The data revolution in life science and healthcare

The Future of Life Science at Engelsberg 2022

STOCKHOLM SCIENCE CITY
FOUNDATION

Introduction

Life science is all about deciphering biological complexity. With the rapid development over recent decades of new technologies and novel methods of data analysis, we are now in a position where integrated biological systems can be studied holistically.

Data has always been the foundation for research, and the number of available data sources and data points is currently increasing at a remarkable rate. In combination with powerful computational resources and advanced methods of analysis, data can today provide new insights into human biology and disease mechanisms at an unprecedented level. This has brought life science into the era of systems biology and enabled the development of precision medicine whereby optimal treatments can be selected with the help of precise diagnostic methods.[1]

There is consensus within the scientific community that development of systems biology and the implementation of precision medicine is hindered, or at least slowed down, by challenges of sharing and utilizing data. This report addresses important questions related to data in life science and healthcare. **In the first section**, examples are provided of state-of-the-art data utilization in research, with the ultimate goal of providing new and better drugs and treatments. **In the second part**, different types of data, their value, and challenges of sharing and taking full advantage of available data are presented. **The final part** provides a broader discussion on the importance of trust, openness and collaboration in order to accelerate scientific advances for the benefit of all. Ensuring transparency in systems in which new knowledge is generated, is vital for the function of societies around the world.

"Today, data can provide new insights into human biology and disease mechanisms at an unprecedented level."

The report builds on presentations and discussions at the symposium *"The future of life science – the data revolution in life science and healthcare"* held at Engelsberg Ironworks, Sweden, in May 2022 (see the appendix for the program with speakers). The aim is not to cover all the topics discussed at the symposium, but rather to focus on key topics to highlight immediate challenges and, most importantly, possibilities for the future.

Engelsberg Ironworks

[1] *Stockholm Science City Foundation. (2020). Report: Stockholm Science City Foundation. (2020). Report: The Future of Life Science – Health data and precision medicine.*
https://ssci.se/sites/default/files/Health%20Data%20and%20Precision%20medicine_0.pdf

Part 1: Implications of the data revolution in life science and healthcare

Molecular mechanisms, systems biology and precision medicine

What is meant by the data revolution in life science? Even though data has always been fundamental to all research, there are several developments that have led to the rise of the term "data-driven life science" during the last few years. Firstly, the advances have been fueled by an explosion in the amount of data from different sources. As **Janne Lehtiö**, Professor of Medical Proteomics at Karolinska Institutet, pointed out during the symposium, there has been an exponential increase in the amount of life science data during the recent decade, and in computing power.

Janne Lehtiö

In combination with increasing interest in cross-disciplinary collaborations, the scope of many biological and medical studies has expanded. Biological processes, mechanisms of disease, and the action of different drugs can now be studied at both a detailed mechanistic level and a systemic level. The result is the emergence of systems biology.

Systems biology is the study of biological systems whose behavior cannot be reduced to a linear sum of the functions of their parts, and therefore focuses on complex interactions within, for example, an organism, tissue or cell. It includes experimental approaches and mathematical modelling. Increased knowledge at the mechanistic and systemic levels has led to the development of precision medicine within healthcare, where the best-suited treatment can be given to a specific patient at the right time, based on individual factors including genes, environment and lifestyle.

What is the status of data-driven life science, systems biology and precision medicine today? It can be argued that the road leading to our current position was paved by the Human Genome Project and the first sequenced human genome 20 years ago.[2] After that breakthrough, several large mapping projects including the mapping of all human proteins (Human Protein Atlas, led from Science for Life Laboratory in Stockholm) have expanded the foundation for understanding human biology.[3] Crucially, research data from such projects can be readily accessed, and this has led to rapid advancements in many areas. **Mathias Uhlén**, Professor of Microbiology at KTH Royal Institute of Technology, Stockholm, told at the symposium that at least half of the data used for research carried out by his group is data already available in the open access resource as part of more than 15 million web pages.

Mathias Uhlén

Cancer treatment is probably the field in which precision medicine has advanced the furthest. Thirty years ago, a cancer type such as lung cancer was classified as one disease. Advancements in immunohistochemistry enabled division into subclasses about 25 years ago, and with the recent introduction of genomic tests, different mutations and genetic profiles are now used as a basis for diagnostics. Based on this knowledge, novel drugs and optimized combination treatments have been developed. As a result, new treatments which are more effective, often with fewer or less severe adverse effects, are now available, leading to increased survival expectance among several groups of cancer patients.

