5 key insights to **medical data sharing**

Martin van Mierloo

luminis
Conversing worlds

When you get, give.

When you learn, teach.

Dr. Maya Angelou

In theory, most researchers, companies and institutions support the idea that sharing research outcomes and research data is good practice. Universities encourage new generations of researchers to reproduce and extend the work of others. A study in 2012 [1] showed that 74% of clinical researchers thought sharing Individual Participant/Patient Data (IPD) should be required. Promoting open science and academic benefit or recognition are the main reasons to share data. However, in reality, medical research data are usually not shared and not available to other researchers.

Various studies [2] [3] report that more than 50% of all clinical trial studies don't lead to published results. Withholding results or data can even cause harm when there are significant side effects of treatments or medicines. A recent study from 2020 [4] showed that only 4.5% of all registered clinical trials have the intention to share Individual Patient Data. And even when Individual Participant Data are available then there can be administrative, legal and IT barriers that can years to overcome before the data can actually be used by other organizations [5] [6].

A reality gap

2020 was a transformational year in many ways. The general public learned that developing and approving a medicine normally takes years. New discussions have started about the current clinical research methods and the approval process of new treatments and medicines driven by the urgency to speed up the time to market.

Especially in the first months of the pandemic there were visible and invisible battles to obtain proper and clean data sets. Even the popular John Hopkins COVID-19 map started by manually entering data from Twitter and Asian hospital reports. There was a lack of vision, coordination, policies and standards for quickly setting up one or more global catalogs with validated data sets.

Researchers, governments, institutions and vendors started all kinds of – mostly decentralized - initiatives and collaborations that accelerated the research efforts to find treatments and vaccines for COVID-19. Some were statistically rigorous, many were not. The rush to find the best treatments and claim academic fame created unprecedented events like the 'Surgisphere scandal' [7]. Data management, compliance and transparency prove to be more important than ever.

