# Responsible Data Management for Personalised Diagnostics

## Report of a series of expert interviews

*Commissioned by the Netherlands Genomics Initiative embedding project ReDaPeD (Towards Responsible Data Management for Personalised Diagnostics)*

**Project Leader:**
Dr Maud Radstake, Centre for Society and the Life Sciences

**In cooperation with:**
Dr Terry Vrijenhoek, Centre for Genome Diagnostics

**Report by:**
Dr Daan Schuurbiers, De Proeffabriek

CSG CENTRE FOR SOCIETY AND THE LIFE SCIENCES

cgd Centre for Genome Diagnostics

De Proeffabriek

**Contents**

## Executive summary

This report summarises the findings of a series of scoping interviews on responsible data management for personalised diagnostics.[1] The scoping interviews mapped respondents' views on the salient issues in storage, access and use of 'big data' as generated by new life science technologies for health care purposes and vice versa. Thirteen experts from a wide range of disciplines were interviewed in depth to map out the opportunities and challenges for data management from their perspective (being (gen)omics, bioinformatics, ICT, clinical research, ethics and social science).[2]

Although some were more precautionary than others, respondents all agreed that the generation, storage, analysis and integration of hitherto fragmented datasets (such as clinical data, experimental data from genomics, proteomics, metabolomics, imaging data and metadata) holds significant promise for the future of both medical research and health care, allowing for advances in diagnostics, screening, prevention and treatment.[3] Data-driven medicine was expected to present disruptive change in several ways: the crossover of medical research data into the domain of health care blurs traditional boundaries between research and care. As a result, it challenges entrenched views on data protection and ownership and the role of patients and users in both research and health care. In the research domain, strong incentives for data sharing as opposed to data protection will affect academic reward structures, intellectual property regimes and privacy regulations. As for health care, blurring boundaries between research and care challenge traditional informed consent procedures and restructure divisions of labour and responsibility among medical experts and patients.

The use of big data in medical contexts was thus seen to present challenges of an ethical, legal and social nature in addition to the 'technological' bottlenecks that still need to be overcome for –omics technologies to become standard practice in the clinical context such as the cost- and time-efficiency of data generation and analysis, sensitivity and specificity of diagnostic tools, and questions of data stewardship and integration.[4] There was wide agreement among respondents that appropriate answers to the broader ethical, legal and social questions are a prerequisite for the societal acceptance of life science technologies in health care.

To address these questions, respondents recommended professionalising data stewardship and integration; revisiting informed consent procedures; assessing public perceptions; enhancing user involvement; and strengthening education and training. Respondents noted that involvement of a broad range of technology *users* (doctors, patients and citizens) will be crucial

[1] These scoping interviews are part of the Netherlands Genomics Initiative embedding project ReDaPeD (Towards Responsible Data Management for Personalised Diagnostics) carried out by the Centre for Society and the Life Sciences in cooperation with the Centre for Genome Diagnostics and De Proeffabriek.

[2] For more details please see the interview guide and list of respondents in the appendix.

[3] Coupled with new opportunities for improved prevention and screening made possible by advances in IT-capabilities (such as electronic health records and mobile health monitoring devices) data integration brings a vision of 'personalised medicine' into view, whereby health care moves from the treatment of disease towards managing one's personal health.

[4] The distinction between 'technological', 'ethical', 'legal' and 'social' is to some extent artificial: all 'technological' decisions harbour moral considerations; 'ethical' considerations eventually solidify in 'technical' decisions; and the 'ethical' 'legal' and 'social' domains are contiguous layers rather than separate realms. For the purposes of this report however, the distinction is relevant: the interviews highlighted different elements for the various topics. The notion of data stewardship for instance integrates questions of a predominantly 'technological' nature with those belonging to the domains of 'ethics' and 'law'.

to public acceptability – although views on the precise form of such involvement varied considerably. Also, concrete examples of actual patient or citizen involvement in pilots or in data integration initiatives were relatively scarce.

In summary, the findings of these scoping interviews suggest that the disruptive potential of big data for both research and health care practices calls for broad user involvement. The technical, legal and moral frameworks for *responsible* data management therefore deserve to be defined proactively in a transdisciplinary context involving researchers from various disciplines, ICT professionals, medical experts, patients and citizens.

# 1. Introduction

The advent of 'big data' in the life sciences is said to hold great promise for the future of both medical research and health care. IT-infrastructures for research are currently being built to store, analyse and share the tsunami of data generated by new -omics technologies and to integrate them with clinical and imaging data. Whereas data integration in the life sciences unveils exciting opportunities for diagnostics, screening, prevention and treatment, it also invites new perspectives on data sharing and data protection. For instance, new standards for sharing research data may clash with traditional ethical standards in clinical practice. A fundamental question is therefore: what happens when open science meets personalised health care? What are the social, ethical and legal boundaries for the use of medical data for research purposes, as fundamental genomics research becomes clinically relevant and clinical data increasingly becomes the object of research?

## 1.1 The ReDaPeD project

The question how to responsibly develop data infrastructures given the tensions identified above is central to the ReDaPeD project (Towards Responsible Data-management for Personalised Diagnostics). ReDaPeD is funded by the Netherlands Genomics Initiative (NGI) and carried out by CSG Centre for Society and the Life Sciences in cooperation with the Centre for Genome Diagnostics and De Proeffabriek. The main aim of the project is to mobilise an interdisciplinary network of partners for future collaboration on responsible data management in personalised diagnostics. Importantly, ReDaPeD aims to identify opportunities for interdisciplinary collaborations and public engagement during the early stages of technological developments.

In addition to the scoping interviews presented here, the ReDaPeD project includes a literature review and a series of workshops held in 2012-2013 with experts from various disciplines. The overall outcomes will lead to a position paper that will be presented during a final workshop late 2013 and will be at the core of future applications for funding of collaborative research.

