Health-RI

The NL Personalised Medicine & Health Research Infrastructure

An initiative of:

February 2016

Preambule

Dear reader,
Attached you will find the specification of the proposal Personalised Medicine & Health Research Infrastructure "Health RI" in 2025. The Personalised Medicine and Health dream is introduced to you by means of this letter.

Today life sciences and medical research in the Netherlands encompass several outstanding basic and translational research programmes directed towards personalised prevention, prognosis, as well as prediction, guidance and monitoring of precision treatment in numerous diseases. In this vision we set course for the medicine and health research infrastructure in the year 2025; with an eye on what will be achieved by the year 2040, and strongly rooted in programmes of today.

Today, medicine has only just left an era that was characterized by treating diseases after the fact, decision making on certain population average and physicians making decisions for their patients. This approach is not future proof: we will move towards a new medicine and health paradigm. This paradigm will be predictive, preventive, personalised and participatory[1]; healthcare will be focused on improving health and striving to help people function as good and long as possible despite potential (chronic) diseases.

Biology and biomedicine as science fields will adopt a systems approach focussing on health and diseases: understanding how biological processes interrelate, how perturbations in a healthy 'personal system' arise, and how interventions (e.g. lifestyle-related, high-precision medication or regenerative medicine) can restore homeostasis. Systems biology (i.e. the knowledge base) and advanced read outs of biology will help transform medicine from reactive into a P4 mode. Combining advanced genetics with non-invasive imaging and longitudinal physiological monitoring locates disturbances in the body at a very early stage, and with great precision. Any intervention still needed is conducted with the highest level of precision and tailored to the needs of the individual. The change to the P4 mode and the rise of personal data collection creates the opportunity to radically shorten the period between clinical and/or scientific investigation and intervention.

By 2040, medicine and health are a fully pro-active, integrated, predictive, preventive, personalized and participatory science *and* healthcare is at affordable cost levels for society. Preventive self-management of citizens as part of their everyday life focuses on improving health and functioning as good and long as possible despite potential (chronical) diseases. We will provide personalized treatment for every patient and empower patients and healthcare workers in clinical decision-making based on full utilization of systems biology based knowledge of mechanisms of disease. Science will have brought the knowledge and the resources needed to do so and made it easily accessible and applicable in society.

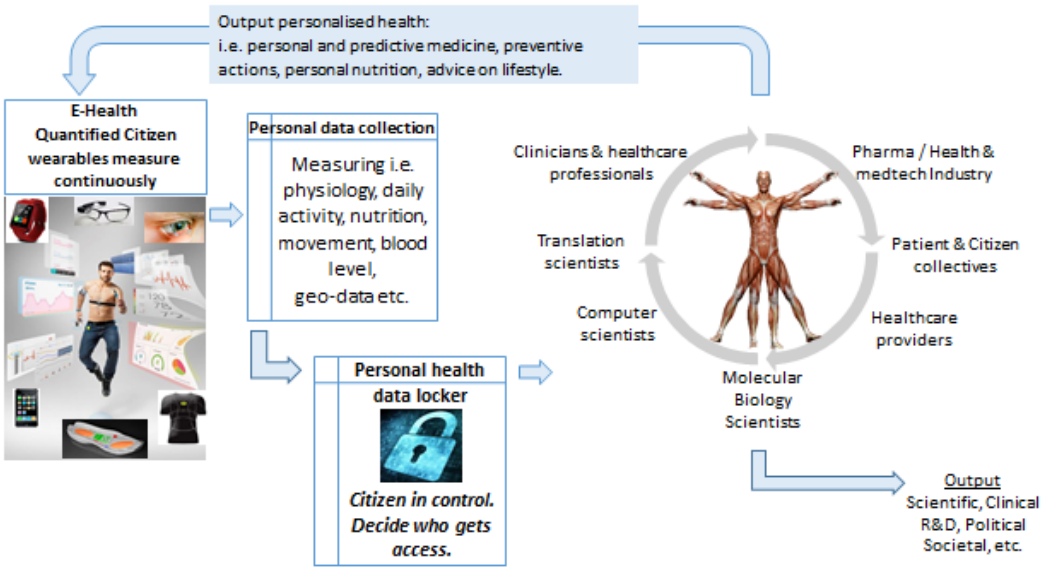
To fulfil this ambition, *by 2025* the medical research community should have access to a research infrastructure that will accommodate all researchers active in areas such as systems genetics, -omics and molecular biology, image sciences, epidemiology, preventive health and clinical medicine. It will have to be common national platform as a strong hub in the international biomedical research network across scientific disciplines, across different types of users and across users-questions.

[1] P4 Medicine Institute (2012). *P4Medicine.* Accessed on the 8th of January 2016, taken from:http://p4mi.org/p4medicine.

This is our perspective on the future

In 2025, Dutch citizens all have their own personal digital biobanks. With help of the available smart and wearable technology, these biobanks continuously collect personal health information (figure 1). Each individual is unique in health and functioning).[2] Thus, the 17 million personal biobanks represent 17 million personal health profiles, influenced by intrinsic and external factors such as genetics, nutrition, lifestyle and (socio-economic) environment. Combined with the information gathered in the biomedical research field on individual patients and the international knowledge base will create a globally unique resource.

In 2025 we will have aligned all major biomedical and health research initiatives in a large-scale unified infrastructure for combined genotyping and deep phenotyping of human diseases *in situ*. Such an infrastructure will combined high-quality molecular and imaging pipelines with nutrition and lifestyle research resources, and feature the ICT and e-infrastructure for sophisticated data integration and systems analyses. By 2025, the infrastructure has assembled the technologies for complete genome analysis combined with detailed phenotyping of the participants based on metabolic and genomic parameters, imaging data, lifestyle information, insight into their microbiome, and information from both electronic patient records and personal health files. Increasingly the research process will depend on direct involvement, data collection and sampling by and from healthy individuals and patients, as well as providing these participants to receive direct feedback on analyses performed with support of mobile electronic devices (e-Health).



In 2025, Dutch bio-medical scientists have massively moved beyond the classical mono-disciplinary and population-based approach into an integrated interdisciplinary systems medicine and personalised health approach. They not only try to understand and model the generic mechanisms underlying health and disease, but also explore the genetic and phenotypic variation, the dynamics of life, and the physiological bandwidth that can help to model the 'health system flexibility' of an individual at (sub-)cellular, organ and organismal level. The technological revolutions in bioimaging, multi-omics molecular profiling of samples, high-tech precision interventions and the availability of longitudinal quantified self, e-health and daily functioning measurements have turned life science into a truly multi-disciplinary and big data-driven science field. This has resulted in a strong connection among medical, life science, technology and computer science communities. They effectively combine their know-how and skills in scientific exploration with social development and business development capacities.

[2] Huber, M. (2011). How should we define health? *BMJ* ; 343 doi: http://dx.doi.org/10.1136/bmj.d4163 (Published 26 July 2011)

In 2025, scientific data should be stewarded in interoperable form and actively shared among scientific groups and disciplines in a fashion that accelerates the construction of a collective knowledge base of rapidly growing value. Dutch citizens massively make their self-collected data available, not only to support their personal self-monitoring and self-exploration, but also to expedite scientific exploration, assured that their privacy is well respected. As full owners of this personal information they remain in control of their personal health data, which will not even need to leave the secure environment of their personal health data locker.

Towards 2025, integrative analysis of multifactorial big data has strongly advanced research into diseases and health. The Netherlands has an infrastructure that offers easy access to (international) sample and data resources, and facilitates scientists to work efficiently in cross-disciplinary studies towards understanding biological complexity and creating the evidence-base to interpret variation among individuals in terms of personal health. Scientists daily use specially designed research-workflows that search the open access data web for information or resources relevant to their particular research question. The results delivered by these '*data-trains'* is processed and validated on the spot and the results are integrated in the research at hand with information retrieved from international reference data collections. Hypotheses are thus produced and tested in real-time with high reproducibility.

In 2025 Dutch clinicians and health professionals have strongly sped up evidence-based medicine and health. In the classical approach of 2015, pre-clinical research, clinical trials, meta-analyses and guidelines ruled. This innovation process simply took too long and could not keep up with the speed of knowledge and technology changes. Rather, a form of health care has taken over that is based upon high-precision, non-invasive and continuous health monitoring, mostly by citizens themselves, and combined with data-driven rapid learning technology available to both professionals and citizens. In many cases where remedies as drugs and surgery are already available, this approach has reduced the care and cure innovation lag time from years to timeframes of weeks or even days, all to the benefit of (daily) functioning for individual persons (patients).

By 2025, health professionals have adopted an integral health approach to offer personalized health management solutions to their patients, based upon a combination of quantified self data collected by the patient, their personal genome profile, and (if required) additional clinical and societal information. Professionals thus guide citizens in their social context to optimise their personal health, easily tapping into pre-selected health models suggested by the international health and disease knowledge base. As health professionals, they make sure the information is correct and that the decision meets patient values and preferences. The 2025 clinician is a translational expert, including the latest research insight to coach their patients in optimising their health.

In 2025, the Dutch field of personalised medicine & health research closely involves all of the above stakeholders, and many more. A ground-breaking research infrastructure connects all these people, citizens and experts, scientists and health professionals. It involves all certified lab facilities and clinics, and all biobanks and data collections. Started as an exclusive life science research infrastructure it has become an invaluable part of the public health domain. As an initial step this proposal is aimed at creating the roadmap towards the ideal infrastructure based on the current and potential strengths of the Dutch life sciences and addressing the emerging needs to sustain and strengthen our scientific leadership position in this sector. It will bridge a broad range of technology and infrastructure initiatives across UMCs, universities and other biomedical research institutes, as well as connect different scientific disciplines: basic sciences, clinical sciences and engineering. The infrastructure is provides the stepping stone towards citizen science participation. This proposal refers to the creation of nation-wide biobanks, state of the art –omics and imaging technology as well as an overarching linked-data infrastructure.

