A nation-wide initiative for brain imaging and clinical phenotype data federation in Swiss university memory centres

Bogdan Draganski[a,b,c,*], *Ferath Kherif*[a,*], *Daniel Damian*[c], *Jean-Francois Demonet*[c], *for the MemoNet consortium*

Purpose of review
The goal of our nation-wide initiative is to provide clinicians intuitive and robust tools for accurate diagnosis, therapy monitoring and prognosis of cognitive decline that is based on large-scale multidomain data.

Recent findings
We describe a federation framework that allows for statistical analysis of aggregated brain imaging and clinical phenotyping data across memory clinics in Switzerland. The adaptation and deployment of readily available data capturing and federation modules is paralleled by developments in ontology, quality and regulatory control of brain imaging data. Our initiative incentivizes data sharing through the common resource in a way that provides individual researcher with access to large-scale data that surpasses the data acquisition capacity of a single centre. Clinicians benefit from fine-grained epidemiological characterization of own data compared with the rest additional to intuitive tools allowing for computer-based diagnosis of dementia. Finally, our concept aims at closing the loop between group-level results based on aggregate data and individual diagnosis by providing disease models, that is, classifiers for neurocognitive disorders that will enable the computer-based diagnosis of individual patients.

Summary
The obtained results will inform recommendations on best clinical practice in all relevant fields focusing on standardization and interoperability of acquired data, privacy protection framework and ethical consideration in the context of evolutive pathology.

Keywords
data federation, dementia, MRI

INTRODUCTION

With increased life expectancy in most economically developed countries, the number of individuals that will potentially become demented is growing almost proportionally. Current estimates report world-wide ca. 48 million people suffering from dementia, which brings the socioeconomic cost of care up to 1% of world's gross domestic product [1]. The fact that brain damage begins years before clinical symptoms underscores the importance of early diagnosis to allow for neuroprotective treatment before critical damage occurs. Early and accurate detection of the underlying causes – neurodegeneration as in Alzheimer's disease, vascular pathology or others, is essential for developing strategies that if not preventing or curing, then at least could delay the onset of disease symptoms [2].

A reality check on the current situation in clinical neurosciences allows for detecting two main issues that, despite the overwhelming progress in understanding the brain, hamper the progress

[a]LREN, Department of Clinical Neurosciences, University Hospital Centre (CHUV) and University of Lausanne (UNIL), Switzerland, [b]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany and [c]Leenaards Memory Centre, Department of clinical neurosciences, University Hospital Centre (CHUV) and University of Lausanne (UNIL), Switzerland

Correspondence to Bogdan Draganski, LREN, Department of Clinical Neurosciences, CHUV, University of Lausanne, Lausanne, Switzerland. Tel: +41 21 314 96 38; e-mail: bogdan.draganski@chuv.ch

*Bogdan Draganski and Ferath Kherif shared first authorship.

Curr Opin Neurol 2019, 32:557–563

DOI:10.1097/WCO.0000000000000721

KEY POINTS

- The CLEMENS data management system provides a viable option for routine acquisition of structured data in the context of a tertiary memory centre.
- Data federation systems represent a viable alternative to data centralization.
- Specific ethical considerations are needed in diseases with progressive course that potentially impacts decision-making.

in developing effective therapies for cognitive decline: poor technical solutions for federation of existing data combined with lack of an integrated system of partnerships between scientists, clinicians and patients [3▪▪]. The increasing pressure for early and accurate diagnosis of cognitive impairment and dementia leads to the cumulation of locally established and validated diagnostic algorithms without concomitant efforts for harmonization and interoperability across specialized centres. Aiming to comply with data protection regulations, individual diagnostic centres set up their own electronic health records (EHR) and picture archiving and communication systems (PACS) to collect patient data, which are currently spread across hospitals, clinics and private practices. The lack of a viable concept for data integration arises not only from data protection restrictions but also can be explained by the scarcity of Information Technology solutions for research on multicentre large-scale clinical data, which hampers collaborative research efforts for personalized medicine at the national and international level. A non-negligible scientific issue is the missing consensus on a diagnostic framework that would allow to map the evolving clinical phenotypes on a multidimensional description of brain damage.