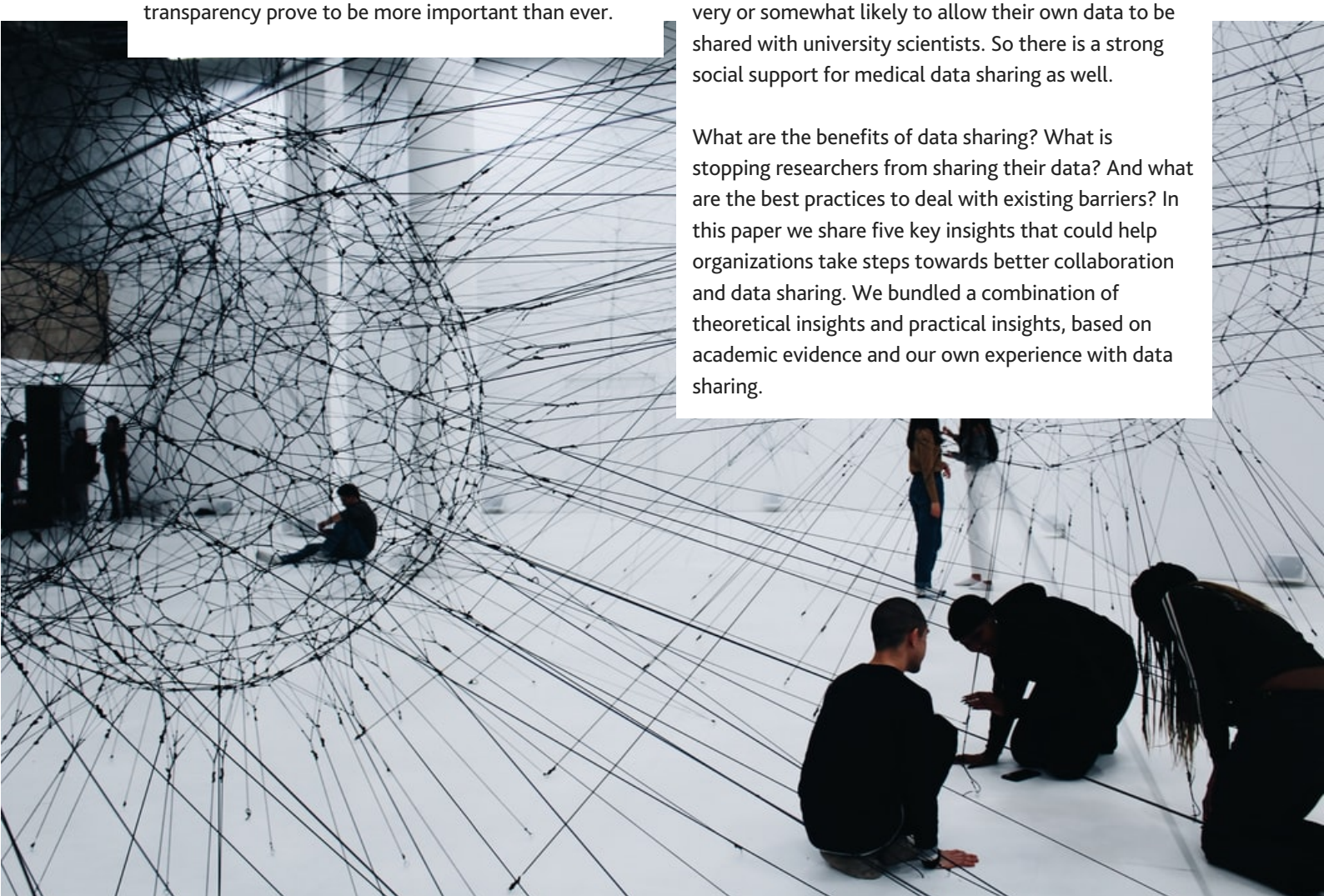
An agenda for the next decade

On a positive note, initiatives like OpenSAFELY and the COVID-19 Symptom Study [8] – to name a few – proved that massive and coordinated data access and data sharing leads to faster and better insights. They also demonstrated the need to include real-time and real-world data to get the best qualitative and quantitive scientific evidence. The results and insights from these initiatives will determine the agenda for collaboration and data sharing in the health sector for the next decade.

Governments (EU, US), policy makers (EU, FD, NIH, PCORI, EMA) and publishers (ICMJE) have all raised the bar for data sharing. Especially over the last 5 years all of these organizations created specific requirements and new policies for data sharing. Not meeting these requirements means not receiving funds, being excluded from projects, and not being published in journals. So there is not only a significant data sharing pull, but also a data sharing push from regulatory institutions.

And what about patients? Mello et al [28] found out that 93% of respondents to their questionary were very or somewhat likely to allow their own data to be shared with university scientists. So there is a strong social support for medical data sharing as well.

What are the benefits of data sharing? What is stopping researchers from sharing their data? And what are the best practices to deal with existing barriers? In this paper we share five key insights that could help organizations take steps towards better collaboration and data sharing. We bundled a combination of theoretical insights and practical insights, based on academic evidence and our own experience with data sharing.



The benefits of data sharing

Medical data sharing takes time, can be complex and costs money. Simply believing in data sharing is not enough to convince coworkers and decision makers to spend time and money on it. So there have to be significant advantages, clear benefits and business cases for medical data sharing.

Since the beginning of the internet in the mid-nineties there have been numerous examples where open standards and co-creation platforms were combined with a massive transformational goal to achieve speed and impact. For example open-source software licenses and platforms such as GitHub allow software engineers to work together on mutual goals in a democratic way. The open-source movement created essential everyday tools such as Linux, Android, MySQL and Wordpress, just to name a few. The scale and impact of open-source software are impressive and lasting. Other famous examples of massive collaboration based on open sharing principles are the Human Genome project (see right), Wikipedia and Kaggle.

Faster innovation, lower costs and trust

At an abstract level, the strategic benefits of data sharing are faster innovation and lower costs. These benefits are caused by three underlying effects:

1. Network effects: with the addition of every node to a network, the structural robustness of the network improves (Percolation theory [9]). Also the overall value of the network grows with every added node (Metcalfe's law [10]).
2. Economic effects: when all parties have equal knowledge the transaction costs of the data will decrease as the number of participants grows. (Coase's theorem [11]). Data in itself is non-rival [12], i.e. data can be consumed by many consumers at the same time. In fact, more and more evidence supports that data is *anti-rival* [13] [14] i.e. its value increases when it's shared and reused.
3. Social effects: sharing and open access foster inclusivity [15], because access to data is not limited by social status, educational status, or geographic location. Another important effect is transparency. Being able to verify data, results and outcomes leads to higher quality and the biggest benefit of all: *trust*.

