

# Chapter 4

## Factoring Big Data into the Business Case for IoT

Anastasius Gavras

### 4.1 Big Data

There is a widespread expectation that big data offers tremendous opportunities. The potential economic value to be generated from the vast amount of data is given in tens or even hundreds of billions of euros per year. However, the cost of big data is rarely mentioned. We can hardly find figures explaining the long-term cost of big data. Yet, there must be a price tag attached to the investments needed for handling big data, including operational expenses for maintaining the data and making them available in the long term.

Citing ex-Commissioner Neelie Kroes' speech at the ICT 2013 event in Vilnius [1]; *every two days we create as much information as was created from the dawn of civilisation to 2003*. In addition she claimed that big data is growing by 40 % per year, a figure that is hard to correlate to the previous statement and in fact even harder to believe if we accept that a similar law like Moore's law applies on big data and that Moore's law will likely continue to be valid until at least 2030 [2], implying that most likely the growth of big data is higher than claimed by Neelie Kroes.

#### 4.1.1 Expectations

The traditional big stakeholders in very large database management systems see a great opportunity and a new market potential. The current hype speaks indeed about disruptiveness in how business is conducted. A large share of the value to be

---

A. Gavras (✉)  
Eurescom GmbH, 69123 Heidelberg, Germany  
e-mail: gavras@eurescom.eu

created will come from new types of data use which are unprecedented. There are tremendous investments to be made in storage, computation and transmission capacity; and there are undoubtedly costs for keeping the systems running.

High expectations are put in non-ICT sectors that so far, although users of ICT had little exposure to big data because it was not easy and it was costly due to highly specialised solutions. Sectors often mentioned include energy, environment, agriculture, health, government and many others. The same applies to purely scientific data repositories. The sectors today are either increasingly engaged in collecting new data, or engaged in opening their archived data under the open access legislative initiatives. Current economic sectors are able to provide a cost-benefit analysis, although the benefits are rather expectations today and depend a lot on yet unknown applications that may emerge, while there is a lot of uncertainty.

It is often neglected that any data that is collected has only meaning in its context. All big data must be contextualised in its scientific or socioeconomic context. In a big enough data set one can discover unrelated correlations and the temptation is high to accept such a correlation as *truth* just because it is discovered in big data. This problem has been impressively demonstrated by the website ‘Spurious Correlations’ [3] and analytically discussed by Calude and Longo [4], who prove that very large databases have to contain arbitrary correlations. These correlations appear only due to the size, not the nature, of data. This phenomenon plays an important role later in the section about the necessary skills to master data sets.

### **4.1.2 Big Data Players**

Open platforms that support knowledge exchange and sharing of largely unstructured data are a strong trend. Several initiatives and companies have picked up this basic idea and try to support organisations in managing their unstructured data and extract knowledge out of the data. However, the landscape is currently very unclear and full of buzzwords.

The fact that new data management companies emerge means that there is a certain complexity which has a cost and through which a profit can be realised. Those who find ways to efficiently manage the data complexity at low cost will be able to sustain a viable business.

Large established IT companies like IBM are offering solutions for managing big data and are advertising these by publishing related use cases [5]. The topics advertised are: Big Data Exploration, Enhanced 360° View of the Customer, Security Intelligence Extension, Operations Analysis and Data Warehouse Modernization. Taking a closer look into the use cases, I can hardly discover disruptive potential. In most cases it is about doing things better and faster which is good enough for a start.

### 4.1.3 Cost Versus Value

Admitting that there is a data management cost due to the sheer volume and complexity of big data, even considering the dropping cost of enabling technology, inevitably raises the question about the break-even point. The following general considerations are meant to highlight the challenge.

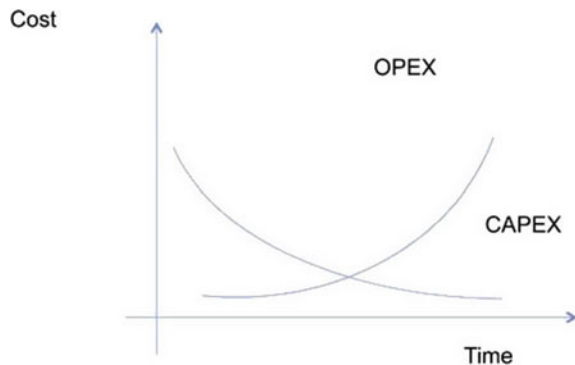
Figure 4.1 illustrates the increasing management cost, which roughly translates to Operational Expenditure (OPEX), versus the dropping technology cost, which roughly translates to Capital Expenditure (CAPEX) over time.

