

# The secondary use of the Finnish national health data repository Kanta – opportunities and obstacles

Emmi Eronen<sup>1</sup>, Alpo Värri<sup>2</sup>, Miika Järvinen<sup>1</sup>

<sup>1</sup> Atostek Oy, Tampere, Finland; <sup>2</sup> Faculty of Medicine and Health Technology, Tampere University, Finland

**Alpo Värri, Dr.Tech., Faculty of Medicine and Health Technology, Tampere University, FI-33014 Tampere University, FINLAND. Email: [Alpo.Varri@tuni.fi](mailto:Alpo.Varri@tuni.fi)**

## Abstract

The Finnish law about the secondary use of health data has made it possible to use also the data in the Kanta national health and social welfare data repository, the Kanta Services for secondary purposes such as scientific research, innovation and development, education, statistics and information management. The Kanta Services contain also a personal health record data repository, the Kanta PHR. The large amount of health data in the Kanta Services makes it an interesting source for various secondary purposes.

This paper studies the opportunities, but also obstacles to the secondary use of Kanta data. The study focuses on two use cases, the Kanta PHR data use and the Patient Data Repository use for disease risk predictor development. The topic is studied by analysing laws, Kanta system documents, scientific literature, current discussions of the topic in public and three expert interviews. The literature search analysed for relevance all articles with keyword Kanta in the Finnish Journal of eHealth and eWelfare since year 2024 and also other scientific literature discussing Kanta Services and secondary use.

The result of the Kanta PHR use case study was that although the EU General Data Protection Regulation would allow the secondary use of the Kanta PHR data based on consent, the current Kanta Services applications do not support this at all. Disease risk predictor development is possible with Kanta patient data repository data in scientific research, but innovation and development activities face administrative difficulties because only aggregated Kanta data is allowed to use for these purposes. Information management can make use of aggregated Kanta Services data and it can be used for benchmarking the quality of a health delivery unit to other similar units in Finland.

While protecting the privacy of the individuals' data in the Kanta Services, improvements to the secondary use law are suggested. These include the easier access to non-aggregated health data for research and development, possibilities to individuals to consent the secondary use of their Kanta PHR data and allowing the use of risk calculations for automatically inviting persons at risk to health check-ups. It is also recommended that lessons are learned from the Finnish secondary use law and the European Union European Health Data Space law allows more flexible use of health data for secondary purposes from those individuals that allow it.

*Published under a CC BY 4.0 license (<https://creativecommons.org/licenses/by/4.0/>).*

**Keywords:** Kanta services, laws, health informatics, health information management

## Introduction

Since its beginning in 2010, the functionality and data content of the Finnish national electronic health services, the Kanta Services has continuously increased [1]. The Kanta Services include, among others, the Patient Data Repository, the Prescription service, MyKanta interface to the patients, and the personal health record, Kanta PHR where users can store some personal health data. Starting in 2018, also the social welfare data of their customers have begun to be included to the Kanta Client Data Repository for Social Welfare Services. The data to Kanta Services come from electronic health record systems of the public and private health care providers as required by the law. The Kanta Patient Data Repository acts as a health information exchange between the health information systems of various health delivery organisations. Kanta and its detailed data content specifications ensure the interoperability of the most common health data of the patients as the connecting systems have to implement the Kanta interfaces for input and output.

The health data in the Kanta Services belongs to the following classes: Essential health information includes diagnoses, risk information, performed procedures, physiological measurements, vaccinations, imaging studies, and laboratory measurements. Some specialities have more data specifications and possibilities to store information than others. These are optometry, oral health care, and emergency care. Medication information and prescriptions are also included. The repository contains also more administrative information about the patients like the assessment of the need for care, appointments, care plans, referral and discharge messages, etc. The Kanta PHR contents, including health and wellness information, have also

been specified but not much such data has been stored to the national database yet. The essential health information is stored using HL7 Clinical Document Architecture Release 2 (CDA R2) structures and the PHR data using HL7 FHIR structures. More detailed information about the contents of the data in these classes can be found from the specification documents in the web pages for system developers [2] and the Kanta PHR specifications [3]. It is worth noting that a part of the information is in narrative text form, but only a part of the information is in structural form, e.g. laboratory results. The proportion of structural data is on the increase.

