13 Developments in Data for Economic Research

Roberto Barcellan, Peter Bøegh Nielsen, Caterina Calsamiglia, Colin Camerer, Estelle Cantillon, Bruno Crépon, Bram De Rock, László Halpern, Arie Kapteyn, Asim I. Khwaja, Georg Kirchsteiger, Vigdis Kvalheim, Julia Lane, Markus Mobius, Luke Sibieta, Joseph Tracy, Frederic Udina, Gugliemo Weber and Lisa Wright

Abstract

There has been a steep increase in empirical research in economics in the past 20–30 years. This chapter brings together several actors and stakeholders in these developments to discuss their drivers and implications. All types of data are considered: official data, data collected by researchers, lab experiments, randomized control trials, and proprietary data from private and public sources. When relevant, emphasis is placed on developments specific to Europe. The basic message of the chapter is that there is no single type of data that is superior to all others. We need to promote diversity of data sources for economic research and ensure that researchers are equipped to take advantage of them. All stakeholders – researchers, research institutions, funders, statistical agencies, central banks, journals, data firms, and policy-makers – have a role to play in this.

13.1 Introduction

The past 20–30 years have witnessed a steady rise in empirical research in economics. In fact, a majority of articles published by leading journals these days are empirical, in stark contrast with the situation 40 or 50 years ago (Hamermesh, 2013). This change in the distribution of methodologies used in economic research was made possible by improved computing power but, more importantly, thanks to an increase in the *quantity*, *quality* and *variety* of data used in economics.

This chapter brings together several actors and stakeholders in these changes to discuss their drivers and implications.[1] *All* types of data are considered. When relevant, emphasis is placed on developments specific to Europe. Sections 13.2 and 13.3 deal with official microdata. Section 13.2 focuses on the

level of access to microdata in Europe and its determinants. Section 13.3 focuses on cross-country data harmonization. Section 13.4 then switches gears entirely and discusses the benefits and costs of large-scale data collection efforts led by researchers, instead of statistical offices. Section 13.5 discusses data produced by researchers, either in the context of lab experiments or in the context of randomized control trials. Both types of data have led to major advances; for the first one in our understanding of human behaviour and the robustness of economic institutions; for the second in our understanding of the impact of policies and the mechanisms underlying them. The chapter closes by discussing new forms of collaborations that researchers are developing with private- and public-sector organizations, with the benefit of access to data of very high quality, as well as the opportunity to contribute to product and policy designs, and what it implies for how research is organized, evaluated and funded.

The basic message of the chapter is that there is no single type of data that is superior to all others. Each type of data is unique and has advantages over the others for a given research question. In many cases, they even complement one another. We need to promote this diversity and ensure researchers are equipped to take advantage of them. All stakeholders – researchers and their institutions, funders, statistical agencies, central banks, journals, data firms, policy-makers – have a role to play in this.

13.2 Organizing Access to Microdata

By Roberto Barcellan, Caterina Calsamiglia, Estelle Cantillon, Vigdis Kvalheim, Luke Sibieta and Frederic Udina

Microdata, that is, data at the individual, household, firm or establishment level, are a rich source for economic research. Their granularity allows researchers to get a better understanding of the heterogeneity of behaviour and outcomes in the population of interest, and thus yields better insights into the potential mechanisms at play. Two types of microdata are of particular interest to economists: survey data, which cover a representative sample of the population of interest and usually follows them over time (for example the Labour Force Survey in Europe), and administrative data, which are collected for administrative purposes.

Survey data were at the forefront of important empirical developments in economics in the 1980s, 1990s and 2000s. The current frontier now is administrative data. Administrative data have many advantages over survey data. They cover a broader set of activities and outcomes. They cover the entire population and track it over time, instead of providing a snapshot on a population sample. Their quality is high: they do not suffer from the kind of attrition, nonresponse and measurement errors that can plague survey data. They are cheap (they

already exist) and recent advances in computing power and data management techniques have made them easier to extract, manipulate and analyse. Last but not least, the possibility of linking different administrative datasets (for example, crime history and education history at the individual level) opens up endless possibilities for new research questions.

But microdata are sensitive. Individuals have the right to have their privacy protected. Firm-level data can contain competition-sensitive information that firms may not want to become public. Providing secure access to microdata and linked microdata is also resource-intensive. These factors make the option of not providing any access attractive for risk-averse or resource-constrained statistical systems. This section describes the legal environment and practical solutions that the European Union and the different Member States have put or are putting in place to reconcile the need for data protection and the promotion of data access for research purposes.

13.2.1 Legal and Technical Background

The European framework for data protection relies on two principles: the protection of personal data as a basic right, and the promotion of the free flow of personal data as a common good.[2] These principles necessarily go hand in hand: data subjects will not accept to have data collected on them if they cannot trust data owners to ensure their confidentiality. In turn, opt out clauses, which are unavoidable if trust is low, reduce the value of the data produced.

Existing European regulations allow (but do not require) Member States to grant access to microdata without the consent of the data subjects (which is typically the case for administrative data) when such data are essential for the pursuit of research and on the condition that they are de-identified.[3] De-identification involves the removal of personal identifiers such as national IDs or names, but can also involve the removal or blurring of other quasi-identifiers such as the address or date of birth.

De-identification does not necessarily remove all privacy concerns, however. Users of the data may be able to use combinations of variables (such as workplace and employment history to re-identify the individuals in the data). This risk of re-identification is heightened with linked microdata. Protocols need to be in place to ensure that confidentiality is preserved. These protocols regulate *who* can access the data and *how* this access is organized.

Under European regulations, access is granted in two steps. First, the institution to which the researcher is affiliated needs to be recognized as a research institution. This is important because it is the institution that eventually guarantees that proper safeguards are in place and the data will remain confidential. Second, access is only granted on a project-by-project basis. Each project (data

request) is evaluated on its scientific merit and the necessity for the confidential data.

Three forms of access are common: secure physical transfer of the data to the researchers, virtual access, and data enclaves (dedicated secure environments). Data enclaves are the safest form of access because the data never leave the room. They are also the most constraining for researchers who need to be able to set aside blocks of time during office hours to go to the data enclave. Virtual access is a remote desktop connection to the institution hosting the data. The data analysis is carried out on the remote desktop and the output is checked before it is sent to the researcher.[4] Secure transfer of the data to the researcher is of course the most convenient form of access for researchers, but it requires trusted researchers and a careful assessment of risks.

Despite a common legal framework at the European level, there are important variations across Europe in laws, and in legal and technical practices for access to microdata. These can be seen as the result of differences in cultural traditions and norms, public attitudes towards research, and quality and resources of local statistical systems. The following Table 13.1 taken from Castellani and Koch (2015) reporting on a comprehensive inventory exercise of data on indicators of competitiveness (the MAPCOMPETE project, described in greater detail in the next section) illustrates the existing heterogeneity in the legal conditions for data access. The indicators of competitiveness reported in this table all need to be built up from firm-level data. The table shows whether the data needed to construct the indicator are available without restriction (++), whether access is possible under some conditions (+) or whether access is impossible (−). The symbol '?' indicates cases where the authors could not get the information on access.[5]

In addition (and not reported in Table 13.1), there are also great differences in nonlegal barriers to access, such as lengthy approval procedures, overly restrictive interpretations of what constitutes '*necessity for research*', and time-consuming or inconvenient access. On this front, revealed preferences by researchers are the best indicators. Nordic countries stand out clearly on this dimension.

13.2.2 The Nordic Leadership

Nordic countries are world leaders in providing access to microdata for research. Their success is partly based on a long tradition of collecting data for administrative and statistical purposes. Virtually all interactions with the government or publicly funded service providers are covered. For individuals, this means for example family composition, medical records (including prescriptions and care), education history, employment status and employer identity, income, and social benefits. For firms, the available data include tax and other

Table 13.1 *Accessibility of selected indicators of competitiveness across EU countries (excerpt from Table 2.4 of Castellani and Koch, 2015)*

	Labour productivity[a]		
	All firms	Exporters	Foreign-owned firms
Austria	−	−	?
Belgium	++	−	−
Bulgaria	+	+	+
Croatia	?	?	?
Czech Rep.	+	+	+
Denmark	+	+	?
Estonia	+	+	+
Finland	+	+	+
France	+	+	+
Germany	+	+	?
Hungary	+	+	?
Ireland	+	+	+
Italy	+	+	−
Latvia	+	+	?
Lithuania	−	−	−
Malta	+	+	−
Netherlands	+	?	?
Poland	+	+	?
Portugal	+	+	?
Romania	−	−	−
Slovakia	−	−	−
Slovenia	+	+	+
Spain	−	−	−
Sweden	+	+	+
UK	+	+	+

[a] Average, median, other moments.

financial statements, sales, ownership, and employee identities and salary histories. Individuals and firm establishments are uniquely identified in all of these administrative registers, making linking possible. Some of the data series start in the early twentieth century. Very few, if any, statistical systems in the world match the comprehensiveness of the Nordic statistical systems.

Data availability and quality are only part of the explanation, however. Ensuring and organizing access to these data is essential too. Trust is the key word here. Nordic countries combine some of the highest levels of protection of personal and business data with the highest levels of access. Projects are screened for their societal interest and to ensure that the data request is legitimate given the research question. Researchers must be affiliated with a pre-approved

research institution (usually in the country), which takes responsibility for any violation of the confidentiality agreement by the researcher. The flip side of these high standards is the shared understanding and public support for securing access to data for researchers. Laws contain provisions to secure legitimate access to data for research purposes. Statistics Finland, Statistics Norway, and the Norwegian Social Science Data Services (NSD), a go-between organization between researchers and data owners, have among their mandates to service researchers.

There is no single '*Nordic model*', however. While all Nordic countries have comprehensive population and business registers that they make available to the research community, research arrangements and organization practices differ.[6] One dimension in which the countries differ concerns how access to data is provided. In Denmark, virtual access prevails. Norway hands out data to researchers but is currently developing virtual access as well. In Sweden, virtual access and physical transfer of data coexist. On-site access is the default option in Finland, but virtual access is also possible. Because these different modes of access entail different risks in terms of data confidentiality and integrity, some data may only be accessible through one channel and not the other.

Irrespective of how access is granted, direct access prevails: the data owners (data registers) are those providing access to the data and thus screening the projects for their compliance with the law and approving researchers. When the data request involves several data registers (for example: education data and employment data), the national office of statistics performs the data merger and its de-identification.

Norway is special among Nordic countries in that it also offers mediated access. The Norwegian Social Science Data Services (NSD) acts as an intermediary between the research institutions and Statistics Norway to provide access to researchers. The NSD screens projects, provides guidance to researchers, negotiates research use and access, and hands out the prepared data to the researchers. The main advantage of mediated access for researchers is its efficiency: Access is fast and free. Not all data can be accessed through NSD, however. NSD focuses on scientific use files (such as surveys) for which the risk of identification is appropriately reduced and simple data requests. With 1200–1400 projects serviced per year, these represent the bulk of microdata requests.[7] Data requests that involve data owners other than Statistics Norway, cover the entire population, or require very detailed and thus identifiable personal information must go through Statistics Norway (direct access).

Despite their success, statistical offices and other data stakeholders in Nordic countries are continuing to push for greater access (including from abroad), while keeping the same high standards of data protection. One of the ongoing projects seeks to develop research on cross-Nordic administrative data. Such research is still rare. The absence of common descriptors (metadata) is a major

obstacle. Other hurdles include a lack of data harmonization, independent and sometimes lengthy application procedures to get access, and organizational constraints with some countries still restricting access on-site or to researchers based in the country. In 2012, NordForsk, the platform for joint Nordic research and infrastructure cooperation, funded a feasibility study on how to enhance cross-Nordic register cooperation. The report (NORIA-net, 2014) advocated the development of a common metadata framework, a common application procedure and a model of joint access to Nordic microdata through the existing remote access systems. This model for cross-Nordic register data access is currently being developed for social data. Economic and medical data will follow.