From the dawn of IT, experts are trying to calculate the optimal equilibrium state of processing, transmission, and storage costs for data. Simply put the question is when it is more economical to transmit large raw data volumes multiple times, versus replicating these data volumes at several locations, versus re-calculating a *result* and transmitting or storing only this result. Of course we know today that this is a moving target and depends on the costs trends of the different related technologies. However virtualization of technology and its use as a service is blurring its classification into the established boxes (OPEX/CAPEX). Using cloud storage induces OPEX, although in the past the purchase of storage arrays was inducing CAPEX. Clever buzzword finders have started to coin the term COPEX (standing for Capital and Operational Expenditure). As of May 2016 this definition for COPEX does not appear yet in the web's largest acronym repository [6].

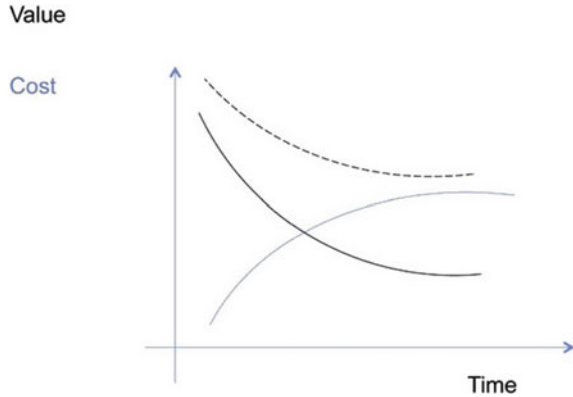
As a concrete example of OPEX for big data, Google posted in 2009 [7], figures about the actual energy cost for a query. According to Google's blog article a single query accounts for 0.0003 kWh or energy. In terms of greenhouse gases, one Google search is equivalent to about 0.2 g of CO<sub>2</sub>. Today Google is striving to reduce its CO<sub>2</sub> footprint to zero [8]. Of course this is achieved by large investments in renewable energy, which means a shift of OPEX to CAPEX. Newer studies are published almost daily that try to pinpoint accurate cost figures for the resource consumption of ICT.

Figure 4.2 illustrates the aggregate cost (CAPEX + OPEX) as compared to the value for society. The dotted line indicates an unanticipated scenario in which the

**Fig. 4.1** Qualitative projection of CAPEX versus OPEX for big data



**Fig. 4.2** Qualitative projection of aggregate cost versus value for big data



total cost of ownership exceeds in the long term the value provided to society or businesses. We still do not have a complete picture of the overall cost of Big Data and the Internet of Things (IoT).

#### ***4.1.4 Societal Cost***

Not all costs have a direct price tag in terms of money or CO<sup>2</sup> footprint. Not focused on Big Data and the Internet of Things alone, the current calculation of the total environmental impact of ICT may have some defects [9], for example by not considering the rebound effect of stimulation of increased demand due to time-saving optimisation, or the software-induced hardware obsolescence and the miniaturisation paradox, which indicate that hardware is getting cheaper faster than it is getting smaller.

In an article, Helbing et al. [10] analyse the impact of advanced algorithms, artificial intelligence and Big Data on the future of society at large, pledging for the safeguarding of fundamental societal values developed over centuries such as freedom and democracy.

#### ***4.1.5 Skills and Enabling Technology***

Miniaturisation and cost reduction of ICT has enabled almost anyone to collect and make accessible digital data, mainly because the technology exists and is affordable. Terabyte hard disks are nowadays in the 50 Euro range, very small scale PCs running free and open source software, such as Linux, are available for less than 30 Euros, and broadband subscriptions, even mobile broadband, are in the 20–40 Euros per month range. With these components at hand each digital native is

capable of building a system that collects, temporarily stores, eventually processes and maybe makes available arbitrary and unstructured data virtually without limits. But is everyone able to assess, organise, preserve and maintain the data in the long term?

As noted earlier the knowledge in the data requires often the correct interpretation by domain experts applying fundamental scientific practices in addition to mining the big data sets, in order to avoid fallacies as discussed in [4]. Relying only on popular wisdom that ‘numbers speak for themselves’ is a dangerous assumption. It is a scientific skill to give numbers their meaning.

Once we have the skills to combine scientific rigour with big data analytics the relevant question about the value of the knowledge that could be extracted from the data and whether this value can be monetized to cover its cost and yield an economic profit and societal benefit should be easier to answer.