This vision has been developed with contributions and endorsements from a large group of scientists:

- Dr. Jeroen Belien (VUMC)
- Prof. Bas Bloem (Radboudumc)
- Dr. Jan-Willem Boiten (Lygature)
- Dr. Luiz Olavo Bonino da Silva Santos (VU)
- Prof. Dorret Boomsma (VUMC)
- Dr. Rolf Bos (TIFN)
- Mr. dr. Jasper Bovenberg (Legal Pathways)
- Prof. Edwin Cuppen (UMCU)
- Prof. André Dekker (MUMC+)
- Dr. Ingrid Dillo (DANS/RDNL)
- Dr. Peter Doorn (DANS/RDNL)
- Prof. Cornelia van Duin (EMC)
- Prof. Chris Evelo (Maastricht University)
- Prof. Dorus Gadella (UvA)
- Dr. Celia van Gelder (Radboudumc, DTL)
- Dr. Richard Finkers (Wageningen UR)
- Prof. Alain van Gool (Radboudumc)
- Prof. Thomas Hankemeijer (LU)
- Prof. Frank van Harmelen (VU)
- Prof. Wilco Hazeleger (eScience Centre)
- Prof. Jaap Heringa (VU)
- Dr. Peter-Bram 't-Hoen (LUMC)
- Dr. Rob Hooft (DTL, NLeSC)
- Prof. Hilleke Hulshoff (UMCU)
- Prof. Bart Jacobs (Radboud Universiteit)
- Dr. Connie Jimenez (VUMC)
- Prof. Folkert van Kemenade (EMC)
- Prof. Judith Klumperman (UMCU)
- Prof. Joost Kok (LU)
- Prof. Wessel Kraaij (TNO/Radboud Univ.)
- Dr. Aad van der Lugt (EMC)
- Prof. Peter Luijten (UMCU)
- Prof. Nico van Meeteren (Topsector LSH Health~Holland)
- Prof. Gerrit Meijer (NKI/AVL)
- Prof. Barend Mons (LUMC)
- Dr. Iris Nagtegaal (Radboudumc)
- Prof. Wiro Niessen (EMC)
- Prof. Ben van Ommen (TNO)
- Prof. Gert-Jan van Ommen (LUMC)
- Prof. Anwar Osseyran (UvA/SURFsara/ RDNL)
- Prof. Marcel Reinders (TU Delft)
- Dr. Marco Roos (LUMC)
- Dr. Hans Roubos (DSM)
- Dr. Jeroen Rouppe vd Voort (ENZA Zaden)
- Prof. Eline Slagboom (LUMC)
- Prof. Ronald Stolk (UMCG)
- Dr. Morris Swertz (UMCG)
- Prof. Cisca Wijmenga (UMCG)
- Dr. Michel Wouters (NKI/AVL)
- Prof. Gerhard Zielhuis (Radboudumc)

Authors and contact information

Name of the infrastructure	**NL Personalised Medicine & Health Research Infrastructure**
Author	Prof. dr. Cisca Wijmenga
Organisation Function	Universitair Medisch Centrum Groningen Professor of Human Genetics and head of the Genetics Department
Address	Dept. of Genetics Antonius Deusinglaan 1 9713 AV Groningen
Telephone Email	+31 50 36 171 00 cisca.wijmenga@umcg.nl
Co-authors	Prof. dr. Gerrit Meijer (NKI) Prof. dr. Barend Mons (LUMC) Prof. dr. Peter Luijten (UMCU) Dr. André Dekker (MUMC+) Dr. Ruben G. Kok (DTL)
Contactperson	Dr. Ruben G. Kok
Organisation Function	Dutch Techcentre for Lifesciences (DTL) Director
Address	Catharijnesingel 54 3511 GC Utrecht
Telephone Email	06-30642350 ruben.kok@dtls.nl

Annex 1: Health-RI in relation to National Science Agenda & Topsector KIAs

Summary

The Health-RI research infrastructure at a glance

In 2025, we envisage a globally unique research infrastructure in the Netherlands that will both drive and support cross-disciplinary research into personalized medicine & health and optimize personalized healthcare.

The overall aim is to enable frontier science and technology development in the field of personalised and high precision medicine and health with high reproducible output. The infrastructure will become the national platform for high-quality experimental design and high-quality measuring with high-quality data stewardship and high-quality data analytics.

To reach this aim, the infrastructure will:

1. Provide technology platforms that allow for complete genome analysis combined with detailed phenotyping of the participants based on metabolomic and genomic parameters, imaging data, lifestyle information, insight into the microbiome, and electronic patient records. Increasingly this will depend on direct interaction, data collection and sampling by and from healthy individuals and patients, as well as providing these participants to receive direct feedback on analyses performed on them supported by mobile electronic devices (e-Health).
2. Drive the collective research and development of novel wetlab and ICT technology, and support the advanced design and execution of medicine and health-related multi-disciplinary research projects.
3. Provide an open platform that serves as the backbone for biomedical engineers to validate and share frontier technology and methodology application, to build services that support the biomedical research of 2025, and to seamlessly connect them to clinical and health practice;
4. Stimulate both academic and precompetitive research, as well as clinical and industrial innovation
5. Involve citizens / patients and their collectives in research (P4 Medicine & Health and citizen science) and stimulate the sharing of their data for research purposes;
6. Provide medical doctors a window on the international knowledge base and an integrated platform to rapidly share and retrieve expertise and information relevant for research and improvements in care methodology to improve the quality and precision of treatments and reduce costs of care;

Incentivise privacy and ownership-preserving data sharing among stakeholders and collectively build a world-class resource of actionable knowledge and information that will serve as the crucial reference base for validation of project outcomes of future health research and health care;

Keywords:
1] predictive, preventive, personalised and participatory (P4) medicine & health; 2] high-precision medicine; 3] biobanks and cohorts; 4] genetics, -omics & bioimaging; 5] FAIR data exchange; 6] e-Health; 7] experimental design and decision support; 8] data quality and reproducibility of research

1. SCIENCE & TECHNOLOGY CASE

1.1 *Science Case.*

We are living in a society that tries to learn how to cope with chronic diseases that affect our vitality, such as obesities and diabetes, cardiovascular and lung diseases, and of cancer and neurodegenerative disorders. The global quest towards understanding both health and diseases is speeding up, and new knowledge is being developed, validated, adopted and implemented at increasing speed every year, not the least through the application of novel disruptive lab technologies such as genomics and bioimaging, and mobile technology in home-care (e-health) and available to citizens to longitudinally measure their 'quantified self'. As these technologies become instrumental in the early prediction and potential prevention of disease, they will pave the way for a socially and economically sound health and medicine system that tailors prevention and care to individual citizens and at lower cost.

There is a strong need to capture and integrate the continuously growing and updating global body of information into better evidence-based models of health and biomedicine. At the same time we need to be able to more rapidly apply the latest life sciences knowledge in clinical practice. While building an integrated knowledge base it is crucial that we close the '*innovation gap*' and reduce the time from proof of concept stage to validation and implementation in personalised prevention and precision healthcare. It is evident from the major investment programmes recently launched in the US (January 2015) and in China (January 2016) that this approach is globally seen as crucial and urgent[3].

To grow towards a system of personalised medicine and health there is a great demand for a next generation infrastructure that can bundle and connect the expertise, methodology, equipment and data resources from specialised molecular, clinical and imaging laboratories, biobank and population analyses, home-care and e-health platforms as well as sources of quantified self-type information on personal nutrition and lifestyle. The richer the collective research resource, the larger the potential for a deep and practical understanding of an individual's health phenotype. Building such an infrastructure includes many disciplines and stakeholders: biomedical scientists and research assistants, biologists, technology experts, computer scientists, healthcare professionals, data and modelling experts, patient organisations, industry and the government.

To support our proposition to build a nation-wide infrastructure for personalised medicine and health research we focus here on the combination of driving biomedical innovations needed (Biomedical Science Case) and on the complementary innovations needed in data and ICT (Computer Science Case).

[3] See: http://www.nature.com/news/us-precision-medicine-proposal-sparks-questions-1.16774 and http://www.nature.com/news/china-embraces-precision-medicine-on-a-massive-scale-1.19108?WT.mc_id=TWT_NatureNews

A] Biomedical Science Case

Sustaining an individuals' health potential

1. Every human being has a unique potential health profile, strongly determined by genetic factors in the developmental stages during a person's lifetime. Environmental factors, disease-causing agents, microbiome composition, social behaviour, lifestyle and nutrition all influence this potential health curve, and: this is different for every individual and changes over time. An excellent example has recently been published by Zeevi et al. in Cell[4], where multiple genetic, physiological and metabolic parameters were measured in a group of 800 people, showing clearly how all subjects responded in a unique manner to certain food intake. We barely understand how these individual differences occur. We do not yet understand the coherence among these factors influencing personal health, and how they cause the gap between potential health and actual health: the Health gap! (see fig. 1).

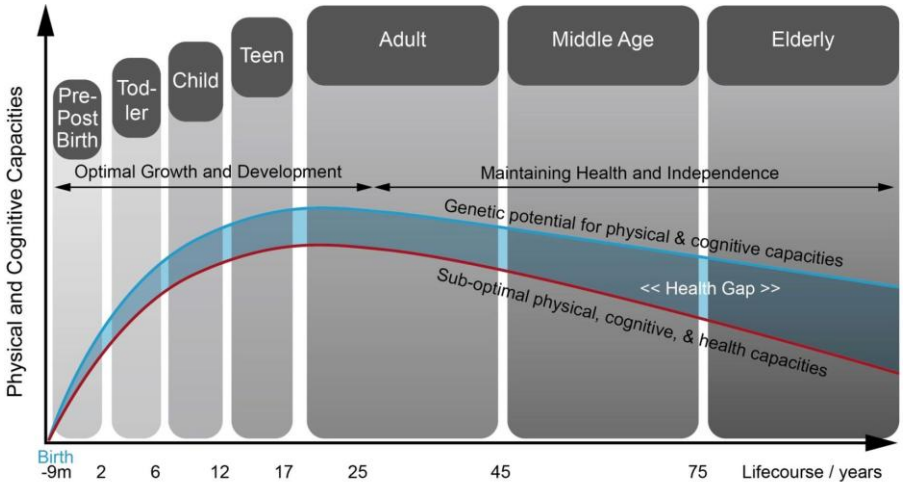


Figure 1 Kaput et al. (Genes Nutr. (2915) 10:12)[5]

Health models needed, based on understanding biological complexity and variation

2. To find out what can be helpful in prevention or intervention when someone deviates from his / her personal health potential, it is crucial to have a basic understanding of an individual's health profile: we need the evidence base for personalised medicine & health. This requires health models based upon an understanding of the complex biological processes that affect functioning and health in an individual. Over the last fifteen years, technological developments have made it possible to measure in many different ways at every known biological level – from molecular, cellular, organ to organism and population level. Based upon these measurements scientists have uncovered a vast complexity in biological systems: a network of interacting factors determines growth, proliferation and resistance to stress inducing factors and negative external influences. We now know that physiology and robustness of a biological system are in part dependent on mechanisms of (locus-specific) regulation, post-translational modification, redundancy and cooperation. What is not yet understood, is how these mechanisms interact: we lack the insight to go from descriptive to predictive modelling in life science, a prerequisite if we want to reach controlled health preservation.

Individually tailored analysis requires new science approaches

3. Translation to personalised medicine & health also requires that we can make individually tailored analyses. To do so we need a much broader knowledge base for reference, and an infrastructure that offers the combined access to for example high-resolution measurements with minimally invasive analytic and imaging techniques and to statistical data-model development tolerant to the complex variation between individual systems and between individuals.

[4] Zeevi et al. (Cell, Vol. 163, Issue 5, 19 Nov. 2015): http://www.cell.com/abstract/S0092-8674%2815%2901481-6
[5] Kaput et al. (Genes Nutr. (2915) 10:12): http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4549339/

High precision technologies to measure with minimal damage and deliver treatment on the spot

4. A range of novel technologies such as image guidance, nano-robotics and genomic mapping strategies will help materialize high-precision medicine. The goal is to maximise effect and to minimise the burden for an individual patient. This requires for instance high-precision treatment and companion diagnosis for the prognosis of disease evolution, to predict treatment efficacy, monitor treatment response and keep patients under surveillance for early detection of disease recurrence. Advanced technologies such as imaging, proteomics and metabolomics will be used for the structural, functional and molecular assessment of disease and for a better understanding of body functions. This comprises research and technical developments in diverse areas such as genomics, hybrid imaging, ultra high-field MRI, light-sheath microscopy, electron microscopy, advanced mass spectroscopy and 'labs on a chip'. Also, tissue engineering and stem cell applications arise for prognostic personalised testing of treatment and in regenerative medicine through highly localised repair and regeneration tailored to specific tissues of individual patients. In all these fields multidisciplinary research (mathematics, physics, chemistry, biology and medical science) will interact closely around new infrastructural facilities for the further advancements in these research domains. These expertise centres have a crucial role in the development of novel technologies and technology applications, and in the design and execution of high-quality measurements to boost the reproducibility of research. Making these expert centres an integral part of the envisaged infrastructure will enable the consolidation of existing infrastructures, help prioritise collective investments, and help tune the combined application among disciplines.

