



Europe Should Promote Data for Social Good

By Alexander Kostura & Daniel Castro | October 3, 2016

Data-driven innovations have the power to address some of the most pressing social challenges in Europe. While many government and non-governmental organizations (NGOs) are using data in their attempts to tackle a range of social issues from high unemployment to the refugee crisis, more can be done. To accelerate progress, public and private-sector leaders should take steps to collect data on disadvantaged populations, facilitate cross-sector collaboration on data projects for social good, and implement policies that encourage data use, reuse, and sharing in support of social goals.

Progress on using data for social good has been uneven across the EU due to resource constraints, digital inequality, and restrictive data regulations. A renewed European commitment is needed to address these challenges.

INTRODUCTION

Changing demographics in Europe are creating enormous challenges for the European Union (EU) and its member states. The population is getting older, putting strain on the healthcare and welfare systems. Many young people are struggling to find work as economies recover from the 2008 financial crisis. Europe is facing a swell in immigration, increasingly from war-torn Syria, and governments are finding it difficult to integrate refugees and other migrants into society.¹ These pressures have already propelled permanent changes to the EU. This summer, a slim majority of British voters chose to leave the Union, and many of those in favor of Brexit cited immigration as a motive for their vote.²

Europe needs to find solutions to these challenges. Fortunately, advances in data-driven innovation that have helped businesses boost performance can also create significant social benefits. They can support EU policy priorities for social protection and inclusion by better informing policy and program design, improving service delivery, and spurring social innovations. While some governments, nonprofit organizations, universities, and companies are using data-driven insights and technologies to support disadvantaged populations, including

unemployed workers, young people, older adults, and migrants, progress has been uneven across the EU due to resource constraints, digital inequality, and restrictive data regulations. A renewed European commitment to using data for social good is needed to address these challenges.

This report examines how the EU, member-states, and the private sector are using data to support social inclusion and protection. Examples include programs for employment and labor-market inclusion, youth employment and education, care for older adults, and social services for migrants and refugees. It also identifies the barriers that prevent European countries from fully capitalizing on opportunities to use data for social good. Finally, it proposes a number of actions policymakers in the EU should take to enable the public and private sectors to more effectively tackle the social challenges of a changing Europe through data-driven innovation. Policymakers should:

- Support the collection and use of relevant, timely data on the populations they seek to better serve;
- Participate in and fund cross-sector collaboration with data experts to make better use of data collected by governments and non-profit organizations working on social issues;
- Focus government research funding on data analysis of social inequalities and require grant applicants to submit plans for data use and sharing;
- Establish appropriate consent and sharing exemptions in data protection regulations for social science research; and
- Revise EU regulations to accommodate social-service organizations and their institutional partners in exploring innovative uses of data.

THE EMERGING OPPORTUNITY FOR DATA INNOVATION

With advances in digital technology, humans and machines are now constantly producing rich streams of data about the systems and processes that make up daily life, and are better equipped than ever to capture, analyze, and act upon this data. In particular, the proliferation of mobile devices and the Internet of Things (a term that refers to the array of network-connected devices with sensors or actuators) allow nearly every aspect of the physical world to generate digital data. This includes data about human health, traffic patterns, environmental conditions, machinery, and supply chains.

The city of Barcelona, for instance, has been a global leader in the smart-cities movement, using the Internet of Things to improve public services with such tools as street sensors that guide drivers to available parking spaces and smart meters that continuously monitor energy use for more accurate billing and better understanding of consumption patterns.³ These data sets, when combined with other sources, including social media data, transactional records, and traditional surveys, can provide even more valuable and unique insights into almost every part of the economy and society.

Technological advances in the collection, storage, analysis, use, and dissemination of data have enabled individuals and organizations from the public and private sector to tap into these datasets to gain new insights into the world's complexities and use those insights to make better decisions and automate more processes. Data analytics, in particular, is a powerful tool that can be applied to many different problems. Data scientists can analyze increasingly large and diverse data sets, including those generated in real time, to uncover new insights. In the past, limitations on data collection often allowed analysts to consider only a sample of the potential available data. Now, the technology exists to allow analysis of entire populations, thereby leading to richer insights and more personalization. Similarly, whereas in the past organizations could not easily share data because of technological limitations, the emergence of cloud computing has made it easier than ever to disseminate large datasets. As a result, many organizations support “open data”—data that is made freely available without restrictions and is released in non-proprietary, machine-readable formats.⁴ The European Data Portal pools and categorizes more than 600,000 open data sets published online by member-state governments.⁵ By allowing civic hackers and entrepreneurs to reuse public-sector data, open data has fueled the creation of new businesses and improved public services. The social impact of open data is difficult to measure, but the McKinsey Global Institute estimates the potential global economic value at \$3 trillion annually.⁶ Many of the opportunities in data for social good detailed in this report are related to the application of advanced data analytics to a unique combination of open data sets.

OPPORTUNITIES TO USE DATA INNOVATION FOR SOCIAL GOOD

Some European governments and nonprofit organizations are innovating with data to improve research, policies, and services for social protection and inclusion. These innovations include information tools for disadvantaged individuals to access social services, analysis of administrative data to help non-profit organizations and government agencies improve services, research that combines and analyzes data

sets for new insights into social challenges, and devices that collect more data on the human experience and augment welfare services.

