

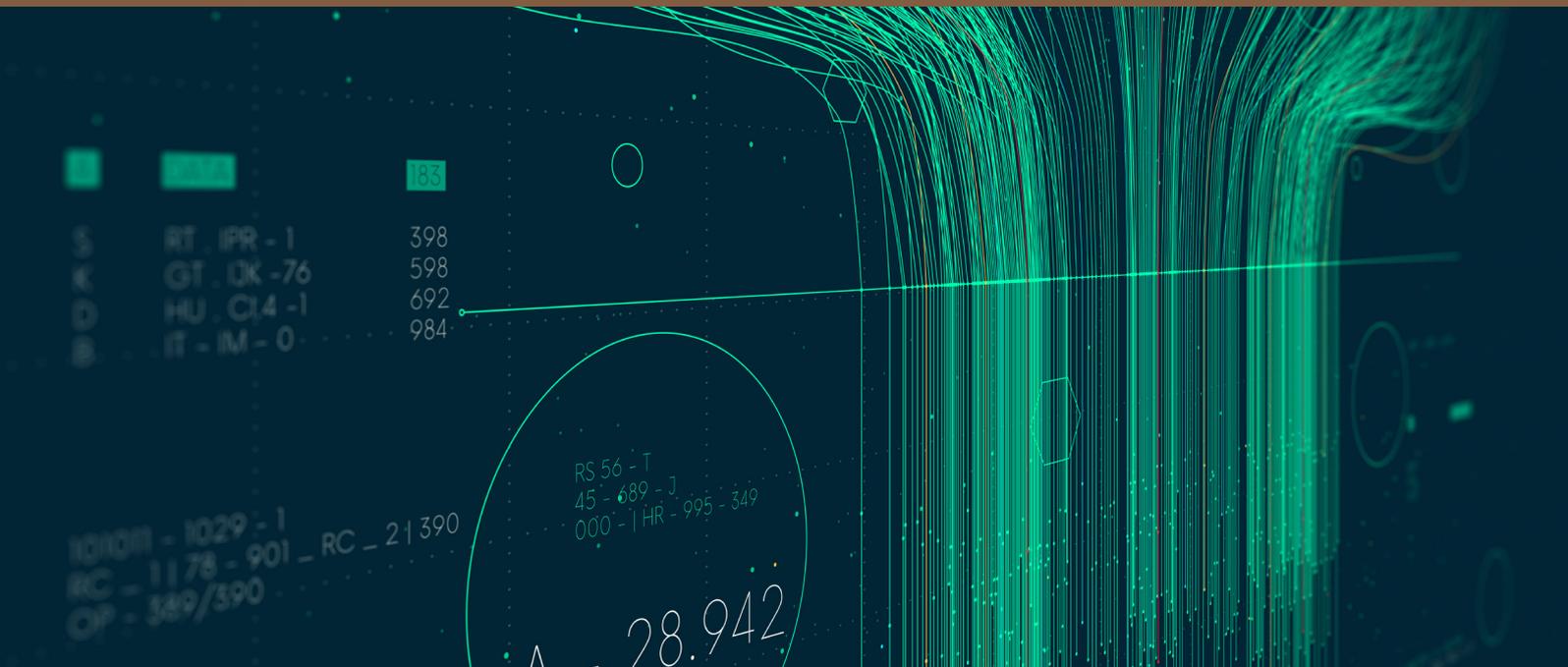


**World Dementia
Council** Leading the Global Action
Against Dementia

Global dialogue on data sharing for dementia research: Transcript

The dementia landscape project

9 June 2021



Organized in partnership with

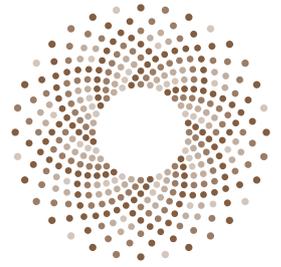


SageBionetworks



ADDI
Alzheimer's Disease
Data Initiative

Co-chairs



Dr Lara Mangravite

Lara Mangravite, PhD, is President of Sage Bionetworks. This organization is focused on the development and implementation of practices for large-scale collaborative biomedical research. Our work is centered on new approaches to scientific process that use open systems to enable community-based research regarding complex biomedical problems. Previously, Dr. Mangravite served as Director of the Systems Biology research group at Sage Bionetworks where she focused on the application of collaborative approaches to advance understanding of disease biology and treatment outcomes at a systems level with the overriding goal of improving clinical care. Dr. Mangravite obtained a BS in Physics from the Pennsylvania State University and a PhD in Pharmaceutical Chemistry from the University of California, San Francisco. She completed a postdoctoral fellowship in cardiovascular pharmacogenomics at the Children’s Hospital Oakland Research Institute.

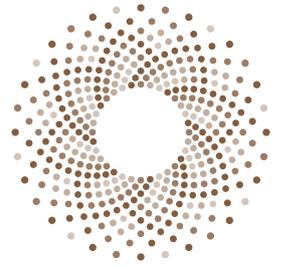


Dr Tetsu Maruyama

Tetsu Maruyama is the Executive Director of the ADDI, where he has the pleasure of working with an exceptional team to enable data relevant to Alzheimer’s and related dementias to reach their full potential. Prior to joining ADDI, Tetsu was the Chief Scientific Officer at the Dementia Discovery Fund, a unique venture capital fund focusing a total investment of \$350 million on creating new treatment paradigms for dementia. Before that he was head of Drug Discovery for Takeda Pharmaceuticals in Japan, after leading the GlaxoSmithKline Centre for Cognitive and Neurodegenerative Disorders in Singapore. He began his industry career at Merck Sharp and Dohme’s Neuroscience Research Center in the UK, after 15 years as an academic neuroscientist at Cardiff University and the University of Minnesota.



Speakers



Dr Angela Bradshaw

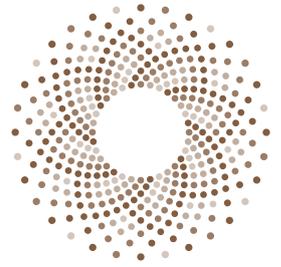
Dr. Angela Bradshaw is a Project Officer at Alzheimer Europe, an umbrella organisation of national Alzheimer's Associations with 39 members across 35 European countries. Alzheimer Europe aims to change perceptions, practice and policy, promoting a rights-based approach to dementia and working to make dementia a European priority. Angela collaborates on several EU-funded projects involving AI and health data, co-authoring the recent policy report "Data sharing in dementia research – the EU landscape". She obtained her PhD in vascular biology at the University of Cambridge, and worked as a lecturer at the University of Glasgow prior to joining Alzheimer Europe.



Professor Andrew Morris

Professor Andrew Morris became the inaugural Director of Health Data Research UK in August 2017. He is seconded from his position as Professor of Medicine, and Vice Principal of Data Science at the University of Edinburgh, having taken up position in August 2014. Prior to this Andrew was Dean of Medicine at the University of Dundee. Andrew was Chief Scientist at the Scottish Government Health Directorate (2012-2017) and has served and chaired numerous national and international grant committees and Governmental bodies. His research interests span informatics and chronic diseases.





Dr Suzana Petanceska

Dr Petanceska joined the National Institute on Aging in 2005, as a program director in the Division of Neuroscience. During her tenure at the NIA she has been overseeing and developing a number of research portfolios and programs in basic and translational research for Alzheimer's disease. She has been instrumental for the development of NIA's AD Translational Research program, the Epigenomics of AD portfolio and the Accelerated Medicines Partnership for AD (AMP AD) – Target Discovery and Preclinical Validation Project. Since 2012 she has been leading a number of NIA's strategic planning activities related to achieving the research goal of the National Plan to Address Alzheimer's: to prevent and treat AD by 2025.



Lenny Shallcross

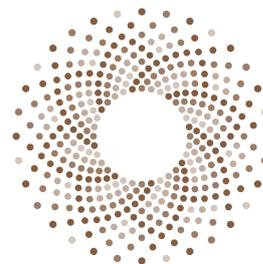
Lenny Shallcross is executive director at the World Dementia Council. Prior to that he was Head of Community Engagement leading programmes across the UK to establish Dementia Friendly Communities. This includes the Dementia Friends programme which is the biggest health social movement campaign delivered by 10,000 volunteers that have recruited 2 million individuals through a community, digital and corporate offer. Before working for Alzheimer's Society he worked in the UK government as a political adviser at the Department for Culture, Media and Sport and the Department of Health, as well as working in Parliament and for the Labour Party.



Global dialogue on data sharing for dementia research

Wednesday 9 June 2021

08:00 PDT San Francisco
10:00 CDT Chicago
11:00 EDT New York
16:00 BST London
17:00 CEST Central Europe



Discussion transcript



Lenny Shallcross
Executive director, World Dementia Council

Welcome everyone. I am Lenny Shallcross, Executive Director of the World Dementia Council. I realise many of you have participated in one of these global dialogues before, or another Council meeting but for those of you who have not, the World Dementia Council was established following the London dementia summit in 2013 hosted by the UK government as part of their G8 presidency.

The Council is chaired by Harry Johns, President and CEO of Alzheimer's Association (US). There are 24 individuals who are members of the Council. Alongside them there are a number of government members. As you know from the invitation and briefing note we sent, later in the year the Council will publish a report looking at the progress the international community has made on the commitments that were announced at the G8 summit. To help inform the report, we want to hear from experts around the world on different aspects of dementia policy. We will launch this report at an in person meeting here in London later in the year – if Covid allows!

This is the seventh dialogue we have held, previous conversations have been on biomarkers, clinical trials, technology among others. 350 global leaders have participated in the dialogues. We have two more conversations over the next few weeks one on the public policy questions around dementia and an ageing society and one on health system readiness.

Chat function

The chat function was available throughout the dialogue for participants to ask questions of presenters and to hold discussion amongst each other. It began just under 20 minutes into the event and is displayed below. It does not necessarily correspond with the adjacent transcript in this document.



Dr Maria Tome

It is the clinicians commitment!



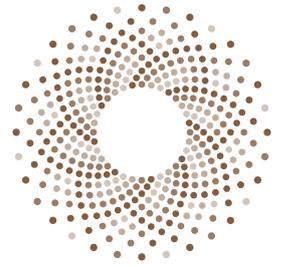
Dr Petr Holub

I would subscribe to that



Dr Maria Tome

The team and the NHS STAFF!!!



After this meeting we will produce a transcript of the meeting. For people who contribute live in this discussion we will check the transcript with you. We will also produce a collection of essays reflecting the flow of the discussion today and the issues raised. I would encourage you to share your thinking either live in the meeting or in the chat conversation. As you will know from the agenda we will kick off with short opening perspectives and then there is an open discussion. To contribute to the conversation just raise your hand in zoom. I am sure you all know this by now but to do that click on the reactions button and select raise your hand.

And with that bit of housekeeping done I would like to introduce and thank the co-chairs of today's workshop: Dr Lara Mangravite is President of Sage Bionetworks and Dr Tetsu Maruyama is Executive Director of the ADDI and also sits on the World Dementia Council. And with that I will now hand over to you Lara to start us off.



Dr Lara Mangravite
President, Sage Bionetworks

Great thanks. It is a pleasure to be here, and I am really excited to have this dialogue with all of you.

To start, the request that has come to us as we put this workshop together is to discuss all of our perspectives on where the state of data sharing is in dementia. There has been tremendous progress since 2013 when this programme started but there are also many challenges that we continue to face. So, I guess as we get started one point that I wanted to begin with was to remind us all that data sharing is a means to an end. And the end is scientific progress.

As we think about what we are using the data sharing processes to do, as with any other methodology, we need to think about how we fit it to the scientific goals at hand. So, some groups use data sharing primarily to enable transparency and reproducibility in research outcomes. Some groups are using it to enable broader access to essential research data resources. This is particularly important in dementia because of the characteristics of the disease and the sort of samples we can get at. And increasingly a lot of groups are using data sharing as a mechanism to aggregate and foster community or collaborative research.

And while all three of these things are important around data sharing, if you are developing a data sharing programme for a particular research community or a particular research problem you may make different decisions on how you set that up depending on what the research question or community you are trying to support and which of those components, as you rank them, are the most important.

And so, I think that what I have seen over the last couple of years is that that has very interestingly created a bit of a tension between generalised approaches we might take to data sharing and the policies we might put in place relative to the sort of individualised needs of the communities, the scientific community who are wanting to use the data and the scientific questions they are wanting to ask that are often particularly variable. So, I think that is a solvable problem, but it is one I see ahead of us right now.



Dr Elisabetta Vaudano

please have a look also at what data sharing at European level can achieve, again COVID focused from IMI project EHDEN <https://www.ehden.eu/covid19-study-athon/>



Inez Jabalpurwala

Can anyone share research on the extent to which existing data sets capture the full spectrum of diversity in our communities?



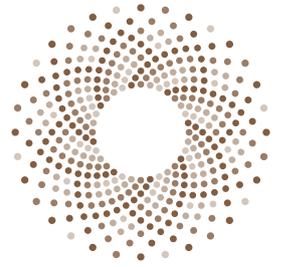
Professor Howard Bergman

How is Open Science and data sharing compatible with patent and commercial protection.



Dr Petr Holub

One major challenge is data quality assurance - which goes all the way to primary data generation and touches incentives



So that is why it is so important that we are starting this dialogue with three individuals who are coming at these questions with three probably overlapping but different perspectives. This will help highlight these sort of tensions, and help us come together thinking through where we might go.

So with that I would like to start by introducing our first speaker Andrew Morris. Andrew is the Director of Health Data Research UK and he is also a professor at the University of Edinburgh. Andrew take it away.

challenges and other aspects as well



Professor Andrew Morris

Director, Health Data Research UK and Professor of Medicine, Vice Principal of Data Science, University of Edinburgh



Professor Mayrann Martone

Have the NIH and UK projects engaged with the FAIR community? It has been my experience that there are some basic practices in FAIR, e.g., use of persistent identifiers, that the details of FAIR have not penetrated many of these projects.



Dr Heather Snyder

What are the tools, needs to inform and educate the global community? Is there an opportunity for us as a community to provide tools to the broader community - those who may have different types of data - on how to share? Where data is to share? What are data sharing considerations?

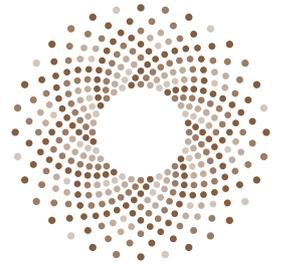


It is a huge pleasure to be here to present to this World Dementia Council dialogue. I am very aware I am not an expert in dementia. But Lenny, Tetsu, Lara, asked me to set the scene today by giving a national case study of data sharing in a pandemic and use it to introduce some of the key themes and issues, in terms of infrastructure, ethics, governance, data standards and policies that are relevant in our quest to accelerate data sharing in dementia internationally. So that is what I am going to do. Data sharing in a pandemic!



Professor Sean Mooney

I think it is important to always



How can we develop and maintain the integrity of robust Global Health Data Research Ecosystems?

HDRUK
Health Data Research UK

- **A national case study**
- **It's a hot topic!**
- **Science Academies of the Group of 7**
 - A Principle Based Governance Infrastructure
 - Operations and Infrastructure
 - Skills – at all levels of data management, data analysis, ethics and governance
 - Establish a commission to achieve these recommendations

31 MARCH 2021

Data for international health emergencies: governance, operations and skills

This Statement has been created by the Science Academies of the Group of Seven (G7) nations. It represents the Academies' view on the need for the G7 countries to realise a better level of 'data readiness' for future health emergencies.

31st March 2021

The exam question for me is how do we develop and maintain the integrity of robust global health data research ecosystems. I will canter through this in six minutes! What I am going to do is give you a national case study. This is a hot topic that is not peculiar to dementia. Indeed, I have shown on this slide the S7 science academies consensus statement, which is the science academies of the G7. It states that for the next pandemic we need to define a principle based governance infrastructure, the technology and operations and the appropriate skills., An independent commission is necessary to achieve these recommendations. Underpinning this is the need to demonstrate trustworthiness and that is going to be a recurrent theme in my presentation today.

collect robust metrics around the value of data - really to identify what's working over time. Helps understand how to drive resources and efforts and metrics could be granular or very high level.



Professor Alan Evans

Is there an inventory of national and international initiatives in open neuroscience.

