technopolis group

March 2021

Harmonizing outcome and impact indicators for TAFTIE/SNB members

Conclusions report

www.technopolis-group.com

Version 1

March 2021

Harmonizing outcome and impact indicators for TAFTIE/SNB members

Conclusions report

FFG (Austria): Rafael Lata

CDTI (Spain): Ascension Barajas Inigo

HAMAG-BICRO (Croatia): Ivona Jerković and Ivana Crnić-Duplančić

EAS (Estonia): Tarmo Puolokainen and Tõnis Tänav

Innosuisse (Switzerland): Adrian Berwert

Innoviris (Brussels): Katrien Mondt

Luxinnovation (Luxemburg): Philippe Mayer

RVO (The Netherlands): Arjan Wolters

SIEA (Slovakia): Peter Adamovský

TA CR (Czech Republic): Petr Horák

ANI (Portugal): João Ferreira

NEDO (Japan): Shumpei Miyajima and Motoshi Kunugi

VLAIO (Flanders): Luc De Buyser

Innovation Norway (Norway): Kerstin Solholm

Facilitated by: Technopolis Group, Matthias Ploeg and Pieter Heringa, with research support from Lennart Velten. Correspondence: Matthias.ploeg@technopolis-group.com

Table of Contents

Tables

Figures

1 Introduction

1.1 Measuring results of innovation policy at TAFTIE

Innovation Policy remains highly in flux, with high expectations from both policy makers and the general public. Innovation is deemed a critical driver of not just productivity, jobs and economic growth, but increasingly also as society's main tool for addressing key societal challenges such as climate change (e.g. clean-tech), inequality (e.g. social and frugal innovations) and affordable high quality healthcare (e.g. ambient assisted living). Over the last few years, the strategic component of national/European capabilities in strategic technologies (e.g. AI, 5G, IoT) further ensures that innovation remains a core policy priority.
However, with such high opportunities come equally high expectations, and as such need for monitoring. Monitoring (and subsequent evaluation) serves not just the traditional use of post-hoc accountability of public spending, but increasingly also for strategic steering of funding and activities. As the pace of innovation activity increases, timelier and frequent indicators and evidence become more critical, in particular to measure short-lived processes, such as entrepreneurship and business dynamics[1]. Failing to capture timely indicators and evidence risks informing policy decisions with information that is no longer relevant. Former EU Commissioner for Research, Science and Innovation, Carlos Moedas, worried that most of the data he used to take policy decisions was outdated[2]. In addition, policy makers increasingly seek insight into impact beyond 'economic' pathways, in particular in line with an increasing mainstreaming of the SDGs).
However, providing robust and relevant insight into the performance of innovation instruments is a challenging affair for a number of reasons. First, innovation processes are inherently non-linear, often diffuse beyond the exact borders of individual companies, causing difficulty in precise attribution. Secondly, impacts of innovation are often highly skewed, with only a small number of companies reaching high economic and/or societal success from the innovations, making it difficult or less relevant to think of 'average impact'. Finally, data on innovation programmes' outcomes is often difficult and/or expensive to collect, and publicly available data through statistical agencies might not always have the right granularity, indicator coverage or frequency for use in monitoring of policy programmes.

Fortunately, the last 2 decades has seen substantial improvements across the board by innovation agencies, scholars and evaluators in improving the ability to assess the effectiveness, efficiency and impact of innovation programmes. TAFTIE, as the premier association covering the large majority of leading European innovation agencies, took a 'thought and practice' leadership role in establishing the structural network for benchmarking (SNB). Benchmarking is highly useful since many potential indicators on effectiveness, efficiency or impact are difficult to value on their own: e.g. is a 2 EUR per invested EUR 'good' or just 'average'? Benchmarking can help provide relevant contextual information for such interpretations, but also, perhaps even more importantly, provide a lot of opportunity to learn and share lessons on successful strategies to improve efficiency.

[1] OECD (2018), OECD Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption, OECD Publishing, Paris, https://doi.org/10.1787/sti_in_outlook-2018-en.

[2] Moedas, C. (2016), "What new models and tools for measuring science and innovation impact?", keynote speech provided at OECD Blue Sky Forum, Ghent, https://ec.europa.eu/commission/commissioners/2014-2019/moedas/announcements/whatnewmodels-and-tools-measuring-science-and-innovation-impact_en

Interestingly, SNB (then still known as the task force TFBIEE) found that benchmarking on effectiveness, efficiency and impact has a lot of prerequisites:

- A similar understanding of the 'archetypical' impact routes (theories of change) of innovation instruments, *so we know what to compare*.
- Agreement on indicators to measure activities, outputs, outcomes and impacts and their precise operationalisation, underlying definitions as well as analytical evaluation techniques (such as counterfactual comparison), *in order to avoid comparing apples and oranges.*
- Where possible, the techniques, operationalisations and definitions should align with best practices and/or international standards, *in order to promote broader comparability and consistency*.

In pursuing the first benchmarking report, SNB also produced, with support from Technopolis Group, a number of key documents:

- **A Reference Model on Indicator Selection,** harmonizing archetypical logical framework for four typical innovation instruments, relevant priority effectiveness and efficiency indicators, in addition to a broader strategy for indicator quality measurement[3].
- **A Reference Model for Evaluation[4],** providing guidance and harmonization in terms of best practises regarding evaluations of innovation instruments, including recommended analytical techniques in how to deal with attribution challenges.
- **A first common indicator framework,** providing definitions for input, activity and output indicators[5].
- **The First Benchmarking Report,** providing templates, frameworks and standardized definitions of key indicators[6].

A second benchmarking report was launched in 2018. With relatively modest updates to the approach, the SNB managed to produce a highly useful report. However, at the same time the SNB ran into a number of new definitions and harmonization challenges**.** Summarized in box 1.1. of the Second Benchmarking Report, these include different definitions of what can be considered a beneficiary, direct vs. indirect contributions from private sector, attribution to specific years etc. One clear lesson is that definitions need to be highly elaborated and precise, as the *devil is in the details*.

---

[3] https://taftie.eu/sites/default/files/Reference%20Model%20Final%20for%20Publication.pdf

[4] https://taftie.eu/sites/default/files/TAFTIE%20%20Evaluation%20Reference%20Model.pdf

[5] https://taftie.eu/sites/default/files/Common%20Indicator%20Framework%20for%20publication.pdf

[6] Public summary available here:
https://taftie.eu/sites/default/files/TAFTIE%20TFBIEE%20Management%20Summary.pdf

1.2 Project 'Measuring Impact and Outcomes'

The original methodology and framework described above only focuses on a primary set of indicators, input, activity and output indicators. These are relatively easy to harmonize as data collection and assessment are typically descriptive and within the domain of the agency themselves. However, comparing outcome and impact indicators also requires a harmonization of monitoring and evaluation methodologies, in particular, around attribution and counterfactual strategies. While the first iteration did provide guidance on best practices for evaluation, full harmonization was not achieved. Since outcome and impact indicators are highly relevant, a review was done of agency monitoring systems, the outcome of which were summarized in November 2019 Report of the SNB 'Monitoring systems in TAFTIE Agencies: outcome and impact indicators'[7]. The main conclusion is that there is a need to 'for building a set of indicators that could be used by most agencies and be included in the TAFTIE Benchmark report'.

A pilot exercise was then carried out without external support by the agencies HAMAG-BICRO & TA ČR (SNB coordinators 2019-2020), for several outcome and impact indicators in 2020 for collaborative R&D programmes.[8].

As such, an SNB project was launched for the period March 2020 – March 2021 to work on harmonization of outcome and impact indicators and related guidance. Technopolis was selected to carry out the project under the guidance of the participating agencies. In total 14 agencies agreed to be part of the project:

1. FFG (Austria)
2. CDTI (Spain)
3. HAMAG-BICRO (Croatia)
4. EAS (Estonia)
5. Innosuisse (Switzerland)
6. Innoviris (Brussels)
7. Luxinnovation (Luxemburg)
8. RVO (The Netherlands)[9]
9. SIEA (Slovakia)
10. TA CR (Czech Republic)
11. ANI (Portugal)
12. NEDO (Japan)
13. Innovation Norway (Norway)
14. VLAIO (Flanders)

More details on these agencies and their activities in the field of innovation policies can be found in report 1 and on the website of TAFTIE: https://taftie.eu/members

[7] https://taftie.eu/content/structural-network-benchmarking-conclusion-report-november-2019

[8] https://taftie.eu/sites/default/files/SNB2020%20Report_FINAL_0.pdf

[9] RVO has indicated to operate as a contributor to the study part on national statistical systems only.

1.3 Structure of the project and this report

The project was structured in three main phases, each working towards a discussion and validation workshop and a specific report:

- **1: Baseline assessment & prioritization of efforts**
  - Baseline of existing monitoring systems (and indicators)
  - Prioritization & selection of indicative list of indicators to be harmonized
- **2: Harmonization and technical specification of indicators**
  - Review of existing possibilities for use of national statistics
  - Proposed technical specification of indicators
- **3: Guidance for Agencies**
  - Guidelines & recommendations for adaptation of existing monitoring systems

This report is the final product of the project, in which the most important findings and conclusions across the earlier reports of all phases are compiled and summarized. It is structured as follows:

Chapter 2 gives an overview on existing monitoring systems at the agencies. Chapter 3 presents different models of using external data to monitor performance of innovation instruments and key benefits and challenges in doing so. Chapter 4 provides an overview of new developments in monitoring, including the potential of open data and new types of indicators, to capture for example the contribution to societal goals. Chapter 5 summarizes the selection criteria for indicators. In chapter 6 we introduce the set of indicators that we have proposed for this project.

