

Data Portal				
Start date: 1 Aug 2014			Completion date: 30 Jun 2020	
<p><b>Overall work package objectives:</b></p> <p>The integration of data across cohorts and single portal data access. This involves the collection, integration, curation and analysis of the substantial volumes of data obtained during the course of project. The final goal is to integrate the cohort data and UK-wide data linkage into a DPUK informatics portal.</p> <ol style="list-style-type: none"> <li>1. UK-wide record linkage, understanding data capability</li> <li>2. Integration of cohort research data and metadata</li> <li>3. Informatics portal established</li> <li>4. Informatics portal dissemination</li> </ol>				
Deliverables	Milestones	Milestone deadline	Work package dependencies	Person(s) responsible
<b>Objective 1:</b>				
D1.1 Prioritising of cohorts for data transfer and facilitation of data transfer	M1.1.1 Readiness of data for transfer established for priority cohorts: Airwave, BDR, Biobank, Cam-CAN, CamPalGN, CAPS, MRC CFAS, CFAS II, Cygnus, ELSA, GS: SFHS, ICICLE-PD, LBC1936, OPDC Discovery, PICNICS, PREVENT, Whitehall II	M1.1.1 Complete	None	CO
	M1.1.2 Facilitation of data transfer initiated	M1.1.2 Complete		
D1.2 Accurate and up-to-date information available to researchers	M1.2.1 Comparison tool created.	M1.2.1 Complete	None	JB
D1.3 Analysis of cohort metadata for DPUK studies	M1.3.1 Primary metadata study to analyse the spread of data available within DPUK	M1.3.1 Complete	None	Oxford analyst with support from Oxford and Swansea (JG, CO)
	M1.3.2 Comparative study with other big data consortia	M1.3.2 Complete		
	M1.3.3 Further analysis of cohort metadata to establish data capability of DPUK with specific attention on variable collection	M1.3.3 Complete		
<b>Objective 2:</b>				
D2.1 Utilise Gap Analysis and contact with Cohort IT Leads to identify cohorts with fewer difficulties Identify cost effective solutions with early adopters Migration of data initiated	M2.1.1 Support options and cost recovery policy in place	M2.1.1 Complete	None	CO, JB, ST
D2.2 Migration of data completed – ongoing development as cohort partners are constantly joining and some cohorts have bespoke sharing permutations	M2.2.1 First data set migrated	M2.2.1 Complete	None	CO, JB, ST
	M2.2.2 Five prioritised cohorts' data on the platform	M2.2.2 Complete		
	M2.2.3 12 Prioritised cohorts' data on the platform	M2.2.3 Complete		
	M2.2.4 16 Prioritised cohorts' data on the platform – change of prioritisation based on cohort circumstances	M2.2.4 Complete		
	M2.2.5 20 Cohorts data on the platform	M2.2.5 Complete		
	M2.2.6 26 Cohorts data on the platform	M2.2.6 Complete		

