Data Ownership: Who Owns 'My Data'?

Ali M. Al-Khouri
Director General, Emirates Identity Authority, Abu Dhabi, United Arab Emirates
Professor, Identity and Security,
The British Institute of Technology and E-Commerce, London, UK

ABSTRACT

The amount of data in our world today is substantially outsized. Many of the personal and non-personal aspects of our day-to-day activities are aggregated and stored as data by both businesses and governments. The increasing data captured through multimedia, social media, and the Internet are a phenomenon that needs to be properly examined. In this article, we explore this topic and analyse the term data ownership. We aim to raise awareness and trigger a debate for policy makers with regard to data ownership and the need to improve existing data protection, privacy laws, and legislation at both national and international levels.

Keywords

Data ownership; big data; data protection, privacy

1. INTRODUCTION

Organizations today have more data than they have ever had previously. Advancements in technology play a critical role in generating large volumes of data. According to a study published by *Information Week*, the average company's data volumes nearly double every 12 to 18 months (Babcock, 2006). Databases are not only getting bigger, but they also are becoming real time (Anderson, 2011; Sing et al., 2010).

Evolving integration technologies and processing power have provided organisations the ability to create more sophisticated and in-depth individual profiles based on one's online and offline behaviours. The data generated from such systems are increasingly monitored, recorded, and stored in various forms, in the name of enabling a more seamless customer experience (Banerjee et al., 2011; Halevi and Moed, 2012; Rajagopal, 2011).

The subject of who actually 'owns' the data or, in other words, the term 'data ownership' has attracted the attention of researchers in the past few years. Data transmitted or generated on digital communication channels become a potential for surveillance. Data ownership issues are thus likely to proliferate. For instance, Facebook's famous announcement that users cannot delete their data from Facebook caused a furor, and Mark Zuckerberg (one of five co-founders of Facebook) was equally famous in his response,*"...It's complicated"*.

Indeed it is! In today's interconnected world driven by the Internet, powered by the gigabyte network operators, we leave a significant and by no means subtle scatter of data trail. The often-asked question, and the issue of discussion today, is- who owns this data? In order to answer this question, it is important for us to step back to examine the very nature of what we call *'data'*.

2. DATA: A MATTER OF INTERPRETATION

There much confusion about what 'data' really is in today's world. The truth is that data are no more than a set of characters, which—unless seen in the context of usage—have no meaning (Wigan, 1992). Data are what one uses to provide some information. The context and the usage provide a meaning to the data that constitute information. Thus, data in the stand-alone mode have no relevance and therefore no value. When there is no value in data, then one would surmise that ownership is not an issue. That is the paradox of data ownership. Figure 1 illustrates a data value pyramid developed by Accenture. The pyramid has three levels, starting from *raw data*, up to the *insights* and then the *transactions* levels. The base of the pyramid features raw, less differentiated,

and thus less valuable data. Moving up the pyramid creates larger value and revenue opportunities.

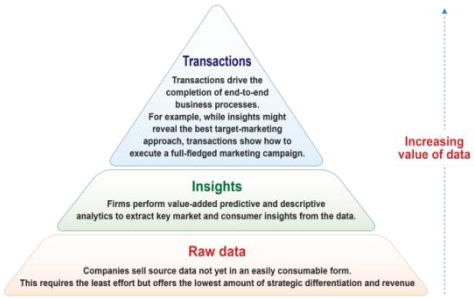

Fig 1: Data value pyramid.
Source: Banerjee et al., 2011

As such, governments and public sector institutions consider data a public utility (WEF, 2011). They tend to label our personal data as 'corporate data' and argue that without data, they cannot function (Holloway, 1988). It is no wonder that the volume of stored data in today's organisations has increased exponentially.