13.2.3 Improving Data Access: Two Case Studies

Beyond Nordic countries, there is also a clear trend in several European countries towards facilitating access to administrative data for researchers. Individual researchers are playing a key role in these developments by leveraging existing laws and working through them to get access and set precedents for other researchers. It would be wrong, however, to view existing developments as being only demand-driven. The UK and the Autonomous Community of Catalonia (Spain) provide two contrasting examples of transition towards greater access in countries that do not a priori benefit from the favourable conditions of Nordic countries. The two cases illustrate the different drivers of this evolution and the form that the transition can take.

A Top Down Initiative with a Central Role for the Research Funding Agency

Two factors drive the current push in the UK for greater access to administrative data. First, recent research using this type of data has changed policies and shown the societal value of such access. For example, the creation of the National Pupil Database, which tracks all pupils in state-funded schools starting from 2002 has led to a better understanding of the drivers of students' outcomes, and to more reliable estimates of the effects of policies. Using this database, Dustmann et al. (2011) and Wilson et al. (2011) have for example shown that ethnic minorities tend to outperform pupils from White-British backgrounds. This changed the way policy-makers view the relationship between ethnicity and student outcomes. The database is also widely used to evaluate the impact of different educational interventions as part of randomized control trials (see Section 13.5).[8] Linking these data to higher education records has shed new light on the policy debate on access to higher education by showing that access is not so much driven by socio-economic characteristics (raising concerns about socially biased admissions) than by prior academic performance (Chowdry

et al., 2013). As another example, Best and Kleven (2015) have used data on housing transactions to document the distortionary impact of several features of the housing tax in the UK, with the policy subsequently changed.

The second driver of change is the concern that the UK may be falling behind in research if it does not grant its researchers access to data.[9] Access to quality data is essential for high-quality research and, worldwide, access to administrative data tends to be limited to researchers based in nationally accredited institutions. Therefore, providing access to UK-based researchers is critical to ensure their ability to carry out research at the highest level. However, access has been variable and difficult to date. There are legal barriers, of course (such as determining whether data access requires consent by the individuals that the data cover). There have also been cultural barriers, such as unwillingness to hand over data even when it is legal to do so, or institutional barriers, such as lack of human resources, preventing the extraction and preparation of the data for research. In addition, intrinsic data limitations, such as the inexistence of national IDs, make linking difficult in the UK.

The UK Administrative Data Research Network (ADRN) was set up in 2014 as a partnership between universities, government departments and agencies, statistical agencies, funders and researchers to foster access to linked de-identified administrative data.[10] It is funded by the Economic and Social Research Council (ESRC) and relies on two building blocks: an administrative data service (ADS) and four administrative data centres (ADRCs).

Ensuring the ethical and lawful use of data and building trust are at the core of the architecture put in place: projects are screened, researchers are trained, the data are de-identified, access is provided in secure environments and all outputs are screened to ensure identification is not possible. The model being put in place is a variation on the mediated access model. The Administrative Data Service helps researchers prepare their proposals and trains them for accreditation to use de-identified administrative data. It also negotiates access to data with the different statistical agencies. Once a proposal is approved and data owners have agreed to provide access, the ADRCs then provide secure access to the data, either on-site or virtual, depending on the preferences of the data owners.

The ADRN is new but it has already received a large number of applications from researchers. This is promising. Challenges remain, however. The existing legal framework for access to confidential data in the UK is fragmented. By law, the data must be destroyed 5 years after the start of the project, raising concerns about replicability. Changes are under way and the ADRN is contributing to the efforts by identifying remaining legal barriers. The main challenge, however, is to increase acceptability by the wider public and data owners. Privacy concerns remain pervasive and understanding of the societal value of research on administrative data is not largely shared.

Partnerships Between the Statistical Agency and Universities Driving Change

Developments in Catalonia illustrate the power of bottom-up initiatives as a source of change when other conditions are ripe to leverage them. The push for greater data access there came from a request received in 2010 by Idescat, the Statistical Office of Catalonia. Two researchers, Caterina Calsamiglia and Maia Guëll, had collected data on school choice (submitted preferences) and school enrolment from the city of Barcelona. The fact that they had obtained these data was in itself remarkable, and the result of lengthy trust-building with the department of education of the city of Barcelona. The researchers had identified interesting patterns in the school application behaviour of Barcelona families and consequently raised the interest of the Department of Education to understand the impact of socio-economic background on this behaviour. This required linking the data with census data.

The data request came at a good time for Idescat. The institute was transitioning from the traditional (stove pipe) linear model of data collection and production of statistics, where different databases covering different aspects of the same statistical object are kept distinct, to an integrated model, where common descriptors in different databases allow for their integration, and new data are seamlessly integrated with preexisting data as they arrive. Moreover, the culture of the organization was receptive to change: the institute is a relatively young and small organization and it had just appointed a new director, Frederic Udina, with a prior career in academia.

The data request was used by Idescat to learn about protocols for secure data access elsewhere (notably Norway and Denmark) and develop them in collaboration with the two researchers. A contract was eventually signed between Idescat and the Barcelona Graduate School of Economics (GSE), to which the two researchers were affiliated. The linking and de-identifying were carried out by Idescat for the researchers.

The contractual and logistical arrangements used for this first exercise served as a template for other data requests. It eventually led to the signature of a framework agreement between the Barcelona GSE and Idescat. The agreement establishes a partnership between the two organizations, where Idescat links and de-identifies the data (possibly first negotiating access with third party owners) and Barcelona GSE provides technical support and some manpower. A scientific committee with representatives from both the Barcelona GSE and other leading research institutions screens data requests and projects and oversees rules. The form of access is decided on a case-by-case basis depending on the re-identification risk that the data contain. Partnerships between Idescat and other research institutes are being developed alongside the same model.

13.2.4 Concluding Remarks

Microdata, and in particular linked microdata, are a goldmine for research. Several chapters in this volume allude to their potential to generate significant research breakthroughs in their respective fields. The good news is that progress is happening, not only in the countries that have traditionally been at the forefront of giving access to researchers, but also elsewhere. When progress takes place, there is usually a common understanding that making data available to researchers generates useful knowledge for policy and society, and provides a leading edge to domestic researchers. Researchers have an important role to play in testing the laws, opening doors, and showing the societal value of the research produced with this type of data. The current European legislative framework for access to microdata is adequate, but, at the time of writing, there are concerns that the new data protection regulation at the EU level could significantly restrict access to personal data. History will tell.

Of course, providing access is costly. In addition, there exist risks of breach of data confidentiality. These factors can lower the appetite of data owners to grant access, with high nonlegal barriers to access as consequences. From this perspective, the creation of a data mediator, such as NSD in Norway or ADS in the UK, whose main mandate is to service data requests by researchers, is particularly attractive. This model also seems particularly scalable in the presence of multiple data owners. Clearly, however, each country will need to develop its own variant that is compatible with the current organization of their statistical system and the incentives in place.

13.3 Data Standards and Cross-Country Datasets

By Roberto Barcellan, Peter Bøegh Nielsen, Bram De Rock, László Halpern, Joseph Tracy, and Lisa Wright

This section discusses the challenges and recent progress towards greater cross-country data harmonization and integration. Cross-country variations in data standards and data definitions are big obstacles to multi-country research. Comparative analyses are nevertheless crucial for better understanding the scope for replicability of policies across borders (i.e., is the experience of country *X* relevant for country *Y*?). Moreover, the recent economic crisis highlighted the high degree of interconnection of world economies and financial markets, and the inability of purely national databases to render an accurate picture of their operations.

Improvements in cross-country data are taking place at two levels.[11] First, there is a policy push in response to the crisis to increase cooperation among central banks and statistical offices to *produce better and more reliable*

macroeconomic indicators at the global level. We discuss how the G20 Data Gaps initiative is driving change at this level. Data access for researchers will, however, be limited to macro indicators, at least in the foreseeable future.

Second, there are efforts to harmonize and link *existing microdata* across different countries. The challenges there are enormous. We start by reporting on the FP7-funded project MAPCOMPETE, which aimed to produce an inventory of indicators of competitiveness, and assess the data accessibility and availability to compute them. We then discuss several ongoing initiatives among statistical agencies in Europe to create harmonized, cross-country, linked microdata. Among other things this will demonstrate that creating the data of the future is not a job that can be done by a few statistical agencies or a small group of researchers. Private data firms also have a role to play in the research landscape to harmonize, link and repackage publicly available data from different sources and countries. We discuss the experience of Bureau Van Dijk, a publisher of global business information, from this perspective.

13.3.1 The Lessons of the Financial Crisis for the Data Environment

While the recent economic crisis did not necessarily result from a lack of proper economic and financial statistics, it nonetheless caught supervisors, policy-makers and investors unprepared to understand its development and impacts in key areas poorly covered by existing datasets. For example, policy-makers and supervisors were unable the weekend before the Lehman Brothers bankruptcy to obtain information on aggregate exposures by each large bank. Banks could not provide this information, partly because of a lack of internal appropriate data systems, but also partly because Lehman consisted of thousands of legal entities. This complicated the planning for the possible Lehman bankruptcy.

The Lehman example illustrates two important challenges for business and economic data. First, firms operate in complex settings but the data structure does not always capture this complexity. This makes it, for instance, very hard to have a precise idea on the aggregate exposure of a given bank or firm. Second, even when the data exist at the micro level, it is essential to be able to link these datasets across countries, time, and legal entities.

Importantly, the data gap identified by the Lehman bankruptcy was not necessarily a problem of *quality* of economic and financial statistics. The problem was the combination of the lack of comparability across countries with the presence of interconnections amongst economies and financial institutions. This combination implied that exposures taken through complex instruments and cross-border linkages of financial institutions were *not* covered by existing data.

The lack of relevant and comparable data, available in a timely fashion, brought statistical authorities under scrutiny. It was clear that they did not

have the capacity to accurately detect, assess and forecast the consequences of the financial crisis. This profoundly impacted the way official statistics are organized. The rest of this section describes several initiatives that were taken in response. The bulk of these initiatives are policy-driven, but the greater awareness for the need to have better and more comparable data also benefits researchers through greater data harmonization and improved linking opportunities.

13.3.2 The G20 Data Gaps Initiative

After the crisis, the G20 countries became fully aware of the need to set up an international framework to address the challenges faced by statistical offices. They called on the International Monetary Fund (IMF) and the Financial Stability Board (at the time called Financial Stability Forum) 'to explore gaps and provide appropriate proposals for strengthening data collection'. This resulted in the report 'The Financial Crisis and Information Gaps', in which the IMF and the Financial Stability Board identified data gaps and presented a set of 20 recommendations to be implemented in the years to come.[12] The 20 recommendations are divided into four different, albeit interconnected groups: built-up risk in the financial sector, cross-border financial linkages, vulnerability of domestic economies to shocks, and improving communication of official statistics. Most of these recommendations concern financial stability indicators, but some of them relate directly to real economic indicators. The objective is to collect policy-relevant data that can, when needed, lead to early warnings.

In parallel, the Inter-Agency Group on Economic and Financial Statistics (IAG) was established to coordinate statistical issues and strengthen data collection. The IAG comprises the Bank for International Settlements (BIS), the European Central Bank (ECB), Eurostat, the IMF, the Organisation for Economic Co-operation and Development (OECD), the United Nations, and the World Bank. The IAG may be seen as a global facilitator. It coordinates international agencies to limit the duplication of efforts at the international level as much as possible. As the crisis had demonstrated a need to enhance communication, the IAG also promotes data provision and dissemination, particularly among the G20 economies. A website, www.principalglobalindicators.org, provides access to comparable economic and financial indicators for the G20 economies and the FSB countries.

