Te Pourewa Mātaki
Inspector-General of
Intelligence and Security

Review of the GCSB's acquisition and use of bulk personal datasets

Public Report

Brendan Horsley

Inspector-General of Intelligence & Security

January 2024

INTRODUCTION

1. Bulk personal datasets are datasets that include personal information relating to a number of individuals (sometimes a very large number), most of whom are unlikely to be of intelligence or security interest, but some of whom might be. A commonly cited, though somewhat dated illustrative example of a bulk personal dataset is a telephone directory. A more contemporary example might be a set of subscriber or customer information for an internet platform.

2. Many bulk personal datasets are publicly available, including as a result of privacy breaches. For intelligence purposes, bulk personal datasets can be a source of possible leads when matched with other information, including other datasets. They can also contribute useful information on targets about whom little is known, particularly at the early stages of investigation.

3. This report is an unclassified version of my report on my review of the GCSB's approach to acquisition and use of bulk personal datasets. It omits some classified details of GCSB operations but my findings and recommendations are as in the classified version.

4. In the past year I also began a review of the Service's approach. I concluded from preliminary research, however, that given some significant changes the Service expects to make in the near future I should discontinue that review, to avoid examining practices likely to be superseded. I will continue to monitor developments in this area for both agencies.

SUMMARY OF REVIEW FINDINGS

5. I have found that the GCSB has a sound approach to the planning and approval of dataset acquisition under warrant, and to the use of collected datasets. The relevant warrant is clear in how the activities may be undertaken. Collected datasets are stored in systems with controlled access that require searches of the material to be justified. All searches are logged, auditable and regularly audited. The GCSB has no policies and procedures specifically addressing bulk personal datasets but I am not concerned about this, at this time, given the level of activities undertaken by the GCSB.

6. I identified a gap in the GCSB's policies and procedures around how s 103 of the Intelligence and Security Act 2017 (ISA) and GCSB data retention policy apply to bulk personal datasets. By definition, bulk personal datasets typically contain significant amounts of information on individuals of no intelligence interest. I found it unclear how the GCSB would decide whether to retain these datasets. I have recommended the GCSB remedies this policy gap and it has agreed to do so.

LEGAL FRAMEWORK FOR DATASET COLLECTION

Acquisition of datasets

7. The acquisition of datasets generally falls within the function of the intelligence and security agencies to collect intelligence, provided it is relevant to New Zealand Government intelligence priorities.[1]

8. Generally the intelligence and security agencies can access, download, retain, and use publicly available information without doing anything unlawful and so with no need for an authorisation under Part 4 of the ISA. The agencies may also lawfully receive bulk personal datasets from partners.

9. In some instances the acquisition of datasets from public internet infrastructures may involve unlawful activities. This includes downloading and retaining hacked and leaked datasets[2].

10. Where the acquisition of datasets involves unlawful activities, the agencies must obtain an intelligence warrant. Typically the necessary warranted powers will be search and seizure. Under s 69 ISA the GCSB may access an information infrastructure when executing an intelligence warrant. This encompasses, for example, accessing a website, social media platform or file sharing service to download data.

Retention and use of datasets

11. The retention and use of information within datasets is governed by ss 102-103 of the ISA, which specify when unauthorised and irrelevant information must be destroyed, and s 104, which specifies when and how incidentally obtained information may be retained and shared. I discuss the details of these requirements below.

Ministerial Policy Statement

12. The Ministerial Policy Statement (MPS)[3] "Publicly Available Information" sets out high-level guidance on the acquisition of information from public sources.[4] It applies to the lawful collection and use of publicly available information, including personal information. Its provisions include:

    12.1. GCSB may collect publicly available information using methods not available to the public, but must ensure any unlawful activity is carried out under an intelligence warrant;[5]

[1] Section 10 ISA.

[2] Downloading and retaining hacked and leaked datasets (including by purchasing) may be a breach of s 246 of the Crimes Act 1961, as receipt of property that has been stolen or obtained by an imprisonable offence. See *Dixon v R* [2015] NZSC 147.

[3] The ISA provides for the Minister responsible for the intelligence and security agencies to issue ministerial policy statements (MPS) providing guidance to the agencies. The Director-General of the GCSB must have regard to MPS when making any decision or taking any action.

[4] "Ministerial Policy Statement: Publicly Available Information" (01 March 2022).

[5] At [23].

12.2. GCSB may collect large datasets which might include personal information relating to a number of individuals. It must have a policy that provides guidance on the collection, use, retention and disposal of this type of information.[6]

13. The MPS "Information management" provides guidance on management of all information collected by the GCSB, including its retention and disposal.[7] This includes:

13.1. GCSB must have policies and processes that provide guidance on the retention and destruction of information that give effect to the ISA, the Public Records Act and the Privacy Act 2020;[8] and

13.2. GCSB must specify proportionate retention periods for intelligence, protective security and cyber information. It must have procedures in place for deciding on the destruction or continued retention of information at the end of a retention timeframe.[9]

ASSESSMENT

14. In my classified report I set out my understanding and assessment of the GCSB's activities, relevant policies and procedures. In this report, I am unable to detail the GCSB's acquisition of bulk personal datasets. I summarise my analysis below.

Clear parameters for action

15. The parameters for the GCSB's acquisition and use of bulk personal datasets are set by a combination of intelligence requirements, warrants, policies and procedures.