Connecting life science, medicine and quantified self approaches

5. Next to the data, facilities and resources built up in life science research groups and health care organizations, *e-health* and *quantified self*-type initiatives gather a wealth of health-related personal information (incl. lifestyle, nutrition, home care), currently manifold initiated or supported by private parties, and geared to empower individuals to steer their personal health. Such initiatives will increasingly generate a vast array of novel longitudinal health data collections of healthy and diseased individuals. *Health-RI* is set up to merge the power of these personal e-health resources with the existing knowledge and information base, and create a goldmine for scientific discovery and product innovation, in an area with direct societal impact: personal health.

Requirement: combining disciplines and connecting distributed resources

6. Life science research resources (facilities, collections, biobanks, databases, etc.) have been built up over recent decades by many disciplines and by scientific institutes, industry and other stakeholders world-wide. They are using a wide variety of advanced techniques and methodology, and the resulting output data are vast in size and of heterogeneous formats. The result is a system in which half (!) of the experiments are hard to reproduce[6]. If we want to get to the next level of understanding of the biological complexity underlying personal medicine & health and rapidly translate this knowledge to health and medical practice we need to be able to better combine the knowledge and information captured in these complementary disciplines and distributed resources. An integrated infrastructure will open up the possibility to easily exchange across scientific and technological disciplines, samples, methodology and data, and to foster systems-level experiments by supporting the whole process from experimental design to high quality analytics to data stewardship.

[6] Freedman et al. (PLoS Bio., 9 June 2015): http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002165

Only with a connected infrastructure for personalised medicine & health research can we move from the traditional reductionist approach to a holistic approach in research, and through a commons-approach break the barriers of distributed incompatible disciplines and resources. This opens up the road to finding answers to fundamental questions in biomedical research related to rapid translation in societal practice, such as:

1. ***How can we functionally combine a variety of advanced technology platforms (e.g. genomics, proteomics, metabolomics, imaging, e-health), and how can we meaningfully combine the readouts derived from these technologies with information derived from other sources, such as quantified-self and e-health data, information on lifestyle and nutritional status, and/or with information derived from patient reports, social media, socio-economic information?***
2. ***How can we minimize patient burden and hospitalization by therapeutic, personalized interventions that can be delivered with hitherto unprecedented precision in a cost-contained manner?***
3. ***How can we understand health and diseases mechanisms in terms of their interactive components at the level of individual persons, and how can we include their development stages to allow evidence-based detection of early and reversible phases of diseases such as obesities, neurodegeneration, psychiatric disorders, cardiovascular diseases and cancer;***
4. ***What is the bandwidth of the phenotypic variation underlying health and functioning between individuals, and how can this variation be explained by the coherence of components in human biology within individual biological systems at cellular, organ and organismal level?***
5. ***What causes the deviation between the actual health profile of an individual person and his/her health potential, and what interventions can we develop to close this' health gap'? Or stated from an individuals' perspective: how can I improve my health, considering my current health status using the current scientific knowledge and learning from the response of phenotypically similar persons. How can we identify and validate the intervention that provides most added value for an individual citizen/patient against reasonable cost and impact?***

B] Computer Science Case

Data stewardship as the basis for sharing and analytics across datasets

7. Scientists in personalised medicine & health, are gradually adopting the active data cycle of experimental design to data analysis and finally to stewardship of data. Unique to this data cycle is the conditional use of data preservation and meta-data attribution for future reference and re-use of the data. Current science policies promote open science and open data resulting from public funding sources, and to a form of data stewardship which ensures that data sets across the world are stored in Findable, Accessible, Interoperable and Re-usable formats: data must be made 'FAIR'[7,8]. Increasingly, these research data are being shared and published making datasets available for others to use, in many cases as part of the public domain. For example, infrastructures such as ELIXIR and BBMRI[9] do not only share methodology and standards, but also facilitate re-use of data across institutions, and the deposition in well-curated international data collections to support science and innovation. Similar initiatives arise across the life sciences and healthcare. Currently it is estimated that just over 10% of international scientific information is deposited in well-curated databases that can be accessed for re-use of the data[10].

[7] FAIR principles (Force11): https://www.force11.org/group/fairgroup/fairprinciples
[8] Wilkinson et al. (*FAIR Data: Guiding Principles for Scientific Data Management and Stewardship, submitted to Nature*)

[9] See: http://www.elixir-europe.org and http://www.bbmri-eric.eu
[10] Read et al. (PLoS One, 10.7 (2015): e0132735): http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0132735

With strong restrictions in (European) regulation on the use of personal (health) information it is even more imperative that a trusted environment is created that helps its participants to build the standards, security protection, certification and rewarding systems to share their data in full concordance with government regulations (and as a driver to improve these) and with respect for intellectual property in industry ('open when possible, closed if needed').

Privacy-sensitive data in healthcare and health research are collected in a decentralised manner

8. Not only data from research projects democratise. In the health sector we are witnessing a strong trend towards decentralisation of privacy-sensitive personal health data. In 2025, every Dutch citizen will hold his / her personal health data in a personal digital locker: the Personal Health File (PGD: 'Persoonlijk Gezondheids Dossier')[11]. This will include data from health professionals, from studies in which the person has participated, self-acquired data from e-health and quantified-self apps, maybe even social media data. Every citizen will be in full control of the use of these data, and next to using the data themselves, many of them will make data available to science, enabling citizen science approaches. Whatever the resource, from the perspective of its owner, the available information will only become valuable if connected with external information, including reference information from the health and biomedical science fields and with enough meta-data in place. From the perspective of the field of personalised medicine & health, there is great value in connecting these personal health data 'lockers' for scientific discovery in a manner that fully preserves privacy. Interestingly, emerging polymorphic encryption technology opens up the possibility for individuals to actively control the contribution of their personal data in specific studies, both for their own interest and for the sake of scientific progress.

From data sharing to distributed big data analytics with in-built 'biological understanding'

9. Big data analysis, rooted in complex algorithms and statistical methods, already shows us how computers are able to recognise patterns in unstructured, highly complex and large-scale data sets and help us test hypotheses in these data. The next steps in life sciences – and in translation to personalised medicine and health – requires an impulse in high-performance big data analysis, not only at the methodological level, but also at the level of compute power and interconnectivity of advanced data and ICT systems. Just like decentralisation of data, data-processing systems will also need to be highly distributed. Some data just cannot be transported over networks because of size limitation or privacy/ownership concerns. The need to combine data sources towards an integrated understanding of health thus gives a strong boost to develop ICT-solutions that allow integrated reasoning among distributed ICT systems. Using semantic web technology, computers can be trained to understand how datasets and disciplines are interrelated. For example: the computer can learn to understand how genetic analysis at the model system "zebra fish" translates to information for quantitative and qualitative model development and hypothesis generation in human biology. Using a data interoperability approach in combination with approaches for distributed learning allows for big data analytics with in-built biological understanding. ICT research and innovation thus finds great inspiration in the distributed data ecosystem of the life sciences, and becomes intertwined with the construction of a world-class 'data web' infrastructure for both ICT and health-related science fields. This connected model will also be at the heart of the future '*European Open Science Cloud*'[12].

Entirely novel information science paradigms will be needed in the future.

10. Current state of the art interoperability techniques rely on fully exposing at least the ontologies, and often also the data items of these systems. However, current paradigms do not allow establishing semantic links between datasets that cannot be fully exposed to each other because of privacy risks. The distributed learning architecture foreseen can be generalised to a generic multi-agent setting (such as those that have been studied in Artificial Intelligence). Each data set and each workflow visiting it (see e.g. Personal Health

[11] See: https://www.npcf.nl/themas/persoonlijk-gezondheidsdossier/

[12] See: *http://horizon-magazine.eu/article/european-science-cloud-horizon_en.html* and *http://ec.europa.eu/research/index.cfm?pg=events&eventcode=749A307D-EB91-87FA-4CB7DC515B17BF5D*

Train[13]) can be regarded as a single agent operating in a multi-agent environment, where data-providing agents and data-consuming agents negotiate over the terms on which data can be exchanged. In this framing, personal health data lockers are all interpreted as (small) individual agents (of which there are then millions), while institutional agents are fewer, but of course much larger. They can be viewed as 'devices' in the Internet of Things, which will soon connect > 40 billion interacting devices.

Novel ICT protocols

11. The ICT systems of the future will clearly need to disseminate the knowledge gain in a faster and more structured way to health professionals and citizens. Rapid learning and validation so that new knowledge is robust enough to be used for an intervention (e.g. in a medical device) requires extreme caution and high quality and security requirements to both the ICT systems Protocols have to be developed that allow for secure and privacy-preserving transactions across data resources and handling systems. The nature and scale of these transactions will seriously challenge the development of novel encryption technology, such as polymorphic encryption and pseudonymisation technology pioneered in the Netherlands that provide security keys at the level of the data plus the transaction/transit process. The system should thus allow to evaluate requests for data access against privacy and ownership/licence concerns in a standardised manner. Also, dynamic consent and rewarding mechanisms should be included that offers active data providers (which might be individual patients, patient groups or companies) with access to other data that is valuable to them in exchange for access. This is in line with modern ideas on data ownership and "data as the new currency"[14]. Tracing the transaction as part of the provenance model of the *Health-RI* infrastructure may require blockchain technology as used in bitcoin systems. Such systems combine full traceability of transactions with protection of privacy concerns.

Overall, development of a connected and distributed infrastructure for personalised medicine & health research poses serious challenges to fundamental computer science topics. A number of key required computer science questions are:

1. **How can we realise semantic interoperability of heterogeneous datasets under limited data-exposure conditions?**
2. **How do distributed data storage, processing, simulation, modelling and analytics applications affect the outcome and performance of analyses and how can this be optimised?**
3. **How can we optimise the architecture of an analytics system of (potentially millions to eventually billions of) personal data 'lockers' containing privacy-sensitive personal information?**
4. **How can encryption technology ensure include dynamic consent policies and at the same time enable automated analytics while preserving security in terms of privacy and ownership of personal and private health data exposed as FAIR data resources?**
5. **How can we secure proper versioning and provenance in an arrangement of an 'internet of data' analytics environment that should be able to grow to global scale implementation (contribution and usage)?**

The above sketched scientific challenges are exemplar drivers to establish a strong national infrastructure that is embedded in the Dutch science, innovation and public health system. The infrastructure will establish the single binding national platform for P4 Medicine and Health research and innovation in the Netherlands, connecting all stakeholders that offer high quality research resources and perform cross-disciplinary biomedical research: research centres and their scientific and technology communities, clinics and healthcare organisations, companies, government bodies and funders, as well as (collectives of) citizens and patients. With close involvement of the frontier NL-ICT sector and computer science and bioinformatics communities, a world-class connecting distributed analytics ICT environment will be developed and implemented as a linked-data

[13] See: http://www.personalhealthtrain.nl
[14] See: http://deloitte.wsj.com/riskandcompliance/files/2013/11/DataCurrency_report.pdf

backbone of the *Health-RI* platform. The infrastructure will become a major driver for frontier research across these science domains (life science – medicine - computer science - data science).

