

Project title:	<i>Health Sense</i>
Project Promoter:	<i>Estonian Ministry of Social Affairs</i>
Project Partner(s):	<i>Health and Welfare Information Systems Center, Tallinn University of Technology, University of Tartu</i>
Donor project partner(s):	<i>Helse West IKT; Western Norway University of Applied Sciences</i>
Total maximum eligible project cost:	<i>2 000 000€</i>
Project grant rate:	<i>100%</i>
Project grant amount:	<i>2 000 000€</i>
Estimated duration:	<i>24 months</i>

a) Project summary

The goal of the predefined project Health Sense is to develop a secure data storage, integration, access and analysis toolkit in order to provide large, complex and detailed sets of health and lifecycle data for the public sector, private sector and R&D institutions. The ultimate aim of the project is to raise the availability of health and lifecycle data in order to enhance the creation of new products, services and solutions by the private sector, novel interventions by the public sector and relevant scientific knowledge by the R&D communities in order to raise life expectancy and prolong healthy life years of people. It is accounted for, that while the user community is likely to broaden even further in the future, then this data is delicate, which means privacy and security of developed/procured components is key. However in addition to technical security, compliance to strict data protection regulations is crucial as well. Thus, a key activity stream will aim to develop data collection procedures, tools and analyse regulations that meet limitations and responsibilities described in the General Data Protection Regulation¹ (GDPR) to ensure usability of developed ICT components and thus ensure project success.

Health Sense is a virtual environment used to collect health, health cycle and lifestyle data for research and service development purposes. In addition, it also provides a toolbox that enables the creation of data analysis-based solutions for prolonging a healthy human life. As the data handled in the environment is sensitive, the tools for safe collection, storage, access and processing of data are an essential part of the provided toolbox.

b) Project background and justification

Estonian governmental information system today has sufficiently good e-ability to collect and handle personal data in different areas including healthcare, social care and labor area that are managed by Ministry of Social Affairs. Mostly these abilities today are used on an operational level to provide services for individuals, and for reporting and simple analyses. The better use of data, to provide the steering process with predictive and calculated supporting data, is

¹ 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. OJ L 119, 4.5.2016, p. 1–88.

planned but not yet implemented. Today a couple of domain-based separated data warehouses are used, and data relations between different areas are built, but only on a data mart level. This leads to limited possibilities in data analysis and AI development.

Today many different countries have the same problem – services are being provided and data is being collected, but separately and by areas with very simple loop back to steering process. Even when the data collected by different areas is gathered to the same environment, the data from different sources is connected on a person level and no additional relations are calculated nor defined. In many circumstances, legal constraints prevent interconnecting person level data.

Estonia is small and dynamic enough to build data integration, analysis models, and solutions to experiment with data warehouse environment architecture and development tools. We have all the necessary electronic data sources for supporting this process. Furthermore, we have years of experience in developing forward-thinking digital solutions and in simultaneously building/updating our regulatory framework for these solutions to be implemented. Estonia, being the near perfect testbed for spearheading new e-governance initiatives, can provide here great value by developing models and solutions, which can later be expanded further to larger data volumes in larger countries, Norway being one of the main considerations.

It is paramount to find committed partners, that are interested in projects of this kind as well as who comply with ethics and the principles of data security. There are several interested groups in the public sector, R&D institutions and private sector. The Department of Data Analysis has been set up under the Ministry of Social Affairs, however the need for health and lifecycle data continuously exists in other public sector institutions in the health and social domain (e.g. Health Insurance Fund, Health Development Institute, Estonian Health Board, Social Insurance Board, Unemployment Insurance Fund) and other public sector domains as well (e.g. Ministry of Defense, Defense Resources Agency, Ministry of Interior, Ministry of Education and Research etc). R&D institutions (e.g. University of Tartu and Tallinn University of Technology and Technology Competence Centers) are interested in using humans' lifecycle data in their research, development and teaching processes, in cooperation with the governmental authorities. Various lifecycle and health data interested communities (e.g. Estonian HealthTech Cluster, Open Knowledge Estonia) are set up so that they connect individuals, research organizations, companies and startups. In addition, in recent years international interest by global (multi-national) companies (mainly pharmaceuticals and health technology) has grown exponentially.

