Experiences with privacy management for DNS 'big data' applications

SIDN Labs Technical Report SIDN-TR-2016-001

Date
1 August 2016

Page
1/16

Classification
Public

Authors
Jelte Jansen, Cristian Hesselman

Contact
Jelte Jansen
jelte.jansen@sidn.nl

Contact
T +31 (0)26 352 5500
support@sidn.nl
www.sidn.nl

Offices
Meander 501
6825 MD Arnhem
The Netherlands

Mailing address
PO Box 5022
6802 EA Arnhem
The Netherlands

Reference: J. Jansen and C. Hesselman, "Ervaringen met privacybeheer voor DNS-'big data'-toepassingen", Privacy & Informatie, volume 19, edition 4, August 2016, page 144 to 155.

Abstract

We discuss the introduction and extension of the privacy framework developed for the privacy-aware analysis of network traffic with a view to the early detection of threats, such as phishing sites and botnets. The extensions made to the framework reflect the experience gained over the last eighteen months. In that time, we have introduced the framework and started to make active use of it in connection with the retention and analysis of the messages that we process in our role as operator of the .nl part of the Domain Name System ('DNS'; the internet infrastructure system that translates domain names into IP addresses). We also set out the main lessons learned, with the aim of helping other organisations that would like to introduce similar privacy frameworks.

This article is a translation of "Ervaringen met privacybeheer voor DNS-'big data'-toepassingen". [16]

SIDN Labs is the research team of SIDN, the company that manages the Netherlands' internet extension .nl. SIDN Labs develops, prototypes and evaluates new technologies and systems that enhance the security and stability of .nl, the DNS and the internet. For details: www.sidn.nl and www.sidnlabs.nl.

1 Introduction

Systems that analyse large volumes of data ('big data') have considerable potential, e.g. as a means of increasing internet security. However, the use of such systems should always be accompanied by measures to protect the privacy of the data subjects. Appropriate steps include, for example, transparency regarding the personal data processed by systems and justification of the analysis methods used [1].

One application of big data is the analysis of internet traffic with a view to the early detection of threats to end users, such as phishing sites and botnets. Traffic from the Domain Name System (DNS) can be used in that way, for example [2]. The DNS translates domain names (e.g. www.example.nl) into IP addresses (e.g. 94.198.159.35) so that browsers, e-mail programs and other internet applications can contact the servers for the domain names in question. Translation is necessary because the internet works with IP addresses, which are often hard for people to remember.

The DNS is part of the 'public core of the internet' [8], along with, for example, the Internet Protocol (IP) and the system that routes messages via the internet (Border Gateway Protocol, BGP). It is a global system, operated by a large number of parties, each of whom is responsible for part of it. As the operator of the internet

extension for the Netherlands, we have responsibility for the .nl part of the DNS. In that capacity, we process a daily total of more than 1.3 billion DNS queries (requests to translate .nl domain names into IP addresses) and we manage the database containing the 5.6 million registered .nl domain names.

In our article of December 2014 we explained how DNS messages can contain personal data, implying a need for an enforceable mechanism for protecting the privacy of the data subjects [2]. The approach that we proposed was a multidisciplinary privacy framework covering the legal, technical and organisational aspects of privacy management. Our privacy framework enables data processors (i) to systematically and transparently strike a responsible balance between detecting threats in DNS traffic and protecting the privacy of internet users; and (ii) to enforce the necessary privacy protection measures within the technical system that performs the analyses. The privacy framework is vital in relation to SIDN's role as the trusted party that operates the .nl extension for the Netherlands. However, the framework is also suitable for use in other jurisdictions and with other types of network traffic.

Our framework was developed in parallel with ENTRADA (ENhanced Top-level domain Resilience through Advanced Data Analysis), the technical system that we use to retain and analyse DNS messages [10], [11],[12]. ENTRADA is designed to retain very large numbers of DNS messages and to analyse them very rapidly [10]. Although we use the system for .nl, it is also suitable for other extensions and other types of DNS operators. ENTRADA is an experimental system developed by SIDN Labs, our R&D team. The software is open source and available to download from http://entrada.sidnlabs.nl.

In this article, we discuss the extensions we have made to the privacy framework in the light of our experience with the framework's introduction and active use within SIDN over the last eighteen months. We also set out the main lessons learned, with the aim of helping other organisations that would like to introduce similar privacy frameworks.

Our privacy framework is suitable for a wider range of applications than the purpose for which it is currently used within SIDN. However, because we wish to demonstrate how it is used in practice, the article begins with a brief explanation of what DNS data is and what DNS messages we retain for analysis (section 2). The privacy framework itself is then described (section 3), before we go on to describe our extensions to it (section 4). The article is rounded off with a summary of the lessons learned (section 5) and our conclusions (section 6).

We wish to thank Karin Vink, Lilian van Mierlo and Simon Hania for reading the draft versions of this article and providing us with valuable feedback.

2 DNS data

The data for which our privacy framework was developed consists of messages from the Domain Name System (DNS). The DNS is a globally distributed system that translates domain names into the IP addresses of the associated servers (subsection 2.1). For example, www.example.nl is translated into 94.198.159.35. In DNS jargon, the translation process is known as 'resolving'. The DNS is actually made up of millions of systems, distributed all over the world and operated by a very large number of different parties. The DNS contains other types of data as well, but it is used mainly for looking up IP addresses. SIDN operates the .nl part of the system and we retain some of the message traffic relating to .nl in ENTRADA (subsection 2.2).