Although the described infrastructure would only be fully functional in 2025, it fits seamlessly in the National Science Agenda recently published by the KNAW, in the 'Implementatieplan Nieuwe Biologie' [15] and in the Knowledge and Innovation Agenda's (KIAs) of several sectors, most prominently Life Science & Health, Agri&Food and ICT, see Annex 1. This already gives a sense of urgency to rapidly start building the infrastructure in the coming years. In this process the infrastructure will become a strong binding factor to link science programmes to innovation and economic and social development across several sectors. ***Health-RI* will become a magnet to attract both human capital and financial capital towards the Netherlands**.

1.2 Expected scientific advantages and breakthroughs

Health-RI as a national infrastructure will be a crucial facility to design and perform systems-level research across disciplines, technology platforms and growing data collections in personalised medicine & health. It will help realise a significant improvement in the reproducibility of biomedical science output and deliver a strongly enriched evidence base for personalised medicine and health research and interventions. This proposal is aimed at creating the roadmap towards the ideal infrastructure based on the current and potential strengths of the Dutch life sciences and addressing the emerging needs to sustain and strengthen our scientific leadership position in this field. If we fail to realise a globally compelling infrastructure such as *Health-RI* we will greatly lag behind in the international arena, especially given the global emergence of major programmes on personalised medicine and health in the US, China, Scandinavia, Switzerland and Australia.

The infrastructure will:

- Boost the development of connected high-quality research resources such as biobanks, population cohorts, advanced technology facilities and data collections across disciplines and stakeholders;
- Continuously push the boundaries in health and medical technologies by driving their frontier development and combining technological innovations among academia and industry
- Stimulate the development of novel and secure technology and tools for data sharing and analysis, starting the 'internet of health and medicine data' as a crystallisation point for a global Internet of (scientific) Things and Research Objects, built on top of the current world-wide-web.

As the collective platform connecting multiple stakeholders, the infrastructure will

- Enable the advanced design and execution of health-related multi-disciplinary research projects, bridging life science and computer/data science disciplines;
- Enable citizens / patients and their collectives to actively participate in biomedical research and to share their (personal) health data for research purposes;
- Provide medical doctors a platform to share their data and exchange expertise and information relevant for personalised medicine and health research;
- Provide industry a standard backbone infrastructure for their health and medical technology innovations driving interoperability across platforms;
- Provide an open platform for biomedical engineers to validate and share frontier technology and methodology and to build crucial infrastructures and services that support the biomedical research pipelines of 2025

Expected breakthroughs of the infrastructure with national and international impact:

- Highly enriched knowledge base for innovations in medicine and health that will turn fatal morbidities into chronic diseases

[15] Implementatieplan Nieuwe Biologie (NIBI, 2013)

- A much better understanding of how our brain actually functions from cradle to grave (developmental disorders, neuro degeneration, brain computer interfacing) and how socio-economic factors may influence our personal health and functioning
- A deep understanding of the composition and role of our microbiome and how it interacts with our body functions, also in relation to nutrition and lifestyle as a co-determinants for health
- An understanding of the fundamental regulatory processes underlying the health potential of individual citizens
- Identified interventions in terms of prevention and early diagnosis to help citizens to retain health and functioning, and to treat patients with high-precision and tailored to their genetic and physiological make-up
- Transition to a P4 medicine and health system in the Netherlands (and globally) leading to a full empowerment of individual citizens to use their personal health data and contribute to science
- Close the innovation gap between science and implementation in P4 medicine and health care
- Retain the value of hundreds of millions worth of research investments in the Netherlands alone that will lead to reusable and reproducible scientific output
- Help provide handles to decrease costs of care, improvement of treatment and quality of life
- Ground-breaking biomedical technologies for high-precision and minimally invasive personalised diagnosis, prognosis and treatment;
- Frontier ICT concepts and technologies for data and protocol encryption, data interoperability and distributed learning across privacy-sensitive data resources
- Launch of the global development of the internet of data

1.3 Health-RI as a distributed infrastructure building upon existing resources

Health-RI is foreseen as a radically new infrastructure in its level of aggregation. Billions of public money collectively invested over the last decades have led to crucial expertise and output, but also to a highly fragmented landscape of local resources, most of them not easily findable or accessible. *Health-RI* establishes the essential connecting layer that makes these local resources an integral part of the collective infrastructure (based on rules of engagement to the platform (see below). The infrastructure helps prioritise and further develop these into national research resources. What results is a world-class facility that will significantly improve experimental design, execution and reproducibility and reduce loss of data in the medicine and health domain.

The initiators behind this proposition, BBMRI-NL, DTL/ELIXIR-NL and EATRIS-NL have already successfully created the first generation of connected resources to pilot the integrated approach of *Health-RI*. In BBMRI-NL[16] for example, over 200 Dutch biobanks and population cohorts and their related data collections have already been assembled, and these resources are being opened up more and more as resources accessible to the broader life science community. Similarly, in DTL[17] strongly capitalising on programmes such as the Netherlands Genomics Initiative, the Centre for Translational Molecular Medicine (CTMM[18]) and the Dutch microscopy community, over 100 technology expert groups have assembled their open research facilities[19] in the areas of genomics, proteomics, metabolomics, medical imaging and advanced microscopy, physiology, bioinformatics, e-science and systems biology. Meanwhile, DTL coordinates the Dutch node in ELIXIR[20], the European infrastructure of core biological reference data resources with the tagline 'data for life'. CTMM-

[16] http://www.bbmri.nl
[17] http://www.dtls.nl
[18] http://www.ctmm.nl
[19] http://www.dtls.nl/expertise-facilities/facilities/
[20] http://www.elixir-europe.org

associated groups are closely involved in establishing the Dutch node in EATRIS[21], focussing on high-quality medical imaging and process management in translational research. With help of the above initiatives, the Dutch UMCs have recently launched the NFU data4lifesciences programme[22] to develop a collaborative ICT and data environment. At the same time there are growing relationships with the federation of Clinical Specialists (FMS) and clinical auditing[23] and initiatives in regional care and e-health[24].

Major other relevant resources and initiatives can be found across the life sciences and ICT sectors, e.g. focussing on the effects of nutrition and lifestyle on health, and manifold involving public-private collaboration (e.g. TIFN: Top Institute Food & Nutrition[25]). In the e-science domain, stakeholders such as SURF and DANS have developed the current-generation infrastructures and protocols for the Dutch science field to perform high performance computational analyses and long-term archiving, respectively, all in close link with international initiatives (EGI[26], EUDAT[27] and PRACE[28]) and RDA[29].

Several of the above initiatives already started the crucial formation and harmonisation of biomedical and e-infrastructures, each with strong connection to peer international community initiatives. Collaboration at the national scale among these initiatives is also growing fast, among scientists, engineers and involved governing boards. This has established a strong basis to effectively build the proposed nationwide infrastructure. Anno 2025, the above initiatives will have aggregated their high quality resources into a single national platform. Health-RI will be a unique and comprehensive infrastructure that enables multidisciplinary teams of scientists and other societal stakeholder groups in experimental design and validation and in overall execution of their frontier research into personalised medicine and health.

1.4 *Health-RI is unique in its level of aggregation .*

As the sense of complexity of life processes rises internationally, it is imperative that research to improve our understanding of human health can be done in cross-disciplinary fashion, and that resources built up over time become more easily combinable and reusable. Visions about realising evidence-based medicine and personal health become obsolete if we do not manage to combine the scientific, technical, ICT and data expertise, and open up the information and research resources for future science and innovation programmes. *Health-RI* focuses entirely on this aspect: realising a single unified infrastructure of connected high-quality resources that can be used in one go in cross-disciplinary studies.

Although this will take a tremendous effort of a large group of stakeholders, the partners behind this proposition and their endorsers are determined to establish this level of integration of the envisaged infrastructure. The alternative would be to stick to business as usual and keep investing in a fragmented field that has only 50% reproducible output anno 2015.

[18] http://www.eatris.eu
[19] http://www.data4lifesciences.nl
[20] https://www.clinicalaudit.nl and http://www.iknl.nl
[24] https://ecp.nl/actueel//4073/ecp-start-met-vitavalley-platfom-langdurige-zorg-met-ict.html and https://www.npcf.nl
[25] http://www.tifn.nl
[26] http://www.egi.eu
[27] http://www.eudat.eu
[28] http://www.prace-ri.eu
[29] https://rd-alliance.org

1.5 Technical case.

Health-RI is a strongly integrated national infrastructure. It operates as a common networked infrastructure of interconnected high-quality research facilities and other essential resources of national and international value in participating university medical centres, universities, research institutes, contract research organisations, e-infrastructure providers, healthcare organisations and a wide range of companies.

All facilities and resources included in the *Health-RI* infrastructure are the result of collective development within respective expert communities and of assessed quality according to latest field standards. All resources adherence to the 'rules of engagement' drawn up by the collective governing board of the infrastructure.

In 2025, the infrastructure consists of (see Figure 2):

- Clinical biobanks, population cohorts, clinical and health-care-related information systems;
- Well-accessible and interconnected facilities in next generation sequencing, proteomics, metabolomics, advanced microscopy, clinical imaging, bioinformatics and computational (systems) biology, computer and data science;
- National-level life science and health-related data repositories plus well-annotated international reference data collections, all in FAIR format to support re-use and validation of biomedical research output in cross-disciplinary studies;
- National-level repositories of e-health data, including aggregated quantified-self data collected by citizen collectives (e.g. parts of Personal Health Files);
- Collective repositories of adaptable research workflows and healthcare decision support tools;
- A next-generation linked-data & workflow exchanging e-infrastructure to allow advanced levels of data and information sharing as well as analytics across distributed resources (Figure 3 below).

