NUS Centre for Banking & Finance Law FG/02

Why more is more – the benefits of requiring publicly listed companies to disclose raw data[1]

Florian GAMPER

Research Associate, Centre for Banking & Finance Law,
Faculty of Law, National University of Singapore

florian@nus.edu.sg

22 April 2016

The views expressed in this paper are those of the author(s). They do not necessarily represent or reflect the views of the Centre for Banking & Finance Law (CBFL), or of the National University of Singapore.

© Copyright is held by the author(s) of each CBFL Working Paper. The CBFL Working Papers cannot be republished, reprinted, or reproduced in any format (in part or in whole) without the permission of the author(s).

The Centre for Banking & Finance Law (CBFL) at the Faculty of Law, National University of Singapore focuses broadly on legal and regulatory issues relating to banking and financial services. It aims to produce research and host events of scholarly value to academics as well as of policy relevance to the banking and financial services community. In particular, CBFL seeks to engage local and international banks, lawyers, regulators and academics in a regular exchange of ideas and knowledge so as to contribute towards the development of law and regulation in this area, as well as to promote a robust and stable financial sector in Singapore, the region and globally. http://law.nus.edu.sg/cbfl/

[1] First presented as a working paper on 29 January 2016

Why more is more – the benefits of requiring publicly listed companies to disclose raw data

Florian GAMPER

Research Associate, Centre for Banking & Finance Law, Faculty of Law, National University of Singapore

22 April 2016

ABSTRACT:

Big data analytics increasingly impacts all aspects of modern life, with one notable exception, namely the disclosure regime for public companies, which has remained remarkably constant since the 1930s. The current disclosure regime emphasises the materiality of information and making information more intelligible. The idea behind the system is to mandate public companies to only disclose information that is material to the average investor in order to prevent an information overload – the idea is to have a system of 'high value low volume information'. This article will argue for the opposite, for 'low value high volume information', namely a system where more data is disclosed but much of that data is not necessarily important or easy to understand.

This article proposes a disclosure regime in which, in addition to the current disclosure obligations, public companies also have to disclose the raw data used to compile the reports. This small change has the potential to have far reaching consequences, in particular it should lead to more accurate share prices and less resources being spend on 'creative' accounting. The reasons for this (i) big data analytics can be applied more easily and (ii) competitive forces can be applied to the process of turning data into information. The price to pay for this improved system is that there will no longer be a level playing field among investors. As only a minority of market participants will have the necessary know-how to analyse the disclosed data. However, the article argues that this is not only a price worth paying, it is actually beneficial.

Why more is more – the benefits of requiring publicly listed companies to disclose raw data

Florian Gamper[1]

ABSTRACT

Big data analytics increasingly impacts all aspects of modern life, with one notable exception, namely the disclosure regime for public companies, which has remained remarkably constant since the 1930s. The current disclosure regime emphasises the materiality of information and making information more intelligible. The idea behind the system is to mandate public companies to only disclose information that is material to the average investor in order to prevent an information overload – the idea is to have a system of 'high value low volume information'. This article will argue for the opposite, for 'low value high volume information', namely a system where more data is disclosed but much of that data is not necessarily important or easy to understand.

This article proposes a disclosure regime in which, in addition to the current disclosure obligations, public companies also have to disclose the raw data used to compile the reports. This small change has the potential to have far reaching consequences, in particular it should lead to more accurate share prices and less resources being spend on 'creative' accounting. The reasons for this (i) big data analytics can be applied more easily and (ii) competitive forces can be applied to the process of turning data into information. The price to pay for this improved system is that there will no longer be a level playing field among investors. As only a minority of market participants will have the necessary know-how to analyse the disclosed data. However, the article argues that this is not only a price worth paying, it is actually beneficial.

Key words: Disclosure regime, big data, securities regulation

[1] National University of Singapore – Centre for Banking & Finance Law; email: lawfg@nus.edu.sg; Tel: +65 6516 1105. This article is still in draft form. In particular, it does not yet contain references for all the prepositions taken from the literature. It must not be cited for any purpose.

Contents

1. Introduction 3
2. The current disclosure regime – depiction of reality rather than transfer of data 4
3. Background to big data 7
   a. Could technology handle the amount of data generated if raw data is disclosed? 7
   b. Big data is more than just a lot of data 8
   c. Data is an asset class 10
4. Description of mandatory disclosure of raw data regime 10
5. Benefits of raw data disclosure 13
   a. Application of big data 13
   b. Applying competitive forces to accounting 14
   c. Less creative accounting 17
6. Objections to disclosing raw data 18
   a. Level playing field among investors not guaranteed 18
   b. Facilitation of trading 19
   c. Information is asymmetric even under current disclosure regime 20
   d. 'Uniformed traders' are not necessary for the sufficient provision of capital 21
   e. Equal access to information is not essential for liquidity 22
   f. Unequal access to information does not exclude laypeople from the profits of capital markets 22
   g. Inefficient accumulation of private information 23
   h. Raw data cannot be disclosed because most of it is confidential 24
7. Does the free market produce enough incentives for companies to disclose raw data? 25
8. Does disclosure of raw data mean that insider trading should be legalised? 28
9. Conclusion 29

1. Introduction

The basic approach of the regulatory regime for public stock markets has essentially remained the same since the introduction Securities Act of 1933 and the Securities Exchange Act of 1934, and this is remarkable. The key constant in the regime is that publicly listed companies must disclose information about themselves. In practice, this means that a public company needs to collect data, organise that data in a meaningful way and disclose the result to the public. For a long time, the only real questions asked of disclosure regimes was how much information should be disclosed, and how to make disclosure more intelligible (i.e. In order to avoid investors being overwhelmed by information, how to ensure (1) that only material information is disclosed, and (2) that information is easy to understand).[2] This unchanged approach is all the more remarkable considering that data processing capabilities today are dramatically different from the 1930s. This is especially apparent in the relatively recent advent of big data, which appears to fundamentally change the way data is being used. Nonetheless, the disclosure regime for public markets has proved remarkably resilient to any change.

This article will propose a disclosure regime which is the exact opposite of the current trend in disclosure. Currently there is a trend to advocate a regime with "lower volume, but an overall higher quality […] disclosure"[3], this article proposes a regime in which raw data must be disclosed i.e. a 'higher volume, but overall lower quality' regime. In this article, such a regime will be referred to as "disclosure of raw data". The idea is that, in addition to the current required disclosure (e.g. publications of annual reports with audited accounts), public companies must also disclose the raw data used in producing these reports. While this may appear to be a small change to the current regime, this article will argue that this change has the potential for profound implications. The primary benefit of disclosing raw data is that it will lead to more accurate stock prices, i.e. the stock price that reflect the discounted future value of a company more accurately. The reason for this are twofold (1) market forces could operate on the level of turning data into information (previously only the issuer could do this), and (2) disclosure of raw data lends itself more easily to big data analytics. An additional benefit is that the incentives for manipulating accounts through creative accounting will be diminished. Yet, in order to achieve a disclosure of raw data regime, it may be necessary to abandon the idea that retail investors investing directly in the stock markets is something desirable. This is because retail investors are unlikely to have access to the technology and know-how required to analyse raw data. Nonetheless, retail investors could still share in the profits of the stock market through investing in intermediaries.

One significant drawback of the disclosure of raw data is that it may lead to a socially inefficient accumulation of private information. This is an issue that requires further assessment. The aim of this article, however, is not to give a detailed assessment of the pros and cons of disclosing raw data. Nor is the aim of this article to give a detailed account of how a raw data disclosure regime would work. The aim

---

[2] Materiality of disclosure and easiness to understand seem to be particular hot topics. For instance, former as former SEC Commissioner Daniel M. Gallagher notes: "Justice Thurgood Marshall warned almost 40 years ago, disclosure requirements with "unnecessarily low" materiality standards risk "simply bur[ying] the shareholders in an avalanche of trivial information—a result that is hardly conducive to informed decision making." [6] When investors are inundated with immaterial information, it increases the likelihood that they will miss key disclosures." In Daniel M. Gallagher, "The Importance of the SEC Disclosure Regime" , Harvard Law School Forum on Corporate Governance and Financial Regulation, http://corpgov.law.harvard.edu/2013/07/16/the-importance-of-the-sec-disclosure-regime/ (accessed Oct 22, 2015)

[3] Ibid

of this article is rather more modest – it is to suggest that further research in this area is worthwhile by providing a rough framework of how a raw data disclosure regime could look, and by pointing out that a cursory analysis suggests that there are significant benefits associated with a raw data disclosure regime.

This article is structured as follows. Part 2 provide an overview of the current disclosure regime. Part 3 gives a short overview of the relevant aspects of big data. Part 4 provide some details on how a raw data disclosure regime could look like. Part 5 discusses the benefits of the proposed regime. Part 6 discusses objections to it. Part 7 analyses whether the free market alone would 'produce' enough disclosure of raw data. Part 8 analyses whether this means that insider trading should be legalised. Part 9 concludes the analysis.