2.1 DNS resolving

Figure 1 shows how resolving works when a user clicks on a URL or enters one into a browser's address bar. Our illustration uses http://www.example.nl/ to represent any given URL. The part between the slashes (www.example.nl) is the domain name, which relates to the server running the site.

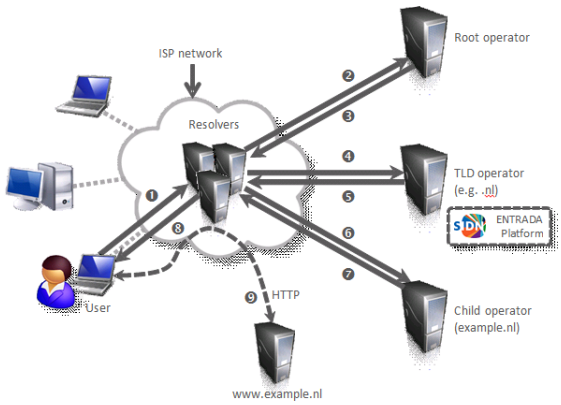

Figure 1. DNS resolving.

To translate www.example.nl into the IP address of the server, the user's machine sends a DNS query to a 'resolver' (step 1 in Figure 1). The resolver is usually a machine operated by the user's internet service provider (ISP), i.e. the company that provides the user with access to the internet. At the browser's request, the resolver looks up the domain name in the global DNS. It starts by approaching a fixed group of 'root servers' (step 2). In the case of www.example.nl, the root servers refer the resolver to the name servers for .nl (step 3). The resolver accordingly contacts a .nl name server (step 4), which duly refers the resolver to the name servers for example.nl (step 5). Next, the resolver sends a DNS query to the name server for example.nl (step 6), which knows the IP address of www.example.nl and sends it back to the resolver (step 7). Finally, the resolver passes on the IP address of www.example.nl to the user's browser (step 8). The browser is then able to retrieve the web page from www.example.nl using HTTP (step 9).

The name servers for .nl receive the messages exchanged in step 4 and send the messages exchanged in step 5. The incoming messages are actually only a proportion ('sample') of the actual number of DNS messages sent by clients, because resolvers use a technique known as 'caching'. Caching involves storing a DNS reply for a certain period of time, so that, if another client asks for the IP address of the same domain name, the resolver can immediately reply from its own cache, without having to contact the name servers in the DNS again. In other words, the resolver skips steps 2 to 7. Caching is an important feature of the DNS, because it is one of the mechanisms that enables the system to keep growing without any loss of performance (scalability).

The infrastructure that we operate for .nl consists of seventy-three name servers, geographically distributed around the Netherlands and abroad.

2.2 Retention and analysis on ENTRADA

The messages that we retain in ENTRADA are the ones exchanged in steps 4 and 5 in Figure 1. An average of about fifteen thousand queries a second are involved, or 39 billion queries (and responses) a month. Under normal circumstances, recording all the data (including IP and Ethernet headers) would require about sixty gigabytes per day per name server.

ENTRADA has so far been operating for twenty months, as an experimental system within the SIDN Labs network. It retains the data from two of the .nl name servers. At the time of writing, that equates to 180,758,031,498 queries and responses, or eight terabytes of stored data. The messages are retained for a maximum of eighteen months, so that we have six months in which to analyse a year's DNS messages. After eighteen months, we aggregate the data and delete the original messages. See [1] for a more detailed description of the retention policy, which we return to in section 3.

We use ENTRADA for a number of experimental purposes, such as botnet detection and the identification of domain names that are suspected of association with phishing activities [10], [11], [12]. We also use the system to monitor the impact of policy changes [13], for the dissemination of statistics on .nl (via https://stats.sidnlabs.nl) and to support applied and academic research by third parties.

3 Privacy framework

The purpose of our privacy framework is to enable us (i) to systematically and transparently strike a responsible balance between detecting threats and abnormalities in DNS traffic and protecting the privacy of .nl users; and