---
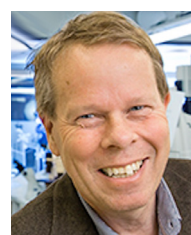
[2] *International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860–921.* *https://doi.org/10.1038/35057062*
[3] *Uhlén M et al. (2015). Tissue-based map of the human proteome. Science 347(6220):1260419.* *https://www.science.org/doi/10.1126/science.1260419*

Recent and ongoing research studies are able to explore how multi-omics approaches (those using a combination of data from different methods) can be used to diagnose cancers and allow even better and more precise selection of treatments (see Fact Box 1 for an overview of different -omics methods). Janne Lehtiö's research group is combining genomics, epigenomics, transcriptomics, proteomics and immunohistochemistry in order to better match each patient with different anti-cancer therapies based on tumor characteristics. Analysis of data from the different sources shows specific patterns that can be used to group the patients. This new segmentation of cancer types results in greater precision in the diagnoses, with rich molecular information on cancer drivers, which is now being used to find novel combinations of treatments, to achieve better outcomes. The next step is to find new protein types, caused by cancer coupled defects in the genome, among the different cancers that can be used as targets for the development of cancer vaccines.

A promising field in the diagnosis of cancer, and probably other diseases, is profiling proteins in blood samples. A project led by Mathias Uhlén has mapped protein and metabolite levels in human blood. The technology that made this possible was recently developed by the Swedish company Olink. Starting with comprehensive blood protein analysis of 200 healthy individuals, it was apparent that each had its own unique protein profile, providing a fingerprint that was stable during the two years of the study. In the subsequent study, 10 000 new blood samples from patients were analyzed, covering 100 diseases and almost 100 patients per disease. Brand new preliminary data showed that there are specific differences between patients with different cancer types. The next question to explore is whether it will be possible to detect cancer-specific changes in the protein fingerprints before any symptoms occur. If so, this type of blood testing could offer a highly effective screening method for the early detection of cancer, which would be instrumental in successful treatment of most cancer types.

"Will it be possible to detect cancer-specific changes in the protein fingerprints before any symptoms occur?"

Other factors that have been important for improved cancer survival are not connected with understanding molecular mechanisms or more effective drugs but rather the processes of patient interaction and procedures for disease assessment and care. In these cases, the utilization of data is important.

Screening programs can be very effective for early detection of cancer but mass screening including all individuals of a certain age range can be problematic. The technologies used for testing may be expensive, and if the test method produces false positives it will cause unnecessary worries and anxiety. By developing smart questionnaires, only individuals who are at risk of disease are called for testing and the screening can be more effective.

FACT BOX 1 – OMICS

All living organisms are made up of cells. Our bodies are built up of hundreds of different cell types with different functions, such as skin cells that provide protection against the environment and respire water and salts when our bodies are too hot; nerve cells that transmit electrochemical impulses that make up our sensory system and enable us to think; and red blood cells that transport oxygen to all the cells in the body. However, every cell in an individual contains the same instruction manual recorded in our DNA, and the genes in our DNA are used to code for the different proteins that carry out functions in the cells. Some proteins direct the synthesis of cell membranes, some transport different substances across these membranes, while a wide range is involved in metabolism, transforming the food we eat and the oxygen we breathe into energy available for the cells.

Since the basic instructions stored in our individual DNA are the same for all cell types in our body, there are processes determining which genes should be activated and translated into proteins in each specific cell type. The processes from activation of genes to synthesis (and ultimately degradation) of proteins take place constantly in a large number of chemical reactions in which different substrates, intermediates and products are present simultaneously. Identities and levels of these molecules can be measured using various techniques. The collections of different types of molecules present in a cell, such as DNA, proteins, sugars and other metabolites as well as substances activating genes are called omes. Examples of omes are the genome (the genes), proteome (proteins), metabolome (metabolites) and epigenomes (chemical substances involved in gene activation and deactivation). The studies of these omes are called omics. Such studies provide basic information on characteristics of different cell types and how the characteristics may differ over time or between individuals, and they also show how the omes may change when the processes in the cells are affected by different diseases.

The constant development of different methods for omics has provided new understanding of human biology and mechanisms of diseases. The following paragraphs briefly describe a number of important types of omics.

Multi-omics: The different omics described here can each provide fingerprints for specific cell types but also specific, cellular processes – sometimes associated with diseases. By combining different omics methods, the biology can be studied in a concerted way. The multi-omics approach enables analysis of data from several omes at once, and advanced computational methods can be utilized to find novel associations of biological entities and new specific patterns in the omics-data associated with certain disease types. Recognition of these patterns can then be used for more precise diagnostics, which also allows better selection of optimal treatments.

