Appendix 3.06.1 Cyber security and grey information

Data that is privately stored in the cloud or accessible intra-web is still subject to being hacked, altered or stolen. China can and has hacked into the US defence communications network and turned it off. China and North Korea can disable the GPS system (only 9 satellites rendering the US military blind). China, North Korea and Russia could disable the 50 or satellites that govern most of the international mobile traffic and entertainment satellites). Anyone can hack into the cloud and render the accounts destroyed or, worse still, alter them. Uber, Facebook, Viber, Snapchat and Tinder can be destroyed – not just the accounts but the whole production system and *modus operandi*. This is a point in risk assessment, sensitivity analysis and the forecast models; and of course, the grey information dependency reports.

The one thing everyone should remember is that online fraud, credit card fraud, banking fraud is increasing exponentially. See Appendix 13.5 for some evidence of such examples and data.

With all the new technologies, it is likely to explode. Your fridge and TV can be hacked. Your router can be hacked and every keystroke you make on your phone or computer can be sent elsewhere. The only fact you need to remember is that there is such a thing as a secure environment. Google's Assistant, Apple's Siri, Amazon's Alexa, and a little behind, Microsoft's Cortana (but it will catch up) can all be hacked. Now this functionality is supposed to be useful – Alexa listening in and suggesting 'helpfully' to buy things when it hears you have run out. Privacy is going to be something of the past… The new generations will not want that sort of privacy. They will grow up with Alexa (or equivalent) interacting with it all the time.

The faster grey information explodes, the higher the number of security breaches, and the potential size of those breaches can occur.

Site attacks

Krish has experienced severe attacks by Russian based centres on US sites during the 2016 Trump presidential campaign. Those same attacks are occurring during the US mid-term elections in late 2018. This is a real threat for every business from hacking your site to denial of service.

Electronics and computer initiated transactions

Even the cosy relationship between management and shareholders is breaking down as computers increasingly make decisions as to buy and sell, with automatic trading systems, software controlling buy and sell decisions, together with dark pools.[1] These computer systems will be analysing all the messages, statements, reports, analysts' information, and press statements, other messages from the company including some personal contact (human to computer, or compute to computer). But they will also tap all information on the net and all media channels (TV, radio, press, magazines), as well as big data, social media and so on. So these trading software systems will also search for all information in real-time. When they start to sell, all such systems will notice and may act in consort either following one another, or because some piece of information has triggered a change. So a rise or fall can be much greater than under human trading. As such this has taken some of the traditional functions of stockbrokers and professional investors. Analysts now have to sell their specialised information and analyses on companies (MiFID II).

Such very simple systems (in comparison to the type we are describing above) gave rise to the 2010 Flash Crash which was either caused by computers trading together with electronic trading exchanges – now prevalent just about everywhere (though in a rather simple form – wait until full AI gets hold of these). Or these algorithms responded badly to a set of external input by a UK trader.

In Michael Lewis's book *Flash Boys*, he documents the struggle to create a fair electronic exchange in New York. Even so, Lewis' contention is that the conventional market is rigged in favour of the large banks and financial institutions which employ high-speed electronic trading firms using their power and computer speed to their advantage to extract billions of dollars from investors. One needs to be aware of the existence of spoofing, layering and especially front-running within dark pools and other exchanges – some of which are, in theory, supposed to be banned practices.

Add to these computer initiated transactions the ability of AI systems to run factories and inventory control systems, among others. Artificial intelligence promises to revolutionise our lives, drive our cars, diagnose our health problems, and lead us into a new future where thinking machines make decisions and do things that we have yet to imagine. But there are risks. Even Elon Musk, who admits he has access to some of the most cutting-edge AI, said that without some regulation 'AI is a fundamental risk to the existence of human civilization'.[2]

[1] Dark Pools are electronic alternative trading systems used by US, UK and European banks and financial institutions, very similar to stock exchanges where trades can be matched. In general many times a second a bank will place thousands of small buy and sell queries and at variety of prices to see what the market price is and n what volumes. The big difference of course is that the orders are dark, meaning that the size and price of the orders are not revealed to other participants.