WP 2 Project report

	M2.2.7 32 Cohorts data on the platform	M2.2.7 Complete		
D2.3 Integrate cohorts fully into DPUK	M2.3.1 Meet with cohorts to explain latest on DPUK and facilitate cohort data transfer	M2.3.1 Complete	None	CO, JB, ST
	M2.3.2 Engage at four cohort events for 2016 and 2017 – forums, seminars, conferences	M2.3.2 Complete		
	M2.3.3 Dissemination of completed DPUK literature for cohort consumption	M2.3.3 Complete		
	M2.3.4 Open up consortium services to cohort groups and individual cohorts for rapid scientific engagement	M2.3.4 Complete		
	M2.3.5 Cohort data services to be initiated	M2.3.5 Complete		
D2.4 Support and facilitate DPUK scientific analysis studies	M2.4.1 5 studies supported and ongoing in Data Portal	M2.4.1 Complete	None	CO, JB, ST SB, MN
	M2.4.2 10 studies supported and ongoing in Data Portal	M2.4.2 Complete		
	M2.4.3 15 studies supported and ongoing in Data Portal	M2.4.3 Complete		
	M2.4.4 35 studies supported and ongoing in Data Portal	M2.4.4 Complete		
	M2.4.5 45 studies supported and ongoing in Data Portal	M2.4.5 Complete		
	M2.4.6 55 studies supported and ongoing in Data Portal	M2.4.6 Complete		
	M2.4.7 First study publications to be facilitated	M2.4.7 Complete		
	M2.4.8 65 studies supported and ongoing in Data Portal	M2.4.8 Complete		
	M2.4.9 100 studies supported and ongoing in Data Portal	M2.4.9 Jun 2019		
<b>Objective 3:</b>				
D3.1 Establish DPUK UKSeRP instance and governance structure	M3.1.1 DPUK UKSeRP instance and governance structure in place	M3.1.1 Complete	None	JB, ST
D3.2 Establish DPUK data portal front-end	M3.2.1 Create DPUK portal website	M3.2.1 Complete		JB, ST, CO
	M3.2.2 Build expression of interest and full application pages for study applications	M3.2.2 Complete		
	M3.2.4 Create policy documents for DPUK portal and procedures for each product (data access, data management)	M3.2.3 Complete		
	M3.2.4 Create policy documents for DPUK portal and procedures for each product (data access, data management)	M3.2.4 Complete		
	M3.2.5 Launch application phases and documentation	M3.2.5 Complete		
	M3.2.6 Fully operational study application system	M3.2.6 Complete		
	M3.2.7 Active links to analysis platform and all other DPUK platforms (imaging, genetic etc)	M3.2.7 Complete		
D3.3 Pilot testing and provision of cohorts to the analysis platform	M3.3.1 First data set tested on UKSeRP infrastructure	M3.3.1 Complete		JB, ST, CO
	M3.3.2 First data set available for analysis	M3.3.2 Complete		
D3.4 Instance of EMIF fingerprinter, DPUK fingerprinter, EMIF variable/participant selection tools and TranSmart to be housed on platform - UNDER REVIEW	M3.4.1 EMIF fingerprinter installed	M3.4.1 Complete	JB, ST, CO	
	M3.4.2 DPUK fingerprinter installed and populated	M3.4.2 Complete		
	M3.4.3 EMIF VST/PST tools installed	M3.4.3 Complete		
	M3.4.4 TranSmart installed	M3.4.4 Closed		

WP 2 Project report

D3.5 Oxford/EMIF staff will be given access to the raw data within the portal uploaded by Swansea to harmonise to the TranSmart data model – CLOSED	M3.5.1 TranSmart installed	M3.5.1 Closed		JB, ST, CO
D3.6 Instances of static and interactive metadata tooling to be installed – Swansea build and install	M3.6.1 Project proposal raised to management and accepted based on data scoping exercise and initial tooling design feedback	M3.6.1 Complete		ST, CO
	M3.6.2 Install of primary components for tooling deployment	M3.6.2 Complete		
	M3.6.3 Deployment of static 3-tiers of information on DPUK Portal website (Cohort Landing Page, Cohort Overview, Data Dictionaries)	M3.6.3 Complete		
	M3.6.4 Swansea metadata tooling to be generated and demonstrated intra-DPUK	M3.6.4 Complete		
	M3.6.5 Swansea tooling to be made live	M3.6.5 Complete		
	M3.6.6 Continual development of tooling according to market research	M3.6.6 Complete		
	M3.6.7 Complete redesign and rework of Cohort Matrix and Cohort Directory – interactive and journey through from matrix to directory	M3.6.1 Complete		
	M3.6.8 Version 1 visualisation tooling installed	M3.6.2 Complete		
	M3.6.9 Install of GAAIN server tooling	M3.6.3 Complete		
D3.7 Link to genetics platform	M3.7.1 Genetics software installed in Swansea	M3.7.1 Complete	CO, ST (Swansea) JF (Digital Health Labs) GM (Cardiff)	
	M3.7.2 Genetics platform remotely linked to Swansea infrastructure	M3.7.2 Complete		
	M3.7.3 First genetic data provisioned to platform	M3.7.3 Complete		
	M3.7.4 Genetic data analysis to begin	M3.7.4 Complete		
D3.8 Link to imaging platform	M3.8.1 Imaging hub established in Swansea	M3.8.1 Complete	CO, ST (Swansea) CM, MS (Oxford)	
	M3.8.2 Imaging nodes to be established across the UK	M3.8.2 Complete		
	M3.8.3 Imaging data to fully link with central hub	M3.8.3 Ongoing		
	M3.8.4 Imaging data to be received from nodes	M3.8.4 Complete		
	M3.8.5 Imaging hub to provision access for centralised data analysis	M3.8.5 Complete		
	M3.8.6 Imaging interoperability with general access procedure finalised	M3.8.6 Complete		
	M3.8.7 DPUK Common ID model concept trialled	M3.8.7 Complete		
	M3.8.8 DPUK Common ID model concept established	M3.8.8 Complete		
	M3.8.9 DPUK Imaging Desktop resource built	M3.8.9 Complete		
D3.9 Link to wearables platform	M3.9.1 Wearables platform software/hardware to be linked to Swansea infrastructure – ON HOLD	M3.9.1 Closed	JA, JC (Manchester) CO, ST (Swansea)	
	M3.9.2 Wearables data to be tested via remote link	M3.9.2 Complete		
	M3.9.3 Wearables data to be used as part of a DPUK study using data portal	M3.9.3 Complete		
D3.10 Link to EMIF-AD cohort platform – CLOSED	M3.10.1 EMIF-AD PST/VST tools to be installed on DPUK environment	M3.10.1 Complete	CO, ST (Swansea), JC (Manchester)	
	M3.10.2 EMIF-AD PST/VST tools to be integrated with DPUK-based data	M3.10.2 Complete		
	M3.10.3 EMIF-AD to be hosted in Swansea and linkable to DPUK via infrastructure and cover remote access for non-hosted data	M3.10.3 Closed		