The work undertaken by the Financial Stability Board with respect to two of the G20 data gaps' recommendations, namely Global Network Connections (R8) and Systematically Important Global Financial Institutions (R9), provide a good illustration of the challenges involved in the task. As the Lehman example made clear, the financial system is increasingly global. It is also less bank-centric, with other increasingly important nonbank intermediaries. However,

the pre-crisis data environment was largely domestic and bank-focused. There was almost no data sharing across national supervisors. This created an increasing mismatch between the reality of financial markets and data structures.

In the middle of the Lehman crisis, national supervisors decided to start collecting risk exposure data for their large banks. In the US, the Federal Reserve Bank of New York cleaned and analysed the data before producing anonymized reports for the supervisors. With time, the accuracy and frequency of the data reports improved, but the anonymized nature of the reports limited their usefulness.

The work of the Financial Stability Board under the Data Gaps Initiative formalizes and expands on these early efforts, with four differences. First, the data now flow from the banks to their home supervisors on to a hub at the Bank of International Settlement. This is a crucial difference since global exposure can only be assessed from merging the data from all supervisors. Second, coverage was expanded to nearly all global systemically important banks. Third, the range of data collected was also expanded, from risk exposure of global systemically important banks in phase I of the project, to funding relationships and indicators for both micro and macro-prudential regulation by the last phase. Fourth, banks and counterparties' identities are not masked in internal reports (subject to jurisdiction confidentiality restrictions). The last phase of the project is scheduled for 2016.[13]

A critical challenge faced by the Financial Stability Board was to set up a multilateral framework for data sharing that would overcome the national supervisors' reluctance to share their data. The agreed upon framework relies on three principles: data ownership by national supervisors, reciprocity (supervisors do not receive reports unless they share their data) and unanimity (all supervisors agree on procedures). Only reports are shared by supervisors; there is no direct access to the pooled data. Given that the ultimate goal is to aid macro-prudential surveillance and regulatory design, the BIS, the IMF and the Financial Stability Board will likely have access to (some of) the data, but access arrangements are still under discussion at the time of writing. Access to researchers is not discussed.

A second important challenge was to properly assess the structure and interconnections in the global financial network in order to adequately identify the risks and vulnerabilities of the system. Special attention was devoted to assessing the links between global systemically important banks but also nonbank intermediaries, and institutions that may not be systemic but are deeply interconnected with several global systemically important institutions. These links can be due to ownership relationships, funding dependencies or other sources of spill-over. Data collection integrates both the exposure side and the funding side for both the consolidated and connected entities, thereby avoiding the earlier biases and gaps due to reporting groups' organizations.

There was also an interesting trade-off with respect to the frequency of data collection, between monitoring requirements (for which monthly data are enough) and crisis management (for which daily data are more adequate). The view held at the time was that if banks designed their data systems for normal reporting, they would not necessarily be able to accommodate daily reporting when needed. A weekly reporting frequency was selected, with the side benefit of forcing banks to significantly improve their internal data analytic capabilities.

13.3.3 Linking Existing Microdata

There is no doubt that the G20 Data Gaps Initiative is leading to better, more comparable, and more timely cross-country data. However, access to researchers is limited to the macro-indicators. In this section, we evaluate existing initiatives and prospects for harmonizing and linking existing microdata across countries.

Data Issues Assessed by MAPCOMPETE

The analysis of the competitiveness of firms is mostly done on the basis of macro-indicators (i.e., measured at the national level). This is partly because these macro data are easy to communicate and the measures can be computed relatively straightforwardly. However, competitiveness of a country (or a sector) is an idle concept, since competitiveness is related to the ability of firms (in a given country or sector) to efficiently produce goods and services. As such, it is crucial to start the analysis from the firm level and take firm heterogeneity into account, in order to fully grasp the impact of, for instance, the financial crisis.

This was the starting point of the FP7-funded project MAPCOMPETE (www.mapcompete.eu). The objectives of MAPCOMPETE were to study and compare existing firm data from 25 European countries and produce new micro (i.e., bottom-up) indicators related to productivity, firm dynamics, international activities and other aspects related to competitiveness. When possible, these micro indicators were linked to available macro indicators. All this allows for the replacement of the macro indicators by the more informative (because they take firm heterogeneity into account) micro indicators.

Although the benefits of these new and comparable micro indicators should be obvious, it is impossible, in the current data environment, to construct them for many countries. MAPCOMPETE identified several reasons for this (Castellani and Koch, 2015). First, there are important differences across countries in terms of availability and accessibility of the data, as well as data definitions (including different units of observations or timing). Second, it is often simply impossible to uniquely link existing datasets inside a given country. Finally,

not all European countries are equally convinced about the potential and benefits of creating these micro indicators of competitiveness. This is problematic because, in the absence of a centralized and coordinated view on data collection, policy recommendations will, by construction, remain seriously limited.

Establishing Internationally Harmonized Statistical Databases

As European business statistics are to a large extent based on common EU regulations, the central business or economic statistics (such as, e.g., structural business statistics, international trade in goods or services statistics and R&D statistics) are de facto harmonized and thus in principle comparable across the 28 member states of the European Union. However, due to the stove pipe production process of official statistics, i.e., each statistic is produced in isolation from one another, these micro data all focus on a limited number of aspects of the firm. To fully analyse the impact of the financial crisis on European firms, or simply get a full picture of firms or sectors, we need to be able to link these existing datasets.

Eurostat has recently funded a series of micro data linking (MDL) projects in order to 'modernize' the European enterprise and trade statistics (the MEETS programme). The basic driver behind this development is twofold. On the one hand, the analysis of cause and effect requires linking micro data, which in turn implies breaking the traditional stove pipe model of statistical production. On the other hand, there is the practical argument of minimizing respondent burden. This not only increases the return on investment from these existing detailed micro datasets, it also means that the statistics can be more adequately used to guide policy-makers. In this sense, these micro data linking projects are clearly complementary to initiatives such as the G20 Data Gaps Initiative, which focus more on new and aggregate indicators.

Figure 13.1 shows the methodology developed for these projects.[14] One of these projects covers, for example, the international organization and sourcing of business functions of firms in nine European countries, for which harmonized datasets based on input data from five different statistical sources in each country were compiled for the period 2008–2012. The resulting data will be used to conduct micro-level economic analyses of essential questions related to social and economic conditions, for example, which factors generate economic growth or how international activities of enterprises influence their economic or job creating performance.

These projects, as well as others such as the project described in Section 13.2 to link administrative datasets across Nordic countries, show the potential of micro data linking for the development of statistical information on the international dimension. Micro data linking serves as an appropriate method to analyse typical research questions on cross border activities ('what kind of enterprise is trading?' instead of 'what do countries trade?'), firm heterogeneity

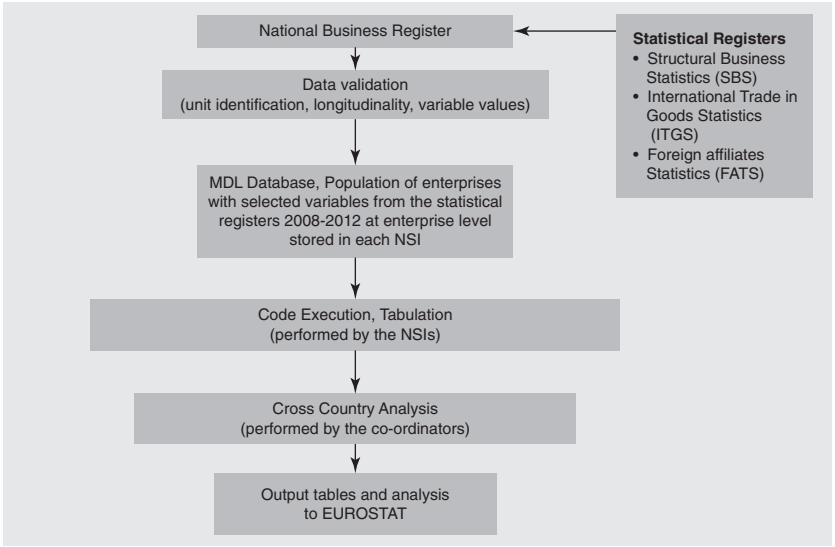

Figure 13.1 Typical organization of micro data linking (MDL) projects.

in systems of national accounts ('what kind of enterprises contribute to GDP?') and the organization of cross-border production processes ('what parts of the business organization move up or down the value chain?').

However, implementing micro data linking to a full extent, both across countries and within countries, is resource-intensive. It also requires coordination amongst and the cooperation of all national statistical offices. As such it is fair to conclude that we still have a long way to go.

The Role of Private Data Firms in Producing Harmonized Firm Data

Statistical offices are not alone in their efforts to harmonize cross-country data. Private data firms have long been involved in collecting, harmonizing and repackaging publicly available business data. Bureau van Dijk (BvD) is one of them. BvD is an established commercial company, collecting financial and Merger and Acquisition (M&A) data on more than 150 million firms worldwide, and a prime example that it is indeed possible to gather, and link, timely available micro data.

The M&A database produced by BvD illustrates some of the challenges involved in creating reliable and harmonized data from publicly available data. An M&A transaction involving two public companies does not pose any difficulties because it follows formal disclosure rules set by the stock exchanges.

Transactions involving privately held firms are trickier. To construct their international M&A database, BvD relies on several data sources, such as official company filings, data from stock exchanges and regulatory bodies, but also news services and M&A advisor data. The main added value of BvD comes from linking all these data sources in a standardized way. This requires the creation of unique identifiers, a mapping of local regulations to international standards and a harmonization of financial formats to allow for international comparisons. All this is done by local experts who understand the local specificities (and languages) and more than 100 specialists active in different fields (company data, scoring companies, stock market info, etc.). This network of experts, combined with the raw data, ensures that the resulting database is both reliable and accurate.

The business model of these private data firms is, of course, not comparable to those of statistical offices. It is nevertheless interesting to note that their work both complements the work of national statistical offices (by covering more data sources and linking them), and substitutes for them (when they sell access to their linked data to researchers). The message to statistical offices should be clear. On a daily basis, and even in the pre-crisis era, data firms like BvD show that it is possible to gather harmonized firm and financial data required for adequately advising policy-makers.

13.3.4 Towards a New Data Environment

The financial crisis has made it clear that the pre-crisis data environment was not capable of adequately informing regulators and policy-makers. As a response, statistical agencies have changed their data collection drastically. This is based on two essential ingredients. The first ingredient is a centralized view on data gathering. Firms and banks operate more and more in a global and interconnected world. To improve the comparability across countries and, equally importantly, to better grasp the complexity of markets, data standards should be the same across countries. The second ingredient, which is related to the previous one, is data warehousing. There already exist many high-quality databases, but one cannot fully exploit their potential because it is simply impossible to link them to each other in a consistent and country-comparable way. In a setting where institutions are often reluctant to share private information, the new data architecture will need a new governance system, such as the one put in place by the Financial Stability Board in the context of the G20 Data Gaps Initiative.

The data environment of the future will require some central coordination and guidelines. The Inter-Agency Group is a nice example of this (and it should continue its efforts). Equally important is to ensure structural funding for this type of data collection and linking.

A final recommendation is related to collaboration with academic researchers. An important motivation for the G20 Data Gaps Initiative was to develop data to enable policy-makers to analyse the impact of the financial crisis. However, the ultimate goal should, of course, be to prevent the next financial crisis (or at least to be able to provide early warnings). As markets keep evolving, researchers can play a crucial role in identifying new indicators of interest and new ways to use the data to better grasp the complexities of financial systems and markets.