[2] http://fortune.com/2017/07/15/elon-musk-artificial-intelligence-2/

Grey, Dark and Deep Web or Net

We need to refine and expand the meaning of grey information:

- The web or World Wide Web is the portion of the web that is available by normal web browsers[3]. This is sometimes called the public network which is available for all to use. Parts of this may be restricted to registered users or users which pay for a service.
    - ⇨ Porn sites are probably less than 10% of the total web sites but they have a disproportionately large number of pages and new content. So a figure of 38% of the web refers to the number of indexed pages.[4]
    - ⇨ You can turn off analytical tracking of who you are, and what you doing by, for example, using Google Chrome's incognito window (top right corner under settings). This is not to be confused with the deep or dark net.
- The surface web is that part of the web that can be found by the search engine web-crawlers/spiders or search engine robots – these systematically browse the web for the purpose of indexing.[5] Some parts of a website or links to other parts of the website may not be easy to find.
- The deep web includes sites and information which is not registered or indexed or read by any search engine. This includes information which can be housed in databases and which is only viewable through dynamic pages generated when the content is requested, and information which resides behind authentication such as on private networks and public networks such as Facebook. This includes data, which may be useful in reporting and auditing. The deep net is approximately 1,000 times the size of the web.
- The dark web or dark net is that portion of the web which cannot be easily reached from the public internet, and usually requires specialised software to access it. Examples of the dark web are the Tor network (most popular) and hidden services, the I2P network, and the RetroShare network. In general, this is not useful for reporting. It may have some use in auditing. The size is much smaller than the web or deep web and is actively hidden.[6] Auditing may have uses for analysing aspects of this, not just illicit use in a company but also fraudulent activities. For example, hundreds of thousands (our estimate) of credit cards are on there for open sale. The going rate for a credit card is around £11 to £50 (for a premium card) and for a corporate American Express card is a lot more. Email accounts go for £90 and corporate email accounts for £350.[7]
    - ⇨ The major uses for the dark web are
        a) Criminal activities such as fraud.
        b) Email spamming and referral spamming (for use in SEO activities).
        c) Sale of semi-legal, illegal, or recreational or other drugs.
        d) Pornography that is semi-legal or illegal such as child pornography.
        e) Dodgy ticket sales.

[3] Such as Microsoft's Internet Explorer, Google's Chrome, Mozilla Firefox, Opera, or Apple's Safari.
[4] Some studies have put this as high as 80% but these are based on a biased sample. Some studies have put this as low as 4% but this is probably an understatement. The 10% and 30% figure is our best estimate.
[5] Web indexing refers to various methods for indexing the contents of a website by search engines making it easier and faster to respond to users search queries.
[6] Tor is the largest but not the only dark or overlay network. Sizes of other dark networks such as I2P and RetroShare are hard to gauge; however based on popularity it would appear reasonable to guess that, in total, the size of dark networks combined is far smaller than the deep web, and highly likely to be smaller than the public Internet.
[7] https://cointelegraph.com/news/dark-web-stolen-credit-card-details-going-for-%C2%A311-a-piece

f) Sale of stolen property.
g) Sales of credit, debit cards and usernames and passwords (for example to bank accounts).
h) Online bank account details.
i) Guns, ammunition, weapons, rocket launchers, explosives, and so on.
j) Other forms of criminal activity.

- A private network is a computer network which is reserved for specific purposes, e.g. company networks, including financial and accounting data.
- An overlay network is a computer network which is built on the top of another network. Nodes in the overlay can be thought of as being connected by virtual or logical links, each of which corresponds to a path, perhaps through many physical links, in the underlying network.