Genomics: Genomics is the study of the genome, the organism's specific set of genes. It also includes the study of interactions of those genes with each other and with the environment in which the organism live. Genomes are made up of DNA (deoxyribonucleic acid) containing the instructions needed to build proteins that build structures or carry out different functions in the cells. Genome-based research has enabled extensive development including better understanding of disease mechanisms, improved diagnostics and more effective therapeutic strategies.

Transcriptomics: For the production of proteins the instructions carried in the DNA of the genes first have to be "read" and transcribed into RNA (ribonucleic acid). These gene readouts are called transcripts, and a transcriptome is a collection of all the gene readouts present in a cell at a specific time. In transcriptomics, analysis of the transcriptome shows when and where each gene is turned on or off in the cells and tissues of an organism. Researchers can thereby gain a deeper understanding of what constitutes a specific cell type and how changes in the normal level of gene activity may reflect or contribute to disease.

Epigenomics: The gene expression in cells and tissues is controlled by a multitude of chemical compounds that "tell" the genome what to do. These chemical compounds make up the epigenome, which differs in different types of cells. In the field of epigenomics, researchers aim to identify the locations and understand the functions of all these chemical tags that mark the genome. Studies also aim to understand connections between diseases and variations in the epigenome.[6]

Proteomics: Proteomics is the large-scale study of proteomes. A proteome is the set of proteins produced in a cell, organism, system, or biological context. As the gene expression differs between cell types and conditions, the proteome is not constant; it differs from cell to cell and changes over time. The proteome depends on the underlying transcriptome, but protein activity is also modulated by other factors in addition to the expression level of the relevant gene. These include rate of protein degradation and protein modifications taking place after the protein is produced.

Metabolomics: Metabolomics is the study of metabolites, which are substrates, intermediates, and products of cell metabolism. The metabolome represents the complete set of metabolites in a biological cell, tissue, organ, or organism. Different cellular metabolic processes give rise to unique "chemical fingerprints" and so metabolomics is a tool, for example to detect cellular processes connected to certain diseases.[7]

Immunohistochemistry: Immunohistochemistry (IHC) is a microscopy-based technique for visualizing cellular components, for instance proteins in tissue samples. Usually, antigens linked to enzymes or fluorescent dyes are used to mark the target proteins. IHC provides visual outputs that can reveal the existence and localization of the target-protein in the context of different cell types, biological states, and/or subcellular localization within complex tissues.[8] The method is used regularly to diagnose diseases, such as cancer, and it is also used to see the difference between different types of cancer.[9] Even though IHC is not an omics method per se it is included in this listing since it is extensively used to localize specific components in tissue and is today often used in combination with different omics methods.

[6] *National Human Genome Research Institute. (2020). Epigenomics fact sheet.*
https://www.genome.gov/about-genomics/fact-sheets/Epigenomics-Fact-Sheet (visited Sept 2022)
[7] *Clish C. B. (2015). Metabolomics: an emerging but powerful tool for precision medicine. Cold*
Spring Harbor molecular case studies, 1(1), a000588. https://doi.org/10.1101/mcs.a000588
[8] *The human Protein Atlas. Immunohistochemistry.*
https://www.proteinatlas.org/learn/method/immunohistochemistry (visited Sept 2022)
[9] *National Cancer Institute. Immunohistochemistry.*
https://www.cancer.gov/publications/dictionaries/cancer-terms/def/immunohistochemistry
(visited Sept 2022)

Epidemiology research increases our understanding of determinants of health

Epidemiology studies the distribution, patterns and determinants of health and disease conditions in populations. Access to a large amount of data is obviously a cornerstone for these studies. The field has gone through a revolutionary development, following the advance in molecular and systems biology and progression in causal inference theory. Epidemiology constantly provides new insights into how different factors affect our health, including the effect of specific interventions on the public health status.