13.4 Researcher-Generated Databases

By Bram De Rock, Arie Kapteyn, Julia Lane and Gugliemo Weber

Partly in response to frustration with the poor cross-country comparability of data, restricted access to data and/or simply the lack of collected data on issues of interest to them, a number of researchers have been involved in large-scale data collection, which they have turned into databases publicly available for the entire research community. This section describes three such highly successful data collection projects and the experiences of the researchers, amongst whom three of the authors, who played leading roles in their creation.

The chosen data projects provide a sample of the diversity in approaches and designs, but also of the challenges that such projects face. The first data project is the Measurement and Experimentation in the Social Sciences (MESS) project, an internet-based survey panel database created by Arie Kapteyn, and Marcel Das and Arthus van Soest from Tilburg University. Kapteyn has in the meantime led other internet-survey panel data projects including the American Life Panel and the Understanding America study. This first case study highlights the advantages of module-based internet surveys.

The second data project is the UMETRICS programme, a large-scale data project involving hundreds of partners in the US to better understand the impact of research funding. Julia Lane is one of the masterminds of UMETRICS to which she is contributing her extensive experience with other data projects, including the Longitudinal Employer-Household Dynamics (LEHD) programme in the US, and the NORC/University of Chicago data enclave. This second case study illustrates the added value of bringing several partners together to tackle large-scale data collection projects.

Our last case study is the Survey of Health, Ageing and Retirement in Europe (SHARE). In contrast to the MESS project, SHARE is a more traditionally organized social survey. The distinguishing feature of SHARE is that data are gathered in 21 countries and one region. SHARE is recognized as a European Research Infrastructure Consortium (ERIC). Guglielmo Weber is country team leader for Italy and deputy director of the SHARE Consortium. This case study

illustrates the challenges linked to organizing such multinational data projects when funding is done at the national level.

The first three sections will, for each specific context and its corresponding objectives, shed light on the benefits and costs of researcher-generated databases. In the final section we discuss the general lessons learned from these projects and provide some recommendations for future data collection.

13.4.1 Measurement and Experimentation in the Social Sciences Project

The MESS project was developed in the Netherlands during the period 2006–2013 with funding from the NWO, the Dutch national research council. It was a collaboration between 8 of the 13 universities in the Netherlands to develop a very innovative and rich social survey on a panel of Dutch households. MESS resulted in the core Longitudinal Internet Studies for Social Sciences (LISS) panel (www.lissdata.nl), containing relatively standard household demographics, to which several modules with specific questions or experiments were added.

Given that the traditional way of collecting household data is time-consuming, and thus expensive, it was decided to save costs in two specific ways. First, the data are collected via the Internet and via innovative tools such as time use apps and internet bathroom scales. These features save time and provide convenience for respondents. They also remove the need for interviewers, which hugely reduces the cost of collecting this type of data. For instance, the cost per interview for the Panel Study of Income Dynamics (PSID) and the Health and Retirement Study (HRS) is between 1,200 and 1,500 US dollars. Data collection in the LISS panel is four times cheaper. It is also two to three times cheaper than traditional social surveys in the Netherlands.

Second, the dataset is constructed in modules. The LISS panel is the core module. It contains the household information typically available in social surveys. In addition, researchers from around the world could suggest extra modules, with for example, extra questions or an experiment. These extra modules were, of course, motivated by the specific research questions of these researchers who hoped to derive a publication from it, but, by design, the collected data were made immediately available to the whole research community. This innovative and flexible system optimally avoids overlap between several questionnaires, thereby reducing the marginal cost of each extra module. Moreover, data from different modules can be linked, allowing researchers to leverage rich data on many dimensions of household behaviour.

The MESS project was highly successful: 131 projects and experiments were conducted during the project, involving 85 different universities and institutes around the world. Unfortunately, it did not receive funding from NWO for the second phase of the project (2013–2023) and was therefore discontinued. The

motivations for the funding decision were twofold. MESS had received infrastructure funding and there was an understanding that the infrastructure should be ready after seven years. In addition, collecting this type of data was no longer a priority. This decision illustrates that funding decisions do not always fully reflect the infrastructure needs of social sciences. In social sciences, collecting data is the main type of needed infrastructure. To capture all dimensions of life, it is crucial to follow individuals and households over a long period. This implies that collecting data is a long term, and actually a 'never-ending' project. While at some point the Netherlands was at the frontier of collecting household data, stopping the funding no longer allowed maintaining the LISS panel and its corresponding modules. A large and attractive infrastructure is wasted.

13.4.2 The UMETRICS Programme

The objective of the UMETRICS programme (iris.isr.umich.edu) is to construct an integrated dataset to measure the impact of public funding on research in the US. The ultimate goal is to answer questions such as 'how much is spent in each discipline and do we get value for money?', or 'does demand for funding by potential science performers imply a shortage of funding or a surfeit of other performers?'.

Although there are several datasets available containing useful information related to the impact of research funding in the US, no single dataset covers all relevant dimensions of both inputs and outputs of research. UMETRICS brings together several US partners (including statistical offices, more than 100 universities and researchers from different disciplines) in order to properly combine existing datasets and/or to create new data. For instance, new text mining approaches and automatic data creation methods were designed in order to create new measures on the impact of 'research output'.

UMETRICS shows that collaboration between research stakeholders can be highly effective. This process starts by identifying the existing gaps (i.e. existing datasets that are not linked or simply missing data). Consequently, funders, universities and the research community need to be convinced to build coalitions to fill these gaps. Building coalitions is essential to obtain the necessary funding and the access to existing data. Finally, a team of dedicated and flexible collaborators has to be assembled to carry out the project. This requires visionary researchers, designing for instance new data mining methods, as well as operational producers implementing these algorithms or adapting existing ones.

Visibility and exchanging ideas were two crucial ingredients during this process. In the development phase, this was done through newsletters, research

workshops, and presentations. Subsequently, several user-friendly products, such as dashboards, short reports and data access, were developed.

UMETRICS has resulted in a unique data source, including novel impact measures, frequently used by researchers and policy-makers. It has generated research published in leading academic journals, such as Science, containing answers to questions similar to the ones mentioned above. It has the potential to be a major source of innovation and evidence-based policy recommendations on how to (re)design research funding schemes.

13.4.3 The Survey of Health, Ageing and Retirement in Europe

Population ageing is one of the important challenges of the twenty-first century for which we need to understand the individual, social and economic impact. SHARE (www.share-project.org) is an ambitious social survey that gathers, in 21 European countries and one region, data on health variables, socio-economic status and social participation for a representative sample of individuals aged 50 or more. Moreover, since ageing is a historical process, the same individuals are also followed through time. As of 2015, 6 waves of data which are, in principle, perfectly comparable across countries, have been completed.

The funding for the first three waves of the SHARE data was mainly provided by the European Commission. Then, in 2011, SHARE was recognized as a European Research Infrastructure Consortium (ERIC). This means that the main funding had to be provided by each individual country. This decentralized system put the sustainability of the SHARE project under severe pressure. For instance, in 2015, 65 different sources had to be combined and four countries did not obtain enough funding to continue to fund the project.

The SHARE data are frequently used by both researchers and policy-makers. Since the start of SHARE in 2005, more than 1,200 papers have been written on the basis of the collected data in one or more countries. SHARE has also provided evidence to support policy in Member States (e.g., retirement age and work conditions in France), at the European level (e.g., longterm projections of the costs of population ageing for DG EcFin) and at the international level (e.g., (re)migration corridors and social support for the World Bank).

The popularity and usefulness of the SHARE dataset are due to its interdisciplinary nature (it involves economists, sociologists, epidemiologists, geriatrics and psychologists) and its multinational and longitudinal aspects. Data that are comparable across countries are important to deal with big societal challenges, such as ageing. Otherwise it is impossible to compare, for instance, the different welfare state policies or to link individual decisions to institutional background variables. Given the many different languages in Europe and the sensitivity of many of the questions in the survey to their actual wording, this is of course not straightforward to implement. The same holds for the longitudinal

aspect, which requires either convincing individuals to keep on participating and/or a complex tracking of individuals. However, the many academic and policy papers convincingly demonstrate the relevance of gathering this type of data.

13.4.4 Lessons from Successful Researcher-Led Databases

We briefly presented three very successful examples of researcher-generated databases. An important common feature of all these data projects is that they are governed by scientists. This ensures the quality of the data and, even more importantly, the relevance of the data. For instance, allowing researchers from around the world to add well-motivated questions and experiments to the core modules of the LISS panel increased its impact and take-up by scientists. Of course, data collection projects should be subjected to scientific review, but experience shows that giving researchers the opportunity to design their own databases can stimulate innovation in research and increase the returns on investment from data collection.

Another important feature in all three examples, and in many other large-scale researcher-led data collection projects, is that the collected data are made freely accessible as soon as possible for researchers. To some extent this should be obvious, since the data gathering is often supported by public funds. But, if we look for instance at the output that the SHARE data generate, it is also clear that this open access policy is essential to stimulate academic research and maximize the return on investment of the funds.

Researcher-led data collection does not, of course, minimize the role of statistical offices. On the contrary, statistical offices play at least two crucial and complementary roles. First, their role in setting data standards is vital in ensuring quality and international comparability. As discussed in Section 13.3, there is a need for central coordination in this respect. Second, there are huge benefits from linking the data gathered by statistical offices to these researcher-generated databases. The Nordic countries and the UMETRICS programme discussed above provide convincing examples that demonstrate that such linking is not only feasible, but boosts the research value of researcher-generated databases.

Data is the main infrastructure for most disciplines in social sciences. However, the nature of this infrastructure is in sharp contrast with, for instance, the natural sciences. Data gathering is, so to speak, a never ending project. This does not mean that one should keep on investing in a specific database, in particular if the data loses its relevance. But, it does imply that funders need a long-term vision and realize that, if one stops funding a database, it often automatically reduces significantly the value of the existing infrastructure. Funding mechanisms for social sciences should take this into account and not tie funding

to today's policy problems. Instead, there should be clear guidelines for review panels analysing renewal requests.

Relatedly, and as the experience of SHARE shows, decentralized funding schemes are a serious threat to the sustainability of longitudinal and international databases. Not all national governments are equally interested in evidence-based research, which implies that it is almost impossible to obtain funding in these countries. This pleads for centralized funding for this type of international data collection projects.

Private funding can play a crucial role during the earlier stages of large-scale data collection projects. In contrast to public sources of research funding, private foundations are more flexible. They are more mission and people-oriented, and are therefore willing to take more risks and accept longer time horizons. Private funding has played a crucial role in the success of UMETRICS. Likewise, the US Longitudinal Employer-Household Dynamics programme (LEHD) had been funded for 20 years by private funds before becoming a national (publicly funded) statistical programme in 2008.

13.5 Data Generation in Controlled Environments

By Colin Camerer, Bruno Crépon, and Georg Kirchsteiger

For a long time, economics has been considered a nonexperimental science, and empirical work was done only with field data. This has changed drastically during the last four decades when more and more economists have begun to conduct experiments. One can distinguish between two types of economic experiments: Laboratory experiments and randomized control trials (RCTs) in field settings. This section describes both methods, their contributions to economics and recent developments in methods and applications of these two kinds of experiments.

13.5.1 Laboratory Experiments

The basic idea of a lab experiment is simple: Participants are put into an artificially-designed and controlled economic situation in which they make decisions. The canonical experimental design creates endowments and induced preferences over outcomes, and specifies a set of rules which compile participants' choices into outcomes. The rules can be very simple, in decision-making experiments, or specify a game-theoretic or market structure in more complicated experiments. The target of discovery is what choices people actually make.