Everything that is not dark but is not publicly accessible can be classified as grey. Although much of the dark web is innocuous, some prosecutors and government agencies, among others, are concerned that it is a haven for criminal activity.[8]

Commercial dark net markets[9], which mediate transactions for illegal drugs and other goods, attracted significant media coverage starting with the popularity of Silk Road[10] and its subsequent seizure by legal authorities. Other sites sell bugs or malware, weapons, and suggest crowdfunded assassinations and hitmen. Sites associated with Bitcoin[11], fraud related services and mail order services are some of the most prolific. See Table 13.6.

[8] A December 2014 study by Gareth Owen from the University of Portsmouth found that the most popular type of content on Tor was child pornography, followed by black markets, while the individual sites with the highest traffic were dedicated to botnet operation. A botnet is a number of Internet computers that, although their owners are unaware of it, have been set up to forward transmissions (including spam or viruses) to other computers on the Internet.

[9] Specialist news sites such as DeepDotWeb and All Things Vice provide news coverage and practical information about dark web sites and services.

[10] Silk Road was an online black market, best known as a platform for selling illegal drugs. As part of the dark web, it was operated as a Tor hidden service, such that online users were able to browse it anonymously and securely without potential traffic monitoring. The website was launched in February 2011; new sellers had to purchase an account in an auction. Later, a fixed fee was charged for each new seller account. In 2013 the FBI shut down the website.

[11] Bitcoin is a payment system where users can transact directly without needing an intermediary. Transactions are verified by network nodes and recorded in a public distributed ledger called the block chain. The ledger uses its own unit of account called bitcoin. The system works without a central repository or single administrator, which has led the US Treasury to categorize it as a decentralized virtual currency.

Table Nomenclature of the Net/Web

Type	Name	Description	Notes and examples
Grey	Surface	Public web	Accessible web via search engines
Grey	Deep Or limited access Or via JavaScript	Unlinked public pages or content Private web Intranet Web archives Cloud	Not indexed
Grey	Non-Http	Phone data, messaging, Emails and other data	iPhone/iOS, Android or other
	Or software	Messaging systems	Facebook (WhatsApp, Instagram & Messenger) Microsoft/Skype Viber
		Social media	Facebook, Twitter, Snapchat, LinkedIn, Google+/YouTube, Yahoo (Tumblr, Flickr), Reddit, Pinterest, Tinder
		Music	Spotify, iTunes, Shazam, Google Play
		Other Apps	close to a million apps
		Organisational data capable of being accessed remotely	
		Other private or commercial data	
Grey	Dark	Indexed via Tor, etc.	Dream market
		Market sites	Silk Road 3 Tochka WallStreet market
	Specialised software	Various	Various

The World Wide Web: The internet

This is perhaps the most important development together with mobile phones. Table 13.7 shows the growth and our forecast of future growth.

Table Growth of the Internet (Surface and Deep web)

	1995	**2000**	**2005**	**2010**	**2015**	**2020**	**2025**	**2030**
	Millions	Millions	Millions	Millions	Millions	Millions	Millions	Millions
						estimated	estimated	estimated
Number of users	16	359	938	1,966	3,300	5,000	7,000	8,000
% of world population	0.40%	5.90%	14.60%	28.70%	45.50%	65.50%	87.60%	96.20%
Number of pages that can be indexed by search engines (e.g. Google)			11,500	30,000	142,000	940,000	3,700,000	14,500,000
Number or URLs				1,000,000	4,700,000	31,000,000	122,000,000	470,000,000
Number of domains	1	10	53	95	300	1,000	1,600	2,100
Number of domains in deep & dark webs				47,500	165,000	600,000	1,120,000	1,680,000

Source: Compiled by authors; predictions also based authors' algorithms

Nearly 50% of the world's population uses the internet. Within the next decade or two, nearly 90% of the world's population will be on the internet.