## 1.2 Scoping interviews

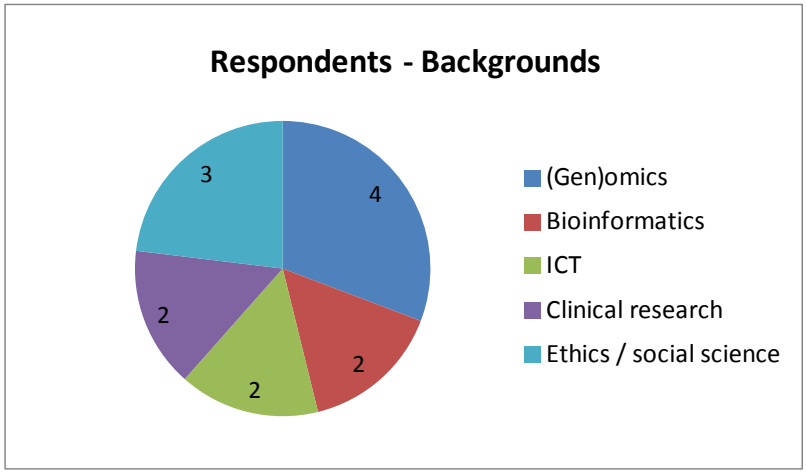
A series of scoping interviews was held to map the technical, ethical, legal and social challenges to the integration of -omics research for personalised healthcare, and to identify opportunities for interdisciplinary approaches and early engagement. While there is a growing body of literature on various elements of that discussion, e.g. legal aspects of informed consent, ethical aspects of next-generation sequencing, public perceptions of genome technologies and so on, the ReDaPeD project aims to zoom in specifically on questions of *data management*. ReDaPeD follows the data as it crosses over from research processes into medical and public domains and back. Accompanying questions of data integration and stewardship are solidifying in specific

contexts (i.e. the research context or the clinical context) but overarching questions on how decisions in one context co-evolve with workflows in other contexts remain relatively unexplored terrain. The interviews therefore aimed to span the range of experts involved in the questions concerning the flow of data: genomics researchers, bioinformaticians, clinical researchers, medical professionals, ethicists and social scientists.

The interviews assessed respondents' views on the opportunities and challenges of the application of life science technologies for medical research and health care. It subsequently went into questions of data sharing, user involvement and ethical, legal and social issues (ELSI) associated with the integration of -omics techniques in the medical domain. Since ReDaPeD aims to explore opportunities for interdisciplinary collaboration within a context of responsible research and innovation, respondents were also asked about their views on responsible innovation and interdisciplinary cooperation. Finally, respondents were asked to identify relevant organisations and initiatives in the field (for further details please see the interview guide below).

### 1.3 Respondents

13 interviews were held with researchers in (gen)omics, bioinformatics, ICT, clinical research and ethics. Given the limited number of interviews, this report does not claim to represent the consolidated views of any of these professional disciplines, or to provide an encompassing overview of all positions in the debate. It rather aims to provide a horizon scan of the various views and opinions out there, pointing towards areas of convergence and divergence across different fields of expertise and identifying overarching questions and areas of shared interest for follow-up research.



## 2. Interview findings

There was wide agreement among respondents that recent advances in the use of -omics techniques and the integration of various datasets in IT-infrastructures (specifically clinical, experimental, imaging data and metadata) could imply momentous change for both medical research and health care. Some respondents were highly enthusiastic, others were more cautionary, but on all sides *data* – and specifically questions of data integration and stewardship – were seen to play an important role:

> *Data is simply becoming more and more important, in terms of both scale and complexity, and more and more we see a desire to integrate a range of data sources – not just experimental research data but various types of information that you can relate to it to do your analysis.* [5]

### 2.1 Personalised medicine

Enthusiasts hinted at the revolutionary opportunities for 'personalised medicine' through the integration of clinical, experimental and personal data. Personalised medicine however is a container term: there were considerable differences in opinion among respondents as to what it actually meant (and, importantly, if and when it is to become a reality). Some respondents equated personalised medicine with genetic self-tests, others with home care systems or various forms of population screening. Yet another way of interpreting personalised medicine is to see it as a term that captures all forms of personalised health advice based on one's genotype plus known environmental influencing factors.

In current medical practice, the notion of personalised medicine predominantly denotes improved patient stratification:

> *What to me is the most important in practice, is that you distinguish groups of patients. That you don't have one size fits all treatment, but that you gear it specifically to the subpopulation that you are looking at. It's too sad for words that horrible chemo is used for many of types cancer while you know that it will only work for thirty percent of the people. Or less. So you have to find ways to identify that thirty percent in advance and not expose the other seventy percent. So that to me is personalised medicine. Tailor made treatment.*

Some respondents however saw stratification as just the beginning of a much larger revolution in health care systems:

[5] Quotes from the interviews have been translated from Dutch to English, staying true to the original statement as far as possible but with minor adaptations for readability.

*I think health care is moving more and more towards personal empowerment, based on qualitatively high science – with a good interface in between.*

Personalised medicine on this interpretation implies highly individualized advice in health care on the one hand, and the direct involvement of patients in medical research on the other. This suggests a view of the 'patient as partner' in the fullest sense, where patients, researchers and medical professionals co-construct new scientific knowledge.

### 2.2 Blurring distinctions between research and health care

Regardless of their views on the precise meaning of the term personalised medicine, all respondents noted that the integration of clinical and experimental data blurs traditional boundaries between medical research and health care. In fact, this was seen to be the driving force for many of the ethical, legal and social considerations identified below. One respondent summarised this recurring theme as follows:

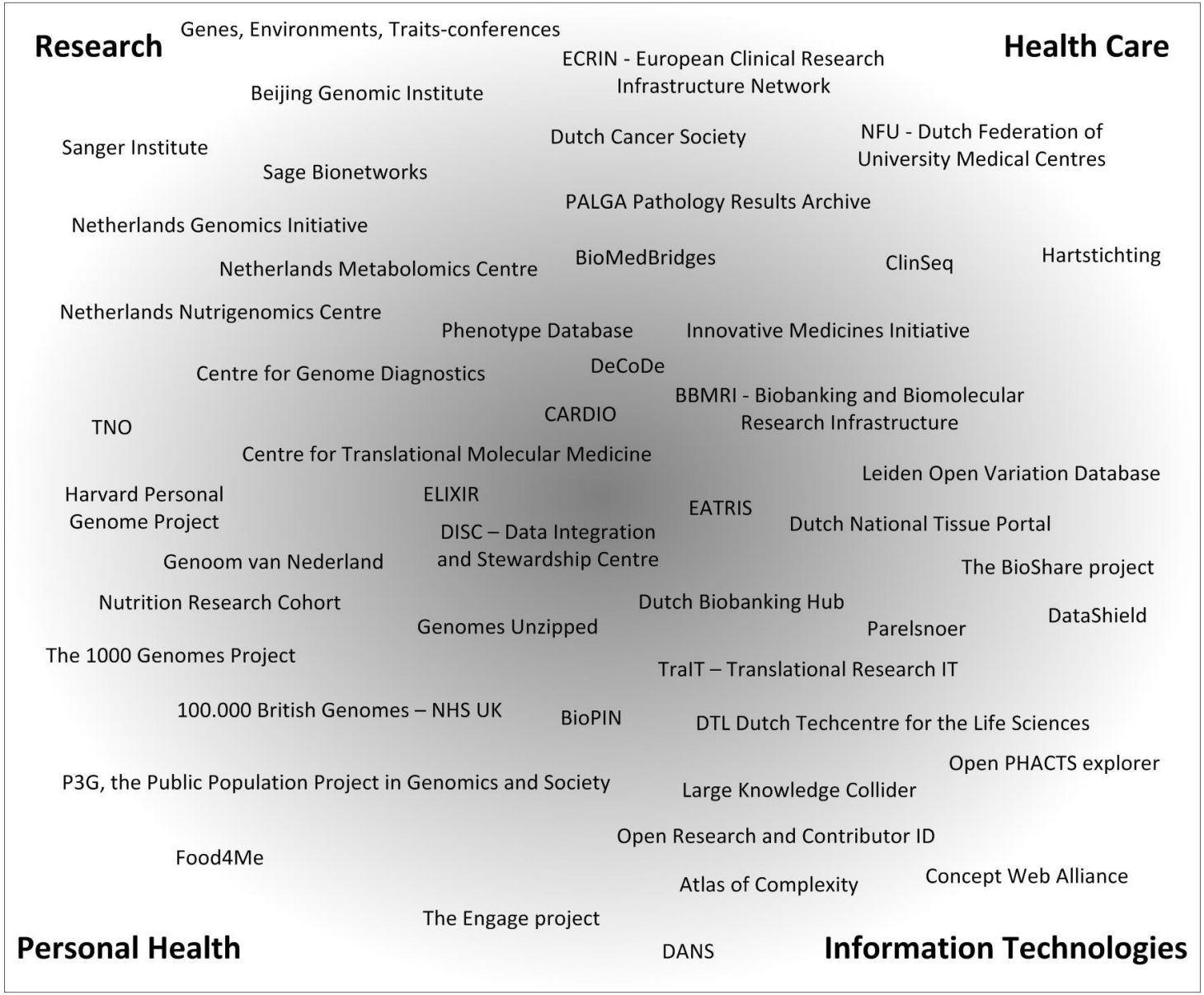
*So the main theme we are talking about for me is "to what extent can health care data be used for research purposes and under what circumstances".*

The reverse question (to what extent research data can be used for health advice) also holds:

*I see a big problem in being able to merge science and health care. I see people, they have cohorts of 11.000 people who are ill and they can't tell them how to change their lives because it will influence the research. On the other hand I see medical professionals who have patients that they can do beautiful research on, but they are not allowed to. I want to fully integrate medical research and science, so I can measure patients and make their data available anonymously, but also to return the group data, so I can de-anonimise and tell the best patients "listen, this is what's going on". That's what I'm working on. Because I can't go into a biobank and then back to the patient. I can't measure one hundred people, do research on them and go back to one hundred individuals. Well, if I want to do personal coaching, personalised medical advice, lifestyle advice, personalised medicine, then I have to be able to do that. So that needs to be the endpoint of our efforts.*

Data integration, then, was deemed to present revolutionary opportunities for medical research and health care while inviting deep questions on the nature and purpose of both. The level of expectation surrounding data integration in the life sciences is reflected by the enormous range of organisations and initiatives that respondents identified as being involved in this endeavour. The figure below indicates the organisations and initiatives that were mentioned during the interviews. Even though this figure lists a significant number of initiatives in the health, research and IT-domains, it probably merely scratches the surface of the scope of the endeavour.

## Organisations and initiatives identified by respondents



## 2.3 Technological challenges

Respondents identified several technological challenges to the realization of data-driven medicine. First of all, while genomics technologies are advancing rapidly, the cost- and time-efficiency of data generation and analysis are still bottlenecks for their widespread use in routine diagnostics, screening or monitoring. Additionally, the sensitivity and specificity of diagnostic tools and systems need to be improved. There is also a need for harmonisation and standardisation of tools, methods and systems and for quality control:

> *The four most important technical challenges are: the errors that the technology still makes; things that the technology doesn't see; the things we do see but don't understand; and how to deal with the various categories of data in a uniform way.*

In relation to the application of next-generation sequencing data for diagnostics, one respondent remarked:

> *It's still mostly a research technology. Every research group have their own machine and develop their own methods because that's what they are comfortable working with. They are all different systems, different tools, and everybody is tweaking and turning the buttons to get the data right. So there is not a standard protocol that says this is how we will do it.*

Hence, the general aim of IT support for translational research is to be able to provide a common platform:

> *The objective is to correlate phenotype-data to some kind of readout, usually something from the genomics area but also often the imaging area. So what you are doing is generating and coupling large datasets, and there is a comparable workflow across the various phenotypes that you should be able to support with a single IT-infrastructure.*

The implementation of research infrastructures is urgently needed given the veritable tsunami of data in the life sciences:

> *Every 8 months the amount of data in the life sciences has doubled. Every 8 months. This isn't sustainable. You know the story of the farmer who asked the king for a grain of wheat on the first square of a chessboard, two on the second, four on the third and so on? Just think what the next 55 years, the next 55 squares will mean in relation to the data that we have today.*