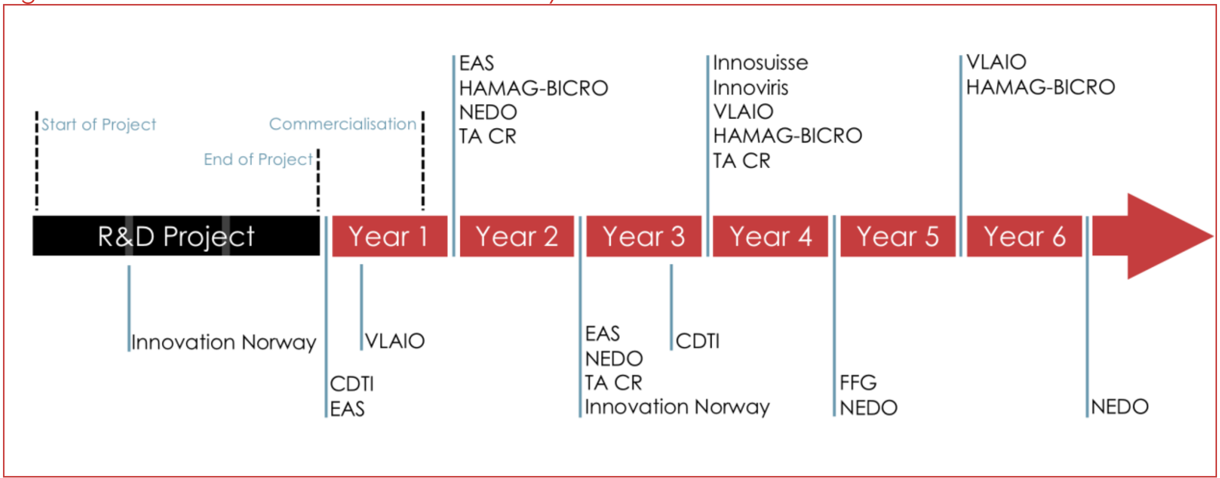
2 Analysis of existing monitoring systems

Before assessing the indicators currently in use and the possibilities to harmonize them, we have first analysed four basic aspects of the (design of) the existing monitoring systems[10]: beneficiaries, the monitoring process itself, the data collection, and the reporting process[11]. This is done in a comparative way, where the differences across the innovation agencies have been analysed. In addition, a cross-cutting analysis is made on differences between the four types of policy instruments: Business R&D grants, Collaborative R&D grants, Competence centres and clusters, and Innovation Vouchers programmes.

Most agencies have not indicated any differences in their monitoring systems between the instruments under consideration (R&D Grants, Cooperative Grants, Innovation Vouchers, Competence Center Schemes), although several agencies measure different indicators for different instruments. SIEA is the only agency that has different approaches for different instruments. This is largely explained by the demands from third parties to the instruments. This leads to differences in indicators measured, but also for example in report cycles and use of surveys.

Most agencies have fixed moments of data gathering, generally related to the finalisation of a project; figure 1 shows the differences in timing and frequency of measurement.

Figure 1 Timeframe of data collection in surveys[12].



Aspects of the monitoring systems per agency are shown in table 1.

[10] Throughout this report, 'monitoring' is understood as all forms of regular data collection on the characteristics of (potential) beneficiaries of subsidy instruments, or the result of those instruments. 'Evaluation' is understood as reviewing an instrument based on global criteria (effectiveness, efficiency), based on a variety of data sources (partly coming from monitoring efforts), and to establish a direct causality between intervention and its results. In practice, the distinction between the two can be blurry. It is mentioned specifically if part of the data collection for example is only done in the context of (ad hoc) evaluations.

[11] The structure of this chapter builds on very useful earlier work by TAFTIE, including *Monitoring systems in TAFTIE Agencies: outcome and impact indicators Conclusions Report* from November 2019. The report also contains more background information on the monitoring system of several agencies. https://www.taftie.org/sites/default/files/ConclusionsReport_SNB19_B_0.pdf

[12] Note that Commercialisation often happens often much later after closure of the project, depending on the technology, sector and context.

Table 1 Aspects of monitoring system per agency.

	Monitoring process (continuous or occasional)	**Baseline**	**Number of Surveys**	**Survey timing (years after project closure)**	**Surveys?**	**Other data sources?**	**Data collection**
Agentschap Innoveren & Ondernemen / Vlaanderen is ondernemen	Both	No	2 per project + additional occasional	Continuous: 0,25 Occasional: 3-5	Yes	Yes	Inhouse and external
FFG Promoting Innovation.	Continuous	No	1	4	Yes	No	External, inhouse support
CDTI	Continuous	No	2	Right after completion of the technical phase + 2 y after commercialisation date	Yes	Yes	Inhouse
HAMAG BICRO	Continuous	Yes	2 - 3	1, 3 and/or 5	Yes	No	Inhouse and external
EAS Enterprise Estonia	Both	No	1 - 3	Immediately and 2 annual surveys	No	Yes	Inhouse
Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation Innosuisse – Swiss Innovation Agency	Both	No	2	Continuous: Project Completion, then 3 years, occasionally 4-6y	Yes	No	Both inhouse and external
innoviris.brussels empowering research	Occasional	No	1	3	Yes	No	Inhouse
LUXINNOVATION Trusted partner for business	Occasional	No	N/A	N/A	Yes	Yes	Inhouse
SIEA Slovenská inovačná a energetická agentúra	Both	Varies[13]	Varies	Varies	Varies	Varies	Varies
TA ČR	Occasional	No	3	1, 2 and 3[14]	No	No	Inhouse
ANI Agência Nacional de Inovação	Continuous	No	N/A	N/A	N/A	No	Inhouse
NEDO	Continuous	No	4	1, 2, 4 and 6	Yes	No	Inhouse
Innovation Norway	Both	No	2	1y and 4y after subsidy	Yes	Yes	External

[13] For SIEA there are differences between different innovation instruments for all of these aspects.

[14] TA CR is now in process of change of its outcome monitoring, from 3 reports (1,2 and 3) it will be changed into 1 report. The timing of this report for each project will be set by beneficiaries.

The analysis of the monitoring systems results in an interesting mix of designs and practices. While some agencies are obviously more experienced in monitoring outcome and impact indicators than others, the analysis shows that there is no clear cut ‘order’ of agencies in the quality or extensiveness of the monitoring systems. Rather, it seems that many agencies stand out in one aspect (e.g. using external data) and still work on setting up practices in another (e.g. integral databases). There are hence ample opportunities to learn from each other. Several agencies have also indicated that they are still in a phase of building up (parts of) the monitoring system.

On the other hand, there are also challenges that many agencies have in common. Most agencies for example do not measure a baseline (yet) when introducing a new instrument or adapting an existing one. The policy making process often does not allow for the time needed to measure a baseline. As a result, the construction of a proper counterfactual also becomes more complicated, and it is hard to attribute the measure outcomes and impact to a specific instrument. Having a good baseline ‘built in’ in the regular data collection system allows for the monitoring of more advanced indicators (allowing to adapt instruments ex durante), and is also very beneficial for evaluations that aim to measure a causal effect.

Another common challenge is that many agencies cannot develop or adopt their monitoring system in isolation. They are bound by strict external requirements (e.g. if an instrument is strongly linked to European instruments like the ESIF funds, that come with their own rules), because they have to align with other national agencies or because of data protection and privacy issues (such as the GDPR). This has to be taken into account when developing recommendations for harmonised indicators.

3 Use of external data

The use of data centrally collected by a provider (such as national statistics agencies or private companies), is a relevant strategy for monitoring innovation instrument performance, as well as a useful tool for data harmonization, as these data sets are typically built on European standard definitions. However, for such sources adequate coverage, favourable access and sufficient frequency are all conditions that need to be in place. Many agencies have experience already with the use of such large external datasets. In this chapter, we present an overview of the experiences that the agencies have in accessing and using microdata from national statistics offices and other (commercial) data providers.

3.1 Overview of access to microdata at national statistics offices

All national statistics offices offer high-level statistical data for public access. Statistical microdata, where individual companies can be identified either directly or indirectly, is (in Europe) provided only for research purposes and to research organisations. An exclusion to this is public microdata, which represents a subset for the public to familiarize themselves with microdata (available are the labour force survey (LFS) and the EU statistics on income and living conditions (EU-SILC)). The perceived strictness and exact definition of a research organisation varies by country. A list of all national statistical offices in Europe can be accessed here.

The protection of data collected for statistical purposes – "statistical confidentiality" - is considered a fundamental principle of official statistics. Statistical confidentiality means that data on individual persons (or business entities) may be used only for statistical purposes and that rules and measures shall be applied to prevent the disclosure of information concerning an individual person or business entity. Further information on the formal rules in Europe is available here.