WP 2 Project report

				), MVS, NH (Janssen)
D3.11 Interoperability of platforms displaying multi-modal data analysis in the Portal	M3.11.1 First multi-modal study ongoing in the Portal (Combination of genetics/imaging and clinical data)	M3.11.1 Complete	None	CO, ST, MS
	M3.11.2 Creation of genetics/omics desktop template	M3.11.2 Complete		
	M3.11.3 Creation of imaging desktop template	M3.11.3 Complete		
	M3.11.4 Link to HDR UK phonemics' initiative created	M3.11.4 Complete		
<b>Objective 4:</b>				
D4.1 Promote the work of DPUK and specifically the informatics portal at UK-wide events, such as conferences, seminars and via UK media	M4.1.1 Posters & representation at DPUK Conference 16	M4.1.1 Complete		CO, ST, RL (Swansea), JG, AM, SB, BS (Oxford)
	M4.1.2 Attendance at UK Biobank Conference 16	M4.1.2 Complete		
	M4.1.3 Posters, presentations & lead infrastructure staff representation at IPDLN Conference 16	M4.1.3 Complete		
	M4.1.4 Attend & represent at Informatics for Health 17	M4.1.4 Complete		
	M4.1.5 Lead position at DPUK Conference 17	M4.1.5 Complete		
	M4.1.6 Attend & represent at AAIC 17	M4.1.6 Complete		
	M4.1.7 Attend & represent at ARUK 18	M4.1.7 Complete		
	M4.1.8 Attend & represent as key speakers at DPUK Conference 18	M4.1.8 Complete		
	M4.1.9 Attend & represent at ADR 18	M4.1.9 Complete		
	M4.1.10 Attend & represent at IPDLN 18	M4.1.10 Complete		
	M4.1.11 Attend collaborative meetings with OBI, GAAIN, and CPAD	M4.1.11 Complete		
	M4.1.12 Attend second round of collaborative GAAIN/CPAD/OBI meetings – up to end of April 2019	M4.1.12 Complete		
<b>Updates on delivery against milestones since last report</b>				
<ul style="list-style-type: none"> <li><b>M.2.4.9 100 studies supported and ongoing in Data Portal</b></li> </ul> <p>79 DPUK study applications are currently being supported and have analysis ongoing in the Data Portal. A further 6 DPUK study applications were supported in the Data Portal and have now completed their analysis.</p> <p>In addition to the above 85 DPUK applications, received via the application form on the website, DPUK is currently supporting in the Portal:</p> <ul style="list-style-type: none"> <li>- 10 separate studies on Memento cohort study-specific datasets, by Memento researchers</li> <li>- 9 separate studies by DPUK Oxford researchers</li> </ul> <p>Therefore, 104 studies have been supported in the Data Portal, with 98 of these studies ongoing (excluding the 6 completed studies).</p> <p><b>Will you complete all your Milestones by June 2020?</b> <i>(If not explain how you plan to finish with dates and how you will report your outcomes to the MRC)</i></p> <p>This milestone can now be considered to have been met – 104 studies have been supported in the Data Portal.</p>				
<b>Team members funded (full or part-time) by DPUK</b> Ronan Lyons – Principal Investigator, Chris Orton – Data Project Manager, Justin Biddle – IT Developer, Laura North – Data Analyst, Mark Newbury – Administrator  <b>Team members involved with the project but <u>not</u> funded by DPUK</b> Simon Thompson – Chief Technical Officer, Swansea University Medical School  <b>Location(s):</b>			<b>Dependencies to and from other work packages, networks and themes</b>  None	