HDRUK
Health Data Research UK

HDR UK's mission is to unite the UK's health data to enable discoveries that improve people's lives

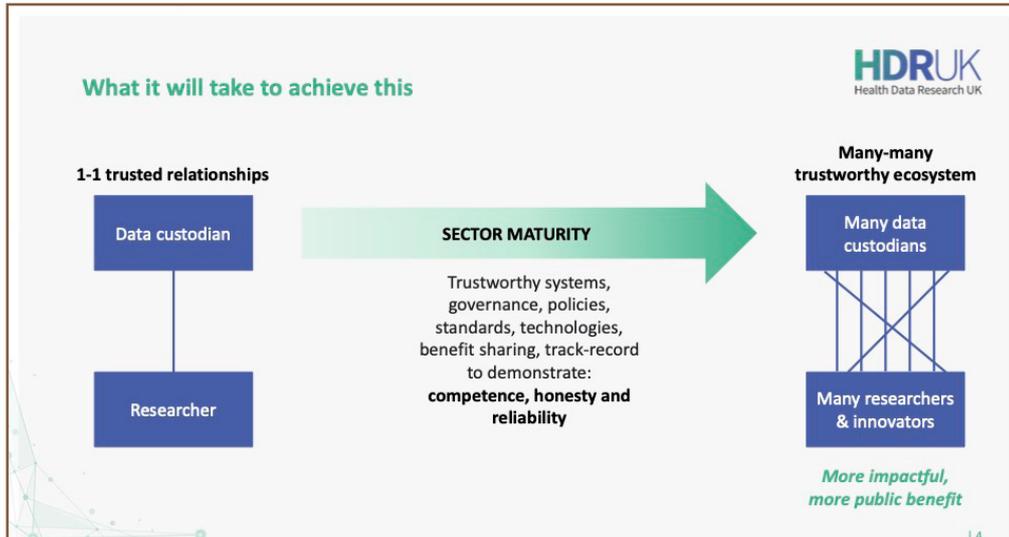
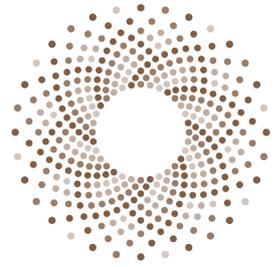
Our 20-year vision is for large scale data and advanced analytics to benefit every patient interaction, clinical trial, biomedical discovery and enhance public health.

Let's flip to the UK. HDR UK – and it is great to see John Gallacher and Ronan Lyons here today because we are all part of the same team – our aim is to unite the UK's health data to enable discoveries that improve people's lives in anticipation of the fourth industrial revolution. To do this will require us to import data science into medicine, health care, and biomedical discovery.



Dr Ivan Koychev

In the light of aducanumab decision, I wonder whether in the coming years datasets holding information on people in preclinical dementia (e.g. amyloid positivity) would have duty to share this information with local interventional studies or clinical services? This has so far not been expected due to lack of disease modification therapies



HDR UK’s goal is to run studies on up to 66 million people. Fragmentation in the system is a major challenge. The slide shows that the current system relies predominantly on trusted bilateral relations. The future will require a trusted “many-to-many” ecosystem. And if we are going to develop that sector maturity, there will be a requirement to standardise the “heavy lifting” to develop the governance, policies, standards and technologies that arguably are domain agnostic. This is key to interoperability which I define as the ability to work across organizational boundaries with no additional effort. Currently a lot of effort goes into the messaging between institutions and systems.



Dr Sean Hill

I agree very much with the principles that have been highlighted: 1. Trust 2. Incentives 3. Governance 4. Interoperability. However, it’s also important to consider the end goal, which is presumably reuse and integration. This requires deep understanding of the scientific meaning and quality of the data and which data it can be combined with.

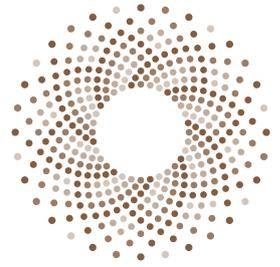


Dr David Sibbald

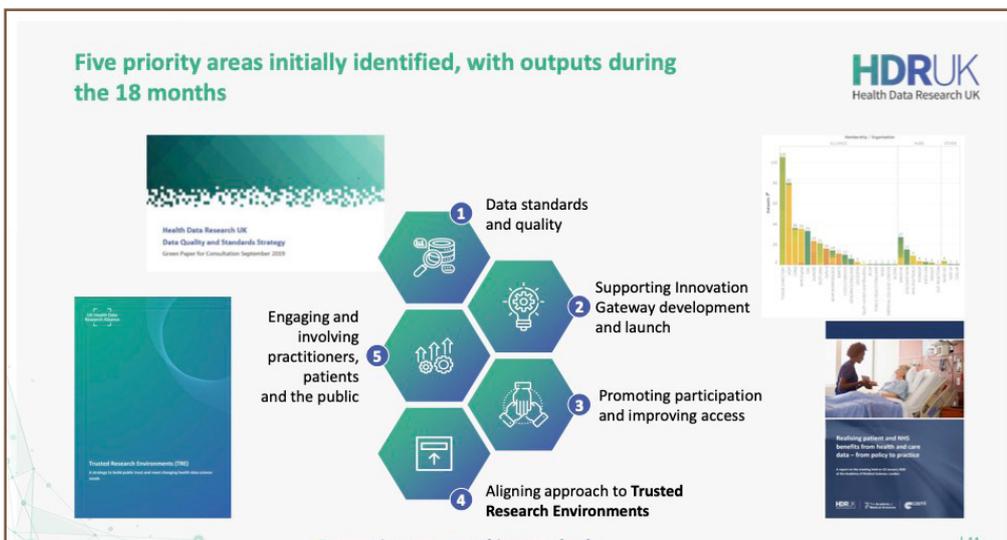
Technical bias here...picking to Angela’s point of ‘no silver bullet’ to address GDPR/ patient consent challenges, I’d suggest Federation, Synthesis and FAIR Data help....also picks up on Heather’s question just now

Independence within the delivery model is key to this. HDR UK is not a funder. It is independent, not a data controller and physically based within the Wellcome Trust. Our purpose is to enable the ecosystem by defining the platforms, the standards, the meta-data the inter-operability, to enable data sharing at scale.

To do that we need to build a community. And this illustrates the members across the UK – and it is growing all the time – with 52 current members of a UK Health Data Research Alliance including charities, universities, private industry, and NHS organizations.



Alliance members have come together to subscribe to the key precompetitive activities in terms of data standards and quality, engaging the public, supporting innovation, but also the rules of the game. How do you work together to support data sharing and data access? And the key role of trusted research environments. Importantly this is all being delivered “the open”.



A concept we think is helpful and the SAIL team in Wales was one of the first to do it is “the five safes framework.” The public understand this concept. Safe people working on a safe project in safe settings with safe data and safe outputs. These safeguards enable trustworthy data sharing.

The second challenge is to improve the quality of the data. We sponsored seven initial data research hubs—I think we are up to 11 now – which are domain specific environments which enable data sharing and improve quality data. Whether it is in respiratory trials (BREATHE), the DigiTrial hub which supported the RECOVERY, or DPUK which is equivalent to a hub to enable data sharing in dementia.

 **Professor Maryann Martone**

What is the current level of data sharing in AD preclinical and clinical? It always seems we are making progress until we consider the entire community which is large I suspect.

 **Dr Lara Mangravite**

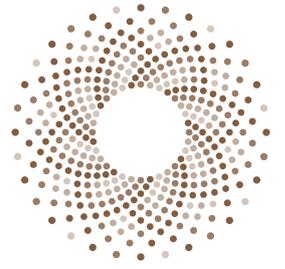
@Sean hill - I completely agree! This can be hard to do at scale and often requires active collaboration.

 **Professor Maryann Martone**

Establishing a culture of data citation is important which is why I emphasize PIDs

 **Dr Petr Holub**

Quality and incentives in addition to FAIR - FAIR-Health concept <https://www.liebertpub.com/doi/full/10.1089/bio.2017.0110>
Federated machine



The Gateway Is a common shop window so that anyone in the world can discover UK datasets and search on the associated metadata. This is in the foothills of where it needs to be, but we currently have around 640 data sets and 11,000 searches a month. Researchers can see who has got what data. Currently a common data request form is being piloted across out trusted research environments.

So that is the theory. Does it work? Let's turn to Covid-19. At the start of the pandemic we were very fragmented and there was an opportunity to work as a team across organizations to bring the UK health data research community into a single endeavour.

learning is one of components (in reply to @Joachim Schultze)



Dr Sean Hill
@Lara Mangravite - We have had success by embedding provenance information of how the data was generated in the metadata itself. This, however, is currently very challenging at scale.

Building the Gateway for researchers and innovators to access data

www.healthdatagateway.org

1,057 registered users
640 datasets
251 access requests
151 tools
10,907 searches in last 30 days

- Access to request over 600 datasets, 149 tools, 195 educational courses and 875 publications
- A dedicated collection of datasets for each Hub – 140 in total
- Facilitated 247 requests to access health datasets, in particular actively supporting the government's National Core Studies into COVID-19 ("Data and Connectivity")

HDRUK
Health Data Research UK

"Really impressed with this resource. I think as a gateway to search by data type and indication, it's a really powerful tool."
David Leather, GSK

I don't expect you to read this slide. But health data research in the UK has produced around 1400 scientific outputs in terms of pre-prints and papers, and we report to the Government every two weeks.



Dr John Langton
I completely disagree with everything you are saying. Frankly I'm shocked

Data & Connectivity National Core Study: COVID-19 dataset availability – 8 June					
Viral genome data now available in Wales. Primary care data now available from ONS as part of the linked data asset (Census-Mortality-HES-GPES)					
Core COVID-19 Datasets available for linkage	Office for National Statistics (Secure Research Service)	England (NHS Digital Data Processing Service)	Scotland (National Data Safe Haven)	Wales (SAIL Databank)	Northern Ireland (Honest Broker Service)
C-19 vaccine data collection	To be made available shortly. Accepting applications now	Vaccines Events & Adverse Reactions	Scottish Vaccination Data	COVID Vaccination Dataset	Data access agreed. Data to be transferred to TRE shortly from Vaccine Management System
COG-UK viral genome	Awaiting data flow from Public Health England	Awaiting data flow from Public Health England (to follow flow into DNS)	COG-UK data available in TRE. Subset also linked to CO-CIN data	COG-UK data available in TRE	Governance agreed, automation of data flow to TRE in progress
Pillar 1 COVID-19 Testing Data	To be linked to Test and Trace data	COVID-19 Second Generation Surveillance System (SGSS)	Electronic Communication of Surveillance in Scotland (ECOSS)	COVID-19 Test Results	COVID antigen testing - Pillar 1
Pillar 2 Testing data (UK Gov)	To be linked to Test and Trace data	COVID-19 UK Non-hospital Antigen Testing Results (Pillar 2)	Electronic Communication of Surveillance in Scotland (ECOSS)	COVID-19 Test Results	Missing results prior to 26 Apr - Data quality issue
Primary Care	Census-Mortality-HES-GPES linked data asset now available (ONS/NEHSD)	GPES extract – 98% practice coverage, large subset of codes. (See items)	Abusoft ESCRO GP Extraction* Prescribing Information System	80%+ coverage of full longitudinal record, with 100% coverage for COVID codes	Enhanced Prescribing Database as proxy
Secondary Care	Census-Mortality-HES-GPES linked data asset now available (ONS/NEHSD)	100% coverage – HES, SUS via DARS extract only, available in TRE soon	100% coverage	100% coverage	Admissions & Discharges
Personal Demographic Service	Internal use only	100% coverage (via DARS extract only)	100% coverage	100% coverage	
Death registry	Provisional Monthly Extract & Linked Census and death occurrence	100% coverage (Civil Registrations - Deaths)	100% coverage	100% coverage	
C-19 Infection Survey (CIS)	Linked to Test and Trace data	N/A	Awaiting DEA accreditation	Awaiting decision on data access	Awaiting decision on data access
COVID-19 Clinical Information Network (CO-CIN)	Being linked to 2011 census	Data for English CO-CIN participants available in Scottish Nation Data Safe Haven	Limited metadata, includes English linked data, and COGUK/CO-CIN data asset	Awaiting decision on data access	Discussions ongoing to collect data in NI
Census 2011	Household structure	N/A	N/A	N/A	N/A
Covid Opinions Survey		N/A	N/A	Awaiting decision on data access	N/A
Business Impact of Covid Survey	c. 5,000 businesses	N/A	N/A	Awaiting decision on data access	N/A
Labour Force Survey	40,000 households, 100,000 individuals	N/A	N/A	Awaiting decision on data access	N/A
Intensive Care data	Preparing data sharing agreement for ICNARC	HES Critical Care (ICNARC available in bare)	SICSAQ (updated weekly)	ICNARC COVID weekly, ICNARC quarterly all admissions and critical care routine data (ECOS) monthly	ICNARC to be acquired
Pillar 3 Testing data (NHS labs)	Captured within Test and Trace data	N/A		COVID-19 Test Results	

Office for National Statistics

KEY

- Custodian engagement
- Dataset available in secure Trusted Research Environment
- Linkages established to other priority datasets (within TRE)
- Datasets available for COVID-19 research via Gateway

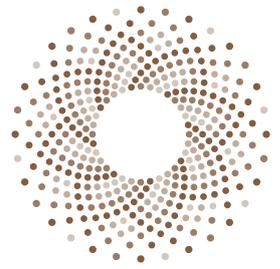
Further information about Data & Connectivity can be found [here](#), along with our [latest monthly sprint report](#)

Data and Connectivity National Core Study webpages and dashboard



Dr Jennifer Yokoyama
As a geneticist & neuroimager, I'd love to hear people's thoughts on how we as a community are also thinking about data such as MRI, EEG, voice recordings, and other potentially extremely identifiable data as we move into this

It is also important to improve data discoverability. Across England, Wales, that's SAIL, Scotland and Northern Ireland we can see what data sets are held where and their state of accessibility. Not just clinical data but genomic data, COVID-29n testing data, labour force data, socioeconomic data and so on.



Status of COVID-19 projects using the data – 8 June

# of COVID-19 Projects by stage (change from previous report)	Office for National Statistics Secure Research Service	England (NHS Digital Data Processing Service)	Scotland (National Data Safe Haven)	Wales (SAIL Databank)	Northern Ireland (Honest Broker Service)	Total
In development	8 (+3)	24 (-16)	23 (-9)	48 (+2)	5 (-)	108 (-20)
Submitted for information	0 (-2)	12 (+2)	4 (-4)	2 (+1)	0 (-)	18 (-3)
Governance approval	6 (+2)	1 (-3)	21 (+9)	2 (+1)	2 (-1)	32 (+8)
Approved but not yet active	6 (+2)	1 (-3)	21 (+9)	2 (+1)	2 (-1)	32 (+8)
Active research projects taking place	26 (-2)	110 (+10)	78 (-)	128 (-2)	2 (+1)	344 (+7)
Completed projects	3	0	3	25	0	31
Active Number of Researchers	367 (+15)	48 (+3)	154 (-85)	233 (-22)	10 (+8)	812 (-81)
Average time from application to active research	115 days	Not available	10 days	7 days	47 days	
Number of disclosure requests in last 4 weeks*	Not available yet	156	53	110	38	
Number of items in disclosure requests*	Not available yet	263	459	639	2843	

* These are requests by researchers to export safe and anonymous data from the Trusted Research Environments. These requests include items such as graphs or data tables

Participation in key UK wide studies:

- PRINCIPLE:** 5,018 participants (+3% in last 2 weeks, with continuing data flow of Pillar 2 COVID +ve test results to support recruitment)
- RECOVERY:** 39,967 participants across 181 active sites (+0.5% in last 2 weeks)
- CO-CIN (ISARIC 4C)**
 - 209,693 Tier 0 (case report) (+1.2% in last 2 weeks)
 - 2334 Tier 1/2
- GENOMICC:** 13,011 participants (+3.3% in last 2 weeks) across 212 ICUs with a total of 5091 intensive care beds
- COVID-19 ZOE symptom study:** 4,661,644

HDRUK
Health Data Research UK

Office for National Statistics

Data Use Registers
For more information on the active projects:

- [ONS Secure Research Service:](#) List of accredited researchers and research projects under the Research Strand of the Digital Economy Act)
- [NHS Digital:](#) Register of approved data releases (includes all access)
- [Scotland:](#) Public Benefit and Privacy Panel approvals
- [SAIL Databank:](#) COVID-19 projects listed on gateway
- [NI Honest Broker Service:](#) Projects currently being carried out

Then it is possible to discover what research is happening in which environment. For example in Wales, they are always ahead of the game, there are 128 active research projects working on that linked Welsh data asset.