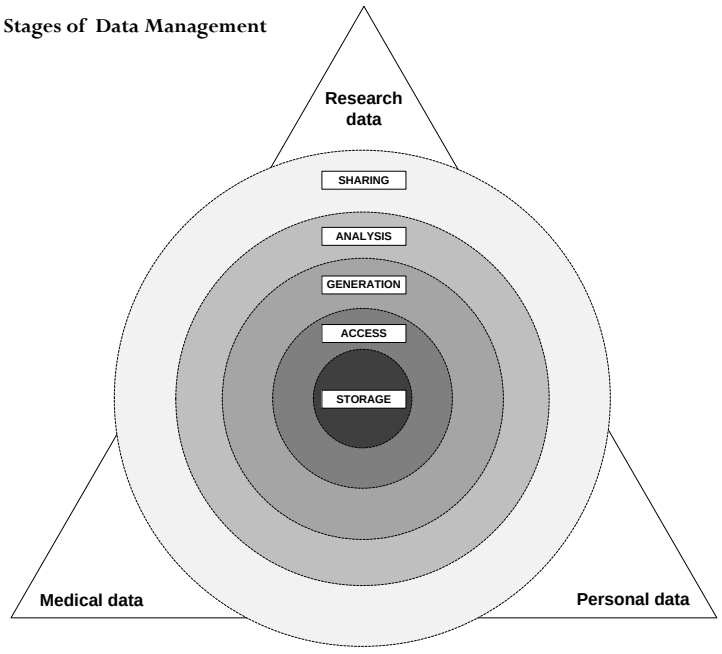
To be able to make sense of the data deluge, integration and stewardship are key:

*The amount of data in the life sciences is growing exponentially. What's more, that data is extremely fragmented, it comes from all kinds of areas in all kinds of formats. They can be protein data, or DNA data, or pattern recognition images, so the diversity is enormous. Much more diverse and complex than for example in physics. So data integration, being able to have the data communicate with each other, is of course essential. So you have to standardize, that's crucial. The quality of the data, that you can check where rubbish comes in, rubbish comes out. And the management itself, right? Storage… that if a PhD student does something and he has his own database and if he leaves it's probably lost, if you don't watch it. … So data stewardship, that's what it all boils down to. That is, the management, the quality, the processing of it all. Yeah, that is getting more and more important. A thread in the whole life sciences.*

There are, then, significant technological challenges in data management. Yet respondents agreed that the organisational and societal challenges involved may prove an even bigger barrier. The following section lists these broader challenges.

## 2.4 Ethical, legal and social challenges

Respondents called attention to ethical, legal and social challenges in relation to data management across different domains (such as research, medical, and personal data) and at various stages of data management (please see the table below for an overview). The figure directly below breaks down the concept of data management according to distinct activities identified by respondents – each with its own set of challenges. Note that it is precisely in the *integration* of various data sources that these challenges emerge.



**Technological, Ethical, Legal and Social Challenges in Data Management - As Identified by Respondents**

| *Data domain* / *Type of activity* | **Research** | **Health care** | **Personal** |
|---|---|---|---|
| **Overarching themes** | ***Blurring boundaries between research and care (E)*** | | |
| | Data Integration and Stewardship (T) | | |
| | Governance - Codes of Conduct (E / L) | | |
| | | Changing doctor-patient relationships (E) | |
| | Education (T) | | Data Literacy (S) |
| **Data generation & analysis** | ***Cost- and time-efficiency (T)*** | | |
| | Sensitivity and specificity (T) | Quality control (T) | Patient participation (S) |
| | Information overload (T) | Right to withdraw (L) | Individual interests and solidarity (S) |
| **Reporting** | ***"Personalised medicine" as Individualised feedback (S)*** | | |
| | Incidental findings (E) | Recontacting duty (L) | Duty to inform / Right not to know (L) |
| | Disclosure policy (L) | Informed consent (L) | Communicating probabilities (S) |
| **Sharing** | ***Organising Access - Data protection & Data sharing (T / E / L / S)*** | | |
| | Intellectual Property Rights (L) | Privacy (E) | Democratising health care (S) |
| | Academic credibility (E) | Access on demand (S) | 3rd party abuse (L) |
| **Storage** | ***Harmonization, Provenance, Interoperability (IT)*** | | |
| | Capacity and cost-efficiency (IT) | | Ownership (L) |
| | Public concern over electronic patient records (S) | | Anonimisation / Coding (IT) |
| | Audit trails (IT) | | |

### *2.4.1 Data sharing and protection.*

A recurring theme in the scoping interviews concerned the question of data sharing. Respondents highlighted different elements according to their specific expertise. Data integration inevitably requires that datasets are made accessible to others – data sharing however invites a host of questions of its own. Respondents agreed that data sharing opens up new and exciting possibilities:

> *What everybody realises is that a lot of non-essential information is collected in everybody's backyard. A lot of information that is non-essential for your own research can be incredibly relevant for someone else.*
>
> ---
>
> *Data might well become the new currency.*

The question of data sharing however has repercussions both for research and health care practices. At the research level, the desire to open up data often clashes with traditional intellectual property regimes. Business models of universities or pharmaceutical companies until recently relied on the *protection* of data in order to be able to patent new technologies. Additionally, traditional academic reward systems depend on the attribution of data to individual researchers:

> *There are two obstacles [to data sharing]. The first has to do with the ethical and legal aspects of the clinical context: what someone measures can have serious implications for the patient's family and surrounding, so this is a very sensitive issue. The other obstacle is simply a research issue: researchers do not want to share their data before they publish. Do not underestimate that.*

Several respondents suggested that data sharing in research requires a complete overhaul of the credit attribution system. One example is to make datasets citeable (instead of the publications in which they appear):

> *So here lies the key. If people notice that they are being referred and this raises their scientific status, they will be more inclined to share their data. Then they'll see the benefit.*

Another suggestion was to design open access systems and raise awareness of their benefit:

> *Open source and open access are often misunderstood as simply putting something in the public domain, whereas in fact it is all about setting the right restrictions, making data*

*available under certain very clear conditions. Open source has become a guarantee that things can be maintained, that they can be shared, that others can build on and expand.*

Apart from these research-related questions, data sharing also invites a range of questions in the clinical context. New data-driven diagnostics potentially disrupt traditional clinical practices. Respondents noted three main areas of attention in this domain: privacy considerations; the need to revisit traditional informed consent procedures; and shifting divisions of labour and responsibility between medical researchers, clinicians and patients.

Respondents were sensitive to a number of ethical questions that the exchange of data between research and health contexts (such as NGS in diagnostics) could bring forth.

*Of course you reach the ethical point: now I know an awful lot about this patient; what does he need to know himself, what can he know, what should he want to know? And if he wants to know, how should you present it to him? Perhaps all you are offering is a heap of insecurity. This needs to be worked out, with ethicists and lawyers, but also with doctors and with patient organisations.*

A returning consideration was that the research component of genome analysis is so large that it does not match with the concept of diagnostics in the traditional sense.

