



HAL
open science

Data Deficits in the Digital Age and How to Fix the Problem: More and Better Statistics to Support Technological Transformation at Work

Nathalie Greenan, Silvia Napolitano

► **To cite this version:**

Nathalie Greenan, Silvia Napolitano. Data Deficits in the Digital Age and How to Fix the Problem: More and Better Statistics to Support Technological Transformation at Work. 2022. halshs-03728530

HAL Id: halshs-03728530

<https://shs.hal.science/halshs-03728530>

Preprint submitted on 20 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

● WHAT COMES AFTER?

BEYOND4.0 supports an inclusive European future via examining the impact of Industrie4.0 and the Digital disruption on the future of jobs, business models and welfare.



POLICY BRIEF

July 2022, Policy Brief #6

Data Deficits in the Digital Age and How to Fix the Problem: More and Better Statistics to Support Technological Transformation at Work

BEYOND4.0 supports the delivery of an inclusive future of decent work and decent lives for European Union (EU) citizens. This Policy Brief focuses on statistics necessary for comprehensive analyses of the socio-economic outcomes of the technological transformation in EU workplaces. It argues that combining data at the employer (as best sources of information on the technological transformation) and the employee (as best sources of information on socio-economic outcomes) levels has the advantage of gathering complementary information that would not be available from only one dataset. It outlines the opportunities and challenges that data users face in such an endeavour and then provides the following recommendations for data providers:

- **Surveys should remain important data sources for evidence-based policies**, and the EU should continue to invest resources in employer and employee level surveys to provide comprehensive information on the technological transformation and its socio-economic outcomes.
- The following actions are needed to **facilitate the combination of data** from multiple sources:
 - Enable the ex-post integration of data and mitigate its costs by improving, as data is generated, the potential for that data to be combined;
 - Harmonise concepts, definitions, measures and classifications across surveys, especially for the key variables to be used in data integration;
 - Extend the data coverage as much as possible by enlarging survey sample sizes or at least improving survey response rates.

- The **measurement framework of technological change in EU surveys should be enriched** so that questionnaires track change over time in a standardised and systematic way by:
 - Setting a common rule about the inclusion of terminology to describe new technologies;
 - Capturing how these technologies are embedded in organisations and used by individuals both for private and working purposes.
- Innovative **data servicing solutions have to be developed**:
 - to ensure data security while facilitating data access to authorised data users;
 - to provide quality data using flexible anonymisation criteria without breaking confidentiality rules;
 - to improve the timeliness of data on technological transformation as well as social statistics targeting socio-economic inequalities and well-being.
- The views and needs of the European social science community and other researchers who use the data are part of the European statistical system. **Consultation with this community** should occur when the EU wants to renew or develop surveys or other data infrastructures.

Background to the Policy Brief #6

Funded by the Horizon 2020 programme, BEYOND4.0 examines the impact of Industrie 4.0 and digital technologies on the future of jobs, business models and welfare in the EU. Its aim is to support the delivery of an inclusive European future that provides decent work and decent lives for EU citizens.

Two particular developments are salient to policymakers: **Industrie 4.0**, which implies the integration of digital technology in production processes, and the **platform economy** involving the digitalisation of work, sometimes referred to as *Uberisation* or platform working. Both are examined by BEYOND4.0 and both offer opportunities and challenges.

This policy brief addresses the statistical needs of research that examines the socio-economic consequences of technological change. Focusing on existing EU-wide surveys, it highlights the potential of the data provided by the European Statistical System (ESS). It also reveals the gaps and challenges in the statistical data in relation to empirically linking technological transformation and organisational practices to their outcomes in terms of work and employment inequalities as well as occupational risks.

Context: Plenty of Data but Not the Right Data to Address BEYOND4.0 Issues From an EU-Wide Perspective

The ongoing digital and industrial transformation corresponds to a new accelerating phase of the technological revolution, sometimes called the 4th Industrial Revolution or the digital revolution. Enabled by increased network connectivity, this phase is marked by large-scale data collection, cloud computing, big data analytics and artificial intelligence, and the integration of digital and analogue or operational technologies, enabling enhanced and extended opportunities for advanced automation in manufacturing and services.