The agencies have different experiences in accessing the microdata of national statistics offices (NSOs) as shown in Table 2 below.

Almost all of them are interested in having access to microdata, and most also have internal capacity to process such data, but only a few are currently granted access.[15] Those who have doubts indicate that it is unclear to what extent the data for impact monitoring would be available in sufficient up-to-date quality, or they lack internal capacity at the moment to make this work. The key to getting access is that the use of the (linked) data has a research purpose (see also the case studies on RVO and EAS). SIEA is recently recognized as research entity and expects to get access soon at Eurostat. Interestingly, about half of the agencies are able to link microdata from sources, such as data from annual accounts to microdata from the national statistics office (or can get this done by an external body). Four agencies do not know if this is an option. If agencies don't know, this is either due to time/capacity constraints, or because they don't know if they can meet the technical specifications and requirements (because they do not have access or because external parties carry out the impact monitoring). About half of the agencies consider data protection as the most important factor that hinders the use of data of the NSO. Other factors are a lack of internal capacity and the lengthy procedures, for example where a new application is needed for every analysis.

Table 2   Overview of experiences with National Statistics Offices

	Access to microlevel firm performance data from NSO?	If not, would you like to have access?	Linking other data to microdata from NSO?	Key factors that hinder use of microdata of NSO	Internal capacity to process and analyze microdata?
Agentschap Innoveren & Ondernemen Vlaanderen					
FFG Promoting Innovation.				Data protection	
CDTI				Data protection	
HAMAG BICRO				Lengthy procedures	
EAS Enterprise Estonia				Limited/poor quality data	
Schweizerische Eidgenossenschaft / Confédération suisse / Confederazione Svizzera / Confederaziun svizra / Swiss Confederation / Innosuisse – Swiss Innovation Agency				Other	
innoviris.brussels empowering research				Data protection	

[15] Only Innoviris states that the data from its ex-post survey is currently enough to cover its needs.

LUXINNOVATION TRUSTED PARTNER FOR BUSINESS				Other	
SIEA SLOVENSKÁ INOVAČNÁ A ENERGETICKÁ AGENTÚRA				Data protection	
TA ČR				Data protection	
ANI AGÊNCIA NACIONAL DE INOVAÇÃO					
NEDO				Lengthy procedures	
Rijksdienst voor Ondernemend Nederland					
Innovation Norway				Data protection	

The procedures for the agencies that work (or used to work) with the microlevel data are also quite similar. After access is granted (based on the criteria for use of data for scientific/research purposes), pseudonymized data becomes available, either through a secured virtual connection or by visiting a secure place at the premises of the NSO. If data is to be linked, the agency (or the external party that carries out the work) uploads the data to be linked to the NSO; it receives back the pseudonymized, linked data. There are always confidentiality agreements in place, and results can never be reported in such a way that they can be traced back to individual firms or organisations (see the case studies for more details).

To give more insight in any differences between the national statistics offices, we present an overview of the policies that the offices have in giving access, based on publicly available information on the websites of the NSOs, see Table 3. This confirms that in general, the NSOs have similar rules, although there are some differences in for example the application process and the cost of getting access to microdata. Eurostat uses three criteria to determine if an organization qualifies for access: the organization should have research as one of its main activities, or be a research department within a broader organization, it should provide evidence of publication of research results, and it should be independent and autonomous in formulating scientific conclusions. Some national offices only mention on their website that applicants should be 'scientific organisations'.

Table 3 Overview of requirements for access to microdata per National Statistics Office.

National statistics office	**Requirements for and organisations with access to microdata**	**Application process**	**Cost**	**Access to microdata (Remote or safe centre)**
STATBEL (Belgium) & National Bank	Public & research institutions, and entities with 'legitimate purpose' have access for scientific/statistical research.	Per project (Online data request)	N.A.	N.A.
Statistik Austria (Austria)	Scientific organisations have access to standardised	Initial application; per project (research proposal	Standardized data sets free of charge, preparation of tailored Task-	N.A. (Level of anonymisation depends on the data protection

National statistics office	Requirements for and organisations with access to microdata	Application process	Cost	Access to microdata (Remote or safe centre)
	datasets and specific datasets (after proposal)	for task specific data)	specific data sets incur costs.	measures that the researcher can put in place)
Instituto Nacional de Estadística (Spain)	Public access to anonymised microdata (high level, not comparable to other countries). Scientific institutions can request more detailed data for research.	Per project (10day response time)	Price list available online in Spanish. Cost depends on type of data.	N.A.
Hungarian Central Statistical Office (Hungary)	Anonymised microdata sets available granted for approved research projects that meet all researcher accreditation criteria (scientific purposes at research organisations).	Per project (Data request)	Preparation of anonymised microdata charged with 60.000 HUF/day + VAT. Research on standard datasets is free of charge, (incl. output checking)	Safe Centre
Eesti Statistika (Estonia)	Legal persons and organisations can use for research microdata held by Statistics Estonia	Per project (excel form, time for review 15 days)	Use of microdata in a safe centre is free of charge. Only charge for preparing the data (€16-55/h).	Safe Centre; remote access (depending on data and contract)
Federal statistical office (Switzerland)	Anonymised personal microdata is only passed on for statistical, research and planning purposes (not for administrative, auditing, fiscal or supervision purposes).	Per project (online form)	N.A	N.A.
Service Central de la Statistique et des Etudes Economiques (Luxembourg)	N.A.	N.A.	N.A.	N.A.
Statistical office of the slovak republic (Slovakia)	Provision of the anonymised microdata sets for scientific purposes follows instructions published on the office's website (Services/Information; not available).	Per project (Research proposal)	Price list available online. Staff rates are set at €15,30/hour.	N.A.
Czech Statistical Office (Czech Republic)	Proof of the scientific activity of the ordering party, a research project	Per project (request form)	Services charged based on online price list (e.g., access to safe	N.A.

National statistics office	Requirements for and organisations with access to microdata	Application process	Cost	Access to microdata (Remote or safe centre)
	description and a pledge of secrecy is required of parties wishing to access microdata.		centre (500CZK initially) or data handling (400CZK/hour.	
Instituto Nacional de Estadística (Portugal)	Statistical data on individuals and enterprises is only be supplied for scientific purposes, in anonymised form to accredited researchers. The accreditation is provided by the Directorate-General for Education and Science Statistics.	Per project (Accreditation application form; Declaration of commitment; Notice of GDPR consent)	N.A. in detail, free provision of services under certain conditions.	Some data only available for access in a safe centre environment.
Statistics Bureau of Japan (Japan)	Limited information. Microdata is made available for: "academic and research purposes"	N.A.	N.A.	Safe centre (encouraged); remote access without the ability to download data.
Statistisk sentralbyå (Norway)	Microdata available to research organisations listed on the offices list of approved organisations. A new institution can apply to be put on that list. Otherwise, it is possible to apply for data based on the origin of the project (e.g., financed by the Norwegian Research Council).	Per project (online available)	N.A.	N.A.
Centraal Bureau voor de Statistiek (The Netherlands)	Access to microdata is granted to authorised institutions with statistical or scientific research as their main activity.	Authorisation applied for five years, data requested online.	Pricing depending on service. Service catalogue online	N.A.
Eurostat (EU)	Microdata is available to institutions listed as recognised entities on the Eurostat website (criteria: scientific activity, data protection measures, publication of research results). Most of these organisations are universities	Authorisation for entry to the list. Research proposals required per project (turnaround time 8 weeks)	N.A.	Safe centre; remotely (only partially anonymised data)

Case summary – Data collaboration between RVO (Netherlands Enterprise Agency) and CBS (Central Bureau of Statistics the Netherlands).

RVO collaborates with the ministry of Economic Affairs and Climate Policy and the national statistics office CBS to develop dashboards and other instruments to monitor and evaluate innovation policy. For this, access to data is limited to specialised teams within RVO and the ministry. Datasets are developed based on current policy questions (e.g., characteristics of firms eligible for potential Covid-19 policy measures).

Variables and data quality

Data is shared between CBS and RVO on a broad basis. RVO contributes by delivering microdata from policy schemes on company level. On its turn, CBS is able to link that data to statistics such as sector classification, size, age or solvability and liquidity of the firm. The resulting data is of high quality, even more so as other authorities (e.g., national tax office) contribute.

Use of data and lessons learned

Data is used based on the policy question at hand, create reference groups or develop an taxonomy of companies (e.g., SMEs). Through the collaboration RVO has a flexible setup at hand that allows for the expansion of available data and possible scientific use. Data protection has emerged as a hurdle of significant importance during the collaboration.

More details on this case study can be found in the report on use of external data.

Case summary – VLAIO (Flanders Agency for Innovation and Entrepreneurship) and the use of CIS (Community Innovation Survey) data

VLAIO uses several external data sources, including information from annual accounts, data from the CIS survey and data from the R&D survey. ECOOM collects and links a large part of the external data.

Variables and data quality

While the CIS offers more data, VLAIO mainly uses firm data on new product or process innovations, and the share of market novelties in the total sales of a firm (proxy for the degree of its innovativeness and success of product innovations on the market). With the larger dataset, it is possible to construct control groups for VLAIO beneficiaries.