It should be noted that the amount of health data about a person in the Kanta Services is dependent on how much health services the individual has used. Every visit to the services generates an entry to the records, but a very healthy individual may have only COVID-19 vaccination entries and dental examination records in the system because most of the people took this vaccination. The paper by Männikö et al. presents quantitative results about the amounts of different types of health information in the Kanta Services [4].

Until now, the patient data in the Kanta Services have been used mainly for the primary use, i.e. for the health care of the individual patients. Since 2019, the Finnish law on the secondary use of health information has allowed access also to the information in the Kanta services for secondary purposes [5] to improve healthcare and support research and innovation activities. These secondary uses include statistics, scientific research, development and innovation activities, teaching, information management, government guidance and supervision of social and health care, and the planning and investigation tasks of the civil servants. At

the same time, the privacy of the patient data and the confidence of the public to the privacy of the data was also an important objective. The secondary uses Kanta and other health data are explained to the public in [6,7], and [8] including information about how the patient can limit the uses of Kanta and other health data in secondary use. This law supplements the European Union General Data Protection Regulation (GDPR) [9] enabling more secondary use of health data than is usual in many other European Union countries. The large amount of data in the Kanta Services makes it a potentially interesting source of training artificial intelligence (AI) applications for health use. The secondary uses are administered by the Finnish Social and Health Data Permit Authority Findata from whom the access requests are applied. The access requests need to be detailed so that Findata can include correct data fields to the result data set to be analysed in a secure computing environment if the access request is accepted. [10]

The objective of the law about the secondary use of health data was to simplify the administrative processes and thus enable the easier utilization of personal-level customer data for the above-mentioned secondary uses. The objectives of this study can be formulated as follows:

1. How well does this law and the present Kanta Services implementation enable the secondary use of health data in practise, particularly in the development of health risk calculators, how does it enable private enterprises to offer products and services based on information obtained from Kanta data, and how does it improve the health care customers' own use?
2. What changes should be made to better achieve the objectives of the law in the areas of objective 1?

When large amounts of patient data are available, it is interesting to calculate various statistics about it. One useful form of statistics is to develop risk predictors based on a disease outbreak and the previous health data of the same patient. Examples of such risk calculators include the FINRISKI cardiac event or stroke risk calculator [11], the diabetes risk score [12], the breast cancer risk calculator [13], and the coronary risk estimator [14].

### Material and methods

The material for this study consists of various national and European Union level laws and materials provided by the Finnish administrative organisations such as the Ministry of Social Affairs and Health, Institute of Health and Social Welfare, the Social Insurance Institution of Finland (Kela), the materials related to the Kanta Services, and some scientific literature. In the beginning, a systematic literature review was not carried out because the topic is so new and the essential literature about the national Finnish Kanta system is so scarce about the topic of this paper that the systematic review would not have brought significant results.

In a late stage of the study, an additional literature search was carried out. As the Kanta Services and the law of secondary use of health data [5] are in Finland, the search focused first to the Finnish Journal of eHealth and eWelfare. Using the broad search term "Kanta" and limiting the time range of the articles to years 2014-2025 (starting five years earlier than the law of secondary use of health data) the search resulted in 192 articles. Three of them were relevant to this study. One more article was found by using Andor literature search requiring both the search terms "Kanta" and "secondary" in the same article.

In addition to the literature searches, current discussions in media, blog texts and comments of the

topic in different platforms, like LinkedIn have been followed to gain insights to the topic. Each source was typically reviewed by one of the authors. The number of comments in the media and blog texts have not been counted.

Three experts were interviewed by the first author. A senior inspector from the Finnish Office of the Data Protection Ombudsman was interviewed by e-mail. A research specialist from Tampere University who was involved in the preparation of the university's statement of the law of secondary use [5] was interviewed about the interpretation and revision of the law. The third interviewee was one of the authors of the medical risk assessment paper [4] because he worked in the same company as two of the authors of this paper. Based on the use of these sources, an overall impression about the opportunities of the secondary use of health data in Kanta was formed, and they are illustrated through the health risk predictor uses case. The thought process and discussions in the research group resulted also in suggestions to improve the situation in favour of improved healthcare services in Finland.