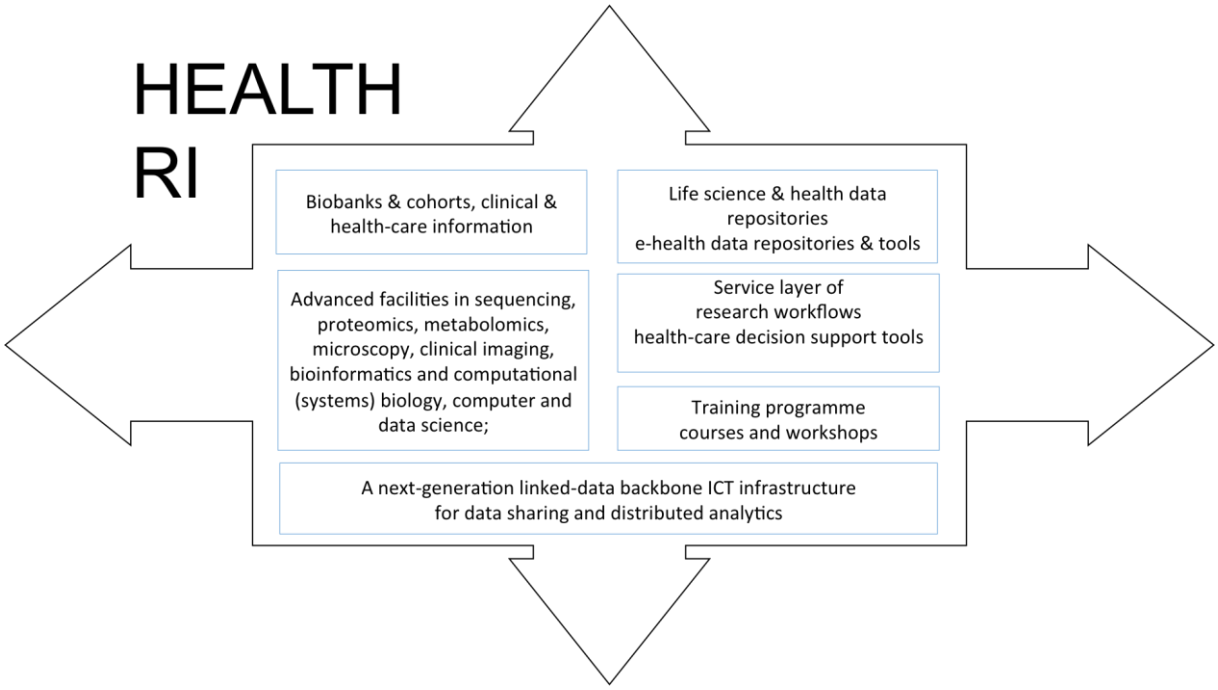


Figure 2, Overview of the Health-RI infrastructure

Besides the physical linked-data backbone, the connective tissue of the *Health-RI* infrastructure will be formed by several process-level elements and by a common training programme:

- Joint governing regulations, quality and data standards and certification, rules of engagement and harmonised user-access guidelines
- A roadmap-process to support the prioritisation in development and integration of core resources within the *Health-RI* infrastructure
- A web-based 'dashboard' that provides access to all core-resources, services and tools, tested methodology, expertise and best practices, and an active programme to help stakeholders prepare novel resources that should be included in the common infrastructure
- A comprehensive training, education and outreach programme built on the capacities of all participants, and tailored to raising the next generation personalised medicine and health researchers.

The collectively governed infrastructure will actively support communities (feeders) to develop and harmonise their research resources or services so that they can be included in the infrastructure or built on top of the core resources. As operational infrastructure, it will provide services to all involved user stakeholders (users) to optimise their personalized medicine and health research process to high levels of reproducibility, from hypothesis generation to experimental design, to high-quality measurements and data analyses, biological interpretation, dissemination and data stewardship for preservation of methodology and output, and for future reuse in experimental validation.

Health-RI is here proposed as the national infrastructure for personalised medicine and health research and innovation, but is conceived as a European hub connected to all relevant international biomedical and technical infrastructures (e.g. BBMRI, ELIXIR, EATRIS, EuroBioimaging, INSTRUCT), as well as to e-infrastructure initiatives such as the European Open Science Cloud and US NIH Big-Data-to-Knowledge (BD2K), making it collaborative across disciplines, borders and industries.

Technical headlines

BioBanks & cohorts

The layer of research resources comprises clinical biobanks and population cohorts and their data repositories, such as pathology collections (e.g. PALGA), large population cohorts (e.g. LifeLines, the Netherlands Twin Register and Generation R); large clinical biobanks (e.g. Parelsnoer Institute), as well as smaller cohorts. These initiatives provide the best formalised and documented biomaterials and annotated clinical information to support medicine and health research.

Technology facilities

This layer of advanced technical facilities bundles the capacities in a number of essential enabling wet-lab technology fields, such as next generation sequencing, mass-spectroscopy for proteomics and metabolomics, advanced light microscopy, electron microscopy as well as clinical imaging (ultra-high-field MRI, PET, NMR, ….) for functional & molecular imaging. Also, expert 'dry-labs' are involved in the fields of bioinformatics and computational (systems) biology, computer science and data science.

The associated facilities comprise a balanced combination of

- Technology Innovation Labs that continuously push technology boundaries to enable truly novel applications in health and medicine-inspired research. These groups pioneer cross-technology integration and (modelling of) datasets.
- Technology Hotels that provide access to high-end instrumentation for other researchers and provide expertise in applying high-end analytical technologies in their biological research projects. The key driver is here to enable others to perform cutting-edge research. Access to the Hotels' advanced

expertise and infrastructure can be offered on a collaborative and/or a cost-recovery basis, depending on the nature of the project and the research group.

- Technology Service Providers, manifold companies, offering technology services at a defined service level, mostly on a fee-for-service basis. The aim here is to provide (cost-)effective access to state-of-the-art technology services for research groups and other companies within the life sciences field.

The facilities must already have reached a strong level of national aggregation and tuning before they are accepted in the Health-RI infrastructure. As part of the Health-RI infrastructure, they connect seamlessly in the cross-disciplinary infrastructure.

Life science and e-Health data repositories

This layer of resources of the infrastructure focuses on the creation and acquisition of well-annotated collections of data and information assembled throughout the Health-RI partnership. Repositories that are highly valuable for re-use in personalised medicine and health research will become part of the core resources of the infrastructure. Through the linked-data backbone, users of the facility will also obtain access to core international reference data collections, such as those collected in ELIXIR[30]

Service layer of workflows enabling personalised medicine & health research

An important functionality of the *Health-RI* infrastructure will be the support of the entire research cycle from hypothesis generation to experimental design and execution of 'data-intensive' research projects, and to data stewardship in FAIR format. To prevent reinvention of existing methodology and strive towards next levels of standardisation in support of experimental reproducibility, *Health-RI* will provide a platform to exchange and construct (standard) experimental workflows and advanced analytics and modelling pipelines, including those used for distributed learning on the *Health-RI* linked-data backbone. A dedicated service layer of the infrastructure will offer these workflows through a web-based 'dashboard'.

Training programme

To support the next generation of scientists active in the future cross-disciplinary, high-tech and data-driven personalised (P4) medicine and health research, it is imperative that a major training effort is made part of the infrastructure. *Health-RI* will involve all its educational partners from academia, universities of applied science and industry to work on a comprehensive programme of education, training and outreach covering the many disciplines at multiple levels of education. Nation-wide research schools and local graduate schools at universities will be important partners to secure advanced levels of education that fit future research skills.

Linked-data ICT infrastructure

The ICT backbone of *Health-RI* will largely be built as a 'life science and health workflow & data exchange". This part of the infrastructure will support seamless access, interoperability, re-use and trust of data among all the above resources contained within the infrastructure. Highly specialised reasoning algorithms will help process data as part of migrating research workflows, making it possible to go beyond observation, theory and simulation into exploration driven science by mining new insights from vastly diverse data sets.

The *Health-RI* linked-data backbone essentially offers a storage, compute and analytics backbone based on distributed learning concepts described in the Computer Science case above. The operational architecture of the infrastructure can be visualised as follows (see Fig. 3):

[30] http://www.elixir-europe.org/services

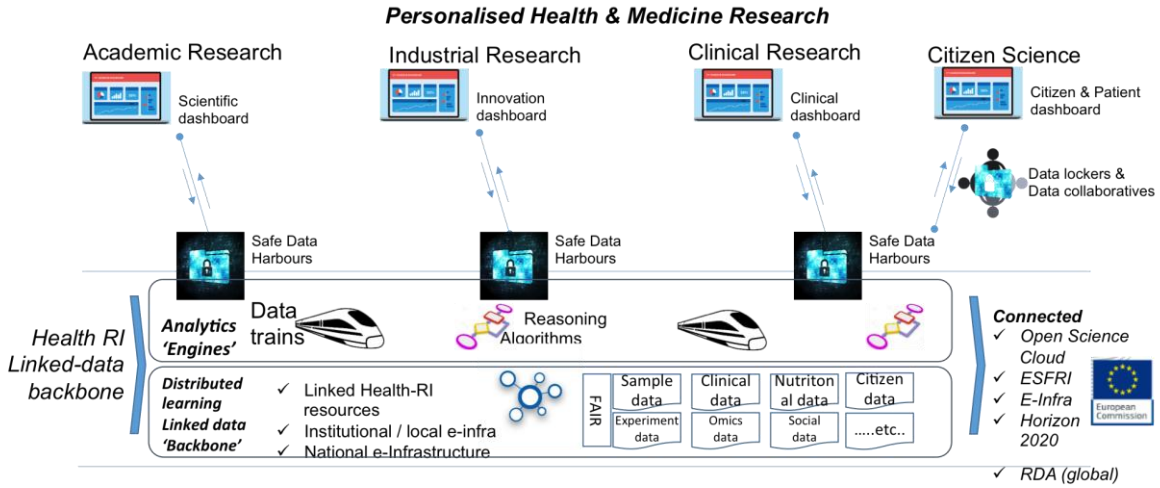


Figure 3: impression of the functional architecture of the linked-data backbone of the Health-RI infrastructure

Resources that want to connect to the *Health-RI* linked-data backbone are supported to make their (meta-)data FAIR and expose these for access by data 'trains' in so-called FAIR Data 'stations'. On the demand side, users are supported to manage their research workflows through a foreseen research 'dashboard' that also provides the overview of connected resources and facilities. *Health-RI* offers full workflow support to execute standard workflows or design and build novel analytical functionalities using existing workflow building blocks as much as possible for standardisation and easy validation purposes. Many of the processing/analysis workflows ('data trains') will visit the data stations instead of moving the data to the processing location, and they will bring back only the results and not the source data to a research project. The e-infrastructure for this linked-data backbone will be modelled after the European Cloud for Open Science and realised with the involved e-infrastructure expert partners.

1.6 Whats new?

1.6.1 Proven aspects

The launching initiatives behind this proposition, BBMRI-NL, EATRIS-NL and ELIXIR-NL/DTL have already proven essential aspects of the envisaged nation-wide infrastructure for personalised medicine and health research, connecting biobanks and accessible research facilities (technology hotels) across life science organisations, and providing the standards for quality experimentation and data stewardship including the principles for the linked-data backbone.

Meanwhile, close links have been already established with related initiatives that cover other essential skills and technologies (e.g. LifeLines, PSI, NFU data4lifesciences, PALGA, IKNL, DICA CTMM/TI-Pharma (Lygature), NMC, NL-Bioimaging-AM, TIFN, ENPADASI, SURF, DANS, NLeSC). The proposed integration receives broad support from within the biomedical and clinical research communities and from the LSH sector. It is especially the harmonisation of protocols and the combination of high-quality experimental design, technology development and high-quality measurements, frontier data analytics and data stewardship that attract strong support. New organisational concepts have meanwhile been established through DTL, a federated platform set up by a broad range of academic organisations and a growing group of industrial partners to connect the various life science disciplines and their resources.

1.6.2 New challenges

In order to realise the infrastructure, there will be technical, societal, scientific and organisational challenges that need to solved. A number of crucial aspects that have not been covered in the biomedical and computer science cases above are highlighted here:

Technological challenges

- Functional combination/integration of advanced technologies in biomedical research
- Combining research data collected in high-tech laboratories with e-health data
- High-performance, scalability and security aspects of distributed data mining and machine learning

Societal challenges

- Social, ethical and legal issues related to ownership and privacy preservation of collected human biomaterials and scientific output based upon these, and in in data/workflow sharing, incl. dynamic and informed consent, accountability, transparency, data protection and data/workflow transfer.
- Effective transition to a P4 Medicine and Health approach, with the citizen / patient more and more in a driving seat with respect to their personal data, and as partners in research

Organisational/funding challenges

- Establishment of an effective organisation open community-based infrastructure that covers a broad range of stakeholders (organisations, communities and disciplines)
- Establishments of a certification role for the infrastructure that secure quality of feeding resources and easy and harmonised access to the broad range of resources that are integrated in *Health-RI.*
- Sustainability and business model for the infrastructure and its resources

2. Embedding

2.1 Health-RI in an (international) perspective

Health-RI is the interfacing infrastructure combining local high-quality research facilities and resources built up over time, and already assembled in international infrastructure initiatives. *Health-RI* will not replace these components, but integrate them in a collective networked research infrastructure that supports next generation biomedical and technology research, and provides a base for clinics and private enterprises to build value added services and applications in medicine & health.

