UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT

OCTOBER 2017

UNCTAD Research Paper No. 10
UNCTAD/SER.RP/2017/10/Rev.1

Digital tools for foresight

Martin Hilbert
Department of Communication, University of California, Davis
hilbert@ucdavis.edu

Abstract

The report illustrates how digital tools, namely big data and computer simulations, can be used to enhance foresight exercises. By using concrete case studies it also shows the feasibility of using these computational methods in developing countries and for the implementation of the 2030 Agenda for Sustainable Development. The report concludes by discussing the considerations that need to be taken into account when doing foresight exercises of a digital future in the context of the implementation of the sustainable development Goals.

Key words: Foresight, sustainable development, sustainable development goals, science, technology and innovation, big data, computer simulations

UNITED NATIONS

Contents

Acknowledgements

This research paper was prepared by a team comprising Martin Hilbert, Department of Communication; University of California, Davis (with the assistance of Suzanne Cheng and Shengling (Wilburn) Wang, University of California, Davis), Bob Bell, Katalin Bokor, and Claudia Contreras (UNCTAD), under the guidance of Dong Wu (Chief, Science, Technology and Innovation Policy Section, UNCTAD). Comments from Professor Ozcan Saritas on an earlier version of this paper are gratefully acknowledged.

Introduction: Foresight for the Sustainable Development Goals in a digital age

The Sustainable Development Goals (SDGs) are 17 sets of goals with 169 targets to be achieved by 2030. As political aspirations, they aim at a future that is yet uncertain. In other words, this future is not deterministically given to us, but has yet to be socially constructed. Therefore, concrete policy tools are needed to deal with and shape this uncertain future. This report argues that the natural tool for doing so is to use foresight methods. Foresight involves bringing together key agents of change and sources of knowledge, to develop strategic visions and anticipatory intelligence (Miles et al., 2008: 11), to shape the future, and it is often executed through participatory consultations. Foresight not only provides approaches and methods about scanning issues that can be measured today (i.e. trends), but also informs policy-makers about future issues or wild cards that are not yet considered in policy design but must be tackled today if we are to develop our societies in a sustainable way. Foresight makes particularly sense in addressing sustainable development challenges. It is vital for any forward planning or policy activity to be able to meet future challenges proactively.

One of the main challenges of current foresight exercises is that we live in a period of technological change. Technological change is exponentially fast-paced, all-embracing, and global in nature. An important driver of this current phase of technological change is digitalization. The digitalization of the entire stockpile of technologically mediated information has taken less than 30 years, as less than one percent was in digital format in the mid-1980s (the rest in analogue format on paper, tape, vinyl, etc.), and more than 99 % in digital format today (extrapolated from (Hilbert & López, 2011)). The exponentially fast innovation cycles of digital technology create high uncertainty; its general purpose applicability embraces all sectors, and its inherent defiance of national borders intermingles the most diverse aspects of a heterogeneous world.

All of this leaves us with the aspiration to shape an extremely uncertain and fast paced future. This paper presents examples of how digital tools can help to implement future foresight (as means to help achieving the SDGs). Digital big data footprint can be used to detect empirical realities, artificial intelligence to extract insights from data, and computer simulations to explore future scenarios that are different from today's reality; to explore the world as we would like it to be; the world where the SDGs are the reality. The paper shows that these tools can be extremely useful to foster foresight studies in developing countries. They are cost-effective, scalable, and sensitive to local contexts. The paper explores various practical applications that use these computational tools in order to implement Sustainable Development Goals.

After this Introduction, Chapter 1 focuses on digital tools, namely big data and computer simulations, which can be used to enhance foresight exercises. It also reviews methods from the field of what is nowadays known as 'computational (social) science'. Chapter 2 presents concrete case studies that show the feasibility of using these computational methods in developing countries and for the implementation of the SDGs. Chapter 3 discusses the considerations that need to be taken into account when doing foresight exercises of a digital future. It places foresight efforts into the challenging context of our global reality, especially focusing on the particularities for implementing the 2030 Agenda for Sustainable Development in developing countries, and addresses a caveat about the role, scope, and limits of foresight exercises.