2. The current disclosure regime – depiction of reality rather than transfer of data

On a general level "[s]ecurities regulation has two main subject areas: the regulation of the securities markets and the securities industry, and the regulation of corporate issuers and information about issuers"[4]. Nonetheless, disclosure of information play an especially important part in the current regime. According to former SEC Commissioner Daniel M. Gallagher "[t]he SEC is first and foremost a disclosure agency"[5]. Historically, the US securities law was created as a response to the events of the 1920s and 1930s. In particular, the goal was to "restore the public's trust in the securities market, which had been undermined by the excesses of the 1920s and the crash of 1929 and the early 1930s [through] the protection of small investors against the abuses perpetrated by knowledgeable insiders"[6]. The idea is that if all market participants have access to the same information, then an investor will not be able to take advantage of another investor (or at least it would be considerably more difficult). The SEC's basic method of achieving equality of information is through the requirement that "[i]nformation 'material' from the standpoint of a reasonable investor must generally be disclosed [by the relevant issuer]"[7]. In practice, this means that companies must periodically (1) file certain forms / reports with the SEC[8], (2) send reports and other communications to shareholders[9], and (3) make certain announcements to the public. This approach is by no means unique to the US, many other securities regulatory regimes follow a very similar approach.[10] On a more abstract level, the regime works as follows: (A) the regulator sets out the type of information which is material and relevant to investors, and (B) the issuer compiles

[4] Donald C. Langevoort, "The SEC, Retail Investors, and the Institutionalization of the Securities Markets", Virginia Law Review, Vol. 95, No. 4 (Jun., 2009), pp. 1025-1083, page 1027

[5] Supra note 2

[6] Luigi Zingales, "The Future of Securities regulation", Journal of Accounting Research, Vol. 47, No. 2, Regulation of Securities Markets: Perspectives from Accounting, Law, and Financial Economies (May, 2009), pp. 391-425, page 391

[7] Henry T.C. Hu, "Financial Innovation and Governance Mechanisms: The Evolution of Decoupling and Transparency" 70 Business Lawyer, 347 – 405 (2015), page 399

[8] For instance companies need to annually file Form 10-K, quarterly form 10-Q and current reports on Form 8-K

[9] Although form 10-K and annual reports may be combined into one documents

[10] See for instance the UK listing regime (http://www.fca.org.uk/firms/markets/ukla) and in particular the Disclosure and Transparency Rules (https://www.handbook.fca.org.uk/handbook/DTR/)

such information and discloses it to the market. Professor Hu notes an important point about this regime. He writes as follows:

> If viewed functionally, the disclosure philosophy [behind the disclosure regime] has, from the start, been implemented largely through a single informational strategy. An intermediary (for instance, the corporation issuing securities) would be required to craft a depiction of reality that met specified quality standards and content requirements and be required to make its depiction available to investors[11]

Hu calls this system "intermediary depiction"[12]. The basic idea is that the regulator is requiring issuers to form a view of what the reality of the company looks like and to disclose this view to the public. To illustrate the point it is useful to draw a distinction between 'data' and 'information'. The distinction between the two is sometimes characterised as follows:

> The term data refers to factual information, especially that [sic] used for analysis and based on reasoning or calculation. Data itself has no meaning, but becomes information when it is interpreted. Information is a collection of facts or data that is communicated.[13]

Although the above statement contains a slight circularity, it is a good illustration of the difference between data and information (NB, it should not be regarded as a definition of these two terms). It is Hu's assertion that the current regulatory system requires issuers to disclose information, and not data. An example of this is that public companies are required to disclose audited accounts.[14] Financial accounting can be defined as "[t]he process of recording, summarising and reporting the myriad of transactions from a business, so as to provide an accurate picture of its financial position and performance."[15] Accounts summarise data rather than transferring the underlying data. For instance, typical accounts report one figure for revenue, or even multiple revenue figures, however, they do not report each individual item sold. Thus, under the current disclosure regime it seems quite clear that, the issuer is required to report "a depiction of the pertinent aspects of [objective reality]"[16] (i.e. information) rather than data. However, why is this a problem, or should it matter at all? After all, when a company discloses it has sold 3 items, how is this different from saying the company sold 1 item + 1 item + 1 item? Hu finds two problems with the system of intermediary depiction:

---

[11] Henry T.C. Hu, "Too complex to depict? Innovation, "Pure Information" and the SEC disclosure paradigm" Texas Law review, 2012 page 1623

[12] Supra note 7 page 347

[13] At dictionary.com http://dictionary.reference.com/help/faq/language/d58.html (accessed Sep 28 2015)

[14] See for instance SEC "FORM 10-K - ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934 - GENERAL INSTRUCTIONS" https://www.sec.gov/about/forms/form10-k.pdf (accessed Oct 22, 2015)

[15] Investopedia, 'Financial Accounting' http://www.investopedia.com/terms/f/financialaccounting.asp (accessed 15 Oct 2015)

[16] Supra note 7 page 381

> i) Modern finance has made the world too complex to be appropriately captured through intermediary depiction;[17]
> ii) The intermediary itself may not adequately understand the reality which it tries to depict and therefore the intermediary's depiction of reality might be flawed.[18]

Even though these two aspects point towards the real problem with intermediary depiction, they are themselves somewhat inaccurate. On the first problem, Hu writes that:

> [m]odern financial innovation is creating objective realities that are far more complex than in the past, sometimes so complex that they are beyond the capacity of the English language, accounting terminology, visual display, risk measurement, and other tool on which all intermediary depiction must primarily rely [19]

There are at least two problems with this claim. First, it seems factually wrong (or at least very unlikely) that financial innovations have become so complex that they can no longer be depicted effectively through a combination of language and mathematics. The combination of language and mathematics is capable of describing exceedingly complex realities. It seems quite a stretch to argue that the tools developed to describe complex realities like quantum mechanics fail in the case of modern finance.[20] Further, the argument that disclosure of underlying data is necessary because of increased complexity of finance puts emphasis in the wrong place. The world before modern finance was not less complex– the world is and has always been an extraordinarily complex place. Even before the advent of derivatives, companies faced an extraordinarily difficult calculation of risks, for instance, a farming business may be exposed to a new kind of pest, a mining company to unknown geological features, pharmaceutical company to value new research, and so on. Derivatives positions may appear more 'imminent' than other risks faced by companies (e.g. a farming business may be completely unaware of the risk of its harvest being wiped out by a pest), however, that does not make these risks less real. Intermediary depiction is not a problem resulting from the complexities of modern finance. Intermediary depiction has some significant shortcomings regardless of the prevailing economic / financial systems.

The second problem of intermediary depiction, highlighted by Hu, points in the right direction. However, by framing the problem as a 'misunderstanding' it again puts the emphasis in the wrong place. Hu's formulation gives the impression that there is such a thing as a correct depiction of reality. Regardless of any philosophical consideration of whether this is actually possible, the point is to create a system where each investor turns data into the information that the investor believes to be relevant.

---

[17] Supra note 7 page 384 and subsequent
[18] Supra note 7 page 385 and subsequent
[19] Supra note 7 page 384
[20] It is, of course, true that it takes a lot of training to understand quantum mechanics, and that most 'depictions' of it are utterly incomprehensible to the layperson. However, this is a problem of how to make depictions of complex realities intelligible to a wider audience, and not whether it is possible to depict these realities. It is also arguably true that quantum mechanics may not actually depict reality. Nonetheless, regardless whether it does or does not actually depict actually reality, quantum mechanics has the potential to depict a complex reality

The problem with the current disclosure regime is that only the issuer, has access to the raw data[21]. Therefore, only one entity is able to turn data into information. If data was open to all investors, each investor could turn it into information. Presumably competitive forces will, over time, develop a superior depiction of reality derived from all each investor's depiction of reality (how this process will work is described more clearly in part 5 of this article). 'Superior' in this context means a depiction of reality which more accurately reflects the fundamental value of the company, not an objective truth. In short, the problem with the current system of mandatory disclosure is that the process of turning data into information is not subjected to market forces but only to regulatory pressure.

3. Background to big data

In order to assess the idea of disclosing raw data, it is advantageous to provide some background on what is commonly referred to as 'big data'. This section argues for three prepositions: first, it will be suggested that from a technological point of view, it is currently possible, or near-possible, to handle the large amount of data that would be created by a disclosure of raw data regime. Second, it will be argued that big data employs a specific method of data analysis and that disclosure of raw data lends itself much more naturally to big data methodology than intermediary depiction. Third, it will be argued that data is best thought of an asset class in its own right.

a. Could technology handle the amount of data generated if raw data is disclosed?

Mandatory disclosure of raw data would generate a vast amount of data. Therefore, a natural objection to a raw data disclosure regime is that it is impossible to handle this much data, and so, regardless of the benefits, such a regime is irrelevant because the amount of data generated could simply not be processed. Although it is difficult to estimate the actual amount of data that would be generated through such a regime, it seems likely that the technology for handling this amount of data already exists or is likely to exist in the near future. However, this article is not about technology (and the author does not claim any particular expertise therein). It may be that handing this amount of data is beyond what is technically feasible, however, the following arguments merely suggest that one should not dismiss the mandatory disclosure of raw data out of hand due to purely technological concerns.

