Transition costs for open science in the Netherlands

Connecting and Accelerating

Robert Consultancy & Technopolis Group, September 2019

Robert van der Vooren
Elma Dujso

Table of Contents

Tables

Figures

Executive summary

This report is the result of a study in the context of the National Platform Open Science on the transition costs to accelerate and connect open science activities in the Netherlands. More specifically, it was asked to map the transition costs for 2020-2023 (4 years), and beyond.

Open science is an important and necessary driver of changing science. Technological and methodological progress is more than ever not exclusively driven by and within a research discipline. During the study we gathered input from 27 experts in the field. Most experts describe the Dutch position on open science as one with a good start in the past yet a steady but too slow progress. There are ambitions to pursue a top position in open science, as it improves the research quality.

Before exploring ambitions and necessary action more closely, we offer a baseline economic model that reflects today's spend on open science in the Netherlands. Together universities spend about 45 to 110 million EUR on open science annually, in relative terms 1% to 3% of their research budget. For the entire public research sector, we calculated an annually spend of 98 to 237 million EUR on open science. Clearly research performing organisations in the Netherlands take an active role in fostering open science. However, with the shared ambition to be among the best, action is needed.

Deeper insights in the ambition have taught us there are 5 crucial challenges ahead. We have found there are 15 actions required to face the 5 challenges. Below we point out each of the challenges with an example of an associated action:

1. Challenge: lack of open science expertise. Exemplary action: more software engineers.
2. Challenge: coordinated and better connections are necessary. Exemplary: interdisciplinary projects.
3. Challenge: standards are lacking. Exemplary: (cross)disciplinary data protocols.
4. Challenge: more is needed to pursue open access. Exemplary: alternative publication platforms.
5. Challenge: raising awareness is needed. Exemplary: campaigning.

The ambitions, challenges and associated actions have been used to design 3 scenarios for the upcoming 4 years between 2020 and 2023. The aim and purpose of scenario 2 has most similarity with the German national programme for research data infrastructure NFDI that was launched early 2019. We have used this comparison to scale down the NFDI funding to a Dutch national level. Below we have briefly described the 3 scenarios and their associated transition costs 2020 – 2023 for the Netherlands.

Scenario: ***Target:***	**Business as usual** ***OA: ~70%*** ***data: findable and accessible***	**In line with the rest of the world** ***OA: ~80%*** ***data: + reusable***	**Ahead of the rest of the world** ***OA: ~90%*** ***Data: + interoperable***
Total annual transition costs	**5,7 – 10,7 million**	**12 – 18 million**	**19 – 37 million**

The total annual transition costs of the scenarios in relative terms (compared to the current spend on open science) ranges between 3-6% of the research budget in scenario 'Business as usual' and 10-20% in scenario 'ahead of the rest of the world'. These are the annual transition costs during 2020 until 2023 (4 years), however we recommend committing to 10 years funding at the same level as the first 4 years. If the Netherlands truly seeks connection with both the ambitions of stakeholders and international developments such as the European Open Science Cloud, we recommend not to pursue the first scenario 'business as usual' but to regain a top position by following one of the other scenarios.

1 Introduction

Open Science is an important and necessary driver of changing science. Technological and methodological progress is more than ever not exclusively driven by and within a research discipline. Change increasingly comes from outside the university and it knows no geographical or disciplinary borders. The challenge for universities now is to how to respond and interact.

Digitisation is an important factor that puts these changes into action. Digitisation also creates new possibilities. Open science is focused on optimally using these opportunities. Open science has an impact on the way research is conducted (open methodology), published (open access), taught (open educational resources), and interacted (citizen science). Besides, it has an impact on how the data is used by researchers (open data).

Open science is seen by some as a social movement that you can either join or not. That is not entirely incomprehensible. Open science is however more than that. Open science, including open data, creates important drivers of change in society. This new way of doing research ask for more sharing and opening data and a good and fair use of research data.

Since the Competitiveness Council conclusions of 2016 Europe agreed to achieve 100% immediate open access of scholarly articles by 2020. The current Dutch government responded by declaring open access and open science the standard. In 2017, 50% of articles with a Dutch affiliated author have been published open access. The share of hybrid open access big deals has significantly increased in the past three years. The Dutch infrastructure for storage and management of research data has central and decentral characteristics. Research performing organisations have different solutions for daily storage needs, for secured international cooperation and for archiving. National Platform Open Science (NPOS) underlines the reusability of research data by using FAIR principles regarding research data that go hand in hand with publications.

On European level there are also relevant developments with the launch of the European Open Science Cloud (EOSC), which is international and transcending scientific domains. The Netherlands is within Europe the initiator of the activities to realise the EOSC in terms of GOFAIR.

1.1 Scope of this study

In June 2019, Robert Consultancy and Technopolis Group were asked to conduct a short study in the context of NPOS on **the transition costs to accelerate and connect open science activities in the Netherlands**. More specifically, it was asked to map the transition costs for 2020-2023 (4 years), and beyond.

NPOS has been composed in 2017 after the Secretary of State of the Ministry of Education and Culture had written a letter to the Second Chamber about open science requesting the creation of the national plan on open science (OCW, 2017). The platform brings together stakeholders to realise the ambitions in the context of open science. One of those ambitions as stated in the plan is to achieve open access publishing and optimal reusability of research data by 2020. At this moment, the platform moves towards an operational programme with an appropriate coordination.

Open science covers a wide range of research topics, in fact so many that choices within this study had to be made. The topic of research data is clearly present. Open access is still important, despite the uptake in recent years. For the purpose of this study we focussed on ambitions, obstacles, actions, coordination, communication, human resources, awareness and training.

The keywords in this report are **connecting and accelerating**. Those are the keywords that have a prominent role in the three scenarios that we developed in the context of this study and for which we mapped the transition costs that are related to the scenarios. Within this context, there is an urgent need for a new impulse, for a national programme for open science, and to make the Netherlands an international frontrunner in the field of open science again. We will elaborate on this when we present the scenarios.

