Dementias Platform UK Imaging Informatics (DPUK-II)

Imaging Informatics Solution Overview

Document Reference: DPUKII-ARC-005

Author(s): Lars Engstrom, Clare Mackay

Date: 16/03/2017

Version: 1.00

Status: Issued

1 Introduction & Background

The MRC Dementias Platform UK (DPUK; www.dementiasplatform.uk) is a multi-million pound public-private partnership, developed and led by the Medical Research Council (MRC), to accelerate progress in, and open up, dementias research. Directed by Professor John Gallacher at the University of Oxford, together with an executive team of investigators drawn from universities around the UK, DPUK's core aims are early detection, improved treatment and ultimately, prevention, of dementias.

DPUK seeks to bring together data from cohorts around the UK to create a platform for 'big data' epidemiology and experimental medicine. Many of DPUK's cohorts contain neuroimaging data, which is often stored separately from the main cohort data due to its size and complexity. The ambition for the DPUK Imaging Informatics programme is to develop a national image sharing and analysis platform for dementia research. The first 'capital' phase of the programme (Feb 2015- March 2016) was funded by the MRC as part of DPUK's CRI award and involved procuring hardware and software for the platform. During this phase we assembled the team, selected and developed the software solution, developed a 'Hub and Nodes' model and procured the hardware. A scope change to the original brief involved developing a specialist node for the UK Biobank to facilitate access to neuroimaging data for DPUK users. The software solution is based on the XNAT product, initially developed at Washington University in St Louis, and adapted by Radiologics Inc. to meet the specific requirements defined by the DPUK Imaging Informatics programme. The second 'implementation' phase of the project was funded by the NIHR and a Wellcome Trust Institutional Strategic Support Fund award (CEM), and involves transitioning the central hub, central node and the UKB node to live service, integrating with the DPUK portal for authentication, and setting up local sites and connecting them to the central hub. The third 'engagement' phase of the programme will involve an exercise to encourage and facilitate cohort custodians to upload data to the platform, and to encourage users to request access to data. The forth 'research tools' phase of the programme will develop a DPUK 'app store' for validated analysis pipelines to be made available.

2   Team

The collaborative team is co-led by Clare Mackay (U of Oxford) and Sebastien Ourselin (UCL) and comprises a Programme Manager/Solutions Architect (Lars Engstrom), and DevOps (Matt South) in Oxford, and substantial expert input from colleagues at UCL (Dave Cash, Ben Yvernault, Miklos Espak, Daniel Beasley, & Marc Modat). We rely on essential contributions from collaborators at each of the DPUK sites, including Guy Williams (Cambridge), Dominic Job (Edinburgh), James Cole (Imperial), Marcus Kaiser (Newcastle), Paul Marsden (KCL), Warren Mittoo (Manchester) and Cyril Charron (Cardiff). Our software partners are the XNAT/Radiologics team, led by Dan Marcus with major contributions from Tim Olsen, Patrick Clough, James Dickson and Will Horton. DPUK imaging informatics is part of both the informatics (lead: Simon Lovestone) and imaging (lead: Paul Matthews) networks of DPUK.

3   Solution Overview & Conceptual Model

The DPUK Imaging Informatics solution specified and delivered as part phase 1 is a Hub and Node model with an additional free-standing node to enable access to the UK Biobank (UKB) image data store. The central hub is located at the Farr Institute at Swansea (alongside the DPUK portal). The functionality includes a cohort catalogue and participant browser that enables users to search against and request access to data. Each collaborating institute (site) has a local XNAT installation for their own Data Management use and the extensions made to the XNAT product under the programme enables researchers to 'freeze' a dataset and upload it to the hub to make it available for searches by registered DPUK users. The UK Biobank node comprises an XNAT interface with an adaptor to plug-in data direct from the UK Biobank image store, subject to approval of a UK Biobank data access request.