The following sections highlight these opportunities as they apply to four areas:

1. Employment and labor-market inclusion
2. Improving prospects for Europe's youth
3. Empowering older adults
4. Integrating migrants and refugees

EMPLOYMENT AND LABOR MARKET INCLUSION

Since the 2008 financial crisis, European economies have struggled with slow growth and high unemployment. As jobs gradually return, policymakers are working to ensure that all people who are eligible to work have access to a paid job and can participate in the economy. Many disadvantaged populations, including the young, untrained workers, people with disabilities, migrants who do not speak the national language, and some native minorities face significant labor-market challenges.

There are at least three ways that data-driven innovation can help address these challenges: by creating smarter information tools for job seekers; improving the efficiency of unemployment assistance from governments and nonprofits; and helping employers implement inclusive hiring practices.

Labor-market data can serve as a powerful platform for tools and services that benefit job seekers. For example, in the UK an online tool called “Where the Work Is” analyzes text from online job postings and open government data on the education, training, and number of people eligible for employment to illustrate where the skills of job seekers may not be matching up with those skills required by employers. The tool presents location-specific data on which occupations have the most openings, what kind of education level is required, and average salaries, so that users may target their job search or plan their education.⁷ With location-specific information on what skills are in demand, job seekers can more effectively identify relevant opportunities and employers can strategically locate job postings.

Similarly, a non-profit organization of data scientists called Bayes Impact is working with the French government to create a digital job-search application that takes advantage of under-utilized and rarely-linked government data sets.⁸ The application combines large amounts of data from across the national government that is pertinent to the job-search process including real-time information on the labor market, social services, training programs, and millions of past examples of career trajectories. Through the use of algorithms, this application serves as a digital career-counseling tool that compares a user's profile

to the data and recommends search strategies and proven career pathways.

The EU is funding the development of a similar tool to help job seekers navigate the labor market.⁹ These innovations boost the efficiency of government unemployment services by supplementing in-person counseling and bringing key information to more job seekers. By combining a large volume of data in one place with tailored recommendations, these tools also help job seekers save time and money in their searches.

Employers can also use socioeconomic data to support more inclusive hiring practices. The British company PiC has developed a technology service that helps hiring managers to assess candidates' qualifications and performance while taking into account their socioeconomic backgrounds. This tool compiles open data on candidates' educational and social background, information that includes the performance of their schools nationally and the number of people from their geographic area who attend higher education, in order to create a "PiC score." By scoring candidates' performance relative to that of their peers and their backgrounds, PiC allows recruiters to more holistically assess applicants that come from disadvantaged circumstances.¹⁰

IMPROVING PROSPECTS FOR EUROPE'S YOUTH

Europe's youth still suffer from the lingering economic impact of the recession with more than 4.5 million young people between the ages of 15 and 24 currently unemployed.¹¹ Data-driven innovations can help address youth unemployment through targeted initiatives based on labor market and education data, improved performance for programs combatting education inequality, and more inclusive digital learning tools.

EU policymakers working on youth unemployment and education can analyze national labor-market data on a larger scale—combining employment statistics, educational attainment by geographic location, and real-time data on employment vacancies—to understand where there are specific skills or education gaps. With these insights, governments and educational organizations can design targeted initiatives like vocational training, apprenticeship programs, or skills-based curriculums to address these gaps among their young people. Private businesses may also use this data to make strategic decisions for recruitment and human capital, such as where to recruit for certain positions or where to invest philanthropic funds in youth skills development.

The European Commission and the European Center for Development of Vocational Training built a data portal combining historical data from all 28 EU member countries on labor-market sectors, occupations, and education. The portal includes a set of analytical tools that deliver

insights like highlighting skills gaps or the underutilization of skills in national workforces, identifying historical trends in educational achievement and worker qualifications, and predicting future job prospects.¹² These insights are useful to both public and private organizations involved in education as well as to individual job seekers.

The key to making a labor market more inclusive for young people is access to education and skills training. According to Eurostat, the statistical office of the EU, the unemployment level for Europeans who have not completed a tertiary education is 17.4 percent, which is significantly higher than the 5.6 percent unemployment level for those with a tertiary education.¹³ Disadvantaged populations such as lower income students do not have the same educational opportunities and thus are less likely to attend universities and be competitive in the labor market. Organizations working to boost educational attainment for disadvantaged children can improve their programming by analyzing the data they collect through their operations and applying insights to future interventions.

For example, the Access Project, a UK nonprofit, analyzes open data on socioeconomic indicators and educational achievement to assess the impact of its program that links up students from low-income areas with volunteer tutors from top universities.¹⁴ The Access Project also recently collaborated with a group of volunteer data scientists at DataKind UK to analyze data from their tutors' reports. The data scientists were able to tap underutilized, operational data to predict the likelihood that a university volunteer will complete the school year with an assigned student. Based on the analysis, the Access Project also made a decision to alter its volunteer reports to collect more information when students cancel tutoring appointments.¹⁵

Finally, advances in data-driven education can support a more inclusive education system through more accessible and effective digital-learning tools that also provide sources of data for education innovation. For example, online learning tools that track student progress can give students more flexibility in the content, pace, and style of their education; and teachers can facilitate learning for different preferences based on individual performance.¹⁶ Teachers can monitor performance data collected by these tools to continuously improve their instruction and meet the needs of individual students. These tools can collect and analyze data about all participating students, providing educators with insights into what is and is not working for students. The data helps administrators and policymakers make informed decisions about how to best allocate educational resources to reach more students.

