

OECD (2014), “Unleashing the Power of Big Data for Alzheimer's Disease and Dementia Research: Main Points of the OECD Expert Consultation on Unlocking Global Collaboration to Accelerate Innovation for Alzheimer's Disease and Dementia”, *OECD Digital Economy Papers*, No. 233, OECD Publishing.  
<http://dx.doi.org/10.1787/5jz73kvmvbwb-en>

OECD Digital Economy Papers No. 233



# Unleashing the Power of Big Data for Alzheimer's Disease and Dementia Research

**MAIN POINTS OF THE OECD EXPERT  
CONSULTATION ON UNLOCKING GLOBAL  
COLLABORATION TO ACCELERATE  
INNOVATION FOR ALZHEIMER'S DISEASE AND  
DEMENTIA**

OECD

Unclassified

DSTI/ICCP/IE(2013)13/FINAL

Organisation de Coopération et de Développement Économiques  
Organisation for Economic Co-operation and Development

17-Mar-2014

English - Or. English

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INDUSTRY  
COMMITTEE FOR INFORMATION, COMPUTER AND COMMUNICATIONS POLICY

Working Party on the Information Economy

UNLEASHING THE POWER OF BIG DATA FOR ALZHEIMER'S DISEASE AND DEMENTIA  
RESEARCH

Main Points from the OECD Expert Consultation on "Unlocking Global Collaboration to Accelerate  
Innovation for Alzheimer's Disease and Dementia"

JT03354446

Complete document available on OLIS in its original format

*This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of  
international frontiers and boundaries and to the name of any territory, city or area.*



DSTI/ICCP/IE(2013)13/FINAL  
Unclassified

English - Or. English

## FOREWORD

This document reports on key issues raised at the OECD expert consultation "Unlocking Global Collaboration to Accelerate Innovation for Alzheimer's Disease and Dementia" held on 20-21 June 2013 at the Harris Manchester College (HMC), Oxford University.

The consultation was organised in response to a call for urgent global action on Alzheimer's Disease and Dementia (AD) by OECD and APEC countries in 2012.<sup>[1]</sup> It brought together academic experts, private sector and policy makers to further scope and obtain perspectives on an international policy and research agenda to accelerate innovation on AD.

The consultation aimed at being forward-looking and action-oriented, and specifically focused on the variety of governance issues and policies that are facilitating or inhibiting new research and developments and implementation of improved care models across OECD countries. Considerable discussion time was devoted to how the OECD could best contribute to securing greater access to data and collaboration globally, in view also of the forthcoming 2013 G8 dementia summit (since held on 11 December 2013).

The consultation was organised in collaboration with the Global Coalition on Aging, Oxford's HMC and the Vradenburg Foundation. Their financial support and that of participating member countries is gratefully acknowledged. This report was written by Elettra Ronchi of the OECD Secretariat based on input from workshop experts, the OECD Committee for Information Computer Policy Committee (ICCP) and the OECD Health Committee. It summarises discussions with a focus on the challenges and opportunities of ICTs and Big Data and incorporates post-consultation feedback received on possible next phases of OECD work. The ICCP Committee approved this report in January 2014 and recommended that it be made available to the general public.

© OECD/OCDE 2014

This paper is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

---

<sup>[1]</sup> OECD-APEC workshop "Anticipating the Special Needs of the 21st Century Silver Economy: from Smart Technologies to Services Innovation", held 12-14 September 2012 in Tokyo, Japan

## TABLE OF CONTENTS

FOREWORD.....	2
EXECUTIVE SUMMARY .....	5
ACCELERATING INNOVATION FOR ALZHEIMER’S DISEASE AND DEMENTIA IS A GLOBAL PRIORITY.....	8
The rationale for early intervention .....	8
ALZHEIMER’S DISEASE: A COMPLEX SCIENCE PARADIGM .....	10
Is Alzheimer’s one disease or a constellation of different diseases? .....	10
Addressing the complexity of neurodegenerative disorders with computational approaches .....	11
UNLEASHING THE POWER OF BIG DATA.....	13
New paradigms of big data use in health research and clinical care.....	14
CREATING VALUE FROM BIG DATA: ADDRESSING THE MAIN CHALLENGES .....	24
Challenge 1: The need for compatible “big” data governance frameworks .....	25
Challenge 2: Financing innovation and sustainability .....	27
Challenge 3: Data sharing and linkage .....	27
Challenge 4: Patient consent .....	28
Challenge 5: Timely dissemination of findings and data.....	28
Challenge 6: Open data strategies to accelerate innovation.....	29
CONCLUSIONS: BIG DATA RESEARCH ON ALZHEIMER’S AND DEMENTIA IS AT A CROSS-ROADS .....	31
ANNEX I : DRAFT AGENDA.....	34
NOTES .....	36
BIBLIOGRAPHY .....	37

### Figures

Figure 1.	Impact of a 5-Year Delay in Onset on Costs, Americans Age 65 with AD (2010-2050) .....	9
Figure 2.	Complexity of Brain Disorders.....	12
Figure 3.	Sequencing cost per genome, 2001-11 .....	14
Figure 4.	Temporal Sequence of FMRI scans.....	16
Figure 5.	Sensor Events in a Residential Facility .....	17
Figure 6.	Tweets containing H1N1, swine flu, or both from May to December 2009 .....	19

**Boxes**

Box 1. Epistatic interactions in Alzheimer’s disease .....	11
Box 2. The Design Council living well with dementia challenge.....	22
Box 3. The new open-data policy memorandum, issued by the Office of Science Technology Policy (OSTP) in the United States.....	29

## EXECUTIVE SUMMARY

An estimated 36 million people lived with dementia in 2010 worldwide - more than two thirds of these individuals suffered from Alzheimer's Disease (AD) and just over half (58%) lived in low- and middle-income countries. Dementia mainly affects older people, although there is a growing awareness of cases that start before the age of 65. After the age of 65, the likelihood of developing dementia roughly doubles every five years, and it is estimated that 50% of individuals over 80 years old will be affected.

There is currently no cure for AD. Available medications can reduce symptoms and improve quality of life, but do not reverse or halt its progression. As people live longer and populations in many parts of the world age, efforts are intensifying to prepare for the significant increase in AD prevalence.

The OECD consultation "Unlocking Global Collaboration to Accelerate Innovation for Alzheimer's Disease and Dementia" was organised to stimulate discussion and expert engagement in setting out an agenda and roadmap for international action.

Questions asked at the workshop included the following:

- What are the inefficiencies in the present approaches to R&D for AD and how can these be best addressed?
- How can countries take full advantage of the Internet and computational power now at our disposal to accelerate innovation of AD?
- What data is available today and what are the key problems in international data sharing and linkage?
- What are the opportunities and challenges of open access and open data platforms?
- How can countries best co-ordinate research and development efforts and address the unevenness of present developments?
- Are there "good practices," particularly in terms of governance that can facilitate accelerated innovation?

This document provides an overview of the key issues raised at the consultation, specifically as they relate to the challenges and opportunities of ICTs and big data. It also incorporates post-consultation feedback received, specifically in relation to possible next phases of OECD work on AD and dementia.

At the consultation, experts concluded that the rapid growth of the range of data collected (behavioural, genetic, environmental, epigenetic, clinical data, administrative, etc.) and the development of large databases and their linkage can create a tremendously powerful new resource for research and evaluation. Big data has the potential to be a "game changer" in global efforts to accelerate innovation in neurodegenerative disease research.

Together with the radical improvements in information technologies, these developments now make it imperative to take stock of global capacity to undertake multidisciplinary, cross-border research and enhance knowledge transfer for the design and implementation of laboratory and clinical studies for improved diagnostics and therapy development and smarter models of care.

There is today a window of opportunity to implement an international agenda for co-operation that can sustain an effective response over the coming decades. The short-term challenge is to attain declining disease and disability rates and improve the quality of life of dementia patients amidst a steep rise in the number of older people. Taking full advantage of the informatics revolution and today's computational power is of central importance to the advancement of this agenda and in line with the principles from the G8 science ministers for publicly funded scientific research data to be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests.

Today the sources and types of data are expanding continuously - there are hundreds of new health and wellness data sources and data networks. Big data is, however, not just a quantitative change, it is a conceptual and methodological change. It will transform the way we do science and the way we deliver care. Participants agreed that the challenge is not just a scientific one.

Creating and using big data to change the future of AD requires careful planning and multi-stakeholder collaboration. Although great promise can emerge from the adoption of big data methods and philosophy, numerous technical, administrative, regulatory, infrastructure and financial obstacles emerge which will need to be hurdled to make this vision a reality. The heterogeneity, complexity and sensitivity of personal health data, in particular, pose special challenges for big data research, inferencing and clinical application.

Reports from the Ontario Brain Institute (OBI), the Alzheimer's Disease Neuroimaging Initiative (ADNI), the European Joint Programme on Neurodegenerative Diseases (JPND), and the US National Alzheimer's Coordinating Center (NACC) indicate that rigorous and compatible data governance frameworks are needed to catalyse the development of an integrated ecosystem of data that can be shared to accelerate innovation.

Data governance refers to the overall management of the availability, accessibility, usability, integrity and security of the data collected and stored. In addition, there must be policies that conform to the governance of the data in order to form a consistent and effective framework.

It was agreed that the OECD can make an important contribution by encouraging multi-stakeholder co-operation to develop compatible and coherent guidance in five key areas:

- **Sustainability** – identify the framework conditions for the development, financing and long-term sustainability of large databases for research on AD.
- **Exchange and access to data** – identify the framework conditions and incentives to facilitate exchange and access to data to strengthen research and evidence-based care.
- **Linkage** – identify policies and good practices that foster co-ordination and complementarity for the effective linkage and use of existing health and other data at national, regional and international levels.
- **Quality and efficiency** – identify practices for enhanced effectiveness of existing data sets.
- **Capacity building and training**– identify incentives to promote education and training of data analysts and bioinformatics experts to use big data effectively.

Following further consultations with experts from member countries, and building on current national and international efforts (for example, International Alzheimer's Disease Research Portfolio in the United States, OBI in Canada, ADNI, or the European JPND programme, which undertook a comprehensive

mapping exercise of the national and European research and infrastructure relevant to age-related neurodegenerative diseases in 2011<sup>1</sup>), a good starting point for the OECD would be to undertake work to identify:

- The most relevant national, regional and international large data sets for research on AD and dementia, including the functions performed and their sustainability.
- The key barriers to data deposition, access, exchange and linkage and areas for international co-operation.
- Policies and good practices that foster and enhance the performance of research for Alzheimer's from the existing large datasets.
- The value to the international scientific community and economic impacts of a tiered networking structure to facilitate the co-ordination of large data sets at national, regional and international levels.



## ACCELERATING INNOVATION FOR ALZHEIMER'S DISEASE AND DEMENTIA IS A GLOBAL PRIORITY

*"The consequences of dementia for societies and economies are devastating everywhere, in high-income countries and low- and middle-income countries alike..." (Hodin, Oxford Workshop 2013)*

In 2012, the World Health Organisation (WHO)<sup>2</sup> reported that 35.6 million people worldwide were living with dementia - more than two thirds of the affected individuals suffered from Alzheimer's disease (AD) and just over half (58%) lived in low- and middle-income countries (Alzheimer's Disease International, 2010 ). As people live longer and populations in many parts of the world expand, this number is projected to rise to 65.7 million by 2030, and 115.4 million by 2050. The fastest growth in the elderly population is taking place in India, and their south Asian and western Pacific neighbours. In China, the burden of dementia seems to be increasing faster than is generally assumed.

AD has a profound economic impact on countries' health and long-term expenditures and places an enormous emotional and financial burden on patients, their family and friends. The costs of caring for patients with AD have been extensively studied (Guterman et al 1999; Moore et al 2001; Small et al 2002; Andersen et al 2003). In the United States, these costs range from USD 157 billion to USD 215 billion annually, making the disease more costly to the nation than either heart disease or cancer (Hurd et al 2013). The human and economic toll extends to families. An estimated 10 million individuals in the United States provide unpaid care to persons with Alzheimer's or other dementias. This is likely to increase as the number of people with dementia rises year on year as demographic profiles change.

