Data Protection Office

Open Data and Data Privacy

Mrs Drudeisha Madhub
Data Protection Commissioner
Data Protection Office
http://www.dataprotection.govmu.org

23-May-17

Introduction

- Governments collect and store a wide range of data that may contain personal and sensitive information. As governments embrace open data initiatives, more of this information may be released to the public.

- Releasing data can increase transparency in government, allow citizens to engage with the public agencies, and empower entrepreneurs to build innovative tools and applications. However, open data comes with inherent risks to an individual's privacy.

- The Data Protection Act in Mauritius regulates the processing of personal data and safeguards privacy rights of individuals. Thus, data controllers including ministries/departments and other public bodies should abide by the Data Protection Act while identifying datasets for open data.

Privacy Challenges/Risks

1. A balanced approach to open data

- The focus of open data is on transparency, accountability of public sector bodies and economic growth. But, it is not on the transparency of individual citizens.

- A balanced approach needs to be followed. On the one hand, rules for the protection of personal data should not constitute an undue barrier to the development of the open data. On the other hand, the right to the protection of personal data and the right to privacy must be respected.

2. Privacy Impact Assessment

- Public sector organisations are recommended to carry out a thorough data protection impact assessment in order to establish whether datasets identified may be made available and if so, under what conditions and safeguards. In carrying out this evaluation the potential impact on the data subjects should be carefully considered.

- A Privacy Compliance Assessment tool is available on the website of the Data Protection Office as a means by which business and government can proactively identify and avoid privacy breaches.

3. The Decision Process

- Whenever possible, the analysis prior to making the data open, should be based on an informed debate and the representation of diverse stakeholders, including:
  - the data controller of the information,
  - data protection authority,
  - Information and Technology authorities,
  - national authorities responsible for freedom of information, and
  - other representatives of different organisations involved.

4. Informing data subjects

- It is good practice, whenever possible, to take a proactive approach and define in advance the datasets that could be made publicly available. Data subjects can then be informed, at the time of collection of the data, whether any data they provide, will become public.

5. Anonymisation and aggregation of data

- Public sector body might convert personal information into anonymised form (usually into aggregated statistical data) and make only such anonymised data available as open data.

- Deciding what level of aggregation may be appropriate and what specific anonymisation techniques to use is a challenging task. Remember, if aggregation and anonymisation are not done effectively, this carries the risk that individuals may nevertheless be re-identified from these datasets. Therefore, account should be taken of all the means likely reasonable to be used either by the data controller or by any other person to identify the said person.

5. Anonymisation and aggregation of data (Cntd)

- Unless data can be anonymised to render it irreversible, data protection law continues to apply. This means, among others, that unless the irreversibility condition is met, the public release of the information must be ‘compatible’ with the initial purposes of data collection under the second principle of the Data Protection Act.

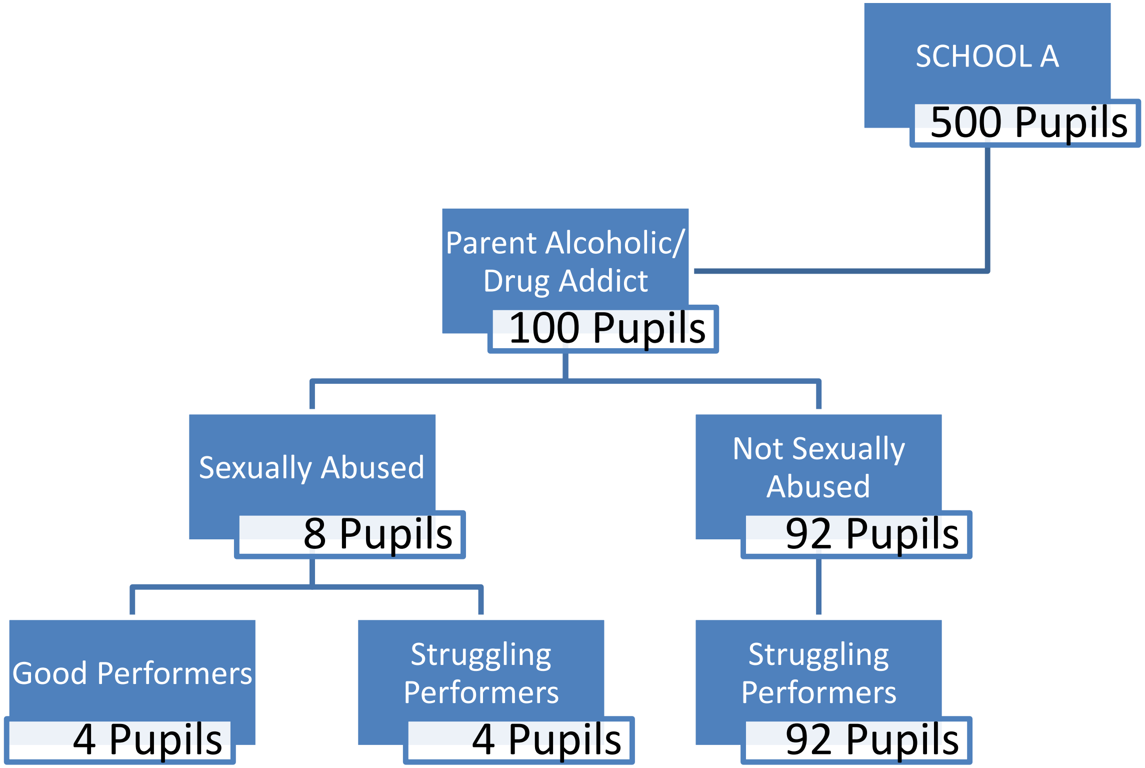
6. Data Linkages and re-identification

- Under the Data Protection Act, there is no “reasonableness” test defined to establish whether anonymised data can be re-identified to personal data. Re-identification risks can increase where one individual or group of individuals already knows a great deal about another individual, for example, a family member, a colleague, contact on a social networking site, a doctor, teacher, a law enforcement agent, or another professional.

- What matters here, however, is not simply whether the individual with prior knowledge can identify the data subject concerned but whether he/she will learn something new from the information obtained through re-identification.

- Let me take an example on a purportedly 'anonymised' research data published with good intentions but without a careful assessment of re-identification risks.

6. Data Linkages and re-identification (Cntd)



- At the school it is common knowledge that XX, a bright and hard-working boy has a difficult family background, and his mother is an alcoholic. He is often bullied by some of his classmates. These same classmates now detect from the statistics re-published in the school paper that XX must fall into amongst the 8 pupils who are sexually abused and are good performers.
- Thus, they have learned new (and in this case very sensitive) information from an ineffectively anonymised dataset.

6. Data Linkages and re-identification (Cntd)

- Therefore, the risk of combining information to produce personal data increases as data linkage techniques and computing power develop, and as more potentially ‘match-able’ information becomes publicly available. Indeed, computational power is doubling every year and data storage, due to the availability of cloud services. Thus, the risk of re-identification through data linkage is unpredictable because it can never be assessed with certainty what data are already available or what data may be released in the future.

Conclusion

- Every stage of the open data lifecycle involves actions and decisions that have consequences for individual privacy. Responsibly opening data to the public involves a process far more complex than just uploading data. Thus, privacy should be considered at each stage of the open data lifecycle

Thank You