Data collection, processing and publication

Conference report

Conference organised by the Club of Regulators in cooperation with the OECD Network of Economic Regulators

University Paris-Dauphine, 4th November 2019

CLUB

Dauphine | PSL
CHAIRE GOUVERNANCE ET RÉGULATION

Conference organised in cooperation with the

OECD Network of Economic Regulators

Table of contents

Data collection, processing and publication

Conference of the Club of Regulators
4th November 2019

With the development of data-driven service provision and innovation, regulators are increasingly reliant on data to assess compliance or quality of service and supervise competitors' behaviours. Access and use of data raise multiple challenges. What are the best practices to access the needed data? How can regulators access third-party data? How can regulators process the data? With what outlook? What could be serve as relevant guidelines to publish them?

Co-organised with the OECD Network of Economic Regulators, this conference is a forum for sharing and discussing specific experiments in national and sectorial contexts, identifying best practices between regulators and operators, as well as documenting opportunities and risks for all various stakeholders.

1st roundtable: Data collection

Eric Brousseau
Scientific Director of the Club of Regulators and of the Governance and Regulation Chair, Dauphine-PSL

In today's data-intensive world, the concurrent development of Big Data – in all its abundance and diversity – and Machine Learning offer a wealth of opportunities for regulators to better understand the world for which they are responsible and enhance the quality of the frameworks which they produce.

However, the effectiveness of data is held back by issues such as difficulty of use, sensitivity, confidentiality, cost equations and integration into business models, all of which will be addressed by this session's speakers.

Introduction

Creating value with data: a holistic perspective

Henri Isaac
French Digital Council & Paris Dauphine University

Sociologist Bruno Latour has called for data – "les données", in French – to stop being referred to as such, as they are anything but "given". He was referring to the struggle which regulators can face in collecting data in their capacity under the law. When companies attempt to avail themselves of information by purchase, in keeping with their own paradigm, they too face multiple roadblocks.

Data come in many forms. They can result from a contract, a user, or a behaviour, often pre-existing any real-world transaction. When produced during transactions, they can be generated by users or by the technical artefacts inherent in the process. Indeed, beyond the data resulting from or related to human action, there exist machine-to-machine data, produced by artefacts when robots communicate with one another, unprompted by a human.

The world of online advertising operates by the millisecond, each an opportunity for billions of data to circulate, according to principles fully grasped perhaps only by Google and Facebook. Whereas the financial markets are under the supervision of a regulator, the advertising arena is not subject to any such surveillance.

Companies are increasingly separating those who work on their IT systems from those bearing responsibility for their data: the skills sets are no longer the same, being technical in the former case and governance-related in the latter. To attract qualified Chief Data Officers, responsible for overseeing data flows and instituting new data-related processes, companies will ultimately need to reorganise.

Data have long been used by companies, for purposes ranging from operational optimisation to cost reduction, provision of information to target populations, or new service offerings. Now, however, they are beginning to see data as a way to transform their business models and ultimately reshape their industries. With this new outlook, data accuracy becomes of crucial importance: while some level of error was tolerable so long as the aim was to diagnose the past, predictive and prescriptive aims demand flawless data, free of both error and bias. Realising that they do not have the skills to build the models they need, companies are also revamping their thinking.

To fully leverage the opportunities before them, they will need: first, to define their data collection strategy; and more importantly, effectively transform data into value. The online video industry offers two contrasting models in this respect, with YouTube far more reliant on data than a player such as Vimeo. While the technical decisions made in this process will be of definite importance, the success of companies embarking on such change will depend on their ability to redesign the processes around data, complete with dictionaries, semantical rules, data owners responsible for quality, and mechanisms for addressing the regulatory and governance-related requirements.

Effective strategies will be those based on an accurate understanding of the issues at stake, and the ability to allocate resources in such a way as to extract the most value possible in each given environment. Today's data organisation modes bear little resemblance to those of the past: filing systems and databases have long given way to data warehouses and, most recently, data lakes – vast repositories of high-quality data kept in their natural or raw format.

To produce value from data through machine learning, predictive models or AI, companies will need to bring together individuals from the IT, data and business worlds – and manage to have them work together effectively. Alongside this, they will need to carry out the painstaking and lengthy process of gathering accurate, complete and consistent data, in a world over-run by "dirty data".

Eric Brousseau

Thank you Henri. I can confirm that before being able to take on the issues of data science and governance, the practical hurdles in gaining access to clean data and compiling it in adequate quantities are already very high.

Initiatives of the Brazilian National Land Transportation Agency

Vanessa da Silva Santos
Regulation specialist, Management of international and market relations, Executive Superintendency, National Land Transport Agency, Brazil (ANTT)

Overview of remit

Brazil's National Land Transportation Agency (Agencia Nacional de Transportes Terrestres, ANTT) is an independent regulator, responsible for inter-state and international land transportation. In this capacity, it is engaged in both technical and economic analysis. It is in charge of renting out and supervising toll roads, supervises passenger transportation (charter, long-distance and semi-urban, i.e., home-school or home-office. The ANTT furthermore has responsibilities in railway regulation, both freight and passenger.

All forms of road freight transportation operating in Brazil must be registered with the ANTT, specifying the name of the transporter and all vehicles slated for use. Road freight accommodates 65% of goods transport in Brazil.