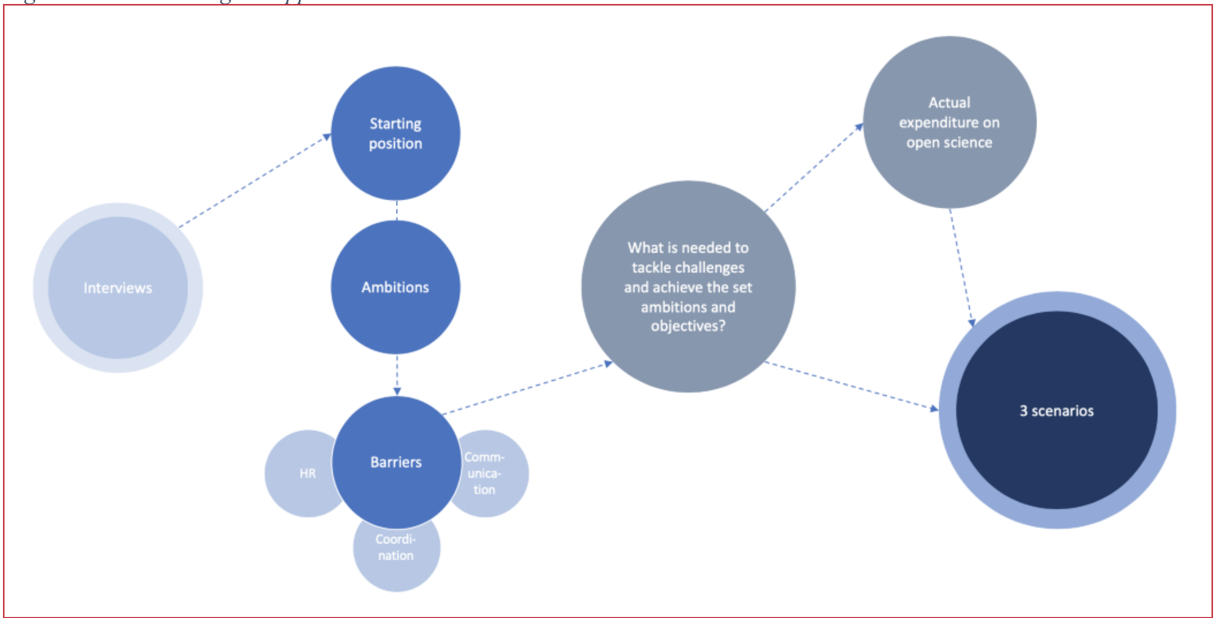
1.2 Methodological approach

In this paragraph we present our methodological approach in which we have integrated the research questions and the scope of the study. We have divided our methodological approach into four phases:

1. Desk study on similar open science programmes with focus on the German NFDI programme, used as a benchmark in actions, ambitions, expenditure and as an anchor for scenarios
2. Interviews with 27 experts for determining starting position, ambitions, obstacles and actions
3. Selecting simplified economic model for mapping current costs and identifying scenarios
4. Concise survey among Dutch universities with focus on inventory of open science activities and use those activities in financial translation applying the simplified economic model.

Those four phases are explained in more detail in following subsections. They contribute to the description of the current state of open science, the ambitions and challenges, in defining activities to overcome these challenges, to estimate the current expenditure, and feeding into three different scenarios for transitions costs (see figure 1 below)

Figure 1 Methodological approach



Robert Consultancy & Technopolis Group

1.2.1 *Desk study*

We conducted desk study, for which we have especially investigated the German NFDI programme (National Forschungs Data Infrastruktur) funding as an anchor and best practice that we could then downscale to the Dutch national level. Two significant reports that preceded the NFDI were published by the German council for research infrastructure RFII. The RFII perceived, as presented in its (2016) position paper, “upcoming disruptive innovations and radical changes that alter research processes across national boundaries.” It was certainly felt that Germany’s research infrastructure and governance needed development. The RFII noticed in its international comparison (2017) a trend towards more top-down research data policies such as national data platforms in the Netherlands.

Despite the positive effects of top down measures, the report questions the sustainability of these platforms, as they do not provide the commonly acknowledged need to integrate scientific users. To solve the integration challenge, the RFII stresses the potential of the consortia model of the national research infrastructure programme NFDI. The NFDI consortia model is based on early stage activation of research communities who will then submit project proposals with partners. Thus, German research seems to be in favour of a balanced approach between bottom-up integration of research communities and top-down measures. The latter clearly underlined in the RFII report that mentions the need and success of binding principles such as the NWO control mechanism to require the submission of data management plan.

The NFDI programme set up in 2019 aims to enhance Germany's research data infrastructure, to network data (inter)nationally and to build a consortia network that provides science driven services. The programme is by no means restricted to hardware solutions. Appendix B describes the facts of the 10-year NFDI programme in terms of its aim, purpose and annual 85 million EUR funding. In line with the RFII reports, the NFDI serves to address the needs of researchers, to facilitate easy access to data and to foster sustainability. Proven practices, communication cross consortia and national networking are key vision elements. It is important to notice NFDI's 10-year funding of around 3 million projects with a mid-term review to prevent ineffectiveness. Projects with to short funding periods cannot be adopted with sustainable funding in time. Other risks for ineffectiveness are the lack of indicators and fragmentation when project funding is to small (RFII, 2017). A programme such as NFDI requires a high tiered governance; in Germany research funder DFG coordinates.

The Dutch royal academy KNAW (2019) concluded that big data in social sciences and humanities will enhance research in the Netherlands if researchers relieve proper facilities such as infrastructure and discipline specific training. Big personal data especially urges researchers to obtain additional expertise and (cross-) disciplinary methods and techniques. The royal academy emphasises the need for an overarching national infrastructure and cooperation within the research community.

Our study initially builds on qualitative data. The quantitative and economic stage is based on a comparison of the German NFDI programme to each of the three scenarios. The scenario with the best match to NFDI serves as an anchor to downscale the German NFDI programme funding to the Dutch national level. To better sense the financial impact of the Dutch transaction costs in the three scenarios, a cost reference is made with the use of a simple economic model (see 1.2.3). This economic model shows today's annual expenditure on the most directly identifiable open science facilities by research performing organisations.

1.2.2 Expert interviews

We conducted interviews with experts in the field of open science in the Netherlands from different research performing organisations. The list of interviewees is attached in Appendix A. During the interviews we have asked the experts about the current state of affairs in the country, the challenges that the Netherlands is facing in terms of open science, what the ideal situation would be, and more specifically focused on the existence of HR, communication, training, policy in the context of open science. The findings from the interviews will be described in the next chapters.