## Results

### *Personal health record data use*

The availability of Kanta PHR data to health and social welfare service providers based on the consent of the individual is scheduled to be available in the beginning of the second half of the year 2025 [15]. Concerning the secondary use, the Act on the Processing of Social and Health Care Customer Data (2023/703) 73§ forbids the other uses of the Kanta PHR data. The possibility of giving a consent to other data processing possibilities is not given in this law. The GDPR does, however, allow the processing of personal data based on consent. This opens possibilities for the development

of services based on the Kanta PHR data. This requires that there exists a mechanism by which this consent can be given easily from the people's user interface to the Kanta PHR.

There are, however, obstacles in making use of the Kanta PHR data for secondary purposes, related to the input of such data to the PHR. Currently, there are two ways of entering data to the PHR: manual self-entry of the height, weight, heart rate, blood pressure, and blood sugar level data to the system and the use of certified well-being apps in entering the data. The people have currently little incentives to manually enter the data to the Kanta PHR because they can only view those data in the PHR themselves and nothing else. A notebook could do the same without the need to log in to anything. Rinnetmäki explained in 2023 why there are currently no wellness apps that would feed data to the Kanta PHR [16]. The main reasons were that the sharing of the data in the Kanta PHR was only intended to health professionals (in future), it was impossible to use the Kanta PHR for children's data, it cannot give alerts in real time and more importantly from the app suppliers' point of view, it is not profitable. Additional costs were introduced to suppliers when the cybersecurity certification became compulsory in 2023, and this resulted in the withdrawal of all participating apps from the Kanta PHR ecosystem. The latest information is that Kanta PHR has been set to "maintenance state" meaning that no new developments are expected to it until the new European Health Data Space (EHDS) regulation [17] enters into force [18].

### *Patient data repository use study through risk prediction use case*

The second idea of using the Kanta data is related to various risk calculations in regional health care settings. These calculations could serve two purposes. First, they could be used to alert individuals

about a heightened risk of a certain disease. This could motivate the individual to change their living habits towards a less risky lifestyle. The second use could have been the prediction of the need for certain types of health services to the regional population based on the estimated risks. Urhonen studied the possibilities of using Kanta data for individual cardiological risk predictions [19]. The risk calculator reached moderate results, but the calculations were often hampered by missing data from the individuals. Although the data set contained 96200 persons – a sample size exceptionally large for Kanta-based scientific studies, there were still not enough cardiovascular disease patients for risk predictor building according to Urhonen. Additional problems were related to the unavailability of death information in Kanta, short length of medical histories and the heterogeneity of the data.

The consumption of alcohol and smoking are often related to elevated risks in many diseases. The Kanta records do not typically contain this information, or it is mentioned in some narrative text section, not in structured form. The habits of the individual may also change along the years and a six-year-old piece of information may no longer be valid. The individual could, in principle, enter the smoking and alcohol consumption data to the Kanta PHR section of the Kanta Services, but this is currently not possible.

Concerning the second use, population morbidity prediction, the situation is somewhat different. The law allows the regional health service to get only aggregated health data for information management and resource planning from the Kanta Services. This aggregated data may not be sufficient for reliable risk calculations. A solution could be that the risk calculators are developed in scientific research based on an even larger dataset than that available in Urhonen's study [19] and that the

Findata computing environment would calculate these risk predictions automatically. The disadvantage would be that a data access request would have to be done to Findata every time that the risk calculations are needed, and a delay of six to ten months would result from Findata handling the request and getting the results. The region might want to get the information near real-time instead.

Two other risk prediction approaches were found from the literature. Immonen et al. envisaged in 2019 that Kanta data could potentially be used for fall risk prediction among the elderly [20]. As the secondary use law [5] was not in force at that time, the authors could not test the idea further. Mänikkö et al. studied the possibility to use data stored in Kanta as a source to the FINRISK [11] risk calculator [21]. They concluded that that population level risk assessment is not (yet) viable based on only the structured data in the Kanta Patient Data Repository (PDR). Additional structured data fields would be needed to the FINRISK calculations, and these fields would need to be filled in, too.