There is clear evidence that combining these effects has led to the rise of, for example, open-source software, social networks and new business models based on the value of data. Whether the organizational model is large scale open innovation, or private/public partnerships, or simply involving citizen scientists: the speed of innovation will increase, and the overall costs of science will be lower.

Success story: The Human Genome project

One of the most important scientific breakthroughs of all time is the mapping of all the genes of human beings. In 1988 the U.S. National Academy of Sciences stated that within 15 years they wanted to have a full mapping of the human genome. After their initial planning stage they published a plan in 1990 called "Understanding Our Genetic Inheritance: The Human Genome Project". Their prediction was that it would take 15 years to complete the project. In April 2003 the results were published in Nature. 99% of the human genome was mapped with a 99,99% accuracy. 3 billion DNA letters were sequenced. In terms of scientific effort this project was compared to splitting the atom or landing on the moon. The project was a huge success that laid the foundation for many new research efforts.

This project would not have succeeded without good intentions, orchestrated collaboration and extensive data sharing. The project started with a 'moonshot'; a belief that the goal could be achieved, even though the amount of work was significant, and the uncertainty was large. More than 2,500 researchers from 6 countries worked closely together. The results of the project came 2 years earlier than expected and the project stayed 10% below budget.

Even today, the National Human Genome Research Institute has an extremely open data sharing policy and remains an example for data sharing. A quote from their data sharing policy: "NHGRI supports the broadest appropriate genomic data sharing with timely data release through widely accessible data repositories". They also make remarks about the type of consent: "Similarly, consent language should avoid restrictions on the types of users who may access the data.". In other words: also the given consent should not limit the use and reuse of data.

Insight 1: privacy and compliance are not the main reasons to not share medical data

There have been many studies investigating the attitudes towards sharing or not sharing data. A study from 2011 [16] revealed some interesting findings: 75% of all respondents agreed that data may be misinterpreted due to the *complexity* of the data. 71% agree that data may be misinterpreted due the *quality* of the data. The main reasons to *not* share data are: 'insufficient time' (54%), 'lack of funding' (40%) and 'insufficient rights' (24%).

A study from 2012 [1] clustered the reasons not to share data in three main categories:

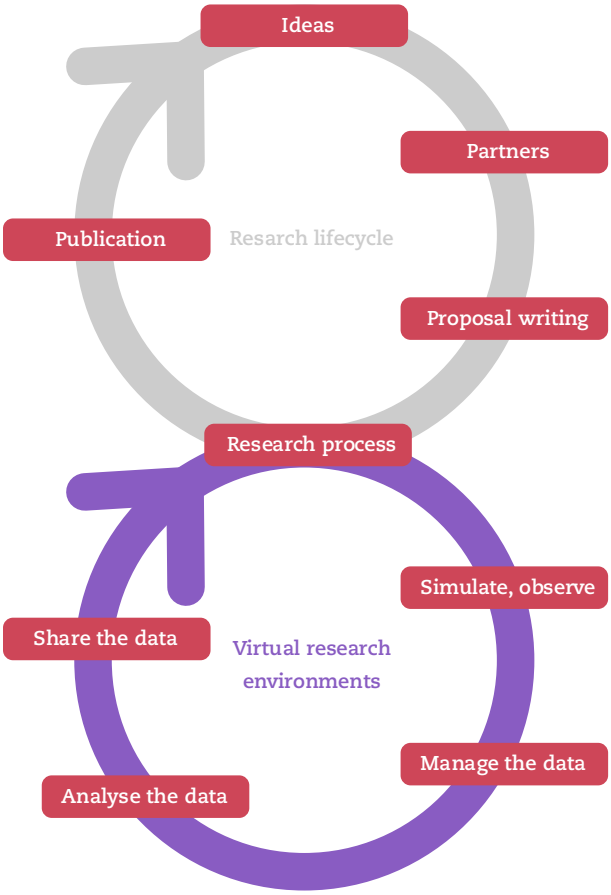
- Concerns related to appropriate data use (n=205)
- Concerns related to investigator or funder interests (n=129)
- Concerns related to protection of research subjects. (n=91)