Interconnecting and informing

The authority's Strategic Plan on Data Collection has two objectives: to mitigate information asymmetry; and to improve the efficacy of regulations. For this purpose, it relies on both internal solutions and data exchange with external parties.

Passenger transportation players are subject to continuous regularity checks, and must comply with operational parameters set by the ANTT. Successful companies are granted either an authorisation (long-distance or charter services) or a permit (semi-urban). As pertains to the fleet, the ANTT requires data on the minimum and maximum fleet age, as well as vehicle registration in its systems. Drivers must not only be listed in the system, but also have passed a background check including criminal history.

One of the lessons learned from the Monitriip initiative was to give more frequent feedback to companies.

These companies are kept informed, through weekly notices confirming due receipt of the requested reporting data to the ANTT, or reminders of missing information. Only those companies having achieved Class 1 status can apply for further markets. There are plans to refine the criteria for classifying the companies, based on their specific features of each type of service (charter, semiurban and long-distance). Having access to the tariffs through Monitriip, the ANTT now has a more accurate idea of each company's revenue.

The SisHAB system is used primarily during the permit application review process, to verify the required data (driver license validity, corporate structure, company registration, etc.). It has helped significantly reduce the time required for verifications, as well as enabled a decrease in the workforce needed, through its connection to other entities' databases.

The Canal Verde Brasil initiative was created to address the severe lack of personnel in managing road freight and passenger transportation services: up to that point, only 500 staff were available for the entire country. 55 portals were built to store vehicle registration

information, contract validity, etc.for the inniciative helped reduce unscheduled deliveries to the Port of Santos, plan enforcement actions and inspections, decrease logistic costs and increase competitiveness.

The National Centre for Operations Supervision monitors roadways, providing information of visibility congestion levels, etc.,

Workstreams to come

Data exchange, while existent, can still be largely expanded, in particular between the regional services, but also between the ANTT and other national-level organisations such as the ONTL, which provides Integrated Information about transport sector and the GSI, focused on safety of the services provided.

Data collection and exchange are being conducted in response to the OECD's recommendations in favour of evidence-based decisions. The responsive regulation project is expected to be fully implemented by 2021.

Data-driven regulation of the transport sector in France: recent history and major developments

Anne Yvrande-Billon
Vice-President, French Transport Regulatory Authority (ART)

Building on a long-existent asset

As fashionable as the term "data-driven regulation" may sound, data have long been the raw material for all regulatory agencies, in infrastructures or beyond. It is the essence of economic independent regulation to be able to collect data and base its decisions on them.

Digitalisation should not be seen only as having an impact on business models. Instead, it must be recognised as an opportunity for regulators to enjoy increased capacity in the fields of data collection, storage and analysis. The infamous black box of monopolies, both figurative and actual, is now on the table and invites stakeholders to see their realities. This transparency can become a highly-effective form of incentive in the hands of the regulator, as well as an invaluable pool of raw material.

To effectively play their role, regulators must not only enjoy the legal power to collect data, but also that to sanction in the event of infringements – all backed up by the adequate human resources.

The ever-broadening scope of the ART

The ART first came into being as the Rail Regulatory Authority. It was empowered to collect data, but only on a one-off basis, for the purposes of a specified investigation. That being, as no infringement policy (especially no sanction committee) was in place, the said inquiries could hardly result in any disciplinary action. This sanctioning power would not be established until several years later, followed by new powers to collect data on a regular basis.

Today, the ART is entitled to act as a trusted third party, in charge of monitoring the relevance of the assumptions on railway undertakings' profitability used by the infrastructure manager to determine the level of track access charges. It can also serve as the dispute settlement authority for any conflicts arising between the various stakeholders of the rail sector., which operated for decades as monopolies and thus do not have the corroborated data sets which other international players might. Toward this end, it is involved in an ongoing process to maximise data quality and reliability.

As of 1 October 2019, the ART became responsible for regulating airport charges. From January 2020, it is set to become the regulator of the Paris rail infrastructure manager (RATP) and is expected to be responsible for controlling the compliance of transport service operators to their obligations to open their transport data and for monitoring the neutrality of the algorithms used to provide multimodal information.

Operating with and for stakeholders: successes and difficulties

Since 2015, the ART's key actions include the organisation of regular data collection campaigns. Run quarterly or annually, they cover all regulated entities, from infrastructure managers to railways undertakings, rail stations managers, coach transport service providers and motorway concessionaries. While this three-year effort has generated significant costs, they are largely offset by the benefits of access to dis-aggregated information. The Authority can thus now inform and monitor as to: volumes and types of traffic (demand and supply); quality of service provided (delays in minutes as well as number of passengers affected); cost; revenues, etc.

The relationship with the operators has proven a definite challenge: already, the ART has initiated two infringement proceedings against rail operators who had not complied with their obligation to provide the regulator with data and responded to two appeals against its decisions. Notwithstanding, it develops fruitful partnerships with academics, runs user surveys that concretely influence service provision decisions, and maintains communication through its reports and benchmarking studies.

Along with the French Energy Regulator, it organised a two-day Datathon structured around the issue of service quality. In that advantageous setting, the participants jointly came up with a tool aiming at helping users to select their route, based on information about past delays and reliability indicators. Other projects for end-users are ongoing to improve awareness and information, including on weather-related variables.