*We are measuring so much that we see a lot that we don't understand. And that is wonderful from a scientific point of view, but from the perspective of diagnostics that is paralyzing. Because when you are honest to a patient you will say: we haven't been able to find anything, but we still have a series of possible leads. And that is… I wouldn't call that diagnostics.*

### *2.4.2 Privacy*

The question of privacy often featured. Opinions were strongly divided. Some respondents felt strongly about the need to protect data to prevent third party abuse.

*I realise full well that in the end one will encounter dilemmas. Genome data are a sort of passport. We need to establish techniques that allow for scientific discovery without being able to identify the patient.*

---

*The more you share, the more valuable it is, but the greater the privacy risks simultaneously become. So it's even harder to see what you got into as a research participant, or a donor, or however you want to call it.*

Some were therefore inclined to take the precautionary approach, arguing that we should pause implementing systems for large scale storage and sharing of clinical data if we cannot oversee the possible consequences in relation to possible abuse. Others considered privacy concerns as being based on outdated conceptions of privacy. They argued that the possible benefits such as improved prevention and diagnosis of disease and more efficient health care systems outweigh the possible risks. Of course, there are many positions in between:

*There are really two sides in this discussion. There are those who say "don't worry guys, we can perfectly anonimise it all and everything will be alright", and those that say "data is never secure, there will always be abuse and we shouldn't even go there". I think these are two extremes that both put too much emphasis on one side. Because in the end it's a cost-benefit analysis, where you gain benefits from research on the one hand and run certain risks on the other – from the personal level to third party abuse – and you need to judge whether the balance doesn't tilt too much towards one end or the other and of course you need to keep your eyes open on both sides.*

### *2.4.3 Informed Consent*

Because of the systems approach to genome analysis, traditional informed consent procedures fall short:

*Informed consent requires that the patient is informed about all possible consequences of the procedure. This is impossible, because we know so much, we can't know in advance what we will find, and if you want to inform the patient you have to explain what the consequences of that might be. So the patient who comes in to be screened for a hereditary form of a certain disease also has to be informed about the consequences of a possible hereditary form of breast cancer in the family, of Parkinson, of all those other things, et cetera. Well, that patient will run out the door screaming, he really doesn't want that. There is a whole range of infinitesimally small changes of incidental findings. How to deal with that? We now ask in general terms if patients want to know about incidental findings or not. I would say we shouldn't enforce a decision, but give everybody a choice. I always say that the most personal of our personal DNA is that it's personal. And that means you have to be able to make a personal choice about it.*

Because of the uncertainty that is inherent in genomics-based diagnostics, communicating the results to patients can prove to be a challenge. As one respondent noted:

*I think the discussion is mostly there where it is vague and difficult. So in other words, the questions you have to ask is "does the patient understand, when a sequence is made, when deep sequencing is done, what that means? Does the patient understand the uncertainty that surrounds it? … How to tell people what it really might mean and where the uncertainties are? How honest are you about it?*

The problem of the return of incidental findings was also mentioned several times:

*Specifically for very rare diseases for which individual hospitals have too little information to find a possible genetic cause, sharing data between hospitals could be a logical step. Patients are urging to please open up those data – they might benefit form data sharing. But what happens once all these genetic components are mapped, and once genomics data can be used to make predictions – and we have more and more. We now have data from several hundred patients, in a few years it might be several tens- or hundreds of thousands. And at that moment we don't even know as much as we might know in 25 years time. And as we are doing research, the genome data will always be there. And in 25 years time we can tell those patients: I am sorry sir, but we found out that you have a serious risk to contract such and such a disease. So that really is incredibly complicated. And you can't consider that from a technological perspective only. But then the dilemma for today is: can we deny patients the opportunity to find new therapies, and can we deal with risk profiles…*

Another issue that becomes more complicated is the right to withdraw:

*Research participants always have a right to withdraw. But in fact, it is near impossible to withdraw research data. The forms often say "you have a right to withdraw", but what they in fact mean is "you have the right to withdraw your consent, but you can't withdraw the information – and in fact we wouldn't know how".*

### 2.5 Addressing ethical, legal and social issues

There was broad agreement among respondents that the ethical, legal and social dimensions of data management are a priority:

*In my opinion, awareness [to address ethical, legal and social issues], at least with geneticists, that field, is incredibly high. The awareness "right, we need to do something with this".*

That said, respondents noted initiatives to tackle these issues were fragmented. While there were individual projects in specific areas where data issues play a role, such as ethical analysis of next-

generation sequencing for diagnostic purposes, there was no evidence of a single concerted effort to 'follow the data' as they move from patients and citizens through research to the clinic and back to the patient:

> *There is much discussion on technique and technical solutions, but as far as I know there hasn't been a lot of conscious thinking about medical ethical aspects. How to communicate with patients? And who owns the data? The well-known question that also applies to biological materials - who owns the sequencing data? If I'm being sequenced, does the data belong to the researcher, or can I expect to be given the data on a disc for personal use? And would I be allowed to give it to other people, because the researcher and I made the data together? I haven't seen those discussions explicitly on the agenda. Some people may be thinking about it, but that is still unstructured – those are the academics who simply reflect on this because of a personal interest in the topic.*

> *In that sense it is timely that you are now putting this on the agenda, the whole field is still wide open. So thinking about: how is this addressed in other countries, what kinds of ideas can you form about this, what things do you need to communicate to whom, etcetera. All that is still entirely open. And the other side of the story, the practical side of 'how are we going to arrange all this' is a discussion that is emerging as we speak.*

### 2.6 What is responsible innovation?

As part of the interview, respondents were asked to define what the term *responsible innovation* meant to them. To the majority of respondents, innovation is a *conditio sine qua non*:

> *I would like to turn the question around: innovation to me is a goal in itself. I think as a society we stand to gain a lot from it, and from an economic point of view innovation is a necessity. So the question should be: when is innovation irresponsible? That's important.*

Responsible innovation, in this view, implies respecting existing normative frameworks while innovating:

> *As far as I'm concerned, innovation is maximizing the opportunities that technologies offer, and responsible means you don't get a negative press because you broke the rules, or transgress commonly accepted norms. So the responsibility-part implies making sure that we know exactly what those frameworks are and do not transgress any ethical, societal or other frameworks, because that does more damage than we could ever hope to fix.*

By implication, all innovations that are realised within accepted normative frameworks would qualify as responsible innovations. 'Responsibility' in this sense thus mainly refers to the

responsible conduct of research, focusing on the characteristics of the actors involved in the innovation rather than the nature of the innovation itself. The question is what happens when the normative frameworks themselves become unstable and contested due to the exploratory nature and the disruptive potential of genome technologies.

