Contents

1. Introduction

The NFU Data4lifesciences (D4LS) program's objective is to improve data findability, accessibility, integration, and reusability (FAIR[1]), and promote sharing of expensive IT facilities in such a manner that the entire data infrastructure will appear to the (international) researcher as a coherent set of high end data services from one virtual 'UMC.nl'. For example, researchers should be able to utilize datasets gathered from any UMC/hospital, use the high-performance compute clusters of all UMCs as well as supporting institutes like SURF and CIT for computer-intensive analyses, and benefit from data handling processes that have been standardized to speed up integration and to improve data quality. The NFU program is organized in work packages according to the below figure.

Figure 1 Overview of work packages NFU Data4lifesciences program

1.1. Original aims stated

Research data must be integer and confidential, patient rights must be respected, and at the same time open access is increasingly demanded. Privacy is a hot topic (*e.g.* European single law: General Data Protection Regulation). Main issues relate to information security & privacy aspects (*e.g.* correct use of informed consent), pseudonymisation, and trusted third party service (TTP) for clinical research.

1.2. Barriers to overtake

A common understanding of the problems is crucial to a successful resolution to the problems. In this paragraph the barriers of data exchange between UMC's and Good Research Practice policy issues are described.

---

[1] *https://www.force11.org/group/fairgroup/fairprinciples*

Data Linkage (BSN use)

The use of personal health information such as medication use in biomedical and clinical research (*i.e.* secondary use of clinical data) requires the protection of the privacy of the individual. Direct exchange of research data using the Burger Service Number (BSN), the unique number for individuals in the Netherlands, is not allowed. Would it be possible to use an encrypted BSN number (or another unique patient identifier) under certain circumstances, and what are the exact prerequisites for this usage?

TTP usage

In the Netherlands researchers must comply with the 'Wet bescherming persoonsgegevens (WBP)' verified by the "Autoriteit Persoonsgegevens" (untill 2016 operating under the name "College Bescherming Persoonsgegevens" (CBP)). In recent years many projects and institutes have started using Trusted Third Parties (TTP) for anonymization and pseudonymisation services looking for privacy-enhancing technology in the balance between good electronic communication of clinical data at the one hand, and the protection of the individual's privacy at the other hand.

Security Policies

Shared use of sensitive data (*e.g.* NFU Parelsnoer hosting) asks for additional security measures such as risk analysis, separate NEN 7510 certification, lean audit process, regular penetration tests, harmonization of data processing agreements ("*databewerkersovereenkomsten"*) and a coordinated approach to the new European GDPR legislation (General Data Protection Regulation).

The Data4lifesciences program has specified the following goals for work package 6 "Good Research Practice" in the initial project plan in September 2014:

1. Data storage and transportation between UMCs and project members of multicenter studies is secure.
2. To facilitate big data sets (international) by standardized TTP services, linked together on a NL and EU scale.

In the next chapter the overall aim, also called the True North, of Data4lifesciense and the updated aims for this work package are described.

2. Deliverables plan

2.1. Overall aim of the Data4lifesciences program – True North

The overall aim or True North of this Data4lifesciences program, as defined by Andre Dekker (WP5 lead) and Jeroen Beliën (WP2 lead), but modified for this WP, is:

What	all clinical, biosample, imaging and experimental data, including metadata (*e.g.* context, definitions of data) of all patients and/or study-participants should be made available
Who	by every UMC of the Netherlands
Why	for every valid research question
When	now and forevermore
Where	in a scalable, secure, and distributed environment
How	as findable, accessible, interoperable and re-usable (FAIR) data, with full protection of privacy of patients, with full control by the owner (*e.g.* UMC; taking into consideration the "privacy" of researchers and research consortia to develop products as result of their research) and without disturbing/impacting clinical care.