Cecilia Magnusson, Professor of Public Health Epidemiology at Karolinska Institutet, pointed out that studies have often shown the power of preventive measures. It is clear that preventive interventions have decreased the number of deaths for several health disorders, such as heart and coronary diseases, and thereby contributing to longer life expectancy. Coronary heart disease mortality has decreased by 70 % during the last 40 years. Studies have shown that prevention accounts for 48 % of this decrease, while the contribution from medical treatment is 42 %.[10] Prevention in this case means both lifestyle changes such as less smoking, but also interventions to lower the blood pressure and to treat pre-diabetes throughout a population.
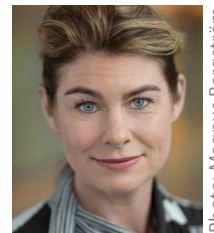
Even though preventive measures evidently are powerful, the lion's share of investment is in research and development to provide new treatments. With the models for financing healthcare systems in most of the world today, prevention is often not commercially viable. One would think that the incentives for government to work more actively on prevention should be strong, since investment in prevention will pay back in lower healthcare costs. However, preventive measures often include actions that affect living conditions and lifestyle. The introduction of requests or regulations that may affect the individual freedoms will encounter resistance and may therefore be less attractive for politicians to implement.

The links between environmental factors affecting our health (as observed in epidemiological studies) and biological mechanisms at the molecular level are often still shrouded in mystery. One field in which the understanding of molecular mechanisms is still very limited is psychiatric and neurodevelopmental disorders. At the same time, there is a growing number of people diagnosed with such disorders worldwide, and more research in these fields is necessary.

Photo: Magnus Bergström

Cecilia Magnusson

"Preventive interventions have decreased the number of deaths for several health disorders, such as heart and coronary diseases"

[10] *Ahmadi, M., & Lanphear, B. (2022). The impact of clinical and population strategies on coronary heart disease mortality: an assessment of Rose's big idea. BMC public health, 22(1), 14.*
https://doi.org/10.1186/s12889-021-12421-0

PART 2: The value of health data and challenges of sharing data

The examples of tremendous advancements in the understanding of human biology and disease mechanisms given in this report have largely been fueled by technology development in combination with sharing of data. At the same time, it is often claimed that the implementation of novel technologies in healthcare is too slow, and that the data generated in healthcare is underutilized due to challenges in sharing data. Restrictions in sharing health data and other sensitive data are necessary to protect personal integrity. However, a common view is that the regulations are no longer fit for purpose and that an updated, and less restrictive, regulation is necessary for research in order to take full advantage of modern technologies aiming to provide better treatments and care.

"It is often claimed that the implementation of novel technologies in healthcare is too slow."

Data collected within healthcare is primarily used to plan, monitor and evaluate the care of individual patients or for planning and evaluation of the way in which care is provided. The use of healthcare data for purposes such as research, product development or community planning, is called secondary use. Today, challenges usually arise when patient data or other types of sensitive data is used by other stakeholders such as university scientists, R&D-departments of pharma or technology companies, or even a different department within the same healthcare provider.

During the last ten years it has also become evident that data from sources other than the laboratory bench or the healthcare system is often valuable for understanding health-related issues. For example, during the Covid pandemic, data from mobile phones was used to analyze how people traveled and how that affected the spread of the infection.

FACT BOX 2 – REGULATION OF HEALTH DATA IN SWEDEN

Data regarding the health of individuals is classified as sensitive personal data and its handling is therefore strictly regulated. In addition to The General Data Protection Regulation (GDPR) we have several laws that apply on the national level in Sweden.

The GDPR applies to everyone who processes personal data relating to EU citizens or residents, or who offers goods or services to people in the EU. The regulations aim to protect the privacy and security of individuals, with a particular focus on personal data. Personal data is any information that relates to an individual who can be directly or indirectly identified. Names and email addresses are obviously personal data. Images, location information, ethnicity, gender, biometric data, religious beliefs, and political opinions are often classified as personal data. Even information that has been encrypted or pseudonymized but can be linked to an individual with additional information, such as an encryption key, is defined as personal data.

An important principle of the GDPR is that you may only collect personal data for specific, specifically stated and justified purposes, and the amount of data should be minimized. Only the data needed for the purposes may be processed and it should be deleted when no longer needed. The data must be protected, for example so that unauthorized persons do not gain access to it.

The purposes for collecting and handling personal data need to be stated explicitly. For example collecting data for "future research" or "improvement of the user experience" is not

allowed since the purpose is not specific enough. In addition, some personal data is defined as sensitive, and in general it is prohibited to handle such data. Sensitive personal data includes data that reveals a person's ethnic origin, political opinions, religious beliefs or trade union membership. Information about health, genetic data and biometric data is also included. Thus, data which is described in this report is normally classified as sensitive personal data. This type of data can however be processed under certain conditions. Explicit consent is one. One can also, according to GDPR, process sensitive data without consent in special circumstances, e.g. health care, social care, statistics, research and for other reasons of important public interest.