## WP 2 Project report

Swansea University	
<b>Summary of plan to deliver on outstanding work (with dates)</b> <ul style="list-style-type: none"><li>• There is no outstanding work to deliver upon based on the milestones listed above. WP2 has supported/is supporting its target of 100 studies since launch.</li><li>• We are continuing development of metadata tools, data ontologies, and platform developments for further omics and imaging integration to fit with the objectives listed above.</li></ul>	
<b>Risks</b> 1) Covid-19 situation has meant that some cohorts are postponing data access committee meetings and/or taking longer to provide decisions on DPUK study data access requests	<b>Mitigation</b> 1) DPUK Swansea as host of the Data Portal has been able to work with no disruption to service, despite all staff managing the Portal remotely. 2) Consequently, the team are in constant contact with applicants and cohorts and doing all they can to expedite decisions and access for researchers
<b>Lessons Learnt</b> <ul style="list-style-type: none"><li>• Cohorts are often overly busy and understaffed. It is necessary to frequently chase groups and cohorts generally appreciate the DPUK Swansea team chasing and reminding them of applications via email. It is necessary to take a strongly proactive approach to contacting cohorts regarding decisions and, where necessary, data upload in order to expedite access to the data for researchers.</li><li>• The Data Portal has been generally underfunded to support deliverables which aim to echo sister or equivalent platforms which have had much higher direct spending. Initiatives which only focus on one aspect of data improvement, such as CLOSER's effort to harmonise cohort data from longitudinal population studies, have received much higher direct funding than a platform, the Portal, which has created an end-to-end user experience and product for streamlined and efficient data analysis from over 40 partner cohorts. The larger investment in DPUK2 for the Portal/knowledge environment will address this learnt lesson.</li><li>• Deliverables should mirror the changing landscape of dementia research to enable refocusing of resources and delivery aims as necessary. The Data Portal for instance was created with the aim of being purely an epidemiology platform, and now houses omics and imaging capability in one system, rather than in a federated network, which came about organically. Deliverables, therefore, although will be overarching to begin with, will need to adapt and be amended to remain relevant throughout the project.</li><li>• Governance in general is a strong part of the DPUK Data Portal and we are setting standards in this area. It does however need to be a point that the MRC bear in mind, that cohorts and data sharing still get entangled in long-winded legal proceedings for data sharing, and the services of the Data Portal should be promoted more to show that the Portal has benefit to not only aid the wider third-party research community, but cohorts themselves.</li></ul>	
<b>Outcomes</b>	
<b>PUBLICATIONS</b>	
<b>Published</b> <ul style="list-style-type: none"><li>• Orton C., et al. (2018). <b>Dementias Platform UK (DPUK) Data Portal - supporting multi-modal data analysis, data linkage and real-world outcomes.</b> International Journal of Population Data Science,</li><li>• North L., et al. (2018). <b>Data Services for Cohort Studies: Increasing the impact of existing research studies and epidemiological readiness.</b> International Journal of Population Data Science,</li><li>• Gallacher J, (2016). <b>What's the good of a science platform?</b> Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 374 (2083),</li><li>• Bauermeister S. et al. <b>The Dementias Platform UK (DPUK) Data Portal (2020)</b> <a href="https://link.springer.com/article/10.1007/s10654-020-00633-4?wt_mc=Internal.Event.1.SEM.ArticleAuthorOnlineFirst">https://link.springer.com/article/10.1007/s10654-020-00633-4?wt_mc=Internal.Event.1.SEM.ArticleAuthorOnlineFirst</a></li></ul>	
<b>OTHER OUTPUTS</b>	
<ul style="list-style-type: none"><li>• DPUK Data Portal Analysis Environment</li></ul>	

