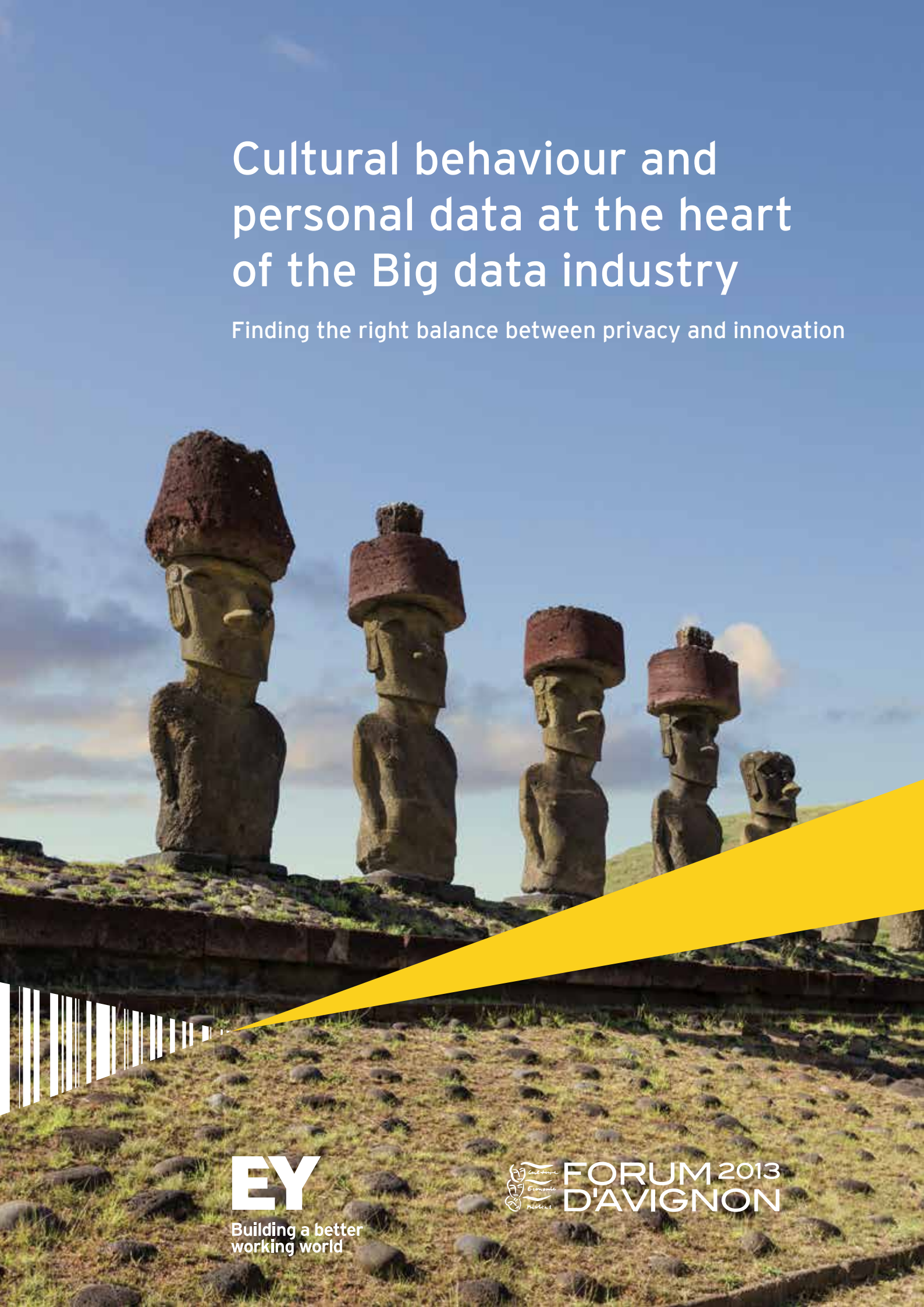
# Cultural behaviour and personal data at the heart of the Big data industry

## Finding the right balance between privacy and innovation



EY
Building a better working world

FORUM 2013 D'AVIGNON

# Contents


1. **Extracting value from Big data: a new power in the media and cultural industry** **p.7**
   - From Big data to Worthy data p.8
   - Personal cultural data at the heart of the emergence of a strategic market p.14
2. **Balanced power and shared responsibilities: the new contours of Big data** **p.19**
   - Regulation – an economic *impetus* for Big data players in Europe p.21
   - What will taxation 3.0 look like? p.33
   - Citizens' responsibilities p.37
3. **Towards a new economic paradigm conducive to innovation and creation** **p.39**
   - Opposing the precautionary principle to Big data: a real risk for innovation p.40
   - Time for action p.42
   - The new cultural data value chain p.47

**Personal cultural data at the heart of Big data** **p.50**
Infographics

**Copies of the study are available on the EY website** (ey.com/mediaentertainment) **and the Forum d'Avignon website** (forum-avignon.org)

# Foreword

## About this report and EY collaboration with the Forum d'Avignon

This report was prepared by EY together with the 2013 Forum d'Avignon.

The Forum promotes social cohesion and economic growth through culture and the creative industries. EY understands the importance of culture and its economic impact, and has supported the Forum d'Avignon since its inception.

EY's partnership with the Forum reflects our mutual commitment to the leaders of these industries.

Since 2008 EY explores the issues at the top of the Forum's agenda and used his recognised expertise and experience to translate these into meaningful and practical lessons for business.

Year after year, EY has observed and analysed the mega trends in the media and creative industries linked to the digital revolution that impacted the Intellectual Property *("Intellectual property in a digital world - The challenges and opportunities in media and entertainment")*, the monetisation *("Monetizing Digital media 2010 - Creating value consumers will buy")* or the tempo between creation and distribution *("Mastering tempo - Creating long-term value amidst accelerating demand")*.

The resulting ecosystem that progressively gained a new balance between the digital champions, the telecom operators and the media companies is significantly challenged by the revolution of Big data. This revolution could result in a limited number of actors able to extract the value of the cultural data and gain the power of control and predictibility.

Our 2013 study *"Cultural behaviour and personal data at the heart of the Big data industry"* examines the issue of finding the right balance between privacy and innovation and discovering. At the heart of this dizzying mass of data, an invaluable data is shining: the personal cultural data.

EY looked across markets and sub-sectors to understand, contextualise and interpret how a new market is emerging, backed by heavy investments. We mobilised media and entertainment practitioners across our network, as well as interviewed executives from some of the world's leading media and entertainment companies. Their insights, along with our market research and proprietary analysis helped shape this report.


## Acknowledgments

EY would like to thank the following experts for their straightforward commentary:

**Roei Amit** (Head of Digital and Multimedia Department, Réunion des musées nationaux - Grand Palais), **Christophe Benavent** (Professor, Head of Master of International Operational Marketing, Université Paris Ouest), **Julien Billot** (Deputy CEO in charge of the Media Segment, Solocal Group - formerly PagesJaunes), **Jérôme Dilouya** (Founder and CEO, Intercloud), **Denis Gaucher** (Executive director Ad Intelligence Europe, Kantar Media), **Alban de Nervaux** (Head of Development and Strategy, Réunion des musées nationaux - Grand Palais), **Alexandra Pelissero** (PR & Communications Manager, Criteo), **Stéphane Ramezi** (Head Publishing Web and Digital Medias, INA), **Benoît Tabaka** (Policy Manager, Google France), **Pierre Geslot** (Head of Digital Reading Projects, France Télécom Group), **Laure Kaltenbach** (Executive Director, Forum d'Avignon), **Olivier Le Guay** (in charge of Editorial Department, Forum d'Avignon).

# Editorial

Fingerprints, wall paintings, figurines, parchments, hieroglyphs and any other form of cultural or artistic production are physical traces left to posterity, whether consciously or not, that have enabled us to reconstruct entire periods of the history of human civilisations and thereby build our knowledge and our culture.

The whole evolution of humanity is indissolubly linked to the traces, fingerprints and other signs we leave behind us.
The advent of the digital era and the emergence of the non-physical traces we are making on the canvas, and which we can now collect, store and analyse *ad infinitum*[1], have given us unprecedented power. The royal prerogative, gradually usurped by the increasing availability of knowledge to the many *via* print, radio and television – and from which the reign of the Internet seems to have liberated us once and for all – is being succeeded by the power to control and predict, thanks to the fact that public and private data concerning us all are now available to a small handful of players who make it meaningful.

Volume, variety, velocity: these three words sum up Big data, expressing the unprecedented volume of data being produced and exchanged through an increasing number of channels (including the web, devices connected to the web and to each other, and platforms), the variety of these data (with a growing proportion of volatile, unstructured data[2]), and finally the ever-increasing velocity of these exchanges.

This all means that Big data is a formidable raw material for anyone who knows how to extract its very substance, offering the key to opportunities for value creation that could feed every sector of the real economy, and especially the media and cultural industry. For at the heart of this dizzying mass of data there shines a category of data of immeasurable value: personal cultural data. When we look at the drastic changes in power relations among internet operators, technological networks and media groups caused by the explosion of traces and information associated with the digital revolution, we cannot help but notice the fact that all these players are making a rush for personal cultural data, the new "open sesame", which should open the doors for them to intimate knowledge of each and every one of us. As both a mirror of our tastes and aspirations and a reflection of the social image we want to convey, personal cultural data are in effect a fragment of our identity. These data are all the more precious in that they create an unprecedented relationship by bringing together fingerprint and calculation[3]: if, like a photograph, digital data preserve the trace of our digital activity, they differ in that they are open to calculation.

1 The forecast capacity of the National Security Agency's new data centre for 2013 (yottaoctet) is a thousand times greater than all data recorded in the world in 2011, *Le Monde,* 12/26/2012

2 Geolocation and event-based data

3 *Faire mémoire des traces numériques*, Louise Merzeau, senior lecturer at the university of Paris Ouest Nanterre, member of the Tactic laboratory, June 2012

Unlike a simple photograph, which, as Roland Barthes noted, "is never distinct from its referent, from what it represents. It requires (the referent's) presence, they are inseparable[4]", digital personal data are detachable and calculable.

Indeed, the revolutionary dimension of the "Big data bang" lies in the autonomisation of the processes whereby huge amounts of data are being constantly and ever more rapidly produced and exchanged. In terms of cultural content, the supply is vast. But when they download, view and exchange cultural content and information, personal data tend to be scattered without control. Although the Internet now appears to be subject to standards and surveillance, there are still some serious questions to be asked about the power of control that permits intimate knowledge of consumers' cultural activities and personal cultural data. And there is also the issue of respect of these personal data and the need to develop concerted responses involving the players in the ecosystem of digital cultural content and alliances around shared values and imperatives[5]: *"Big data is ethically neutral, the use of Big data is not"*[6].

So, if Big data appears to mark a major break that will cause us once and for all to leave behind an era whose limits are being revealed by the exhaustion of resources, and enter a promising knowledge-based economy, then it is a matter of urgency that we learn to preserve the fragile resource of personal cultural data, whose survival depends upon delicate balances and shared responsibilities, which will prepare the ground for this new and rapidly developing market.

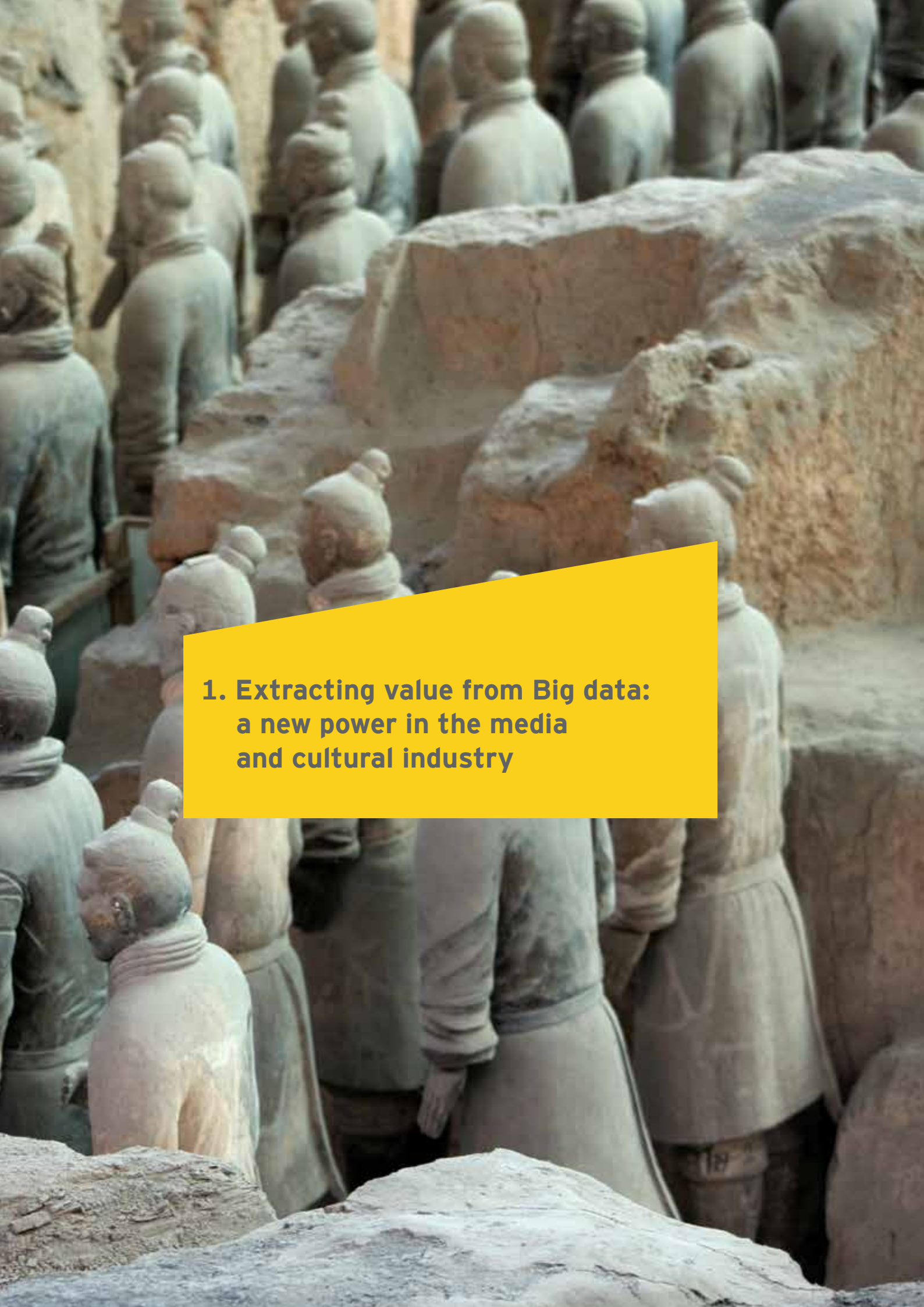
If they protect this ecosystem by creating trust, its players will be able to maintain control – control that also implies not submitting to the tyranny of data in attempts at profiling which would, if taken to the extreme, bind each individual to the results, thus leaving no room for serendipity and at the risk of falling under the domination of a single culture…

Fears that have, in their time, been associated with print, or with Latin during the Renaissance[7]…

In addition to the anticipated gains in efficiency and performance, by providing in-depth knowledge of the behaviour patterns of its public and users and therefore making it more able to anticipate their expectations, personal cultural data represent a formidable opportunity for the media and cultural industry to revolutionise its economic model and stimulate creation – *ad infinitum*.

**Bruno Perrin**
Partner EY
Telecom, Media, Technology
Market Segment Leader,
France

4 *La chambre claire,* Roland Barthes, Gallimard, 1980
5 Brand, reputation, trust, sustainability, transmission
6 *"Ethics of Big data – balancing risk and innovation"*, Kort Davis, September 2012
7 *Des pyramides du pouvoir aux réseaux de savoirs - Tome 1,* René Trégouët, senator, information report 331 - 1997/1998 *Commission des Finances* (Finance Committee)



# 1. Extracting value from Big data: a new power in the media and cultural industry

# From Big data to Worthy data

## From digital traces to the DNA of the digital self

### An explosion of digital traces

Over the past few years, digital data have grown exponentially and form a huge mass of data known as "Big data".

The increase in this mass of data can be attributed to three factors, known as the three Vs:

- The boom in the **volume** of data emitted: as increasingly nomadic individuals, we are producing more and more data, leaving behind us increasingly fresh digital traces in real time. At the same time, storage capacities are also increasing exponentially, which means a history can be kept of all these digital traces. In 2013, the volume of data generated or processed will exceed 4 zettaoctets[1], which is equivalent to a pile of DVDs reaching a height of 2 million Empire State buildings.
- The **variety** of connected devices is expanding: from laptop to refrigerator, from tablet to smart car to smart TV, the number of connected devices is exploding, as evidenced by the increasing demand for IP addresses. Capable of communicating with each other, these devices form the "Internet of things" and can trace their users and send precise information on their movements and habits, even when they are not being used. This variety also applies to the unstructured mass of data produced by these multiple sources, which emit data in just as many different codes, languages and formats.
- The **velocity** of information, that is, the speed at which data are processed simultaneously, is also increasing: these data, emitted by increasingly interconnected and interdependent sources, are travelling at ever-increasing speeds in networks that function less and less as silos.