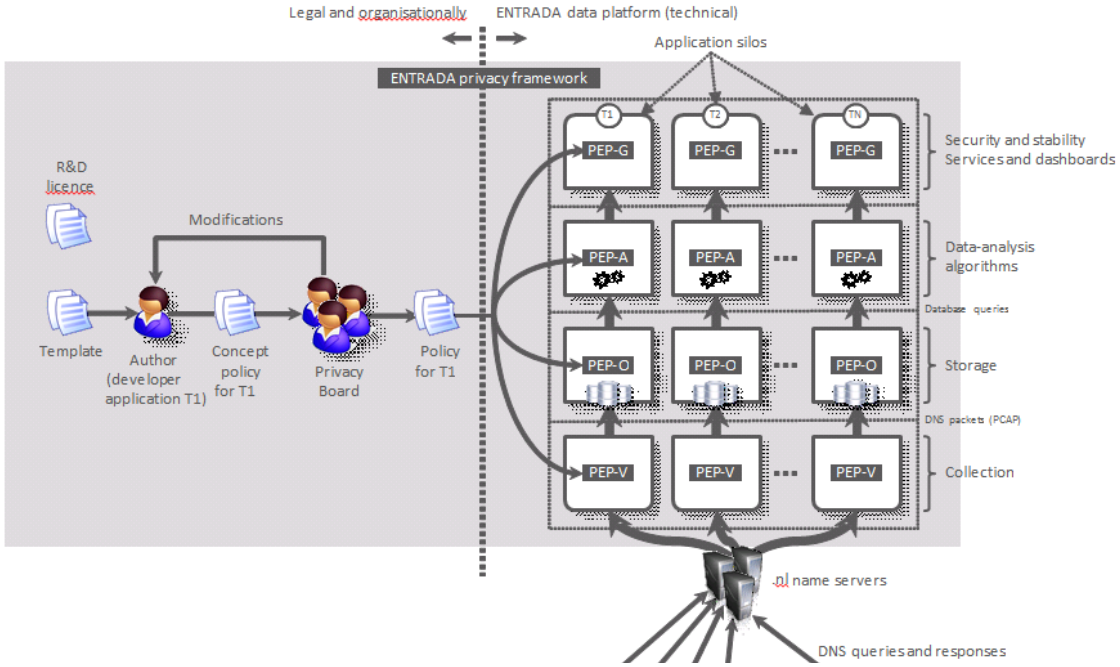

Figure 2. ENTRADA privacy framework.

(ii) to enforce the necessary privacy protection measures by technical means.

Figure 2 depicts the privacy framework, which is summarised in this section. First, the key concepts are explained: personal data (subsection 3.1), privacy policies (subsection 3.2), the Privacy Board (subsection 3.3) and policy enforcement points (subsection 3.4). The objects illustrated to the right of the dotted vertical line in the figure are the objects making up the technical system (ENTRADA). For a more detailed explanation of our framework and the underlying thinking, see [1].

3.1 Personal data

At an early stage in the development of ENTRADA, we realised that some of the data that ENTRADA would be processing could include personal data. Two types of personal data are involved: the IP addresses of resolvers (see subsection 2.1) and the domain names that resolvers look up for end users.

The IP addresses of resolvers are not necessarily personal data, because an ISP will often make a resolver available to a group of users as part of a service package. However, some people run their own resolvers, or query our servers directly, e.g. following infection by a botnet that has its own resolver. Under such circumstances, an IP address is an item of personal data. Because we cannot tell in advance whether a resolver is acting for a group or an individual, we treat all IP addresses as potentially personal data.

We take a similar approach with the domain names that users look up. Although they are usually very general (e.g. 'google.nl'), they can be quite specific or even unique (e.g. 'client461.my.bank.nl'). In the future, DNS traffic will contain fewer domain names that need to be regarded as personal data, because resolvers will send less information [14]. Instead of receiving a query about 'client461.my.bank.nl', our name servers will merely be asked about 'bank.nl'. However, we expect that it will be some time before all resolvers on the internet employ data minimisation technology.

See [1] for a detailed legal explanation of why we treat resolvers' IP addresses and looked-up domain names as

personal data.

3.2 Privacy policies

At the heart of the privacy framework is the privacy policy. A privacy policy defines the data that an ENTRADA application processes, the purpose of the processing and the filters applied to the personal data. Consider the example of an academic study into the patterns discernible in the domain names that are looked up during a nationwide computer security incident. IP addresses would not be relevant to such a study, so the associated privacy policy would provide for them to be filtered out. Alternatively, suppose that a service is offered to internet service providers, which involves analysing DNS data to detect virus-infected computers on an ISP's network and alerting the ISP so that remedial action can be taken. In that case, only the IP addresses are important, not the domain names looked up from those addresses. The privacy policy for the service would therefore state that data is to be shared only with the relevant ISP and that the domain names looked up by the user are to be filtered out.

Filtering is an operation (e.g. pseudonymisation or aggregation) performed on personal data in order to conform to the principles of proportionality and subsidiarity by preventing the excessive or unnecessary processing of the data. Filters are an essential feature of our privacy framework, because they involve the verifiable technical enforcement of privacy policies (see subsection 3.4).

An application receives only the data provided for in the associated policy, and without a privacy policy an application is not permitted access to any personal data. A privacy policy also ensures that matters such as purpose limitation, legitimate basis and data protection are verifiable, both internally and publicly. Within the ENTRADA privacy framework, each application has a privacy policy.

A privacy policy is a text document whose structure resembles that of the personal data processing reporting form used by the Dutch Data Protection Authority (DPA) [15]. Its main elements are:

- Purpose: the purpose for which the personal data is processed by the ENTRADA application. In the case of a data traffic analysis application, for example, the purpose might be the automated detection of malicious botnets within the .nl zone with a view to enhancing the security of the .nl domain.
- Personal data: the particular personal data that the application processes. Where ENTRADA is concerned, that may be IP addresses, looked-up domain names or both (see subsection 3.1).
- Filters: the particular data filters provided for by the policy, the circumstances in which they are to be applied and the personal data to which they are to be applied. A filter might involve the pseudonymisation or complete deletion of personal data.
- Retention: how long the personal data required for an application will be retained. At the end of the retention period, the ENTRADA platform deletes or anonymises the retained data.
- Access: the people or systems that have access to the data and the circumstances under which they may access it. If other systems have access, the policy must also describe the technical access arrangements and state the security measures by which the access is controlled.
- Type: we distinguish between two types of application: research and production. Privacy policies for research applications may require less strict data definitions, because it is not always entirely clear what data will ultimately be needed for the research. However, stricter access controls and data sharing controls may be needed. In the case of a production application, the data to be used must be defined very precisely and access to the data must be very strictly limited.
- Other security measures: any security measures not already referred to in the other sections.