Our ability to handle large amounts of data has dramatically improved in the recent past. According to Hal Varian "[b]etween the dawn of civilization and 2003, five exabytes of data was created; now that amount is generated every two days"[22]. Further, our ability to process data has also increased dramatically. Hilbert and López suggest that "global storage capacity grew at an annual rate of 23 percent over that period (to more than 290 exabytes in 2007 for all analog and digital media), [while] general-purpose computing capacity, a measure of the ability to generate and process data,

[21] Please see section 4 of this article for the discussion on 'raw data'

[22] Richard Topham, "Data is new raw material of business – almost on a par with capital, labour", available at https://www.experianplc.com/media/news/2012/data-is-new-raw-material-of-business-almost-on-a-par-with-capital-labour/ (accessed Sep 29 2015

grew at a much higher annual rate of 58 percent."[23] According to Dean and Ghemawat in 2009 Google alone processed more than 20 petabytes of data per day[24]. Facebook is estimated to process more than 500 terabytes per day.[25] All of this does, of course, not show that the amount of data generated by mandatory disclosure of raw data could be handled. However, it does suggest that the world is already processing a vast amount of data. Regulators are often concerned that investors cannot handle lots of data[26]. However, although it is certainly true that not all investor can handle a large amount of data, some enterprises can handle a very large amount of data indeed. This at the very least suggests that further research into mandatory disclosure of raw data is justified.

b. Big data is more than just a lot of data

Another important aspect of big data is to distinguish it from 'just a lot of data'[27]. Big data analytics uses a methodology quite different from traditional data analytics. According to Williams "the first main difference with Big Data is that we no longer have to rely on sampling to determine the likely outcome of a population".[28] The difference is probably best illustrated by comparing to traditional (non-big data) statistics. In a traditional statistical analysis, a hypothesis is created which is then tested against data sampled from a population. However, big data analytics aims to sidestep the concept of sampling, and instead directly analyses the entire population. This means that sophisticated statistical methodology to draw conclusion from samples to the population would not be required anymore, one can simply use the entire population.[29] Another difference can be expressed as follows – big data puts data before hypothesis. In a traditional statistical analysis, the hypothesis is created first and tested against the available data. Whereas, in big data analytics data is often

---

[23] Martin Hilbert and Priscila López "The world's technology capacity to store, communicate, and compute information," Science, April 2011, Volume 332, pages 60-65, in Bauer, Ranade and Tandon "Big data and the opportunity it creates for semiconductor players" available at file:///C:/Users/lawfg/Downloads/Big_data_and_the_opportunities_it_creates_for_semiconductor_players.pdf (access Sep 29, 2015)

[24] Miller Rich, Data Knowledge Centre, http://www.datacenterknowledge.com/archives/2008/01/09/google-processing-20-petabytes-a-day/ (accessed Oct 16, 2015)

[25] Tam Donna, "Facebook processes more than 500 TB of data daily",http://www.cnet.com/news/facebook-processes-more-than-500-tb-of-data-daily/ (accessed Oct 16, 2015)

[26] Former SEC Commissioner Gallagher writes: "When investors are inundated with immaterial information, it increases the likelihood that they will miss key disclosures. Even more likely is the possibility that investors, despairing about the voluminous compilations of corporate minutiae contained in company filings, will never even look at disclosure documents. In either case, the result is that investors are left less informed when making investing decisions than they would be if presented with a document that didn't require a magnifying glass to read and a PhD to understand. The irony that the vast expansion of the Commission's mandatory disclosure regime may help incentivize investors to throw their hands up and simply ignore company filings is not lost on me or, I'm sure, all of you."

[27] Williams, David, "If 'Big Data' Simply Meant Lots of Data, We Would Call It 'Lots of Data" in Forbes, http://www.forbes.com/sites/davidwilliams/2012/09/19/if-big-data-simply-meant-lots-of-data-we-would-call-it-lots-of-data/ (accessed Oct 16 2015)

[28] Ibid.

[29] For instance, Williams phrases it like this: "Instead of summarizing a population into finite segments to drive differentiated treatments and relevance, we can go down to a much finer level of granularity, ultimately the individual level" Ibid

collected first, and thereafter interesting correlations are observed in the data.[30] An example may help to illustrate this point. According to the Economist:

> In 2004 Wal-Mart peered into its mammoth database and notice that before a hurricane struck, there was a run on flashlights and batteries, as might be expected; but also on Pop-Tarts, a sugary American breakfast snack. On reflection it is clear that the snack would be a handy thing to eat in a blackout, but the retailer would not have thought to stock up on it before the storm.[31]

In summary: big data analytics aims to collect as a much data as possible, without postulating whether or not the data collected is useful. After this collection, big data analytics attempts to find interesting correlations in the data, without necessarily claiming anything 'deeper' about the correlations discovered, e.g. causality. However, it is important to keep in mind, that there is no clear dividing line between traditional data analytics and big data analytics, rather it as a spectrum. One may be tempted to argue that, even if all of this is true, the distinction between big data and traditional statistical analysis may be of interest to statisticians and data scientists, but that it does not have wider implications. However, certain commentators have suggested that the shift from traditional data analytics to big data analytics may have very profound implications indeed. For instance, former WIRED magazine editor in chief Chris Anderson goes so far as to claim that big data could mean the end of the scientific method. He argues as follows:

> Scientists are trained to recognise that correlation is not causation […]. Once you have a model, you can connect the data sets with confidence. Data without a model is just noise. But faced with massive data, this approach to science — hypothesize, model, test — is becoming obsolete. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyse the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.[32]

This claim, is probably an exaggeration[33], however, it illustrates that the difference between the two approaches has the potential to have immense repercussions.

---

[30] For a more detailed discussion of this idea see for instance Mayer-Schönberger Viktor and Cukier Kenneth "Big Data: A Revolution That Will Transform How We Live, Work, and Think" 2014, Eamon Dolan/Mariner Books; Reprint edition (March 4, 2014)

[31] The Economist, "A different game – Information is transforming traditional business", Feb 25, 2010, http://www.economist.com/node/15557465

[32] Anderson Chris "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", Wired, 2008, available at http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory (accessed Sep 29, 2015)

[33] This author certainly does not believe that big data means the end of the scientific method as we know it

c. Data is an asset class

Big data is a raw material and data itself is increasingly viewed as its own asset class. For instance, the Economist writes "data is becoming a new type of raw material that's on par with capital and labour"[34], according to European Consumer Commissioner Meglena Kuneva personal data is the "new oil"[35], and the World Economic Forum believes that within personal data we are witnessing the "emergence of a new asset class"[36]. Not only does data appear to be an asset class / resource it appears to be rather valuable one. For instance, the McKinsey Institute estimates the value of big data to the US health care sector alone to be $300bn annually.[37] Thus, it seems to be quite clear that from a commercial point of view, data is an asset / raw material.[38] For most other assets / raw materials, the decision related to how they are to be produced and employed is left to the market and not regulators.[39] However, under the current disclosure regime regulators determine how data / information is produced and distributed. This, of course, does not in itself mean that the current disclosure regime needs to be changed. It merely suggest that one should ask the question why regulators, rather than the market, decide how to distribute this asset / raw material.[40] In the 1930s, information was simply viewed as one of the inputs into the investment decision making process. From this perspective, it may make sense to create a system which aims to allow each market participant to access the same information. However, if data is regarded as an asset / raw material, and transformation of data into information is a key aspect of the investment process, then one may be more inclined to argue for regime where data is disclosed rather than information.

4. Description of mandatory disclosure of raw data regime

Before it is possible to analyse the disclosure of raw data from a normative point of view, it is necessary to provide a rough sketch of what the disclosure of raw data actually entails. As mentioned in the introduction, this article merely aims to provide a rough outline and not provide all the details required of a raw data disclosure regime. On a general level the idea of disclosing raw data is quite simple. Hu puts this as follows:

> Advances in computer and web technologies now make for easier an approach more focused on the "transfer" of objective reality – or, more precisely, information that is highly mimetic of objective reality and exists independently of any observer […] the intermediary need not always stand between investor and objective reality, recounting to the

[34] The Economist, "Data, data, everywhere – a special report on managing information", 2010
[35] World Economic Forum "Personal Data: The Emergence of a New Asset Class", 2011 page 5
[36] Ibid.
[37] McKinsey Institute "Big data's potential for businesses" available at http://www.ft.com/intl/cms/s/0/64095dba-7cd5-11e0-994d-00144feabdc0.html#axzz3nBqJ5b5J (accessed Sep 30 2015)
[38] Although the law still finds it more problematic to regard data as an asset, see for instance the English case of Oxford v Moss (1979) 68 Cr App Rep 183 i
[39] Regulators, of course, have a role to play for instance if the use of a raw material is associated with externalities
[40] Although the law still finds it more problematic to regard data as an asset, see for instance the English case of Oxford v Moss (1979) 68 Cr App Rep 183

> investor what the intermediary sees. […] With the transfer mode, the entity is involved only with respect to mechanical task of, in effect, transmitting pertinent aspects of objective reality in the form of pure information. This information can be downloaded, observed, and analysed by market participants.[41]

Hu's description points towards what the raw data disclosure entails, however, this description needs to be more refined to make it meaningful. The first problem is that there is almost an infinite amount of raw data 'attached' to each piece of "objective reality". To illustrate, suppose a bakery is required to disclose raw data for its sales of bread. It could, for instance, disclose the identity and price of each item sold in a certain period. However, it could also disclose the date and the time the item was sold, the location of the shop where it was sold, the shelf where the item was stored, the name of the shop assistant who sold it, the name (or loyalty card number) of the customer who bought the item, etc. The list could be almost indefinitely extended. One could, in addition, require the bakery to disclose all its data, not just data in relation to its sales. The bakery could be required to disclose how many sick days each worker took over a certain period, what coffee they buy for the cafeteria, how often new flour is purchased, and so on. To require a company to disclose "objective reality" without limitations requires the company to collect and disclose an infinite (or almost infinite) amount of data. This is, of course, impossible. Thus, "objective reality" must be limited in some way to make the collection of data practical, and to make the concept meaningful.

There are at least three alternatives of what disclosure of raw data could mean in practice. Before these alternatives are discussed, it is important to state that throughout the discussion it will be assumed that disclosure of data will only be required where there are no laws that would (1) prevent such disclosure (e.g. privacy laws), or (2) provide the company a right not to disclose the data (e.g. intellectual property rights). Such laws could seriously limit the amount of data that could be disclosed. Nonetheless, it is important to emphasise that this article does not advocate a regime, which would override privacy or other concerns. Further, this article will only consider alternatives that do not require the company to collect data it would not otherwise collect (i.e. the regime advocated would not require companies to collect more data than they already do). For instance, if a company does not collect data on the height of its employees, then a proposal that would require a company to collect this data will not be considered by this article. The reason for limiting the analysis in this way is because it is assumed that it would be inefficient to require companies to collect additional, and potentially superfluous, data. However, this is only an assumption, further research may well indicate that collection of certain additional data is optimal.