Yet, the persistence of sluggish growth in Western economies after the Internet boom indicates an economic and social transformation that is still incomplete. Moreover, the automation possibilities offered by new technologies have revived concerns about job losses and rising inequality similar to those expressed in the 1990s. Part of this persistent puzzle lies in how new technological opportunities shape the decisions of actors at all levels, particularly in firms and public organisations, influenced by market pressures, industrial relations, public policies and wider institutional arrangements in a globalised world.



The BEYOND4.0 project proposes a new way of analysing technological transformation, first, by conducting in-depth analysis of how new technological opportunities are combined with other tangible and non-tangible investments and are directed towards the production of new goods and services; and, second, by assessing its socio-economic outcomes – an issue that is overlooked in the monitoring of Europe's digital decade.

To analyse such a complex and rapidly changing landscape at the EU level and provide the necessary empirical evidence to support policy development, rich, harmonised and up-to-date information is needed. Indeed, it is widely recognised that data is a key asset in the digital age; it is the fuel for artificial intelligence, Industrie 4.0 and the platform economy. Moreover, data-driven decision making is seen as a new way to tackle future scientific, economic and social challenges.

However, the BEYOND4.0 project has revealed a serious data deficit that not only has challenged this project's objectives but also has important ramifications for EU policy development. We find that, **paradoxically, in a world of data abundance, the statistical data necessary for the EU to understand the technological transformation and its socio-economic outcomes is currently both limited and of poor quality.**

Critique of the Current Data Situation

A **first key point** in this observation is that we need to use statistical data and both employer-level data (which holds the best measures of technological transformation) and employee, worker or household-level data (which provides the best measures of socio-economic outcomes). Currently, analysis is undertaken using one or the other source. Rarely are these two levels of information linked. This is because the EU does not administer such data, making most of the analyses that inform EU (and Member State) policymaking partial.

This partiality matters. On the one hand, employer views about socio-economic outcomes are likely to be one-sided and incomplete. They are not the best informants of the socio-economic outcomes of technological change for employees. On the other hand, few employees have a comprehensive view of technological developments within their organisations. Most have insightful understanding about what is happening in their direct work environment but do not master the linguistic categories of emerging technologies, business models or management concepts. This is why employee, worker or household level surveys can only collect basic information about the most widely diffused technologies and their uses.

A **second key point** is that technological transformation is not just about adopting new technology. When companies choose to implement or adopt new technologies, they also face strategic choices about how to embed them into the organisation in order to maximise the opportunities these technologies offer. Effective implementation requires the combination of this technology with management and working practices, and the skillsets within the organisation to create new processes, products or business models. Surveys about technology diffusion typically do not include items about these complementary dimensions but which are central to delivering technological transformation. Indeed, these surveys are restricted to the description of the introduction of a set of technologies, as the Community Survey on ICT usage and e-Commerce in Enterprises demonstrates.

There Are *Two Possibilities* for Addressing These Key Points:

1. Develop new datasets that collect employer and employee information at the workplace or enterprise level ex ante and organise their linkage in the survey preparation phase.
2. Integrate existing employer and employee level datasets at an intermediate level through an ex-post linkage based on common and harmonised variables across datasets.

Some datasets or surveys already incorporate the first possibility, but only at the national level. At the EU level, the EU-funded MEADOW project has provided a measurement framework and guidelines for collecting and interpreting internationally harmonised data on organisational change (including the adoption of new technologies) and its economic and social impacts on both private and public sector organisations. It advocates that a survey that links data from employers and employees is the best option for measuring the dynamics of organisations and work.