1.2.3 Simplified economic model for mapping current costs and identifying scenarios

Having understood and mapped the Dutch open science starting position, ambitions, obstacles and actions based on the interviews, the study's objective is to identify possible scenarios. Three scenarios are distinguished in a two dimensioned (ambition and coordination) plot.

In developing the scenarios, we have considered the different roles and responsibilities for open science of research performing organisations, research funders and the government. Typically, research performing organisations must lead the actual transition, whereas research funders have the role to set conditions for change in their research funding programmes. The government has a role in bringing together the different actors by introducing national policy aims and instruments.

Different macro or micro economic models can be used to study the costs of a sector transition. However, the sector of research performing organisations can hardly be labelled homogeneous. Micro economic models such as data management budget tools are not applicable to the many different research disciplines. There seems to be more consensus regarding macroeconomic approaches to research transition costs. Best practices of large-scale transitions, such as the in 2019 launched German NFDI programme for research data , are relevant references for developing accurate macroeconomic scenarios in the Dutch context.

Within the framework of this study, costs are defined as transition costs. Thus, they should be understood as costs that are different from fixed costs but rather temporary, transitional and project based. Transition costs aim to speed up and accelerate the open science developments, to connect different stakeholders in the field, and to pursue the open science ambitions.

Typically, such costs emerge temporarily as long as the transition lasts. For a transition it is crucial to know the starting position, the strategic aim and the required actions or interventions, sometimes necessary to resolve obstacles. These questions are in the backbone of this study. Our macroeconomic model for transaction costs begins with a survey among the many stakeholders in the research sector.

1.2.4 Concise survey

We sent out a concise survey among Dutch universities. The findings of this survey were used for our simplified economic model. As we wanted to capture the current expenses of universities to get an idea of how much is spent on open science and how this relates to the investment scenarios for the transition, we sent out a concise survey to a few universities with the idea to extrapolate the findings.

The method used to map the annual expenditure exists of two steps: 1) inventory of open science activities and 2) financial translation of the open science activities. We have separated those two steps for a more efficient execution. We are avoiding financial questions to universities and focus on what they know, more quantifiable and whether certain elements are existing. For instance: the interviewees or experts usually do not know how much a data steward costs per year or how much a repository costs, but they do know how many data stewards there are or whether there is a central repository for their articles.

After the inventory we make the financial calculation based on exemplary annual costs within a bandwidth. For instance, a 1.0 FTE data steward position is estimated between 70.000 and 90.000 EUR (annually). The costs of a repository (management, maintenance, replacement included) range between 400.000 and 700.000 EUR (annually).

For the inventory were selected: Delft University of Technology, Utrecht University, University of Groningen and Wageningen University & Research. Based on the findings of the survey, we have made a financial calculation of the current situation, that we have consequently used as a baseline on which we’ve built our scenarios.

1.3 Structure of this report

This study is structured in the following way. Chapter 2 describes the current state of affairs in a series of patterns that were derived from expert interviews. This chapter also analyses the ambitions and challenges identified by the different actors and synthesises them into a shared (national) ambition as well as actions that are linked to the identified challenges. In chapter 3 three scenarios are outlined for a transition towards open science. Chapter 4 connects the findings of the previous chapters in a set of recommendations.

2 The current state of open science in the Netherlands

This chapter presents the current state of affairs with regards to open science developments in the Dutch research and science landscape. It also describes the main collective ambitions and challenges for the development of open science in the Netherlands. For this, we interviewed various open science experts in the Netherlands. In Appendix A, we have included a list of the interviewees.

2.1 Open science in the Netherlands

Interviewees indicated that the necessary transition towards open science is a task primarily belonging to research performing organisations. One could speak of open science as an intrinsic ambition to develop science. Investments and costs related to open science developments are therefore secondary to these intrinsic ambitions. As an intrinsic challenge, it seems that a large part of the progress of the open science related ambitions and plan are thus difficult to concretize into specific actions. Therefore, there is a need for **more national coordination** to avoid everyone inventing the wheel all over again. Instead, interviewees indicated that it is relevant to **share knowledge** as to what activities related to open science the stakeholders in the field are involved in.

We observed an increase in **bottom up initiatives and pilots**, also citizen science initiatives, such as a pilot in one of the universities for coding assistants which is coordinated from the university library where faculties hire someone to assist and support in coding and writing for four consecutive weeks.

We also observed that research performing organisations currently invest more in **data steward tasks** rather than data steward positions. Researchers spend time on data steward tasks as an add-on to their research tasks. Related to this, we see a trend in growing skills in the field of data. Currently, training with regards to data skills is organised through separate modules but also in mandatory modules on RDM and FAIR data. However, also in software we see an emerging relevance, as 80-90% of researchers use software. Software has become just as relevant as data in the open science trend. Therefore, software expertise is needed. Some experts indicated that software sustainability experts are necessary just as much as data stewards.

In terms of supportive data management, the concept of local **Data Competence Centres** is introduced. Wageningen University for instance already introduced such Data Competence Centres some years ago. These competence centres are the point of contact for researchers. Simultaneously, they connect, coordinate and support everything and everyone in relation to research data management and value creation such as big data analysis. However, what remains a challenge is the different levels of awareness regarding open science across scientific disciplines. There needs to be a **balance between the willingness and ability** to change. Recent research from CWTS and Elsevier (2019) shows that researchers understand the necessity and value of sharing data, but in practice data is published considerably less.

2.2 Expenditure on open science activities

Based on a survey conducted at a small number (4) of Dutch universities, we mapped the current expenses on open science activities. Subsequently, we extrapolated the findings of this survey to the Dutch public research sector.

The survey consisted of questions related to the following:

1) the number of data stewardship as a position or as a task at the research performing organisation,

2) the number of people involved in the central support team for open science,

3) the existence of a central system for research data,

4) the number of data storage facilities and systems,

5) the existence of open science programmes,

6) the existence of a repository for scientific articles, research data, books,

7) the existence of an open access fund and the annual budget,

8) the existence of a facility for publishing support and

9) other open science facilities.

