

## VIEWS ON DATA IN HEALTH

Danish health data is of exceptional quality and represents a great research and innovation potential. In connection with several projects, however, the Innovation Fund has identified major challenges in accessing and using health data and thus in realizing these potentials. Examples include the fact that it has taken several years to obtain the necessary permissions to obtain and link different data sources.

Therefore, the Innovation Fund has taken the initiative to organize the workshop "Data in health", where we have invited significant stakeholders around the use of Danish health data to achieve the following:

- Overview of the circumstances that make access to and use of data in the health area difficult.
- Exposure of expected future health needs.
- Proposal for a plan for solutions that ensure better utilization of the Danish data relevant to health.

The purpose of this document is to create reflection among the participants prior to the event, so that the work at the workshop can take place more efficiently and get more benefits.

We have therefore sought in a short form to provide a common understanding of data sources, data purposes, challenges and needs regarding the use of health data.

As part of this, we here raise a number of questions about the identified challenges, which are sought to be solved at the workshop. These issues are based on a number of expert interviews and are divided into three categories: **data logistics, data security and value creation from data.**

### OVERVIEW OF DATA AND ACTORS

High quality health data has great and wide value because they allow us to make new insights and make important decisions on a solid, factual basis. An ever-growing amount of data is produced and consumed by a complex and growing group of public and private actors. These actors have very different interests, focus and motivation in their work with data, and thus there is a great need for coordination and conflict potential built into the field.

The various actors have many different purposes of collecting and using data. The most common ones are:

- Better diagnostics (accurate and earlier)
- Better treatment (custom and better timing)
- Better public health and prevention
- Better individual health
- Epidemiological insights - both correlation and causality
- Improved health production, including the use of data to predict demand and resource consumption
- Quality assurance and development
- Commercial use in research, development and marketing
- Citizens' access to their own data
- Third Party Access to Data

### DATA LOGISTICS

### HOW DO WE CREATE DYNAMIC DATA?

When researchers create new knowledge with data and, for example, Purifying dataset so that they are possible to work with, these refined data should not be born back into the system in this improved version, regardless of whether they could be of great benefit to other researchers. Thus, there is a large loss of potential in that data, often publicly funded, is not available to others through the official channels.

### HOW DO WE USE DYNAMIC DATA?

The Danish registry data is built around static data. The technological advances now make it possible to work with dynamic data to a far greater extent than just a few years ago, and this allows the use of real-time data for decision support in the clinic. The use of dynamic data requires attributability to be relevant in a health promotion context. The common practice is now to allow data to be viewed retrospectively one year at a time, which is both labor intensive and heavy and in practice prevents the use of dynamic data.

### HOW DO WE MAKE ACCESS TO DATA EASIER?

In general, there is dissatisfaction with the process through which permission to use personal data for research is obtained. It is widely believed that it is very difficult and unmanageable to get permission to work with data, and that the processing time is unnecessarily long and very disruptive. Lack of competence to address the often complex issues of the authorizing authorities is pointed out by several as a major weakness.

It is shown that no guiding help is available while the application is being designed. Therefore, much time is spent on understanding the application requirements, and a lot of working time and waiting time is lost, since feedback comes after the application is submitted.

Another common criticism is that the individual authorities do not state that an additional license (s) must be applied for by the other or other authorities.

Today, a large number of complex demands are made on the applicant, including:

- Science ethical permission
- Permission for use of data registers and possibilities for interconnections
- Sharing data with different purposes
- Cross-border sharing
- Sharing with the citizen
- How clear a hypothesis should be

Ideally, the authorities could offer application support to meet these requirements and the above-mentioned challenges, in the form of:

- Updated knowledge about the legal basis - where is the legal basis for the various data her
- Information, guidance and help with completed applications
- Contact with other relevant parties
- Acceleration of case processing time
- Deep expertise and ability to understand specific professional issues

- Possibility of personal help during the application process
- A much clearer online interface for executing applications

#### HOW DO WE AVOID DATA NEPOTISM?

It is pointed out that data potentiation is a pronounced phenomenon where a data controller prevents others equal access to secure their own monopoly on data. It is shown to be the case when data controllers themselves research the data. This is not in the public interest.

#### HOW DO WE AVOID THAT DATA IS LOCKED BY SINGLE ACTORS?

There are clear examples of data lock-in from both public and private actors. There are examples of private computer system providers that lock data into formats and the page requires extra payment if they are to be used for purposes other than originally intended, eg. research. This is a real obstacle to the free movement of data.

It is also shown that there are challenges with data sharing between actors and administrative levels and borders - eg. regions.

#### HOW DO WE CREATE AN OPTIMAL SYSTEM TO GET AND USE DATA?

Part of the broad health data discussion has covered the need for common IT systems, including having a common Danish EPR system.

However, consensus among experts in the field is that there is a need for interoperability rather than everyone having to use the same system. There are several reasons for this:

On the one hand, technological development is so rapid that large metastatic systems are too heavy and almost obsolete when implemented; and secondly, there will be a continuous need for special systems, e.g. X-ray must be able to work with the other systems.

It is demonstrated by experts that today it is possible to link different systems as long as access to source data is provided. Thus, there seems to be a broader attitude shift towards the challenges of different data standards being handled as long as source data is available. This corresponds to the paradigm shift that has taken place in other business areas, where one is rapidly moving away from large monolithic systems and towards loosely coupled databases controlled by data management middleware. This achieves a much greater agility in relation to incorporating new data sources in a dynamic context. However, working with common standards and interoperability is an advantage in the future.