Health-RI will involve the Dutch nodes in all relevant *international* research infrastructures.

√ BBMRI / BBMRI-NL (Biomedical collections and population/cohort studies):
√ ELIXIR / DTL/ELIXIR-NL /BioSB (multi-omics facilities, bioinformatics and life science data/ICT resources and international reference data collections and standards)
√ Eurobioimaging - EATRIS/EATRIS-NL (medical imaging, neuroimaging) – NL-BioImaging-AM advanced microscopy) – Emerging frameworks within the neuroimaging and electron microscopy communities.
√ INSTRUCT / Netherlands Proteomics Centre, Netherlands Proteomics Platform (structural biology/proteomics),
√ Netherlands Metabolomics Centre (metabolite analyses)
√ ECRIN (clinical trials)
√ PRACE – EGI – EUDAT – Research Data Alliance / SURF – RDNL (e-infrastructure facilities)

In addition, numerous other national stakeholders will be involved. Physically, the linked-data infrastructure of *Health-RI* offering distributed analytics is modelled after the European Open Science Cloud[31].

[31] See: https://ec.europa.eu/digital-agenda/en/news/european-science-cloud-horizon-horizon-magazine

2.2 Relation to existing infrastructures

As stated above *Health-RI* builds on a wide range of existing cross-institutional and international biomedical facilities and resources assembled in the Netherlands, often Dutch nodes in international infrastructure frameworks (ESFRI/e-infra). These initiatives form essential building blocks for *Health-RI.*

The *Health-RI* infrastructure is non-discriminate to the type of partners that can hook up. All local resources will feed into the collective infrastructure based upon dedicated *rules of engagement*, covering aspects such as adhering to the *Health-RI* stakeholder governance model, adopting generic accessibility to the local facilities or resource, quality and transparency of operations, data access, legal and ethical standards, etc., see Chapter 3.

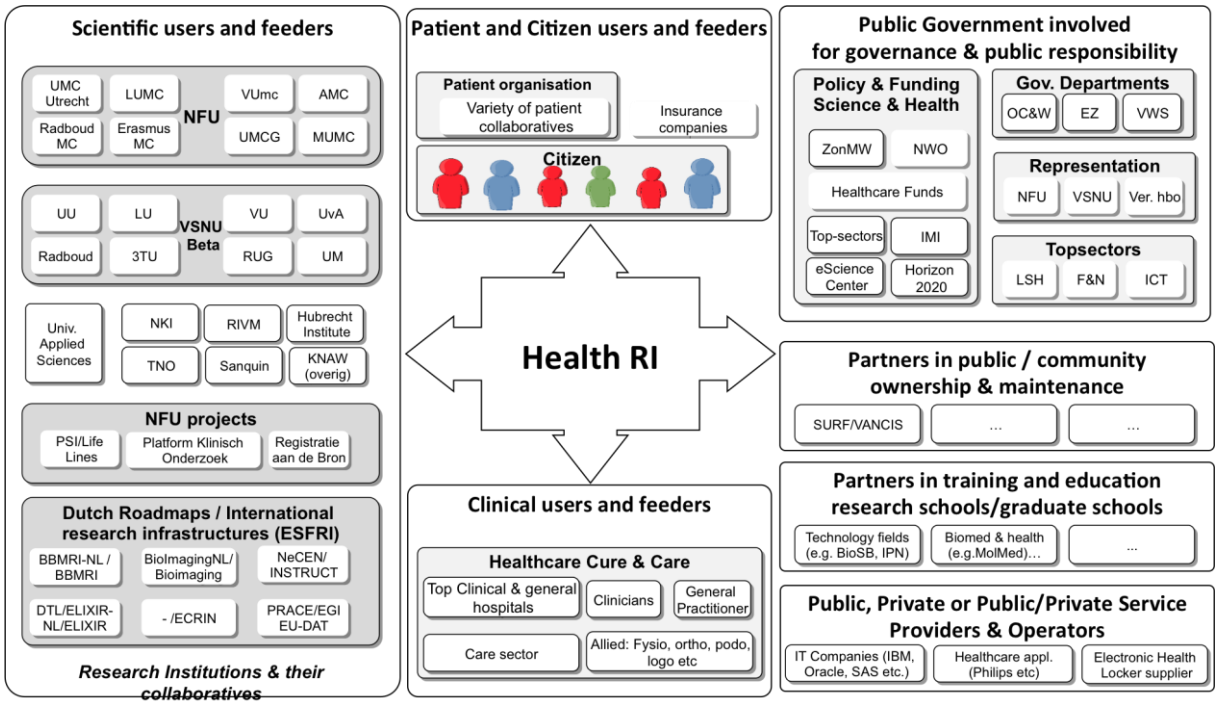


Figure 4 Health-RI links existing research and health organisations and their current infrastructure initiatives

2.3. National access.

Health-RI is foreseen as a publicly driven infrastructure that connects a broad network of diverse stakeholders, including industry. User access will be harmonised across the infrastructure, but depending upon aspects of running costs for individual services/resources. These aspects will be included in the rules of engagement set by the governing board of the infrastructure.

Guaranteeing access

The resources in the *Health-RI* infrastructure are in principle openly accessible as public services to all research projects, following strict user guidelines and regulation to access resources. Primary access for 'external' users (public or private) to research facilities and other capacity-demanding resources will always be conditional to an evaluation of scientific, ethical and/or technical feasibility. This may already be (partly) covered through review of project proposals organised externally or as part of the infrastructure access process.

For the sake of the sustainability of the resources, we foresee a necessary user contribution. This may have various forms, and could take shape through scientific collaboration, through simple citation of use (incl. data citation) in publications, with a reasonable access fee or otherwise, depending on the type of project or type of user(public/private/national/international). The nature of this contribution will be worked out in the detailed design-phase of the infrastructure. The foreseen nation-wide character of the infrastructure includes the option to agree with science/project funders a standard access fee in project proposals that pass their review process. This would greatly stimulate the use and sustainability of the infrastructure and its contribution to the

overall goals of national science and innovation programmes.

2.4 *Health-RI* and NL research strengths?

The Netherlands is very well positioned to play an internationally leading role in design and construction of the proposed integrated nation-wide infrastructure for personalised medicine and health research.

Compared to most countries in the world, we have a very well organised science and innovation field that has already established many cross-connections at levels relevant for this infrastructure: life science fields, public or public-private, biobanking initiatives, technology and e-health initiatives and data/ICT initiatives (see above). Development of the proposed infrastructure would therefore capitalise on hundreds of millions of science and innovation investments made in the last decade as it bundles infrastructure efforts across sectors of programmes that have all received excellent reviews: e.g. FES funded programmes that feed into Life Science & Health, Agri&Food and ICT: CTMM, TI-Pharma, genomics and technology initiatives developed under the Netherlands Genomics Initiative, COMMIT, TI-Food & Nutrition, the cross-UMC programmes data4lifescience and the Dutch nodes of emerging European infrastructures (see above).

Notably, the recent preparations towards a European Open Science Cloud infrastructure strongly converge towards the *Health-RI* model of a linked-FAIR-data backbone allowing high performance distributed analytics. The Netherlands is in a unique position to showcase such an approach, as a tangible way to implement Open Science as strongly propagated by our government.

BBMRI-NL, EATRIS-NL and ELIXIR-NL/DTL have reached a strong level of consensus on how to work together towards realising an integrated research infrastructure for personalised medicine and health research building upon their respective international strengths. The collective university medical centres, as united in the NFU, have meanwhile prioritised the proposed research infrastructure.

2.5 Advantages of *Health-RI* as national infrastructure.

Large-scale research facilities are important for scientific progress and for the positioning of strengths in the Netherlands, as already highlighted in the science case of this proposal. *Health-RI* will offer access to a globally unique networked infrastructure of interconnected resources, thereby stimulating frontier biomedical and technological development, and training the next generation of experts in biomedical and computer science. Already in the building phase, the infrastructure will be able to contribute to the reproducibility of Dutch personal medicine & health research by addressing key issues that have been show to lead to loss of reproducibility: better experimental design, high-quality measurements and excellent data handling/data stewardship.

The realisation of *Health-RI* will have major societal and economic advantages:

1. Societal advantages: the infrastructure will greatly facilitate the participation of citizens/patients in frontier P4 Medicine & health research, and they will benefit directly or indirectly through their cooperatives from the output of this research.
2. Economic advantages: The infrastructure will contribute to lower costs and higher quality in the Dutch healthcare and biomedical research system. Technology programmes will manifold closely involve industry (large industry, SME's, start-ups) in the development and application of advanced biomedical technologies. This will greatly stimulate Dutch business development in a huge global market.

2.6 Existing comparable infrastructures

The Netherlands is well positioned to start development of this infrastructure. Huge initiatives have recently been launched in the US and in China, Australia, and Scandinavia. It is most likely that these initiatives will converge into trans-national and even global networks, such as already happens at the level of European ESFRIs and e-infrastructures.

None will have the same progressive approach level of integration as the infrastructure here presented.

3. Organisation & Finances

3.1 Organisation.

As explained in Chapter 2 and depicted in Figure 4 above, a broad field of stakeholders and initiatives is already involved in the development of the vision assembled in this proposition, all offering important building blocks, expertise and experience for its realisation, in crucial areas of biomedical and ICT technology, biomedical, clinical and ICT research, building and operating research infrastructures and/or in aspects of professional data stewardship. These stakeholders and many more will be involved in the realisation *Health-RI* as the Dutch national infrastructure for personalized medicine and health research.

Key initiators are BBMRI-NL, EATRIS-NL and ELIXIR-NL/DTL, who have reached a strong level of consensus on how to work together towards realising an integrated research infrastructure for personalised medicine and health research building upon their respective experiences in organising cross-disciplinary infrastructures in adjacent fields. The collective university medical centres, united in the NFU, have meanwhile prioritised the proposed research infrastructure. DTL involves the great majority of Dutch academic and e-infrastructure organisations and the DTL partners[32] and DTL board[33] strongly support the construction of the proposed infrastructure. The model of a networked but strongly connected research infrastructure for personalised medicine and health with a connecting linked-data backbone may well be implemented in other sectors and science fields as well.

Health-RI should develop as a public-driven infrastructure supporting Dutch science and innovation. Like railroads, water and energy supply we foresee the infrastructure to be a public good, governed under an authority that answers to the public. This does not preclude that private or public/private parties may participate in the design, construction, validation and operational phase of the infrastructure. In fact, the infrastructure could create a open innovation environment and marketplace for such parties to develop and offer services within the infrastructure. All contributing parties will have to adhere to the rules of engagement set by the governing board of the infrastructure partnership. These rules of engagement are foreseen as a set of conditions, rules of behaviour, maybe even laws and other principles that will govern the way parties may participate in *Health-RI.*

3.2 Possible organisation structure

3.2.1 Community-based organisation

Given the public nature of the infrastructure and the broad field of disciplines and stakeholders involved, we anticipate an organisation and governance structure for *Health-RI* that can support the creation of a 'commons' environment. This structure will need to secure transparency and inclusiveness across scientific disciplines,

[32] See: http://www.dtls.nl/community/partners/
[33] See: http://www.dtls.nl/about/governance/board/

institutions, citizen collectives, industries and government. This is crucial to secure public accountability and the sustained involvement of all essential stakeholders as contributors or as users of the infrastructure.