Based on these 9 questions and the subsequent findings, we were able to build an economic model that shows the **current** state of affairs in terms of **annual costs of open science** (as shown in table 1). The activities and facilities in table 1 are directly related to open science, for instance data stewards, data storage and funds for open access Article Processing Charges (APC). The model uses the number of activities and facilities counted in the survey to determine the number at the level of all universities (VSNU). Each of the activities and facilities ranges from a minimal unit cost price to a maximum unit cost price. From this stage all costs are expressed in annual costs. Quantity of activities multiplied by unit cost prices determine the minimal and maximum total costs.

For this model, we assumed the four universities to be representative for the VSNU, excluding the university medical centres and research institutes. In order to include all research performing organisations, however, we used data on the domestic expenditure on R&D by public sector (PSERD) to determine the multiplier between VSNU and the total public research sector.

Table 1 Annual open science costs of the VSNU and the public research sector in the Netherlands

Activity/facility	Quantity survey n=4	Quantity VSNU	Minimal unit costs	Maximum unit costs	Minimal total costs	Maximum total costs
	#	#	*amount in EUR*		*amount in mln. EUR*	
Datastewards (fte)	36	126,0	65.000	95.000	8,2	12,0
Central research supporters (fte)	48,9	171,2	65.000	95.000	11,1	16,3
Central Research Information System	4	14,0	300.000	800.000	4,2	11,2
Data storage facilities	31	108,5	100.000	300.000	10,9	32,6
Training	11	38,5	20.000	70.000	0,8	2,7
Repositories	8	28,0	200.000	400.000	5,6	11,2
APC funding	2	7,0	50.000	300.000	0,4	2,1
Publishing support facility	1	3,5	1.000.000	5.000.000	3,5	17,5
Other	3	10,5	100.000	400.000	1,1	4,2
Total annual costs VSNU, absolute terms					**45,6**	**109,7**
Total annual costs VSNU, relative terms					**1,3%**	**3,2%**
Total annual costs Public Sector, absolute terms					**110,4**	**265,4**

Robert Consultancy & Technopolis Group

The model offers a baseline in the study, as a reminder of today's annual costs of open science and a reference for comparing the scenarios in chapter 3. VSNU spends about **45 to 110 million EUR on open science annually**. In relative terms this means that VSNU spends between **1% and 3%**. For the calculation in relative terms, we assumed the research budget to be 50% of the total universities' revenue. Subsequently, we used the same ratio for the total public research sector.

VSNU does not include all research performing organisations in the public sector. We have calculated that the Dutch **public research sector annually spends 110 to 265 million EUR** on open science activities and facilities[1], assuming the same relative expenditure.

[1] - Stats.oecd.org, 'gross domestic expenditure on R&D by public sector (PSERD)'. Most recent PSERD year is 2017.

The results in the model indicate that research performing organisations in the Netherlands not only endorse European open science recommendations. They also fund the implementation of necessary actions to foster open science, with the objective to make it a norm.

2.3 Open science ambitions

Having mapped the current open science affairs in the Netherlands, in this paragraph we present the open science ambitions, based on the interviews with experts, desk research and our consultant expertise.

First, it was indicated that the Netherlands has the aim to become the frontrunner in open science again. The Netherlands used to have this position, with the Amsterdam Call for Action on Open Science in 2016, setting up the National Plan Open Science and signing the Declaration Open Science in 2017, and relevant organisations signing the San Francisco Declaration on Research Assessment: an international initiative that intends to stop the irresponsible use of bibliometric indicators in 2018 and 2019. Internationally there is a variety of methods for measuring the relative number of open access publications. Different sources report different open access scores per country. However, it is safe to conservatively conclude the Netherlands has a top 5 open access position globally since 2016. Experts indicate open access nowadays is still growing, but recent figures remained behind expectations and additional efforts might be needed.

The *mission* that seems to bring together all the stakeholders in the Dutch open science landscape is a **better, transparent, impactful and replicable science**. Science should be understood as **science for society**: sharing data is relevant to tackle societal challenges. Therefore, **big data skills** are a must for all the different disciplines regardless of their familiarity with FAIR data. Research performing organisations want **more data stewards** at more and different levels to support this development. However, experts indicated that it is important there should be a professional role or position for data stewardship with a **clear career perspective**. Such new scientific profiles should be created that receive **appreciation**. Furthermore, a shift from solely one researcher to multidisciplinary research teams with data stewards and software engineers is ongoing. Citizen science as part of open science has been underexposed, yet it is a trend that universities consider in their new policy plans.

In order to move towards these ambitions, **more central coordination** is necessary. With coordination we do not refer to the governance structure that has recently been improved. What we do mean is the need for a more active coordination to support the ambitions we heard during the interviews.

2.4 Challenges for the transition

The potential challenges for the transition to open science in the Netherlands can be synthesised into **five** main challenges.

1. Lack of open science expertise

In our interviews we see consensus about a need for more expertise and more recognition of open science activities. Open science expertise is a broad definition, in many cases (cross) discipline specific about skills such as software engineering, data curation, etcetera. Many interviewees indicated that there are no incentives for sharing data, for involving non-academics in research and that there is a lack of expertise in open science. Therefore, there should be more appreciation and recognition for positions such as data stewards, as well as data managers and software engineers. This goes hand in hand with more awareness. Universities are financing their data stewards on their own, but the ingrowth of data stewards is not going rapid enough. Furthermore, in terms of data facilities, there is also a huge distrust in commercial cloud services, which keeps the costs of data management too high.

- vsnu.nl, facts and figures. Universities' total revenue in 2017 was 6.932 million.

2. Coordinated and better connections are necessary

As stated in the introduction, connecting is one of the keywords of this report, in addition to accelerating the open science transition. Most interviewees agreed that connecting with other stakeholders is key in achieving our ambitions and key in sharing existing knowledge. For that to happen it is relevant to ask the question who would need the knowledge or insights that one has and how this could be communicated with others.

For all this to be made possible, it was indicated that better coordination between the various national parties is needed. The National Platform Open Science is a step in that direction, but it could be further enhanced. Experts mention that universities and research performing organisations more broadly should build an open science expertise more intensively and more professionally, and that the government needs to put instruments in place to accelerate this.