To promote this kind of data-driven innovation in education, the EU funded the Learning Analytics Community Exchange (LACE) for educators and policymakers to share evidence of how mining educational data improves education outcomes. The LACE project

launched the Evidence Hub, an online platform that collates and categorizes results from learning analytics projects around the world to support evidence-based decision making in education.¹⁷

EMPOWERING OLDER ADULTS

Europe has been a leader in data-driven healthcare by funding research that promotes data sharing, interoperability, and nontraditional data sources.¹⁸ With the goal of supporting active aging, whereby older adults remain independent and contribute to society for as long possible, the EU is also pursuing better collection and use of data for improved health and well-being of this population.

By significantly increasing data collection on all aspects of life for older Europeans and applying advances in data analytics and artificial intelligence, Europe's medical researchers can build a better understanding of the aging experience to inform policies meant to assist the older generation. Opportunities include coordinating the collection and sharing of health data specifically related to aging, machine-learning tools for early disease detection so older adults can take active roles in their treatment plans, and applications of the Internet of Things to assisted living.

In order to specifically address issues related to active and healthy aging, the EU relies on the Survey of Health, Ageing, and Retirement in Europe (SHARE), a cross-national longitudinal survey about the health, socioeconomic status, and social networks of approximately 123,000 individuals aged 50 years or older from 20 European countries.

In 2016, the coordinators of SHARE announced an extension of the survey to the remaining eight countries of the EU, making it the largest European social-science panel study.¹⁹ Because SHARE collects unique data on the experience of aging and is made available to researchers free of charge, the data is cited in approximately two studies per week.²⁰ Updating this traditional survey data and making it available free of charge in machine-readable format has opened up new research opportunities to evaluate public policies for active ageing. For example, researchers in Germany linked SHARE data to government data on pension claims to evaluate whether retirement assistance is reaching those who need it.²¹

In addition to supporting research, the ability to access, combine, and analyze sources of health data, such as electronic health records, contributes to new decision-support systems for older patients and caregivers. For example, scientists at Vrije University Medical Center in the Netherlands trained machine-learning algorithms—a branch of artificial intelligence in which computer programs can automatically and iteratively interpret new data—to analyze magnetic resonance imaging (MRI) scans for signs of Alzheimer's disease. MRIs reveal subtle indicators about brain chemistry that can help diagnose patients earlier,

providing them the opportunity to have a role in designing their treatment plans before they are too sick to do so. Machine learning trains the software to evaluate the images immediately following the MRI scan and to provide a reliable recommendation to the doctor as to whether or not a patient's scans show early signs of the disease.²² To develop these important applications of artificial intelligence, researchers must have access to large amounts of patient data to test and improve the accuracy of innovative analytic techniques. Researchers at Vrije, for instance, developed their tool with a data set of 260 patients, but reported that data sets with thousands of patients would help improve the accuracy of the tool.²³

Another key opportunity in data innovation to improve quality of life for older adults is the Internet of Things for managing medical conditions, and generally assisting elderly individuals in their homes. New mobile and sensor technologies can provide 24-hour monitoring, and even automated care, for elderly patients living in a connected home. Wearable devices can alert emergency responders when an elderly person has fallen, and smart pill bottles can help individuals adhere to their medication regimens.

These technologies support opportunities for data innovation by making it easier than ever to collect, combine, and share granular health and lifestyle data, ultimately yielding insights that can create social and economic value to European countries. Older adults are able to stay in their homes longer, boosting quality of life and supporting better health outcomes. Through funding mechanisms like the Active and Assisted Living Joint Programme and the Innovation Partnership on Active and Healthy Aging, the European Commission has supported the development and scaling up of many data-driven health technologies, such as monitoring devices to detect falls, brain-recovery game systems, and wearable sensors to collect detailed data on patients with Parkinson's disease.²⁴

INTEGRATING MIGRANTS AND REFUGEES

Europe is in the midst of a refugee crisis, with more than 1.3 million people seeking asylum in the EU in 2015 from places like Afghanistan, Syria, and Yemen.²⁵ Lacking the financial and material resources to support their families and the necessary language skills to gain employment, immigrants, especially refugees escaping perilous conditions in their home countries, are particularly vulnerable. Europe's government agencies and charities that provide services to migrant populations can make their services more efficient and effective through analyzing and sharing operational data.

Charities distributing food, for example, can use operational data to improve their distribution networks and to deliver the necessary amount of fresh food to the populations that need it most. The French Red Cross

recently shared its food-aid distribution data with a group of volunteer data scientists at Bayes Impact. Analyzing the distribution of food items at 800 distribution sites and supermarkets, the data scientists developed an algorithm to predict demand at certain sites, a map showing where the Red Cross should build new food-aid centers to meet demand, and a monitoring platform for the amount and type of food at each site.²⁶

Additionally, Tesco, the UK's largest supermarket chain, and a network of charities are using an application called FoodCloud to match businesses and their surplus food with charities that can distribute it directly to populations that need it. Participating businesses input information about the food they wish to donate, and the app alerts charities in the area of the opportunity. The interested charity then collects the donation directly from the business. By crowdsourcing data on surplus food and dispersing it in real time to nearby organizations, the app helps to eliminate food waste.²⁷

Sharing data can also enable organizations to coordinate service delivery. RefAid is an information sharing app used by official aid workers to coordinate services and to inform refugees of services near their new homes.²⁸ RefAid seeks to address a lack of coordination within and among charities and other non-governmental organizations (NGOs) that provide assistance to refugees, such as legal services and food banks. Aid organizations, such as the Red Cross of Italy, input data on the type of aid they offer, the location, the availability, and other key details so that aid workers know what's available in a given location. Refugees can then use the app to understand what services are available.²⁹