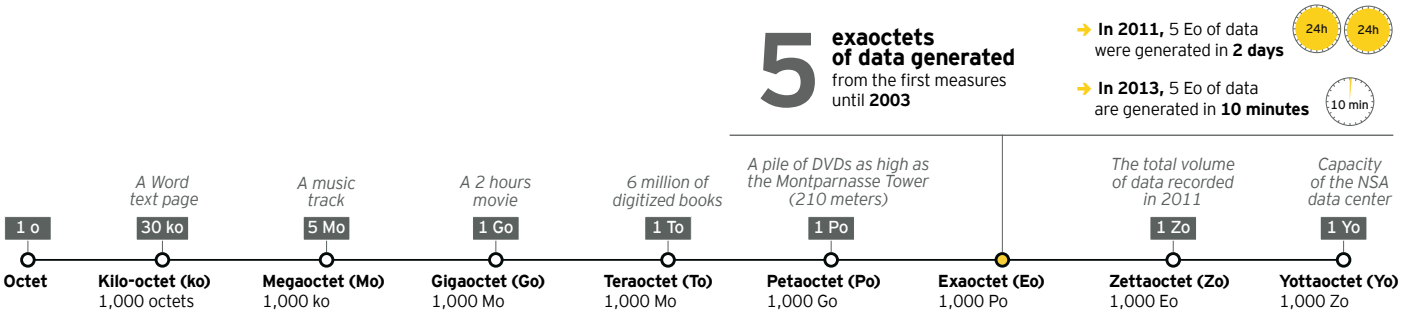
When users browse the internet, they leave traces of various kinds. These data are collected in accordance with various operating methods and are attached to users either *a priori* or *a posteriori*, depending on whether users identify themselves voluntarily or are identified through cross-check (*e.g.* cookies, IP address, MAC address).

In addition to the three Vs of Big data, there is the **value** that these data represent, both to businesses and to users. Use of these digital "traces" can in fact make browsing very easy for users and provide them with quality services (provision of a mailbox, obtaining free apps, free games, etc.), at no direct cost. Aware of the high value of these data, businesses seek to establish or manage a "give and take relationship", the aim being to reward users who transmit their personal data by offering them products or services to encourage them to reveal their consumer preferences and main interests through, for example, comments on social networks. However, offering additional services in return for the data surrendered implies a delicate balance for businesses: consumers, who are not always aware of the extent of this "give and take relationship", may prove reluctant to be seen as monetisable products. Hence the importance of making them aware of what they can do to control their personal data. ▸ See p. 37

Has this mass of digital data, like money, reached a sufficiently critical threshold to underpin a fully-fledged economic system? Is there a hierarchy of value among these digital traces? What is the value of this digital data?

1 *Big data, nouveaux défis, Revue de l'Association des Anciens de Telecom Paris (Newsletter of the Telecom Paris Alumni Association), July 2013*

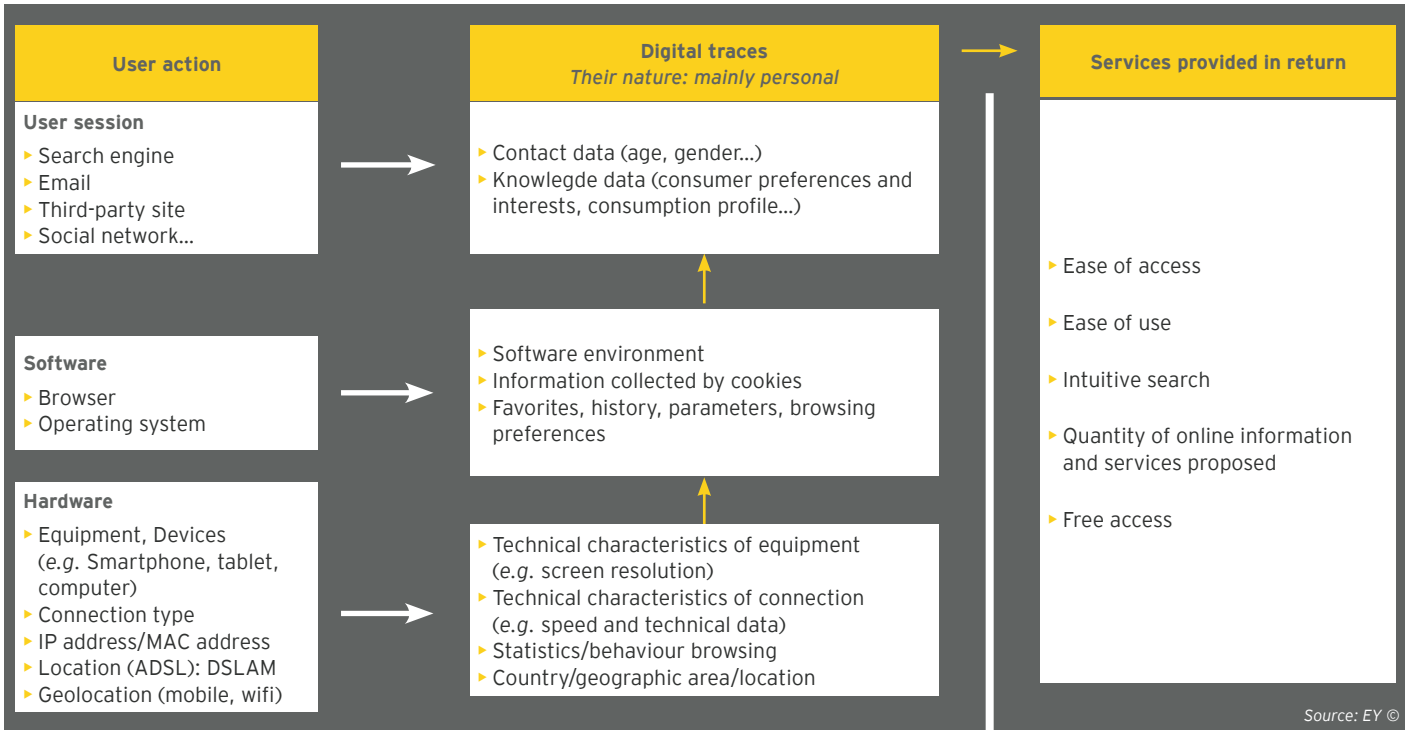
## From octet to yottaoctet, the data scale

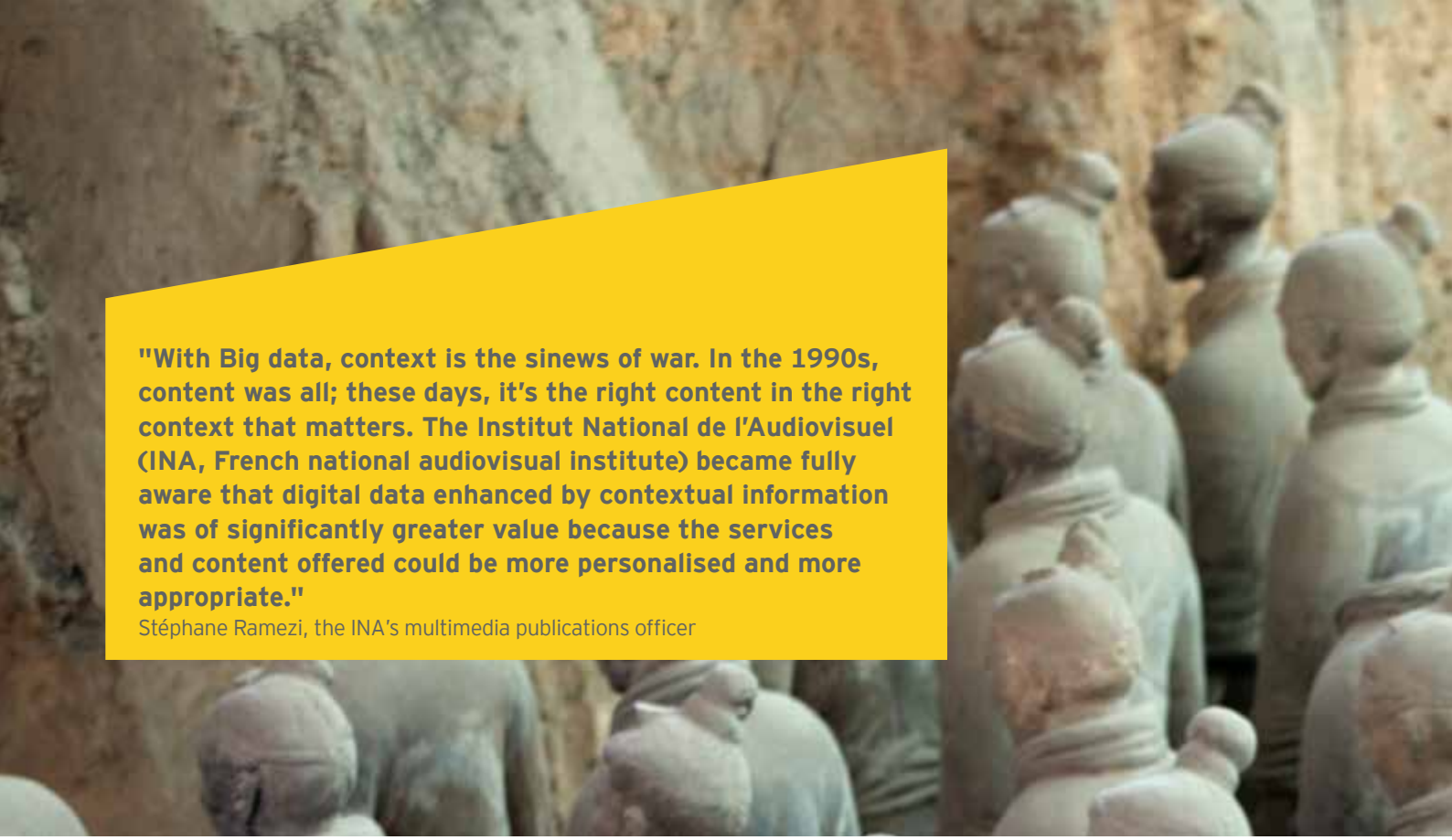



*Source: CNRS*

## What traces are left by a simple cultural search on the web?

*Example of an entry reservation for an exhibition at the Grand Palais*





**"With Big data, context is the sinews of war. In the 1990s, content was all; these days, it's the right content in the right context that matters. The Institut National de l'Audiovisuel (INA, French national audiovisual institute) became fully aware that digital data enhanced by contextual information was of significantly greater value because the services and content offered could be more personalised and more appropriate."**
Stéphane Ramezi, the INA's multimedia publications officer

## Give coherence to digital traces

Individual, heterogeneous, multiple and scattered – taken individually, the digital traces collected in real time have no value. Their value comes from the meaning drawn from them, in terms of correlation (cause and effect) or predictability.

It is possible to make sense of these data by attaching them to their common cause: the behaviour of a human being. In this way, one might hope not only to understand the behaviour of an individual from the traces he or she leaves behind but also then reconstruct the DNA of his or her "digital self".

There are times when the physical and digital images of an individual overlap. The simple purchase of a transport card is an example: a person who at six o'clock every evening, other than at the weekend, walks through the security gate of a Paris metro station with his or her pass will also at the same time be leaving traces of Internet searches through his or her mobile, leaving enough indicators for his or her habits to be deciphered: this is undoubtedly his or her journey after leaving work. But sometimes the physical and digital selves project different images. It is, for example, possible to misinterpret a piece of information transmitted on a social network, such as "Like" or "With": taken in isolation, such a piece of information will not only be too fragmentary to allow a behaviour or personality to be deciphered, but will also, above all, not necessarily correspond to the physical self's identity in terms of social level, education, etc.

In all cases, whether or not the image projected by the digital self matches the real image of the physical self, the digital image will be of real value only if it has coherence over time. If this is the case, then the traces left by the digital self may give statistical models a predictive quality, and thus generate value.

## Reconstructing the digital self's DNA through increasingly specialised algorithms

In addition to the collection and storage of data, algorithmic intelligence is a vital tool for giving meaning to the mass of data created by the "traces" left behind each Internet user. The purpose of this algorithmic intelligence is to collate and compare digital data from various sources to construct and characterise the digital self's DNA in order to decipher and predict that self's behaviour. It is, therefore, a question of contextualising the data which would, alone, have no meaning and which derives its entire value from being connected with a whole range of other data. It is this power of contextualisation that will enable a business to offer the right product to the right target at the right time through the right channel – in brief, to enhance, personalise and increase the value of its supply of content and services.

## Predicting the behaviour of a digital self

One of the aims of algorithmic intelligence is to establish cause-and-effect links to enable the digital self's behaviour to be more accurately predicted. At the same time, or in addition to the correlation, behaviour analysis models are used to seek predictability, and this is what the computer scientists and data analysts who draw up the algorithms are working on.

These predictive models are already being widely used. The recommendations made by online sites selling cultural goods and services, such as Netflix or Amazon, are based on models that can predict what an individual might like on the basis not only of his or her previous purchases, but also of similar purchases made by other consumers, so that appropriate products can be offered.

### The changes in computing techniques

Big data implies the processing of massive amounts of data (numerous historical sources, correlation bases, etc.) in a reasonable amount of time, even in real time. Very often, a combination of traditional statistical methods (descriptive statistics, segmentation, scoring, etc.) and computing solutions makes it possible to solve these difficulties.

For example, parallel calculations repeat the same calculations on separate data groups, or sequences, before compilation so that they can be completed more quickly overall. This calculation method is combined with statistical estimators to reach the most correct possible response in the allotted time.

It should be noted that the types of descriptive statistics being generated today are more pure than they were when we had only data samples that had to be extrapolated (because of the cost of collection, storage and processing). This is why the amount of data available and their processing are no longer a limitation, making it possible to work on more exhaustive data.

## The value of a digital trace: a balance between future cash-flow and the cost of collection/analysis

The deciphering of a digital self's DNA has a cost (collection of data, storage, deciphering information using high-performance algorithms).

The main challenge therefore lies in the high costs and investment represented by the infrastructures for processing massive amounts of data, and particularly in finding a means of monetising and increasing the value of the new analyses made possible by Big data. The key is to optimise the cash-flows generated by correct predictions of digital behaviour, whilst still controlling the cost of developing algorithms and purchasing, collecting and storing information.

Faced with a tidal wave of data and the rush for this new digital gold, we have to remember that not all of these data have the same value. This means that a scale of value needs to be established, with the value of data varying according to what they do or do not reveal about the digital self's DNA. Although, for example, an address or telephone number might have been of value in the era of mass telemarketing, they are now much less valuable than data on an individual's main interests or latest internet searches. This is why, in this tangled mass of all kinds of data, a promising nugget is beginning to shine out as a potential source of "worthy data": that nugget is personal cultural data.

# Personal cultural data: a golden resource

## The cultural behaviour patterns of *homo conexus* that give rise to personal cultural data

Over the past few years, the amount of time spent consuming cultural content has been constantly increasing – a phenomenon that is amplified by the fact that consumers are now *homo conexus*: connected, mobile and multitasking. This phenomenon is the source of the explosion of personal cultural traces left by *homo conexus,* together with the increase in data generated by the Internet of things, whether they are produced by:

- Digital cultural behaviour: search engines, transactions concerning cultural goods and services (tickets, books, music, etc.), accessing and sharing streamed videos, music and series, visiting social networks, media portals, etc.
- Or physical cultural behaviour: a trip to the cinema or an exhibition might leave traces via search engines, images and comments posted on social networks, or geolocation. As might a simple evening in front of the television: 52% of comments on Facebook currently concern programmes broadcast on television.

## High demand for digital personal cultural data

Far more than the consumption of goods and services, which satisfy basic primary needs, our consumption of cultural goods and services reveals a great deal about our preferences, our main interests and our aspirations. It concerns our very identity, not just in terms of its more intimate aspects, but also in terms of its more social, communal dimension: the cultural goods and services we consume, or our cultural behaviour patterns, have a large part to play in the image we want to project in our social environment. This is why we are happy to share our experience and opinion of a book, film, game, concert or exhibition with our online friends or communities, as well as with the editors of online content.

Personal cultural data embraces powerful contextual information, and makes it possible not only to ascertain fairly accurately the digital self's spending power but also to predict its behaviour. This is why, because of its intrinsically high value, it is now the centre of attention for all the Big data players.

There is already a demand on the market for personal cultural data, mostly led by groups such as Fnac and Amazon, which use these personal cultural data to package, adjust and target what they have to offer. The demand might be there, but what about the supply?

**Benchmark of data: the personal cultural data stands out from the crowd**

*Source: EY ©*

| | Public cultural data[1] | Transaction data[2] | Marketing data[3] | Personal cultural data[4] |
|---|---|---|---|---|
| | ★☆☆ | ★★☆ | ★★☆ | ★★★ |
| **Degree of digitizing** | Low | Medium | High | Medium |
| **Variety of sources** | High | Medium | Medium | High |
| **Variety of formats** | High | Low | Medium | High |
| **Volume of data** | Low | High | Medium | High |
| **Intimacy** | Low | Medium | Medium | High |

## An increasingly exhaustive output of data for greater understanding of the digital self's cultural behaviour

Although, on the supply side, personal cultural data has long been limited to the analysis of payments, businesses and establishments in the media and cultural industry are now extending the cultural experience on both sides of payment transactions. These players are aiming to enhance their supply of cultural services by offering a continuum of services.