The Patient Data Act is the Swedish legislation that allows healthcare providers to process certain sensitive personal data for various purposes within their own operations without patients' consent. This may involve using the data for treatment, follow-up, quality assurance, administration and planning. The Patient Data Act also allows for different care providers to give and receive access to each other's patient data when necessary for preventing, investigating and/or treating illnesses and injuries for that specific patient. However, this exchange requires consent from the patient.

The Patient Data Act also regulates internal confidentiality within healthcare organizations, stating that only those individuals who need the data to be able to carry out individual care or in order to perform their work may have access to individual patient data. The Patient Data Act does not support the collection and handling of personal data from several different healthcare providers in the same or in different municipalities or regions, except in regional or national quality registries which are subject to specific regulation in the Patient Data Act (see below). A municipality or region may process data for follow-up and quality assurance, but only from its own organizations. The law does not support gaining access to the data of private contractors, even if those contractors are financed by the region or municipality.

The Patient Data Act further regulates how personal data may be handled for secondary use, for example in regional or national quality registers. The law gives the patients the right to receive information about the registration and the right to remove themselves from the register. Quality registers may handle information about health, such as genetic information, since it is normally considered to be part of a patient's health status in a quality assurance context and not as something deviant or separate from someone's health.

The Ethical Review Act regulates how health data may be used in research. The law applies to research involving humans and human biological material. It covers research involving both physical intervention and the processing of sensitive personal data. In order to carry out registry studies, approval is needed from the Ethical Review Board, often without the need for consent from the individuals included in the research. Anonymized data may be used for research without this approval but only if data is completely de-identified (i.e. when identification is not possible through other data or an encryption key). Disclosure of data for research also requires that the organization issuing the data conducts a risk assessment in accordance with the **Public Access to Information and Secrecy Act**. The risk assessment seeks to investigate whether the data can be disclosed without risk to security or damage to the person to whom the data relates or those close to them.

Currently, there is a governmental review of how the regulations for secondary use of health data work in practice and how they may be updated. The results from the review are to be presented in September 2023.

James Wilson, Professor of Philosophy at University College London, highlighted what he thought was the most important question to address: *How can health data be made more valuable for the purposes of clinical care, health research, and commercial companies, while maintaining high levels of public trust?* To start with, he presented a model in which the data sources were divided into three groups:

1. **Data collected by health professionals** in the course of providing healthcare, or by health researchers (such as hospital records, clinical trial data and omics data).
2. **Data self-collected by individuals** for the purposes of monitoring or enhancing their health or wellbeing (such as data from health and wellbeing apps and smartwatches).
3. **Data collected for reasons other than health** that allows inferences about personal health or wellbeing to be drawn.

James Wilson

Professor Wilson further pointed out that governments can exercise most regulatory control over the first group. Hence, this data type is most intensely discussed by governments, even though it may not be most important overall. But failure to maintain high levels of trust in the last two data types – which may be hard to control by governments – may also affect public trust in the first type of data, and hence the health care system and public bodies. Therefore, both public organizations and commercial companies are responsible for maintaining high levels of trust in the whole system.

When discussing the value of data, it is important to remember that this will be different for different stakeholders and that the whole value may not be adequately captured in financial terms. Often, there will be win-win situations. This applies, for example, when data is given with consent from patients to researchers who use the data to develop new and more effective treatments, which benefit the patients, or when hospitals curate their administrative data and provide it to companies that develop better planning tools. However, values will sometimes be competing or conflicting. The value the patient places on their medical confidentiality differ from how researchers or innovators value a large dataset. If citizens' choices are valued by giving them control over secondary uses of "their" health data, the value of the data for research or health system planning may be degraded. Professor Wilson argues that the conflict between competing values, as in the example of citizen choice and completeness of datasets, becomes more extreme if public trust diminishes.

Part 3: The matter of trust

James Wilson argued that trust among stakeholders is crucial in order to share and to fully utilize the value of health data. Sharing health data, and probably other types of personal data, can be viewed in different ways. Firstly, health data can be considered as an ordinary *market commodity* that can be bought and sold, refined and transformed. Health data may also be seen as a *gift*, given without charge, as an act of generosity that will benefit others, and sometimes also the donor in the long run. Health data may also be shared as an *act of solidarity* where the data is used for the good of society or a community.

All of these types of "transactions" have their ethical considerations and risks stakeholders' mistrust. On a commercial market there is always a risk that the seller or the buyer feel that they have been misled. Is the price fair and what do the paragraphs in the contract mean? On the other hand, if the data is given as a gift, it is most probably important for the donor that it is clear that the data is to be used for something that is beneficial for individuals or society, otherwise the willingness to donate data will be most likely to decrease over time. If data is given as an act of solidarity, there is a risk of perceived exploitation if the distribution of risks and benefits is seen to be inequitable.