What can we do with linked multi-modal datasets? Examples include covid-19 vaccine effectiveness and safety – this is a Scotland wide example linking vaccination, primary care, covid-19 testing, hospitalisation and mortality records. Aziz Sheikh and the team were the first people in the world to publish, in the Lancet, the first dose effectiveness at 28 days. And in Nature Medicine today they published risk in terms of thrombocytopenia and thromboembolic events.

space —do we move towards ENIGMA-like “common preprocessing” pipelines? The worry here would be that innovation in processing methodology could suffer...catching up in the chat - perhaps federated data is the way around this?



Professor Ronan Lyons
Open Science and Open Data are not necessarily compatible due to GDPR etc. Open Science is best achieved through Federation of Accessible Data through a network of Trusted Research Environments.



Professor Caleb Webber
How do we reward data citations? We reward publications only which could disincentivize.



Professor David Llewellyn
Data sharing is

Impact Case Study 1: Vaccine Effectiveness and safety in 5.4M

Challenge

- To confirm the real-world effectiveness and safety of COVID-19 vaccines to establish the success of the roll out is urgent

Solution

- Enabling access to the EAVE II dataset, which links vaccination, primary care, COVID-19 testing, hospitalisation and mortality records for 99% Scottish population (5.4M people), through the Scottish National Safe Haven
- This setting provided access for approved researchers in a safe and secure way and allowed rapid analysis to data, with publication of findings within weeks of data collection

Impact

- Analysis showed that the Pfizer and AstraZeneca vaccines reduced hospital admissions by 85% and 94% respectively (and this result was maintained for adults >80 years)
- The results provided early insights that the COVID-19 vaccination is working on a population-wide level and provides confidence in the continuing roll out

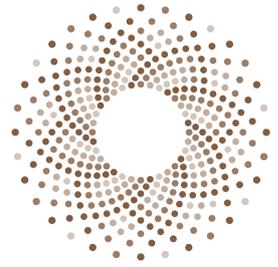
Interim findings from first-dose mass COVID-19 vaccination roll-out and COVID-19 hospital admissions in Scotland: a national prospective cohort study

Lancet May 1st

First-dose ChAdOx1 and BNT162b2 COVID-19 vaccines and thrombocytopenic, thromboembolic and hemorrhagic events in Scotland

Nature Medicine 9th June

What about non-Covid related harm? This is Cathy Sudlow’s work in the English TRE of 56 million people and 4 billion records.



Impact Case Study 2: Non COVID related harm in 54M

British Heart Foundation Data Science Centre
Led by Health Data Research UK

HDRUK
Health Data Research UK

Data type	Country			
	England (NHS Digital cardiovascular TRE)	Scotland (National Data Safe Haven)	Wales (SAIL)	Northern Ireland (Honest Broker Service)
Population size (approx)	58 million	5.5 million	3.2 million	1.6 million
Hospital data (e.g. HES, SMR, PEDW)	Available	Sep-20	Oct-20	TBC
COVID lab testing data (e.g. SGSS, ECOS)	Available	Sep-20	Oct-20	TBC
Primary care data	Available	Sep-20	Oct-20	TBC
Death data	Available	Sep-20	Oct-20	TBC
ITU data	TBC	Sep-20	TBC	TBC
ITU/HDU admissions (CHES)	Available	N/A	N/A	N/A
Prescribing/Dispensing data	Sep-20	Sep-20	Oct-20	TBC
NICOR CVD data	Oct-20	N/A	Dec-2020	TBC
Stroke audit data	Oct-20	Sep-20	Dec-2020	TBC
National Vascular registry	Nov-20	N/A	TBC	N/A

GP Data Exploration

Jenny Cooper
10 September 2020

GP Data (gdppr_dars_nic_391419_j3w9t)

Overview

The GP Data provides a line per clinical recording for a patient. There are therefore multiple rows per patient. The record_date ranges from 1793-12-07 to 2019-12-18. The plot below shows completeness of data from 2015 to present.

Figure 1: Total number of records using Record Date from 2015 to present (monthly)

4 billion records
56.5 million patients

The team showed the catastrophic fall in admission for heart disease when covid-19 struck last March in the UK and the slow rate of recovery.

COVID-19 and cardiovascular disease: The 4C initiative driver project

British Heart Foundation Data Science Centre
Led by Health Data Research UK

HDRUK
Health Data Research UK

4C initiative Overview Cardiac Cerebrovascular Other vascular Tabulate About

Clinical Care for CVD in the COVID-19 emergency

Submit data +

Hospital statistics

Select hospital statistics

Total A&E visits

Total hospital admissions

Total COVID admissions

Graphical options

Show individual hospitals/trusts

Show percentage change

Show 4-week average

Select period (mid-week dates)

11/09

Data from collaborators

Mean no. of activity

Date

— 2019-2020 — Total A&E visits
--- 2018-2019 --- Total hospital admissions

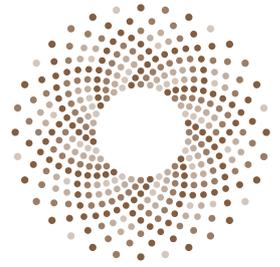
In terms of trials, colleagues in Oxford not only looked at ascertainment of outcomes but used the acute testing data to fuel recruitment, into the PRINCIPLE trial led by Chris Butler.

essential, and accessibility remains a major barrier to data scientists and AI specialists in the dementia field. To give a practical example, few datasets are suitable for Kaggle-style hackathons and competitions. Innovation is being stifled as a result, and ethical concerns about disclosure need to be balanced by ethical concerns about impairing progress.

 **Dr Lara Mangravite**
@Maryann - PIDs are being used in the projects that I'm aware of. I agree with you that they are not as pervasive as they should be.

 **Professor Maryann Martone**
Cross domain sharing is also likely essential

 **Dr Joachim Schultze**
If you are interested, AI, using Medical



Impact Case Study 3: Data enabled recruitment to PRINCIPLE



- **Developed and enabled near real-time, data-driven, ethical approach for enhancing recruitment into the PRINCIPLE trial** - by contacting people with a recent COVID-19 positive test result from the central NHS Test and Trace service within 24-48 hours of their test result.
- **Rapid Public Patient Involvement & Engagement survey** (>90 respondents) to gain insights on public perceptions - demonstrated an efficient process to rapidly inform the trial team's understanding and recruitment approach
- **Multi-disciplinary team** formed to solve the complex issues required to provide a near-real time feed of potential trial participants from test result data to the research team.
- **IMPACT: Rapidly increased weekly recruitment rate** allowing more rapid evaluation of treatments and increased safety for participants. PRINCIPLE has now recruited **4860 participants** (from <1000 pre-data enablement)
- **Now employing approach across other National Core Studies and Urgent Public Health trials** – Roll out of this approach to support AZ trials. AZ have reported **global interest in approach**.

This is the UK approach to building a trustworthy data research ecosystem. A key question is whether this model is scalable. Currently we are trialling this with colleagues at Minderoo and the Gates Foundation in the ICODA project with an initial focus on international covid-19 data sharing.

data and still focusing only on joined insight sharing rather than data sharing can be technically achieved. For more information, just see how we envision this: <https://www.nature.com/articles/s41586-021-03583-3>

International COVID-19 Data Alliance (ICODA)



ICODA is an open and inclusive global collaboration of leading life science, philanthropic and research organisations that have come together to harness the power of health data to respond to the COVID-19 pandemic. ICODA is convened by [Health Data Research UK](#), the national institute for health data in the UK.

VISION

To unite international health research data to enable discoveries that benefit everyone, everywhere, by reducing the harm of COVID-19; and enable an efficient data response to future pandemics and other health challenges.

MISSION

To build an open international partnership that demonstrates trustworthiness to support a rapid response to COVID-19 and a long-term alliance for making data accessible to researchers and scientists around the world.

Launched 6th July 2020

ICODA has the mission, the same values and the same approach to public engagement.



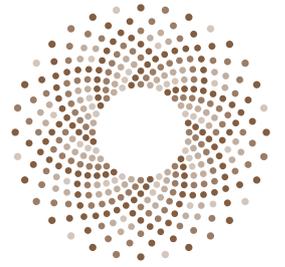
Dr John Langton

I agree data sharing is essential. There are easy ways of addressing “green” concerns and I think that is a trial red herring on the issue. How do we validate insights? With data. How do we find new insights? Data. There is no replacement for data sharing. It is essential.



Professor Craig Ritchie

Great discussion - agree with Art - the motivation to share needs to be incentivised for the PI, there are many points of resistance like the need to publish for e.g. REF



Through a broad global Alliance and Workbench

HDRUK
Health Data Research UK

Our Partners...and growing

The African Academy of Sciences, aridhia, BREATHE Health Data Research Hub, CERTARA, Cytel, DNASTACK, generation SCOTLAND, Genomics england, Global Alliance for Genomics & Health, Health Data Research Network Canada, IDDO, ISARIC, NICE National Institute for Health and Care Excellence, NOVARTIS, RDA RESEARCH DATA ALLIANCE, SAIL DATABANK, SeRP, SHIVOM, Wellcome

Our Therapeutics Accelerator Funders

BILL & MELINDA GATES foundation, Chan Zuckerberg Initiative, Mastercard, Microsoft AI for Health, RIVERSIDE FOUNDATION, Wellcome

Convened by

And many of the same partners.

How to develop and maintain the integrity of trustworthy health data research ecosystems?

HDRUK
Health Data Research UK

- It's about **coordination and connectivity** not **command and control**
- **Demonstrating trustworthiness** with the public through consultation and good governance on data sharing and analysis
- Importance of a **governance infrastructure, standards and FAIR data services**
- A commitment to a **multi-disciplinary endeavour** public, service providers, government,
- A commitment to **team science**
- A determination to **advance open science practices** internationally.

We are not there yet!

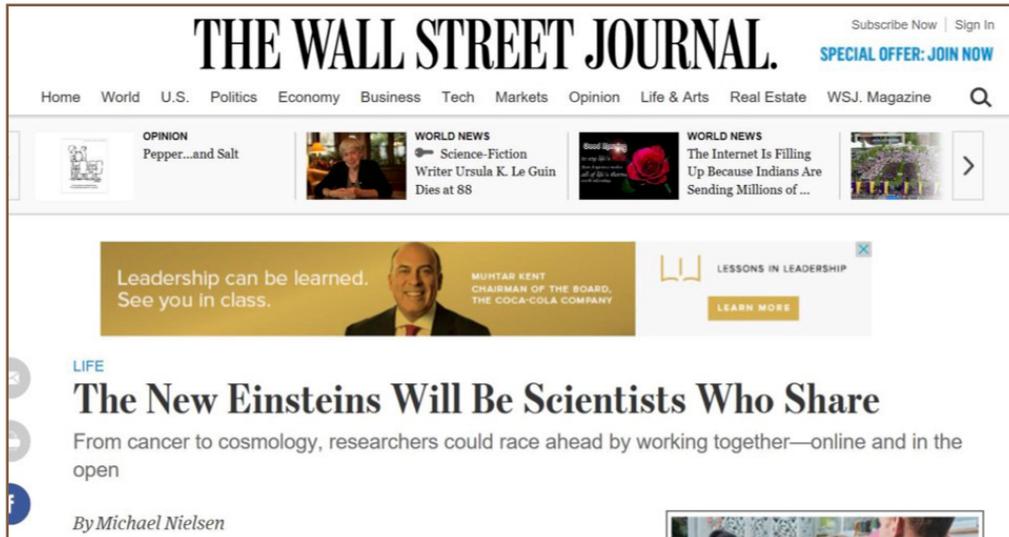
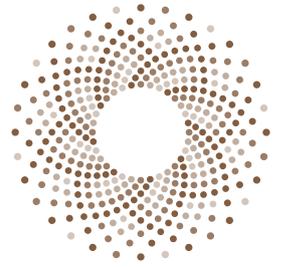
So, to conclude Lara, the question is how to develop and maintain the integrity of trustworthy health data research ecosystems?

I would suggest it is about coordination and connectivity, not command and control. Independence is important. We need to demonstrate trustworthiness in everything we do. The public are fantastic in terms of guiding the science and guiding the research. It is about the policies and standards for a governance infrastructure and technology infrastructure and FAIR data services. But the key to this is a team effort and a team that subscribes to advancing open science practices internationally.

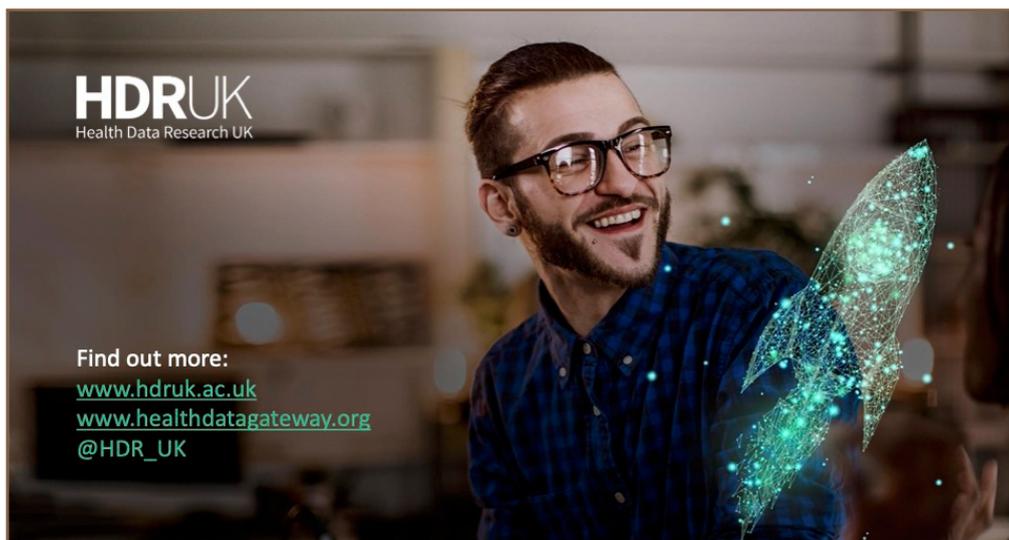
in a competitive space and (more importantly) most grants do not give the necessary resources to prepare data into a format that eases sharing. Finally - with consent - most participants expect data to be shared but the consent states it 'may' be shared - perhaps funders could insist that consent states it 'will' be shared and articulate exactly onto which platform. Also let not forget the vast amount of data and samples sitting untouched in industry vaults and freezers and continue to look at motivators to share/ make available their data.

 **Professor Maryann Martone**
+100 @john

 **Dr Saskia Sivananthan**
agreed omlpletely @ johnlangton



We are not there yet. We are in the foothills of where we need to be but I do subscribe to the headline in the Wall Street Journal that “The New Einsteins will be the scientists who share.” I applaud the efforts of the World Dementia Council.



So, thank you very much I will stop there.



Dr Lara Mangravite
President, Sage Bionetworks

Thank you very much Andrew I appreciate that. So next we are going to hear from Angela Bradshaw who is a project officer at Alzheimer’s Europe and co-authored their recent report on data sharing policy and dementia. Angela



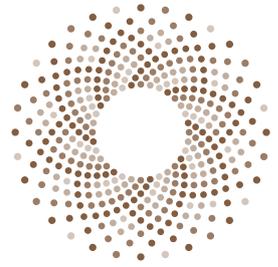
Dr Angela Bradshaw

Speaking to Joachim’s point, there are technical approaches that can work for certain types of data science, in which models are shared beyond firewalls rather than data changing hands. These models can then be trained on data behind the firewall, and returned to the source. It won’t work for everything - sharing of raw or individual patient level data is still necessary and important - but there are ways to overcome some of the challenges associated with federated data networks, in specific situations.



Professor David Llewellyn

Developing rich simulated data based upon real data could be really useful if new insights can still be derived from the ‘fake’ data. A vital workaround that



Dr Angela Bradshaw
Project Officer, Alzheimer Europe

Alzheimer Europe

Data sharing in dementia research - the EU landscape

Angela Bradshaw
Project Officer
Alzheimer Europe

www.bit.ly/DataSharing_AE

circumvents privacy concerns?



Dr Heather Snyder

@Maryann- what do you see the role of journals and funders in implementing these principles around sharing?