There is, however, a steadily growing number of successful data repository and meta-analysis initiatives, such as the *Alzheimer's Disease Neuroimaging Initiative* (ADNI), the *Enhancing NeuroImaging Genetics through Meta-Analysis* consortium (ENIGMA), the *Innovative Medicines Initiatives – European Medical Information Framework – EMIF, European Prevention of Alzheimer's Disease – EPAD*, which are used by many researchers world-wide. Similarly, recent medical informatics solutions provide promising outlook in the context of large-scale data management and standardized preprocessing [4,5].

The critical impact of cognitive aging on our social-security systems and economy requires a paradigm shift that brings together multiple stakeholders through a system allowing the creation of stable collaborations. The Association of Swiss Memory Clinics – SMC (www.swissmemoryclinics.ch), a tightly woven network that links specialized memory centres in Switzerland represents such an integrated system that has the potential to facilitate the roll-out of innovation and technical standards in clinical neurosciences. Recent SMC recommendations for diagnostic work-up in patients with memory complaints paved the way for harmonization and interoperability of acquired data across different clinical domains [6].

The five Swiss university memory clinical and research centres proposed a strategy for data capturing in the clinical routine, coupled with a flexible data federation framework that allows for integration of 'real-world' clinical data across SMCs. The locally deployed Information Technology modules are using state-of-the-art technology for automated feature extraction from brain imaging data to then allow creation of local aggregates of de-identified clinical information and multidomain biomarkers. The local aggregates can be queried following specific research question and then analyzed at the federated level using statistical methods customized for distributed data. Our strategy builds on using existing data to reach a data-driven consensus on classification of cognitive impairment and to establish priority neuropsychological assessment tests, cerebro-spinal fluid (CSF) and imaging biomarkers. This is followed by creation of machine-learning based models of cognitive impairment that are used at the local centres for individual's computer-based diagnosis. Closing the loop between knowledge generated from large-scale multidomain data and individual's diagnosis by experts, will allow for true personalized medicine approach with steadily improvement of disease models accuracy.

SYSTEM COMPONENTS

Data capturing

We build on the readily available data capturing system CLEMENS, that led to the creation and maintenance of a state-of-the-art patient registry [7▪▪]. The CLEMENS concept is a translational information management system allowing for capturing multimodal data including clinical phenotype and neuropsychology test results, brain imaging (MRI, PET) and blood/CSF biomarkers that are acquired for diagnostic purpose on a routine basis. Data is collected and managed with an immediate short-term benefit of facilitating monitoring key performance indicators for clinical practice and a long-term strategy of reuse of clinical patient data in scientific research into memory-related disease(s) that

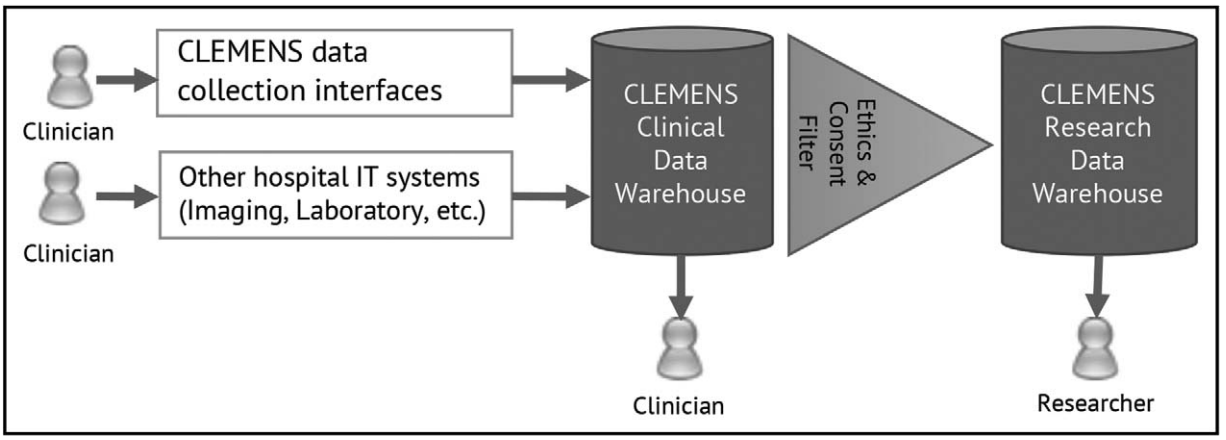

FIGURE 1. Data management. Schematic representation of the CLEMENS data management workflow with interface to the primary sources of information – clinicians and hospital IT system to create protected clinical data warehouse and after filtering privacy sensitive information – research data warehouse.