Moving forward with perspective

Data-driven regulation requires new skills sets, tools and ownership of technologies. It must be at the core of the organisation, and not merely a data processing mode like any other. As this form of regulation develops, the scope of the data explored could grow, to include for instance users' mobile phone GPS history to characterise common demand flows.

As these new opportunities emerge, however, it is essential that they be seen in perspective, as both appealing and potentially dangerous, and that each decision made be duly substantiated, in process and purpose. In highly-subsidised sectors where political interference is likely, such as that regulated by the ART, data-driven regulation is a means of heightening rationality. Some of the digital operators now coming on the scene are not involved in transport per se, but provide services that could directly benefit users and must thus be kept on regulators' screen.

Jen Ablitt

How do you verify the neutrality of algorithms? Do you take a strategic or rather technical approach?

Anne Yvrande-Billon

The law entrusting us with these powers has yet to be passed. In the meantime, I can say that our aim is to ensure that the algorithm is free of bias.

Data Collection: Horses for courses

Cristina Cifuentes
Commissioner, Australian Competition & Consumer Commission (ACCC)

Scope and principles

The Australian Competition & Consumer Commission (ACCC) and Australian Energy Regulator (AER) are independent entities operating under the same umbrella agency. The ACCC's scope includes such areas as telecommunications, data structure, energy and water. It has just completed its review of digital platforms, focusing on Google and Facebook, and examining the algorithms underpinning their operations.

The ACCC engages in data collection and analysis to ensure market efficiency and transparency. It does so in line with the principle of "horses for courses": choosing the best response to a given situation, and the best means to achieve a specific end. It operates consistently in compliance with the regimes on access, publication, use, etc., in each of the sectors with which it deals.

Operational principles and uses of data

Some data are collected primarily for monitoring purposes. This is the case with petrol, a topic of evidently significant interest, as each report published by the ACCC on the sector causes a spike in media resonance and public response.

The Commission can also use data to produce benchmarking studies of market participants: to help determine pricing, measure revenue levels or identify the most efficient entities as a means of guiding others. A panorama view of the broadband speeds offered across the market can also spur lagging players to improve their capacity.

User- or consumer-driven data in the form of complaints or reports are gathered by the ACCC to determine its enforcement priorities. From these various sources, it can also provide data that enable commercial negotiations to take place on a standard basis. The ACCC has used the open survey format to gather data from energy players and consumers about the most valuable services to develop.

The definition of data applied by the ACCC is much broader than that used by other entities: more than numerical indicators, it also includes behavioural information, financial trends, market analysis, technical capabilities, etc., all of which can also be used to incentivise players, beyond the regulatory needs to which they alert.

Understanding the stakes and responding accordingly

Data collection can be automatic, as in the case of price monitoring data, based on voluntary participation by players, or in some cases, required pursuant to the law. In the latter case, provision is expected in a specific format and non-compliance is handled as a criminal offence. These data are intended for specific purposes, stored in a secure, private web application, and subject to limitations on usage and retention time.

All data collected are subject to comprehensive cleaning procedures, and where possible, guaranteed by a senior executive of the originating organisation. Third-party or proprietary data cannot offer the same assurances, by definition. Considering the broad impact which the regulator's decisions have, they must be evidence-based and supported by robust analysis. When deemed beneficial to society, confidential information can even be disclosed.

Alongside the issues at stake in strategic analysis, the main challenges faced in data collection are its resource- and time-intensiveness.

Data collection and publication. Trends across regulatory authorities

Anna Pietikainen
Senior Policy Advisor, Regulatory Policy Division, Public Governance Directorate (OECD)

The OECD's Network of Economic Regulators (NER) brings together over 70 regulators across sectors, countries and regions of the world. This broad reach enables it to break down barriers and initiate discussion between representatives of different networks and industries on cross-cutting issues. Its central focus is that of governance and performance, and the link between the two. In effect, the NER touches on many of topics raised by speakers thus far: the capacity of regulators to process data; their powers to collect the said data under the law, and further, to enforce their regulations; the need for coordination mechanisms between economic regulators and other authorities, as well as with international organisations.

The Indicators on the Governance of Regulators survey consisted of 77 questions on 3 topics: independence, accountability and scope of action. It covered five sectors: energy, e-communication, water, air and rail. Regulators in 38 countries were addressed and 169 sector questionnaires were received. In total, 13 000 answers were processed by the OECD Secretariat.

- *What data do regulators collect?*

Most regulators across sectors do collect data on industry and market performance; only 9% do not. E-communications players tend to be the best-in-class. Air sector regulators tend to collect somewhat less data, possibly as a result of their institutional set-up as ministerial regulators in some countries.

Most regulators collect data on the economic performance of their sector, but less than on industry and market performance. As this information tends to be more macro in nature, pertaining to competition, market concentration, investment outcomes, etc., it may be considered less within the scope of the regulator.

Most regulators do collect data on their own financial performance. However, they are less likely to do so when it comes to their own organisational/corporate performance. The same trend applies to data on the quality of regulatory processes. The OECD would flag the latter as an area of importance for regulators: impact assessments, stakeholder engagement processes, and outcomes of regulatory activity in general.

Eric Brousseau

What does the expression "organisational and corporate performance" imply?