The following elements will need to be taken care of in the organisational architecture of *Health-RI*:

√ The infrastructure will need an open and flexible partnership of contributors, inclusive to all organisations relevant to field of personalised medicine and health research and subscribing to the rules of engagement for contribution to this public infrastructure

√ The rules of engagement are a set of conditions, rules of behaviour, maybe even laws and other principles that govern the way parties act and interact on *Health-RI*. These rules synthesise the key scientific, economic, ethical, moral and legal guiding principles of *Health-RI* that are community endorsed and can be enforced to all subscribing parties

√ The governance will need to secure a fair and effective representation of the major stakeholder organisations (contributors and users) as a ring of authorities that collectively report to the public and hold responsibility for the development and public accountability of the *Health-RI* infrastructure.

The following bodies are foreseen:

√ Governance
- Board responsible for the overall strategy, output, trust and public accountability of the infrastructure, representing all major stakeholder groups (academia, clinics, public health sector, public ICT infrastructure, topsector TKI / Boards, industry collectives, science policy & funding partners, governmental departments
- Executive Committee responsible for the design, prioritisation and planning of the development and operations of the common infrastructure, with representation of all major nation-wide infrastructural initiatives that establish the *Health-RI* platform
- Operational Teams responsible for design, construction and operation of the essential components of the infrastructure (see Fig. 2) and their interconnection, as part of international frameworks (e.g. ESFRI), involving technical, engineering and service experts in relevant disciplines
- Contributor Committee(s), with a representation of individual organisations contributing resources to the collective infrastructure and advising the Executive Committee and Operational Teams in the scientific and technical integration of resources in *Health-RI*

√ Advisory committees
- Scientific and technical advisory committee(s), with independent (national and international) representation of the major scientific, translational and technical disciplines involved in the international field of personalised medicine & health research
- Clinical & healthcare advisory committee(s), with representation of major clinical and healthcare disciplines
- ELSA advisory committee(s), with experts from the fields of ethics, privacy, law, technology assessment, social sciences
- Education advisory committee(s), representing educational initiatives in the *Health-RI* framework
- Funders-advisory committee(s), representing science, innovation and infrastructure funders
- Industrial advisory committee(s), with representation of the major industrial
- User advisory committee(s), with representation of the major user groups
- Citizen science committee(s), with representatives of citizen/patient cooperatives

3.3 Costs and Funding.

Below, we sketch the phasing of the realisation of the *Health-RI* infrastructure in the next decade. A detailed roadmap and prioritisation of *Health-RI* components will be worked out in the coming year with the field of stakeholders at level of sectors, NL infrastructures, research institutions, clinicians & healthcare providers, citizen/patient organisations, industry, and government/funders.

3.3.1 Cost for development phase 1

Phase time span: 2 year (2016 – 2017)

1	Conceptual design, cross sectoral commitment	*1.000.000*
2	Integration and consolidation of BBMRI-NL, ELIXIR-NL/DTL, EATRIS-NL and (part of) NL-Bioimaging-AM and SURF/DANS (biobanks, facilities, e-infra);	*15.000.000*
3	1st development phase of linked-data ICT backbone (10^2 connections);	*12.000.000*
4	Implementation strategy en programming. Prep. LS-tech & ICT R&D programmes	*1.000.000*
5	Build organisation, governance, coordination and communication	*1.000.000*
	Total first phase (€)	*30.000.000*

3.3.2 Cost for development phase 2 and first construction phase

Phase time span: 5 years (2017/18 – 2022/23)

1	Full development of existing resources (biobanks, facilities, e-infra)	*50.000.000*
2	Integration of other infrastructures, incl. e-Health	*20.000.000*
3	Run 1st LS-tech and ICT R&D programmes	*40.000.000*
4	2nd development phase of linked-data ICT backbone (scale up to 10^4 connections), incl. data hosting	*50.000.000*
5	Launch workflow/tool development programme	*10.000.000*
6	Launch training & education programme	*2.500.000*
7	Develop business & sustainability model	*2.500.000*
8	Organisation	*5.000.000*
	Total second phase (€)	*180.000.000*

3.3.3 Cost for construction phase 2 and first exploitation phase

Phase time span: 2-3 years (2022/23 - 2025)

1	Exploitation of all major infrastructures included in *Health-RI*	*30.000.000*
2	Run 2nd LS-tech and ICT R&D programmes	*20.000.000*
3	3rd development phase of linked-data ICT backbone (full scale up to 10^8 connections), incl. data hosting	*50.000.000*
4	Run workflow/tool programme fully translated in operational service layer	*10.000.000*
5	Run full *Health-RI* training & education programme	*4.000.000*
6	Implement business and sustainability model, generating fist revenue stream	*3.000.000*
7	Organisation; fully embed *Health-RI* as public good in open science and public health sectors	*3.000.000*
	Total third phase (€)	*120.000.000*

3.3.4 Exploitation

Total estimated annual cost for exploitation of *Health-RI* is €50.000.000, which is 5% of the expected annual Dutch budget for science & innovation in the health sector. These are costs for maintenance, innovation and use of the infrastructure.

The business model of the infrastructure aims to make the infrastructure sustainable after 2025, not only

covered from public science & innovation budgets, but also from citizen participation, health funds, industry and other (international) users of the infrastructure.

3.3.5 Business case rationale

Next to the scientific and health-care output advantages as described extensively in previous parts of this proposal, the *economic* rationale for *Health-RI* is rooted in 1] revenue-generation, 2] innovation advantages and 3] reducing loss of costs in irreproducible scientific output.

Ad 1. Revenue generation

As described above under 2.3 (Access to *Health-RI*), we foresee a user contribution by scientists from academia and industry and by other users of the infrastructure. This would greatly support the financial sustainability of the offered resources and ascertain a user-driven development of the infrastructure. A reasonable access fee is foreseen, especially for capacity-intensive services and for training, and this may vary for particular types of users (public/private/national/international). Such access fees may well be covered as a budget item in research grants funded though NWO. Alternative models, such as a basic subscription model and/or based upon exploitation of IP generated in the infrastructure may add to a revenue stream partially covering cost of the infrastructure. It is too early to predict the level of revenues that could be generated through this approach, but the income could be substantial.

Ad 2. The entrepreneurial state – stimulating innovation and industry development

Health-RI is a public-driven infrastructure, developed as a public research resource for open science and public health. Apart from the direct involvement of industry in the development and construction phases of the infrastructure, it will generate frontier technological concepts ready to be picked up in dedicated innovation pipelines of large and small industry and in clinical development.

In close relationship with several topsectors (LSH, Agri&Food, ICT, HTSM), it can be expected that new products, business concepts and industries will develop based on the frontier technological concepts and the huge market for innovative personalised medicine & health as well as generic ICT service propositions. The latter relates to a globally front-running position in linked-data analytics and e-infrastructure technologies. This is clearly an area that is more generic than biomedical science, and opens up great chances if the Netherlands acts as an entrepreneurial state[34].

Ad 3. Reducing cost of scientific output loss, increasing reproducibility and re-use of data

Research in the US shows that 50% of pre-clinical research is not reproducible [35]**.** About 25% of public and privately funded research in this field does not lead to solid output due to meagre or untraceable experimental design and bad data handling. In the study presented, this accounts for a loss of about 15 billion (!) US$, see Figure 5 below. Additionally, a preliminary study shows that 88% of US biomedical research datasets is not deposited in a well-curated and accessible public data repository, making the problem of improving the discoverability and re-use of biomedical datasets significant [36].

[34] The Entrepreneurial State concept stipulates the value of public driven (fundamental) science and development as the source of private industry innovations and development of industrial sectors. http://marianamazzucato.com/the-entrepreneurial-state/

[35] Freedman LP, Cockburn IM, Simcoe TS (2015) The Economics of Reproducibility in Preclinical Research. PLoS Biol 13(6): e1002165. doi:10.1371/journal.pbio.1002165, june 2015

[36] Read, Kevin B. et al. "Sizing the Problem of Improving Discovery and Access to NIH-Funded Data: A Preliminary Study." Ed. Vincent Larivière. *PLoS ONE* 10.7 (2015): e0132735. *PMC*. Web. 5 Jan. 2016.

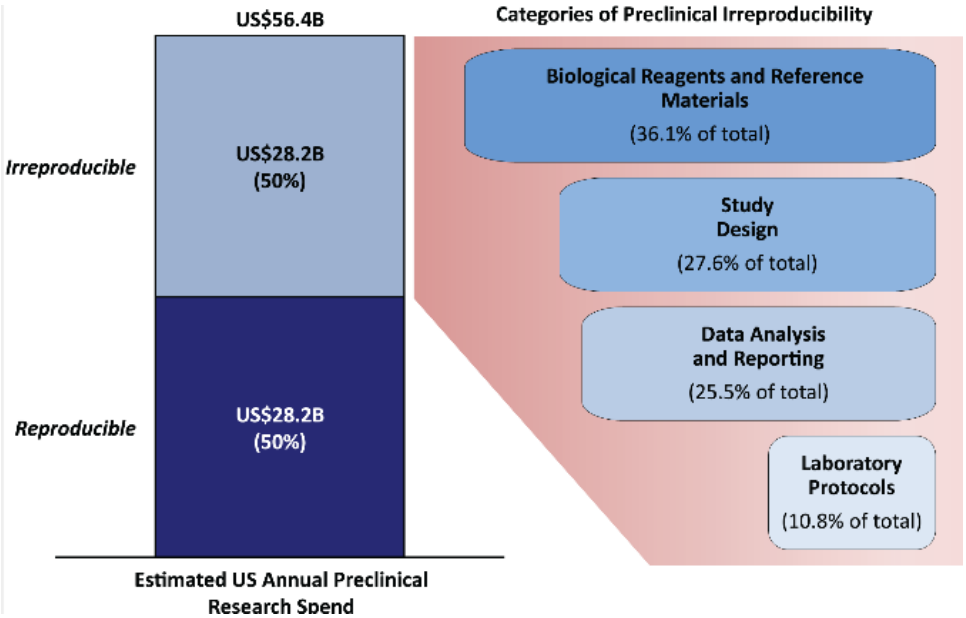


Figure 5: Irreproducibility of research and contributing factors

There is no reason to believe that this situation is much better in other fields of life science, and indeed in many other fields of science, nor that the US biomedical science would be much less reproducible than Dutch research. Indeed, fragmentation of initiatives and loss of methodology and data leads to a significant loss of effective budget spending in life science research. The cost burden of this unnecessary loss in our current scientific system is huge and mostly hidden.

The *Heath-RI* approach addresses exactly these components. It will make biomedical research much more reproducible by focussing on high quality experimental design and execution, and on effective re-use of research data. With an estimated 1 billion of annual publicly funded research projects in the biomedical life sciences [37], the current cost burden of irreproducible Dutch research output would amount 250 M€. Even if *Health-RI* would only improve scientific reproducibility and data handling by 20%, the annual projected *Health-RI* exploitation costs of 50 M€ would be covered.