#### HOW DO WE ENSURE THAT NEW AND ADVANCED DATA SOURCES CAN BE WORKED AND USED IN THE HEALTH WORK?

Molecular biological data of various kinds (omics) are attributed to an increasing importance for health. Not least in light of the Danish ambitions for personal medicine and the establishment of a national genomic center. Use of these data in disease situations is likely to require the compilation of data from multiple countries. There are discussions in and outside Denmark about models to achieve this. In order to work optimally in a clinical context, it requires the possibility of working with attributable data. It contradicts the public announcements in several countries, which point out that anonymous data is only shared.

## DATA SECURITY

### HOW DO WE HANDLE PATIENT AND CITIZENS ACCEPTANCE?

It is an open question whether the Danish model for consent is good enough. There are several alternative models of consent to choose from: generally open consent, individual specific consent to each action, some form of dynamic consent or the Data for Good model.

### HOW DO WE HANDLE DIFFERENT REQUIREMENTS FOR CONTROL AND SECURITY?

Data security and security is an important issue. It is clear that most actors are accustomed to a systemic approach to control and security. There are players who are tempted to get personally identifiable data for fear of security and the requirements that are set, perhaps especially with regard to the uncertainty surrounding GDPR. GDPR is mostly perceived negatively, though by some as a progress. Basically, the need for control and security can be divided into three categories:

#### 1. Systemic control

- a) Various authorities
- b) Public and private organizations

#### 2. Personal control

- a) Transparency and traceability; the individual's right to insight into all personal data, as well as the right to follow their use and to be informed where they are
- b) Accessibility and usability; requirements for interoperability so that data cannot be locked, access to own data and the ability to use them as desired

#### 3. Patient and citizen control

- a) Security in the form of data not being hacked, stolen, misused or published
- b) Security in the form of external actors not changing data without being detected (leading to incorrect data entry processing)
- c) Security in terms of use to data to get the safest possible treatment

### HOW CAN WE DIFFERENT DATA ACCESS BASED ON DATA SENSITIVITY?

Much of the data produced in hospitals is production data that is not directly sensitive to persons. However, these data have great value in the development and optimization of the hospitals. Here they find application in relation to creating better workflows and higher quality, as well as for predicting load and resource needs.

In the debate on and the management of data access, there generally does not seem to be a distinction between personable data and production data. It is considered important to draw attention to this distinction and to adapt the acquis accordingly, so that full and immediate value can be extracted from these production data.

### WHO SHOULD BE RESPONSIBLE FOR NON-CLINICAL DATA?

The system is composed so that the health system has responsibility for action (is the responsibility) in relation to the data defined as health data (clinical data).

It slows the zeal of the established system to work with other data sources, regardless of perceived needs, as one cannot overlook and perceive the responsible aspect.

### **CAN WE RECONCILE THE DIFFERENCE BETWEEN RESEARCH AND CLINIC?**

There is a clear dividing line between research and clinic, which most are perceived as an artificial barrier that prevents data and knowledge flow. This has the negative effects of hampering the transformation of research into clinical practice, the exercise of specific clinical research, and the performance of real-time work.

The dividing line is, among other things, a consequence of the fact that a great deal of identification protection is required in relation to research, but not in relation to clinical treatment, where data must be able to be easily attributable. This creates a distinction between building clinical and research use according to the same principles. A further comment on this is the fact that we are attributable in our online life, such as on social media and various digital services.

### **HOW DO WE GET TO PERSONAL MEDICINE AT THE SAME TIME THAT WE ENSURE THE ANONYMITY FOR THE INDIVIDUAL?**

Generally speaking, one can say that one cannot work predictively with personal health without access to attributable data that creates a better understanding of the individual disease. Therefore, a real discussion of anonymity and de-identification is required as to the extent to which data can be attributed.

Working with causality also requires directly attributable data sources, which make it possible to see which events are associated with the individual. Although a hidden name and artificial identification protects the individual's identity, it will at all times be possible to break due to compound unique data points. Therefore, it may be necessary to allow only algorithms to find causality, so that "natural persons" do not fish in sensitive data. In this way, an actual anonymity is ensured compared to that researchers and developers cannot see the individual persons.

### **VALUE CREATION FROM DATA**

#### **HOW DO WE SECURE PRIVATE INNOVATION BASED ON PUBLIC DATA?**

One of the burning platforms is the challenge of creating a dynamic public-private partnership. In this context, there are public voices that demonstrate that business insists on accessing data without responding to how their access to data contributes to a higher social value rather than simply improving the individual company's product in one. business purposes. Thus, a challenge is thrown to the private actors to demonstrate social value in the event that they gain access to data.

#### **SHOULD WE INCLUDE HEALTH DATA CREATED BY NON-PUBLIC ACTORS?**

If we do not include health-relevant data that is not produced by the health system / public actors or in the established health regime of private actors, then these data must be assumed to come on the market from other large international players. There is an explosion in the area that will hit all countries, including Denmark. In this context, digi.me, which already works with Icelandic health data, as well as companies such as LunaDna and Nebula, use blockchain to ensure users full traceability as well as earnings from genomic data and health data. We also see the four major US - Facebook, Google, Amazon and Apple - investing heavily in healthcare and similar developments in China with Tencent, Alibaba, Huawei and Baidu. These actors are big and everyone works with the composition of data across - and this will also have an influence in Denmark.