European Translational Information and Knowledge Management Services

eTRIKS Deliverable report

Grant agreement no. 115446

Deliverable D7.9

Proposal for a Data Sharing Federation model for secure sharing of biomedical data for secondary research purposes

(Title in DoW: Report on feasibility and recommendations for cross-study analytics for phase 3 implementation)

Due date of deliverable: Month 36

Actual submission data: Month 50

Dissemination Level		
PU	Public	PU
PP	Restricted to other programme participants (including Commission Services)	
RE	Restricted to a group specified by the consortium (including Commission Services)	
CO	Confidential, only for members of the consortium (including Commission Services)	

DELIVERABLE INFORMATION

Project	
Project acronym:	eTRIKS
Project full title:	European Translational Information and Knowledge Management Services
Grant agreement no.:	115446
Document	
Deliverable number:	D7.9
Deliverable title:	Proposal for a Data Sharing Federation model for secure sharing of biomedical data for secondary research purposes
Deliverable version:	V1.1
Due date of deliverable:	Mth 36 - September 2015
Actual submission date:	Mth 50 – December 2016.
Leader:	Fabien Richard (CNRS)
Editors:	Fabien Richard (CNRS), David Henderson (Bayer), Neil Fitch (BioSci), Robert Irmisch (Sanofi), Chris Marshall (BioSci), Charles Auffray (CNRS)
Authors:	Fabien Richard (CNRS), David Henderson (Bayer), Neil Fitch (BioSci), Robert Irmisch (Sanofi), Chris Marshall (BioSci), Charles Auffray (CNRS) Corresponding author: Fabien Richard (frichard@eisbm.org)
Reviewers:	ESAB members: Pim de Boer, Bartha Knoppers, Emmanuelle Rial-Sebbag, Annamaria Carusi, Susanna Palkonen
Participating beneficiaries:	CNRS, BioSci
Work Package no.:	WP7
Work Package title:	Ethics for eTRIKS platform data
Work Package leaders:	Charles Auffray (CNRS), David Henderson (Bayer)
Work Package participants:	Fabien Richard and Charles Auffray (CNRS), Robert Irmisch (Sanofi), David Henderson (Bayer), Neil Fitch and Chris Marshall (BioSci).
Estimated person-months for deliverable:	4
Nature:	PP
Version:	1
Draft/Final:	Final
No of pages (including cover):	25
Keywords:	Data sharing, Data protection, Privacy, Personal Data, Re-use of data, eTRIKS

Table of Contents

Executive Summary

We are at the dawn of a deep transformation of Life Science towards an open science that requires not only radical changes in our way of sharing Data, but also an engagement of all public and private stakeholders. The document shows the challenges of a global and open Data sharing that fully addresses the data protection requirements (section 1), but also the concepts, tools and procedures that have already been developed and implemented in other economic and scientific fields, and eTRIKS recommends to use for the implementation of its proposed Data sharing model: The Data Sharing Federation (section 2). The use cases described in section 3 illustrate how the Data Sharing Federation would operate across countries. Aside the technical aspects, the main stakeholders (i.e. data users, data providers, data subjects) should comply with the good practices of Data sharing and rules of the Data Sharing Federation (section 4). Finally, eTRIKS recommends that all the stakeholders including national authorities, publishers, mathematicians, statisticians work together on several challenges that are beyond the Data Sharing Federation, but need to be addressed for fully enabling a global, open and secure Data sharing (section 5).

Inputs and outputs from related deliverables

The main deliverables inputting into D7.9 are D7.8, which established the data ethics and protection framework, D7.6, which established the data security measures, and D3.6 (standard starter pack), which recommends standards for the data curation.

Also relevant here is the recommendations made by the Ethics and Security Advisory Board whole report (D7.5), which provided guidance to the WP7 group in determining priorities, risks and a way forward in handling these matters.

The output is beyond the time life of eTRIKS.

The release of the deliverable D7.9 has been delayed due to a significant reduction of resource during Period 3 and Period 4, and the prioritisation of other deliverables following an extensive consultation process. The learning and development from other deliverables have also been valuable in building an understanding of how the operability of the model can be realised.

Proposal for a Data Sharing Federation model for secure sharing of biomedical data for secondary research purposes

(version 1, December 13 2016)

Authors:
Fabien Richard, David Henderson, Robert Irmisch, Neil Fitch, Chris Marshall, Charles Auffray.

Corresponding author:
Fabien Richard (frichard@eisbm.org)

Acknowledgements:
The authors would like to thank the ESAB members Pim de Boer, Bartha Maria Knoppers, Emmanuelle Rial-Sebbag, Susanna Palkonen, Annamaria Carusi for their very helpful inputs.

Restrictions of the document release:
Only eTRIKS partners can use this document. **Sharing it with a third party requires a written authorisation of the eTRIKS WP7 team "Data Privacy and Security"**.

Aim of the document

The present document proposes a data sharing model that 1) circumvents the roadblocks that prevent or slow down the re-use of personal *data concerning health* and *genetic data* (hereafter called 'Data'), 2) safeguards the *data subject*[1]'s rights, and 3) maintains the level of data protection required by the European General Data Protection Regulation (GDPR)[2].

Scope of the document:

The scope of the present document covers only the Data sharing for archiving or secondary scientific research purposes, which means that the Data that are shared and re-used for archiving or secondary scientific research purposes have been already collected and processed lawfully (i.e. after data subjects have given their informed consent) for the purposes of the original project (i.e. the primary purposes). This document focuses on the Data that have been collected from European citizens and/or are processed by European *controllers* and/or *processors*. The challenges of Data sharing as well as the proposals to address them are based on the requirements and constraints set by the European GDPR that will replace the European Data Protection Directive on May 25th, 2018. The technical and operational measures related to the proposed data sharing model follow the principles of data privacy by design or by default, as required by the GDPR, as well as the FAIR principles[3].
This document does not cover:

- The data processing for the original project, as described in the data subject's informed consent
- The secondary use of samples
- Technical details of the data sharing model
- Any measures improving data interoperability and comparability, although these are acknowledged to be a pre-requisite for an effective sharing and re-use of health data.

[1] A data subject is a natural person who participates in a study and provides a controller(s) with personal data for the purposes of the said study set by the said controller(s).

[2] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL%3A2016%3A119%3ATOC

[3] Data are FAIR when they are Findable, Available, Interoperable, Re-usable. For more details, see Wilkinson MD et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18.

Definitions:
Unless specified in the notes of this document, the *italic* terms are defined in the Article 4 of the GDPR.