Dr Sean Hill

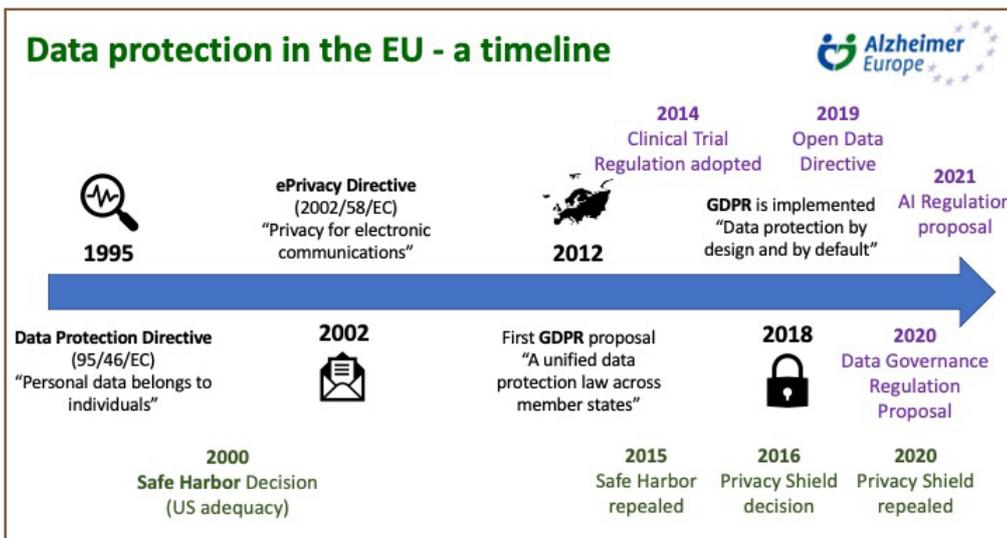
In principle, federate learning sounds very appealing as a way to inform models that can be shared without sharing the data. However, it still requires that the data that is the informing the model is actually compatible. Having harmonized data produced from consistent protocols remains an enormous challenge.



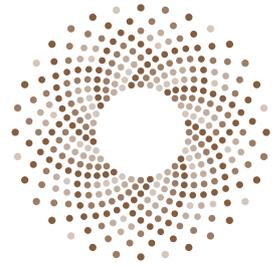
Professor Alan Evans

Data-sharing and insight-sharing are not mutually exclusive. Sharing data from different regions allows us to ask questions that are impossible to

Hi everyone and thanks for inviting me to speak. I am going to talk today about some of the key findings from a recent report we published earlier this year in February on data sharing and dementia research. And actually, the starting point for this point was the value of data sharing which Andrew has aptly demonstrated in the last few minutes. And also the importance of sharing data to honour the contribution of research contributors and patients. So, in the report we really wanted to identify some of the key barriers but also the enablers for data sharing. So, in today's brief presentation I am going to be focussing on data protection, privacy, and the patient perspective.



This slide shows a very brief timeline of data protection regulations in the EU starting with the Data Protection Directive in 1995. This was followed by the ePrivacy Directive in 2002. The first GDPR proposal came in 2012 and it took about six years of wrangling to



finally get an implementable GDPR which happened in May 2018. Alongside this timeline we have had continuous innovation and evolution not just in terms of our ability to use and reuse and technologically process data but also in the regulatory environment.

address from a local context, e.g. why is vascular dementia more prevalent in China and AD more prevalent in the West ?



Again, you can bring large data together in swarms without sharing the data. We are technically there. So there is no need to put data from different data owners together, which for re-use purposes in a GDPR Environment is really difficult.



@Maryann - thanks! I think you just named it a little bit in terms of data credit system.



There is a disitributed provenance model developed under ISO TC/276 as ISO 23494 - which goes from assigning PIDs

The General Data Protection Regulation



"Data protection by design and by default"



- Expands and updates the 1995 Directive, without fundamentally changing the data protection approach
- Ensures the free movement of data throughout the EU
- Greater harmonization and certainty across Member States
- Guaranteeing the right to personal data protection within AND beyond the EU
- Increasing **trust**: additional safeguarding measures for sensitive personal data about health
- The "research exemption" (Art.89): safeguards and derogations related to processing for scientific research or statistical processes

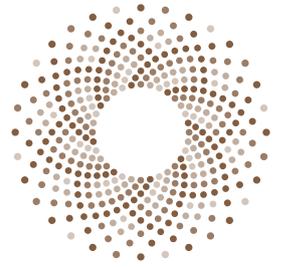
So, in the US we have had the safe harbour decision which was the first US adequacy decision. It was signed in 2000. It was repealed 15 years later and replaced by the Privacy Shield and [indistinct] unfortunately dismantled that last year in July. Alongside you can see we have got a plethora of regulations related to data protection and privacy. The Clinical Trial Regulation was adopted in 2014 although it is not yet implemented. The Open Data Directive was implemented in 2019 and we have recently got proposals for an AI Regulation and a data governance regulation.

So, all of these layers of EU regulation, and this is overlying state by state, so member state by member state, differences and derogations, mean that researchers have quite a complex landscape to navigate when they try and make sure they conduct clinical studies in a way that protects the rights of the individuals and the data protection of individuals.

Criticisms and benefits of the GDPR



- **Room for interpretation**: no clear lawful basis for secondary data use
- **Regulatory divergence**: lack of guidance & alignment on pseudonymization and consent
- **Cross-border transfers**: complexities of data sharing beyond the EU
- **Creation of a risk-averse environment** : calls for a case-by-case identification of DP issues, controller bears primary responsibility for data breaches
- ✓ **GDPR and broad consent**: recitals 29 and 33 include provisions that support broad consent for data sharing and re-use
- ✓ **Greater awareness of data protection rights**: 70% of Europeans have heard of the GDPR
- ✓ **Improved trust in data sharing**: 33% of DRG survey respondents were less concerned about sharing their health data with pharmaceutical companies



This is the GDPR. I am not going to go through it in any detail but basically GDPR expands and updates the 1995 Directive, but it does not fundamentally change the data protection approach. We could see the GDPR coming and in some ways it is not as much of a surprise as it is made out to be. So, the GDPR has three main aims: To ensure the free movement of data; To ensure greater harmonization and; Fundamentally to guarantee the right to personal data protection both within and beyond the EU.

Andrew mentioned trust and trustworthiness and that is going to be an underlying theme of my presentation as well. So, the underlying goal of GDPR was to increase trust. There are safeguarding measures for example the protection of sensitive personal data.

And there are also provisions made to enable research and one of these is the research exemption. So article 89 lays out some safeguards and derogations related to the processing of data for scientific research.

(which can be used for citations as well) to metadata needed for assessing quality = fitness of data for a particular purpose.



Inez Jabalpurwala

@Heather Snyder, I would add we also need to think about how to meaningfully include Low- and Middle-Income Countries, to capture biological and non-biological variability around the world.



Dr Heather Snyder

@Inez - absolutely! And provide them the opportunities to engage. Sharing insights only would restrict who can be part of these discussions as well.



Dr Sean Hill

Google has made a knowledge graph to organize data for the World Wide Web - establishing an Alzheimer's knowledge graph for sharing data

Patient and participant perspectives on data sharing




- Most research participants would like their data to be shared for societal benefit, and to advance research
- Patients and research participants often hold strong views on **who** they would like their data to be shared with

- Data scandals have damaged public **trust** in data sharing; patients and research participants are concerned about **loss of privacy** and data misuse
- **Greater transparency:** many research participants would like to be involved in decisions about data sharing

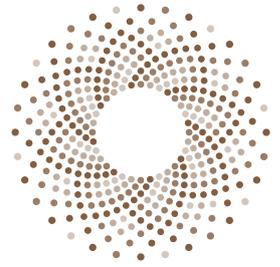


“Research participants are not patients in the traditional sense of the word. It really should be looked at as a partnership.”

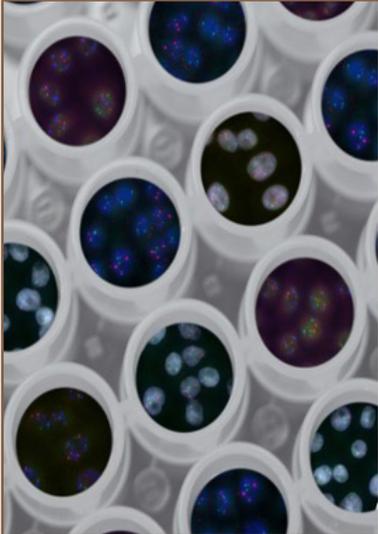
So, although the GDPR was not intended to impede research there have been several valid criticisms levelled at GDPR by researchers and these are laid out here. Some say that the GDPR offers too much room for interpretation and diversions. It needs to do this to a certain extent to allow member states to have their own data protection rules and use their own framework.

And in particular researchers have criticised GDPR for not providing enough guidance and perhaps not enough alignment or harmonization on really key issues for research such as pseudonymization and consent. A number of researchers in the US have highlighted problems with cross-border transfer. It can be hard to share data beyond the EU. Researchers need to be able to identify a legal basis for example to share the data and there is a lot of responsibility placed on the data controller. So now we have data protection officers. We have potentially quite large fines for breaches of data regulation. All this has created quite a risk-averse environment for data sharing.

However, on the other hand GDPR has been to a certain extent beneficial for data sharing. So, the architects of GDPR did foresee the potential for the need for broad consent, so the provisions that support broad consent and data re-use. And a really



important point to make is the GDPR was not designed as a data protection regulation for research it is a broad regulation that goes way beyond the limits of scientific research. So, from the perspective of helping Europeans and people trust that their data is going to be used in a way that is safe and respects their rights it has had some success. Europeans have a greater awareness of data protection rights and surveys show an improved trust in data sharing even with companies that traditionally haven't had perhaps the best reputation when it comes to using data.



Barriers to data sharing:

- **Regulatory divergence** and legal uncertainties around data protection rules for clinical research
- **Commercial sensitivity** issues for private sector organisations
- **Financial and time cost** of data sharing; lack of technical support
- **Motivational concerns:** the 'reputation cost' of data sharing for researchers
- **Privacy concerns** for patients and research participants
- **Lack of digital skills and data literacy**, impeding trust in data sharing



can provide an integration point that can be searched, reused, provenance tracked and used to drive an incentive system.

 **Dr Saskia Sivananthan**

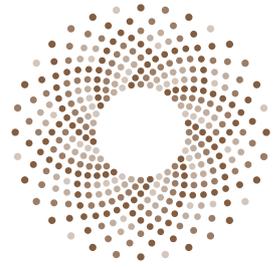
@Inez also the diversity in the data in the G7s. Canada I would argue is extremely poor in collecting or mining this information to answer questions for racialized communities

 **Professor Caleb Webber**

Most molecular dementia researchers get their data interpretations from papers of very varying quality. Providing more robust, integrated and collected data interpretations would be the game changer for the majority of molecular dementia researchers. Knowledge graphs are a great way forwards.

To finish, I'd like to link the points I made around GDPR and trust to the patient perspective. It is clear from many surveys that the majority of research participants want their data to be shared for societal benefit even if they are not necessarily going to benefit themselves. And also participants want to advance research. They also have very strong views on who they would like their data to be shared with. But they also have a keen awareness of what is happening around them. So currently in the UK there is a bit of a furore about GP records being shared and there are scandals such as Cambridge Analytica. And the general public and patients are keenly aware of these scandals, and this has to some extent damaged public trust in data sharing. The consistent concern of participants in dementia research is loss of privacy and data misuse.

And this is where GDPR comes in. Because the GDPR is really there to help reinforce trust and make sure that people's data protection rights are respected and protected. What we need going forward is more transparency. So, research participants would like to be involved in decisions about data sharing. We have all heard the phrase nothing about us without us. Research participants are not really patients in a traditional sense of the word. They are partners and they should be viewed as such.



Enablers for data sharing:

- **Improved regulatory alignment** and best practice guidelines for GDPR compliance
- **Greater financial and technical support** for researchers to share data, including after project funding ends
- **Methods to attribute data authorship** and monitor data sharing & reuse
- **Education, training and transparency:** showing people how their data are used and shared
- **Involving people** with dementia in the design and conduct of research



George Vradenburg

<https://www.davosalzheimerscollaborative.org>



Dr Sean Hill

@Caleb Webber - fully agree



Dr Petr Holub

@David - indeed. But for federated learning, reliable data quality is even more important



Dr Sean Hill

@David - reliable, harmonized and understood data is required for federated learning



Professor Sean Mooney

It occurs to me that we should have better mechanisms to reward data sharing. The Research Parasite awards <https://researchparasite.com/> come to mind, but those reward recipients. How to recognize the value of the data provider

Going forward what we would like to advocate for as an organization is the value of involving patients and research participants in finding solutions to the remaining challenges of the GDPR. For example, co-creating some principles for broad consent and the secondary use of data. This needs to involve patients, the general public, research participants in order for those principles to be endorsed and adopted more broadly.

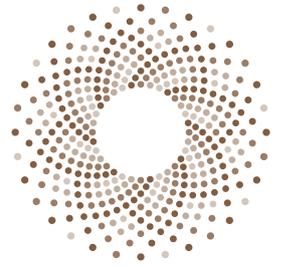
And finally, to conclude in our report what we wanted to really reinforce was there is no silver bullet to solve challenges in data sharing. There are many issues. These come in different shapes and target different stakeholders within the data sharing ecosystem. Problems with GDPR issues with regulatory divergence are only one facet of the problem. And we really need to address all of these issues in concert if we are to make a noticeable improvement and difference in data sharing.

Thank you all for listening and I will stop there.



Dr Lara Mangravite
President, Sage Bionetworks

Thank you Angela, I appreciate it. We are doing reasonably well on time I appreciate it and thank you to our speakers. So now let me introduce our third speaker Suzana Petanceska. Suzana directs the office for strategic development and partnerships in the Division of Neuroscience at the National Institute on Ageing. Take it away Suzana.



Dr Suzana Petanceska
Program Director, Division of Neuroscience,
National Institute on Aging (NIA)

and those who take the efforts to make their data more FAIR?



Inez Jabalpurwala

That's exactly it @Saskia and @Mia; our definition of "quality" data must include whether the data is inclusive of different communities.

Thank you Lara and thank you for the invitation to share our Institute's approach to enabling open science practices



Professor Maryann Martone

I agree @sean M. Very important to start providing accolades to those who share the data.



Dr Petr Holub

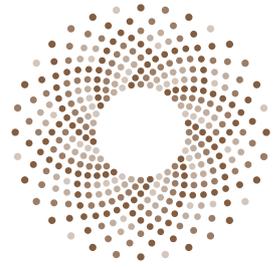
= Dealing with quality issues and with data heterogeneity in federated learning is one of major open topics as it is much more difficult than in pooled data/learning



Dr Sean Hill

@maryann @sean m Fully agree!

So, our work in this domain is guided by the input from the large multi-stakeholder community participating in the NIH Alzheimer's Research Summits starting in 2012.