All these groups are interested in having access to person's lifecycle data. Some of them produce human lifecycle data, and through that expand the variety of data collected. Due to the high level of IT coverage in Estonia, many international companies have become interested in using health and lifecycle data in their R&D efforts. This can provide value as additional resources to the project.

Implementing the modern methods of person's lifecycle-based data collecting, analysis and AI tools will lead to a longer and healthier life and reducing costs per unit in health- and social care. Thus, supporting the goal of the **welfare technologies** concept – staying independent in activities of daily living while receiving tailored support from various care services when needed.

Health Sense is a central (root) project for the redevelopment of data utilization in health, work and social care area, where the policies and services are intertwined and thus should be combined to provide maximum effectiveness and cost-efficiency. The latter is currently hindered by the scatteredness of data in these areas which leads to deficiencies in accessing and utilizing data for revising and building services. Thus, the toolkit developed through Health Sense would enable to modify, create and provide invisible and/or preventive services to citizens based on evidence gathered through complex data analysis/findings. The project will be implemented as a predefined project because it requires the coordination between several public and private actors, many of which have very specific knowledge in the project area. Without the successful implementation of this project today, the push for combined evidence and data-led improvement of human social and health wellbeing, will be delayed for years.

c) Objective, expected outcome(s) and list of outputs

Project activity streams:

- 1.1. Project management
- 1.2. Data collecting environment
- 1.3. Data storage with pseudonymization and obfuscation tools
- 1.4. Warehouse data structures and ETL process
- 1.5. Data Analysis and visualization environment for third parties (e.g. private sector, R&D institutions)
- 1.6. Building data flows in cooperation with third parties (e.g. private sector, R&D institutions)
- 1.7. Analysis of legal regulations to foster the results of the project

For more detailed activities see timetable/milestone plan (e). For budget overview see sheet accompanied with this document.

Expected outputs:

- data collection environment from different sources (governmental and non-governmental)
- data pseudonymization, data serialization/deserialization, obfuscation and long text extraction tools
- partly encrypted data storage with encryption/decryption tools and methodology
- data flows from runtime information system to data warehouse (built on needs of third parties). Illustrated with more detail in Appendix 2 and 3.
- user rights and consent management system
- data analysis environment for third parties
- Analysis of legal regulations to ensure project outcomes
- data warehouse environment for Ministry of Social Affairs and other governmental institutions (for reporting and analysis purposes)

Expected outcomes:

- to create a data management platform for developing innovative solutions for healthcare and welfare service providers to develop and experiment new innovative products and services;
- to encourage the emergence of new health, welfare and lifestyle businesses and data driven business models;
- to create a launch pad for new healthcare and welfare services for their emergence to the market in cooperation with SMEs and start-ups;
- to promote scientific thinking and provide new inputs for research, development and innovation in SMEs and start-ups in the delivery of health and welfare services and educate highly skilled workforce in cooperation with universities.

Expected project result	Indicator	Unit of measurement	Source of verification	Frequency of reporting	Baseline value	Baseline year	Target value
Health Sense data management platform developed and operating	Data analysis procedures/tools in place and documented	Binary	Documentation available on HWISC website, software hosted at HWISC or state cloud infrastructure	Semi-annually	No	2018	Yes
	Data collection and publishing procedures/tools in place and documented	Binary	Documentation available on HWISC website, software hosted at HWISC or state cloud infrastructure	Semi-annually	No	2018	Yes
	Data collection, and publishing environment architecture and data analysis environment architecture developed and published	Binary	Documentation available on HWISC website, software hosted at HWISC or state cloud infrastructure	Semi-annually	No	2018	Yes
	Number of data mining/Artificial Intelligence pilot projects successfully finished	Number	Documentation outlining projects and results available on HWISC website	Semi-annually	0	N/A	3
	Number of professional staff trained	Number	Attendance sheets	Semi-annually	0	N/A	10

Project sustainability

All results created during the project will be maintained by Health and Welfare Information Systems Centre and Ministry of Social Affairs. All software created during the project will be hosted and maintained by Health and Welfare Information Systems Centre. The necessary resources to maintain and host results will be applied from the annual state budget of Estonia.