In Europe, the project "European Collaboration on Dementia – EuroCoDe" estimated in 2010 the overall cost of illness for people with AD at EUR 160.3 billion (EUR 71.7 billion for direct costs and EUR 88.6 billion for informal care). In a number of European countries, discussions are underway on the future financing of long-term care in general and dementia care in particular.

At global level, costs of care were conservatively estimated in 2010 at USD 604 billion a year, or more than 1% GDP (Alzheimer's Association Report, 2012). In the absence of a breakthrough, prevalence and costs will soar (Ferri, et. al., 2005).

### **The rationale for early intervention**

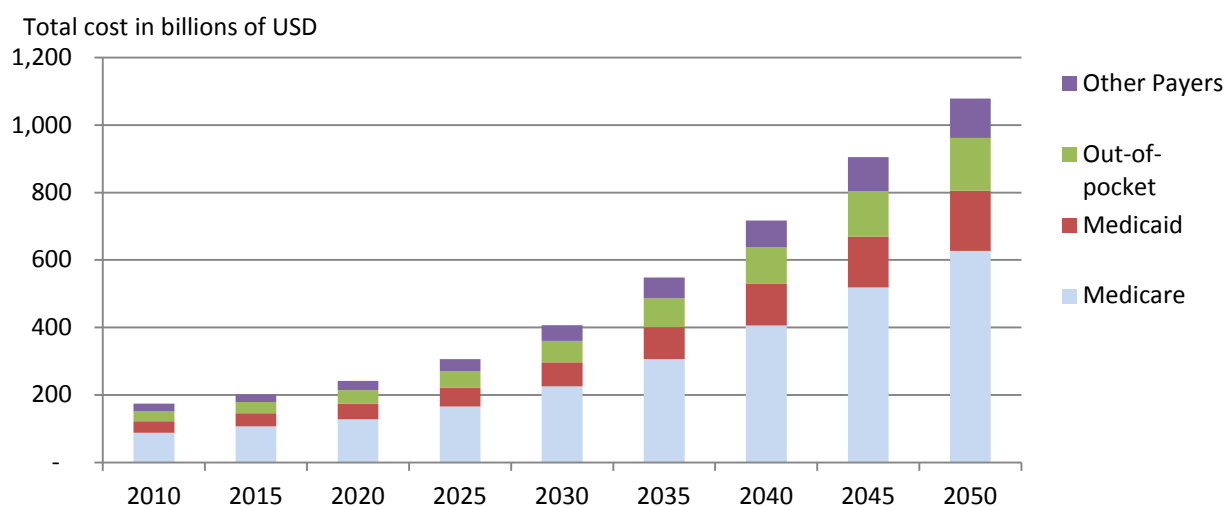
As nations around the world prepare for the significant increase in AD prevalence, efforts are intensifying to better understand the natural history and progression of the disease and develop new approaches for diagnosis, treatment and prevention. The challenge is to attain declining disease and disability rates and improve the quality of life of dementia patients amidst a steep rise in the number of older people.

It has been recently estimated that AD is related to seven potentially modifiable risk factors (Barnes and Yaffe, 2011) (diabetes, midlife hypertension, midlife obesity, smoking, depression, cognitive inactivity or low educational attainment and physical inactivity). A reduction in all seven examined risk factors could significantly reduce the number of cases of AD.

Economic modelling studies also now point to the significant returns to families and communities when Alzheimer's is identified early and disease management to delay onset and slow progression begins immediately (Weimer and Sager, 2009; Alzheimer's Association Report 2012) .

Research modelled on United States prevalence and cost of care suggests that a disease-altering therapy that would delay onset of 5 years would decrease the total number of Americans age 65 and older with AD from 5.6 million in 2010 to 4 million by 2020. Within this scenario, by 2020, five years after the introduction of the treatment in 2015, total costs to all payers for the care of people with the condition would be USD 50 billion less than would be expected without the breakthrough (Figure 1). By 2050, the reduction in total costs to all payers would be USD 447 billion; decreasing from an expected USD 1.078 trillion to USD 631 billion with the breakthrough (Alzheimer's Association 2010).

**Figure 1. Impact of a 5-year delay in onset on costs, Americans age 65 with AD (2010-2050)**



Source: Adapted from: Alzheimer's Association (2010)

The cause and pathophysiological correlates of disease progression are, however, still poorly understood. Results of disease-modifying drug trials in established Alzheimer's dementia and mild cognitive impairment have been disappointing, perhaps because these states are too far advanced for effective intervention.

What has emerged over the past three decades is the potential to monitor people at high risk for Alzheimer's disease with biomarker profiles and brain imaging. These studies have led to accumulating evidence that the neurodegenerative phase of Alzheimer's disease starts, perhaps, more than a decade before clinical symptoms of mild cognitive impairment emerge.

The disease is marked by continuous neuronal deterioration and applying a threshold to determine clinical status is therefore arbitrary and ambiguous.

*“ The clinical construct of AD as we know it is probably obsolete and a new model is needed – one that relies on measures that are sensitive at the earliest point of deflection in prodromal phases of the disease – before clinical symptoms appear “* (Middleton, Oxford Workshop 2013).

A complete understanding of how the pathology of AD leads to its symptoms will also depend on answering fundamental questions related to the mechanisms of normal cognitive function.

## ALZHEIMER'S DISEASE: A COMPLEX SCIENCE PARADIGM

*The multi-factorial nature of Alzheimer's disease requires sophisticated computational capabilities for pattern analysis of the large streams of data and observation points (behavioral, genetic, environmental, epigenetic, clinical data, etc.), the development of databases and global data sharing, federated networks. Of great promise are new smart approaches to model and assess cognitive and functional ability for early diagnosis and treatment of AD, as the primary impact of this disorder is a gradual and progressive decline in mental abilities, which ultimately results in functional incapacity.*

Alzheimer's disease (AD) is today considered the prototype problem for the Grand Global Challenge in healthcare and a proxy for a number of chronic, complex neurodegenerative diseases (Khachaturian, 2012). For decades a range of competing theories have struggled to clarify causation of AD including: genetic; toxic; infectious agents; immune function and trauma. Significant evidence supports each of these models, which are not considered mutually exclusive (Bradley, 1990). Despite decades of intensive research, the causal chain of mechanisms behind AD has, however, remained elusive as reflected in recent failures of well-designed clinical trials on promising investigational new drugs (Callaway, 2012).

### **Is Alzheimer's one disease or a constellation of different diseases?**

Much evidence today suggests that AD is the result of numerous interactions between age and gender, genetics and epigenetics, environment and lifestyle. Modifications of gene activity and expression, for example, can accumulate with age, leading to an altered response to stress and to an enhanced susceptibility to the disease.<sup>3</sup> Interactions between two or more genes and environmental factors can also play an important role in the patho-physiology of this disorder (Box 1). Individual risk factors may not only precipitate the development of dementia but also affect the course of the disease itself and survival of the affected individuals.

The pathological and clinical expression of AD reflects a systems failure in a complex neural network (Khachaturian, 2012). There is also growing and general consensus among researchers that the neurobiological underpinnings of Alzheimer's syndrome start many years before the appearance of any clinical signs (Kozauer and Katz, 2013).

The challenge of replacing the traditional paradigms of the disease with a more complex or comprehensive model of AD that takes account of all potential genetic, environment and lifestyle risk factors, the non-symptomatic phases of the disease and is still approachable therapeutically is formidable. *"This model precludes simple animal model extrapolation ; new generations of mouse models and in silico approaches are needed"* (Saïdo, Oxford Workshop, 2013). Achieving statistical power in clinical trials also requires large numbers of people (sample sizes), "pooling" samples, or combining patient data from numerous studies, and will critically depend on the access to and use of large data sets, advances in ICTs, data mining, computational and bioinformatics methods to sift and sort through the massive amounts of biological data, etc. *"It requires embedding research into clinical practice"* (Stuss, Oxford Workshop,

2013). A complete understanding of how the pathology of AD leads to its symptoms will also depend on answering fundamental questions related to the mechanisms of normal cognitive function.

#### **Box 1. Epistatic interactions in Alzheimer's disease**

Epistasis is the interaction between two or more genes to control a single phenotype. Combarros et al. (2009) confirmed 27 different significant epistatic interactions for Alzheimer's disease, which were grouped into five categories: cholesterol metabolism, beta-amyloid production, inflammation, oxidative stress, and other networks. Some interactions were synergistic, while others were antagonistic. The synergistic interactions indicate that the pair of involved genes together increases the risk of Alzheimer's disease. Meanwhile, the antagonistic relationships indicate a protective relationship between two genes. The strongest interactions involved the pairing of apolipoprotein E4 with three different genes: alpha(1)-antichymotrypsin,  $\beta$ -secretase, and butyrylcholinesterase.

Diabetes is another complex disease that is influenced by both epistatic and environmental factors. Only in rare cases does the disease appear to be monogenic, and generally, multiple genes seem to be involved (Florez et al., 2003). While it is known that diabetics have insufficient levels of insulin and high blood sugar levels, the specific factors underlying disease susceptibility are still being researched. (Cox et al., 1999; Wiltshire et al., 2006).

Evidence also exists that epistasis is involved with other complex diseases, including cardiovascular disease, hypertension, autism, schizophrenia and other neurological disorders, as well as sporadic breast cancer, bladder cancer, and other types of cancer (Combarros et al., 2009).

#### **Addressing the complexity of neurodegenerative disorders with computational approaches**

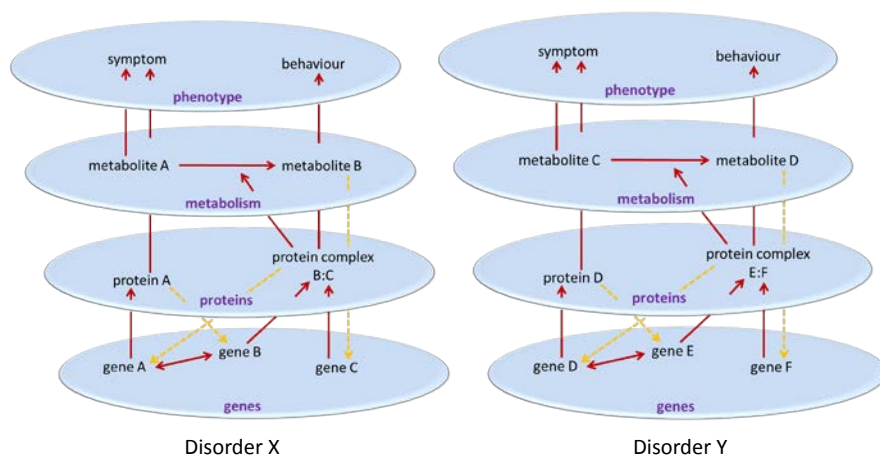
*The Etiology and physiopathology of AD are neither linear nor additive but like a ballet choreographed interactively over time, involving genetic, gene expression, epigenetic and a multitude of environmental factors [...]. Any framework proposed for explaining functional loss must take account of processes at multiple levels of organization - from the molecular to the behavioral. (Middleton, Oxford Workshop, 2013).*

AD is not an exception in central nervous systems (CNS) research and development. The successful development of novel first-in-class therapeutic agents in the CNS has been lagging with respect to other disease areas. Only 8% of CNS drugs that enter clinical trial phase I are approved, with about 65% of the failures due to lack of efficacy or sufficient differentiation in phase III. This high degree of failure is caused by the extreme complexity of the human brain neurobiology and the increasing realisation that the clinical outcome is driven by emergent properties of neuronal circuits, rather than by a single target (Kola and Landis, 2004; Arrowsmith, 2011).

The challenge in understanding AD is to be able to map what is known about the biochemical and cellular pathology to the abnormal behaviours characteristic of the disease. It is to move from the current situation, in which the "standard model" for a disease process is a flow chart of interconnections between brain regions, to a conceptual model that integrates the wealth of information at the molecular, cellular, network, systems, and behavioural levels (Figure2). The complexity of brain disorders make it, therefore, necessary to use and combine multiple, heterogeneous data sets.