complies with informed consent and data protection regulation. The key point here is that CLEMENS fills the gap between nonresearch-oriented clinical routine and research-oriented structured data.

CLEMENS represents a server-based application that can be deployed on servers within existing intranet networks or can be hosted on remote servers and accessed via secured connections (Fig. 1). All users of CLEMENS access the software via standard browser with a secured and authenticated connection according to the most up-to-date industry standards. Building on top of generic open-source software platforms, the system organizes data management through a series of microapplications that cover various aspects of a memory clinic's activity, such as recording of incoming patient referrals, outcomes of multidisciplinary evaluation, clinical observations, neuropsychological tests and caregiver status (Fig. 2). The system allows for site-specific settings, while being conform with the concept of a 'minimal dataset' common to all participating sites.

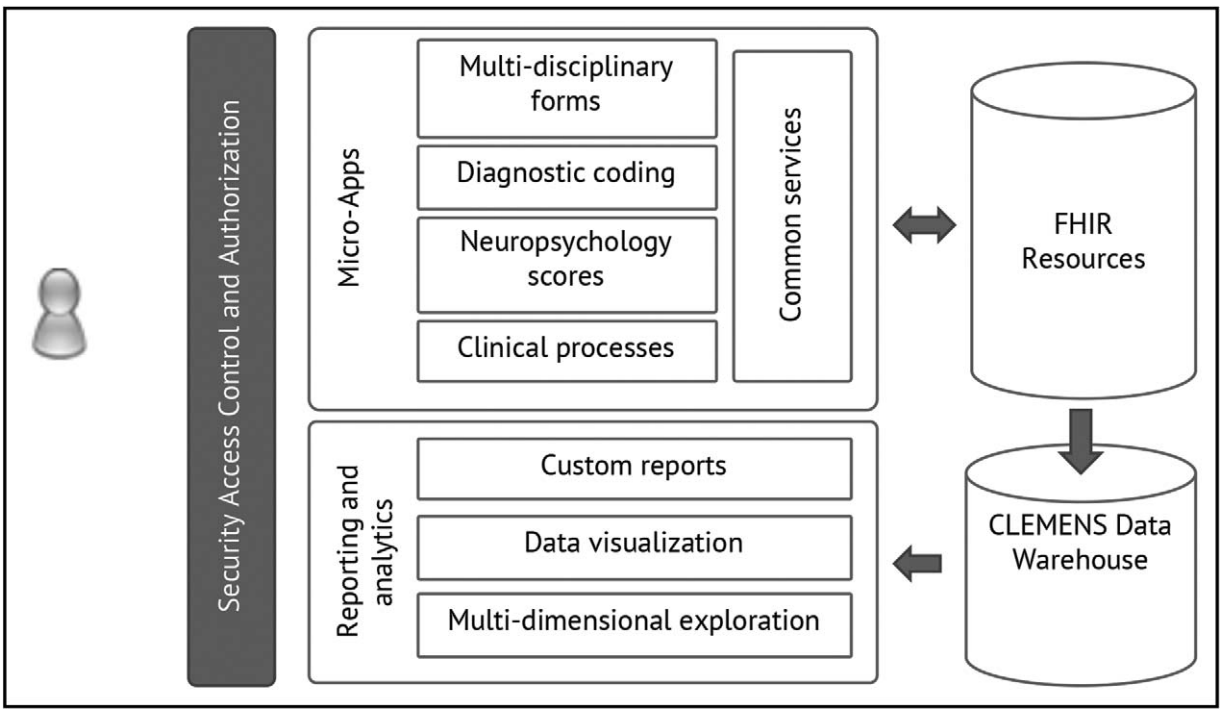

FIGURE 2. CLEMENS system. Schematic representation of the CLEMENS system components consisting of microapps tapping into specific domains of clinical work-up to provide common services that pass through the CLEMENS data warehouse and allow for visualization of query results and multidimensional exploration.