The MEADOW guidelines were implemented in four European countries, driven by individual Member States between 2010 and 2015. In 2013, Eurofound's European Company Survey (ECS) also adopted some indicators developed in the MEADOW employer survey. The ECS is not a linked employer-employee survey but a dual voice survey that includes interviews with employer and employee representatives. It does, however include many of the dimensions needed to build a comprehensive picture of workplace technological transformation, e.g.

technology use, work organisation, human resource management, skills utilisation, direct employee participation and social dialogue. Thus, to date, this EU-wide employer survey has gone the furthest in approaching the technological transformation. Nonetheless, as with most employer level surveys, the ECS does not provide a comprehensive and reliable approach to measuring socio-economic outcomes.

The second possibility, and the one pursued by BEYOND4.0, is to jointly analyse the technological transformation and socio-economic outcomes with employer and employee level datasets integrated at an intermediate level. This approach entails an evident loss of information by substituting individual (or micro) data with aggregate (meso or macro) data. Still, it is a good alternative when data characteristics do not allow for record linkage or statistical matching. We have explored how to optimise the integration of several EU surveys and data sources focused on specific dimensions of the technological transformation or specific socio-economic outcomes:

- For employer data, in addition to the ECS, we selected three EU surveys coordinated by Eurostat: the Community survey on ICT usage and e-Commerce in Enterprises, the Community Innovation Survey (CIS) and the Continuing Vocational Training Survey (CVTS);
- For worker/employee/household data, we selected Eurofound's European Survey on Working Conditions (EWCS) and two surveys coordinated by Eurostat: the Community survey on ICT usage in households and by individuals and the EU Labour Force Survey (EU-LFS).

We focused on an intermediate level of integration, using a common cell given by a size group in a sector within a country in a given year. To implement this approach, the key variables needed to build this common cell have to be present and harmonised in the available datasets. The next step was to aggregate information at the intermediate level and then integrate the resulting datasets. Problems however remain with this data usage.

The surveys administered by Eurofound are easy to access at the individual level, they are well documented and include harmonised key variables for integration. The data provided by Eurostat are problematic:

- Although they are conducted EU-wide, the Community employer level surveys are either non-accessible at the individual level or accessible in such a limited way that it is not possible to use them for EU-wide comparative research. The main reason is that the data are anonymised to protect business confidentiality and this prevents linkages to other datasets.
- Data access is bureaucratically burdensome, requiring Eurostat to approach each Member State with details of the applicant research project and to seek approval. This situation holds even for EU-funded projects such as BEYOND4.0 that have their own projects ethics approval and research design general agreement through DG Research.
- Information of this bureaucratic process and other data access limitations are not openly flagged. The information on the Eurostat website, for example, does not clearly point to the significant limitations and challenges securing access to the micro-data.

As a result, none of the selected Community surveys at the employer level had accessible microdata to support BEYOND4.0's integration strategy. In the end, the CVTS was discarded, and our integration strategy proved feasible only by using the already aggregated data available on Eurostat's website for the Community surveys on ICT usage and e-Commerce in Enterprises and for the Community Innovation Survey. These aggregates have two main drawbacks: they are black boxes (for instance, the reason for a missing value is not explained) and they are not designed to answer specific research questions (for instance research questions could lead to the construction of a composite indicator that has not been foreseen in the standard aggregation step).

Eurostat's employee-level micro-datasets have fewer user problems than employer-level ones but can still be challenging to use. For example, in the survey on ICT usage in Households and by Individuals, there is a problematic deletion of some response items such as the distinction between employees and self-employed or, in the Labour Force Survey, there is the loss of the second digit of the NACE rev2 sector classification.

Furthermore, the weaknesses of the EU survey instruments used to approach the digital revolution for research purposes are noteworthy. Indeed, although a yearly Community survey on ICT usage and e-commerce in enterprises has existed in the ESS since 2002 with the aim to collect and disseminate harmonised and comparable information on the use of ICTs at the European level, Eurostat still does not provide access to its micro data for research purposes.