Intellectual property

Unless stipulated otherwise in the agreements between project promoter and project partners, intellectual property rights of the results of the project shall be owned by the party generating it. When results are generated jointly, it shall be jointly owned, unless the parties in question agree on a different solution. The method of defining joint ownership shall be based on their contributing efforts towards the result in shares proportionate to their scope of work/budget for that result.

After completion of the project, project promoter shall have the non-exclusive, royalty-free right to use, modify, copy, transfer or make derivative results of the project. The rights of the project promoter shall not be limited with regard to certain rights or to the purpose, term, territory, extent, manner or means of using the results of the project.

d) Information on the project promoter and project partner(s)

Ministry of Social Affairs is a governmental institution that operates in the field of social security. For ministry have been set five strategic objectives - ensure people's economic prosperity and their good work, ensure people's social coping and development, support the well-being of children and families, promote people's mutual care, equal opportunities and gender equality, and ensure people's long and high-quality life. Ministry of Social Affairs emerges objectives through compiling development plans for the fields being under control of ministry, as well as organization-based development plans.

Ministry of Social Affairs as project promoter, will ensure seamless project coordination.

Health and Welfare Information Systems Centre (HWISC) is info technology competence centre of Ministry of Social Affairs. HWISC is involved in the development, and management of ICT services in the health, social and labour fields within the area of government of the Ministry of Social Affairs. HWISC is a representative of the Ministry of Social Affairs in the „Health Sense“ project as the project manager.

HWISC will take upon the leading role in most of the project activities including procuring the necessary licensed software and hosting as well as designing the technical architecture of components and maintaining the solutions developed during Health Sense project. HWISC will assist in project coordination and in organizing meetings as well as in analysis of legal regulations to ensure the success of the project.

Tallinn University of Technology (TalTech), the second largest university in Estonia, is international university teaching the knowledge and skills necessary for today's international and competitive marketplace. TalTech is the top centre of science in Estonia. With strong emphasis on internationalization, TalTech has a strong multicultural student body, many international professors (11% foreign staff members), great cooperation with the world's top universities (e.g. Stanford, Technische Universität München, Del University of Technology and KTH Royal Institute of Technology in Europe) and with international companies (e.g., Microsoft, ABB, IBM. etc.). There are over 30 fully accredited international degree programmes (4 Bachelor programmes, 18 Master programmes and 10 PhD programmes) that are available fully in English.

TalTech will develop the appropriate model for digital health care data that could be used for data collection and storage in new generation Estonian Health Information System with the feasibility to implement the same model in other large scale Electronic Health Records (EHR). TalTech will develop the data model for digital medical data capture and storage to achieve semantic interoperability and develop universal continuity of care process standard.

University of Tartu (UT) is Estonia's leading centre of research and training. It preserves the culture of the Estonian people and spearheads the country's reputation in research and provision of higher education. UT belongs to the top 1.2% of world's best universities.

As Estonia's national university, UT stresses the importance of international co-operation and partnerships with reputable research universities all over the world. The robust research potential of the university is evidenced by the fact that the University of Tartu has been invited to join the Coimbra Group, a prestigious club of renowned research universities.

UT includes four faculties. To support and develop the professional competence of its students and academic staff, the university has entered into bilateral co-operation agreements with 79 partner institutions in 31 countries.

University of Tartu will undertake the development of obfuscation toolkit.