*"There are compelling reasons for attempting to develop models and simulations of the disease pathways - not as a replacement for other approaches, but as a means of new insight into pathogenesis and treatment"* (Lovestone-Oxford Workshop, 2013).

Simulations and computational approaches provide a means of thinking about complex systems, particularly those with rich interactions between levels.

**Figure 2. Complexity of brain disorders**

17

Source: D. Stuss (Oxford Workshop, 2013)

Thanks to the computational power available today, vast amounts of complex data can be analysed holistically to identify new findings or set forth new hypotheses.

Computational modelling allows, for example, analysis of the role of network dynamics in the initiation and progression of the neuropathology in Alzheimer's disease. They can enable the simulation of the entire spectrum of clinical phenotypes of complex brain disease and thereby accelerate the discovery/validation of new generations of therapies.

Although modeling will never be able to simulate the complete human brain in detail, it is a possible solution to increase the chances of clinical success. Modeling and simulation techniques are also useful tools for clinicians and researchers to establish the appropriate population to treat or narrow the efficacy range needed for a drug in early stages of development to be commercially viable. Dosing mismatch is an area in clinical drug development that has significantly improved between 1995 and 2004, likely as a consequence of improved pharmaco-metric modeling and simulation (Hurko and Ryan, 2005).<sup>4</sup>

The field of AD research is at a critical juncture to benefit from modeling: biomarker-based screening measures are under development as are therapies that may modify the underlying neuro-pathological disease process. Full acceptance and understanding of this technology necessitates, however, investments in infrastructures and a new generation of scientists and data analysts with a broad understanding of network physiology, pharmacology, biology, and drug discovery.

## UNLEASHING THE POWER OF BIG DATA

*Big data is part of the current health care and innovation landscape and has been now for some time. Advances in information technologies and cyber infrastructures have created a virtual deluge of new types of health and wellness data ranging from new data through digital imaging, sensors, and analytical instrumentation to new ways of collecting genetic, biological and behavioural information and to combining data from different sources, such as clinical and administrative records.*

Biomedical researchers, health care professionals and patients are generating huge amounts of data of great value to AD research from an array of devices such as genomic sequencing machines, high-resolution medical imagers, electronic health records, ubiquitous sensing devices, and smart phone applications that monitor patient health. It is predicted that more medical information and health and wellness data will be generated in the next few years than ever before (McKinsey&Company, 2013).

The remarkable expansion of digital health data is largely driven by the confluence of important technological developments, notably the increasing ubiquity of broadband access and the proliferation of electronic health records, smart mobile devices and smart ICT applications such as sensor networks and machine-to-machine (M2M) communication. The large decrease in Internet access costs over the last 20 years has also been a significant driver. Advances in magnetic disk technology have dramatically decreased the cost of storing data (Grochowski and Halem, 2003). For example, a one-terabyte disk drive, holding one trillion bytes of data, costs around USD 100. Cloud computing has also played a significant role in the increase in data storage and processing capacity. These advances are driving the increase in the volume of digital records and databases contemplated and being developed today for genetic, neuro-imaging and epidemiological research.

*“The Internet has made measurable what was previously immeasurable: The distribution of health information in a population, tracking health information trends over time, and identifying gaps between information supply and demand.”*(Eysenbach, 2006)

All these new sources of data can create a tremendous resource to accelerate innovation for neurodegenerative disease. The potential public health uses of big data, extend today from genomics to clinical care, environmental and behavioural data to derive better insights into these diseases. Comparative-effectiveness researchers are mining government and clinical databases for proof of the best, most cost-effective treatments—information that could transform health care policy. Researchers now have access to human genetic data and genomic databases they can combine to study treatment outcomes.

According to McKinsey & Company, with the right tools, big data could be worth USD 9 billion to United States’ public health surveillance alone and USD 300 billion to American health care in general, by improving detection of and response to infectious disease outbreaks, and the latter largely through reductions in expenditures.

*“Big data can be transformative , the possible uses can, however, be difficult to anticipate at the time of initial collection. Big data is not just a quantitative change , it is a conceptual and methodological change. It will transform the way we do science and the way we deliver care. ”* (Rossor, Oxford Workshop, 2013)

For example, the discovery of Vioxx’s adverse effects, which led to its withdrawal from the market, was made possible by the analysis of clinical and cost data collected by Kaiser Permanente, a California-based managed-care consortium. Had Kaiser Permanente not collected and connected these clinical and

cost data, researchers might not have been able to attribute 27 000 cardiac arrest deaths occurring between 1999 and 2003 to use of Vioxx.

### New paradigms of big data use in health research and clinical care

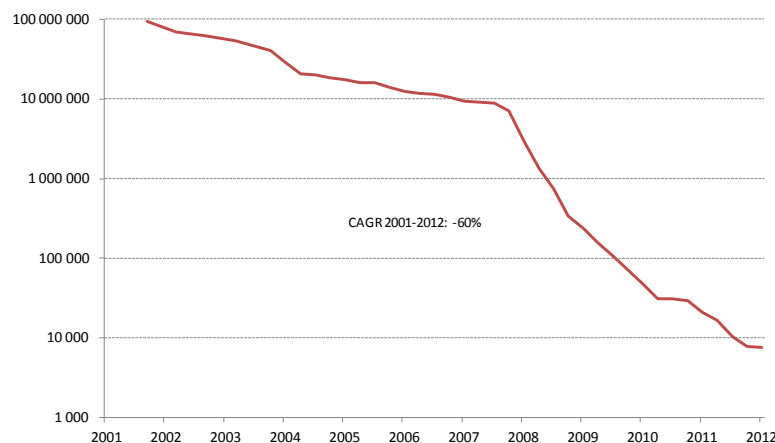
The next sections list a range of examples of sources of big data and approaches, which provide new ways to measure various aspects of disease progression and health for improved diagnosis and care delivery, as well as translational and clinical research.

#### *Whole genome sequencing to advance the study of Alzheimer's disease*

The study of AD genetics is complicated by the likelihood that the risk of late-onset AD is influenced by many genes. Identifying these genes requires analysing the genomes of large numbers of people.

Over the past two decades the power of genetic sequencing has increased by one million-fold enabling this kind of research. No previous technology in history has increased in power that fast. DNA sequencing machines can now read about 26 billion characters of the human genetic code in less than a minute, and the sequencing cost per genome has dropped by 60% a year on average from USD 100 million in 2001 to less than USD 10 000 in 2012 (Figure 3).

**Figure 3. Sequencing cost per genome, 2001-11**



Source: OECD based on United States National Human Genome Research Institute ([www.genome.gov/sequencingcosts/](http://www.genome.gov/sequencingcosts/)).

Whole genome sequencing involves many terabytes of data and translating that information into useful information is beyond the scope of any single entity. It requires ambitious co-ordinated efforts between different types of organisations. Successful collaboration on this scale requires sophisticated platforms, processes and tools. The Alzheimer's disease genetic initiative was, for example, established in 2009 by the United States National Institute of Ageing to support the use of large-scale, high throughput genetic technologies by researchers studying late-onset AD.

Recent efforts include a public-private research project launched in 2012 and completed in July 2013 by the Alzheimer's Association and the Brin Wojcicki Foundation in collaboration with the Alzheimer's Disease Neuroimaging Initiative (ADNI), and a range of US academic centres, which performed whole-genome sequencing on more than 800 people. The genome sequencing data – estimated to be 200 terabytes, will be housed in and available through the United States Global Alzheimer's Association

Interactive Network (GAAIN), a planned massive open access network of Alzheimer's disease research data.

### *Neuro-imaging for the early detection of Alzheimer disease*

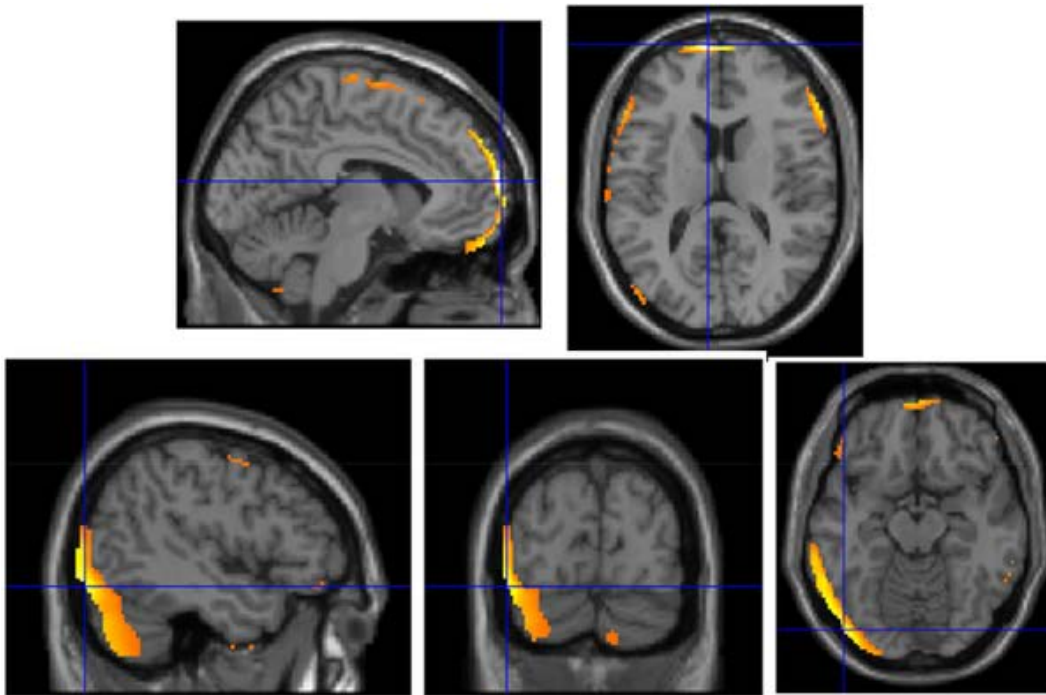
Conventional structural neuroimaging, such as computed tomography (CT) or magnetic resonance (MR), has long played a supportive role in the diagnosis of memory disorders and is today recommended for the routine evaluation of AD. However, because structural changes may not be detected at visual inspection until late in the course of the disease, more contemporary structural imaging techniques have emerged that aid in detection of subtle changes not readily apparent on routine images obtained at a single time point. These include positron emission tomography (PET), single photon emission CT (SPECT), and functional MR imaging (fMRI).

Functional magnetic resonance imaging (fMRI), in particular offers the promise of revolutionary new approaches to studying human cognitive processes, provided we can develop appropriate data analysis methods to make sense of the huge volumes of data. fMRI measures brain activity by detecting changes in blood flow and blood oxygen levels ( the ratio of oxygenated haemoglobin to deoxygenated haemoglobin in the blood with respect to a control baseline), at many individual locations within the brain (Figure 4). It is widely believed that blood oxygen level is influenced by local neural activity, and hence it is generally taken as an indicator of neural activity.

A twenty-minute fMRI session with a single human subject produces a series of three dimensional brain images each containing approximately 15 000 voxels, collected once per second, yielding tens of millions of data observations. Each voxel contains hundreds of thousands of neurons.

Accurate quantification of regional brain volumes is time- and labour-intensive. If this limitation of fMRI-based methods could be solved via automation of scan analysis, such methods are almost certain to become useful tools for the early detection and monitoring of Alzheimer disease in patients.



**Figure 4. Example of fMRI scan**

Source: Rasgon NL, Geist CL, Kenna HA, Wroolie TE, et al. (2014) Prospective Randomized Trial to Assess Effects of Continuing Hormone Therapy on Cerebral Function in Postmenopausal Women at Risk for Dementia. PLoS ONE 9(3): e89095. doi:10.1371/journal.pone.0089095

### ***Sensor-based systems to monitor behavioural changes***

Sensor-based systems can also be leveraged to provide clues on emerging physical and mental health problems. Ubiquitous sensors have, for example, an increasingly important role to play in integrating a novel and less-biased window of cognitive and behavioural monitoring of older patients (Figure 5). These systems can also provide assistance in increasing the independence and security of people who have problems of memory, planning, and carrying out tasks of everyday life.