"Trust among stakeholders is crucial in order to share and to fully utilize the value of health data."

How can trust be created and maintained so that people are willing to share their data for research and development? The question of trust is presumably much broader than just an issue of trust linked to the sharing of personal data. The trust for governments, healthcare systems, academic research organizations or commercial companies obviously varies between individuals and citizens of different countries, as well as subgroups within society. In Sweden, trust in science and researchers has traditionally been high. In a public survey carried out in 2018 by the organization Forska! Sverige, 95 % of respondents were positive about sharing their health data for research and purposes aiming to promote health.[11]

Public level of trust for research and new knowledge is a matter of debate. **Arne Jarrick**, Professor Emeritus of History at Stockholm University, said at the symposium: "Some friends of knowledge say that we no longer live in a knowledge society." However, he pointed out that such a statement would be hard to prove since historical data is not available and therefore any changes in the level, and distribution, of knowledge in society are impossible to identify. What history can tell us is that scientists are usually the pioneers when it comes to embracing, or at least accepting, new knowledge, and that the public follow with some delay.

Arne Jarrick

Knowledge is instrumental for a democracy to work. Åsa Wikforss, Professor of Theoretical Philosophy at Stockholm University, even claimed that "democracies require that citizens accept expertise and research" and that today we live in a very unreliable information environment. Information technology platforms such as social media allow both true facts and false claims to be spread worldwide in an instant. So, even though the amount of accessible information is greater than ever, it is probably also more difficult than ever to distinguish true from false information.

Fake 'facts' can be spread for political or commercial purposes, and it is clear that false news often spread faster, longer and recieves wider attention on social media platforms than true facts.[12] Sometimes it is even difficult to discern what science to believe. **Adam Rutherford**, Honorary Senior Research Associate at University College London, pointed out that historically there are several examples of how science has been used for political purposes, such as to legitimize racism in general and more specifically to justify eugenics programs during the first half of the 20th century. Even today and during the last two

Adam Rutherford

[11] *Research! Sweden. (2018). Healthcare Data – High Time for Action*
http://www.forskasverige.se/wp-content/uploads/SR-eng-agendarapport-varddata-2018.pdf
[12] *Vosoughi, S., Roy, D., Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151. https://www.science.org/doi/abs/10.1126/science.aap9559*

decades, alleged linkage between nationality, ethnicity and IQ received much attention,[13] even though it has been shown by several scientists that the data on which these hypothesis rely, are incomplete and the analysis is flawed.[14]

Should responsibility for distinguishing true from fake information be assigned to individuals? Will the problems be solved if people learn to think more critically? Professor Wikforss claims that this would be helpful, but that it is not enough. Studies on what is shaping our thinking and how our beliefs are formed, show that we tend to believe in 'evidence' that supports what we already believe in and that our thinking is very much dependent on our social environment.

This means that it is most likely that we will believe in the same things tomorrow as we do today and that we are likely to believe what our friends or role models tell us. When we are surrounded by a lot of disinformation (false information that is deliberately spread to deceive people) our critical thinking will be undermined. Hence, **Åsa Wikforss** suggests that stronger legislation is necessary, which means that providers of technology platforms such as social media are to a larger extent responsible for the content. In addition to what Åsa Wikforss presented, Arne Jarrick put forward two important factors leading people to dismiss new knowledge even though it is supported by strong evidence: the new knowledge may be counterintuitive, and the new knowledge may challenge your usual behavior.

Åsa Wikforss

People generating and communicating new knowledge have an important mission today. Studies on how new knowledge has been accepted by the public throughout history may provide a few hints on what is likely to work. Arne Jarrick has identified four different strategies that scientists have used to achieve acceptance of their ideas:

- **The low-key strategy.** Downplay the novelty of the idea and try not to provoke.
- **Be over-confidential** and emphasize the fantastic features of the idea.
- **Convince a specific target group.** Identify a relevant group of people and focus on getting acceptance from them.
- **Be honest.** Be transparent about what you do and what you don't know. Also show the limitations of the idea and if there are any circumstances where the idea is difficult to prove.

There is no historical evidence to show that any strategy is better than another, but it is a safe bet to suggest that you should understand your target group before choosing your strategy.