This enables them to prolong their clients' cultural experience and form a committed relationship with them – a relationship that will provide precious client information, provided they are able to amass and effectively combine an unprecedented quantity and variety of information generated in different places and at different times.

**Cultural behaviour patterns, more than traditional socio-economic groups, give value to digital profiles.**

1 We consider a "public cultural data" a data produced or sourced by a cultural public institution with an effective cultural activity. *(Source: Camille Domange, rapporteur, "Guide Data Culture", n°2013-01, mars 2013, ministère de la Culture et de la Communication - secrétariat général)*
*E.g. inventory of an art collection, exhibition catalogs*

2 Data linked to financial transactions of cultural goods and services.
*E.g.* entrance in theatres, music sales, membership cards

3 Qualification and contact data collected by media and culture players.
*E.g.* suscribers to the "personal access" for the Louvre Museum, suscribers to a media newsletter, etc.

4 Data providing insights on the consumer cultural preferences.
*E.g.* discussions on social networks, forums, comments on artworks, exhibitions, artists, etc.

# Personal cultural data at the heart of the emergence of a strategic market

## A developing market backed by heavy investments

The high value of digital personal cultural data is leading to significant movement among the players, as well as heavy investment. A highly strategic market is in the process of emerging as the result of a dual phenomenon of cooperation and convergence, with:

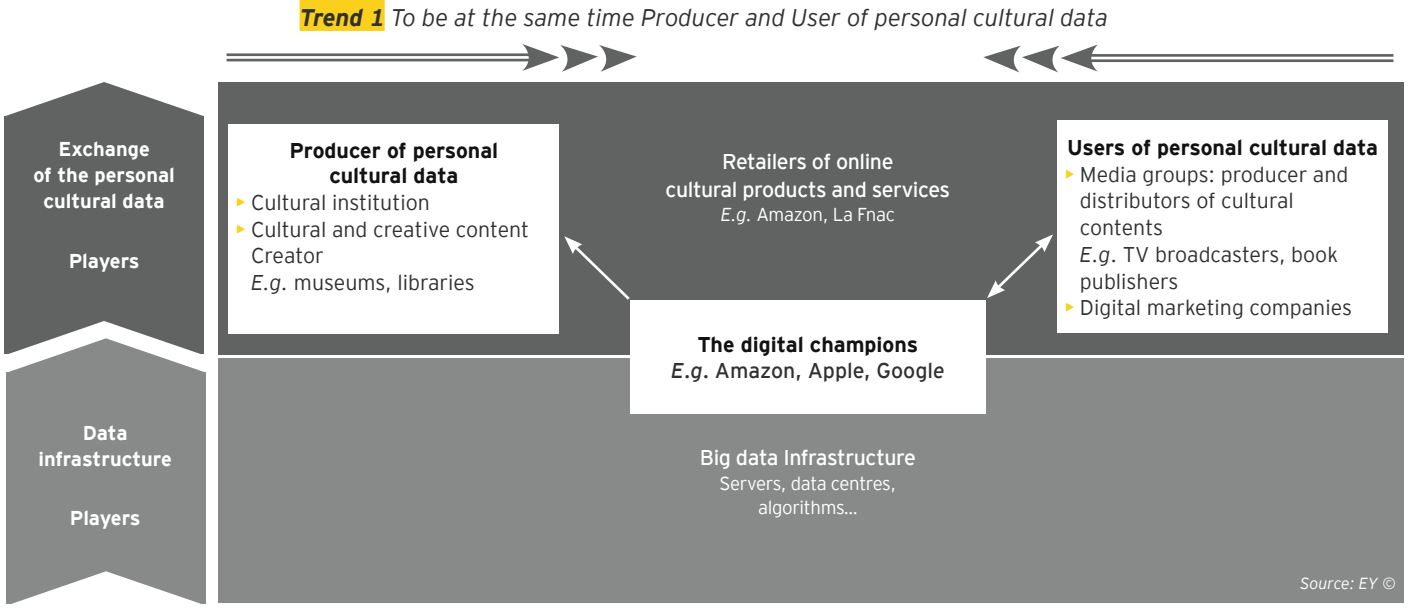
- On the one hand (Trend 1), media and cultural businesses and establishments which are innovating in order to offer a continuum of services to enhance and prolong the cultural experience by gathering precious information on preferences concerning the consumption of cultural goods and services.

*Example: under the partnership agreement signed by TF1 and Facebook in October 2013, the social network will share its data with the French audiovisual group, which will benefit from Facebook tools dedicated to tracking and analysing conversations about TV programmes. These tools will enable TF1 not only to measure the number of posts around a specific subject and analyse them, but also, while programmes are being broadcast, to post up and measure the number of conversations in real time on the social network, which brings together 26 million users in France and generates "52% of comments on television[1]".*

- On the other hand (Trend 2), digital leaders which already have data infrastructures and are users of personal cultural data are investing in becoming more involved in the production of personal cultural data and in producing and broadcasting cultural content themselves.

*Example: The Google Cultural Institute involves various projects, including the Google Art Project, which allows users to walk through the online galleries of 151 museums and cultural centres around the world on line (Versailles, Quai Branly, MoMa, The National Gallery in London, Nelson Mandela Foundation, etc.). This service relies on databases using various technologies, such as Google Maps, and collecting personal cultural data at international level, with a strong multiplying effect through social networks. The aim is to develop shared or compatible standards for museums throughout the world. This invitation to the world of culture is being complemented by the group's investment in videos (purchase of YouTube for 1.65 billion dollars in 2006), books and audiovisual content (films and series from the catalogues of the big American studios or television channels such as the BBC, etc.). Amazon is also investing in the cultural world. In August 2013, the group launched a fine art platform that sells more than 40,000 œuvres from 150 galleries around the world, with exceptional but also affordable artwork. The aim is clearly to reach a larger audience.*

### A new strategic market is structuring




1 *"Social TV : Facebook partagera ses données avec TF1 et Canal+", Le Monde,* 10/07/2013

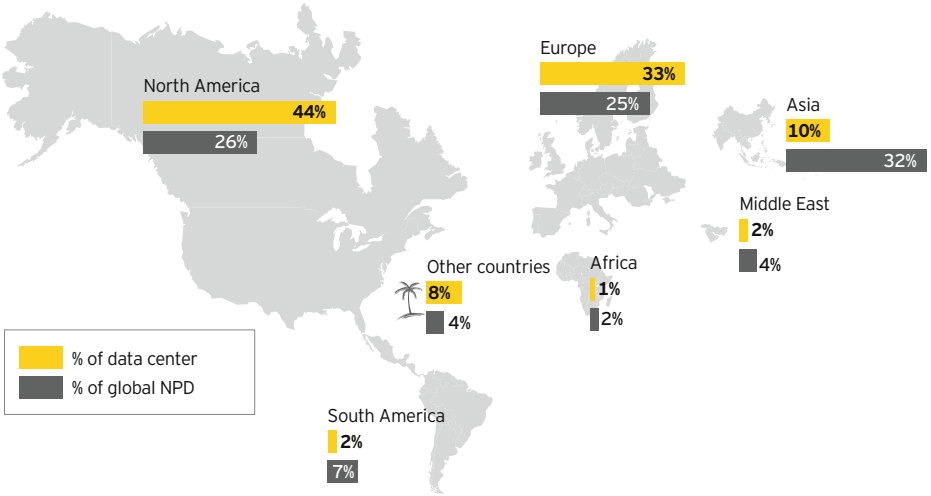
## Big data infrastructure: what are we talking about?

Information on the geographical location of data centres is very difficult to obtain and considered as strategic for the main players. Using the public information available for the data centres selling services, we have drawn up an anamorphosic map of data centres, which shows that the USA is leading the field in terms of data-storage capacity, by comparison with Europe and other parts of the world. Furthermore, the Big data collectors' storage capacities are much greater than required in the very short term, as illustrated by the fact that the National Security Agency's new data centre has a capacity of 1 Yottaoctet[2].

Furthermore, analysis of GAFA[3]'s costs and revenues shows that R&D and the storage and processing of data account for a significant proportion of costs, and that direct use of Big data accounts for only a small proportion of revenue[4], with the bulk of revenue coming from advertising or distribution.

### Anamorphosic map of data centres

*Source: EY ©*



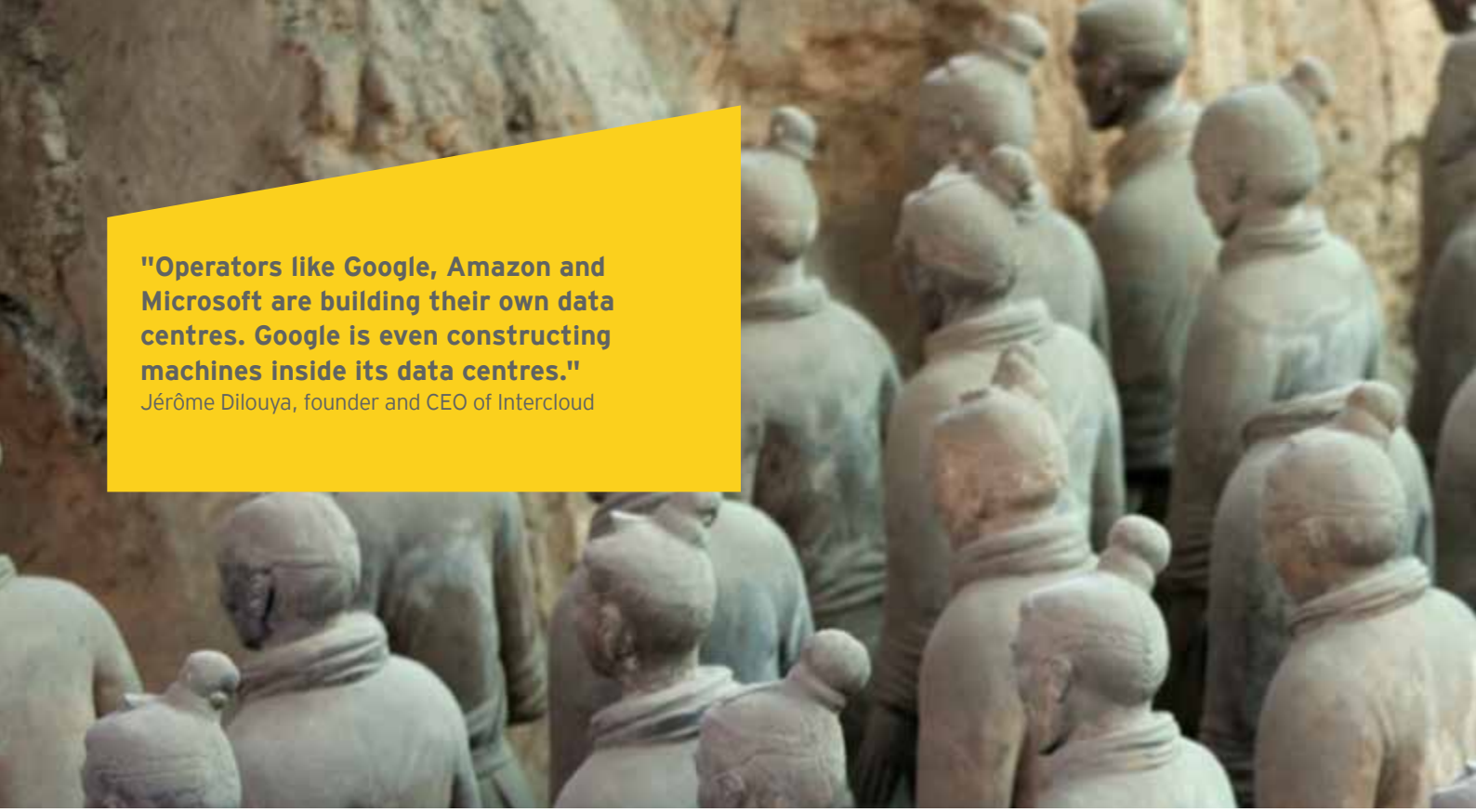

If we compare, by major geographical zone, the percentages of data centres[5] in the world to the percentages of global NPD, we can see that some regions are over-represented in terms of the number of data centres in relation to their economic weight (% of their NPD/global NPD). North America, for example, generated 26% of global NPD in 2012 and was home to 44% of data centres offering commercial services.

2 *"Vertigineux Big data"*, Sarah Belouezzane and Cécile Ducourtieux, *Le Monde,* 26 December 2012

3 GAFA is the acronym for the Big Four of the digital world: Google, Apple, Facebook, Amazon

4 *Direct use of Big data come from the sale of services for the storage, processing and analysis,* Jeffrey Kelly, Big Data Vendor Revenue and Market Forecast 2012-2017, October 2013

5 These are data centres offering commercial services. Sources: datacentremaps.com (visited on 09/27/13), CIA world factbook



**"Operators like Google, Amazon and Microsoft are building their own data centres. Google is even constructing machines inside its data centres."**
Jérôme Dilouya, founder and CEO of Intercloud

## Features of the personal cultural data market

If we are more clearly to define the Big data market and understand the rules of the game, it is useful to analyse the features of the main forces at play.

### Marginal costs decreasing gradually

Initial investments mainly concern the "Big data infrastructure": servers, data centres, algorithms. The massive collection of personal data, and the development of algorithms to make use of that data, require heavy initial investment to provide the service to the first client. However, this investment quickly pays for itself, because of the low cost of providing the service to further clients.

### A market in which only a few players will be able to survive in the medium term

Not every business has the means to invest in this infrastructure, whose quality will determine the quality of data use. This means that a new player's chances of success are far from certain, particularly since it is at risk of incurring significant "sunk costs".

Also, the player with the largest mass of information (social network, app, operating system, etc.) and most powerful algorithms will attract more clients and collect all the more additional reliable data. This will have the result of enhancing the quality of its analyses and algorithms and, therefore, strengthening its position on the market.

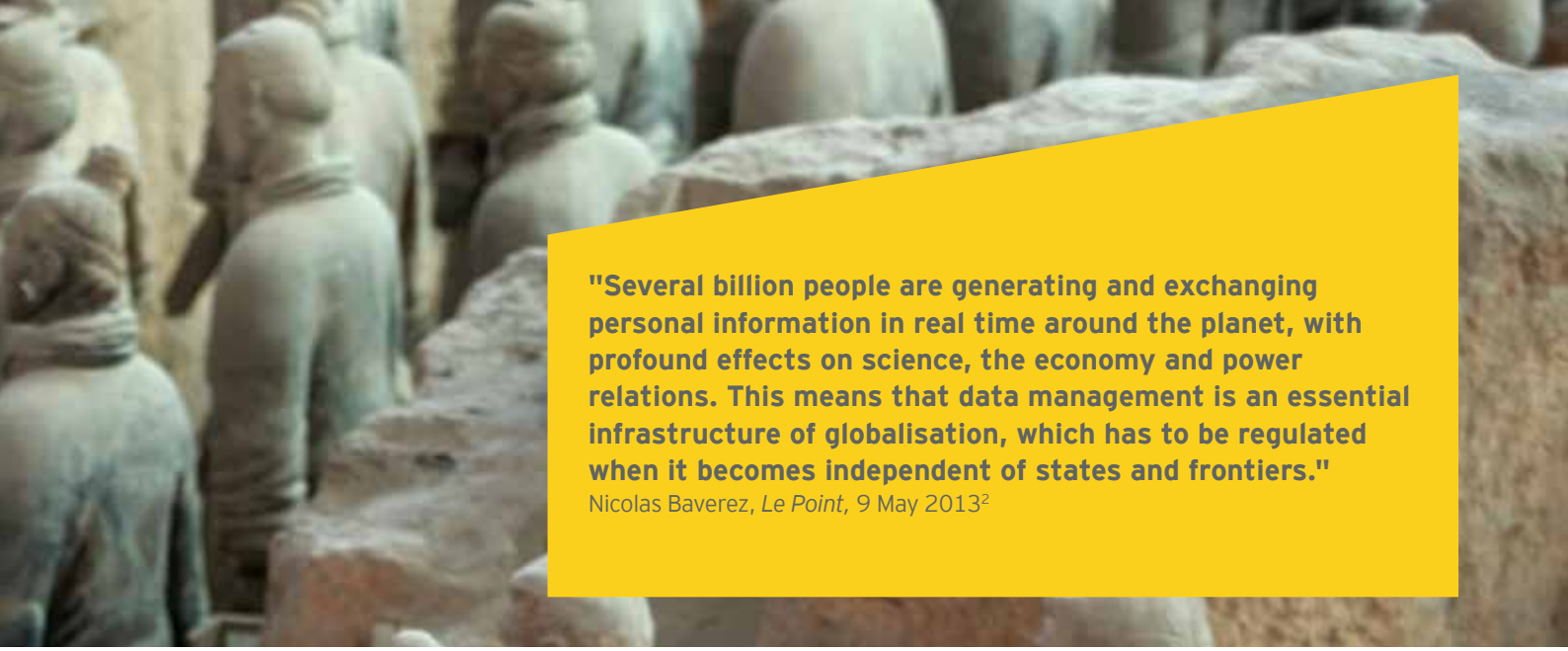
This snowball or winner-takes-all effect, added to the obstacle of initial cost, significant economies of scale and considerable "sunk costs", inevitably tends to concentrate the market in the medium term.