1. Challenges of sharing Data for secondary and scientific research purposes

1.1. Harmonisation of the data protection in the European Union (EU)

The GDPR advances the harmonisation of the requirements for the protection of personal data (hereafter called 'data protection' or 'DP') between the EU member states (hereafter called 'member states'), and considers the Data (i.e. health and genetic data) as a special category of data (also called 'sensitive data'). However, the DP requirements remain to be harmonised for the following points:

- The GDPR allows member states to provide for derogation of the data subject's rights referred to in Articles 15, 16, 18 and 21, if Data are processed for scientific research purposes and such rights are likely to render impossible or seriously impair the achievement of those purposes (Article 89 (2));
- The GDPR allows member states to maintain or introduce further conditions, including limitations, when processing sensitive data e.g. the Data (Article 9 (4)), which may hamper the *cross-border processing* within EU (despite the GDPR recommendations (Recital 53));
- The GDPR does not cover Data of a deceased person, and lets the member states decide whether and how to protect those Data (Recital 27).

Moreover, the GDPR does not rule on the special status of genetic data. The case of Henrietta Lacks (the individual from whom the widely-used HeLa cell line was derived) has raised an important privacy issue: the release of genetic data of a living or deceased person may violate the privacy of their family because of the potential to identify relatives and/or reveal some of their personal information (e.g. a disease risk) without their consent[4].

1.2. The anonymisation issue

Since 2000 and now in the era of big data, the amount and the diversity of data that are collected for a data subject have enormously increased, which makes or strengthens the uniqueness of the subject's dataset. This trend keeps growing with the use of cheaper technologies (cost of a human genome sequencing reaching less than 1000 US dollars) and mobile health-monitoring devices. Currently, as shown in several reviews, a genomic microarray or sequencing experiment provides information on millions of genetic variants per data subject, while only 30-80 of those variants are enough for re-identifying a data subject[5]. Similarly, the dates and locations of four purchases were enough to identify 90 percent of the people in a data set recording three months of credit-card transactions by 1.1

[4] https://www.nih.gov/news-events/news-releases/nih-lacks-family-reach-understanding-share-genomic-data-hela-cells

[5] Lin, Z.; Owen, A.B.; Altman, R.B. Genomic research and human subject privacy. Science 2004, 305, 183; McGuire, A.L.; Gibbs, R.A. No longer de-identified. Science 2006, 312, 370–371

million users[6]. Moreover, the current de-identification[7] methods are not often suitable for the health data mainly for three reasons:

- As explained above, the data diversity hampers or prevents effective anonymisation procedures;
- Some qualitative data cannot be "blurred" (e.g. female and male);
- Perturbation methods are available to add noise to experimental results that are already, by nature, noisy data, which may reduce the utility of those results.

Finally, the personalised medicine requires to link back to data subjects in order to inform them (e.g. increased susceptibility to a disease) or propose them adapted treatments.

Even if complete anonymisation of Data may not be possible, de-identification methods should be applied to data such as geographic location, age and all recorded dates in order to reduce the risk of re-identification, as required by the GDPR.

1.3. Access to health and/or genetic data

There are currently two main ways to access existing Data: through public access or through controlled access. As explained below, neither method is suitable for a broad, efficient and safe Data sharing. Recently, the Global Alliance for Genomic Health (GA4GH) has proposed the use of "registered data access" as an intermediate tier between open and controlled access[8].

1.3.1. Public access

Whether most data subjects agree on sharing openly their Data remains debated[9,10]. However, Robinson *et al* have shown that 54% of participants who agree on sharing openly their genetic data could not initially recall with whom they had agreed to share their genetic data or did not understand that open access data sharing meant that their genetic data could be accessed and used by anyone on the internet without restriction[11]. Robinson *et al* have suggested that educational training should be provided to data subjects before they authorise an open sharing of their Data, which is one of the requests of several patient advocacy groups (personal communications at the eTRIKS Brussels meeting on Oct 20th 2016). Although needed, increasing the data subjects' awareness on the risk of a public Data access will require time and resource, which may hamper efficient, open data sharing.

[6] Montjoye Y.-A., Radaelli L., Singh V. K., Pentland A. S., Unique in the shopping mall: On the reidentifiability of credit card metadata. Science 347 (6221), 536-539. DOI:10.1126/science.1256297 (2015)

[7] 'Data de-identification' means a process of rendering data pseudonymous or anonymous. De-identified data are pseudonymised or anonymised data.

'Anonymisation' means a process of removing all elements enabling the identification of an individual person (i.e., of rendering data anonymous). Anonymised data are not personal data.

'pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person (Article 4 (5) of the GDPR). Pseudonymised data are personal data.

[8] Dyke S.O.M., Kirby E., Shabani M., Thorogood A., Kato K. and Knoppers B.M. Registered access: a 'Triple-A' approach. Eur J Hum Genet. 2016 Sep 28. doi: 10.1038/ejhg.2016.115

[9] Robinson, J.O.; Slashinski, M.J.; Wang, T.; Hilsenbeck, S.G.; McGuire, A.L. Participants' recall and understanding of genomic research and large-scale data sharing. J. Empir. Res. Hum. Res. Ethics 2013, 8, 42–52

[10] http://www.dailymail.co.uk/health/article-3838493/Trust-doctor-don-t-people-worry-NHS-share-personal-data.html

[11] Robinson, J.O.; Slashinski, M.J.; Wang, T.; Hilsenbeck, S.G.; McGuire, A.L. Participants' recall and understanding of genomic research and large-scale data sharing. J. Empir. Res. Hum. Res. Ethics 2013, 8, 42–52

Moreover, public sharing of genetic data raises an issue with regard to privacy of the data subject's family.

1.3.2. Controlled access

Data access is controlled by a Data Access Committee (DAC) who may request approval by an Institutional Review Board (IRB), also known as an Independent (research) Ethics Committee (IEC), in some limited circumstances. Moreover, the requester must agree to a Data Use Agreement (DUA). This data sharing model has several drawbacks:

- Data access is restricted to a few users
- Obtaining data access is time and resource consuming
- Even if data access and transfer are well controlled at the legal level, the data sharing model relies on a trusted network. In other words, the IRB/IEC, the DAC and the controllers have no technical means to monitor or enforce protection once a requester has been granted access to individual data and/or received them in their server. They cannot ensure that:
    - The requester will respect the data subject's rights to objection, rectification or erasure (assuming the data are personal);
    - Someone (i.e. a malicious authorised user) will not re-identify and/or disclose individual data;
- Although the risks are reduced, malicious re-identification using aggregated data remains a risk.
- The risk of data breach for a given dataset statistically increases with the number of users who are granted access to the said dataset.

1.4. Privacy notices

When Data are processed for scientific research or archiving purposes, the GDPR requires that the controllers provide data subjects with an understandable and accessible privacy notice that informs the data subjects about who processes their Data, how, and why they are processed (Articles 12-14). Moreover, the GDPR addresses patient advocacy groups' needs and broadens the data subject's rights: they have the rights to Data access, rectification, erasure ("to be forgotten"), portability, to restriction of processing, and to object (Articles 15-21). Even if controllers are allowed to not inform data subjects when Data are not obtained directly from data subject and the provision of such information renders impossible or seriously impair the achievement of the objectives of that processing (Articles 14 (5 (b)) and 89 (2) of the GDPR), this waiver will be soon invalid because of the hyper-connectivity of data subjects through internet or applications installed on their mobile phones, and their growing desire to be engaged in research by dynamically controlling the use of their Data.