Helse Vest IKT and Western Norway University of Applied Sciences are cooperating in building infrastructure for Intromat (<https://intromat.no/>), a project on introducing e-health in mental health. As Intromat requires designing of architecture, API-s, interfaces, pseudonimization/synonymization tools similar to the aim of Health Sense proposal, then it provides an ample opportunity to build synergies for both projects and future collaborations. Helse Vest IKT supplies all IT services to specialized health services in the counties of Rogaland, Hordaland and Sogn og Fjordane. Helse Vest customers include the public hospitals in Helse Vest, and several private health service providers in the region. Both Helse Vest IKT and Western Norway University of Applied Sciences have extensive experiences in designing and building software architecture.

Bilateral partners will provide consultations and best practice exchange on the following topics: warehousing architecture (e.g. component architecture, processes,

relations with other warehouses, ETL principles and relations with data providers, standards), licensed components (database environment, ETL tools, analytical tools, data virtualization environments) and developed tools (synonymization tools, obfuscation tools, tools for long text extraction, data encryption/decryption). The bilateral partnership is crucial to ensure, that solutions developed in Estonia are state-of-the-art, internationally relevant and to gauge the potential for sharing components with Norwegian counterparts.

e) Timetable / Milestone plan *)

Extensive timetable is also visualized in a separate excel sheet

Activity (see attached timetable)	Activity stream	Leader	Timeframe (week from project starting date)
Project kick-off: 1st February 2021 (or as soon as possible)			
Staffing the project team	1.1	Ministry of Social Affairs, Health and Welfare Information Systems Centre	1-40
A communication campaign and negotiations to select third party projects	1.1	Ministry of Social Affairs, Health and Welfare Information Systems Centre	1-70
Building overall system architecture	1.1 – 1.7	Health and Welfare Information Systems Centre (incl. project consortium*)	1-23
Analysis of legal regulations to ensure project success	1.7	Ministry of Social Affairs, Health and Welfare Information Systems Centre	40-102
Selecting and describing projects for development with third parties (private sector, R&D institutions)	1.1, 1.5-1.6	Health and Welfare Information Systems Centre	6-63

Building hardware architecture	1.1 – 1.7	Health and Welfare Information Systems Centre (incl. project consortium*)	4-9
Describing the criteria and restrictions for the Initial software components to be procured (ETL, DW database, PostgreSQL. Encryption/decryption, obfuscation, pseudonymisation, user permissions, virtualization etc.	1.1 – 1.7	Health and Welfare Information Systems Centre (incl. project consortium*)	1-25
Installation of initial environment to infrastructure	1.1 - 1.6	Health and Welfare Information Systems Centre	8-11
Component, installation, and testing (Data collecting environment – ETL, BIW tools)	1.1, 1.4, 1.5	Health and Welfare Information Systems Centre	9-14
Building virtual working environment for third parties (closed environment for ETL, data marts, and analyse tools)	1.1, 1.5 - 1.6	Health and Welfare Information Systems Centre	12-20
Tenders of licenced tools	1.1 – 1.5	Health and Welfare Information Systems Centre	9-16
Completing initial environment (tender result installations) –	1.1 – 1.5	Health and Welfare Information Systems Centre	17-21

central DW and data lake environment			
Testing of environment	1.1 – 1.5	Health and Welfare Information Systems Centre	22-25
Description of initial environment, user manuals	1.1 – 1.5	Health and Welfare Information Systems Centre	9-25
Development/detailing of architecture	1.4	Project consortium*	5-75
Building BIW for IS-s	1.1 – 1.5	Health and Welfare Information Systems Centre	26-90
Development of logical data structures for DW	1.4	Tallinn University of Technology	2-64
Building data flows from BIW to central DW (ETL)	1.1 - 1.5	Health and Welfare Information Systems Centre	17-90
Development of user rights environment	1.1 – 1.5	Health and Welfare Information Systems Centre	22-34
API, REST etc. interfaces to access data	1.5	Health and Welfare Information Systems Centre	40-60
Third party projects testing the built toolkits and environment	1.1 - 1.6	Health and Welfare Information Systems Centre	26-102
Development of tools for collection of aggregated data	1.5 - 1.6	Health and Welfare Information Systems Centre	52-60