The regional health delivery organisation might want to use the risk calculator to call individuals to health check-ups based on automatic risk calculations running in the background. The law allows only the data collected by the region to be used as input to such a service, not the data in the Kanta Services. Moreover, there is another obstacle for this kind of activity. This kind of automatic decision-making of calling only some patients to health check-ups may infringe the rights of those individuals who are not invited to check-ups and such an activity is not allowed in this form [22].

When the data of a single individual is used in the prediction of the health outcome of that individual, the system that makes the prediction is a medical device according to the definition of a medical device in the European Union Medical Device

Regulation (MDR) [23]. This requires that the organisation that produces the risk calculator to the health care organisation must have a quality system in place. If the predictor is based on artificial intelligence, the situation gets even more complicated. The medical device AI systems which are required to undergo a third-party conformity assessment, are classified as high-risk AI systems in the European Union AI Act [24]. This, in turn, implies that the AI system provider needs to fulfil the requirements of the standards covering 10 AI Act related areas, specified in the European Commission's standardisation request [25]. This may be discouraging to some providers but if the healthcare customers are willing to cover the costs, a solution may become available.

The AI system providers face another difficulty in developing the predictor application because the law of secondary use of health data does not allow the use of individual health data to innovation activity. Only aggregated data is accessible for innovation activity [5] but such data is typically not useful for AI system training. The use of the non-aggregated personal information is only possible for scientific research. The AI system provider might have to rely on scientific research projects to develop the prediction algorithm, but this is problematic from the MDR's point of view because the scientific researchers don't typically work under a certified quality system which the MDR requires. Particularly, if the medical device is classified to a higher risk class than one, the notified body assessing the company's conformity to MDR may want to have access to the development data of the AI system. This is typically not possible if the Findata officers see too many risks in releasing the data out of the secure processing environment even after anonymisation. This would be problematic with respect to the transparency requirements of the AI Act.

### ***Patient data repository use for service quality improvement***

The law of secondary use of health data [5] allows the use of aggregated Kanta data for information management, to improve the services by comparing the health region's data to the data from other regions. This kind of benchmarking can be useful, and it may reveal quality problems in treating some diseases in the region. Even the delays in getting access to the results data caused by Findata processes are not that harmful as the care quality development processes are typically not time critical.

Two examples can be found in the literature encouraging the development of new similar services. The cholesterol values of diabetic patients improved in the city of Äänekoski when more attention was paid to them from Kanta data [26]. Frondelius et al. studied the feasibility of using Kanta data as input to quality registers [27]. Despite the limited amount of structured information and deficiencies in the completeness of the information, the relevance of the information in the Kanta Services met the information needs of the national quality registers.

### ***Improvement suggestions***

Some of the problems related to the law of secondary use of health data [5] have been recognised, and work is undertaken to make improvements to the law. It is understandable that the privacy of the individuals is highly protected, but improvements are possible without causing harm to the individuals. A general suggestion is to interpret the law in a more research and development friendly way.

An essential challenge in the secondary use of health data is the varying quality of the data. The secondary use would benefit for more systematic use of structured registration of the data instead of

free-form narrative text. The needs of additional structured data fields came up also in the studies of Männikkö et al. [21] and Frondelius et al. [27]. The problem of using free-form text has been identified in the Institute of Health and Welfare [28] and the attempt to improve is one of the focal points in the Kanta development plans for 2025 [29].

The role of Findata should be rethought in the revision of the law of secondary use of health data. Findata has the monopoly in the pseudonymization and anonymisation of health data. It should be reconsidered if these tasks were also possible to other organisations in order to remove the Findata bottleneck and shorten the delays to data access. This might also lower the cost of scientific studies in some cases. The criteria of scientific research should be made visible so that it would be easier to the applicants of Kanta data use to predict the outcome of the request.

Currently, Finnish research organisations have faced difficulties in participating international multi-centre health research projects because of the requirements of the secondary use law. The needs of these studies should be recognised, and the law should be adjusted accordingly. The anonymisation of the data should be easier thus enabling the processing of the data outside the secure processing environment which has also been felt to be expensive. The use of the national supercomputer resources should be administratively enabled because the advanced development of AI algorithms sometimes requires this.