1.5. Data subjects' engagement

An increasing number of data subjects want to be actively engaged in research by 1) being informed about the use of their Data and 2) giving or changing their authorisation of using their Data. Interestingly, 75-85% of data subjects are willing to share their Data if their authorisation is asked[12,13], while only 28% of them are willing to share their Data if their authorisation is not asked[18]. It seems that a such dynamic approach ('data access on demand') will be a way of enabling an easier and more transparent communication between stakeholders. Associated with a data user[14]s' commitment on communicating the results of their Data processes to data subjects, this approach will contribute to reward the data subjects' engagement, which may result in stimulating the data subjects' engagement for a better and broader Data sharing[15]. That is why the project EnCoRe[16] was initiated, and several secure platforms/dashboards such as the platform Reg4All[17] or the Platform for Engaging Everyone Responsibility (PEER)[18] have been implemented. In the same spirit, the National Health Service Blood and Transplant has recently launched a new initiative in England in order to encourage blood donation: blood donors receive a text on their mobile phone when their blood is used for saving lives[19]. The number and the size of these platforms are foreseen to increase, since 1) the Article 20 of the GDPR gives data subjects the rights of data portability (i.e. data subjects have the right to request a copy of their Data and use them as they want), 2) the Articles 12-14 of the GDPR require that a data user (defined as a controller in the GDPR) informs data subjects regarding processing of their Data, and 3) the increasing connectivity of data subjects will invalidate the information requirement waiver (for details, see section 1.4).

1.6. Mindset about data sharing

Many data providers are still reluctant to data sharing despite increasing pressure from regulatory authorities, government agencies, some publishers, funders, patient advocacy groups, and a part of the scientific community for open Science. Several reasons explain their reluctance:

- Sharing 'their' data puts at risk their own research: they could be scooped by other groups, which may result in losing the exclusivity of their publications and ultimately funding sources[20,21];
- Sharing data is a burden. Making data interoperable and re-usable according to the FAIR principles requires investment of significant resources, time and money;
- Sharing data is not yet well recognised and rewarded, and few journals such as 'Scientific Data' enable scientists to publish only clean and standardised data;

[12] Presentation of R. Sheldon at the GAPP conference. http://med.stanford.edu/gapp/events/gapp-conference-2016-videos.html

[13] Tarini BA1, Goldenberg A, Singer D, Clark SJ, Butchart A, Davis MM. Not without my permission: parents' willingness to permit use of newborn screening samples for research. Public Health Genomics. 2010;13(3):125-30. doi: 10.1159/000228724. Epub 2009 Jul 11.

[14] A data user is a natural person who defines the secondary and scientific purposes of a process (i.e. s/he is a *controller*) and processes Data for those purposes. Data users are mainly scientists or clinicians.

[15] Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Genet. 2015 Feb;23(2):141-6. doi: 10.1038/ejhg.2014.71. Epub 2014 May 7.

[16] http://www.hpl.hp.com/breweb/encoreproject/about.html

[17] https://www.reg4all.org/more.php

[18] https://www.peerplatform.org/portal/

[19] https://www.blood.co.uk/news-and-campaigns/news-and-statements/blood-donors-texted-when-their-blood-goes-to-hospitals-to-save-lives/

[20] Longo DL & Drazen, JM. Data Sharing. N Engl J Med 2016; 374:276-277January 21, 2016DOI: 10.1056/NEJMe1516564. *http://www.nejm.org/doi/full/10.1056/NEJMe1516564#t=article*

[21] Why don't scientists always share their data? http://blogs.nature.com/naturejobs/2016/10/21/why-dont-scientists-always-share-their-data/#comment-4553

- Sharing data with a third party may be a security issue. Data providers are responsible for assessing that the third party will process the shared data in compliance with the requirements of the national data protection laws. They, however, often do not have the means and resources to do such an assessment, and even if the third party is legally responsible for processing the shared data, a data breach will negatively impact their reputation.

Reducing the risks of data sharing and increasing its benefits will certainly foster a shift in the mindset of data providers towards a more open data sharing.

2. Recommended Data sharing model

In order to 1) address the requirements of the GDPR and the national data protection laws, 2) circumvent the current blocking factors that hamper broad Data sharing, 3) provide technical means for a more controllable data protection, and 4) increase the engagement of data subjects in research, eTRIKS recommends the implementation of a federated data sharing system (hereafter called 'the Data Sharing Federation') that is built on the **TRUST principles**:

- **T**ransparency. Data subjects are informed of data users' requests if they wish (Articles 13 and 14 of the GDPR), and data breaches when required (Article 33 of the GDPR);
- **R**eciprocity and reward. The contribution of stakeholders (data subjects, data providers[22], and data users) is acknowledged or rewarded in a study;
- **U**niversality. The use of Data is open to any registered data users if that use is authorised by a national law and/or a data subject;
- **S**ecurity. Data are processed in a controlled environment. Data users and their requested processes are recorded for auditing purposes;
- **T**iered data use. The authorisation of data use depends on the data type, the analysis purpose, the data user's profile, the analytical algorithm that a data user wants to use, and the data subject's will.

Similar federated systems have been successfully implemented. For instance, BioMart[23] enables biologists to access, visualise and analyse public data stored in several connected databases through a unique web portal. The International Cancer Genome Consortium also uses BioMart in a controlled environment in order to facilitate data sharing in the cancer research community[24]. Recently, the consortium GA4GH has proposed to also adopt a federated system approach[25]. However, all the current data sharing environments with

[22] A data provider is a natural or legal person who collects and/or holds Data of a data subject(s) under the informed consent(s) of the said data subject(s). Data providers can be hospitals, research laboratories of academic organisations, pharmaceutical companies, public repositories, or data subject's organisations (not exclusive list).