Figure 1 shows the conceptual model of the XNAT solution implemented and deployed under the DPUK Imaging Informatics programme.
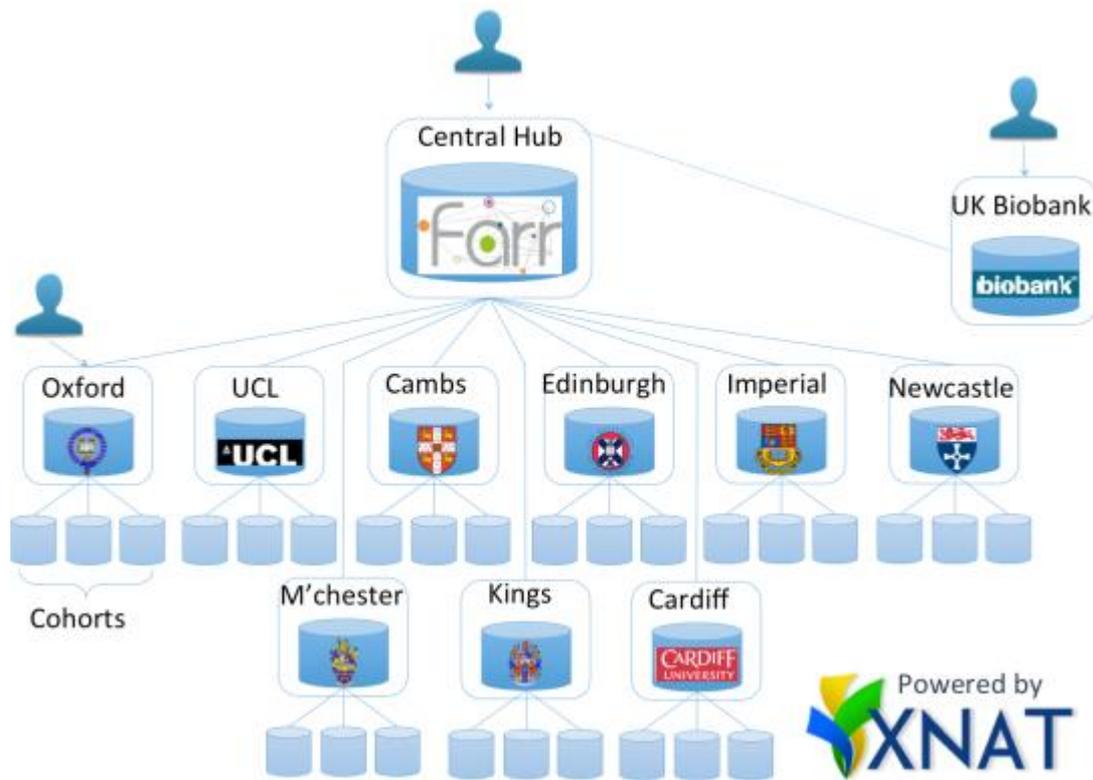
Figure 1: DPUK-II Conceptual Model

4   Implementation

This section outlines the implementation of the XNAT solution developed for the DPUK programme both from a software as well as an infrastructure perspective.
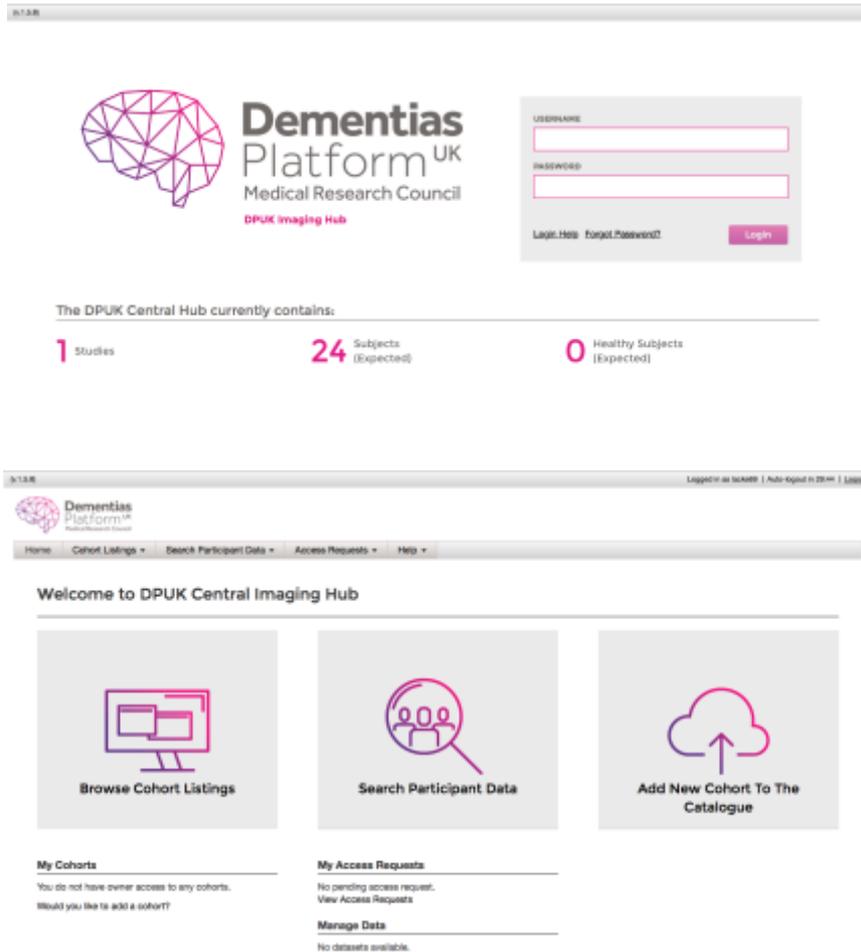
4.1   Central Hub