The decision of the deputy data protection ombudsman concerning the search of individuals in risk of a disease to be called to health check-ups calls for a change in the law to enable this apparently useful activity. While waiting for the law to change, consents could be sought from individuals to allow this kind of activity. The involvement of a

health professional in the call process decision-making could also be sufficient in making the activity possible although it would mean extra costs.

The forbidden use of health data of individual subjects in innovation activities is a major source of disappointment towards the secondary use law. The law should take the requirements of the MDR and AI Act into account and enable the development of algorithms for better decision support for the health professionals. At the same time, more education about the characteristics of the AI systems should be given to health professionals. This would also equip them better to carry out the human oversight tasks mentioned in the AI Act.

The right to the secondary use of the Kanta PHR data should be included to the secondary use law. The right for the individual to forbid this use by a simple button click in the PHR user interface could be sufficient to protect the individuals most concerned about their privacy. At the moment, the individual can opt out of the secondary use of his/her health data by using the Findata portal or by filling in a form and sending it to Findata [30].

## Discussion

The initial enthusiasm of the possibilities offered by the law of secondary use of health data and the access to Kanta data [5] have changed to disappointments to some extent. The study revealed that there are various obstacles which make the leveraging of Kanta data to secondary uses less interesting than originally thought. Although the access to health data in the Kanta services is possible according to the law, the law restricts the use of the Kanta health data substantially. The use of the non-aggregated personal information is only possible for scientific research and the delays and costs related to this has also negatively surprised the scientific community. For this reason, the community welcomes

the revision of the law and hopes that the second version will correct the encountered problems at least partly. The innovation and development activities have benefited only very little about the law so far. The possibility to use only aggregated data for product development is often insufficient. Wang et al. have also discussed the problems relating to the access to health data by private companies [31]. They recommend the policymakers to ease the processes for companies to access health data for research and development purposes. The MDR requirement to store and make the data used in product development available to a notified body's inspection is also in conflict with the Findata's system to keep the data available only a limited amount of time.

This study has brought forward the problems related to the secondary use of personal health record data in the Kanta Services, but it has also suggested solutions to improve the situation without incurring substantial additional costs. Although the suggestions would be implemented, they do not solve all the Kanta PHR related issues, like the profitability of the PHR apps to their providers if the economic models of financing these apps do not change, too.

The health risk predictor use case presented in this paper made the opportunities and obstacles in using the Kanta data visible. Although some commonly missing pieces of information like the alcohol consumption and smoking data make some risk predictions unreliable, there are opportunities to predict other risks like pneumococcal disease risk when the Kanta information request contains the relevant variables [32]. It looks promising to continue to investigate risk prediction possibilities with Kanta data, but the data set should be larger than expected [32].

The service quality improvement use case applying information management is a relatively clear beneficiary of the secondary use data because it can make use of aggregated data. The missing death data from the Kanta patient data repository could be compensated using another registry to which Findata has access.

Education is one of the allowed secondary uses of health data. Our literature searches did not, however, find any examples of such uses. The overwhelming majority of approved data permits are related to scientific research and only some statistics project can also be found easily from the catalogue of approved data permits [33]. Scientific research is not limited to academic research institutions because some private companies have also obtained data permits to scientific research.

The solutions to the observed problems require adjustments to the law of secondary use of health data. In the European Union level, the new European Health Data Space (EHDS) regulation defines also conditions for the secondary use of health data [17]. As the goal of EHDS is also to support researchers and innovators, we hope that our observations about the obstacles to the secondary use of Kanta data are noticed and avoided in the EHDS. Particularly the medical device manufacturers can look forward to the EHDS because it does not limit their access only to aggregated health data.

## Conclusions

The possibilities of using Finnish health and well-being information in Kanta services were explored by studying laws, materials from national actors such as the Kanta services, scientific articles, media discussions about the topic, and conducting few expert interviews. As a result, a number of challenges related to the use of health and well-being information in potential application areas were



identified. Many of the identified challenges were related to legislation hindering or preventing the development of planned solutions. It was also noticed that the present law does not allow the secondary use of well-being data. Other factors hindering development include the poor quality of Kanta data, missing data fields needed in risk some calculations, and the unfinished state of the Finnish Personal Health Record (PHR) concept.