Unobtrusive monitoring based on statistical pattern recognition and machine learning promises to contribute to our ability for early diagnosis of AD and prevention. The applications range from basic vital sign monitoring that includes environmental, behavioural and physiological sensing, to computational modelling. For individuals with chronic conditions, unobtrusive home-based monitoring could also result in better patient outcomes by allowing the physician to monitor compliance with pharmaceutical regimens and activity level guidelines; to better understand the range of variation of patient outcomes.

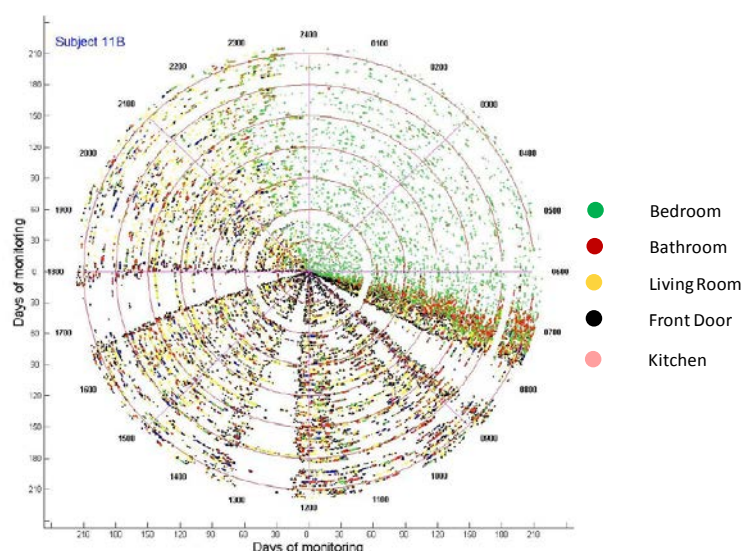
For older patients, sensor-based devices can also be utilised to monitor falls and near-falls physical activity, socialization, or even overall mobility. For example, wearable fall detectors that include accelerometers are a good example of information technology for assisted living at home (Brown, 2005). In most of these systems, a periodic report from the sensors is sent back via wireless communication, to a local base station.

The biggest hurdle to overcome in making these approaches useful for AD research and improve the quality of care is the development of efficient and user-friendly data-flow processing and effective

conversion of the sensor events into clinically actionable knowledge. Context-awareness imposes significant demands on the knowledge maintained by these systems.

Progress will, hence, depend on the development of robust algorithms and computational models that can fuse and derive meaning from the diverse sets of information. Key factors influencing scalability, therefore, include: i) seamless integration and interoperability of the technology; ii) reliability of message capture, translation and delivery to healthcare professionals and the amount of information transmitted per patient and iii) frequency of monitoring and transmission and context awareness.

**Figure 5. Sensor events in a residential facility**



Source: Misha Pavel, OECD-NSF Workshop, 2011.

### ***System's biology to model the complex molecular mechanisms underlying AD***

Network and systems biology strategies today offer a powerful means to explore the complex molecular mechanisms underlying AD (Chen et al. 2006; Liu et al. 2006). Two of the most surprising discoveries from the genome sequencing projects are that the human gene repertoire is much smaller than had been expected, and that there are just over 200 genes unique to human beings (International Human Genome Sequencing consortium, 2004). As the number of genes alone does not fully characterise the biological complexity of living organisms, research efforts are now increasingly directed to better understanding protein–protein interactions and gene regulatory interactions. This has propelled network and systems biology to the frontier of biomedical research (Vidal, et. al., 2011, Barábasi, et. al. 2011).

In humans, the potential complexity of the resulting networks — the human interactome — is daunting: with ~25 000 protein-coding genes, ~1 000 metabolites and an undefined number of distinct proteins and functional RNA molecules, the number of cellular components that serve as the nodes of the interactome easily exceeds 100 000. The number of functionally relevant interactions between the components of this network, representing the links of the interactome, is expected to be much larger. With so much data available, the challenge is to integrate that information into a meaningful single interaction network.

The highly interconnected nature of the interactome also means that, at the molecular level, it is difficult to consider diseases as being consistently independent of one another. Indeed, different disease molecular mechanisms can overlap, so that perturbations caused by one disease can affect other diseases (Barabási, et. al., 2011).

The systematic mapping of such network-based dependencies, between patho-phenotypes and their disease modules has culminated recently in the concept of the “diseasome”, which represents disease maps whose nodes are diseases and whose links represent various molecular relationships between the disease-associated cellular components.

Although of great promise, progress towards a reliable network-based approach to neurodegenerative disease is currently limited by the incompleteness of the available interactome maps and the need for powerful visualisation tools as well as statistical methods that are reliable in the context of interconnected environments (Barabási & Oltvai, 2004).

### *Using social media for health care research*

Web- and mobile-based applications of social media are emerging as useful new approaches for the dissemination and collection of health and lifestyle information as they can reach a broad audience in a very short period of time, are easy and affordable to access and use, and cater to a large variety of people.

Online social communities provide, for example, a vehicle for individuals with chronic diseases to share information on therapies and disease progression. Participants contribute personal stories that provide learning experiences for other participants who may be contending with a similar health problem. Some online communities are moderated by health care professionals who can offer expert advice via message board posts or synchronous chat sessions.

There is growing recognition that online communities not only provide a place for members to support each other, but also contain knowledge that can be mined for public health research, monitoring, and other health-related activities. By harnessing the power of global, widely used user-generated content, social media is playing an increasingly critical role as an important communication platform on health and disease and as an opportunity to collect data on patients’ experiences to guide policy and communication planning”.<sup>5</sup>

For example, the social network PatientsLikeMe developed a lithium-specific global data collection process to capture information about individuals suffering from amyotrophic lateral sclerosis (ALS) that were registered with the network and who began taking the drug off-label via their physician (Wicks, et. al. 2011).

ALS is a chronic condition where both randomised trials and nonrandomised clinical studies have yet to provide an effective therapy. It is a rapidly fatal neurodegenerative disease causing progressive weakness and muscle atrophy; median survival from symptom onset is 2–5 years. In 2008, a small study suggested that lithium carbonate slowed ALS. Once that study was published, hundreds of ALS patients on PatientsLikeMe began taking the drug and a few used freely available tools such as Google Spread sheets to ‘crowd source’ their own study. In response, PatientsLikeMe upgraded its tools and developed new analytical techniques to evaluate whether lithium was effective (Wicks, et. al., 2011).

PatientsLikeMe has recently started to post aggregated information on experience from patients (or caregivers) who have reported AD as well as early onset dementia. Personal opinions from patients of their own memory and organisational skill lapses could serve as one of the earliest signs of eventual dementia. If so, future studies that look at new treatments, therapies and lifestyle changes to stave off dementia would likely find this information useful.

The social network also regularly imports the complete dataset from ClinicalTrials.gov to let its membership know (free of charge) about the 30 000+ active trials for which they may be eligible. The biggest issue for PatientsLikeMe remains the feasibility of recruiting and then engaging patients themselves on the online platform, which for older patients has complications such as requiring the ability to use a computer, remember passwords, navigate a complex website, etc. PatientsLikeMe also does not yet have custom user interfaces for caregivers but plan to have this in the next couple of years, which should benefit dementia caregivers including family members.

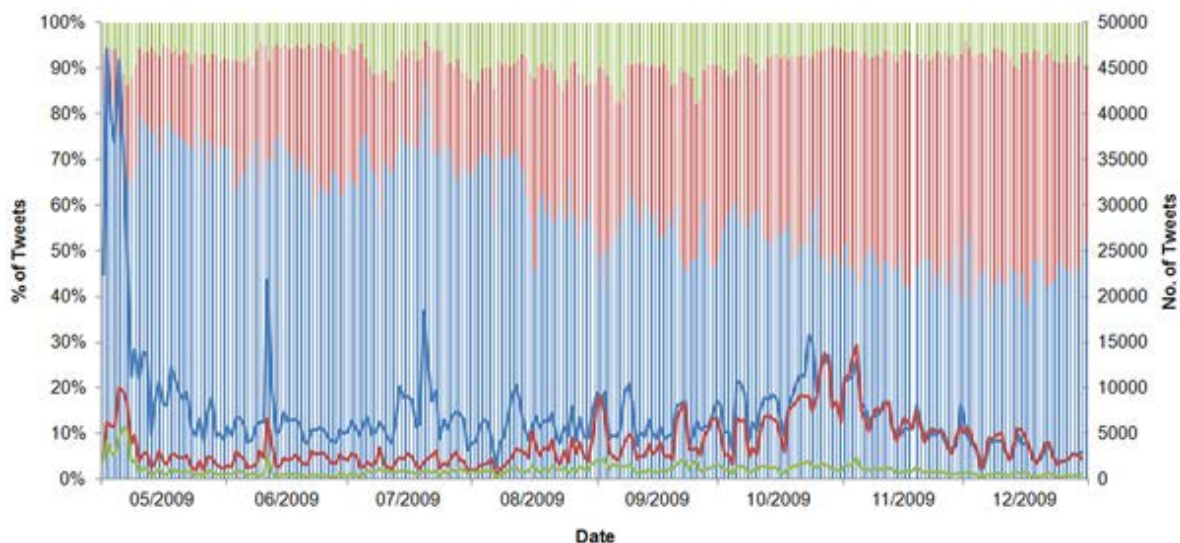
The Banner Alzheimer's Institute in the United States is leveraging the power of the social networks to develop an "Alzheimer's prevention registry", to help scientists test new treatments. By 2013 more than 9 300 individuals had joined the effort. Most are cognitively normal adults age 50 or older—the Registry's target population— and 60% report a family history of Alzheimer's disease.

Government agencies are also using social networks to engage the public (for example, during product recalls and in H1N1 flu pandemic preparations) and as a source of behavioural measures.<sup>6</sup>

Twitter ([www.twitter.com](http://www.twitter.com)) is also emerging as a suitable platform for this purpose. Twitter allows users to send and read short text-based messages limited to 140 characters, which contain a wealth of data. Mining these data provides an instantaneous snapshot of the public's opinions and health-related behavioural or other responses. Longitudinal tracking allows identification of changes in opinions or responses. In addition to quantitative analysis, twitter also permits qualitative exploration of likely reasons why sudden changes have occurred (e.g. a widely read news report) and may indicate what is holding the public's attention.

Twitter content has been studied recently for tracking flu epidemics (Chew & Eysenbach, 2010) (Figure 6), to assess public misunderstandings surrounding antibiotic use (Scanfield, et. al., 2010; Signorini et al., 2011; Sadilek et al., 2012), and more recently (Robillard, et. al., 2013) to gain insights on how online users share information about dementia and the type of information shared.

**Figure 6. Tweets containing H1N1, swine flu, or both from May to December 2009**



Source: Chew C, G. Eysenbach (2010), "Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak". PLoS ONE 5(11): e14118. doi:10.1371/journal.pone.0014118)

There is also a large amount of literature proposing methods to extract useful information from online searches data generated each day, including through Yahoo! and Google (Eysenbach, G. 2006).

A Centers for Disease Control study conducted with Yahoo! in 2005 suggested that Internet searches for specific cancers correlated with their estimated incidence (Spearman rank correlation,  $\rho = 0.50$ ,  $P = .015$ ), estimated mortality ( $\rho = 0.66$ ,  $P = .001$ ), and volume of related news coverage ( $\rho = 0.88$ ,  $P < .001$ ).<sup>11</sup> The authors concluded that “media coverage appears to play a powerful role in prompting online searches for cancer information” (Cooper, et. al., 2005).

Although current Internet search query data are no substitute for timely local clinical and laboratory surveillance, recent studies indicate that the intensity of certain web queries on influenza and influenza-like illness follows the same pattern as the laboratory and sentinel reports for influenza, and that they can be used as additional input data for estimation models (Hulth, et al., 2009; Eysenbach, 2007).