1. Digital tools for Foresight

Information and communication technologies (ICT) are currently transforming the way information is obtained, knowledge is created, and insights are derived. This revolutionizes the way research is being done. The applications reach from inductive empirical inquiry (e.g. 'big data'), to deductive theoretical scholarship (e.g. 'computer simulations'). The resulting opportunities are very broadly labelled as "computational social science" (Cioffi-Revilla, 2014; Conte et al., 2012; Lazer et al., 2009).

Computational science tools have given new wind to a variety of mixed method approaches used in foresight exercises (e.g. Alexander & Maiden, 2005; Haegeman, Marinelli, Scapolo, Ricci, & Sokolov, 2013; Hansen, Rasmussen, & Jacobsen, 2016; Kwakkel & Pruyt, 2013). Empirical evidence allows to ground foresight into reality, while theoretical models allow to explore futures that are desired, but for which no empirical data exists. With this paper we argue that it is useful to make an explicit effort to adopt digital methods as an integral part of foresight methods. This does not argue to completely replace any human involvement in foresight exercises with algorithms, but to complement human planning with data-driven decision making and the formal development of scenarios.

Given that ever more of human conduct is taking place in digital networks, and given that digital conduct inevitably leaves a digital footprint, the social sciences currently have access to an unprecedented amount of data on the most diverse aspects of our reality and its dynamics (Lazer et al., 2009). The catch-phrase here became 'big data' (Manyika et al., 2011; Mayer-Schönberger and Cukier, 2013). Big data considers the accumulation and analysis of greatly increased information resources, beyond the storage and analytical capacity of earlier hardware and software resources. Big data is usually used to describe a massive volume of both structure and unstructured data that it is difficult to process with traditional database and software techniques due to its magnitude (UN Global Pulse, 2012: 13). It is made possible by increases in both data storage capacity and the range of available data sources.[1] Its impact on the social sciences has been compared with the impact of the invention of the telescope for astronomy and the invention of the microscope for biology (providing an unprecedented level of detail about the system of interest). Confronted with such increase in the level of perceivable granularity in social dynamics, analysts and policy makers have an inevitable obligation to make use of it to inform future planning.

This digital deluge is nowadays not only used to predict stock market behaviour, commercial consumption patterns, and traffic jams, but also to predict epidemics, medical necessities, environmental disasters, and poverty levels (Hilbert, 2016a; Manyika et al., 2011; Mayer-Schönberger and Cukier, 2013). For example, using simple metadata digital footprints like call duration and call frequency, it has been shown that one can predict socioeconomic, demographic, and other behavioural trades with 80-85% accuracy (J. E. Blumenstock & Eagle, 2012; J. E. Blumenstock, Gillick, & Eagle, 2010; Frias-Martinez & Frias-Martinez, 2014; Frias-Martinez, Frias-Martinez, & Oliver, 2010; Raento, Oulasvirta, & Eagle, 2009). This basically allows to reverse-engineer much of a traditional household survey or census (Frias-Martinez and Virseda, 2013), which is information often lacking in least developed countries. In a combination of machine learning and publicly available data, including high-resolution daytime and night-time satellite imagery, researchers recently were able to explain up to 75% of the variation in local-level economic outcomes in Nigeria, Tanzania, Uganda, Malawi, and Rwanda (Jean et al., 2016). In other words, they were able to predict the distribution of poverty better than any other existing approach.

It seems not only natural to take advantage of these computational science methods for foresight purposes, but also imperative. This is especially relevant in the light of the lack of existing data to inform specific development goals in developing countries.

For practical purposes, it is important to recognize that the big data paradigm includes a heavy focus on analyses, often involving machine learning, but also other forms of artificial intelligence. For one, the catch-phrase 'big data' refers to new sources of data. The digital footprint created with each digital communication and transaction can replace traditional data sources (like surveys) with proxy indicators that correlate with the variable of interest. The benefit is the low cost and real-time availability of the digital proxy indicator. The epitome is Google's illustrious use of 50 million most-common search terms as a proxy for the spread of the