A solution plan was presented to address these challenges, mostly by revising the law. The law should allow consent-based use of the PHR data. Various health risk calculators could be implemented, but even more were possible if additional health data fields were included to the Kanta data structures and these fields were filled in. The law should allow also the access to non-aggregated data to innovation activities, e.g. for AI system development. The law should enable participation to international multi-centre studies with data from

Kanta services. The Findata-induced bottleneck to data access should be removed by allowing other secure routes to Kanta data, as well.

### Conflict of interest declaration

The authors 1 and 3 declare the following financial interests/personal relationships which may be considered as potential competing interests: Atostek Oy provides an eHealth API Gateway service which enables healthcare and social welfare providers a fast and easy connection to Finland's national Kanta services. The publication is related to the quality of data stored in the Kanta archive and improving the quality of the data could have a positive impact for Atostek's services. We do not consider this to be a conflict of interest as improving the quality of the data in the national Kanta archive is possible only through constructive criticism of the current situation.

### References

[1] Jormanainen V. Large-scale implementation and adoption of the Finnish national Kanta services in 2010–2017: a prospective, longitudinal, indicator-based study. *FinJeHeW*. 2018;10(4):381–395. <https://doi.org/10.23996/fjhw.74511>

[2] Kanta. Keskeisten terveystietojen määrittelyt. [Internet]. Kanta-Palvelut, Kansaneläkelaitos; 2025 [cited 2025 Jan 19]. Available from: <https://www.kanta.fi/jarjestelmakehittajat/keskeisten-terveystietojen-maarittelyt>

[3] Simplifier.net. Kanta. Finnish PHR R4. [Internet]. Simplifier.net; 2025 [cited 2025 Jan 19]. Available from: <https://simplifier.net/FinnishPHRR4/~introduction>

[4] Männikkö V, Förger K, Kujanen H, Tikkanen J, Antikainen S, Munukka J. Overview of Finnish national patient data repository for research on medical risk assessment. *FinJeHeW*. 2024;16(3):254–268. <https://doi.org/10.23996/fjhw.146124>

[5] Finlex. Laki sosiaali- ja terveystietojen toissijaisesta käytöstä (2019/552). [Internet]. Helsinki: Finlex 2019 Apr 26 [cited 2025 Jan 19]. Available from: <https://www.finlex.fi/fi/laki/ajantasa/2019/20190552>

[6] Kanta. Kanta-tietojen toisiokäytön avulla edistetään hoidon vaikuttavuutta ja helpotetaan palvelujen seuranta [The secondary use of Kanta data supports the effectiveness of treatment and facilitates the monitoring of services] [notice, internet]. Kanta-Palvelut, Kansaneläkelaitos; 2022 Dec 8 [cited 2025 Mar 24]. Available from:

[https://www.kanta.fi/tiedote/-/asset\\_publisher/cf6QCnduV1x6/content/kanta-tietojen-toisiok%C3%A4yt%C3%B6n-avulla-edistet%C3%A4%C3%A4n-hoidon-vaikuttavuutta-ja-helpotetaan-palvelujen-seurantaa](https://www.kanta.fi/tiedote/-/asset_publisher/cf6QCnduV1x6/content/kanta-tietojen-toisiok%C3%A4yt%C3%B6n-avulla-edistet%C3%A4%C3%A4n-hoidon-vaikuttavuutta-ja-helpotetaan-palvelujen-seurantaa)

[7] Kanta. Tutkimus ja tiedolla johtaminen [Internet]. Kanta-Palvelut, Kansaneläkelaitos; 12.3.2025 [cited 2025 Mar 24]. Available from: <https://www.kanta.fi/tutkimus-ja-tiedolla-johtaminen>

[8] Ministry of Social Affairs and Health. Secondary use of health and social data. [Internet]. Helsinki: Ministry of Social Affairs and Health [cited 2025 Mar 24]. Available from: <https://stm.fi/en/secondary-use-of-health-and-social-data>

