Chapter 2
What Data Science Means to the Business

Abstract Big data have been associated with some common misconceptions so far, and this chapter will help the reader in identify and understand those fallacies. It is going to be then shown the best data deployment approach, followed by an ideal internal data management process. A four-stages development structure will be provided, in order to assess the big data internal advancements, and a data maturity map will summarize a set of relevant metrics that should be considered for an efficient big data strategy.

Data are quickly becoming a new form of capital, a different coin, and an innovative source of value. It has been mentioned above how relevant it is to channel the power of the big data into an efficient strategy to manage and grow the business. But it is also true that big data strategies may not be valuable for all businesses, mainly because of structural features of the business/company itself. However, it is certain that a data strategy is still useful, no matter the size of your data. Hence, in order to establish a data framework for a company, there are first of all few misconceptions that need to be clarified:

(i) **More data means higher accuracy**. Not all data are good quality data, and tainting a dataset with dirty data could compromise the final products. It is similar to a blood transfusion: if a non-compatible blood type is used, the outcome can be catastrophic for the whole body. Secondly, there is always the risk of over fitting data into the model, yet not derive any further insight—"if you torture the data enough, nature will always confess" (Coase 2012). In all applications of big data, you want to avoid striving for perfection: too many variables increase the complexity of the model without necessarily increasing accuracy or efficiency. More data always implies higher costs and not necessarily higher accuracy. Costs include: higher maintenance costs, both for the physical storage and for model retention; greater difficulties in calling the shots and interpreting the results; more burdensome data collection and time-opportunity costs. Undoubtedly the data used do not have to be orthodox or used in a standard way—and this is where the real gain is locked in—and

F. Corea, *Big Data Analytics: A Management Perspective*,
Studies in Big Data 21, DOI 10.1007/978-3-319-38992-9_2

they may challenge the conventional wisdom, but they have to be proven and validated. In summary, smart data strategies always start from analyzing internal datasets, before integrating them with public or external sources. Do not store and process data just for data's sake, because with the amount of data being generated daily, the noise increases faster than the signal (Silver 2013). Pareto's 80/20 rule applies: the 80 % of the phenomenon could be probably explained by the 20 % of the data owned.

(ii) **If you want to do big data**, **you have to start big**. A good practice before investing heavily in technology and infrastructures for big data is to start with few high-value problems that validate whether big data may be of any value to your organization. Once the proof of concept demonstrates the impact of big data, the process can be scaled up.

(iii) **Data equals Objectivity**. First of all, data need to be contextualized, and their "objective" meaning changes depending on the context. Even though it may sound a bit controversial, data can be perceived as objective—when it captures facts from natural phenomena—or subjective—if it reflects pure human or social constructs. In other words, data can be *factual*, i.e., the ones that are univocally the same no matter who is looking at them, or *conventional/social*—the more abstract data, which earn its right to representativeness from the general consensus. Think about this second class of data as the notions of value, price, and so forth. It is important to bear this distinction in mind because the latter class is easier to manipulate or can be victim of a self-fulfilling prophecy. As stated earlier on, the interpretation of data is the quintessence of its value to business. Ultimately, both types of data could provide different insights to different observers due to relative problem frameworks or interpretation abilities (the so-called *framing effect*). Data science will therefore never be a proper science, because it will lack of full objectivity and full replicability, and because not every variable can be precisely quantified, but only approximated.

Let's also not forget that a wide range of behavioral biases that may invalidate the objectivity of the analysis affects people. The most common ones between both scientists and managers are: *apophenia* (distinguishing patterns where there are not), *narrative fallacy* (the need to fit patterns to series of disconnected facts), *confirmation bias* (the tendency to use only information that confirms some priors)—and his corollary according to which the search for evidences will eventually end up with evidences discovery—and *selection bias* (the propensity to use always some type of data, possibly those that are best known). A final interesting big data curse to be pointed out is nowadays getting known as the "Hathaway's effect": it looked like that where the famous actress appeared positively in the news, Warren Buffett's Berkshire Hathaway company observed an increase in his stock price. This suggests that sometime there exist correlations that are either spurious or completely meaningless and groundless.