### 4.1.6 *Streams of Data*

At a certain point in time the cost of processing and querying big data stores may become economically unaffordable. At this point in time we must have found ways to intelligently distil useful knowledge out of a passing stream of unstructured data and just drop the rest. The challenge lies in identifying what we can extract from this stream of raw data that is produced by the Internet of Things, and which will provide also in the future opportunities for yet unforeseen use of historical big data.

Stream data processing is also necessary in cases where even structured data would be meaningless without their temporal characteristics, implying their processing in real time or capturing the full context in space and time. Obviously stream data processing makes most sense at the edge of the network, basically at the location where data are generated. However due to the extreme decentralisation of processing this may result in increased security problems, since *traditional* security measures—as will be discussed later—are not appropriate for securing every single data source.

## 4.2 The Internet of Things

The Internet of Things (IoT)—used herein in a broader scope—promises many benefits in terms of new applications and in particular new opportunities for a substantial change in societal behavioural patterns. And indeed we have witnessed many exciting new technologies and applications that are enabled by the IoT. Considering IoT from a narrow point of view and looking to the currently deployed sensors in smartphones we could actually question whether there is actually an Internet of Things that is different than what we have today. This is because it looks

like we have already multi-billion sensors connected to the network, yet the vendors and operators manage to keep the networks up and running and new services emerging daily. It is time for a reality check and to ask the question about the sustainability of the current approach as well as to examine future directions for new business cases enabled by IoT.

### ***4.2.1 Cost for Connectivity***

In terms of cost for connectivity, we have not progressed much in the last few years and perhaps we have even done a few steps backwards. Today virtually all IoT applications are based on some sort of client/server principle in which there is a substantial computing capacity in a virtualised computing environment to which each single sensor and actuator is connected somehow. This means that beyond the investment needs for the IoT enabled world (CAPEX) in the frontend, there is a substantial OPEX and CAPEX in the backend support for IoT-enabled applications. Assuming an average lifetime of three to five years for the hardware delivering the computing capacity we are facing a disconnect with respect to the lifetime of other supporting infrastructures, such as networking where the depreciation time of infrastructure investments is in the order of 10–20 or more years, although shrinking. This means that the cost of supporting and serving hundreds of billions of smart objects in the long term is considerable and may not be included in the current assumptions about Total Cost of Ownership (TCO).

### ***4.2.2 Security and Trust***

In terms of security and trust, there exists to date no future proof concept; even less a concept that is economically viable at the anticipated scale. The traditional model in which IT domains are organised and protected centrally by some gatekeeper will not work for the IoT world for reasons induced by the extreme decentralisation of most IoT-enabled systems. We cannot protect each smart object individually on an economically viable basis.

In the area of trust falls also the practice of many vendors delivering smart products that collect data about the behaviour of their customers in the hope that these data may be an exploitable asset. This is a serious and to date underestimated problem. In many cases individuals may not care (even if they knew) about the practice. However enterprises care a lot and sometimes have the means to discover and resist the practice. At least in Europe the trend in legislation is to strengthen the rights of citizens and businesses with respect to the control of their data.

Furthermore, a discovered vulnerability is a product defect and must be fixed. Traditional industries have been hit very hard economically in cases where recalls are necessary (e.g. the car industry). In the ICT industry the strategy is to distribute

software updates, which nevertheless puts an increasing cost to the long-term maintenance of IoT-enabled applications. The fact that many new IT devices have a short expected lifetime motivates vendors to drop older devices from their maintenance roadmaps so that these devices do not receive security updates anymore. How will such devices be protected in the future? Do we need to replace our energy smart metres, digital doors locks and smart connected cars every three or five years when the vendors stop delivering security updates?

### 4.2.3 *Future-Readiness of IoT*

In *Cyber Physical Systems* (CPS) IoT devices are directly connected to real-world artefacts and have a substantial influence on their properties. Where homes, buildings and factories are smartened, it is a reasonable expectation by the customers that the devices used in this context have an expected lifetime in the same order of magnitude to be future proof. This is in the order of 30–50 years and beyond. In many cases different generations of products, that appear every 2–3 years, have different maintenance requirements and different ways of servicing. How will vendors and suppliers be able to cope with the long-term cost of maintenance? Dumping the long-term cost on the customers will not work, because when the early adopters discover that the TCO of a smart fridge is an order of magnitude higher than that of the stupid old fridge, they will start to question the added value of the smart world.