While questions about robots and Artificial Intelligence (AI) applications have entered the survey in 2018 and 2020, respectively, the EC has commissioned in 2020 an ad hoc European enterprise survey on the use of technologies based on AI to feed into the Digital Economy and Society Index (DESI). However the methodology used in this survey is weak. We are also likely to lose the little information available on what is happening in the European workplace as Eurofound is probably discontinuing the ECS or at least not planning to collect data in the future due to a lack of budget.

In short, some high-quality data paid for by EU citizens are not openly available, have restricted access and, in some cases, suffer from variable or item response deletion before being made available to researchers, making the released data both limited and poor. Moreover it looks likely that some of the sources of better data are about to disappear.

We Need to Stop Groping Around in the Dark and Have More and Better Data Produced by the ESS

Eurostat is changing and developing a new strategy for data collection and data services, but in the short-term there may not be much improvement in data generation and service to data users. Evidence-based policy is needed to deliver effective technological transformation. If this evidence is to be provided, researchers need to stop having to grope around in the dark for data and instead have more and better data generation and access from the ESS.

Policy Recommendations

Our experience with BEYOND4.0 leads us to make a number of recommendations that will help create more and better data to support the EU becoming a prosperous, inclusive and sustainable digitalised economy and society.

1. Surveys should remain important data sources for evidence-based policies

Web-scraping and machine learning techniques are thought to avoid much of the need for survey data collection. However, there are still significant problems with these techniques, as the collection of web-based vacancy data has shown. It would be more useful to combine survey data in innovative ways with unstructured data from the web or other private and public sources. In any case, it should be recognised that data from different sources are uneven. Recommendations to stakeholders should therefore be systematically preceded by an assessment of the strengths and weaknesses of data sources.

The EC and Member States argue that in the social sciences and humanities, data collection is the task of statistical agencies. However, national and international statistical agencies sometimes feel that they do not have sufficient funds to ensure researchers' access to data, and they may ask to cover their additional costs with research funds, putting pressure on the limited budgets allocated to the social sciences. There is therefore great potential for improvement, particularly in the development of data infrastructures, to cope with the socio-economic consequences of technological transformation.

2. The EU needs to maintain its investment in EU-wide employer and employee level survey

Within this investment, effort should be directed towards two aims. First, facilitating the combination of data from multiple sources and, second, enriching the measurement framework of technological change to take account of the constant renewal of technologies and complementary intangible investments.

2a. Combining data from multiple sources

A linked employer-employee survey would best provide the information needed to capture the technological transformation and its socio-economic outcomes. On the one hand, employer representatives would be asked about how technologies are developed together with new management, organisational and working practices. On the other hand, small samples of their employees would be asked about their uses of technology, work experiences and outcomes.

In the absence of such a survey, the best data configuration derives from the implementation of an ex-post linkage between existing EU-wide surveys or datasets at the employer and employee levels. In BEYOND4.0, we apply the most flexible approach of data combination, which consist in identifying a common level to which information can be averaged up and then investigated jointly.

Linking employer and employee data sources requires preliminary data harmonisation work and to date, much can be done to facilitate this approach and make it more efficient. In existing EU employer and employee surveys, the main obstacle to the reliable integration of data is the available categorisation of common variables at a detailed level for the sector (Nace Rev. 2



classification of economic activity), the region (NUTS classification), the enterprise size category (no standardised classification) or the professional status (in particular the distinction between employees and self-employed). The identification of the sector is also generally more accurate from employer surveys than from employee surveys because the production units directly provide the information while employees could not be fully aware about it. It is essential to ensure the comparability over time and between countries of these categorisations by double

coding when classifications are updated and by documenting country level procedures to support it. The problems needing to be addressed is both the lack of category standardisation and the overly restrictive anonymisation rules that, surprisingly, are often stricter for employer data than employee data.

Sample sizes can also be an issue in a data combination approach that aggregates information at a common level. Samples should be large enough to have a minimum number of observations at the level of aggregation. Hence, consideration should be given to facilitating the ex-post integration of datasets in the development of EU surveys in order to mitigate downstream integration costs and increase the analytical usefulness of the combined datasets.