Jean-Yves Ollier

The response may not have been clear to the survey respondents.

A Participant (Alex)

The quality of regulatory process is clearly described as referring to the "tools and processes used to take regulatory decisions, such as impact assessment, stakeholder engagement, and ex-post evaluation". Examples are provided, including measurement of accuracy, timeliness, accessibility, risk analysis, etc.

Anna Pietikainen

- *What do regulators do with this data?*

Regrettably, information on regulators' use of data is less abundant. Broadly speaking, regulators draw on the data collected to correct market asymmetries, make correct decisions with regard to pricing, investment needs, etc., and provide transparent information to consumers. Data on market performance specifically would enable a foresight function: as economic regulators have their "finger on the pulse", they are very well placed to provide this type of advice on upcoming issues in emerging technologies and the digital transformation.

Most regulators furthermore publish reports on their activities to show the value of their agency to the external community, even when not required to do so by law. Energy sector regulators lead the way in this respect, followed by those responsible for e-communication and air.

Regulators also generally publish performance information and data online. When they choose not to do so, they most frequently cite reasons of confidentiality or lack of relevance of the information at their disposal. Indeed, many of those surveyed do not understand how their data could be useful and refrain from reporting.

Only half of the regulators surveyed present a report on their activities to Parliament. This is an area of opportunity in many jurisdictions, for a more structured and steady relationship between regulators and parliamentarians to ensure reporting on the regulator's activities.

Debate

From the floor

Is the quality of the financial regulations produced dependent to any extent on the accuracy and reliability of the data gathered? Is this possible connection explored by any studies, or substantiated by examples?

Annegret Groebel

I am not aware of any such studies, whether in the economic or financial field. We had to engage in extensive discussion before coming to an agreement on the type of financial documents required of reporting companies, as each document is drawn up but also reviewed with a distinct purpose. The conclusions which regulatory authorities draw from depreciation periods and even methods are very different from those of businesses.

Eric Brousseau

Financial regulators are also vested with a mission to protect the public and use this as a criterion in their review process.

From the floor

Mme Yvrande-Billon, the SNCF publishes a great deal of data on its open-access platform. Is this the result of regulatory incentives, and is this corpus large enough for your purposes?

Anne Yvrande-Billon

If only we had so much power! We collect some types of data, requesting specifically that they be disaggregated, and where possible, "raw", so that we can recalculate as needed. The SNCF reports its punctuality only down to the minute, such that a delay of 5 minutes 59 seconds will be recorded as only a 5-minute delay. Similarly, cost subsidies and revenues are not always portrayed neutrally.

Jean-Yves Ollier

What are the main challenges in measuring the quality of regulatory processes or internal reporting? How is quality translated into a metric that enables each regulator to report to the public and the Parliaments?

A Speaker

This matter is on tomorrow's agenda.

From the floor

Do you take an interest in new techniques offered by Google and Apple, such as homoeomorphic encryption?

Anne Yvrande-Billon

The question is technical, when we are always a bit behind the market and its players, due to an unfortunate lack of resources and over-abundant agenda.

Cristina Cifuentes

The use of data is not new. The difficulty today lies in the amount of data being released and the lack of disaggregation in it.

Anna Pietikainen

This can be referred to as the "pacing problem": the cycle of laws and policies is far slower than that of businesses as they reorganise. It is important that regulators close the gap by finding the right talent and being able to offer them compensation on par with that which they might earn with a private sector firm.

Eric Brousseau

Generally speaking, governments struggle to compete with Google and other large companies on the salary front. To rebalance these asymmetries, regulators could for instance choose to release data as a means of informing the public. Due to resource constraints, however, we are far from ready to create this data-intensive environment.

Cristina Cifuentes

The ACCC partners with private-sector organisations to collect data, but also to institute data standards on rights such as portability of data, analyse the data and prepare reports.

From the floor

To what extent are potential public-private partnerships today being held back by conflicts of interests?

Eric Brousseau

Collecting data today should be based on API and made mandatory.

From the floor

How do the data gathered for investigative purposes differ from those used in monitoring? How are they seen by policy makers ?

Anne Yvrande-Billon

Most data are collected on a regular basis within the framework of market monitoring. We are ex ante regulators, but also expected to prevent illicit behaviour.

Vanessa da Silva Santos

The lack of ex post evaluation of stakeholder engagement is regrettable.

2nd roundtable : Data processing and publication

Miguel Amaral
Senior economist, Regulatory Policy Division, Public Governance Directorate, OECD

One of the core objectives of data processing and publication can be to reduce the information asymmetries but also to improve the design and delivery of regulations. This raises different challenges for governments and regulators as the need for expertise and data privacy and security issues. This session will shed light on these different challenges, and in order to sketch them and initiate the discussion, let me give the floor to Jean-Yves Ollier, Member of the French Counsil of State.

Introduction

Data processing and publication

Jean-Yves Ollier
Member of the French Council of State & Coordinator of the Working Group on Regulatory tools within the French general consultation for the Regulation of Digital Economy

The previous round table offered a valuable summary of the issues common to processing and collecting data. Data processing and publication form the core of evidence-based and transparent economic regulation. Informational challenges are a critical element of market failures, which regulators contribute to answering by rebalancing informational asymmetries, measuring externalities, and understanding the preferences of captive consumers and users.