In a study by the National Institute of Health [17], 190 respondents were questioned about their perceptions and attitudes towards data sharing. A quote from their conclusion: "*While some technological challenges do exist, many of the barriers to sharing and reuse are* ***social*** *in nature, arising from researchers' concerns about and attitudes toward sharing their data.*"

The answers revealed that the interviewed researchers found it important that some kind of acknowledgement is given when their data is (re-)used, for example co-authorship, citation or recognition. Other social factors include statements like "It isn't customary to share data in my research field" (35%). Another common factor is priority or urgency: "I haven't had an opportunity to do so" (45%). But also functional/technical reasons were mentioned: "I don't know any repositories that accept the kind of data I produce" (45%) or "I don't know how to prepare my data for sharing with others" (30%).

What may be surprising is that there are fewer concerns about privacy or commercial interests than there are concerns about control, misinterpretation, acknowledgement and data quality. And there are practical implications such as time, priority and costs.

In order to overcome the practical implications Tenopir et al [16] proposed an extension to the classical academic research loop in which data sharing is a second result next to publication:



What all these studies boil down to is that even though there are practical barriers, the real change has to come from changing the habits of researchers and finding solutions to work on the trust and control aspects of data sharing.

The good news is that over the last 10 years proven solutions have been created to create better transparency. There are good solutions to deal with privacy, see insight 2. Building data ecosystems creates systems of trust, see insight 3. Metadata is essential for good data quality and data provenance, see insight 4. And 'data visitation' as mentioned in insight 5 is a decentralized data architecture that should create better transparency, control and, ultimately, trust.

Insight 2: there are proven and compliant solutions for sharing data

As we learned on the previous page, compliance is not the biggest issue when deciding on whether or not to share data. However, words like 'privacy', 'security' and 'compliance' pop up in every conversation about data sharing. Of course this has to do with the privacy of patients and the fact that everyone has to comply with existing privacy laws.

Privacy laws like GDPR use principles like 'purpose limitation', 'data minimization' and 'storage limitation'. This 'less is more' approach seems to conflict with a broader data sharing vision which encourages the reuse of original data at a scale aimed at unknown or unforeseen new applications. The good news is that the same law offers solutions as well. These vary from a generic 'privacy by design' principle to common tools like anonymization and pseudonymization.

There are three main steps to protect the privacy of patients when sharing data outside your organization:

1. Remove all records without full and proper patient consent. This means all patient consent fields must be stored in a digital way and have to include basic attributes such as personalia, date of signature and possible limitations (expiry date).
2. Remove any (reference to) personal identifiable data (e.g. name, date of birth etc.). The strictest way to achieve this is by applying anonymization. This means altering the data in such a way that a data subject can no longer be identified directly or indirectly. The simplest way to achieve anonymization is simply remove or omit any data fields with personal identifiable information. A more advanced method is pseudonymization. This method replaces personal identifiable information with placeholders. Placeholders can be unique to a record ('DP Johnsen' is replaced with 'Patient_93829') or can represent a range of data. For example all birthday fields are replaced with a an 'age' or 'year of birth' field, addresses are replaced by cities, regions or countries). There are many tools and methods for pseudonymization available. Best practice is to anonymize where possible and to apply pseudonymization to all other fields.
3. One could say that applying 1 and 2 should be sufficient. But especially with medical data there is a chance that individual persons can still be identified by using other linked data sets. Suppose a data set contains fields such as 'age', 'ZIP code' and 'Disease'. When there is a second data set with research data about diseases per zip code it could turn out that unique records can be derived by combining the two and thus making it possible to identify unique patients. In 2002, Latanya Sweeney came up with a mathematical way to prevent this by using 'k-anonymity' [18]. A few years later a second method called 'ℓ-diversity' [19] was added to even further strengthen the guarantee that persons cannot be identified. Use a solution that supports and applies these methods in order to guarantee the privacy of your patient.