Advances in mobile phone technology are especially useful in refugee assistance services, given the high mobile phone ownership rates among refugee populations. The Office of the United Nations High Commissioner for Refugees (UNHCR) estimates that 84 percent of the world's urban refugee population and 57 percent of the rural refugee population own mobile phones.³⁰

REFUNITE, a technology-focused nonprofit based in Copenhagen, has created a platform to help refugees trace and reunite with their family members after journeys and asylum processes that last several years and cross many borders. The platform gathers information from users who opt in and creates a searchable database that is available online and via a mobile phone. Because many refugees do not own Internet-capable phones, REFUNITE utilizes both short messaging service (SMS) and unstructured supplementary service data (USSD), a technology that allows mobile phone users to communicate with a program application via basic text messaging.³¹

BARRIERS TO DATA FOR SOCIAL GOOD

While the previous examples illustrate the potential to use data for social good, relatively few European governments and social-service agencies are taking advantage of these opportunities. First, nonprofits and government agencies typically lack a “culture of data” or organization-wide strategies and processes around the collection and analysis of data for a range of functions. Second, the European labor market suffers from a shortage of data scientists, data-literate managers, and even employees with basic ICT skills. Third, European policies and practices can impede the kind of data collection, use, and reuse that is critical for innovative approaches to social inclusion and protection.

THE MISSING CULTURE OF DATA

Governments and non-profit organizations generally lag behind private industry in adopting the digital technologies that are essential for data innovation. One particular area where this is evident in the EU is the slow growth of e-government—Internet-enabled government services—despite advances in digital technology.

In its annual assessment of EU member states’ digital technology use, the European Commission found uneven progress in the availability and use of e-government services. In 2015, only 26 percent of EU citizens submitted official forms online, with the highest usage reported in Estonia, Denmark, and Finland and less than 10 percent usage reported in Czech Republic, Bulgaria, and Romania. The Commission also found that Internet users in the majority of EU countries are more likely to shop or bank online than they are to access government services online.³²

Non-profit organizations are similarly slow to digitalize, and many have not yet fully embraced the role of data in improving their operations. For example, a survey of 467 non-profit professionals in the U.S. showed that most nonprofits collect data but do not know how to use the information in practice. Of the survey respondents, 46 percent reported that they did not regularly use data to inform their decisions.³³ Furthermore, those nonprofits that do regularly collect and use data tend to do so as part of marketing, fundraising, and communications efforts. A similar survey of UK nonprofits found that organizations were focusing their use of digital technologies and data in communications and fundraising, with only a small portion using technology in human resources, finance, or operations.³⁴

European nonprofits and government agencies miss out on opportunities to analyze operational data to improve their performance due to leadership’s lack of awareness and understanding of how the data they collect can be used across the organization. In addition, nonprofits report limited time and personnel to pursue data-driven projects.³⁵ Managers of social-service organizations often lack a strategy for how to

integrate data into their programs and operations, and as a result, do not know when to hire personnel with data training or to form institutional partnerships with data-analytics experts. The missing culture of data explains why many of the examples of progress in this report are driven by volunteer groups of data scientists rather than by the organizations themselves.

DIGITAL AND DATA SKILLS GAPS

To effectively formulate and implement data-driven strategies to tackle societal challenges, Europe needs more trained data scientists and data-literate managers across the public and private sectors. Workers with training and experience in digital technologies and managers who understand applications of data science to their fields are vital to supporting the strategic use of data within governments and social service organizations.

In 2014, The European Commission estimated that 32 percent of EU workers had insufficient digital skills and forecasted a deficit of 825,000 digital workers by 2020.³⁶ While not every employee needs to be a data scientist to use these tools, advanced data skills are crucial to innovating with data in the first place. Another Commission report measured the supply and demand of EU “workers who collect, store, manage, and analyze data as their primary, or as a relevant part of their activity,” and estimated a gap of 396,000 unfilled positions in 2015.³⁷ Because of this shortage, private companies recruit data scientists with some of the highest salaries in the labor market.³⁸ For government agencies and social-service providers who cannot compete with those salaries, the shortfall of data scientists is even greater.

Additionally, the digital skills gap and uneven progress in data for social good reflects the “digital divide” in the EU. Eighteen percent of the EU population has never used the Internet, and those without access tend to be more concentrated in a few countries, such as Romania, Bulgaria, and Greece. These countries also tend to have the lowest performance in measures of digital human capital like science, technology, engineering, and mathematics (STEM) graduates, workers with basic ICT skills, and ICT specialists.³⁹ As a result, progress in data-driven innovation may not reach these countries as quickly as it will the UK, Finland, and Denmark, which score higher on these measures. This is especially alarming when one considers that the refugee crisis and high unemployment are especially acute in Greece, where the digital skills gap is among the widest.⁴⁰

POLICIES LIMITING DATA SUPPLY

Another barrier to applications of data for social good is the restricted supply of reusable data in the European economy and society. Data-driven innovation for social good depends on the ability of government agencies, social-service providers, academic researchers, and private

companies to access data that may have been originally collected for different purposes.