The following is a list of potential options for the disclosure of raw data regime, arranged from a regime closest to the current regime (with the least amount of raw data disclosed), to the regime furthest from the current regime (with the most amount of data disclosed).

i) **Current disclosure regime as the starting point:** Under this regime a company would be required to disclose all data it used to compile the

[41] Supra note 6 page 388

reports required by the current disclosure regime. If a company is required to disclose its accounts, then, in addition to disclosing the accounts themselves, the company would also have to disclose all the data it used to compile the accounts.

ii) **Designate certain types of raw data:** Under this regime, the regulator would designate certain types of raw data a company would need to disclose (this could be varied by industry, company size, and so on), provided that the company collects that data in its normal course business.

iii) **All raw data a company has obtained, unless an exemption applies:** This regime simply requires public companies to disclose all data it has obtained, unless an exemption is granted. For example: an exemption could be the level of granularity for salary data (i.e. a company would not be required to disclose how much each employee is paid). Such exemptions would be a necessity, as few commercial enterprises could survive if all their data is disclosed. Of course, if too many exemptions are granted, this system would be very similar to the current disclosure regime. Nonetheless, the default position is that disclosure would be required for all data, and only under special circumstances could data be withheld.

The obvious advantage of regime iii) is that more data would be disclosed than in i) or ii). A major disadvantage of regime iii) is that it may impact a company's choice of data to collect. If every piece of data a company collects must be disclosed, this will be an additional variable for a company to add to its decision making process on data collection. Conversely, under alternative i) and ii), the requirement for the company to disclose raw data should have less of an impact on the amount of data the company decides to gather. This is because, under alternative i), the company must collect certain data because of its obligation to create regulatory reports. Under alternative ii), while the company's choice of data to collect may be influenced (e.g. if a regulator requires that a certain type of data needs to be disclosed, the company may decide not to collect that type of data), the impact is much less severe than regime iii). This is because the company knows what data requires disclosure. The downside of regimes i) and ii), is that regulators would have a large say in the decision what type of raw data is disclosed, instead of leaving it up to the market. This is especially true of regime ii), which would require regulators to develop an understanding of the type of data the market desires. While this is not impossible, it is a very difficult task.

Regime i) is the least 'radical' of the three, as it is closest to the current disclosure system. This regime merely requires the company to disclose the data it already uses to compile its regulatory reports. This regime has two advantages. First, its nearness to the current regime is more likely to be accepted by regulators and the industry. Second, there is less risk of unforeseen consequences. The change from the current system merely consists that disclosure previously done in summary only also to be disclosed in non-summary form, unless a specific exemption applied. An exemption could apply if data in non-summary form would result in significant harm to the issuer. For instance, an issuer may have no problem disclosing the total amount of salaries paid, however, may not want to disclose individual salaries as this could be

harmful to team harmony. Regime i) could also be treated as a stepping stone towards regimes ii) or iii). If it turns out that disclosure of raw data is a positive development, the model could be gradually extended. For the time being, however, option i) appears the safest and least disruptive. A word of caution, however, the details of the obligation to disclose all data used in compiling the regulatory reports still requires a lot more analysis, and there are many unanswered questions. However, one should not dismiss a regime merely because the questions surrounding it cannot immediately be answered. Perhaps the approach to disclosure of raw data should be thought of as a 'common law statute',[42] where courts and regulators are tasked with developing law along certain parameters.

In summary the regime proposed is one where the company still has to comply with all the current regulatory requirements. Additionally, the company needs to disclose the data and the calculations used to compile the regulatory reports, unless the data is privileged, for instance names of bank clients. Privileged data being limited from disclosure is potentially quite a serious limitation on the amount of data disclosed. However, it is topic for further research how exactly to commensurate objectives like bank confidentiality and the desire for greater openness.

5. Benefits of raw data disclosure

Disclosing raw data brings with it a number of distinct advantages, which will be discussed below. However, it is important to keep in mind that the below are initial thoughts only – some of the details are still missing.

a. Application of big data

That disclosing raw data lends itself more naturally to big data analysis than intermediary depiction may seem obvious, simply because more data is being disclosed. However, the more important point is that the structure of the current disclosure system is based on traditional data analysis. As discussed in part 3.b. of this paper, big data uses a methodology that puts ‘data before hypotheses’. Whereas, the current disclosure system is based on traditional data analysis, putting ‘hypotheses before data’.

In the current disclosure system regulators form a hypothesis that determines what is be best format for information to be disclosed, and which information is most useful for investors (e.g. revenue and costs are a good indicator of future company performance). Big data analytics would approach the problem the other way around. Rather than postulating which data is useful, it would collect the data first and, based on the data, determine what is useful. If raw data were disclosed, investors could derive an advantage by finding interesting, previously unknown, correlations in the data. Through this process, it would emerge which data is important and which is not. At the moment, investor are unable to use this process as they do not have access to the data. Instead, investors must rely on regulators to determine what data is significant.

It is important to emphasise that the distinction between the raw data disclosure regime and the current regime is not binary, but a spectrum. Data disclosed under the current system can still be used for big data analytics. One could, for instance, analyse all annual reports and look for interesting correlations. It could turn out that sales data is correlated with positive returns, or it turns out that the number of “;” used in an

[42] For a discussion of ‘common law statues’ see for instance Margaret H. Lemos, “Interpretive methodology and delegations to courts: Are “common law” statues different?” 2013, http://scholarship.law.duke.edu/faculty_scholarship/2545/ (accessed Oct 19, 2015)

annual reports correlates with positive returns. Big data analytics is already being applied to stock markets. Ruiz, Hristidis, Castillo and Gionis, measured "the correlation of the stock market events with these features, using Twitter as a data source."[43] They show that "that even relatively small correlations between price and micro-blogging features can be exploited to drive a stock trading strategy that outperforms other baseline strategies"[44]. Curmea, Preisb, Stanleya, and Moatb "find evidence of links between Internet searches relating to politics or business and subsequent stock market moves. In particular, […] an increase in search volume for these topics tends to precede stock market falls".[45] The problem with this kind of analysis is that the way big data analytics is applied to the stock market is in order to measure investor sentiment (e.g. as revealed by Twitter feeds), and not at the level of the fundamental value of the company. Using Keynes' analogy of the beauty contest[46], big data is presently applied to the stock market to find out what investors think the best investment is, not to what the best investment actually is. If one believes that there is some tendency for a company's stock market returns to be in line with its fundamental value, then raw data disclosure would help to move the analysis back to finding the best investments. This is because raw data would make it easier to conduct big data analysis of a company's profit drivers. As long as money can be made from this analysis, the market will undertake it. Therefore, eventually, the share price of a company should reflect more accurately reflect its fundamental value.

b. Applying competitive forces to accounting

Another major benefit of raw data disclosure is that the process of turning raw data into information is subject to competitive market forces, and as such improves the quality of the output. The argument is that, in the context of public companies' disclosure, the process of turning raw data into information is not the sort of process that should be done by regulators, but instead should be decided through market forces. Sutton, van Zijl and Cordery observe that originally accounting was an 'atheoretical' activity, they write:

> Accounting evolved as an atheoretical technology for recording financial transactions. From its historical foundations [General Purpose Financial Reporting (GPFR)] has become central to the development of socialized capital, bridging the information gap between owners and managers of capital. […]. Although the first joint stock company was formed in the UK in 1553 and the first modern corporation (the East India Company) in 1600, prior to 1775 commercial accounts were generally kept for owners' own use, with no external users to consider […]. Financial reporting in the UK, as the progenitor of the corporation

[43] Eduardo J. Ruiz, Vagelis Hristidis, Carlos Castillo, Aristides Gionis, Alejandro Jaimes, "Correlating Financial Time Series with Micro-Blogging Activity", WSDM '12 Proceedings of the fifth ACM international conference on Web search and data mining, Pages 513-522, 2012,

[44] Ibid

[45] Chester Curmea,b,1, Tobias Preisb, H. Eugene Stanleya,1, and Helen Susannah Moatb, "Quantifying the semantics of search behavior before stock market moves" Proceedings of the National Academy of Sciences, 08/2014, Volume 111, Issue 32

[46] Keynes John Maynard, "The General Theory of Employment, Interest and Money", 1936, Palgrave Macmillan, chapter 12

> in its modern form, was substantially unregulated in the nineteenth century.[47]

It is precisely the task of "bridging the information gap between owners and managers of capital" which creates the need for a system in which accounting standards are administered by a *de facto* "government mandate[d] […] regulated monopoly"[48]. The appeal of disclosing raw data is that it makes a system possible where owners and managers of capital are separate, the information gap between them is bridged and yet there is no need for a government-regulated monopoly. It is obvious that creating an effective accounting standard is a challenging undertaking. Accounting is full of tough questions. Hu, for instance, gives the example of how difficult it is to account for derivatives[49]. There are numerous other examples e.g. hedge accounting, accounting for pensions, or the debate on historical costs vs. mark to market accounting, to mention just a few. Disclosing raw data would help to 'solve' some of these difficult accounting issues, in the sense that they would lose some of their relevance. Special Purpose Entities ("SPEs") and Structured Investment Vehicles ("SIVs") provide a good example of how disclosing raw data could help. For an SPE to achieve its goal, an "[o]ff-balance sheet treatment for financial reporting purposes is generally necessary" [50]. Off-balance sheet financing provides varies benefits to the sponsoring entity, e.g. lowering the leveraging ratio. The key to achieve an off-balance sheet treatment is for the SPE and the sponsoring entity to be effectively separated. In particular, in needs to be the case that there is no recourse from the creditors of the SPE to sponsoring entity.[51] In the wake of various scandals, like Enron, the criteria for off-balance sheet treatment have been made increasingly strict. "To overcome this challenge, the focus turned to another financing vehicle, the Structure Investment Vehicle."[52] However, Amoruso and Duchac note the following:

> At the heart of the SIV, however, was an implicit guarantee by the vehicle's sponsor of the SIV's short- and medium-term creditors. This piece is critical. The lack of an explicit guarantee combined with an equity tranche owned by independent third parties allows the sponsor to avoid consolidating the SIV, because the sponsor does not technically share in the risks and rewards of the vehicle […] Even where the bank does not invest in the capital, the relationship with capital note investors

---

[47] Sutton David, van Zijl Tony, Cordery Carolyn, "Twentieth century academic accounting's role in the failure to develop a coherent theory of accounting", July 14, 2010, SSRN http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1761660 (accessed Oct 19, 2015)

[48] Emerson, Karim, Rutledge, "Fair Value Accounting: A Historical Review Of The Most Controversial Accounting Issue In Decades" Journal of Business & Economics Research, Volume 8, Number 4, 2010 page. 21

[49] Supra note 7, pages 385 - 388

[50] Amouruso J. Anthony, Duchac Jonathan, "Special purpose entities and the shadow banking system: The backbone of the 2008 financial crisis", Academy of Accounting and Financial Studies Journal. 18.2 (Apr. 2014)

[51] Amouruso and Duchac list the following five criteria, which must be fulfilled that must be satisfied to achieve off-balance sheet treatment for an SPV. "1. It must be a separate legal entity from the sponsor, 2.The entity is "bankruptcy remote", 3. It is created to carry out a fairly specific activity, and 4. It is thinly capitalized, (i.e., heavily leveraged), with the residual equity ownership held by a third party other than the sponsor" ibid page 108

[52] Ibid page 113

> may be such that it behooves the bank to avoid losses to capital note investors to protect that relationship"[53]

In theory, it may be possible for regulators to create a standard whereby it can be determined whether an implicit guarantee has been given by the sponsor to the SIV. Yet, it is considerably more likely that the question of whether an implicit guarantee exists will be better answered through disclosure of raw data. The question whether a guarantee was given (implicit or explicit) does not necessarily have a binary answer. Also, the existence of a guarantee might depend on factors outside the relationship between SIV and sponsor. The point is that whether an effective guarantee exists, or not, is as much of a complex question as the assessment as of whether a new product will be successful or whether a company is well run. A company's management should be able to provide its views on this matters, nonetheless, investors should be able to form their own views. Raw data disclosure would make it considerably easier for market participants to form their own views on whether an implicit guarantee has been given, and, based on such views, market participants could better address their investment decisions. Thus, the question whether an implicit guarantee was given would be answered by people who have a stake in the outcome and not by regulators, who have no stake in the outcome. Furthermore, it is already the case that analysts routinely restate financial statements.[54]

One may be tempted to argue that the foregoing is only true for complicated modern finance – that these issues would not arise in a world without derivatives, SPVs, and their ilk. This is essentially Hu's argument.[55] However, the problem set out above applies to any kind of standardised accounting system. For instance, it applies to the debate on when to capitalise, when to expense a cost, and the related debate on lease accounting[56]. How to record an expense is a fundamental part of an accounting system, and not a feature of modern day 'esoteric' finance. Disclosing raw would allow each market participant to formulate her own view. Generally speaking, the problem with a system that mandates the disclosure of standard accounts is threefold.

First, the nature of standardisation requires that certain dissimilar situations are treated alike. For example, IFRS (or other accounting systems, like US GAAP) may allow a certain cost to be capitalised, however, on closer inspection it may turn out it that it should be expensed. This does not mean that IFRS got it wrong, it is simply in the nature of any standardised system that some cases are addressed inappropriately. Second, a benefit of disclosing raw data is that people with a stake in the outcome decide how transactions should be recorded. Presently, standard setters are not directly affected by their decisions – they are not market participants who stand to gain or lose money if they get it right or wrong. Although, of course, standard setters, are 'indirectly' affected by the correctness of their decisions, because getting it right helps to justify the value of their existence, while repeatedly getting it wrong would question the value of having standard setters at all. The third benefit is that more people have

---

[53] Ibid page 115

[54] For financial statement recasting in the context of credit risk assessment see for instance George Batta, Ananda Ganguly, Joshua Rosett, "Financial statement recasting and credit risk assessment" Accounting and Finance 54 (2014) 47–82

[55] Supra note 7 page 384

[56] For a review on the literature of lease accounting see for instance, Barone Elisabetta, Birt Jacqueline and Moya Soledad, "Lease Accounting: A Review of Recent Literature, Accounting in Europe", 2014, 11:1, 35-54, DOI: 10.1080/17449480.2014.903630

the chance to analyse the data and form an opinion, and through competitive forces in public markets this should be accurately reflected in the price of the securities traded.

c. Less creative accounting

The main way that disclosure of raw data will reduce costs is the resulting reduction in the incentives to structure transactions to achieve a particular accounting treatment, and the corresponding reduced expenditure on accounting. Transactions are structured to achieve a certain accounting treatment for a number of reasons, inter alia, to influence the amount of tax payable, to prevent a breach of certain debt covenants, to comply with applicable regulations. Even under a disclosure of raw data regime these reasons for structuring transactions would remain. These practices will endure due to the incomplete nature of contracts, legislation and regulations. However, many transactions are structured for the purposes of market perception. Arguably, a dramatic example of this was Enron. Schwarcz's argues as follows:

> It now appears that Enron engaged in of manipulative accounting transactions, devoting much more energy to creative accounting than to making a profit to account for. Its primary motivation was to minimize financial-statement losses and which could have hurt Enron's credit rating and thereby damage its credibility in the energy trading business.[57]

Manipulating accounts is by no means restricted to Enron alone. Oliveras and Amat argue that "it is widely recognised that accountants can use their knowledge of accounting rules to manipulate the figures reported in the financial statements".[58] Griffiths states that:

> Every company in the country is fiddling its profits. Every set of published accounts is based on books that have been gently cooked or completely roasted. The figures, which are fed twice a year to the investing public, have all been changed in order to protect the guilty. It is the biggest con trick since the Trojan horse [...]. In fact this deception is all in perfectly good taste. It is totally legitimate. It is creative accounting[59]

While the above quote might be a bit of an exaggeration, it seems quite clear that companies invest considerable time and effort in order to achieve specific accounting treatments. If companies had to disclose their raw data, the incentives to engage in creative accounting for the purposes of market perception would be significantly reduced. If market participants are able to use the underlying data to assess the

---

[57] Schwarcz, Steven L, "Enron, and the Use and Abuse of Special Purpose Entities in Corporate Strcutures", The Fianacier, Vol 9, Nos 1-2, 2002, page 23. Also note that, Schwarcz argues that Enron's motivation was to influence its credit rating, which could be considered an instance of the incomplete contract problem. However, the key point is that this was done in order to preserve "its credibility in the energy trading business".

[58] Ester Oliveras and Oriol Amat, "Ethics and creative accounting: some empirical evidence of accounting for intangibles in Spain" http://repositori.upf.edu/bitstream/handle/10230/806/732.pdf?sequence=1 (accessed Oct 23, 2015)

[59] Ian Griffiths "An Introduction to Business Ethics" edited by George D. Chryssides, John H. Kale, Cengage Learning EMEA, 1993, page 360

profitability of a company, there is little point in manipulating accounts to make a company appear more profitable. One of the benefits of this regime is freeing up resources. All the wily accountants currently occupied with creative accounting could focus their talent on other (more productive) enterprises. In addition, all the slick buy-side analysts could spend less time restating financial statements and more time analysing a company's the real profit drivers. Essentially, the cat-and-mouse game between issuers (who try to make themselves look as profitable as possible) and investors (who try to reverse-engineer the numbers) would be greatly reduced. Nonetheless, as stated above, the incentive to structure transactions would not completely disappear because creative accounting due incompleteness of contracts or legislation would remain.

Whether costs incurred by regulators will increase or decrease is ambiguous. Because the market could review a company's methods of drawing up its accounts, arguably, less resources would be needed to police accounts – ensuring that they are drawn up properly. However, as the regime proposed in this paper still requires companies to comply with all the current disclosure requirements, regulators would still need to ensure these obligations are fulfilled. In addition, if disclosure of raw data is mandatory, regulators would have the additional requirement of checking that the raw data is properly disclosed. Also, there will be the additional costs of making raw data accessible to the public. One may also be tempted to count the costs of analysing raw data as an additional cost. However, this would be incorrect. No one is forced to analyse the data, therefore, investors would only analyse the data if they regard it as worthwhile i.e. the expected profits are less than the expected costs.

6. Objections to disclosing raw data

The above analysis suggests that, while there may be some debate about the extent of benefits associated with the disclosure of raw data, that there are some benefits seems quite clear. However, disclosing raw data may also have some drawbacks. In order to make a comprehensive assessment it is necessary to consider some of the potential objections to raw data disclosure in more detail.

a. Level playing field among investors not guaranteed

The main objection to making disclosure of raw data mandatory, is probably the concern that it could lead to an uneven playing field among investors. Investors who have access to the technology and know-how to analyse the disclosed raw data will have a significant advantage over investors who do not. For instance, Hu, although, in principle sympathetic to the idea of disclosing raw data, writes:

> The task of overserving objective reality never goes away. It is shifted from the business entity that has the requisite expertise […] to market participants. Market participants will not be on a level playing field[60]

That all investors should have access to the same information is one of the central pillars of regulating public markets. According to the SEC, the regulation of public markets is based on this principle:

[60] Supra note 7 page 391

> The laws and rules that govern the securities industry in the United States derive from a simple and straightforward concept: all investors, whether large institutions or private individuals, should have access to certain basic facts about an investment prior to buying it, and so long as they hold it. To achieve this, the SEC requires public companies to disclose meaningful financial and other information to the public. This provides a common pool of knowledge for all investors to use to judge for themselves whether to buy, sell, or hold a particular security. Only through the steady flow of timely, comprehensive, and accurate information can people make sound investment decisions.[61]

Under disclosure of raw data regime, while all market participants would have access to the same *data*, it is certainly not the case that they would have access to the same *information*. Disclosure of raw data is in direct contrast to the SEC's mission statement. However, to assess whether an unequal playing field is a strong enough argument against a raw data disclosure regime being introduced, one needs to evaluate why having a level playing field among investors is desirable in the first place. From a conceptual point of view, at least two reasons appear to explain why equal access to information is important: i) facilitation of trading[62] ii) fairness[63]. The following sections will analyse these reasons in more detail, arguing that the only reason why disclosure of raw data would not be desirable from a societal point of view is because it may lead to socially inefficient levels of high investment in private information.

b. Facilitation of trading

The argument that equal access to information facilitates trading often comes in the form of the 'liquidity argument'. According to Leuz and Wysocki, the liquidity argument consists of the following:

> In essence, an uninformed investor fears that an informed investor is willing to sell (buy) at the market price only because the price is currently too high (too low) relative to the information possessed by the informed trader [...]. As a result, the uninformed investor lowers (increases) the price at which he is willing to buy (sell) to protect against the losses from trading with an informed counterparty. […] Adverse selection problems fold back to the point at which the firm issues shares.