The number of domains that can be found by search engines such as Google and Bing (Microsoft) is already around 100 million and the deep and dark net (defined later) has about 1,000 times that number and is rising. So the number of pages which are examined (or indexed) by the search engines is already in excess of 1 trillion and should reach half a quadrillion by 2030 – if growth continues at the current rate..

Big data, social media and grey data

Big data is sometimes defined as extremely large data sets, including all grey data that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions. The distinguishing feature is that in order to find meaningful patterns, much IT investment is necessary to manage and maintain that data. Wikipedia defines big data as:

> Big data is data sets that are so voluminous and complex that traditional data-processing application software are inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualisation, querying, updating, information privacy and data source[12].

In theory, social media are computer-mediated technologies that facilitate the creation and sharing of information, ideas, career interests and other forms of expression via virtual communities and networks. In practice, they have split down to Facebook type sites, Twitter type sites, messaging (WhatsApp, Messenger etc.) and then specialist sites dealing with photos, music, dating, or a host of other functions.

Private Eye[13] poked a gentle tease of the Big Four in their big data endeavours:

> The Facebook scandal again exposes how the big accountancy-cum-consultancy firms' sprawling, conflicted businesses can be relied on to overlook shady practices.
> Last year, it has just been revealed, an audit by PricewaterhouseCoopers (commissioned by the US's Federal Trade Commission) of Facebook's privacy practices between 2015 and 2017 revealed that, er, there was nothing to see. 'In our opinion, Facebook's privacy controls were operating with sufficient effectiveness to provide reasonable assurance to protect the privacy of covered information', PwC declared. Never mind that Facebook's use and abuse of privacy settings was simultaneously enabling the exploitation of around 50m users' data for nefarious purposes – as the Cambridge Analytica episode has now exposed.
> The world's second largest accountancy firm, which is meant to root out trouble rather than bury it, happens to have plenty of interests of its own in 'big data'. Its consulting division advertises how it will help businesses with 'data: your new superpower'. It can 'boost profits through strategic pricing models and data-driven marketing'. PwC puts it: 'The result? Maximised profits.'

[12] https://en.wikipedia.org/wiki/Big_data
[13] Private Eye, 4 May 2018 and it is in Issue No. 1469

> Among the firm's own more profitable business lines is a formal alliance with that other great data monster, Google. 'PwC and Google for Work bring to market both innovative disruptive technology and deep industry and process knowledge required to help clients reinvent, innovate and transform their businesses,' drones the firm. 'Together we have developed joint solution offerings to help our clients drive profitable outcomes and reinvent their business.'
> All PwC's 'Big Four' rivals are climbing aboard the data gravy train too. EY, for example, has a burgeoning line in the use of healthcare data. 'When the human body is the biggest data platform,' it asks in its not-all-that-sinister marketing blurb, 'who will capture value?'

Ecommerce
Electronic commerce is a way of doing business over large electronic networks such as the internet. Also called e-commerce, electronic commerce greatly facilitates transactions between companies and consumers (B2C), between one company and another (B2B), and between individual consumers (C2C) – such as ebay.

The issue to absorb is that the rate of growth and exponential rise of e-commerce is unstoppable. And the number of sites trying to sell within the UK from outside the UK is growing even faster. Amazon may be the major e-commerce site but many smaller sites also sell via Amazon's market place. Amazon has about 16% of the market followed by Tesco (9%) and then eBay (8%). For a list of the top e-commerce sites in the UK see:
https://ecommercenews.eu/top-500-ecommerce-retailers-uk/

For the US the following statements are true:

- E-commerce is the only trillion-dollar industry growing at a double-digit percentage rate each year
- There are 110,000 e-commerce websites generating revenue of meaningful scale on the internet
- More than 12% of the 100,000 highest-traffic websites are e-commerce, and that density clearly declines to about 10% for the long tail. E-commerce websites make up approximately 12% of the internet.
    - The largest e-commerce sites on the internet make up about 1% of the total population and generate 34% of the total revenue.
    - A distinct middle tier of e-commerce sites makes up 51% of the total population and generate 63% of the total revenue.
    - Small e-commerce sites make up 48% of the total population and generate 3% of the total revenue.