The CLEMENS package uses a standardized data exchange model based on FHIR (Fast Healthcare Interoperability Resources), a standard created by HL7. FHIR structures allow interchange of data between Information Technology systems in the healthcare domain and allow CLEMENS to be easily adapted to communicate and exchange information with existing systems in complex hospital Information Technology environments. The use of FHIR structures also facilitates encoding of data using other clinical or third-party standards such as ICD-10, LOINC, SNOMED-CT and others. The system uses the latest technologies in enterprise system security to ensure correct identification of its users and their associated roles, as well as state-of-the art cryptography protocols for the encryption of all communication with the system.

Data federation

Data federation allows to bring data from multiple locations together without moving or centralizing the data where data can be queried and analyzed in real time regardless where it is stored. Our data federation system represents a software solution that connect patients' data and provides set of methods ranging from automated feature extraction to statistical methods for data exploration, statistical modelling, predictive machine learning and visualization of the results. The data deployment model is hybrid representing a composition of two distinct infrastructure types – private, that is, protected and community execution environments. In hospital execution environments, patients' data are de-identified, extracted from the medical records and other data sources, stored as multidimensional arrays to be then analyzed. The instance(s) running in community execution environment(s) orchestrate the distributed execution of data analytic algorithms, provide result pooling and aggregates those partial results in the information commons built on the large cross-centre multidataset patient population.

The data federation system has local modules for processing, storing and analyzing de-identified and harmonized neuroimaging, neuropsychological, biological and demographic data (Fig. 3). The federation system is built with privacy-by-design – only the results of the queries or the analyses are extracted, whilst the original data remain in their primary location [8▪]. The system includes user interface for selecting data, building models, sharing machine-learning models with distributed model training and cross-site validation and visualization of the results. This web application provides the