[9] EUR-Lex. Consolidated text: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). EU; 2016 [cited 2025 Jan 19]. Available from: <http://data.europa.eu/eli/reg/2016/679/oj>

[10] Findata. Finnish Social and Health Data Permit Authority Findata. [Internet]. Helsinki: Findata – Finnish Social and Health Data Permit Authority [cited 2025 Jan 20]. Available from: <https://findata.fi/en/>

[11] Institute of Health and Welfare. FINRISKI-laskuri. [Internet]. Terveyskirjasto; 2021 [cited 2025 January 20]. Available from: <https://www.terveyskirjasto.fi/pgt00013>

[12] Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care*. 2003 Mar;26(3):725-31. <https://doi.org/10.2337/diacare.26.3.725>

[13] U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute. The Breast Cancer Risk Assessment Tool: About the Calculator: About the Breast Cancer Risk Assessment Calculator (The Gail Model) [Internet]. NIH, National Cancer Institute; updated 2024 Nov 19 [cited 2024 Dec 11]. Available from: <https://bcrisktool.cancer.gov/about.html>

[14] Petrazzini BO, Chaudhary K, Márquez-Luna C, Forrest IS, Rocheleau G, Cho J, Narula J, Nadkarni G, Do R. Coronary Risk Estimation Based on Clinical Data in Electronic Health Records. *J Am Coll Cardiol*. 2022 Mar 29;79(12):1155-1166. <https://doi.org/10.1016/j.jacc.2022.01.021>

[15] Kanta. Julkaisuaikataulu, Kanta-palvelujen julkaisuaikataulu palvelujen kehittämisestä. [Internet]. Kanta-Palvelut, Kansaneläkelaitos; 2025 [cited 2025 January 19]. Available from: <https://www.kanta.fi/jarjestelmakehittajat/julkaisuaikataulu>

[16] Rinnetmäki M. Hyvästi Omatietovaranto. [Internet, Blog text in LinkedIn]. LinkedIn; 2023 Oct 1 [cited 2023 Oct 16]. Available from: <https://www.linkedin.com/pulse/hyv%C3%A4sti-omatietovaranto-mikael-rinnetm%C3%A4ki>

[17] European Commission. European Health Data Space Regulation (EHDS). [Internet]. European Commission, Directorate-General for Health and Food Safety [cited 2025 Jan 20]. Available from: [https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space\\_en](https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en)

[18] Kela, Kanta. Omatietovarannon tilannekatsaus. [Presentation, internet]. Kanta-Palvelut, Kansaneläkelaitos; 28.1.2025 [cited 2024 Mar 24]. Available from: <https://www.hl7.fi/wp-content/uploads/OTV-tilannekatsaus-28.1.2025.pdf>

[19] Urhonen H. Evaluation of the suitability of Finnish national patient data repository for

cardiovascular disease risk prediction. [Master of Science Thesis.] Tampere: Tampere University, Faculty of Management and Business; 2024. <https://urn.fi/URN:NBN:fi:tuni-202404033279>

[20] Immonen MS, Similä H, Lindholm M, Korpelainen R, Jämsä T. Technologies for fall risk assessment and conceptual design in personal health record system. *FinJeHeW*. 2019;11(1-2):53-67. <https://doi.org/10.23996/fjhw.73258>

[21] Männikkö V, Kujanen H, Munukka J, Förger K. Finnish national patient data repository as data source for FINRISK risk calculator. *FinJeHeW*. 2024;16(4):489–500.