Lab experimental situations are 'real' in the sense that the participants' decisions have an impact on the rewards they receive for participation in the

experiment. Monetary reward is typically used because almost everyone is motivated to earn more money, and marginal reward is positive (there is no satiation, as might occur with points, public reports of success, or other non-monetary rewards). In order to test economic theories, the actual decisions of the participants are recorded and compared to the decisions predicted by economic theory. Careful control of endowments and preferences, and explanation of the rules determining outcomes, create conditions under which theoretical predictions should apply. Experimental results contradicting established economic theories are also used to guide the development of new, better theories.

Experimental data generated in a carefully designed and controlled environment have many advantages compared to field data. First of all, the researcher observes directly the variables of interest – no proxies are needed. All relevant variables can either be induced by design, or measured. One can also easily compare the results of treatments differing only in one aspect of the experiment. This allows for a direct test of a causal relationship. Experiments can also be purely replicated – that is, replicated with the intent of reproducing original experimental conditions as closely as possible. Pure replication checks whether the results of the initial experiment were a false positive, are robust to inevitable small changes, or resulted from biases in investigator reporting or journal selection practices. To allow for replications, experimental papers typically include the written instructions given to the participants. Going even further, it is common practice, and is also required by some leading journals, that researchers make their experimental data and analysis code available for the purpose of making replication easy. Note that RCTs in field sites typically require access to a group of participants, and may be more difficult to purely replicate in comparison to lab experiments.

Two major concerns about results of lab experiments are the robustness and the generalizability of experimental findings. Fortunately, robustness can be easily checked by doing experiments using theoretically motivated variations of the initial experimental design. Indeed, robustness explorations have been done in many areas, particularly when initial results contradicted received theory. Another important concern is whether lab results generalize to particular field settings. First note that most lab experiments test theories which purport to apply generally to many field settings. And since field settings differ on many dimensions, if one desired to maximize generalizability of a lab experiment there is no obvious metric for doing so. However, in many cases, lab designs have taken special care to match corresponding field settings as closely as possible. In general, when the design in the lab closely matches the conditions in the field, observed behaviour is usually very similar, as judged by responses to changes in variables such as prices or description (Camerer, 2015).

Ideally, there is a dynamic complementarity between field data, experimental data research, and economic theory. Take ‘gift exchange’ as an example:

Akerlof's 1982 theory of involuntary unemployment assumed that workers and firms are prone to exchange gifts, in which firms pay higher wages and workers repay the wage gift with high effort. In theory, this mechanism is used by firms to induce workers to provide reciprocal effort in situations otherwise characterized by moral hazard. This hypothesis was confirmed by many experiments (beginning with Fehr et al., 1993), which in turn led researchers to develop theories to explain reciprocity (e.g., Dufwenberg and Kirchsteiger, 2004) and to look for similar kinds of reciprocity in the field (see e.g., Falk, 2007). The interplay between the lab and field data can also go the other way. For example, the winner's curse was first observed in field data on oil-lease bidding (Capen et al., 1971). It was then extensively analysed in lab experiments (summarized by Kagel and Levin, 2002).

Lessons from Lab Experiments

During the last 30 years, lab experiments have become well-established and central to economics research. This is reflected by the number of articles reporting experimental results that are published in economic journals. Between 2006–2010, about 100 experimental papers were published in leading general journals (*American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Review of Economic Studies*, and *Economic Journal*), and more than 350 papers were published in specialized journals (such as *Games and Economic Behavior*, *Journal of Economic Behavior and Organization*, and *Experimental Economics*).[15]

While there are also some experimental studies in macroeconomics and political economy, the huge majority of the papers (about 95%) investigate four broadly defined fields: Individual decisions, social preferences, markets, and games. This reflects the close relation of lab experiments with (micro)economic theory. Most of the experimental findings are quite robust up to plausible variations of the experimental design. These robust findings include:

Markets. In market experiments with induced supply and demand for a homogenous good, observed prices and quantities converge quickly to the market clearing equilibrium, in particular when trade is organized in a centralized manner ('double auctions'). This result holds for very thin markets (even with only three traders on each market side), for a huge variety of different demand and supply conditions, for multiple connected markets operating at the same time, for different subject pools, etc. (for an overview see part 1 of Plott and Smith, 2008).

Games. In many experimental games, the observed outcome coincides with the Nash equilibrium prediction. Systematic deviations from the Nash equilibrium prediction have induced the development of level-k and quantal response theory, which can (most of the time) explain the observed deviations from Nash.

Auctions. In experimental auctions, one typically observes the winner's curse in common value auctions, and overbidding in auctions where bidders' values are privately known and unaffiliated.
Social preferences. Participants do not simply care about their own earnings. Fairness, reciprocity motives and social image motivations also have an impact on their behaviour (see e.g., Fehr et al., 1998).
Public goods. In games of voluntary provision of public good, participants typically start with a contribution level of 50 per cent of their endowment. This is followed by a decrease of the contributions to 10 per cent. If participants have the opportunity to punish each other at a personal cost, contributions start and stay at high levels because low contributions are sometimes punished, with important cultural variations (Herrmann et al., 2008).
Monopoly. In experimental markets with a monopolist, the actual price demanded by the monopolist is below the monopoly price because buyers withhold demand, which disciplines the monopolist to lower prices.

In all these cases the experiments were typically replicated 10–100 times, either as pure replications or with an initial replication followed by changes to the original design to test the robustness of the findings. A typical example is provided by the public good experiments, with and without punishment opportunities (see Herrmann et al., 2008).

On the other hand, some experimental findings are not as robust, due to endogenous expectations, local norms, or other reasons. An example of this is provided by experimental financial markets. Some, but not all, of these markets experience price bubbles and, when bubbles occur, their size differs substantially.

Lab experiments are also used in teaching economics. More and more introductory classes in economics use experiments to bring theoretical concepts like the impact of incentives, or market equilibration, or social dilemmas, to life. This increasing use of simple economic experiments arises from the fact that the results are reliable. The dependability of economic experiments is similar to other experimental demonstrations, such as visual illusions that are used to teach principles of perception in cognitive psychology, and chemical reactions used in basic chemistry. On the other hand, experimental economics still has limited influence on the core courses in economics graduate training. Even at institutions such as Caltech, where most economics faculty do some experiments, PhD students can graduate without knowing anything at all about the methods or findings.

New Developments and Outlook for Lab Experiments

A new development in lab experiments is the use of biomedical tools. This started with the use of eye-tracking tools and the measurement of response

time. functional magnetic resonance imaging (fMRI), electroencephalograms (EEG), and causal administration of bioactive substances have also been used (typically in co-operation with neuroscientists) to get a better understanding of the detailed biological mechanisms underlying the observed behaviour, and of individual differences (e.g., Camerer, 2013). So far, the impact of this research on mainstream economics has been limited. It is unlikely that the economics profession will begin to use these new tools as enthusiastically as it has taken up 'conventional' (choice-based) lab experiments. However, as for the impact of psychology experiments, biological evidence may have some impact on theory, in understanding emotions, self-control, addiction, and other topics, even if those data are collected by noneconomists (or in occasional collaborations with economists).

Another important development is the use of advanced technology to recruit volunteer participants outside of the usual constraints of a college campus. Why should participants have to come to a physical lab, when modern technology enables them to participate remotely, perhaps even on mobile phones? To this end, a large number of experiments increasingly use online 'labour markets' such as Amazon Mechanical Turk ('*Mturk*'). An even more dramatic step is to conduct abstract experiments outside of the places where the researchers work, a method called lab-in-the-field (pioneered in economic anthropology, for example, Henrich et al., 2005 and see Haushofer et al., 2014). Lab-in-the-field experiments can address the important concern that experimental social science has traditionally oversampled highly educated and rich subjects from industrialized countries while striving to make generalizations about everyone on earth (Henrich et al., 2010).

13.5.2 Randomized Control Trials

Economic experiments can also be used to evaluate the impact of (proposed) policy measures. The basic idea of such Randomized Control Trials (RCTs) is the following. A policy measure is proposed in order to achieve a certain goal. Before the measure is implemented broadly, randomly chosen potential 'recipients' of the measure receive the measure ('treatment group'). Some other randomly chosen potential recipients do not get the measure, but are observed with respect to the variable(s) of interest ('control group'). Since the treatment and the control group are randomly selected, there is no systematic difference between the two groups except for being subject to the policy measure or not. Therefore, any observed difference between the two groups, after the measure has been implemented, can be attributed *causally* to the policy measure and the result is an unbiased predictor of the impact of the proposed measure. This allows assessing the impact and efficiency of the measure and to improve its design before it is rolled out generally.

The provision of summer jobs to disadvantaged teenagers in order to reduce their criminality provides an illustration of the approach. Some disadvantaged youngsters (the treatment group) get summer jobs, while others do not (the control group). To assess the efficiency of this programme, the crime rates of both groups are compared. To make sure that the observed differences in the crime rates are indeed due to the summer job, and not due to other systematic differences between the two groups, both groups are randomly chosen. In an RCT study conducted among 1634 disadvantaged high school youth in Chicago, Heller (2014) found that such a programme reduced violence by 43 per cent over a period of 16 months and 3.95 fewer violent-crime arrests per 100 youth.

RCTs have some important advantages over field data to evaluate the impact of policy measures. The main advantage is the ability to establish and assess the causal link between the policy measure and the outcome of interest. Results are obtained in a clear, understandable and transparent framework. The policy decision is a complicated and long process. Thanks to the transparency and palpable scientific rigour surrounding their use, RCTs provide results which can inform the policy-making process efficiently about the right programmes and their impacts.

Lessons from RCTs

There has long been a demand by policy-makers and institutions for scientific evidence about the impact of policies. The first RCTs were implemented in the US to this effect by large nonprofit consultancies. Famous examples include employment programmes or changes in the unemployment insurance system (for an overview, see Meyer, 1995) or, more recently, changes in the health insurance system (the Oregon experiment, see Finkelstein et al., 2015). In some cases the results of these RCTs have led to major changes in policy, for example, the National Job Training Partnership Act Study (see Bloom et al., 1997).

Use of RCTs is not restricted to the US, however. In France, anonymous resumes have been proposed as a measure against discrimination on the job market. Behaghel et al. (2014) used an RCT to measure its potential impact and found it resulted in worsened outcomes for minorities. This led the French employment agency to abandon the idea. Martin Hirsch, the French High Commissioner for Youth, set up an 'Experimental Fund for Youth' in 2008, to encourage innovative programmes for youth as well as their rigorous testing and evaluation. In the context of this fund a large number of RCTs have been launched to address youth policy questions in education, health, housing, and employment.[16] The impact of RCTs on policy implementation has been strongest in developing countries. There also, there is a high demand for evidence. This has given rise to close co-operation between researchers, NGOs,

and aid donors, with RCTs used to evaluate programmes in fields like health, education, gender and agriculture.

The development of RCTs has achieved several outcomes. One first outcome of RCTs is their ability to update and, when needed, correct beliefs about the effectiveness of different policies. Education in developing countries is an example. Measured by additional years of schooling per 100$ spent, policies informing parents about the returns on schooling and policies on deworming of primary school children have turned out to be far more effective than other measures like cash transfers, merit scholarships, or free school uniforms.[17]

Microcredit provides another example where RCTs have led to a shift in beliefs about which policies are effective. Due to the widespread belief that microcredit is a strong tool to alleviate poverty, politicians as well as aid donors in many parts of the world have supported microcredit institutions. However, RCTs conducted in six countries (Bosnia and Herzegovina, Ethiopia, India, Mexico, Mongolia and Morocco) have found that while microcredits have a positive impact on the scale of activities, they do not significantly improve earnings of beneficiaries (Bauchet et al., 2011). These results changed beliefs about the promises of microcredit. Microcredit should be seen as one tool, among many others, to fight poverty.