2.2. Specific aims and deliverables for WP6

This work package aims to achieve the prerequisites of 'good research practice' within the program as a whole. Privacy is a controversial and important issue in scientific research. The patient or healthy participant must be reassured that his or her data will not be made public. This issue is becoming increasingly important as a result of the technological developments in 'big data'. This work package implements procedures and technologies from privacy and security perspective in accordance with the guidelines specified in HANDS (Handbook for Adequate Data Stewardship; D4LS WP1). This may include for instance the pseudonymisation of personal data via Trusted Third Parties, the encryption of citizen service numbers, and implementation of security standards.

The work package has several adjoining initiatives:
- WP1: describes practical implementation of the guidelines in HANDS.
- WP2: defines the D4LS IT architecture principles to which practical tools within multicenter study must be compliant
- WP4: creates the conditions (procedures, resources, tools) to enable the sharing and analysis of research data within multicenter study (Digital Research Environment)
- WP 9: provides the human network between the institutions to share information / documents on 'Good Research Practice'

Because of the diversity of activities, this work package has been divided into a number of topics (themes), each with its own expert group. The work package leader is responsible for the overall coordination and liaison between the expert groups, as well as towards other work packages and adjoining initiatives outside the program.

The topics of the work package are:

Theme Toolbox
- Development of a D4LS 'data classification' toolbox (framework) which supports end-users with *Good Research Practice* information, and based on the data life cycle described in HANDS (WP1).
- The toolbox enables research data management within multi-center studies (*i.e.* data exchange) and contains a decision tree to support end-users (researchers, data managers) in their selection process to decide which tool, solution or service is suitable for the specific research process. The collection and publishing of national and local UMC initiatives within the toolbox is in collaboration with WP1 (HANDS) and WP9 (Access to experts, trainings and support).

Theme TTP service
- Policy-making and implementation on the topic of safe and secure data linkage.
- Strategy development and political lobby for efficient exchange of research data within the UMCs and related organizations (*e.g.* use of social security number).
- Investigate existing and new development methods for secure data linkage solutions.
- Develop and deploy the functional roadmap for the national TTP service including an expert network and involvement of organizations with related initiatives such as LCRDM (Landelijke Coördinatiepunt Research Data Management), CBS, Surf, BBMRI-NL, Coreon and Mondriaan

Theme Electronic Lab Notebook (ELN)
- Support developments in the field of digitization of data collections in research laboratories and the need for centralizing knowledge on this topic. The collaboration consists of exchanging information about procedures, maintenance organization, pricing models and agreement about the position of ELN within the research domain/reference architecture (collaboration with WP2).
- Identifying opportunities for joint procurement of a ELN tool or additional tools and functionalities, in collaboration with SURF.

Approach

For each described theme a working group is composed, existing of representatives from the UMCs.

The work package leader is responsible for the coordination and organization of required meetings. For the TTP service or 'Secure Data Link' subject a working group has been started in conjunction with LCRDM. This Special Interest Group (SIG) consists of a core team of subject experts and reports to a larger research community, as well as the LCRDM (*click here* for more information).

Deliverables

The work package 6 specific aims and associated activities/deliverables contributing to the True North are listed in the table below.

Q: quarter of a year
H: half of a year

Theme D4LS Toolbox			
Number	Activity/Deliverable	Approach	Due date
T1	Toolbox design	Functional and technical design of the data classification toolbox. Technology platform will be chosen.	Q3 2017
T2	Toolbox pilot	Development and deployment of the toolbox for a single UMC.	Q1 2018
T3	Toolbox beta release	Toolbox framework is available for all the UMCs. Members of the UMCs can share (upload, download) information (documentation).	Q3 2018
T4	Toolbox maintenance release	Feedback comments of end-users based on beta release is processed.	Q4 2018

Theme TTP service			
Number	Activity/Deliverable	Approach	Due date
S1	Technologies	Investigate existing and new development methods for secure data linkage solutions etc. contributing to a shared and harmonized (transition) architecture between the UMCs. Share within community meeting 2017.	Q1 2017
S2	Policy development	Create a position paper to easily share information within and outside the working group community about save and secure policy of data linkage.	Q2 2017