Policy authors use the ENTRADA policy template whenever they write new policies. Consequently, ENTRADA policies have a uniform structure, and their content is standardised as far as possible. Use of the template simplifies policy formulation, policy evaluation by the Privacy Board (see subsection 3.3) and policy publication. Policy authors include ENTRADA application developers and researchers.

The privacy framework was designed on the assumption that it would be used infrequently, typically when a new ENTRADA application was developed or an existing one modified. So, for example, the policy template form was designed for manual completion. However, we were also guided by the belief that it is appropriate to give privacy active consideration, which is not encouraged by comprehensive automation.

3.3 Privacy Board

The Privacy Board is the body that is responsible for evaluating privacy policies. The Board considers whether the privacy policy is satisfactory. To that end, the Board assesses whether the purpose is explicitly defined, whether the application has a legitimate basis, whether the personal data to be used is actually needed for the defined purpose, and whether the filters and protection measures are adequate.

The Privacy Board also considers whether the purpose of the application justifies the means. That involves assessing the contribution that the proposed ENTRADA application is likely to make to the stability and security of .nl and weighing it up against any potential impact on the privacy of .nl users.

In line with the privacy framework's technical, legal, and organisational scope, the SIDN's Privacy Board includes a technical expert, a legal expert and a member of the management team with insight into organisational matters.

3.4 Policy enforcement points

A Policy Enforcement Point (PEP) is a software component within the ENTRADA platform that enforces a privacy policy by technical means, particularly filtering (see subsection 3.2). A filter may be employed at various junctures: prior to data collection, prior to data recording, prior to processing (within ENTRADA itself), and prior to the communication of data to an application or research user. One filter that we have implemented involves the deletion of IP addresses. Applications that do not require IP addresses are given access only to filtered data and therefore cannot see the addresses.

4 Extensions

The ENTRADA privacy framework was introduced and first put to active use within SIDN in 2015. At the time of writing, we have completed and implemented two policies and the Privacy Board is considering a further five policies. One application was found not to require a privacy policy, because the data to be processed did not constitute personal data.

On the basis of our experiences over the last year, we have, at the Privacy Board's suggestion, extended our privacy framework in a number of ways. A privacy policy evaluation process has been defined (subsection 4.1), a number of new elements have been added to our privacy policy blueprint (subsection 4.2), the remit of the Privacy Board has been extended (subsection 4.3), the concept of the evaluation report has been introduced (subsection 4.4) and we have made a declaration to the DPA about our activities and the privacy framework (subsection 4.5).

4.1 Evaluation process

Fout! Verwijzingsbron niet gevonden. illustrates the newly defined privacy policy evaluation process and the interrelationships amongst the various concepts within the framework.

The process begins with the formulation of a privacy policy by a policy author, e.g. a researcher who wants to undertake a new study, or a developer who wants to set up a new ENTRADA platform application. The policy author formulates the policy using the template referred to subsection 3.2 and submits it to the Privacy Board for evaluation.

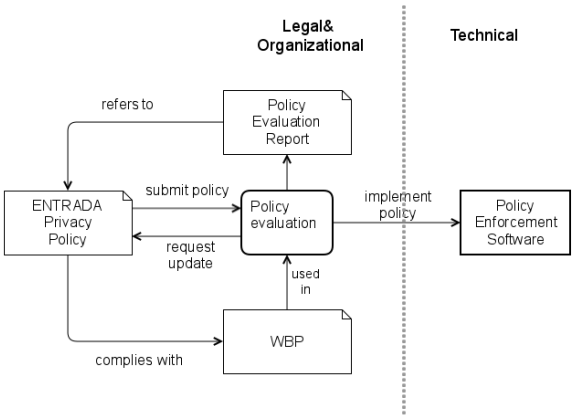

Figure 3. Evalution process.

The Privacy Board evaluates the policy using the policy evaluation report template (sebsection 4.4) and informs the policy author of its conclusions. If the policy is approved, the policy author proceeds to implement the associated filters in the ENTRADA platform by creating corresponding PEPs (see subsection 3.4). The Privacy Board also publishes the policy on our intranet. The policy is published along with the Board's evaluation report, so that readers can see why the Board considers the policy to be acceptable. In the near future, we will start publishing approved policies on our public internet site as well.

If the Board rejects a policy, the author receives an evaluation report explaining the reasons for the policy's rejection. The author then has the opportunity to revise the policy and resubmit it. The process is iterative and ends when the Board approves the policy or rejects it absolutely. At the author's request, the Privacy Board can advise the author on how to improve the policy so that it is acceptable.