Another outcome of RCTs is that, in some instances, it has been possible to test important aspects of economic theories. One example is the issue of cost-sharing. Many programmes involve some cost-sharing, where the recipient of a certain programme has to share some of the costs. Bed nets, which are one of the most effective tools to fight malaria, are one such example. They are usually sold at a subsidized price. The reason for the subsidy is the belief that otherwise many people would not buy it, so the demand would be too small. On the other hand, bed nets are not given for free because of the belief that people have to pay something to value a product. So the question is twofold: Is the demand price-elastic? If yes, does giving the bed nets for free reduce their use? To answer these questions, Cohen and Dupas (2010) ran an RCT where participants were first asked whether they were willing to buy the bed net at a price that was randomly determined. This allowed the researchers to determine the price-elasticity of the demand. In the second step, some participants received their bed nets for free, and use of the bed nets was compared between those who paid for it and those who received it for free. The results were clear-cut: the answer to the first question is yes, demand is highly price-sensitive. This replicates other studies looking at the price-elasticity of water disinfectant (Ashraf et al., 2010) or deworming drugs (Kremer and Miguel, 2007). And, interestingly, the answer to the second question is no: the use of bed nets was not influenced by whether it was given for free or not.

A second example of the use of RCTs for testing theories concerns displacement effects in labour markets. A concern about active labour market policies,

that is, policies used to bring unemployed back to the labour market, is that the benefits obtained by the beneficiaries of these policies come at the expense of workers who do not have access to these programmes. This concern has been around for long but, until recently, the presence of this effect or its magnitude had not been measured. Crépon et al. (2013) developed and implemented an experiment to address this question. The design was based on a double randomization. In a first step, markets were selected to develop or not the programme, generating 'test' and 'control' markets. In a second step, a fraction of the potential participants in test markets were randomly assigned to the programme and some to a control group. Comparing potential participants in the control group in test markets and control markets has shown that there is indeed a displacement effect and that it is substantial. Results showed a substantial improvement in the employment situation of beneficiaries, but no improvement of potential participants in test markets as a whole compared to control markets.

A final example is related to price incentives. One general belief is that for a price policy to have an impact, it has to substantially change financial incentives. For example, if it comes to subsidizing a product, the subsidy has to be large. RCTs have shown that this is not necessarily the case: large impacts can be obtained from very small incentives. One example is immunization programmes. These programmes often face the challenge that participation is low. Too few people start the immunization and/or too many drop out. In the context of an immunization programme for children in rural India, Banerjee et al. (2010) found that even a reward as small as one kilo of lentils considerably increases participation. The authors tested two different treatments. In 30 villages they installed reliable immunization camps and in 30 other villages they installed the same type of immunization camp combined with small incentives (one kilo of lentils for each child for each show-up). 74 villages served as control group. The authors found that twice as many children finished the immunization programme when it was combined with small incentives relative to when it was not.

New Developments and Outlook for RCTs

Faced with a policy design question, policy-makers have many ways to try to answer it. All these potential answers are based on ideas about mechanisms at play and on views about what matters or not. Simultaneously, there is a demand from policy-makers to learn about the impact of their programmes. RCTs are increasingly used to meet this demand. They have shown that not all the solutions work, that some beliefs about programme efficiency and underlining mechanisms are incorrect. They have also shown that the consequences of mistakes can be of first order importance.

RCTs allow accumulating two types of knowledge: about programmes that work or not, but also about mechanisms at work or not. This knowledge

enriches the information set of policy-makers when thinking about new policies. Because of their versatility and ease of implementation, RCTs have also allowed policy-makers to test new ideas and policies cheaply and are therefore contributing to innovative policies.

Challenges remain, however, regarding their external validity. This is an active and interesting area of current research. One external validity challenge for RCTs is that the selection into the treatment or the control group might have a direct impact on the behaviour of subjects (the so-called '*Hawthorne and John Henry effect*') and this might therefore lead to biased estimates of the policy impact. Aldashev et al. (2015) show how such effect can be minimized by a careful communication about the randomness of the selection process. Another concern is that a general introduction of a policy measure might trigger general equilibrium effects that cannot be captured by RCTs. The likelihood of such effects is small when the policy measure concerns only a small fraction of the population and/or has a relatively minor impact on the recipients. Most of the RCTs fulfil one or the other of the two conditions.

An open question for many RCTs is the generalization of findings found in one country to other countries, contexts and cultures. While some of the studies discussed in this section are cross-country, most of the RCTs concentrate on a single country and context. Replication in other countries would be desirable, but is often not possible due to financial constraints. Another open question concerns the (absence of) corruption. When a treatment is tested, the researchers make sure that the rules are followed and that no corruption occurs. But when the policy is rolled out to the general public, corruption, for example, of the civil servants involved, might actually reduce the effectiveness of the tested programmes relative to the results found in the RCTs. The next section will return to some of these issues.

13.6 The Changing Face of Public and Private-Sector Collaborations in Economic Research

By Estelle Cantillon, Liran Einav, Asim I. Khwaja and Markus Mobius

An emerging trend in economic research is the development of new forms of collaboration between researchers and private- and public-sector organizations. One form that such collaborations have taken is closer relationships with private firms for access to their proprietary data. A complementary form has been collaborations between researchers and policy-makers, where the focus is not only on data, but also on helping design and recalibrate policy interventions. In both cases, these collaborations are providing researchers with unmatched data access and data quality, as well as opportunities to investigate novel research questions or existing open questions in new ways.

Figure 13.2, taken from Einav and Levin (2014), illustrates some of these trends. The figure shows the percentage of papers published in the American Economic Review (AER) that obtained an exemption from the journal's data availability policy, as a share of all papers published by the AER that relied on some form of data (excluding simulations and lab experiments). Almost 50 per cent of empirical papers published in the past two years benefited from an exemption, a big jump from the situation less than 10 years earlier. Of course, proprietary data include other types of data than those obtained through close collaborations with public and private organizations, but it does provide a sense of the speed at which these changes are taking place.[18]

This section describes several examples of fruitful public and private research collaborations, and uses them to discuss their potential and their implications for how we organize, evaluate and fund research in economics.

13.6.1 New Private-Sector Collaborations in Economic Research

More and more data are generated by private firms. Firms now routinely collect data on their interactions with customers, on the activities of their employees, or on their suppliers. These data are often stored in electronic format reducing the cost of their retrieval and handling. For some of these firms (the 'data-driven' firms), the management and exploitation of these data are at the core of their business. Examples include Google, Microsoft, or Yahoo!, to name some of the most famous. For others, data represent a way to tailor their marketing efforts, optimize their pricing, or simply improve their organization. For yet some others, these sources of data are left unexploited for lack of expertise or awareness of their potential.

The richness of these data is a goldmine for researchers. Data that cover the internal workings of firms offer a chance to understand their operations like nothing before. For example, Tucker (2008) examined the roll out and the determinants of adoption of a new video messaging technology in an investment bank, using data on their employees and the 2.4 million calls that they made over a three year period. The data contain information about the hierarchical relationships between employees as well as their position within the informal communication network. This allowed the researcher to measure 'whose adoption matters' for the decision of an employee to adopt.

Because their coverage is different from other data sources, private proprietary data also allow researchers to look at new issues. For example, Adams et al. (2009) used transaction-level data from a US auto sales and financing company to document consumer behaviour in the subprime market. This unique dataset, covering 500,000 applicants and 150,000 loans originated between 2001 and 2004, as well as subsequent payments, default, and recovery outcomes, provides a great lens into the behaviour of the poor and unstable (e.g., undocumented workers), who are often under-represented in survey or

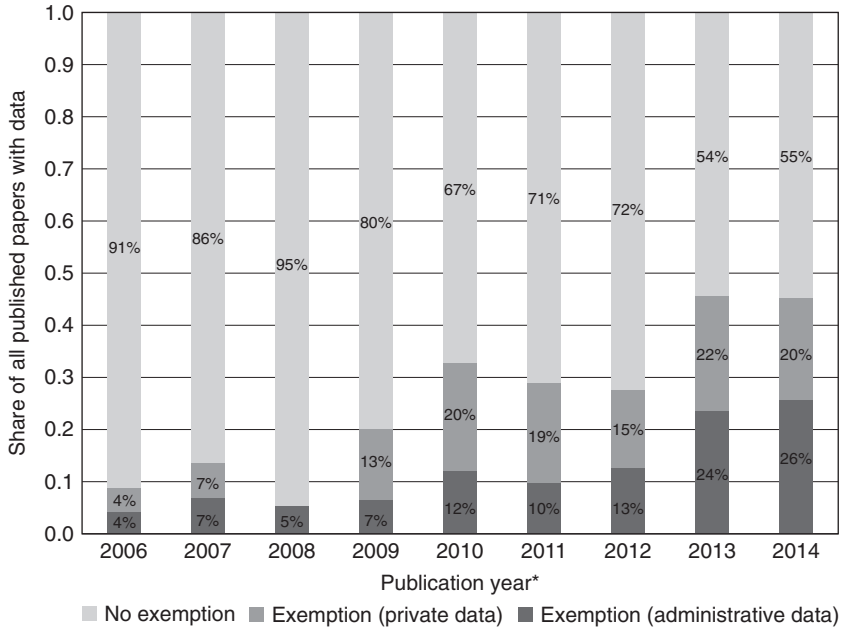

Figure 13.2 The rising use of nonpublicly available data in economic research (Einav and Levin, 2014).

administrative data. The data allowed the researchers to document the level of adverse selection and moral hazard in this understudied market that happened to be important in the outset of the subprime crisis in 2007.

A third advantage of these data is that their higher granularity can be useful to identify natural experiments arising from micro-level variations, thereby providing stronger identification. For example, Einav et al. (2014) used detailed browsing and purchase data on the population of eBay customers in the US to study the effect of sales taxes on internet commerce. In the US, online retailers must collect sales taxes only on purchases by residents of the same state. No taxes are collected for interstate purchases. The authors identified thousands of items and millions of browsing sessions in which individuals clicked on an item only to find out whether they were subject to the sales tax (if they lived in the same state as the seller). The difference in behaviour between the two groups of buyers then allowed the researchers to estimate the tax elasticity in a way that would be hard to measure from aggregate data.

These data have their own issues, however. They are collected for the needs of the business, not for research. As a result, the data may not cover all the

variables one would ideally want. Their format may not be optimized for research use, and in fact can rapidly change over time as internal IT systems are upgraded. The data may not be kept for a long time, reducing the potential time-span that one can study. Documentation is occasionally poor. Last but not least, they can be highly sensitive for the firm.

This means that access to these data tends to be more effort intensive, and involve a longer and riskier process than other sources of data.[19] Relationships are crucial here. Successful projects have often had a cheer-leader inside the firm. Relationships are crucial at the beginning of the project because a minimum level of trust needs to be established between the researchers and the firm giving them access to its data. They are crucial during the project because the firm will often need to continue to devote time and internal resources to provide access to the data, explain the data, and/or extract additional data, and because preliminary results may force the researchers and the firm to reorient the research question. Finally, they are crucial at the end of the project because the firm will often request to first clear the paper before it is circulated publicly.

Data-driven firms are special in the context of private-sector collaborations with researchers. Data form the core of their business and they already make heavy use of data scientists to put them to good use. However, they are often interested in research that can help better harness the power of their data and inform the design of their products. To foster research on questions of relevance for them, many of these firms – including Microsoft, Yahoo! and eBay – have experimented over time with different forms of engagement with researchers. This has gone from sponsoring papers and conferences on selected themes to establishing their own research labs that recruit economists on the academic job market and have a visitor programme to encourage academic economists to work on issues of relevance to them. Athey and Mobius (2012), Celis et al. (2014), and Blake et al. (2015) are examples of empirical research papers that have grown out of these collaborations.

