Ethical Use of Big Data

Definition: The opening up of government, corporate and private data as massive online data sources is of great potential value, from apps that show us crime hot spots to databases that enable diseases to be understood and cured and victims of disasters to be located. However, alongside the benefits, there is also the threat of abuses, such as the creation of an infrastructure of permanent surveillance, the discovery of the whereabouts of victims, disclosure of socially stigmatising diseases or conditions or the use of Big Data by fraudsters for targeting victims.

The decreasing costs of data storage and processing power facilitate the economical processing of huge quantities of largely unstructured data. This makes consideration of the ethical use of Big Data *urgent*, it is also *important* because processing of Big Data unethically may lead to a variety of societal ills.

Governments, businesses and other organisations are collecting vast amounts of data purely because they might want to use it at some time in the future, rather than just collecting data for specific purposes (as was the case in the past). These largely unstructured sets of data may well contain data that is personally identifiable and data sets that, when combined with other data, can be de-anonymised. The public is generally unaware of this and it can represent both an invasion of privacy and an asymmetry between the interests of the individual and those of the state/ organisations. Much of the data collected by governments is done under statute and the individual has no option but to provide it. The unconstrained sharing of government-sourced data can be harmful to the individual and undermine public trust in government.

Background/Context: Traditionally data has been numbers, such as observations of the natural world and accounting figures. This data has been collected and put into structured formats for analysis. If the data sets contained personal data they have been subject to the provisions of the Data Protection Act 1998 (DPA). The DPA is based on eight principles which state that personal data must be:

1. Processed fairly and lawfully;
2. Obtained and only used for specified and lawful purposes;
3. Adequate, relevant and not excessive;
4. Accurate, and where necessary, kept up to date;
5. Kept for no longer than necessary;
6. Processed in accordance with the individual's rights;
7. Kept secure;
8. Only transferred to countries that offer adequate data protection.

The DPA places additional restrictions on the processing of 'sensitive data' such as that concerned with medical conditions, religious beliefs, ethnicity, sexual orientation, trade union membership, or criminal conviction history.

The distinguishing feature of Big Data is that it is largely unstructured, including free form text and video. It is also frequently: neither peer reviewed, nor audited, nor necessarily accurate, nor kept in a single global jurisdiction. It is often collected and kept 'in case it is needed in the future' rather than for a specific purpose to address an identified problem or test a hypothesis. This means that if Big

Data includes personal data, it is likely to fall foul of principles 2, 3, 4, 5, 6 and 8 of the DPA. Even if the original data sets do not contain personally identifiable data, it is likely that when several of these data sets are combined and 'mashed up' identity discovery will be possible.

Key issues: the key areas of concern in the rise of Big Data relate to: trust in the use of internet; transparency; the balance of costs and benefits (risks and rewards) associated with the use of Big Data and with whom those lie; the efficient delivery of services; the boundaries of the use of data for individual and societal good; intentional or unintentional discrimination; exploitation of the vulnerable; abuse of power; the security of data and citizens. The ethical issues raised by the collection, combination and analysis of Big Data revolve around:

- The value and monetisation of data;
- Data protection and censorship;
- Privacy and surveillance;
- The quality of data and fitness for purpose;
- The analysis of Big Data as a predictor of individual actions in the future.

BCS position on the key issues:
BCS believes that there must be an ethical approach to the use of Big Data. This is summarised as follows:

- Transparency, control and consent are fundamental to the ethical use of Big Data. BCS will press for all organisations (public, private and third sector) to ensure that individuals are made aware and give their consent when data related to them is being collected, combined and analysed. The BCS will push government, corporations and the not for profit sector to be open about their collection, use and monetisation of Big Data. The OECD's seven principles for the protection of personal data (1980) on which the DPA was based remain core to the necessary transparency in relation to Big Data (although it can be argued that some of the principles are becoming ineffective or impossible in the Big Data era). The principles are:

    1. Notice - data subjects should be given notice when their data is being collected;
    2. Purpose - data should only be used for the purpose stated and not for any other purposes;
    3. Consent - data should not be disclosed without the data subject's consent;
    4. Security - collected data should be kept secure from any potential abuses;
    5. Disclosure - data subjects should be informed as to who is collecting their data;
    6. Access - data subjects should be allowed to access their data and make corrections to any inaccurate data; and
    7. Accountability - data subjects should have a method available to them to hold data collectors accountable for following the above principles.

- BCS believes that analysts and policy makers must understand the limitations associated with the use of massive largely unstructured data sources and ensure that they derive evidence based policies from them in a way that is both scientifically and statistically correct, fair and ethical to contributors and non-contributors to those databases alike. This includes ensuring they engage with all demographic groups to understand their views on the consumption of Big Data by both businesses and government. They also need to consider the dangers related to the use of Big Data for the purpose of predicting individual actions;
- We feel that more research needs to be undertaken to understand better how mass observation and constant surveillance affects the quality of data and whether certain types of policy and service decisions are better made with smaller rather than bigger quantities of data;
- Users of Big Data must be made aware of the likelihood that attribute combination will de-anonymise Big Data sets in whole or in part and that when such data sets are de-

anonymised, they become personal data and must be treated as required by the Data Protection Act. This is particularly important where sensitive data is concerned;

- BCS endorses the push by governments and business to grow the online economy and ensure that the UK is in the forefront of deriving benefits from the online world.
- BCS will push for ethical scrutiny of the collection and analysis of massive open data online so that it can be used for social good, in a way that is proportionate to any potential individual privacy concerns;
- BCS will strive to educate the general public about the value of their personal data and how it is or may be used or abused by users and organisations, so that the public can make their own decisions about the balance of risks and rewards for themselves.
- BCS supports the view that the internet should remain an open platform available to all and will resist attempts at Balkanisation;
- BCS will promote the European Union and OECD approach to internet governance encapsulated in the COMPACT acronym, which states that the internet is a space of:
    - **C**ivic responsibilities,
    - **O**ne un-fragmented resource governed via a
    - **M**ulti-stakeholder approach to
    - **P**romote democracy and Human Rights, based on a sound technological
    - **A**rchitecture that engenders
    - **C**onfidence and facilitates a
    - **T**ransparent governance both of the underlying internet infrastructure and of the services which run on top of it.

A fuller account of these issues can be found in the associated background paper: Background to BCS Position Statement on the ethical use of Big Data – Louise Bennett 2014

Useful documents:

BCS Personal Data Guardianship Code
BCS Aspects of Identity Yearbooks
Feedback to Parliament and Internet Conference, October 2013
Ethics of Big Data Position and background papers
Patient Information Governance and Caldicott 2 Review
Keeping your online health and social care records safe and secure
Hospital Data and Data sets collection consultation response
NAPC paper – Patient Data Governance – is an armistice in sight? Summer 2013

Louise Bennett, Chair Community of Expertise (author)
Andrew Cormack, Community of Expertise (spokesperson)
Justin Whatling, Chair BCS Health (spokesperson)