Governance of Big Data development ecosystems

Author: Diego Silva
2018

Abstract

This document pretends to describe the current challenges faced by national governments, and consequently society and the global community, with the increase of information production and the manifold possibilities to use it, disseminate it and regulate it. Governance of Big Data means recognizing the inherent complexity of its development ecosystems and proposing arenas for the articulation of actions pursuing a better use of information and increment benefits from this use to society. Currently discussions on development issues have highlighted the importance of information and data management as crucial for improvements in quality of life, transparency of governments, entrepreneurial ecosystems and environmental protection; so sustainable development under current contexts required profound conversations and immediate actions on data ecosystems where Big Data plays a key contemporary role.
1. Data Revolution: contextualizing current data challenges

Information has been part of human communities for most part of their existence. Before the existence of written language information was kept in artifacts, drawings or symbols. Later written documents became the best way to retain information and save it from the past of time. Far ahead, books, potentialized by the invention of the printing press became the ultimate storage technique for information and so libraries the terminals of human knowledge. Information was power, societies knew it and therefore encourage its production, protect it and use it. With the development of computing techniques in the last centuries, but particularly with the massive explosion of computing terminals in the last decades of the twentieth century (initially with personal computers and later with mobile devices), information and its impact on society has change on a dramatical way hard to fully grasp.

The enormous amount of information that flows through one person’s pocket every single day is larger than ever imagined before. People consume and produce information constantly and when multiplied by millions of users the growth is exponential and the potential uses are infinite. Although societies have been preparing for this change, the flow has been hard to canalize in the initial years, on one hand capacity to storage, analyze and utilize information was shorter than the amount of information generated and on the other hand many people, institutions and governments have not completely see the picture and organize themselves to engage with the current demands for adequate use of the information. Challenges are huge therefore actions are required to propose possible future uses of information capable of promoting sustainable development via monitoring and decision making while at the same time allowing some degrees of preventing the whole ecosystem getting out of control.

The basic challenge is to understand the possible future scenarios and propose mechanisms to take advantage of the use of information. Hopefully reducing the power of the few that have been able to control some flows of information, by funneling it towards its interests such as large information corporations and allowing to expend the benefits of information usage to the public, to small enterprises, to social initiatives and to the daily lives of people. The next section contextualizes the Data Revolution and the demands faced by the world and nations under current context.
a. **UN Data Revolution Document (A world that counts)**

In November 2014 the United Nations presented a document requested by the General Secretary to an Expert Advisory Group, title "A world that counts. Mobilising the Data Revolution for Sustainable Development". This document discussed the role of data in sustainable development while identifying the challenges faced by national governments to promote the use of information and data for their specific development policies. It was an urgent call for actions required to start developing an international ecosystem for adequate use of data and demanded: (1) the development of principles and standards on legal, technical and moral arenas of data use, (2) sharing technology and innovations for the common good via a network of exchange, (3) promoting equalizing scenarios via resources investments, conferences and cooperation aimed to reduce international disparities. The document was aligned with the 2030 Agenda for the sustainable development and set a navigation route for the coming years in terms of the complexity of the discussion of designing and building adequate information ecosystems.

Since the document was agreed and published, the United Nations, multiple international organizations, information companies and stakeholders, and of course national authorities have been constantly discussing the new possible paths and exploring initiatives that will allow to enrich the evolving information scenario of the world. There has been an international public arena for debate and some countries have been able to use it to start implementing interesting and relevant initiatives to learn how to funnel the vast flow of information towards the benefit of their communities.

b. **Information as the new oil: The upsurge of information companies**

Although many actors in the social arena did not perceive the importance of changes brought by the exponential growth of information and data, some others took advantage of their awareness and began to get ready long time ago. Information became for many the new oil, and some few have been pioneers therefore capable of taking a robust power position in the complex scenario. As with any other essential resource there are many that might think that it is correct to have it accessible to everyone, but on the other side there are always those that control the access to the resource and will be willing to concentrate their power by holding their capability to access the resource. Information as the new oil means two things: (1) that it is a new valuable
resource that can be use, exploited, sell and resell among many other incentives that will guarantee its existence and (2) that there will be competition for access to information, and that powers will arise around the emerging tensions. Consequently, information, as many other essential commodities of modern life, becomes crucial for the debates and conversations on sustainable development.