(iv) **Your data will reveal you all the truth**. Data on its own are meaningless, if you do not pose the right questions first. Readapting what DeepThought says

> in *The Hitchhikers' Guide to the Galaxy* written by Adams many years ago, big data can provide the final answer to life, the universe, and everything, as soon as the right question is asked. This is where human judgment comes into: posing the right question and interpreting the results are still competence of the human brain, even if a precise quantitative question could be more efficiently replied by any machine.

The alternative approach that implements a random data discovery—the so-called "*let the data speak*" approach—is highly inefficient, resource consuming and potentially value-destructive. An intelligent data discovery process and exploratory analysis therefore is highly valuable, because "we don't know what we don't know" (Carter 2011).

The main reasons why data mining is often ineffective is that it is undertaken without any rationale, and this leads to common mistakes such as false positives, over-fitting, ignoring spurious relations, sampling biases, causation-correlation reversal, wrong variables inclusion or model selection (Doornik and Hendry 2015; Harford 2014). A particular attention has to be put on the causation-correlation problem, since observational data only take into account the second aspect. However, According to Varian (2013) the problem can be solved through experimentations.

In a similar fashion as in Doornik and Hendry (2015), it is here claimed the importance of the problem *formulation*, obtained leveraging theoretical and practical considerations and trying to spot what relationship deserves to be deepened further. The *identification* step instead tries to include all the relevant variables and effects to be accounted for, through both the (strictest) statistical methods and non-quantitative criteria, and verifies the quality and validity of available data. In the *analytical* step, all the possible models have to be dynamically and consistently tested with unbiased procedures, and the insights reached through the data interpretation have to be fed backward to improve (and maybe redesign) the problem formulation (Hendry and Doornik 2014).

Those aspects can be incorporated into a lean approach, in order to reduce the time, effort and costs associated to data collection, analysis, technological improvements and ex-post measuring. The relevance of the framework lies in avoiding the extreme opposite situations, namely collecting all or no data at all. The next figure illustrates key steps towards this lean approach to big data: first of all, business processes have to be identified, followed by the analytical framework that has to be used. These two consecutive stages have feedback loops, as well as the definition of the analytical framework and the dataset construction, which has to consider all the types of data, namely data at rest (static and inactively stored in a database), at motion (inconstantly stored in temporary memory), and in use (constantly updated and store in database). The modeling phase is crucial, and it embeds the validation as well, while the process ends with the scalability implementation and the measurement. A feedback mechanism should prevent an internal stasis, feeding the business process with the outcomes of the analysis instead of improving continuously the model without any business response (Fig. 2.1).