2b. Enriching the measurement framework of technological change

Technological change is dynamic and particularly so in the case of the current digital and industrial transformation. Technologies are continuously upgraded and renewed at a pace that sometimes accelerates. Surveys capture the adoption and use of technologies in enterprises or by individuals through direct measures. However, surveys struggle to grasp the technological dynamism. They are better suited to measuring static or slowly changing phenomena. For example, the ICT Usage and e-Commerce in Enterprises Survey questionnaire changes every year to take into account technological change and so collect timely information. However, the way a given technology enters or exits the questionnaire is not systematic. A common rule should be set.

Furthermore, technological transformation occurs through combined technological and socio-organisational changes involving the implementation of new management, organisational and working practices and skillsets. Survey questionnaires should be optimised to track over time - in a standardised and systematic way - the emerging technologies likely to be diffused as well as how they are embedded in organisations and used by individuals for private or working purposes. For example, a list of technologies and management, organisational and working practices to be covered could be developed based on a qualitative assessment of the field, supplemented by a quantitative analysis of the frequency with which these practices are mentioned in professional journals or Internet searches plus a glossary to collapse them together as relevant.



3. The EU needs to invest in innovative data servicing solutions

High data access costs for data users and poor data quality are unhelpful both to researchers and policymakers reliant on robust evidence. If confidentiality issues are important, privacy-preserving technologies need to be developed to ensure data security while facilitating access for authorised data users. Innovative solutions also need to be developed to provide quality data without breaking confidentiality rules. An online application could be used to tailor aggregations taking into account the needs of data users in terms of combined datasets. This online facility would allow data users to make their own trade-offs in choosing the key variables for the common cell that defines the aggregation level.

The recently adopted EU regulation on the Integrated Social Statistics stresses the need to improve the timeliness of statistics feeding into the social scoreboard. In terms of socio-economic outcomes, early estimates of poverty, inequality and reliable and internationally comparable indicators of well-being are called for. The timeliness of data on technological transformation also needs to be improved to facilitate meaningful futuring exercises and better anticipate its socio-economic consequences.

The European data process (development, access, documentation) should better take into account the views and needs of the EU-wide community of social scientists and other researchers using the

data. Consultation within this community needs to be improved institutionally. The EU-funded InGRID1&2 projects provide inspiring examples of ways to strengthen the European dimension of data infrastructures that ultimately serve evidence-based policies on inclusive growth. For instance, by harmonising standards for questionnaires and classifications, integrating and archiving data on a European level and stabilising a distributed research infrastructure for EU-wide data easily accessible and responsive to the needs of a growing science community.

Conclusion

Technological transformation is affecting the world of work. However, there is currently insufficient data and good access to data to understand this technological transformation and its socio-economic outcomes. As a consequence, policy making suffers and so the data deficits need to be plugged. This policy brief has outlined the problems with current data within the ESS and points out what needs to be done to improve this situation and how more and better statistics can be generated to support a technological transformation in the EU that delivers an inclusive future of decent work and decent lives.

Authors

Nathalie Greenan and Silvia Napolitano, both of CNAM CEET in Paris.

Useful Readings

Abowd, J.M., & Kramarz, F. (1999). The analysis of labor markets using matched employer-employee data. *Handbook of Labor Economics*, 3: 2629-2710.

Cardenas Rubio, J. and Warhurst, C. (2022). Big data and the shift towards real time labour market information. *New Direction*, May, 48-52.

Corrado C.A. & Hulten C.R. (2010). How Do You Measure a 'Technological Revolution'? *American Economic Review*, 100(2): 99-104.

Economic and Social Council (2016). Transformative agenda for official statistics, *Report of the Secretary-General, Statistical Commission*, United Nations, E/CN.3/2016/4.

