



This document is downloaded from the VTT's Research Information Portal https://cris.vtt.fi

VTT Technical Research Centre of Finland

Data-driven precision medicine ecosystem - stakeholder needs and opportunities

Lähteenmäki, Jaakko; Ervasti, Mari; Fagerström, Richard; van Gils, Mark; Pajula, Juha; Ruutu, Sampsa; Sigfrids, Anton; Valovirta, Ville; Ylén, Peter

Published: 29/08/2018

Document Version Publisher's final version

Link to publication

Please cite the original version:

Lähteenmäki, J., Ervasti, M., Fagerström, R., van Gils, M., Pajula, J., Ruutu, S., Sigfrids, A., Valovirta, V., & Ylén, P. (2018). *Data-driven precision medicine ecosystem - stakeholder needs and opportunities*. VTT Technical Research Centre of Finland. VTT Research Report, No. VTT-R-03318-18 http://www.vtt.fi/inf/julkaisut/muut/2018/VTT-R-03318-18.pdf



VTT http://www.vtt.fi P.O. box 1000FI-02044 VTT Finland By using VTT's Research Information Portal you are bound by the following Terms & Conditions.

I have read and I understand the following statement:

This document is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of this document is not permitted, except duplication for research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered for sale.



RESEARCH REPORT

VTT-R-03318-18 29.8.2018

Data-driven precision medicine ecosystem

stakeholder needs and opportunities

Authors:

Jaakko Lähteenmäki (editor), Mari Ervasti, Richard Fagerström, Mark van Gils, Juha Pajula, Sampsa Ruutu, Anton Sigfrids, Ville Valovirta, Peter Ylén

Confidentiality: Public



Report's title				
Data-driven precision medicine ecosystem - stakeholder needs and opportunities				
Customer, contact person	Order reference			
Outi Tuovila	3458/31/2017			
Business Finland, Pork				
Project name	Project number/Short name			
Business ecosystem fo	115511 / PreMed			
Author(s)		Pages		
Jaakko Lähteenmäki (e	editor), Mari Ervasti, Richard Fagerstrom,	74		
Ville Velovirte, Deter V	ajula, Sampsa Ruutu, Anton Sigtrids,			
Keywords	en	Report identification code		
nrocision modicino, por	reanalized medicine, secondary use	VTT P 02219 19		
artificial intelligence de	nome technology biobank business	VTT-R-03318-18		
ecosystem	chome teenhology; blobank, business			
Summary				
This report summarizes the	ne results of the first phase of VTT's PreMed p	roject. It provides an overview of		
the Finnish precision med	licine landscape and business opportunities op	ened by secondary use of health		
related data as enabled	by biobanks and other data sources. The obs	servations are based on openly		
available information and	Interviews of key stakenolders. The report des	piobanks) entities currently in		
development (e.g. the S	Service Operator and Genome Center) and i	related national activities, most		
notably the FinnGen pro	pject. It is concluded that the Finnish infras	tructure is well-developed and		
internationally competent	, but there are worries among the stakeholders	concerning the delays in setting		
up the components under	er development. Analysis of international dev	velopments reveals that a large		
access to data and proce	essing tools. Although the international marke	t potential is high most Finnish		
SME's in the precision r	nedicine domain still have difficulties in gene	erating profitable business. The		
related reasons and bottle	enecks are analysed and new business opportu	inities are analysed in the report.		
The potential of Artificial I	ntelligence is specifically addressed as a techn	ology powered by the increased		
availability of health data	for secondary use. The report provides an ar	alysis of business opportunities		
selection and direct-to-co	nsumer denomics. The analysis provides polic	ies and strategies which should		
be adopted to remove the	e main barriers. It is widely understood that su	ccessful business in data-driven		
precision medicine relies	on networked stakeholders. In order to und	erstand the mechanisms of the		
business ecosystem an ir	nitial ecosystem model has been constructed ir	the PreMed project. The model		
enables simulation of diffe	erent alternative development paths for the pre	cision medicine ecosystem. The		
to be completed by April 6	runner work planned to be carried out in the Pre 2020	eivied project, which is scheduled		
	2020.			
Confidentiality	Public			
Espoo 29.08.2018	Poviewod by	Accord by		
whiten by	Reviewed by	Accepted by		
Jaakko Lähteenmäki	Kari Kohtamäki	Jari Ahola		
Principal Scientist	Research Team Leader			
VTT's contact address				
Jaakko Lahteenmäki, jaakko.lahteenmaki@vtt.fi				
Distribution (customer and	d VTT)			
Draft for comments: Business Finland, VTT (project group), participants of PreMed interviews				
and discussions.				
Final version available	at VII website (<u>www.vtt.fi</u>).			
The use of the name of VTT Technical Research Centre of Finland Ltd in advertising or publishing of a part of this				
report is only permissible with written authorisation from VTT Technical Research Centre of Finland Ltd.				



Preface

The PreMed project is motivated by the increasing possibilities to exploit data resources in providing better health and social care. Data resources are opening the way to precision medicine, referring to customization of healthcare to the individual patient. Precision medicine opens business opportunities to companies operating especially in pharma and healthcare domains. Disease prevention is a growing application area of precision medicine and offers opportunities to companies developing products for personal wellness. It is clear that co-operation of all stakeholders including companies and public organisations is needed in order to fully exploit the potential of data-driven precision medicine.

This report includes the results of the first phase (1.5.2017-31.8.2018) of the PreMed project financed jointly by Business Finland and VTT. In order to built an overall understanding of the environment a series of stakeholder interviews and a review of precision medicine related activities at international scale has been carried out. Based on the results, opportunities for Artificial Intelligence technology in data-driven precision medicine have been identified and business opportunities have been analysed. Finally, applying system dynamics approach an initial ecosystem model has been build. The model has been used in preliminary simulation of ecosystem development scenarios.

The authors wish to thank the external experts for their contributions by participating the interviews and informal discussions and by commenting the draft version of the report. The inputs have been essential for completing the report.



Contents

Pre	Preface				
Со	Contents4				
1	Introduction				
2	Scope and definitions				
3	Hea	Ith data exploitation infrastructure	9		
	3.1	Primary data sources	9		
	3.2	Centralized data storages	9		
	3.3	Centralized services	10		
	3.4	3 4 1 General Data Protection Directive	11 11		
		3.4.2 Finnish national privacy legislation	12		
		3.4.3 Legislation on secondary use of health and wellness data	12		
		3.4.4 Biobank Act	12		
	35	3.4.5 Genome Act	13		
	3.6	Isaacus project	13		
	3.7	FinnGen project	14		
	3.8	MyData	14		
	3.9	Finnish health and social services reform	15		
	5.10		10		
4	Rev	ew of international activities	18		
	4.1	Data collection	18		
	4.Z 4.3	Current precision medicine initiatives	10		
	4.4	Overview of key data sources, storages and centralized services	20		
		4.4.1 Data access policies	22		
		4.4.2 Available data analysis tools and computing resources	23		
	45	4.4.3 Services for companies	24 24		
F			27		
Э	inter		20		
	5.1	Methods	26		
	5.2 5.3	Observations from interviews and discussions	20		
	0.0	5.3.1 Pharma and diagnostics companies	28		
		5.3.2 Research and expert services	28		
		5.3.3 Medical device manufacturers	28		
		5.3.4 Bioinformatics software and services	28 29		
		5.3.6 Genetic test interpretation and counselling services	29		
		5.3.7 Healthcare service providers	29		
		5.3.8 Healthcare information systems and applications	29		
		5.3.9 Insurance services	30		
	5.4	Bottlenecks and challenges	30		
	5.5	FinnGen opportunity	33		
	5.6	Stakeholder views on secondary data use	33		
6	Com	pany perspective to data exploitation	34		
	6.1	Operating environment	34		
	6.2	Data exploitation roles	35		



		6.2.1 Data controllers and brokers	35			
		6.2.3 Research service providers	36			
		6.2.4 Service and technology providers	36			
	6.3	GDPR considerations	37			
7	Artifi	cial Intelligence Opportunity	38			
	7.1	Artificial Intelligence use cases	38			
	7.2	Potential technologies	39			
	7.3	Future opportunities	41			
8	Busi	ness strategies	43			
	8.1	Development of new therapies	44			
	8.2	Health risk assessment and diagnostics	47			
	8.3 8.4	Direct-to-consumer generation	48 50			
~	0.4 East		- 1			
9	ECOS	system model	54			
	9.1	System dynamic modelling	54			
	9.2	Simulation results	54			
40	9.5		50			
10	10 Conclusions					
Re	References61					
Те	Terms and acronyms6					
Ap	Appendix 1. Precision medicine initiatives65					
Ap	Appendix 2. Non-profit precision medicine organisations69					
Ap	Appendix 3. Data access, information sharing and collaboration structures					
Ap	Appendix 4. Data access guidance comparison72					
Ap	ppendix 5. Experts involved in interviews and informal discussions					
Ap	Appendix 6. Interview questions75					



1 Introduction

Personal data is increasingly collected and used in the context of digitalized services. Consequently, the amount of personal data stored in information systems is increasing exponentially. Besides the primary purpose, collected personal data is used for so called secondary purposes, such as for monitoring the quality of provided services, in the development of new services or products and for scientific research. In general, secondary use is permitted under certain conditions by the existing privacy legislation, in particular the General Data Protection Regulation (GDPR). Additionally, specific legislation addressing secondary use of health data (e.g. in the context of biobanks) has been implemented in some countries, including Finland^{1,2}.

In healthcare, exploitation of data enables *precision medicine*³, referring to customization of healthcare, with medical decisions, treatments, practices, or products tailored to the individual patient. This is expected to be cost effective in comparison to the prevailing "one-size-fits-all approach" as interventions can be targeted to those individuals, who really benefit from them. Precision medicine is often based on genetic, molecular and cellular analysis of the individual and involves techniques such as molecular diagnostics, genetic sequencing, imaging and data analytics. Precision medicine is increasingly data-driven by integrating multi-omics data, clinical data and other real-world data influencing the health or the disease state of the individual.

Precision medicine is expected to have a remarkable role especially in cancer care as tumours have particular genetic signatures correlating with their response to treatments. Genetic tests can help to decide which treatments a patient's tumour is most likely to respond to, sparing the patient from receiving treatments that are not likely to help. Benefits are foreseen also in other clinical domains. Metabolomics of certain drugs has been shown to depend on certain gene variants, which means that information of these variants can be used when selecting the drug to be used in a particular case. Furthermore, the personal risk of many common diseases, including many cancers and cardiovascular diseases as well as autoimmune diseases, depends on the genome. Knowing these risks through genetic tests is helpful in identifying and motivating individuals to preventive interventions.

Data-driven precision medicine applications in healthcare and pharma domains are based on sophisticated analysis of data and increasingly exploiting Artificial Intelligence (AI) technologies. Examples in healthcare include advanced data-driven Clinical Decision Support systems (CDSS), automation of customer service processes by chatbots and support of self-care by automatic feedback to patients.

For the pharmaceutical product development data-driven precision medicine opens new opportunities in several phases of drug development. In the preclinical phase genome and other individual level data provides understanding of disease mechanisms and helps in identifying the drug targets and molecules. The availability of individual-level health data is highly valuable for the clinical trials and for Real World Evidence (RWE) studies.

The recently started FinnGen project⁴ is expected to have a big positive impact on precision medicine in growing a national reserve of genome data and in developing the required infrastructure and competences needed to exploit it. In the framework of the project bio samples of 500 000 Finns will be collected and combined with clinical registry data. The data and samples will be made available to biobank research during and after the project.

¹ <u>https://media.sitra.fi/julkaisut/Muut/International_review_secondary_use_health_data.pdf</u>

² <u>https://www.biopankki.fi/en/laws-and-regulation/</u>

³ https://en.wikipedia.org/wiki/Precision_medicine

⁴ <u>https://www.finngen.fi/</u>



Precision medicine has a remarkable business potential. The PreMed project (Business ecosystem for Finnish precision medicine platform) is targeted towards identifying and increasing the business opportunities for companies based on data-driven precision medicine. In particular, the project aims to create an ecosystem model describing how various types of companies and public organizations can co-operate and make business.

This report includes the results of the first year of the PreMed project financed jointly by Business Finland and VTT. In order to built an overall understanding of the environment a series of stakeholder interviews and a review of precision medicine related activities at international scale has been carried out. Based on the results, opportunities for AI technology in data-driven precision medicine have been identified and potential business strategies have been analysed. Finally, applying system dynamics approach an initial ecosystem model has been build. The model has been used in preliminary simulation of ecosystem development scenarios.

The report forms the basis for further work to be carried out in the PreMed project scheduled to be completed by April 2020. The detailed plan for further work will be compiled in cooperation with participating companies. Participation is open to all interested stakeholders whether or not they were involved in the project during the first year.



8 (75)

2 Scope and definitions

Precision medicine' is a broad term and lacks a clear and stabilized definition¹. It is often used interchangeably with *'personalized medicine'* (Nimmesgern et al. 2017). The scope of the report is *data-driven precision medicine*, which emphasizes the role of individual-level health data in increasing our understanding of human health and behaviour as a whole and thereby opening new opportunities for new products and services. In particular, we will look at the opportunities offered by *secondary use*² referring to applications where data is used beyond its primary purpose (Danciu 2014). We especially look at secondary use enabled by public, centralized data sources.

The use of biobank materials and data is in this report considered to fall to the secondary use domain. Even if samples and data are in many cases collected specifically for the purpose of biobank research, still a substantial part of the collections originate from normal healthcare processes and are transferred to biobanks for further (secondary) use.

In this report, we consider widely the exploitation of any kind of health and wellness related data - including both clinical and non-clinical data. For example, a person's use of social services provides useful information related to holistic wellbeing of the individual. For practical reasons we have used the term *health data* to refer to any health *related* data. Our main focus is in exploiting *individual-level* data as opposed to population-level aggregated data.

¹ <u>https://ghr.nlm.nih.gov/primer/precisionmedicine/precisionvspersonalized</u>

² Deloitte: International review on the secondary use of health and social care data and applicable legislation



3 Health data exploitation infrastructure

The Finnish infrastructure related to secondary use of health data is depicted in Figure 1. The figure shows data sources on the left and data users on the right. In between, centralized data storages and services support collection of data and making it available for data users. Components of the infrastructure are further described in the sections below.



Figure 1. Finnish health data exploitation infrastructure

3.1 Primary data sources

The main data sources shown in the left side of Figure 1 fall into four main groups. Patient and customer registers refer to EHR's (Electronic Health Records) and other operational information resources used in the provision of health and social care services. Other online services refer for example to wellness applications, which are not part of the healthcare system, but independently used by the individual. Research registers are collected for the purposes of various types of research studies. Depending on the obtained user consent, research registers may be maintained beyond their use in the original study and can be used later for other studies. Biobanks collect bio samples and related clinical information under the Biobank Act¹ for the purpose of clinical and population research.

3.2 Centralized data storages

Centralized data storages combine information from local and regional systems. Kanta services² include a national EHR archive, which enables accessibility to patient data and social care customer data by multiple organisations under consent by the individual and as needed in service processes (Hyppönen 2015). All healthcare service providers are required by

¹ <u>http://www.finlex.fi/en/laki/kaannokset/2012/en20120688</u>

² https://www.kanta.fi/en/system-developers/kanta-architecture



legislation¹ to connect their EHR systems to the Kanta archive and to transfer patient data to the archive for all healthcare service events. Besides healthcare, gradually, also social services events will be covered by the Kanta system². Kanta also includes a service (MyKanta) which allows citizens to view their own clinical data, provide their consent to share documents between registrars, view data usage logs and set restrictions on data usage. MyKanta is currently being extended in functionality to also support insertion of own health data manually or via personal devices. This functionality is supported by an open API (application programming interface), which allows certified applications to be connected with MyKanta and to exchange personal health data under consent of the individual.

Another centralized data storage will be the genome database, which is to be operated by the national Genome Center³ currently in preparation. When ready, the genome database will provide a secure storage location for all genome data. This will considerably improve the availability of genome data both for healthcare processes and for research purposes.

National statistical registers collect data needed essentially for statistics and research (Gissler 2004). Many of the registers are based on national legislation. These registers are hosted by:

- National Institute for Health and Welfare (THL)
- The Social Insurance Institution of Finland (Kela)
- Population Register Centre (VRK)
- Finnish Medicines Agency (Fimea)
- Finnish Institute of Occupational Health (FIOH)
- Statistics Finland
- Tax Administration

THL maintains several registers needed for providing statistics on the Finnish healthcare systems. For example, the Care Register for Healthcare collects data on the activities of health centres, hospitals and other institutions providing inpatient care and on the clients treated in them as well as on home-nursing clients for the purposes of statistics, research and planning. **Kela** maintains several registers e.g. with information on the use of social benefits and purchase of reimbursed medicines. **VRK** maintains the population information system containing basic information (e.g. name, date of birth, native language, address and family relation) of Finnish citizens and foreign citizens residing permanently in Finland. **Fimea** maintains registers on biological and carcinogenic exposure as well as several research registers with data on occupational health and wellbeing. **Statistics Finland** provides a wide range of statistics and information services. In health research domain, e.g. the *Causes of death* register is widely used. The **Tax administration** maintains for example the National Incomes Register.

3.3 Centralized services

Centralised services for supporting secondary use of data are emerging rapidly. Especially important will be the Service Operator, which aims to be a one-stop-shop where health data can be accessed and used in a secure environment. The operator is planned to maintain all related service processes and components, including:

³ <u>http://stm.fi/en/artikkeli/-/asset_publisher/genomikeskuksen-perustaminen-etenee-tyoryhman-arviomuistio-lausunnoille</u>

¹ <u>https://www.finlex.fi/fi/laki/ajantasa/2007/20070159</u> (in Finnish)

² https://www.kanta.fi/en/professionals/client-data-archive-for-social-welfare-services



- Authentication and authorization
- Data search and discovery
- Data access application and acceptance processes
- Secure, remote data processing environment
- User support

The Service Operator concept was initially developed in the framework of Sitra's (The Finnish Innovation Fund)¹ Isaacus project (Section 3.6). The Service Operator activity will be enabled by a new law in preparation (Section 3.4.3). The law defines THL as the legal authority responsible for accepting data access applications as well as organizing the set-up and maintenance of the Service Operator processes. The operative function of the Service Operator is estimated by THL to start by the end of 2019, but the schedule is still not confirmed and depends on the timing of the Secondary Use Act coming into effect.

The Finnish biobank co-operative (FINBB) created in 2017 is aimed at providing a single channel to the Finnish biobank resources². The organisation currently has an on-going project for developing an availability database, which enables biobank customers to find information about available biobank resources. FINBB also provides legal, commercial and communication support for its members.

3.4 Legislation

The use of health and social services data is governed by several national laws and EU regulation. Complete description of this legislative framework is outside the scope of this report. Below, we have briefly described the main legislation from the point of view of secondary use of health data.

3.4.1 General Data Protection Directive

The General Data Protection Regulation (GDPR) (EU) 2016/679 is a regulation in EU law on data protection and privacy for all individuals within the European Union. The GDPR substantially increases the obligations of data controllers in ensuring the privacy of data subjects. While increasing obligations to data controllers the GDPR also extends the opportunities for secondary use of data. The lawfulness of data processing for secondary purposes without explicit consent is based on Article 6 (paragraph 1, point e): 'processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller'. Article 9 limits the use of 'special categories of personal data' such as data concerning health, but enables the use of data for various secondary purposes. These include management of health and social care systems and services, public health, ensuring healthcare and medicinal products or medical devices, archiving, statistical purposes and scientific or historical research. Usage of data for such purposes is subject to several measures to be taken by the data processor to protect the data subject's privacy and rights.