Information as a valuable good, similar to oil, becomes a resource with enough incentives for entrepreneurship, capital allocation, resources investment and productive innovations. This means that new markets will emerge by companies providing solutions to needs and demands, and as any market there will be some that will gain access to them and some others will not. Information companies have new oceans of possibilities for innovation and product placement, and consumers will be amazed about the multiple new possibilities that will emerge in the coming years as the ones that they have been experiencing in the last decades. So, information does not only provide support to other economic sectors and public policy dimensions but also becomes a development potential itself. The remaining question will be how to address these new emerging markets, how new moral and ethical behavior will be demanded for fair competition and better market practices and how will these markets will be regulated to guarantee collective benefits.

The fact that information can be considered as a new valuable merchandise poses challenges to the social debate and demands collective agreements that will foster the incentives that allow innovation, entrepreneurship and production while developing mechanism of fair markets, reduce of inequalities and guarantee similar opportunities to most of the people, if not all.

c. Challenges for the non-information companies: Government and Public sector, Civil Society and Academia

For multiple years information, specifically statistical, was a monopoly of national states. This due to the cost and large resources demanded for information production. National Statistical Offices (NSO) where in charge of commanding large information operatives such as censuses or surveys and managing many administrative records to provide information for better decisions. Nowadays information production costs have decrease and many others have the power to produce relevant and valuable information challenging what has been called sovereignty of information by nation-states. The emerging question here is what the role of the public sphere on information production and diffusion in a decentralizing information world is.
The first answer to this complex question has been the move of NSO to a coordination role of the information ecosystems rather than the centralization one that had for some many years. The development of National Statistics Systems (NSS) has become a mechanism of broadening the scope and reach of data production and dissemination towards multiple actors in a so-called ecosystem. First by regulating the information production among many other governmental institutions, second by setting guidelines and standards for information production focusing in protecting data quality, validity and trust, and third by promoting strategies to acquire technical and human capacities to face new challenges of the information sectors. An additional task is the coordination role of the NSO within the complex information scenario particularly between the information private sector and the users: governments must take seriously the emerging challenges of regulating the relationships between supply and demands of information and propose possible ways of interaction between the members of the information ecosystems.

It is a moment of change and NSOs must conceal traditional statistical approaches with modern technologies and capabilities in order to develop new ways of interaction and propose new forms of information use guaranteeing quality and confidence in the imminent new national data ecosystems. Decentralization processes of information production and the incremental process of information diffusion and use potentializes the capacities of individuals and collectives that were formerly restricted to groups with priority access to information sources. As a result sustainable development will be promoted via the constant improvement of information access and guaranteeing quality standards of production while allowing data for openness and transparency.
2. Big Data: More than a simply buzz word

Many people think Big Data as a new technology or as a new tool. And this may be partial fault of the approach companies have been giving to their marketing strategies and their product placement. But Big Data rather than a technology is a process, some have call it a phenomenon, that emerges from the huge quantity of information constantly produce by contemporary society. This vast amount of information is produced, collected, use and disseminated by private companies, governmental institutions, social organizations, academia and individual. This process called Big Data is probable similar to the development of the information age that we have faced in the last decades but now the intensity and size has change modifying the attributes of the challenges, so Big Data at the end of the day is nothing more than Data, but a lot of it. And we are called to think on how to manage it.

Big Data has been defined as the rapid acceleration and the expanding volume and velocity of diverse types of data. Therefore, is usually defined by the use of three Vs: (1) volume, (2) velocity, and (3) variety. Later many other attributes (and other Vs) have been added to the definition but basically this elementary definition allows us to understand that the profound changes of Big Data are in the intensity attributes of data, which only resulted when some thresholds where reach by the immersion of new technological capabilities in the use of information by individuals and an exponential change was faced all over the information ecosystems.

