Privacy Policy

SIDN data platform

Date
1 January 2024

Classification
Public

Author
SIDN

Page
1/5

SIDN BV

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 Netherland

Registration
Arnhem 88 77 28 96

Title of application/study	SIDN data platform
Policy start date	1 January 2024
Project description	The SIDN data platform will be a central storage facility where large volumes of unmodified data are collected and stored. The stored data may be processed and used for various analytical purposes. The data files will be stored in phased zones: raw, standardised, integrated and consumable, so that users of various types can use the data in its various forms as they may require.
Purpose of application/study	The data platform has the following purposes: 1. Enhancement of data quality and consistency: by using a data lake, we can standardise the way that data from each source is gathered and structured. That reduces the risk of erroneous assumptions and saves time. 2. A single version of the truth: the data lake (house) serves as a central location for all information matters and ensures that everyone is working with the same data. Source system dependencies within informational processes are reduced, minimising informational vulnerability. Furthermore, if a source system is replaced, there are no implications for the

structure of the processes that utilise the information. That in turn helps to assure continuity.

3. Data source combination opportunities: by using a data lake (house), we can gather data from multiple sources. The subsequent combination of data from different sources can yield valuable insights. For example, registrars can be given useful information about their portfolios.
4. Reduced management and maintenance burden: managing and maintaining local servers is increasingly time-consuming. The set-up of new applications and/or reporting modules is also labour-intensive. The need to assign capacity to such activities means that less capacity is available for innovation. By using a central data lake in the cloud, we can reduce our product teams' workloads.
5. Enhanced security: a data lake (house) is excellent for the secure containment of data. Centralised data storage enables us to ensure that individual users have access only to the data that they are authorised to view. It also simplifies the task of secure data management, since only one location has to be managed.

The following data sources are accessed for the following data processing purposes:

- Domain Registration System (DRS): for internal reporting on registrar transactions, generating DRS reports so that Support personnel have appropriate information about registrars and domain names at their disposal, and generating SIDN Insights to provide registrars with performance insights.
- DMAP: for giving registrars information about their performance in relation to Registrar Scorecard (RSC) variables through SIDN Insights, and for internal reporting on progress with technical standards and internet trends.
- Valimon: for giving registrars insight into DNSSEC-related matters and for internal progress monitoring.
- Google Analytics/PiwikPro: for obtaining information about registrars' online conversion rates in the context of the domain name suggestion tool.
- Netcraft: data for identifying domain names associated with abuse, detecting differences amongst registrars and measuring resolution speeds.
- Business Central: for helping registrars to understand their invoices.

The data platform will use data from the sources listed above to generate reports and populate dashboards for use by SIDN staff and customers such as registrars, ICANN and CENTR. The reports in question are all currently being generated using multiple applications. The data platform will assume the report generation functionalities of the other applications currently in use.

This policy relates exclusively to the data platform. Separate privacy policies must be defined for applications and studies that make use of the data platform but fall outside the scope of this policy.

Personal data

The project will involve the processing of the following personal data:

- All domain name transaction data
- All registrar transaction data (name, e-mail, phone number, address)
- Registrar classifications (segments)
- All registrant data (name, e-mail, phone number, address)
- Domain names associated with abuse
- Abuse classifications
- Abuse timings
- IP addresses of domain names associated with abuse
- Domain names looked up in the Whois
- Domain names validated by Valimon (domain name, error code where relevant)
- Trade Register number (DMAP)
- Registrars' billing data (e-mail, invoice content)

Under certain circumstances, the processed data may include personal data.

Legitimate basis

By centralising data access, you reduce the pressures that drive unnecessary, excessive duplication. The definition of roles and permissions is also simplified, as are data management and the supervision of data access and use. That in turn contributes to the security, transparency and manageability of the organisation.

By combining data from various sources, we will be able to provide customers with additional insight into their portfolios and contribute to goals such as growing .nl. Our ability to exert a positive influence on the security of the zone (through anti-abuse activities and the RSC) will also be enhanced.

Filters	No filters are applied.
Retention	A separate retention period will be specified for each source or business object. The default retention periods will be 2 years for raw data and 5 years for aggregated data.
Access	The data platform will have multiple layers. Authorisations will be defined separately for each layer. Raw data: this layer will contain raw data, as received from the source (with added metadata). It will be accessible only to automated processes (pipelines) that serve to pipe data to the next layer. Standardised data: this layer will contain data in the form of files and tables. The historical sequence will be determined and data will undergo technical transformations such as compression and standardisation. The data will be queryable, and will be accessible only by authorised members of the data team and data scientists within SIDN Labs working in a secure environment, by means of a username-password combination and a second authentication factor. Integrated data: this layer is where the technical implementation of business objects takes place. The data will be queryable, and will be accessible only by authorised members of the data team working in a secure environment, by means of a username-password combination and a second authentication factor. Consumable data: in this layer, the data will be functionally optimised by means of filtering, aggregation and the application of specific business rules. The products of this layer will be made available in 'datamarts' and/or datasets. The marts will be accessible to authorised SIDN personnel for the compilation of reports. Authorisations may differ from one dataset or datamart to another. Export layer/API: specific layer for customer access to reports, datasets, etc (e.g. by CENTR, ICANN).
Publication/sharing	Data from the consumable data layer will be shared with personnel in the Business & Support Department, the Finance Department, SIDN Labs and the MT. As a matter of principle, no

personal data will be shared, unless necessary for the performance of professional duties.

Via this layer, reports and datasets will be made available to external users such as registrars and ICANN.

The data in question may include personal data in circumstances where information regarding a particular domain name or registrant is required, such as where a Support worker is assisting a customer with an enquiry, or where a registrar has a domain-level question.

Aggregated data (statistics) will be used in various publications. Such aggregated data will not include personal data.

Type	Production
Other security measures	The data management module will provide insight into data use, for audit purposes.
Summarised project description	The data platform will be a central storage facility where large volumes of unmodified data are collected and stored. The stored data may be processed and used for various analytical purposes. The data files will be stored in phased zones: raw, standardised, integrated and consumable, so that users of various types can use the data in its various forms as they may require.