Applying and automating these three steps lays the foundation for safe and controlled data sharing. In order to fully meet all rules and regulations there are of course more requirements that you have to meet, such as having a privacy policy, assigning a Data Privacy Officer (DPO), performing data protection impact assessments and other criteria that are out of scope for this paper. Also remember that there could be national laws or other regulations that require additional measures.

All these steps obviously require some work and constant attention. The good news however is that there are excellent solutions for compliance requirements. Being afraid of non-compliance shouldn't be a reason not to share data. Before you start building your own solutions for compliance, our advice would be to look for existing and proven software solutions that offer compliance checks and functionality.

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
	Fully closed	**Internal ecosystem**	**Trusted partner ecosystem**	**Managed ecosystem**	**Open access ecosystem**
Access management	Manual access control Role based access control	Role based access control Attribute based access control	Virtual private network Federated access SAML	Central directory	Open
Data management focus	Data quality	Data architecture Master data management Metadata management Data integration	Data security & compliance	Licensing Data interoperability Data lineage/provenance	Attribution

Insight 3: there are many types of data ecosystems

The ultimate goal behind open science is an all-access world in which all data, metadata, insights and knowledge are openly available to everyone in the world. Of course this doesn't happen overnight.

Our advice is to take a stepped approach towards data sharing. The figure above shows five stages in which an organization can grow from a fully closed organization to becoming part of an open access ecosystem. This figure shows that the various stages create a different focus for access management and data management.

An important first step is the transition from closed data silos to an in-company data ecosystem (Stage 2). Combining data from multiple sources in, for example, a data lake, data warehouse or data platform forces organizations to think about their data architecture and data integration. It also forces them to think about metadata management and other types of access control. Even though this can be challenging from a technical, functional and organizational perspective, it will lay the foundation for every other form of data sharing.

The second step is to start sharing data with one or more trusted partners. There have been many examples of public/private collaborations based on exchanging data. For example, farmers exchanging data with suppliers, logistics companies, buyers and researchers [20], or manufacturing companies sharing lots of data with suppliers in order to optimize their supply chain.

Research from McKinsey [21] shows that 40% of the companies that created an ecosystem for sharing believe they are on a path to create revenue from the ecosystem. 10% actually make more than 5% of their revenue from ecosystem plays.

McKinsey [25] also mention three ways data ecosystems provide value to companies:

1. Growth: new value propositions or even lines of business can create new revenue streams
2. Productivity: new insights from combined data can lead to a higher efficiency
3. Risk reduction: data can be used for risk analysis and fraud detection.

The next step is to open up the ecosystem for other parties. In this managed ecosystem (Stage 4), new organizations can sign up to become a node in the network. Most managed ecosystems use one or more central platforms to exchange data.

An example of managed ecosystem is a Healthcare Information Exchange (HIE). There are three types [22]:

1. Directed exchange: connected institutions can send and receive information to and from each other.
2. Query-based exchange: more focus on storing data and being able to query consolidated datasets
3. Consumer mediated exchange: ability for patients to access and managed their health information

These managed ecosystems or HIEs provide immediate and significant value as they not only improve the efficiency of the daily healthcare operation, but can also act as primary source for data discovery and exploratory analysis.

Starting with data sharing internally and growing towards a data ecosystem helps organizations grow in data maturity and benefit from a range of opportunities.

