Calculating the Risk to Data Subjects

This paper describes a method used to calculate the risk (impact and probability) of a 'personal data breach' associated with a range of systems in an organisation. The purpose is not to end up with a pure mathematical measure of risk but to allow an organisation to understand the relative risk between systems and inform prioritisation. It is loosely based on the methodology described in ISO/IEC 27005:2011.

It has lots of limitations, it is a work-in-progress, please feel free to re-use and provide feedback.

Step 1 – Identify relevant data subjects

The organisation identifies the main classes of data subjects whose data is processed by the organisation. For a commercial organisation, this is typically workers (and ex-workers), customers and prospective customers.

Step 2 – Develop a catalogue for all systems

Each system is reviewed and the following matrix completed. Company specific clusters should also be defined where data elements are typically processed together.
(Matrix is show completed with an example system).

System Name	Salesforce CRM	
Data Subjects	Customers	Workers
Number of data subjects whose data is stored	2 million	50
Number of data subjects whose data is processed / transmitted annually	400,000	50
Data Types Processed (Y/N?		
General Personal Data *Names, Addresses, Demographic*	Y	Y
Financial Data *Bank account or payment card numbers*	Y	N
Transactional Data *Orders / Order history*	Y	N
Special Category Data *As defined by Article 9(1), excluding health*	N	N
Health Data *Data relating to a person's physical or mental condition*	N	N
Criminal Data *Criminal records, allegations*	N	N
Location data *Physical location, IP addresses GPS data*	Y	Y
Employment and performance data *Appraisals, salary, disciplinary, performance metrics, timekeeping*	N	N
Image data *Photographs, Video, CCTV*	N	N

Step 3 - Estimate the Impact of a Personal Data Breach on each class of Data Subject

Using an organisation-wide scale (example below) estimate, for each class of data subject assess:

A. The worst-case impact of as breach of a record/file – as typically there will be a few records in a system that if breached would have a much higher impact on an individual's fundamental rights and freedoms than most of the records. A good example is an HR system where a few records may contain particularly sensitive (in the ordinary sense of the word) information, whereas most do not.
B. The impact of a breach of a typical record / file

Score	1	2	3	4	5
Short Description	Minor	Low	Medium	High	Critical
Breach of Confidentiality, integrity or availability of personal data resulting in					
Privacy	Disclosure of travel itinerary Disclosure of address / contact information	Limited disclosure of financial or special category data	Theft of identity / criminal use of identity. Risk of criminal use of payment instrument	Irretrievable control over data that the data subject would consider sensitive.	Life changing damage to career, personal life or reputation
Financial	financial loss up to £100	financial loss up to £1000	financial loss between 1K and 10K	financial loss up to £10K	Loss of assets, financial loss over £10K
Mental integrity	Short-term (day/ few days) stress	Worry, anxiety Temporary effects on mental state	Medium term effects on mental state (< 1 year)	Long term effects on mental state (> 1 year)	Permanent effects on mental state
Physical integrity	Physical discomfort	Hospital outpatient required	Hospital in patient required	Chronic condition	Death or life changing injury
Employment	Worry about discrimination in role	Discrimination in role	Significant discrimination in role	Loss of employment	Loss of employment/ inability to secure next role

It can be useful to create a baseline impact score for each data subject / cluster combination.

It is important that impact (and probability) are determined consistently because this exercise is used for prioritisation, so relative results are more important than absolutes.

Step 4 - Estimate the Probability of a Breach of many records

Using the organisation-wide scale estimate the probability of a breach affecting many/all of the data subjects whose data is processed in the system

Score	1	2	3	4	5
Short Description	Unlikely	Possible	Likely	Probable	Certain
Probability of an event occurring within 24 months	< 10%	11% - 40%	41-70%	71%-90%	>90%

Step 5 – Calculate the Population Breach Risk Score

For each system / data subject combination calculate a population risk score using the following variables.

A) Worst impact of a breach in confidentiality of a single record (1-5)
B) Typical impact of a breach of confidentiality (1-5)
C) Number of records
D) Total 'population' impact of a typical breach in confidentiality of the system (B^2*C)
E) Probability of a breach in confidentiality of the system (1-5)
F) Total population risk (total population impact x probability D*E)

And if applicable consider availability and integrity
Would a lack of availability or integrity also affect someone's fundamental rights and freedoms? This tends to be context specific

Worked example
A system used to send emails to ask for customer feedback that contains 1 million records containing personal contact data, summary of last order and date

Measure	**Scores**	**Reason**
Worst impact of a breach in confidentiality of a single record (A)	2	Privacy
Typical impact of a breach of confidentiality (B)	1	Privacy
Number of records (C)	1,000,000	
Population impact (D=B^2 * C)	1,000,000	
Probability of population breach (E)	3	Data sent by unencrypted email to unreliable third party
Population risk (F=D * E)	3,000,000	

Step 6 – Rank systems in order of priority

Systems should be ranked by Population Risk Score in groups of data subjects. For example:

Data Subjects: Customers

System	Worst Case Impact	Population Risk Score
System A	4	958,000,000
System B	4	623,000,000
System C	3	432,000,000
System D	3	281,000,000
System E	3	178,100,000
System F	3	178,000,000
System G	2	178,000,000
System H	1	80,000,000
System I	1	70,000,000

Data Subjects: Workers

System	Worst Case Impact	Population Risk Score
System Z	3	3,920,000
System Y	5	2,250,000
System X	4	700,000
System W	5	400,000
System V	4	400,000
System U	3	400,000
System T	3	400,000
System S	3	360,000
System R	5	230,000
System Q	3	200,000
System P	4	180,000

Remember that the absolute score isn't the important figure in this methodology – it is just designed to record / analyse priorities.

Limitations / Issues

- Balancing the risk of a single record containing 'worst case' data with over-assessing the whole system – currently manually adjusted by visual scan of spreadsheet looking at the worst-case column.
- Assuming a public health approach to consider total population risk 'accepts' risk of significant harm to a small number of individuals in the total population. Is this fair?
- Dealing with environments where some systems have 20 million records vs. a few thousand – how to balance 20 million low risks with a thousand high risks – especially across different groups of data subjects (other the case for systems holding customer data vs. colleague data).
- Tends to focus on breach of confidentiality – but availability and integrity are also important – no consideration for other privacy risks (eg risk assessment based on impact/probability of identified LINDDUN threats materialising).
- How granular to make this – it could be a rabbit-hole of work that doesn't tell you much more than a superficial exercise but which generates lots of impressive-looking paperwork.
- Inconsistent assessment by assessors even with training and a single organisational scale.
- Hard for ex-information security people to focus on fundamental rights and freedoms and the effect on individuals rather than the organisation.
- How do you assess impact with a range of data subjects. A £1,000 financial loss to one data subject is inconvenient, to someone else it is the end of the world because it is many times their net worth. How do we account for this?
- What does risk acceptance look like – how can a DPO set a threshold value?
    - What would a regulator consider reasonable/appropriate/pragmatic?
- How do you translate this back to an enterprise risk register that is based on risk to the organisation, not risk to the data subject (because believe me, you will be asked to do this).

Note

This brief practical paper was initially presented at a January 2018 practitioners' workshop on privacy impact metrics at De Montfort University that was kindly organised by Professor Eerke Boiten. It's not meant to be an academic study but a description of how I've been doing simple privacy impact assessments.

As an industry, we're at a very early stage of privacy impact assessment maturity. So please, join the discussion, contribute and help build the capability.