---

[61] See SEC, "The Investor's Advocate: How the SEC Protects Investors, Maintains Market Integrity, and Facilitates Capital Formation" http://www.sec.gov/about/whatwedo.shtml (accessed Oct 23, 2015)

[62] See for instance Christian Leuz and Peter Wysocki, "Capital-Market Effects of Corporate Disclosures and Disclosure Regulation", 2006, Commissioned by the Task Force to Modernize Securities Legislation, page 192. They also mention "potentially improve corporate governance and managers' investment decision" as a benefit of disclosure (Ibid page 192). This benefit is not specifically discussed in this article. However, it should be clear that disclosure of raw data should be beneficial here as well

[63] See for instance SEC, Proposed Rule: Selective Disclosure and Insider Trading - SECURITIES AND EXCHANGE COMMISSION - 17 CFR Parts 230, 240, 243, and 249 - Release Nos. 33-7787, 34-42259, IC-24209, File No. S7-31-99 - RIN 3235-AH82 - Selective Disclosure and Insider Trading https://www.sec.gov/rules/proposed/34-42259.htm#P22_10578 (accessed Oct 22, 2015)

> […]. This effect implies that the firm must issue more shares to raise a fixed amount of capital.[64]

Leuz and Wysocki go on to explain that mandatory disclosure mitigates this problem in two ways:

> First, more information in the public domain makes it harder and more costly for traders to become privately informed. As a result, fewer investors are likely to be privately informed […]. Second, more disclosure reduces the uncertainty about firm value, which in turn reduces the potential information advantage that an informed trader might have.[65]

Under disclosure of raw data, some market information will be distributed unequally, and this will, following the Leuz and Wysocki argument, raise the cost of capital. On the other hand, a countervailing tendency exists because disclosure of raw data reduces "the uncertainty about firm value". Which of these two effects will come to dominate is not clear. Yet, even if it is assumed that the former effect dominates, it does not necessarily follow the liquidity argument leads one to conclude that disclosure of raw data is not beneficial. The following are a few preliminary considerations to support the notion that disclosure of raw data is beneficial even if it creates asymmetry of information. These not fully fledged arguments, merely suggestions. These are discussed, not to suggest that these arguments are correct but, to indicate the wrongness of dismissing disclosure of raw data out of hand purely because of asymmetric information. Ultimately the question of whether asymmetric information is fatal for a disclosure of raw data regime can only be settled through further research.

c. Information is asymmetric even under current disclosure regime

Under the current regime, some traders (e.g. professional traders) already enjoy huge advantages over others (e.g. retail traders). Hedge funds and other professional traders spend large amounts of money on the best technology and access to information, in addition to hiring the best graduates and experienced professionals. It is somewhat implausible to postulate that traders at highly sophisticated hedge funds use the same information as 'mom and pop' investors. In 1973, Homer Kripke was already urging the SEC to abandon "the myth of the informed layman"[66] and to focus the disclosure regime on the needs of professional investors.[67] He essentially argued that securities are too complicated to be understood by the general public, and the disclosure regime should acknowledge this by catering to the needs of professional investors rather than laypeople. Thus, it is arguably already the case that information is unequally distributed in the market. However, under the current system, society does not reap the benefits that disclosure of raw data would bring.

---

[64] Supra note 123, page 191
[65] Ibid page 193s
[66] Kripke Homer, "The Myth of the Informed Layman" The Business Lawyer, Vol. 28, No. 2 (January 1973), pp. 631-638
[67] Kripke argues that "[d]isclosures should be oriented to disclosing what the informed investor may think is important" ibid. page 637

d. 'Uniformed traders' are not necessary for the sufficient provision of capital

One could, however, argue that although it is true that sophisticated investors, like hedge funds, enjoy an informational advantage over retail investors, sufficient capital (and thereby sufficient liquidity) in the market is assured because the market caters for both professional and retail investors (i.e. retail investors are a large enough segment of the market, without which not enough capital would be supplied). However, this is not the case. For instance, the Kay review produces the following historic figures in relation to UK equity markets:

Historical Trends in Beneficial Ownership (Percentage Held)[68]

	1963	**1975**	**1981**	**1991**	**2001**	**2008**	**2010**
Rest of the world	7	5.6	3.6	12.8	35.7	41.5	41.2
Insurance companies	10	15.9	20.5	20.8	20	13.4	8.6
Pension funds	6.4	16.8	26.7	31.3	16.1	12.8	5.1
Individuals	54	37.5	28.2	19.9	14.8	10.2	11.5
Other	22.6	24.2	21	15.2	13.4	22.1	33.6

What is striking about this table is that the proportionate shareholding of individuals has declined dramatically from 54% in 1963 to 11.5% in 2010. The above figures, however, are potentially misleading because shares are often held by intermediaries, which leads to a blurring of the notion of 'ownership'. As the Kay review itself notes "[b]ecause of this ambiguity in the meaning of ownership, data such as that presented in [the] Table […] should be treated with care."[69] Notwithstanding the foregoing, a report by the Cass Business School also puts the figure for individual shareholdings to 11%[70]. The change in the structure of shareholding from shares were previously held is not just a UK phenomenon. Luigi Zingales notes:

> What has changed the focus is not only the success of the 1930s securities regulation but also the increase in institutional ownership (from less than 10% in the 1930s to more than 70% today), which has made the protection of unsophisticated investors from fraudulent securities and stock market manipulation outdated.[71]

In a similar vein, Langevoort argues that modern securities markets have become institutionalised[72]. Evans adds that "recent New York Stock Exchange ("NYSE") data reveals that trades by individual investors represent, on average, less than 2% of NYSE trading volume for NYSE-listed firms […]. There is no question that US securities markets are now dominated by institutional investors"[73]. It appears to be factually wrong to claim that retail investors play an important role in the provision of capital in the current securities markets. In the 1930s, the notion that disclosure should have

---

[68] THE KAY REVIEW OF UK EQUITY MARKETS AND LONG-TERM DECISION MAKING – FINAL REPORT, 2013 page 31
[68]

[69] Ibid page 31 (also the holding by the rest of the world is overrepresented)
[70] Ibid page 32
[71] Luigi Zingales, Journal of Accounting Research, Vol. 47, No. 2, Regulation of Securities Markets: Perspectives from Accounting, Law, and Financial Economies (May, 2009), pp. 391-425, page 392
[72] Donald C. Langevoort, "The SEC, Retail Investors, and the Institutionalization of the Securities Markets" Virginia Law Review, Vol. 95, No. 4 (Jun., 2009), pp. 1025-1083
[73] Alicia Davis Evans, "A Requiem for the Retail Investor?", Virginia Law Review, Vol. 95, No. 4 (Jun., 2009), pp. 1105-1129, page 1105

the goal of ensuring that laypeople have access to the same information may have been plausible. However, the current situation is drastically different. The point is that disclosure of raw data favours professional investors over retail investors. However, this is not a problem for market liquidity because retail investors make up a tiny minority of the market.

e. Equal access to information is not essential for liquidity

The argument is that for markets to reach a price equilibrium it is not necessary for all market participants to be well informed, or even that a majority of market participants are well informed. Even if only a small group of market participants have access to the relevant information, this is sufficient for markets to reach equilibrium. Gilson and Kraakman have pointed out that "rapid price equilibrium does not require widespread dissemination of information, but only a minority of knowledgeable traders, who control a critical volume of trading activity"[74]. Applying this to the disclosure of raw data means that, even if only a handful of investors can make use of the raw data, provided they "control a critical volume of trading activity" a price equilibrium will emerge fairly quickly. Once an equilibrium price is established, the market should be fairly liquid. Although there are still some unanswered questions as how exactly the mechanism described by Gilson and Kraakman works (as Gilson and Kraakman are willing to admit[75]), the intuition behind the mechanism is quite compelling. It goes something like this: traders must make some profit from their trading activities, otherwise, they would not engage in any trading activity. When the market is out of equilibrium, traders can make profits. When markets are in equilibrium, persistent profits are impossible. Thus, trading on prices which are out of equilibrium, is not the antithesis of an efficient market, but a necessary part of it.[76] There is no sharp distinction between markets with informed traders and uninformed traders. All markets need some level of uniformed traders to function. On the other hand, if there are too many uninformed traders, liquidity will dry up. Thus, the relevant issue is not whether disclosure of raw data will create traders which have superior information (it certainly will), but whether there will be sufficiently large number of informed traders controlling "critical volume of trading activity" to ensure liquidity. This is ultimately an empirical question (and a very important one at that). Further research is needed to answer this question. For the purpose of this article, it suffices to state the following: For stock markets to work it is not necessary for there to be a majority of informed traders. Even with a minority of informed traders, stock markets can still function. Therefore, a regime that creates both informed and uninformed traders (such as the disclosure of raw data regime), does not necessarily lead to the drying up of liquidity in the market.