3. Standards are lacking

A third challenge that we identified is the lack of standardisation. Interviewees indicated further development of FAIR data principles requires leadership, debate and most of all research protocols. The latter at least within a discipline but preferably cross-disciplinary. In the KNAW report (2019) about big data in social sciences and humanities it is indicated that cross-disciplinary practices are needed as more and additional big data expertise and technique is required. Certainly, this is the case for sensitive data in the context of data protection legislation. However, for all disciplines, improving FAIRness of data contributes to research quality. It was felt among interviewees that it would help share more actively and on a national scale best practice.

4. More is needed to pursue open access

The 100% open access is not a requirement but an ambition. According to experts, progress has been made in terms of open access. However, compared to the years 2015 – 2017 experts signal less growth in open access. In addition, experts express concerns as current and future negotiations that largely will determine the success towards 100% open access. This ambition strongly depends on the negotiations in 2019/2020 with publishers for immediate open access (big deals, see latest EUA report). Alternative strategies and instruments are not fully settled. However, progress made in 2018 and 2019 for a more scalable green route to become a serious possibility (requiring additional efforts/funding). Several initiatives for sustainable =alternative publications platforms are explored as well. Furthermore, as experts indicated, open access for books, conference proceedings and native journals also need a push, next to stimulating smaller publishers in their transition to open access or a scalable and efficient way ('long tail'). Copyright retention would strengthen the negotiating position of libraries and make it easier for researchers to share articles open access. So far most within the research community fear the current vendor lock-in cannot be easily solved.

This transition to open access is a work in progress and transition costs to push open access towards (near) 100% are to be considered as well.

5. **Raising awareness is needed**

All change starts with awareness. We still see a difference across the scientific disciplines in terms of knowledge and awareness of open science developments. Particularly in the transition phase, and in a time where there is stigmatisation of the older generation researcher, it is crucial to raise awareness around open science. In addition to awareness through for instance campaigning, more training of both older and newer generation researchers is key. The aim of awareness raising is to provide more facts and figures as there still are inaccuracies and misunderstandings.

2.5 How can these challenges be prevented or overcome?

The table below advises on actions that can be undertaken to prevent or overcome these challenges. Some of those actions have already been mentioned throughout the discussion of the challenges. The table below provides an overview of this.

Table 2 The 5 challenges in the Netherlands and 15 actions to solve them

Challenge	Actions to prevent or overcome challenges
Lack of open science expertise	1. More by means of (more) data stewards at research performing organisations 2. More data managers at research performing organisations 3. More software engineers at research performing organisations 4. More incentives for data sharing 5. More appreciation and recognition for the abovementioned data tasks 6. A national plan and more cooperation on (meta) data infrastructure and services
Coordinated and better connections are necessary	7. Sharing knowledge among stakeholders in the open science field 8. Better national coordination in making this exchange possible 9. More platforms, more cooperation, and projects across stakeholders
Standardisation is lacking	10. Standards and best practices for FAIR data
More is needed to pursue open access	11. Financial support for reaching OA for scientific disciplines that have no OA deals or means in order to reach those 12. Alternative publication platforms 13. Open Access for books, conference proceedings, native journals
Raising awareness is needed	14. Campaigning with best practices, ambassadors and factsheets 15. Training of researchers

Robert Consultancy & Technopolis Group

Based on this analysis that was built on the current state of affairs, the ambitions, the challenges that go hand in hand with these ambitions and the actions that are needed to overcome these challenges, we have identified three scenarios that on the one hand provide elements of open science within the scope of our study that are related to the actions presented above, and on the other hand define the transition costs understood as costs to accelerate the developments in the field of open science in the Netherlands. These scenarios will be presented in the next chapter.

3 Scenarios

3.1 Methods used to define scenarios

Whilst defining the scenarios to tackle the challenges in open science and to include the transition costs for the open science developments in the Netherlands, we divided this into three stages:

1. Designing scenarios by ambition and coordination,
2. Incorporating challenges and actions into scenarios,
3. Determine costs associated with each scenario.

During the first stage we set the main criteria to systematically distinguish between scenarios. In this study we use **ambition** and **coordination** as the main criteria. These two criteria have also been applied to the interview guide and thus generate a lot of information about how participants feel about the transition in terms of aims, coordination and obstacles. Although interview participants do not share a clear ambition, we have been able to observe different levels of ambition. The same goes for coordination.

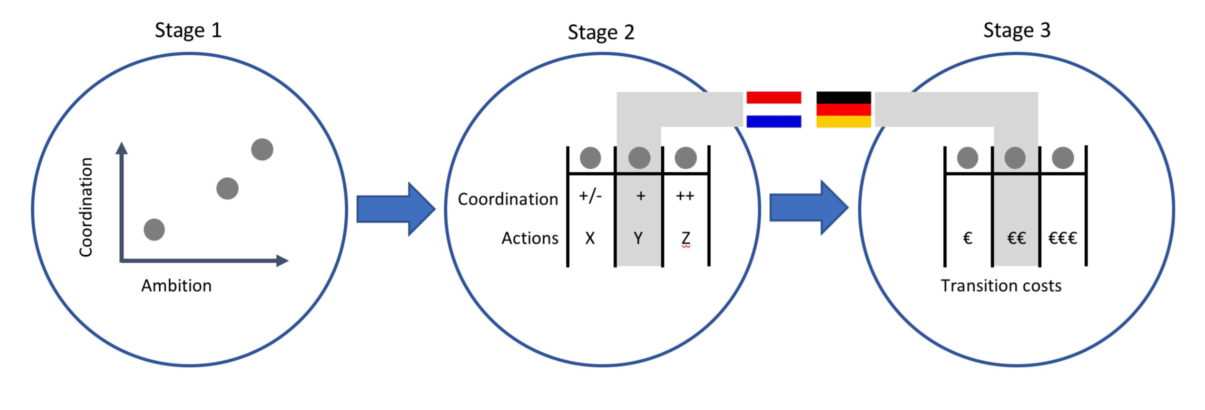
In the second stage the necessary open science actions from the previous chapter and corresponding coordination are incorporated in each of the scenarios. We have identified the following scenarios:

- Business as usual,
- In line with the rest of the world,
- Ahead of the rest of the world.