The hype about IoT and CPS has triggered many ‘innovations’ in the market for which the added value for the customer is questionable. Vendors may think in terms of better maintenance and service for the benefit of the customer or about warranty tracking for the benefit of their own supply chain optimisation. But a smart coffee brewer, a smart toaster, or a smart water boiler, do not provide added value to the customer unless they make better coffee, better toast or better hot water! Certainly this observation does not devalue many useful applications of the new technology, but it is always a simple and clear value proposition that decides about the market success of an IoT-enabled product or service.

All above deficiencies culminate to an uncomfortable truth that the current business models around IoT might be broken. Of course this is a pessimistic view. An optimistic view formulates the problem such as that no one has yet found a viable and sustainable business model for the large scale. In order to progress the search for viable business models, the discussion of the broader concerns above, gives us an indication that we have to face issues in four dimensions; namely (i) technology, (ii) business, (iii) policy and (iv) last but not least customer. All four dimensions are a source for requirements that have to be satisfied at the same time. Smart city experiments around the world are a good starting point to learn to deal with all these requirements, since all dimensions are prominently present in most

scenarios. However all smart city pilots that exist to date are just that: pilots! None of these experiments claims a self-sustainable operation that is ready for the long term in the city scale.

In order to develop sustainable business models, they should not be designed around the traditional understanding of value chains, but should rather be designed with the flexibility to cope with value networks that emerge in digital business ecosystems.

#### ***4.2.4 Legal Frameworks***

In January 2012, the European Commission proposed a comprehensive reform of data protection rules in the EU. On 4 May 2016, the official texts of the Regulation and the Directive have been published in the EU Official Journal in all the official languages. While the Regulation will enter into force on 24 May 2016, it shall apply from 25 May 2018. The Directive enters into force on 5 May 2016 and EU Member States have to transpose it into their national law by 6 May 2018 [11].

A further regulatory reform may be triggered by the question on product liability. Traditionally product liability is limited to *products* in the form of tangible personal property. In the future the correct functioning of a device (e.g. IoT, medical sensor) includes a functioning network and backend service. Smart (connected) devices will have a far reaching impact on manufacturers, service companies, insurers and consumers, since legally a product or service may become defective upon network or service failure (even temporal) or upon discovered security vulnerabilities.

The new European data protection rules and potential evolution of the legal framework on liability will have a significant impact on how businesses deal with data, products and services and it is reasonable to expect that it will induce a higher long-term cost.

#### ***4.2.5 Cannot Sell the Solution***

Many companies are developing solutions that are marketed as products, e.g. for smart cities or large corporate customers. But a city is unlikely to buy a solution which is prone to all uncertainties about the economic viability and correct functioning of the solution in the long term as described above. Of course the cities are eager to provide more services to their citizens and to find ways to more efficiently manage the city as a whole. But the value is in the information and the knowledge that can be extracted from the data a solution provides, once it is deployed and operational. Buying the information or knowledge means buying sanitised data that are extracted by specialists and who can perhaps attach a *quality label* on the data.



### 4.2.6 *The Value Is in the Data*

This means that the business cases for IoT and Big Data is in fact a business case for Quality Data. Quality guarantees could be of substantial value to customers, whether public customers (cities, government) or private corporations. From a financial and investment point of view such a customer would have no CAPEX and would only subscribe to a service that provides Quality Data, hence paying a subscription fee. Such a fee can much easier be assessed with respect to the value it provides to the customer and consequently a decision to subscribe could be easier and faster.

Remains the question of who should build the infrastructure for IoT and who should extract the knowledge from the Big Data. The dilemma looks like the recent dilemma of the telecoms industry for which the telecom operators are building and operating the infrastructure, however other—so called over the top (OTT)—players earn the profits. However this is not entirely true since as hinted earlier many grass root initiatives are emerging and growing, which are building or retrofitting IoT infrastructures and are offering interfaces to access the information which can be utilised to extract Quality Data. Some of the observed properties of the shared economy should be examined in this context to identify how a value network could be build that allows for a reasonable compensation of all stakeholders. This compensation may not be of financial nature in all cases.

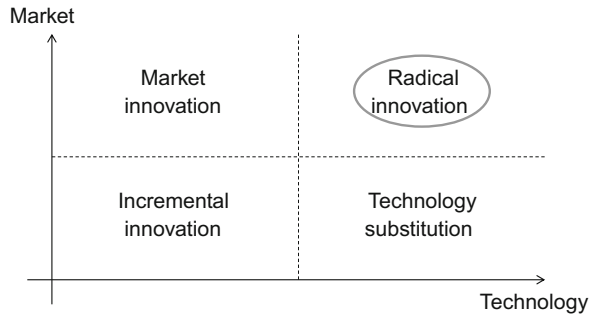
A more traditional and therefore perhaps more credible assumption is that new entrants, in fact they should be called *market creators*, will appear in the market and who will deploy IoT and Big Data solutions directly delivering *radical innovation*. They would own and operate the solutions and take the full risk (and profit) of this approach. They will offer Quality Data as a Service (QDaaS) for anyone that is willing to subscribe to a real-time feed of Quality Data in the same way as we subscribe to a newsfeed today or subscribed to a newspaper in the past. Most importantly some market creators may augment such data with vertical sector knowledge which will increase the Value of the Data and will render them more useful for sector applications and services.