[23] Guberman J.M. et al. BioMart Central Portal: an open database network for the biological community. Database, Vol. 2011, Article ID bar041, doi:10.1093/database/bar041 (http://database.oxfordjournals.org/content/2011/bar041.full.pdf+html)

[24] The International Cancer Genome Consortium. International network of cancer genome projects. Nature Vol 464j15 April 2010jdoi:10.1038/nature08987

[25] Global Alliance for Genomics and Health. A federated ecosystem for sharing genomic, clinical data. Science. 2016 Jun 10;352(6291):1278-80. doi: 10.1126/science.aaf6162

public, registered[26], or controlled data access rely on trusted users and organisations. Moreover, verifying that the all data users are bona fide researchers or clinicians (assuming the criteria required to demonstrate a bona fide researcher or clinician are defined) does not exclude that a bona fide user discloses sensitive data by mistake, or the personnel of an organisation neglects security measures. Finally, the risk of data breach increases with the number of data users who are granted access to data. To our knowledge, the Data Sharing Federation model we describe here is the only data sharing model that:

1) enables an open Data sharing while guaranteeing a high level of data privacy and security,

2) addresses the above objectives,

3) follows the above principles, and

4) leverages on the following and existing solutions:

- Virtual data enclaves[27] such as the ones provided by organisations such as the U.S. Census Bureau, the federal statistical research centres[28], or Health Care Cost Institute[29] for analysing data in economy, demography, or health care. In its data sharing policy and implementation guidance[30], the National Institute of Health (NIH) **“***recognizes that the sharing of data from clinical trials and under other situations may require making the data anonymous or sharing under more controlled means, as through a restricted access data enclave. Sharing though data enclaves would grant access only to researchers who agree to preserve the privacy of subjects and provide means to protect the confidentiality of the data.*”
- Privacy-preserving analytical and visualisation tools and procedures;
- Web 2.0 tools that enable a dynamic and higher level engagement of data subjects in research.

2.1. Data Sharing Federation

2.1.1. Definition

The Data Sharing Federation is a community of data users, providers, brokers and subjects who work together on enabling Data sharing across countries in a privacy-preserving manner and in compliance with the requirements of European and national data protection laws.

2.1.2. The structure

As shown in Figure 1, the Data Sharing Federation enables sharing Data among the data subjects, data providers and data users across countries, and has six types of components: 1) the web portal of the Data Sharing Federation (hereafter called ‘DSF portal’), 2) servers of data providers, 3) servers of data brokers[31], 4) platforms of data subjects, 5) computers of

[26] Dyke S.O.M., Kirby E., Shabani M., Thorogood A., Kato K. and Knoppers B.M. Registered access: a ‘Triple-A’ approach. Eur J Hum Genet. 2016 Sep 28. doi: 10.1038/ejhg.2016.115

[27] A data enclave is a controlled, secure environment in which eligible individuals can perform analyses using restricted data resources (NIH definition: http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#enclave)

[28] https://www.census.gov/about/adrm/fsrdc/locations.html

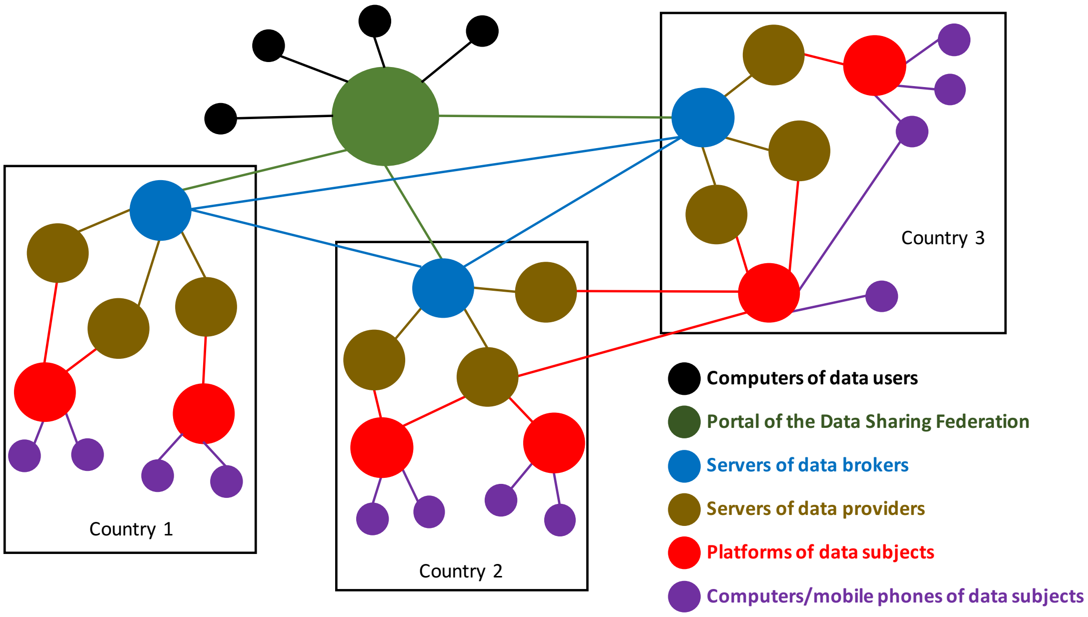
[29] http://www.healthcostinstitute.org/files/About%20the%20HCCI%20Data%20Enclave%20fact%20sheet%20-%20February%202015.pdf

[30] https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

[31] A data broker is a legal person who is responsible for providing an environment compliant with the requirements of national data protection laws and the Data Sharing Federation where data users can process Data for secondary and scientific purposes in a privacy-

data users, and 6) computers or mobile phones of data subjects. Each country has its own federated system that controls the data use requests of the data users based in the said country, and the use of data that have been hosted in the said country. Each federated system shares Data with the other federated systems only through its data broker (blue lines).

Figure 1. The overall structure of the Data Sharing Federation:


Country 1
Country 2
Country 3
Computers of data users
Portal of the Data Sharing Federation
Servers of data brokers
Servers of data providers
Platforms of data subjects
Computers/mobile phones of data subjects

This overall structure enables the following data flow across countries:

- A data user registers and connects to the Data Sharing Federation only though its web portal (black lines);
- The DSF portal redirects the data user on a data broker of their country (hereafter called ‘the master data broker’) (green lines);
- The master data broker sends the data user’s request to all the data providers directly (brown lines) or through data brokers of other countries (blue lines);
- The data providers inform the platforms of data subjects what Data are going to be processed (red lines);
- The platforms of data subjects authorise (or not) data providers the Data processing (red lines) based on the consent information they receive from the data subjects (purple lines);
- If authorised by data subjects, the master data broker receives data from the data providers of their country (brown lines) and/or the data brokers of other countries (blue lines);
- The data users analyse and visualise the received data by using the computing power of the master data broker’s servers. The parameters and result of an analysis are recorded in an analysis report file that is stored in the master data broker’s servers;