The GDPR recital 159 includes a statement that 'scientific research' should be interpreted in a 'broad manner', including 'for example technological development and demonstration, fundamental research, applied research and privately funded research', as well as 'studies conducted in the public interest in the area of public health'. This definition enlarges the potential for data exploitation. The Finnish Personal Data Act (523/1999) contains provisions for usage of data in scientific research and statistics, but not for 'technological development and demonstration'.

¹ <u>https://www.sitra.fi/en/</u>

² <u>https://finbb.fi/</u> (in Finnish)



Besides direct impact on data exploitation, the GDPR may also have indirect favourable effects. It strengthens the data subject's rights to obtain copies of his/her own data from the data controller. In particular, the data controller shall provide data for the individual in 'commonly used electronic form' (GDPR/article 15). This will considerably increase the possibility for the individual to reuse personal data and to share it with trusted parties, e.g. to support research and development activities.

3.4.2 Finnish national privacy legislation

The Personal Data Act (523/1999)¹ nationally implements the regulation outlined in the EU Data Protection Directive (Directive 95/46/EC). Since May 2018 the GDPR has been the primary privacy legislation in all EU countries. However, national supplements to the GDPR will be allowed. For Finland, such complementary regulation is currently being processed by the Parliament of Finland² and is expected to be in force during autumn 2018. As relevant for secondary use of health data, the government's bill for new privacy legislation complements the GDPR concerning legal basis of using personal data and the specific requirements pertaining to the use of data for scientific and statistical research.

3.4.3 Legislation on secondary use of health and wellness data

This law, referred in this report as *Secondary Use Act*, regulates the use of health and wellness data for secondary purposes³. It is currently in the acceptance process by the Parliament of Finland. As currently estimated, the Secondary Use Act will be in force by the end of 2018. It is closely related to the Biobank Act, but addresses essentially data resources instead of biological samples. The objective of the Secondary Use Act is to enable efficient and controlled access of data resources for legitimate exploitation, while maintaining the rights and privacy of the data subject. A central component of the law addresses the legal entity which will be in charge of granting access to data for data users and of establishing and running the related processes. According to the bill the legal entity is the National Institute of Health and Welfare (THL), which may outsource the operational processes to be managed by an external organization, referred in this report as the *Service Operator*.

According to the Secondary Use Act specific consent from the data subject will not be mandatory when the data is used for scientific or statistical research, which enables the traditional 'registry based research' approach. Usage of data for technical development and innovation will be enabled by the law only under specific consent by the data subject or in case the data set is fully anonymised. The legislation does not explicitly define the concepts 'scientific research' and 'technical development and innovation'. Consequently, the judgement of the categorisation will need to be made case-by-case by the respective data controller.

3.4.4 Biobank Act

The Finnish Biobank Act entered into force in 2013. The act provides the legal framework in which biobanks operate (Soini 2016). It defines Valvira (National Supervisory Authority for Welfare and Health) to supervise the activities of the Finnish biobanks. In particular, the objective of the Biobank Act is to ensure the rights of the donors and to protect their privacy.

The Biobank Act is currently being renewed in order to better align it with GDPR and the national legislation on secondary use of health and wellness data. The new Biobank Act will

¹ <u>https://www.finlex.fi/fi/laki/ajantasa/1999/19990523#L6P27</u> (in Finnish)

² <u>https://www.finlex.fi/fi/esitykset/he/2018/20180009#idp451117520</u> (in Finnish)

³ <u>https://www.finlex.fi/fi/esitykset/he/2017/20170159</u> (in Finnish)



define in more detail the conditions for using data for technical development and innovation purposes.

3.4.5 Genome Act

The Genome Act is currently in the acceptance process by the Parliament of Finland¹. The law will regulate the usage and management of genome data. In particular, it will regulate on the operation on the National Genome Center to be established under THL as the legal authority. The genome centre will promote the use of genetic information in healthcare, research and innovation activities in the health sector. The Genome Act is expected to come into force on 1.1.2019.

3.5 National strategies

The value of healthcare data has been recognized in several national strategies.

The Finnish National eHealth and eSocial Strategy 2020^{2,3} by the Ministry of Social Affairs and Health (STM) highlights the potential of data in supporting research, innovation, industrial and commercial activities. It brings up the need for legislation to enable secondary use of social welfare and healthcare data.

The Health Sector Growth Strategy⁴ published by the Ministry of Employment and Economy provides recommendations for systematic development of research and innovation activities in order to achieve economic growth in the healthcare sector. The strategy report recommends seamless joint access to personal health data and patient documents to be enabled for research purposes.

The National Genome Strategy⁵ recommends the establishment of a national genome centre and the development of the related legal framework. The genome center would bring together all parties in the field of genomics and it would have the responsibility of developing and maintaining a reference database of genomes for the needs of clinical care and research. Following the recommendations of the Genome Strategy a specific legislation for governance of genome data is currently in preparation⁶.

3.6 Isaacus project

The Isaacus project⁷ was started by Sitra in 2015 with the objective of setting up an organisation to take responsibility of providing access for secondary use of health and wellness data. The activity was carried out in parallel with the preparation of the corresponding legislation. The Isaacus project included a set of pre-commercial development activities chartered to develop components needed by the Service Operator. The components created as output from Isaacus were:

¹ <u>Government bill for Genome law</u> (in Finnish)

² <u>https://pdfs.semanticscholar.org/presentation/4c8d/ed88912008cfe7a6bd791e4599a764d29d2a.pdf</u>

³ <u>http://stm.fi/julkaisu?pubid=URN:ISBN:978-952-00-3548-8 (in Finnish)</u>

⁴ Health Sector Growth Strategy

⁵ National Genome Strategy

⁶ <u>https://stm.fi/en/artikkeli/-/asset_publisher/genomikeskuksen-perustaminen-etenee-tyoryhman-</u> arviomuistio-lausunnoille

⁷ https://www.sitra.fi/artikkelit/isaacus-national-health-data-hub/



14 (75)

- A common metadata model for describing data sets to be searchable via a centralized service (activity led by THL)
- A secure remote environment for data processing (activity led by Statistics Finland)
- Common processes and tools for biobanks for preparing data sets for research use (activity led by BBMRI)
- A 'one-stop shop' permission service where researchers can apply access to data resources (activity led by National Archives of Finland)
- Data lakes gathering patient data to be accessible for secondary uses (activities of two hospital districts HUS and VSSHP and the City of Kuopio)

The Isaacus project was running until mid 2018. During the last year Isaacus has been in the mode of preparing for the operational phase to be led by THL under the Secondary Use Act. The objective is that the components developed by Isaacus could be largely exploited in the operational phase of the Service Operator.

3.7 FinnGen project

FinnGen is a large precision medicine project launched in Finland during 2017. It is a unique study combining genome information with healthcare data. FinnGen differs from many other initiatives in collecting a majority of the samples from diseased individuals (with a diagnosed disease) using the Finnish hospital biobank network. This is imperative when studying the genetic links to diseases. The project aims to improve human health through genetic research and to identify new therapeutic targets and diagnostics for treating a large variety of diseases. In the framework of the project, blood samples of 500 000 Finns will be sequenced. The samples will be collected by the Finnish biobanks. Currently, already 180 000 samples have been collected. The project is lead by the University of Helsinki. It is funded together by Business Finland and seven international pharmaceutical companies: Abbvie, AstraZeneca, Biogen, Celgene, Genentech (a member of the Roche Group), Merck & Co. and Pfizer.

During the project, the collected samples and corresponding phenotype data will become available for biobank research. Thereby, the project is expected to have a remarkable impact in stimulating genome-based research in Finland and internationally. Besides ambitious scientific research, the project specifically addresses also needs of businesses and aims at involving SME companies in joint add-on projects.

3.8 MyData

MyData refers to the paradigm of enabling the individual to have full control over his/her personal data. The MyData approach has gained considerable interest during recent years and it has been followed by many services, e.g. by exposing API's which allow user data to be retrieved in structured form enabling the data to be used in other applications under consent by the data subject. The GDPR may reinforce this trend, as it explicitly requires service providers to release data to their customers in open reusable formats.

Despite of the favourable trend, a viable, common MyData architecture has not yet been achieved although several related studies and preliminary specifications have been published and some early concepts have been presented. Interoperable Personal Health Records (PHR's) (Li 2017) can be considered as first MyData implementations. The idea of an interoperable PHR is to provide a user-controlled data repository where service providers can store personal data in a common format. The Taltioni service was the first implementation of this model in Finland, but did not meet its business goals and is now terminated. The Kanta





PHR¹ being developed is essentially based on the same model, but has better chances to succeed as it is a part of the national healthcare infrastructure, potentially enabling it to connect with EHR systems.

The MyData architecture designed and documented in the DHR² (Digital Health Revolution) project is based on a centralized consent management enabling the individual to share his/her data between application without implementing any centralized database³. The IHAN (International Human Account Number)⁴ system proposed by Sitra is another initiative targeted to enable the individuals to be in control of their data. In particular IHAN aims at an ecosystem with common rules and concepts for data exchange. The MyData model is highly interesting from the PreMed project perspective as it would enable a convenient way to obtain the data subject's consent for using the data in the cases where direct secondary use of personal data is not permitted.

3.9 Finnish health and social services reform

A major health and social services reform is being implemented in Finland⁵. A key objective of the reform is to assign the responsibility for providing public healthcare and social services to 18 autonomous regions. These regions are considerably larger than municipalities, which currently are responsible for arranging the services. An important long-term objective is to achieve considerable savings through new operating models and efficient processes.

The reform will be accompanied by considerable investments in ICT systems. In addition to the basic functionalities, such as care documentation in EHR systems, there will be increased need for monitoring the quality and efficiency of service processes. This need is reinforced due to the fact that the healthcare customer will be given more freedom in selecting the healthcare provider. Both public and private social and health services centres will be available to choose from. In order to use the freedom to choose, the customer needs clearly expressed information on the quality of the available services. Graphical summary information can be created by software applications which can access a wide range of up-to-date data. Consequently, the health and social services reform is likely to increase the need for secondary use of data.

3.10 Data users

Data users are entities which directly or indirectly benefit from the secondary use of data. Direct data users access original individual level data. Indirect data users refer to entities, which cannot or do not want to access individual level data directly, but benefit from data analytics results and aggregated summary data provided for them.

Universities and research institutes have a long tradition of secondary use of data. The use of registry data is a cost-efficient alternative compared to experimental studies and enables very large populations to be studied. Usage of registry data for scientific research is clearly covered and supported by the corresponding laws, including the GDPR.

Public and private health service providers are increasingly showing interest towards secondary use of patient data - not only for traditional academic research, but increasingly also for monitoring and improving healthcare processes and quality. Analysis of patient data would also be useful when selecting patients for targeted preventive interventions. Moreover, it enables data-driven clinical decision support systems (CDSS), where patient's health parameters are compared to similar reference patients. Development of such systems requires

¹ <u>https://www.kanta.fi/en/web/guest/professionals/kanta-phr</u>

² <u>http://www.digitalhealthrevolution.fi/</u>

³ http://hiit.github.io/mydata-stack/

⁴ https://www.sitra.fi/en/topics/human-driven-data-economy/#what-is-it-about

⁵ http://alueuudistus.fi/en/social-welfare-and-health-care-reform/about-the-reform



secondary use of data even if the actual CDSS system would be used only in the context of patient care and thus categorized as *primary use* of data.

Public authorities are an important data user group. It is commonly accepted that public investments and political decisions should be based on knowledge and evidence, which in most cases requires real-world data analytics. National statistical registers have been available for such use for long, but there is a need to access more detailed data and data in a wider scope. There is also a need for easy-to-use and understandable user interfaces providing actionable information for the users, as is the case in the MIDAS project (Meaningful Integration of Data, Analytics and Services)¹.

Citizens benefit from secondary data indirectly in the form of services provided for them. Healthcare statistics are publicly available and can be accessed by individual citizens. More typically, citizens consume such information in the form of newspaper articles and other media content prepared by journalists and bloggers.

Several types of companies are potential direct or indirect data users. We have identified the following groups:

- Pharma and diagnostics
- Research and expert services
- Bioinformatics software and services
- Sequencing, genotyping and diagnostics services
- Genetic test interpretation and counselling services
- Medical device manufacturers
- Healthcare information system and application providers
- Healthcare services
- Insurance services
- Food and nutrition

Pharma and diagnostics companies have a long tradition of using patient data under patient consent in various phases of drug development and in Real-World Evidence (RWE) studies. Most typically pharma companies subcontract these studies to be carried out by external **research and expert services** companies such as contract research organisations (CRO's).

Bioinformatics software and services providers are typically SME's, which have developed software for genome or other omics data analytics and offer related services for their customers, such as pharma companies or research service providers.

Sequencing genotyping and diagnostics providers are companies with laboratory facilities and capability to provide genetic diagnostic tests and genome sequencing services.

Genetic test interpretation and counselling services refer to various types of services supporting the exploitation of the raw genome data in healthcare e.g. as in pharmacogenomics (business-to-business, B2B) or independently by the consumer to support preventive self-care (business-to-consumer, B2C).

Medical device manufacturers are potential users of health data to support their R&D activities. In a typical case, the device manufacturer is interested in collecting data from the use of their existing products. External data resources could potentially be used to support the R&D towards new products.

¹ <u>http://www.midasproject.eu/</u>



Healthcare information systems and applications are increasingly including advanced decision support features providing guidance to clinicians and patients. Developing, tuning and further development of such features requires individual level health data to be available.

Healthcare services companies - similarly to public healthcare providers - need data for monitoring and improving healthcare processes and quality.

Insurance services have strong interest in health data. Ethical issues and the fear of loosing customers prevent insurance companies from using individual level health data. However, insurance companies can indirectly benefit from aggregated health data, e.g. in getting a better understanding of the mechanisms leading to disablement, and of the related preventive interventions.

Food and nutrition companies need data on citizen's lifestyles and health e.g. in the development of products for personalized nutrition. Besides targeted questionnaires, the public data resources collected e.g. by THL are potentially attractive for the R&D activities.



4 Review of international activities

In this Chapter, we provide a global review on national initiatives and organisations in the field of precision medicine (PM). As precision medicine ecosystems largely rely on exploitation of genomic and health data, we introduce internationally available genomics and health data sources, centralized data storages, data services, and related resources. We also provide a brief outlook on the overall market size.

As the constantly declining costs of genome sequencing make widespread application of genetics increasingly plausible¹, its business potential and promises of better healthcare become more apparent. Recently governments world-wide have implemented major initiatives to facilitate national growth of genomic research, markets, and clinical implementation of PM innovations. Alongside governments, research institutes, industry and collaborative organizations intend to share, harmonize and consolidate genomics and health data by establishing biobanks and databanks, data infrastructures, data standards and computing services. It is becoming increasingly evident that efforts towards precision medicine marks a key global development in several omics, data and healthcare related sectors.

4.1 Data collection

In order to map the precision medicine landscape, we made a review to locate recent global initiatives, major national and international organisations, and commercial companies that promote collaboration and data sharing in the precision medicine domain. We aim to demonstrate prominent, up-to-date and recent precision medicine efforts in Europe and globally. We have included a variety of efforts such as research projects, collaborator organisations, investments to build data infrastructures and databases, and strategic plans to support genomic science, innovation, and economy. Precision medicine related projects were searched by combining keywords such as 'precision', 'personalized', 'genomic', 'initiative', 'medicine' and names of countries or continents. The initial findings were used to locate references to major organisations not found in the primary searches. We narrowed the findings by including only up-to-date organisations, and excluded initiatives with outdated online information. To locate commercial companies, we used the member-lists of two major precision medicine associations, one in Europe and one in the USA.

4.2 Market outlook

There is a variety of commercial organisations providing services and products to the precision medicine markets. For example, current members of the *Personalized Medicine Coalition* and *The European Personalised Medicine Association*, include a number of commercial clinical laboratory testing services, diagnostic companies, biotechnology and pharmaceutical companies, healthcare insurance companies, IT and informatics companies, service providers, research tool companies and venture capital investors².

The commercial services within data-driven precision medicine range from analysing data, to generating data platforms and cloud services, and to utilizing the data in services targeted for the individual consumer. DNA sequencing is often outsourced to specialized companies, as is in the case of Genomics England, which uses Illumina to sequence the collected DNA samples in the 100 000 genomes project. Large-scale computing services provide platform services, as in the case of Amazon Web Services that hosts a precision medicine platform for the American Heart Association. Data analysis and informatics companies such as SevenBridges or Syapse provide not only platform services, but also detailed analytics and infrastructure services to businesses. Potentially disruptive companies have emerged and might emerge in areas targeting individual consumers, as in the case of 23andMe or FitnessGenes that base their

¹ <u>https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/</u>

² <u>http://www.personalizedmedicinecoalition.org/Members/Current_Members</u>



product on interpreting individual DNA samples to provide information on lifestyle and health risks.

Market research organisations estimate a significant growth in overall precision medicine markets for the coming years. Frost and Sullivan estimate that global precision medicine markets will grow from 38 billion dollars in 2015 to 134 billion dollars in 2025 (Frost&Sullivan 2017a), and that oncology related markets will cover 40-50% of the market (Frost&Sullivan 2017b). Roughly similar estimates are made by other market research companies, estimating that the compound annual growth rate of global precision medicine markets will be between 9,7 % to 14 % in the coming seven years¹.

4.3 Current precision medicine initiatives

This section presents the aims and goals, estimated funding, and sequencing targets of a number of key initiatives, strategies and research projects in the field of precision medicine. We list national and academic projects currently actual in several countries throughout the world. We found that the surveyed initiatives have roughly similar overall aims and goals, but the sequencing targets and economic investments vary significantly. A list of the initiatives and collaborative organisations are found in Appendices 1 and 2 respectively.

Precision or genomic medicine initiatives, research programs and strategies are currently in progress on a large scale in several countries. In the last year alone, new initiatives in Europe have been launched in Denmark, Finland, Sweden, and Switzerland. Since 2015, China, Japan, South Korea, New Zealand and the US have initiated similar efforts. Earlier started projects include the international H3Africa project, which launched in 2010, and the Estonian Biobank which has been collecting samples for almost 20 years.

Funding of national projects vary from initial investments of a few millions to more ambitious and long-term investments counted in hundreds of millions. For example, Denmark has made an initial investment of 13 M€ for their national PM strategy during 2017-2022, and Switzerland will allocate around 60 M€ for PM development projects during 2017-2020 through their Personalized Health Network. In comparison, South Korea invests 48 M€ in a single cohort study aiming to sequence 10 000 cancer patients. More ambitious investments ranging from 250 to 670 million euros have been made in US, France and UK to support national PM initiatives. These are large scale sequencing projects that aim to sequence from 100 000 to 1 million individuals. China is in unique in comparison, as the government aims to support precision medicine projects nationally with about 8000 M€ in order to sequence up to 100 million genomes in the population in the coming 15 years.

The data collection methods and amount of genome sequencing plans in research projects and population sequencing initiatives vary in a similar manner. Genomic Medicine Sweden aims to sequence 25 000 genomes annually in the future, while the French aim at 235 000 annual genomes. The FinnGen research project uses samples collected by Finnish biobanks and aims by 2023 at a total of 500 000 sequenced genomes combined with phenotype data available from national registries. Genomics England plan to sequence 100 000 genomes related to 75 000 individuals, while also collecting lifestyle and health data by using electronic health records. While the emphasis in data collection is often in understanding the genomic variation and utilization of genomic knowledge in healthcare, national strategies also highlight the use of various data sources, big data and data sharing as a means to advancing the main goals of precision and preventive healthcare. Examples of research projects using various data

¹ Frost and Sullivan (2017a) estimate a CAGR of 13,3 % during 2015 to 2025. Mordor Intelligence estimate that the precision medicine markets will grow at a CAGR of 9,7 % during 2018-2023. <u>https://www.marketresearchfuture.com/reports/precision-medicine-market-925</u> Market Research Future estimate a CAGR of 12,48 % between 2015-2022 <u>https://www.mordorintelligence.com/industry-reports/precision-medicine-market</u>

Persistence Market Research predict a CAGR of 14 % from 2016 to 2024 https://www.persistencemarketresearch.com/market-research/precision-medicine-market.asp





is FinnGen, Estonian Genome Project or Genomics England, all of which will produce extensive and longitudinal databases that combine genomic, lifestyle and health data.