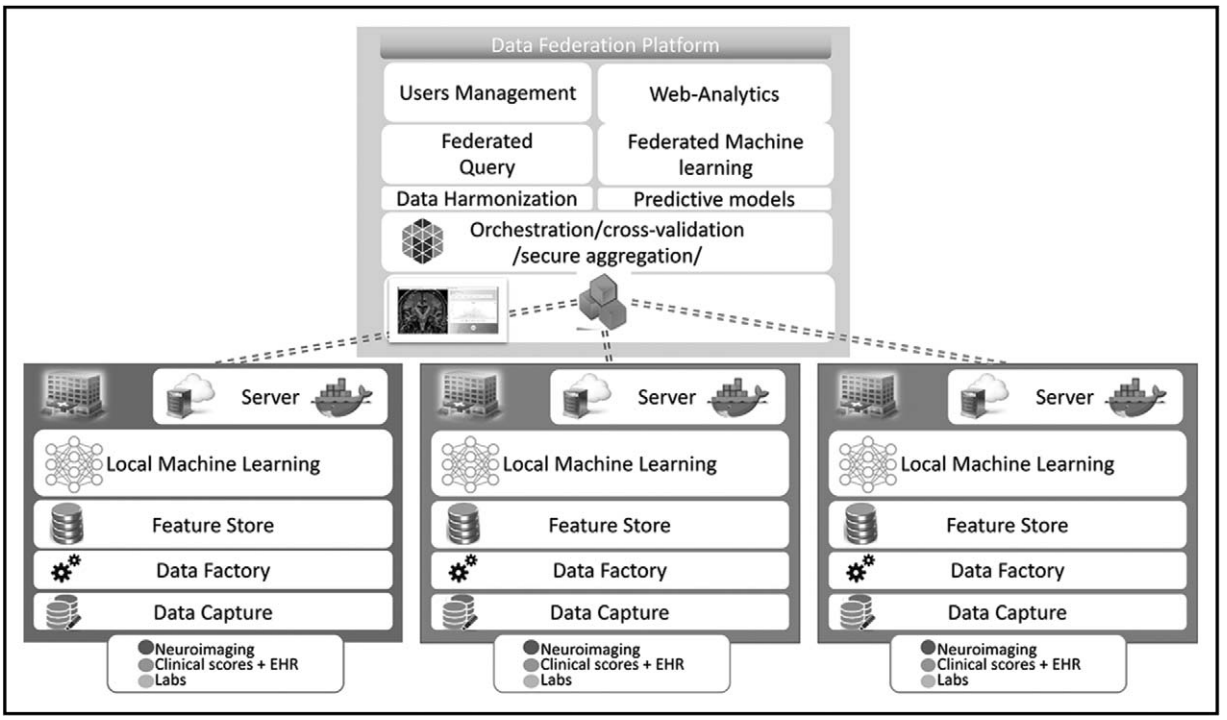

FIGURE 3. Data federation. Schematic representation of the Data federation framework with centre-specific privacy protecting environment interacting with research data warehouse content to extract features and allow for centre-specific statistical analysis of queries submitted to the central data federation platform. This central platform orchestrates and aggregates the centre-specific meta data.

functionality to search, select and classify data elements as variables, co-variables and filters for configuration of the statistical or machine learning models. The interface allows collaborations between clinicians from the different SMCs. The system includes an automated end-to-end software installation, configuration and system monitoring, which minimizes on-site Information Technology efforts. In addition, the federation system is based on microservice architecture, which enables agile continuous integration and continuous deployment of new components. The docker-based microservice deployment architecture is configurable to provide compliance to the specific hospital security policy and data centre architecture. Data capture, data processing and dataset containers can run in the red data centre zone, whereas the web service containers could run in the zones with less restrictive access. OAuth-based authentication and authorization in combination with https access to the transport layer security-encrypted data provide secure web access.

MRI processing and harmonization

The MRI data processing subsystem is built using the 'data factory' solution of the data federation system: the data factory allows stacking of data preprocessing tasks (DICOM to nifti conversion, image preprocessing, feature extraction) into pipeline and workflows. Local brain features – for example, grey matter volume, cortical thickness, are measured using SPM12 – an open source software package written in Matlab (www.fil.ion.ucl.ac.uk/spm, Wellcome Trust Centre for Neuroimaging, UCL London, UK).

The conventional radiological assessment of neurocognitive disorders includes a qualitative estimation of structural brain changes based on isotropic 3D anatomical T1-weighted MRI data. This evaluation is enriched by investigation of white matter changes, parenchymal defects and microbleeds using dedicated MRI protocols. Given the overwhelming heterogeneity of applied imaging protocols across hospitals in Switzerland, there is a pressing need for standardization and harmonization that we systematically address in three iterative phases: neuroradiological evaluation of MRI data from all project sites to systematically investigate the quality necessary for neuroradiological assessment; establishing recommendations for standardized neuroradiological assessment and structured reporting for MRI evaluation in neurocognitive disorders; statistical comparison between standardized automated postprocessing and analysis with the results of the experts' 'structured reports'.

VALIDATION APPROACH

Our initiative for nation-wide data federation has the potential for a breakthrough in one of the most significant public-health and socioeconomic challenges for our ageing society – cognitive decline leading to dementia. The short-term goal of the initiative is to reach a data-driven consensus on taxonomy in neurocognitive disorders that will allow for an informed selection of clinical variables from patients' eHR. Building on the readily available data capturing federation systems, we will stratify homogenous subgroups of patients within each participating site that follow well established or new classification schemes [9,10▪▪]. These emerging disease models will be tested within and across centres for consistency to then look for a particular classification scheme for neurocognitive disorders that shows the best explanatory power for the winning model. Building on the obtained results, we will follow-up with an iterative phase of improving the diagnostic accuracy of the models for neurocognitive disorders stemming from the previous step that are validated on the basis of experts' opinion. Disease models created at the federated level from aggregate data will be compared for diagnostic accuracy with within-centre models.