The wide variety of data types, from the purely quantitative to the behavioural, reflects the many types of potential users of this data. On one end of the spectrum are masses of wholesale market data, which must be sorted into researchable cases, using algorithms. On the other, tools such as retail market surveys and consumer polls produce more refined, qualitative data. Data processing, too, occurs in multiple forms, as illustrated by market surveillance data, ranging from the fully-automated to that sought by carefully-formulated surveys.

The resulting landscape is characterised by a complex combination of sometimes contradictory requirements: regulators work in the name of transparency and open data, while companies cling tight to their confidentiality and intellectual property.

Already in 1998, one of the first reports on public data in France explored the conflicting needs of the public sector for indefinitely available data and the individuals' rights to set retention times. Today, data management is an inescapable issue well beyond the technology sector, as virtually all businesses must maintain and, if possible, monetise information. For the regulator, this need raises fundamental questions about access to data, discrimination in the processing of data, interoperability and portability.

In some cases, data become the substance of regulation and a matter of inter-regulation.

- This was the case in the energy sector when the decision was made to phase out the regulated tariff available to corporate customers. Should competing players be given access to key data about those customers upon market opening? Were those data protected by any form of intellectual property rules, or did competition law command access to these data?
- The availability and reliability of financial data are becoming of increasing importance.
- Data relating to renewable energy raised an interesting regulatory challenge in France. Whereas data publication was considered a matter of general good governance by the regulator, reflecting the long-term commitments of the French State on renewables, their release may have triggered the re-negotiation of half of France's pledge (80 billion euros over a twenty-year period), in favour of offshore wind farms.

Each type of data can be connected back to a type of use, from market surveillance to tariff setting, market opening, and reporting (to the Parliaments and general public). Mirroring the challenge of mass data processing is the ability to define limited sets of relevant indicators in certain cases, especially incentive regulation.

The main challenges in adapting processes and resources include:

- the legal framework of the information collection and processing, where the regulator is caught between the duty to be transparent and open, and criminal liability in handling confidential information;
- IT security, particularly when it comes to trading data;
- the increasing role of algorithms in data processing; human resources, particularly internal resources for the regulators in the new technologies;
- the audit and qualification of data, potentially by an in-house department.

Evidence-based activity planning – why we do things, don't do other things and explain it all to the public

Jen Ablitt
UK Office of Rail and Roads, Deputy Director Safety Strategy and Policy, Head of the UK Delegation to the Channel Tunnel Safety Authority

The Office of Rail and Road, the market regulator in Britain, also holds powers relating to competition and consumer protection. This presentation will be based more particularly on its work in the field of health and safety. It seeks to achieve regulation that is independent from industry, risk-based and beneficial to its target populations. Toward this end, it is able to draw on an amount of data far larger than most of its counterparts in Europe. As such, its strives not to cope with a lack of data, but to effectively explain to the public how it uses data to inform decision-making and plan its regulatory activities.

A three-year project is underway, to improve our risk-based activity-planning. Our strategic planning involves a series of steps: identifying the data to be used; applying an analytical risk assessment method to determine safety risk priorities; publishing its conclusions as Strategic Risk Chapters on its website. The last step in this process is aimed at transparently informing the public about its understanding of risk, but also prompts industry players to adjust their own decision-making, as well as guides yearly work planning for the 80 health inspectors working at the Office. Further reporting takes place, both to the internal board and the external community.

The types of data released can be divided into two categories:

- "lagging", i.e. outcomes, such as the number of deaths, number of train accidents, etc.
- or "leading", i.e., near misses, precursors, the strength of risk management processes and culture.

The data further vary in degree of accessibility, ownership, frequency of release and coverage by sector. The Office welcomes any data useful to regulatory decision-making and planning.

Its analytical risk-ranking and risk-evaluation method is divided into five stages:

- simple risk assessment of the adverse event,
- study of possible outcomes,
- evaluation of degree of existing control,
- assessment of likelihood of improvement or, to the contrary, worsening,
- inventory of the powers at play, external pressures or public perception of the risk.

The process is based on a quasi-delphi method, with the assistance of experts, chosen for their balanced perspective and experience with data on the industry. The published Strategic Risks Chapters cover a range of topics, including a review of the safety capabilities, safety management, and operational or technical risk topics.

The aforementioned three-year project to develop this activity is divided into three annual plans:

- The first year was dedicated to a review of the business processes, using feedback from the team leaders and some early, simple improvements, as well as early proof of concept work on the use of artificial intelligence (text mining, inspection reports, data management) to solve difficult questions such as human performance, linking regulatory activity with industry safety performance.
- The second year is being used to develop the analytical method and risk-ranking procedure, as well as the business intelligence and planning processes.
- The third year will be angled at building on the proof of concept work to develop any intelligent systems or tools to support our risk assessment processes.

In order for the public to understand safety risk, regulators must be seen as a trustworthy competent and reliable public body. This requires the publication of high-quality information intelligible to the public. The ORR will therefore seek to grasp the public's understanding of safety, through focus groups, before reshaping its communication in a way that is meaningful and contextualised to respond to captures customer concerns.

From the floor

How do you measure the degree to which a safety culture has developed?