In the third stage, we focus on the transitional costs associated with the necessary actions and coordination. Moreover, as explained in the introduction, we used the NFDI programme as a reference for the costs on FAIR data. We used an extrapolation of this programme for the three scenarios that are defined below. The NFDI programme was however not used as a reference for open access. Nevertheless, we indicated the costs on open access based on what is currently spent by the VSNU and what has been spent on open access by NWO in recent years. However, the expenditure on open access also depends on strategic decisions that will be made in the future.

From a programmatic perspective, it will be explained how funding of transition costs for the next ten years can shift from an extra funding source to research performing organisations.

Figure 2 Defining scenarios in 3 stages



3.2 Scenario 'Business as usual'

Aim

Within the scenario of business as usual, there will not be a significant change in policy, coordination or actions for the next 4 years. This position will likely result in a moderate uptake of open access (educated guess: 70%), improvement on findability of data, but not so much on interoperability and reuse.

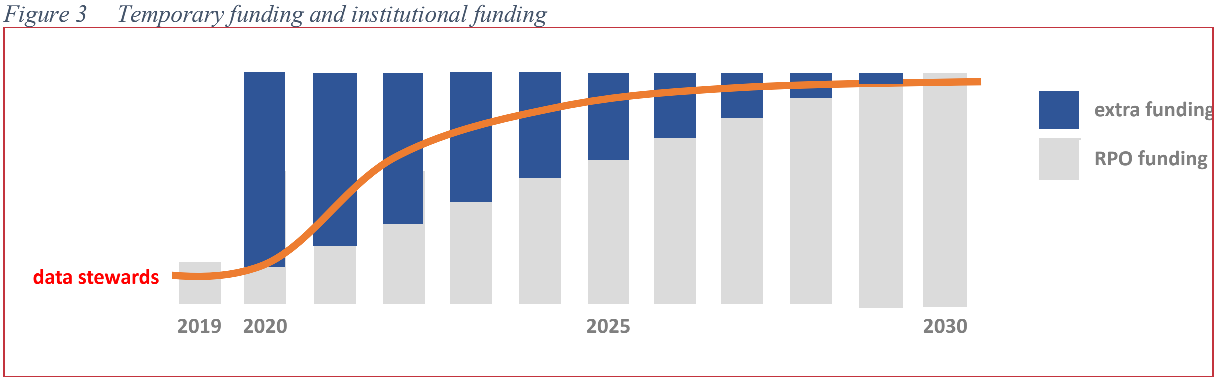
Actions

For 'Business as usual', the coordination consists between the NPOS partners through functional working groups such as the national coordination of research data management (LCRDM) and a coordination structure of the National Plan Open Science as we know it today. The actions addressed in this scenario to overcome the five challenges are an **impulse to data stewards (A(ction) 1 previous chapter)**, **sharing knowledge (A7)** and **campaigning (A14)**. To strengthen the data infrastructure a **national plan for data storage and services (A6)** can be helpful. The national plan mainly serves as a position paper and a checklist for research performing organisations. The value added of such a national plan is the making of it by a large group of experts with the function of a think tank. It will also help to clarify the Dutch position on closely related matters such as the European Open Science Cloud. In 'business as usual' the VSNU/UKB open access policy remains the same. Main actions for open access are to foster open access big deals and facilitate repository deposits with acceptable embargoes.

Costs

Additional funding of the national effort during the transition will not exceed the 2019 NPOS budget of 220.000 EUR for board and secretary positions and 400.000 EUR for NPOS national coordination projects. Funding of the action costs in scenario **'Business as usual'** is **5 to 10 million annually**. Most of this funding is necessary to accelerate growth of data stewards. At this point is it important to explain the financial mechanism for growth of data steward positions. In chapter 1 we defined transactional costs as necessary to accelerate or connect. Applied to data steward positions it is suggested not to increase OCW funding of research performing organisations but to fund data stewards for a maximum ten years. Temporary funding enables research performing organisations to create significantly more data steward positions on short notice now that there is an urgent demand for it. A longer funding period will help research performing organisations to phase out the temporary funding and bear the costs of the appointed data stewards from regular OCW funding.

Figure 3 Temporary funding and institutional funding



3.3 Scenario 'In line with the rest of the world'

Aim

In line with the rest of the world requires additional actions and a policy change. It is likely that open access to publications can be pushed towards 80% with funding to compensate open access publications in journals that are not covered by the VSNU open access big deals. The aim for research data in line with the rest of the world is to foster reusability of data with the prerequisite of findability addressed in the first scenario.

Actions

A more coordination is required to facilitate this scenario **(A8)** that introduces pilot project funding for data reuse and stimulation **funding for open access in humanities, social sciences and arts (A11)**. Funding open access in these disciplines seems legitimate in the context of the Plan S feedback. Besides, the coordination structure must be more demanding in order to coordinate all actions described in the previous scenario and additional actions. The first additional action suggested is to invest in **data managers (A2)**. The past few years experts identified data skills in different data positions for different domains. The policy domain requires data stewards, the tactical and strategic domain must have data managers and the research domain needs (cross)disciplinary data expertise and software (=data) engineers. To improve a national structure for sharing expertise on **data training (A15)** and **best practice (A7)** a **national centre** is suggested. Unlike the functional organisation of LCRDM the suggested national centre will be a legal entity. The value added by this national centre is the dedicated task to network the different data services (SURF, eScience centre, and others) and support the coordination structure. Project funding for calls is suggested to stimulate national collaboration (no competition) and to learn from innovative best practices with FAIR data. The suggested kind of projects have a dual purpose: they contribute to **best practices of data reuse (A10)** and they help **raising awareness (A14).** Initially, our field work included exploring concrete national projects with the aim to share best practices in data reusability. However, we found out that interviews can be inspirational for exploring projects[2] but it requires a different approach identify project to share best practices. A more common approach is to launch a call for project proposals. In such a call NPOS can clearly outline its intent that researchers need to address in their proposals.

The national plan for data infrastructure in this scenario is more than a position paper. It is also a national term of reference applicable to (consortium) procurement for commercial data services (to be adopted by VSNU, NFU, VH and others).