Use of data and lessons learned

It is only since two years that the system has the current setup. The main challenge for the future is to expand it further, more systematically and ideally also refine it. More specific data with a higher response rate would allow for deeper analysis what specific measures work for what target groups etc.

More details on this case study can be found in the report on use of external data.

3.1.1 Reflections and discussion on the use of microdata at statistics offices

There are clear benefits in using data from national statistics offices as a strategy to monitor innovation instrument performance. The data is usually well defined and harmonized, as the data sets are typically built on European standard definitions. The use is especially attractive if the data can be linked at firm level to specific data that is collected among the beneficiaries of the innovation instrument.

Several lessons can be drawn when considering the harmonization of impact and outcome indicators. The data from NSOs is only to a limited extent used for monitoring. Its main use is in enabling evaluations, for example by generating systematic information on non-users. If it is used for monitoring, this is usually on general indicators for firm performance, not on R&D indicators. It might be that other agencies do use external datasets that are focused more on such indicators (for example patent data). The time lag before the (processed) data becomes available for use at the NSO, also might make this kind of data less attractive for (continuous) monitoring.

The systems described are generally rather advanced and have taken years to build up and evolve. This is a long-term process that is not always easy to imitate. The hurdles are not so much in the technical availability of data, but rather in cleaning and properly linking the datasets (especially for larger firms with complicated business structures). A major hurdle is that access is needed to the (pseudonymized) microlevel data of the NSOs, and they should allow for linking datasets, e.g., based on company number. It seems that although the NSOs use similar rules, there are differences in whether or not such use is considered to be part of the research purposes for which use of microlevel data is allowed. Even if the interpretation allows for the use of microdata, trust needs to be built between the exchanging organisations to collaborate efficiently which takes time and effort. On top of that, it needs to be made sure that linked datasets are compliant with data protection measures and for example tax secrecy, etc.

It is also important to strike a proper balance between minimizing the administrative burden and measuring the indicators in a way that is relevant for all stakeholders involved. In principle, the use of data of NSOs can help in decreasing the administrative burden for firms. However, the advanced systems often involve many actors, all with their own goals and ambitions. Linking datasets can help to decrease the administrative burden, but it is often time-consuming and complicated to build up (see above).

3.2 Using other external data providers

Apart from the use of microdata from national statistics offices, several agencies also use other external data sources, such as private data providers, sometimes in combination with data from NSOs. Depending on the source, this can also be standardized and harmonized data, as some providers have databases with world-wide coverage based on standardized definitions, often well aligned with for example OECD and European standard definitions.

Case summary – External data for outcome indicators at Innovation Norway

Every three or four years, Innovation Norway partly outsources the process of combining and analysing datasets as input for the development of outcome-indicators via public tender. The current contract was won by a private firm, yet the previous one by Statistics Norway.

Variables and data quality

Data on beneficiaries is collected by Innovation Norway and delivered to the party that analyses the data. In Norway, all limited liability firms are obliged to upload their data, which also includes information on value added, turnover, profit, number of employees and productivity. In addition to this data, there is also data available on other kinds of public support the firm received (or did not receive), as several Norwegian agencies share their data. The data is linked through the VAT number of the firm. Innovation Norway uses a customer survey before and after subsidy, with a three-year time interval to investigate output-indicators.

Use of data and lessons learned

Innovation Norway uses the data to construct control groups and measure the effect of cluster participation. In addition, the data forms the basis of more general studies (e.g., a study comparing profitability, growth in sales, value added, employment and productivity in firms that received support from Innovation Norway in 2016). As the system in place is able to link several data sets, it allows for integrated analysis of a broader set of policy instruments. The stability over time also helps to analyse long term effects.

More details on this case study can be found in the report on use of external data.

Case summary – EAS and the use of external data

The Estonian agency EAS combines data from the e-commercial register, the tax and customs boards and the national statistics office, and integrates it with own data. A central characteristic of the system is that it is designed to minimize the administrative burden for firms by reusing (similar) data that is already available.

Variables and data quality

EAS makes use of variables such as turnover, value added, and number of employees, all provided by the e-commercial register. The microdata is freely available for public organisations (although a confidentiality agreement is required). As this data is based on e-annual reports of companies, the data quality is high. The tax and customs board provides more up-to-date data on turnover, added value, paid taxed and employee numbers when required. Lastly, the national statistics office provides data on export (at an aggregated level). Similar to the e-commercial register, the data is of high quality.

Use of data and lessons learned

The data is used by EAS for monitoring and ex ante analyses and for the comparison of beneficiaries with control groups. In most cases external parties conduct the analyses, however, there are more ad-hoc cases where EAS analysis the data itself. The system in place is very sophisticated system and would likely take very long for others to adopt. An important driver behind the system of (re)using data is a relatively strong resistance against increasing administrative burdens in the public opinion.

More details on this case study can be found in the report on use of external data.

3.2.1 Reflections on the use of other external data providers

Many of the benefits of using data from national statistics offices as a strategy to monitor innovation instrument performance also hold for using data from external data providers: the data is (depending on the provider) often well-defined and harmonized. The agencies that use such data often do so in combination with the use of other data sources (such as data from NSOs). As the case of Innovation Norway shows, sometimes private providers and NSOs can even be used more or less interchangeably. Many of the lessons that can be drawn from cases hence hold both for the use of data from NSO and from other data providers. Also, with these cases it should be kept in mind that the cases in which external data providers are used are generally advanced systems that have taken years to build up and align. These processes are not easy to imitate (in some cases probably even impossible as they are adapted to local circumstances). Hurdles are often in cleaning and linking datasets properly, and in finding datasets that have both relevant variables and good coverage of relevant firms.

If other data providers are used, their data is often based on annual accounts. This ensures a high reliability of the data; however, it also limits the availability of indicators. The available data are often mainly of interest as background variables (to define groups of firms for example, or to construct control groups for evaluations). Also, the coverage of the datasets may be smaller than for example from NSOs; either because a sample is used, or because specific groups (such as small firm are excluded. Another disadvantage in using the data for continuous monitoring is the potential time lag; data from annual accounts in several case becomes available only once per year (and only after the accounts on the financial year are settled, which often takes at least several months). This kind of data does have a lot of potential for ex ante analysis, for example in defining target groups for new innovation instruments (as the case of EAS shows).

An advantage of using data by external data providers is that it may limit the administrative burden for beneficiaries of the innovation instruments. Another way to achieve this is by using open data. In the next chapter we illustrate its potential with several possible use cases.

Table 4 Key (dis)advantages of using different sources of external data

Data source	Advantages	Disadvantages
National statistics office	• Data well defined and harmonized (often built on European/OECD standards) • If access to microdata is possible, data can often be linked to firm level data of beneficiaries. • High quality and reliability of data	• Available variables often more useful for evaluation than monitoring • Substantial time lags, less useful for continuous monitoring. • Getting access (to micro-data) is often a hurdle; data need to be compliant with data protection measures. • Linking data can be complex and time consuming.
Other (private) data provider	• Data (depending on provider) often well-defined and harmonized (often built on European/OECD standards) • If access to microdata is possible, data can often be linked to firm level data of beneficiaries. • Usually, high quality and reliability of data • Sometimes very useful for ex ante analysis. • Can potentially lower the administrative burden for beneficiaries.	• Available variables often more useful for evaluation than monitoring • Substantial time lags, less useful for continuous monitoring. • Coverage of datasets may be smaller than with e.g. NSOs. • Getting access can be costly. • Linking data can be complex and time consuming.
Open data	• Can extend the coverage of smaller datasets. • Potential for predictive/ex ante analysis. • Unlocks possibilities for indicators in new fields, such as societal impact.	• In early phase of development, exploratory work needs to be done; harmonization across different languages for example might be complicated. • Appropriate training material can be a hurdle.

Data source	Advantages	Disadvantages
	• Minimizes administrative burden for (potential) beneficiaries.	• So far, analytic reliability seems to decrease with longer input texts. • Useful for specific types of indicators.

3.3 Other (novel) data sources

The availability of various kinds of data relevant to innovation policy is rapidly increasing, including the ease of access to curated forms of those data. Whereas Chapter 4 discusses in detail how agencies can themselves create their own data sets based on web scraping and other data extraction methods, the table below briefly presents an overview of a number of relevant available data sources for indicators. For more in-depth analysis, combination of different datasets, smart matching algorithms and machine-learning based categorisation are often important elements and may require specific programming capabilities.

Table 5 Overview of the new types of data relevant for innovation policy.