Jen Ablitt

We apply a risk management maturity model, which considers all aspects of a management system, using criteria from different levels, including culture. It can be found on our website.

Anne Yvrande-Billon

How do you coordinate with the other branches regulating the rail system, bridging for instance the economic and the health-related?

Jen Ablitt

We joined forces in 2006 and have developed a fruitful relationship since then. Although we share many corporate and support functions between our economic and safety regulatory activities, we have found that we require specific safety specialism for some skills such as data and statistical analysis.

From the floor

How do you shape public perception, when it is so multi-faceted and often refractory?

Jen Ablitt

Ultimately, the ability to regulate is based to some extent on the reputation of the regulator. Public opinion of the important issues thus does matter. Often, regulators need to be ready to explain why we are NOT doing more on topics of public concern – either because

we are not able to, best placed to, or because the evidence shows it is not of greatest risk to the public. For instance, crowding on trains is believed to be a safety concern, when in reality, it has little impact on safety (although there may be long term effects on well being).

Nevertheless, we need to consider also channels of communication to the public, and who we mean by "public". It is naïve to expect that even 1% of the travelling public will read any document produced by the regulator. It is nonetheless vital that the information be available. The most difficult concepts on which to communicate are risk and uncertainty, to which the public is generally quite averse.

Publication of data – Chances, challenges and limitations

Annegret Groebel
Director of International Relations at the German Regulatory Authority for Electricity, Gas, Telecommunications, Post and Railway (Bundesnetzagentur)

Multiple powers, conflicting injunctions

The Bundesnetzagentur is authorised by law to request information, as well as to enforce regulations and inflict sanctions. It is also subject to obligations in transferring data to other national bodies or European bodies, sometimes under set conditions.

Its four main areas of activity are:

- market regulation,
- market definition and analysis,
- marketing monitoring and statistics,
- quality and price comparison for consumers.

The BNA defines the format and degree of detail expected of the information reported by market participants.

Further to the Net Neutrality Regulation (of the Telecoms Single Market), NRAs are tasked with monitoring bits and bytes to ensure processing takes place in a non-discriminatory manner. In the event of suspicion or consumer alert, the BNA must also request information, for instance on traffic management, capacity and operations – all of which can be considered highly-sensitive from a commercial or strategic standpoint.

Thus, as economic regulators are expected to ensure a level playing field so that market participants can enjoy success with their business models, they must also process and publish the data in a way that does not distort competition, which sometimes may lead to tensions with the principle of transparency .

Forms of publication

Register-based publications use existing records to provide valuable reference materials to potential users. Compiled information on passive infrastructures, for instance, can help focus the site selection process for operators looking to roll out new fibre networks, as well as enable tremendous cost reductions. The BNR also collects the location of all solar panels installed in Germany; acc. to a court ruling those refusing to report will then be ineligible for special tariffs or subsidies.

The Energy Market Register, containing all the fundamental data from electricity and gas generators, is open to the public. Lastly, SMARD provides real-time energy market data on the authority's website, using new data processing capabilities to create transparency for the market and the general public.

Data from Agency Decisions form another source of potential information, provided they meet neutrality requirements. The Bundesnetzagentur's decision to publish the efficiency levels it had calculated in the efficiency benchmarking, was deemed by a court decision a violation of private players' commercial data and had to be ceased.

Similarly, the Agency is bound not to publish any data relating to the special tariff applications requests received of the dominant market player, as this could distort or impact competition on the market. Naturally, in the name of equality, the same rule must also apply to former monopoly-holders, as privileged as their past status may have been. In competitive markets, operators to refrain from disclosing their pricing strategies.

Obstacles and needs

Four main principles tend to limit the publication of data today:

- confidentiality, i.e. business and trade secrets;
- harm or danger to the operator;
- personal data protected under the GDPR.

Government agencies must be empowered to request the relevant information; today, they are required to seek permission for each type of data use planned. The continuous increasing of the amount of data collected will not replace market activity.

The question remains as to how to identify the data that will best illustrate and substantiate the work of the regulator, as well as create the certainty and trust to which it aspires.

Using modern data tools to unlock the history of pipelines in Canada

Sandy Lapointe
Executive Vice President , Integrated Energy Information and Analysis, Canada Energy Regulators

The Canada Energy Regulator (CER) regulates oil & gas pipelines and electrical power lines that cross a national, provincial or territorial border. It takes care to do so in a manner that protects the public and environment, while supporting efficient markets.

As a life-cycle regulator, it is responsible for :

- reviewing applications for new projects or upgrades to existing infrastructures,
- engaging in supervisory activities such as inspections and audits,
- determining maximum service pricing for companies,
- improving the import and export of natural gas and export of oil,
- providing energy statistic, analysis and information to the public,
- and engaging with the public and indigenous peoples in a meaningful manner.

This broad remit gives the CER access to a vast range of documents, reports and information, not only of a technical or commercial nature, but indeed including knowledge of the indigenous peoples, the environment, socio-economics, communities and economics.

Established in 1959, it has been involved in over 700 different hearings and related supervisory activities for the infrastructure it oversees. The vast majority of the resulting records have been received in pdf or paper format, resulting in 28 million pages of information, and 0.5 terabytes of transcripts and evidence. So long as it was unstructured, it could not be accessed, searched or analysed.

