*All gists in the following extract are double-underlined

Last updated: December 2014

BULK PERSONAL DATA: GUIDANCE ON THE AUTHORISATION PROCESS

INTRODUCTION

It is very important that we get this process right every time because:

- **It's the law:** The data assessment and authorisation form show that SIS has considered carefully whether what we are doing is consistent with our statutory functions and is necessary and proportionate to achieve the purposes of national security, economic wellbeing of the UK or the detection/prevention of serious crime. It is usually the only record on file that sets out these considerations.

- **It's the right thing to do:** The power of bulk data analysis carries with it an obligation to use it responsibly. By holding and exploiting bulk data, SIS has the *potential* to intrude on a wide scale into the privacy of individuals who are not of intelligence interest. The authorisation process is one way that we show we are taking this responsibility seriously and that we are minimising intrusion as much as we possibly can. This is about mastering data in a way that is appropriate and sustainable in a liberal democracy.

- **We are subject to external oversight:** The Intelligence Services Commissioner spot checks the process to ensure we are following it correctly. He scrutinises and challenges our necessity and proportionality arguments. He also monitors the standard of our record keeping. To keep his confidence in us, we must maintain the high standards we have set. If we don't show we are behaving responsibly, there is a risk that our ability to use bulk data could be constrained in future.

- **It leads ultimately to operational impact:** Operational teams depend on us to process data from acquisition to exploitation as quickly as possible, so they can begin realising the benefit from the data. Every one of us must strive to complete our part of the process in the most efficient way we can. Attention to detail therefore helps to avoid unnecessary delays.

GENERAL POINTS TO REMEMBER

The form should aim to be a standalone document. Imagine the Commissioner is reading the form. Will he understand what you have written without any other paperwork? Avoid using SIS jargon (or at least expand in brackets what you mean) and spell things out clearly, even if they seem obvious to you. The closest parallel is how we write submissions for the FCO in plain language that require no background knowledge.

When we talk about 'intrusive data', remember that all bulk data is inherently intrusive. We should instead talk about 'personal data that SIS deems to be particularly intrusive'. These are the categories where we are required to make an additional justification to retain and exploit the data.

[redacted]

Always include both your staff number and designation when you sign the form – the Commissioner has been known to spot omissions of either. This is about easily identifying the individual officer who made the judgement.

'ACQUISITION CASE' SECTION

Please ensure you fill out all the boxes under section 2 ('Acquisition Case'), rather than expect someone else to complete them later. That includes the file reference next to the dataset name. If in doubt you can find it in the markings section of the workflow or ask the relevant team.

'Brief Description of Dataset Content' – Try to make sure this box is detailed enough. Imagine you were the Commissioner: you don't normally see the assessment document or the immutable comments, so you're relying on what is contained in the authorisation form to understand exactly what the data is.

'Date of Acquisition' – This is important and needs to be exact. The moment an SIS officer takes possession of data, we are deemed to be 'processing' it under the Data Protection Act.

'Classification' – This should match up with 'Can this Data be Shared?'. Consider whether data could/should be shared with partners and annotate the relevant form to explain your reasoning.

'Media Serial Number' – Make sure this is not blank (or annotate to confirm that there is no removable media involved). If in doubt, check with the relevant team.

'Necessity' – Fairly self-explanatory, but if completed by an operational team, encourage the person filling out this box to consider the wider benefits for the Service of holding and exploiting the data beyond their own immediate area. For example, the data might assist on their local pol/mil requirements, but might it also be useful for CT, CP, CI or other cross-cutting issues?

'Risk Assessment' – Consider all the angles and complete this as fully as possible. The relevant team can offer further guidance.

FACTORS TO CONSIDER WHEN JUDGING LEVEL OF RISK

- Political implications
- Impact on GB prosperity
- Impact on providers.
- Impact on SIS reputation
- Current sensitivities and regional context
- Operational equities
- Litigation

RISK LEVEL DEFINITIONS

Low when revelation would:

- be unlikely to cause HMG political discomfort because (for example), the data is available commercially or offered voluntarily;
- have no or minimal impact on investment in GB;
- have no serious impact on GB diplomatic relations with other nations;
- not alienate GB public opinion because it would be expected that we would gather this data in the National interest;
- present no residual risk to operational equities.

Medium when revelation could:

- cause HMG some political discomfort but be manageable in the longer term because (for example) we might be expected to gather data on the target set (e.g. CT);
- complicate/delay foreign investment in GB;
- [redacted];
- be understood and appreciated as being in the National interest by the majority of GB public;
- [redacted];

High when revelation would:

- cause HMG serious embarrassment [redacted];
- jeopardise foreign investment in GB;
- [redacted];
- alienate public opinion and degrade HMG reputation;
- have an adverse impact on SIS reputation within HMG and the public;
- damage SIS equities or reduce SIS's ability to operate effectively.

DATA ASSESSMENT AND 'DATA INTRUSIVENESS' SECTION

If you tick any box saying 'Yes' or 'Unable to identify', make sure you write in the free text to explain what you have found or why you weren't able to find it. Remember that the Commissioner doesn't usually receive your full data assessment, so you may need to repeat some key points.

'Contains Data on UK Nationals?' – There is no legal requirement for us to identify this (in fact under the European Convention on Human Rights we should not discriminate on the grounds of nationality unless there is a good reason). But this is politically sensitive and so we have taken the policy decision to keep track of what we hold. [redacted].

'Contains Data on Minors (under 16s)?' – Again, there is no legal requirement for us to identify this but we have a policy that we should treat data on children as more sensitive. [redacted].