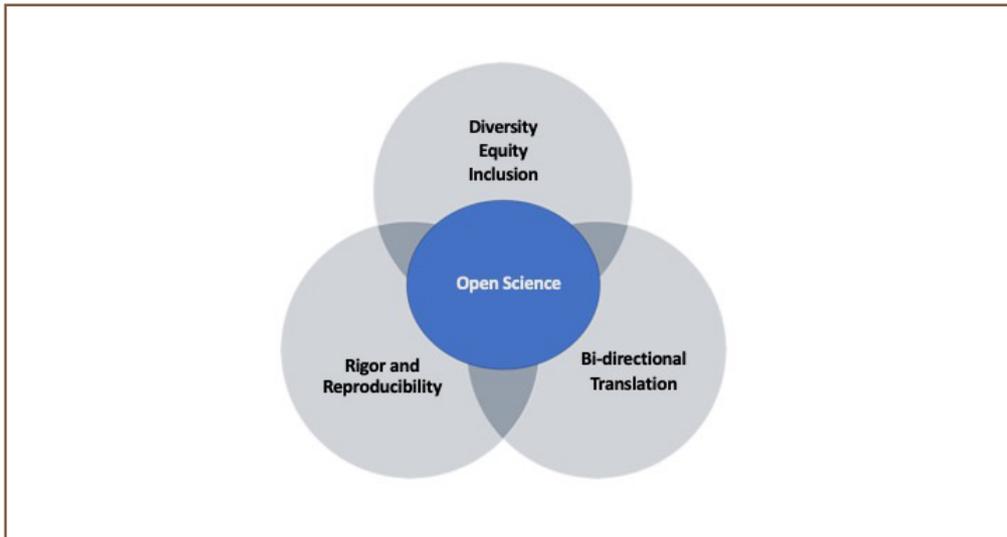


Dr Jennifer Yokoyama

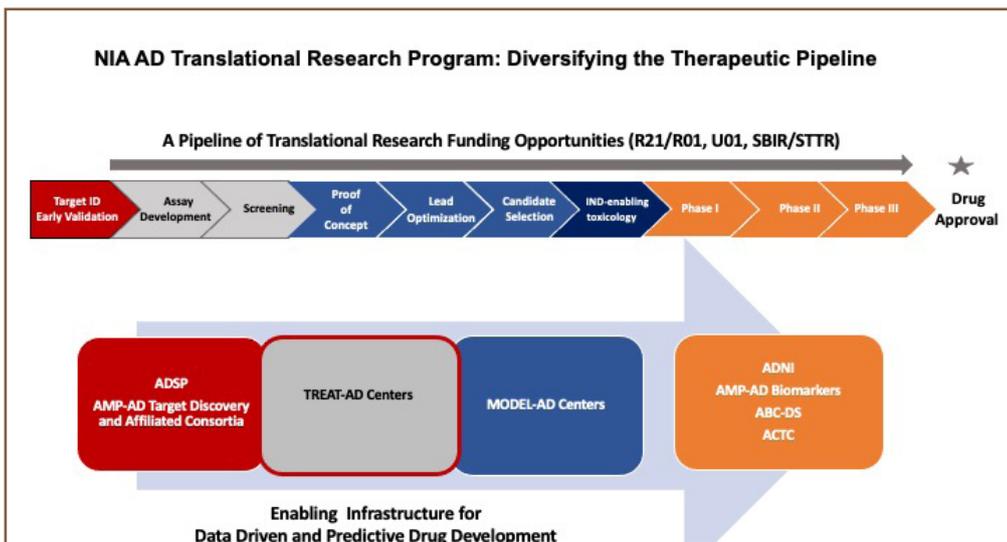
@Inez and @Heather, agree – meaningful inclusion of researchers from less resourced regions would require enhancing tools for data sharing, providing access to larger, broadly representative datasets, AND augmenting computational resource (taking into consideration that even having decent bandwidth is a major limitation in many regions)

Dr Lara Mangravite

There's a whole host of technical solutions designed to overcome the issues of sharing data that can not be made fully open - for reasons of privacy, regulation, or commercial interest. We have found that none of these is a universal solution - but together they give us an arsenal of options. Again,

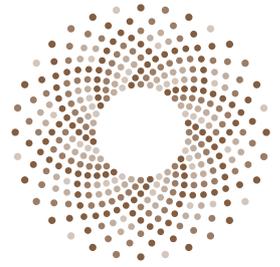


Throughout these four summits open science practices have been recognised as essential for improving research rigor and reproducibility, for enabling bi-directional translation and for increasing diversity, equity and inclusion in research on dementia.



Over the past fifteen years or more we have been developing a robust translational research program for Alzheimer's that has two major components. The first is a pipeline of funding opportunities that spans the space from target ID early validation through to late-stage clinical drug development. And through these funding initiatives we support the development of new clinical drug candidates for a diverse portfolio of therapeutic targets.

The second component on the bottom of the slide is a set of enabling infrastructure programmes. Through these we help the community of drug developers more effectively execute each step of translation.



we need to tailor our solution to the scientific goals of the users.



Dr Petr Holub
Another issue with federated learning is that the participating sites have to have sufficient compute infrastructures to actually allow deployment of federated training



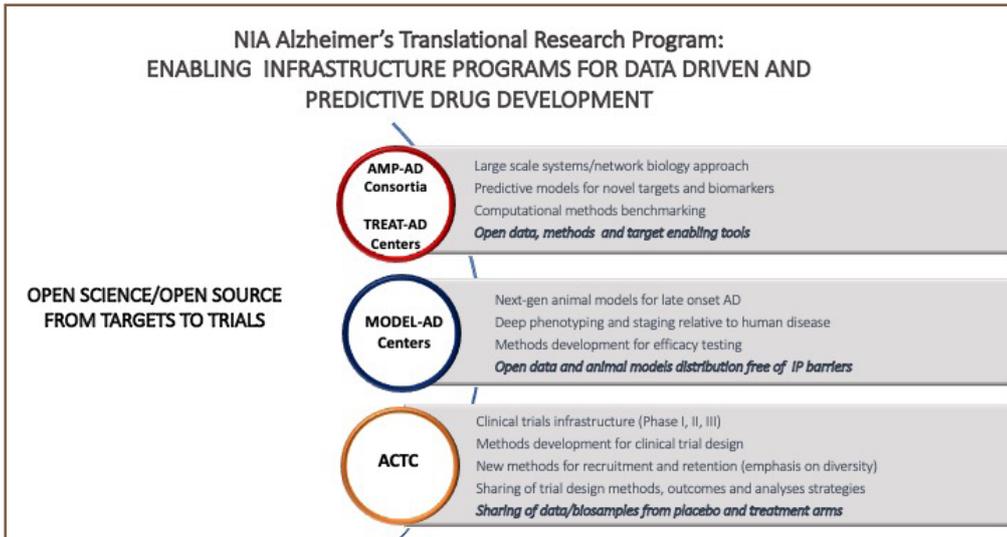
Dr Sean Hill
@Lara Mangravite - very much agree - similar experience.



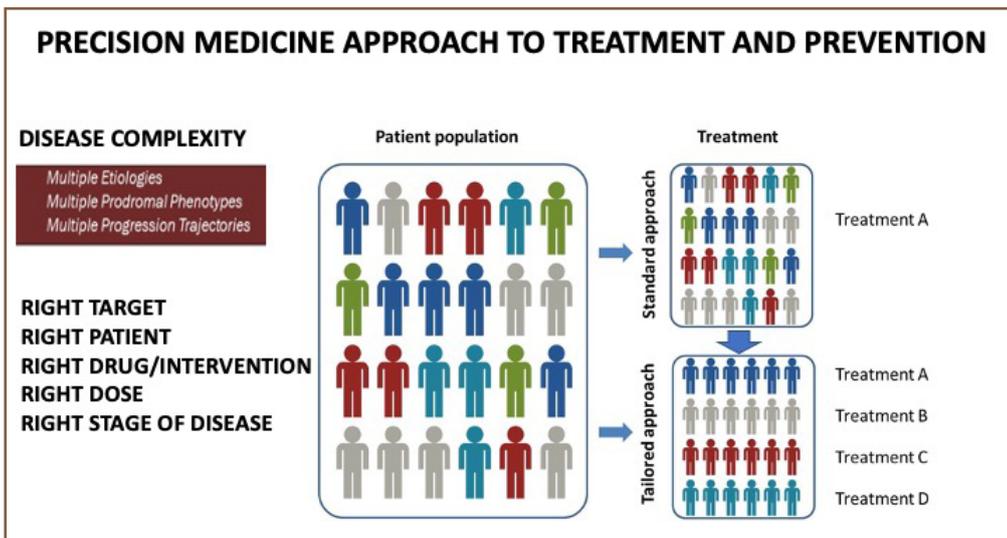
Professor Alan Evans
Very much agree with the idea of supporting remote processing, as data gets too big to move. May also allow us to skirt restrictions to trans-jurisdictional data transfer.



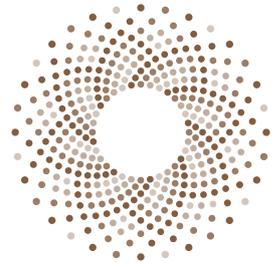
Professor Sean Mooney
@Petr, et al. - I'm very positive toward federated approaches. They



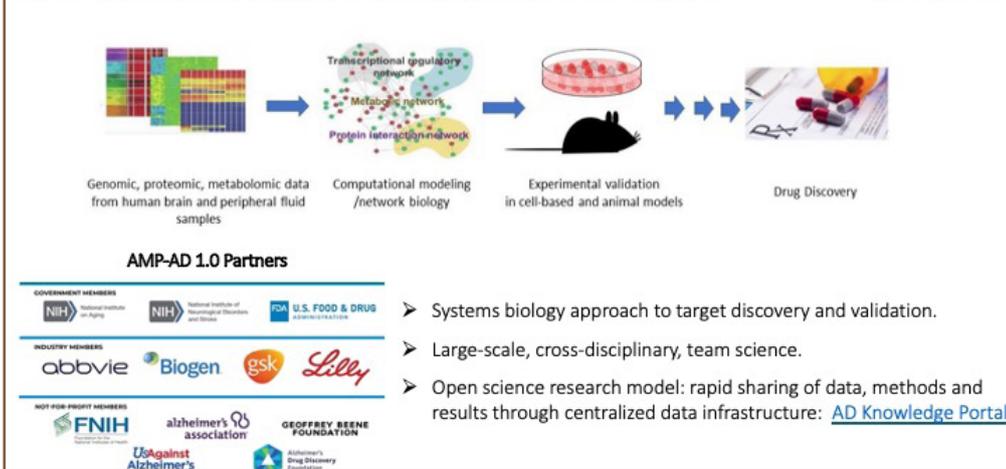
Our vision behind the development of translational infrastructure programmes under open science/open source principles is to accelerate the delivery of data resources, knowledge and research tools that are needed to overcome some of the major factors that contribute to the attrition of candidate drugs in the clinic



And importantly to help the research community implement a precision medicine approach to AD therapy development.



**ACCELERATING MEDICINES PARTNERSHIP FOR ALZHEIMER'S DISEASE (AMP-AD)
TARGET DISCOVERY AND PRECLINICAL VALIDATION PROJECT** Launched in 2014



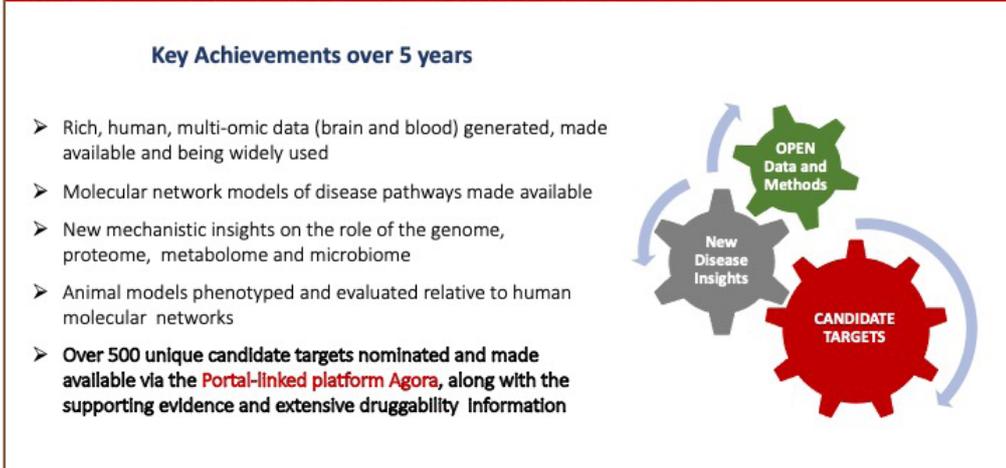
don't solve all problems but have some advantages in certain use cases. Again requires standards / harmonization that also puts additional burden on the federated sites in addition to computing infrastructure.



The International Brain Initiative (IBI) has working groups on federation and data governance. Sean H and I are both on this committee. Would be good to engage on a wider scale on the issues of data governance for neuroscience.

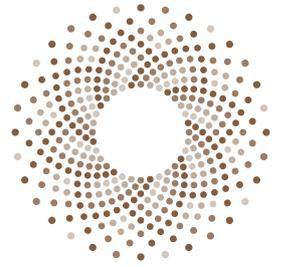
So, the first component of the open science translational infrastructure is the Accelerating Medicine Partnership for Alzheimer's Disease. Established over seven years ago as a public-private partnership, this is a large-scale team science program that used a system-biology approach within an open science research model to discover and characterise novel candidate targets, derived from human data.

**ACCELERATING MEDICINES PARTNERSHIP FOR ALZHEIMER'S DISEASE (AMP-AD)
TARGET DISCOVERY AND PRECLINICAL VALIDATION PROJECT**



Linked to Liz' point on data portability, enabling the data subject to view how their data travels between groups, and ultimately how it delivers insights, is a potentially valuable

The rapid sharing of data was a major contributor to the excellent productivity of the programme. Over the first five years the AMP-AD teams delivered a wealth of high quality human multi-omic data that has been widely used, enabled many new disease insights, and made publicly available more than 500 unique candidate targets along with the supporting evidence and druggability information.



AMP AD 2.0: Enabling a Precision Medicine Approach to Target and Biomarker Discovery
Launched: March 2021

- Expand multi-omic profiling in samples (brain, CSF, blood) from diverse cohorts
- Generate longitudinal immunologic profiling data across diverse cohorts
- Expand the existing sn/sc molecular profiling efforts to multiple brain regions, and in samples from diverse cohorts

AMP AD 2.0 Partners

The new iteration of the programme was recently launched with an expanded research scope focussed on enabling a precision medicine approach to target and biomarker discovery by generating multi-omic data across diverse cohorts.

way to enhance trust in data sharing. IMI projects such as EPAD are best practice exemplars on informing and involving participants, with learnings that could be shared on enabling trust and engaging with participants.

Propagating the Open Science Research Model

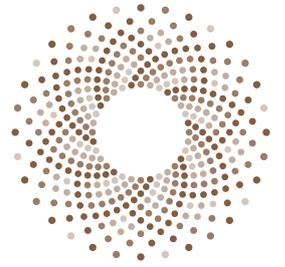


I very much agree with Alan. We need to make sure that the compute technology locally where the data are produced get's better, rather than trying to move data around.



Also the International Neuroinformatics Coordinating Facility is doing a lot on training and coordination around data standards, data management and data sharing.

So, to propagate this research model since the launch of AMP-AD we have established a constellation of affiliated consortia that are using a systems/network biology approach, to gain a deeper understanding of the complex and heterogeneous etiology of AD. The teams participating in these programs are sharing all their data and analytical outputs through the AD Knowledge Portal, a centralised big data infrastructure initially built for the AMP-AD programme.



Translational Centers: Capacity Building for Data Driven Drug Development

<https://treatad.org>



- Open-source, target enabling tools for studying disease biology and as seeds for drug discovery for novel targets



- 50 new mouse models (knock-in) of Late Onset AD (LOAD)
- Multi-modal phenotyping and aligning with human AD phenotypes
- Pipeline for rigorous efficacy testing
- Open data and model distribution free of IP barriers

<https://model-ad.org>

Adjacent components of the open science translational infrastructure are the MODEL-AD and the TREAT-AD translational centres. These centres are building on the output of the AMP-AD consortia and other big data programmes supported by NIA like ADNI and the AD Sequencing Project. They are delivering well phenotyped animal models for LOAD and a whole array of target enabling tools necessary to study the disease biology, and also to advance novel targets into drug discovery.



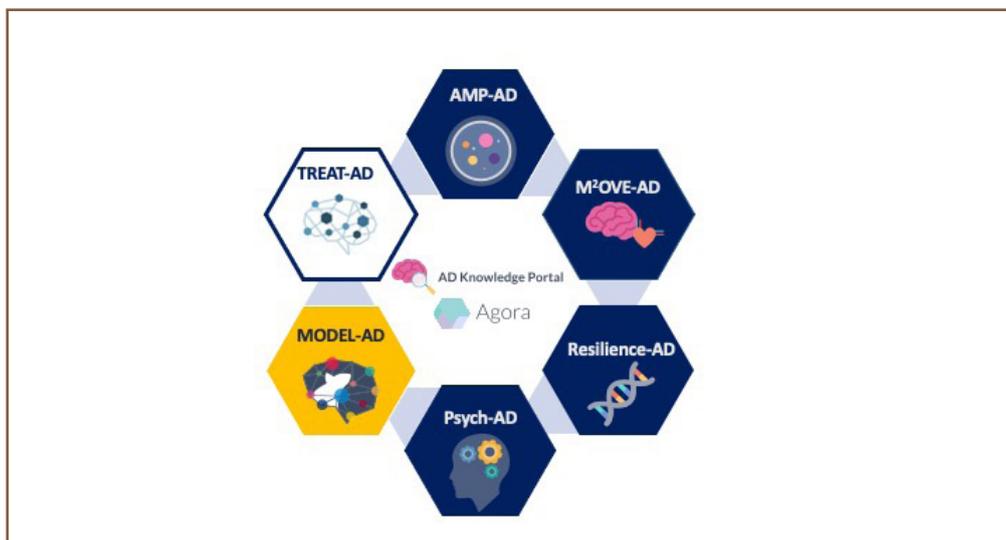
Dr Sean Hill

The challenge of heterogeneity of data requires a common data model, minimal metadata and ideally provenance information about how the data was produced.