### A market linked to the "common good"

The personal cultural data market is sensitive because information on a population, and data about cultural behaviour patterns and their use, are often drawn from public statistics or surveys and processed/analysed at national level by public statistics institutes. We should also remember that cultural intellectual property falls into the public domain after a long period of use by the authors and producers of these data/contents[1].

**"This is the challenge we are up against: who is going to win the race for CRM of the web? (...) Whoever has acquired the critical mass of data will win the race for digital CRM."**
Roei Amit, assistant director responsible for digital affairs, Réunion des musées nationaux - Grand Palais (Rmn-GP)

1 *"La propriété intellectuelle à l'ère du numérique – Défis et opportunités pour le secteur Médias et Divertissement", EY, November 2011*



**"Several billion people are generating and exchanging personal information in real time around the planet, with profound effects on science, the economy and power relations. This means that data management is an essential infrastructure of globalisation, which has to be regulated when it becomes independent of states and frontiers."**
Nicolas Baverez, *Le Point*, 9 May 2013[2]

## Market characteristics similar to those of an essential infrastructure?

Significant initial investment followed by low marginal costs, high "sunk costs", significant economies of scale, "winner-takes-all" effect, a domain concerning the common good, traditionally managed by public bodies, etc.: these elements, which are characteristic of natural monopolies, are described at length in the economic literature.

The analogy between the Big data market and essential infrastructure raises the question of the upstream and downstream markets in which its players are involved.

Indeed, as regards the major trends that have marked some historical sectors, such as telecommunications and railways, the challenge is to identify the markets upstream and downstream of the natural monopoly and to encourage competition and a diversity of players on these markets.

From this perspective, do the infrastructure and exchange of data constitute a relevant market? Is this an essential infrastructure? Are there dominant players on this market that need to be regulated?

Although the market which is currently taking shape around personal cultural data seems to meet some of the criteria typical of a natural monopoly, it nonetheless remains a fragile ecosystem whose balance needs to be maintained, at the risk of seeing the resources promised by cultural data eventually running dry. The law seems to offer a framework suited to this ecosystem, providing regulatory instruments whose aim is to encourage competition and create an environment that is propitious to innovation. Furthermore, at the heart of this developing market, it is data on billions of internet users that are at stake, generating zettaoctets of digital data being collected and used by a relatively small number of economic players. Here too, with a concern for long-term equilibrium, regulation has a role to play in encouraging transparent, trusting relationships between businesses and individual users. It is by giving them a strengthened right to access and control their digital identity that these digital selves will continue to exist and produce data that create value.

### The characteristics of an essential infrastructure

**A network or an infrastructure presents the following characteristics**

- Technical characteristics
  - Technical indivisibility (lumpiness of investment)
  - Long life
  - Not widely traded
  - Exclusion could be technically difficult
- Economic characteristics
  - Subbadditive cost function (scale economies)
  - High sunk costs
  - Network externalities (winner-take-all effect, "snow ball effect" on the number of users, importance of a critical mass of customers, auto-realised anticipations)
  - War of the standards, standardization *de facto*)
- Socio-economic characteristics
  - Centralised planning and coordination
  - Traditionally public owned
  - Traditionally linked to the common good

2 Nicolas Baverez, « La révolution Big data » , *Le Point*, 9 May 2013

# 2. Balanced power and shared responsibilities: the new contours of Big data

**"If they are to create sustainable innovation that is not rejected by users, then businesses must offer guarantees as regards the protection of personal data. It isn't a cost, it's an investment."**
Isabelle Falque Pierrotin, *« Data, la nouvelle ruée vers l'or »* (Data – The New Gold Rush), *Enjeux Les Échos*, March 2013

The mass of digital traces available is now enough to allow us to track and model rational, coherent "digital selves". In the race to extract and analyse pertinent data, we can see an entire market taking shape: the digital personal cultural data market.

Although it offers some promising and extraordinary economic prospects, this new market nonetheless needs to be monitored as it emerges and develops.

Indeed, if we are to guarantee its harmonious development in the medium and long terms and encourage a new economic ecosystem that is innovative and creates lasting value, then we have to ensure that the massive exchange and use of personal cultural data respects the privacy of each individual (the first producer of personal cultural data), that the investment businesses make in collecting and organising data is protected and, finally, that markets are structured in such a way as to promote healthy economic competition.

# Regulation – an economic *impetus* for Big data players in Europe

## Protection of personal data as a matter of trust

### The significant need for trust in the management of personal data

Social networks often pick up on breaches of privacy or information security and set up a buzz around practices that are in breach of data protection law. This suggests that the protection of personal data is a central concern for consumers, who are increasingly questioning businesses and those responsible for processing their data about the guarantees and security they can provide in relation to the personal data entrusted to them. The questions most frequently asked concern the location of these data, the security measures set in place to protect them from unauthorised access, how they are being used by those responsible for processing them, who are their recipients, etc.

In an environment in which issues of confidentiality and privacy are attracting increasing interest from all the parties involved, more and more businesses are including this dimension in their communications with their clients, thus seeking to demonstrate the importance they attach to the protection of privacy.

In addition to the legal aspect, compliance with the regulations on personal data protection is becoming an effective way for businesses to convey their ethical and social commitment. Respect of good practices as regards the protection of data of a personal nature certainly represents a competitive advantage for businesses, in any industry, and also helps them to guard against any risk to their reputation.

Taking it as axiomatic that retaining a client requires considerably less investment than acquiring one, businesses use various strategies to maintain and foster client loyalty. This loyalty depends essentially on trust, which can be built only on the basis of the two fundamental notions of knowledge and recognition. The notion of knowledge reflects "the client's history", that is, his or her interactions with the business's various services, plus information on products or services purchased. The notion of recognition implies a personalised response to each client: meaning using knowledge to offer the client a targeted response, or even to anticipate this or her needs.

The constant technological advances that have characterised the past decade, together with the still too common discrepancy between undertakings, whether voluntary or imposed, as regards data protection and data-processing practices, are leading consumers to make stronger demands concerning trust and transparency.

## How to define data of a personal nature?

The notion of "data of a personal nature" is difficult to define because of its changeable nature – sometimes subjective, sometimes relative, but primarily contextual.

Directive 95/46/EC, which is the fundamental European text on the protection of personal data, defines "personal data" as "any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."

Other definitions are possible: "we should no longer be talking about personal data, but rather about relational and transactional data. We have to abandon the notion of personal data as a defined thing.[1]"

Indeed, data that, taken independently, might seem insignificant are collected and analysed to create personal profiles or to define digital identities based on actual personality traits.

Sometime only a date and place of birth can allow the identification of an individual. But use of social networks, search engines and messaging services can provide a clear picture of that individual as indicated by his or her tastes, habits, plans or beliefs. The risk is that all these data can be combined to identify an individual, without even knowing his or her name. The amount of data now being collected, processed and stored on each individual makes it possible to go much further and contributes to a very precise profiling of individuals and the possibility – depending on the type of data being processed – of producing probability models to learn more about their religious beliefs, political opinions, life style, sexual orientation, and many other aspects of their personal private life.

This raises the question of the need to redefine the types of data that should be considered sensitive and therefore be more strictly regulated by law. Data collected on the books an individual might be reading, the music he is listening to or the movies he is watching can, in fact, provide indications concerning political orientation, religious beliefs or even sexual orientation, and could therefore be subject to the law on the protection of sensitive data. We have no choice but to accept that, whether or not they meet the definition of sensitive data, cultural data should be protected because they relate to the personal life and intimacy of individuals.

This issue is all the more crucial in the context of the development of Big data and cloud computing which, by economy of scale, might lead to an increase in all the risks raised by the following questions:

- Where are personal data stored?
- Are personal data secured?
- Do individuals still have control over their data?
- How can individuals prevent the processing of their data?
- How can individuals recover their data?

The issue of trust becomes all the more fundamental when it is a question of building a lasting relationship with users and clients, because of the need to prevent doubts and fears that could obstruct the development of what seems likely to become a major lever of value creation in the media and cultural industry.

Together with these private initiatives, public players need to take the necessary action to guarantee adequate protection of personal data and support the growth of the industry, in line with the policies adopted in many countries to provide a framework for e-commerce.

1 *Vie Privée à l'Horizon 2020*, Dominique Boulier, Cahier IP n°/ 1, CNIL, November 2012

# Toolbox

## How are personal data collected and processed?

### EVALUATING RESPECT OF PRIVACY

Privacy Score gives a score to websites based on their awareness and respect of privacy.

www.privacyscore.com

### TRACKING THE TRACKERS

*The Guardian* has introduced an online application to enable internet users to understand how they are tracked online and by whom. The red circles are the ten most prolific tracking companies; the blue circles are the 100 most popular sites which use them.

www.theguardian.com

### WHAT IS PERSONAL DATA WORTH?

The *Financial Times* site enables you to calculate how much businesses charge for personal data on the basis of numerous criteria, including age, occupation, marital status, potential activities or illnesses.

www.ft.com

### CONTROLLING INFORMATION – CONFIGURING YOUR WEB BROWSER

All internet browsers allow you to block tracking cookies.

Some browsers, such as Mozilla Firefox, allow you to choose a specific option, which comprises informing websites that you do not want to be tracked by advertisers and other third parties. Respect of this parameter is optional, since websites are not required to respect internet users' wishes. It is also possible to set up additional plug-ins to monitor all the data gathered through your devices and the way in which they are processed, and potentially block any collection of data.

## USA/Europe: two different approaches to personal data protection

In terms of public policy, different approaches can be taken to tackle the question of privacy and personal data protection. The USA and Europe, for example, have radically different definitions of the concepts of "protection of privacy" and "data of a personal nature".

The first difference between these two approaches lies in the specific nature of the American system, in which federal laws coexist with the laws of each of the fifty states, which puts the USA at the top of the list of countries that have passed the most laws on the protection of privacy and personal data.

The second difference lies in the fact that the USA does not have a single body of regulations protecting privacy (with the exception of the Privacy Act of 1974, which gives citizens the right to know what information the federal government holds on them, and to demand correction or obtain compensation if their data are used in an unauthorised way). In the USA, protection of privacy is specific to particular sectors of activity, industries or other groupings. For children, for example, there is the Children's Online Privacy Protection Act (COPPA) of 1998, which protects the personal data of children from being collected or misappropriated on commercial internet sites. The financial sector is covered by the Financial Services Modernisation Act or "Gramm-Leach-Bliley Act" of 1999, which governs the conditions for providing third parties with personal information held by financial institutions; and there are laws to protect health data, such as the Health Insurance Portability and Accountability Act (HIPAA) of 1996, which establishes standards for the electronic exchange of medical information in order to protect patient privacy. There are numerous laws to protect data, but they are sector-specific rather than universal. For the media industry, for example, there are the Cable Communications Policy Act, the Telecommunications Act and the Videotape Privacy Protection Act.

Beyond the legislative dimension, the main difference between privacy law in the USA and the European Union is of a philosophical nature. While privacy law in the USA is based on consumer protection and aims to achieve a balance between privacy and effective business, the European Union perceives the respect of privacy as a fundamental citizen's right that is more important than any commercial interest. American law on the protection of privacy is intended to protect the consumer rather than the individual, and sanctions deceptive or misleading actions rather than failure to respect the rules protecting civil liberties.

Lastly, another significant difference is of course the particular emphasis placed by US legislation on the protection of data security, especially in the form of the obligation to notify breaches of security. Several US states have long-established laws obliging organisations to notify security breaches. So the great majority of American states have laws obliging businesses to implement specific procedures in the event of the theft or loss of personal data. These regulations oblige companies to enhance their internal security measures in order to avoid costly procedures that would also damage their brand image if breaches were to come to the public's attention.

In Europe, the obligation to notify security breaches is not yet part of substantive law on data protection under Directive 95/46/EC, though it does already exist for electronic communication providers under Directive 2002/58/EC on privacy. However, the draft regulation on data protection published by the European Commission, which will recast the European framework for data protection (see insert, "Forthcoming changes in Europe as regards the protection of personal data: enhanced rights for individuals, greater accountability of organisations, technologies that respect privacy, and more sanctions"), provides for a similar obligation that will apply to all those responsible for processing data and all subcontractors in Europe.

It is, in this respect, worth pointing out that the differences between these two approaches are bound to be reduced because the burgeoning rise in transatlantic data flows requires interoperability.

Indeed, just one month after the European Commission published its draft regulation on data protection on 25 January 2012, the White House published its bill on respect of consumers' privacy.

Comparison of these two proposed reforms perfectly illustrates the drive to harmonise the two approaches.

"The European Union and American legislation are beginning to use the same language as regards the law on data protection, both in terms of the proposed legal definition and in terms of the major principles involved[1]."

So, although the two proposed reforms come from different legal sources, they are converging, with the emergence now of some common denominators and, in particular, consideration of the importance of ensuring that operators behave responsibly (the concept of "accountability", which is to be introduced into European law) and of the need to provide adequate guarantees for people in order to control the threats to privacy deriving from technologies.

## US REGULATION

**Sources**

- No single body of regulation protecting privacy but a variety of laws specific to particular sectors of activity.
- Different laws in the 50 States regarding IT security and privacy protection issues.

**Philosophy/Objectives**

- Based on consumer protection and aims to achieve a balance privacy and effective business.
- Particular emphasis placed on data security, especially in the form of the obligation to notify security breaches imposed to all companies.

## EU REGULATION

**Sources**

- A single body of rules protecting privacy and personal data.
- Transposition of the directives in each of the 28 Member States.

**Philosophy/Objectives**

- Respect of privacy as a fundamental right more important than any commercial interest.
- The taking into account of the risks relating to data security and therefore IT security requirements is not fully harmonised, notably concerning the notification of security breach which is an obligation only for providers of electronic communications services.

1 Free translation of European Integration Realities and Perspectives: EU and US Data Protection Reforms. A Comparative View, Gabriela Zanfir

**Forthcoming changes in Europe as regards the protection of personal data: enhanced rights for individuals, greater accountability of organisations, technologies that respect privacy, and more sanctions**

**The draft regulation (proposed by the European Commission and subject to amendments by European Parliament) on the protection of personal data will enhance citizens' rights by introducing the following rules:**

- Creation of a "right to oblivion" to help citizens manage the risks concerning online data protection. When the person concerned no longer wants his or her data to be processed, and when there is no legitimate reason for processing to continue, data will be erased. The aim of these rules is to enhance the rights of individuals. It is not a question of suppressing past events or restricting freedom of the press.
- Creation of a right to transfer personal data from one service provider to another.
- Strengthening the principle of "consent", which, where required, must be given explicitly.
- Obligation on businesses and organisations to inform people in good time of breaches of security concerning their personal data which have the potential to cause them injury. They will also have to inform the competent data protection authority.
- Improving administrative and judicial review procedures in the event of violation of data protection rights.
- Increased responsibility and application of the principle of accountability ("the obligation to give account to the parties concerned") to those responsible for processing data – with data protection risk assessments, data protection officers, and application of the principles of "privacy by design" and "privacy by default".

Application of the principles of "privacy by design" and "privacy by default" will also enhance the effectiveness of European rules on data protection – which means that data protection guarantees will be embedded in products and services, and that parameters that comply with personal data protection will become the norm – for example on social networks. These rules will strengthen people's rights in a practical manner.

By introducing these stricter data protection rules, the European Commission is hoping to contribute to increasing trust in online services, so that citizens are able to use new technologies with more confidence and benefit from the advantages of the internal market.

**The other advances under the draft regulation are as follows:**

- A single body of rules on data protection, valid throughout the EU.
- A single contact: in principle, companies will have to deal with just one data protection authority at national level, that is, the competent authority in the Member State in which they have their main establishment.
- People will have the right to refer to their national data protection authority, even when their personal data are being processed outside their country of origin.
- The European Union rules also apply to companies not established in the European Union, if they are offering goods and services within the European Union or if they are monitoring citizens' online behaviour.
- Increased responsibilities and the introduction of the principle of accountability for those responsible for processing personal data.
- Administrative constraints, such as systematic notification requirements for businesses responsible for processing personal data, will be removed.

*Source: European Commission – the draft is still under discussion (joint decision-making process by European Parliament and Council)*

## The protection and development of personal data as a business asset

Personal data are now at the centre of online innovation and advertising and are "a type of asset for businesses"[1]. Expected to become one of the digital economy's driving forces, they are the focus of massive investment.

Paradoxically, the legal instruments available to businesses to protect and enhance the value of their data currently seem to be too limited to take adequate account of this strategic asset.

**Big data and protection of cultural personal data**

**The issue of personal data protection lies at the heart of the cultural Big data market, because personal cultural data reflect an individual's personality.** Cultural data are the "data of intimacy", and yet they are not specifically governed by substantive law (European directive of 1995) and are not mentioned in the draft regulation on the protection of personal data currently under discussion in Brussels (vote to take place in the first half of 2014, eventually postponed to 2015 following the European Council of October 24-25th). Also, and this is another fundamental issue, it is important for the law to recognise the asset value of these data.