In November 2008, Google Flu Trends was launched as an open tool for influenza surveillance in the United States. Engineered as a system for early detection and daily monitoring of the intensity of seasonal influenza epidemics, Google Flu Trends uses internet search data and a proprietary algorithm to provide a surrogate measure of influenza-like illness in the population (Olson, et al., 2013; Yin, et al., 2012). The algorithms may still need to be refined though, as the journal *Nature* reported in February 2013 that Google’s Flu Trends data was significantly over-estimating the number of influenza cases. Some researchers suggested that widespread news coverage led to spikes in influenza-related searches by people who were not ill.<sup>7</sup>

### ***Harnessing patient data from electronic health records***

As Electronic Health Records (EHRs) and other health informatics devices become increasingly widespread, harnessing health and clinical care data may yield valuable information. EHR data can provide rich contextual data about patients, including their demographics, clinical parameters (e.g., blood pressure, BMI, haemoglobin A1c, and lipids), medications, patterns of care, and how health is changing over time and along clinical pathways. EHR-based surveillance systems capable of extracting, analysing, and reporting quality of care and disease-specific indicators to public health departments in almost real time are emerging. For example, information can be gathered today to find predictors for diseases or adverse effects of treatment that would otherwise have gone unnoticed by most traditional approaches.

EHRs are also likely to become increasingly important sources for public health as more practices adopt EHRs with specific functionalities to satisfy care and health reporting requirements. Though challenging, capturing and delving into this data for AD is worth the effort, particularly as new data-mining, analytical and computational tools and techniques are emerging.

Physicians at Harvard Medical School and Harvard Pilgrim Health Care have, for example, recently demonstrated the potential for computer algorithms to analyse EHR data to detect and categorize patients with diabetes for public health surveillance. The researchers used algorithms to review 4 years of EHR data from HPHC, a large, multisite, multispecialty ambulatory practice, and flag patients suspected of having diabetes based on lab results, diagnosis codes, and prescriptions (Klompas, M., 2012).

The algorithms successfully identified patients who had been clinically diagnosed with diabetes, and were able to distinguish between patients with Type I and Type II diabetes. The algorithms were more effective at identifying patients with diabetes than reviewing disease diagnosis codes alone.

A similar approach was recently used to identify patients potentially at risk for AD and dementia and validate a diabetes-specific dementia risk score. Using data from two longitudinal cohorts of almost 30,000 patients with Type II diabetes aged 60 and older, with 10 years of follow-up, the researchers developed and

validated a diabetes-specific dementia risk score. The predictive test uses information about metabolic events, cerebrovascular disease, depression, education, age, cardiovascular disease, microvascular disease and diabetic foot, and assigns each a value related to their association with dementia. Combining these, an overall score can be assigned to the patient, indicating the 10-year probability of developing dementia (Exalto, et. al., 2013).

### ***Crowd-sourcing innovations for Alzheimer's disease***

Small initiatives will never achieve the benefits that scale brings to a complex science project. Crowdsourcing is a process that trains the power of the Internet and of large groups on a specific problem and holds the promise to help accelerate innovation for AD. Crowdsourcing can process data quickly and on unprecedented scales and with better quality-control than any individual or small research group can attain.

The “hit rate” for solutions from crowd sourcing is quite high, in some cases up to 40% –which is remarkable, especially since many problems are generally put out on the web because they are, by definition, beyond the problem-solving ability of the organisation posting them.

The core concept in crowdsourcing has been around for some time. For example, open source software development can be considered a type of crowd sourcing. Popular software such as Firefox, Apache, and Linux are the results of crowd sourcing. User-generated content including Wikipedia, YouTube, Yahoo Answers, and social bookmarking are also good examples of productive crowdsourcing.

Given its open, informal structure, crowdsourcing is cross-disciplinary by design. In some cases, even gifted amateurs and people without direct experience with the problem provide valuable insights and solutions. This form of mass collaboration has been successfully used by corporations for product design and marketing.

Acknowledging the success of the approach, the Design Council and the UK Department of Health launched in 2011 a 12-month national design challenge to create practical product and service solutions to improve the quality of life for people living with dementia and launched them as real initiatives (Box 1).

In the United States, The Geoffrey Beene Foundation Alzheimer's Initiative (GBFAI) launched in April 2013, a USD 100 000 Geoffrey Beene Global NeuroDiscovery Challenge whose goal was to identify male/female differences in the pathogenesis and presentation of AD in the pre-symptomatic, early symptomatic and late stages of AD. The objective of the challenge is to generate novel hypotheses addressing the causes and consequences of such differences and provide a rationale supported by available data and a detailed research plan describing how to test them. Of particular interest are hypotheses that can be validated in a secondary phase by mining existing AD databases for analysis of variables associated with biologic male/female differences in AD.

### **Box 2. The Design Council living well with dementia challenge**

Dementia is a huge issue for the United Kingdom and the world, but it's also a real opportunity for social innovation.

Acknowledging the scale of the issue, the Design Council and Department of Health launched a 12-month national design challenge to create practical product and service solutions to improve the quality of life for people living with dementia and launched them as real initiatives

Five innovative solutions were selected:

Buddiband ([www.buddi.co.uk](http://www.buddi.co.uk)): A comfortable, discreet and waterproof wristband personal alarm that can send alerts from anywhere to buddi's support services, enabling people with dementia to get out and about with confidence.

Dementia Dog ([www.dementiadog.org](http://www.dementiadog.org)): A service providing assistance dogs to people with dementia, helping them lead more fulfilled independent and stress-free lives.

Grouple ([www.grouple.cc](http://www.grouple.cc)): A secure, private online social network helping people share the responsibilities of caring for someone with dementia.

Ode ([www.myode.org](http://www.myode.org)): A fragrance-release system designed to stimulate appetite among people with dementia.

Trading Times ([www.tradingtimes.org.uk](http://www.tradingtimes.org.uk)): a web-based service that matches careers with local businesses for flexible paid work, providing opportunities to earn and stay connected with society.

The teams behind these solutions included designers, entrepreneurs and service providers, as well as experts in nutrition, dog training and olfaction.

The five solutions demonstrate the vast potential of innovative ideas in an under-served market and show how crowdsourcing can play a key role in confronting a major social challenge.

### ***Accelerating innovation with citizen science***

*“Machine learning creates new opportunities for human participation”* (Simpson, Oxford Workshop 2013)

Citizen science initiatives are also growing in importance internationally and in the bioscience field. Generally ‘citizen science’ refers to a network of people, many of whom may have no specific scientific training, performing research-related tasks such as recording specific observations over time to reveal patterns and trends. The approach leverages what Clay Shirky called “cognitive surplus”, Clay, S., 2010 which describes the vast amount of time that people, collectively, spend on activities such as watching TV.

Citizen science projects often involve non-professionals taking part in one or more of the following:

- Crowdsourcing
- Mass-participation
- Data collection
- Data analysis



*“The human race spends 16 years every hour playing Angry Birds. There’s a lot of brainpower out there and what we try to do is take that brainpower and make it more useful to researchers.”* (Simpson, Oxford Workshop, 2013)

Zooniverse,<sup>8</sup> for example, works with researchers to design sites that take their data and presents it into a format that will let the crowd help them to achieve their objectives. Zooniverse has a community of more than 850 000 people, who have taken part in more than 20 citizen science projects over the years. These initiatives support a form of ‘scientific democracy’, where data can be shared among and utilised by investigators in public and private sectors, policy makers, and the public.

Foldit is another popular online citizen-science initiative, in which individuals are scored on the structure of proteins that they have folded. The game records the structure and the moves that the players do, and scientists can capture the data that is then used to improve the game in every aspect, from the quality of the scientific results that are coming back to how long people play the introductory levels that are supposed to teach the game. The whole game is like an ongoing, continuous experiment.

In Foldit puzzles, for example, players are rewarded for solving clashes and voids, places where the protein is not consistent with known biochemical patterns. Players are able to build a hypothetical protein and see how it works in the game. The game’s score is based on a proxy for how well the protein would work in a laboratory, whether it can catalyse some reaction that the scientists are interested in, or how well the protein sticks to some part of a virus, or even in the case of the Symmetry puzzles, how well the protein sticks to itself. Solutions that are promising are then synthesized in the laboratory.

Foldit was successfully used to remodel the backbone of a computationally designed enzyme that catalyzes the Diels-Alder reaction, which brings together two small molecules to form a particular kind of bond that the scientists were interested in making (Eiben et al, 2012). This catalysis can be useful for building other kinds of small molecules, such as drugs and chemicals. Scientists went back and forth with the players and in the end, designed an enzyme that was about 20 times more efficient in catalysing the reaction than the one the scientists had started with. Tapping into the vast cognitive surplus online and incorporating crowdsourcing in Alzheimer’s research to both enlist the public’s help and engage public interest holds tremendous promise for accelerating innovation.



## CREATING VALUE FROM BIG DATA: ADDRESSING THE MAIN CHALLENGES

*A data revolution is now taking place, but it merely scratches the surface of what can be done with information technology, as the key is not so much the volume of the data or the technology but the knowledge derived. To successfully create value from big data, it is important to focus on the breadth of the data, signal extraction, and deep insights (R. Stuss, Oxford Workshop, 2013).*

The crucial studies needed to validate alternative models of prevention and care and develop therapeutic strategies aimed at slowing disease progression will require an unprecedented level of data collection, storage and processing and new investments in research and infrastructure. No one nation has all the assets to pursue this type of research independently. The business case for global collaboration and for harnessing big data to accelerate innovation for Alzheimer's and dementia is today undisputed, but a range of challenges needs to be addressed. These challenges revolve around the three "Vs" of big data: volume, velocity, variety.

Big data implies enormous volumes of data. Now that data is generated by machines, networks and human interaction (for example, on systems like social media) the volume of data to be analysed is massive. The growth of 'big data' is literally forcing health organisations to revisit and rethink their infrastructure, capacity and business plans or else 'drown under the weight of cost and compliance'. The amount of information being collected and the pace (velocity) at which data is accumulating is so huge that modern database management tools are becoming overloaded and obsolete. Increasing the scale of data also adds analytic complexity. New analytical approaches are needed to produce the necessary insights.

Variety refers to the many sources and types of data. Big data is heterogeneous (structured, unstructured, text etc.), reflecting the traditional silos across care settings, industry/research, scientists/clinicians. Data collected in the various silos are often "fuzzy" — incomplete, of poor quality and inconsistent. As health data currently stands, it's often hard to know which information is accurate and which is out of date. Unreliable data usually add "noise" and masks effects.

*"Non-valid data provide non-valid results despite large sample size or any level of statistical significance obtained. This is a familiar challenge in any sector and health care is no exception. Validity and veracity of data are two additional "Vs" that data scientists, researchers and clinicians need to be concerned with. (Kukull, Oxford Workshop, 2013).*

The harvesting of large data sets, their linkage and the use of analytics also raise specific privacy concerns. The task of ensuring data security and protecting privacy becomes harder as information is multiplied and shared ever more widely across sectors, organisations and borders. Privacy and data protection laws are premised on individual control over information and on principles such as data minimization and purpose limitation. Yet it is not clear that minimising information collection is always a practical approach to privacy in the age of big data. The principles of privacy and data protection must be balanced against additional societal values such as public health, national security and law enforcement, environmental protection, and economic efficiency. A coherent framework would, ideally, be based on a

risk matrix, taking into account the value and social benefits of the different uses of health and personal data against the potential risks to individual autonomy and privacy (OECD, 2013a).

A strong consensus emerged at the consultation on the need for standards and guidance to address these issues. Comprehensive and compatible privacy frameworks are necessary to make a real dent in accelerating the pace of AD innovation. Promoting data standards and creating well curated and maintained scientific data repositories would also go a long way towards accelerating research. Lastly, it will be necessary to integrate data deposition and sharing requirements in grant funding to make real progress.

The next sections focus on the range of challenges raised by experts at the consultation. This summary does not represent an exhaustive list, but rather a compilation of the main issues identified.

