementing a supplementary security layer was highlighted as part of the critical thinking and risk assessment under the evolving NENC CHC's SILVER and Palliative Care CPPs (see CPP reports). The presence of such a layer opens up the possibility of automatically taking data from a sensitive data environment (e.g. an HIE) holding named-data in an HSCN-compliant setting, pseudonymising it – and, if required, linking it – and then transferring it to a new analysis repository that still sits behind whatever firewalls are seen as necessary by the data custodian. In this new repository, the data can be analysed behind a DataSHIELD front-end thus preventing users from seeing or copying the individual-level data while allowing fully efficient analysis (as if one had full access to the individual-level data). At the same time, intrinsic DataSHIELD disclosure controls - tailored and locked-down by the data custodian – are used to mitigate the risk of disclosure via inferential analytic attack.

As it is based on what is now one of the most widely used analytic platforms for health care analytics (the R environment), is being used commonly in large-scale health research settings and is entirely based on open source freeware, DataSHIELD is without doubt a strong candidate to take on a substantive role in providing a supplementary security layer.

Goals under CHC

(1) DataSHIELD functionality to be added into the TRE as an additional goal of the ARCHIE subproject.

Work with the EU H2020 project EUCAN-Connect (eucanconnect.eu), which is the main source of resources to develop DataSHIELD to:

(2) Complete all code checking and other QA for all basic DataSHIELD functions and set up continuous testing environment as a precursor to releasing new version of DataSHIELD.

(3) Ensure the next major release of DataSHIELD (version 5.0) proceeded smoothly, and could be achieved before the end of formal CHC-NENC funding (30 September 2019). This goal was achieved in mid-September so CHC is now jointly credited as enabling release of DataSHIELD version 5: which has proven to be very successful.

(4) Enhance international understanding and uptake of DataSHIELD by running international workshops in Newcastle upon Tyne in September 2018 and 2019. Both meetings were very successful, and CHC was badged as a key sponsor of both meetings. It has also been acknowledged on all slide sets when DataSHIELD has been presented both nationally and internationally

Installation in the TRE

Version 4 of DataSHIELD was installed on a Linux virtual machine in the TRE. It was configured so one of our users was the full administrator of the VM and could access it via SSH and the web interface. The other user was specified as just a normal user, i.e. they had fewer access privileges as would be expected in a real-world environment. We demonstrated the system worked, and could therefore be used in an HSCN setting on sensitive data if required.

Major DataSHIELD release

In order to make DataSHIELD easier to use we have added significant functionality to it, and we have released version 5 (see github.com/datashield/dsBase/releases/tag/5.0.0 and github.com/datashield/dsBaseClient/releases/tag/5.0.0). As part of this work we also upgraded the version of DataSHIELD in the TRE to v5.0.

FHIR integration

The frustrations with the HIE meant that we had no FHIR based service to build against, so we failed to achieve DataSHIELD FHIR integration.

Overarching difficulties

One particular issue we encountered in the development of what was needed was a reluctance of many IT professionals in the region to consider deploying specific solutions outside of their familiarity. The most usual data sharing solutions in the NHS are based on enterprise-level architecture, with specific accommodation for local sharing provided by peer-peer, often bespoke or system to system sharing solutions. Federation Architecture solutions deploying infrastructural and middleware solutions common to multiple organisations are less common, and open up complicating issues of joint ownership of infrastructure, and difficult problems of information sharing (also complicated by new GDPR regulations in the middle of the project. This work under CHC-NENC has allowed us to recognise the need to carefully address this understandable perspective of NHS IT professionals and we now take a much more cautious and better explained approach to exploring the potential use of DataSHIELD in health service settings.

Impact and sustainability of work on DataSHIELD

The goal to develop and apply DataSHIELD to health service data advanced greatly during the CHC-NENC project. DataSHIELD is being used increasingly more widely in international medical research consortia and given work with international collaborators and in partnership with AIMES (see below) we are now confident that it can be rolled out in parallel for uses in HSCN settings. For example, we have developed close links with the MIRACUM project which is developing and applying DataSHIELD to data from four hospitals in Freiburg, Germany. In addition, we are working with the Conception IMI project to use DataSHIELD with data from child health services across Europe. Both of these projects have raised useful suggestions for improvements to reflect health service needs that we plan to incorporate into the architecture of DataSHIELD. These include: (1) using ‘pull’ rather than ‘push’ mechanisms to transmit analytic commands into servers holding health service data (work being undertaken by MIRACUM project); and (2) the idea of enabling analysis to be entirely controlled by a script file that can be written ahead of time, formally agreed to by all data controllers, and then locked down until it is run so that no modifications can be made. This proposal was put forward by the IMI Conception project. We are also working actively with the Born in Bradford project (Prof John Wright, CHC-Yorkshire) to enable secure DataSHIELD-based analysis of data they currently hold on servers in the Bradford NHS Foundation Trust.

Finally, we are working with AIMES to seek funding under appropriate National Innovation Grants to formally integrate DataSHIELD into the commercial version of their next generation TRE.