> *I think as a researcher you have your own responsibility to keep thinking about what you're doing and whether you can still justify it with your conscience. So precisely at the moment that you enter a terrain where ther aren't many regulations yes, that you keep thinking about it. … So I think as a researcher you have a responsibility to think about what you're doing.*

Starting from a similar interpretation of responsibility (as the responsibility of the innovating actor), several respondents argued that responsibility in relation to data management not only implies *refraining* from morally reprehensible actions but also a moral *duty to act*: to them, the very ownership of valuable data to an extent entails a moral obligation to share that data:

> *Responsible data management is often translated as 'what I can and cannot do with my data, who has access to it and who doesn't?' My point is: what is your responsibility if you have data? To share them. Having valuable data and not sharing them, now that is irresponsible.*

The moral duty to share data was raised in relation to researchers, but some even raised this argument against research participants: if your genome holds secrets that may provide a clue to the prevention of disease, then you have a moral obligation to help your fellow citizens and share that data.

In addition to the individual, personal dimensions of responsibility, some respondents introduced additional – product and process-dimension-related[6] – elements of responsibility, for instance in relation to the product dimension of the innovation:

> *Responsible innovation means that you use the knowledge that is being developed to contribute to the development of society. Things that make people happier, that people find useful, that bring them closer together, that make the environment a little healthier or greener…*

Several respondents connected the procedural dimension of stakeholder involvement to the notion of responsible innovation:

[6] For the distinction between the product and process dimensions of responsible innovation see Von Schomberg (2012) 'Prospects for Technology Assessment in a framework of responsible research and innovation' in: M. Dusseldorp and R. Beecroft (eds). *Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methoden*.Wiesbaden: Vs Verlag, pp 39-61.

*I think I would consider 'supported' to be responsible. 'Supported' in the sense that stakeholders should have a voice.*

Another respondent noted:

*History is full of Collingridge-like dilemmas, so I think responsible innovation has predominantly to do with a certain reflective process in which sufficient attention has been given, on the one hand, to efficient technology development, that there has been sufficient reflection on the moral values, and the connected moral judgments - that that isn't purely theoretical but that important groups have been involved as well - and thirdly, that the scenarios and the impact of a technology has been assessed.*

## 2.7 Interdisciplinary collaboration

The interviews also enquired for experiences with interdisciplinary collaborations. Respondents mostly reported collaborations with legal and ethical experts. For most of them, the form of collaboration that sprang to mind was ethical or conceptual analysis performed in parallel with technological research projects:

*An ethicist could give a theoretical reflection on… what health care professionals could do, morally speaking, with data that is being generated or shared. And what a patient could do with it.*

Yet some saw cause for earlier involvement:

*Once you start using this in a pipeline towards the patient, you have to integrate ethical issues. But if the ethicist is involved late in the process, there is a chance that they won't be able to catch up.*

Indeed, earlier involvement of ethicists in technological design seems to emerge in several places. As one ethicist noted:

*Part of a constructive collaboration to me is that I try to think along on the front lines of research and offer my expertise, and that my colleagues on the other hand offer their expertise when I have a question.*

This approach sits well with the desire for practical, hands-on approaches as expressed by technology developers:

*Sooner or later there will be a list with identified problems and we'll need help in addressing them. And that's legal, it's ethics, it's societal discussions, it's laws and regulations, and so forth. But it would be só good if we would do it starting from practice. … To identify practical problems, and collaborate to solve them – not just to solve them, but to get better healthcare.*

At any rate, case-based work was seen as a precondition for successful interdisciplinary collaboration:

*Well, I think there is a knowledge gap between every other scientist. So between lawyers and ethicists as well. And if we talk in the abstract then they will stay in their own field of expertise. So I would try to make it very concrete. With very concrete assignments. If I have to cooperate with a lawyer and we're working on an abstract level, I don't understand the principles of the code of law or the reasoning behind it or whatever. But if we have to decide together "are these identifiable data, and explain why", then the lawyer will say "I'll have to find this specific article and this set of criteria" and I will give my input. My idea is that we will find each other much quicker. Certainly the abstract philosophers, I think the technologists completely drop out of that discussion. But if you sit them together on a very concrete question…*

## 3. Recommendations

This section summarises the recommendations that can be derived from the interview findings, including respondents' own comments and suggestions on how to shape responsible data management. The following priorities were identified to address ethical, legal and social challenges: professionalising data stewardship and integration; assessing public perceptions; enhancing user involvement; and strengthening education and training. These recommendations are discussed below.

### 3.1 Professionalising data stewardship and integration

The notion of data stewardship and integration was flagged across the range of expertise of respondents, suggesting different connotations of the term. First of all, stewardship has strong 'technical' connotations, as a requirement that data are accessible, that they are harmonized, that they can be used across platforms and systems, that they are stored safely and securely. Second, there is a normative connotation in terms of assuring research integrity – stewardship implies that researchers who submit data follow the norms of scientific conduct: that data can be traced back to the experiments in which they were generated, that the procedures for obtaining them are clear, uniform and transparent, that researchers keep appropriate records of data collection procedures, etc. A third element of stewardship concerns scrupulousness in granting access: in addition to preventing outright data theft or third party abuse, data stewardship requires clear and transparent procedures for data sharing between researchers, medical professionals, patients and citizens. The question of sharing brings us to a fourth and final dimension of stewardship: the question of ownership. Several respondents indicated that the rights and responsibilities of data donors merits further consideration – the permanence of rights (or the release thereof) is an integral part of the question of data stewardship and integration:

> *If you, as a research participant, donate data, you don't in fact give it up forever. Regardless of the legal question of who owns the data, in a very real sense they remain your data. In a way researchers have the data 'on loan'. And the question of stewardship is what you can and cannot do with the data as long as you have it on loan. So I think* responsible *data management is characterised by how well the terms of the loan are morally regulated.*

The question of data stewardship thus simultaneously engages questions of a 'technical' nature with normative questions. What's more, because of the novelty of procedures for data collection, storage and use, both the technical and moral norms are, in a sense, objects of negotiation themselves. The 'technical' dimensions of data stewardship (*"does everyone follow the procedures for data collection and storage? Is it clear where they come from and how they've been generated?")* thus address

but one part of the question – beneath the surface lies the deeper moral question *what are the morally appropriate terms and conditions for having these data on loan from their donors?*

The recommendation to be derived from this is that ***the debate on data stewardship and integration should engage technical and moral dimensions in unison***.