[1] Report of the Secretary General on Information and Communication technologies for inclusive social and economic development, p.8

seasonal flu (Ginsberg et al., 2009). Secondly the notion of big data goes beyond data itself and focuses on methods of data analytics to inform intelligent decisions. Independent from the specific giga-, tera-, peta-, or exabyte scale, the big data paradigm argues to systematically place analytic treatment of data at the forefront of intelligent decision-making. The process can be seen as the natural next step in the evolution from the "Information Age" and "Information Societies" to "Knowledge Societies" (for more see Hilbert, 2013). Building on the digital infrastructure that led to vast increases in information during recent decades, the big data paradigm focuses on converting this digital information into knowledge that informs intelligent decisions. Continuing with the previous example, Google processed an impressive 450 million different mathematical models in order to test for correlations between online search terms and flu outbreaks reported by official data. Eventually, 45 search terms were identified that outperformed traditional models of flu outbreak with real-time predictions (Ginsberg et al., 2009).

While the opportunities of big data are enormous, especially for developing countries in which traditional statistics are scarce (Hilbert, 2016; Letouzé, 2012; WEF, 2012), the very same Google flu trend case also exemplifies the ultimate limitation of (big) data analytics: the fact that all data is from the past. The application of the same Google flu trend algorithm became increasingly out of sync with actual flu epidemics over time (Lazer et al., 2014). The reason is straightforward: reality (in this case search behaviour) is changing over time, which requires a different input and therefore changes the results if the same algorithm is used. The big data algorithm was optimized for a reality that does not exist anymore, and applying it to the new reality will be deceptive.

What the best big data analytics can do is to update estimations in 'real-time': so-called 'nowcasting'. Both the terms are actually misnomers, since the very act of recording converts it into 'past-time' data. Data is always from the past, per definition. Therefore, it can only detect patterns that have occurred in the past. When the past and the future follow the same stationary logic, data analytics is extremely useful in predicting future patterns. In other words, if the generative mechanism that produced the past data continues to be valid in the future, then the past data can tell us something about the future. However, if significant changes occur in the system's dynamic, empirical statistics based on past data are at best limited to predicting the future, if not deceiving (Hilbert, 2014a, 2015b). In a technical sense, the key concept here is the 'stationarity' of the data source. This is usually in the fine-print of most empirical data analysis (big or small data) and refers basically to this idea: if the generative mechanism is 'stationary', then past data is useful to predict the future. If not, mere descriptive statistics cannot help us, because they were produced from a different reality. Most statistical forecasting methods are based on the assumption that the time series can be rendered approximately stationary through the use of mathematical transformations.

This argument is crucial for our purposes since the 2030 Agenda for Sustainable Development[2] is an aspiration that explicitly aims at creating a world that is different from today's reality. A world with no poverty (SDG 1), zero hunger (SDG 2), clean water and sanitation (SDG 6) and affordable and clean energy (SDG 7) is decisively different from the world in which we lived in 2015 when the goals were adopted. It is the definition of development goals to change the 'stationarity of the system'. It is the ambition to replace the dynamic that produces poverty with a new one that does not. The data footprint left behind by the old mechanism can only help us to understand the old mechanism. This is a useful first step, but it cannot automatically tell us what will happen if we change it. We have never lived in a world with zero hunger and affordable and clean energy. Who could tell us what it would be like and how to get there? Data from the past alone cannot.

In order to predict a future that has never been, theory-driven models are necessary (Hilbert, 2015b). These allow variables to be adjusted with values that have never existed in statistically observable reality. Digital tools also act as a game changer in this challenge. Computational simulations allow to setting up theory-driven models that greatly expand the scope and level of sophistication of traditional 'paper-and-pen' models. While traditional models are only able to handle a very limited number of variables (at most a dozen or so), today's computational power allows creating mathematically formalized models with thousands and even millions of dynamic variables. Such computer simulations of artificial societies have no conceptual

[2] See A/RES/70/1.

limitations on the achievable level of detail and accuracy (in practice they do face computational limits). Most recent simulations are based on individual agents ('agent-based models') (Epstein, 2005; Epstein and Axtell, 1996; Helbing and Balietti, 2011; Miller and Page, 2007; Schelling, 2006; Wilensky and Rand, 2015), which allow to simulate the interactions of individuals (not merely aggregated factors or variables). In essence, they are similar in nature to popular computer strategy games, such as SimCity. The researcher sets up a reality (often based on existing modules of behaviour and structure) and then lets the model run in order to observe how the myriad of interactions often produce surprising and counterintuitive outcomes. Agent-based computer simulations are essentially equation based, just that differential equations are both programmed into the flexible behaviour of the agents and result from their interaction. Compared to traditional equation-based model, "the agent-based model is more versatile and therefore preferable" (Sukumar and Nutaro, 2012: 6).

