MS DATACONNECT
– CONNECT DATA, CONNECT PEOPLE –
www.msdataconnect.com

We aim to transform the mostly population based management of MS of today into an individualized, personalized and precision level management. We believe the key to achieve this next level is “FAIR” data.

FAIR data for next generation management of Multiple Sclerosis
A FOUR C-PLAN FOR SUCCESS

F.A.I.R. data
Findable, Accessible, Interoperable and Re-usable

Step 1: COLLECT
Develop data collection procedures and tools to create data that is FAIR (Findable, Accessible, Interoperable and Re-usable)

Step 2: CONNECT
Develop IT solutions to allow (temporarily) pooling and linking of FAIR datasets

Step 3: COMPLETE
Develop statistical methods to define minimal requirements for datasets

Step 4: CONSTRUCT
Develop new analytical methods for optimal mining of connected and pooled FAIR datasets

An intuitive representation of this 4C plan is presented. Data is collected all over the world by different stakeholders (step 1: COLLECT). This results in an extensive amount of data and datasets. Every dataset is represented using a puzzle of a face. Many insights could be discovered when these datasets could be pooled and connected (step 2: CONNECT). Methods to identify the minimal requirements for common datasets are required (step 3: COMPLETE). When sufficient overlap between the databases involved is secured and powerful analytical methods are developed to cope with the imperfections of datasets featured by different layers of missing data, these data sets can be optimally mined to create new insights for MS management (step 4: CONSTRUCT).

Proof-of-concept is provided by building a multidisciplinary repository for MS in Belgium

COLLECT
Data is collected 1st in the hospitals or laboratories using the primary systems 2nd by patients 3rd by independent practitioners using the electronic medical dossier (EMD). The data is registered and captured by optimizing and expanding 3rd existing open source IT software (HD4OP, HD4PROM, HD4Pyc) (= 0 & 1 in fig)

CONNECT
Data is securely transferred, encrypted and pseudo-anonymized by a Trusted Third Party. The use of the National Registry Number as identification number in all data collection steps makes it possible to connect all data belonging to the same patient (2 in fig).

COMPLETE
The HD4RES software is used by the researchers for monitoring, follow-up and managing the data collections (3 & 4 in fig)

CONSTRUCT
Data is stored in a secure data warehouse. Researchers are granted access and analysis of the data is possible in the data warehouse using standard statistical tools. Data is used for research, for example to build personalized decision support systems and to evaluate and construct composite outcome score measures. The repository is used to provide Health Statistics for all stakeholders involved (5, 6 & 7 in fig)

An online catalogue listing the desired content of the repository is defined. This list contains, among others, following variables:

1° patient specific data
2° disease specific data
3° treatment strategies
4° paramedical data
5° clinical data
6° patient reported outcomes data
7° biological sample specific data
8° patient and sample phenotyping