There are three challenges with data supply. The first is government policies on open data. A majority of the innovative uses of data for social good discussed above rely on open government data. For organizations to effectively reuse this data for social innovation, governments need to publish data online, in machine-readable format, and in a timely manner. This is best enabled through explicit open-data policies at all levels of government.⁴¹ One such policy is the 2013 G8 Open Data Charter, in which G7 countries and Russia committed to the following principles: release open data by default, ensure high quality and quantity of data, make data usable, and release data for improved governance and innovation.

The Center for Data Innovation conducted a review in 2015 of open-data policies in these countries and found that France, Germany, and Italy had made little progress when compared to the UK, the leader.⁴² Likewise, the European Commission's Open Data Monitor reports broad disparities among national governments in the absolute number of datasets published, the number with open licenses, those in machine-readable format, and whether they meet minimum standards for metadata—data like author and date that provide important context for the data set. For example, in 2015 France had published a total of 15,653 data sets (47 percent of which were available in machine-readable format) while Poland had 378 (39 percent of which were machine-readable).⁴³

The second challenge is that governments have not always pursued the collection of data relevant to addressing social protection and inclusion. Disadvantaged populations in Europe will only benefit from data for social good if organizations collect and use data specific to these groups. Otherwise, certain social and demographic groups cannot benefit from data due to a “data divide”—the social and economic inequalities that result from a lack of collection or use of data about an individual or community.⁴⁴

These gaps in official data on minority populations often stem from biased data collection practices such as national surveys that only get distributed to numbered addresses. For example, the Roma—a nomadic, ethnic minority group settled across Europe—are a population that lives in data poverty in many European countries. The EU recently completed “The Decade of Roma Inclusion 2005-2015,” an intergovernmental effort to address discrimination and inequality in Roma communities. In 2015, the Decade of Roma Inclusion Secretariat Foundation produced the Roma Inclusion Index to assess the state of data collected on Romani peoples and found that the availability of official and non-official data varied greatly among countries.⁴⁵ These

kinds of inconsistencies in data collection across Europe obstruct efforts for social inclusion and result in under-informed policy decisions.

Similarly, migrants suffer and risk marginalization because of data poverty. European policymakers dealing with the challenges of increased immigration are limited by a lack of up-to-date information on migrants. Without timely data, governments cannot propose effective policies to assist migrants seeking a better life within their borders.

Because data on the movement of humans is inherently difficult to collect and maintain, governments often have no record of migrants until they go through months-long entry processes. Furthermore, the International Organization for Migration (IOM) determined that different organizations and countries were collecting data without sharing it effectively with the entire community of stakeholders working on migration issues, including national and local governments and aid organizations. To address that, the IOM established the Global Migration Data Analysis Centre in Germany, with the mandate to collate timely migration data from different sources and to disseminate data and expert insights.⁴⁶

The third challenge relates to restrictive data-protection laws. Specifically, the EU's General Data Protection Regulation (GDPR), adopted in April 2016, imposes strict requirements on organizations using personal data—information that could be used to identify an individual. This creates administrative burdens and higher compliance costs, discouraging organizations from reusing or sharing personal data.

For example, the GDPR includes a requirement for “data minimization” where organizations using personal data for purposes beyond the reason for which it was initially collected must obtain informed consent from each individual in the data set. This type of requirement creates high compliance costs that are likely to deter organizations from innovative reuses of data. Re-obtaining consent from more than 100,000 survey respondents for a benign new application of their data that could generate valuable social benefits would be an enormous task for any organization but especially so for smaller, resource-constrained non-profit organizations that support welfare services. The consent requirements could be particularly costly to medical research and inhibit potential public-health innovations that rely on large data sets (like personal health records) because it would stop researchers from accessing data from deceased patients, or others who cannot provide their consent a second or third time.⁴⁷

The GDPR also extends legal obligations on data security to both the “data controller,” who determines how personal data is to be used, and the “data processor,” who processes the data on behalf of the data controller. This creates new legal liabilities for organizations that provide data services like storage or processing. As a result, businesses and

nonprofits that specialize in data science may be hesitant to partner with social-service organizations that deal with personal data. The essential “data processors” in these partnerships may be dissuaded by the reputational risk of working with a non-profit organization that does not fully understand the new laws and the regulatory risk of high fines if they are found in violation of the GDPR.

Finally, the GDPR may discourage social-service organizations, especially smaller nonprofits, from even pursuing data-driven projects. The law allows organizations only two years to make all the necessary changes, creating a fairly steep learning curve for organizations that are not already using or sharing data and have thus not developed the administrative structures to meet the requirements. The GDPR also threatens organizations that do not comply with fines as high as four percent of annual turnover.⁴⁸ While not all data used by social service providers is personal data, some organizations may still be deterred from experimentation due to their lack of experience.

RECOMMENDATIONS

In order to replicate and scale the success of EU data-driven innovations for social good across Europe, policymakers in both Brussels and other national capitals, along with leaders in the private sector, should make data innovation a higher priority. This will mean focusing more on eliminating data poverty, encouraging cross-sector collaboration, and supporting regulations that enable responsible experimentation and use of data for social good. Doing so will not only help address key social policy challenges in Europe but will also help support the EU’s budding data economy.

First, EU agencies and national governments should work to address data poverty by reaching out to marginalized populations, like recently arrived immigrants, in data collection programs. If certain populations are unable to supply the necessary data due to issues of access to technology, digital literacy, or even biased data-collection practices, then these communities will not be able to benefit from data-driven innovations designed for social protection and inclusion.