f. Unequal access to information does not exclude laypeople from the profits of capital markets

The liquidity argument deals with the question addressing whether or not the market will function. A different objection is that, although markets may function even without retail investors, a stock market with only professional traders is unfair because

[74] Ronald J. Gilson and Reinier H. Kraakman, The Mechanisms of Market Efficiency, Virginia Law Review, Vol. 70, No. 4, Fifty Years of Federal Securities Regulation: Symposium on Contemporary Problems in Securities Regulation (May, 1984), pp. 549-644, page 569

[75] Ibid page 570

[76] See also Sanford J. Grossman and Joseph E. Stiglitz "On the Impossibility of Informationally Efficient Markets", The American Economic Review, Vol. 70, No. 3 (Jun., 1980), pp. 393-408

it deprives ordinary investors from the possibility of participation in the profits generated by stocks. This is especially worrisome because shares have historically had the highest long return of all asset classes.[77] Thus, disclosure of raw data could potentially deprive retail investors of a very important source of returns. In a similar vein Jaron Lanier's argues that big data analytics, in general, has a tendency to lead to a concentration of wealth in the hands of the few.[78] The argument is that disclosure of raw data is potentially unfair to those people who lack the ability to analyse the data. Nonetheless, this argument is wrong. Even if there are only professional investors in the market, the general public can still profit from the stock market by investing through such professional investors or otherwise in investment vehicles (e.g. via mutual funds, index tracker funds, ETFs, and otherwise). Many would already argue that the best way for a layperson to invest in the stock market is through such investment vehicles, rather than stock picking.[79] To argue that a market consisting only of professional traders deprives the general public from sharing in the gains of the stock market relies on a very antiquated idea of how a layperson should invest. There is, of course, the question of how profits should be split between professional investors and the public. It might turn out that the split is grossly unfair, requiring regulatory intervention. As such, it is at this level that regulatory intervention should occur.

g. Inefficient accumulation of private information

Hirshleifer[80] pointed out that under certain circumstances "private information acquisition for speculative gains in securities markets is socially wasteful."[81] In a pure exchange economy (i.e. an economy without production) the accumulation of information to forecast future events is always socially wasteful. Further, even in an economy with production, the value of private information is less than the social value of information. Hirshleifer, writes as follows:

> Public information as to [the future] is indeed of social value in a regime of production and exchange. However, it remains true that the value of private foreknowledge is enormously greater to any individual than the value to him of public foreknowledge. […] Thus, the incentive for the use of resources to generate private information remain excessive[82] (original italics removed)

Disclosure of raw data requires investment in technology and data analytics. Following Hirshleifer's argument, potentially a socially sub-optimal high level of investment

---

[77] See for instance, BlackRock, "Asset class return – A 20-year snapshot", (https://www.blackrock.com/investing/literature/investor-education/asset-class-returns-one-pager-va-us.pdf), accessed Oct 22, 2015

[78] The argument is that big data analytics has two effects: i) the profits of big data tend to only go to a few, amongst other reasons because the technology / know-how is expensive, ii) those who do not share in the profits of big data lose income because big data makes many traditional jobs redundant or less profitable (e.g. translators, musicians etc.) See Jaron Lanier, "Who Owns the Future?" Simon & Schuster; Reprint edition, March 4, 2014

[79] See for instance, Laura Shin, "How should a lay person invest in the stock market" Forbes, SeP 26, 2014, http://www.forbes.com/sites/laurashin/2014/09/26/7-steps-to-stock-investing-without-too-much-risk/ (accessed Oct 22, 2015)

[80] Hirshleifer, Jack, The Private and Social Value of Information and the Reward to Inventive Activity, The American Economic Review, Vol. 61, No. 4 (Sep., 1971), pp. 561-574

[81] Leuz Christian, Wysocki Peter "Capital-Market Effects of Corporate Disclosures and Disclosure Regulation", 2006, Canada Steps Up, Volume 2, Research Studies, Evolving Investor Protection, page. 198

[82] Supra note 53, page 567

could take place. This is a powerful objection to the idea of disclosing raw data because, in essence, Hirshleifer's argument says that rather than encouraging more investment into analysing stock markets, less investment is better.

However, there are some counterarguments against the idea that Hirshleifer's argument shows that disclosure of raw data is wrong. First, it is not necessarily clear that disclosure of raw data will lead to more costs for society as a whole. It is likely that some market participants will find it profitable to allocate more resources to analysing stock markets. However, there are also some cost savings associated with disclosing raw data. It is not clear which effect will dominate. Essentially, one could argue that there are socially inefficient costs associated with both disclosure regimes, but at least under disclosure of raw data, society receives more accurate forecasts. Second, as it was argued by Coffee[83] that "private monitoring creates free-rider problems by conferring uncompensated benefits on other investors"[84]. This argument can be extended to suggest that accumulation of accurate private information about a company has positive externalities for the economy as a whole. Some aspects of the global financial crisis provide a good illustration of this problem. Arguably, had investors been in a better position to calculate the true risk exposure of Lehman Brothers, Bear Stearns, AIG, as well as other large financial institutions before 2007 / 2008, this would have provided valuable information to the market as a whole, and arguably for the entire financial system and world economy. The argument is that if one investor finds out valuable information about a company, this information will soon become public knowledge, which enables other market participants to adjust their action. This does not necessarily mean that they will 'trade' on this information, however, they adjust other activities. For instance, if it becomes common knowledge that a bank is close to bankruptcy, people will not just stop buying the bank's shares. They may also refuse to accept the bank as counterparty to a derivative, and the latter is where the real benefit of the externality lies.[85] It seems that that in highly interdependent markets (as modern financial markets appear to be), the externalities mentioned by Coffee are more important than in highly fragmented markets. The argument regarding externalities does not, by and of itself, show that there is no problem with inefficient acquisition of private information. However, it does suggest that the acquisition of private information has important externalities. Whether (1) the value of the externality, or (2) the wastefulness of the private information, has greater effect is an empirical question that only further research can settle. Nevertheless, it is important to recognise that Hirshleifer's argument presents a powerful argument against disclosure of raw data.

h. Raw data cannot be disclosed because most of it is confidential

In his arguments against disclosure of raw data Hu states the following:

> Much of the pure information may be confidential, proprietary, or otherwise not appropriate for public disclosure. Existing uncertainties as to when such pure information [i.e. raw data] would be available to the public also make difficult the full deployment of a transfer mode strategy[86]

---

[83] Coffee, John C. Jr, Market Failure and the Economic Case for a Mandatory Disclosure System, Virginia Law Review, Vol. 70, No. 4, Fifty Years of Federal Securities Regulation: Symposium on Contemporary Problems in Securities Regulation (May, 1984), pp. 717-753

[84] Supra note 64, page 198

[85] Arguably, Hu tries to argue for a similar point when in his analysis of JP Morgan and the London Whale scandal at pages 385 – 388 supra note 7

[86] Supra note 7, page 391

It is certainly true that parts of raw data will be confidential or private. What proportion of the disclosable raw data will be privileged in some form (e.g. through confidentiality or privacy legislation) is an empirical question, to which the answer is not currently available. However, even if one assumes (as Hu seems to do) that a substantial proportion of the raw data will be privileged, it does not follow that disclosure of raw data should not be implemented. There are two reasons for this. First, even if only a small proportion of data is disclosable it would still provide a benefit, albeit a smaller one. Second, advocating a model of raw data disclosure is intended to urge regulators to adopt a framework that favours the disclosure of raw data over the disclosure of summary data. This means that regulators should adopt a mind-set that encourages disclosure of data as close as possible to raw data e.g. if it is not possible to disclose the names of counterparties to a derivative positions, the positons potentially could be disclosed in anonymised form. This would undoubtedly be a serious limitation to the disclosure of raw data, however, it would still be better than the current disclosure system.[87] Ultimately, the regimes of raw data disclosure and intermediary depiction may actually be quite similar in practice. However, the potential similarities between the two regimes does not mean that one should not strive for the better regime. In fact, arguing that a disclosure of raw data regime might potentially be quite similar to the present regime may give comfort to those that fear that this dramatic regime change would introduce too much risk into the system.

7. Does the free market produce enough incentives for companies to disclose raw data?

An implicit assumption in the above argument is that the disclosure of raw data regime would require mandatory disclosure, rather than just permitted disclosure. It could be argued that if disclosing raw data is such a great idea, companies would engage in it voluntarily. The debate on whether disclosure needs to mandatory or whether a prohibition against untrue disclosure is enough, has, of course, a long history. Thus, one might be tempted to think that the analysis of whether disclosure of raw data should be mandatory is essentially the same analysis as whether disclosure should be mandatory in the current system. However, it turns out that the reasons why disclosure should be mandatory are different, and that the case for mandatory disclosure is actually stronger in the case of raw data disclosure than in the case of intermediary depiction.