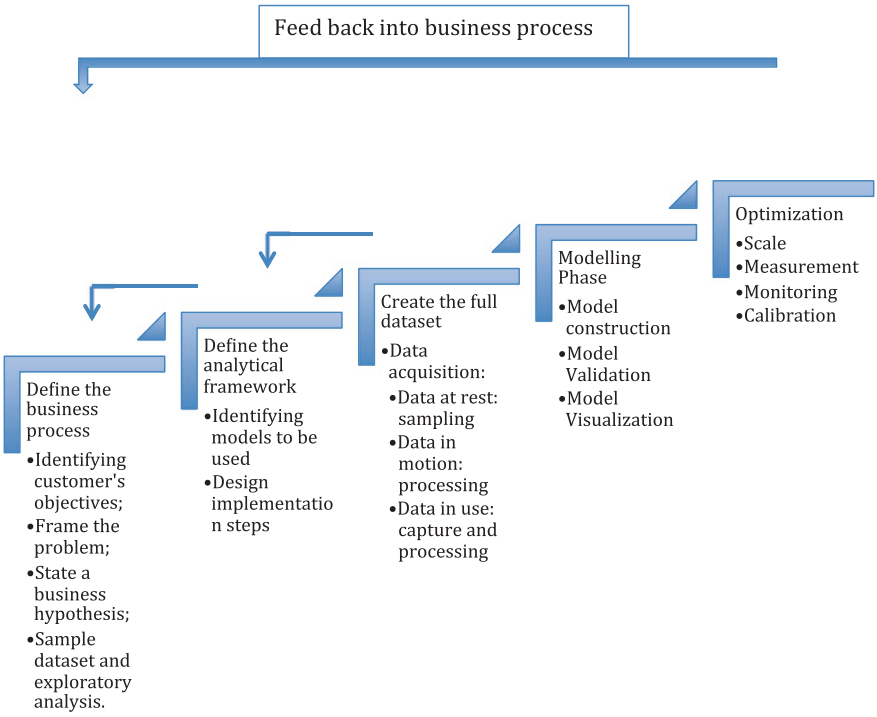

Fig. 2.1 Big data lean deployment approach

In light of all the considerations made so far, data science should then reach a compromise between the two approaches, because "*in medio stat virtus*": a specific problem should be tackled using a structured process, and an accurate question has to be posed at the beginning, but it is essential to be open and flexible to follow new unexpected paths and managing unanticipated consequences based on what data are telling. Big data are increasing the accuracy of predictions made, and enhancing the comprehension of many phenomena and human behavior. Ultimately, it appears to reduce the world complexity, providing an answer to any question posed. What is really going on instead is that they are providing multiple solutions, solutions that sometimes are so groundbreaking that they call for the question itself to be updated or amended. They disclose infinite new possibilities, which actually results in greater complexity—an intricacy that it is though manageable with a low change resiliency. Data science works as feedback-loop, and unlocking the data potential may involve a fully mind-shifting, which is important to be understood before embarking on it.

Data allow to grasp things that elude human's attention, but since they are not good or bad per se, they should never be blindly trusted. Data identification and interpretation is where the additional business value lies, and also how

the mistakes or frauds occur. Value-generation is a three-steps process: it results from first determining who are the recipients of data, then correctly enquiring, and finally providing user-friendly outcomes to the right audience (and sometime great visualizations are not useful for the sake of clarity), and translating those results into actionable points.

The four misconceptions about big data summarized above seem to be the most common traps for businesses moving into this area for the first time. Extra care is suggested when big data are approached in the first place. However, even for field veterans, implementing a successful data strategy may be cumbersome, due to a set of problems experienced at different levels (technical, business, or operational) and with different degrees of intensity. On top of everything the greatest complication is the cultural issue, and how C-level professionals perceived big data projects. The top management may indeed not be aware of the potential impact of data science for their business, so they have to be instructed through a proof of concept, i.e., a short, high-value, and low-resource-consuming internal project that can persuade them on incrementing the functional area of analytics. Moreover, a second imperative is the creation of a *golden record*, which is a unique and well-defined version of every data entity, and the identification of the correct technology and architecture. This is basically a theoretical matching issue, and it has to be thought as choosing the best unambiguous key in order to understand and validate one entity and separate it from other similar ones. In this respect, it is therefore essential to establish a solid internal data procedure, which has to consist of at least four main pillars: aggregation, integration, normalization, and finally validation, as summarized in the following figure (Fig. 2.2).

Data need to be consistently aggregated from different sources of information, and integrated with other systems and platforms; common reporting standards should be created—the master copy—and any information should need to be eventually validated to assess accuracy and completeness. Finally, assessing the skills and profiles required to extract value from data, as well as to design efficient data

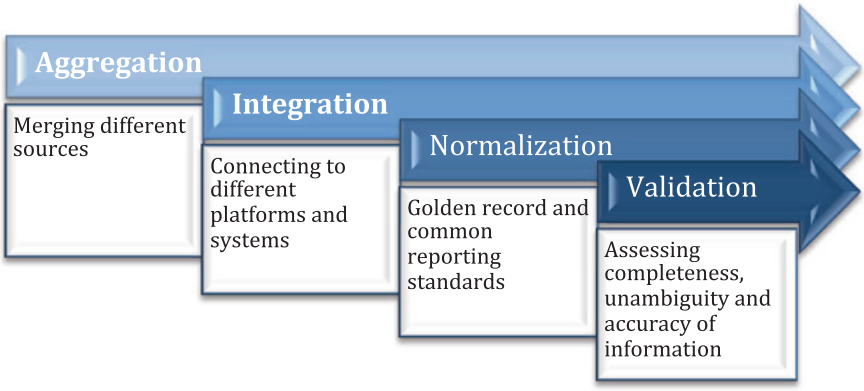

Fig. 2.2 Internal data management process

value chains and set the right processes, are two other essential aspects. Having a solid internal data management, jointly with a well-designed golden record, helps to solve the huge issue of *stratified entrance*: dysfunctional datasets resulting from different people augmenting the dataset at different moments or across different layers.