European governments should ensure that official statistics and surveys include all population groups, with a particular focus on including historically uncounted populations like migrants and Romani people. In addition, European governments should digitize their civil registration offices to ensure that data from key legal documents, such as birth, death, and marriage certificates, can be integrated into national vital statistics. EU agencies and foundations should support social-science research projects that collect or use data about minority populations that may be excluded from official statistics, including the collection of data that can be disaggregated by categories like ethnicity, socioeconomic status, and gender. This type of data is most relevant to

the challenges of social exclusion and poverty that the EU is trying to resolve.⁴⁹

In order to further address data poverty, the EU should adopt the European Commission's current policy proposal to improve the collection and analysis of social statistics—data from social surveys.⁵⁰ The proposed policy combines multiple surveys and institutes standards for social data collection so as to eliminate duplication and enable interoperability or data linking.

By improving interoperability and standards, this policy will better enable researchers and government agencies to combine social-survey data with other data sets to explore complex social-science questions, such as the relationship between income and health. The policy also aims to diversify the sources of data for national statistics offices and encourage these government agencies to pool from different data sources.⁵¹ Ideally, this proposal will make it possible to integrate data on populations such as refugees that might not be collected via official surveys. Formalizing this proposal would also have the overall effect of ensuring that social data collected by EU agencies is structured, comprehensive, and comparable so as to support use and reuse for social policies and programs.

Second, European governments should support non-governmental efforts to use data for social good, such as through public-private partnerships and by encouraging broad reuse of open government data. For example, government agencies should pursue partnerships with NGOs, businesses and research universities, granting them access to otherwise restricted government data in order to address public goals.

These collaborations can be mutually beneficial. The private partner gains access to government data, resources, and experiences that can potentially improve their own products and services, and the government brings data and issue-based expertise from the private sector to work on a public problem. To accomplish this, governments can design legal agreements to encourage partnerships that share risk, protect proprietary or sensitive data, and allow the private partner to monetize additional related services or inventions. For example, U.S. federal government research labs use Cooperative Research and Development Agreements to collaborate with non-governmental partners on large-scale projects such as making massive government data sets on the environment available to the public through cloud-based services. Through these agreements, the private partners are permitted to bring research results and inventions to the market.⁵²

Government agencies should also sponsor data challenges or hackathons, where the government convenes civic hackers, civil-society organizations, and academics in a competition to develop innovative uses of open data. By offering prizes in the form of monetary support or

resources to bring innovations to scale, government agencies can incentivize experimentation with their data and encourage social entrepreneurship. For example, the French government agency in charge of open data and public information, Etalab, sponsors the “DataConnexions” competition, which rewards innovative applications, services, or data visualizations that reuse government data. Teams must utilize at least one public data set and may combine it with other data sets for a project that serves the public interest.⁵³ Previous winners include an app called “medication” that uses public data from the Ministry of Health and regional health offices to educate French consumers about available medications and their side-effects.⁵⁴

Third, EU policymakers should ensure that all programs targeting social inclusion and protection make effective use of data. The agencies that make policies, control funds, or plan programs related to social protection and inclusion should specifically encourage data initiatives in these fields. For example, the European Commission’s research funding mechanism, Horizon 2020, has in its 2016-17 funding cycle opportunities related to big data analysis of social inequalities. But the EU should mandate that all applications for funding related to research on social inclusion and protection require explanations on how grantees plan to use and share data. Furthermore, the EU should create a platform for intergovernmental collaboration and encourage national governments to share innovative uses of data in social programs and policies with other countries. This would also support the EU’s broader interest in evidence-based policymaking, in which research results are shared across related policy fields.⁵⁵

Fourth, EU and national governments should consult closely with both the research community and social service providers as they implement and enforce the requirements of the GDPR concerning the use, reuse, and sharing of personal data. Due in part to a concerted effort from health researchers across the EU, the GDPR includes an exemption for the reuse of personal data related to scientific research that could be interpreted broadly to include social-science data innovation.⁵⁶ Because member states and their courts will be responsible for defining this exemption in practice, policymakers should understand and advocate for allowing social service providers and their institutional partners to analyze and reuse personal data for research purposes. Additionally, these exemptions come with extra requirements to implement technical and organizational safeguards in processing personal data, including techniques such as anonymization where data is stripped of personally identifiable information and processed in isolation so as to prevent re-identification.⁵⁷ Member states interpreting these exemptions should work closely with researchers to create data-sharing practices that balance data protection and social innovation.

Finally, EU regulators should recognize that the GDPR may obstruct innovative uses of personal data in social services by placing strict requirements on organizations that generally have less experience using data to advance their goals. Through cross-sector collaboration, European governments at all levels should educate NGOs working on social inclusion about their legal responsibilities under the GDPR, and create opportunities for public consultation on possible exemptions for socially beneficial data uses.

Education is not enough, however. EU and member state policymakers should lower fines for social-service nonprofits working with personal data and allow more time for organizations to comply. One particular area of concern is the potential for legal obligations on both “data collectors” and “data processors” to have a chilling effect on data innovation that relies on the involvement of outside expertise. Policymakers should consider exemptions for “data processors” in projects with public missions and encourage partnerships that allow private sector data to be made available for public benefit.

CONCLUSION

Innovative uses of data can greatly increase actionable insights that will help European governments, businesses, nonprofits, and individuals to build a more inclusive, welcoming, and prosperous society. However, these benefits will not be widespread unless Europe adopts a coherent strategy for using data for social good. This includes ensuring both that Europe’s social programs better integrate data innovation and that other policies, particularly the GDPR, do not act as major barriers to data for good projects.