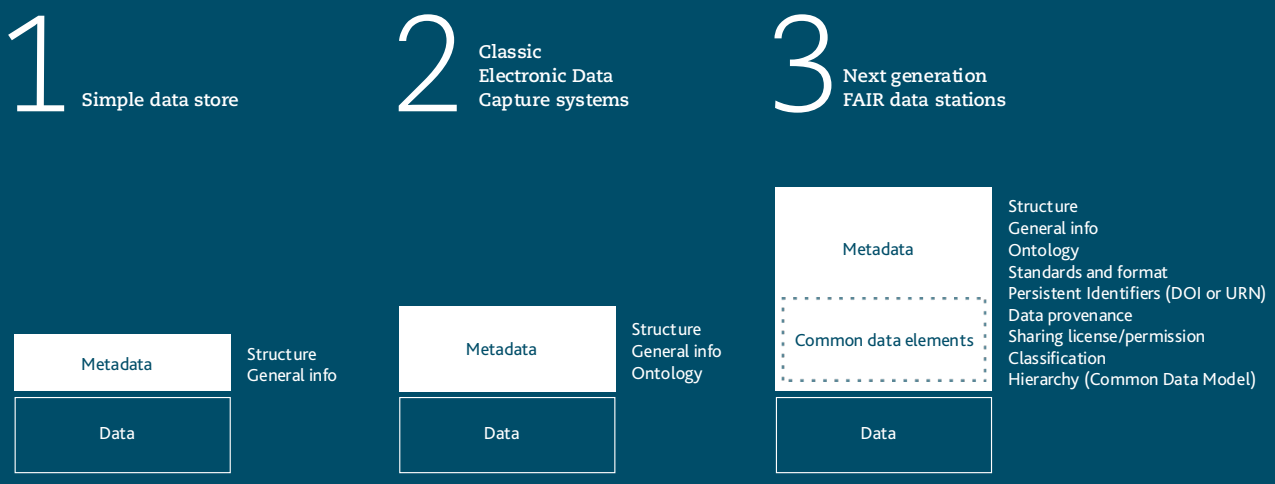
Insight 4: Metadata is more important than data

In many publications and in conferences there is a lot of focus on data and how to get new insights from data. But data has no meaning and no context without proper metadata. In many types of data stores metadata is often limited to the structure of the data (type of fields, length of fields etc.) and general properties (author, version etc.). Modern data storage and sharing solutions require a more detailed and extensive set of metadata. There are three main reasons for this:

1. Compliance has become top priority over the last decade and therefore has to be part of the metadata as well. Aspects such as data provenance, data lineage and use of personal data are highly relevant for third party users of the data. Not only are they part of Good Clinical Practices for example, but in a broader sense they convey trust. And, as we learned from the reasons that people do not share data, trust is one of the largest barriers to overcome and prevents many researchers from sharing their data. Being able to trace back the data to its origins and not having to worry about consent or privacy will help data sharing in the long run.
2. Information is more and more linked. Instead of using natural language to convey information, standardized ontologies offer more formal relations between subjects and entities. The use and application of ontologies for medical purposes (e.g. UMLS, SNOMED-CT, LOINC), copyright (e.g. Creative Commons Rights Expression Language) and many ontologies for specific areas such as genetics, preclinical, nutrition etc. is growing. The FAIRsharing website (https://fairsharing.org/) contains a database of available standards, databases, policies and collections. By applying standards and ontologies as part of the metadata researchers improve the findability and the interoperability of their data.
3. Machine readability has become more important. In order to benefit from technologies such as AI and smarter information retrieval systems it is imperative to offer data and datasets in machine readable formats. This basically means that all data is available in a digital, readable and open data format (e.g. XML, JSON, CSV), using standards and ontologies (see previous point) and accessible without human intervention (no manual checks and procedures for example).

These trends and the increased attention to data management will require more time from researchers and data managers and the requirements for data stores will increase. Many clinical research data are stored in simple file formats (such as Excel) or in Electronic Data Capture (EDC) systems. Next generation data repositories that fully support data sharing offer better data management and support the workflow to achieve this. Another trend is the use of Common Data Elements. Data are grouped in common data elements in a logical way with its appropriate metadata. This simplifies re-using data structures and metadata and creates better classification of data sets.

Applying these and other FAIR principles to your own data infrastructure will support a future in which 'personal health trains' [23] and 'deep neural nets' [24] can autonomously search and retrieve the right publications and data sets and propose next steps for future research. This may sound like sound like science fiction, but the technology is already there!



Insight 5: There are many data sharing models; choose the right one!

There are several models and dimensions when it comes to data sharing. The basic questions are *why* you would like to share, *what* you would like to share and with *whom* you would to share. Another fundamental question is of course *how* you would like to share data. Ad-hoc data sharing – sharing a data set after a request - could be a start, but it's not a structural solution.