It is in this context that ownership of data needs to be considered. As data are generated, then data are stored. When we speak about data ownership, we refer to the storage process. If so, then the ownership of data storage resides with the owner of the storage. Thus, we as individuals, the government as our governing agent, law enforcement agencies and the courts, security agencies, our service providers, and our network operators who enable us to move our data are all our data owners. They own the storage systems and thereby the data held within such systems. In addition, the emergence of customer data integration (CDI) and of master data management (MDM) technologies has enabled the integration of disparate data from across multiple silos into commonly defined, reconciled information accessible by a range of systems and business users (Dyché, 2007).

We would like to pause here and examine the concept of 'My Data'. Just what is 'My Data'? Do we consider information of friends and family that we hold to be ‘My Data’? Do bank statements and credit statements sent by banks qualify as 'My Data'? Would the financial statement sent by a company to me as a shareholder qualify as 'My Data'?

We would argue no. 'My Data', in its strict sense, comprise just our personal attributes—no more. This is the data that I own. I use 'My Data' as information to identify myself for my personal gains, whether physical, logical, or emotional. 'My Data' are thus in the open and either ***implicitly*** or ***explicitly*** shared. When I share the data, I delegate the ownership. Thus, ***my data*** have multiple owners, and the number of owners increases with each share. Figure 2 provides an illustration of this viewpoint.

Fig. 2: Potential owners of 'My Data'

As the number of transactions increases with 'My Attributes', 'My Data' grow and, in turn, increase the number of data owners. What essentially is happening is that with every transaction, information is shared as data. Each time information is shared, new data are generated. As new data are generated, new ownership is created. The diagram (Figure 2) illustrates just a tip of the proverbial iceberg of data generation from 'My Data'!

3. PERSONAL DATA ECOSYSTEM AND OWNERSHIP

Typically, organisations can capture different personal data in a variety of ways (Marc et al., 2010):

- Data can be "volunteered" by individuals when they explicitly share information about themselves through electronic media, for example, when someone creates a social network profile or enters credit card information for online purchases;
- "Observed" data are captured by recording users' activities (in contrast to data they volunteer). Examples include Internet-browsing preferences, location data when using cell phones, or telephone usage behaviour;
- Organisations can also discern "inferred" data from individuals, based on the analysis of personal data. For instance, credit scores can be calculated based on a number of factors relevant to an individual's financial history.

Each type of personal data (see Figure 3)—volunteered, observed, or inferred—can be created by multiple sources (devices, software applications), stored, and aggregated by various providers (Web retailers, Internet search engines, or utility companies), and then analysed for a variety of purposes for many different users (end users, businesses, public organisations).

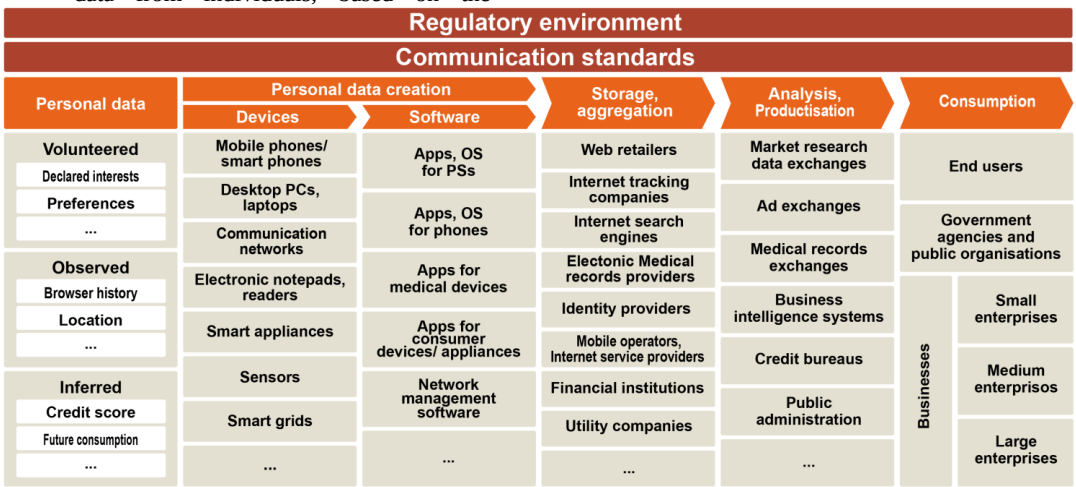

Fig. 3: The Personal data Ecosystem: A complex WEB from data creation to data consumption. Source: Bain & Company. WEF, 2011