Type of Data	Examples of Providers	Available Relevant Indicators
Patents (and other IP)	PATSTAT (EPO)	Patents, Co-patenting, Tracing Research Results, New Product Registration (Trademarks etc.)
Bibliometrics	Scopus Insight, Web of Science	Industry – Academia Co-publications; Industry-Academia Staff Mobility
Publications related to clinical trials	Pubmed, Europepmc	Allows linking publications back to clinical trials, offering also the possibility to distinguish between the different stages of a clinical trial or to specific diseases
Social Media	Meltwater/LexisNexis	Company Communications, e.g. on new products, services, or SDG-oriented communication
Policy Uptake	Overton	Citation of Research Outputs in Policy Documents (e.g. social innovations)
Integrated databases	Dimensions, Lens	This kind of databases allows for example to track down which publications came out of a project, how these were used in patents, policy documents and (social) media
Career/Staff	LinkedIn	Analysis of R&D-capacity at firms, R&D staff careers
Regulatory Datasets	E.g. EMA, REACH, EFSA	Approval of new innovations in specific sectors, use of certain components related to the innovation project

4 New developments in monitoring

4.1 Using open data

As we have seen in the previous chapters, there are various ways to use data from external data providers to monitor (and evaluate) the outcome and impact of innovation instruments, although the implementation can be complicated and time consuming. Since about two decades, there is a growing interest among both social scientists and policy makers in another option: using the potential of machine learning and other quantitative analysis and data collection methods. This section will start with showcasing some use cases that could be of interest to innovation agencies. This is followed by a review of the big data collection methods and machine learning analyses methods.

4.1.1 Use cases for innovation agencies

There are several use cases with relevance for innovation agencies. The first is predicting innovation indicators. Developing innovation indicators for a larger sample of firms is an interesting use case for the application of 'new' quantitative methods, as it overcomes sampling constraints of current databases like CIS. Several authors have tried to expand the coverage of CIS and other datasets by developing a classification algorithm that uses the text from a company's website for estimating whether the company in question is likely to be innovative or not.[16] Evaluation of the performance of these models shows that the generated innovation indicators correlate quite well[17] with the responses provided in surveys such as the CIS. Another use case is assessing if the impact of projects align with national societal missions of for example the SDGs. Several efforts have already taken place for this specific use case, by amongst others the OECD that developed an SDG classifying tool[18] and online hackathons that zoomed into specific SDG and its indicators[19]. The main challenge in these efforts in the finding of well-suited training data. A third potential use case is the prediction of breakthrough patents. The quality and use of patents (and scientific publications) is very heterogeneous. Several authors have developed classification models that predict early on whether a patent will be a break through patent.[20] The performance of these models is significantly lower[21] than in the use cases described before, but can nevertheless provide some ex ante insights, that especially at the macro level could present some useful innovation indicators.

---

[16] Axenbeck, J., & Breithaupt, P. (2019). Web-based innovation indicators: Which firm website characteristics relate to firm-level innovation activity? *ZEW Discussion Papers, No. 19-063*. Retrieved from https://www.econstor.eu/bitstream/10419/213351/1/1688826920.pdf and Kinne, J., & Lenz, D. (2019). Predicting Innovative Firms using Web Mining and Deep Learning. *ZEW Discussion paper*. Retrieved from https://madoc.bib.uni-mannheim.de/51168/1/dp19001.pdf

[17] https://madoc.bib.uni-mannheim.de/51168/1/dp19001.pdf report a precision of 0.81, a recall 0.64, and an f1 score of 0.71.

[18] OECD. (2019). Linking Aid to the Sustainable Development Goals – a machine learning approach. *OECD Development Co-operation Working Papers*. Retrieved from https://www.oecd-ilibrary.org/development/linking-aid-to-the-sustainable-development-goals-a-machine-learning-approach_4bdaeb8c-en

[19] ZINDI. (2018). Sustainable Development Goals (SDGs): Text Classification Challenge. Retrieved from https://zindi.africa/competitions/sustainable-development-goals-sdgs-text-classification-challenge

[20] Hain, D. S., & Jurowetzki, R. (2020). Introduction to Rare-Event Predictive Modeling for Inferential Statisticians. Retrieved from https://arxiv.org/abs/2003.13441

[21] The study by Hain and Jurowetzki reports an accuracy of 90%, a recall of 52% and precision of 3%.

4.1.2 *Big data collection methods*

Many different datasets can be potentially used by innovation agencies. With the growing number of projects over the years, the size of the existing datasets with structured data about projects and beneficiaries has often outgrown the available capacities for a manual review of complete datasets. Furthermore, the current datasets often lack coverage of some relevant innovation indicators. This can be expanded with for example scientometric data (on patents and publications). While this data is nowadays better structured (and in case funding identifiers are clearly specified relatively easy to link to internal databases), accessing this data comes at considerable costs and its relevance depends on the indicators and sectors assessed. Another option are company websites, which often list information on innovation projects, publications and collaboration partners. Scraping this information, structuring it and linking it to internal databases can provide data on several relevant complementary innovation indicators. Also social media data can provide insights, on expertise available, but also for example collaborations and labour mobility flows.

Considerations for data collection methods

Discussions regarding the above-described data collection methods could benefit from including the following three considerations:

- Discuss for each indicator which data sources are available and consider factors such as quality, coverage, costs and accessibility.
- Make data linking easier by asking organisations to provide relevant keys (such as the URL of their website(s), twitter handle(s)) that will enable linking internal databases to external collected data. Similarly, ask beneficiaries to clearly list funding IDs when publishing anything to which the funding has contributed so that it can eventually be linked back to the project.
- Think about the legal (GDPR) and technical possibilities for sharing the available data with third parties in case these will be involved in the analysis.

4.1.3 *Machine learning analysis methods*

Although many of the datasets collected with the methods described above can be analysed with traditional methods, machine learning quickly gain ground, as their potential increases due to higher computational power and large datasets. Most of the algorithms of relevance for studying social science phenomena can be categorised in one of the following two groups:

- Supervised learning algorithms: these predictive algorithms are mostly used for classification tasks in which data is classified to existing predefined categories (hence supervised), such as predicting whether a firm is innovative or attaching SDG labels to innovation projects
- Unsupervised learning algorithms: these algorithms aim to detect un-predefined structures in the data (hence unsupervised), such as segmenting all start-ups in a country into N different groups in which the algorithm is tasked with minimising the within-group-variation

While for both categories many potential use cases for innovation agencies could be identified, supervised learning has gained the most traction for social science applications. In the report on external data are more details on the fundamentals of the supervised learning, the technological implementation, model performance, and further considerations for applying these methods in practice.

Case summary – FFG (Austrian Research Promotion Agency) and the use of Aurelia data and open data

FFG in Austria recently invested in access to data from the Aurelia dataset, a database with company information on firms in Austria. This is combined with the e-CORDA database. This is then used to extract the URLs from company websites. A web scraper is used to scrape the text of websites by companies that are involved in H2020 projects.

Variables and data quality

The data used by FFG includes variables like turnover, employment, shareholder information and a NACE-classification of the firm. For these, the coverage is good. The basis for this web scraper is the Scrapy framework, written in Python. This is an open source and collaborative framework for extracting data from websites[1]. A text mining algorithm is trained to classify if the collected passages contain relevant information on green deal aspects or not. For the text mining algorithm, the NLTK (Natural Language Toolkit) is applied, also in Python. NLTK is also open-source software. The algorithm classifies websites to the distinct green deal areas, based on a probability; it shows the probability that was assigned to a web page being relevant for a green deal area.

Use of data and lessons learned

The project is still in a very early phase, more lessons will likely follow in due time. However, exploring the possibilities of text mining in combination with linked data sets and company websites may be potentially very interesting for other agencies too, as it allows for completely new ways to gather information, without increasing the administrative burden for firms.

More information on this case study can be found in the report on the use of external data.

4.1.4 Reflections on the use of open data

The use of open data allows for the analysis of new indicators, that would result in a much higher administrative burden for (potential) beneficiaries, and in some cases a lower quality of data if they were collected in other ways. One the advantages is that (depending on the amount of available data) it is less sensitive to sampling issues than data from for example the CIS survey, and that analyses can be very targeted to specific sectors, regions or groups. It can be used also to expand the coverage of other datasets. Moreover, it allows for the assessment of new impact indicators that are problematic with 'classic' datasets, such as the alignment of projects with societal missions or SDGs. Something similar is currently explored by FFG to assess green deal aspects of projects. In general, a main challenge for such applications is to find suitable training data for the machine learning models employed in such data analysis. Moreover, at the current state of methodology, the analytic performance of such models reduces when the length of inserted text increases. It has also broader potential applications, for example in ex ante prediction of patent quality. The potential increases even further when linked to other datasets.

4.2 New indicators for societal impact

There is growing attention among policy makers and other stakeholders for the societal impact of innovation; this is reflected by an increasing demand for statistical indicators that shed light on the societal effects of innovation projects. Seven of the participating agencies in the TAFTIE project for example currently have an indicator that attempts to measure (parts of) the environmental effect of the innovation projects.

A major driver of current global statistical data collection has been the world leaders' endorsement of the 17 Sustainable Development Goals (SDGs). Monitoring and measuring the contribution of STI to implementing the SDGs still needs new and more indicators. For example, analysis based on detailed budget allocations may provide more detailed information on STI "input" commitments to the SDGs, e.g. those relating to poverty alleviation, equality or clean energy. Intermediate output indicators such as publications, patents and number of firms in SDG related domains, can be used for better STI monitoring and evaluation.

4.2.1 UN-based indicators

The UN is making efforts to conceptualise and substantiate data and indicators related to member states' STI performance and alignment with SDGs. The Table below presents examples of relevant indicators within the UN system.