### **Challenge 1: The need for compatible “big” data governance frameworks**

The complexity of AD and dementia impose large data collections, which implies the need for multiple collaborations and exchanges of data. Although existing national population-based cohorts and/or studies offer valuable data, they do not individually fulfil all of the specific needs for: a) prospective validation studies; b) developing new computational algorithms/models, which will require large databases from varied sources and domains of measurements (e.g., biomarkers, genetic, behavioural, imaging and other information on co-morbid conditions), in order to discover/establish probability profiles for predicting the relative risks for memory disorders/dementia in asymptomatic populations; and c) furthering the fundamental understanding of the heterogeneity in the prevalence-incidence of Alzheimer’s disease as well as the knowledge about people at greatest risk for developing Alzheimer’s disease. (Kachaturian, Oxford Workshop, 2013)

International collaboration through partnerships or federated global networks, or more structured virtual international organisations can support studies with sample sizes large enough to achieve “definitive” results, promote spinoff research projects, and yield faster “translation” of results into clinical and public health applications.

In neurodegenerative research, there are many national organisations today building data repositories and patient registries. In the United States alone, the Alzheimer’s disease Neuroimaging Initiative and the Parkinson’s disease (PD) Progression Markers Initiative gather brain images and biological fluids from people with or at risk for AD and PD, respectively. The US National Alzheimer’s Coordinating Center has collected longitudinal records from more than 25 000 people, and recently started assessments for fronto-temporal dementia as well. Records from those who inherited an AD-linked gene are part of the Dominantly Inherited Alzheimer Network. The US Consortium to Establish a Registry for Alzheimer’s disease (CERAD) <sup>9</sup> has functioned as a vehicle for a wide range of studies and as a mechanism for developing and testing dementia-specific instruments. Many other OECD countries are putting in place similar national databases, although an inventory or detailed information in regard is still largely missing.

In recent years a substantial number of multi-site federated networks and regional collaborative consortia have also emerged. Examples include the Ontario Brain Institute (OBI) , the Alzheimer’s Disease Neuroimaging Initiative (ADNI), the Joint Programme on Neurodegenerative Diseases (JPND), and the US National Alzheimer’s Coordinating Center (NACC) .

*The need to build on and integrate these existing initiatives worldwide to avoid reinventing the wheel and creating new stand-alone duplicative resource-intensive infrastructures is of critical importance for progress in AD (Kachaturian, Oxford Workshop, 2013).*

This requires creating compatible data governance frameworks- which today vary significantly. Without compatible data governance frameworks the ability to authenticate, use, process and share data is compromised. Data governance refers to the overall management of the availability, accessibility, usability, integrity and security of the data collected and stored. In addition, there must be policies that conform to the governance of the data in order to form a consistent and effective framework (Table 1).

**Table 1. Key attributes of data governance**

<b>Attribute</b>	<b>Description</b>
Availability	The conditions by which data is available to users when needed.
Accessibility	Agreements on who should have access to data and where it should be located (researcher group members only? Should registration be required? Third party distribution permitted?).
Interoperability/Integration	Data must be semantically and syntactically interoperable across systems (what form should the data take?).
Traceability	There must be a trail of data from its source to its destination (metadata should be available).
Quality	The data must be accurate and complete.
Privacy and security	The data must be kept secure.
Ownership	The rights to data use and its possible commercial exploitation must be agreed.

Source: OECD adapted from Kukull and Stuss, Oxford Workshop 2013

Reports from the OBI, ADNI, the JPND and the US NACC indicate that rigorous and compatible data governance frameworks are needed to catalyse the development of an integrated global ecosystem of data that can be shared to accelerate innovation (Oxford Workshop, 2013).

Large-scale collaborations can be highly effective, but only if researchers agree in advance on standards and protocols so that data can be pooled and compared easily. Information must also be gathered consistently if it is to permit effective linkage and secondary analysis of data.

Interoperability is a central issue to be addressed. Without interoperability, researchers are forced to enter into bilateral data sharing arrangements, as there is no assurance that the mechanisms used between any two parties will extend to any other parties. In a networked environment, interoperability means common protocols defining the basic mechanisms by which users and resources negotiate, establish, manage, and exploit sharing relationships. It means sharing not only data but anything that connects to the data production and processing including computing tools, applications, methods, software, metadata, workflows across different platforms and even communication. A standards-based open architecture

facilitates extensibility, interoperability, portability, and code sharing; standard protocols make it easy to define standard services that provide enhanced capabilities.

The experience of established data networks provides an important knowledge-base on which to develop a coherent vision and guidance on good practices to be shared internationally.

## **Challenge 2: Financing innovation and sustainability**

For many big data projects, networks or federated research platforms, once the initial funding runs out, the most significant challenge is developing a sustainable business model. Long-term sustainability and financing appear to be the most challenging and, in most cases, unknown aspects of the many database, networks and other big data initiatives in existence today.

Development and maintenance of databases is a costly activity. As an illustration, the cost of the implementation phase of the Canadian longitudinal study on aging - which includes the first wave of data collection, follow-up on the initial cohorts and management - was estimated at CAD 23.5 million. The generation of big data (e.g. imaging, microarray, phenotypic, etc.) can include costly processes, requiring expensive consumables as well as specialised equipment and personnel for their generation. If for financial reasons, these networks or databases are unable to perform their tasks under conditions that meet the requirements of scientific research, scientists will either see valuable information lost or being transferred into a strictly for-profit environment. While many of the speakers discussed their project's progress and success in moving from planning to implementation, most could not clearly forecast its long-term sustainability or revenue models.

There is no magic bullet today with respect to the options or strategies required to achieve long-term financial sustainability. Financial sustainability is a critical issue for all initiatives, even those that are relatively more mature and directly funded by public sources.

As on-going financial support is uncertain, quite often large data networks such as ADNI must seek out multi-source financing.

*“Big data AD research requires durable investment strategies. In this respect, networking grants or research grants, which are often awarded as seed funding to establish these networks or databases, do not assure long term sustainability or operational financing.”* (Frisoni, Oxford Workshop, 2013)

Some networks currently charge registration fees to those who want to gain access to a specific data set and associated database. Varying fee structures can be applied for access depending on the nature of the data, the status and constraints. Fixed-cost subscriptions might also be paid by public funding agencies or not-for-profit organisations in respect of their long-term fixed demand for information to allow access to their grantees or employees at a nominal cost. Another model that appears to have great potential in being successful towards the prolonged financial sustainability of databases is “public-private partnership”.

The changes needed to address the uncertainties require an active role and financial commitment of both public and third party private payers.

## **Challenge 3: Data sharing and linkage**

Data sharing and linkage issues need to be addressed in order to effectively harness technological progress and ensure that available data will be turned into useful and actionable health information to accelerate research for AD.

The value of data explodes when it can be shared and linked with other data, thus data integration is a major creator of value. Building capacity for research using data linkage however, requires, investment in structures and frameworks for governance; standards, including for privacy and security; ethics and community involvement; information technology and information management; data sets and metadata; methods and tools for data linkage and analysis of linked datasets; and human capacity.

Linkage of a large number of clinical and research data and information about individuals raises numerous privacy and confidentiality issues. Researchers often find coded samples more valuable (even critically so, in certain types of research) than unlinked or unidentified samples because the linkage with identity provides a way to follow up with individuals in longitudinal studies. To ensure privacy and confidentiality of databases with coded material, technical and procedural computer security (including control and monitoring of access to, and transport of, data) may be essential. An important element, therefore, is the measures that should be undertaken to ensure the protection of the data and information contained within registries and data networks.

A recent OECD study on country practices (OECD, 2013b) in allowing linkage of data to monitor the quality of health care indicates that well-intended policies to allay concerns about breaches of confidentiality and to reduce potential misuse of personal health information may be limiting data use. A survey of 20 OECD countries explored the extent to which countries have developed and use personal health data and the reasons why data use may be problematic in some. The study concludes that to develop international studies comparing health care quality and health system performance, actions are needed to address heterogeneity in data protection practices.

The resources required to comply with legislative and policy requirements to enable data linkages is an additional related problem, as is the cost of developing the technical capacity to undertake the work.

#### **Challenge 4: Patient consent**

An important obstacle to data collection and linkage of relevance to AD research is the issue of obtaining patient consent. Informed consent is one of the most complex issues for AD research. Informed consent has become the pillar for protecting autonomy in research involving human subjects. Within the medical/scientific field, informed consent generally presumes the ability to indicate clearly to the participant the use and purpose of the particular research activity. The very nature of AD and dementia renders the provision of this type of information particularly difficult. All research on AD is, therefore, usually granted under the provision of surrogate consent.

Much variability, however, exists in Institutional Review Board (IRB) surrogate consent practices in studies involving incapacitated adults. This variability may have adverse consequences for needed research on AD and dementia (Gong and Winkel, 2010). Evidence of good practice is needed to guide policymakers and IRBs (Powell, Oxford Workshop, 2013).

A particularly controversial issue is whether the designation of a legal representative who can give consent on behalf of persons with impaired capacity is needed or whether a waiver on consent is ethically acceptable. The issue is determined by the laws of the jurisdiction in which the research is to be conducted and practices vary widely across OECD countries.

#### **Challenge 5: Timely dissemination of findings and data**

Extensive recent research has shown that the results of studies are often not shared publicly in a timely way and that between 25% and 50% of clinical trials remain unpublished even several years after completion (Chan, et. al., 2013; Decullier, et. al. 2005; Von Elm, et. al., 2008; Turner, et. al., 2008; Rising et. al., 2008). These studies also suggest that half to two thirds of all government-funded studies are

published two or more years after completion of the clinical trial (Ross, et. al., 2012) , that is, after completion of enrolment and observation. Much of the data from clinical trials in AD and dementia are currently not available to the scientific and clinical communities.

In a sample of NIH funded clinical trials registered in ClinicalTrials.gov, a 2012 study found that fewer than half of the trials were published in a peer-reviewed biomedical journal indexed in Medline only within 30 months of trial completion.

There are various reasons for not sharing data, writing up an article or not submitting it. These include pressure from research sponsors, instructions from journal editors, patent application submission, protection of scientific lead, patent negotiation, time for resolution of intellectual property ownership and slow dissemination of undesired results. Other reasons include policies and practices at universities that place a premium on patenting over publishing and weak incentives for researchers to share data. This can also act as a barrier to the replication and validation of scientific experiments (OECD, 2012).

A number of measures may help reduce the existence and impact of data dissemination delays and bias, including changes to the policies on publication of publicly-funded research (open access policies), electronic publishing, the prospective registration of interventional clinical trial studies.

#### **Challenge 6: Open data strategies to accelerate innovation**

There has recently been a great deal of interest in the concept of open science, and open data sharing mechanisms in AD research. The idea here is that no single entity can afford to generate and manage the necessary data required for adequate biological study of AD. It is also assumed that data can be shared and managed without devaluing the data or eliminating the ability to patent novelty. The dialogue is on-going, with several specific business models being proposed, (Hughes, 2009; Chesbrough and Garman, 2009) and there are indications that this strategy may be cost-effective, since the value proposition is no longer in the acquisition of data but in their interpretation.

Governments, as key funders of public research, play an important role in developing policies to foster greater deposition, sharing and access to and use of scientific research data. For example, public policies and guidance from research funding agencies can facilitate the sharing of and access to data resulting from publicly funded research (OECD, 2007) (Box 3).

#### **Box 3. The new open-data policy memorandum, issued by the Office of Science Technology Policy (OSTP) in the United States**

The new open-data policy memorandum, issued by the United States OSTP and the Office of Management and Budget, in 2013 requires federal department and agencies:

- to collect information in a way conducive to open publication;
- to build information systems that support information accessibility;
- to strengthen data management and release practices;
- to take steps to ensure privacy is maintained; and,
- to incorporate the openness requirements into their core agency processes.