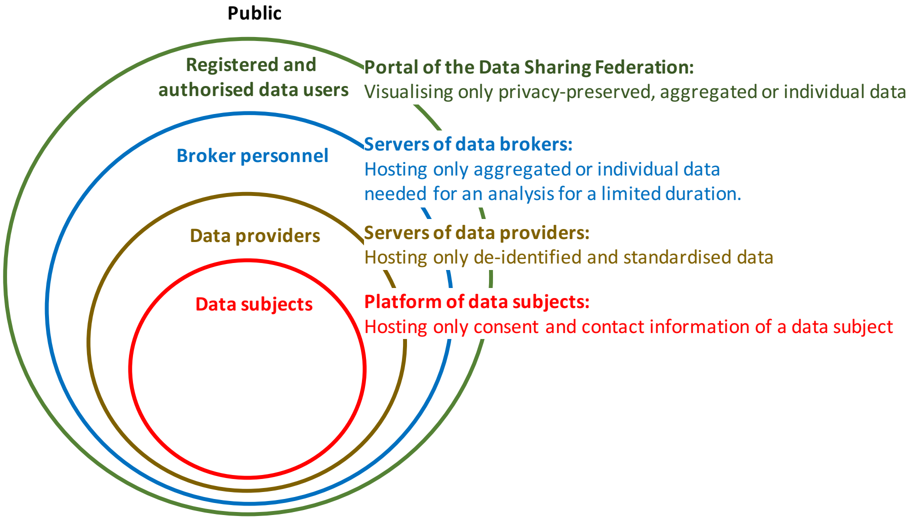
preserving manner and in compliance with the requirements of European and national data protection laws. Data brokers are usually not-for-profit organisations that have a strong IT infrastructure and expertise.

- Once the results of the data user's analysis are published in a peer–reviewed journal,
  - the analysis report file and the publication references are stored in a server of the Data Sharing Federation, and are available through the DSF portal, and
  - a brief report of the analysis results written by the data user in a lay English is sent to the data subjects who wanted to be informed about the positive outcomes of the use of their Data.

This overall structure has several advantages:

- It enables queries and analyses across countries in compliance with the GDPR and national data protection laws;
- The information flow is not broken between the data users and the data subjects, which allows Data sharing to be transparent, fast, dynamic, easy and rewarding;
- Each component of a national federated system is autonomous for the management of its server(s), and accountable for its compliance with the legal data protection requirements of its country;
- Data providers are connected to one data broker of their national federated system. Thus, a data provider does not need any more to sign data use or transfer agreements with many data users, and can focus only on auditing the data broker's security measures that must comply with requirements of the law(s) and/or authority of its country, which provides a higher security level;
- Data are compartmented in each component of a national federated system and Data sharing between components is minimised. Moreover, the components of a national federated system form concentric and independent security shields[32]. The data compartmentation, data minimisation, and the independent security shields reduce together the risk of a data breach to its minimal level (Figure 2).

Figure 2. Data compartmentation and security shields.


Public
Registered and authorised data users
Portal of the Data Sharing Federation:
Visualising only privacy-preserved, aggregated or individual data
Broker personnel
Servers of data brokers:
Hosting only aggregated or individual data needed for an analysis for a limited duration.
Data providers
Servers of data providers:
Hosting only de-identified and standardised data
Data subjects
Platform of data subjects:
Hosting only consent and contact information of a data subject

'public' means here any persons who are not registered in the Data Sharing Federation servers.

[32] A security shield means a collection of operational and technical measures of security.

It also implies that:

- The data brokers act as data hubs, and their servers have the same features, architecture and level of security across countries;
- The authorities of the countries that have federated systems agree on transferring Data between data brokers when authorised by data subjects;
- All the stakeholders comply with the requirements of the Data Sharing Federation (for details, see section 4).

2.1.3. the characteristics

The Data Sharing Federation has the following characteristics:

- It provides data users with a unique web portal (i.e. the DSF portal);
- It allows a registered and authorised user to query and analyse Data across countries by using common privacy-preserving tools hosted in the data brokers' servers;
- It prohibits any data users to directly access to, copy, or export individual and de-identified data;
- It prohibits the visualisation of individual and de-identified data that are not needed for an analysis (data minimisation, Article 89 (1) of the GDPR);
- It automatically manages the data subject's authorisation of data use without the approval of an ethical or data access committee, since:
    - Data are protected as described in the above bullets (for details, see section 3);
    - Storage and secondary use of Data for archiving and scientific purposes is compatible with any initial purposes set in a consent form (Article 5 (1 (b)) and (1 (e)) of the GDPR);
    - Data subjects may object to a Data use when notifying them is possible.

2.2. The components of a national federated system

A national federated system has four types of components: 1) the data providers' servers, 2) one data broker's server, 3) the data users' computers, and 4) the data subjects' platform(s).

2.2.1. The data users' computers

Data users who could be not informatics-savvy researchers should be able to use Data with no or a minimal installation of any specific software in their computers. Only an internet connection and browser are required to connect to the DSF portal.

2.2.2. The Data Sharing Federation web portal (the DSF portal)

The DSF portal is the unique entry point to the Data Sharing Federation for anyone who wants to register, access to public information such as the data users' publications and privacy-preserving analytical and visualisation tools, submit a new analytical and visualisation tool, design data queries and analyses, or visualise the query or analysis results. The server of the DSF portal hosts only data users' credentials and public information. Very importantly, it does not host analysis-needed Data, analysis results and data users' profiles, or provide computing power for analyses. Those features are provided only by the master data brokers' servers.

2.2.3. The data providers' servers

The server(s) of a data provider has the following characteristics:

- It is dedicated to the national federated system and is isolated from other servers of the data provider;
- Its connections to the data broker's server and the data subjects' platform(s) are encrypted;
- It hosts a database that contains only pseudonymised and/or anonymised data that are standardised and written in English, and for which data subjects have authorised their processes/uses for secondary and scientific purposes;
- Each pseudonymised dataset has one randomly generated key that links this dataset to a data subject, and is unique in the data provider's database;
- The data model of the database is the same as or mapped to the common data model of the Data Sharing Federation. This concept of a common model is required for enabling data queries across countries, and has been successfully implemented by other multi-party organisations such as the ICGC[33], the Shared Health Research Information NEtwork (SHRINE)[34], or the Observational Medical Outcomes Partnership (OMOP)[35];
- It performs tasks that require a moderate computing power.

2.2.4. The data broker's server

The data broker's server provides data users with a virtual data enclave in which:

- Registered data users can query Data in the whole Data Sharing Federation
- Authorized data users can analyse and visualise data only with the approved, privacy-preserving tools that are hosted in the data broker's server. However, registered users are allowed to submit their own tools to the Data Sharing Federation (for details, see section 3.5). The analyses are performed by the high-performance computing machines of the data broker;
- Authorised data users can export privacy-preserving reports of their analyses;
- Registered data users have their user space for recording the parameters of their queries and analyses;
- Security is the 1st rule:
    - No data users can copy or export individual data (i.e. no download of individual data into the data users' severs or computers). If data users want to analyse their Data with other's ones, they must import and record them in a database of a data provider based in their country (for details, see section 3.3);

[33] The International Cancer Genome Consortium. International network of cancer genome projects. Nature Vol 464j15 April 2010jdoi:10.1038/nature08987