4.2 Privacy policies

When evaluating the first privacy policies submitted to it, the Privacy Board reached the conclusion that the format of the privacy policy template was such that authors were not providing all the information needed for a proper evaluation. A number of elements were accordingly added to the template (see below). In adding the new elements, we sought to strike an optimum balance between our need for conciseness and the Board's need for sufficient information for a thorough evaluation.

Privacy policies need to be concise in order to minimise the administrative burden on policy authors (e.g. researchers and developers) and to ensure that there is no deterrent to following the process. Brevity also helps readers, including those outside the organisation, to understand what data is to be processed for what purpose.

On the other hand, the policies must include all the elements that the Privacy Board requires to satisfy itself that the policy conforms to the Dutch Data Protection Act, e.g. with regard to data limitation and retention periods. In addition, policies need to describe the technical control measures (e.g. filters) sufficiently precisely to enable the Privacy Board and the general public to judge whether the measures are adequate.

We have therefore added the following items to the privacy policy template form whose structure was described in subsection 3.2:

- **Legitimate basis**: The legitimate basis for processing the data, as referred to in Section 8 of the Data Protection Act. For example: in order to fulfil a contract with the data subject. Where ENTRADA is concerned, the legitimate basis will usually be to protect the reasonable interests of the data subject and/or the general public. Where that is not the case, one of the other legitimate bases referred to in the Data Protection Act will need to be cited.
- **Publication/sharing**: If the application's output is to be shared with a third party, the privacy policy needs to specify what data will be shared, with whom and subject to what conditions. Publication/sharing is distinct from providing direct access to data, as described in the original version of the policy template (subsection 3.2).
- **Title of application**: The study or the application to which the policy applies must be clearly identified.
- **Date**: The date of the policy's submission for evaluation must be stated.

A specimen privacy policy based on the completed

template form – with the original sections and the new ones referred to above – can be found in 0.

4.3 Privacy Board

When we introduced our privacy framework, it became apparent that the Privacy Board needed to be given additional responsibilities in order to implement the privacy framework within the organisation. The additional responsibilities required were:

- Organisational: defining, documenting and setting up the processes and internal communication channels (e.g. an e-mail address) for the framework
- Technical: evaluating the feasibility and impact of the technical measures for the protection of personal data within the ENTRADA platform
- Legal: declaring personal data processing activities to the Data Protection Authority (see subsection 4.5).

The Privacy Board also has a number of recurring responsibilities:

- Evaluating privacy policies submitted for consideration and re-evaluating existing policies
- Naturally, evaluating new policies is the Board's main ongoing task. However, because projects and services sometimes take a new direction or come to an end, existing policies regularly need to be updated as well. Under such circumstances, re-evaluation by the Privacy Board is necessary.
- Publishing active privacy policies
- Transparency is one of the aims of the privacy framework. It is therefore our intention to publish not only general descriptions of our data processing activities, but the text of each privacy policy implemented.
- Regularly re-evaluating the framework itself
- The framework is a new concept and the applicable legislation is subject to change. Our privacy framework therefore requires regular re-evaluation and adaptation where necessary.

4.4 Evaluation report

The purpose of the evaluation report is to set out why the Privacy Board has approved or rejected a policy. Authors require such information in order to improve their policies, and documenting the Board's decisions enables the organisation to build up an archive for future reference. The reports also enable the general public to understand the rationale behind the approved policies.

An evaluation report supplements the associated policy. Some elements of the report serve to specify the Privacy Board's considerations and reasoning or contain references to the Data Protection Act. Other elements are merely yes/no answers, provided to confirm that the Board has considered the corresponding aspects of the policy. In each section of the report, the Privacy Board states its grounds for concluding that the relevant requirements have or have not been met. If one wishes to know only what data is processed for what purpose, it is sufficient to read the privacy policy on its own. However, if one wishes to understand why such processing is considered reasonable, one can find out by referring to the evaluation report.

The evaluation report is divided into a number of sections, as follows:

- **Title**: title of the evaluated policy
- **Date of evaluation**: date that the policy was evaluated
- **Applicability of Data Protection Act**: why the Privacy Board believes that the Data Protection Act is applicable, including details of the data to be processed, why that data is considered to be personal data, and whether the processing is to be automated or to involve personal data contained in a file
- **Purpose**: whether and, if so, why the Privacy Board considers that the purpose of the processing provided for in the policy is specific, explicitly defined and legitimate
- **Legitimate basis**: whether the Privacy Board considers the personal data processing regulated by the privacy policy to have a legitimate basis and, if so, which section of the Data Protection Act is therefore considered applicable
- **Purpose limitation**: whether the Privacy Board considers that the privacy policy provides for adequate measures to ensure that no data will be