Environment for publishing data sets and statistical results (Open Data)	1.5	Health and Welfare Information Systems Centre	58-70
Development of pseudonymisation toolset	1.3	Health and Welfare Information Systems Centre	9-73
Development of obfuscation toolset	1.3	University of Tartu	9-97
Development of long text extraction tool	1.3	Health and Welfare Information Systems Centre	42-97
Development of serialization/deserialization tool	1.3	Health and Welfare Information Systems Centre	42-102
Meetings with partners (in Estonia)	1.1	Health and Welfare Information Systems Centre, Helse-West IKT, Western Norway University of Applied Sciences	43, 66, 96
Meetings with partners (in Norway)	1.1	Helse-West IKT, Western Norway University of Applied Sciences, Health and Welfare Information Systems Centre	56, 76, 86
Final descriptions of environment, user manuals	1.1	Health and Welfare Information Systems Centre	90-100
Promotion of project results	1.1	Health and Welfare Information Systems Centre	95-102
End of project: 31th January 2023			

*Project consortium includes project promoter and all project partners (incl. donor partners).

f) Budget

Included in separate Excel sheet.

g) Overview of risks:

1. Legal frame obstacles to be followed in order to merge the different datasets.
 - a. Probability: Likely
 - b. Impact: Moderate
 - c. Response: Development of legal framework (incl. revising current regulations) will go in strong cooperation with the project team and project objectives. Revising regulations is part of not only major, but also minor e-government initiatives in Estonia. This is done regularly, routinely and simultaneously for all new governmental ICT-related initiatives as well as for minor modifications in existing digital infrastructure.
2. Breach of principles and laws regarding privacy, data safety and ethics.
 - a. Probability: Unlikely
 - b. Impact: Major
 - c. Response: Apply strict data security measures and methods throughout data storage and transfer infrastructure. Provide third-party access to only pseudonymized data only within the secure environment with access to analytical tools. Log all actions, which are done with the data in the secure environment. Allow only aggregated/analysed data to be exported out of the secure environment after logs have been checked for any suspicions. Additional principles regarding data privacy and trust to minimize this risk are detailed in Appendix 1.
3. Procurements for hardware and/or software fail or will be delayed that leads to difficulty to support requested activities.
 - a. Probability: Unlikely
 - b. Impact: Major
 - c. Response: Thorough preparation of the procurement documents
4. Lack of interest from SMEs and/or academia that leads to underuse of the Health sense capability.
 - a. Probability: Unlikely
 - b. Impact: Moderate
 - c. Response: Engage additional entities to promote the capability. Provide additional operational support from governmental agencies (e.g. datasets that can be used to solve specific problems).
5. Reputation damage to the project owner and project partners
 - a. Probability: Unlikely
 - b. Impact: Minor
 - c. Response: Carefully plan, prepare and staff the project before the official start to mitigate planning fallacies. Carefully select partners based on their showcased skills, interest, past commitments and adherence to data privacy and ethical principles.
6. Postponement of welfare technology open call projects
 - a. Probability: Likely
 - b. Impact: Moderate
 - c. Response: Include also projects/partners with reliable funding outside of welfare technology open call in addition to the open call projects.