Professor Sean Mooney

Side comment: Data quality important, completely agree. If the providers of health care truly sees the value of data, the data will continue to improve in quality and scope. Might not improve sharing but our healthcare partners are increasingly investing in data for their own needs that is also aligned with researcher needs. I expect this trend to continue and AD and dementia to be a part

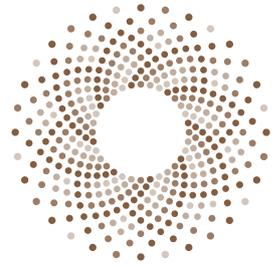


These centres are highly interconnected with the AMP-AD consortia and they are using the same data infrastructure to share their data and tools output; collectively these programmes serve as a powerful discovery engine.



Dr Heather Snyder

@Sean - many use CDISC developed by CPATH



INCENTIVIZING DATA/RESOURCE SHARING

Clear and enforceable expectations for rapid and broad sharing of data and research tools:

- ❖ data sharing expectations stated in funding opportunities and used as terms and conditions for continued funding
- ❖ no publication embargo for data release and data reuse
- ❖ no requirements for co-authorship for data/tools reuse
- ❖ no IP barriers for sharing research tools

FAIR data infrastructure that can support different types of end-users (data scientists, biologists, clinical researchers/users from academia and industry/users with varied access to compute resources):

- ❖ streamlined data governance
- ❖ streamlined data contribution process
- ❖ attribution for data/resource generators

Funding for data sharing/data management activities

Training and community building



Dr Sean Hill

@Heather - agree CDISC can be an important part of the solution - FHIR also is valuable.



Professor Maryann Martone

Yes. Both INCF and IBI are working on a common metadata model. Would be good to work through these general organizations as well as solving the problems for this particular community.



Dr Sean Hill

@maryann Agree!



Dr Lara Mangravite

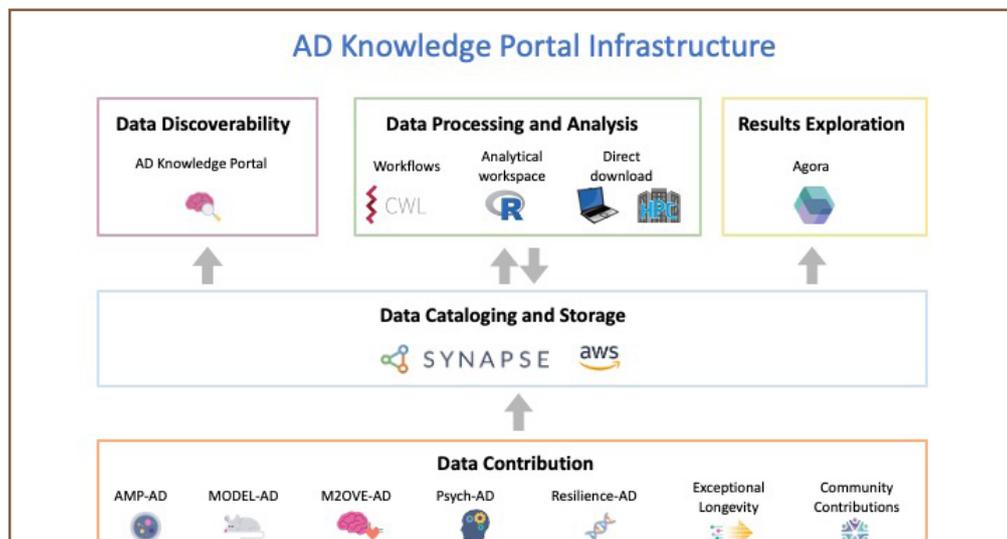
+1, @maryann and @sean



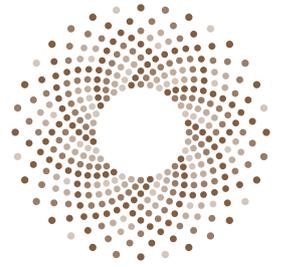
Professor Caleb Webber

For data standards, the UK DRI has set up multiple Data Working Groups for different data types defining best

So how did we incentivise data sharing and open science practices across these programmes? Some of the key elements that have proven to be effective include: i) clear and enforceable expectations that are articulated both in funding opportunities and used as terms and conditions for future funding, ii) investing in the development of FAIR data infrastructure that can support different types of end-users and can give attribution to the data contributors, iii) providing funding for data sharing and data management activities, and iv) support for training and community building.



Few words on the AD knowledge portal FAIR data infrastructure. It is designed to support the hosting and sharing of many and varied data types throughout the entire data set, from raw data to results and to serve the needs of different end-users; those with and without bio-informatics expertise, users coming from small and large academic labs and researchers from the biotech and pharmaceutical industry.



practices, required metadata and data production standards.



We really need to start an integrative initiative where there is real incentives, value and reward for integrating data into a common searchable research (e.g. the knowledge graph mentioned earlier). @Ivan - absolutely agree!



@Sean - I think many funding orgs are moving in that direction. We certainly have ...

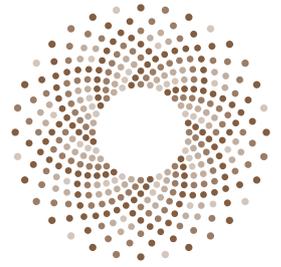


@Heather - fantastic
@Ivan - engaging patients is tremendously valuable and they are often the prime advocates for data sharing
@Ivan - it also create an engagement in care

Making complex biological data sets discoverable and accessible in useable formats and growing a global userbase is a constant work in progress. It is a very human intense effort.



I want to acknowledge that this has been superbly executed by a dedicated team of engineers, data scientists, governance experts, community managers at Sage Bionetworks as part of the NIA-supported AMP-AD Data Coordinating Center.



that is powerful. I also believe that for data quality - clinicians have to be deeply engaged in using the data as part of assessing patients and decision support



Professor Yves Joannette

Exciting and inspiring discussion. But a concept is missing, though very much in the backdrop of all we are discussing: the CULTURE of data sharing (whatever the approach: data shifting, federated approach, insight). Funders, Universities and Industry have to contribute to the shift from a « personal data richness culture » to a « collective/ collaborative data culture ». We need tools of all sort, but we should not underestimate the fact that the culture of data sharing is not shared by most, and generally not valued by universities or funders in their

Inside NIA A Blog for Researchers

Open science delivers a wealth of AD/ADRD research data to a portal near you
November 18, 2020

Raising the bar on data and biosample sharing from AD/ADRD clinical trials
December 09, 2020

Data management and sharing requirements: Tips and tricks to plan ahead
March 24, 2021

We continue to raise the bar on data sharing and research sharing across all our programme including clinical trials; with us is my colleague Laurie Ryan and she can comment on this during the discussion.

NIH AD Research Summits: Change academic, publishing, and funding incentives to enable collaborative, transparent and reproducible research

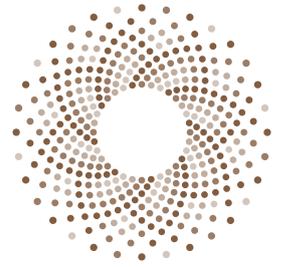
NIH National Institute on Aging

And lastly, I'd like to emphasize that while funding agencies have a very important role to play in incentivising open science practices, they are only one leg of the three-legged stool of stakeholders responsible for creating incentives around data sharing and open science practices more broadly; academic institutions and publishers also have an equally important role to play in this.



Dr Lara Mangravite
President, Sage Bionetworks

Thank you Suzana. That was a great way to get us started. We heard about a project in global infrastructure to support broad data sharing across the health ecosystem.



We heard about policies that are being implemented to promote trustworthy and safe data sharing. And we heard about a community specific effort that is focussed on clear research outcomes. And I do think I heard the theme of trustworthiness come up across all these programmes as absolutely important.

And Importantly Suzana touched on this question of incentives: we can build all this great infrastructure but how do we incentivise people to use it? The pandemic was a great incentive across many of our data sharing programmes and are there ways we can build on that moving forward?

So, to open the discussion I will turn it over to Tetsu who is going to moderate that.



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Great thanks Lara. And thanks very much to the speakers for the content, the variety, the relevance and also sticking very close to time which is nearly unprecedented. I think this discussion will work best the less you hear from me. I will just set some ground rules. We would love to have an open and free flowing discussion, but I may tread on the brakes a bit if I feel we are heading in the wrong direction. We are monitoring your chat so if you want to make comments by chat, or if you don't mind us pulling out some of your comments that has come in through chat, we can do it that way.

But before I open it up to the field just to re-iterate some of the things that Lara mentioned that I thought I heard running through the presentations. Really the issues are about how are we are building communities that share data and use shared data. Because sharing data in principle is something that everyone generally supports but actually getting the data shared, getting people to use that data, is what brings value to it. And that is about how we build and maintain trust while doing that sharing and also to some extent and considering the data that is most useful to share, the integration of the data that is most useful to share, and its goal. One thing I think about a lot because I am not a data scientist is: is all this sharing just for data scientists right now because if is not, and I hope the answer is that it is not, how do we make it so that people who are not data scientists can we make the most use of the data to generate the scientific advances that we are all looking for.

If you would like to put your hand up before you speak that might help us to manage things a bit more but if no one else is talking and you want to jump in, please do so. And just one more housekeeping thing this is not a Q&A about the presentations. If you would like to direct a question towards one of the presenters or about the presentation that is fantastic. But everyone who is here is here because you are a stakeholder in the data sharing research community, in the Alzheimer's community, and I am sure you have your own ideas and topics that you would like to bring to everyone's attention.

So Heather's hand is up.

typical evaluation processes, for example. As we move forward on all the fronts discussed in this workshop, we also need to address this critical backdrop that will determine the success or failure of the enterprise.



Dr Magali Haas

Great discussion and list of topics - would love to see these clustered. Additional topics:



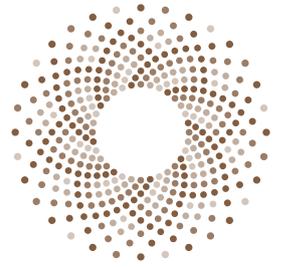
Dr Tetsu Maruyama

Thank you, Yves, I agree that this is central



Professor Craig Ritchie

What Ivan points to is that we will feed back data on 'incidental findings' that are actionable and therefore worth notifying research participants of. As 'science' and (maybe) drugs come on line then what is 'actionable' may change? This can't be a new problem in other disease areas.



Dr Heather Snyder
Senior Director, Medical & Scientific Relations,
Alzheimer's Association

I chatted my question but maybe I figured I'd say it as well. Hi everyone, it is great to see so many familiar faces. I am thinking about the presentations, and the different types of tools that are out there. I know there are many on this dialogue that are representing different ways of data sharing and I am wondering what is rising above all of that. What are the tools we as a community need to develop or be thinking about for the entire community? It doesn't really matter where you are sharing, or what kind of data you are sharing, because I know there are different tools for sharing different data. But where are the ways where we can inform someone about what are the considerations of GDPR, what are the considerations if you are in this country or that country and want to collaborate with that country. What are the things that you need to be thinking about? How do you understand as a researcher what some of those needs are or steps are? So thinking above all the different tools and platforms that are represented here what does that look like for us? I think you Andrew said the next generation of Einsteins are those that share. So how do we help give the next generation those tools? It is not really a question for anyone; I would welcome people's thoughts on.



Dr Tetsu Maruyama
Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Great question Heather. And by tools you are referring specifically to those legal and governance tools?



Dr Heather Snyder
Senior Director, Medical & Scientific Relations,
Alzheimer's Association

Maybe. It could be that. It could be educational tools. It could be a 10-minute YouTube video that you watch. Sometimes we make it really complicated. And perhaps it needs to be with the legal and governance tools. But sometimes a quick video or a short tutorial or something. I know Art Toga is on the phone and a lot of what we do with his team with GAAIN and LONI are done more on a one to one connection on linking groups together, and following up with those groups individually on their needs. But what are those bigger picture opportunities that can link teams together?



Dr Tetsu Maruyama
Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Art as you were called out would you like to jump on that one?



Inez Jabalpurwala

Patient engagement also means involving those with lived experience in establishing research questions and priorities.



Dr Sean Hill

@Yves - I fully agree - there is a need to change the INCENTIVE system



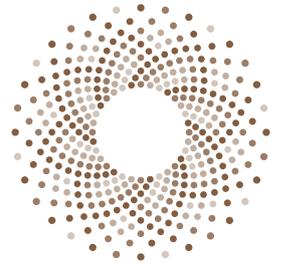
Professor Maryann Martone

@Yves at our NASEM workshop on Changing the Culture of Data Management and Data sharing, we had a broad variety of perspectives on this topic and several of us who have been working in this area for a long time noted a palpable shift in the conversation from why should I to "how do I" and "I need more tools."



Dr Magali Haas

Inter-operability across platforms;



Professor Arthur Toga

Provost Professor of Ophthalmology, Neurology, Psychiatry and the Behavioral Sciences, Radiology and Biomedical Engineering, USC

Sure. I am going to answer that slightly obliquely. I think when thinking about the notion of trust the investigators need to be included. All of us have experienced investigators having specific requirements about how they want to share. That may be the opposite of what the funder wants, or what the community wants. But we have to engage them in a way that provides them with a motivation to share. And that means whatever rules they want to impose we have to somehow address those. And that may be by informing them it is not necessary or respecting those things. And I think that is a very important thing for encouraging sharing from a project-by-project basis which is often overlooked. Science is competitive by its very nature in terms of funding and publication and career advancement. So, mechanisms to address that need to be considered by whatever solutions we envisage here.

Common Data Model for Brain? ; Citizen Scientists - how can they be engaged appropriately? Data Curation - how do we make this easier (refer back to Data Model/Standards);



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

And that would be included in any toolkit we are mentioning here. There are tons of hands up and I would also ask that you put your hand down after you have spoken. I think Joachim was next.



Dr Sean Hill

@Inez - fully agree, and find it to be extremely valuable.



Dr Joachim Schultze

Founding Director of Systems Medicine, German Center for Neurodegenerative Diseases (DZNE)

Thank you very much. I am Joachim Schultze from DZNE in Bonn, new to the Alzheimer's field. I am coming from the omics field bringing this to DZNE. You will all know Pierluigi Nicotera [Scientific Director of DZNE] who asked me to join today. I would like to challenge today's theme, the open science and data sharing. I think these two things are often brought together but if you look at GDPR these do not fit all the time. If you go into medicine, and I am a medic by training, we have physician-patient privilege. Doctors don't share data; they share insights into diseases. I am not sure why we are trying to adopt something from computational sciences, which is data sharing, rather than insight sharing.



Dr Magali Haas

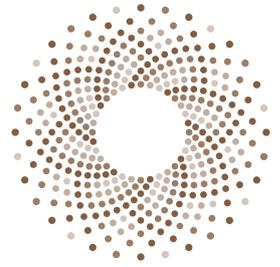
How do we advance THIS conversation beyond AD — why isn't this a WHOLE BRAIN discussion?



Professor Maryann Martone

So I think that many more-scientific societies, research administrations-are willing to engage compared to before. But they need training, worked examples and tools. More than BRAIN @ magali. Whole body.

If you do insight sharing than data protection is a different game. Because you will hold the data locally and this would follow much more closely the GDPR rules. You just have to force the computational science field to develop the tools which would allow us to share insight across nations rather than foster data sharing only. This is what we are doing, we have a paper in Nature about Swarm Learning where we really dissect this and make the case: it is not about data sharing in the future, it is about knowledge and insight sharing. And by the way data sharing also means data duplication, data transport, both are against green computing because you need more energy when duplication and moving data.



So, if you conceptually move from data sharing to Insight sharing you can change the computational infrastructure long term, and this will help bring computational science and medicine better together. Because then you are working on the traditions of medicine, which is learning together. Medical people always learn from each other, and with each other. And data from patients is not shared it is only the insights that are shared. So, I would just like to throw that open to the floor.