The Covid pandemic has often been called an infodemic. Many felt overwhelmed by all the available information. There was, and still is, a large amount of false information in circulation regarding how the virus is spread, the impact and adverse effects of the vaccines, and political ulterior motives associated with different strategies. A strong polarization was observed on how best to fight the pandemic, and in some cases that tipped over into very strong emotional expressions and even direct threats. Sadly, there are examples of scientists who moved away from the public debate, not daring to share their knowledge and opinions.

Fortunately, there were also outcomes that were much more positive. Firstly, it is clear that open sharing of research data on the virus and its spread resulted in an extremely rapid increase in new knowledge that led to better understanding on how to best treat the patients and the development of new vaccines at unsurpassed speed. Secondly, news coverage of the pandemic has given many people outside the scientific community a better understanding of how research works – that research is an iterative process where ideas supported by data is presented and then tested against new data. Some hypotheses are dismissed, and others are modified, strengthened and refined until there is enough evidence to call it a fact. No single data point can tell us the truth, but when the route to new knowledge is transparent, it will be more easily accepted.

[13] *Lynn, R. and Vanhanen, T. (2002). IQ and the wealth of nations. Westport, CT: Praeger. ISBN 0-275-97510-X*

[14] *Thomas Volken. (2003). IQ and the Wealth of Nations. A Critique of Richard Lynn and Tatu Vanhanen's Recent Book, European Sociological Review, 19(4), 1 September 2003, 411-412 https://doi.org/10.1093/esr/19.4.411*

Moving forward

Shifting the focus back to the use of data in life science and healthcare, it is obvious that data-driven life science is already providing new means to treat severe diseases such as cancer – and hence saves lives. To continue and accelerate this development, effective sharing and utilization of data is key.

Data sharing and utilization is a question of technology, legislations and trust. In several countries investments are ongoing to build the technological infrastructure needed for data-driven life science and precision medicine. Efforts are also made to modernize legislation in order to facilitate development while at the same time protecting personal integrity. In the EU the first steps on the road to the so-called European Health Data Space (EHDS) have already been taken. The idea is to build an EU common infrastructure for health data, including rules, common standards and practices, and a governance framework. The aims are to provide citizens with increased access to and control over health data, and to provide "a consistent, trustworthy and efficient set-up for the use of health data for research, innovation, policy-making and regulatory activities".[15]

In May 2022 the Swedish government decided to set up a review of opportunities for secondary use of health data and how legislation may be updated to directly or indirectly strengthen health care.[16]

New legislation will presumably include measures for maintaining trust among different stakeholders. Clear regulations will diminish uncertainties and the risk of conflicts. However, addressing the question of trust over and above the legislation is important. Ensuring high levels of trust in our research institutions, healthcare systems, public organizations and companies will most likely increase citizens' willingness to share health data, especially if it is shown to be beneficial for society, for other people and for themselves.

"Ensuring high levels of trust will most likely increase citizens' willingness to share health data, especially if it is shown to be beneficial for society, for other people and for themselves."

People working in, or close to, research have a responsibility to safeguard and improve levels of trust. Explaining how new knowledge is generated, being open about what you do and don´t know, as well as why you have gained the knowledge is probably a good start. Hopefully threats against some scientists during the Covid pandemic will not scare other experts away from the public debate. Instead, it is vital that they ensure that their voices are heard.

[15] *European Commission. European Health Data Space.*
https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en.
(Visited Sept 2022)
[16] *Regeringskansliet. (2022). Hälsodata som nationell resurs för framtidens hälso- och sjukvård Dir. 2022:41.*
https://www.regeringen.se/rattsliga-dokument/kommittedirektiv/2022/05/dir.-202241 (Visited Sept 2022)

Appendix: Program for symposium

The Future of Life Science: The data revolution in life science and healthcare
May 12-13, 2022

Session 1: From data to insights

The data revolution in life science and healthcare

- Mathias Uhlén: Professor of Microbiology, KTH Royal Institute of Technology

Genes, proteins, and precision medicine

- Janne Lethiö: Professor in Medical Proteomics, Karolinska Institutet

What determines our health?