One of the major arguments against mandatory disclosure of any kind is the so called 'unravelling argument'.[88] The unravelling argument holds that a prohibition against untrue statement is sufficient to induce companies to disclose all relevant information. A rule that actively requires companies to disclose relevant information

[87] The reason why this article does not support Hu's idea of a hybrid mode (see supra note 7 pages 391 and subsequent) is that, apart from the question whether the solutions he proposes actually work (and this author thinks that they do not), they are only a solutions for very specific problems in the disclosure regime for financial institutions. One of the main points this article tried to stress is that disclosure is a problem for all companies, not just financial institutions. Therefore, a more general solution is required.
[88] See for instance S. J. Grossman and O. D. Hart, "Disclosure Laws and Takeover Bids" The Journal of Finance, Vol. 35, No. 2, Papers and Proceedings Thirty-Eighth Annual Meeting American Finance Association, Atlanta, Georgia, December 28-30, 1979. (May, 1980), pp. 323-334, or

is unnecessary. The intuition behind the unravelling argument is rather simple. Grossman and Hart give the following example:

> [C]onsider a seller of oranges who states that a box of oranges contains "at least 5 oranges." If I know that the seller knows the exact number of oranges, then I know there must be exactly 5 oranges per box, since if there were 6 per box, then the seller would have stated that there are "at least 6 oranges per box.[89]

According to the proponents of the unravelling argument, a similar mechanism is at work in the stock market. The highest quality issuer (i.e. the most desirable company from an investor's point of view) has an incentive to reveal its quality to the market. However, after the highest quality company has revealed its quality to the market, the second highest quality company then has an incentive to reveal its quality to the market. After that, the third highest quality company has an incentive, and so on. All that is needed, from a regulatory point of view, is a prohibition against lying – mandatory disclosure is not necessary.

Traditionally arguments against the unravelling argument depiction usually fall into three categories:

i) the unravelling argument depends on very specific assumptions;[90]
ii) mandatory disclosure acts a commitment device;[91]
iii) disclosure creates externalities;[92]

These arguments apply equally to intermediary depiction and raw data disclosure. However, apart from the externalities argument, the arguments against the unravelling argument are not very strong. Regarding the first argument, some of the assumptions that the unravelling argument relies on are: disclosure must be costless; product quality must be comparable; consumers need to know that producers have private information; as well as others. Without these assumptions, the full disclosure equilibrium might not prevail. However, Leuz and Wysocki argue that "even if these assumptions [the assumptions on which the unravelling argument is based] are violated, the general spirit of the unravelling argument still applies". [93] The intuition behind mandatory disclosure as a commitment device is that the management of a company may have an incentive *ex-ante* to commit to disclose information in the future, however, management may have an incentive to negate on this commitment *ex-post* (e.g. management may not want to disclose information on poor performance). However, Leuz and Wysocki argue that, "the owners of the firm ultimately bear the cost of not providing a commitment to disclosure (as well as any residual agency problems). For this reason, managerial agency problems are per se not a sufficient reason for mandatory disclosures."[94] For this argument to work, it needs to be shown that mandatory disclosure is a more efficient commitment device than a system that could be devised by the market. That disclosure has positive externalities provides a more convincing argument for mandatory disclosure. Usually the argument is made

[89] Ibid page 324
[90] Supra note 123 page 197
[91] Ibid
[92] Supra note 123 page 198
[93] Supra note 123 page 197
[94] Supra note 123 page 198

that disclosure by one firm provides useful information about other firms too.[95] For instance, disclosure by Coca-Cola also provides useful information about Pepsi. However, Coca-Cola can only capture the benefits associated with the information that relates directly to it and, therefore, discloses too little information.

There is probably some truth in the argument from externalities. However, for raw data disclosure there is an additional argument for disclosure to be mandatory. H. S. Shin pointed out that "the unravelling argument is extremely sensitive to what the informed party actually knows"[96]. In applying this to raw data disclosure this means that the market cannot draw the conclusion from a company not disclosing data that the data conveys negative information because the company does not know what information the data conveys. The unravelling argument presupposes that the company is able to rank the information it could truthfully disclose. In the above example, the company needs to know that disclosing at least 6 oranges is better (i.e. will raise demand) than disclosing at least 5 oranges. Grossman and Hart extend their example to cases where the seller is unsure how many oranges there are in the box.[97] In their example, the box may contain 10 oranges, 20 oranges, or otherwise, there is a 50% chance of 100 oranges or a 50% of 75 oranges. Yet, even in this case the company is able to rank the outcomes. However, this is not a good illustration of big data analytics. A better illustration is to suppose that a company has access to lot of data, and from that data the company draws certain conclusions. The company also knows that if others could gain access to the same data, they will be able to draw different (and maybe better) conclusions from the data. The question then becomes whether the company has an incentive to disclose the underlying data. If not, then it cannot be assumed that not disclosing means the company has something to hide. To illustrate, let's continue with the above example. Assume, somewhat fancifully, that there is also a second box (box B) which contains a piece of paper that either reads +10%, 0% or -10%, with equal probability.[98] The company does not know what the paper says but it can decide whether to disclose the content to the market. After the buyer bought a box of oranges box B will be opened and depending on what the paper says, she will be given 10% more oranges, will lose 10% or the number of oranges will remain unchanged. Box B in this example is intended to represent information derived from raw data. The question now becomes whether the company has an incentive to disclose the content of box B. If the consumers are risk neutral then the orange seller will be indifferent to disclosing the content of the second box. If the consumers are risk averse they will prefer box B to be opened before they buy the box of oranges. For the seller the opposite is true. If the seller is risk averse she will prefer the box B not to be opened. What this means for a disclosure regime is the following. By disclosing raw data, the company knows it increases the accuracy of the forecasts of its fortune, thus, assuming investors are risk averse, one would expect its costs of equity to decrease, giving the company an incentive to disclose the raw data. On the other hand it is risky for the company to disclose raw data (because the company does not know whether the data contains good or bad information). Assuming the company's management is risk averse (which is a standard assumption) and investors to be risk neutral (or at least close to being risk neutral because they can diversify their investments) this will provide an incentive to the company not to disclose the raw data. This argument does

[95] For a discussion of this idea see for instance Frank H. Easterbook and Daniel R. Fischel, "The Economic Structure of Corporate Law", Harvard University Press, 1991 pages 276 - 315

[96] Shin H.S., "News Management and the Value of the Firm", The Rand Journal of Economics, Vol. 25 No 1 (Spring 1994), pp 58 – 71, page 59

[97] Supra note 149 page 324

[98] Arguably, to make the example more similar to the big data case, one should assume that what is says on the paper is uncertain and now probabilities can be attached to it. However, the argument does not change if one assumes that in cases of uncertainty, equal probabilities are attached to all the outcomes.

not, in itself, suggest that there disclosure of raw data should be made mandatory. All that the argument suggests is that it is unlikely that the free market provides sufficient incentives for companies to disclose raw data voluntarily.

8. Does disclosure of raw data mean that insider trading should be legalised?

The above arguments may lead one to conclude that if they are true, this would could also mean that insider trading should be legalised. After all, the above argument suggests that there is no problem when some investors have superior information to other investors. Thus, one may be tempted to conclude that if this is true, this also means that there is no problem if certain people, such as corporate insiders, trade on superior information. However, although the above argument weakens some of the arguments for criminalising insider trading, they do not necessarily suggest that insider trading should be legalised. The two main arguments against criminalisation of insider trading were summarised by Bainbrideg (which in turn summarises Henry Manne's arguments)[99], as follows:

> First, […] insider trading causes the market price of the affected security to move toward the price that the security would command if the inside information were publicly available. If so, both society and the firm benefit through increased price accuracy. Second, […] insider trading [is] […] an efficient way of compensating managers for having produced information. If so, the firm benefits directly (and society indirectly) because managers have a greater incentive to produce additional information of value to the firm.[100]

According to Bainbridge, the first argument fails on empirical grounds because the information generated by insider trading is too small to move the share price.[101] The second argument is stronger.[102] Nonetheless, it is important to note that the second argument specifically deals with the relationship between managers (i.e. people who have some power over the company) and the owners of the company. Thus, this argument cannot apply to the disclosure of raw data because in this regime the 'inside information' is held by people who do not have control over the company. Inside information could indeed be an efficient way to compensate managers. However, the well-known problem with this argument is that an entrepreneur also receives inside information even when her performance was sub-standard.[103] If short-selling is possible, the entrepreneur is indifferent between positive and negative information about the firm, and therefore has no incentive to increase performance. Under a raw data disclosure regime data too, superior information is a reward. However, under disclosure of raw data, the award of superior information goes to the person with superior ability to analyse data (and superior data analysis arguably, should be

[99] Stephen M. Bainbridge, "Insider Trading: An Overview", SSRN, 1998, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=132529 page 5
[100] Ibid.
[101] Supra note 160 page 6
[102] Henry Manne, who originally proposed these two arguments, also through that it is the stronger of the two argument. See supra note 160 page 7
[103] Supra note 160 page 9

incentivised, regardless of whether the analysis itself reveals positive or negative news).

A similar argument applies if one considers the arguments in favour of criminalising insider trading. Again, Bainbridge summarises the arguments as follows:

> (1) insider trading harms investors and thus undermines investor confidence in the securities markets; (2) insider trading harms the issuer of the affected securities; and (3) insider trading amounts to theft of property belonging to the corporation and therefore should be prohibited even in the absence of harm to investors or the firm.[104]

The first argument is similar to the argument that disclosure of raw data will lead to less liquidity in the market. As discussed above, this is unlikely to be the case. However, the other two arguments deal with the relationship between managers and the company, and therefore, do not apply to the disclosure of raw data. The point of this discussion is to suggest that disclosure of raw data, like the arguments for decriminalising insider trading, assert that there is no problem if information is distributed unequally in the market. However, the former and the latter differ on who should have superior information.

9. Conclusion

There are of course still many questions unanswered in relation to disclosure of raw data. However, this article has identified the potential for significant benefits. Thus, at the very least it is worthwhile to conduct further research into this area. The area that needs particular study is how the raw data disclosure regime would work in practice (i.e. what data needs to be disclosed by companies, in what format should it be disclosed, what are the costs associated with this regime). On the normative side, whether the accumulation of private information in public markets is inefficient from a societal point of view or not requires further analysis. It may turn out that a raw data disclosure regime is not as dissimilar from the current regime as one may expect. Nevertheless, as stock markets have the power to create and destroy a sizable amount of wealth, even a small improvement may provide a large benefit.

[104] Supra note 160 page 6