Despite of organisational differences between the type of the initiatives (e.g. national strategy, biobank, research project), the objectives and aims of each initiative are roughly similar. Many of the initiatives aim for scientific advancement and practical implementation of personalized medicine through supporting national research infrastructures, collaborations, data sharing and data harmonization nationally and globally, and kick-starting precision medicine innovations and markets. These goals are similar to what Dubow and Marjanovic (2016) found in their study comparing 30 national sequencing initiatives.

The recent years have also seen global collaborations and advocacy groups emerge (Appendix 2). For example, the Global Genomic Medicine Collaborative and the International Consortium for Personalised Medicine (ICPerMed) were established in 2016. Older collaborations around precision medicine include the Personalized Medicine Coalition (2015) and the European Personalised Medicine Association (2009). These organisations are formed to create and support networks and collaborations, raise awareness and highlight benefits of PM in order to accelerate its adoption and development. A more data driven approach is offered by the Global Alliance for Genomics and Health that promotes responsible genomics data sharing by developing technical standards and framing related policies. They offer free toolkits for sharing genomics data according to ethical and regulatory frameworks, data security and genomic data analysis.

4.4 Overview of key data sources, storages and centralized services

In this section we introduce widely available genomic data sources, data tools, structures and services. We surveyed websites and documents from precision medicine related organisations to map resources available for commercial and other non-academic actors in precision medicine (see Table 1 and Appendices 3 and 4). We also studied how organisations explicitly state to support SMEs, the industry or otherwise commercial use.

Table 1 includes a list of organisations that provide different kinds of health or genomics data related services. We found 3 types of key resources. These include *data access* and restrictions by which access is controlled; *computing and analysis* resources such as tools and testing environments; and *other resources*, such as education, industry support strategies including specialized SME programs, or forums to facilitate commercial collaboration.



Project	Data access			Tools		Other resources			
	open	restric ted	access fee	links to data	data analysis tools	testing environ- ments	industry support program	training or education	collabo- ration resources
The Estonian Genome Center		х	х		х				
UK Biobank		Х	Х						
Canadian Open Genetics Repository (COGR)	х								
Hartwig medical foundation (NL)		х	Х						
European Genome Phenome Archive (EGA)		х							
International Cancer Genome Consortium (ICGC)	х	х			х	х			
The National cancer Institute (NCI), Genomic Data Commons (GDC) (USA)	х	х		х	х	х			
European Bioinformatics Institute (EMBL- EBI)	х	х		х	х		х	Х	х
PrecisionFDA (USA)		х			х	х			х
AHA Precision Medicine Platform (USA)		х	х		х	х			х
The National Genomics Infrastructure (NGI) (SWE)				х	х				х
ELIXIR (Europe)				Х	Х		Х	Х	Х
The 100 000 Genomes Project (Genomics England)		х					х		х
FinnGen (FIN)		Х							
LifeGene (SWE)		Х	Х						
The Personal Genome Project (International)				х					х
Global Alliance for Genomics and Health (GA4GH)					х				х

Table 1. Available data sources, tools, and other resources



While biobanks (Appendix 4) and some research projects are organisations that mainly provide access to data or samples, other platform-type organisations provide wider, often free-of-charge services, including computational infrastructures and tools for analysing, testing, sharing, submitting or downloading data (Appendix 3).

4.4.1 Data access policies

Table 1 depicts organisations that offer organisations and researchers access to a variety of data sources (see Appendices 3 and 4 for detailed information). Half of the organisations presented host datasets containing health, lifestyle and genetic data. Four organisations focus solely on molecular level biological or cancer-related data, and three of the surveyed organisations are collaborative data infrastructures that function as catalogues to external data sources without providing their own data.

There are different ways to gain access to the data. Data access restrictions depend on access policies, the type service provided, or anonymity of the data. *Open access* datasets means the data is freely available for download without restrictions. There are various approaches for providing or supporting open data access. For example, the *Canadian Open Genetics Repository* (COGR) pools data from medical diagnostics laboratories in Canada to create a publicly available, open access repository of human gene DNA variants. A similar, but more international effort is the *Personal Genome Project*, which is a coalition that supports publicly shared genome, health and trait data¹. Its participants are encouraged to create voluntary, non-anonymous (i.e. researchers do not promise anonymity for participants), open-access data. The data is freely available on the participant's websites and proves that even non-anonymous health and genome data is publicly available.

With the exception of the Personal Genome Project, only anonymous, molecular level data was available for open access in our review. Organisations maintaining open access data such as The Genomic Data Commons (GDC) and the International Cancer Genome Consortium (ICGC), hold cancer specific data on genomic, epigenetic and transcriptomic levels. The European Bioinformatics Institute (EMBL-EBI) provides a variety of biological data sources, including DNA and RNA, gene expressions, proteins structures, systems, chemical biology and related ontologies. In contrast, organisations that hold combined clinical, lifestyle and 'omics' datasets, such as the UK Biobank, Estonian Genome Center and LifeGene research project, restrict access to data.

Restricted data access refers to datasets that are controlled by a data access committee (DAC), that have an access fee, or otherwise set conditions for accessing data. The DAC usually requires a research proposal or other documents to legitimize the use of the data. Restricted access is common for organisations that sustain datasets that vary in the level of anonymity, and it is common that access to less anonymous datasets is limited. For example, the *Genomic Data Commons* (GDC) makes non-identifiable high-level genomic data publicly available through a data portal, but access to individually identifiable data is granted by a DAC². *The European Bioinformatics Institute* (EMBL-EBI) and the *International Cancer Genome Consortium* (ICGC) follow similar practices.

As biobank research typically involves rich individual-level data sets and samples it is clear that diligent data access management need to be applied. The UK Biobank ensures appropriate use of all their data by requiring registration and application through an access management system³. From the data user perspective it is important that clear guidance and data access process is available. Appendix 4 indicates some user guidance approaches taken by biobanks and other data providers. It can be seen that guidance and prices are largely published and some biobanks also have forms for data applications and data availability

¹ <u>https://www.personalgenomes.org/</u>

² https://gdc.cancer.gov/access-data/data-access-processes-and-tools

³ <u>http://www.ukbiobank.ac.uk/register-apply/</u>



requests. It is remarkable that most collections are globally available, although some of which require collaboration with a local partner. Suitable data and sample contents can be exploited worldwide, which considerably enlarges the likelihood for data users to find resources matching their needs.

Access fees. It is common for biobanks to charge a fee to cover the service costs for providing access to data. Other organisations such as the *AHA* state that they provide open access to cardiovascular datasets, but they take a fee for the cloud computing capabilities used¹ while analysing data on the platform. Some organisations have different fees for academic research organisations and commercial use of the data. For example the *Hartwig Medical Foundation* takes a fee when accessing data for commercial purposes².

Apart from research proposals and access fees, some biobanks and research projects have conditional data access. For example, LifeGene (Sweden), the Danish National Biobank and Hunt (Norway) require applying organisations to be associated with national research institutes or universities. Some organisations directly deny data access from most external parties by involving exclusively their core research partners or data service providers. The 100 000 Genomes Project allows full data access for partner researchers and institutions, while offering a limited data access to commercial parties through a 'discovery forum'³.

Access to external data. The National Genomics Infrastructure (NGI) in Sweden and Elixir are examples of data infrastructures that bring together a variety of resources in order to coordinate and facilitate genomics research, provide access to training, tools and external data resources. These organisations provide lists of datasets to be accessed via external partners. *The Personal Genome Project* is an international collaborative platform for researchers sharing publicly accessible genomics, trait and health data and data is accessible via involved institutions.

4.4.2 Available data analysis tools and computing resources

Many organisations provide computational infrastructures for analysing, testing, sharing, submitting and downloading data. While the platform-type organisations might function as valuable assets for research oriented SME's, their primary function is to support the research community and often basic research. The tools and testing platforms offered are mostly free of charge resources set up in order to support analysing and consolidating the vast amount of genetic and health data.

Several of the surveyed organisations provide access to various resources created in order to analyse, test and develop data analysis. There is a plethora of freely available tools for different analysis, visualisation and testing functions. For example, *The European Bio Informatics Institute* (EMBL-EBI) lists up to one hundred different tools to help analyse data from the molecular data resources⁴, and *Elixir* provides a registry for browsing and searching for available software and data resources⁵.

Providing software tools can be used as a means to standardize processes of analysing and reporting data. *The Global Alliance For Genomics And Health* (GA4GH) shares toolkits in order to push the organisation's mission for setting standards, enable responsible sharing of genomic data and frame policies accordingly. Thus they list tools that enable sharing, storing and retrieval of data, but also tools for data security and even ethical toolkits. The Automatable Discovery and Access Matrix for example, is a tool that standardizes the way consent and

¹ <u>https://precision.heart.org/tos</u>

² <u>https://www.hartwigmedicalfoundation.nl/wp-content/uploads/2017/03/Rules-of-Procedure_ENG.pdf</u>

³ https://www.genomicsengland.co.uk/working-with-industry/

⁴ https://www.ebi.ac.uk/services/all

⁵ <u>https://www.elixir-europe.org/platforms/tools</u>



conditions of data use is represented, enabling easier digital communication, automation and sharing¹.

A few organisations also provide sandbox-type testing and developing platforms. *The International Cancer Genome Consortium* (ICGC) has tools for enrichment analysis, cohort comparison and visualising genetic alterations². They also provide a Jupyter sandbox tool for secure experimenting and exploring. The same tool is used by the AHA Precision Medicine Platform. A similar type of testing environment is available in the *precisionFDA* platform that has sandbox features for experimenting, data sharing and for creating apps³.

4.4.3 Services for companies

While all of the surveyed organizations aim, in one way or another, to support the research community and genomics research, some organizations also explicitly aim to support commercial use of their data and tools. For example, *Genomics England* facilitates collaboration and engagement between industry, academia, NHS and others through their 'Discovery Forum'⁴. *Elixir* provides similar resources with their 'industry programme' set up to support industry, stimulate innovation, and reduce barriers between company research partnerships⁵. Elixir also has an "*Innovation and SME programme*" that hosts events for SME's in national nodes across Europe.

The European Bioinformatics Institute facilitates an Industry Programme aimed mostly at maintaining contact with major pharmaceutical companies using EMBL-EBI in their R&D⁶. The organization reports that they support SME's through free data access, tools and infrastructure. They also organize networking events in collaboration with other organizations such as Elixir SME forum. Both Elixir and EMBL-EBI also maintain training and education programs accessible to both academic and private enterprise researchers.

4.5 Implications for Finnish companies

Innovations in the 'omics' sector, pharmaceutical R&D, clinical care, IT and big data, and personal health services, build largely on the capacity to form meaningful outputs and uses from heterogeneous data sources. Thus, several countries invest large sums to sequence populations in order to create data, to form national data infrastructures facilitating data use, to support research utilizing the data and to develop innovative products and services. The question for policy makers, research organisations and companies remain how to collect standardized and meaningful data in an ethically sound way, and how to efficiently exploit the data in science and business, thereby unlocking the potentials of precision medicine.

As shown in Section 4.4, national and international data resources are widely available. Some of these organisations host significant data sources including 'omics' data, clinical health data, and lifestyle data. Typically, anonymous genomic sequence data or aggregate health data is openly available. Such data may be highly valuable to SME's especially in the bioinformatics domain.

Non-anonymised data and data combining various genomics, health and lifestyle information is mostly available only for scientific research studies. Such data sets - combining genotype and phenotype information - are available mainly through biobanks and university hospitals. Access is restricted to scientific research studies and in some cases limited to be used in co-operation with local research partners. Still, the precision medicine field in Finland can benefit from the fact that research organisations can access high-quality international data resources

¹ <u>https://www.ga4gh.org/ga4ghtoolkit/regulatoryandethics/</u>

² https://dcc.icgc.org/analysis

³ <u>https://precision.fda.gov/docs/creating_apps</u>

⁴ https://www.genomicsengland.co.uk/working-with-industry/

⁵ https://www.elixir-europe.org/industry

⁶ https://www.ebi.ac.uk/industry



in addition to national data sources. Apart from universities and research institutes, benefits can be foreseen to come through internationally oriented companies, which can make quality research complying with scientific standards and which can use this knowledge in providing services for other companies. Networks and ecosystems need to be formed to connect research providing companies to the heterogeneous group of companies with research questions and information needs.

As revealed by this review, various organisations provide data analysis tools, or help commercial entities by providing education or forming collaborations. Such resources and infrastructures can be very useful for Finnish SMEs. For these companies the biggest benefit appear to come from platform type organisations offering data, analysis tools, testing environments and collaborative structures.



5

In order to get an understanding of the current precision medicine landscape in Finland, interviews where performed with various stakeholders during the first phase of PreMed in 2017-2018. The focus was to understand the organisations in general, their value chain(s) and, the meaning of health data for them. We also wanted to know about the stakeholders' expectations on the FinnGen project and its contribution to the Finnish precision medicine ecosystem. The results of the interviews collect general views of the players as well as their wishes and doubts.

5.1 Methods

In order to get key persons from the organizations for the interviews a flexible approach was taken. Most interviews where performed face-to-face at their premises, but some by Skype or by phone in case of scheduling problems. The duration of the interview was typically 1-2 hours. The interview was recorded in some cases, after given oral permission of the interviewed person. The interviewed experts are listed in Appendix 5.

A set of questions was used as the basis for the interviews (Appendix 6). However, due to the heterogeneity of the target group, all listed questions were not relevant for all interviews. Therefore, freedom was given to follow the flow of the interview and take up issues during the discussions. In all cases, the direct comments have been kept anonymous in reporting the results.

5.2 Interviewed organisations

It has been estimated that the Finnish health technology sector contains over 400 companies, most of which are fairly small (over half of them have less than 10 employees)¹. Only some of these where selected for interviews based on their relevance with respect to exploitation health data and genome technology. Other stakeholders interviewed include e.g. biobanks, hospitals (private and public), pharma companies and various service providers. In addition to the formal interviews we also contacted additional pharma companies for informal discussion, especially to get a wider perspective for planning the next project phases.

As listed in Table 2, altogether 32 interviews and 6 informal discussions were carried out. Table 2 also indicates four earlier interviews (carried out Autumn 2016), which we exploited in the study. The used classification of organisations is approximate, as some companies may belong to several classes.

¹ <u>Suomen terveysteknologia-alan nykytila ja haasteet (340/207, Tekes)</u>



Organisation category	Organisations
Biobank	Auria biobank, FINBB, Blood Service Biobank, Helsinki Biobank**
University, hospital district, research institute	VSSHP, University of Eastern Finland , THL, Sitra
Research and expert services	Medaffcon, Medengine, CrownCRO*
Bioinformatics software and services	BC Platforms, Euformatics, Genevia, Medisapiens
Sequencing genotyping and diagnostics services	Blueprint Genetics*, Zora Biosciences*
Genetic test interpretation and counselling services	Abomics, Diagfactor
Medical devices manufacturer	GE Healthcare, Modulight
Pharma and diagnostics	Orion, Abbvie, Pfizer, Novartis**, Roche**, Bayer**, Boehringer Ingelheim**, Janssen**
Healthcare information systems and applications	Duodecim, BCB Medical, Kaiku Health, Odum, Quattrofolia, Mediconsult*
Healthcare services	Mehiläinen, Terveystalo, Docrates, Aava
Insurance services	Lähitapiola
Food and nutrition	Fazer, Raisio

Table 2. Organisations involved in interviews and informal discussions

* Interview carried out before PreMed project context in Autumn 2016.

** Informal discussion

5.3 Observations from interviews and discussions

In general, the interviews reflected the growing need of companies to exploit health-related data. The FinnGen project was considered as a great opportunity to increase the exploitation of genotype combined with phenotype data and to increase the visibility of Finland in this field. The dramatic drop of sequencing cost was noted in several interviews to be an important driver for large-scale exploitation of genome technologies. A common hope was that - besides benefits for academic research - the FinnGen project would also bring benefits and opportunities for companies.

Many of the interviewed companies highlighted the need for various types of health data, including personal health monitoring and life style data. Genome data was considered only as one data type among the others, although its growing importance was recognized. In this respect, the opportunity to access data through a single service point (Service Operator) as defined in the Secondary Use Act was warmly welcomed.

The legislation, in particular the Secondary Use Act and the Biobank Act, where considered to give competitive advantage to Finland along with the Finnish social security number, the genetically unique population and high-quality health data resources.

The current status and needs concerning data and genome technology exploitation for each company category is summarized in the following subsections. Related bottlenecks and challenges are addressed in Section 5.4. The observations are based on the interviews and informal discussions with stakeholders listed in Table 2.



5.3.1 Pharma and diagnostics companies

Pharma and diagnostics companies have a long tradition of using patient data in various phases of drug development and in Real-World Evidence (RWE) studies. Data analytics is carried out in-house and in the context of extensive academic collaborations. RWE studies are complementary to clinical trials carried out under informed consent of the study subject. They are highly important to pharma companies for example in the process of getting drug reimbursement decisions from the authorities. RWE studies are also important in providing information on the use of the drug in the context of adaptive pathways¹. Adaptive pathways refer to bringing drugs faster to markets by reducing the need for clinical trials before marketing authorization. Adaptive pathways are applicable in the context of life-threatening diseases where there are no existing treatment options.

Access to health data registries is not only important for RWE studies, but also for patient stratification in preparation for clinical trials. Additionally, omics data has an increasing potential in the early phase of drug development as it enables an understanding of the underlying biological mechanisms and possible adverse effects as well as in finding new biomarkers for medical diagnostics.

Multi-national pharma companies are continuously evaluating candidate countries for drug development, clinical trials and RWE studies. Availability of high-quality data efficiently and within predictable time schedules is an important selection criteria. The interviewees felt that Finland is in a good position thanks to statistical registers systematically collected over long time periods, but the processes for accessing the data are too slow. Consequently, pharma companies have high expectations on the Service Operator and biobanks in being able to develop systematic and efficient processes for data access. It was also brought up that Finland is a small country and should invest in collaborative analytics in international context to maximize benefit of its data resources to the scientific and medical communities.

5.3.2 Research and expert services

Research and expert service companies are carrying out clinical trials and RWE studies under contract with pharma companies. They are in a key role in accessing the data sets for the studies, and therefore the needs to improve the data access processes were pronounced. In the interviews, also the need for clear legislation and rules concerning data access was raised. Besides the pharma sector, also healthcare organisations are an important customer group and there are signs of other emerging customer groups, such as medical device manufacturers.

5.3.3 Medical device manufacturers

There is a clear trend of increasing intelligence to medical devices and, in particular, collecting data from networked devices (Internet of Things, IoT) to be jointly analysed. This trend applies especially to monitoring systems at intensive care units (ICU) and wards in hospitals. The objective of patient data analysis is to obtain a holistic, real-time view of the patient for the clinicians and initiate alarms as needed. The combination of various types of data enables also predictive models to be created. Development and tuning of such models requires availability of monitoring data sets. Getting permissions for data access separately from different hospitals is laborious. Consequently, there is a high demand for a centralized services covering not only statistical registers, but also detailed patient data in hospital data lakes.

5.3.4 Bioinformatics software and services

Bioinformatics software and services companies carry out data analytics and interpretation on individual-level data. The data is typically provided by the customer in the context of a work

¹ EMA Guidance



assignment. Consequently, challenges of accessing data are not pronounced in the interviews. The data analytics projects typically exploit in-house software, which is also available to be purchased by the client. A new opportunity is seen in embedding data analytics software modules in medical devices. The interviewed companies are globally oriented and already have a remarkable share of their business with international customers.