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

It is a great point and hasn't come up before. Clearly insight sharing would be part of any community of data sharing or open science. But would anyone like to stick up for data sharing? Pawel I think you have had your hand up for sometime and then Miia.



Dr Paweł Świeboda

Director General, Human Brain Project

Thank you Tetsu. I am Pawel from the EBRAINS Research Infrastructure and Human Brain Project. What I wanted to draw our attention to is the distinction between findability and actual re-use. Because very often attention is focused on the availability of data sets but much less so on re-use itself. And at the end of the day it is re-use that matters. So, it starts from the fact that people should deposit data in data repositories that other people want to use. Easier said than done, but that is extremely important, because we need valuable data sets that other researchers would want to re-use.

Speaking now as someone who leads the project developing tools and services to work with data. I think what is important is to get close to people sharing data early on and accompany the journey. The way we do this is we develop personalised workflows. We have high level support team. We try and work with the researchers to understand the question which they are struggling to answer.

The third element for me is the need for high end analytics. We need to extract value from data in other words and the only way to do that is to deploy a range of tools at our disposal. The way we do this is providing supercomputing and HBC facilities, on to curation and annotation services all the way to personalise virtual brains that enable, that generate, the predictive power. So, for me, to sum up, it is about creating this virtuous cycle. It is reuse that leads to standards and standards that facilitate reuse. But we need to accompany the journey and amplify what the researcher is working with through high-end analytics.



Dr Lara Mangravite

@ Yves - yes. We see this with data sharing in the pandemic. Once incentives aligned, the research and health communities were able to easily share data using the existing protocols and systems we have. Outside of the pandemic, our scientific reward system creates incentives that are often perceived to be at odds with data sharing.



Dr Elisabetta Vaudano

Agree very much @ Magali!



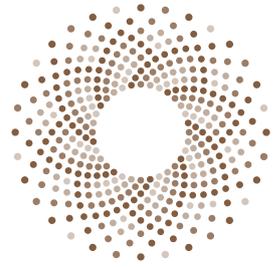
Dr Magali Haas

@Maryann that's where the interoperability model comes in but shouldn't we organize 'ourselves' first?



Inez Jabalpurwala

@Magali - 100%. The brain is one system within the



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Thank you Pawel. I called on Miia and I will get to you, but Maryann has had her hand up vigorously for some time so if we can go in that order.



Professor Maryann Martone

Professor Emerita, UCSD

I had a few comments. I think in the chat John Langton is making the case for why it isn't just sharing insights but it is sharing data. One needs to share both in a computable and FAIR form. But the idea that individual insights working with limited data sets are leading to anything significant is the thing we are trying to work against because that has been the paradigm for the last several decades. And so, I think the idea is one can scale up. We are in the age of compute and human assist. They work in tandem. Computers can do things that human cannot. They can integrate data at a scale we cannot. At the same time human insight and human skills can do things that computers cannot. So, we develop our systems and machines so we can do both in tandem.

Secondly there was a couple of points I have made in the chat. One of the was about FAIR. There is a lot of talk of incentives and other sort of things. But my experience working with a lot of projects, especially coming from the clinical domain, is that when we talk about FAIR we are especially focussed on harmonization and reuse which is really critical. And the sort of wider open FAIR community they focus a lot on findability and accessibility. And that means making sure we have provenance meta-data, and we have persistent identifiers, [and full citation metadata] that can tap into a system of data citation that is being developed slowly but is coming into use. So, I would encourage to make sure there is engagement with all of these resources that are being developed by that other part of the community that is largely publisher driven and library driven. To make sure that we are creating these data resources in a way that supports this whole system of data citation that can help us to learn to treat data as a primary product of research so we can start to develop a credit system around it.

I was also very struck at the National Academy workshop that I co-chaired, and Lara was at that as well, if you look at data sharing as a whole, it is about 1% of the papers that were published that had data associated with them. So, getting some sense of perspective is important. Having being doing this for years, there is always palpable excitement about how much is being shared until you match that against what potentially could be shared. And understanding in the community about how pervasive it is, how well supported it is, are critical. This has been brought up before. I think from both the basic sciences and clinical sciences is really important to know. And having those types of metrics is really important.

And finally, I remember vividly a few years ago I was brought in a UCSD on a potential project that was on Alzheimer's and I kept saying don't you want data from things that are not Alzheimer's and they said no they were just going to focus on Alzheimer's data.

human system and impacted by the environment. This adds to the complexity, but also the promise that data sharing can yield broader impact.



Dr Angela Bradshaw

@Ivan, a very important point with associated ethical challenges that needs to be addressed head-on, in consultation with people who have lived experience of dementia.



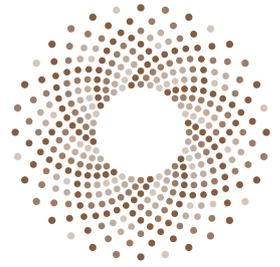
Dr Lucy Stirland

Wholeheartedly agree with Elisabetta that we need to consider dementia in the context of a person's overall health (>80% of people with dementia have at least two other conditions)



Dr Sean Hill

@tetsu - fully agree



I think that is a real lost opportunity because we know the problems of comparing Alzheimer's to controls. But really being able to broaden this out to multiple domains is where the power of data sharing and machine learning is really going to come into force.



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Great thank you. Miia if you have something to share to that because I know you have been collecting data that goes broadly beyond Alzheimer's disease. And then we have a tremendous backlog of hands which I think is very exciting.



Professor Miia Kivipelto

Professor in Clinical Geriatrics, Karolinska Institutet (KI)

Thank you. First, I want you to thank all the speakers for the very interesting talks. This is such an important topic so thank you for the invitation. I have been working with many European projects and studies that have not been primarily designed for Alzheimer's but other chronic diseases and also worked on several global initiatives. I have three short reflection, based on lesson I have learnt so far.

The first is related to what Heather mentioned about the tools. We have learnt so much it is fair to say to know that data sharing is the future, but it takes time. So how can we help researchers get there in a quicker time? It takes time so when we are planning the timescales in the studies it is important to take this into account. And now in many of the European calls [indistinct] we are having dedicated packages for data sharing. So that we should use this so not each teach has to plan work from scratch. So, let's put this as a focus area taking, the lessons learnt and dedicated tools or tool pots for the researchers what we have learnt about data sharing, so there are facilitators and carriers in that sense.

The second point I would like to make is about the data quality. I am increasingly thinking: why do we want to share, what is the purpose? We need to keep that as the primary call and focus on the data quality of what we want to share. It is not always possible to focus on the quality if the data is already collected but in the cases we can effect that, now we have the WW-Finger with forty countries and we have the prosepct of data harmonization where we can get data that is easier to share and compare. So that is something I am thinking about more, to have more joint initiatives and really plan what data we are collecting and sharing. This is not only clinical data but biomarker data, how we are collecting blood samples and so on.

The final point is back to the willingness to share. I think someone mentioned that patients and participants are willing to share if they know how and there is the trust. How can we have the mindset for the next generation of researchers that says data sharing is something we should do? Again, I think it comes down to why? In some cases



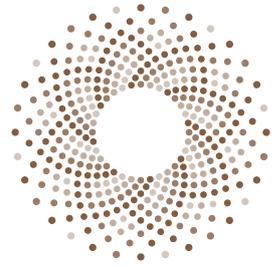
Dr Ivan Koychev

Thanks Craig - exactly; I think the issue of biomarker confirmed preclinical/prodromal dementia goes beyond incidental finding as it often is the primary goal of the study; I believe there is a risk that biomarker-confirmed participants are seen as a super valuable but depletable resource for the researchers and are at risk of not being informed of potential opportunities for disease modification either through approved or investigational agents



Professor Jane Roskams

Agreed! a dementia patient is a human first - with multiple confounding variables. Gathering these critical pieces of information in a humane ethical way will be what we need



it is easy to see how you can increase the power to do some sub-group analysis. Quite simply things like sex differences have not yet been studied a lot because we are not always powered in our own studies. So, there is much more we can do. It would be so important for diversity. And also for the data from studies that are not primarily for Alzheimer's disease; helping these long term follow up studies that are not so easy to start. In conclusion, we should reflect on what we have learnt from covid-19 and how we can take that mind set for Alzheimer's as well.

to understand all the symptoms of "the dementias"....



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)



Dr Sean Hill

We use W3C provenance quite successfully.

@petr - agree with your points. Critical to use provenance to understand how data is produced, not only used.

Wonderful thank you Miia that is very helpful. Maybe we can go with David and then Liz.



Dr David Sibbald

CEO, Aridhia



Professor Maryann Martone

+1 for automated provenance

Thanks Tetsu and thanks to everyone who has been presenting and speaking. I want to throw in a couple of technical infrastructure approaches. Some of it has come up in some of the previous conversations. Thinking about Angela's comment that there is no silver bullet to address the GDPR requirements for data controllers. I think there are three things that I would call out technically if you like that as a community we should be spending more time on.



Dr Joachim Schultze

But this is similarly true for shared data. It is not specific for federated - or as I like to say - decentral data.

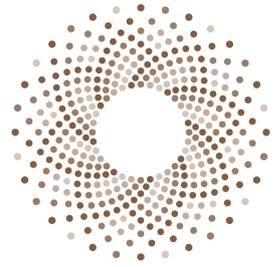
The first is the federation of data. This point about restrictions on data transfer. A federated approach is one that basically allows the data computational to take place at site and the results to travel back for consolidated analysis. So we are not shifting the data; record level data, patient level data is not leaving the site. There is a lot of work being done for many years by the Global Alliance for Genomic and Health on a federation, ADDI is pushing that field quite considerably and that is something as a community we need to rally around. Secondly, the applicability of data synthesis particularly for training of AI models is really important if we are looking to bring the next generation of data science and techniques in. We know those models need a broad data set to be trained and retrained on. The applicability of data synthesis as a means to do that is extremely important. The third area, which has come up a couple of times is the real focus on FAIR data services. And in particular I would suggest enhanced meta-data services that really allow the open publication of meta-data to a level that allows the research community to do sensible pre analysis prior to launching a data access request workflow.



Dr Pieter Visser

A challenge in data sharing is the resources needed for that at the cohort level. So funding should not only be provided for data collection itself but also data

So I think from a technical infrastructure perspective picking on federation, picking on data synthesis and picking on enhanced meta-data services from a fair data services point of view would be well worth our time spending more time and effort and energy on.



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Just briefly to comment on David's comment and thank you for that. It feels like it is clear some of the issues that are coming out – and I am trying to scan the rapidly evolving chat while we are doing this as well. The issue of trust keeps coming though. The issue of incentives. How are we rewarding researchers? How are we creating and maintain incentives for research participants? How can we share data as well as sharing insights? Some of these are really going to come down to technical solutions. But some of them are frankly about sociology. They are about understanding the community and how to reach into it in a trusted and effective way. But things like federated data, synthetic data are potentially technical fixed that might get us to where we want to go without putting data at risk without disenfranchising any of the research participants.

I don't know if any of the speakers I want to call on want to address those or move the discussion to another point. But we are going to go to Liz then Jane and James please.



Liz Brandt

Co-founder and CEO of Ctrl-Shift

Thank you for the invitation to take part in the meeting. I am also not a dementia specialist! I am a personal data specialist. My company called Ctrl-Shift have been working on the sustainable use of personal data to enable growth and value. Where that has really focussed has been on is how the individual can use their data to create new value for themselves but also to enable that data to be used by multiple organizations. In some of the simplest forms that could be a user journey. But some of the more complex forms are where you are into research. We have been doing that for 11 years.

We worked initially on the Mydata programme in the UK and then on the development of GDPR which got an interesting overview earlier on. The main bit of GDPR we focussed on was data portability which I noted didn't really get much of a mention and doesn't generally get much of a mention at all. But data portability gives the individual the right to access their data and make that data freely available as they see fit. And I think it is something that is overlooked massively because there are a lot of barriers to enabling that. We did a piece of work for the UK government on data portability and the opportunities for growth for the UK economy which has been widely used by multiple countries across the world. This work looked at some of the challenges and barriers to do that which have been very much reduced now. And many of the people who have spoken has spoken about some of the ways that those challenges are being reduced.

We also work with the EU DG connecting [indistinct] who are one of the people who are designing some of, the next, regulation in the EU around data. They have designed the proposed Data Governance Act. This is putting some governance structures and quality structures around data intermediaries. Those intermediaries act on behalf of the individual to enable that data portability to happen. Which means that those data intermediaries can make that data available. I would go back to what Joachim was saying.

sharing later on. A nice example is the Netherlands Consortium of Dementia cohorts that provided funding for datasharing- although with funding for 5 years.



Dr Magali Haas

@Alan – completely agree - pipelines should be 'accessible' too! VETTED too!



Professor Maryann Martone

They should be FAIR tools



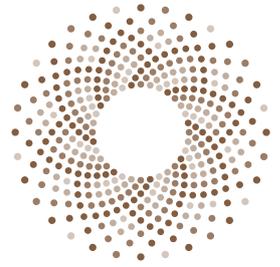
Dr Heather Snyder

I'm sorry that I have to step off. Thanks for the discussion Tetsu and Lara, and thank you Lenny and Josh! Look forward to our continued dialogues together. Take care.



Dr Sean Hill

@Alan - fully agree on University level recognition of open science



It is not necessarily moving the data around, but it is making the data available. So, the data could stay in the original location and be accessible in a linked data form using those data intermediaries. The data intermediaries provide consent and use and access capabilities.

So, we have been working with them on that. And I will just go through some of the other things that we have been working on. So, we are on the working group for the open banking. Which may sound a divergence from this conversation but actually it a structure that allows data to be shared from banks, from retail bank accounts, about individuals, and making that available across the piste in a very structured way in a very highly governed way by the Financial Conduct Authority. But that has spawned something called GOFCoE, which might be interesting for people to look at. The Global Open Finance Centre of Excellence. But what they have created is a longitudinal data set of people's financial data that enables multiple organizations to look at the impact of things like policy making on the economy. So that is a really interesting example that might be really useful for people to go and have a look at. We have been involved in the design of that as well.

We have also been involved in the design of Open Finance which is opening up pension and insurance data and also in the design of smart data that is opening up energy and telephone data and also looking at health data. And all of that is very useful for this conversation. A lot of the sharing of financial data is very similar and a lot of the power that you can get from this data can be seen in some of these examples. So, some of the tools that are being created to do that actually enable some of things that you want to do with research around dementia. We are also working with the working group from the All Party Parliamentary group on longevity on what is called open life data which plays to what Miia mentioned. It is not just health data but other data outside health setting. So, data from my finances, from my activity, from sleep, data from my life basically. And we are looking at how to create standards and governance for that data across clinical and non-clinical settings. We are also working with the EU common data health space team to look at how we create common data health spaces in Europe. And we are working with the World Economic Forum on a code of conduct for personal data sharing. So, you can see that there is a theme behind this which is personal data sharing under the control of the individual. We are building out a number of Sandboxes and services associated with that.

One is a mental wellbeing sandbox which we worked on with Facebook, HSBC, Public Health England and togetherall, which is a mental health service provider. That sandbox looks at, and this is where it all starts to come to life, it looks at how you enable and empower individuals to get to access to multi-sector data: that might be activity data it might be finance data it might be media consumption data. This data enables them to understand what is happening with their mental state. And it comes back to something another contributor, I think it was David, was saying around sharing of insights. Where the individual gets to share, to analyse what is going on with their life, act on it and/or share that insight at a moment in time. But that insight can to multiple places and can do multiple things for that individual including being able to be shared into research. What we are looking at that is how do you create value for the individual, what sort of services you can create. But also how you can create sustainable data sharing and a data flow with a business model that sits behind that reflects that.



Dr Jennifer Yokoyama

@Alan nodding vigorously at the point of EMPOWERING those developing tools as well — from all regions of the world



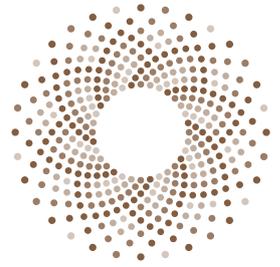
Professor Maryann Martone

An academic culture based on Research Objects. Creating the culture involves broad engagement along with mandates. Let them be active participants in defining this culture.



Dr Elisabetta Vaudano

Thank you for very inspiring conversation while you wrap up you are all warmly invited to the IMI impact on Dementia event next Tuesday 2-4pm Brussels time! You cabn register here <https://www.imi.europa.eu/news-events/events/imi-impact-dementia-event>



One other point I would make is on the trust point which I know we have been focussing on. We have been working on trust metrics as well and how you create trust metrics. And we are working with Which, a consumer champion group in the UK, on creating a trust framework to enable this data to flow.