S3	Pseudonymisation for dummies	Create a pseudonymisation for dummies document where the policy and some practical use cases are described	Q3 2017
S4	National environment design	Create a design of the National Data linkage environment consisting of more than one solution and be part of the reference architecture (WP2)	Q4 2017
S5	National environment implementation	Support and advice the implementation of the National Data linkage environment.	H1-2018

Theme Electronic Lab Notebook (ELN)			
Number	Activity/Deliverable	Approach	Due date
L1	Knowledge sharing	Information sharing about digitization of data collections in research laboratories and the need for centralizing knowledge on this subject.	Q2 2017
L2	Research Domain	Positioning of ELN tools within the research domain/ reference architecture (collaboration with WP2 and WP4)	Q3 2017
L3	National Procurement	Investigate the opportunities for a joint procurement of purchasing a ELN tool (in collaboration with SURF)	Q2 2017
L4	National Procurement	Support the National Purchase of an ELN when decided to start a joint venture (L3); exchange functional requirements and procurement information within the working group	H2-2017/ H1-2018
L5	Standardization	Support developments in standardization of data collection and data exchange.	H1-2018

3. Time schedule

This paragraph gives an overview of the long, mid-term and short term timelines of which the deliverables as presented in paragraph two (with more detailed timelines) are part of.

3.1. Long term (2020+)

In the long term, to get to the program's True North, a large scale facility for (bio)medical (translational) research data in the Netherlands is needed (as also pitched at the KNAW-Agenda "*Grootschalige Onderzoeksfaciliteiten*").

3.2. Medium term (2018-2020)

- Merge with the broader national research infrastructure for personalized medicine & health research, Health-RI, from the perspective of secure data linkage, Electronic Lab Notebook registrations, and easily sharing Good Research Practice information between UMCs.
- Acquire structural and innovation budget(s) for sustaining and enhancing the 'Good Research Practice' Policy of the Dutch research infrastructure for biomedical research.

3.3. Short Term (2017-2018)

- Good Research Practice Policy is embedded/part in the National Data Infrastructure of the NFU by means of:
  - o The data classification toolbox is the place to collect, publish and share Good Research Practice information within and between the UMCs.
  - o Data linkage functionalities are part of the National Data Infrastructure, including an expert network.
  - o Working methods of ELN are shared between the UMCs.
  - o The opportunities for a joint procurement of an ELN tool in collaboration with SURF are investigated.

4. Budget plan

4.1. Resources/deliverables expected from or created in collaboration with other WPs

The following resources, deliverables can be drawn from, or will be (co-)created in close collaboration with other work packages within the program:

Personnel/Products contributing to the D4LS WP6 output
General management and administrative support (central/WP0)
Data stewardship guidelines (WP1) - Development of HAND containing guidelines for research data management - Collection data classification toolbox information from the UMCs
Harmonization of IT processes and architecture (WP2) - Match the position of ELN within the research domain (map on reference architecture)
Coordinate access to experts and support (WP9) - network of experts that (can/will) contribute to services in the toolbox
In-kind contributions from projects and organizations, like NFU, TraIT, Mondriaan, PSI and BBMRI-NL
SURF contributes to the joint procurement of purchasing an ELN tool.