  See: https://blog.rjmetrics.com/2014/06/18/how-many-ecommerce-companies-are-there/

The other big driver in our view is big data which includes social media and the chat apps. These can influence the purchase decision. TripAdvisor has become mandatory whether you believe the reviews or not. Often such big data apps can strike fear into the heart of a small hotel owner, who will often bend over backwards to avoid a negative report. Conversely, the reverse can happen and the sheer volume of big data is important. For example, this might have

led to Edgar House in Chester to the title of the most romantic hotel in the world. (TripAdvisor 2017 survey) with 589 reviews for a seven bedroom boutique hotel.

KPMG's Astrus system looks promising and is now linked with IBM's Watson AI system. It links to more than 40,000 data sources with 88 languages and can be used by firms as well as KPMG's forensic teams.

> Businesses face increased pressure to have tighter, more systemised third-party risk management programs. Astrus has the potential to enhance an organisation's competitive edge to effectively assess their risk.
>
> The Astrus solution now includes two main product offerings: Astrus Enhanced Due Diligence Reporting – typically commissioned to respond to regulatory requirements – and Astrus Monitoring, designed to alert organizations to changes in their business partners' risk profiles.
>
> Astrus is a cost-effective solution to third-party risk management that provides organizations with a more robust and actionable approach for managing the escalating cost and complexity of compliance[14].

Of course not all is plain sailing.

[14] https://home.kpmg.com/uk/en/home/services/advisory/risk-consulting/forensic-landing/third-party-risk-management-landing/kpmg-astrus-improving-third-party-risk-management.html

Use of grey information and leveraging these IT developments

Eccles and Krzus[15] claim that there is no reason why companies and their audiences cannot use big data and analytics with cloud computing and social media to improve the creation, distribution, and consumption of integrated reports. They also say that when the power and collaborative benefits of cloud computing are brought to bear on big data analytics' applications, using information generated from many different sources, companies can significantly improve their integrated reporting and integrated thinking.

There are several issues with this statement:

1) Integrated reporting is not a done deal. Where it has started and is being used, it is far from perfect or unbiased. It is frequently not truly audited and is biased in both opinion and context. See the previous chapters – especially Chapter 11.
2) Eccles and Krzus ignore our definition of grey data which makes up probably well over 95% of all data.
3) Even big data, analytics, and other developments have yet to make any dent into financial or non-financial reporting; and even less on auditing. Where these developments have been used, they form just a tip of the iceberg.
4) Compliance and filing requirements in a largely regulatory-driven corporate reporting world have reinforced a paper-based paradigm for decades (this is acknowledged by Eccles and Krzus).

Big data does not replace traditional data and analytics. Conventional data sources, especially financial data, will still reign and be important. The new and grey data are additions which need to be handled with care. That said this will lead to a change in priorities and challenges.

A Gartner report[16] (from 2013) identified the types of data analysed within an IT and global business environment – not necessarily an accounting or reporting viewpoint. The results were:

Transactions	70%
Log data[17]	55%
Machine or sensor data	42%
Emails/documents	36%
Social media data	32%
Free-form text	26%
Geospatial data	23%
Images	16%
Video	9%
Audio	6%
Others	12%

However, this is already out-of-date. Video and Images are growing fast, emails and social

[15] Op cit page 269
[16] Survey Analysis: Big Data Adoption in 2013.
https://www.gartner.com/doc/2589121/survey-analysis-big-data-adoption
[17] Data generated by any activity on a website such as a click. This usually has a time stamp, IP (internet protocol) address, location, network paths, device, operating system and so on.

media have had the fastest growth but may level out. Many but not all of these types of data can be used in within our wider definition of reporting.