[34] https://open.med.harvard.edu/wiki/display/SHRINE/High-Level+Architecture

[35] http://omop.org/

    - No data users can import or modify analytical or visualisation tools without the approval of the Data Sharing Federation. They may be authorised to change some parameters. If data users want to use a new tool, they must submit it on the DSF portal (for details, see section 3.5);
    - All the requested analyses are recorded (e.g. data users' IDs, analysis parameters, data sources, dates) for audits done by the national authorities;
    - Only analytical and visualisation tools approved by the Data Sharing Federation are installed in the servers of the data brokers;
    - Individual Data stored in the server of the data broker may be encrypted if the encryption does not reduce significantly the analysis performance;
    - The connections to other data brokers' servers, the data providers' servers and the data users' computers are encrypted.
- Data minimisation is the 2nd rule:
    - When registered data users query Data, the data broker only informs them the presence or absence of the queried Data (for details, see section 3.2);
    - When authorised data users analyse Data, the data broker provides them with only the Data needed for the analysis (for details, see section 3.4).
- Proportionality is the 3rd rule: the data broker stores the Data needed for the analysis only for the time of the authorised data user's connection or the time of the analysis if the latter last several hours or is in a queue. After that time, the Data used for the analyse are erased in the data broker's server;
- Respect is the 4th rule: the data broker allows a data user to analyse Data only if this analysis is authorised by the data subjects (for details, see section 3.4) and/or a national law(s);
- Transparency is the 5th rule: the data broker informs the data subjects about who uses their Data and for what purposes if 1) it is possible to inform them, and 2) the data subjects want to be informed.

Data enclaves hosted by federal statistical agencies have not led to any known security breaches up to now.

2.2.5. The data subjects' platforms

As explained in section 1.5, an increasing number of data subjects want to be asked to be engaged in research by 1) being informed of the use of their Data, and 2) giving or changing their authorisation of using their Data. To this, they register to data subject's platform (also called dash boards) such as the platform Reg4All[36] or the Platform for Engaging Everyone Responsibility (PEER)[37]. In a national federated system, similar platforms need to be put in place. Two cases are possible:

- The platform stores the data subjects' Data as well as their consent information. Thus, it is considered as a data provider, it is securely connected to the server of one data broker, and should follow the technical requirements listed in the section 2.2.1;

[36] https://www.reg4all.org/more.php

[37] https://www.peerplatform.org/portal/

- The platform stores only the consent data of data subjects. Thus, it has to be securely connected to the server(s) of a data provider(s) where the Data of data subjects who are registered in the platform are stored.

Those platforms host personal data and, thereby, must comply with the national data protection laws and the GDPR.

They also should contribute to increase the awareness of data subject regarding Data sharing, and provide information that explain the benefit of sharing Data for research and, ultimately, data subjects, its risks, and the data subjects' rights and duties. Based on those information, data subjects would take an informed decision on sharing their Data for secondary and scientific purposes.

They also should play an important and active role in rewarding and strengthening data subjects' engagements: They should translate the data users' reports on the outcomes of the Data analyses, and individually send them to the data subjects who are associated with those outcomes and want to be informed on the benefit of the use of their Data.

Finally, those platforms should provide data subjects with a secure and easy access to their Data through mobile applications and/or internet.

3. Use cases

3.1. Registration

Data users must register their profile before being allowed to use Data. To do this, they first have to enter their information on the DSF portal (e.g. their name, the name of the organisation they belong to, their function in the said organisation, the country of their residency, their professional or personal address, phone number, email address, their ORCID if they are publishing researchers). Secondly, they have to agree on the terms and conditions of data use that are indicated in a concise way and in lay language:

- The data user is a controller according to the GDPR, and has to comply with the requirements of the GDPR and the data protection law of his/her country;
- The data user does not attempt to obtain individual data from the Data Sharing Federation without authorisation;
- The data user does not disclose any individual data obtained from the Data Sharing Federation without authorisation;
- The data user never attempts to re-identify data subjects;
- The data providers are named as co-authors or collaborators in the data user's publication(s) that show the result of an analysis performed on the data obtained from the said data providers (if the data providers agree).

Thirdly, data users are asked to activate their account in order to receive a username and password, and finalise their registration. Once their accounts are activated, a master data broker is associated with their account. The selection of the master data broker depends on the location of the data user's organisation: The closest data broker within the data user's country is the master data broker for the said data user. The master data broker hosts analysis-needed Data, analysis results and data users' profiles, or provides computing power for data user-requested analyses.

The registration is free and open to anyone agreeing with the rules of the Data sharing federation.

3.2. Data query

A Data query is defined as checking the availability of Data. The Data query is automated, and its steps are the following:

1. A data user designs a Data query on the DSF portal;
2. The master data broker sends the Data query to all the data providers directly and through the other data brokers (hereafter called 'the requested data broker');
3. The master data broker collects the answers sent by the data providers and/or the requested data brokers, and calculates the number of data subjects for which Data are available;
4. The master data broker gives the data user the following information that are visualised on the DSF portal: 1) a "yes, they are available" or "no, they are not available", 2) a range number of data subjects e.g. "1-50", "51-100", "101-300", "> 300". The exact number of data subjects and Data provenance are not indicated for privacy reasons.

At that stage, there is no dissemination of Data. Therefore, registered data users can perform any Data queries without informing data subjects. The GA4GH project 'Beacon' follows a similar procedure[38]. The queries are free, and registered data users can save the query parameters in their user space.

3.3. Data submission

Registered data users may want to analyse other's Data together with Data they have collected. Thus, they first have to submit "their" Data. If their organisation is not a data provider of a national federated system, they have to submit the Data to a data provider who is based in their country and belong to a national federated system. If their organisation is a data provider of a national federated system, they submit the Data in their organisation. In both cases the Data submission must comply with the requirements of the Data Sharing Federation (for details, see section 4.1) and be approved by the national data broker.

3.4. Data analysis and visualisation

In the following sections:

- All the data flow steps are automated. Unless specified, there is no human intervention in the below processes;
- An analysis is run in a national federated system for the sake of simplicity. Nevertheless, the approach remains the same when an analysis is run across countries: the requesting data broker receives either individual or aggregated data from the requested data brokers after the data providers have checked that data analysis is compatible with the data subject's consents (when it is required);
- The data brokers provide data users with a same library of approved, privacy-preserving analytical tests that prevent the re-identification of subjects from

[38] http://beacon-network.org/#

individual or aggregated data by using transformation or differential privacy algorithms. The latter add a small noise in the analysis result without reducing significantly the data utility[39].The concept of using a tool library such as the ones in R[40], Galaxy[41], Bioconductor[42]for data analysis is already broadly accepted by bio-informaticians, bio-statisticians and biologists. However, those tools need to be modified in order to become privacy-preserving analytical tools (for details, see section 3.5);

- The privacy-preserving analytical algorithms are moved from a data broker to the data providers' servers when the computation can be parallelised and is not too computation-demanding. The mobile computation is commonly implemented in other domains, and prevents transferring individual data and increases the analysis performance and the data security[43];
- The data brokers provide data users with a same library of approved, privacy-preserving visualisation tools that prevent the re-identification of subjects from individual data (e.g. no scales are shown in figures or histograms);
- The data broker and the data providers respectively record the analysis parameters and data sources needed for the analysis for auditing and data provenance. The latter allows a data user(s) to cite the data contributors[44] who have provided, de-identified, and curated[45] the data needed for the analysis; and
- Data subjects want to be informed when their Data are processed.