*“Data is non-rivalrous, ownership is not a good concept when it comes to data. We have a problem in understanding human genetic variation for AD and it’s pretty apparent that we don’t have the infrastructure, norms, or practices for making sense of genomic and other information. What we’ve got are technologies that will deluge us with data quickly, but no real structures for systematically making sure we efficiently understand it and make it useful. If data aren’t shared, progress will be slow.”* (Cook-Deegan, Oxford Workshop, 2013)

Open data strategies have the potential to enhance the efficiency and quality of research by reducing the costs of data collection, by facilitating the exploitation of dormant or inaccessible data at low cost and by increasing the opportunities for collaboration in research as well as in innovation.

Open access to data for drug development is based on the fundamental premise that a proprietary attitude to all parts of drug discovery is extremely costly and inefficient. The fundamental problem is that industry collectively focuses too many resources on proof-of-concept studies for too few targets, and the studies are done in a proprietary way, with little collective learning. Further, because failure in proof of concept is never enough to dissuade others, these studies encumber the limited resources of industry for years, thereby limiting the ability of industry to pursue new and potentially relevant drug targets (Edwards, et. al., 2009).

The Structural Genomics Consortium (SGC) is an example of a successful open-access clearinghouse. The SGC is a public-private partnership with funding from governments, foundations and industry working in the area of drug discovery. The SGC has adopted an open-access policy. Under this policy, the SGC will not seek, nor permit its affiliated scientists or collaborators (including from industry) to seek, patents that would grant exclusive rights over its research outputs. Further, it works with its funders – including major pharmaceutical companies and foundations – to persuade others using the research outputs to similarly forgo patent rights.

Another example is the World Wide Alzheimer's Disease Neuroimaging Initiative (WW-ADNI), a collaborative effort of scientists from around the world for neuroimaging initiatives being carried out in North America, Europe, Japan, Australia, Chinese Taipei, Korea, China, and Argentina.<sup>10</sup> Data from WW-ADNI are quickly made available to the scientific community at no cost so researchers can use the information when designing or evaluating their own research. ADNI data is expected to play a key role in identifying effective treatments for Alzheimer's, as well as methods that may prevent the disease or slow its progression.

Open source drug discovery also offers potentials for accelerating drug development. The concept borrows two principle aspects from open source computing (i.e., collaboration and open access) and applies them to pharmaceutical innovation. By opening a project to external contributors, its research capacity may increase significantly. The non-profit research organisation, Results for Development Institute (“R4D”), has recently undertaken a high-level review of open source drug discovery projects aimed at neglected diseases.<sup>11</sup>

There are lessons to be learned from these initiatives, their successes and failures, in order to generate greater understanding of how much-discussed open data strategies work in practice and may be implemented to accelerate innovation in AD and dementia.

## **CONCLUSIONS: BIG DATA RESEARCH ON ALZHEIMER'S AND DEMENTIA IS AT A CROSS-ROADS**

Alzheimer's disease, as a chronic brain disorder, is the prototype problem for the Grand Global Challenge in healthcare. The scientific, public health and policy challenges are multifaceted, ranging from health economics and healthcare delivery, to health technology and biomedicine innovation. No one nation, agency, institution, company or industry has all the assets to pursue this type of research independently. The complex nature of AD disease requires large streams of multidimensional data and observation points – integrated multidisciplinary programmes and new sophisticated computational capabilities

As acknowledged by G8 Ministers at the time of writing this report,<sup>12</sup>, OECD countries are approaching a cross-roads in research for AD and leading world academics, business representatives and governments believe that there is a window of opportunity for international co-operation and coherent policies to help accelerate innovation.

On the one hand, advances in the biomedical sciences, data generation and processing and ICTs are predicted to have enormous potential for the prevention, diagnosis, treatment and care of AD. On the other, because the current R&D process is fragmented, costly, unpredictable and inefficient, these advances cannot optimally deliver improved diagnostics, health products and care processes to patients.

Researchers in industry, hospitals and universities continue to make significant contributions to scientific understanding. However, without better data sharing, interpretative capacity, and co-ordination of knowledge, we can make only limited progress in our understanding of the molecular basis of neurodegenerative diseases or whether treatment or intervention work.

The explosion of promising new fields of research, technological opportunities and data generation will not automatically translate into new products and care solutions for AD. In order to deliver this promise, these new developments will have to be accompanied by organisational, infrastructural and governance changes throughout the health innovation system.

The good news is that a large number of new approaches are already being used by governments, industry, and the medical community to increase efficiency, predictability, and rate of success, as well as promote international collaboration, public-private partnerships and data sharing to lower the risks and cost of research.

A wide range of national, regional networks and international initiatives have recently emerged for greater sharing and access to data in line with the principles of publicly funded scientific research data to be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests. The need to build on and integrate these existing initiatives worldwide to avoid reinventing the wheel and creating new stand-alone duplicative infrastructures is of critical importance for progress in AD.

This requires creating compatible data governance frameworks. Without compatible data governance frameworks the ability to authenticate, use, process and share data is compromised.

On the basis of speakers' presentations, roundtable discussions, guidance documents and scientific papers released by experts at the OECD Consultation on "*Unlocking Global Collaboration to Accelerate Innovation for Alzheimer's Disease and Dementia*", five interlinked areas were identified as urgently requiring co-ordinated international action:



1. **Sustainability:** Identify the framework conditions for the development and long-term sustainability of large databases and networks for research on Alzheimer's:
  - Consider a review of the existing national, regional and international large databases and networks for research on Alzheimer's, including the functions performed and their sustainability.
  - Identify the most relevant public data portals and databases with their respective links.
  - Consider funding strategies that can be used to establish and support the development of large databases to accelerate innovation on Alzheimer's.
  - Consider developing agreed guidance to reduce unnecessary duplication in holdings and services.
2. **Exchange and access to data:**–Identify the framework conditions and incentives to facilitate exchange and access to data:
  - Consider criteria to encourage the timely deposition of key data so as to realise the benefits of mutual research investment.
  - Consider harmonised criteria for the safe and scientifically based national, regional, and international distribution of data, based on a rational assessment of risks. These would include agreement on appropriate standards, and mutually agreed privacy and security policies.
  - Foster the development of policies to harmonise the operational parameters under which large databases function, including those governing access to data as well as their exchange and distribution, taking into account relevant national and international laws and agreements.
3. **Linkage:** Identify policies and good practices that foster co-ordination and complementarity of existing datasets sets at national, regional and international levels:
  - Consider policies and good practices that foster and enhance the performance of research for Alzheimer's from the existing large databases.
  - Facilitate international co-ordination among national databases and networks by identifying the framework conditions and good practices to ensure linkage.
  - Consider the value and economic impacts to the scientific community of a tiered networking structure to facilitate the co-ordination of large data sets at national, regional and international levels.
4. **Quality and efficiency:** Identify practices for enhanced effectiveness of existing databases:
  - Consider the scientific and technological tools and management systems necessary to meet the future needs of the emerging databases including the required minimum level of infrastructure that would ensure the highest quality of data sharing with respect to research, services, information and material.
  - Consider incentives to facilitate the sharing of relevant metadata for a more comprehensive and useful information system.

5. **Capacity building and training:** Identify incentives to promote education and training:

- Consider policies and good practices to support training and education to maintain and extend the necessary expertise to maximise the benefits of the emerging big data field.

## **ANNEX I: DRAFT AGENDA**

### ***EXPERT CONSULTATION ON UNLOCKING GLOBAL COLLABORATION TO ACCELERATE INNOVATION FOR ALZHEIMER'S DISEASE AND DEMENTIA***

**20-21 June 2013**

**The Harris Manchester College | Oxford, United Kingdom**

#### **Opening of the Meeting and Introductory Remarks**

- ***Ralph Waller*** - Principal of Harris Manchester College; Pro-Vice-Chancellor elect of the University of Oxford
- ***Elettra Ronchi and Jacqueline Allan*** - Senior Policy Analysts (OECD)

#### **Why accelerating innovation for Alzheimer's and dementia is a global priority**

- ***Michael Hodin*** - Executive Director, Global Coalition on Ageing, United States

#### **Biomedical Innovation for Alzheimer's disease and dementia: latest advances**

- ***Jean-Noël Octave*** - President Institute of NeuroScience, Catholic University of Louvain, Belgium.
- ***Takaomi Saido*** – Senior Team Leader, Laboratory for Proteolytic Neuroscience, RIKEN Brain Science Institute, Japan.

#### **From Bench to Bedside : meeting the needs of patients, academia , and industry**

- ***Richard Johnson*** - CEO, Global Helix LLC; National Academy of Sciences Board on Life Sciences; Advisory Council, Global Coalition on Aging, United States
- ***Martin Rossor*** - Vice-Chair, JPND Scientific Advisory Board.

#### **Addressing the Big Data Challenges**

- ***Donald Stuss*** - President, Ontario Brain Institute, Canada
- ***Robert Simpson*** - Researcher and Developer Zooniverse, Oxford, UK

#### **What data can be shared today? Thinking about the tomorrow.**

- ***Robert Cook Deegan*** - Director, Center for Genome Ethics, Law and Policy, Institute for Genome Sciences & Policy, Duke University, United States
- ***Simon Lovestone*** - Director NIHR Biomedical Research Centre for Mental Health, King's College, UK

## **Opening of Day 2 of the Meeting**

### **Major policy trends and new regulatory paradigms for fostering the translation of biomedical innovation for Alzheimer's disease**

- **Philippe Cupers** - Head of Sector, Neuroscience, DG Research and Innovation, European Commission
- **Mario Romao** - Senior Policy Manager, Intel Corporation, Belgium

### **Fostering Open Access for Alzheimers and dementia research**

- **Walter Kukull** - Director, National Alzheimer's Coordinating Center, United States
- **Giovanni Frisoni** - Neurologist and Deputy Scientific Director, IRCCS Fatebenefratelli, The National Centre for Alzheimer's Disease, Brescia, Italy.

### **Towards an integrated data ecosystem for new smart models of care and research**

- **Tia Powell** - Professor of Clinical Epidemiology & Population Health, Albert Einstein College of Medicine Yeshiva University, United States
- **Mehdi Khaled**, Vice President, Healthcare & Life Sciences, Oracle, Singapore

### **Innovative partnerships to facilitate the translation of biomedical innovation for Alzheimer's disease**

- **Lefkos Middleton** - Professor of Neurology, Neuroepidemiology and Ageing School of Public Health, Imperial College London, United Kingdom.
- **Zaven Khachaturian** – President, PAD2020, United States

### **How can the OECD contribute to moving the international agenda forward?**

- Moderators: **Elettra Ronchi and Jacqueline Allan** - Senior Policy Analysts (OECD)

### **Concluding Remarks -**

*End of meeting*

## NOTES

- 1 JPND Research at: [neurodegenerationresearch.eu/for-researchers/about-the-mapping-exercise/](http://neurodegenerationresearch.eu/for-researchers/about-the-mapping-exercise/)
- 2 WHO Fact sheet 362: April 2012. [www.who.int/mediacentre/factsheets/fs362/en/](http://www.who.int/mediacentre/factsheets/fs362/en/) .
- 3 <http://www.nature.com/scitable/topicpage/epistasis-gene-interaction-and-phenotype-effects-460>.
- 4 Quantitative and Systems Pharmacology in the Postgenomic Era : New Approaches to discovering drugs and understanding therapeutic mechanisms- An NIH White Paper by the QSP Workshop Group 2011. <http://isp.hms.harvard.edu/wordpress/wp-content/uploads/2011/10/NIH-Systems-Pharma-Whitepaper-Sorger-et-al-2011.pdf>)
- 5 Funding outlook for NCI: news from the experts. 2006 <http://www.apa.org/science/about/psa/2006/05/nci.aspx>
- 6 H1N1 Web and Social Media Metrics Cumulative Data Report April 22, 2009 – December 31, 2009, Division of eHealth Marketing (DeHM) National Center for Health Marketing (NCHM) Centers for Disease Control and Prevention (CDC).
- 7 Declan Butler (2013) When Google got flu wrong ; US outbreak foxes a leading web-based method for tracking seasonal flu Nature News, 13 February 2013, at: [www.nature.com/news/when-google-got-flu-wrong-1.12413](http://www.nature.com/news/when-google-got-flu-wrong-1.12413)
- 8 <https://www.zooniverse.org/>
- 9 Consortium to Establish a Registry for Alzheimer’s Disease: <http://cerad.mc.duke.edu/>.
- 10 World-wide Alzheimer’s Neuroimaging initiative: [www.alz.org/research/funding/partnerships/ww-adni\\_overview.asp](http://www.alz.org/research/funding/partnerships/ww-adni_overview.asp)
- 11 Masum H and Harris R (2011) Open Source for Neglected Diseases: Magic Bullet or Mirage? Center for Global Health R&D <http://healthresearchpolicy.org/assessments/open-source-neglected-diseases-magic-bullet-or-mirage>
- 12 [www.gov.uk/government/publications/g8-dementia-summit-agreements](http://www.gov.uk/government/publications/g8-dementia-summit-agreements)