The classic pattern for medical data sharing – and for most data sharing in organizations for that matter – is a centralized model. All Individual Patient Data from all parties is stored and consolidated at one central location. This central location is where all the analysis takes place. The main advantage of this model is that all detailed data are available for analysis and data only flows downstream to a central data lake or data warehouse. The centralized model has several significant disadvantages though:

- It is difficult to manage in terms of privacy and security. The impact of a data breach increases with every new node.
- Maintaining a consistent data quality level is difficult. Every node has to guarantee a specific data quality level.
- It will be difficult to handle and consolidate different types of data.
- It is difficult to scale. It will work with up to dozens of nodes, maybe even hundreds, but as the volume increases it will be more difficult to manage in terms of processing, computation and analysis. There is one single point of failure.

There are solutions and workarounds for most of these disadvantages, but these disadvantages are inherent to the centralized model.

Therefore it is certainly interesting to look at other, more decentralized models. Schiebner et al [26] described various models for decentralized data sharing. Their model for 'site level meta-analysis' uses aggregated data sets that can be distributed. Other nodes can create their own analyses based on aggregated data. A master node can do meta-analysis based on the local site analysis. This model doesn't share Individual Patient Data on a central location and is therefore safer and easier to manage. The biggest downside is that the granularity of the aggregated data prevents detailed level analysis.

A model that is very suitable for machine learning applications is the federated learning model. In this model nodes are training local models, which are transferred into a larger central or global model.

One of the upcoming models is the 'data visitation' model [27]. This model uses a combination of a central broker function and nodes that can share data with other nodes directly, or with 3rd parties. The broker takes care of permissions, licensing/contracts and encryption. It can also act as a catalog or index for the FAIR metadata. All nodes will always remain in control of their own data. From an architecture and standards point of view this model looks a lot like the architecture of the internet itself. It is infinitely scalable, open and highly adaptable. Therefore we believe this is the most promising architecture for the next decade.

Epilogue

In this paper we have barely touched the surface of a complex topic such as data sharing. By providing you with different angles and relevant insights supported by recent academic research we hope to convince you that medical data sharing is a topic that is worth exploring. There will be definitely be challenges but any form of data sharing will contribute to open science and will ultimately increase the speed of medical research and development of better medicines and treatments.

If you would like to discuss this topic with one of our specialists then you can always contact us. We are happy to support any organization that wants to support medical data sharing.

Also any form of feedback or discussion regarding the contents of this paper is always welcome.

Martin van Mierloo

martin.vanmierloo@luminis.eu
+31 (0)6 – 14 84 15 19

About the author
Martin van Mierloo works as a product manager for Luminis Technologies and is responsible for the product strategy and roadmap for data solutions such as InformationGrid and MediGrid. He has more than 20 years of experience with developing innovative digital and data driven solutions of which more than 10 years in the health sector. Martin regularly publishes about digital and data strategies.

About MediGrid
Developed in close collaboration with the University of Liverpool, MediGrid was developed to consolidate and analyze data from various British hospitals. MediGrid is not only used for smarter analysis but can also be used to share data in safe and structured way. MediGrid is very well equipped for the application and analysis of real world data (RWD) and contains a user friendly interface to create your own reports and dashboards.

Website: https://medigrid.io

References

[1] V. Rathi, "Sharing of clinical trial data among trialists: a cross sectional survey," 2012. [Online]. Available: https://doi.org/10.1136/bmj.e7570.

[2] R. Chen, "Publication and reporting of clinical trial results: cross sectional analysis across academic medical centers.," 2016. [Online].

[3] N. J. DeVito, 2020. [Online]. Available: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)33220-9/fulltext.

[4] E. E. Statham, "Primed to comply: Individual participant data sharing statements on ClinicalTrials.gov," 2020. [Online]. Available: https://doi.org/10.1371/journal.pone.0226143.

[5] S. J. Nevitt, "Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review," 2017. [Online]. Available: https://doi.org/10.1136/bmj.j1390.

[6] V. M. Schünemann, "Obtaining and managing data sets for individual participant data meta-analysis: scoping review and practical guide.," 2020. [Online]. Available: https://doi.org/10.1186/s12874-020-00964-6.