These models can and should be calibrated by empirical data. "A good complex systems model both begins and ends with data: Low-level data [from the micro-scale] is used to formulate the assumptions about the building blocks of the model, and both high [macro-scale] and low-level data is also used to test whether the resulting emergent phenomena properly correspond to those observed in the real world" (Farmer, 2012). However, they go beyond mere data, and require theoretical assumptions of behaviour and interactions. This also allows transferring insights from one case to the other, often with surprising results. Theoretical models allow us to transfer local and small-scale evidence into different contexts. A developed Africa will not simply be a statistically extrapolated version of Europe's development trajectory. The statistical model of a fitted curve to some data points cannot give us an in-depth insight here. However, we can take bits of empirical evidence from different instances and piece together an assumed theoretical model of Africa's development trajectory. This would show us how Africa would develop 'in theory'. Both theory and empirical data are needed. The combination of both empirical (big) data analysis of the digital footprint and theoretical computer simulation models of possible scenarios provide a wide variety of opportunities to inform foresight exercises.

2. Examples of how Big data and computer simulations can inform foresight studies

Several UN agencies (including the UN Statistical Commission and UN Global Pulse) have recognized that the potential of big data is an important dimension of ICTs' contribution to the monitoring and assessment of SDGs. In the following, we will review several case studies that provide evidence on how big data methods can be used to produce intelligence to inform foresight studies. The examples presented as well as the Sustainable Development Goals that they are focused on were selected with a view to illustrate our argument and does not mean that these Goals are more relevant than others. For additional cases showing the use of big data for development see for instance Hilbert (2016a) and Letouzé (2012).

SDG2: Sustainable food production in Colombia

Focus area: SDGs 2: End hunger, achieve food security and improved nutrition, and promote sustainable agriculture, Target 4: By 2030, ensure sustainable food production systems and implement resilient agricultural practices that increase productivity and production, that help maintain ecosystems, that strengthen capacity for adaptation to climate change, extreme weather, drought, flooding and other disasters and that progressively improve land and soil quality.

The International Center for Tropical Agriculture is an organization that strengthens agricultural technologies, innovations, and new knowledge that helps small farm owners improve their crop yields, incomes, and usage of natural resources. Scientists collaborated with the Colombian Government, Agriculture and Food Security, and Colombia's National Federation of Rice Growers to collect a big volume of weather and crops data in last decade in Colombia. The initiative predicted upcoming climate changes in Córdoba, a major rice-growing area in Colombia. The results are highly localized. In the town of Saldaña, for example, the analysis showed that rice yields were limited mainly because of solar radiation during the grain-ripening stage. Meanwhile, in the town of Espinal, the team found that it suffered from sensitivity to warm nights. Solutions do not have to be costly, as such information can help farmers to avoid losses simply by sowing crops in right period of time. The climate change foresight helped 170 farmers in Córdoba avoid direct economic losses of an estimated $ 3.6 million and potentially improve the productivity of rice by 1 to 3 tons per hectare. To achieve this, different data sources were analysed in a complementary fashion to provide a more complete profile of climate change. So-called 'data fusion' is a typical big data technique. Additionally, analytical algorithms were adopted and modified from other disciplines, such as biology and neuroscience, and were used to run statistical models and compare with weather records. With support from national and international organizations such as the World Bank and the Fund for Irrigated Rice in Latin America, the initiative has started to approach rice growers associations in other countries, including Nicaragua, Peru, Argentina and Uruguay.[3]

SDG 8: Global networks against forced labour

Impact area: SDGs 8: Decent work and economic growth, Target 7: Take immediate and effective measures to eradicate forced labour, end modern slavery and human trafficking and secure the prohibition and elimination of the worst forms of child labour, including recruitment and use of child soldiers, and by 2025 end child labour in all its forms.