The main part of the solution developed under the DPUK Imaging Informatics umbrella for the DPUK Imaging Network is the Central Hub (https://hub.dpuk.org). The Hub is the location where all cohorts for the imaging network are listed and data is shared. The hub is Integrated with the DPUK Portal authorisation and authentication model and users access the hub to:

- Add new cohorts to the Cohort Catalogue
- Browse cohort listings
- Search participant data across cohorts with shared data from any collaborator

- Request access to subjects corresponding to a search query (inclusion / exclusion criteria)
- Provide approval of data access requests

The Central Hub is built on top of an extended version of the XNAT 1.6.5 code base. All projects and images are persisted in this XNAT instance and the Hub web front end provides a user-friendly search interface to the end-user. User authentication is managed by the DPUK portal user directory and presented to the hub via an LDAP interface.

Figure 2 shows screenshots from the Central Hub showing the login and landing page once a user has authenticated.

Figure 2: DPUK-II Central Hub Screenshots

4.2   Local nodes

XNAT nodes have been procured for each of the 9 collaborating DPUK imaging centres in the network (Oxford, UCL, Cambridge, Edinburgh, Imperial, KCL, Manchester and Cardiff). Each collaborating institution was, as part of the main grant, awarded an amount of funds based on their local infrastructure requirements to commission a hardware infrastructure to deploy the imaging node solution to. Node XNATs are similar to the open source version of XNAT 1.6.5 but with additional capability for creating data 'freezes' to upload to the Central Hub.

4.3   Central Node

The central node is an XNAT node that is hosted on the same hardware infrastructure as the Central Hub. It is not part of the hub solution itself, but instead treated as a "Local Node" hosted on the central infrastructure at Swansea. This node will have same capabilities as any of the other nodes deployed at each collaborating institution and is connected to the Hub to enable data uploads by using the freeze functionality provided. The rationale for standing up this node is to enable national studies, that are just starting to load data, the convenience of persisting the data in a central location straight away rather than keep data locally at any of the collaborating institutions XNAT nodes.

4.4   UKB Node

The UK Biobank is currently scanning 100K individuals, and DPUK has committed to rescan 10K of these individuals. This provides a neuroimaging dataset of an unprecedented scale, meaning that multiple downloads are problematic from a bandwidth perspective, and prohibitive for users without massive institutional computing resources. The UKB node is a specialist node that is based on the open-source version of XNAT 1.6.5. An adapter has been built to enable XNAT to access images from the UKB data store by using the exiting data access utilities developed by UKB for this purpose. This node provides features to upload and persist the "key" provided by UKB after a successful data access request to gain access to the images approved for use by a specific project. XNAT will provide a common user interface for users wanting to access the UKB image store and it is hiding the command line interface (CLI) from the end-users that otherwise must be used for this

purpose. The solution also allows users to execute pipelines against the image dataset they have been approved to use, meaning that the pipelines can be brought to the data rather than the data being brought to the pipeline.

4.5 Infrastructure

The DPUK imaging solution is deployed in several different physical locations. The local nodes are hosted at each collaborating institution's location of choice. The central hub and the central node is deployed and hosted at the Farr institute in Swansea while the UKB node is hosted at the Advanced Research Centre (ARC) at Begbroke, Oxford.

4.5.1 Locally at each institution

Local teams were given flexibility in relation to the hardware to fit in with local infrastructure, as long as it supports the pre-requisites of the solution developed. The central team from Oxford is able to support the collaborating institutions with their installations of the software solution as required.

4.5.2 Farr Institute in Swansea

The Farr Institute in Swansea is the hosting provider for the central hub and the central node infrastructure. This infrastructure is the key hosting component for the solution and it will be monitored by both the operational team at Swansea as well as the operational team directly associated with the imaging programme (this group is in the process of being setup).
The initial capital investment allocated to the imaging informatics programme enabled procurement of the following hardware resource for the central infrastructure.
At the FARR institute we have 512Gb RAM and 116 cores for application and compute backed by 200TB of ZFS storage that's mirrored in two locations on campus for resilience.