Costs

In a comparison of the Dutch scenarios with the German NFDI programme a match was found with the scenario "In line with the rest of the world". The match on actions and central coordination is a crucial step to identify the transition costs. In the methodology paragraph we mentioned the economic approach to downscale the German NFDI programme funding for transitional costs to the Dutch situation. The total annual funding in the NFDI programme is 85 million EUR, for direct project costs 70 million and for indirect costs 15 million (22%). Here we do not follow the NFDI program. For the Dutch scenarios we replace the generic 22% cost share with more accurate budget information of the current nationally funded coordination of NPOS. The current coordination budget is the one that is enough for a scenario 'Business as usual'. To determine the nationally funded coordination for the two other scenarios we

[2] Some generic descriptions of projects that came up during the interviews:

- a project aimed at creating a national thematic data cloud of high societal relevance, for instance the combination of medical data, environmental data and social data,
- a project aimed at researching added value of software engineering to disciplines that do not yet use such expertise,
- an empirical study about alternative research evaluation.

apply linear extrapolation. On open access we could not follow the NFDI, because this programme has a focus on data infrastructure and not publications. We have used information of past funding by VSNU and NWO on open access to calculate transition costs.

The transition costs of scenario **'In line with the rest of the world'** range between **10 to 15 million annually**, open access and coordination excluded. To fund a national scale project for best practices in reuse of data require between 3 and 5 million EUR. In this scenario we suggest an annual call for proposals to fund 3 projects, initiated by researchers in consortia with other stakeholders. The costs to compensate open access in disciplines that cannot benefit from the VSNU open access big deals range between 1 and 2 million. Coordination costs are estimated at 1 million.

3.4 Scenario 'Ahead of the rest of the world'

Aim
Most ambition can be found in a scenario we describe as 'Ahead of the rest of the world'. In this scenario, data interoperability will receive significant impulses in all disciplines. Open access publications will be pushed towards 90%.

Actions
Coordination in the scenario 'Ahead of the rest of the world' requires a programme organisation. In many countries a research funder, experienced with project calls, has responsibility for this type of programme management. Unlike the coordination in the previous scenarios, we suggest not to have NPOS managing the project funding but NWO.

On top of funding for data stewards and data managers we suggest investing **in software engineers (A3)**. Software engineers can make a significant contribution to (cross)disciplinary teams. Big personal research data can benefit from the expertise of software engineers.

It has been recognised that research evaluation is focussed on output that does not include data the same way it does for instance publications. We suggest promoting projects aimed at introducing FAIR data efforts in criteria for research evaluation **(A4 and A5)**. Existing interdisciplinary projects or new projects can both be applicable for funding as long as they contribute to making a case for the added value of data reuse and interoperability.

In the previous scenario we suggested a national open science expertise centre as a legal entity with the task to network and to support the coordination structure. In the scenario 'Ahead of the rest of the world' we additionally suggest establishing an **open data institute**. As a *research* institute it will be complementary to organisations such as SURF, eScience centre, and others.

Pushing open access towards 90% likely requires a change in the VSNU policy. Since 2015, the VSNU open access aims at immediate open access to be negotiated with publishers. We suggest opening the policy by introducing project **'seed' funding** to promote **alternative open access venues** for publications and books. Seed funding is aimed at promoting start-up open access venues. A similar approach has been adopted by France in a strategy to support more biblio-diversity. The French mandate requires cost reduction on subscription big deals with publishers, and the saved money being reinvested in alternative open access venues.

Costs
The total cost of scenario **'Ahead of the rest of the world'** is **19 – 37 million annually**. Table 3 contains an overview of costs associated with coordination, data and open access for the three scenarios.

Table 3 Scenarios transition costs 2020 – 2023 (4 years)

Scenario: ***Target:***	***Business as usual*** ***OA: ~70%*** ***data: findable and accessible***	***In line with the rest of the world*** ***OA: ~80%*** ***data: + reusable***	***Ahead of the rest of the world*** ***OA: ~90%*** ***Data: + interoperable***
FAIR data	data stewards	+ data managers + national expertise centre	+ software engineers + OS research institute
Data infrastructure	national plan data services	+ terms of reference	+ thematic data solutions
Awareness	campaign plan	+ reuse pilot projects	+ rewards pilot projects
Citizen science	bottom-up initiatives		
Subtotal annual costs open science	5 – 10 million	10 – 15 million	15 – 30 million
Open access	more open access big deals	+ APC compensation + B(ooks)PC compensation	+ alternative venues
Subtotal annual costs open access	budget neutral	1 – 2 million	2 – 5 million
Coordination	NPOS 2019	NPOS +	NPOS + and NWO
Subtotal annual costs coordination	0,7 million	1 million	2 million
Total annual transition costs	**5,7 – 10,7 million**	**12 – 18 million**	**19 – 37 million**

Robert Consultancy & Technopolis Group

We end this chapter with two conclusions. Our first conclusion concerns the annual transitional costs compared to the current public sector spend on open science (chapter 2). Annual transition costs in scenario 1 'business as usual' equals 3% to 6%. In scenario 2 'in line with the rest of the world' between 7% and 9% and in scenario 3 'ahead of the rest of the world' between 10% and 20%. In terms of temporary additional funding these relative shares seem 'reasonable'[3].

The second conclusion is about the duration of transition funding. The study objective accordingly we have calculated transition costs for 4 years between 2020 and 2023. The actions in each of the scenarios require 4 years of funding at the level mentioned in table 3. A considerable amount of actions needs longer term funding. For instance, funding to accelerate the growth in data stewards can only be scaled down over more than 4 years (see figure 4). We will come back to this point in the next chapter with recommendations.

[3] In order to present clear relative figures we replaced the bandwidth of public research sector spend (chapter 2) with a median spend, that is 187,9 mln. EUR.

4 Recommendations

In this study it was important to balance between the broad spectrum of open science and the level of thoroughness on each of the transition challenges. We learned that both open access and FAIR data have a prominent role in the transition to open science. At the same time, it has also become clear to us that both transitions have a very different character with different types of challenges and investments. FAIR data is a new way of reusing and managing data that requires a new way of working, new skills and related investments. In the case of the transition to open access, the efforts and transaction costs depend on the chosen ambition level.

Open access is a delicate topic given the complexity of publisher negotiations and the pressure to make a swift transition envisioned by PlanS. The next significant Dutch growth in open access depends on the current and future negotiations with the largest publisher. For this study we used the current VSNU open access strategy to develop the scenarios. The same strategy has been applied by other European countries with mandates that allow for a no-deal if a publisher refuses to comply. We recommend considering reinforcing the Dutch mandate the same way.