[7] C. Offord, 2020. [Online]. Available: https://www.the-scientist.com/features/the-surgisphere-scandal-what-went-wrong--67955 .

[8] C. H. Sudre, "Attributes and predictors of Long-COVID: analysis of COVID cases and their symptoms collected by the Covid Symptoms Study App," 2020. [Online]. Available: https://doi.org/10.1101/2020.10.19.20214494.

[9] V. K. Shante, "An introduction to percolation theory," 1971. [Online]. Available: https://doi.org/10.1080/00018737100101261.

[10] B. Metcalfe, "Metcalfe's Law after 40 Years of Ethernet," 2013. [Online]. Available: https://doi.org/10.1109/MC.2013.374.

[11] R. Coase, "The Coase Theorem and the Empty Core: A Comment," 1981. [Online]. Available: https://www.journals.uchicago.edu/doi/abs/10.1086/466980?journalCode=jle.

[12] C. I. Jones, "Nonrivalry and the Economics of Data," 2019. [Online]. Available: https://doi.org/10.3386/w26260.

[13] I. Kubiszewski, "The production and allocation of information as a good that is enhanced with increased use," 2010. [Online]. Available: https://doi.org/10.1016/j.ecolecon.2010.02.002.

[14] N. Pekka, "Will the data markets necessarily fail? A position paper PDF Logo," 2019. [Online]. Available: http://hdl.handle.net/10419/205201.

[15] R. Vicente-Saez, "Open Science now: A systematic literature review for an integrated definition," 2018. [Online]. Available: https://doi.org/10.1016/j.jbusres.2017.12.043.

[16] C. Tenopir, "Data Sharing by Scientists: Practices and Perceptions," 2011. [Online]. Available: https://doi.org/10.1371/journal.pone.0021101.

[17] L. Federer, "Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff," 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0129506.

[18] L. Sweeney, "k-anonimity: a model for pretecting privacy," 2002. [Online]. Available: https://doi.org/10.1142/S0218488502001648.

[19] A. Machanavajjhala, "L-diversity: Privacy beyond k-anonymity," 2007. [Online]. Available: https://doi.org/10.1145/1217299.1217302.

[20] S. Celik, "Data sharing in the agricultural sector," 2019. [Online]. Available: https://eudatasharing.eu/examples/data-sharing-agricultural-sector.

[21] V. Chung, "Ecosystem 2.0: Climbing to the next level," 2020. [Online]. Available: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/ecosystem-2-point-0-climbing-to-the-next-level.

[22] HealhIT.gov, 2020. [Online]. Available: https://www.healthit.gov/topic/health-it-and-health-information-exchange-basics/what-hie .

[23] B. Mons, "The VODAN IN: support of a FAIR-based infrastructure for COVID-19," 2020. [Online]. Available: https://doi.org/10.1038/s41431-020-0635-7.

[24] C. Cai, "Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making," 2019. [Online]. Available: https://doi.org/10.1145/3290605.3300234.

[25] M. Aaser, "Four ways to accelerate the creation of data ecosystems," 2020. [Online]. Available: https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/four-ways-to-accelerate-the-creation-of-data-ecosystems.

[26] J. Scheibner, "Revolutionizing Medical Data Sharing Using Advanced Privacy Enhancing Technologies: Technical, Legal and Ethical Synthesis," 2020. [Online]. Available: https://arxiv.org/pdf/2010.14445.pdf.

[27] C. Sun, "A Privacy-Preserving Infrastructure for Analyzing Personal Health Data in a Vertically Partitioned Scenario," 2019. [Online]. Available: https://doi.org/10.3233/shti190246.

[28] M. M. Mello, "Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing," 2018. [Online]. Available: https://www.nejm.org/doi/10.1056/NEJMsa1713258.

Photo credits:
Cover: Elaine Casap - https://unsplash.com/photos/qgHGDbbSNm8
Page 3: Alina Grubnyak - https://unsplash.com/photos/R84Oy89aNKs
Page 6: Chris Yang - https://unsplash.com/photos/1tnS_BVy9Jk
Page 10: Jamie Templeton - https://unsplash.com/photos/6gQjPGx1uQw