The following cases are ranked form the lowest to highest re-identification risk, which depends on the user's choice regarding:

- The data selected for the analysis;
- The privacy-preserving analytical test; and
- The visualisation of the results.

That is why, before starting an analysis, the registered data user must provide the following information in lay terms and English language:

- The analysis-related therapeutic area;
- The purpose of analysis.

3.4.1. Case 1: analysis of aggregated, anonymised data

1. Set the analysis. The data user sets the selection of Data and chooses a privacy-preserving analytical test in the data broker.

[39] Cynthia Dwork's presentation at the GAPP conference, March 2016: http://med.stanford.edu/gapp/events/gapp-conference-2016-videos.html

[40] https://www.r-project.org/

[41] https://galaxyproject.org/

[42] https://www.bioconductor.org/

[43] Dave Maher's presentation at the GAPP conference, March 2016: http://med.stanford.edu/gapp/events/gapp-conference-2016-videos.html

[44] 'Data contributor' means a natural person who collects, produces, de-identifies, curates, and/or manages data.

[45] 'Data curation' means a process of cleaning (i.e. removing inconsistency and misspelling, completing), transforming (e.g. converting numerical values with international units), and standardizing data (i.e. reporting, vocabulary and format standards). Curated data are data that have been through the curation process.

2. Selected data. The data broker sends the data selection parameters, the analysis-related therapeutic area, and a unique analysis identification number to all the data providers. The data selected from a given study dataset do not allow to single-out a data subject in the said study dataset, and thereby are considered as anonymised data.
3. Authorisation: not required. Each data provider's server that contains the data needed for the analysis in a given study dataset records the provenance of the selected data by associating the data broker-sent information with the ID # of the said study dataset and the pseudonymisation keys.
4. Analytical test. The privacy-preserving analytical algorithm is run on the selected data in the servers of data providers and/or the data broker, and the analysis result is aggregated data.
5. Visualisation and report. The data broker displays and reports only the analysis result and an approximate number of data subjects who have been selected for the analysis. The data provenance and the exact number of data subjects are indicated only if the data user is publishing the analysis result and the number of data subjects is greater than 20.

- Information to data subjects. The data providers who provided data for the analysis inform data subjects the processing of their anonymised data.
- Re-identification risk: almost zero.

3.4.2. Case 2: analysis of aggregated, personal data

1. Set the analysis. Same procedure as the one described in case 1.
2. Selected data. The data broker sends the data selection parameters and a unique analysis identification number to all the data providers. The data selected from a given study dataset allow to single-out a data subject in the said study dataset, and thereby are considered as personal data.
3. Authorisation: may be required in some countries. The provider's server that contains the data needed for the analysis in a given study dataset, requests the analysis purpose, its scope, and the data user's profile from the data broker, and assesses whether the requested information is compatible with the data subject's consent recorded in the platform of data subjects. The data providers send the data broker the numbers of data subjects for which the analysis is authorised and those for which the authorisation needs to be asked. The data broker displays an approximation of those numbers. If the data user decides to analyse more data and wait for the replies of the data subjects for which the authorisation needs to be asked, then the data providers send an authorisation request to those data subjects. Each data provider who has found selected data in a given study dataset for the authorised analysis records the provenance of the selected data by associating the data broker-sent information with the ID # of the said study dataset and the pseudonymisation keys.
4. Analytical test. Same procedure as the one described in case 1.
5. Visualisation and report. Same procedure as the one described in case 1.

- Information to data subjects. The data providers who provided data for the analysis inform data subjects the processing of their personal data.
- Re-identification risk: extremely low.

3.4.3. Case 3: analysis of individual, anonymised data

1. Set the analysis. Same procedure as the one described in case 1.
2. Selected data. Same procedure as the one described in case 1.
3. Authorisation: Same procedure as the one described in case 1.
4. Analytical test. The privacy-preserving analytical algorithm is run on the selected data in the server of the data broker, and the analysis result is individual data.
5. Visualisation and report. The data broker displays and reports only transformed, individual data and an approximate number of data subjects who have been selected for the analysis. The data provenance and the exact number of data subjects are indicated only if the data user is publishing the analysis result, and the number of data subjects is greater than 20.

- Information to data subjects. Same procedure as the one described in case 1.
- Re-identification risk: very low.

3.4.4. Case 3: analysis of individual, personal data

1. Set the analysis. Same procedure as the one described in case 1.
2. Selected data. Same procedure as the one described in case 2.
3. Authorisation: Same procedure as the one described in case 2.
4. Analytical test. Same procedure as the one described in case 3.
5. Visualisation and report. Same procedure as the one described in case 3.

- Information to data subjects. Same procedure as the one described in case 2.
- Re-identification risk: low.

3.5. Submission of new analytical or visualisation tools

A data user submits the script of a new analytical algorithm in the DSF portal. A bio-informatician team modifies the script of the submitted algorithm in order to make it a privacy-preserving analytical algorithm, and searches for the presence of back door or hidden and malicious script. Once the modifications are done and tests are passed, the analytical algorithm is approved and included in the library of privacy-preserving analytical tools that all the data brokers use.

4. Compliance requirements of the Data Sharing Federation

In order to make the Data Sharing Federation effective, every stakeholder (i.e. data providers, data brokers, data subjects and data users) has to follow the TRUST and FAIR principles, and thereby comply with the requirements of the Data Sharing Federation.

4.1. Data providers

As mentioned in the previous sections, when data providers accept to be part of the Data Sharing Federation, they also accept to be accountable for making Data open and sharable/interoperable in a secure and privacy-preserving manner, while respecting the data subject's will as well as the data contributors' work. To this, the data providers have to:

- De-identify Data (for details, see eTRIKS de-identification recommendations in the D7.8 document) (privacy);
- Store de-identified Data in an isolated and dedicated server that only the data provider and broker can access (for details, see eTRIKS security recommendations in the D7.6 document) (security);
- Curate de-identified Data by using eTRIKS curation guidelines and internationally-adopted reporting, vocabulary, format standards (which includes translating original Data in English. For details, see eTRIKS Standard Starter Pack) and map them with the common data model of the Data Sharing Federation (interoperability);
- Associate each de-identified and curated dataset of a data subject with the conditions of use set by the said data subject. When requested by the data subject, inform him/her the processes of his/her Data, and ask his/her authorisation for those processes (respect to data subject's will); and
- Associate each de-identified and curated dataset with the data contributors who have worked on the said dataset (respect to data contributors' work).