In addition to traditional data collection, the CER aims to be more transformative in its approach to data, with the aspiration to achieve more meaningful goals:

- using these decades of information to perform enhanced trend analysis on pipeline systems, through data analytics and visual analysis opportunities;
- making strong data-driven decisions, thanks to better more comprehensive data;
- enabling long-term efficiencies and enhanced global competitiveness, in particular through automation of processes and the ability to streamline findings;
- achieving better public information levels and enhanced engagement, to build stakeholder trust and regulator credibility;
- creating more accessible data, useable in ways which the people will value.

The CER must respond to external pressures as well. The 2014 Open Government Directive instituted the obligation that all Canadian government agencies make the vast majority of

their data available to Canadians in an open format.

To become a "more modern" regulator, capable of using advanced data techniques to find, share and use information, the Regulator thus made three key commitments, in a move toward step-level change:

- allocate the appropriate resources, within the limited funding available;
- bring in the necessary specialised skill sets, creating a special department internally as well as building partnerships with Universities and non-profit organisations;
- think in a new way, open to experimentation – as well as to some failure.

As of today, the Canada Energy Regulator has fifty on-going data projects and more planned for the future.

Unlocking environmental and socio-economic data: three examples of projects at the CER

Jim Fox
Vice President, Integrated Energy Information and Analysis, Canada Energy Regulators

Providing the keys to understanding environmental risk

Observing the divergences between public perception of risk and that considered significant by the authorities, the CER decided to use advanced data techniques to unlock environmental, socio-economic information that would help reshape public thinking. Though publicly accessible, its 80 volumes of documents – on such topics as soil, vegetation, water foliage, fish habitat, wildlife species – required expert understanding.

Technology was used to decompose the data from 2 424 environmental and socio-economic assessments, extract tables, pictures, geographic information, using an algorithm, and ultimately make it sortable and searchable. The resulting structured corpus will not only enable tracking for 20 to 30 years into the future, but also be useable by stakeholders, academics, environmentalists and other government departments. Furthermore, the code used to generate the content will be published and made adaptable by the public to produce its own useful datasets.

Canada's indigenous peoples

Over the course of the CER's history, it has accumulated more than 6 000 hours of hearings. The transcripts contain inestimable information not only to legal specialists, but in one area of particular importance to Canada: the nation's relationship with its indigenous peoples.

The unique insight and information about the culture, traditions and history of these groups, all encapsulated in these hearings, have been made accessible through a partnership with a University. They are now searchable by speaker, hearing type and date. In addition, Natural Language Processing (NLP) has been used to model the topics discussed within the transcript, yielding yet further layers of searchability. The result is a compendium of knowledge and wisdom over time.

Transforming information and data in an enhanced manner

For the last 25 years, Canada has published an annual Energy Futures Report, covering multiple aspects of the import, export and production of energy. It has proven of interest to economists in particular, but difficult to parse for the average Canadian.

Thanks to the application of new technologies, it has become a database that is easy and comfortable to search. A mass of data as complex as the regulatory conditions for securing authorisation to build or expand a pipeline is now tagged by project type, company and location, reflecting trends and offering answers on detailed points of interest to any reader.

Sandy Lapointe

While much remains to be accomplished, many opportunities also lie ahead. The CER is pleased to be able to look back on such accomplishments, as well as excited about its new areas of skill and partnerships.

Data processing and publication: what for? ERSAR perspective

Ana Barreto Albuquerque
Member of the Executive Board of the Water and Waste Services Regulation Authority (ERSAR)

With the Government's reform of the water sector and release of new public policy requirements in 1990, the expectations around data reached a new level of acuteness. From the national-level strategic planning then taking shape, to the arrival of an institutional framework including regulation, the need to establish a single definition of quality drinking water and the importance of user awareness and available information, all the lines seemed to converge into the need for strong information-driven regulation.

To fulfil its goals – consumer protection, operator protection, and environmental protection – ERSAR must collect and process data in multiple areas.

The ability to produce relevant information is a core asset. A complete information system has been established in Portugal for the water services, using reliable information, both to support the definition of public policies and business strategies and to evaluate the service that is actually provided to society, to be able to convey a reliable and regularly updated overview of the sector.

The regulator has the responsibility to collect regularly, validate, file and process information about the sector, following a standardised format, which is sufficiently complete and easy to interpret.

Dissemination of information takes place on two levels. The first is the national level, which is most useful for defining policies and development strategies. The second takes place at the level of the utility and is especially useful for the operation of the systems. The relevant information is available to users in a more accessible format.

The regulatory model has two areas: structural regulation of the sector and regulation of utilities' behaviour. All these competencies of the regulation model should be perfectly articulated so that a coherent model can be achieved.

Structural regulation must contribute to a better organisation of the sector. The regulatory authority cooperates in the development of the national strategies, proposes measures for resolving shortcomings, monitors and periodically reports on the evolution of the sector and the implementation of the strategies. At the same time, ERSAR proposes new legislation and changes to the existing legislation. It also adopts new regulations, binding for the operators. This kind of regulation implies direct control over the external context and indirect control over the operators, reducing or eliminating the possibility of undesired behaviour. It strongly influences the form, content and nature of the direct regulation of operators' behaviour, and for this reason, it is complementary to it. Complementary activities are also included in this area, such as the preparation and disclosure of sound information accessible to all stakeholders and the technical support to the operators through intense publishing and training activities, usually in cooperation with academic and research centres. ERSAR also has the duty to answer all queries on solid technical grounds in due time.