So, in all this chaos of data generation and delegated ownership, where does true ownership lie? The answer to this question is found in truth, veracity, and therefore verifiability. Each time information is generated, a set of data related to this information is created. This data, when relayed further, should stand up to scrutiny and verification. The contention is that the source that can verify this data and confirm the veracity of the information is the 'True Owner' of the data. Figure 4 presents a conceptual model to illustrate the source of truth and data ownership.

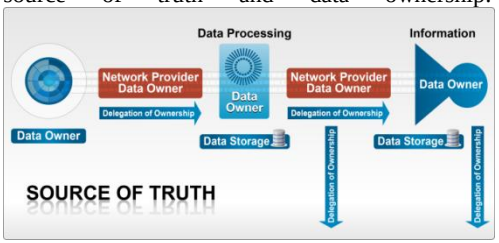

Fig. 4: Source of Truth of Data Ownership

Even in the context of complex Web transactions, this statement would remain valid. Intimately linked to the growth of 'Big Data' are such

technology trends as the growth of mobile technology and wireless devices, the emergence of self-service channels, the broad adoption of cloud-based services, and the expansion of social networking and remote collaboration (SAP, 2012). For instance, Google is synonymous with its search engine and provides a host of services that require a user to login. When a logged-in user searches the Web, data are generated related to the user's search patterns. While the search information itself does not belong to Google, the data collected on the search patterns do. Any analysis based on these search patterns can be traced back to Google's search data.

As depicted in Figure 5, "Big Data" companies collect and analyse massive amounts of data under the argument that they can spot trends and offer users niche insights that help create value and innovation much more rapidly than conventional methods. This generates more data, the analysis of which is—more often than not—as useful as the original information. This analytical information now belongs to the person and/or the organization that performed the analysis. This brings up the critical issue of data usage and information usage.

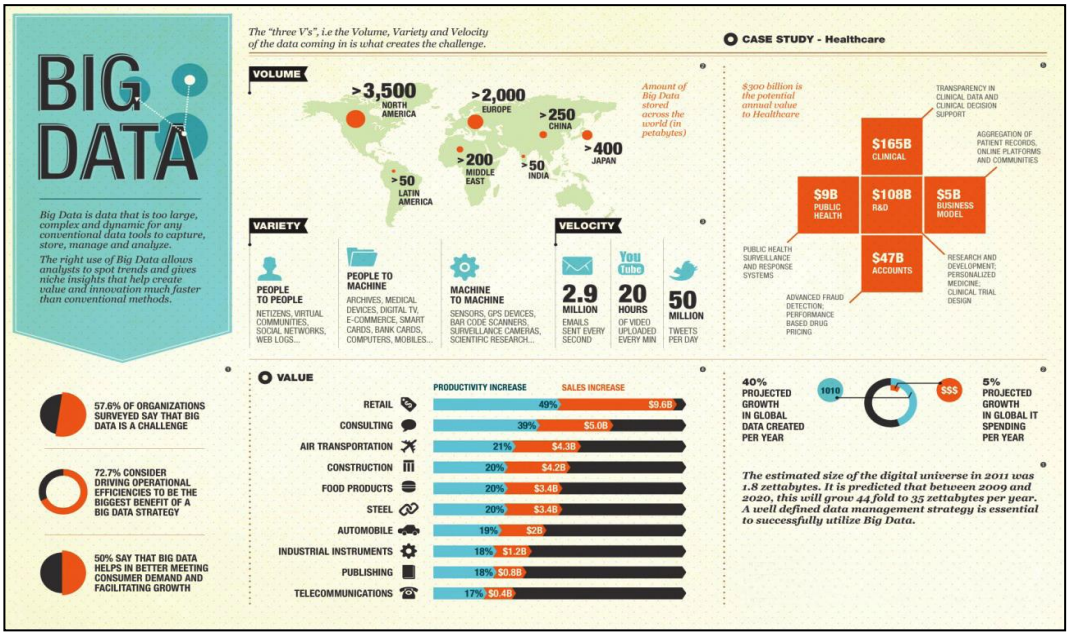

Fig. 5: Big Data. Source: http://www.brainflash.com/2012/10/18/big-data-is-that-a-career/