Citizen science has not yet been included in most policies, but there are indications it is gaining ground. We recommend conducting further study to prepare for citizen science.

We systematically constructed 3 scenarios. Each scenario has a unique set of actions and associated annual transition costs. Although the study distinguished between the upcoming 4 years and thereafter, we concluded in chapter 3 that annual transition costs of many actions remain also after 4 years. We recommend commitment to fund annual transitional costs for a period of 10 years. Furthermore, we recommend to fund during this period at the same level as foreseen in 2020 – 2023. We do note that priorities after 4 years may need reconsideration. From another perspective it is important to commit to 10-year funding, because there is evidence that short term project funding does not bring sustainable solutions. It is recommended to introduce a midterm review of the advised approach. This midterm review must be completed within the fourth year of the programme funding.

It is important to sense the most appropriate scenario after the fieldwork interviews, the economic modelling and the lessons learned internationally. We sense that it would fit the Netherlands to pursue an ambitious scenario that cannot be found in scenario ‘business as usual’.

Typically for this kind of study the number of scenarios and their composition can be questioned. However, questioning the scenarios must begin with a shared ambition and policy aims. It is thus important that all stakeholders agree on these ambitions and policy aims.

Within each of the scenarios, we recommend prioritising the suggested actions. Prioritising growth of data stewards, data managers or software engineers and determining associated costs depends on a more detailed approach than the one provided in this study. It is therefore relevant to map these positions and associated costs in more detail.

The German NFDI programme served as an important best practice for FAIR data and as a useful source of experience with large scale data programmes for this study. We recommend continuing extensive exchange of knowledge and experiences.

In this report we calculated the current annual open science costs and the transitional costs in 3 scenarios. Another important research question about the long term costs for open science yet remains unanswered. It is important to note that this question cannot be answered precisely as long as it is uncertain what strategic decisions research performing organisations will make. In the meantime, the best answer to this question is a reference to the study of the Commission High-Level Expert Group on the European Open Science Cloud (2016) that argues that on average about 5% of research expenditure should be spent on FAIR data.

Appendices A List of interviewees

Table 4 Interviews conducted during this study

Name	Institution
Alastair Dunning	4TU, Delft University of Technology
Anne Besse-Lototskava	Wageningen University & Research
Arjan Schalken	Vrije Universiteit Amsterdam
Barend Mons	Leiden University Medical Center and GOFAIR
Celia van Gelder	Dutch Techcentre for Life Sciences (DTL)
Erik Fledderus	SURF
Frank Allgoewer	University of Stuttgart
Frank Miedema	University Medical Center Utrecht
Frank Seinstra	eScience
Franke van der Molen	Radboud University
Hans de Jonge	Netherlands Organisation for Scientific Research (NWO)
Huib de Jong	Amsterdam University op Applied Sciences (HvA)
Huib Pols	Netherlands Organisation for Health Research and Development (ZonMw)
Ingeborg Verheul	National Coordination Research Data Management (LCRDM)
Ingrid Dillo	Data Archiving and Networked Services (DANS)
Jan-Willem Boiten	Netherlands Federation University Medical Centres (NFU) and Health RI
Jeroen Sondervan	Utrecht University
Karen Maex	University of Amsterdam
Margreet Bloemers	Netherlands Organisation for Health Research and Development (ZonMw)
Melanie Imming	SURF
Melle de Vries	Royal Netherlands Academy of Arts and Sciences (KNAW)
Patrick Schelvis	Ministry of Economic Affairs and Climate Policy (EZ)
Ron Dekker	CESSDA
Ronald Stolk	University of Groningen (RUG)
Ruben Kok	Dutch Techcentre for Life Sciences DTL and Health RI
Thom Palstra	Dutch University Association (VSNU) and University of Twente
Wilma van Wezenbeek	Delft University of Technology

Robert Consultancy & Technopolis Group

Appendices B Factsheet NFDI

An important reference for this study is the German national research data infrastructure (NFDI). The NFDI, established late 2018, provides a structure (collaboration between the national research funder DFG and the Joint Science Conference GWK) and a funding programme. The DFG website provides comprehensive information about NFDI (https://www.dfg.de/en/research_funding/programmes/nfdi). Below a selection of facts.

Aim

- Systematically manage scientific and research data, provide long-term data storage, backup and accessibility, and network the data both nationally and internationally.
- Bring multiple stakeholders together in a coordinated network of consortia tasked with providing science-driven data services to research communities.

Purpose

- Establishment of data handling standards, procedures and guidelines in close collaboration with the community of interest.
- Development of cross-disciplinary metadata standards.
- Development of reliable and interoperable data management measures and services tailored to the needs of the community of interest.
- Increased reusability of existing data, also beyond subject boundaries.
- Improved networking and collaboration with partners outside the German academic research system with expertise in research data management.
- Involvement in developing and establishing generic, cross-consortia services and standards in research data management together with other consortia.

Type and Extent of Funding

- The German federal and state governments envisage funding up to 30 consortia.
- A total of up to €85 million is available per year to fund the consortia in the final development phase; this amount includes a 22 percent programme allowance for indirect project costs.
- The amount available to fund direct project costs thus totals approximately €70 million annually.

As a rule, an individual consortium may receive between €2 million and €5 million, which includes the programme allowance for indirect project costs and €1.6 million to €3.9 million for direct project costs.

Appendices C References

LCRDM (2019) - Data stewardship on the map: A study of tasks and roles in Dutch research institutes

NFDI (2019) - website infromation, https://www.dfg.de/en/research_funding/programmes/nfdi

RfII (2016) - German Council for Scientific Information Infrastructures: Enhancing Research Data Management: Performance through Diversity. Recommendations regarding structures, processes, and financing for research data management in Germany

RfII (2017) - German Council for Scientific Information Infrastructures: An International Comparison of the Development of Research Data Infrastructure. Report and suggestions

HLG EOSC (2016) Realising the European Open Science Cloud: First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud

Rathenau Institute (2019) - Factsheet: Open access van wetenschappelijke publicaties

CWTS, Elsevier and University Leiden (2019) - Open data: The researcher perspective

KNAW (2018) - Big data in wetenschappelijk onderzoek met gegevens over personen