## WP 2 Project report

- DPUK Data Portal website and tools – <https://portal.dementiasplatform.uk>
- DPUK Data Portal: Data Resource Profile – <https://www.biorxiv.org/content/10.1101/582155v1>
- Portal website has been relaunched with far heavier traffic and use for information and applications for DPUK studies. Marketing campaign and discovery awards have contributed to this, as well as presence at international conferences between December 2017 and July 2020. The dissemination of information at conferences by means of demonstrations, posters, and talks on best practise within this research genre continued for the remainder of the grant at other established events alongside further user guides for the DPUK Data Portal (now completed and online).
- DPUK Data Portal and DPUK contribution to the following dataset (SAIL Dementia eCohort): <https://ijpds.org/article/view/1121>

### **PROTOCOLS & PROCEDURES**

- Established data access procedure following relaunch of DPUK Data Portal site in December 2017

### **ENGAGEMENT ACTIVITIES**

#### *2020*

- ADDI Project Working Group – Miami, February 2020
- HDR UK Scientific Strategy Board – January 2020

#### *2019*

- HDR UK Scientific Strategy Board – November 2020
- DPUK Datathon – November 2019, Swansea
- DPUK Datathon – October 2019, UEA
- ADDI Project Working Group – London, October 2019
- ADDI Project Working Group – Los Angeles, July 2019
- Orton C. AAIC 2019 – Portal poster presentation, Los Angeles, July 2019
- DPUK Datathon – May 2019, Exeter
- Orton C. – IASIS: Big Data for Precision Medicine, Plenary Meeting, March 2019
- Orton C. – EU Joint Programme - Neurodegenerative Disease Research, ADDITION Consortium Kick-off meeting, regarding the DPUK Data Portal supporting the research analysis, February 2019
- Orton C. – Health Data Research UK Site Meeting - January 2019

#### *2018*

- Bauermeister S – Oxford Dementia Research Day 2018 Poster: The MRC DPUK Data Portal: Bringing Researchers to Data: The Research Journey:
- Bauermeister S – DPUK Data Portal Workshop Presentation at the DPUK Annual Conference
- IPDLN 2018 DPUK Data Portal presentations (Chris Orton + Laura North)
- DPUK Datathon – October 2018, Turing Institute- 42 registered participants and users
- North L. – International Population Data Linkage Network Conference - Banff, Canada, September 2018 - Data Services presentation 2018
- Bauermeister S - ARUK Technical Workshop
- Orton C. – Global Alzheimer's Disease Data Innovative - Interoperability Working Group, Toronto, December 2018
- Orton C. – Nature Neuroscience Interoperability Meeting, Toronto, December 2018 - Data Portal presentation and panel appearance
- Orton C. – International Population Data Linkage Network Conference - Banff, Canada, September 2018 - Data Services presentation
- Orton C. – International Population Data Linkage Network Conference - Banff, Canada, September 2018, Data Portal presentation

## WP 2 Project report

- Orton C. – Administrative Data Research Network Conference - Belfast, June 2018
- Orton C. – N S Digital Collaboration Meeting
- Orton C. – DPUK Annual Conference - Data Portal presentation session
- Orton C. – Alzheimer's Research UK Conference - London - Poster Presentation
- Orton C. – DPUK Cohort Workshop
- Orton C. – Brains for Dementia Research Study Day
- Orton C. – Member of the governance, design and delivery strategy groups for the MS Society
- Orton C. – Participant at the International Cohorts Summit
- Orton C. – DPUK MRI-PET Imaging Network Launch
- Orton C. – Lecturing as part of Oxford University's MSc in Non-communicable diseases - neurodegenerative conditions with a focus on dementia
- Orton C. – Talk at a workshop with AMED on Ageing Society
- Orton C. – Datathon 2018 DFP Cohort?
- Lyons R – 2/10/18 - Technology Sector Symposium - Developing a UK Wide Health Data Research Platform
- Lyons R – 5/9/18 - Design Thinking Workshop
- Lyons R – 29-30/5/18 - International Clinical Phonemics' Workshop
- Lyons R – 1 /5/18 - Swansea University HDR UK & NHS Digital Data Workshop