For example, in a well-publicized incident that occurred in August 2006, America Online published a dataset of search results. These data were collected from the searches conducted by users and were intended to provide analytical material to researchers. The data published were anonymous, without any reference to the users who carried out the searches. The searchers' identities were distinguished as numbers. Five days later, however, *The New York Times* was able to locate one of those searchers by linking her search history to other public data, such as the phonebook (Barbaro and Zeller, 2006).

4. THE NEED TO REDEFINE THE ECOSYSTEM

So who possesses the right to use the information and data that truly do not belong to one's self? This is an issue that transcends borders of commerce, ethics, and morals, leading to privacy issues and the protection of privacy. It is trivial

that the current personal data ecosystem is fragmented and inefficient (WEF, 2011). For many participants, the risks and liabilities exceed the economic returns. On the other hand, personal privacy concerns are inadequately addressed, and current technologies and laws fall short of providing the legal and technical infrastructure needed to support a well-functioning digital infrastructure (ibid.). Instead, they represent a patchwork of solutions for collecting and using personal data in support of different institutional aims, and subject to different jurisdictional rules and regulatory contexts (e.g., personal data systems related to banking have different purposes and applicable laws than those developed for the telecom and healthcare sectors).

It is of importance that governments play a more active regulatory role in modernising their existing policy frameworks to protect personal data from the unlawful processing of any data (Robinson et al., 2009). The government should move away from a regulatory framework that measures the adequacy of data processing by measuring compliance with certain formalities, and towards a framework that instead requires certain fundamental principles to be respected, and that has the ability, legal authority, and conviction to impose harsh sanctions when these principles are violated (ibid.).

A recent report published by the World Economic Forum recommended that the personal data ecosystem be debated and redefined (WEF, 2012). This prompts all stakeholders to come to a consensus on some key areas, including the security and protection of data, development of accountability, and agreements on principles or rules for the trusted and allowed flow of data in different contexts. See also Table 1

.**Table 1. Key principles to guide the development of the personal data ecosystem.**

Guiding Principle	Description
Accountability	Organizations need to be held accountable for appropriate security mechanisms designed to prevent theft and unauthorized access of personal data, as well as for using data in a way that is consistent with agreed upon rules and permissions. They need to have the benefit of "safe harbour" treatment and insulation from open-ended liability, when they can demonstrate compliance with objectively testable rules that hold them to account.
Enforcement:	Mechanisms need to be established to ensure organizations are held accountable for these obligations through a combination of incentives, and where appropriate, financial and other penalties, in addition to legislative, regulatory, judicial, or other enforcement mechanisms.
Data permissions:	Permissions for usage need to be flexible and dynamic to reflect the necessary context and to enable value-creating uses, while weeding out harmful uses. Permissions also need to reflect that many stakeholders— including but not limited to individuals—have certain rights to use data.
Balanced stakeholder roles:	Principles need to reflect the importance of rights and responsibilities for the usage of personal data and strike a balance between the different stakeholders—the individual, the organization, and society. They also need to reflect the changing role of the individual from a passive data subject to an active stakeholder and creator of data. One perspective that is gathering momentum, though it is far from being universally accepted, is that a new balance needs to be struck that features the individual at the centre of the flow of personal data, with other stakeholders adapting to positions of interacting with people in a much more consensual, fulfilling manner.
Anonymity and identity:	The principles need to reflect the importance of individuals being able to engage in activities online anonymously, while at the same time establishing mechanisms for individuals to effectively authenticate their identity in different contexts, so as to facilitate trust and commerce online.
Shared data commons:	The principles should reflect and preserve the value to society from the sharing and analysis of anonymised datasets as a collective resource.