5.3.5 Sequencing, genotyping and diagnostics providers

Sequencing, genotyping and diagnostics providers are companies with laboratory facilities. The companies range from SME's with focused test product offerings to large international companies with wide spectrum of laboratory services. In general, these companies gain from increased demand for omics studies in the context of biobank research and the use of biomarker-based tests in healthcare. Access to data is not of high priority for companies focusing in laboratory services, but some of the companies in this group have also activities in the data analytics domain.

5.3.6 Genetic test interpretation and counselling services

These SME companies base their business on supporting the use of genome data in healthcare or by individuals. A typical example is pharmacogenomics (Ji 2016). Metabolism of certain drugs is dependent on known gene variants, which has to be taken into account when selecting the appropriate drug and dose. This motivates the healthcare service provider to order the respective gene test and interpretation service. Interpretation and counselling services are also provided directly to consumers. In this case, the gene test typically addresses health risks or hereditary diseases. In offering these services the SME's do not directly need data resources. However, genome sequencing of large populations may have a favourable effect as it potentially increases demand for interpretation and counselling services. On the other hand, data would be needed for RWE studies to reveal evidence on the accuracy and impact of the tests.

5.3.7 Healthcare service providers

Healthcare service providers have strong interest towards secondary use of health data from many perspectives. At university hospitals and other clinical service providers, there are healthcare professionals carrying out clinical or registry-based research. Hospital data lakes and biobanks are quickly emerging and are already in active use. The Service Operator and joint biobank services are expected to provide further benefits in accessing data and sample resources of different controllers and in providing the necessary processes for finding and getting permissions to access the resources.

On the other hand, access to health data is needed for monitoring and analysis of healthcare quality and processes. Another example, brought up in the interviews was monitoring of wellbeing of employees as part of an occupational health service. Such applications are based on the data controlled by one organisation and, in these cases the Service Operator is not expected to bring additional value. The usage of genetic testing is still relatively modest in the day-to-day healthcare practices. According to the interviews, more evidence on the benefits are needed and the guidance should be incorporated in the care guidelines.

5.3.8 Healthcare information systems and applications

In this category, there are large and small companies providing information systems and applications to healthcare professionals and citizens. Information systems and applications are not anymore only for data management and viewing. Instead, there is a need for converting the data into useful information through advanced decision support, reporting and visualisation functionalities. The development of such advanced functionalities relies heavily on access to data. On the other hand this category of companies benefit from the growing demand for quality registers and biobank information systems powered by secondary use of data.



5.3.9 Insurance services

Insurance services are interested in using health related data e.g. in designing new products motivating individuals for health maintenance. Population-level information e.g. concerning geographical differences and dependency between lifestyles and diseases are highly interesting and could be used in new insurance products and in insurance pricing strategies. Insurance companies are extremely careful in maintaining their reputation towards customers. Consequently, there seems to be no plans for secondary use of individual-level health data directly.

5.3.10 Food industry

Companies in food and nutrition industry are focusing R&D efforts towards new products with positive health effects. There is an increasing interest towards working together with employers in order to manage health risks of employees. Such activities include targeted interventions for risk groups and combining the elements of nutrition and physical exercise. The understanding of personalized effects of food is of high importance and has motivated several research activities. The future is foreseen where optimum diet for an individual can be constructed by using individual-level data, including genome and microbiome data. Especially, the effect of microbiome on health and the possibility of modifying the microbiome with the help of food are currently under extensive research.

Access to individual-level data storages is an important opportunity to increase the understanding of individuals' differences in response to food. A big challenge for research is how to reliably collect information about the human behaviour - e.g. intake of food.

5.4 Bottlenecks and challenges

As indicated in Section 5.3 there are high expectations towards data-driven precision medicine and the recent efforts in improving access to health data are widely appreciated. However, the interviews revealed also a number of bottlenecks and challenges. The main issues have been collected in Table 3 divided into groups: healthcare, pharma, SME companies, data access and public policy.

Healthcare and pharma development are the main drivers for data exploitation. Many SME's seek co-operation with larger organisations, but face difficulties. For example, in healthcare, the use of genome and other omics tests has not increased as fast as hoped, and consequently the demand for such services stays low. The companies also feel that access to Finnish hospital processes for real life testing of new inventions or for clinical studies is not always easy. In some cases, testing abroad is easier and faster, and innovations are adopted faster outside Finland. This has led to companies establishing units abroad and, consequently economic benefits both for our domestic healthcare as well as our export business have been lost or delayed. Furthermore, many SME's feel that FinnGen does not offer concrete options for collaboration for them.

In pharma sector, the main bottleneck is that drug reimbursement decisions and regulatory approvals take too much time, which causes the drug development costs to increase. Also access to data should be faster and better organized. Although the Service Operator, joint biobank services (FINBB) and the Genome Center are considered as positive initiatives, many of the interviewees were worried about the slow progress. Similar concerns related to coordinating biobank activities are also faced in other countries (Chalmers 2016).

Concerning public policies, it was pointed out that financing of R&D has been falling during several years, which implies that Finland's advantage is quickly reducing. Lack of coordination, standard interfaces and competences were other issues slowing down the development.



The collaboration of the Finnish Biobanks was regarded crucial by many of the interviewed stakeholders. The newly established FINBB was considered to be important in this respect, but its role was considered still somewhat unclear with respect to the individual biobanks.

The pricing of data and samples seems to be a challenge in many respects. Public biobanks benefit from public resources in collecting data and samples. Therefore, most of the companies feel that data should be released for low price or even free of charge. On the other hand, some experts feel that releasing data without proper payment prevents a viable operating environment and markets from being developed.

Bottleneck / challenge	Primarily affected stakeholders	Description			
Healthcare					
Lack of clinical guidelines, education and decision support systems.	Genome test interpretation servicesHealthcare providers	Clinical guidelines and education concerning the use of new methods and technologies in healthcare are missing, which decreases the use of genetic tests in healthcare (e.g. pharmacogenomics). Embedding guidelines in decision support systems would help the clinicians.			
Research evidence is missingGenome test interpretation servicesHealthcare providers		Research evidence on the benefits of genome tests and new pharmaceutical products would be needed in order to motivate their use and related costs in healthcare.			
Genome testing services missing	Healthcare providers	High quality genomic profiling and results interpretation services for cancer care are missing from Finland and services need to be searched from international markets.			
Public healthcare providers in Finland are slow in adopting new technologies	All companies developing new healthcare applications and devices	Hospital districts and municipalities do not invest in adopting and piloting new products and technologies.			
Pharma					
Slow drug reimbursement decisions and regulatory approvals	Pharma companies	Drug reimbursement decisions and regulatory approvals take too much time. This increases drug development costs and causes RWE studies to be done in many cases outside Finland.			
Biobanks					
Motivating donors	• Biobanks	Finland has a reputation that citizens are <i>good donors</i> of samples and data. Motivation of the donors should be better maintained. For example, FinnGen should offer genome analysis results for the donors. It was pointed out, that a biobanks' "donor portal" is already being planned.			
Biobank consent	• Biobanks	<i>Opt-in</i> approach for biobank consent is inefficient and laborious. <i>Opt-out</i> consent would allow faster and more efficient data collection for biobanks.			
Biobank co-operative	• Biobanks	The Finnish biobank co-operative (FINBB) currently leaves out private biobanks. This is unfortunate concerning biobank co-operation and because public monopolies should be avoided. The role of FINBB is still unclear with respect to the services provided by the individual biobanks.			
SME companies					

Table 3. Bottlenecks and challenges hindering data-driven precision medicine.



Bottleneck / challenge	Primarily affected stakeholders	Description			
Companies are not paid for data they collect from the user.	 SME's providing health monitoring solutions. 	SME's are expected to connect their systems with centralized services (e.g. Kanta PHR) and to enable monitoring data to be used by other services. SME's are reluctant to release data without compensation.			
Participation of SME's in FinnGen is considered difficult	• SME's	FinnGen has taken the perspective of big pharma companies and academic research. There are not many real opportunities for SME's to co-operate.			
Lack of co-operation between SME's and large companies in Finland	• SME's	The amount of large companies in healthcare and pharma sector in Finland is small. It is difficult for SME's to find markets and to initiate co-operation with large companies to create successful products.			
In-house development of software in public sector	Companies developing healthcare applications	Companies feel that public sector organisations should not develop SW <i>in-house</i> . Instead, they should co-operate more with industry and adopt existing solutions of companies.			
Risk of over-pricing data	Companies	If access to data is over-priced the interest of companies (especially SME's) may drop.			
Data access					
Access to data is slow and unpredictable	Research service providers and other direct data users	The service processes of register-owners and biobanks is slow. Especially, large international pharma and medical device companies need fast and predictable access to data resources. If this can not be provided in Finland the studies will be executed elsewhere. Also it would be important to access more detailed level data than currently made possible in FinnGen.			
Centralized services for data access are still missing	 Research service providers and other direct data users 	Despite of the centralized services, it seems that data remains distributed and difficult to exploit still for some time. Also, the division of responsibilities between the centralized and biobank-level services remains unclear and it is not yet clear which data resources will be available through centralized services (for example there is interest towards the hospital data lakes and primary healthcare data).			
It is difficult for companies to access data	 Companies (including private healthcare) 	Availability of individual level data is restricted for scientific research. SME's do not have the required competences and resources and can't afford to use the help of research service providers. Furthermore, data access conditions are more strict for private healthcare than for public healthcare researchers.			
Personal data should be better exploited.	 Healthcare providers Pharma companies Food industry 	The usage of personal health data (apart the normal clinical data) is not in-built in the healthcare process and only used in the minority of cases. Personal health data has growing importance also in clinical research. Nutrition research would benefit from detailed lifestyle data. A cost-efficient consent mechanism is needed.			
Healthcare data not in exploitable form	 Research service providers and other direct data users 	A large part of data in patient information systems is in free text and therefore difficult to exploit. Secondary use is properly taken into account in quality registers, but these have not been developed fast enough as compared to other Nordic countries.			
Inclarity of regulation	Research service providers and other direct data users	Legislation (Secondary Use Act, Biobank Act and GDPR) leaves open how data can be used ('scientific research' vs. 'data use for development and innovation'). There are no common criteria for anonymisation in use.			
Public policy					
Public funding for R&D and infrastructures decreasing	All companies	Public funding for R&D and infrastructures has been decreasing. Finland should focus public funding in selected areas to increase impact.			



Bottleneck / challenge	Primarily affected stakeholders	Description
Risk that data is released for large international companies without proper compensation	Finnish companies.Biobanks	It is not considered appropriate to release data without proper compensation. Instead, results derived from the data should be offered. The interests of Finnish companies should be taken into account appropriately.
Lack of coordination in standards and platforms	All companies	Efficient use of data is only possible, if information systems (e.g. hospital data lakes) are connected via standard interfaces based on international coding systems and vocabularies.
Lack of competences in specific areas	All companies	There is a lack of experts with competences e.g. in data analytics, simulation techniques and epidemiology.

5.5 FinnGen opportunity

In all interviews, the FinnGen project was brought into discussion with the objective of understanding its importance for companies. As expected, FinnGen project was seen most useful by the pharma companies, especially those participating the project, but also others, which will have the possibility to use FinnGen data via *add-on projects*. Some Finnish SME companies were involved in the project, e.g. providing software platform components, and found the project beneficial for them. Otherwise, the companies outside FinnGen could not see direct benefit for themselves, although the project in itself was mostly considered to be positive and to bring good visibility for Finland.

Some SME's had ideas for contribution in FinnGen in enhancing the connection between the project and the donors. An opportunity was seen in offering the donors a views for their genome data, personal risk factors and pharmacogenomics analysis results. Such activity would be highly important in bringing value to individuals and in motivating more people to join the project as donors. So far, such service have not been implemented.

5.6 Stakeholder views on secondary data use

The interviews raised interesting company perspectives on data-driven precision medicine. In general, most interviewed companies were expecting to benefit from the availability of individual-level data in some way. As expected, the pharma industry had most concrete ideas on data exploitation, and many of the companies already had experience on biobank research as well as using data from hospital data lakes. Especially RWE studies were mentioned as an important application area both by the pharma industry and by research service providers. Potential benefits of secondary data were also identified for the healthcare area, but the related business cases are not yet very clear. By default, public and private healthcare providers focus their resources in providing healthcare services using evidence-based clinical methodology. It seems, that especially in the public healthcare, there are not resources for experimenting new innovative methods in co-operation with companies. In private sector, there seems to be more activity towards exploiting data analytics to support care provision. Especially, the occupational health domain offers interesting new business opportunities.

An interesting observation is that healthcare providers are primarily interested in accessing their own data repositories. Wider access to health data is 'nice to have' but not a priority, especially if excessive efforts are needed. This perspective is manifested by the rapid emergence of hospital data lakes as well as in the development of in-house data-driven solutions to support occupational health.



6 Company perspective to data exploitation

Based on the interviews and informal discussions it is clear that availability of individual-level data for secondary use has a positive impact to society in several ways. Most evidently, positive impacts are seen in academic research and drug development where studies can be based on larger subject groups with individuals characterized by rich genotype and phenotype data. Besides pharma industry, also other business domains such as healthcare services, medical device manufacturers and food industry, are foreseen to gain from secondary use data. Moreover, there is a growing number of companies providing products and services to support data exploitation.

Based on the interviews and informal discussions with stakeholders the objective of this chapter is to clarify the overall operating environment and the conditions under which secondary use of health data is possible for companies.

6.1 Operating environment

The operating environment for health data exploitation is depicted in Figure 2. Four main roles with respect to data exploitation have been identified: (1) data controllers, (2) data users, (3) research service providers and (4) service and technology providers.

The organisation roles and their relations are described in the following subsections. It is important to note that in Figure 2 one company may appear in several roles. For example, a pharma company is in the role of *data user* when using services of *research service providers*. On the hand, the research service provider role may also be covered by the pharma company's internal research unit.



Figure 2. Operating environment for health data exploitation.



- 6.2 Data exploitation roles
- 6.2.1 Data controllers and brokers

Data controllers and brokers refer to centralized data operator services (FINBB or Service Operator) as well as individual biobanks and other data controllers which possess health data available for secondary use. As described in Section 3, the centralized services are still under development and in the current state data is obtained for research through the data controllers directly. These include the biobanks and all other organisations responsible for maintaining data registers, e.g. KELA, THL and Statistics Finland. Also health service providers are data controllers and may directly disclose data for secondary use. The legal framework for providing access to data and/or samples from biobanks is established by the Biobank Act. Correspondingly, data from the Service Operator will be accessed under the terms provided in the Secondary Use Act.

As permitted by the related legislation, the three main approaches for user consent are (see Figure 2):

- 1. Data use with **specific consent** by the data subject
- 2. Anonymised use of data (without specific consent)
- 3. Data use for **scientific research** (without specific consent)

In Case 1, a specific consent shall be requested from the data subject, after which corresponding use of the data is allowed and is not limited to scientific research. The challenge of this model is that it is extremely difficult to collect specific consents especially from large groups of individuals. Therefore, it is not foreseen to be a viable approach for most of the anticipated secondary use scenarios.

In Case 2, the data controller or broker carries out the anonymisation of the data to be disclosed. Also, in this model, the usage of data is not limited to scientific research and can be used for a large variety of innovation and development purposes as well as for monitoring organisational processes and quality. The challenge in this model is that it is extremely difficult anonymise data properly. When the disclosed data sets are large and contain several individual level health parameters, it is highly likely that persons can be identified.

In Case 3, the data is used without specific consent, but shall be carried out according to scientific research practices. This is the prevailing model used in biobank research, where only a generic consent is required and in registry based research where no consent is required. Following scientific research practices implies that a research plan with valid research objectives need to be documented and results shall be published.

6.2.2 Data users

Data user refers to companies using data directly in-house or indirectly via research services. In the latter case the data user is not directly accessing the individual-level data, but merely data analytics results, aggregated data and visualisations.

Currently, direct exploitation of health data mainly takes place in pharma and diagnostics companies. Data accumulated in healthcare registers is useful on one hand in the early phases of drug development, when identifying patient groups for new drugs and studying the related disease mechanisms. On the other hand, health data is needed in Real World Evidence (RWE) studies targeted to monitor the efficacy of pharma products.


As indicated in Figure 2, several other groups of companies besides pharma and diagnostics companies are potential information users. There is a growing number of companies providing services related to genetic testing. Business to consumer (B2C) and Business to business (B2B) services offer genetic tests and related interpretation and guidance. Typical examples are health risk analysis (B2C) (Su 2013) and pharmacogenomics assessments (B2B) (Brown 2017). Such services are based on clinical evidence (published scientific knowledge). Access to data enables building this evidence and monitoring of the performance of the provided service.

For emerging value-based healthcare it is important for healthcare service providers to monitor the efficacy and quality of the healthcare delivery process. Monitoring is based on secondary use of patient data. This is especially relevant when new interventions, drugs and care models are adopted to clinical practice. A good example is the use of genetic tests in order to support care planning and decisions on medication.

Healthcare information systems and medical device manufacturers are increasingly accompanied by artificial intelligence based software components providing advanced functionalities, such as clinical decision support. The development of such components requires health data to be available for model tuning and verification purposes. Also, data repositories can be used in assessing the number of certain types of patients, which may be important information in product development.

Insurance providers have various types of information needs. An important area is insurance pricing. The usage of individual-level health data in pricing would not be ethically appropriate. However, aggregated health data could reveal, for example, health differences depending on living area, which could be taken into account in pricing.

6.2.3 Research service providers

Research service providers have the required capabilities of carrying out scientific research studies. In addition to specialized companies, this kind of services is traditionally provided by universities and research institutes. This role is important, since scientific research methodology is a pre-condition for using individual-level data without consent as indicated in Section 6.2.1. Many of the companies in data user role do not have the required competences and prefer not to commit themselves to data analysis activities using scientific research methodology. Instead, they may prefer to outsource these activities to research service providers, which are expected to provide key results in a clear and directly usable form.

Research service providers are currently serving mostly the pharma industry. However, it is anticipated that this role needs to expand in the future to better serve the wide range of information users as depicted in Figure 2. Based on the interviews it seems clear that the group of data users is rapidly growing and getting more heterogeneous. As a result there is a growing need for versatile research services addressing not only the needs of pharma companies, but also the other business areas, such as healthcare services, medical device manufacturing and food industry.

6.2.4 Service and technology providers

Service and technology providers are a heterogeneous group of companies supporting datadriven precision medicine. By default, large amounts of data are involved leading to the need for specific data management services. In many cases also the services of specialized data analytics companies as well as software providers are needed. In the case of biobank research, there is often a need for various types of laboratory and analysis services targeting a specific area of interest, such as genomics, proteomics and metabolomics jointly referred as *omics*. The expanding activities around data exploitation is expected to increase the demand for service and technology providers.



6.3 GDPR considerations

While potential business benefits of releasing data for secondary use are clear, it is important to take into account privacy and other fundamental rights of the data subjects (Ayday 2015). Secondary use is particularly critical, since in many cases it is not possible to obtain explicit consent from the data subjects. This is the case especially when analysis of large data masses is needed. As noted in Section 3.4.1. the usage of data without explicit consent is in line with the GDPR when the data is used for scientific research or when the data is fully anonymised. Anonymisation may be a feasible solution for data sets, which contain only a limited set of data for each individual and where *rare phenotypes* are not present. When a large spectrum of individual-level data is handled, full anonymisation is often not possible or it would require a substantial amount of information to be removed from the data set. The challenge is expected to become even more prominent in the future as the dimensionality of biobank data grows (Polašek 2013).