## BIBLIOGRAPHY

- Alzheimer's Association (2010), *Changing the Trajectory of Alzheimer's disease: A National Imperative*, [www.alz.org](http://www.alz.org).
- Alzheimer's Association Report (2012), "Alzheimer's disease facts and figures", *Alzheimer's & Dementia*, 8, 131–168.
- Alzheimer's Disease International (2010), World Alzheimer Report 2010.
- Andersen C.K., Lauridsen J., Andersen K., et al., (2003), "Cost of dementia: impact of disease progression estimated in longitudinal data", *Scand J Public Health*. 31 (2):119–25.
- Arrowsmith J. (2011), "Trial watch: Phase II failures: 2008–2010", *Nat. Rev. Drug Discov*. 10 (5):328–329.
- Barabasi, A.L., Gulbahce, N. and Loscalzo, J. (2011), "Network medicine: a network-based approach to human disease", *Nature Rev Genetics*: 12 (1): 56-68.
- Barabasi, A.L. and Oltvai, Z.N. (2004), "Network biology: understanding the cell's functional organization", *Nat. Rev. Genet*. 5 (2), 101–113.
- Barnes D.E., Yaffe K., (2011), "The projected effect of risk factor reduction on Alzheimer's disease prevalence", *The Lancet Neurology*, 2011;10(9):819-828.
- Bradley, W.G., (1990), "Alzheimer's disease: "Theories of causation"", *Advances in Experimental Medicine & Biology*, 1990, 282, 31-8.
- Brown, G. (2005), "An Accelerometer Based Fall Detector: Development, Experimentation, and Analysis", Internal Report, University of California at Berkeley, July 2005.  
<http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.128.6988>.
- Callaway, E. (2012), "Alzheimer's drugs take a new tack", *Nature* . 489 (7414), 13-14.
- Chan A.W., Hrobjartsson A., Haahr M.T., Gotzsche P.C., Altman D.G. (2004), "Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles", *JAMA* 291(2):2457-65.
- Chan AW, et al. (2013). Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*; 291:2457-65
- Chen J.Y., Shen C., Sivachenko A.Y. (2006), "Mining Alzheimer disease relevant proteins from integrated protein interactome data", *Pac Symp Biocomput* 11: 367–378.
- Chesbrough, H.W. & Garman, A.R. (2009,) "How open innovation can help you cope in lean times", *Harv. Bus. Rev.* 87 (12): 68–76, 128.

- Chew C., Eysenbach G., (2010), “Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak”, *PLoS ONE*. 5 (11): e14118.
- Clay S., (2010), *Cognitive Surplus: How Technology Makes Consumers into Collaborators*, The Penguin Press, Penguin Books Ltd, England.
- Combarros, O., et al. (2009), “Epistasis in sporadic Alzheimer's disease”, *NeurobiolAging*, 30 (9): 1333-1349.
- Cooper C.P., Mallon K.P., Leadbetter S., Pollack L.A., Peipins L.A. (2005,) “Cancer Internet search activity on a major search engine, United States 2001–2003” *J Med Internet Res*. 7 (3):e36.
- Cox, N.J., et al. (1999), “Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans”, *Nature Genetics* 21 (2): 213–215
- David L.W. and Mark A.S., (2009), “Early identification and treatment of Alzheimer’s disease: social and fiscal outcomes”, *Alzheimer’s & Dementia* 5 (3) : 215 – 226.
- Decullier E., Lheritier V., Chapuis F. (2005), “Fate of biomedical research protocols and publication bias in France: retrospective cohort study”, *BMJ* 331(7507):19
- Eiben CB, et al. (2012), Increased Diels-Alderase activity through backbone remodelling guided by Foldit players *Nature Biotechnology* 30 :190-192 [www.nature.com/nbt/journal/v30/n2/full/nbt.2109.html](http://www.nature.com/nbt/journal/v30/n2/full/nbt.2109.html)
- Edwards, A.M., et al. (2009), “Open access chemical and clinical probes to support drug discovery”, *Nature Chem Biology* 5 (7) 436-440.
- Exalto L.G , Biessels, G.J. et al., (2013), “Risk score for prediction of 10 year dementia risk in individuals with type 2 diabetes: a cohort study”, *The Lancet Diabetes & Endocrinology* - 1 (3): 183-190
- Eysenbach G. (2006), “Infodemiology: tracking flu-related searches on the web for syndromic surveillance”, *AMIA Annu Symp Proc* 244-248.
- Eysenbach G., (2007), *Five-year prospective study harvesting search and click data from Google 2004-2007*, presented at AMIA Annual Fall Symposium, Chicago [www.jmir.org/2009/1/e11/](http://www.jmir.org/2009/1/e11/)
- Ferri C.P., Prince M., Brayne C., et al. (2005), “Global prevalence of dementia: a Delphi consensus study”, *Lancet* 366(9503):2112-7.
- Florez, J.C., et al.(2003), “The inherited basis of diabetes mellitus: Implications for the genetic analysis of complex traits”, *Annual Review of Genomics and Human Genetics* 4, 257–291
- Gong M.N., Winkel G., et al. (2010), “Surrogate consent for research involving adults with impaired decision making: survey of Institutional Review Board practices”. *Crit. Care Med*. 38 (11): 2146-54.
- Grochowski H. and Halem, R.D. (2003), “Technological Impact of Magnetic Hard Disk Drives on Storage Systems”, *IBM Systems Journal* 42(2): 338-346
- Guterman E.M., Markowitz J.S., Lewis B., et al., (1999). „Cost of Alzheimer’s disease and related dementia in managed-Medicare”, *J Am Geriatr Soc*. 47 (9):1065–71.
- Hughes, B. (2009), “Harnessing open innovation”, *Nat. Rev. Drug Discov*. 8 (5): 344–345.

- Hulth A., Rydevik G., and Linde A. (2009), “Web Queries as a Source for Syndromic Surveillance”, *PLoS ONE* 4(2): e4378. [www.plosone.org/article/info:doi/10.1371/journal.pone.0004378](http://www.plosone.org/article/info:doi/10.1371/journal.pone.0004378);
- Hurd, M.D., Martorell P., et al., (2013), “Monetary costs of dementia in the United States New England”, *Journal of Medicine* 368 (14): 1326-1334.
- Hurko O., and Ryan J.L. (2005), “Translational research in central nervous system drug discovery”, *NeuroRx*. 2(4): 671–682.
- International Human Genome Sequencing Consortium (2004), “Finishing the euchromatic sequence of the human genome”, *Nature* 431 (7011): 931–945.
- Joseph S.R., et al., (2012), “Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis” *BMJ* 344:d7292
- Khachaturian, Z.S., (2012), “Perspectives on Alzheimer’s Disease: Past, Present and Future”. In: Carrillo, M.C. and Hampel H (2012) (eds.), *Alzheimer’s Disease : Modernising Concept, Biological Diagnosis and Therapy*, *Advances in Biological Psychiatry*, Basel, Karger, vol 28, pp 179–188.
- Klompas M., Eggleston. E. (2012), “Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data”, *Diabetes Care* 36(4): 914-921
- Kola I., and Landis J., (2004), “Can the pharmaceutical industry reduce attrition rates”, *Nat. Rev. Drug Discov.* 3 (8):711–715.
- Kozauer N., Katz , R., (2013), “Regulatory Innovation and Drug Development for Early Stage Alzheimer’s Disease”, *N Engl J med* 368 (13): 1169-1171.
- Liu B., Jiang T., et al., (2006), “Exploring candidate genes for human brain diseases from a brain-specific gene network”, *Biochem Biophys Res Commun* 349 (4): 1308–1314.
- McKinsey&Company, (2013), “The big data revolution in health care: Accelerating value and innovation “ McKinsey&Company
- Moore M.J., Zhu C.W., and Clipp E.C., (2001), “Informal costs of dementia care: estimates from the National Longitudinal Caregiver Study”, *J Gerontol B Psychol Sci Soc Sci.* 56 (4):S219–28.
- OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD Publishing. doi: [10.1787/9789264034020-en-fr](https://doi.org/10.1787/9789264034020-en-fr).
- OECD (2012) *OECD Science, Technology and Industry Outlook 2012*, OECD Publishing. doi: [10.1787/sti\\_outlook-2012-en](https://doi.org/10.1787/sti_outlook-2012-en)
- OECD (2013) *Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges*, OECD Health Policy Studies, OECD Publishing. Doi: [10.1787/9789264193505-en](https://doi.org/10.1787/9789264193505-en)
- Olson D.R., Konty K.J., Paladini M., Viboud C., et al. (2013), “Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales”, *PLoS Comput Biol* 9(10): e1003256. [www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1003256](http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1003256).



- Rising K., Bacchetti P., and Bero L.A. (2008), “Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation”, *PLoS Med* 5 (11):e217.
- Robillard J.M., Johnson T.W., Hennessey C., Beattie B.L., et al. (2013), *Aging 2.0: Health Information about Dementia on Twitter*. PLoS ONE 8(7): e69861. doi:10.1371/journal.pone.0069861.
- Ross J.S., Mulvey G.K., et al., (2009), “Trial publication after registration in ClinicalTrials.gov: across-sectional analysis”, *PLoS Med* 6 (9):e1000144.
- Ross J.S. et al. , (2012) Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis *BMJ* ;344:d7292
- Scanfeld D., Scanfeld V., and Larson E.L. (2010), “Dissemination of health information through social networks: Twitter and antibiotics”, *American Journal of Infection Control* 38 (3): 182–188.
- Sadilek A., Kautz H., Silenzio V. (2012), “Predicting disease transmission from geo-tagged micro-blog data”, *Conf Proc 26th AAAI Conference on Artificial Intelligence* 136-42.
- Signorini A., Segre A.M., Polgreen P.M. (2011), “The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic”, *PLOS One* 6(5):e19467.
- Small G.W., Donohue J.A., Brooks R.L. (1998), “An economic evaluation of donepezil in the treatment of Alzheimer’s disease”, *Clin Ther.* 20 (4):838–50.
- Turner E.H., et al., (2008), “Selective publication of antidepressant trials and its influence on apparent efficacy”, *N Engl J Med* 358 (3):252-60.
- Vidal, M., Cusick, M. E. and Barabasi, A.L. (2011), “Interactome networks and human disease” *Cell* 144:986–998.
- Von Elm E., Rollin A., Blumle A., Huwiler K., Witschi M., Egger M. (2008), “Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee”, *Swiss Med Wkly* 138 (13-14):197-203.
- Weimer DL and Sager MA. (2009), “Early identification and treatment of Alzheimer’s disease: social and fiscal outcomes.” *Alzheimer’s and Dementia* 5. 215 – 226.
- Wicks P., Vaughan T.E., et al, (2011), “Accelerated clinical discovery using self-reported patient data collected online and a patient matching algorithm”, *Nature Biotechnology* 29 (5): 411-414.
- Wiltshire, S., et al. (2006), “Epistasis between type 2 diabetes susceptibility loci on chromosomes 1q21–25 and 10q23–26 in northern Europeans”, *Annals of Human Genetics* 70(Pt6):726–737.
- Yin J., et al., (2012), “ESA: Emergency Situation Awareness via Microbloggers”, *CIKM 2012*  
<http://www.ict.csiro.au/staff/jie.yin/files/de0418-yin-CIKM12.pdf>