<https://doi.org/10.23996/fjhw.146762>

[22] Tietosuojavaltuutetun toimisto [The Finnish Office of the Data Protection Ombudsman]. Apulaistietosuojavaltuutettu: Tietosuoja huomioitava ennakoivan terveydenhuollon toimintatavoissa. [Internet]. The Finnish Office of the Data Protection Ombudsman; 2022 [cited 2024 Aug 23]. Available from: <https://tietosuoja.fi/-/apulaistietosuojavaltuutettu-tietosuoja-huomioitava-ennakoivan-terveydenhuollon-toimintatavoissa>

[23] EUR-Lex. European Union. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices. [Internet]. EU; 2017 [cited 2025 Jan 20]. Available from: <http://data.europa.eu/eli/reg/2017/745/2025-01-10>

[24] EUR-Lex. European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. [Internet]. EU; 2024 [cited 2025 Jan 20]. Available from: <http://data.europa.eu/eli/reg/2024/1689/oj>

[25] European Commission, eNorm Platform. C(2023)3215 - Standardisation request M/593 Standardisation request to European Committee

for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence. [Internet]. European Commission; 2023 [cited 2025 Jan 20]. Available from: [https://ec.europa.eu/growth/tools-databases/enorm/mandate/593\\_en](https://ec.europa.eu/growth/tools-databases/enorm/mandate/593_en)

[26] Tolonen R. Kanta-palvelut tarjoavat tutkimukselle tietosisältöä, jonka avulla sote-palveluita voidaan kehittää vaikuttavammiksi. [Internet]. Kansaneläkelaitos (Kela), Tietotarjotin; 2023 May 26 [cited 2025 Mar 24]. Available from: <https://tietotarjotin.fi/uutinen/2473313/kanta-palvelut-tarjoavat-tutkimukselle-tietosisaltoa-jonka-avulla-sote-palveluita-voidaan-kehittaa-vaikuttavammiksi>

[27] Frondelius A, Kinnunen UM, Jormanainen V. The Significance of Information Quality for the Secondary Use of the Information in the National Health Care Quality Registers in Finland. *Methods Inf Med*. 2025 Jan 29. <https://doi.org/10.1055/a-2511-7866>

[28] Achte P, Mykkänen J, Konttinen R, Rätty T, Sundman J. Tiedon laadun kehittämiskohteet. Sote-tietoarkkitehtuurin ohjausryhmä 7.12.2021 [presentation, Internet]. THL; 2021 [cited 2025 Jan 20]. Available from: [https://www.julkari.fi/bitstream/handle/10024/148374/4%20Tiedon\\_laadun\\_kehittamiskoh-teet\\_20211207\\_soteTA\\_%20ohry\\_Achte-Mykkanen.pdf?sequence=6&isAllowed=y](https://www.julkari.fi/bitstream/handle/10024/148374/4%20Tiedon_laadun_kehittamiskoh-teet_20211207_soteTA_%20ohry_Achte-Mykkanen.pdf?sequence=6&isAllowed=y)

[29] Kanta. Kanta-palvelujen painopisteet vuodelle 2025 on vahvistettu [Internet]. Kanta-Palvelut, Kansaneläkelaitos; 2024 [cited 2024 Dec 18]. Available from: [https://www.kanta.fi/tiedote/-/asset\\_publisher/cf6QCnduV1x6/content/kanta-palvelujen-painopisteet-vuodelle-2025-on-vahvistettu](https://www.kanta.fi/tiedote/-/asset_publisher/cf6QCnduV1x6/content/kanta-palvelujen-painopisteet-vuodelle-2025-on-vahvistettu)

[30] Findata. Oikeudet tietoihisi. [Internet]. Helsinki: Findata – sosiaali- ja terveysalan tietolupaviranomainen; 2025 [cited 2025 Mar 24]. Available

from: <https://findata.fi/tietoa-findatasta/oikeudet-tietoihisi/>

[31] Wang F, Mohamed M, Ahokangas P, Karppinen P. Balancing stakeholder interests and paradoxes in health data sharing within health ecosystems. *FinJeHeW*. 2024;16(1):6–22. <https://doi.org/10.23996/fjhw.138290>

[32] Männikkö V, Turunen J, Åhman H, Harju E. A large-scale risk assessment and classification model

for pneumococcus using Finnish national health data. *Healthcare Analytics*. 2025;7:100382. <https://doi.org/10.1016/j.health.2025.100382>

[33] Findata. Myönnetyt luvat. [Internet]. Helsinki: Findata – sosiaali- ja terveysalan tietolupaviranomainen; 2025 [cited 2025 Mar 24]. Available from: <https://findata.fi/luvat/myonnetyt-luvat/>