Source: WEF, 2012

These principles should be global in scope, but also applicable across sectors and focused beyond merely minimizing data collection, storage, and usage of data to protect privacy. The principles need to be built on the understanding that to create value, data must move, and moving data requires the trust of all stakeholders. Organisations will need to develop and implement a comprehensive data governance program that should be based on these guiding principles. This should help organisations to design and implement more comprehensive structures and to put in place solid accountability that altogether establishes a coordinated response to key issues of trust, transparency, control, and value.

5. CONCLUSION

The term 'data ownership' is likely to attract more attention from both practice and research fields. The private sector will continue to use data as a source of competition and growth. Advocators will always justify their practices that this contributes to productivity, innovation, and competitiveness of entire sectors and economies (Manyika et al., 2011). Governments will need to play a more active role to protect citizens' privacy rights, in light of the evolving world we live in today.

Governments will inevitably need to redesign and enforce data protection privacy laws and legislations. This will require establishing policies at both the national and international levels. As such, governments will need to open up dialogue to establish comprehensive data protection and privacy laws that could be implemented globally. This should be followed by a clearly articulated set of standards, policies, procedures, and responsibilities regarding data ownership and data-related activities that may minimize any detrimental outcomes in an event of a data breach (PTACT, 2010). Governments should also focus on enforcing transparency. Public education programs might be a good initiative to support understanding of how individuals can protect their personal data, and how such data are being stored and used (Manyika et al., 2011).

As time passes, we are likely to see increasing public concerns about privacy and trust in today's interconnected online environments. Governments will need to help the public to understand where they should position themselves within this spectrum. It will be challenging times for governments to keep up with the pace of technology development, and those lagging will have a hard time indeed.

6. REFERENCES

[1] Alstyne, M.V., Brynjolfsson, E. and Madnick, S.E. 1994. Why not One Big Database? Principles for Data Ownership. http://dspace.mit.edu/bitstream/handle/1721.1/2516/SWP-3695-31204002-CISL-94-03.pdf?sequence=1 [Accessed October 2, 2012].

[2] Anderson, J.H. 2011. Real Time Systems Resource Management. Real Time Systems. Vol. 47. No.5. pp.387-388.

[3] Anderson, R. and Roberts, D. 2012. Big Data: Strategic Risks and Opportunities: Looking Beyond the Technology Issues. http://www.crowehorwath.net/uploadedFiles/Crowe-Horwath-Global/tabbed_content/Big%20Data%20Strategic%20Risks%20and%20Opportunities%20White%20Paper_RISK13905.pdf [Accessed October 2, 2012].

[4] Babcock, C. 2006. Data, Data, Everywhere. Information Week. http://www.informationweek.com/data-data-everywhere/175801775 [Accessed October 3, 2012].

[5] Banerjee, S., Bolze, J.D., McNamara, J.M. and O'Reilly, K.T. (2011) How Big Data Can Fuel Bigger Growth. Accenture. http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Outlook-How-Big-Data-can-fuel-bigger-growth-Strategy.pdf [Accessed October 12, 2012].

[6] Barbaro, M. and Zeller, T. 2006. A Face Is Exposed for AOL Searcher No. 4417749. http://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted=all&_r=2& [Accessed October 15, 2012].

[7] Davenport, T.H., Eccles, R.G. and Prusak, L. 1992. Information Politics. Sloan Management Review, pp. 53-65.

[8] Dyché, J. 2007. A Data Governance Manifesto: Designing and Deploying Sustainable Data Governance. http://www.siperian.com/documents/WP_Jill Dyche_DataGovernance_June07.pdf [Accessed October 7, 2012].

[9] Evans, B.J. (2011) MUCH ADO ABOUT DATA OWNERSHIP. Harvard Journal of Law & Technology, Volume 25, Number 1. pp. 70-130.