Due to the inherent problems of full anonymisation, it seems that the scientific research approach is most appropriate in many cases. This approach is attractive also because the GDPR adopts a broad definition for 'scientific research' covering also 'development and demonstration of technology'. On the other hand, the broad definition leaves open, what actually can be included under the term¹. Unfortunately, the current draft of the complementary national privacy legislation² does not include a clarification for the definition of 'scientific research' either. The draft Secondary Use Act³ explicitly defines the term 'development and innovation activity' as an application area for secondary use on the condition that explicit consent is given by the user.

Evidently, detailed conditions for accessing data are not yet clear and probably different interpretations of the GDPR will be seen in different EU countries. For the time being, the best guess is that without explicit consent, individual-level data will be released for studies carried out using scientific research methodology. This requires that a valid research question needs to be formulated and a corresponding research plan needs to be presented by the data recipient. There may also be requirements to publish the study results. In any case, the recipient of the individual-level data becomes the controller of the data set and needs to ensure GDPR compliancy in managing the data set and in ensuring the rights of the data subjects. Pseudonymisation of the data does not free the data processor from the obligations stipulated by the GDPR (Chassang 2017).

The responsibility for managing and processing individual-level data sets can only be taken by organizations, with the required competences, processes and information systems. Pharma companies typically have such capability. Additionally, such services are provided by specialized SME's, universities and research institutes. Taking into account the growing importance of health data for pharma development and other business sectors, it is clear that the demand for research services will grow. Especially, the new emerging data users, such as healthcare providers, medical device manufacturers and food industry need external services in order to exploit individual-level data. In typical scenarios, companies have research questions related to their product development and marketing. They need answers to the questions, but are not willing to invest their own resources for carrying out the actual research activity.

As discussed above, access to individual-level health data will remain strictly controlled even if improved processes and platforms enabled by centralized services as described in Section 3.3 are expected to enable more efficient and predictable access to data.

¹ <u>https://www.law.kuleuven.be/citip/blog/scientific-research-under-gdpr-what-will-change/</u>

² https://www.finlex.fi/fi/esitykset/he/2018/20180009

³ https://www.finlex.fi/fi/esitykset/he/2017/20170159



7 Artificial Intelligence Opportunity

7.1 Artificial Intelligence use cases

From Artificial Intelligence (AI) technology point of view, the interviewed stakeholders pointed out many topics of interest. Due to the large variety in business fields and company profiles, the found AI-related topics covered a broad range of use cases and application areas. Table 4 summarizes these application use cases and related technologies.

Table 4. Application use cases with AI potential¹

Application use case	Related technologies and AI options
Online health services, Virtual Clinic, eHealth services	chatbots, data integration, intelligent visualization, intelligent sensor data processing, data fusion, NLP, DSS, ML, NN
Virtual coach; Questionnaire based mobile app with recommendations for maintaining health and medicine usage	large database handling, suggestion engines, natural language processing
Digital interventions, companion app; intelligent applications to support medicine usage and increase adherence	DSS, data integration, free text interpretation
Continuous detection of unhealthy lifestyle and preventive actions, providing adaptive personalized feedback	multi-domain personal representations, profiles, lifestyle 'fingerprints', ML, traditional clustering methods, time-series analysis (LSTM)
Intelligent support for use of medicine and integration of care process in combination with e.g. lifestyles, demographic background and use of medicine	DSS, data integration, natural language processing, intelligent visualization, intelligent sensor data processing, data fusion, clustering, high-dimensional data exploration, understanding, exploitation
Personalized analytics for (occupational) health	data integration, understanding, exploration, DSS, time-series prediction
Personalized services for patients with severe illnesses (e.g. cancer, trauma).	data integration, understanding, explication, intelligent visualisations
Data exploration from complex data combined from health and other domains	DSS, multi-domain classifier, ML, DNN
Utilization of self measured health information and use of it with traditional health care data.	MyData approach and data fusion
Image processing for medical data	Image segmentation, automated measures and feature extraction, ML, DNN, CNN, traditional algorithms
Update of clinical guidelines using AI methods based on medical findings and known literature. Using wider perspective efficiently in health care.	DSS, data integration, free text interpretation, ML, DNN, NLP
Finding new use targets for existing medicines.	simulations, cohort detection
Early-failing in drug development with help of genetics. Al helping to detect the harmful genes and/or efficient medicine.	simulations, cohort detection, classifiers, clustering

¹ AI = Artificial Intelligence, NLP = Natural Language Processing, NN = Neural Networks,

ML = Machine Learning, DSS = Decision Support System, DNN = Deep Neural Networks,

CNN = Convolutional Neural Networks, LSTM = Long Short Term Memory, DL = Deep Learning



Application use case	Related technologies and AI options
Support for fast cycle medicine development (6-8 months), cohort detection and selection, Preclinical studies for medical treatment based on genetics.	classifier and clustering, time-series analyses
Risk prediction at personal, meso and population level from genomic, lifestyle and demographic data	classifier and clustering
Impact assessment, taking objective pros and cons of interventions into account in decision making	DSS, data integration, natural language processing, intelligent visualization
Large scale market segmentation of insurance customers	classifier and clustering, ML, traditional clustering methods
Preventive forecasting for life insurances, risk assessment	classifier and clustering, ML & NN

Specific areas of interest for AI applications can be detected from Table 4. The first one is the eHealth service related ideas to bringing clinicians in contact with patients via computers or mobile devices. These services can speed up the contact to all patients while only severe illnesses require actual visit to hospital. Virtual AI-based assistants can provide automatic feedback and decision support e.g. for chronic disease patients at home thereby enabling independent self-management of the disease. In more severe cases, even real-time surveillance with AI-based detection of the deterioration of the patient condition can be achieved. For such applications, predictive models based on large data sets on earlier patient cases are needed.

The second clear interest area concerns drug and medicine based therapy development. The data with AI technologies could enable faster drug development and better cohort detection for drug tests. It could also enhance drug based treatments when the drug to be used could be selected based on genetics (pharmacogenomics) or cohort based knowledge. Less try outs would be needed to identify the best possible drug treatment for the patient.

The third interest area is personalized services for healthcare and health maintenance, e.g. provided by occupational health providers or insurance companies. Such approaches are based on holistic understanding of the individual's health status and how it can be best supported with available resources, including both legacy healthcare services and eHealth services. Al technologies provide various tools to implement such applications, but significant amount of data is needed to create reliable results.

7.2 Potential technologies

The previously identified topics include opportunities for various AI related technologies, where the greatest challenges are related to the data access and quality. As a single topic, AI is extremely wide and it is better to consider AI technologies as an umbrella which includes various methods under it. Often AI refers to machine learning (ML) methods or neural networks (NN) with deep learning (DL) approaches (Meyer 2018). The greatest modern advances with AI titled technologies are based on deep learning methods, which can be considered as one type of machine learning with neural networks. Common to all of these are that they need a large amount of well curated data.

Recently, natural language processing (NLP) has been used in many application areas, including healthcare. A typical NLP application is the interpretation of natural language with NN methods using DL approaches. NLP is a convenient tool for processing non-structured contents such as free text reports or user feedback. This type of data is abundant in healthcare settings. More advanced approaches use NLP to create chatbots to homepages or social



media platforms. Chatbots are successfully used e.g. in healthcare service provider settings, to answer simple questions, and help with making reservations for appointments. In recent approaches, NLP technology is typically combined with speech recognition and synthesis, which has enabled assisting technologies like Amazon's Alexa¹ or Apple's Siri². In practice, all big cloud providers have their own implementation of these technologies.

The current development in cloud services has brought basic AI technologies within reach for every company and even for individual developers, without needing in-depth knowledge about AI technology and paradigms itself. All major cloud providers have their own pre-trained APIs for users to utilize text, image, sound and video recognition services:

- IBM BlueMix Watson³
- MS Azure Al⁴
- Google Cloud Al⁵
- Amazon AWS Machine Learning⁶

The pre-trained API's with low costs enable a large variety of basic analytics to be implemented in different services without virtually any cost for AI development. In practice, the case is not simple as the processing of health domain data involves significant security and privacy requirements. This typically causes limitations for using commercial APIs. Often it is needed that the computations are ran on the premises where the data physically resides or, more typically, on virtual private networks (VPN) with no connection to the public Internet. Another trend is calculation within devices (e.g. mobile phones) and only provide summarized data to the outside world (edge and fog computing) (Chen, 2018).

An obvious question concerns the decision between developing an AI or ML system in-house to create a custom application and using existing commercial APIs. There is no single answer for this, but one significant challenge is data and system security. Many data controllers want to ensure that all data processing happens inside Europe, which can reduce the options to use commercial APIs. Furthermore, for security reasons the data controllers may not allow data to be processed outside their own network, which also causes limitations for the use of commercial APIs. Cloud providers have options to create secure VPN environments for this kind of cases, but it requires custom work to set up such processing environments. Another side is accuracy, if the data is specific for a certain group and the case is not common it is almost mandatory to create a custom AI solution for the case.

One significant challenge in data utilization is the GDPR, which implies the need for anonymisation in many application areas. When data is fully anonymous the GDPR does not apply. The challenge is that even when the data seems to be anonymous, it may still be possible to identify individuals by combining the data with other data sets. When using large datasets with a high number of measures and variables it is hard to confirm that ML methods cannot identify single individuals from the data. For this reason, many new privacy preserving methods are under development. For example, Google researchers have proposed an ML method where a model is trained in portions and at the end, the partial models are combined as whole model without any original data.

¹ <u>https://developer.amazon.com/alexa</u>

² <u>https://developer.apple.com/sirikit/</u>

³ <u>https://console.bluemix.net/catalog/?category=watson</u>

⁴ <u>https://azure.microsoft.com/en-us/overview/ai-platform/</u>

⁵ <u>https://cloud.google.com/products/machine-learning/</u>

⁶ <u>https://aws.amazon.com/machine-learning/</u>



7.3 Future opportunities

The use of AI technologies are in a transformative state. During the previous AI hype it was believed (by some) that many areas of medical work, such as analysing radiological images, could become entirely automated by AI technology. Currently, the consensus is that AI merely relieves time-consuming routine work and supports human users to make better decisions and to perform more precise work faster¹. For example, Google researchers released their article stating that AI will not replace the human doctor, but can assist them to make better decisions (Li 2018). Most of the specialists in AI related fields have agreed this already for a long time and in practice most of the current AI applications are actually supporting the human actions. The better term for artificial intelligence would be assisting intelligence, or augmented intelligence, sometimes referred as IA, (Intelligence Augmentation/Amplification) (Jarrahi, 2018).

Al solutions can be better than human equivalent when the task is limited (narrow intelligence is needed) and variance within the cases under the task are within certain limits. Detecting details from images or expressions from texts are this kind of tasks (Li 2018). A natural continuation for this process would be systems which combine various AI technologies in series and parallel setups to create more complete picture from the case under investigation. Even in these cases, the main aim would be to enable human specialists to use more time in actual work instead of manually collecting and interpreting increasing amounts of information.

In the application use cases listed in Section 7.1 there are clear examples where AI can speed up processes and help to make more accurate decisions in health domain. NLP technologies could be used to structure the information content and even create clinical data classification from free text descriptions. Also free text entries in Electronic Health Records can be translated to coded structural information. On the other hand, the same approach could also be used to generate a written standard report from basic medical imaging cases. The intelligent system could first segment the image and then measure standard measures form the segmentation. When going further on this direction the machine learning system could even try to predict the outcome of the patient based on data given as input. One example has shown by the Google researchers with chest x-ray imaging algorithm². The first deep learning powered medical image processing tool from Arterys is another good example³. In each of the cases the doctor would check the results and approve them before they would be used in clinical practice.

One interesting topic for the health domain is user group segmentation to identify patients, which need more attention and monitoring than average patients. An advanced AI application case would include optimization of the healthcare path through primary care, hospitals and homecare aiming at improvement in terms of healthcare quality and cost-efficiency. This would be possible with a wide enough view to the healthcare system and good modelling of different care paths throughout the healthcare system. Clearly, these cannot be solved with a single AI method, but they need a well tested system utilizing multiple methods.

A similar challenge to the healthcare path optimization is within drug development and usage. With large data sets and proper AI methods it would be possible to identify which drugs are most suitable to different individual patients based on their treatment history, demographics and genetic burden. This would enable the precision medicine approaches in drug therapy and could be combined with lifestyle changes supported by AI powered coaching tools. AI technologies can also support remote care by monitoring a rich combination of data from the patient during daily activities and vital signs at home and when needed give a warning to the health care personnel to check the status of patient early enough before the health status of the patient is getting down. In this kind of scenario the system would need both real time

¹ <u>MIT review: Google x-ray project shows AI won't replace doctors any time soon</u>

² HealthImaging: Google AI algorithm may improve chest x-ray interpretation, radiologist efficiency

³ PR Newsware: Arterys Receives First FDA Clearance for Broad Oncology Imaging Suite with Deep Learning



measurements but also the patient history and some classification what kind of signs are emphasizing the risk. Humans can monitor single cases on these kind of systems but manual observation of various real-time signs is challenging.



8 Business strategies

In this chapter we identify and analyse emerging opportunities for precision medicine related products and services and develop recommendations for business strategies and public policy measures to accelerate formation of a dynamic business ecosystem in Finland.

Four opportunity areas have been identified from existing research literature and market studies:

- 1. Development of new therapies
- 2. Health risk assessment and diagnostics
- 3. Selecting therapy
- 4. Direct-to-consumer genomics

Each opportunity area covers value creation functions that precision medicine applications may provide to patients, consumers, businesses and society. Value is a benefit or positive impact to user which may be created through various functions: for instance, improving health impact of treatments, more efficient and precise identification of diseases and health conditions, acceleration and streamlining of drug development, or provision of wellness related information that consumers are willing to pay for.

For each opportunity area we analyse demand conditions and to whom the value will be potentially created. Value may be generated to individual patients or consumers, healthcare organisation, or companies such as pharmaceutical industry. Products and services can, to a large extent, be organized in value chains. For example, a medical device manufacturer provides a genotyping or sequencing technology to a diagnostic service provider, which performs genomic tests to health care organisations (public or private) providing clinical care to individual patients. Between each customer and provider (seller and buyer) there needs to be a reasonable proposition where the value of product meets the customer's willingness to pay.

In many cases it is also likely that value can be distributed between several types of actors. For instance, a pharmacogenomics test service might improve patient health by reducing adverse drug reactions while simultaneously creating savings for the healthcare organisation by avoiding provision of extra treatments to deal with undesired side effects. Value creation has, in these kind of cases, systemic properties which extend beyond one-to-one relationship between user and supplier. The systemic value creation potential is particularly large for data which can be used by several users for various purposes and combined with other data multiple times without its value being lost. Data sharing capabilities are thus of particular interest for this analysis.

Third, we review supply of new products, services and solutions which claim to deliver the proposed value. The review is based on analysis of existing and recently introduced products and services, as well as emerging new solutions for which information is publicly available (e.g. scientific publications reporting results from ongoing trials). The review does not aim to be a comprehensive overview of all relevant products and companies currently active in this enormous field. Instead, we present examples of interesting cases on emerging technologies and new business models.

The fourth element in our analysis is drivers and barriers for market creation, value generation and business growth. Drivers may range from economic factors (e.g. cost-effectiveness, high cost of treating chronic diseases) to topical trends (e.g. consumer preferences) and social factors (e.g. safety, privacy). As for barriers we aim to identify bottlenecks which hamper business growth such as factors which are curbing down latent demand (e.g. lack of high quality evidence or missing clinical guidelines), reducing market scalability (e.g. lack of data



interoperability), or hindering and slowing down market evolution (e.g. regulatory requirements not keeping pace with fast developing market developments).

Each section concludes by a discussion on possible business models and by proposing business strategies and public policy measures to accelerate business growth and value capture for Finland. These may include removal or lowering of identified barriers (e.g. regulatory reforms), setting up shared resources (e.g. data platforms), or improving preconditions for collaboration in research and development.

8.1 Development of new therapies

Perhaps the most obvious opportunity area for precision medicine is development of new therapies, particularly novel pharmaceutical drugs. Collection and analysis of genomic and other omics data has the potential to generate several benefits to new drug development process. It may allow more precise targeting of research efforts and accelerating finding of novel mechanisms and associations thus reducing R&D costs. Data may enable more precise patient stratification for clinical trials speeding up the process and reducing costs. According to some estimates genomic information used in identifying target molecules could double the success rate of drug discovery process. Data can also be used to control for the response of drug therapies with different populations around the world. Large opportunities are assessed to be gained when genomic data is combined with other health data such as clinical data describing phenotypes and treatments. Electronic health records (EHRs) contain relevant information to be used for establishing new patient stratification principles and identifying disease correlations (Jensen et al. 2012). Detailed phenotyping is expected to contribute to patient recruiting for clinical trials, and over time, start automating the recruiting process. Linking genetic data with other health data offers thus large untapped opportunities for new drug development.

The opportunity of using health data in development of new products actually extends beyond the pharma companies. The food industry is increasing efforts towards developing personalized food products with favourable effects on health. Such product development can also get considerable benefit from the access to genome, microbiome and phenotype data.

Demand. The demand landscape is divided in two distinct stages along the value chain. First, pharmaceutical companies source various technologies and R&D services to support the drug development process. These include genomic sequencing and genotyping devices and services, bioinformatics, software, computing, analytics, and research services to manage clinical trials. Second, once a new therapy has been introduced on the market its demand is conditioned by regulatory approval, therapeutic and cost-effectiveness assessments and price negotiations with reimbursing organisations. The final demand is hinging on approval from healthcare providers and reimbursing organisations, whose awareness of new therapy opportunities and their clinical impacts affect speed of uptake.

Supply. Data can be supplied by non-commercial and commercial sources as described in Section 4.4. Public or not-for-profit biobanks and large-scale research projects generate noncommercial sources of data. A current trend in the medical research is formation of precompetitive research communities (Palotie & Ripatti 2017). The communities undertake research projects without commitments to immaterial property rights. These sources are expected to provide a wealth of relevant data for precision medicine discoveries and applications.

Commercial sources of data are private clinics, some of which have established their own biobanks, and specialised data providers. Also companies offering direct-to-consumer genetic testing services have been making contracts with pharmaceutical companies about genetic data provision for R&D purposes.



A variety of business opportunities for firms are available in supplying research services to pharmaceutical companies performing drug development projects. There is demand for contract research organisations (CRO's) specialising in genome-based research, data analytics and interpretation services, as well as information systems (Palotie & Ripatti 2017).

In the technology space, the market leaders in advanced sequencing technology are Illumina and Thermo Fisher. In addition, there are several smaller companies engaged in the technology development race. An interesting opportunity is the supply of data management platforms. This involves integration of genomic data, other omics data, and data retrieved from operative healthcare information systems. BC Platforms and BCB Medical are promising Finnish companies in this market segment.

New business models are experimented by companies which compensate individuals for licensing their genomic data to pharmaceutical companies. Genos, based in San Fransisco, provides a service to sequence its customers' exome for a reasonable price (399 US dollars in 2018). Customers can then license their genomic data and receive a financial compensation. The business model of Genos is thus not based simply on sale of genetic testing but rather brokering access to genetic information between individuals as data providers and companies as data users (Roberts et al. 2017).

Drivers. The primary driver behind use of genomic data for drug development is the potential to find new therapies for hereditary diseases, new treatments for cancer, and many other severe diseases. Treatment of severe diseases is costly, creating a strong economic driver for up taking new drugs.

An important technological enabler is the rapidly falling cost of sequencing. In little more than ten years the price of sequencing the whole human genome has come down from 3 billion during the Human Genome project to less than 1000 US dollars currently. The low price of sequencing makes it feasible to gather genomic data from a very large number of people. These kinds of initiatives are now taking place all over the world as described in Section 4.3. One of the large scale projects is the Finnish FinnGen which aims to collect genetic data from 500 000 individuals. Investments in collecting massive genomic data sets in several flagship initiatives around the world are accumulating a wealth of data which drives research and development. Many of these sources operate under non-commercial principles which means that anonymised data is to some extent open for exploitation.