The change generated by the development of the Big Data process can be explained as if suddenly all water dams use for water collection will begin to receive ten to twenty times more water that they can hold; we will not be able to manage water with the existing pipelines systems that we have, and a lot of water will be lost during the years of reorganization. Humans have been developing new technologies that allowed new innovations and decentralized the way we interact with technology, suddenly millions of people were producing new ideas and the amount of information surpasses our capacities and now it is necessary to reorganize the system to take advantage of the new resource and do not waste it.
Governance of Big Data ecosystems is one of the major challenges that society is currently facing as a response to the specific technological stage of the world and due to its important relation with sustainable development elements and processes. Above all developing countries must confront it to be ready and organize themselves for an adequate use of possibilities of change and improvements under the new information scenario. This is the reason why governance emerges as a central concept that will allow to reach desirable objectives. Thus, Governance can be understood as the form how society organize to make decisions; many people accept as true that governance can be assimilated by the concept of government, but it is broader and includes multiple and variegated actors that surpasses the public sector. In terms of Big Data ecosystems, governance will be the result of adequate interactions among relevant actors in charge of multiple dimensions of information and data management.

Before discussing primary elements and attributes required for Big Data ecosystems it is necessary to clearly identify components of the ecosystem. But, identifying ecosystems is not about listing the numerous actors that play a role on the information arenas but also trying to understand the incentives each one has that makes them operate and function. We can call this process as ecosystem mapping. Once ecosystems are mapped it is possible to start thinking on mediating and coordinating the emerging relations within the ecosystem. Some links may be broken by the lack or absence of some incentives while other links may be malfunctioning due to information asymmetries or power abuses of suppliers or users; in both cases some mediation is demanded to solve conflict and allowing better results of the decision-making process. Each country, city or community will have different ecosystems; therefore, governance needs to be adopted according to the interactions among existing actors and potential future ones that might be possible to identify previously.

Governance of information production, use and diffusion is a must for incremental benefits during the process of Big Data consolidation and future improvements. Making adequate decisions under scenarios of coordination, articulation of initiatives, cooperative work and flexibility is the desired scenario for taking
advantage of the manifold positive effects of Big Data ecosystems for society. National Statistics Offices have discussed in the last decades the necessity for consolidation of strong and durable National Statistical Systems (NSS) capable of surpassing the specific role of NSOs and allowing the interaction of multiple social actors. So, successful National Statistical Systems, in their more sophisticated versions, become good examples of information governance. NSS, and governance, require the emergence of standards and regulation for the adequate information production but also to promote interoperability and exchange of data within the variegated actors present at the ecosystem. Identifying the general uses of information and the role that each actor has will allow to propose and agreed behavioral rules and conditions for an adequate governance interaction among actors.

Thus, governance should not only be promoted at the public and governmental spheres but also at the individual and private spheres of the society. The question that arises is who might take the lead on Big Data governance, how should the leadership be assumed and what should be the major goals of any attempt for Big Data governance. In the following lines we will try to discuss these main questions and propose potential actions that will help to elaborate some ideas on the requirements and needs that an ideal governance demand.

a. The role of public investment and regulatory capability of governments

Governments have been in charge of leading public policy in multiple realms. Reorganization of society via democratic changes, expanding markets and individual agency has changed the role played by governments during the last decades. Although governments do not play the central and unique role in governance they do play a significant, relevant and functional one. Governments prioritize how a large portion of social resources are allocated, therefore decide the focus of many of the actions that society will undertake in the coming years. For instance, if a government embarks on an infrastructure package, will generate incentives for construction companies and financial institutions to actively participate in the process. So, allocation of public resources on strategic sectors will not only generate a direct impact but multiple indirect ones that will allow to transform economic dynamics and trends. The expenditure function of governments is crucial for the generation of incentives and play a key role in the development of Big Data ecosystems.

Additionally, governments still retain their power of deploying laws and legal frameworks under which a society perform. The legal apparatus is a powerful tool due to the capacity of modeling and stimulating
potential scenarios. While a law may motivate multiple actors to participate in a program or invest their resources in a certain type of initiative, a law may also ban certain behaviors or disincentive the allocation of resources in specific arenas. Contemporary legal governmental power does not focus on prohibiting rather than regulating the behavior of private sector organizations, individuals, academia and NPOs. The regulatory capacity of states allows the establishment of guidelines, roadmaps and strategies that will allow the consolidation of certain initiatives relevant for social change.