processed for a purpose other than that defined in the policy
- **Retention period**: whether the Privacy Board considers that the retention period provided for in the privacy policy is justified and no longer than necessary for the purpose of the application or study
- **Data set limitation**: whether the Privacy Board considers that the data set whose processing is provided for in the privacy policy is limited to that necessary for the purpose of the application or study
- **Data reliability**: what assurances there are that the data to be processed is accurate (check item to verify that the data used derives from software-generated service reports and in principle must therefore be correct, as it is the data used to provide the service)
- **Data processors**: what data processors are to be used (check item to verify that there are no processors other than those identified in the privacy policy)
- **Data security**: what is to be done to secure the data (check item to verify that the data is adequately secured)
- **Special personal data**: whether any special personal data is to be processed (check item to verify that no special personal data is to be processed)
- **DPA declaration**: whether the data processing provided for in the privacy policy has been declared to the DPA (check item to verify that the processing is included in the ENTRADA activities declared to the DPA; see subsection 4.5)
- **Protection of subjects' rights**: whether the data subjects' rights are protected, including references to the relevant provisions of the Data Protection Act (check item to verify that data subjects' rights are respected, as required in the Data Protection Act)
- **Data outside the EU**: whether any personal data will be shared with any person or organisation outside the EU (check item to verify that no data will leave the EU without adequate safeguards)
- **Evaluation**: the Privacy Board's conclusion as to the acceptability of the privacy policy, including, where appropriate, any conditions that the Board may choose to attach to its approval of the policy (e.g. that a contract is to be closed to regulate the sharing of data)

A specimen evaluation report can be found in 0.

Our evaluation report template was developed by drawing up a list of all the passages of the Data Protection Act that are relevant in this context. The list was then translated into a checklist of criteria for the evaluation of a policy. Various technical considerations were added, so that the checklist provided a basis for ascertaining that the technical security measures are sound and comprehensive.
Originally, the checklist was intended merely as a reference tool for the Privacy Board to use when evaluating a policy. However, we quickly realised that it was useful to record the Board's conclusions regarding each point. Then, instead of a plain declaration of approval or rejection, there would be a clear statement of the Privacy Board's reasoning. We therefore converted the checklist into a second document template, which the Privacy Board uses to compile its evaluation reports.

The Privacy Board's evaluation reports are published with the associated policies, but are standalone documents. Consequently, the policies remain concise (see subsection 4.2), while detailed argumentation and references to the Data Protection Act are available in the evaluation report. If a privacy policy is not approved, and not therefore published, the approach we have adopted means that the policy author can see exactly why the policy has been rejected, and we are able to maintain a comprehensive internal archive of evaluations.

4.5 DPA declaration

The final extension to our privacy framework is a DPA declaration. Section 27 of the Data Protection Act [4], requires that all personal data processing activities not included in a list of specific exceptions set out in the Data Protection Act Exemptions Decree are declared to the DPA. The exceptions include activities such as maintaining a members' address list for a club and

making direct contact with data subjects.

Because transparency is one of the principles on which our framework is based, we decided to declare all our processing activities to the DPA, regardless of whether they are covered by the Exemptions Decree or not. Our declaration to the DPA is can be viewed at [6].

Declaration will cease to be mandatory when the General Data Protection Regulation comes into effect, but for the time being we continue to follow the existing legislation.

5 Lessons learned

Over the last year, we have learnt a number of important lessons, which (i) have led to the extension of our framework (see section 5 and (ii) may be relevant to other organisations interested in adopting privacy frameworks, whether based on ours or not.

First, we have found that developing and implementing a privacy framework was accompanied by an increase in general privacy awareness within SIDN. Such awareness is important within any organisation governed by the new legislation in this field. However, the implementation of a privacy framework creates a particular need for heightened awareness concerning personal data processing.

We also found that people within SIDN soon started approaching the Privacy Board about privacy-related matters, so that the Board acquired other responsibilities in addition to evaluating privacy policies (see subsection 4.3). One initially subordinate task that quickly became an important part of the Privacy Board's activities was informing the organisation about privacy and answering questions on the subject. While most people at SIDN already had a good sense of what was possible, permissible and responsible, it became apparent that greater clarity was required, certainly regarding the legal aspects.
The complexity of the legal situation was underlined by experience of formulating privacy policies on the basis of the template. While authors found that completing the template form was mostly straightforward, many were challenged by the section where they had to state the legitimate basis for processing (the one field that relates directly to the Data Protection Act). We have therefore decided that a future version of the policy template will make it clear that at least one of the legitimate bases listed in Section 8 of the Data Protection Act must be applicable. Instead of an open input field, the template form will in future provide a list of options from which authors can choose.

6 Conclusions and plans fort he future

An enforceable privacy framework for the retention and analysis of 'big data' from the Domain Name System (DNS) is vital in relation to SIDN's role as the trusted party that operates the .nl extension for the Netherlands. The ENTRADA privacy framework provides us with the means to fulfil that role, enabling us to transparently and systematically strike a responsible balance between detecting threats in DNS traffic and protecting the privacy of internet users.

Experience gained from the introduction and active use of our privacy framework over the last year has enabled us to improve our framework further, e.g. by adding an evaluation report template. Furthermore, we have learnt various lessons, which may be relevant to other organisations.

Our framework serves as a comprehensive, practical basis for making use of potentially personal data in circumstances where it is not possible to obtain the data subjects' direct consent, in order to further enhance the security and stability of .nl through the analysis of DNS messages. Our framework facilitates transparency regarding the processing of personal data and is more precise and more comprehensive than the straightforward publication of a single all-embracing privacy statement. It therefore enables us to go beyond what is required by the Data Protection Act. We hope that our approach can serve as an example to other organisations that process personal data or are considering doing so.
In the near future, we will start publishing our approved privacy policies. We also intend to regularly

re-evaluate our existing policies and the framework itself. We are monitoring the progress of the planned EU Data Protection Regulation (DPR) and, although the DPR was taken into account when our framework was developed, we will review the framework when the DPR comes into force.