REFERENCES

1. “Demographic Analysis,” European Commission - Employment, Social Affairs, and Inclusion, accessed July 29, 2016, <http://ec.europa.eu/social/main.jsp?catId=502>.
2. Lord Ashcroft Polls, June 24, 2016, <http://lordashcroftpolls.com/2016/06/how-the-united-kingdom-voted-and-why>.
3. Laura Adler, “How Smart City Barcelona Brought the Internet of Things to Life,” *Harvard Kennedy School Ash Center for Democratic Governance and Innovation*, February 18, 2016, <http://datasmart.ash.harvard.edu/news/article/how-smart-city-barcelona-brought-the-internet-of-things-to-life-789>.
4. Daniel Castro and Travis Korte, “Data Innovation 101,” Center for Data Innovation, November 2013, <https://www.datainnovation.org/2013/11/data-innovation-101/>.
5. “European Data Portal,” last modified September 15, 2016, <http://www.europeandataportal.eu/en>.
6. James Manyika, et al, “Open Data: Unlocking innovation and performance with liquid information,” (McKinsey Global Institute, October 2013), <http://www.mckinsey.com/business-functions/business-technology/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>.
7. “Where the Work Is,” Burning Glass Technologies, accessed August 2, 2016, <http://wheretheworkis.org>.
8. “Helping millions of people break out of unemployment,” Bayes Impact, accessed July 30, 2016, <http://www.bayesimpact.org/focus/unemployment>.
9. OPENSIMR website, accessed August 2, 2016, <http://openskimr.eu>.
10. PiC Company Website, accessed August 2, 2016, <http://www.pic.is>.
11. “Youth Employment,” European Commission, accessed August 2, 2016, <http://ec.europa.eu/social/main.jsp?catId=1036>.
12. “Skills Panorama,” European Centre for the Development of Vocational Training, accessed August 2, 2016, <http://skillspanorama.cedefop.europa.eu/en>.
13. “Europe 2020 indicators - employment,” Eurostat, last updated July 19, 2016, http://ec.europa.eu/eurostat/statistics-explained/index.php/Europe_2020_indicators_-_employment.
14. “The Access Project,” accessed August 2, 2016, <http://www.theaccessproject.org.uk>.
15. “Improving Access to Education by Supporting Tutors,” DataKind, accessed August 2, 2016, <http://www.datakind.org/projects/improving-access-to-education-by-supporting-tutors>.

-
16. Inclusive Learning website, accessed August 2, 2016, <http://www.inclusive-learning.eu>.
 17. “Evidence Hub,” Learning Analytics Community Exchange, accessed August 2, 2016, <http://evidence.laceproject.eu>.
 18. Paul MacDonnell and Daniel Castro, “Europe Should Embrace the Data Revolution,” (Center for Data Innovation, February 2016), <http://www2.datainnovation.org/2016-europe-embrace-data-revolution.pdf>.
 19. Survey of Health, Ageing and Retirement in Europe, “Mutual Learning: Joint pan-European research improves the understanding of ageing societies,” press release, July 18, 2016, <http://www.share-project.org/home0/news/article/press-release-mutual-learning-joint-pan-european-research-improves-the-understanding-of-ageing-so.html>.
 20. Ibid.
 21. Axel Börsch-Supan, et al, “Early retirement for the underprivileged? Using the record-linked SHARE-RV data to evaluate the most recent German pension reform,” in *Ageing in Europe - Supporting Policies for an Inclusive Society*, edited by Axel Börsch-Supan, Thorsten Kneip, Howard Litwin, Michal Myck, and Guglielmo Weber (Berlin, Boston: De Gruyter, 2015), 267-268.
 22. Joshua New, “10 Bits: the Data News Hotlist,” *The Center for Data Innovation*, July 15, 2016, <https://www.datainnovation.org/2016/07/10-bits-the-data-news-hotlist-79>.
 23. Ed Gent, “Artificial Intelligence Could Help Catch Alzheimer’s Early,” *Live Science*, July 7, 2016, <http://www.livescience.com/55313-artificial-intelligence-alzheimers-early-detection.html>.
 24. “EU funded societal challenges projects,” Digital Single Market - European Commission, accessed August 5, 2016, <https://ec.europa.eu/digital-single-market/en/lab-market-what-happens-after-projects-end>.
 25. “Migrant crisis: Migration to Europe explained in seven charts,” BBC, March 4, 2016, <http://www.bbc.com/news/world-europe-34131911>.
 26. Data For Good France, accessed August 5, 2016, <http://www.dataforgood.fr/#projets>.
 27. “What is Food Cloud?,” accessed September 6, 2016, <http://food-and-community.tesco.ie/home/supporting-local-communities/food-cloud/what-is-foodcloud>.
 28. Refugee Aid App website, accessed July 30, 2016, <http://refugeeaidapp.com>.
 29. Ibid.
 30. “Connecting Refugees,” (UNHCR, June 2016), <http://www.unhcr.org/5770d43c4.pdf>.