APPENDIX 1 – Privacy and trust requirements and related responses

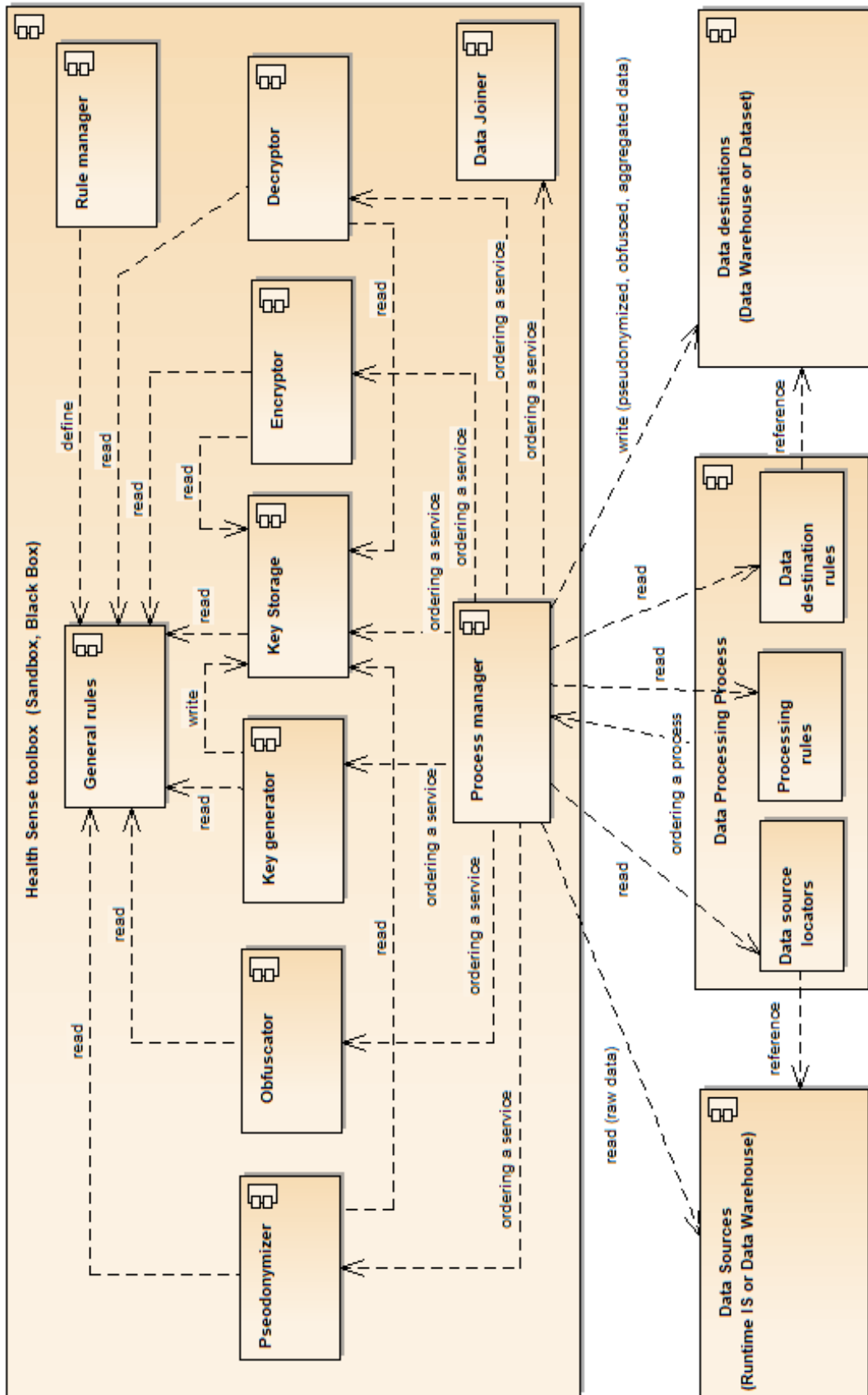
1. Requirements

- a. The results of the project require the application of high-level data protection and sensitivity rules to prevent data misuse and privacy breaches.
- b. All data processing activities must be verifiable at a later stage.
- c. Data processing must take place on the basis of legislation.
- d. Only deal with ethical and technically capable partners who can provide a secure data storage environment and trusted team.