### 2017

- Orton C. – Presentation at Database Resources for Large-Scale and Complex Data workshop, Engineers' House, Bristol, 21st June 2017
- Orton C. – Presentation on DPUK as a use case of UKSeRP at the Administrative Data Research Network Conference, Royal Surgeon's Hall, Edinburgh, 2017
- Orton C. – DPUK Data Portal Live Demonstration, Swansea University, at the DPUK Annual Scientific Meeting 2017
- Thompson s. – DPUK Infrastructure Presentation at the EMIF DPUK Joint Conference, February 2017
- Bauermeister S – Poster presentation: The MRC DPUK Data Portal: Bringing Researchers to Data: The Research Journey: Alzheimer's Research UK Conference 2017
- DPUK infrastructure twice presented at ADRN Conference 2017 2<sup>nd</sup> June
- DPUK Attendance as an exhibitor and featured session at AAIC 2017 19<sup>th</sup> July
- Orton C. – Anti-Microbial Resistance Databank collaboration meeting
- Orton C. – AAIC 2017 - Dementias Platform UK Showcase presentation
- Lyons R – Presentation to MRC Council - Centre for the Improvement of Population Health through E-records Research (CIPHER)

### 2016

- Orton C. – DPUK Cohort Collaboration Meeting
- Orton C. – DPUK Presentation at the International Population Data Linkage Network Conference, Swansea, August 2016
- Orton C. – DPUK Cohort Seminar
- Orton C. – Outline presentation of work package 2 of DPUK, the informatics stream, at the time focusing on the build and purpose of the DPUK Data Portal. The poster featured a description of the envisaged tools, data management and planned activity for the coming year.
- Lyons R – Big Data – next big thing: evaluation. Presentation to 1st NHS Wales Digital Health and Care Conference

**Please tell us the most successful outcome and what it means to dementia research**

## WP 2 Project report

DPUK provides a technical platform, which allows for a variety of virtual desktop sizes supporting multi-modal research into Dementia. To date, this has supported 104 studies. The platform is increasingly supported by a range of tools to help researchers explore metadata prior to application and continues to grow in popularity.

### **Project narrative**

The DPUK Portal has been set up to provide a secure, single-portal data access environment for Dementia research of data across cohorts. This has involved and continues to involve the collection, integration, curation and analysis of the substantial volumes of data obtained during the course of project. The final goal is to integrate the cohort data and UK-wide data linkage into a DPUK informatics portal. The four main project objectives were:

- UK-wide record linkage, understanding data capability
- Integration of cohort research data and metadata
- Informatics portal established
- Informatics portal dissemination

This has been hugely successful in creating a data repository resource and facilitating cross-cohort analyses within a world-leading secure analysis environment.

Initially, a DPUK governance structure was established along with the DPUK Portal website, which included the Cohort Directory and Cohort Matrix data comparison tools for researchers to explore and compare cohort data. Latterly, the Cohort Explorer was created for this purpose giving the Data Portal a data visualisation capability.

Cohort data transfer has been continually facilitated and we have built a network of 42 cohorts whom we work with closely and have agreed upon data sharing mechanisms, either directly to the Portal for onward sharing aided by the Portal operations team, or for upload to the Portal on a project-specific basis whose data is available for application from researchers accessing the Portal.

As of the 15th April 2020, 104 Dementia studies have been supported in the Portal, including multi-modal studies analysing genomics and imaging data, and involving the full range of cohorts who hold data on the DPUK Portal. 309 applicants have applied to access DPUK cohort data from 80 institutions in 19 countries.

We have developed metadata tools, data ontologies, and are making developments for further omics and imaging integration to fit with the objectives listed above.

We continue to look to develop into the emergent data science areas of omics and imaging research and have actively positioned the Portal to be ready for extra capability in these fields as DPUK1 ends and DPUK2 begins. Significant legal and governance challenges have been overcome, even with some yet to be fully answered, and DPUK and the Data Portal has created best practise for data sharing, management, and analysis within the dementia and neuroscience landscape.