The answers to these problems are not trivial, and we need a frame to approach them. A *Data Stage of Development Structure* (DS2) is a maturity model built for this purpose, a roadmap developed to implement a revenue-generating and impactful data strategy. It can be used to assess the current situation of the company, and to understand the future steps to undertake to enhance internal big data capabilities.

Table 2.1 provides a four by four matrix where the increasing stages of evolution are indicated as *Primitive*, *Bespoke*, *Factory*, and *Scientific*, while the metrics they are considered through are *Culture*, *Data*, *Technology*, and *Talent*. The final considerations are drawn in the last row, the one that concerns the financial impact on the business of a well-set data strategy.

Stage one is about raising awareness: the realization that data science could be relevant to the company business. In this phase, there are neither any governance structures in place nor any pre-existing technology, and above all no organization-wide buy-in. Yet, tangible projects are still the result of individual's data enthusiasm being channeled into something actionable. The set of skills owned is still rudimental, and the actual use of data is quite rough. Data are used only to convey basic information to the management, so it does not really have any impact on the business. Being in this stage does not mean being inevitably unsuccessful, but it simply shows that the projects performance and output are highly variable, contingent, and not sustainable. The second Phase is the reinforcing: it is actually an exploration period. The pilot has proved big data to have a value, but new competences, technologies and infrastructures are required—and especially a new data governance, in order to also take track of possible data contagion and different actors who enter the data analytics process at different stages. Since management contribution is still very limited, the potential applications are relegated to a single department or a specific function. The methods used although more advanced than in Phase one are still highly customized and not replicable. By contrast, Phase three adopts a more standardized, optimized, and replicable process: access to the data is much broader, the tools are at the forefront, and a proper recruitment process has been set to gather talents and resources. The projects benefit from regular budget allocation, thanks to high-level commitment of the leadership team. Step four deals with the business transformation: every function is now data-driven, it is lead by agile methodologies (i.e., deliver value incrementally instead of at the end of the production cycle), and the full-support from executives is translated into a series of relevant actions. These may encompass the creation of a Centre of Excellence (i.e., a facility made by top-tier scientists, with the goal of leveraging and fostering research, training and technology development in the field), high budget and levels of freedom in choosing the scope, or optimized cutting-edge technological and architectural infrastructures, and all these bring a real impact on the revenues' flow. A particular attention has to be especially put on data lakes,

Table 2.1 Data stage of development structure

Drivers/stages	Primitive	Bespoke	Factory	Scientific
Culture	• No leadership support • Analytics as an IT asset • Conveying information (reporting, dashboard, etc.) • No budget • Descriptive analytics	• Leadership interest and midlevel management backing • Analytics used to understand problems • Specific application/department • Funding for specific project • Tailored modus operandi (not replicable) • Predictive analytics	• Leadership sponsorship • Analytics used to identify issues and develop actionable options • Alignment to the business as a whole • Specific budget for analytics function • Advanced data mining • Prescriptive analytics	• Full executive support • Data-driven business • Cross-department applications • Substantial infrastructural, human, and technology investments • Advanced data discovery • Automated analytics
Data	• Absence of a proper data infrastructure • Disorganized and dispersed silos • Duplicated information	• Data marts (lack of variety) • Internal structured data points • Data gaps or incomplete	• Virtual data marts • Internal and external data, • Mainly structured data • Easy-to-manage unstructured data (e.g., texts)	• Data lakes • Any data (unstructured, semi-structured, etc.) • Variety of sources (IoT, Social media, etc.) • Information life cycle in place
Technology	• Absence of data governance • No forefront technology (spreadsheet for reporting) • Low investments	• Integrated relational database (SQL) • Improvements in data architecture • Setting of a golden record • Scripting languages	• Pioneering technologies (Hadoop, MapReduce—see Appendix I) • Integration with programming languages • Visualization tools	• Centralized dataset • Cloud storage • Mobile applications • APIs, internet of things, and advanced machine learning tools
Talent	• Dispersed talents • Few people with few data analytical skills	• Mix of few full-time and some part-time data scientists • Proper data warehouse team • Strategic partnership for enhancing capabilities	• Well-framed recruitment process • Proper data science team • IT department fully formed and operative • Supporting of IT to data team	• Centre of excellence • Dominion experts and specialists • Training and continuous learning • Active presence within the Data Ecosystem
Impact	*No return on investments (ROI)*	*Moderate revenues, that justify though further investments*	*Significant revenues*	*Revolutionized business model (blue ocean revenues)*