Made in a Free World is a non-profit organization that tracks and identifies raw materials and goods associated with slave labour, using data uploaded by the companies. Using digital footprints, the organization's marketing analysis program FRDM (Forced Labor Risk Determination & Mitigation) helps businesses and organizations to look into their supply chains to identify slave and child labour, or products that come from conflict zones. The database contains over 54,000 products, services, and commodities from the United Nations Standard Products and Services Code. Business products and services have been broken down into material and parts, and traced to their origin. Country risk is generated using internationally recognized and peer-reviewed reports. The organization also partnered with Ariba Network, the largest business-to-business trading platform that connects more than two million companies with their global supply chains. The tool is sponsored by the US federal government and used by federal contractors and independent companies. When a company uploads purchased data, the software compares the data to its Global Slavery database. An accompanying action plan aims at strengthening vendor agreements, improving supplier code of conduct, and providing suggestions to companies on effectively getting rid of slavery in their supply chains.[4]

SDG 3: Big Data for ending Tuberculosis in Brazil

Impact area: SDG 3: Ensure healthy lives and promote well-being for all at all ages. Target 3: By 2030, end the epidemics of AIDS, tuberculosis, malaria and neglected tropical diseases and combat hepatitis, water-borne diseases and other communicable diseases.

The Brazilian Cadastro Único (Unified Registry in English) is a database that aims at identifying all existing low-income families for inclusion in social welfare and income redistribution programs. Through a computerized system, the federal government of Brazil consolidates the data collected in the database to formulate and implement specific policies that contribute to the reduction of social vulnerabilities that these families are exposed to. The system is a gateway to 20 public policies. To register, family members must have a monthly income of up to half the minimum wage per person (less than R$ 394, or about USD 125) per person. As the social program expanded, extensive efforts were made to ensure all potential beneficiaries were included and to verify their information –with over 1200 crews visiting homes across the country and

[3] Sources: (The Guardian, 2014; CCAFS, 2015).
[4] Sources: (The Australian Business Review, 2015; How Big Data is driving sustainability, 2016; Business Wire, 2015)

even trekking deep into the Amazon by canoe. As of 2016, the Cadastro Único was tracking the information of over 103 million people.[5] The system also captures tuberculosis cases and collects information that distinguishes people that have received cash transfers from those who have not. The information contained in this massive database has helped to forecast the prevalence of tuberculosis in the country until 2025. It is estimated that if the related social programs were to stop today, the Tuberculosis prevalence in 2025 would be approximately 6% higher than if the programs were to continue. By contrast, an expansion of the cash transfers to all tuberculosis cases would contribute to a faster decrease in the prevalence of the disease, with a 33% lower prevalence in 2025. Policy makers and administrators are better informed by these foresights and can make more effective policies related to tuberculosis, regulate respective health and pharmaceutical markets more efficiently, and use these insights to explore the side-effects of tuberculosis on other aspects of the Brazilian economy and society, including family policies, labour rights, and living conditions. [6]

SDG 4: Universal education in India

Impact area: SDG 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all; Target 1: By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes.

Akshara Foundation is a Non-Government Organization in India that aims at narrowing the gaps in the universalization of pre-school and primary education in India. The organization has been collecting primary data from 40,000 schools and over one million students since 2006. The organization itself did not have sufficient capacities to fully make use of these data. Motivated by the abundance of primary data available, Hewlett Packard approached Akshara Foundation with a fresh analytical perspective to turn the data into useful information identifying the links between education resources, facilities and Indian children's education experiences. Data scientist from HP's Analytics and Data Management Practice helped to clean and organize previously messy and incomplete dataset of Karnataka, India. Together they created a custom dashboard that provides insights of the ratio of students and teachers and the adequate number of books per student. Machine learning algorithms were used to look for hidden patterns in the data. One surprising finding is that the introduction of separate bathrooms reduces drop-out rates in school. Such insights allow administrators and policy makers to set any foresight and policy planning exercise on a more comprehensive outlook of reality. Sources: (Gutierrez, 2015; *Datanami*, 2015).

SDG 6: Water management in Taiwan province of China

Impact area: Goal 6: Ensure availability and sustainable management of water and sanitation for all; Target 5: By 2030, implement integrated water resources management at all levels, including through transboundary cooperation as appropriate.

Despite the Province's high annual rainfall, only about 20% of the rain can be used as a water resource due