The privacy framework and the associated processes are now part of everyday working practice at SIDN, but we anticipate that they will require continuous refinement. We are open to further feedback and suggestions, and are happy to speak to anyone interested in our approach.

7 References

[1] Wetenschappelijke Raad voor het Regeringsbeleid, "Big Data in een vrije en veilige samenleving", april 2016, http://www.wrr.nl/fileadmin/nl/publicaties/PDF-Rapporten/rapport_95_Big_Data_in_een_vrije_en_veilige_samenleving.pdf

[2] C. Hesselman, J. Jansen, M. Wullink, K. Vink en M. Simon, "Een privacyraamwerk voor 'DNS big data'-toepassingen". Privacy & Informatie, afl. 6, december 2014, https://www.sidnlabs.nl/downloads/whitepapers/PEI_2014_6_Hesselman.pdf

[3] P. Mockapetris, "Domain Names – Concepts and Facilities", IETF, November 1987. https://www.ietf.org/rfc/rfc1034.txt

[4] Wet bescherming persoonsgegevens. geldigheidsdatum_02-10-2015, 2000. http://wetten.overheid.nl/BWBR0011468/

[5] Vrijstellingsbesluit WBP, http://wetten.overheid.nl/BWBR0012461/geldigheidsdatum_02-10-2015, 2000

[6] "Enhanced Top-level domain Resilience Through Advanced Data Analysis", CPB-melding 1591862, https://www.collegebeschermingpersoonsgegevens.nl/asp/ORDetail.asp?moid=8585898d8b88, Mei 2015

[7] M. Müller, "SIDekICk: SuspIcious DomaIn Classifcation in the .nl Zone", M.Sc. thesis, Universiteit Twente, juli 2015, https://www.sidnlabs.nl/downloads/publications/Muller_Master_Thesis_EIT_SP.pdf

[8] M. Davids, "ENTRADA-koppeling met AbuseHUB", blogpost, septermber 2015, https://www.sidnlabs.nl/a/weblog/entrada-koppeling-met-abusehub

[9] "De publieke kern van het internet. Naar een buitenlands internetbeleid", WRR-rapport nr. 94, maart 2015, http://www.wrr.nl/publicaties/publicatie/article/de-publieke-kern-van-het-internet-1/

[10] M. Wullink, G. Moura en C. Hesselman, "ENTRADA: a High Performance Network Traffic Data Streaming Warehouse", IEEE/IFIP Network Operations and Management Symposium (NOMS16), Istanbul, Turkije, april 2016, https://www.sidnlabs.nl/downloads/sidn-noms2016_EN.pdf

[11] G. Moura, M. Müller, M. Wullink en C. Hesselman, "nDEWS: a New Domains Early Warning System for TLDs", IEEE/IFIP International Workshop on Analytics for Network and Service Management (AnNet 2016), Istanbul, Turkije, april 2016, https://www.sidnlabs.nl/downloads/presentations/sidn-annet2016.pdf

[12] M. Wullink, M. Müller, M, Davids, G. Moura en C. Hesselman, "DNS Big Data Applications with ENTRADA", Symposium on Electronic Crime Research (eCrime'16), Toronto, Canada, juni 2016, https://www.sidnlabs.nl/downloads/whitepapers/sidn-ecrime.pdf

[13] M. Wullink, "ENTRADA: The Impact of a TTL Change at the TLD Level", presentatie, DNS-OARC Spring 2016 workshop, Buenos Aires, Argentinië, april 2016, https://www.sidnlabs.nl/downloads/presentations/DNS-OARC-Buenos-Aires-final.pdf

[14] S. Bortzmeyer, "RFC7816: DNS Query Name Minimisation to Improve Privacy", DNS Query Name Minimisation to Improve Privacy, march 2016, https://tools.ietf.org/html/rfc7816

[15] Autoriteit Persoonsgegevens, “Melden verwerking persoonsgegevens”, https://autoriteitpersoonsgegevens.nl/nl/melden/melden-verwerking-persoonsgegevens

[16] Ervaringen met privacybeheer voor DNS-‘big data’-toepassingen https://www.sidnlabs.nl/downloads/papers-reports/SIDN-TR-2016-001-NL.pdf