The new European regulatory framework concerning personal data will certainly still be better and more ambitious than the regulations in force in the rest of the world. It could contribute to the development of Big data players in Europe, in that it will enhance the levels of security and trust that European players will be obliged to offer to those whose data they process.

### Existing legal instruments: the protection of databases

Databases are defined as collections of works, of data arranged systematically or methodically and individually accessible by electronic or other means.

Databases enjoy dual legal protection:

- Copyright protects the structure of the database if it can be deemed to be original in terms of organisation of the choice of sections and their layout. It is not intended to protect the information content of the database.
- The *sui generis* right of database producers allows for protection of the investment made in compiling the content of the database and could, therefore, to a certain extent make up for the limited nature of copyright. However, this protection is limited in scope in that it offers only the possibility of preventing the extraction or reuse of a quantitatively or qualitatively substantial proportion of the database's content.

These two systems protect different things (the originality of the "container", that is, the structure and presentation of the database vs. the investment in the content, that is, in the make-up of the database).

Yet, even taken together, they do not make it possible to cover in any real sense the full value of information, even when it is structured within a database:

- Copyright protects an empty shell – the structure and organisation of the database, not the information it contains (unless this information can be protected independently of its inclusion in the database – for example, titles of works protected by copyright).
- *Sui generis* right allows for the indirect protection of information, in that it grants the producer of the database the right to prevent its use by a third party, but only subject to the producer providing evidence of the (qualitatively or quantitatively) substantial nature of the reutilisation or retrieval of the database's content and, particularly, of the composition of the prior investment made to establish the database.

This means that the law governing databases does not allow for the protection of all types of information, with potentially strategic information not necessarily being eligible for the protection provided by the law on databases.

1 "Competition and personal data protection", Joaquin Almunia, Vice President of the European Commission responsible for Competition Policy, Privacy Platform event: Competition and Privacy in Markets of Data, Brussels, 26 November 2012, SPEECH/12/860

## Existing legal instruments: the protection of information through know-how

Know-how is a variable notion whose definition can differ significantly from one country to another.

In France, for example, there is no precise legal definition of the notion. This means it is merely a value which can be the subject of licensing or transfer agreements with third parties wanting to take advantage of the knowledge involved. As such, it cannot be directly protected by an industrial property title, which would be easy to protect. Know-how is an economic good that can be given a value as part of a company's assets. But it is not a legal good and cannot be considered in isolation as the object of an exclusive right.

Know-how, understood as technical knowledge that can be passed on but is not directly accessible to the public, is covered by the legal texts only in the very specific criminal case of trade secret violation. This notion, however, covers only secrets that can be used in the industrial sphere, in the form of methods, processes or materials (Cass. Crim. [French court of criminal appeal], 24 June 1985, n°83-92.873). It does not include commercial methods or commercial secrets (clients' files, for example).

The economic value deriving from investment in developing know-how can therefore be protected only under common law, in suits for unfair competition or parasitism, which are likely to raise two types of difficulty:

- The need to provide evidence of an offence, a prejudice and a causal link between the offence and prejudice (by contrast, violation of an intellectual property right requires only proof of the existence of the right and of the act of violation).
- The granting of damages and interest that very rarely cover in full the prejudices suffered by the holder of the know-how, despite the fact that, once the know-how has been divulged, it loses a large proportion of its value and will no longer benefit from any protection.

The weakness of the system to protect know-how or innovations not eligible for protection under copyright or industrial property right is an obstacle to developing the value of companies' intangible assets. However, TRIPS agreements, whose purpose is to define minimum regulations to protect intellectual property to be implemented by WTO members, are bringing the notions of "secret" and "know-how" together and cover the broader notion of "undisclosed information". This notion allows for the introduction of a general ban on the theft and use of confidential information or techniques and processes with a commercial value that are neither generally known nor readily accessible.

While the French concept of "secret" is very narrow and criminal proceedings are upheld only in the case of trade secrets, other countries have adopted a notion that is closer to the letter of TRIPS agreements[1], thus allowing for more extensive protection of companies' information assets. For example:

- Italy: protection of information concerning the company and its technical and industrial experience, including its commercial experience, if they are secret and have an economic value.
- USA: protection of financial, economic or commercial information.

1 Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS)

## New legal instruments to be developed to protect companies' assets

We need to conduct a general analysis of the legal protection of investment to develop the value of companies' information assets in order to draw up legal instruments that enable Big data players, including businesses in the cultural sector, to develop the value of their investment in Big data to gain a better knowledge of their clients and offer new services.

### A copyright approach to develop the value of assets deriving from personal data?

In addition to the issue of purely legal protection, various analyses have been conducted at government level concerning the current distortion as regards the actual and accounting values of personal data.

From the government viewpoint, this distortion is currently creating a problem as regards the taxing of GAFA, whose economic model is essentially based on making use of their users' data, and whose profits are not always correctly accounted for in the countries in which they provide their services.

Nicolas Collin and Pierre Colin, the authors of the report on digital taxation, recommend that personal data protection be modelled on the copyright system.

To recap, copyright has two distinct dimensions:

- Moral rights to authorise or prevent the distribution of a piece of work.
- Economic rights guaranteeing authors due remuneration for the commercial use of their works.

On this basis, the report recommends recognising the economic dimension of personal data by allowing those concerned to perceive the real value of the data concerning them as an exchange for the "free" services available to everyone on the internet. The moral rights component, for its part, would be modelled on the regulations governing personal data protection. This means that companies which were able to demonstrate respect of information law and liberties would be able to benefit from a reduced tax rate.

▸ See p. 37

# Competition law and market structures based on the exchange of digital personal data

## A "two-sided market", according to the European Commission

To date, from the point of view of the good functioning of markets, the point of convergence between personal data and competition law has rarely been discussed. In fact, the European Commission used the word "asset" to refer to personal data for the first time when Google and DoubleClick merged in 2008[1].

For the European Commission, players in the online advertising market, such as search engines, are operating in a "two-sided market"[2]. On one side of the market, the online service provider is offering free services (search engines, emails, content, etc.) to users. On the other side of the market, the online service provider is offering paid services to advertisers. In other words, on the first side of the market, users benefit from free services in exchange for the collection of their data and, on the other side of the market, these data are monetised for advertisers[3]. The profits of the new digital players do not come directly from the services they provide for users, but derive from their advertising or distribution activities.

▸ See p. 15

## A market to be analysed in accordance with competition law

One of the current major issues is that the vast majority of personal data is collected by a small handful of online service providers which, as leading innovators on their market, are able to gather huge amounts of data.

Legislation on personal data protection and competition law do not pursue the same ends, since data protection regulations are intended to protect privacy, while competition law is intended to guarantee effective competition on the market.

As a result, the possibility of counterbalancing the rising power of the main online service providers as regards the collection and use of personal data needs to be sought both in competition law and in data protection law.

Various practices could be qualified, under competition law, as anti-competitive agreements or abuse of a dominant position. Such practices may occur when personal data are acquired or when a company prevents others from acquiring those data. The fact that a company holds and reserves for its own exclusive use data that are deemed vital to other operators (in accordance with the definition of "essential infrastructures"), possibly citing intellectual property law and, particularly in this case, the specific law on databases, might also qualify as abuse. Lastly, as "assets", personal data can be taken into account in competition impact analyses of mergers of two or more companies, known as "merger control analyses".

1 *EU Commission, Case No COMP/M.4731 – Google/DoubleClick*, 11 March 2008

2 *EU Commission, Case No COMP/M.5727 – Microsoft/Yahoo! Search Business, 18 February 2010, §100*

3 "Personal data, will Competition Law collide with privacy? – Competition law and personal data: Preliminary thoughts on a complex issue", D. Geradin and M. Kuschewsky, Concurrences n°2-2013

## Anti-competitive practices in the acquisition of data: preventing other operators from acquiring data

Two examples of anti-competitive practices might be cited: firstly, exclusivity agreements signed between online service providers and editors, whose effect is to prevent other online service providers from gathering data and, secondly, preventing users from transferring their data from one online service provider to another.

### Exclusivity agreements

Online service providers can sign agreements with editors whereby the editor uses only the provider's services, such as a search engine, on its internet site[3]. Under these agreements, the editor cannot make use of the services of other online service providers. As a result, particularly when a multitude of exclusivity agreements are signed between editors of services and a single online service provider, other online service providers are prevented from gathering data from editors who, bound by their exclusivity agreements, cannot enter into other agreements[3].

These exclusivity agreements may also be signed between online service providers and software or hardware providers when the online service provider's services are bound by default to their software or hardware[4].

### Prevention of data portability

Prevention of data portability refers to the practice implemented by online service providers whereby they prevent users from transferring their personal data to another online service provider. Data portability is one of the new notions introduced by the draft regulation on general data protection (Article 18)[5].

Prevention of data portability could prevent advertisers from exporting data from an advertising campaign from one platform to another[3].

**Merger analysis**

In the case of the merger of Google and Doubleclick, the European Commission examined the effect of the greater quantity of personal information obtained by the ensuing entity.

In this case, the enquiry revealed that combination of the information on search behaviours and web browsing behaviour would not offer a competitive advantage in the advertising sector since it could be reproduced by other competitors with access to similar data on use of the web[1].

Although this issue was considered and examined in the analysis of the merger operation involving, in particular, Google, it did not lead to the conclusion in this case that the merger might have an anti-competitive effect.

4 These exclusivity agreements are likely to exclude competitors on the market, particularly when they are signed by companies in a dominant position, and even more so if a number of such agreements have been signed. "Personal data, will Competition Law collide with privacy? – "Competition law and personal data: Preliminary thoughts on a complex issue", D. Geradin and M. Kuschewsky, Concurrences n°2-2013, p. 37

5 Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), 2012/0011, 25/01/2012

## How to promote balance?

Where dominant players are identified on a market perceived as an "essential infrastructure", certain instruments are available to regulate the situation and have been tested extensively in other sectors, such as telecommunications.
For example, an authority can regulate the market *ex-ante* by requiring the publication of reference offers, fixing some tariffs (or fixing them in the medium term), encouraging the entry of players on niche markets, promoting data portability, or requiring the dominant players to separate some of their functions.

The regulator might also focus on users, requiring players to compete in terms of quality. To do this, the competition authority will have identified measures to enhance the use of personal data as a differentiating factor in terms of competition:

- The right to data portability might limit the risk of being blocked by promoting cumulative or alternative membership of several social networks, for example.
- The length of time for which data are stored must be reasonable and proportionate to the objective being pursued by whoever is processing them; limiting this period could reduce the barriers to entry.
- Transparency as regards the nature of data collected, the purpose of their processing, and the recipients of the information would provide users with the means of comparing offers on the basis of the criterion of personal data protection, giving them the power to control use of their personal data[1].

1 *"Données personnelles, le droit de la concurrence doit-il prendre en compte la protection de la vie privée? – Le point de vue de l'Autorité française de la concurrence", Bruno Lasserre, President of the French competition authority, Concurrences n°2-2013, p. 28*

# What will taxation 3.0 look like?

Beyond the challenges raised by the need to adapt international taxation to the realities of the digital economy, the subject of the taxation of data is now clearly on the table.

The questions that come up in this respect are many and complex:

- Should there be a direct link between taxation and data, and if so why? (economic justification, budgetary constraint, protection of competition, emergence of national champions, sharing of value among operators and creators).
- Do we need to introduce special taxation for the digital world in general and data in particular? If so, how do we tax data?
- Does taxation really have to be used as a means of providing checks and balances in the digital era? Or should it simply be adapted to the new realities?

## 2013 – an extraordinary *momentum* for taxation of the digital economy

Never have the efforts made by States to restore the equilibrium of their public finances been so intense, being reflected in general tax increases in most developed countries.

The digital economy has been particularly in the firing line, with States clearly demonstrating a wish to see tax receipts realigned with where revenue is generated, or even with the destination of digital goods and services.

At the same time, budget deficits are leading to cuts in public funding for culture in general, and for creative development in particular.

The feeling now is unanimous: international tax regulations have to be adapted to the new realities of the digital world and new technologies, particularly as regards the rules of territoriality governing corporation tax and indirect taxes.

Indeed, the combined effect of the systematic dematerialisation of goods and services and of bipolar business models[2] (based essentially on intangibles and information technologies) means there is often no longer any sufficiently clear nexus to allow the allocation of taxation rights to States in whose territory services are delivered or revenue is generated by ever more connected consumers[3].

2 Business model comprising a "free" activity backed by a paid activity, most often located in different territories

3 *Mission d'expertise sur la fiscalité de l'économie numérique* set up by Nicolas Collin and Pierre Colin (January 2013)

Furthermore, the differences in taxation systems (rates, bases, preferential tax regimes, etc.) from one country to another, sometimes within the same economic zone (*e.g.* European Union), lead to tax disparities between players in the digital economy depending on their size (multinationals versus SMEs/ETIs) and location (USA, Europe, emerging and rapidly growing countries). For States, these differences are also reflected in significant losses of receipts, whose effect is aggravated by increasing tax competition among these same States, and sometimes also by the optimised practices of globalised operators.

States have been reacting vigorously since 2012 and, in 2013, agreed an ambitious short-term action plan, including in particular:

- The OECD's 15-point BEPS Programme to combat erosion of the tax basis and profit shiftings (June 2013), which specifically includes consideration of the taxation challenges posed by the digital economy at the top of the list.
- Harmonization within the European Union of the VAT rules governing services provided electronically (new unified regime as of 1 January 2015, making tax payable in the State of consumption).

The international and political scope of this approach means that a certain amount of time will pass between the launching of these action plans, the reforms implementation and their actual impact on the real economy.

Similarly, harmonisation within the European Union of VAT rules continues to raise certain problems concerning application, and it has been impossible to plan for its entry into force any sooner than 2015, even though these problems have been evident for a long time. The issue of the digital economy was recently on the European Union Council's agenda at a meeting held on 24 and 25 October 2013.

These theoretical and practical difficulties can be explained mainly by the complex and specific nature of business models in the digital economy, which still largely remain beyond the grasp of traditional taxation concepts and instruments.

In this respect, an increasingly strong body of opinion is emerging among players and experts in the digital economy, contesting the need for special taxation for the digital world, since digital technologies are in fact now a part of every sector of the economy (industries, services, market and non-market sectors, private and public sectors) and therefore require a global approach, possibly under common law.

## A broad consensus for not taxing data is now emerging

Several public players have fought for the introduction of special taxation measures for the digital economy and/or more specifically aimed at the international internet giants.

The ends pursued by proposed taxation measures have been many and varied:

- Funding culture, for countries with a pro-active taxation policy on this matter.
- Restoring public finances.
- Balancing the effective tax rate for national and multinational players.
- Restoring the direct link between where tax is payable and where economic value is generated in the digital age.

France is without a doubt the most dynamic think tank on the matter, as revealed by the study of Politiques Fiscales dans le domaine de la Culture (Taxation policies in the cultural domain) conducted by EY for the Forum d'Avignon over the period 2009/2012, which identifies nearly 50 tax incentives and 15 taxes specific to the cultural domain.

The question of the taxation of data gave rise to a lively debate, which is now being echoed at international level.

After the rapidly abandoned concepts of tax based on advertising revenue generated on the internet[1], a bandwidth tax, "click" tax, or tax based on the manufacture and sale of mobile phones[2], it is the concept of a tax based on data, which was launched in France, that has been the most innovative and certainly the most controversial at international level.

Contemplated by Nicolas Collin and Pierre Colin in their report, this concept is intended to create specific taxes linked to the use of data generated by the regular, systematic tracking of users' activities in a given geographical territory.

Resting on recognition of the importance of the collection and use of data in digital operators' value chain and revenue, this concept, on paper, has the merit of being based on a simple taxable event and guaranteeing the neutrality of the levy.

More innovatively, this proposal was intended not so much to maximise the volume of new tax receipts but, highly ambitiously, to strengthen individual freedoms and competition by encouraging virtuous behaviours on the part of operators using data (by the application of reduced or progressive tax rates, depending upon the behaviours observed in relation to a predefined criterion of compliance[3]).

However, although intellectually appealing, this concept of the "virtuous" taxation of data was widely criticised at international level, particularly because, over and above the ideological debates, it posed extreme, unresolved difficulties as regards application and implementation and may also result in increasing the final cost for end customers.

It was therefore not adopted by the Conseil National du Numérique (French national digital council) (September 2013) in its opinion nº 2013-3, which gave preference to a strategy of political negotiation for an international corporation tax reform.

The subject might return to centre stage, particularly at European level, for example, on the basis of a justification for taxing exports of data from the European Union.

1 Tax on advertising revenue proposed by Senator Marini (France) in 2010

2 Tax on smart mobile phones recommended by the Mission Lescure Report in 2013 (France), but not adopted by the French government

3 For example: amount of information provided to users on their personal data protection rights, collection of consent and ease with which users can exercise their rights via the interface, openness to competition and new services, third party access to data, etc.

## What will taxation 3.0 look like for the economy of the digital age?

A recent EY study[1] noted that the problems and issues associated with the growing use of cloud services in the digital economy are still being widely ignored or sidelined, in particular:

- Inappropriate or complicated tax systems in many countries, no standardised rules.
- The definition of revenue generated via the cloud in terms of the rules governing withholding taxes.