repositories that store data in native formats: they are low costs storage alternatives, which supports manifold languages. Highly scalable and centralized stored, they allow the company to switch without extra costs between different platforms, as well as guarantee a lower data loss likelihood. Nevertheless, they require a metadata management that contextualizes the data, and strict policies have to be established in order to safeguard data quality, analysis, and security. Data have to be correctly stored, studied through the most suitable means, and to be breach-proof. An information life cycle has to be established and followed, and it has to take particular care of timely efficient archiving, data retention, and testing data for the production environment.

A final consideration has to be spared about cross-stage dimension "culture". Each stage has associated a different kind of analytics, as explained in Davenport (2015). Descriptive analytics concerned what happened, predictive analytics is about future scenarios (sometime augmented by diagnostic analytics, which investigates also the causes of a certain phenomenon), prescriptive analytics suggests recommendations, and finally automated analytics are the ones that take action automatically based on the analysis' results.

Some of the outcomes presented so far are summarized in Fig. 2.3. The following chart shows indeed the relation between management support for the analytics function and the complexity and skills required to excel into data-driven businesses. The horizontal axis shows the level of commitment by the management (high vs. low), while the vertical axis takes into account the feasibility of the project undertaken—where feasibility is here intended as the ratio of the project complexity and the capabilities needed to complete it. The intersection between feasibility of big data analytics and management involvement divides the matrix into four quarters, corresponding to the four types of analytics. Each circle identifies one of the four stages (numbered in sequence, from I—*Primitive*, to IV—*Scientific*). The size of each circle indicates its impact on the business (i.e., the larger the circle, the higher the ROI). Finally, the second horizontal axis keeps track of the increasing data variety used in the different stages, meaning structure, semi-structured, or unstructured data (i.e., IoT, sensors, etc.). The orange diagonal represents what kind of data are used: from closed systems of internal private networks in the bottom left quadrant, to market/public and external data in the top right corner.

Once the different possibilities and measurements have been identified—in the Appendix II, a Data Science Maturity Test (DMST) is provided, and it can be used to understand what stage a firm belongs to—it would be also useful to see how a company could transition from one level to the next. In the following figure (Fig. 2.4) some recommended procedures have been indicated to foster this transition.

In order to smoothly move from the *Primitive* stage to the *Bespoke*, it is necessary to proceed by experiments run from single individuals, who aim to create proof of concepts or pilots to answer a single small question using internal data. If these questions have a good/high value impact on the business, they could be acknowledged faster. Try to keep the monetary costs low as possible (cloud, open source, etc.), since the project will be already expensive in terms of time and