Appendix A Specimen privacy policy

Title of application/study	ENTRADA, general
Purpose of application/study	The ENTRADA (ENhanced Top-level domain Resilience through Advanced Data Analysis) platform is a platform for the retention and analysis of DNS query data. It exists to support the development of new services and applications that enable us to further enhance the security and stability of .nl, and isolated investigations into incidents with the potential to threaten the stability of .nl. This policy covers the platform itself, prototypes and studies. Separate privacy policies are defined for production systems, applications and studies outside the scope of this policy.
Personal data	Because ENTRADA is a general platform for research and development, it is not possible to say exactly what data that will and will not be required. DNS query data is therefore retained. As explained in the policy paper *A privacy framework for DNS big data applications*, the personal data that is retained and processed consists of IP addresses and looked-up domain names.
Legitimate basis	The purpose of the platform is to support research into and the development of applications that enhance the security of .nl and its users. The legitimate basis for use of the data is that it serves a legitimate interest.
Filters	No filters are applied.
Retention	Data is retained for eighteen months. The retention period has been chosen so that we have sufficient time to analyse a year's data.
Access	The data is accessible to SIDN Labs staff and SIDN's DNS operators. The data can be accessed only from SIDN Labs' internal network, using the HTTPS protocol. Access is controlled by password or by Kerberos authentication. SIDN Labs staff and SIDN's DNS operators have been instructed on the responsible use of data.
Publication/sharing	The data is not shared. Published research results do not contain specific personal data. Data processed under this policy is not shared with third parties. Separate, specific privacy policies are formulated for projects and services that do involve the sharing of data with third parties.
Type	R&D research
Other security measures	N/A

Appendix B Specimen evaluation report

Policy		
	Title of policy	ENTRADA, general
	Date of evaluation	5 January 2016
Purpose limitation		
	Applicability of Data Protection Act	Will any personal data be processed? *The Privacy Board believes that the studied query data could be traced back to individual IP addresses and that a proportion of IP addresses could be traced back to natural persons. Hence an IP address can sometimes be an item of information regarding an identifiable natural person (Data Protection Act, Section 1a). The Privacy Board accordingly recommends that the data be treated as containing personal data. Looked-up domain names should similarly be treated as containing personal data.* Will personal data be processed on an automated or semi-automated basis, or will personal data contained in a file be processed manually? *Yes. In light of the provisions of Section 2, subsection 1, of the Data Protection Act, the Act may be deemed applicable to the processing.*
	Purpose	Is the purpose specific, explicitly defined and legitimate? *Yes, the Privacy Board believes that the privacy policy defines the purpose of the processing in specific and explicit terms, as referred to in Section 7 of the Data Protection Act. The Privacy Board also considers increasing the security and stability of .nl to be a legitimate purpose.*
	Legitimate basis	Is there a legitimate basis for the processing? *The Privacy Board believes the processing serves the legitimate interests of both SIDN (whose objects include increasing the reliability and security of .nl and of the internet as a whole) and third parties (the users of .nl). Hence, there is a legitimate basis, as referred to in Section 8f of the Data Protection Act.*
Safeguards and control measures		
	Purpose limitation	Are there adequate safeguards to ensure that personal data does not undergo further processing that is inconsistent with the purpose for which it was obtained? *Yes, the Privacy Board believes that Section 9 of the Data*

		Protection Act is complied with, insofar as ENTRADA is used exclusively for internal research and adequate measures are taken to control access to it. *Wherever data is to be shared with an outside party, a separate privacy policy for the application or study in question is drawn up and submitted to the Privacy Board for evaluation.*
	Retention period	Are there adequate safeguards to ensure that personal data is not retained for any longer than necessary for the defined purpose? *Yes, the Privacy Board believes that, in keeping with Section 10 of the Data Protection Act, a retention period of eighteen months is realistic in that it allows six months for the study of a year's data.*
	Data set limitation	Are there adequate safeguards to ensure that processing is limited to the data that is necessary for and relevant to the defined purpose? *Yes, the Privacy Board believes that, in keeping with Section 11 of the Data Protection Act, the data set used is the minimum required for the fulfilment of the defined purpose of the processing. ENTRADA-based research requires the use of a complete data set.*
	Data reliability	Are there adequate safeguards to ensure that the gathered data is accurate? *Yes, the Privacy Board believes that, in keeping with Section 11 of the Data Protection Act, the data used may safely be assumed to be accurate, because it is gathered by SIDN itself using its own systems. Access to those systems is controlled, preventing third-party interference with the data.*
	Data processors	Are there adequate safeguards to ensure that data is processed only on the data controller's instructions? *Yes, in keeping with Section 12 of the Data Protection Act, data is processed exclusively by SIDN Labs staff and SIDN's DNS operators, i.e. employees of the data controller who require access in order to carry out their duties.*
	Data security	Are appropriate technical and organisational measures in place to secure the data? *Yes, the Privacy Board believes that, in keeping with Section 13 of the Data Protection Act, access is adequately controlled.*

Other		
	Special personal data	Is any special personal data processed? *No, the Privacy Board has taken external professional advice and is of the opinion that no special personal data, of the kind referred to in Section 16 of the Data Protection Act, is processed.*
	DPA declaration	Has the requirement to declare personal data processing activities to the Data Protection Authority, as contained in Section 27 of the Data Protection Act, been met? *Yes, processing is covered by declaration number 1591862.*
	Protection of subjects' rights	Do the activities meet the information obligations created by Sections 33 and 34 of the Data Protection Act? *Yes, the Privacy Board believes that Section 34 is applicable.*
	Retention within EU	Is any data transferred to any country outside the EU, necessitating compliance with Section 76 of the Data Protection Act? *No, ENTRADA data is processed exclusively by SIDN employees.*
Conclusion		
	Evaluation	*The Privacy Board approves the privacy policy entitled 'ENTRADA, general'.*