These market creators would then leverage on the current communication, computing and storage infrastructures provided by the incumbent ICT stakeholders, regardless of them being network operators or cloud providers and will provide value based on Quality Data beyond networking and services.

### 4.2.7 *Radical Innovation*

This subsection provides an interpretation of *radical innovation* as was used in the previous subsection. Innovation per se, can be categorised according to several

**Fig. 4.3** Partitioning innovation in technology versus market dimensions



aspects. For example strategic innovation, which has a business oriented focus, and process or product innovation, which are concerned with the improvement of existing, or the introduction of new processes or products. Herein innovation is perceived along the axes *Technology* and *Market* as illustrated in Fig. 4.3. *Incremental innovation* denotes the case where existing but improved technologies are used to improve existing services and products. *Technology substitution* denotes the case where new technology is used to create new products and services in existing markets. *Market innovation* denotes the case where existing but improved technologies are introduced in new ways into the market, effectively creating a new market segment. Finally *Radical Innovation* denotes the case where new technologies are used to create new markets, effectively introducing disruptions on both axes.

*Big Data*, and the *Internet of Things*, taken alone, introduce relatively clear paths to either *Market Innovation* or *Technology Substitution*. However, since the first endeavours to introduce disruptive ways of creating knowledge from existing data and the latter endeavours to create new technology for the enablement of new products and services, in combination they likely lead to *Radical Innovation*.

### 4.3 Conclusions

This paper names and analyses some of the expectations and possible long-term effects in relation to Big Data and the Internet of Things. It spells out the hidden cost of the new technologies that are emerging very fast and fuel the imagination of entrepreneurs and investors. Certainly it does not aim to diminish the success stories achieved so far by early adopters. Its purpose is rather to point out the areas that need further attention in the future, such as security at large, or to deliver quality data, as well as to sketch a possible future in which radical innovation takes place for the advantage of the brave entrepreneurs and early adopters and the benefit of the society in the long term.

## References

1. [http://europa.eu/rapid/press-release\\_SPEECH-13-893\\_en.htm](http://europa.eu/rapid/press-release_SPEECH-13-893_en.htm) (Dec 2015)
2. Powell JR (2008) The quantum limit to Moore's law. Proc IEEE 96(8). <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4567410> (May 2016)
3. Spurious Correlations, <http://www.tylervigen.com/spurious-correlations> (May 2016)
4. Calude CS, Longo G (2016) The deluge of spurious correlations in Big Data. In: Foundations of science, pp 1–18. doi:10.1007/s10699-016-9489-4, <http://www.di.ens.fr/users/longo/files/BigData-Calude-LongoAug21.pdf> (May 2016)
5. Big Data at the speed of business; the 5 game changing big data use cases, IBM. <http://www-01.ibm.com/software/data/bigdata/use-cases.html> (May 2016)
6. What does COPEX stand for? <http://www.acronymfinder.com/COPEX.html> (May 2016)
7. Powering a Google search, Jan 2009. <https://googleblog.blogspot.de/2009/01/powering-google-search.html> (Dec 2015)
8. Google green, our footprint: beyond zero. <http://www.google.com/green/bigpicture/> (May 2016)
9. Gavras A (2008). Why there is no such thing as “green ICT”, Eurescom mess@ge magazine, Nov. 2008. <http://archive.eurescom.eu/message/messageNov2008/Why-there-is-no-such-thing-as-green-ICT.asp> (Dec 2015)
10. Helbing D, Frey BS, Gigerenzer G, Hafen E, Hagner M, Hofstetter Y, van den Hoven J, Zicari RV, Zwitter A (2015) Digitale Demokratie statt Datendiktatur, published Dec. 2015. <http://www.spektrum.de/news/wie-algorithmen-und-big-data-unsere-zukunft-bestimmen/1375933> (May 2016)
11. European Commission, Protection of Personal Data. <http://ec.europa.eu/justice/data-protection/> (May 2016)