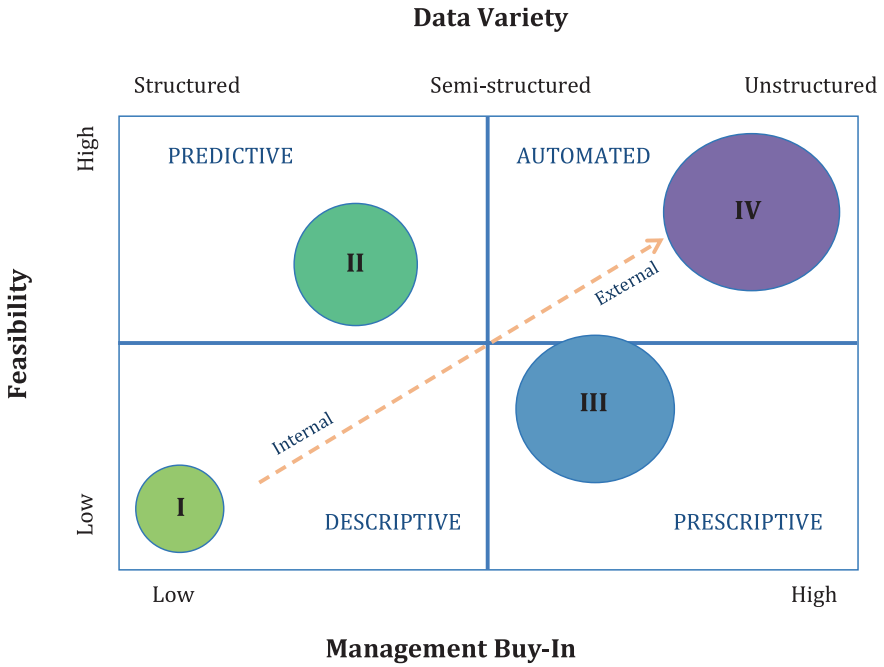

Fig. 2.3 Big data maturity map

manual effort. On a company level, the problem of data duplication should be addressed. The transition from *Bespoke* to *Factory* instead demands the creation of standard procedures and golden records, and a robust project management support. The technologies, tools, and architecture have to be experimented, and thought as they are implemented or developed to stay. The vision should be medium/long term then. It is essential to foster the engagement of the higher-senior management level. At a higher level, new sources and type of data have to be promoted, data gaps have to be addressed, and a strategy for platforms resiliency should be developed. In particular, it has to be drawn down the acceptable data loss (*Recovery Point Objective*), and the expected recovered time for disrupted units (*Recovery*

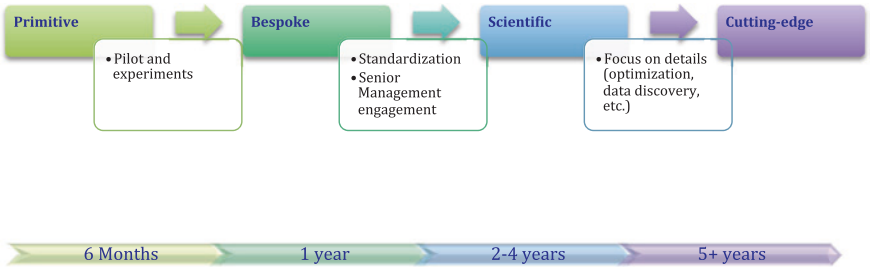

Fig. 2.4 Maturity stage transitions

Time Objective). Finally, to become data experts and leaders and shifting to the *Scientific* level, it is indispensable to focus on details, optimize models and datasets, improve the data discovery process, increase the data quality and transferability, and identify a blue ocean strategy to pursue. Data security and privacy are essential, and additional transparency on the data approach should be released to the shareholders. A Centre of Excellence (CoE) and a talent recruitment value chain play a crucial role as well, with the final goal to put the data science team in charge of driving the business. The CoE is indeed fundamental in order to mitigate the short-tem performance goals that managers have, but it has to be reintegrated at some point for the sake of scalability. It would be possible now to start documenting and keeping track of improvements and ROI. From the final step on, a process of continuous learning and forefront experimentations is required to maintain a leadership and attain respectability in the data community.

In Fig. 2.4 it has also been indicated a suggested timeline for each step, respectively up to six months for assessing the current situation, doing some research and starting a pilot; up to one year for exploiting a specific project to understand the skills gap, justify a higher budget allocations, and plan the team expansion; two to four years to verify the complete support from every function and level within the firm, and finally at least five years to achieving a fully-operationally data driven business. Of course the time needed by each company varies due to several factors, so it should be highly customizable.

Few more words should be spent regarding the organizational home (Pearson and Wegener 2013) for data analytics. We claimed that the Centre of Excellence is the cutting-edge structure to incorporate and supervise the data functions within a company. Its main task is to coordinate cross-units activities, which embeds: maintaining and upgrading the technological infrastructures; deciding what data have to be gathered and from which department; helping with the talents recruitment; planning the insights generation phase, and stating the privacy, compliance, and ethic policies. However, other forms may exist, and it is essential to know them since sometimes they may fit better into the preexisting business model (Fig. 2.5).