An interesting aspect of these data-intensive firms is their ability to experiment with different product designs or prices easily and at a low cost. In fact, many online platforms already routinely carry out experiments on a small share of their operations and then implement the successful ones. It is a small step to the development of carefully crafted field experiments for research purposes (in the spirit of RCTs discussed in Section 13.5). Blake et al. (2015) designed such an experiment to measure the returns on paid keyword search in search engines and found them to be much lower than previously understood.

13.6.2 New Public-Sector Collaborations in Economic Research

New forms of collaborations are also developing between academics and policy-makers. The drivers there are different: this is not just about greater data

availability – though that is often part of this as well – but more about a genuine interest, on both sides, to go beyond the traditional interactions that characterize consulting relationships or policy evaluations towards a deeper collaboration where researchers can directly contribute to policy design.

This form of collaboration has been particularly salient in developing and emerging economies which, often lacking internal research capacities, have been relatively more willing to engage researchers in this manner. The Evidence for Policy Design (EPoD) research group at the Harvard Kennedy School has been at the forefront of these developments. The group is involved in long-term policy research engagements with government agencies in several developing countries, including Pakistan (education, taxation, and reform of the civil service), India (air pollution), and Indonesia (welfare programmes). Their early experience served to develop a systematic '*Smart Policy Design*' approach to such policy research engagements. In this approach, researchers and policymakers start by identifying questions of common interest, then jointly diagnose the underlying factors that are at play. Next, they use theory to collaboratively design potential mechanisms that could address these factors, followed by empirically testing the assumptions and implications of these mechanisms as the policies are rolled out, and finally use these feedback loops to carry out regular policy recalibrations.

The research upside of this approach is significant. The joint definition of the research question ensures its policy relevance and thus the potential impact of the research. The ability to carry out large-scale policy experiments addresses existing concerns about the scalability and external validity of 'standard' randomized control trials (see Section 13.5). The continuity of the relationship allows researchers to address the dynamic aspects of policy-making, including the need to recalibrate based on new information, which are often absent in 'standard' research.

Like for private-sector collaborations, relationships are crucial. Trust is built over a long period of time, through engagement at multiple levels of governments and agencies, where researchers have to demonstrate the value they can bring. In the case of the Smart Policy Design programme, early institutionalization between faculty members at Harvard and the relevant administrations and agencies, as well as the development of executive education programmes targeted at senior civil servants in these countries contribute to further trust.

One example of the potential of such work is an ongoing multi-year (now in its sixth year) collaboration between researchers at Harvard, LSE and MIT and the Excise & Taxation department in Punjab (Pakistan). The researchers have been involved in three distinct projects to date. The first was the design and testing of different pay for performance schemes for tax collectors (Khan et al., 2014). The second is using merit-based transfers and posting as a way of rewarding civil servant performance. The third project involves credibly linking

property tax payments to better provision of local public goods and services in order to rebuild tax morale and the citizen's social compact with the state. Each of these projects arose from specific questions that were of interest to both policy-makers and researchers, introduced policy mechanisms that were informed by theory, and involved large scale RCTs in order to produce plausible causal estimates of the impact of these mechanisms. Each project started with a pilot phase where initial design was recalibrated using small-scale experiments. Access to proprietary data (tax collector level collection rates etc.) was also critical. The projects were natural follow-ups to one another. Such an intense, theory and data driven process would not have been feasible without having built the trust and mutual interest needed for a long-term relationship.

Duflo et al. (2014) provide another example. Their research builds on a long-term relationship between the researchers and the Gujarat Pollution Control Board (GPCB) in the Indian state of Gujarat. Their paper combines an RCT approach designed to change environmental auditors' choice and frequency of the firms to audit, with 5 years of administrative records of correspondence between regulated firms and GPCB. The implementation of the RCT changed GPCB's financial resources dedicated to environmental audits as well as some staffing and managerial processes, in a way that would not have been possible without the full backing and active involvement of GPCB. The data allowed the researchers to estimate the cost of regulation in this setting and quantify the benefits of discretion in the selection of the firms to audit.

Other applications, this time to the design of welfare programmes, include the work of Alatas and coauthors with the Indonesian government on the re-design of their social welfare programmes (Alatas et al., 2012a, Alatas et al., 2012b, Alatas et al., 2016) or Muralidharan's collaboration with the state of Andhra Pradesh (India) on the use of biometric identification cards for enhancing the effectiveness of welfare payments (Muralidharan et al., 2015).

13.6.3 Risks, Challenges and Outlook

We have so far mostly outlined the research benefits of these new forms of collaboration between researchers and private firms or governments. They are not without risks, however.

A significant risk for researchers is that the collaboration breaks down, after much effort has been expended, but before the research outputs are cleared and the results published. A common reason for such break-down is staff turnover. A manager or department head, who was initially supporting the project, changes jobs or moves firms/departments, leaving the project without an inside cheer-leader. Another reason is that the results are too sensitive. Adams et al.'s 2009 work on the subprime market described above is an example. The paper was first circulated during the outset of the subprime crisis and

drew much attention in the trade journals of subprime lending. Given the regulatory risks associated with the results, the firm that gave them the data froze the dissemination of follow-up work for two years, after which the authors managed to negotiate a 'termination' deal and got their results published (Einav et al., 2012 and Einav et al., 2013).

Maintaining trust and interest, over the long run, is therefore essential. Identifying research questions that are of interest to both parties is useful. A suitable definition of the research question may also prevent tensions ex-post when the results turn out to be unacceptable – for regulatory, public relations, or competitiveness reasons. Designing the relationship for multiple research outputs is another (complementary) solution. Initial research outputs build appetite for the following ones, reduce the risks borne by researchers who can already publish some results, and alleviate the tension between the short-term horizons of policy-makers and firms, and the longer-run horizons of researchers. More generally, successful collaborations require a change of mindset from researchers, who need to be more problem-driven (identifying a question of relevance to the other party) than solution-driven (identifying a dataset that best suits the effect the researchers are interested in).

An important hurdle for many researchers is nondisclosure agreements (NDAs) which not only involve them and the data owners, but also their institutions. Given the sensitivity of their data, data owners often require stringent conditions to guarantee confidentiality and high penalties in case of breach, which legal offices of research institutions are reluctant to accept.[20] Over time, best practices that are acceptable to all will emerge based on the more successful experiences.

An obvious risk arising from such a necessarily close relationship between the researchers and the data owners is the risk of conflict of interest and the loss of scientific integrity that comes with it. The nondisclosure of the data only exacerbates this risk (though going through a rigorous academic review process does help). Financial independence is useful here. It frames the relationship as one between equals, rather than consulting, and provides credible walk-off threat points.[21] Several leading journals in economics are now requesting authors declare any financial interests, including funding, related to their research before it is published.

In the end, however, the risk for scientific integrity is probably not much higher than for other empirical work. Biased reporting (due to conflict of interest) may be a bigger issue but, on the other hand, the involvement of parties with a direct interest in the research actually provides greater discipline against other types of scientific fraud, such as cooking the results. Indeed, the firm or department will not continue the research or policy reform without being convinced about its usefulness. The difficulty of replication may not be such a big issue either. Even when data are available, pure replications are rare in economics because the culture of replication does not exist (Sebransk et al., 2010).

If the results are sufficiently important, similar databases can be used as a substitute for replication. Moreover, to the extent that the implications coming out of the work get adopted by the private or policy partner, this may allow for an even higher level of continuous testing and field 'replication'.

A final challenge is the complexity of handling and analysing data produced by some of the data-intensive firms for which the standard statistical software packages are inadequate. Most economists do not have the programming skills and technical training (aggregation algorithms, machine learning, etc.) to take advantage of some of these data. In fact, some of the projects reported here have data scientists as coauthors for exactly these reasons. Things are changing. Some graduate programmes in economics now offer courses on statistical learning and other methods for 'Big Data'. In the UK, the Economic and Social Research Council (ESRC) is sponsoring the development of doctoral courses to handle new forms of data such as internet data, satellite and aerial imagery, and geolocation and other tracking data.

Scientific disciplines always adapt to scientific opportunities and practices. New forms of research collaborations will be no different. Journals will continue to adapt their practices to the needs of the profession as they have done in the past for rules on collaborative work, data sharing and financial interest disclosure. Likewise, larger and possibly multi-disciplinary teams of researchers are likely to eventually impact how we organize and fund research in economics, and how we evaluate individual contributions for co-authored research.[22]

Despite these risks and challenges, we are optimistic regarding the research potential of these new forms of research collaboration. There is a lot of terrific information lying in private hands or simply outside of official data and much of it does not have good publicly available substitutes. The possibility of jointly designing and recalibrating policies can catalyse really exciting and novel work. We should simply be aware of the constraints and optimize around them.

13.7 Concluding Comments

An easy prediction to make is that economics will be more data-intensive in the future, and that both existing and new sources of data will continue to contribute to significant research breakthroughs. This chapter covered many different types of data and argued that each has its unique benefits. It is important for economic research to acknowledge the benefits of this diversity and the potential complementarity among data producers.

Each data type comes with its own constraints and challenges. They were described in detail in the corresponding section. If one general message stands out it is that all stakeholders have a role to play in improving the production, quality and accessibility of data for economic research. **Researchers** are, of course, at the heart of this. They can build trust and support for greater data

access by showcasing the value of their research based on these data. They can contribute to data innovation when they act as academic entrepreneurs in large-scale data collection efforts. They can develop new methods to generate and leverage data that can increase our understanding of human behaviour and the economy. **Funders** need to design flexible funding instruments that meet the needs of the diversity and specificity of data in economics. They can also play a role in federating researchers' interests for training and promoting data access. **Statistical agencies and central banks** are essential to ensure the quality of data, define standards, and develop metadata to promote data harmonization and linking across countries. **Data firms** will continue to provide value by harmonizing and linking firm data that lie outside of the scope of official data. **Research institutions** need to establish protocols to guarantee the integrity of the data entrusted to their researchers and build the required ethical and legal expertise to support their researchers' ventures into new data sources. **Journals** are important to maintain the highest standards of scientific integrity. The leading journals in the profession have in the past accompanied changes in the way research is organized and produced. They should continue to do so. Last but not least, **governments and policy-makers** are essential because they provide the political impetus that makes changes possible. Their leadership will be determining for the likely developments in key areas for research such as access to microdata, cross-country data harmonization and linking, and research funding for data infrastructure. At the European level, this means:

1. ensuring that the current revision of the Data Protection Directive does not reduce access to personal data for researchers,
2. promoting the introduction of legal provisions in European and national legislations to secure legitimate access to data for researchers, as in done in several Nordic countries,
3. promoting the introduction of mandates for statistical offices, including Eurostat, to service researchers,
4. clarifying the legal framework for the access of confidential data across borders,
5. mandating data harmonization and linking of existing business data across Member States, and developing access to such data for researchers,
6. reforming the current funding mechanism for data infrastructure to meet the needs of data infrastructure in the social sciences, including securing stable European-level funding for cross-national data collection efforts.

Notes

1. This explains why this chapter has an unusual number of coauthors. It is based on the presentations made and discussions that took place at the COEURE workshop on 'developments in data and methods for economic research' in July 2015. Authors of individual sections are indicated under the title of the section. Reference to this

chapter can be made to the chapter as a whole by citing all the authors or to an individual section by citing only the authors and the title of that section. This chapter represents the views of the authors and not of their institutions.