### 3.2 What do 'we' find acceptable? Public perceptions studies

The scoping interviews looked for 'expert views' on responsible data management. Similarly, detailed analysis of ethical and legal aspects of data-driven medicine has been largely performed by 'ethical and legal experts'. But since these developments potentially affect all citizens through changes in the health care system, several respondents identified the need to know how the public thinks about these developments. At the moment, there is limited evidence on public perceptions on the integration of genome data in health care, or on the ethical frameworks that are currently being developed. Indeed, several respondents identified a need to empirically test the validity of ethical frameworks:

> *The theoretical framework for the return of findings has been developed. I can imagine that a sociologist or a medical anthropologist could now empirically validate that framework, for instance by setting up a study that interviews all patients included in parallel with their inclusion in the sequencing protocol, in-depth interviews or a questionnaire for example that tests whether the patients have understood the information. Did they understand, or did they say they understand? What are their preferences?*

---

> *We should build case studies that are based on real situations, research and diagnostics of hereditary diseases, but also cardiovascular, and cancer, and it doesn't even matter what technology you discuss, it could also be proteomics or something else. And we should confront the public with this, in the right way, so that we come to understand both the first reactions and the underlying motivations, so we get a public judgment on ethical aspects and we can test whether our academic viewpoint actually corresponds to societal views. I think it's very important to verify ethical frameworks against societal practices.*

Several respondents noted that the measures we take today in fact concern future generations – so their views and preferences ought to be incorporated:

> *The next generation will think differently about things than we do. The world around us changes. What recently has changed significantly is the way we deal with information and access to information. New generations deal with information, finding and using information, in entirely different ways. This is what we forget in our discussions, because*

*we are having these discussions for the next generations, right? What we now invent is something that the next generation will need to deal with. And this is where we still fall short, because most academic discussions are based on the past and fail to incorporate future ways of dealing with information. And that might completely change the discussion on informed consent. Maybe we should do that entirely differently, and say: just have a look on this website, and that will be enough. While that is different at the moment. I don't know.*

In other words, ***there is a need to further assess public perceptions of data-driven medicine and empirically validate legal and ethical frameworks.***

### 3.3 User involvement - discussing acceptability

In addition to assessing perceptions of the public at large, respondents pointed to the need to involve actual users in the concrete development of data management infrastructures. The term 'user' was found to encompass patients, but also broader categories such as medical practitioners and the public at large. Patient involvement was mentioned several times as a way to smoothen the transition towards medical application:

*I think we should really see things form a patient's perspective. Because we often reason from the standpoint of the general public who are healthy. But it's the patients who mostly benefit from what we do. And the general interest very quickly and very often opposes the patients' interests. I don't know whether that means you have to involve medical specialists and patient organisations, but…*

Medical specialists were seen as another group that can convey information on patient perspectives:

*To involve medical specialists is very important, because they see those patients very often, and of course, every patient is different, you have to take that into consideration, but generally speaking patients respond very differently from the way we expect them to.*

The patient's perspective however constitutes a particular perspective. Some respondents instead advocated broader user involvement:

*How to arrange safety, security and those sorts of things? It's all being handled very technically at the moment, but such a discussion with a wider forum… could be very meaningful I think.*

---

*In relation to questions of the return of findings for example, I think now is the time [for patient involvement]. The theoretical premises have been defended and elaborated in the international literature. So they know now what types of policy there are and what your moral… what types of argumentation there is for it. And I think it's now time to test this empirically with research subject and patients. Just to see "do they understand it? What are their preferences? I'm making all these assumptions, but they need to be tested at some point." … So I think that social research, to quantitatively measure preferences, I would be very happy with that.*

Some argued that broad societal involvement is a necessary condition for social acceptance as it satisfies public demand for transparency:

*You can see that public pressure is increasing, that people are saying "I want to be better informed, I want to know what happens to clinical material" And not just radio or TV shows but people who are actually saying "I want to know what happens with my material, but also with my data." And if people are being sequenced they also want to know.*

There seems to be an underlying assumption that the pervasiveness of these developments merits a broader societal debate. This leads to the following recommendation: ***the potential impact of the use of genome data in diagnostics for patients and citizens alike requires the involvement of patients, medical professionals and the public at large in the development of data management procedures.***

### 3.4 Strengthening education and training

Nearly all respondents recommended strengthening education and training, although there was considerable variation in the topics and target groups identified. Some noted that researchers should be better educated about the practical use of their research findings in the clinic. Others emphasized training medical specialists on the implications of genome data:

*That also goes for the way we educate our medical practitioners, which is a very conservative way, because it's the conservative people like me as it were, and those above me, who provide that education. While those are precisely the people who will have to address the patients question like "well, I've had my genome passport determined, and I've read this and that, and I wonder what to do with it".*

Several respondents also stressed the need to educate the public at large, arguing that increased knowledge of genetic information might increase support for its use in diagnostics (and perhaps lower unrealistic expectations). Then again, some concerned doubt:

> *I'm not sure yet, how to prevent [public concern]. One of the reflexes one always sees is that people say "well, we have to explain it better to the public." But I always doubt whether that will work because a considerable majority will think "if they're going to such great lengths to tell us all this, what do they have to hide?"*

So while respondents broadly shared a general concern for education and training, no specific priorities emerged from the range of target groups and learning objectives proposed, leading to the recommendation that ***priorities for education and training need to be established.***

## 4. Conclusion

The interview findings paint a variegated picture of the opportunities and challenges for data integration in the life sciences. Some respondents emphasize revolutionary opportunities for diagnosis, prevention and treatment – others stress concerns over privacy, ownership or informed consent. The use of genome data in a clinical context is still in its infancy (genome technologies aren't part of routine diagnostic procedures just yet, and 'personalised medicine' is largely a synonym for improved patient stratification). Yet respondents agreed that the disruptive potential of big data calls for attention to its ethical, legal and societal dimensions.[7] They shared an interest in questions of data integration and stewardship, the revision of informed consent procedures, the social acceptability of foreseen developments and education programmes, although there was considerable variation in their preliminary answers to these questions. The recommendations above reflect these findings and suggest ways forward.