Another major economic driver is the high cost of new drug development. A common estimate in the pharmaceutical industry is that development of new pharmaceutical therapy has an average price tag of roughly 1000 M€. Large part of the costs are accrued in the clinical trial stages. Any opportunities to shorten time to market are highly interesting due to the big associated financial benefits.

Barriers. While there are very strong economic drivers for advancing precision medicine, there are still barriers, many of which are institutional and social by nature. The first is obtaining and managing patient consent for data gathering in reliable way, which meets regulatory requirements and satisfies people's concerns about privacy. The second is combining genomic data with other health data which is still cumbersome as access to electronic health records is restricted and the data not readily in a usable form. Data integration also suffers from legacy information systems with limited support for interoperability standards. Third, there is shortage of highly qualified research service providers who can cater for more comprehensive research management needs of pharmaceutical firms.

An economic issue has come up with many new drugs in the precision medicine space due to their very small target group of patients. New advanced drugs might target rare diseases having only a thousand patients or less in one country. The large R&D investments used to develop the drug and bring it on the market need to be charged from this small population of patients and their healthcare providers or insurers. Revenue should also simultaneously cover



expenses from other unsuccessful projects which did not exit from the R&D pipeline. Recent examples include blindness gene therapy by Spark Therapeutics with a list price of 850 000 US dollars per treatment. Several earlier gene therapies (Provenge, MACI, Glybera, Chondrocelect) introduced on the market have been already withdrawn due to their limited commercial success.

Business models. Development of new drugs sets challenges for conventional business models of pharmaceutical companies as many of the drugs target small populations of patients with rare diseases such as genetic disorders. Covering the high costs of R&D creates a pressure to charge very high unit price per treatment. Pharmaceutical companies increasingly argue for value-based care and outcome-based reimbursement models. Under this approach, high unit price are justified on the basis of the total value provided including e.g. other care costs avoided. At the same time, there are expectations that utilization of genomic data can bring down drug development costs, in which case the pressure to change the current business model would become less stringent.

As for companies that are providing research services for pharma industry, there are numerous business opportunities to cater for their needs. Business models can be based on comprehensive research services to manage clinical trials, analyze associated data and provide real-world evidence. Or they can be provided as specific service inputs to the process on-demand basis.

Bioinformatics, analytics and software solutions can be provided both as 'turnkey' solutions or more targeted inputs to the client's process. As internal capabilities of pharmaceutical companies vary, it is likely that various business models may thrive.

In addition to public healthcare institutions also private healthcare service providers are establishing biobanks. They are exploring opportunities to become data providers to pharmaceutical firms. As an asset, they can combine other health data with genomic data from their registers. Whether this will become a viable business in the future remains to be seen.

Policies and strategies. In order to overcome the barriers there are various policies and strategies that can be adopted and reinforced:

- Regulation to permit and manage secondary use of health data; the respective laws processed by the Parliament of Finland will need to be accompanied by guidance in order to be efficiently applied in practice.
- Establishing a one-stop shop for accessing data from biobanks either via FINBB or the Service Operator (the FINBB has been established, but does not yet have resources to establish such service).
- Sufficient resourcing of the Service Operator enabled by the secondary use legislation.
- Developing capabilities for big data management, bioinformatics, analytics, and artificial intelligence across all stakeholder groups involved; support to small and medium sized companies to specialise in these domains to serve pharmaceutical companies not finding it necessary to build that capability in-house.
- Developing low cost patient consent management services (e.g. as API service)
- Protocols and codes of conduct for sustaining privacy particularly addressing secondary use of health data.
- Establishing and agreeing on standards for data interoperability including metadata and common data models supporting secondary use of data.
- Fast and timely undertaking of cost-effectiveness studies and health technology assessments (HTA's) to accelerate market access for new therapies.



• Development and use of outcome-based reimbursement models and systematic generation of associated outcome data.

8.2 Health risk assessment and diagnostics

Access to genomic data provides opportunities to identify health risks and diagnose genetic diseases. Traditional genetic tests have been long available to find genetic diseases, many of which are rare in number among the whole population. Hereditary diseases, such as Huntigton's disease, are typically single gene mutations which can be tested with targeted genetic tests. The modern next generation sequencing has become recently cost-effective to link person's genotype with predisposition to illnesses or health conditions. This allows for identification of more diverse sets of risks across the entire genome.

Demand. The demand for health risk assessment and diagnostics comes from multiple sources: public health care providers, private care providers, and consumers. Clinical diagnostic tests are typically ordered by a patient's medical service provider. They are used to identify or rule out a specific genetic condition. One example is clinical diagnostics for the BRCA1 and BRCA2 genes affecting hereditary breast cancer and ovarian cancer syndrome (Tandy-Connor 2016).

Comprehensive screening of populations or sub-populations by genotyping is not yet widely undertaken due to insufficient cost-effectiveness. For many diseases and health risks clinicians have conventional means to identify risks which have been found to be more cost-effective than genetic tests. These include queries about the family disease history and conventional measurements such as weight or blood tests etc.

Supply. Examples of diseases for which genetic tests are available are breast cancer and prostate cancer. New genetic tests are becoming available also in Finland. An example is FinnScreen provided by Dextra which includes analysis of 40 Finnish genetic disease risks. The test is a 'carrier screening test' intended for couples for controlling their risk for passing inherited disorders to their children. The test is designed for the Finnish population to cover the most common genetic disorders.

Drivers. A technological driver is evolution of sequencing and genotyping devices from heavy laboratory facilities towards lighter and portable devices. It appears likely that the devices will soon enter point of care and thus closer to clinical settings where diagnostics can be quickly performed. An example of a company developing a hand held sequencing device is Oxford Nanopore from the UK.

Another potential driver is accumulation of genomic data in data repositories and electronic health records. Once the genomic data becomes readily available its unit cost for reuse is likely to drop making it economically more feasible to use genetic data in clinical decision-making as one source of information. While the data is not yet widely available, it is likely that over time it will start accumulating and become available.

A potential institutional driver could be provided by outcome-based healthcare financing models. The proposed transition in the Finnish healthcare reform towards capitation based financing model combined with outcome-based targets might create new incentives for more comprehensive screening of genetically-conditioned risks. Particularly for common illnesses, such as cardio-vascular diseases or diabetes, whose occurrence has a more complex association with genetic variation, more comprehensive responsibility of health status over a regional population could provide a driver for more systematic predictive testing.

Barriers. A barrier to widespread diffusion of genomics-based risk assessment services is divergence of views over providing people with information about risks which might or might not actualise over a long time. Many fear it might create unnecessary agony for factors which



are not under person's influence. It is likely that controversy around the topic will continue for some time. However, recent studies indicate that many people do appreciate information potentially impacting their future health conditions. This appears to apply also to children. According to one recent study a majority of adolescent children would prefer to know the results of unanticipated findings found in genomic testing even if the findings were not medically actionable until adulthood (Hufnagel et al. 2016).

An economic barrier is prevailing doubts about the clinical utility and cost-effectiveness of comprehensive screening programs. In order to roll out comprehensive screening of population strong evidence is needed as cost-effectiveness remains an issue. Moreover, in some cases the association of genotype with a disease applies only for certain sub-population. Breast cancer is a case in point, as it has been assessed cost-effective only for people from particular ethnic groups (D'Andrea et al. 2016).

Business models. As the price of genetic tests and even comprehensive genomic sequencing continues to drop with high speed, diagnostic services are gradually moving closer to the point of care context. Medical device manufacturers' revenue is increasingly coming from consumables and services related to operating devices. In industry forecasts it is argued that sequencing the whole genome of a patient on a routine procedure by healthcare institutions is not very far off in the future. The data would be stored in electronic patient record and retrieved on-demand basis. This would create a large opportunity for genetic test providers.

Policy and strategies.

- Improve the evidence base to demonstrate the clinical utility of genetic tests with costeffectiveness studies.
- Develop capabilities to use genomic data as one source of information and combine it with other health data in the clinical decision-making. Further investments on solution development are needed in order to create clinically sound results in more complex health conditions.

8.3 Selecting therapy

There is now increasing amount of evidence that a patient's genetic profile may affect efficacy of drug therapies. There are currently at least 80 medications whose efficacy to treat specific disease has been demonstrated to be associated on patient's genotype (Relling & Evans 2015). Less than 20 genes out of all 20 000 genes have been found to be affecting drug metabolism. Unsuitable drug or incorrect dosing can cause side effects, which in some cases can be severe, even leading to death. In the United States it has been estimated that more than 100 000 deaths annually are caused by adverse drug reactions. In Germany 3.25 % of all hospitalizations are due to adverse drug reactions generating treatment costs of 434 million Euros per year (Plöthner et al. 2016). Controlling for patients' genotype thus has a potential to increase drug therapy response, reduce adverse drug reactions and decrease health care costs.

One of the first drugs with associated response depending on patient's genetic profile is warfarin which is used for blood thinning. Varying response may lead to too low dosing for some patients and too high dosing for others. Both may have serious health consequences and cost implications.

There is now a large number of diseases for which evidence has been generated: cardiovascular diseases, stroke, epilepsy, depression, high cholesterol, rheumatoid arthritis, tuberculosis, and various types of cancer. New findings are published in accelerating speed.

In cancer treatment, use of genomic information extends beyond simple patient genotyping to genetic profiling of tumours. A majority of cancers are driven by genomic alterations that



with cancer.

It is expected that point of care decisions will become increasingly customized to the unique genomic and proteomic features of a patient's tumour. One example is treatment of thyroid cancer for which identification of benign nodules with help of genomic data can avoid close to 70 % of patients from unnecessary surgery.

A highly specific use case for genetic data is controlling for genetic compatibility in organ transplantation.

Demand. The conventional demand for genetic test diagnostics is related to pre-natal screening and various hereditary diseases. Many new discoveries concerning genetic predisposition to diseases and associations between genotype affecting metabolism and drug response are not yet expressed as tangible demand for genomic data and analytics. As demand is emerging for more advanced precision medicine solutions, the channels through which the demand will materialize as business opportunities for companies offering related products, will become more clear only in the future.

Supply. The supply of services to cater for selecting the most effective therapy taking into account the patient's genotype include stand-alone genetic tests, analytics and interpretation of data, data storage services, integration with electronic health records, and clinical decision-support systems. In Finland there are several firms active in this space such as Abomics and Euformatics.

Drivers. The overriding driver for applying genetic information for selecting the therapy and dosing of drugs is the potential to improve quality of care while simultaneously avoiding costs of dealing with adverse drug reactions or unresponsiveness to therapy. Potential for cost reduction provides a major impetus for applying pharmacogenomics data in clinical use.

Barriers. Cost-effectiveness is not yet confirmed for all discovered associations between person's genotype and response to therapy. Retrospective research would benefit also companies in their marketing efforts by providing evidence-based claims about the value of their products and services. However, technology development and market scale-up are likely to bring the unit cost down over time, improving the cost-effectiveness of pharmacogenomics tests in clinical use.

Interesting findings were provided by Verbelen et al. (2017) who found out that for half of the cases in which association of a drug was associated with patient's genotype, a related pharmacogenomics test was found cost-effective. Moreover, they concluded that when genotype information will be readily available at negligible cost from the electronic health records in the future, a clear majority of economic evaluations would support pharmacogenomics-guided treatments. In other words, once collecting genotype information becomes a standard procedure in the health care process instead of distinct on-demand service request, the cost-effectiveness of pharmacogenomics tests increases significantly and is likely to become a routine operation for most of the identified cases.

Another hindrance is lack of care guidelines addressing the value of genetic data in relation to selecting the therapy and dosing (Heliste et al. 2016). Clinical guidelines are currently developed by a collegium of medical professionals on the basis of a selection process based on relevance assessment. While clinical guidelines can be powerful means to diffuse best practices across the medical profession, lack of guidelines may slow down adoption of new technology and practices.

A barrier for business development based on supplying diagnostic and pharmacogenomics services to public healthcare is a tendency of some public healthcare institutions to develop



these capabilities in-house. Particularly for non-technological parts of the process, such as interpretation and counselling services, non-commercial practice emerging inside the healthcare institution, or by a subsidiary company established by a public hospital, can slow

A further difficulty experienced by precision medicine solution providers is shortage of opportunities to pilot new solutions with public healthcare providers such as university hospitals and demonstrate the value of solutions. Some innovation platforms have been experimented recently, which might open up new possibilities, but practices are still unsystematic and opportunities mostly on ad hoc basis.

down market growth for private supplier firms operating on commercial principles.

Business models. Using genomic information to guide selection of therapy and dosing of drugs creates need for decision support systems to assist clinicians whether a drug is suitable for a patient. Some providers seek close integration with healthcare and hospital information systems. Another ones provide comprehensive services to deliver test results commissioned by clinicians. As development of healthcare information systems with regard to pharmacogenomics is still in an early stage, it is early to assess which model is likely to prevail on the long run. Particularly in oncology it is becoming increasingly relevant to develop companion diagnostics which are tests used as a companion to a therapeutic drug to assess its appicability to a patient.

Policy and strategies

- Increase training and awareness raising among clinicians and health care practitioners.
- Support and accelerate creation of clinical guidelines for pharmacogenomics tests and disseminate related information.
- Produce more evidence about cost-effectiveness of pharmacogenomics tests in real world settings.
- Develop information systems to improve integration of genomic data with patient health record data and to provide advanced decision support for clinicians
- Develop and adopt standard data exchange interfaces to enable interoperability between sources of data, information systems, health care institutions, and centralised electronic health records and databases.
- Build capabilities and secure protocols to store genomic data which can be reused. As a person's genome remains unchangeable over time genomic data can be reused. This will bring down unit cost making it cost-effective to utilise it for various purposes. However, protocols are required to ensure personal data protection and maintain patients' confidence on privacy. The national Genome Centre is planned to provide this kind of centralised capability.
- Develop and adopt platform-based analytics solutions which automate data storage, integration, and analytics.
- Support to experiments and pilots. There are many funding alternatives to financing pilots: R&D grant funding, challenge-based contests, or public procurements of demo solutions.

8.4 Direct-to-consumer genomics

Consumer genetic testing services have been available for some time already. The first services were targeted at studying person's ancestry and soon followed by health-related tests. Recently introduced consumer services now extend beyond medical health domain into wellbeing, nutrition and sports. An example is a test provided to assess optimal sports rehearsal and nutritional program based on a person's genetic type.





Demand. Marketing and sales of genetic testing services are not simply responding to preexisting demand on the market, but also actively aiming to create demand and establish a market (Udesky 2010). The market of direct-to-consumer genetic tests is estimated worth 70 million US dollars in 2015 and expected to grow five-fold by 2022.

The proliferation of direct to consumer genetic services has led to increased demand for genetic counselling services. Their use is recommended by clinical professionals to improve interpretation of results and avoid unnecessary anxiety. The US labour authorities have estimated that the genetic counselling field is expected to grow by 29 percent through 2024. It is also becoming more professionalised with accreditation mechanisms being developed. In Finland, an example of a genetic counselling service is provided by DiagFactor to assess genetic risks for obesity, heart illnesses, and nutritional guidance. The service also covers interpretation of the results. The marketing claim for the service is promotion of proactive healthcare. Another company active in the field is Negen.

Supply. A review published in 2016 found 136 companies that offer consumer genetic testing covering a large variety of tests including paternity tests and child talent tests (Phillips 2016). The typical operating model is provision of genetic tests directly to consumers, mainly via the internet subscription and delivery. The service usually includes provision of a sample kit, return of the sample via post to the service provider, and reception of the results via email or online web portal. Some of the most well known companies providing direct-to-consumer genetic testing services are 23andMe, FamilyTreeDNA, Ancestry.com, and MyHeritage. Services from these companies have been available a little over a decade (since 2007).

In the United States, the Food and Drug Administration (FDA) has approved the first genetic health risk tests to be marketed directly to consumers in 2017. The approved tests cover ten diseases or conditions including Parkinson's disease, late-onset Alzheimer's disease and hereditary thrombolia. In their approval letter the FDA emphasizes that while the tests may provide genetic risk information to consumers, the tests cannot determine a person's overall risk of developing a disease or condition. There are also many other factors that contribute to the development of a health condition, including environmental and lifestyle factors.

Perkin Elmer has recently introduced genetic tests for 59 genes which have been confirmed to play a role in one of 34 health conditions, to all of which there are treatments available. The tests are primarily targeted at healthy individuals. The company will be selling the tests through Helix, a consumer marketplace, bringing together a variety of genetic tests for health and other purposes.

The business model of Helix is based on a platform strategy connecting users and genetic test providers on a single marketplace. For 80 US dollars fee, Helix will sequence the full exome (the portion of the genome responsible for coding proteins) and additionally some noncoding regions of the genome. Helix stores the information in digital form. The customer can then authorize any of Helix's multiple service providers to provide genetic tests. Whereas the first genetic test companies entering the consumer market (e.g. 23andMe) use more limited genotyping technology Helix uses DNA sequencing which yields about 100 times more information. Helix also collects and sequences a DNA sample only once and stores it so that the customer can access additional DNA analysis products without providing another saliva sample. Helix online test shop covers services from health and diet to ancestry, family planning, fitness and even entertainment applications (e.g. wine recommendations based on client's genotype).

Drivers. An essential driver of consumer demand for genetic tests is personal concern about risk for genetically-dependent diseases and health conditions. It has been argued that there is a trend towards decentralization, consumerization and democratization of care delivery models. Access to information over the internet leads to increased awareness and active engagement of people over their health. Another powerful trend impacting consumer behaviour



is increasing desire to measure physical characteristics and underlying factors. The so called 'quantified self' movement is an extreme manifestation of this kind of thinking.

The main technological driver is decreasing cost of testing which has rapidly brought also more extensive whole genome sequencing services to a reasonable price category for larger groups of consumers. Combined with business models where the full cots of testing is not charged from the consumer, but covered by the service provider in return for future revenue for additional sales makes it possible to offer very attractive prices for consumers.

Barriers. The regulatory environment for testing companies has been uncertain during the first decade of their operation. There has been inclarity about how genetic tests for consumers are to be classified and whether they need to comply with regulations for medical devices. The market has taken off initially in the United States where regulation has been rather permissive. In Europe each country has their own regulatory framework in place. Finland has ratified the European Convention on Human Rights and Biomedicine which also regulates use of genetic testing. The agreement specifically prohibits any discrimination against a person on the grounds of genetic heritage and performance of predictive tests for genetic diseases for other reasons than health purposes. It also excludes the modification of the human genome for other than preventive, diagnostic or therapeutic purposes. However, the current legal interpretation of the agreement's application does not cover services provided on the consumer market but only actions performed by healthcare providers and public authorities. In the draft proposal for Genomic Law, genetic tests sold on the consumer market are proposed to fall under consumer legislation and general data protection regulation (GDPR).

It has been argued that since the majority of companies do not currently offer whole-genome or whole-exome scans it means that the utility of test results for the persons tested is inherently rather limited (Phillips 2016). Another criticism is that risk predictions might have no clinical relevance, and might give false alarms or false assurance (Udesky 2010). Results might be also difficult to interpret and should be done together with qualified professionals. Also, concerns have been raised about the quality of some of these tests (Udesky 2010, Tandy-Connor 2018). There has been controversy about the accuracy of the tests as well as how to interpret results correctly (Phillips 2016). A recent validation study found that 40 % of consumer genetic tests gave false positives related to raw genotyping data (Tandy-Connor 2018).

Also data security and personal identity are issues which are seen as potential risks slowing down their widespread use (Phillips 2016). There is a lack of common security standards for ethical concerns and privacy. Potential risks of compromising the privacy protection include loss of donor anonymity and data leaks (Ayday et al. 2015). Also the fact that a person's genotype is shared among close relatives creates an ethical issue that also family members could be identified from the genetic profile data.