Given the lack of group-to-individual generalizability of research-derived inferences, we provide a closed-loop system that extracts knowledge from aggregated large-scale clinical data to offer this as a computer-based diagnostic tool to experts in the field that validate its diagnostic accuracy on individual patients and enable pragmatic trials at large-scale. The proposed browser-based interactive platform represents a major incentive to clinicians that can now benefit from knowledge that is usually restricted to specialized fields of research and provide better service to their patients. Most importantly, the empirical evidence gathered at the federated level will guide in informed way further initiatives for data standardization and harmonization in the relevant clinical domains across SMCs. The incrementally increasing data interoperability will render more high-quality data to refine and diversify the created disease models. The augmenting diagnostic accuracy of the provided disease models will attract steadily growing number of clinicians that validate the model accuracy to provide unique opportunities for personalized treatment and monitoring of clinical outcome. Our readily available solutions for data capturing and federation are scalable and low maintenance, which renders our project feasible within the timeframe of the project. In parallel, we will test the flexibility of our system to expand beyond the Association of SMCs in other clinical domains.

The intention here is to bring descriptive and predictive machine-learning modelling into the clinics. Bottom up models – that is, predictive models, for example, 'can brain features determine the diagnosis?', are not usually available to clinicians. This is because of limited access to required data that are as stored in eHR/PACS, the necessity for standardized data preprocessing for complex imaging data, the necessity for appropriate, specialized tools and know-how and knowledge of data mining – that is, automated data analysis to uncover specific patterns. We provide a wide range of tools for data exploration and model builder to allow testing patterns detected through bottom-up data mining in a top-down approach.

PRIVACY AND ETHICAL CONSIDERATIONS

Collecting data from patients with neurodegenerative or ageing-associated cognitive disorders is an ethically complex task that requires outmost ethical safeguards. There is a wide consensus that informed consent is pivotal for respecting the autonomy of patients and research participants. However, there is uncertainty with regard to which consent mechanism is best suited for older people with cognitive disabilities. One major challenge associated with persons with cognitive disabilities is determining whether they have the necessary competence for consent. Cognitive skills notoriously deteriorate over time with advancing disease progression, making informed consent difficult if not impossible to obtain. The assessment of competence, which is a prerequisite for consent, cannot rely exclusively on an assessment of cognitive performance (e.g. MMSE, MoCA) but requires additional criteria, such as a capacity to understand and retain information and an ability to communicate a decision. As Alzheimer Europe recommends, whenever informed consent cannot be obtained directly from the older person because of the underlying disease or a lack of competence, indirect consent should seek either through advance directives or via proxy decision making [11]. In 2019 Swissethics and the Unimed-Suisse developed a template for a General Consent document that is expected to provide a Switzerland-wide resource to request patients to use their health-related personal data and samples for medical research [12]. Through the General Consent, patients who receive treatment in a Swiss hospital and/or clinics can consent to the further reuse of their data and samples for research purposes.

CONCLUSION

Our initiative will provide a proof-of-concept for a framework for personalized health in the domain of ageing associated neurocognitive disorders by addressing the noninteroperability of patient clinical data to then link them with all biomedical and other health-relevant data. The goal here is to make these data available to researchers for collaborative research. The far-reaching goal of our initiative is to bridge the gap between basic research and clinical practice by presenting research results in form of descriptive, predictive and prescriptive disease models that can be used by clinical researchers for individual diagnosis, prognosis and for pragmatic clinical trials.