European Commission (2021). The European Pillar of Social Rights Action Plan, DG Employment, Social affairs and Inclusion, COM(2021) 102 final.

Greenan N., Lorenz E., Allan S., Amossé T., Archiburgi D., Arundel A., Bejerot E, Bellmann L., Bressé S., Coutts A., Csizmadia P., Ester P., Forth J., Gjerding A.N., Green F., Guillemot D., Rasmussen J.G., Hamon-Cholet S., Härenstam A., Hesselholdt A., Houtman I., Huys R., Illéssy M., Isidorsson T., Jansen W., Kirner K., Kraan K., Lazaric N., Le Guel F., Lund R., Lundvall B-Å, Makó C., Nielsen R.N., Nielsen P., Ouchhi S., Pahnke A., Ramioul M., Román A., Sirilli G., Sirilli S., Som O., Van Cruysen A., & De Vroome E. (2010). The MEADOW Guidelines, 978-2-7466-1809-1. halshs-01362486.

Greenan N. & Napolitano S. (2021). Why do employees participate in innovations? skills and organizational design issues and the ongoing technological transformation. In Klaus F. Zimmermann (ed.) *Handbook of Labor, Human Resources and Population Economics*, Cham: Springer.

Greenan, N. & Walkowiak, E. (2010). La dynamique des changements à long terme. *Réseaux*, 4: 231-273.

Lenaerts, K., Ramioul, M., Nelson, K., Steinmetz, S., Esteve, A., Gábos, A., György Tóth, I., Besamusca, J., Articus, C., Münnich, R., & Shlomo, N. (2021). Potential pathways for the future development and sustainability of the InGRID research infrastructure, InGRID-2 project 730998 – H2020.

Mathieu C. & Warhurst C. (2020). *Platform work*, Policy Brief #3, Beyond 4.0, [Beyond 4 0 EU Policy Brief 3 Platform Work.pdf \(beyond4-0.eu\)](#)

Seghir, M. (2021). Integrated database linking international employer and employee surveys on work organisation, job quality and performance, InGRID-2 project 730998 – H2020.

Warhurst C., Mathieu C., Dhondt, S. (2020), *Industrie 4.0*, Policy Brief #2, Beyond 4.0, [Beyond 4 0 EU Policy Brief 2 Industrie 4.0.pdf \(beyond4-0.eu\)](#).

Project Details

Project name	Inclusive Futures for Europe BEYOND the impacts of Industrie 4.0 and Digital Disruption — BEYOND4.0
Coordinator	Prof. Dr Steven Dhondt (scientific coordinator), Dr Peter Oeij (project coordinator). Nederlandse Organisatie Voor Toegepast Natuurwetenschappelijk Onderzoek TNO, Netherlands
Consortium	Department of Social Research, University Of Turku, Finland Institute for Employment Research, University of Warwick, UK Institute for the Study of Societies and Knowledge, Bulgarian Academy of Sciences (ISSK-BAS), Bulgaria Le CNAM-CEET, France Nederlandse Organisatie Voor Toegepast Natuurwetenschappelijk Onderzoek TNO, Netherlands Technische Universitat Dortmund, Sozialforschungsstelle Dortmund (sfs) (TUDO), Germany UCL Institute for Innovation and Public Purpose (IIPP), London, UK University of Helsinki, Finland University of the Basque Country - Sinergiak Social Innovation, Basque Country – Spain
Duration	2019 - 2023
Funding Scheme	Grant Agreement number: 822296 — BEYOND4.0 — H2020-SC6-TRANSFORMATIONS-2018-2019-2020/H2020-SC6TRANSFORMATIONS-2018
Budget	EUR 2,999,970.00
Website	www.beyond4-0.eu
Further reading	www.beyond4-0.eu/publications
For more information	contact@beyond4-0.eu, peter.oeij@tno.nl
Disclaimer	The contents of this publication are sole responsibility of the BEYOND4.0 project Consortium and do not necessarily reflect the opinion of the European Union.