Information as a valuable good becomes an important asset for the private sector and its privatization diminishes the potential benefits for the public realms. It is necessary that novel forms of access to private information are proposed and developed. One, and probably the obvious and maybe naive, is the voluntary openness of private databases by firms in order to allow the better understanding of a phenomenon enabling multiple actors the proposal of solutions for an issue such as mobility, health or education. A second possibility is the collaboration and alliance between information owners and information analyzers that may be capable of proposing strategies for the use of information without the lost of property rights by the private companies or even without even letting the data leave their private fences but allowing the use of the intangible benefits of information use. A third one is the carefully thought regulatory framework, especially when dealing with public good such as telecommunications or health and education systems, that will demand private operators to adequately report information for the benefit of the majority. And many other solutions can be proposed and implemented in the coming years, the only request is institutional will.

In the case of Big Data environments’ development governments play a key role by allocating resources that will produce public goods such as infrastructure, initial major investments, human capacities and knowledge and innovation processes. On top of this incentives governments are required to design, collectively with other actors, an adequate legal framework that promotes investments, motivates the development of entrepreneurship and innovation while protecting individual and collective rights of the population. Although national governments are not the unique player, they still retain the role of defending the majority desires and have the power to act as a generator of public capacities within the country.

b. The role of the private sector

In many opportunities the private sector receives numerous responsibilities that are beyond its scope and realm. Sometimes the private sector is perceived as responsible for environmental issues or poverty reduction, but these are tasks that correspond to other part of social organization. Private companies have the responsibility to allocate resources
(at their own risk) in certain arenas that they consider might be profitable with the promise that will be receiving back the profit of their actions. This simple social agreement is the one that allows the emergence of competitive markets, efficiency allocation of resources and constant innovation in production, commerce and capital investments. The private sector demands and requires economic incentives to operate and react to these incentives allocating resources that will produce what is demanded.

So, the private sector has the important role of growth of Big Data ecosystem in multiple arenas of its development process. Private sector must provide products required for the adequate functioning of contemporary Big Data ecosystem that range from technological equipment to information intangible services. Moreover, a constant innovation dimension is required to continuously provide more and better goods and services, that can be arranged together with government and non-government institutions but that required the entrepreneurship and capacities of private sector to develop. The constant transformation of markets and emerging of new markets for technological and information goods and services demands the interaction of private incentives with society to adequately provide services that will increase the interaction within Big Data ecosystems. The role of private sector happens in decentralized interactions but allow the emergence of generalized benefits.

c. The relationship of Big Data and society

In terms of governance the third sphere of interaction is the one that emerges among individuals, non-governmental organizations and academia. Big Data ecosystems required a group of interested citizens that will constantly demand improvements and will be capable of paying for them. The focal engine for incentives is the collective social consumption of goods and services that emerge in the consolidation process of Big Data ecosystems. Individuals, organizations and academia require the generation of public goods and services that the private sector will not allocate on its own, but at the same time demand goods and services that the private sector allocates. So, society requires to actively participate in any process of Big Data governance either direct but also indirect in order to accomplish better decisions.

The interaction with individuals, organizations and academia might be complicated due to its variegated forms and multiple agents but there must be mechanisms of recognition and instruments for participation that will allow them to be part of the decision arenas. Governance based on the needs and demands of users that generates incentives for private enterprise and promote an adequate participation of government allow the emergence of positive dynamics of Big Data governance within contemporary societies. Understanding the role and promoting that each actor play its role adequately is the central challenge of successful governance.
4. Future challenges for Big Data environments

Governance is not a one solution equation. It requires the participation of multiple actors trying to include the majority of them if not all. Each actor has its own interests, capacities and agendas, and no one should be attempting to modify them but to put them into service of the collective benefit. Incentives are at the core of adequate governance and identifying them allows designing proper interactions that will be sustainable and will not demand artificial catalyzers. The public realm carries a big responsibility of reproducing public goods that are produce with the resources of all and benefit all (or at least attempt to do it), while private sector has to make decisions based on investments and returns that will provide services needed but constrain to the actual profit generation capacity of a market, therefore constraint. Social interaction will allow to identify multiple actors with a manifold of incentives providing new possibilities for governance, which at the end means new possibilities for decision making. Identifying goals, recognizing the social structure and deciding where to allocate resources and efforts will allow the development of specific governance mechanisms for each society attempting to promote Big Data ecosystems.