Table 6 Examples of STI-based indicators within the UN system

Result	Indicator	Specification	Source(s)
Societal Impact	R&D expenditure as a proportion of GDP	SDG 9	UIS
Societal Impact	Number of researchers per million inhabitants	SDG 9	UIS
Societal Impact	Ecological sustainability composite index	SDGs 13, 14 & 15	GII
Societal Impact	GDP/unit of energy use	SDGs 13, 14 & 15	GII
Societal Impact	Environmental performance	SDGs 13, 14 & 15	GII
Societal Impact	ISO 14001 environmental certificates/bn PPP$ GDP	SDGs 13, 14 & 15	GII

4.2.2 OECD-based indicators

Outside the UN system, the OECD plays the leading role in the collection, harmonisation and interpretation of innovation data. While most of the OECD STI indicators are focused on economic objectives, monitoring alignment with SDGs is of increasing relevance. In the Daejeon Declaration on STI Policies in 2015, ministers across OECD and non-OECD economies highlighted the essential role of STI in meeting global and societal challenges, such as environmental sustainability, food security and healthy ageing, and in achieving the SDGs.

Additionally, the OECD has developed a classification of environment-related technologies using patent class systems. This classification provides multiple categories and sub-categories ranging from environmental management to water-related adaptation and climate change mitigation technologies. The classification systems explored by the OECD allow for a rich characterisation of relevant technologies by describing the engineering features of an invention and its applications at a very fine level of detail. For instance, the International Patent Classification (IPC) system includes over 70,000 separate technological classes, and the Cooperative Patent Classification (CPC) system has over 200,000 classes. Therefore, these data enable the identification of very specific environmental technologies. In addition, each patent can be allocated to multiple classes allowing to classify categories which are horizontal in nature.

4.2.3 Firm-level indicators

The increasing amount and quality of firm-level data being generated and collected is creating new opportunities for evidence-based policy analysis in the area of innovation. Numerous public and private datasets are tracking an increasing number of firms and gathering information about them in a larger variety of dimensions and formats. Examples of common dimensions include financial and employment data, R&D and capital investments, strategic collaborations, etc… Additional sources and formats include textual information describing firms, biographies of management teams, information scraped from their websites or firm-specific news. These types of data are being increasingly used in policy and innovation research. [22]

In the context of SDGs, using textual sources describing firms can be of particularly high potential, because traditional industrial classifications are typically not disaggregated enough to identify potential societal goals. In the case of climate action, private databases like Crunchbase or Pitchbook are creating specific tags in environmental categories based on text describing firms' activities and using keywords. In Pitchbook, examples of new categories in emerging spaces include "Sustainable fashion" and "Sustainable tourism" in the area of B2B, "Desalination tech" and "Smart waste management" in the area of materials and resources, or "Renewable ocean energy" and "Carbon capture and removal" in the area of energy. These categories enable the creation of indicators reflecting innovative activity in very specific environment-related sectors, tracking, for example, number of new firms, amount of capital raised, venture capital deals as well as mergers and acquisitions.

Table 7 Examples of possible indicators within the Pitchbook and Crunchbase

Result	Indicator	Specification	Source(s)
Societal Impact	Number of new firms, deals or investments in "Sustainable tourism"	SDGs 12, 13, 14 & 15	Pitchbook
Societal Impact	Number of new firms, deals or investment in "Desalination tech"	SDGs 6 & 14	Pitchbook
Societal Impact	Number of new firms, deals or investment in "Carbon capture and removal"	SDGs 7, 13 & 15	Pitchbook
Societal Impact	Number of new firms, deals or investment in "Green technology"	SDGs 7, 13, 14 & 15	Crunchbase
Societal Impact	Number of new firms, deals or investment in "Environmental consulting"	SDGs 7, 13, 14 & 15	Crunchbase
Societal Impact	Number of new firms, deals or investment in "Green buildings"	SDGs 6 & 11	Crunchbase

[22] See e.g. Dalle, J., M. den Besten and C. Menon (2017), "Using Crunchbase for economic and managerial research", OECD Science, Technology and Industry Working Papers, No. 2017/08, OECD Publishing, Paris, https://doi.org/10.1787/6c418d60-en.

5 Selection criteria for indicators

TAFTIE has built a rich school of thought on the design, selection and harmonisation of indicators for innovation policy in recent years. In 2015, with the support of Technopolis Group, TAFTIE formulated a set of criteria for good indicator quality, see Table 8 below. These criteria have been used also to select the indicators to harmonise (see next chapter).

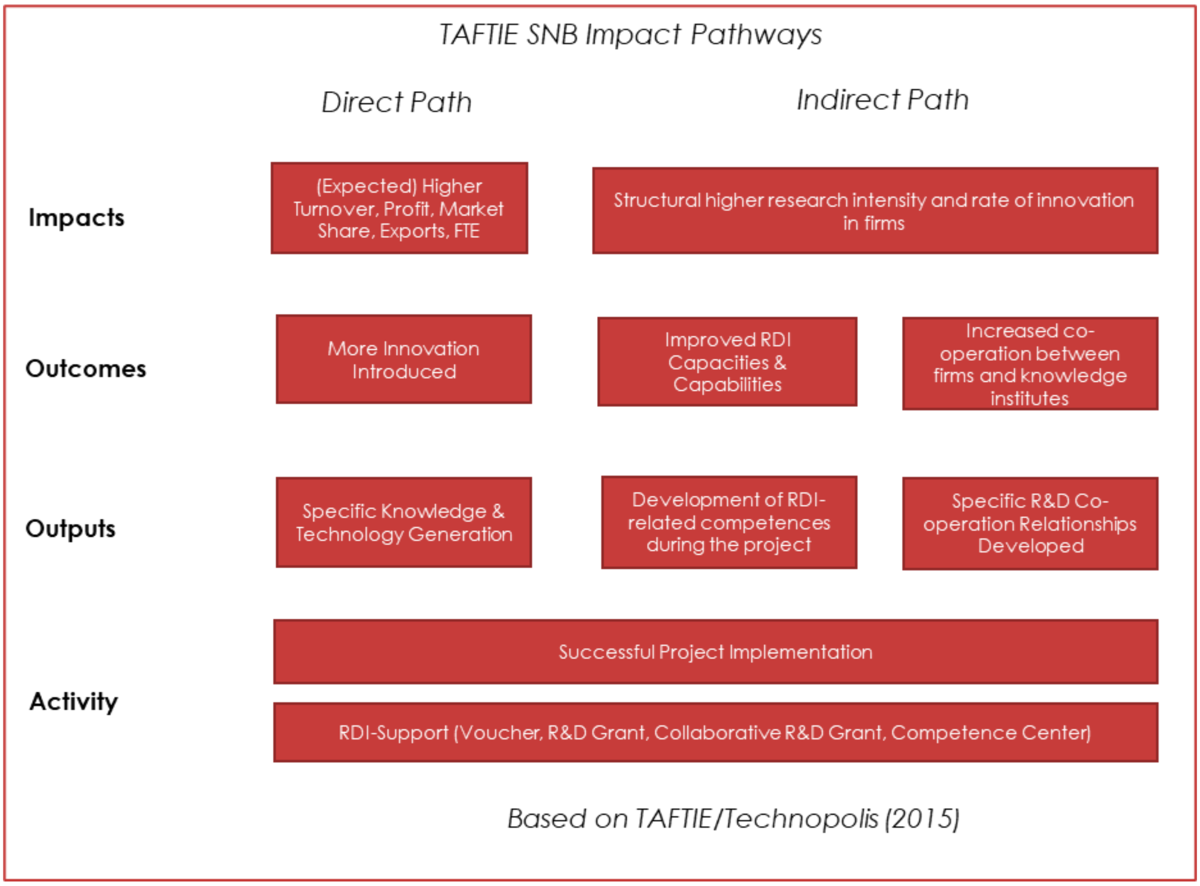
Table 8 Criteria for good indicator quality

Identification and selection	Operationalization	Sourcing & Collection
• Derivation of relevant indicators through a logical framework analysis • Integration of indicator identification and selection in the policy cycle • Integration of indicator selection and identification into an integral monitoring • Indicator selection supports triangulation but avoids duplication Indicators selection takes into account the operationalization & sourcing requirements	• Indicators operationalization should have attention for counterfactual aspects • Indicators have to be acceptable to stakeholders while maintaining relevance and independence • Indicators should align with international standards and take into account opportunities for benchmarking. • Indicators should be designed in a robust way The relevance of the selected indicator should be guarded throughout operationalisation	• The data collection/sourcing should be carried out in a cost-effective manner • Data collection is carried out in context of a long-term data strategy (together with key partners) • Indicators collection should be sourced from a credible, reliable and independent data source • The data collection process does not cause unnecessary or disproportional burden on beneficiaries Data collection provides feedback to indicator operationalisation and selection

Source: TAFTIE 2015

Another element of the 2015 TAFTIE report was a simplified logical framework to represent common outputs, outcomes and impacts of innovation instruments. In line with the innovation literature, two causal paths are distinguished: a direct (R&D leads to concrete innovations) and an indirect pathway (R&D improves absorptive capacity, dynamic capabilities, networks; see Figure 2 below and report 1 on monitoring systems).

Figure 2 2015 TAFTIE logical framework: TAFTIE SNB Impact Pathways



Source: TAFTIE 2015

The indirect pathway is especially relevant for understanding and measuring long term impact, and both pathways should hence be covered in the indicator set, although most indicators currently available revolve around the direct pathway.