16. The GCSB does not distinguish bulk personal datasets from other datasets. It applies its general rules for all data. I have found that the GCSB's general parameters for warranted activities and data retention establish clear processes for acquisition, assessment, and use of bulk personal datasets. This is supported by reasonably robust processes set out in the relevant warrants and "operational documentation", which compiles and summarises the procedures for how the GCSB will conduct activities under a warrant.

17. A GCSB audit in 2022 recommended formalising the operational documentation for the relevant activities into standard operating procedures (SOPs). I agree that this will be a useful step for controlling activities.

Parameters for retention and ongoing use of datasets as a whole

18. Section 103 of the ISA requires the GCSB to destroy "irrelevant information" obtained under an authorisation as soon as practicable. Irrelevant information is defined in s 103(1) as information that:

[6] At [32].

[7] "Ministerial Policy Statement: Information management" (01 March 2022).

[8] At [30].

[9] At [31].

a. is obtained by an intelligence and security agency within the scope of an authorised activity; but

b. is not required, or is no longer required, by the agency for the performance of its functions.

19. On its face, s 103 of the ISA sets an expectation that when an agency obtains information under an intelligence warrant, any information not required for the performance of the agency's functions must be destroyed as soon as practicable. For example, when an agency intercepts communications, only those communications that are relevant will be kept.

20. The Bureau's approach to the acquisition and retention of bulk personal datasets acquired under warrant is, broadly:

20.1. To determine whether there is a reasonable suspicion that a dataset includes information relevant to an intelligence requirement. If so, it may be acquired.

20.2. To assess within a set period after acquiring a dataset whether it does in fact contain relevant material. If it contains any relevant information it may be retained; if not, it must be destroyed.

20.3. To keep the dataset under assessment for ongoing retention, according to its utility for GCSB functions.

21. The section 103 requirement is difficult to apply to bulk personal datasets. Most of the information in these datasets will likely be irrelevant, for intelligence purposes, at the time of acquisition or the initial retention decision, and will likely continue to be irrelevant. It will not always be clear to the agency what information in the dataset is relevant, and in some cases relevance will emerge in time, in light of other information. This is because bulk personal datasets are often used as reference datasets, against which other data may be matched to determine something of intelligence value. For example a dataset may only become of use when specific communications are intercepted of a person suspected of having links to organised crime, at which point it might reveal further selectors that can be investigated.

22. The GCSB Data Retention policy provides a good framework for determining retention for all data, recognising that the case for keeping it must be demonstrable and more than merely speculative.

23. I recognise that determining the relevance of a bulk personal dataset to agency functions is a matter of degree. There will often be uncertainty about relevance at the time of acquisition and initial retention. The GCSB should be able to describe and categorise the anticipated value of a dataset, however. This enables later review. The obligation to retain only what is necessary and relevant is ongoing. Initial uncertainty should resolve over time.

24. Continued retention of data within a bulk personal dataset should be justified by a demonstrable value for the GCSB's functions, not the mere promise of potential value. This should be enforced by a default presumption for deletion of irrelevant data after a certain time, unless a persuasive case is made for continued retention. This expectation is reflected in the

GCSB's policy on data retention and destruction. Good record-keeping of assessments and subsequent reviews of bulk personal datasets are essential to this process.

25. The details of a dataset and its context will be relevant to a robust assessment. If a dataset is relevant to a foreign government entity of high priority for intelligence gathering, for example, it may be easier to set out why currently unusable data may be relevant in the future. A broader dataset containing a mixture of information may necessitate some data being carved off and deleted.

26. The GCSB advised my review it would develop further guidance for its staff on how s 103 and the Bureau's data retention and destruction policy should be applied to bulk data sets. I agree this would be a good step.

27. I **recommend** the GCSB develops guidance on how s 103 ISA and the GCSB data retention and destruction policy apply to bulk personal datasets.

Review and improvement processes and "fit for purpose" compliance systems

28. Although my review found the GCSB has a sound general framework for acquiring and using bulk personal datasets, I noted that it does not yet have a structured or methodical approach to carrying out this activity. This is not a pressing concern, but as work in this area develops I would expect to see the Bureau developing a more strategic approach. I do not think this requires a recommendation from me at this stage.

Amenability to oversight

29. As my office does not have full direct access to all GCSB systems, I am unable to independently examine all aspects of the GCSB's ongoing acquisition and use of bulk personal datasets. Much of this activity occurs on operational systems I can examine on request, or have demonstrated, but not access independently at any time. This can be mitigated by good documentation of the Bureau's activities that I can request, or access in the GCSB's central records system. My office is notified of relevant self-reported compliance incidents and has access to GCSB audit records, both of which assist oversight.

30. I considered whether the GCSB should implement a register of certain datasets acquired under relevant warrants, as well as any datasets transferred to GCSB from partners. The GCSB questioned whether the work involved in setting up and maintaining a register would provide significant benefit for oversight compared with my office requesting information when it is needed. I concluded that a register does not need to be set up at this time, but will continue to review relevant warrant applications and revisit the need for a register in the future if necessary.

CONCLUSION AND RECOMMENDATION

31. I have found that the GCSB has a sound approach to the planning and approval of dataset acquisition under warrant, and to the use and retention of collected datasets. The relevant warrant is clear in how the activities may be undertaken. Collected datasets are stored in

systems with controlled access that require searches of the material to be justified. All searches are logged, auditable and regularly audited.

32. I have identified some issues that can arise for the retention of bulk personal datasets and the requirements of s 103 and the GCSB’s Data Retention and Destruction policy. The GCSB has advised that it will look to develop guidance on how to apply s 103 and the policy and to bulk personal datasets and I recommend it does so.