[37] Number per annum, based on the figures of IBO report 2012. Research project spendings from 2nd, 3rd and 4th 'geldstroom' appr. €950mio. http://www.rijksbegroting.nl/system/files/12/2012ibouniversitairmedischecentrarapport_0.pdf

4 Steps towards development of Health-RI.

4.1 Roadmap.

The programme line in development of the *Health-RI* infrastructure has been sketched above. To bring this ambition to a broad consensus on a collective roadmap to set up the organisation and start development, it will be important to use the great synergy already present in the field and receive active KNAW support to reach further alignment of the broader field of Dutch stakeholders around the *Health-RI* proposition.

Figure 6 below provides a possible roadmap strategy to develop the infrastructure step by step, striving for a mixed model of project and infrastructure funds with overlapping timelines and regulations.

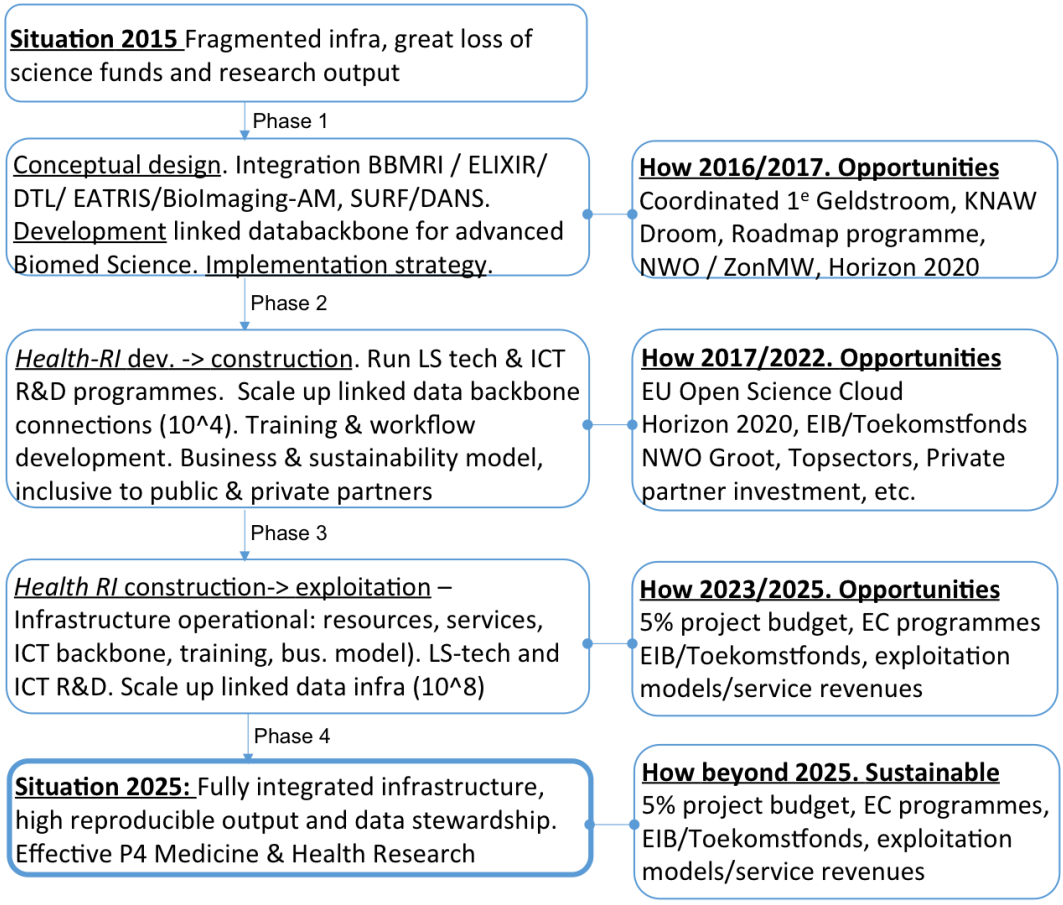


Figure 6: Possible funding roadmap strategy to realise the Health-RI infrastructure

Annex 1: to the Health-RI proposal

Topsectors Knowledge and Innovation Agenda's 2016-2019

KIA Life Sciences and Health (LSH / 'Health~Holland')

LSH identifies three key challenges in life science related to our increasing ageing population, non-communicable (Chronic) diseases and resulting disability burden:

1. ***Maintain health and functioning, focus on prevention (including personalized primary prevention and prognostic pathways)*** – assist people to remain vital and functioning without medical needs, or supportive care.
2. ***Maximise effect, minimise burden*** – if disease or disability occurs, support people to maintain and/or regain vitality and functioning as much and as fast as possible.
3. ***Manage health and disease outside the hospital*** – if disease and/or disability persists enable people to adapt, self manage and function at their best in their living environment.

Although this LSH agenda addresses the coming four years, it may be expected that elements of the proposed nation-wide infrastructure for personalised medicine and health research will already assist in the scientific breakthroughs necessary to realise this agenda. In addition, *Health-RI* components can help in aligning industry and academia of this sector around the construction of a collective infrastructure of growing value, that guides high quality experimental design and execution and assists in the data stewardship aspects of LSH programmes. The connection with LSH stakeholders will also assure the optimal alignment with initiatives in care, home-care and e-health.

KIA Agri&Food 2016 – 2019

Roadmap Agro and Food	Link Health-RI
Topic: Nutrition and Health **Description.** Nowadays, consumers are able to measure the effects of their nutrition on health and they can use this to make reasoned choices regarding their diet. This can contribute significantly to vitality and well-being, healthy life and reduce healthcare costs. There is a sub-optimal health because of our modern lifestyle, our eating behaviour and demographic changes which is expressed by, among other things, reduced well-being and increase in chronic, diet-related diseases such as obesity, cardiovascular disorders, effects on mental functioning and diabetes. This puts great pressure on public health care costs, economic growth and other social factors. Research into the relationship between nutrition and should focus more on being able to determine the specific impact of food on individuals and groups and less on the average impact on groups of people. In the future, this should lead to a simple and accessible system maximize their personal dietary choices.	In 2025 a participant will have its own personalised diet, because it is generally known that standardised diets can achieve opposite effects for some people. Individuals do not only respond different to diseases and medication due to i.e. genetic elements and lifestyle, but they can also respond different to identical foods due to differences in their microbiomes. A personalised diet, in combination with a healthy lifestyle can bring individuals closer to their potential health curve.
Topic: Consumer and Chain **Description.** Understanding consumer behaviour helps to achieve a more sustainable food production, consumption and a healthier diet. The market demands to provide high quality and safe products. Citizens want that these products are created in a sustainable manner. Interventions aimed at sustainable and healthy food production and consumption should reflect as much as possible in habits, motivations and desires of consumers, eg by making advice personally. That makes more likely that eating habits and food procurement will actually change. Personal advice can reduce healthcare costs and improve quality of life and wellbeing.	Research into the relationship between eating habits, consumer manners, personal advice on nutrition and the effect on the potential health curve.

KIA ICT 2016 - 2019

The ICT sector creates crossovers with many sectors, including LSH and Agri&Food. Focussing on the relevance for personalised medicine and health the sector has clearly identified important innovation topics relevant to the creation of the *Health-RI* infrastructure. The ICT KIA 2016-2019 (termed 'Dutch Digital Delta') has selected Healthcare Systems and Services (e-health) and Life Sciences and Health as major fields for ICT innovation and it supports the need for better data stewardship and data interoperability based upon the FAIR principles, as developed with in the life sciences. It is clear that the proposed ICT infrastructure that will serve as the linked-data backbone among the *Health-RI* resources will require a strong involvement of the ICT research community. In this respect, it is a very positive sign that the topic of personalised medicine and health is one of the pillars in the whitepaper COMMIT2DATA4LIFE, sketching the contours of a cross-sector ICT research programme in the Netherlands. Especially topics such as distributed learning across local data resources, high-performance and privacy-preserving analytics of (personal) health data and digital security are of paramount importance to the proposed nation-wide integrated infrastructure for personalised health research and care. These topics are in full alignment with the above-sketched computer science case.

THE NATIONAL SCIENCE AGENDA LINKS TO HEALTH-RI VISION ON SEVERAL OF ITS "ROUTES"

Route number	Agenda	Link Health-RI
081. How will genetics play a role in understanding, screening and the treatment of rare diseases?	Many diseases have a genetic component. Often multiple genes play a role. Finding the exact cause of the disease, and identifying possible treatments is very complex. In addition to genetics, environmental factors may also determine whether a disease is expressed.	Infrastructure makes it easier to gather different research outcomes. The availability of information about different aspects of the human life (genetics, environment, daily activities) will create a more comprehensive understanding of a patients journey and situation over time.
095. How can the health sector, be more focused on the uniqueness of a person?	The effectiveness of drugs and therapies is not always as desired. The importance of a person-centered approach to health care is widely shared and manifests itself increasingly in both policy and research: from average patient to personalized medicine.	Personalised Health. Health-RI will create the possibility for scientists to find ways to target diseases on a tailor made way for each individual
098. How can we make breakthroughs in basic medical examinations and better translate them to the development of new drugs?	Upcoming breakthroughs in biomedical research & drug development will greatly affect us in a positive way: - Next generation DNA sequencing to get information about patients can be gathered more quickly; - Developments in advanced therapies, such as gene therapy and immunotherapy, stem cells and organoids; - Better knowledge of biomarkers; - Progress in the field of ICT makes learning faster and better. A key challenge for the coming years is to ensure rapid patient access to innovative therapies at socially acceptable costs.	No more unnecessary repeating of research-> lower costs. Availability of different data resources can support medical breakthroughs.
102. How can we develop new drugs and modes in order to stay vital and healthy as possible?	It is possible to deal with age-related diseases as a group rather than as individual diseases. Early diagnosis is essential. In addition, new drugs and treatments are needed to make the population healthy and age actively. Molecular understanding of disease is at the heart of new treatments. The future lies in translational research into human beings and their environment. The interaction between heredity, environment and lifestyle of the individual determines health.	Translational research-> access to enormous amounts of data, makes it possible to examine complex relationships between human beings and environment. Minimize the health gap and maintain optimum health.
105. How will Big	The big data revolution also provides the ability to accelerate	*Health-RI* offers the mechanism to link

Data and technological innovation play a role in health?	medical research by combining information from all the possible data files with each other. A major challenge is to make these data in patient care and research FAIR: Findable, Accessible, Interoperable and Reusable. This requires issues such as standardization, ease of use, durable storage, property and privacy are well organized throughout the health sector. With these developments, it is necessary to develop new concepts, methods and software to analyse and interpret the multitude of health information and knowledge well, and can translate into personal health advice.	crucial research resources that that are highly scattered today and makes them more easily findable, accessible, interoperable and reusable (FAIR) through collective data portals: clinical biobanks and population cohorts, high-end lab facilities and sensory networks (eHealth), local, national and international scientific reference data resources, as well as high-capacity compute & storage facilities
135. How can we have a better understanding of the features, functionalities and the interaction of molecules in living systems?	We still do not understand how interactions between molecules can lead to living organisms. However, we do know that the molecules of life recognize each other, react with each other to new molecules, enter into interaction with each other and together forming structures by means of many complex processes. Knowledge of such complex molecular systems and the ability to steer it in a desired direction, offer clues to address social issues in the fields of environment, energy and health.	*Health-RI* will offer the possibility to get knowledge of i.e. Complex molecular processes due to the availability of different research resources.