It is crucial for the success of indicators that they are acceptable to stakeholders. In the framework of this harmonisation effort, this also implies that the selected indicators should be relevant to the majority of the involved agencies. They have given a first preliminary indication of their priorities in an informal prioritization exercise during the first workshop (see report 1 for more details).

There should be a good balance between relatively simple and more complex indicators, as the current practices are very divergent among agencies. An important element of this is the attribution of outcomes to a specific subsidy or project. While direct surveys by innovation agencies often focus on the results of the specific intervention, datasets that are provided by external providers such as national statistics offices often measure the innovative performance of a firm in a broader sense. Both methods have their own strengths and challenges, and both are used by many agencies. We hence propose at least one indicator that requires attribution and one indicator that does not per result area. It may be useful to reflect also on the possibility of measuring additionality of specific programmes. Some agencies already include direct questions regarding the attribution of subsequent effects, which can be very helpful in subsequent evaluation. However, for some agencies it would require major changes in their data collection system to include an indicator that can measure the additionality of an intervention, as they currently do not have information on organisations that did not receive the subsidy or intervention, and it is not required for monitoring. That is why the list of indicators does not require the use of control groups.

When developing harmonized indicators, it is important to take into account external standards such as those developed by the OECD (Frascati), Eurostat (e.g. CIS-surveys), EC (e.g. ERDF frameworks) and national statistical agencies. These indicators are not only typically robustly designed and tested, they are more likely to be consistently available and difficult to change. Of course, indicators must be relevant and useful to the agencies' objective first.

Given the long time lag between programme implementation and final (economic or other) impact of R&D and innovation activities, it is also useful to include a 'lower outcome' -level indicator, focused on the behavioural follow-up of an innovation project, as these are considered good predictors of the likelihood of further outcomes. In addition, such information can also help to better target monitoring efforts that take place long after the project had ended, i.e. by focusing on those beneficiaries that did take follow-up steps after the project. That is why indicator A1.1 is included, that measures the share of projects for which potential innovations stemming from R&D in the project are still actively pursued.

In addition, given the increasing prominence of mission-driven innovation policies and the SDG-framework, we have also included a common indicator for such (expected) societal impacts, see also the annex for further elaborations on potential indicators to measure aspects of societal impact.

The indicators are selected across five different result areas (building on Figure 2) innovation, improved RDI capacities and capabilities, increased cooperation between firms and knowledge institutes, economic impact, and societal impact, see Section 6).

Harmonisation of measurement

The harmonisation of indicator measurement not only requires the selection of indicators to be measured, but also detailed definitions and alignment of data collection (e.g. in the timing of the measurement). It is unfeasible that all agencies unify their measurement systems. It would also be unrealistic to expect that definitions etc can be adjusted immediately. That is why we have proposed to agree on 'ideal type' definitions for each indicator, yet also to leave room for the agencies to gradually converge their existing measurement systems to the newly proposed definitions over time.

Final selection of indicators

The selection of indicators has been an iterative process along the project. We have made a first exploration based on the analysis of monitoring systems. During the second workshop, a first -indicative prioritization was made, based on the preferences of the agencies. That served as input for an elaboration in more detail, based on our experience, some desk research, and the existing materials of the agencies. This has been discussed during the last workshop. After that, we have made a further finetuning of the definitions based on the feedback during and after the workshop and conducted a last validation round (by mail).

6 Proposed set of indicators

Based on the selection criteria and the discussions and input during the workshops, we have proposed the set of indicators that is presented in the table below. They are divided over two result levels (outcomes and impact), in five result areas (Innovation, Improved RDI capabilities, Increased cooperation between firms and knowledge institutes, Economic impact and Societal impact).

Table 9 Overview of harmonised indicators across two levels and five result areas

Result Level	Result	Result Sub-Area	Portfolio level-indicators	Measurement	Methodological notes[23]
Outcomes Measurement Year: 2-3 years after Project Completion Baseline Year: Year of Completion of the Project	**A: Innovation**	A1: Innovation Potential	A1.1 Share of projects from which directly-linked potential innovations are still actively pursued through concrete investment by supported firms (time/money)[24,25]	Self-reporting (Survey)	Self-reported attribution
		A2: Introduction of Innovation	A2.1: Share of projects that resulted in at least one innovation for supported firms (new/significantly improved product/service on the market or new/significantly improved process implemented)	Self-reporting (Survey)	Semi-Compatible with SFI CCR 01. Self-reported attribution
			A2.2 Share of supported SMEs[26] that have introduced one innovation (new/significantly improved product/service on the market or new/significantly improved process implemented) in the last 3 years[27]	Self-reporting (Survey) OR external monitoring	Compatible with SFI CCR 01; CIS-survey. No attribution

[23] The codes that begin with "SFI" refer to indicators used in the European Regional Development Fund (ERDF), one of the European Structural Funds (SF).

[24] I.e. count for how many of all projects that were supported the SME beneficiary still actively tries to turn it into an innovation (by investing time/money).

[25] For about half of the indicators, we propose to include SMEs only, to improve comparability and reliability of data. Large firms are excluded of indicators based on non-attributable surveys, due to the low expected direct relationship between the R&D project and performance. In the European definition, companies are considered SME with a staff headcount <250, or a turnover of 50M€. For A1.1, A2.1, A2.2 and E1.1 it may be considered to measure both SMEs and all firms, to see after the first round what differences are.

[26] 'Supported SMEs' here means: SMEs that had a supported project that ended 3 years or more ago

[27] The reported innovations do not need to stem from supported projects

[redacted]	**B: Improved RDI Capacities and Capabilities**	B1: Increase in R&D Investments	B1.1 l Winsorised average[28] increase (%-point) of total R&D expenditures of supported SMEs between baseline and measurement year of supported SMEs	Self-reporting (Survey) OR external monitoring	Compatible with RALLX from CIS. No attribution
			B1.2 Total increased R&D expenditures (EUR) of supported SMEs per mEUR support[29]	National Statistical Databases; Self-reporting (survey)	No direct attribution
		B2: Increase in R&D Staff	B2.1: Winsorised average increase (%-point) in R&D staff (FTE[30]) between baseline and measurement year of supported SMEs	Self-reporting (Survey) OR external monitoring	No direct attribution
	C: Increased Cooperation between Firms and Knowledge Institutes	C1: Continued collaboration with KI after project	C1: Percentage of supported firms that still actively collaborate on research and innovation with one or more knowledge institutes involved in the project[31]	Self-reporting OR patent/web-scraping analysis	Compatible with CEO68 from CIS No direct attribution
Result Level	Result	Result Sub-Area	Portfolio level-indicators	Measurement	Methodological notes
Impact Measurement Year: 3-6 Years after Completion of Project Baseline Year: Year of	**D: Economic Impact**	D1: Increase in Turnover	D1.1 Winsorised average increase in Turnover (%-point) of supported SMEs between baseline and measurement years	National Statistical Databases or commercial database; Self-reporting (survey)	No attribution, filter for SMEs to improve interpretation
			D1.2 Total increased turnover (EUR) for supported SMEs per mEUR support	National Statistical Databases or	No direct attribution

[28] This refers to a 'winsorized mean', meaning the mean where the values of the top 5% and bottom 5% are replaced by the 5% and 95% percentiles. This is similar to a truncated mean, where outliers (e.g. above 95% and below 5%) are excluded, but with winsorising the values are not excluded but replaced by a 'reasonable maximum/minimum', here the values of the 5th and 95th percentiles.

[29] If only research partners are supported: per MEur of support to the research partner (the indicator is about ratio between support given and increase in private R&D).

[30] If only headcounts are available, a typical conversion ratio can be used to calculate FTE.

[31] this indicator is currently limited to continuation of collaboration with knowledge institutes already involved in the project. If enough agencies have data on this, it might be considered to broaden it to new collaborations with knowledge institutes as a result of the project. We recommend to discuss this after the pilot round.

				commercial database; Self-reporting (survey)	
			D1.3 Winsorised average increase in Turnover (%-point) as a result of the project of supported SMEs between baseline and measurement years	Self-reporting (survey)	Self-attribution
			D1.4 Total increased turnover (EUR) of supported SMEs as a result of the project per mEUR support	Self-reporting (survey)	Self-attribution
		D2: Increase in Exports	D2.1 Winsorised average increase in Export (%-point) of supported SMEs between baseline and measurement years	National Statistical Databases or commercial database; Self-reporting (survey)	No direct attribution
			D2.2 Winsorised average increase in Export (%-point) of supported SMEs as a result of the project between baseline and measurement years	Self-reporting (survey)	Self-attribution
		D3: Increase in FTE	D3.1 Winsorised average increase in total FTE of supported SMEs between baseline and measurement years	National Statistical Databases or commercial datbase; Self-reporting (survey)	No attribution; Compatible with CCR13
			D3.2 Winsorised average increase in total FTE among supported SMEs as a result of the project between baseline and measurement years	Self-reporting (survey)	Self-attribution, Compatible with CCR13
		D4: Increase in productivity	D4.1 Winsorised average increase (%) in value added[32] per FTE for supported SMEs between baseline and measurement years	National Statistical Databases or commercial database; Self-reporting (survey)	No direct attribution

[32] It is recommended to use the Eurostat-definition and calculation of value added. See https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=DSP_GLOSSARY_NOM_DTL_VIEW&StrNom=CODED2&StrLanguageCode=EN&IntKey=16619885&RdoSearch=BEGIN&TxtSearch=value%20added%20at%20factor%20cost&CboTheme=&IsTer=&IntCurrentPage=1&ter_valid=0#:~:text=Definition,operating%20subsidies%20and%20indirect%20taxes.