A dataset is available in the Data Sharing federation only after the data broker has approved that the said dataset is compliant with the above-described requirements.

4.2.Data brokers

The data brokers have a pivotal role in the Data Sharing Federation: they are the 'glue' within and between national federated systems not only at the infrastructure level but also at the operational level. They not only provide data users with the same user-friendly, privacy-preserving and secure environment for querying, analysing and visualising Data across countries, but also are the compliance guardians and the support to the stakeholders to reach that compliance. To this, the teams of the data brokers have to work together on:

- Aligning the standard operation procedures regarding data security, privacy, and interoperability;
- Assessing or providing new privacy-preserving analytical or visualisation tools;
- Addressing issues that stakeholders have reported;
- Providing documentation and training material in order to help stakeholders to be compliant with the requirements of the Data Sharing Federation;
- Informing data contributors and data subjects their contribution in the data users' publications with the collaboration of data providers.

In addition to this global work, each data broker has to help stakeholders of its national federated system to be compliant with the national data protection laws, when requested by the stakeholders.

4.3. Data users

The data users play a critical role in incentivising data providers, data contributors, and data subjects to share Data. That is why, when they register in the Data Sharing Federation, they accept to:

- Cite the data contributors as authors in the publications that show the analyses of the Data linked to the data contributors;
- Acknowledge the Data Sharing Federation in their publications;
- Send the abstracts of and the links to their publications to the Data Sharing Federation. This information will be communicated to the relevant data providers, data contributors, and data subjects.

4.4. Data subjects

- The data subjects are the core of any data sharing model, since data sharing depends first on their consent. To make their consent machine readable, the data subjects' platforms have to adopt and use standard consent codes such as the ones proposed by the GA4GH[46]. The data subjects' platforms are also accountable for communicating data users' information (e.g. data request, published results) to data subjects in their native language.

5. Conclusion and recommendations

We are at the dawn of a deep transformation of Life Science towards an open science that requires not only radical changes in our way of sharing Data, but also an engagement of all public and private stakeholders. We have shown in the previous section that Data sharing can quite easily leverage on concepts, tools and procedures that have already been developed and implemented in other economic and scientific fields that have undergone a similar transformation. The engagement of public and private stakeholders is a nascent but not yet sufficient movement for an efficient and global Data sharing. The Data Sharing Federation would enable and foster this engagement. However, some aspects of this engagement are beyond the scope of the Data Sharing Federation, and requires to be further developed or improved. This is why eTRIKS recommends the relevant stakeholders to urgently work on the following points:

- Rewarding the work of data providers and data contributors. Based on the principle of reciprocity, publishers should request the first authors of a submitted publication to provide the data providers and data contributors' names that should be read in the abstracts either as authors or as data providers/contributors;
- Data standardisation goes hand in hand with data privacy and interoperability. Standard organisations that work on terminologies, ontologies, format standards and reporting standards should increase the coverage, the granularity, and interoperability of their standards in order to improve data privacy and interoperability: the more data providers adopt a standard, the more Data are interoperable, the less a malicious user is able to know the data provenance.

[46] Dyke SO et al. Consent Codes: Upholding Standard Data Use Conditions. PLoS Genet. 2016 Jan 21;12(1):e1005772. doi: 10.1371/journal.pgen.1005772. eCollection 2016.

- Improving and increasing the privacy-preserving analytical algorithms. Mathematicians and statisticians should work in collaboration with clinicians and biologists on developing more differential privacy algorithms suitable for the analysis of Data;
- Harmonisation of national data protection laws regarding the processing of Data of alive and deceased data subjects. As explained in the section 1.1, national data protection laws regarding the processes of Data of alive and deceased data subjects are not yet harmonised between the EU member states, which severely hampers the cross-border processing within EU. Data Sharing Federation reduces the burden of the cross-border processing of Data. However, simplifying the cross-border processing within EU is urgently needed and requires a harmonisation of national data protection laws;
- Establish a contract of Data sharing between the data subject and the controller. eTRIKS considers that sharing Data with the whole scientific community is not only an ethical obligation for the good of all, but also has to follow the FAIR and TRUST principles (in particular, the principle of reciprocity towards the data subjects who give the controller their time, effort, data and samples). To address these principles and the data subjects' will of being engaged in research, eTRIKS recommends data subjects to leverage on their rights set in the GDPR, and add conditions of processing their Data for the initial purposes of the controller's study in their consent form where the controller commits to share usable data subject's Data with the scientific community for secondary research purposes with the authorisation of the data subject, when required (for details conditions, see annex).

The implementation of the proposed Data Sharing Federation and the measures that address the above points will greatly foster a seamless, global data sharing, and a powerful, open science for the benefit of the whole humankind.

6. Annex

Proposed Data sharing consent form.

The conditions of processing the data subject’s Data for the initial purposes of the controller’s study in the informed consent are the following:

1. The controller commits to make data subject’s Data fully re-usable for archiving and secondary scientific purposes according to the FAIR principles, which means that the controller:

   1.1. De-identifies data subject’s Data, standardises them by using internationally-adopted standards, and indicates what standards have been used before processing data subject’s Data for the purposes of the controller’s study;

   1.2. Securely keeps data subject’s Data in a database for an unlimited time or until I decide their erasure;

   1.3. Makes the de-identified and standardised Data of a data subject available to the Scientific community through the Data Sharing Federation or any other secure data sharing system that enables their use with the data subject’s permission and a minimal risk of data breach, one year after the end of the controller’s study;

2. Portability of the data subject’s Data (Article 20 of the GDPR). The controller commits to securely send the data subject a copy of electronic files that can be read by a free software, and contains the standardised and not standardised Data of the said data subject, one year after the end of the controller’s study;

3. Rights of data subjects to object:

   3.1. Right to object (Article 21 of the GDPR). The controller informs the data subject any use of their Data for secondary scientific purposes, and wait for their permission of data use for seven full days. After this delay, the controller is allowed to consider that the data subject did not object the use of their Data;

   3.2. Right to object the processing of genetic data when the data subject is vulnerable or deceased. Following the recommendations of the GDPR (Recitals 38 and 75) and the Article 8 of the GDPR regarding vulnerable data subjects, and in order to avoid other judicial cases ‘Henrietta Lacks’[47] where genetic data raise privacy issues regarding relatives with whom a data subject shares common ancestors or from whom the data subject is their ancestor, the controller commits to not process the personal, genetic data of the data subject without the written authorisation of the data subject’s legal representative when the data subject is unable to give their authorisation (i.e. the data subject is very ill or deceased).

[47] https://www.nih.gov/news-events/news-releases/nih-lacks-family-reach-understanding-share-genomic-data-hela-cells