5. Organization plan

5.1. Project team

Name	Home Institute/Project	Project Role
Regular		
Jan Jurjen Uitterdijk	UMCG, Mondriaan	WP lead
Theme Toolbox		
<t.b.d.>		
Theme Secure Data Link		
Erik Flikkenschild	LUMC	Coordinator
Irene Nooren	SURFsara	
Jan Willem Boiten	Lygature, TraIT	
Harry Pijl	UMCU, Mondriaan	
Sandor Schmikli	Julius Centrum	
Wim Schaasberg	CBS	
Lucy Overbeek	Palga	
Marike de Bruijne	CentERdata	
Daniel Theunissen	MUMC+	
Theme Electronic Lab Notebook		
<t.b.d.>		Coordinator
Diederik Kuster	VUMC	
Joost van Kempen	Radboudumc	
Wim Dijkema	UMCG	
Petra van Overveld	LUMC	
Paul van der Schaft	Erasmus MC	
Hanneke den Breeijen	Erasmus MC	

6. Communication plan

Information on the progress and results of the Data4lifesciences and the Good Research Practice work package will be actively and openly communicated within the NFU, related stakeholders/partners/projects like BBMRI-NL, TraIT, as well as the wider community. This section describes the information plan for people outside the direct scope of this work package.

6.1. WP Stakeholders

Organization	Contact person	Min. frequency	Type
D4LS	Mentors	Monthly	Personal[2]

6.2. D4LS Stakeholders

Organization	Contact person	Min. frequency	Type
NFU-Overall	Project manager	Monthly	Personal[3]
Operational team		Monthly	F2F meeting or TC

6.3. NFU Stakeholders

Organization	Contact person	Min. frequency	Type
UMCs		bi-monthly	Flash reports

---

[2] Personal = *E.g.* a face to face meeting or video- or teleconference with the contact person
[3] Personal = *E.g.* a face to face meeting or video- or teleconference with the contact person

7. Risk plan

Below a risk matrix is given which identifies the major risks as can be foreseen at this moment in the project. The risk plan is to formally review the risk list every 6 months so that new risks can be added, hazard can be re-estimated and actions be taken.

Risk description	Probability	Impact	Hazard (P*I)	Action
1. Not enough funding	6	8	48	Contingency - Redraft and/or reprioritize (parts of) project plan
2. Business case for individual UMCs too weak: local IT infrastructure projects prevail above a harmonized national IT infrastructure	6	6	36	Mitigation - Prioritize and continue with those that do support/contribute
3. Not clear how many hours can be spent and are needed per participant.				
4.				

8. References

http://datafairport.org/ and *https://www.force11.org/group/fairgroup/fairprinciples*

http://www.enrio.eu/codes-guidelines-3/national-codes (*http://www.enrio.eu/home*)

http://www.data-archive.ac.uk/create-manage/life-cycle

Aronson & Rehm, Nature, 15 October 2015, doi:10.1038/nature15816 *http://www.nature.com/nature/journal/v526/n7573/index.html#insight*

9. Glossary and abbreviations

- *BBMRI-ERIC*   Biobanking and BioMolecular Research Infrastructure Europe
- *BBMRI-NL*   Biobanking and BioMolecular Research Infrastructure Netherlands
- *CBS*   Centraal Bureau voor de Statistiek
                      (Statistics Netherlands)
- *CIT*   Centrum voor Informatie Technologie
- *CTMM*   The Center for Translational Molecular Medicine
- *CTMM-TraIT*   A Sustainable Infrastructure for Translational Biomedical Research
- *ELIXIR H2020 Excelerate* EU project on facilitate the integration of Europe's bioinformatics resources, supporting all sectors of life-science R&D
- ELN   Electronic lab notebook
- *FAIR*   Findable, Accessible, Interoperable and re-useable
                      see also *http://datafairport.org/*
- GDPR   General Data Protection Regulation
- *HANDS*   Handbook of Adequate Natural Data Stewardship (WP1 D4LS)
- *LCRDM*   Landelijk Coördinatiepunt Research Data Management
- *NFU*   Nederlandse Federatie van Universitair Medische Centra
                      (Netherlands Federation of University Medical Centres)
- SIG   Special Interest Group
- *SURF*   ICT-samenwerkingsorganisatie van het onderwijs en onderzoek
- TTP   Trusted Third Parties
- UMC   University Medical Center