The figure shows different combinations of data analytics independence and business models. It ranges between business units (BUs) that are completely independent one from the other, to independent BUs that join the efforts in some

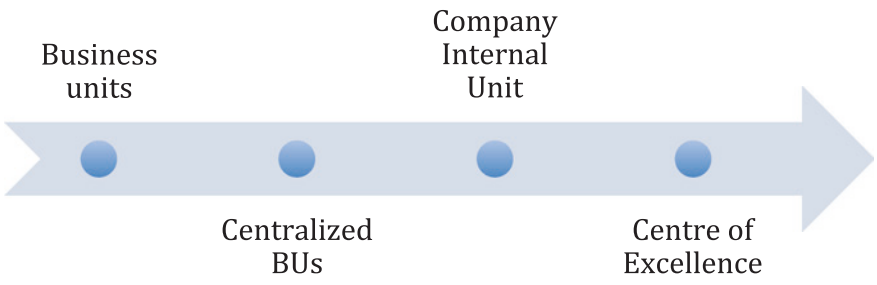

Fig. 2.5 Data analytics organizational models

specific projects, to an internal (corporate center) or external (center of excellence) center that coordinates different initiatives.

In spite of everything, all the considerations made so far mean different things and provide singular insights depending on the firm's peculiarities. In particular, the different business life cycle phase in which the company is operating deeply influences the type of strategy to be followed, and it is completely unrelated to the maturity data stage to which they belong (i.e., a few months old company could be a *Scientific* firm, while a big investment bank only a *Primitive* one).

Hence, there are at least two different approaches that have to be analyzed, i.e., the *prospective* and the *retrospective* one. The prospective concerns mainly start-ups, i.e., companies that are running on since few time and that are not producing huge amount of data (yet). They will begin producing and gathering data though, so it is extremely relevant to set an efficient data strategy form the beginning. The second case instead, the retrospective, is about existing businesses that are overwhelmed by data, and they do not know how to use them—or they may have specific problems, as centralized integration. It is clear the difference between those two circumstances, and it is then needed to explore it further.

Firstly, a startup is completely free from any predetermined structure, and it can easily establish a strong internal data policy from the beginning adopting a long-term vision, which would prevent any data related issue in the future. This is a matter to not be underestimated, and it requires an initial investment of resources and time: if the firm does it once and well, it will get rid of a lot of inconveniences and bothers. A well-set data policy would indeed guarantee a lean approach for the startup throughout any following stage. Moreover, young companies are often less regulated, both internally (i.e., internal bureaucracy is lower) and externally (i.e., compliance rules and laws). They do have a different risk appetite, which pushes them to experiment and adopt forefront technologies. Nonetheless, they have to focus on quality data rather than quantity data to start with.

A mature company instead usually faces two main issues: they have piles of data, and they do not know how to use them, or on the other side they have the data and a specific purpose in mind, but they cannot because of poor data quality, inadequate data integration, or shortage of skills. In the first case, they are in the *Primitive* stage, meaning that they have data but no clue on how extracting value from them. Since big institutions usually have really tight and demanding job roles, it is sometime impossible to innovate internally—in other words, they are "*too busy to innovate*". Some sector is more affected by this problem (fintech for instance) with respect to others (biopharma industry), as showed in Corea (2015). They should then hire a business idea generator in a first place, meaning an experienced high-level individual who can provide valuable insights even without owing a strong technical background, and afterwards a proper data scientist (or outsourcing the analysis phase). The second scenario, the one in which the data are present but useless, mainly two solutions can be adopted for each of the problems above identified: either the firm implements from scratch a new platform-team-culture, or it outsources it to intermediaries. Whereas in the first case the marginal utility has to compensate for the implementation and running costs of the new platform

and the salaries for the new employees, using specialized startups, universities, or expert consultants could be quite useful due to their high specialization and flexibility. The first case is also challenging because sometime the ability to assess data is fairly poor, and the database are so disorganized and low-quality that the decision whether to invest a lot of money in something that can—but also cannot with a good probability—have a return in five years time is really terrifying.

When it comes to choose whom to outsource to, the universities often represent a preferred avenue by big corporations: universities always need funding and above all data for running their studies (and publishing their works). They cost far less than startups, they have a good poll of brains, time, and willingness to analyze messy datasets, and universities that pursue pure theoretical research can be integrated by real impactful business questions. Startups instead are revenue-generating entities, and by definition they will cost more to big incumbents, but they often gather the best minds and talents with good compensation packages and interesting applied researches that universities cannot always offer. In both cases, the biggest issue is anyway about data security, confidentiality, and privacy: what data does the company actually outsource, how the third parties keep the data secured, how do they store them, and the HiPPO (i.e., *highest paid person's opinion*) concern, are the most common issues to deal with. Another relatively new and interesting way for big corporations to get some analysis for free are meetups and hackathons, that can be exploited as window-dressing for the firm, and in the meantime used to get good analysis and insights virtually for free.