This means that, more than ever before, in the globalised and highly technological context of the digital economy, we need to promote coordinated action at international level which will make it possible to adapt and harmonise tax regulations.

At local level, and without curbing the activities of national legislators, it seems clear that efforts have first to focus on, on the one hand, simplifying tax regimes that are currently far too complex and, on the other, improving the security of operators who are facing constant changes in legislation.

Without denying the real need to improve and harmonise tax regimes and restore States' public finances, digital operators, both large and small, are still waiting for clear, legible taxation policies that focus on the long term and make it possible to guarantee the development of tomorrow's champions.

Rather than creating new taxes or new rules, do we not need quite simply to apply the existing ones and, if necessary, content ourselves with adapting and improving them?

The OECD line of approach now should be, in the short term, to prioritise updating the fiscal definition of a permanent establishment, and reforming the international standards concerning transfer pricing.

In the medium term, States may probably want to revisit the principles of territoriality as regards corporation tax on digital services, given that States will logically be tempted, at national level, to prefer taxes to be applied at those services' destination.

All the challenge ofhe works now being undertaken at OECD level will therefore be to take into account the digital realities so as to create, at last, the taxation 3.0 for the new digital world, hopefully under a single coordinated framework so as to avoid multiple and different approaches of at national vs. international level.

1 *EY Global Survey Report "Tax Considerations in Cloud Computing" (March 2012)*

# Citizens' responsibilities

## Controlling one's own data

Increasing control of the data being transmitted by individuals or their connected devices means we need to be more aware, though without going so far as to give up all digital activity, which is possible for only a tiny proportion of the population.

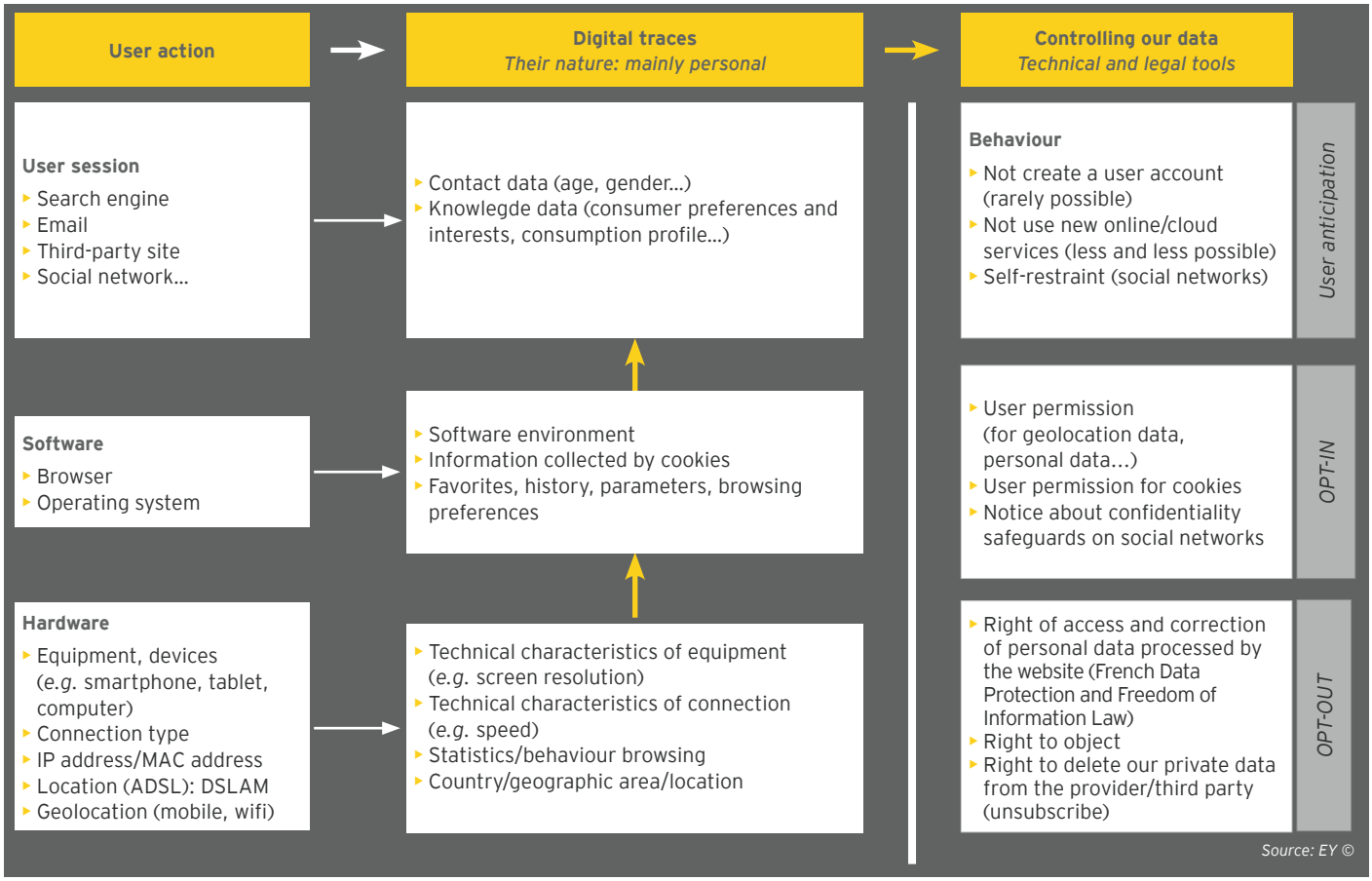
The graph below shows how users can control their digital traces and data, both *a priori* and *a posteriori*.

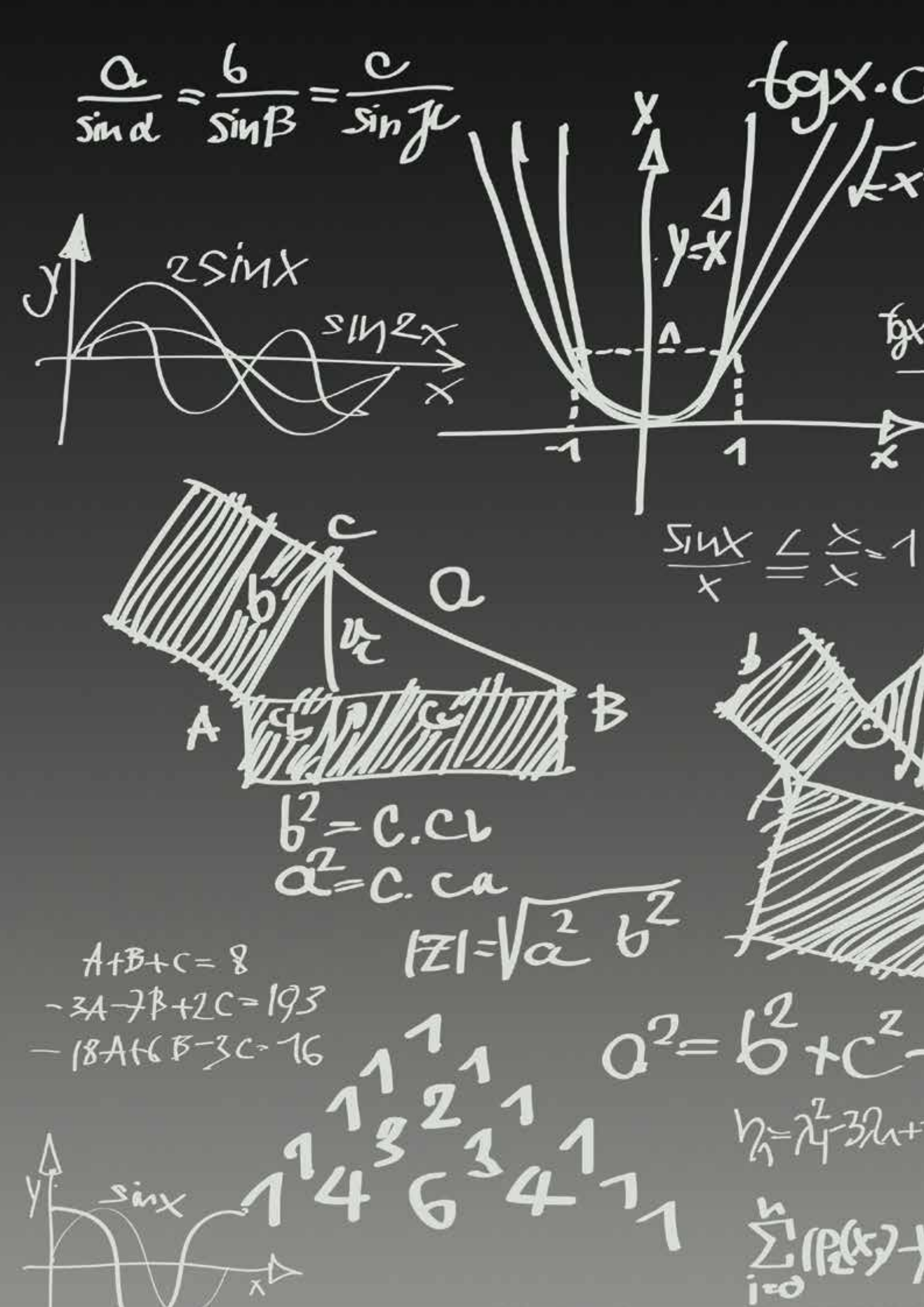
The European Commission's recommendations as regards the right to oblivion, strengthening the principles of consent, and improving administrative and judicial review procedures are moving in the direction of controlling control.

The recommendations and reinforcement of the notion of the "accountability" of "internet access/service operators" go hand in hand with trust and e-reputation – a key asset for all players on the market. ▸ See graph below

The combined effects of a keener awareness of the need to control their own data on the one hand and of the draft European regulation on the other, will give citizens more confidence and power of control again.

**How to control one's data *a priori* and *a posteriori*?**




a/sinα = b/sinβ = c/sinγ
y
2sinx
sin2x
x
tgx·c
x
y=x
-1
1
x
sinx/x ≤ x/x = 1
c
b
a
A
B
b² = c.cb
a² = c.ca
|Z| = √(a² b²)
A+B+C = 8
-3A-7B+2C = 193
-18A+6B-3C = 16
1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
a² = b² + c²
y
sinx
x

# 3. Towards a new economic paradigm conducive to innovation and creation

# Opposing the precautionary principle to Big data: a real risk for innovation

## Critical information mass and levels of analysis: is the economic exploitation of data really incompatible with respect of privacy?

### Thermodynamics

Thermodynamics, which enabled the development of industry in the nineteenth century with analyses based solely on "macro" indicators, seeks to explain energy and heat exchanges. It is rich in practical applications that we use every day: engines, refrigerators, turbines, reactors, etc. The models need very few magnitudes to describe the behaviour of systems and their development, mainly entropy, temperature, pressure, and volume.

The accuracy of these macroscopic predictions is quite astonishing when one realises that, at microscopic level, these behaviours are caused by the component particles of gases and liquids and that each particle is defined as a minimum by its speed vector (3 data) and its position (3 data), which involves billions of essential data.

The accuracy of the macroscopic analysis derives from the fact that a large number of microscopic states are possible for a single macroscopic state. Also, temperature is the result of the agitation of molecules, which might be found in billions of possible configurations for the same temperature measurement.

This means that thermodynamics teaches us that it is possible to make decisions on the basis of a measurement of a "macro" indicator such as temperature without necessarily having to measure all the magnitudes at "micro" level (positions and movements of billions of molecules giving rise to a certain temperature). For example, it is possible to pasteurise apple juice by maintaining a temperature of 75°. There is no need to model what is happening at the level of each molecule ("micro" level) to conduct this operation successfully.

**By analogy with Big data, it is possible to make decisions by analysing the "macro" behaviour patterns of a group of individuals without having to identify precisely the intimate behaviour of each member of the group.** Do we need to know everything about Mr. or Mrs. X to make predictions concerning a large population group? In other words, do we need to link all the information collected to a clearly identified individual (name, address, etc.) in order to offer that person the most appropriate products?

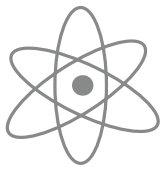
One might imagine that the magnitudes of measurement need to be "macro" and very precise to allow for the perfect adaptation of supply to demand. However, demand on the macroscopic level is the result of numerous individual (microscopic) demands which might remain anonymous. A return to the "micro" level for the act of purchase can be analysed in the same way: an advertising campaign with good messages sent to the appropriate population group could be more effective than individually targeted advertising.

## What sciences teaches us

### Thermodynamics

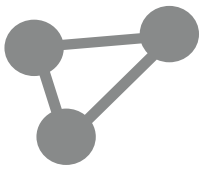
it is possible to make decisions on the basis of a measurement of a "macro" indicator such as temperature without necessarily having to measure all the magnitudes at "micro" level.

### Quantum physics



Physics teaches us that microscopic analysis, the domain of twentieth-century quantum physics, is random in nature and that infinite precision of measurement is pointless.

### Genomics



If we respect the confidentiality of individual decodings that, when combined, make it possible to make discoveries and decisions, then a balance between progress and respect of privacy is achieved.

**Application to Big data**

Big data and respect of privacy are reconcilable in a regulated and secured ecosystem.

*Source: EY ©*

### Quantum physics

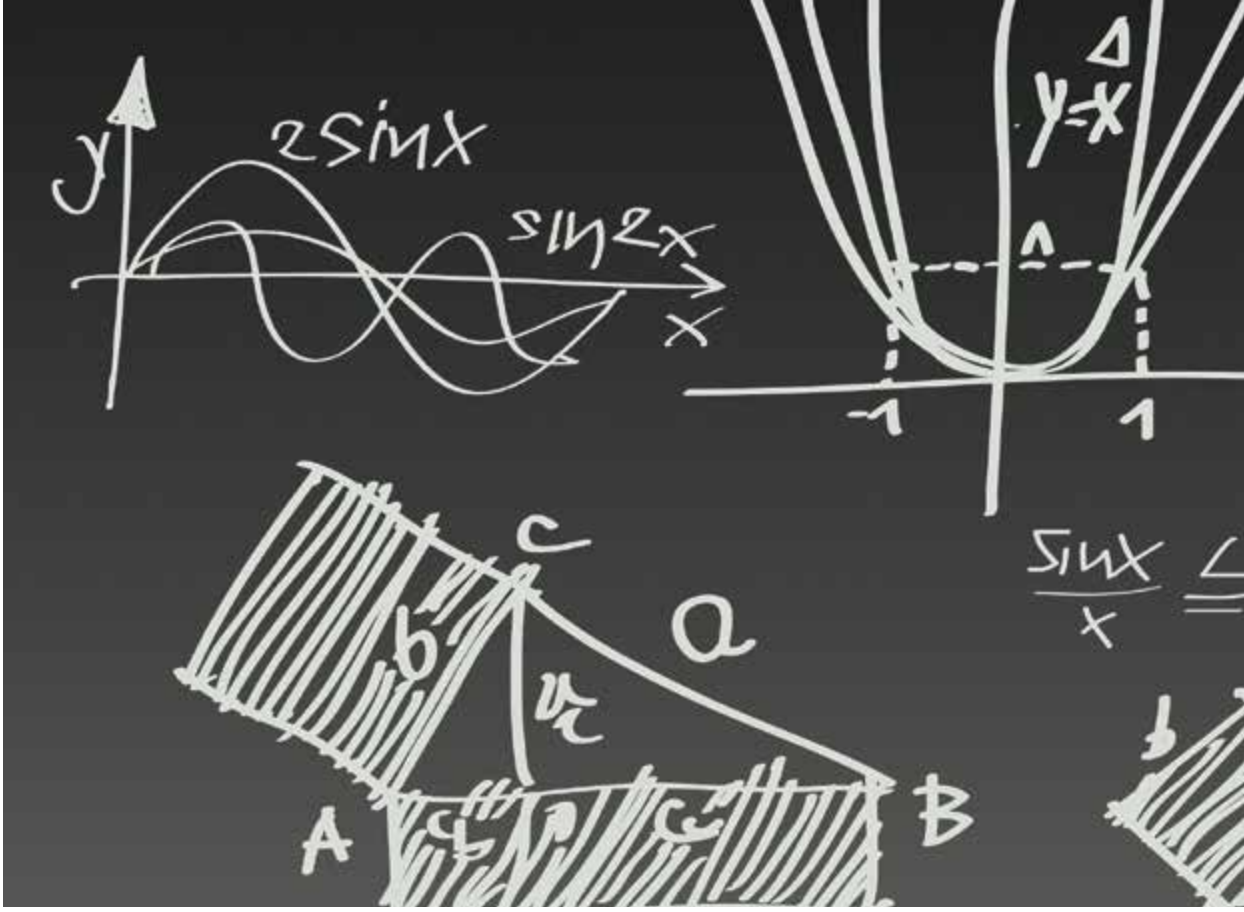
Physics teaches us that microscopic analysis, the domain of twentieth-century quantum physics, is random in nature and that infinite precision of measurement is pointless (it is impossible to know simultaneously the position and speed of a particle). Finally, at this level, observation interferes with experience.