In 2004, the regulation of the quality of service made important steps through the implementation of a system of assessment of the performance of operators, which has evolved into an evaluation process of the quality of the water services provided to consumers. Targets for the quality of service were specified based on indicators relating to the user interface, the sustainability of utilities and environmental sustainability and have been continually monitored by the water services regulatory authority.

The regulatory authority compares the results of these evaluating mechanisms with the results of other operators acting in distinct geographical regions, thus producing an objective measure of benchmarking which reflects a concern with education and increased value. This system was upgraded in 2011 and 2017. The results of the evaluation process of the quality of service have public disclosure in the Annual Report on Water and Waste Services in Portugal (RASARP), pushing operators to increase efficiency. The information thus produced is nowadays the basis for investment decisions, policy analysis, national and European statistics about the Portuguese water services and for monitoring of the compliance with the human right to water and sanitation. Information produced by the regulator is feeding national reports in the framework of the Sustainable Development Goals, namely for «Goal 6: Ensure access to water and sanitation for all».

Like other stakeholders, the regulator contributes to the formulation of the Portuguese public policy for water services, towards its rationalisation and the resolution of any malfunctions regarding the regulated services and towards the organisation of the sector, promoting for example an increase in the efficiency and effectiveness of water services and the search for economies of scale, scope and process. It also monitors the strategies adopted for the sector throughout their implementation and regularly reports on their evolution and their limitations. The allocation of funding for investment in the water sector is based on the information provided by the quality of service indicators developed by the regulatory authority. It also contributes towards improving the transparency of these services, bringing more responsibility and scrutiny and creating a more demanding society, and positive pressure for improving the performance of the sectors.

This information on each operator is publicly available, on website, as well as via smartphone app.

The development of reliable indicators is more than a matter of data processing. It also requires organisational preparedness, with information systems capable of yielding answers, the ultimate aim being to grow closer to operators and better understand each group of players. One department at ERSAR deals with data processing alone. Quality of service and quality of water have been combined in the same department. The teams are formed and operate in pursuit of a set, over-arching aim.

From the floor

Do you publish results about the “worst-in-class” as well?

Ana Barreto Albuquerque

Yes, we do so in particular through our Claims Analysis reports. Spreadsheets are also maintained on each operator.

Anne Yvrande-Billon

What is the financial cost of this important work on facilitating access and visibility to users, considering the mass of raw pdf data?

Jim Fox

The scanning process is slow, but not as daunting as it may sound. We also rely heavily on partnerships. Through our relations with one University, we were able to have a senior undergraduate class take on the entire process, creating a model based on criteria we specified.

There also exists a nationwide system that matches up charities and government institutions in Canada with high-tech industry individuals wishing to take a sabbatical. For a relatively small amount, such organisations can have a highly-qualified individual work with them for 18 months.

Anne Yvrande-Billon

Do the benchmarking study results have any impact on the value of the players included ?

Ana Barreto Albuquerque

The players also try to improve as the results are released. As to financial value, no public companies are listed on the market. If they refuse to have their data published, we request an explanation.

From the floor

Have you developed Coverage indicators on water supply ? If so, who defines the related goals ? How is the operator-declared information audited ?

Ana Barreto Albuquerque

We publish a guide containing a definition of all the data requested, to minimise confusion. During the process we carry out risk analysis to decide which operators should be audited. Before publication of results an assessment report is sent to the operators for review.

Eric Brousseau

How can the information provided to stakeholders change behaviours ?

Jim Fox

We are still early in our process and will adjust based on our observations and players' responses.

Sandy Lapointe

Increased transparency has reduced the amount of complaints in several instances. Ever since companies have been required to place their emergency management manuals on their websites, the public seems to feel reassured and complaints have declined.

From the floor

How do you respond to players that refuse to share data on the grounds that it may not be of any quality ?

Jean-Yves Ollier

The response depends also on type of powers granted: some regulators may collect data, but not sanction for failure to report.

Ana Barreto Albuquerque

We were greatly assisted when the entity responsible for releasing European funds in Portugal began using our indicators to determine eligibility for grants and subsidies. There is far greater willingness to provide information than in the past.

Annegret Groebel

In the face of obvious unwillingness to provide data, we have had to develop our own cost-based models. Benchmarking data can also provide the necessary information. We have also had to issue very detailed requirements, when textbook accounting rules were already in use.

Miguel Amaral

Do you have any insight about how to help regulators at the national and international level ?

Annegret Groebel

I recommend experience-sharing about new technologies and possible avenues for data processing. Actual data sharing would be more difficult, due in part to confidentiality requirements.

Jean-Yves Ollier

Resources being very scarce, discussion can focus at least on methodology.

Jim Fox

Within our teams, the most junior staff have far more sophisticated data skills, as well as the perceptiveness and analysis that go along with them. They thus already offer an opportunity for valuable experience-sharing.

Chaire Gouvernance et Régulation
Fondation Paris-Dauphine
Place du Maréchal de Lattre de Tassigny - 75016 Paris (France)
http://chairgovreg.fondation-dauphine.fr