A barrier to wellbeing applications is the fact that genetic data alone is not likely to be valuable, but combined with other personal data it is likely to provide one important input to various counselling services. Further product and technology development will be needed to improve capabilities to integrate various types of health data from different sources to enable tailormade and personalised analyses.

Business models. Early genetic tests to consumers have been based on a simple fee for service revenue model. Recently more complex business models are emerging. Some may offer subsidized test prices in return for the right to use personal genomic data. Others experiment with solutions enhancing data security and privacy with new technology such as blockchains. An interesting business model is the platform model where consumers and genetic test providers are linked together on a marketplace. The genetic data submitted by consumers is analyzed and stored by the platform company. The consumer may then authorize any of the service providers to get the data and perform tests ordered by the consumer.



As direct-to-consumer genetic test services are provided mostly online and traded overseas there is an emerging need for genetic counselling to ensure correct interpretation of findings. Some of the firms provide counselling with a fixed staff, others connect a large pool of experts on a part time basis. Professionalisation of genetic counselling is expected. Also some forms of 'gig' economy business models among medical experts might appear in this space.

Policies and strategies

- Establish and apply quality standards for direct-to-consumer tests in a way which easily communicates the reliability and relevance of a test to the consumer. This might also involve more clear distinction between scientifically qualified health tests from other tests with primarily entertainment or lifestyle function.
- Provision of interpretation services of genetic test results by qualified healthcare professionals having skills to assess the results in the context of several other factors such as personal and family medical history.
- Taking advantage of the emergence of platform-based business models for genetic testing; either as a platform owner, or service supplier operating on the platform enabling fast scalability opportunities.



9 Ecosystem model

9.1 System dynamic modelling

In order to examine alternative development paths of a data-driven precision medicine ecosystem, a system dynamics model was developed. System dynamic modelling is an approach to study the behaviour of complex socio-technical systems. In the approach, the structure of the system (i.e. chains of cause and effect and feedback loops) are first conceptualized. Then, the behaviour over time of the system is examined using computer simulation.

Systems described by system dynamics models are typically large, complex, nonlinear sets of differential equations. Theoretical system analysis usually fails in these cases, mainly because linear approximations cannot be used. Top-line structural analysis can be performed (such as analysis of dominating positive and negative feedback loops and their interactions) but theoretical analysis is beyond the grasp of the modeller.

This lack of theoretical analysis tools leaves the theoretical modeller somewhat running on empty; the main methods left are simulation and experimental analysis of data. Nevertheless, the theoretical analysis of structure gives important insight to the potential behavioural patterns of the system, may they be desired or unwanted (e.g., exponential growth, death spiral, rise and collapse, oscillation, balancing buffer and inertia, to name but a few). Structural analysis is a tool for evaluating different behavioural modes and alternative futures. Computer simulations give more insight to dominance of behavioural patterns, leverage variables and quantitative metrics (Ylén and Hölttä 2007)





9.2 Model of precision medicine ecosystem

In the emerging data-driven precision medicine ecosystem, the main stock variables that change through time are the *amount and quality of data*, the level of *research and entrepreneurial activity*, and the size of the *market* based on data-driven precision medicine. These stock variables are key factors for developing a new innovation and business ecosystem for precision medicine. They determine the critical mass for breakthrough in different future



scenarios and their interdepencies create drastically different dynamics. For instance, the amount of data determines the value and usefulness of the database, the level of entrepreneurial activity is the seed for a Finnish precision medicine ecosystem and the size of market correlates to the profit potential in different business models. The limits of usable data and business activity came up in the interviews.

The causalities of ecosystem factors are described in feedback loops, which are shown in Figure 4.



Figure 4. Stock-flow system dynamics model

The model includes the following reinforcing feedbacks related to the growth of an ecosystem based on health data:

- R1: Entrepreneurial and research activities: New research and business activities depend on the availability of consolidated data. An increase in the amount of data can result in an increase in research and business activities. In the model, it is assumed that an increase in research and entrepreneurial activities makes the ecosystem increasingly attractive for other companies, thus increasing the level of research and entrepreneurial activities in the ecosystem further.
- **R2: Market creation:** The creation of a new market (i.e. the emergence of new B2C services based on health data) is dependent on the existing entrepreneurial activities in the ecosystem. Once a market for the new services exists, firms can invest profits from sales further. This increases the level of research and entrepreneurial activities in the ecosystem leading to further growth of the market.
- **R3: Data generation:** Once a market has emerged, it is possible to collect more data from end users in the market. This can then boost new research and business activities and the growth of the market further.



Various stakeholders are prominent in different parts of the ecosystem model:

- Data-owners boosting or hindering the creation and use of data.
- Research service providers are the enablers in research activity.
- Start-ups and established companies provide the entrepreneurial activity.
- Data-users determine the market size.

9.3 Simulation results

A first question addressed using the simulation model is how to speed up the identified reinforcing feedback loops related to investments in business activities, market creation, and the generation of data.

In the model, resources can either be used for acquiring new health data or supporting new businesses in the emerging ecosystem. In the simulation experiments, the effects of different levels of resource allocation are compared. As shown by the results, either a too low (0.1) or too high (0.9) fraction of resources that are invested into data acquisition can slow down the growth of the ecosystem. This means that by boosting dominantly either data acquisition or entrepreneurial activity only affects one part of the ecosystem and the neglected part left for evolutive natural selection develops significantly slower hindering the overall development of the ecosystem.

In addition, in the figure below the dotted lines show how the results change if external support of the ecosystem ends at year 15. In the long term, it is essential that the ecosystem is not dependent on external support (external resources in Figure 4). These regulatory and incentive based support systems are described in detail in *policies and strategies* of Chapter 8.



Figure 5. Effect of resource allocation to data acquisition and supporting business activities

In the next simulation experiment, we consider a situation in which there are two competing ecosystems based on health data. We assume that in ecosystem 1, data acquisition is faster, which can reflect a difference in legislation related to data collection. However, in ecosystem 2 there are more efficient mechanisms for market creation for the use of health data as well as more research service providers. These differences are reflected in different values of the parameters 'Effect of external factors on data acquisition' (impact on data acquisition rate in model), 'Effect of entrepreneurial activity on market growth' (impact on market growth rate in



model), and 'Research service providers' (impact on new research and business activities rate on model).

	Ecosystem1	Ecosystem2
Effect of external factors on data acquisition	0.05	0.03
Effect of entrepreneurial activity on market growth	1	1.5
Research service providers (between 0 and 1)	0.9	1

Ecosystem 1 describes a situation in which data acquisition is boosted by, e.g., legislation and financial benefits but the start-ups and businesses are left on their own whereas in ecosystem 2 data acquisition is driven by market and business forces and the entrepreneurial research activity is dominating factor. Ecosystem 2 corresponds to the consumer driven business scenario described in chapter 8.4.

As indicated by the simulation results, the relative benefit of faster data acquisition results in a faster growth in research and entrepreneurial activity of ecosystem 1 during the initial phases of ecosystem development. However, in the longer term the efficient creation of a market is necessary for sustained growth. This is why in the longer term the research and entrepreneurial activity of ecosystem 2 is higher.



Figure 6. Competition of two rival ecosystems

A third question addressed is how the ecosystem development and value creation is affected by multiple data sources beyond health data. As indicated by the simulation results, reducing the resources allocated to health data and allocating them to non-health data acquisition and business support can increase the growth of the ecosystem that is based on health data. The reason is that due to the reinforcing feedback R3, an emerging market for non-health data can also help increase the health data that is collected. This case also corresponds closely to chapter 8.4.

One example of this is leisure and sport data. All this personal data can be used for precision medicine and by fusing the data and as for instance genetic data is collected for sport it can also be collected for medical use.





Figure 7. Effect of non-health data on ecosystem growth

The system dynamics model developed was used to study how the development of a datadriven precision medicine ecosystem depends on different factors. Specifically, factors affecting the rate of data acquisition, new business and research activities, as well as the growth of a new market were examined. The presented results are initial examples showing how system dynamics models can be used for ecosystem simulation. In future work, the model will be validated and improved based on available data sources. In addition, the model will be used for simulating the effects of alternative policies that aim at accelerating the emergence of the data-driven precision medicine ecosystem.

A more detailed system dynamic model can be created for the patient care level. The model can be used for evaluating effects of more precise diagnosis, tailored treatment solutions, screening procedures, etc. to patient care paths, outcomes, costs, quality of life to name but a few.



This report provides an overview of the Finnish infrastructure, stakeholders and business potential in the field of data-driven precision medicine. In particular, the report focuses on opportunities opened by secondary use of health related data as enabled by biobanks and other data sources. The observations are based on openly available information and interviews of key stakeholders, including representatives of both companies and public sector organisations.

There are several on-going public sector activities targeted at supporting data-driven precision medicine. These include new legislation (e.g. the Secondary Use Act, new Biobank Act and Genome Act), centralized data storages (Kanta services, genome database), services supporting data access (Service Operator, Finnish Biobank co-operative) and collaborative research activity, most notably the FinnGen project. Overall, the Finnish infrastructure is well-developed, internationally competent, and the on-going development activities are widely appreciated. At the same time, there are worries concerning the delayed time schedules of the public activities in setting up the infrastructure. It is also still unclear for the stakeholders, what kind of data and under which conditions will be available for secondary use. Additional guidance for interpreting the legislation will be needed.

The review of international developments in precision medicine revealed a large number of ongoing efforts targeted at collecting genome data via sequencing projects, supporting cooperation between stakeholders and improving access to data resources and processing tools. Also a large amount of genome and other omics data is openly available from several sources. However, open access was observed to normally apply only to molecular level data: access to more useful data resources, which combine omics and phenotype data is restricted to scientific research. Such restrictions could be avoided in the future by portals, which enable individuals to donate their data for open use, although clear evidence of the feasibility of this model is still missing.

Although the market potential of precision medicine is high, most Finnish SME's in the domain are still unprofitable although some of them show growth in turnover. Many SME's have difficulties in initiating co-operation with larger organisations and, in general, the companies felt that public healthcare service providers should be more active in testing and adopting new technologies and innovations introduced by the industry. Apart from a few exceptions, most SME's also could not see direct benefits from the FinnGen for themselves, although the project's high value for pharma industry, academic research and for the visibility of Finland was largely recognized. The pharma sector has a long tradition of exploiting health data e.g. in Real World Evidence (RWE) studies and, therefore, were more positive about getting benefits from biobanks and the Service Operator. It seems evident that the SME's currently providing research services for pharma companies could enlarge their service offerings towards companies in other sectors. The interviews revealed that there is a high potential for exploiting health data in sectors such as healthcare services, medical device manufacturing and food industry, but support from experienced research service partners is needed.

Better availability of health data for secondary use opens up opportunities to Artificial Intelligence based applications. We have identified several potential use cases for AI technology. AI can speedup the data dependent processes and enable professionals to make more accurate decisions faster. It can be used also to find bottlenecks of the processes and to find solutions for solving the problems.

Based on the interviews, it is clear that companies are still in the early phase of developing precision medicine based business. In order to accelerate this process, we have analysed business opportunities in four areas covering new therapy development, health risk assessment and diagnostics, therapy selection and direct-to-consumer genomics. The analysis included estimation of the demand and supply factors in each opportunity area as well as identification of the related drivers and barriers. The analysis resulted in policies and



strategies, which should be adopted to remove the main barriers. Many of the companies would benefit from actions that would accelerate adoption of new precision medicine solutions by users, create preconditions for market shaping and consolidation, and support scaling-up. Various potential actions were identified, such as more systematic practices for undertaking pilots and experiments with healthcare organisations, developing clinical guidelines, improving interoperability and integration capabilities, and generating evidence about cost-effectiveness of precision medicine solutions.

The observations of this study enforce our preconception that successful business in datadriven precision medicine relies on networked stakeholders. In order to understand the mechanisms of the business ecosystem an initial ecosystem model was constructed. The model enables the simulation of different alternative development paths for the data-driven precision medicine ecosystem. Initial simulation results reveal the importance of balanced investment for increasing the amount of available data and for new business and market creation. The early results show how system dynamics models can be used for such ecosystem simulations.

This report provides a basis for further work planned to be carried out in the PreMed project during September 2018 - April 2020. The further activities are planned to include a research study targeted at collecting evidence on the benefits on applying precision medicine, e.g. pharmacogenomics tests, in healthcare. In addition to the scientific objectives the study is expected to provide information about the functionality of data access processes of biobanks and the Service Operator and to pinpoint related bottlenecks. Additionally, the ecosystem simulation model will be validated and improved based on available data sources and new models will be added for simulating the impact of precision medicine applications in specific patient care paths. The project will collaborate closely with the FinnGen project as well as related projects to be started in the framework of the Personalized Health and Artificial Intelligence programs of Business Finland.



References

Ayday, E., De Cristofaro, E., Hubaux, J., Tsudik, G. 2015. Whole genome sequencing: revolutionary medicine or privacy nightmare? Computer 48 (2), 58-66.

Brown, L.C., Lorenz, R.A., Li, J. & Dechairo, B.M. 2017. Economic Utility: Combinatorial Pharmacogenomics and Medication Cost Savings for Mental Health Care in a Primary Care Setting. Clin Ther., 39(3), p. 592-602.

Chalmers, D. & al. 2016. Has the biobank bubble burst? Withstanding the challenges for sustainable biobanking in the digital era. BMC Med Ethics., 17(1):39.

Chassang, G. 2017. The impact of the EU general data protection regulation on scientific research, Ecancermedicalscience, 11, 709.

Chen, M. & al. 2018, Edge cognitive computing based smart healthcare system, Future Generation Computer Systems, Vol 86, p. 403-411.

Danciu, I. 2014. Secondary use of clinical data: The Vanderbilt approach. Journal of Biomedical Informatics, Vol. 52, p. 28-35.

D'Andrea, E., Marzuillo, C., De Vito, C., Di Marco, M., Pitini, E., Vacchio, M., Villari, P. 2016. Which BRCA genetic testing programs are ready for implementation in health care? A systematic review of economic evaluations. Genetics in Medicine 18 (12), 1171-1180.

Dubow, M. & Marjanovic S. 2016. Population-scale sequencing and the future of genomic medicine. Learning from past and present efforts. The RAND Corporation, Santa Monica, California, and Cambridge, UK.

Frost & Sullivan 2017a. Global Precision Medicine Growth Opportunities, Forecast to 2025 -Strategies and Tactics for Accelerating Growth in a Transforming Market. January 2017 Global Transformational Health Research Team at Frost & Sullivan.

Frost & Sullivan 2017b. Global Precision Medicine IT Solutions for Oncology—Companies to Action - Early Movers Expect to Facilitate Evidence-Based Cancer Care. November 2017 Global Transformational Health Research Team at Frost & Sullivan.

Garraway, L. 2013. Genomics-driven oncology: framework for an emerging paradigm. Journal of Clinical Oncology 31 (15), 1806-1814.

Gissler, M., Haukka, J. 2004. Finnish health and social welfare registers in epidemiological research. Norsk Epidemiologi, 14(1), p. 113-120.

Heliste, J., Elenius, K., Niemi, M., Elenius, V. 2016. Miten hyödyntää genomiikkaa potilastyössä? Farmakogenetiikka saapuu klinikkaan. Duodecim 132, 1561-1568.

Hufnagel, S., Martin, L., Cassedy, A., Hopkin, R., Antommaria, A. 2016. Adolescents' preferences regarding disclosure of incidental findings in genomic sequencing that are not medically actionable in childhood. American Journal of Medical Genetics 170 (8), 2083-2088.

Hyppönen, H., Hämäläinen, P., Reponen, J. 2015. E-health and e-welfare of Finland. Checkpoint 2015. Report 18/2015, National Institute for Health and Welfare.

Jarrahi, M. H. 2018, Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making, Business Horizons, 61(4), p. 577-586.

Jensen, P., Jensen, L., Brunak, S. 2012. Mining electronic health records: towards better research applications and clinical care. Nature Reviews Genetics 13, 395-405.



Ji, Y. & al. 2016. Preemptive Pharmacogenomic Testing for Precision Medicine: A Comprehensive Analysis of Five Actionable Pharmacogenomic Genes Using Next-Generation DNA Sequencing and a Customized CYP2D6 Genotyping Cascade. J Mol Diagn., 18(3), p. 438-445.

Li, J. 2017. A Service-Oriented Approach to Interoperable and Secure Personal Health Record Systems. Proceedings of the 2017 IEEE Symposium on Service-Oriented System Engineering (SOSE), p. 38 - 46.

Li, Z. & al. 2018, Thoracic Disease Identification and Localization with Limited Supervision, CoRR, abs/1711.06373.

Meyer, P & al. 2018, Survey on deep learning for radiotherapy, Computers in Biology and Medicine, Vol 98, p. 126-146.

Nimmesgern, E., Benediktsson, I. and Norstedt, I. 2017. Personalized Medicine in Europe. Clinical And Translational Science, 10: p. 61-63.

Palotie, A., Ripatti, S. 2017. Suomi luomassa genomiikan ja terveystiedon internetiä. Duodecim 133, 771-775.

Phillips, A. 2016. Only a click away - DTC genetics for ancestry, love... and more: a view of the business and regulatory landscape. Applied and Translational Genomics 8, 16-22.

Polašek, O. 2013. Future of biobanks – bigger, longer, and more dimensional. Croatian Medical Journal, 54(5), p. 496–500.

Plöthner, M., Ribbentrop, D., Hartman, J. 2016. Cost-effectiveness of pharmacogenomic and pharmacogenetic test-guided personalized therapies: a systematic review of the approved active substances for personalized medicine in Germany. Advances in Therapy 9, 1461-1480.

Relling, M., Evans, W. 2015. Pharmacogenomics in the clinic. Nature 526, 343-350.

Roberts, J., Pereira, S., McGuire, A. 2017. Should you profit from your genome? Nature Biotechnology 35 (1), 18-20.

Soini, S. 2016. Biobanks as a Central Part of the Finnish Growth and Genomic Strategies: How to Balance Privacy in an Innovation Ecosystem?, 44(1), p. 24-34.

Su, P. 2013. Direct-to-Consumer Genetic Testing: A Comprehensive View. Yale J Biol Med. 86(3), p. 359–365.

Tandy-Connor, S., Guiltinan, J., Krempely, K., LaDuca, H., Reineke, P., Gutierrez, S., Gray, P., Davis, B. 2018. False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirmation testing for appropriate patient care. Genetics in Medicine, April 2018.

Udesky, L. 2010. The ethics of direct-to-consumer genetic testing. Lancet 376, 1377-1378.

Verbelen, M., Weale, M., Lewis, C. 2017. Cost-effectiveness of pharmacogenetic-guided treatment: are we there yet? The Pharmacogenomics Journal 17, 395-402.

Ylén, J.-P.; Hölttä, V., 2007. System dynamics - a tool for designing and analysing complex processes. Simulation News Europe, 2007. Vol. 17, nro 1, pp. 27-31.