By analogy with Big data, it emerges that "microscopic" observation of an individual can modify his or her behaviour. **If people know they are being "observed", then they will tend to modify their behaviour, making it less predictable.** Hence the importance of providing an environment of safety and trust, so that people feel free to decide their own behaviour patterns and are not reluctant to exchange or communicate their data.

### Genomics

Although mapping an individual's genome makes it possible, at "micro" level, only to declare probabilities (illnesses, etc.), at "macro" level, the combining of these individual probabilities makes it possible to reveal the real risks and opportunities for a large population group, thus justifying major investment that makes it possible to improve individuals' health (R&D, vaccines, treatments).

We can see that, if we respect the confidentiality of individual decodings that, when combined, make it possible to make discoveries, take decisions, implement public policies, **then a balance between progress and respect of privacy is achieved.**



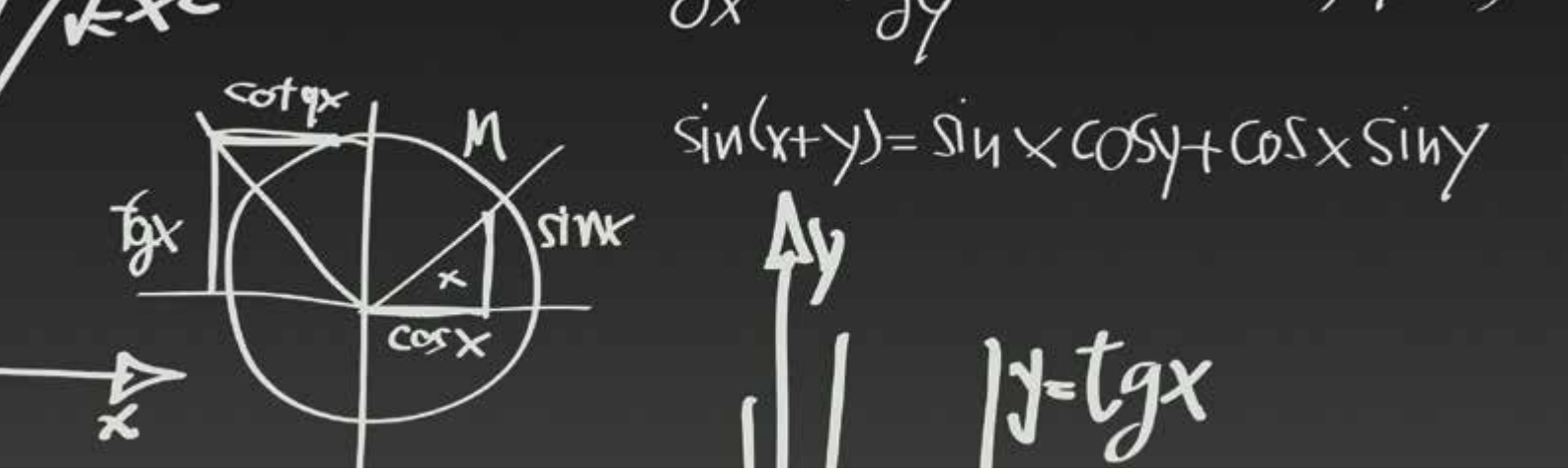

# Time for action

It would be fanciful to imagine that we can ignore the "Big data bang" or barricade ourselves against it: the exponential speed with which the mass of data generated is increasing and being stored, the advancement of the means of analysing and combining these data, and the potential and interest generated by the results make a moratorium unrealistic. Indeed, although the future is not entirely clear, it is nonetheless certain that those who are unable to face the challenge, right now, will be the big losers.

**It is therefore urgent that we reconcile the "principle of innovation" with the "precautionary principle".**

Invoked because of the risks, certainly real, that the use of Big data might pose to our privacy or identity, the "precautionary principle" does not, however, seem to be an appropriate response.

We have to pass through a necessary phase of "test & learn" experimentation, and the errors, even abuses, that are characteristic of this transition phase should raise awareness, particularly among those who have had the unpleasant experience of having their "private" life exposed online. These pitfalls will sometimes force people to exercise their ability to bounce back and grow a new skin, sloughing off their initial digital identity and any traces that might be associated with it.




# Entering the Big data era: where to begin?

## 1. Making a diagnosis

- Drawing up a map of the available data, both internal (business data) and external (websites, social networks, open data).
- Assessing the quality and pertinence of the data as regards development priorities, *e.g.* funding creation, innovation, services, research or segmentation of public/clients, increasing the value of data, operational efficiency...

## 2. Establishing a decision-making structure and key skills

- Determining appropriate organisation and governance for the use of data by thinking in terms of multi-disciplinary organisation;
- Enhancing analytical, mathematical, statistical and sociological skills for the management and use of data – whether by growing organically, purchasing digital companies or setting up partnerships. In other words, businesses should provide themselves with "data scientists", experts who are "able" to process the flood of data and draw out its entire quintessence in terms of decision-making and management[1].

## 3. Establishing a strategy and action plan

- Identifying and implementing means of enhancing client knowledge, particularly by:
  - encouraging unknown clients (readers buying at point of sale, theatre and cinema audiences, etc.) to identify themselves – by means of a loyalty programme, encouraging internet purchases;
  - analysing client behaviour patterns: content generated on social networks, client web browsing, history of transactions and content consulted.
- Identifying the level of means appropriate to the strategy: it is essential to identify clearly which data are useful in order to avoid the costly systematic storage of unusable data, and to give priority to the "test & learn" approaches proven by the big digital players, consisting of setting up test sites and regularly and systematically measuring their effectiveness in order to improve them rapidly or suspend them if necessary.

## 4. Developing and promoting an in-house data culture

In the media and cultural industry, the level of maturity as regards "data" differs from one player to another: those who have direct access to their public, with client or member bases (newspaper groups, events organisers...), already have a good knowledge of these groups and a direct marketing culture. The challenge of a Big data approach is to consolidate and use data in a context of increasing volume, velocity and variety of sources – particularly digital sources – which are difficult to reconcile with historical data.

It is an opportunity, but also a sizeable challenge, for players such as audiovisual broadcasters and content creators who traditionally had no direct link with their audiences.

The development of a client data-use culture and the concomitant skills also varies from department to department: marketing departments have already integrated these approaches, whereas editorial and creative teams have more difficulty in adopting them – and might even be reluctant to do so, because they might perceive the use of data as being incompatible with their ethical rules. Tools for analysing audiences and identifying trends have to be seen as complementary sources whose purpose is to enhance their content, not to replace their work on researching and identifying priorities. The success of computer graphics, which are heavily shared on social networks, shows that audiences are receptive to visual quantitative information.

Implementing this internal change alone will probably not be enough to extract the full essence of the data. It is by forging alliances that players in the cultural ecosystem will be able to capitalise on personal cultural data in order to innovate and create new value products.

1 *"Big data – nouveaux défis", Telecom ParisTech, N°169, July 2013*

### Acquisition of rights

**A company such as Netflix combines two sources of information to determine its rights purchasing strategy for its video subscription service: analysis of the most watched content (themes, actors, formats, etc.) and of the most pirated content.**

## Enhancing the data collected

**Data management and use in "Big data mode" lie at the heart of the economic model of the big global digital players.**

Despite the value inherent in making full use of personal data, French players are lagging behind the global digital players, which have positioned themselves as the reference distributors or broadcasters of creative content. One example is Amazon, which uses data (web pages consulted, purchasing history, wish lists...) to recommend content and products to its clients: 35% of its sales are linked to this mechanism. Or there is Apple, leading music distributor with iTunes, which analyses its users' music libraries (composition, most listened-to titles...) to recommend new artists or CDs.

As in other industries, players in the media and cultural industry can consider making full use of personal data to refine their knowledge of their public by, for example:

- Logically analysing the spread and content of messages (positive or negative) on social networks, as has been done by the transport industry (Quantas Airlines, RATP) and banking sector (Bank of America); identifying opportunities for broadcasting creative content, on the basis of the main interests and tastes expressed by users or communities.
- Anticipating behaviour, as in the case of analysing visits to a tourist area or concert hall.
- Cross-analysing external data (social networks, etc.) and internal data (browsing data, consultation/purchase histories, main interests, etc.) to gain a better knowledge of expectations and integrate them in the design of new content and services.

**By adopting a "data approach" as a means of optimising client value, players in the cultural industry will be able to meet these three challenges represented by the relationship with public, the broadcasting and distribution of creative content, and the financing of creation.**

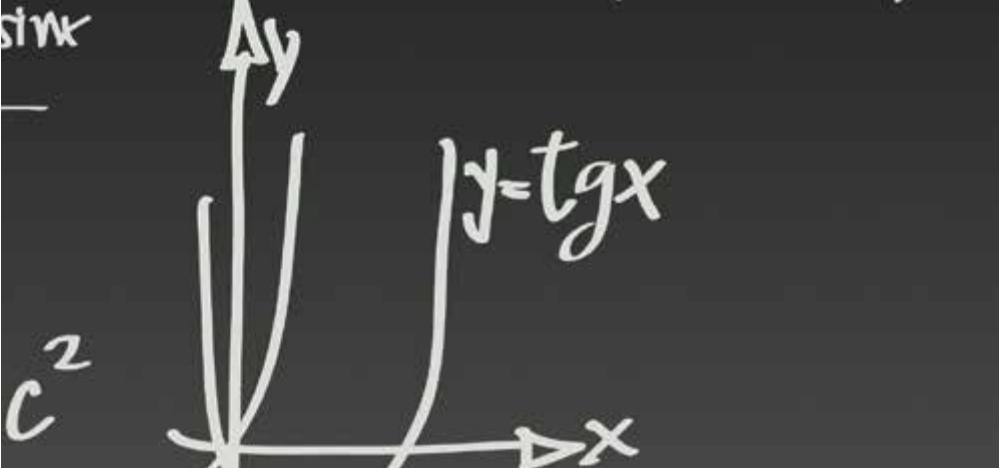
### 1. Enable the broadcasting and distribution of creative and manage the interaction with public

In the media and cultural industry, development of "client knowledge" makes it possible to respond to the major challenge of establishing a privileged, enduring relationship with public. This is particularly true in the case of event production and content creation, in order to create projects and content most appropriate to the target audience.

In the media, tools to analyse trends, such as Trendsboard, use the data generated by social networks and blogs to identify subjects that are beginning to make the news, thus making it possible for editors to anticipate a "buzz" or cover certain events which would previously have been much harder to predict.

Offering personalised interaction on the basis of the data analysed therefore seems to be a key way of creating or strengthening a close relationship with your public:

- The data management is a lever to promote cultural diversity within both cultural spots and digital media.
- Offering contents through the analysis of interests, such as recommendation websites, allows content to meet their public.
- Collecting and using personal data on your public makes it possible to prolong the relationship beyond a given event (live theatre, museum visit, film screening...) by creating communities and offering additional services and content;
- In the audiovisual industry, content offered on digital supports makes it possible to develop and take advantage of an individualised knowledge of audiences.




## 2. Obtaining a global view of your public's cultural activities

The tools used to collect and use data are often inappropriate and/or fragmented because activities are separated into silos. As a matter of fact, at the moment, a public establishment like the RMN-GP[1] has no way of knowing whether a person who bought a certain book in one of its 40 boutique bookshops might be interested in an exhibition at the Grand Palais or the Musée du Luxembourg, or any other service, such as a subscription to a newsletter or a themed tour concerning a specific artist. In fact, the information that would make it possible to make these links or predictions is not currently available, either because it is not being collected, or because it is scattered around various databases managed by different information systems which are not yet interoperable.

Therefore appears a dual challenge: the media and cultural industry needs not only to invest to provide itself with the necessary CRM skills, but also to ensure cross-referencing between the services involved in the collection and processing of data.

This global view is vital if the media and cultural industry players are to be able to offer a more personalised and contextualised experience and thereby hope to prolong the experience in the context of a committed relationship.

Furthermore, cultural spots (galleries, museums…) which use traffic data, such as many public transport projects, become more accessible, and will be able to formulate tiered pricing policies, based on the traffic data. Visitors will also get real time pieces of information, such as the waiting time to an exhibition.

Access to and the reutilisation of public cultural data might also help to create a global view of a public's cultural activities. Open cultural data has to be included among the actions taken by players in the cultural and creative industries to use and reap full benefit from their data. Offering third-party players, and particularly start-ups, the possibility of making use of cultural players' data will considerably broaden the field of possibilities, particularly for creating services for users or clients.

## 3. New funding for creation?

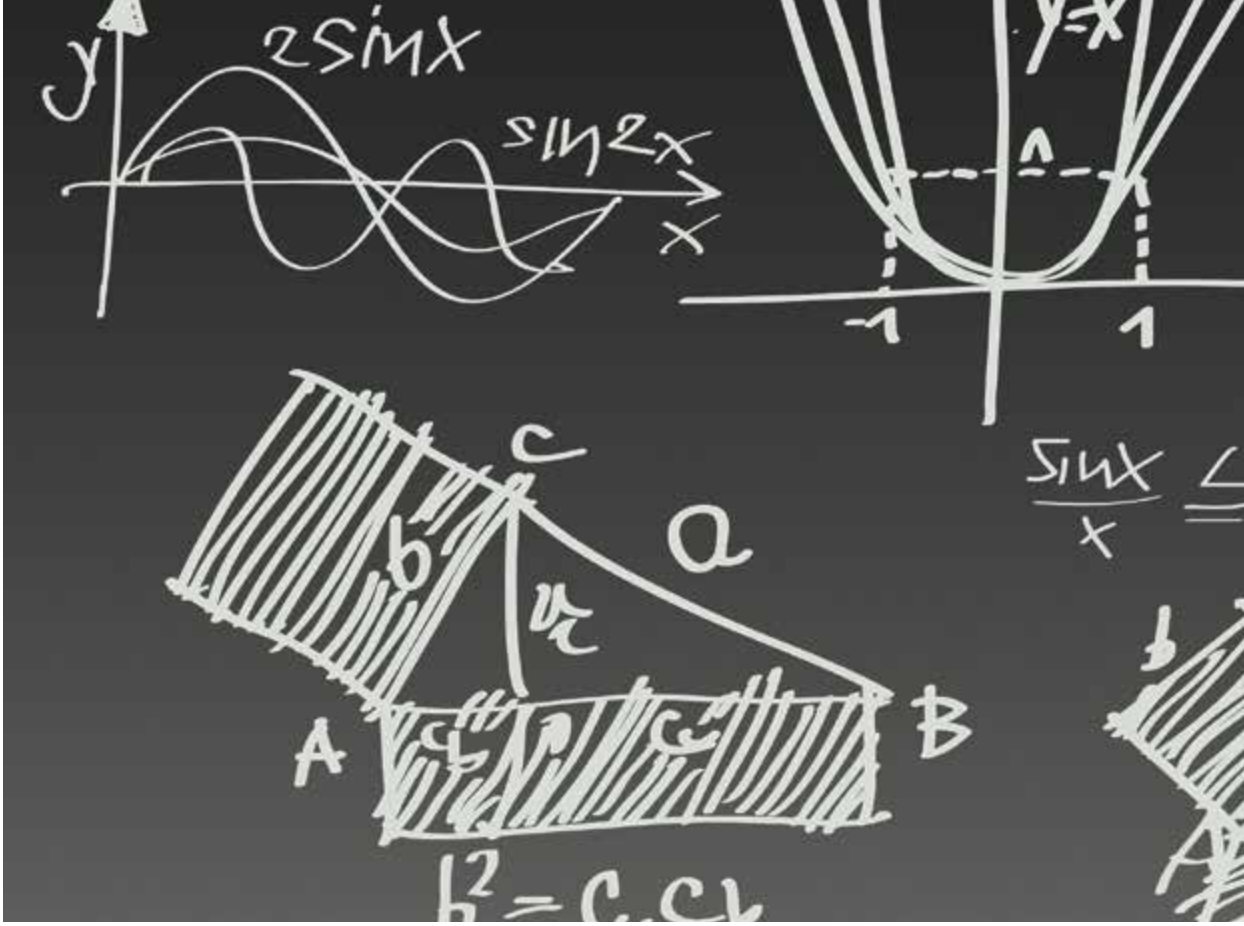
Data can be made available to other players in the cultural industry, distribution, social networks, etc. so that they can use them as they are or integrate them in other data combinations, in an 'open data' approach or with a commercial approach. This means that commercial use of data would make it possible to offer free or partially free content thanks to advertising. For advertisers, audiences have all the more value if they are defined so as to allow for effective targeting, hence the importance of recent data, even if they are anonymous.

Personal cultural data thus have a major role to play in an environment in which the cultural sector is seeking and inventing new funding models.

Even now, data lie at the heart of shared funding, whose model is based on the creation of identified communities, of connected individuals who make a personal commitment to funding a creative project. Crowd funding generates data (of a type similar to that generated by social networks) and, although sites do not monetise this data on an advertising model, they might use it to improve the site, make recommendations and help to increase funding.

In the future, personal cultural data will certainly be a new funding channel in an environment marked by a scarcity of public funding. Plans for funding film or recording projects could, for example, include taking full advantage of any new data collected: an executive producer would give a co-producer the right to manage the community of fans of the work created, and any advertising revenue generated.