2. The main legal texts at the EU level are regulation 45/2001 on data protection, and regulations 223/2009 and 557/2013 on the access to confidential data for scientific purposes, as well as their translation into national laws and regulations.
3. At the time of writing this text, the Data Protection Directive of 1995 is being revised with ongoing discussions between the Council, the European Commission and the European Parliament. Some of the proposed changes risk restricting access to personal data without consent. See, for example, the position statement issued by the Wellcome Trust with the backing of hundreds of European research agencies and academic associations: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp059364.pdf.
4. A less user-friendly version of virtual access, sometimes referred to as remote execution, requires researchers to send their codes without seeing the data; the codes are applied to the data and the output is checked for risk of confidentiality breach before it is sent to the researchers.
5. Castellani and Koch (2015) also identify barriers in terms of the ability to link different datasets. This is not reported here.
6. Annex II of the NORIA-net (2014) report describes the legal and organizational conditions under which access to microdata including biobanks and register data is organized in Nordic countries.
7. As a point of comparison, Eurostat services about 300–400 survey-based projects per year.
8. See https://educationendowmentfoundation.org.uk/evaluation/.
9. This echoes a similar concern in the US for facilitating access to administrative data. See, for example, Card et al. (2010).
10. The ADRN is one component of the Big Data Network initiative (http://www.esrc.ac.uk/research/our-research/big-data-network/). A second component is the Business and Local Government Data Research Centre that seeks to make data collected by business and local government available to researchers. A third component focuses on the third sector and social media data.
11. Because of its motivation, the focus of this section is on business data.
12. 'The Financial Crisis and Information Gaps', October 29, 2009, available at: https://www.imf.org/external/np/g20/pdf/102909.pdf.
13. See http://www.financialstabilityboard.org/what-we-do/policy-development/additional-policy- areas/addressing-data-gaps/ for details.
14. An online Handbook on Methodology of Modern Business Statistics available at http://www.cros-portal.eu/content/handbook-methodology-modern-business-statistics was developed and serves as a template for future exercises. Funded projects are described at http://ec.europa.eu/eurostat/statistics-explained/index.php/MEETS_programme_-_towards_more_efficient_enterprise_and_trade_statistics.
15. See Charles Noussair's website: www.slideshare.net/charlies1000/laboratory-experiments.
16. See http://www.experimentation.jeunes.gouv.fr/.
17. See http://www.povertyactionlab.org/policy-lessons/education/student-participation.

18. Proprietary data also include data that the researcher had to purchase from a commercial data vendor and administrative data.
19. There are also occasional data dumps where firms give out data to researchers because they know one of the researchers personally or through contacts (Cohen and Einav, 2007 is one such example). Access in this case is fast and easy but the one-shot nature of the relationship prevents any follow-up or additional data extraction based on initial results and is not free of legal risk about data disclosure.
20. One of us was involved in a NDA that required her research institution to accept responsibility for any accident or death caused by her presence on the firm's premises.
21. Interestingly several funding agencies are moving towards demanding that the data collected as part of funded projects be publicly available, in sharp contrast with the funding needs for the type of research described in this section.
22. In an analysis of articles published in three leading economics journals (*American Economic Review, Journal of Political Economy, Quarterly Journal of Economics*), Hamermesh (2013) noted a steady increase in the number of coauthors over time with the first four-author paper published in 1993 and the first five-author and six-author papers published in 2011. Currently, there is very little penalty for publishing coauthored papers in economics.

References

Adams, W., Einav, L., and Levin, J. 2009. Liquidity Constraints and imperfect Information in Subprime Lending. *American Economic Review*, **99**, 49–84.

Akerlof, G. 1982. Labor Contracts and a Partial Gift Exchange. *Quarterly Journal of Economics*, **97**, 543–569.

Alatas, V., Banerjee, A., Chandraskhar, A., Hanna, R., and Olken, B. 2012a. Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia. NBER Working Paper 18351.

Alatas, V., Banerjee, A., Hanna, R., Olken, B., and Tobias, J. 2012b. Targeting the Poor: Evidence from a Field Experiment in Indonesia. *American Economic Review*, **102**, 1206–1240.

Alatas, V., Banerjee, A., Hanna, R., Olken, B., Purnamasari, R., and Wai-Poi, M. 2014. Self-targeting: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy*, **124**, 371–427.

Aldashev, G., Kirchsteiger, G., and Sebald, A. 2015. Assignment Procedure Biases in Randomized Policy Experiments. *forthcoming, The Economic Journal*.

Ashraf, N., Berry, J., and Shapiro, J. 2010. Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia. *American Economic Review*, **100**, 2383–2413.

Athey, S., and Mobius, M. 2012. The Impact of News Aggregators on Internet News Consumption: The Case of Localization. Mimeo available at http://markusmobius.org/sites/default/files/localnews.pdf.

Banerjee, A., Banerji, R., Duflo, E., Glennerster, R., and Khemani, S. 2010. Pitfalls of Participatory Programmes: Evidence from a Randomized Evaluation in Education in India. *American Economic Journal: Economic Policy*, **2**, 1–30.

Bauchet, J., Marshall, C., Starita, L., Thomas, J., and Yalouris, A. 2011. Latest Findings from Randomized Evaluations of Microfinance, Access to Finance Forum, Report by CGAP and its Partners. Available at: http://povertyactionlab.org/publication/latest-findings-randomized-evaluations-microfinance.

Behaghel, L., Crépon, B., and Le Barbanchon, T. 2014. Unintended Effects of Anonymous Resumes. IZA Discussion Paper 8517.

Best, M., and Kleven, H. 2015. Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the UK. Mimeo London School of Economics.

Blake, T., Nosko, C., and Tadelis, S. 2015. Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment. *Econometrica*, **83**, 155–174.

Bloom, H., Orr, L., Bell, S., Cave, G., Doolittle, F., Lin, W., and Bos, J. 1997. The Benefits and Costs of JTPA Title II-A Programmes: Key Findings from the National Job Training Partnership Act Study. *The Journal of Human Resources*, **32**, 549–576.

Camerer, C. F. 2013. Goals, Methods and Progress in Neuroeconomics. *Annual Review of Economics*, **5**, 425–455.

Camerer, C. F. 2015. The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Reply to Levitt and List. In: Frechette, G., and Schotter, A. (eds), *The Methods of Modern Experimental Economics*. Oxford Univ. Press, 2015.

Capen, E. C., Clapp, R. V., and Campbell, W. M. 1971. Competitive Bidding in High-Risk Situations. *Journal of Petroleum Technology*, **23**, 641–53.

Card, D., Chetty, R., Feldstein, M., and Saez, E. 2010. Expanding Access to Administrative Data for Research in the United States. In: Schultze, C. L., and Newlon, D. (eds), *Ten years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas*. National Science Foundation.

Castellani, D., and Koch, A. 2015. Mapping Competitiveness with European Data. *BRUEGEL blueprint 23*.

Celis, E., Lewis, G., Mobius, M., and Nazerzadeh, H. 2014. Buy-it-Now or Take-a-Chance: Price Discrimination through Randomized Auctions. *Management Science*, **60**, 2927–2948.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., and Vignoles, A. 2013. Widening Participation in Higher Education: Analysis using Linked Administrative Data. *Journal of the Royal Statistical Society Series A, Royal Statistical Society*, **176**, 431–457.

Cohen, A., and Einav, L. 2007. Estimating Risk Preferences from Deductible Choice. *American Economic Review*, **97**, 745–788.

Cohen, J., and Dupas, P. 2010. Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment. *Quarterly Journal of Economics*, **125**, 1–45.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. 2013. Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment. *The Quarterly Journal of Economics*, **128**, 531–580.

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. 2014. The Value of Regulatory Discretion: Estimates from Environmental Inspections in India. NBER working paper 20590.

Dufwenberg, M., and Kirchsteiger, G. 2004. A theory of Sequential Reciprocity. *Games and Economic Behavior*, **47**, 268–298.

Dustmann, C., Machin, S., and Schoenberg, U. 2011. Ethnicity and Educational Achievement in Compulsory Schooling. *The Economic Journal*, **120**, F272–F297.
Einav, L., and Levin, J. 2014. Economics in the Age of Big Data. *Science*, **346**, 715–721.
Einav, L., Jenkins, M., and Levin, J. 2012. Contract Pricing in Consumer Credit Markets. *Econometrica*, **80**, 1387–1432.
Einav, L., Jenkins, M., and Levin, J. 2013. The Impact of Credit Scoring on Consumer Lending. *RAND Journal of Economics*, **44**, 249–274.
Einav, L., Knoepfle, D., Levin, J., and Sundaresan, N. 2014. Sales Taxes and Internet Commerce. *American Economic Review*, **104**, 1–26.
Falk, A. 2007. Gift Exchange in the Field. *Econometrica*, **75**, 1501–1511.
Fehr, E., Kirchsteiger, G., and Riedl, A. 1993. Does Fairness Prevent Market Clearing. *Quarterly Journal of Economics*, **108**, 437–459.
Fehr, E., Kirchsteiger, G., and Riedl, A. 1998. Gift Exchange and Reciprocity in Competitive Experimental Markets. *European Economic Review*, **42**, 1–34.
Finkelstein, A., Hendren, N., and Luttmer, E. 2015. The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment. Mimeo MIT.
Hamermesh, D. 2013. Six Decades of Top Economics Publishing: Who and How? *Journal of Economic Literature*, **51**, 162–172.
Haushofer, J. et al. 2014. A Methodology for Laboratory Experiments in Developing Countries: Examples from the Busara Center. Mimeo available at https://princeton.edu/~joha/publications/Haushofer_Busara_2014.pdf.
Heller, S. 2014. Summer Jobs Reduce Violence among Disadvantaged Youth. *Science*, **346**, 1219–1223.
Henrich, J., Boyd, R., Bowles, S., Camerer, C.F., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Hill, K., Gil-White, F., Gurven, M., Marlowe, F., Patton, J.Q., Smith, N., and Tracer, D. 2005. Economic Man in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies. *Behavioral and Brain Sciences*, **28**, 795–855.
Henrich, J., Heine, S. J., and Norenzayan, A. 2010. The WEIRDest People in the World? *Behavioral and Brain Sciences*, **33**, 61–83.
Herrmann, B., Thöni, C., and Gächter, S. 2008. Antisocial Punishment across Societies. *Science*, **319**, 1362–1367.
Kagel, J. H., and Levin, D. 2002. *Common Value Auctions and the Winner's Curse*. Princeton: Princeton University Press.
Khan, A., Khwaja, A. I., and Olken, B. 2014. Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors. *Quarterly Journal of Economics*, **131**, 219–271.
Kremer, M., and Miguel, E. 2007. The Illusion of Sustainability. *The Quarterly Journal of Economics*, **122**, 1007–1065.
Meyer, B. D. 1995. Lessons from the U.S. Unemployment Insurance Experiments. *Journal of Economic Literature*, **33**, 91–131.
Muralidharan, K., Niehaus, P., and Sukhtankar, S. 2015. Building State Capacity: Evidence from Biometric Smartcards in India. Mimeo University of California at San Diego.
NORIA-net. 2014. Joint Nordic Registers and Biobanks, A Goldmine for Health and Welfare Research. NordForsk Policy Papers 5.
Plott, C., and Smith, V. 2008. *Handbook of Experimental Economic Results*. Elsevier.

Sebransk, N., Young, L., Kelner, K., Moffit, R., Thakar, A., Raddick, J., Ungvarsky, E., Carlson, R., Apweiler, R., Cox, L., Nolan, D., Soper, K., and Spiegelman, C. 2010. Make Research Data Public? – Not always so Simple: A Dialogue for Statisticians and Science Editors. *Statistical Science*, **25**, 41–50.

Tucker, C. 2008. Identifying Formal and Informal Influence in Technology Adoption with Network Externalities. *Management Science*, **54**, 2024–2038.

Wilson, D., Burgess, S., and Briggs, A. 2011. The Dynamics of School Attainment of England's Ethnic Minorities. *Journal of Population Economics*, **24**, 681–700.