Term / acronym	Description
AI	Artificial Intelligence
anonymous data	Anonymous data are defined as information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable (Chassang 2017).
B2B	business-to-business
B2C	business-to-consumer
BBMRI	Biobanking and BioMolecular resources Research and Infrastructure
biobank	Biobanks are collections of biological samples and data gathered with the donor's consent for future medical research and product development for healthcare and health promotion purposes ¹ .
CDSS	Clinical Decision Support System
CNN	Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
DSS	Decision Support System
EHR	Electronic Health Record
FDA	Food and Drug Administration
Fimea	Finnish Medicines Agency
FINBB	Finnish Biobank co-operative
FIOH	Finnish Institute of Occupational Health
GDPR	General Data Protection Regulation
HUS	The Hospital District of Helsinki and Uusimaa
ICT	Information and Communications Technology
Isaacus	Project initiated by Sitra with the objective of setting up an organisation to take responsibility of providing access for secondary use of health and wellness data in Finland ² .
Kela	The Social Insurance Institution of Finland
LSTM	Long Short Term Memory
NLP	Natural Language Processing
NN	Neural Network

¹ <u>https://www.biopankki.fi/en/</u> ² <u>https://www.sitra.fi/artikkelit/isaacus-national-health-data-hub/</u>



omics	Refers to a field of study in biology ending in -omics, such as genomics, proteomics or metabolomics. Omics aims at the collective characterization and quantification of pools of biological molecules that translate into the structure, function, and dynamics of an organism or organisms ¹ .
PHR	Personal Health Record
РМ	Precision Medicine
pseudonymisation	Pseudonymisation means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. Pseudonymised data remains personal data but being protected through coding or encryption (Chassang 2017).
secondary use	Secondary use of health and social care data relates to information collected in the course of providing healthcare or social services, but being used for purposes other than the patient care, i.e. research and development. ²
Secondary Use Act	Refers to the national legislation on secondary use of health and social data (currently in acceptance process of the Parliament of Finland).
Service Operator	National operator for providing access to data for secondary use as enabled by the Secondary Use Act (currently in set-up phase).
Sitra	The Finnish Innovation Fund
STM	Ministry of Social Affairs and Health
THL	National Institute for Health and Welfare
Valvira	National Supervisory Authority for Welfare and Health
VRK	Population Register Centre
VSSHP	Hospital District of Southwest Finland

 ¹ <u>https://en.wikipedia.org/wiki/Omics</u>
 ² <u>Deloitte: International review on the secondary use of health and social care data and applicable legislation</u>



Appendix 1. Precision medicine initiatives

Country	Project	Budget	Sequencing goals	Aims and objectives
African continent	Human Heredity and Health in Africa (H3Africa) ¹ 2010– ongoing	~32 M€ (2011- 2016)	N/A	 To create a pan-continental network of laboratories Ensuring access to relevant technology Clinical integration Facilitating training
China	China Precision Medicine Initiative ² 2016-2030	~8000 M€	2000 individuals (initial), aims at 100M in 2030	 To fund national precision medicine projects
Denmark	National strategy for personalised medicine ³ (2017-2020)	~13 M€ (initial invest.)	N/A	 Establish a foundation for the development of better and more targeted health care for patients, through the use of new technologies and new knowledge Strengthen the ethical, legal and safety aspects related to the use of genetic in-formation in health care Establish a joint governance structure and strengthen collaborations across the country – both in healthcare and research Establish a cooperation about a safe, joint and coherent technological infrastructure Initiate relevant research – and development projects
Estonia	Estonian Biobank EGCUT ⁴ (2000-ongoing)	5 M€ (in 2018)	100 000 in 2018, already contains 52 000	 International ambition to be a leading genome center in Europe Advancing science, conducting research and education Gather population health data and hereditary information to improve public health Be the main force in the development of personalised medicine in Estonia

⁴ <u>https://www.geenivaramu.ee/en</u>

¹ https://h3africa.org/

 ² <u>https://www.weforum.org/agenda/2017/11/3-ways-china-is-leading-the-way-in-precision-medicine/</u>
 ³ <u>http://www.sum.dk/English/Healthcare-in-Denmark-An-Overview/Personalised-Medicine.aspx</u>

https://www.genomeweb.com/genetic-research/estonia-invests-5m-genotype-100k-people-2018-partpersonalized-medicine-project

https://www.geenivaramu.ee/en/about-us/egcut-development-plan-2015-2021



Country	Project	Budget	Sequencing goals	Aims and objectives
Finland	Finland's genome strategy implementation ¹ (2015-ongoing)	N/A	N/A	 Make Finnish healthcare more effective through making genomic data more widely used for and interconnected between individuals, healthcare and research. Transform Finland into an attractive environment for research and business in genomics
Finland	FinnGen ² (2017-2023)	59 M€	500 000 individuals by using samples from national biobanks	 Produce medical innovations Support Finland to become a pioneer in biomedicine and personalised healthcare. Create a co-operation model between public sector and healthcare industry. Provide early access to new personalised treatments and health innovations for all Finns.
France	The French Plan for Genomic Medicine 2025 ³ (2016-2025)	670 M€	235 000 patients annually by 2020	 Position France as a leading country in PM PM integration into the clinical pathway and management of common diseases Create a national genomic medicine framework that facilitates innovation, capitalization and economic growth
Japan	Japan Initiative on Rare and Undiagnosed Diseases (IRUD) ⁴ (2014-2018)	~10 M€ (2014)	N/A	To create a comprehensive medical network and an internationally compatible data-sharing framework
Netherlands	Hartwig Medical Foundation ⁵ (2015-ongoing)	N/A	Whole genome sequencing data 40 per week, currently 2700	 To enable personalized treatment for cancer patients Set up a databank to increase diagnosis and treatment quality, reduce costs of care and support clinical integration

¹ <u>https://media.sitra.fi/2017/02/28142539/Improving health trough the use of genomic data.pdf</u>

² https://www.finngen.fi/en

³ http://solidarites-sante.gouv.fr/IMG/pdf/genomic_medicine_france_2025.pdf

⁴ <u>https://www.amed.go.jp/en/program/IRUD/</u>

https://www.nature.com/articles/ejhg2017106

https://www.ncbi.nlm.nih.gov/pubmed/28794428

⁵ <u>https://www.hartwigmedicalfoundation.nl/en/</u>



Country	Project	Budget	Sequencing goals	Aims and objectives
New Zealand	Precision Driven Health Initiative ¹ (2016-ongoing)	~22 M€ (7year investment)	N/A	 Accelerate research Utilize various data sources, incl. big data Support self-management technologies
South Korea	The South Korean Cohort Study ² (2017-ongoing)	~48 M€	10 000 Cancer patients	 To develop targeted therapies for cancer patients Developing a cloud-based hospital information system
Sweden	Genomic Medicine Sweden (GMS) ³ (2017-2023)	~390 k€ (pre analysis phase)	25 000 genomes annually	 To build an infrastructure to facilitate PM integration into the Swedish healthcare system
Switzerland	Swiss Personalized Health Network (SPHN) ⁴ (2017-2020)	~60 M€ (2017- 2020)	N/A	 Accelerate development of PM Facilitates research projects, data sharing and infrastructure projects
UK	Genomics England ^{5,6} (2013-ongoing)	~515 M€	100 000 whole genomes from ~70 000 individuals	 Implements the 100 000 Genomes Project Create a new medicine service for NHS Accelerate medical research
UK	The Scottish Genomes Partnership ⁷ (2015-ongoing)	~17 M€ (initial investment in 2015)	1377 genomes sequenced in May 2018	• The partnership leads four research areas: cancer, genetic diseases, a Scottish population study, and it collaborates with Genomics England
USA	Precision medicine initiative and All of Us Research Program ⁸ (2015-ongoing)	~1252 M€ (2018- 2028)	1 million volunteers	 To build a national research participant group To develop precision medicine in oncology To start the research project called All of Us.

- ³ http://swelife.se/vi-erbjuder/fornyelse/gms-genomic-medicine-sweden/
- https://ki.se/en/mmk/new-national-initiative-for-precision-medicine-genomic-medicine-sweden
- https://www.equalis.se/media/127583/lucia-cavelier_genomics-medicine-sweden.pdf

¹ <u>http://www.precisiondrivenhealth.com/</u>

² <u>http://www.koreabiomed.com/news/articleView.html?idxno=1267</u>

⁴ <u>https://www.sphn.ch/en.html</u>

⁵ https://www.genomicsengland.co.uk/

⁶ https://cnfl.extge.co.uk/display/SDKB/100%2C000+Genomes+Project+-

⁺Frequently+Asked+Questions

⁷ <u>https://www.scottishgenomespartnership.org/</u>

⁸ <u>https://ghr.nlm.nih.gov/primer/precisionmedicine/initiative</u>

https://allofus.nih.gov/



Country	Project	Budget	Sequencing goals	Aims and objectives
USA	Million Veteran Program ¹	N/A Currently ~600 000 (aims	Investigate how genes affect healthBuild worlds largest database with	
	(2009-ongoing)		at 1 million)	blood samples and health information

¹ <u>https://www.research.va.gov/mvp/</u> <u>https://www.sciencedirect.com/science/article/pii/S0895435615004448</u>



Appendix 2. Non-profit precision medicine organisations

Country	Project	Aims and objectives
Europe	International Consortium for Personalised Medicine (ICPerMed) (2016) ¹	 Facilitates collaboration in personalized medicine research, implementation and funding. Accelerate research by coordinated approaches Support research into benefits of personalized medicine with emphasis on healthcare systems and citizens. Support PM approaches for citizens
Europe	European Personalised Medicine Association (2009) ²	 Promotes adoption of and access to personalized medicine, promotes advanced diagnostics
Europe	European Alliance for Personalised Medicine (2012-ongoing) ³	 Promotes development, adoption of and access to personalized medicine, promotes diagnostics
USA	The Personalized Medicine Coalition (PMC) ⁴ (2005)	 PM advocacy group; raising awareness, education for healthcare workforce, policymakers. Influencing regulatory frameworks, public policies to invest in PM products
Global	Global Alliance for Genomics and Health ⁵ (2013)	 Non-profit alliance working to create frameworks and standards for responsible, voluntary and secure sharing of genomic and health related data
Global	Global Genomic Medicine Collaborative ⁶ (2016)	 Create networks and collaborations Provide a knowledge base for best practices and genomic medicine activities Assist in clinical implementation and evidence generation

¹ <u>https://www.icpermed.eu/</u>
² <u>http://www.epemed.org</u>
³ <u>https://www.euapm.eu/</u>
⁴ <u>www.personalizedmedicinecoalition.org</u>
⁵ <u>https://www.ga4gh.org/</u>
⁶ <u>https://g2mc.org/</u>



Appendix 3. Data access, information sharing and collaboration structures

Organisation	Content	Data
AHA Precision medicine platform ¹ (USA)	Cloud-based data resource for cardiovascular diseases. Provides open access to cardiovascular data sets, tools and forums for collaboration.	Clinical health data, lifestyle data, genetic data, cardiovascular and stroke data.
Canadian Open Genetics Repository (COGR) ²	Collaborative effort for the collection, storage, sharing and robust analysis of variants reported by medical diagnostics laboratories across Canada.	Information related to human gene DNA variants and their relationship to disease.
ELIXIR ³ (Europe)	Provides infrastructure to enable public sharing of genetic data, services for permanent archiving and distribution of personally identifiable genetic and phenotypic data.	Genetic and phenotypic data.
European Bio Informatics institute (EMBL-EBI) ⁴	Shares data (e.g. nucleotide sequences) and tools originating from life science experiments.	Biological data sources covering: DNA&RNA, Gene Expressions, Proteins, Structures, systems, chemical biology, and ontologies.
European Genome- phenome Archive (EGA) ⁵	An EMBL-EBI service for permanent archiving and distribution of personally identifiable genetic and phenotypic data resulting from biomedical research projects.	Personally identifiable genetic and phenotypic data.
Global Alliance for Genomics and Health (GA4GH) ⁶	Non-profit alliance that works to create frameworks and standards to enable sharing of genomic and health related data. Develops standards, tools, and frameworks that are designed to overcome technical and regulatory hurdles to international genomic data sharing.	-
International Cancer Genome Consortium (ICGC) ⁷	Coordinates a large amount of research projects globally to generate comprehensive catalogues of genomic abnormalities. Provides tools for visualizing, querying and downloading the data released quarterly by the consortium's member projects.	Cancer genomes at the genomic, epigenetic and transcriptomic levels.
PrecisionFDA ⁸ (USA)	PrecisionFDA is a research sandbox that provides the genomics community with a web portal where they can experiment, share data and tools, collaborate, and define standards for evaluating analytical pipelines.	-

¹ <u>https://precision.heart.org/</u>

- <u>https://www.ehkh-europe.org/</u>
 <u>https://www.ebi.ac.uk/</u>
 <u>https://www.ebi.ac.uk/ega/home</u>
 <u>https://www.ga4gh.org/</u>

- ⁷ <u>http://icgc.org/</u>
 ⁸ <u>https://precision.fda.gov/</u>

² http://opengenetics.ca/

³ <u>https://www.elixir-europe.org/</u>



Organisation	Content	Data
The National Cancer Institute (NCI) Genomic Data Commons (GDC) ¹	The NCI's Genomic Data Commons (GDC) provides the cancer research community with a unified data repository that enables data sharing across cancer genomic studies in support of precision medicine.	Cancer genomic data including mutations, copy number variants, mRNA and miRNA sequence data, post-transcriptional modifications.
The National Genomics Infrastructure (NGI) ² (Sweden)	Provides access to technology for massively parallel/next generation DNA sequencing, genotyping at all scales and associated bioinformatics support to researchers based in Sweden.	-
The Personal Genome Project ³ (International)	A coalition of projects across the world dedicated to creating public genome, health, and trait data.	Genome, health, and trait data.

https://gdc.cancer.gov/
 https://ngisweden.scilifelab.se/about
 https://www.personalgenomes.org/


Appendix 4. Data access guidance comparison

Organisation	Content	Guidance for data access
Auria Biobank ¹	Provides biological samples collected in connection with normal health-care and medical examinations for use in medical research.	Instructions and forms available for applying data.
Helsinki Biobank ²	Helsinki Biobank provides samples with associated clinical information, for medical research and R&D purposes.	Instructions available for applying data (forms for availability requests).
Turku Clinical Research Centre ³	Clinical informatics service responsible for maintaining the hospital district's patient information and making it available for research purposes.	Instructions for access published.
THL Biobank ⁴	THL Biobank hosts a remarkable collection of population and disease-specific samples for research purposes.	Instructions and prices available for applying data.
China Kadoorie Biobank⁵	Population based biobank with questionnaire data and physical measurements, blood samples and health outcome data.	Instructions and criteria for access published.
UK Biobank ⁶	Collection of questionnaire data, blood, urine and saliva samples and healthcare follow-up data of 500 000 individuals. Includes EHR, lifestyle, health and genotype data.	Instructions and costs for access published.
Norway HUNT ⁷	Questionnaire data and biological specimens from the general population.	Instructions and costs for access published. Norwegian research partner required.
Norwegian Institute of Public Health ⁸	Joint service for accessing data from health registries and health studies as well as samples from Norwegian biobanks.	Instructions and costs for access published.
Cancer registry of Norway / Janus Serum Bank	Population based biobank for cancer research.	Instructions for access published. Norwegian research partner required.
Danish National Biobank	National biobank with large sample collection and information about samples elsewhere in the Danish health system.	Instructions for access published. Danish research partner needed.
Biobank Sweden	Centralized service point for Swedish biobanks.	Instructions and forms available for applying data (application shall be directed to the specific biobank owning the sample/data).

- ¹ <u>https://www.auriabiopankki.fi</u>
- ² https://www.terveyskyla.fi/helsinginbiopankki/en
- ² https://www.terveyskyla.fi/heisinginbiopankki/en
 ³ http://www.turkucrc.fi/en
 ⁴ https://thl.fi/en/web/thl-biobank/for-researchers
 ⁵ http://www.ckbiobank.org/site/
 ⁶ http://www.ukbiobank.ac.uk/
 ⁷ https://www.ntnu.edu/hunt/data
 ⁸ https://www.fhi.no/en/



Organisation	Content	Guidance for data access
LifeGene Sweden ¹	Prospective cohort study including periodically collected questionnaire data, physical measurements and samples. Also includes health registry data.	Instructions and costs for access published. Association with Swedish university or research institute required.
The Estonian Biobank (University of Tartu) ²	Population based biobank including a cohort of > 50 000 people. Collects i.e. health, lifestyle, demographic, and genetic data.	Instructions for access published.
Hatwig Medical foundation (Netherlands) ³	A database of biopsy samples and clinical parameters of patients recruited for studies of the Center for Personalized Cancer Treatment (CPCT). Collects genetic and clinical data from cancer patients.	Instructions for access published.
Biobank Graz ⁴	Comprises both population-based and disease- focused collections of human biological material.	Instructions and forms available for applying data.

https://www.lifegene.se/
 https://www.geenivaramu.ee/en
 https://www.hartwigmedicalfoundation.nl/en/
 http://biobank.medunigraz.at/en/



Appendix 5. Experts involved in interviews and informal discussions

Jaana Ahlamaa, Medaffcon Johanna Arola, Finnish Biobank Klaus Breitholtz, Genevia Henrik Edgren, MediSapiens Päivi Eriksson, University of Eastern Finland Jari Forsström, Abomics Merja Auvinen, Diagfactor Jarmo Hahl, Medaffcon Kari Harno, University of Eastern Finland Kirsi Heikola, Novartis Reini Hurme, Zora Biosciences Pekka Ihalmo, Janssen-Cilag Johanna Hemdahl, Boehringer-Ingelheim Lila Kallio, Auria Biobank Antti Karlsson, Auria Biobank Jarmo Kaukua, Boehringer-Ingelheim Ossi Koskinen, Lääkärikeskus Aava Ilkka Kunnamo, Duodecim Kalle Kuusisto, Lähitapiola Marika Laaksonen, Fazer Ossi Laukkanen, Mehiläinen Tatu Laurila, Novartis Joel Lehikoinen, Kaiku Health Anne Lehtonen, AbbVie Tommi Lehtonen, Blueprint Genetics Amy Levál, Janssen Nordic Anssi Linnakivi, Roche Pekka Luukkainen, Lähitapiola Kristina Hotakainen. Mehiläinen Christer Nordstedt, Orion Harri Okkonen, Quattro Folia Seppo Orsila, Modulight Arto Pakkalin, Bayer Jaakko Parkkinen, Pfizer Jukka Partanen, Finnish Red Cross Blood Service Eero Punkka, Helsinki Biobank Christophe Roos, Euformatics Ville Salaspuro, Mediconsult Jaana Santaholma, Novartis Tero Silvola, BC Platforms Jukka Suovanen, Odum Tuula Tiihonen, Sitra Ilpo Tolonen, Docrates Juha Tuominen, Terveystalo Ingmar Wester, Raisio Hanna Viertiö-Oja, GE Healthcare Tom Wiklund, Docrates Petteri Viljanen, BCB Medical Arho Virkki, Turku University Hospital Maija Wolf, Medaffcon Arto Vuori, National Institute for Health and Welfare (THL) Anneli Vuorinen, Crown CRO Tero Ylisaukko-Oja, MedEngine





Appendix 6. Interview questions

Tentative interview questions (list adapted case-by-case for the interviewed organisation).

Understanding Company

Customers, products and price? Who are your main competitors or partners? What is your strategic direction and timeline for growth? What geographical areas are you planning to operate in? How do you market your products or services? What is your core IPR/IP? What is the role of scientific research or research collaboration in your Company? Value Chain How do you define the value chain in this area now and in the future? What is your Company's position in the value chain? How do you see the market size in the future? Is the market formation dependent on public or private sector? Are there sufficient industrial actors in Finland? What is missing? Is there enough knowledge transfer between users and industry? What about science and industry? What needs and assets does your Company have for growth in precision medicine?

Health Data

Do you or are you planning to exploit personal health data and/or genome data? Which data is exploited and how (e.g. in R&D, as part of final product, ...)? How do you access (or plan to access) the data it needs? Which are the main obstacles currently in accessing data (legal, IT, permission process, competence..) Do you expect to benefit from the planned legislation on the secondary use of health and wellness data? How? What partnerships do you need for exploitation of health data (clinical, biobanks, research organisations...)?

Do you have any plans for the analysis of data, or data enrichment?

What kind of trends are foreseen in 3-5 years related to health data exploitation?

FinnGen

How would you understand 'Finnish precision medicine platform'? Importance of FinnGen platform for future business development What stumbling blocks or showstoppers can be foreseen? What competence and type of resource will be needed? How should the Finnish ecosystem be formed and strengthened? How should FinnGen be utilized?