[10] Grant, J. and Kirchmaier, T. 2004.Corporate Ownership Structure and Performance in Europe. London School of Economics. http://eprints.lse.ac.uk/19960/1/Corporate_Ownership_Structure_and_Performance_in_Europe.pdf [Accessed October 12, 2012].

[11] Halevi, G. and Moed, H. 2012. The Evolution of Big Data as a Research and Scientific Topic. Research Trends: Special Issue on Big Data, Issue 30. http://www.researchtrends.com/wp-content/uploads/2012/09/Research_Trends_Issue30.pdf [Accessed October 12, 2012].

[12] Hart, D. 2000. Data Ownership and Semiotics in Organisations, or Why "They're Not Getting Their Hand on My Data!. http://www.pacis-net.org/file/2000/377-388.pdf [Accessed October 14, 2012].

[13] Holloway, S. 1988. Data Administration, Gower, Aldershot.

[14] Khan, S.M. and Hamlen, K.W. 2012. Anonymous Cloud: A Data Ownership Privacy Provider Framework in Cloud Computing. http://www.utdallas.edu/~hamlen/khan12trustcom.pdf [Accessed October 21, 2012].

[15] Loshin, D. 2001. Enterprise Knowledge Management: The Data Quality Approach (The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann.

[16] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H. 2011 Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights and pubs/MGI/Research/Technology and Innovation/Big Data/MGI_big_data_full_report.ashx [Accessed October 11, 2012].

[17] Marc, D., Martinez, R., and Kalaboukis, C. 2010. "Rethinking Personal Information – Workshop Pre-read." Invention Arts and World Economic Forum.

[18] PTACT.2010.Data Governance and Stewardship. e Privacy Technical Assistance Center. http://www2.ed.gov/policy/gen/guid/ptac/pdf/issue-brief-data-governance-and-stewardship.pdf [Accessed October 13, 2012].

[19] Rajagopal, S. 2011. Customer Data Clustering Using Data Mining Technique. International Journal of Database Management Systems. Vol.3, No.4, pp. 1-11.

[20] Robinson, N., Graux, H., Botterman, M. and Valeri, L. 2009. Review of the European Data Protection Directive.

[21] SAP.2012.Harnessing the Power of Big Data in Real Time through In-Memory Technology and Analytics. http://www3.weforum.org/docs/GITR/2012/GITR_Chapter1.7_2012.pdf [Accessed October 13, 2012].

[22] Schnarch, B. 2004. Ownership, Control, Access, and Possession (OCAP) or Self-Determination Applied to Research: A Critical Analysis of Contemporary First Nations Research and Some Options for First Nations Communities. http://www.research.utoronto.ca/ethics/pdf/human/nonspecific/OCAP%20principles.pdf [Accessed October 12, 2012].

[23] Shields, G. 2010. Addressing Security and Data Ownership Issues when Choosing a SaaS Provider. http://www.quest.com/quest_site_assets/whitepapers/wpw_saas_shields_us_mj.pdf [Accessed October 11, 2012].

[24] Singh, Y.J., Singh, Y.S., Gaikwad, A. and Mehrotra, S.C. 2010. Dynamic management of transactions in distributed real-time processing system. International Journal of Database Management Systems, vol. 2, no.2. pp.161-170.

[25] WEF.2011. Personal Data: The Emergence of a New Asset Class. World Economic

Forum. http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf [Accessed October 16, 2012].

[26] WEF.2012. Rethinking Personal Data: Strengthening Trust. http://www3.weforum.org/docs/WEF_IT_RethinkingPersonalData_Report_2012.pdf [Accessed October 21, 2012].

[27] Wigan, M.R. 1992. 'Data Ownership' in Clarke R.A. & Cameron J. (Eds.) 'Managing the Organisational Implications of Information Technology, II' Elsevier / North Holland, Amsterdam.

[28] Woodbury, C. 2007. The Importance of Data Classification and Ownership. http://www.srcsecuresolutions.eu/pdf/Data_Classification_Ownership.pdf [Accessed October 9, 2012].