Figure 3 outlines on a high-level the current use of the central infrastructure resource allocated for the DPUK Imaging Informatics solution.
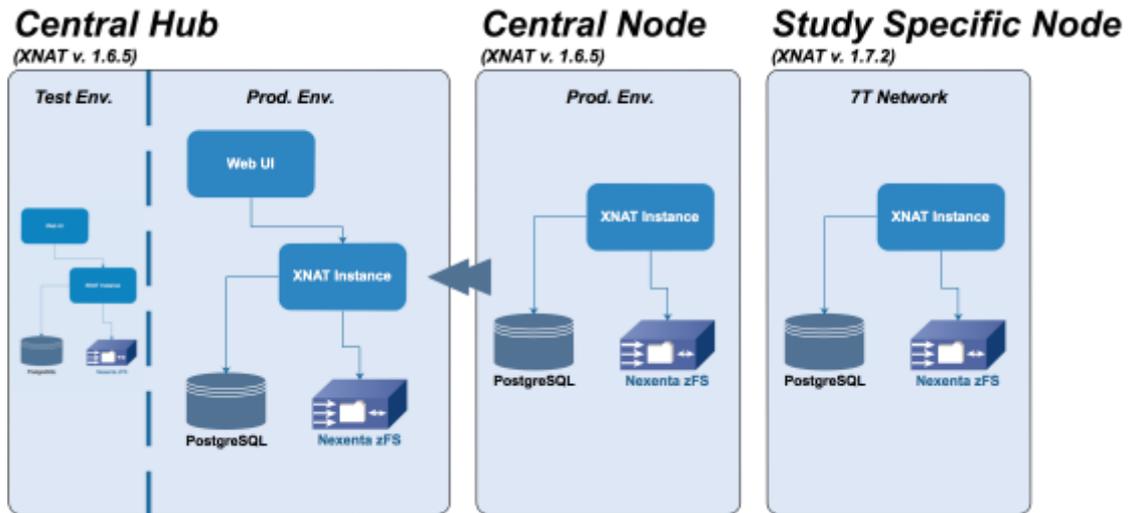
Figure 3: Farr Institute Deployment

4.5.3 ARC (Advanced Research Computing) at Begbroke, Oxford

Advanced Research Computing (ARC) is a central resource available to any Oxford University researcher who needs high performance computing (HPC). Systems and storage are located in a secured, purpose-built data centre at Begbroke.

The DPUK imaging informatics programme utilises ARC as the hosting capability for the UKB XNAT node infrastructure. The key rationale for using ARC for the UKB XNAT node is that the UKB solution also is hosted there and hence the solutions would be able to utilise the same internal networking in order to reduce latency and therefore improve performance.

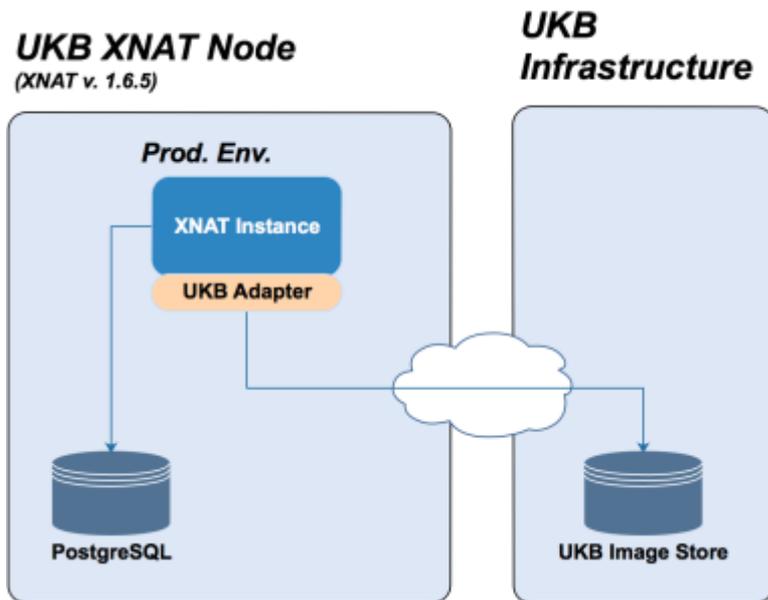At the ARC we have 596Gb RAM and 80 cores for application and compute backed by 100TB of ZFS storage.

Figure 4: ARC Data Centre Deployment

5 Current Activities (Feb 2017)

The team are nearing the end of the 'implementation' phase and are preparing for the 'engagement' and 'tool development' phases of the programme. The current focus is to disseminate the solution and transition all institutions into live service. Once that has been achieved the focus moves to the usage of the platform by making sure that cohorts are uploaded and shared. In addition, a number of new multi-centre studies have specified the DPUK imaging infrastructure for their data management plans (MRC UK7T partnership, MRC PET-MR network partnership, MRC/NIHR Deep and Frequent Phenotyping Project). The team are supporting each of these projects.

The operational and sustainability model for DPUK imaging informatics has not yet been established. We are working with our partners to develop a long-term model and exploring various options for funding.