Dr Tetsu Maruyama
Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Wonderful thank you very much I appreciate it. Jane was next, I think.



Professor Jane Roskams
Professor of Neuroscience, Centre for Brain Health,
University of British Columbia

Right thank you Tetsu I was going to pick up on a couple of points that Liz just brought up but also bring it back to Heather's original question around what we need and what tools do we need. We have a great collection of people on the screen right now who have very different skills to bring but who have very different levels of understanding as to what dementia data is, are, where they came from, how they were obtained. I do think there is a really need for clarity on that front if we are going to engage the talents of even the people on this call. So, Heather I agree. If we are going to get to that goal of getting the most out of the data that can be shared and made available, we should be creating roadmaps for people. Whether it is videos, clickables, and so on. Here's how MRI works and here is how it looks in an Alzheimer's patient versus not. So that a data scientist who knows nothing about the brain can look and begin to apply their expertise to that.

I do think there is room for Joachim suggestion of insight driven analysis. With the [indistinct] everyone might not have been standardised in the same way. The human genome project showed us some of the ways to do this. Most of the progress in that wasn't led by geneticists it was led by mathematicians and computer scientists and informaticians before we had this field of health data science or genome data science. We need to let those people in. And in order to let those people in we need to do the kind of thing that Heather suggested which is show the value of how we find the data and provide the data that is made available.

And last piece that ties these things together and all of us who are involved in different brain data sharing initiatives have to tackle on a daily basis is this whole idea of a data quality. The brain images have their standard. Just because you fall beneath that does it mean the data that is available can't be analysed? The whole idea of data inspection protocols that could be applied regardless of what country your data happens to be sitting in. So, the idea of the quality of data for analysis gets to the forefront is absolutely crucial if we are going to be able to move the needle. It is not exclusive of the kind of insight that Joachim came up with. But I do think we need to bring it back to that because the whole idea of generating the most out of the data we have means that we are working with the best data, and we can apply brand new tools that people who don't understand one brain region from another can help us develop because they are the people who are going to crack this and hopefully before the rest of us develop dementia.



Professor Jane Roskams

Thanks Josh - and Tetsu and Lara - for convening this. Let hope we can come up with an Action Item list to take this internationally to a different level of impact



Professor Maryann Martone

Thank you so much for inviting me



Professor Alan Evans

Excellent session !



Dr Sean Hill

Thank you very much for convening! Very happy to have taken part.



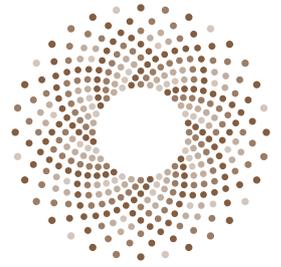
Dr Saskia Sivananthan

Thanks all, fantastic discussion



Dr Elisabetta Vaudano

Bye



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Absolutely and I think these are great points. And bearing in mind as we go through these discussions that we really are working to build a community of users as well as the community of data. We often think about the data and how we want to share that. I often think, as I have mentioned before, as a sort of scientist and a completely data naïve person that we need to building out a way to bring non-data scientist in to use the data better but at the same time we need to find a way to enable data scientist that are not specialists in the field to use that. James, I called you then Ivan, Elisabetta and Rebecca.



Professor James Rowe

Professor of Cognitive Neurology and Director of Cambridge Centre for Frontotemporal Dementia and Related Disorders, University of Cambridge

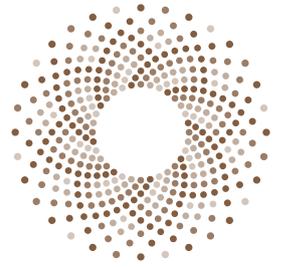
Thank you. James Rowe in the UK. I think there has already been a major paradigm shift in the research community who now see it as a win-win for data sharing and as we have heard today there are increasing tools and platforms to make it easy, and solutions to issues which is a credit assignment. But we must not forget the patient and public stakeholders. We have to make sure we bring them with us. But they are different. I think most patients are supportive of data sharing. They are often surprised we don't do it already and in the consent process they question why we are making such a fuss when they want the most benefit to come from their participation: "why we don't just get on with it?". The general public are different, perhaps especially in the UK, and maybe others countries as well. They are more sceptical. So, there is a major job of work to do to close the gap between trustworthiness (which is high) and trustedness (which I think is still low). But unless we close that gap, the data won't be representative and inclusive. Unless we have representative and inclusive data then all the advances we get in diagnosis and new therapeutics we will only serve a minority of the population, and not reach out to all those who are at risk of dementia. Limited representativeness of data would put a cap on the value of repurposing of data. We also need new data from communities that haven't traditionally engaged with research studies.



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Absolutely. Thank you James for bringing that up and it has been running though some of the chat throughout the session as well. When we are talking about inclusiveness, we are talking about both research participants but also scientists who are coming from communities that are not typically being reached. That is all very important. Ivan, I called you out next I believe.



Dr Ivan Koychev

Clinical academic psychiatrist, University of Oxford

I am Ivan and also in the UK and similar to James I work with Dementia Platforms UK. I wanted to pick up on a point which I didn't hear which is we are talking a lot about the benefit of sharing data for the benefit of researchers and research in general. But as a clinician and an academic I also want to be mindful of the benefit of data sharing for the benefit of patients. And I think this is particularly relevant now that we have the aducanumab decision. So that we can end up being in a situation where a lot of cohorts have data that is relevant to people's health care. So, we have data on biomarkers for example amyloid positivity. At some point we need to be very clear whether a research registry has a duty to inform individuals about their amyloid status or to link them up to interventional studies that are local. And that is not really a topic I am hearing that much about, and I did wonder whether anyone else has any thoughts about it. My feeling is we want to be ahead of the game rather than people realising that someone has this potentially very important information about their healthcare for which there is now a disease modifying treatment that they haven't been informed about.



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

So, if I understand what you are raising here the further data is shared the more data that could be extracted from it the more the ethical burden on the relationship between data and patients.



Dr Ivan Koychev

Clinical academic psychiatrist, University of Oxford

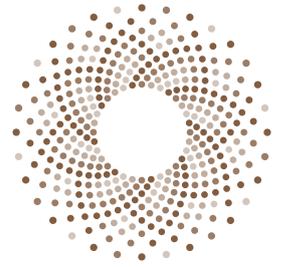
If I could just clarify. So far there hasn't been a burden on researcher registers to share data on pre-clinical dementia because there hasn't been a disease modifying treatment. But we are in a new world now where suddenly there is this option and I think we have to as organizations have a very clear policy. We are holding this biomarker data at what point does it have to be shared either with the patient or healthcare provider or with intervention studies?



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Very interesting point. Maybe we can circle back to this a little bit afterwards. Elisabetta, Rebecca, Alan.



Dr Elisabetta Vaudano

Principal Scientific Manager and portfolio manager for the neuroscience area at the Innovative Medicines Initiative (IMI)

I want to come back to a fundamental question. We talking about sharing data but what is the aim? It is to get to knowledge isn't it? What are the different elements that altogether we need to consider to get from data to information and finally knowledge? Different elements have been mentioned in this conversation. Of course, one is the willingness of people. So that is incentive for individual but also incentive for the group. So, we need to find elements that can be recognised from both or if they are individual need to be considered. But also, we need to consider all the elements that have already been mentioned. Like data quality. But also, the point that there is a strong element of education that needs to be integrated at all levels. You need to go right back to the universities. There should be education on data sharing and on the value of data sharing that is part of education from the very beginning. There should then be something in the career path to reward this aspect.

Another thing I wanted to say briefly is that we have a lot of learnings, a lot of infrastructures for data sharing already. There is a lot out there. And the problem for everyone is we are drowning in information. It will be very important to get a catalogue of what has been already done and not just re-invent the wheel every time. Because this is a constant danger we face.

I think we need not only to talk about Alzheimer's data because find me an Alzheimer's patient that is a "pure" Alzheimer's patient. He/she will most probably be a patient with multiple morbidities, cardiovascular, diabetes and so on. And there is a lot of learning and data in this case. Last but not least, we should also look at other disciplines, not just computer science but finance is another, something like block-chain for example comes to mind. We must not work in silos we have to try and work but consider new disciplines that we thought were far away and get them in the conversation. And don't forget the regulators.



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

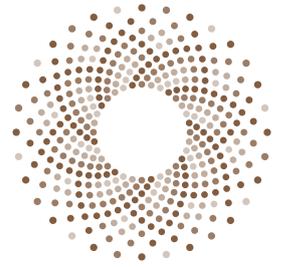
Thank you very helpful. Rebecca.



Dr Rebecca Li

Executive Director of Vivli and on faculty at the Center for Bioethics at the Harvard Medical School

Thank you so much. I will feed off your remarks. I do wish data sharing education would start much earlier in university. I do think that would be helpful. I represent a global data sharing platform that shares not just Alzheimer trial data but other dementias and other diseases as well in a secured way using managed access. And really, I would



enthusiastically endorse the idea of data sharing including access to de-identified or anonymised data in order to in some cases to pressure test other researcher insights, and in some cases to drive new scientific insights to drive progress and support reproducibility. And this is what we are seeing when this data is shared. So, we are really seeing the fruits of data sharing coming out at the other end in terms of publications.

We are seeing people integrating data not just with Alzheimer's and dementia but other disease areas as well. So, I would really echo what Elisabetta said. They are really taking Alzheimer's data and integrating it with other diseases. What we see are the insights are found when we integrate that data with other closely related diseases and clearly you are saying, and I am not an Alzheimer's specialist, that you rarely get purely Alzheimer's patients without co-morbidity. They are looking at the integration of Alzheimer's data and covid data. So, this is all driving new insights in ways we never imagined.



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Rebecca absolutely and thank you, and Elisabetta, for your contributions. They really highlight for me one of the key aspects of data sharing. It would be great if everyone would look at that same single study again through the same lens and re-analyse it. But we wouldn't be building out our knowledge base as much as would if we have the opportunity to integrate data that people haven't put together before to gain insights they wouldn't have gotten. The more we think we know about a particular disease I think the less we know about it, and the more we bring in insights from healthy brain function as opposed to brain disease, bring in cardiovascular and inflammatory etc, the more we get those insights. And I think that is a real goal of data sharing.

Petr and then Alan and we will gliding into a landing here at the end of our time.

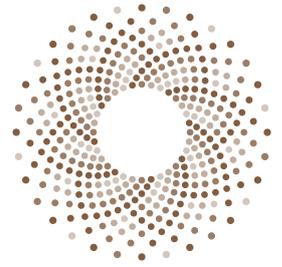


Dr Petr Holub

Associate Professor of computer science, Masaryk University

Thank you I will be very quick. I am no means an Alzheimer's specialist I am a computer scientist. I would like to react to two topics that have been raised.

One of which Joachim raised on sharing insights and knowledge. And one thing I think as a computer scientist is we actually need the data sharing but the data sharing is the transference [indistinct] into the model sharing. Eventually what you are sharing at the end of the process is the trained models, the AI models, machine learnt models and also other tools that are another form of knowledge. But for that we need high quality data. A number of people highlighted federated learning here. But from a computer science perspective a huge challenge with federated learning is quality of the data and the consistent semantics of the data. Because when you do it you can't revisit the data manually and investigate whether there are any doubts and problems. It has to run on the data that has consistent semantics and quality.



This brings me to the second point which is distributive provenance. Were as [indistinct] and I were working or leading an iso-standardization of distributive provenance that also aims to deal with the privacy issue by supporting opaque components as part of the provenance. So, you don't need to reveal privacy sensitive parts of the provenance but still have them in a trusted way. So, there is someone who can provably go back to the source. This has many benefits. It has benefits for the patients because this is also a mechanism that can be used for tracking back for incidental findings. And eventually only the authorised person, which is in all likelihood the primary care institution, is the one that can resolve which patient the finding entails. And it could also help with the incentives because then you can trace back how the data is used you can track data citations. The provenance can't be manual. And that is why we are heading for the iso-standardization. We have to develop ways where the data is automatically generated because the provenance information has to be generated automatically as well because otherwise we will never have documentation of the data quality at scale.



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Critical trust issues. Thank you for that great insight. Alan lastly but very much not least.



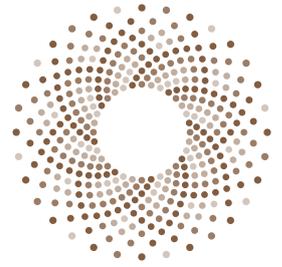
Professor Alan Evans

Professor of Neurology and Psychiatry, McGill University

Thank you very much everyone. It was a pleasure hearing Andrew Morris saying Wales was leading the pack in terms of innovation! I am here representing the Canadian open neuroscience platform and there are two points I would like to make.

First, this should not be just about data sharing, which has been adequately justified, but also tool sharing. We want to empower a large community across the globe in low and middle countries such that they can get at data and analyze it for themselves, rather than just being providers of data who never hear about it again. I think tool sharing and empowering a wider community is a central tenet of open science.

The other point I would like to make, it has been mentioned but I think it deserves emphasis. We have to change the culture of the universities. Right now, promotion committees often don't understand the importance of open science or a data paper. When one student puts out one paper that is analysed by 100 different groups, they still only get recognition for one paper. There is a culture change that has to take place at the level of the university system. One thing that we could do is to put out a position paper for broad dissemination that emphasises (i) the importance of open science and (ii) the need to incorporate open science factors such as "data papers" within traditional university mechanisms for recognition and advancement.



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Thank you Alan I think that is a great point and data sharing is part of an ecosystem of open science. And it is going to require a realignment of incentives. And it will not be easy. There is a lot of them in the chat and in our discussion people have touched on how we do this with industry as well? How do we square data sharing with the need for proprietary development of programmes? Obviously, this is a big issue that we are dealing with every day as well. I would love to take some more comments, but I am afraid we are about to wrap up here. I will say the chat has been a fabulously rich and Lenny and Josh inform me that we will as part of the offline record of this have access to the chat so we can go through it more deeply. And I think it will be a very valuable tool.

Things that I heard coming though, both the great presentations and then the discussion group into things like: Culture. How we are going to create a culture that enables data sharing; About tools and more general technical approaches that are required as well both in terms of governance and allowing some trustworthy structures such as federated data to become actually useful for people; Inclusiveness as refers to both data research participants to researchers themselves. And how in order for our data sharing to becoming information sharing we are going to need to be inclusive in all those difference ways. And keeping in mind the quality of the data is going to be key to maintaining and furthering that trust.

And finally, all of those things build on each other. There is not that one thing we can say if we do that one thing then everything will be solved. Each one of these aspects feeds into another aspect. So that more trusting will allow more data sharing and diversity and inclusiveness. More inclusiveness leads to more trust. More sharing leads to more tool development as we think about what we need to develop to make things safe etc etc.

So, Lara if you have any further comments you would like to make or otherwise just nod vigorously.



Dr Lara Mangravite

President, Sage Bionetworks

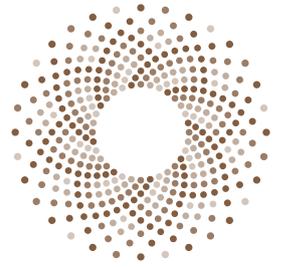
I think you have covered it well Tetsu.



Dr Tetsu Maruyama

Executive Director, Alzheimer's Disease Data Initiative (ADDI)

Fantastic. In which case I will pass it back to Lenny to close the meeting.



Lenny Shallcross
Executive director, World Dementia Council

Thanks Tetsu and thanks Lara for chairing it I would also thank our opening presenters. And of course all of you are participating especially as I notice for some of you in Australia it is about 3.30am in the morning which is an awful time to be awake. We will send you a transcript shortly and as I mentioned we will be producing a collection of essays that reflect the rich and interesting conversation that we had today both live and in the chat. And with that I wish you everything from a good morning to a good night. Thank you all.

The World Dementia Council (WDC) is an international charity. It consists of senior experts and leaders drawn from research, academia, industry, governments and NGOs in both high-income and low- and middle-income countries, including two leaders with a personal dementia diagnosis. The WDC has an executive team based in London, UK.

worlddementiacouncil.org

© 2021 World Dementia Council
UK charity registration number: 1170743

Cover image editorial credit: Shutterstock.com