	E: Societal Impact	E1: Contribution to Environmental Challenges (SDG 11-13)	E1.1 Share of concrete innovations as result of funded projects by supported SMEs (A2) that contribute to environmental goals	Self-reporting (survey);	Indirect self-attribution through A2; Compatible with CIS Question 13
			E1.2 Share of supported SMEs that refer to keywords related to SDG11-13 in their online communications	Web-scraping	No attribution: Requires development of standard methodology (e.g. informed by OECD classification tool)

6.1 Availability of indicators at agencies

We have analysed the possibilities for the agencies to measure the proposed indicators. In every result area there is at least one indicator where a group of agencies can start benchmarking. As expected, the most complicated area is societal impact; this is an emerging field where the results in the first round are probably more exploratory but given the policy developments in this area it has large potential to start harmonizing measurements here.

For a few agencies it will be complicated to actively participate in the pilot round, as they currently do not monitor outcome and impact indicators for the innovation instruments under consideration. However, for them it is probably useful and informative to learn from the steps that the other agencies take.

It is not necessary that every agency starts measuring each indicator. We suggest starting the pilot round for every indicator that can be measured by at least three agencies; that allows for a meaningful comparison of values and measurement methods; over time more agencies may be able to join.

For all outcome indicators, there are at least three agencies per indicator who can deliver values for the pilot round without large caveats. For a few impact indicators, there are less than three agencies per indicator who can deliver values without caveats. For D2.1 and D2.2 (increase in export) almost none of the agencies can deliver values without adaptations of the system (only EAS can deliver D2.1 without any adjustments). It should hence be reconsidered after the first round if these indicators should be included in the list, and if there are any alternatives. The same goes for D1.3 and D4.1: although there are some agencies that can measure these indicators, it may turn out after the first round that it increases the administrative burden too much to implement this broader as quantitative indicators (with a harmonized definition). This should be considered after the first round. Also, the indicators on environmental impact (E1.1 and E1.2) are relatively complicated. Two agencies can measure E1.1, and several agencies have a similar indicator fits reasonably well. E1.2 can not be delivered immediately. Given the more exploratory nature of this indicator, we recommend explore its possibilities further nonetheless, and use it to feed the discussions within TAFTIE about both appropriate indicators for societal impact and use cases to exploit the potential of open data. The question of ecological and societal impact will probably gain much more importance in the future. We think it would be worth to explore in a later phase how these indicators (or maybe better available ones if they appear in the future) could be expanded over time and try to find a balance between aggregated impacts (ecological and societal) and some details about the concrete impact.

7 Guidelines and recommendations for implementation

Most agencies currently have a well-established system of monitoring (see report 1'Comparative report of difference monitoring systems' for more details) that required years to build up and evolve over time, and is often interwoven with other systems, data providers, etc. It will hence also take time to adapt the systems in such a way that the data becomes comparable. In this chapter, we describe steps to be taken in the coming years to work towards harmonized indicators.

7.1 Benchmarking round

The first step will be to conduct a first benchmarking round. The first pilot exercise was carried out without external support by the agencies, HAMAG-BICRO & TA ČR (SNB coordinators 2019-2020), for several outcome and impact indicators in 2020 for collaborative R&D programmes.[33] Final results are presented in the third Benchmark Report 2020. Even while the data may not be fully harmonized in the first years, benchmarking will already be a very insightful and relevant exercise. In this round, every agency tries to measure a first value for each indicator it can deliver, and specifies the details of this measurement, including:

- Source/datasets used for the measurement and (if applicable) for the baseline. If the indicator is based on a survey question: the exact phrasing of the question.
- Moment(s) of measurement
- Alignment with other systems, such as ESFI.
- Caveats with respect to the sample used (for example the exclusion of firms <9 people in the CIS dataset).
- Frequency in which this indicator can be measured (including information on time lags, if for example data from annual accounts is used).
- Any other (technical) challenges encountered while measuring.

7.2 Overcoming technical issues

In doing so, the agencies will probably encounter many technical issues. The second step is then to make an overview of all these issues and decide on ways to solve or overcome them, to ensure that the next round will be less prone to measuring differences. Below we sketch the issues that already emerged in our analysis, and we provide suggestions for solutions.

- **Timing of measurement**: As the proposal already suggests, to some extent it is possible to work with ranges; measuring the outcome indicators at two or at three years after project completion probably does not result in large systematic differences. It is also possible to use correction factors. This is no exact science: other confounders, like conjunctural developments have probably a much higher impact on variables like turnover, export and job growth than these kind of measuring differences.
- **Measurement units**: we have tried to avoid differences in measurement units by using standardized units (e.g., %-points). However, in some cases it is inevitable to use other measurement units, such as FTEs or Euro (as a currency), and conversion factors will have

[33] https://taftie.eu/sites/default/files/SNB2020%20Report_FINAL_0.pdf

to be chosen. For currencies, we suggest using one fixed conversion rate in each measurement round. For jobs, we suggest to use FTE, and convert headcounts when needed (see the the Frascati manual[34]).

- **Statistic measures**: we propose to use the winsorised mean as measure for central tendency (the mean where the values of the top 5% and bottom 5% are replaced by the 5% and 95% percentiles).[35] Examples, in a uniform distribution of 1,2 … 99, 100, all values above 95 would take the value 95, all values below 5 would be replaced with the value 5. Afterwards, the mean is taken. This value can be communicated as 'corrected average' for general communication and is more intuitive than a median value. Although the winsorized mean is less known than other statistics for central tendency, it is very robust and easier to interpret than a median (of course, it is possible to report the medians too). Winsorizing basically decreases the impact of possible extreme outliers (caused either by noise/errors in the data or because there are for example firms with an exceptional profit from an innovation that may otherwise distort the picture). It can easily be computed in several software packages[36].
- **Instrument comparability:** Although instruments might have similar names, there can be substantial differences in terms of concrete objectives, implementation method, selection criteria etc. that may strongly affect the expected outcomes along the indicators defined. It is important to take these differences into account when benchmarking, in order to ensure comparing like-with-like as much as possible.
- **(Systematic) differences in sample**: several agencies have indicated that they cannot include all their beneficiaries for some indicators, as they would have to calculate it based on data from external data providers (that in many cases do not have full coverage of firms). We suggest discussing this after the pilot round has taken place.
- **Attribution**: it is important to ensure that no systematic differences exist in the way effects are attributed to a specific project or measure. If questions are asked in another context than measuring performance and impact of innovation policy instruments, this should be mentioned in the reporting of the benchmark.
- **Alignment with other systems**: it is important that the proposed indicators are aligned with international standards where relevant.
- In the survey questions, there may be (small) **differences in the phrasing of the questions**. This is especially the case for less traditional indicators, such as the ones on societal impact.
- **Organisation types**: some innovation instruments do not only have firms, but also knowledge institutes. We propose to measure the indicators for firms; these are often the most important beneficiaries, and in most cases, they are the target group of the instrument. To further improve the comparability of data, for some indicators we propose to look at SMEs only.

---

[34] OECD (2015), Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris, https://doi.org/10.1787/9789264239012-en.

[35] See for a an example of a study where this statistic is explained and applied: Kisseleva, Katja and Mjøs, Aksel and Robinson, David T., The Returns to Early-Stage Investment in Innovation (August 24, 2020). ESMT Berlin Working Paper, Available at SSRN: https://ssrn.com/abstract=3680129 or http://dx.doi.org/10.2139/ssrn.3680129

[36] For an explanation on how it can be computed in Excel for example, see https://www.real-statistics.com/descriptive-statistics/outliers-and-robustness/ or https://www.statology.org/winsorize-data-excel/

7.3 Equivalence and convergence over time

It is very likely that several iterations are required to come to a system where the ways of measuring are well-harmonized and comparable over time and across agencies. The steps above should hence probably be repeated several times. The comparability of the data can of course grow over time. Even while the data may not be fully harmonized in the first years, benchmarking will already be a very insightful and relevant exercise.

We recommend using this report as a living document, to keep track of changes in measurement over time, and update it for example once every year. Many agencies revise (the specifications of) their indicators after a couple of years. The lessons from the pilot round and subsequent rounds can also be used as input for such revisions, to come to more robust indicators.

It is important to keep in mind that the aim is not to come to identical measurement systems; this is also impossible given the heterogeneity and complexity of the systems. The aim is rather to come to equivalence, and make sure the methods of measuring are harmonized to such an extent that indicator values can be reasonably compared. As always with indicator selection, it is also important to strike a balance between minimizing the administrative burden (for beneficiaries, and also for the agencies themselves and maybe other stakeholders) and creating data that is fit for purpose.