- Cecilia Magnusson: Professor of Public Health Epidemiology, Karolinska Institutet

Session 2: How is data used in life science and healthcare today

Modeling infectious disease outbreaks

- Tom Britton: Professor in Mathematical Statistics, Stockholm University

Precision medicine in oncology

- Albrecht Stenzinger: Professor of Molecular Tumor Pathology, Head for the Center for Molecular Pathology, University Hospital Heidelberg

Panel discussion

- Including Anna Sandström: Head of Science Policy and Relations Europe, AstraZeneca

Session 3: The impact of health data on science and society

The value of health data

- James Wilson: Professor of Philosophy, University College London

Ai technology in healthcare - practical applications and lessons learnt

- Annalisa Pawlosky: Senior Research Scientist, Google Applied Science, Google Research & Google Cloud

The impact of data on healthcare

- Clara Hellner: Director of Research and Innovation at Region Stockholm

Session 4: Opportunities, ethics and the way science contribute to our common knowledge

Facts and opinions – what is shaping our thinking

- Åsa Wikforss: Professor of theoretical philosophy, Stockholm University
- Arne Jarrick: Professor Emeritus of History, Stockholm University

(Mis)use of science and the role of scientists in the public debate

- Adam Rutherford: geneticist, science author and broadcaster, Honorary Senior Research Associate at UCL, UK

Contact us!

Stockholm Science City works based on needs to strenghten Stockholm's competitivness within life science. We develop concepts, create relationships and spread knowledge.

Welcome to contact us:
info@ssci.se
www.ssci.se
08-522 110 40

Follow us:

ssci.se/en/newsletter

linkedin.com/company/Stockholm-Science-City

instagram.com/SthlmSciCity

facebook.com/SthlmSciCity

Ylva Williams is the CEO at Stockholm Science City

STOCKHOLM SCIENCE CITY
FOUNDATION

Stockholm Science City, Wenner-Gren Center, Sveavägen 166, 113 46 Stockholm

The financiers of the Foundation

Karolinska Institutet
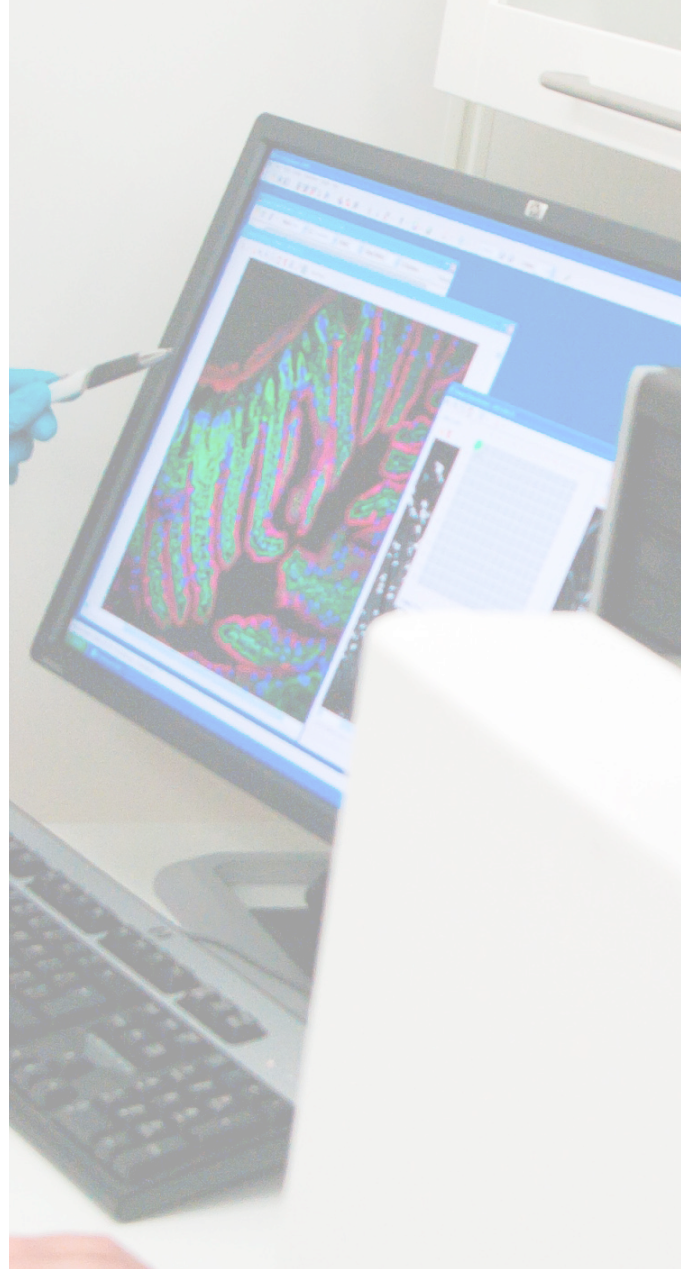
Kungl Tekniska Högskolan
VETENSKAP OCH KONST

Stockholms universitet

Stockholms stad

Region Stockholm

Photos: Staffan Eliasson

STOCKHOLM SCIENCE CITY
FOUNDATION