Acknowledgements

MemoNet consortium: J.M. Annoni (Neurology service, Cantonal hospital Fribourg, Switzerland), M. Bürge (BESAS Berner Spitalzentrum für Altersmedizin Siloah, Bern, Switzerland; Association Swiss Memory Clinics, Bern, Switzerland); J. Ghika (Neurology service, Hôpital du Valais Sion, Switzerland; Leenaards Memory Centre, Department of Clinical Neurosciences, University Hospital Centre (CHUV) and University of Lausanne (UNIL), Switzerland); G. Frisoni (Memory clinic, Department of psychiatry, University of Geneva and Geneva University Hospitals, Switzerland), A. Gietl (Institute for Regenerative Medicine, Centre for Prevention and Dementia Therapy, University of Zurich, Switzerland); C.h. Hock (Institute for Regenerative Medicine, Centre for Prevention and Dementia Therapy, University of Zurich, Switzerland,); S. Klöppel (Old age psychiatry service, Department of psychiatry, University of Bern, Switzerland); A.U. Monsch (University Department of Geriatric Medicine Felix Platter, University of Basel, Switzerland); O. Rouaud [Leenaards Memory Centre, Department of Clinical Neurosciences, University Hospital Centre (CHUV) and University of Lausanne (UNIL), Switzerland]; J.P. Thiran [Signal Processing Lab, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland; Department of Radiology, University Hospital Centre (CHUV) and University of Lausanne (UNIL), Switzerland].

The authors of the study express their gratitude to Dr Effy Vayena for her contribution on the ethics part of the manuscript.

Financial support and sponsorship

B.D. is supported by the Swiss National Science Foundation (NCCR Synapsy, project grant Nr 32003B_159780 and 33CS30-148401) and the Leenaards Foundation. L.R.E.N. is very grateful to the Roger De Spoelberch and Partridge Foundations for their generous financial support.

Conflicts of interest

There are no conflicts of interest.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- ■ of special interest
- ■■ of outstanding interest

1. Association A. 2018 Alzheimer's disease facts and figures. Alzheimers Dement 2018; 14:367–429.
2. Gorelick PB, Furie KL, Iadecola C, *et al.,* American Heart Association/American Stroke Association. Defining optimal brain health in adults: a presidential advisory from the American Heart Association/American Stroke Association. Stroke 2017; 48:e284–e303.
3. ■■ Frisoni GB, Boccardi M, Barkhof F, *et al.* Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. Lancet Neurol 2017; 16:661–676.

This study provides a clear vision for the harmonization and use of biomarkers for diagnosis and prognostication in dementia.

4. Das S, Glatard T, Rogers C, *et al.* Cyberinfrastructure for open science at the Montreal Neurological Institute. Front Neuroinformatics 2017; 10:53.
5. Ming J, Verner E, Sarwate A, *et al.* COINSTAC: Decentralizing the future of brain imaging analysis. F1000Research 2017; 6:1512.
6. Bürge M, Bieri G, Brühlmeier M, *et al.* Recommendations of Swiss Memory Clinics for the diagnosis of dementia. Praxis (Bern 1994) 2018; 107: 435–451.
7. ■■ Damian D, Rouaud O, Draganski B, *et al.* Memory center: the Lausanne model. Clin Transl Neurosci 2018; 2:1–8.

Comprehensive description of a data management system in the context of a tertiary memory centre in Switzerland – rich on technical details and implementation in the daily routine.

8. ■ Melie-Garcia L, Draganski B, Ashburner J, Kherif F. Multiple linear regression: Bayesian inference for distributed and big data in the medical informatics platform of the Human Brain Project. bioRxiv 2018; 242883. doi: https://doi.org/10.1101/242883.

Up-to-date rare example of statistical framework for embedding data federation systems.

9. Bailly L, David R, Chevrier R, *et al.* Alzheimer's disease: estimating its prevalence rate in a French geographical unit using the National Alzheimer Data Bank and national health insurance information systems. PloS One 2019; 14:e0216221.
10. ■■ Jack CR, Bennett DA, Blennow K, *et al.,* Contributors. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement J Alzheimers Assoc 2018; 14:535–562.

One of the most recent attempts to provide a neurobiological definition of dementia because of Alzheimer's disease.

11. Alzheimer Europe - Ethics - Ethical issues in practice - 2011: Ethics of dementia research - informed consent to dementia research [Internet]. [cited 2019 Apr 28]. Available at: https://www.alzheimer-europe.org/Ethics/Ethical-issues-in-practice/2011-Ethics-of-dementia-research/Informed-consent-to-dementia-research
12. Generalkonsent für die Forschung: Unimedsuisse - der Verband der Schweizer Hochschulmedizin [Internet]. [cited 2019 Apr 28]. Available at: https://www.unimedsuisse.ch/de/projekte/generalkonsent