'Protected Characteristics and Confidential Information' – These are categories of 'personal data that SIS deems to be particularly intrusive'. Although we sometimes talk about 'intrusive fields', really the language we should use is that there is 'personal data deemed by SIS to be particularly intrusive' (intrusiveness is not a legal term and could be a more subjective judgement). By default, SIS deletes all personal data deemed to be particularly intrusive from a dataset, unless a specific case has been made for why it is necessary and proportionate to retain and exploit it. So if any box is ticked in this section, the exploitation officer will need to make a special case in their section of the form if we want to keep and use that data. If you tick one of these boxes, make sure that you give a description in the free text box underneath to explain the exact nature of the information that is personal data deemed to be particularly intrusive, remembering that the Commissioner will probably not see your assessment document.

DETAILS OF CATEGORIES OF PERSONAL DATA THAT SIS DEEMS TO BE PARTICULARLY INTRUSIVE

These definitions sometimes require discussion and what follows is not meant to be exhaustive. If in doubt, ask the relevant team or Legal Advisor.

SIS uses section 2 of the Data Protection Act 1998 as the basis for most of these categories, with some others being derived from the RIPA CHIS Code of Practice or selected for policy reasons. The SIA is looking to standardise these categories as far as possible.

Religion
The definition in law for sensitive personal data includes "religious beliefs or other beliefs of a similar nature".
Be aware of minority religious sects or factions as well as the obvious larger religious faiths. [redacted]

Political opinions
This includes affiliation with or membership of a political party, grouping or movement and donors to political parties.
NB - trade union membership is effectively singled out in law as a separate sensitive category. At present, this is covered by the 'Political' category, but may well become a category in its own right on the form.

Racial or Ethnic origin
Be aware of data that mixes up nationality and ethnicity/race. We must distinguish between the two.

Disability/Medical Condition
The definition in law for sensitive personal data includes "physical or mental health or condition".
This is information about a condition itself, as opposed to confidential information shared with a doctor (see below). In addition to medical conditions, this category includes blood group; physical characteristics (hair/eye colour); and biometrics (iris/fingerprint scans etc).
[redacted]

Sexual Orientation
The definition in law for personal sensitive data refers to "sexual life".

Financial
This includes transactional data that allows you to understand the financial activity of an individual. It would also include obviously private data or a combination of information which would allow a person to leverage goods and services. [redacted]

Criminal Activity
The definition in law of sensitive personal data includes "the commission or alleged commission...of any offence" and "any proceedings for any offence committed or alleged to have been committed by him, the disposal of such proceedings or the sentence of any court in such proceedings".
[redacted]

Legally Privileged Info
This refers only to information that would be confidential between a lawyer and their client. So public court proceedings would not necessarily be included (but this would probably fall into the category of Criminal Activity).

Journalist Info
This refers only to information that would be confidential between a journalist and their source. [redacted]

Medical Info
This refers only to information that would be confidential between a doctor and their patient.

Spiritual Counselling

This refers to information that would be confidential between, for example, a priest and an individual giving confession. The definition in the CHIS Code of Practice states, "spiritual counselling means conversations between a person and religious authority acting in an official capacity, where the individual being counselled is seeking or the religious authority is imparting forgiveness, absolution or the resolution of conscience in accordance with their faith".

'EXPLOITATION CASE' SECTION

See below for guidance on the issue of actual and collateral intrusion.

Justification – This is the box where we make the argument for why it is necessary and proportionate to retain and exploit the data. There can be an overlap with the 'Necessity' box completed by the acquisition officer, but the proportionality angles are not covered elsewhere in the form.

Necessity means "why is it necessary for SIS to retain and exploit the data in order to fulfil its statutory functions?". The three purposes for which SIS exercises its functions are: in the interests of UK national security; economic wellbeing of the UK; the detection/prevention of serious crime. So consider mentioning one of these three purposes as well as the NSC/JIC priority.

Spell out the obvious – e.g. this data would allow us to identify new and higher quality targets for recruitment/disruption against NSC priority X and would be in the interests of Y [UK national security / UK economic wellbeing / detection or prevention of serious crime].

Remember that, by default, SIS deletes any information in a dataset that is personal data deemed to be particularly intrusive – unless we can make a specific argument for why it would be necessary and proportionate to retain and exploit it. So any category of personal data that has been ticked in the assessment section will require particular mention in the justification. This also includes the presence of minors, which requires a specific justification. You could consider making an argument to address the presence of UK nationals if they represent a high proportion of the individuals in the data.

Proportionality means "is your proposal to retain and exploit the data in proportion to the desired outcome?". We should not take a sledgehammer to crack a nut. Could we achieve the same objectives you have mentioned in a less intrusive way? If not, be explicit in explaining why this is not possible.

The description of the dataset on the database: if in doubt about what this should be, flag up in the comments that this needs to be considered further or discussed.

If you mention that the Service audits the use of the database, it might be worth spelling out that this is done in order to reduce the risk of misuse of the data.

GUIDANCE ON ACTUAL AND COLLATERAL INTRUSION

This is a judgement - there is no right or wrong answer. But we should try to make our judgements as consistent as possible and be ready to justify why we made those judgements.

The box on the authorisation form entitled 'Intrusion of Exploitation' should address in free text how you came to the judgement on the levels of intrusion. This shows the Commissioner and other oversight bodies that we have considered all angles.

We need to consider the potential level of intrusion of a dataset across the whole course of its life cycle. This includes its potential to be intrusive when combined with other data - even if we don't yet have that data, but could acquire it in future.

There is a link between actual and collateral intrusion. It would be odd to say the actual intrusion of a dataset is high but the collateral intrusion is low.

[redacted]