The alternative to the outsourcing scenario is the buy-in mentality, which looks at buying and integrating (horizontally or vertically) anything that the company does not develop in-house. It is definitely more costly than other options, but it solves all the problems related to data privacy and security. Incubators and accelerators can offer a substitute way to invest less in more companies of interests that deal with several useful subjects without fully buying many companies, and this is why it is becoming so popular nowadays. The disadvantage of this fragmented investment business however is that new companies have a high-risk of failing—and the big firm loses not only the amount invested but also business opportunities and competitiveness—and companies need to invest in a team to select and support the on-boarded ventures.

Hence, it could be useful to integrate all the solutions provided so far and identify a solution in the middle. What we propose here is a two-steps approach: in the first phase, universities can be used to run a pilot, or the first two to three worthy projects that can drive the business from a *Primitive* stage to a *Bespoke* one. Then, the results are used to persuade management to invest into data analytics. The perfect hybrid option would be to create an internal data analytics center that is completely both physically and administratively disconnected from the main company. Hence, in a different building, the team should be run as it was a proper startup, and has to be charged by fully autonomy and freedom of means and thinking.

The conclusions of this chapter are drawn by final note for big corporation and their data approach, as well as a list of reasons of why big data projects may fail.

It is not clear when a company should start worrying about switching or going for a big data strategy. Of course there is not a unique standard answer, because the solution is tightly related to business specificities, but broadly speaking it is necessary to start thinking about big data when every source of competitive advantage is fading away or slowing down, i.e., when the growth of revenues, clients acquisitions, etc., reaches a plateau. Big data are drivers of innovation, and this approach could actually be the keystone to regain a competitive advantage and to give new nourishment to the business. However, it should be clear by now that this is not something that may happen overnight, but it is rather a gradual cultural mind-shift that requires many small steps to be undertaken.

Concerning how and why a big data projects may fail, there could be several different reasons. There are though some more commons mistakes made by companies trying to implement data science projects. It happens often indeed that the scope is inaccurate because of lacking of proper objectives or too high ambitions. On the other hand, the excessive costs and time employed in developing efficient project result from high expectations as well as absence of scalability. Managing correctly expectations and metrics to measure the impact of big data into the business is essential to succeed in the long term.

References

Carter, P. (2011). *Big data analytics: Future architectures, Skills and roadmaps for the CIO*. IDC White Paper. Retrieved from http://www.sas.com/resources/asset/BigDataAnalytics-FutureArchitectures-Skills-RoadmapsfortheCIO.pdf.

Coase, R. H. (2012). *Essays on economics and economists*. University of Chicago Press.

Corea, F. (2015). What Finance Can Learn from Biopharma industry: an innovation models transfer. *Expert Journal of Finance, 3*, 45–53.

Davenport, T. H. (2015). The rise of automated analytics. *The Wall Street Journal*, January 14, 2015. Retrieved October 30, 2015 from http://www.tomdavenport.com/wp-content/uploads/The-Rise-of-Automated-Analytics.pdf.

Doornik, J. A., & Hendry, D. F. (2015). Statistical model selection with big data. *Cogent Economics & Finance, 3*, 1045216.

Harford, T. (2014). *Big data: Are we making a big mistake*? Financial Times. Retrieved from http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#ixzz2xcdlP1zZ.

Hendry, D. F., & Doornik, J. A. (2014). *Empirical model discovery and theory evaluation*. Cambridge, Mass.: MIT Press.

Pearson, T., & Wegener, R. (2013). *Big data: the organizational challenge*. Bain & Company White paper.

Silver, N. (2013). *The Signal and the Noise: The Art and Science of Prediction*. Penguin.

Varian, H. (2013). Beyond *big data. NABE annual meeting*. San Francisco, CA, September 10th, 2013.

http://www.springer.com/978-3-319-38991-2

Big Data Analytics: A Management Perspective

Corea, F.

2016, XIII, 48 p. 7 illus. in color., Hardcover
ISBN: 978-3-319-38991-2