1 *Réunion des musées nationaux - Grand Palais*



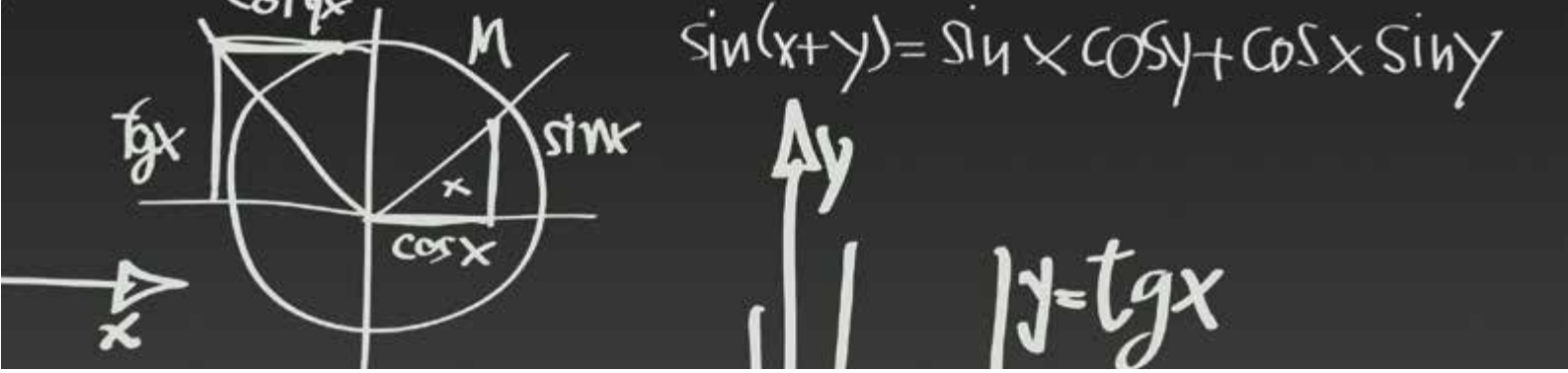

**In France, several public establishments are up to speed on the open publication of cultural data**

Since the opening in 2011 of the data.gouv.fr portal, the French Ministry of Culture and Communication has been working on drawing up an inventory and providing developers with various data sets, such as a list of public cultural establishments together with their addresses and publications, media-library data on architecture and heritage (list of buildings protected as historic monuments, list of publicly-owned properties classified as Historic Monuments), and also about thirty data sets held by the Centre national du cinéma et de l'image animée (CNC – French national film and animation centre), including, for example, the international receipts of French films and a list of film establishments. Many towns and regions have also started offering open access to public data, which might be (anonymous) demographic statistics, as well as lists of film sets and statistics on the consultation of works in libraries.

Open access to these data, cross-referenced with the main interests of individuals or communities, provides an opportunity not only to develop new services and uses, but also to facilitate the promotion of and access to culture.

There are already many applications based on open cultural data, including:

- Interactive maps, such as the module developed by the French Ministry of Culture and Communication, which represents all of France's cultural sites in the form of a galaxy, or the "Cartographone", a map showing film sets in Paris.
- Data visualisation interfaces, with, for example, an article from the newspaper, Le Monde, containing a module that makes it possible to visualise interactively the statistics on the consultation of works in Paris's libraries.
- Practical applications making it possible, for example, to find the nearest library, information on local museums (such as Musambule in Marseilles), or even multi-editor cultural programmes (such as Cibul in the Provence-Alpes-Côte d'Azur region).
- Enhanced visits: as Culture Clic, offering not only practical information on French museums but also a catalogue of 900 works that can be viewed in enhanced reality.




# The new cultural data value chain

## New players

The personal cultural data market is developing apace: throughout the data value chain, producers, collectors, aggregators and users of personal cultural data are shifting and positioning themselves (production/collection and use of data, and even storage in some cases).

This process is likely to continue and lead to the emergence of new players and new jobs: pure players specialising in the production/sourcing of data, others specialising in the collection/analysis of data (cross-referencing data from various sourcing specialists and with other data in order to contextualise them and provide users with the analysis they require) and users of personal data who will develop targeted services and applications with higher added value.

All these services will need a Big data infrastructure, thus bringing in another category of players.

The smooth functioning of the whole ecosystem of personal cultural data depends upon what we shall call the "trust filter".

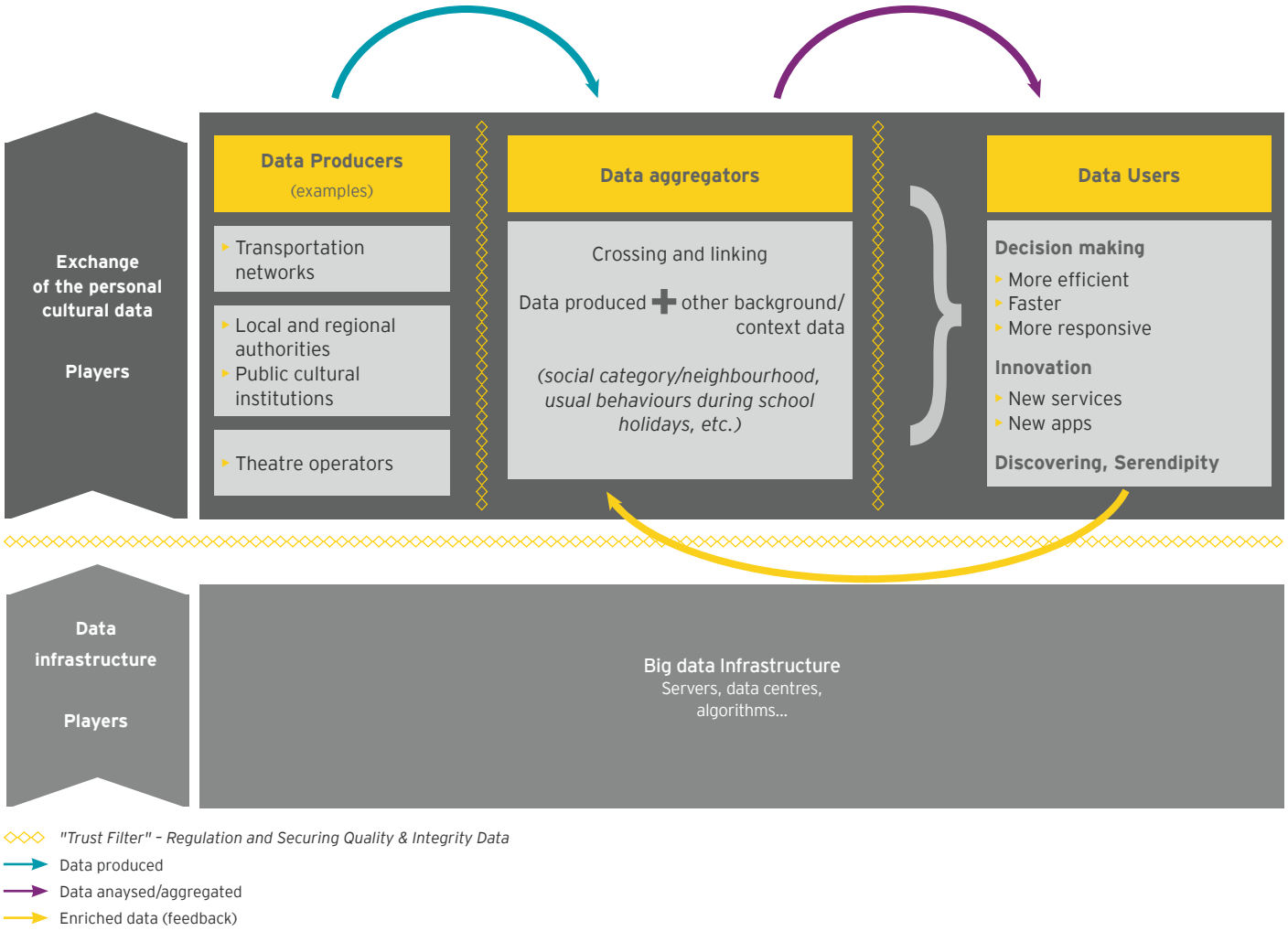
This diagram illustrates the development of a market in which data and their use will be separated; this would have the three-fold advantage of providing a response to a market that would otherwise move towards a natural monopoly, meeting the challenge as regards the nominative data collected/exchanged, and freeing up all the data's potential in terms of innovation.

**The players in the new data value chain**

- **Data producers:** who will build genuine "enhanced-value" databases and, capitalising on their trust capital, will be able to sell the data on to third parties.
- **Data aggregators:** who will combine, cross-reference and contextualise these data for analytical purposes.
- **Data users:** who, using the analyses obtained, will develop new services and new applications. These users might be media companies and start-ups developing specific, targeted applications for media groups or cultural establishments (Hopper application).
- **Big data infrastructure operators:** new digital players, telecommunications operators, etc.
- **Infrastructure managers:** who might, in addition to their storage capacities, rent calculation capacities to companies with only occasional needs as regards the processing of Big data.
- **Trust managers:** who ensure that the integrity/quality of data are maintained from sourcing to use, thus reassuring the end user and consumer. This phase is essential to guarantee, for example, that the electronic version of a work downloaded online (e-book) is the same as the original work.

**The players of the new data value chain**

*Source: EY ©*




## Positioning oneself in this value chain

**Some players will seek** to expand their presence in the data value chain, either horizontally (from data production to data use) or vertically (from Big data infrastructure to services for end users and clients). This is already happening in the case of the Internet giants, either through internal growth or through external growth/**alliances with pure players**.

Some **innovative business models (pure players)** are going to emerge:

- **Sourcing**.
- **Specialisation** in the development of highly targeted mobile applications and services (for example on the mobile video games market).
- **Certification and auditing associated with data security throughout the value chain**, to respond to consumers' and users' requirements as regards transparency and trust, but also to prevent reputational risk.

# Personal cultural data at the heart of Big data

## Infographics

# The personal cultural data, a golden resource

## A new strategic market is structuring around the personal cultural data

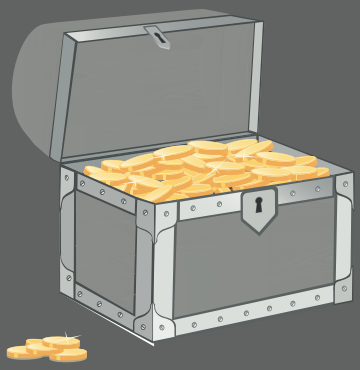
*Source: EY ©*

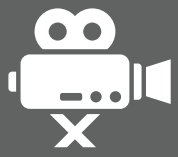
***Homo conexus***
2.3 billion
in the world in 2013

**Internet of things**
20 billion objects connected
in the world in 2013

**Explosion of digital traces**
4 zettaoctets of data produced in 2013,
a pile of DVDs as high as 2 million Empire State buildings

Cultural
institutions





Media and entertainment
industry

**Personal cultural data**
Data of intimacy generated
by the cultural behaviour

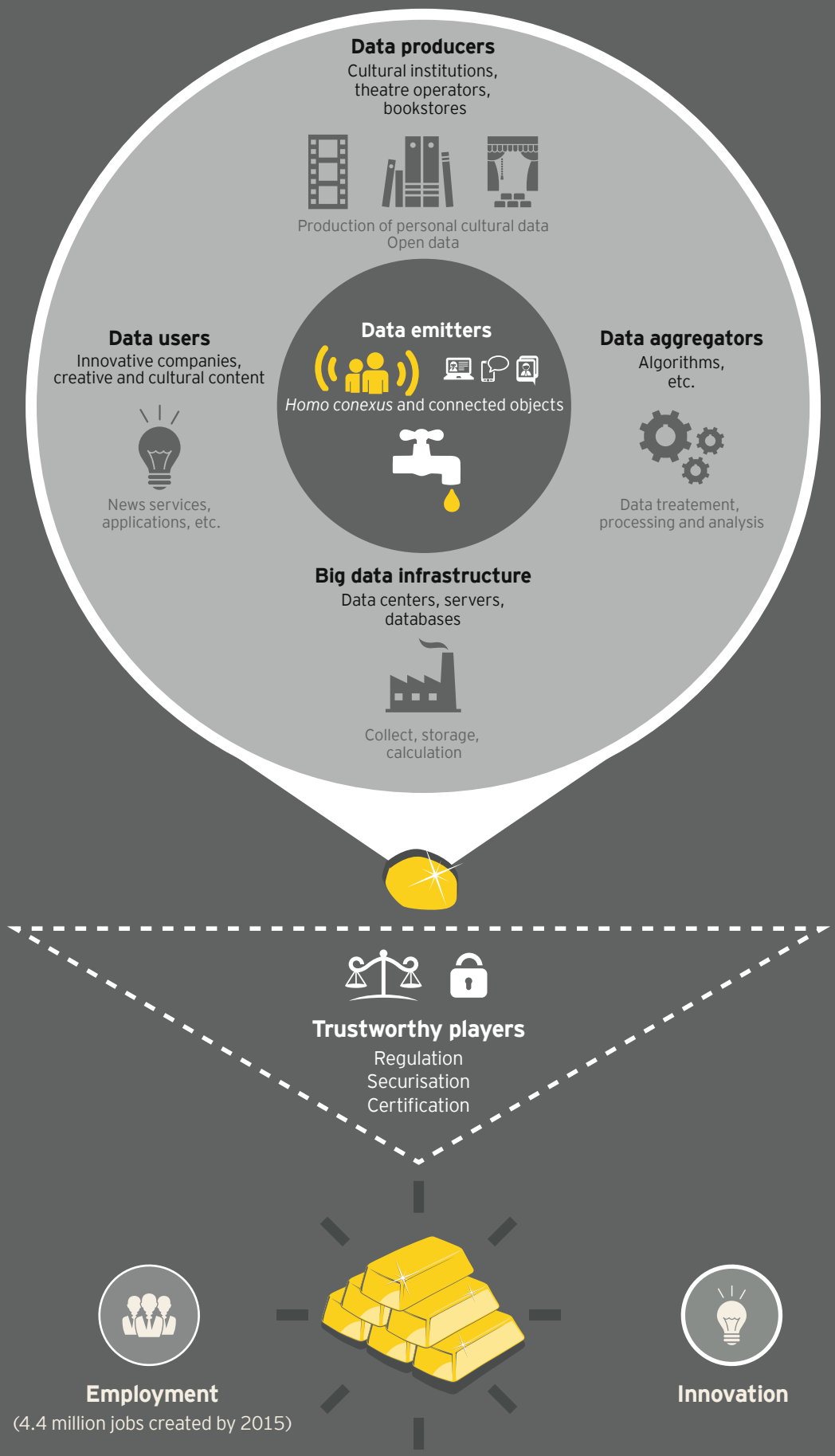
GAFA*

* GAFA: Google, Apple, Facebook, Amazon

# From Big data to Big value

## Towards an ecosystem based on trust

*Source: EY ©*



## About EY

EY is a global leader in assurance, tax, transaction and advisory services. The insights and quality services we deliver help build trust and confidence in the capital markets and in economies the world over. We develop outstanding leaders who team to deliver on our promises to all of our stakeholders. In so doing, we play a critical role in building a better working world for our people, for our clients and for our communities.

EY refers to the global organization, and may refer to one or more, of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. For more information about our organization, please visit ey.com.

© 2013 Ernst & Young et Associés
All Rights Reserved.

Studio EY France - 1309SG178
SCORE France N° 13-032
Photos: © Fotolia - © Gettyimages - © EY
ED: none

This is a free translation from the French original.

This material has been prepared for general informational purposes only and is not intended to be relied upon as accounting, tax, or other professional advice. Please refer to your advisors for specific advice.
The opinions of third parties set out in this publication are not necessarily the opinions of the global EY organization or its member firms. Moreover, they should be viewed in the context of the time they were expressed.

ey.com

## Contacts

**Bruno Perrin**
Partner, Ernst & Young et Associés
Telecom, Media, Technology Leader in France
Tel.:+ 33 1 46 93 65 43
Email: bruno.perrin@fr.ey.com

**Fabrice Naftalski**
Partner, Ernst & Young Société d'Avocats
Tel.: + 33 1 55 61 10 05
Email: fabrice.naftalski@ey-avocats.com

**Éric Fourel**
Partner, Ernst & Young Société d'Avocats
Tel.: + 33 1 55 61 11 22
Email: eric.fourel@ey-avocats.com

**Solenne Blanc**
Executive Director, Ernst & Young Advisory
Tel.: + 33 1 46 93 59 17
Email: solenne.blanc@fr.ey.com

**Louisa Melbouci**
Media and Entertainment Marketing Manager
Tel.: + 33 1 46 93 76 47
Email: louisa.melbouci@fr.ey.com

## Press contact

**Velislava Chichkova**
Tel.: + 33 1 46 93 45 06
Email: velislava.chichkova@fr.ey.com

This EY report was realised under the supervision of Bruno Perrin, Fabrice Naftalski and Régis Houriez, with the contribution of Marie-Pierre Bonnet-Desplan, Solenne Blanc, Vincent Placer, Louisa Melbouci, Pierrick Vaudour, Guillaume Marcerou, Sébastien Bardou, Aurèle Tabuchi, Colin Garnier, France de Roquemaurel and Sandrine da Cunha.