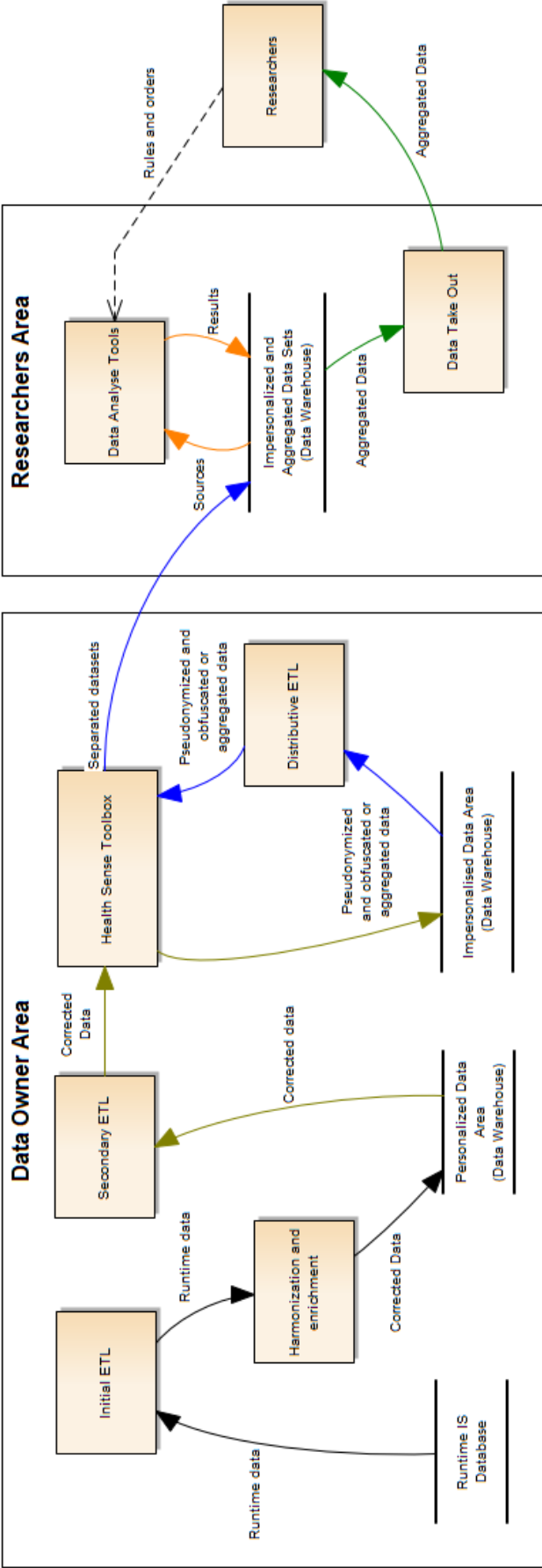
2. Responses

- a. Data storages are divided to two securely separated areas: data owner area and researchers area; personalized data can be stored only at the data owner area; at the researchers area data can be stored in pseudonymized and aggregated manner; pseudonymized and aggregated data can also be stored at the data owner area. (requirement a)
- b. Data owner area is also divided to two separated areas: personalized data area and impersonalized data area. This allows better to separate users who need personal information from users who do not need them. (requirement a)
- c. Access to the personal and pseudonymized data can only be done through secure environments using highest-level authorization tools (corresponding to EIDAS “high”). (requirement a)
- d. Each dataset must be compiled using the different pseudonym for the same person. (requirement a)
- e. Along with data pseudonymizing, obfuscation is always applied. (requirement a)
- f. “Aggregated data” is in most cases the synonym of “open data”; however, this is only possible if a sufficient amount of data is aggregated on one line; the minimum number of aggregated rows must be defined for each data set (requirement a)
- g. It is mandatory to create isolated sandbox-like environments where researchers can analyse (anonymized or aggregated) data; only aggregated (statistical) data can be extracted from these environments. (requirement a)
- h. All activities with the data must be stored in logs; it is necessary to build a log analysis environment to detect misuse of data. (requirement b)
- i. At the first stages of the project, it is necessary to find solutions for the creation of data processing activities in accordance with current legislation; at the same time, it is necessary to identify regulatory shortcomings that prevent data processing in the interests of individuals and initiate changes to these laws. (requirement c)
- j. The background, intent and goals of all partners must be verified every time when data usage permission is granted; all data usage is monitored on the basis of log analysis; when the misuse of data is detected, the misusers accounts will be closed until the decision is made (requirement e)

APPENDIX 2 – Health Sense overall component architecture



APPENDIX 3 – Health Sense overall data flow chart



ETL - Extract, Transfer and Load (data transfer)