-
31. “Mobile making a difference in the lives of refugees,” infobip website, last updated May 25, 2015, <http://www.infobip.com/en/blog/sms-and-ussd-technology-making-a-difference-in-the-lives-of-refugees/>.
 32. “Europe’s Digital Progress Report 2016,” European Commission, May 20, 2016, <https://ec.europa.eu/digital-single-market/en/news/europes-digital-progress-report-2016>.
 33. “The State of Data in the Non-Profit Sector” (everyaction and nonprofit hub, March 18, 2016), <https://act.everyaction.com/2016-nonprofit-data-whitepaper>.
 34. “Business transformation and the role of Heads of Digital” (Eduserv, 2016), <http://www.eduserv.org.uk/insight/reports/Business-Transformation-and-the-role-of-Heads-of-Digital>.
 35. “The State of Data in the Non-Profit Sector.”
 36. “Digital Agenda Scoreboard 2015,” European Commission - Digital Single Market, accessed August 6, 2016, <https://ec.europa.eu/digital-single-market/en/human-capital>.
 37. Gabriella Cattaneo et al., “European Data Market Report: Data Workers and Data Skills Gaps” (second interim report from IDC and Open Evidence, presented to DG Connect-European Commission, June 9, 2016), <http://datalandscape.eu/study-reports>.
 38. Gil Press, “The Supply and Demand of Data Scientists: What the Surveys Say,” *Forbes*, April 30, 2015, <http://www.forbes.com/sites/gilpress/2015/04/30/the-supply-and-demand-of-data-scientists-what-the-surveys-say>.
 39. “Digital Agenda Scoreboard 2015,” European Commission.
 40. Ibid.
 41. Daniel Castro and Travis Korte, “Data Innovation 101,” (Center for Data Innovation, November 2013), <https://www.datainnovation.org/2013/11/data-innovation-101/>.
 42. Daniel Castro and Travis Korte, “Open Data in the G8: A Review of Progress on the Open Data Charter,” (Center for Data Innovation, March 2015), <http://www2.datainnovation.org/2015-open-data-g8.pdf>.
 43. “Open Data Monitor,” accessed August 22, 2016, <http://opendatamonitor.eu/frontend/web/index>.
 44. Daniel Castro, “The Rise of Data Poverty in America,” (Center for Data Innovation, September 2014), <http://www2.datainnovation.org/2014-data-poverty.pdf>.
 45. Decade of Roma Inclusion Secretariat Foundation, Roma Inclusion Index (Budapest: Decade of Roma Inclusion Secretariat Foundation, 2015), 12-13, http://www.romadecade.org/cms/upload/file/9810_file1_roma-inclusion-index-2015-s.pdf.
 46. Global Migration Data Analysis Centre, accessed August 30, 2016, <http://iomgmdac.org>.

-
47. Travis Korte, “Proposed EU Data Protection Regulations Could Impede Medical Research,” (Center for Data Innovation, October 2014), <https://www.datainnovation.org/2014/10/proposed-eu-data-protection-regulations-could-impede-medical-research/>.
 48. Dawn Varley, “Why you should know what the GDPR is - and what you can do now” *Purple Vision Blog*, April 26, 2016, <http://purple-vision.com/why-you-should-know-what-the-gdpr-is-and-what-you-can-do-now/>.
 49. Roma Initiatives, “No Data–No Progress” (Open Society Foundations, June 2010), <https://www.opensocietyfoundations.org/sites/default/files/no-data-no-progress-20100628.pdf>.
 50. European Commission, “Question and Answers: Towards better social statistics for Europe,” fact sheet, August 24, 2016, http://europa.eu/rapid/press-release_MEMO-16-2868_en.htm.
 51. Ibid.
 52. Alexander Kostura and Daniel Castro, “Three Types of Public-Private Partnerships that Enable Data Innovation” (Center for Data Innovation, August 2016), <https://www.datainnovation.org/category/publications/in-depth>.
 53. “DataConnexions,” Etalab, accessed August 22, 2016, <https://www.etalab.gouv.fr/dataconnexions>.
 54. “DataConnexions #5 Palmarès & Retour en images sur la finale,” Etalab, accessed August 22, 2016, <https://www.etalab.gouv.fr/dataconnexions-5-palmares-retour-en-images-sur-la-finale>.
 55. “ESN Evidence in Social Services,” European Union, last modified August 21, 2015, http://europa.eu/epic/news/2015/20150819-esn-social-services_en.htm.
 56. Chapter II, Article 5(1)(b), Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), European Parliament and the Council 2016/679 (27 April 2016).
 57. Chapter IX, Article 89, General Data Protection Regulation.

ABOUT THE AUTHORS

Alexander Kostura is the 2016 Google public policy fellow at the Center for Data Innovation. Kostura is passionate about information and communications technologies as tools for inclusive economic growth, good governance, and social welfare. He has most recently conducted research in corporate data sharing for social good, specifically in international development and humanitarian response. Kostura holds a B.S. in foreign service from Georgetown University and an M.A. in law and diplomacy from the Fletcher School at Tufts University.

Daniel Castro is director of the Center for Data Innovation and vice president at the Information Technology and Innovation Foundation. His research interests include health IT, data privacy, e-commerce, e-government, electronic voting, information security, and accessibility. Previously, Castro worked as an IT analyst at the Government Accountability Office where he audited IT security and management controls at various government agencies. He has a B.S. in foreign service from Georgetown University and an M.S. in information security technology and management from Carnegie Mellon University.

ABOUT THE CENTER FOR DATA INNOVATION

The Center for Data Innovation is the leading global think tank studying the intersection of data, technology, and public policy. With staff in Washington, DC and Brussels, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as predictive analytics, open data, cloud computing, and the Internet of Things. The Center is a nonprofit, nonpartisan research institute affiliated with the Information Technology and Innovation Foundation.

contact: info@datainnovation.org

datainnovation.org