The overall conclusion to be drawn from these findings is that the potential impact of the application of life science data in the clinical context (and vice versa) mandates broad user involvement. As one of the respondents noted, *responsible* data management requires that data infrastructures (specifically the normative decisions contained in their architecture) are widely *supported*: by patients, researchers, medical professionals, and the public at large. To achieve such support, it is imperative that broader ethical, legal and societal concerns are addressed at early stages of technological development, so that the outcomes of the discussion can still shape technological decisions.

In response to this outcome, the ReDaPeD project plans to organise a transdisciplinary forum involving researchers from various disciplines, ICT professionals, medical experts, patients and citizens to proactively define the technical, legal and moral frameworks for *responsible* data management for personalised diagnostics. This approach is in line with the concept of responsible research and innovation as it is emerging in national and European science policy debates, focusing on the early engagement of stakeholders in technological developments. On this approach, ethical aspects are not considered as constraints, but as drivers for the innovation process. 'Responsible data management' involves taking the needs, concerns and values of different stakeholders into account in the design of infrastructures for the storage, integration and use of data collected for diagnostic or health monitoring purposes.

[7] As Roy Amara noted, we tend to overestimate the effect of a technology in the short run and underestimate its effect in the long run.

# Appendix

## List of respondents

| | |
|---|---|
| Dr Jan-Willem Boiten | Project Manager, TraIT (Translational Research IT) |
| Dr Annelien Bredenoord | Assistant Professor Medical Ethics, Julius Centre, University Medical Centre Utrecht |
| Prof Edwin Cuppen | Principle Investigator, Hubrecht Institute / Professor of Genome Biology, Department of Biology, Utrecht University / Professor of Human Genetics, Head of Research, Department of Medical Genetics, University Medical Center Utrecht |
| Dr Peter Doorn | Director, DANS (Data Archiving and Networked Services) |
| Dr Ruben Kok | Coordinator, Dutch Techcentre for Life Sciences / Managing Director, Netherlands Bioinformatics Centre |
| Dr Colja Laane | Director, Netherlands Genomics Initiative |
| Dr Jeantine Lunshof | Ethics Consultant, Personal Genome Project at Harvard Medical School / Assistant Professor, Section Molecular Cell Physiology, Faculty of Earth and Life Sciences, VU University Amsterdam |
| Prof Gerrit Meijer | Head, Department of Pathology, VU University Medical Center / Principal Investigator, TraIT |
| Prof Frank Miedema | Professor of Immunology, University Medical Centre Utrecht / Dean and Vice chairman of the Executive Board, University Medical Centre Utrecht |
| Prof Barend Mons | Professor of Biosemantics, Leiden University Medical Centre / Scientific Director, NBIC |
| Prof Ben van Ommen | Principal Scientist, TNO / Director of the NuGO, the Nutrigenomics Organisation |
| Dr Rene von Schomberg | Scientific Officer / Policy Officer, European Commission |
| Dr Terry Vrijenhoek | Programme Manager, Centre for Genome Diagnostics |

## Interview guide for the Scoping Interviews

*[translated from Dutch]*

### Introduction

1. What is the purpose / mission / context of the organisation you work for?
2. What is your role within this organisation? Could you briefly describe your current work?

### Assessing life science technologies and applications

3. According to you, what are the most important technological developments in *[respondent's specific research area]* for medical research and health care?
4. To what extent are these technologies applied in practice, or could they be (more extensively) applied (in diagnostics for example)?
   a. What are the biggest opportunities?
   b. Outside of clinical diagnostics, are there applications for public health (prevention)?
   c. What are the challenges that could impede the use of this technology in (clinical) practice?

### Data issues

5. To what extent is data sharing an important issue in the development and use of this technology for both research and in clinical practice? To what extent are data being shared? Why (not)?
   a. Who are sharing data (researchers, physicians, industry, patients, government)
   b. What opportunities does sharing of (research and medical) data offer? To whom?
   c. What risks could be connected to data sharing? For whom?
6. To what extent do biobanks offer a model for questions of data management?

### Users and stakeholders

7. Who are the most important users and stakeholders in data obtained by way of life science technologies? *[doctors, patients, industry, government, citizens]*
8. How are these technologies received by (envisioned) users and beneficiaries (doctors, patients, industry, government)
9. What will change for doctors and patients if genomics technologies are applied on a large scale for medical / diagnostic purposes?
10. To what extent are users of these technologies involved in the development of the technology?
    a. Do you think their involvement would be useful? Why (not)? How could involvement look?

### Ethical, Legal and Social Issues

11. To what extent do you foresee ethical, legal or social challenges for the use of data obtained by -omics data in a clinical context? *[checklist: informed consent, privacy, access, third party abuse, reimbursement, ownership, direct to consumer testing]*

12. To what extent do those ELSA-issues affect your own work? Is it something you address during work? If so, how?
13. Who identifies these challenges? Why?
14. Who is affected by these challenges? When and where?
15. Where do you recognise opportunities to address (or prevent) these ethical / legal challenges? Who is responsible for addressing them?

**Interdisciplinary collaborations**

16. To what extent could it be useful to cooperate with ethicists or social scientists during the early stages of development of technological platforms and standards for storage, sharing, access to data? Why is or isn't that useful? Under what terms and conditions?
17. Do you have personal experience of collaborations with ethicists, philosophers or social scientists? If so, could you describe that experience? How would you rate those collaborations?

**Responsible Research and Innovation**

18. The title of this project is: **responsible data management** for personalised diagnostics; this refers to the notion of responsible research and innovation that features in Horizon 2020 and in the Dutch Top Sector Policy.
    a. Are you familiar with that term?
    b. What does 'Responsible Innovation' mean to you in the context of life science technologies and data management?

**Personalised medicine**

19. The title of this project is: responsible data management **for personalised diagnostics**; the term personalised medicine is often used. What do you understand by this term? What role do life science technologies play here?

**Networks and references**

20. Which individuals and organisations cannot be left out in the discussion on responsible data management in the context of life science technologies? *[in the Netherlands / Europe / globally]*
21. Which issues need to be addressed in that discussion? What is the most important purpose of that discussion in your view?
22. What are the key references on data issues in life science technologies? *[scientific and grey literature, policy documents, funding opportunities, congresses and symposia]*

23. Are there any other topics or issues that you would consider to be relevant for this project?