

# Data stream

---

In connection-oriented communication, a **data stream** is the transmission of sequence of digitally encoded coherent signals to convey information.<sup>[1]</sup> Typically, the transmitted symbols are grouped into a series of packets.<sup>[2]</sup>

Data streaming has become ubiquitous. Anything transmitted over the Internet is transmitted as a data stream. Using a mobile phone to have a conversation transmits the sound as a data stream.

## Contents

---

### Formal definition

### Content

### Usage

### Integration

### Data sources visible

### Formats

### GDPR

### References

## Formal definition

---

In a formal way, a data stream is any ordered pair  $(s, \Delta)$  where:

1.  $s$  is a sequence of tuples and
2.  $\Delta$  is a sequence of positive real time intervals.

## Content

---

Data Stream contains different sets of data, that depend on the chosen data format.

- **Attributes** - each attribute<sup>[3]</sup> of the data stream represents a certain type of data, e.g. segment / data point ID, timestamp, geodata.
- **Timestamp** attribute helps to identify when an event occurred.
- **Subject ID** is an encoded-by-algorithm ID, that has been extracted out of a cookie.
- **Raw Data** includes information straight from the data provider without being processed by an algorithm nor human.
- **Processed Data** is a data that has been prepared<sup>[4]</sup> (somehow modified,

validated or cleaned), to be used for future actions.

## Usage

---

There are various areas where data streams are used:

- **Fraud detection & scoring** – raw data is used as source data for an anti-fraud algorithm (data analysis techniques for fraud detection). For example, timestamp or amount of cookie occurrences or analysis of data points are used within the scoring system to detect fraud or to make sure that a message receiver is not a bot (so-called Non-Human Traffic<sup>[5]</sup>).
- **Artificial intelligence** – raw data is treated like a train set and a test set during AI and machine learning algorithms building.
- **Raw data** is used for profiling and personalization to customize user profiles<sup>[6]</sup> and divide them for segmentation, e.g., per gender or location (based on data point).
- **Business intelligence** – raw data is a source of information for BI systems, used for enriching user profiles with detailed information about them, e.g., purchase path or geodata. This information is used for business analysis and predictive research.
- **Targeting** – processed data by data scientists improve online campaigns and is used for reaching the target audience.<sup>[7]</sup>
- **CRM Enrichment** – raw data is integrated with customer-relationship management system. CRM integration allows to fill the gaps in users' profiles with demographic data, interests or buying intentions.

## Integration

---

Core integrations with data streams are:

- Data streams are integrated with systems such as customer data platform (CDP), customer relationship management (CRM) or data management platform (DMP) to enrich users' profiles with external data. It is possible to expand the knowledge about existing users by using external sources.<sup>[8]</sup>
- Data streams are used to enrich business intelligence systems and make analysis more precise and conclusions more accurate.
- In the case of content management system (CMS) integration, Data Stream is used to identify the users and personalize their visit, even if it's their first one. By data analysis, the actual content of the website is adapted to the user.
- Data streams are integrated with demand side platform (DSP) within programmatic advertising ecosystem. Parties (e.g., advertisers) can exchange the users' IDs and concatenate with them existing profiles.
- Data streams are used to choose respective user segments (e.g., people interested in the automotive industry) and use them in an online campaign. Segments are enriched with more user characteristics out of data stream and then sent to DSP.

## Data sources visible

---

In a data stream it is visible what device has been used by the user side – it is visible on user agent:

- **mobile** – when a user uses a mobile browser to explore, it has narrow screen resolution and mobile app version, respectively;
- **desktop** – when a user uses a desktop browser or app version.

The following information is shared out of used device:

- Actual URL to the visited website, where an event occurred
- User Agent
- Geolocation
- Internet Protocol (IP)

## Formats

---

A **data point** is a tag that collects information about a certain action, performed by a user on a website. Data points exist in two types, the values of which are used to create appropriate audiences. Those are:

- 'event' with information about occurrences of the specific event (e.g., click on a link or displaying ad)
- 'attribute' with numerical or alphanumerical values.

**Segment** is a logical statement, built on specific Data Points using AND, OR or NOT operators.<sup>[9]</sup>

**Hybrid data** – raw data out of both Data Point and Segment data formats.<sup>[10]</sup>

**URLs** – is a set of information about a particular URL that has been visited.

## GDPR

---

Information gathered out of websites are based on user behavior. Data providers deliver both personal or non-personal information. There are two types of user data available in data stream:

- **Personally identifiable information (PII)** – information that allows clearly or by combining with data identification methods identify a person. Examples of PII are: insurance ID, email address, phone number, IP address, geolocation, biometric data.<sup>[11]</sup>
- **Non-personally identifiable information (non-PII)** is information that can't be used to identify a person or to track a location. A cookie or a device ID is an example of non-PII.

## References

---

1. "Federal Standard 1037C *data stream*" ([http://www.its.bldrdoc.gov/fs-1037/dir-010/\\_1451.htm](http://www.its.bldrdoc.gov/fs-1037/dir-010/_1451.htm)). Archived ([https://web.archive.org/web/20070413122107/http://www.its.bldrdoc.gov/fs-1037/dir-010/\\_1451.htm](https://web.archive.org/web/20070413122107/http://www.its.bldrdoc.gov/fs-1037/dir-010/_1451.htm)) from the original on April 13, 2007. Retrieved April 4, 2007.

2. "Data Stream" (<https://www.techopedia.com/definition/6757/data-stream>). *techopedia.com*. Archived (<https://web.archive.org/web/20190424083723/http://www.techopedia.com/definition/6757/data-stream>) from the original on April 24, 2019. Retrieved April 24, 2019.
3. "Attribute" (<http://www.businessdictionary.com/definition/attribute.html>). *businessdictionary.com*. Archived (<https://web.archive.org/web/20190424083729/http://www.businessdictionary.com/definition/attribute.html>) from the original on April 24, 2019. Retrieved April 24, 2019.
4. "What constitutes data processing?" ([https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-constitutes-data-processing\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-constitutes-data-processing_en)). *ec.europa.eu*. Archived ([https://web.archive.org/web/20190424083732/https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-constitutes-data-processing\\_en](https://web.archive.org/web/20190424083732/https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-constitutes-data-processing_en)) from the original on April 24, 2019. Retrieved April 24, 2019.
5. "Non-Human Traffic [NHT]" (<https://theonlineadvertisingguide.com/glossary/non-human-traffic/>). *theonlineadvertisingguide.com*. Archived (<https://web.archive.org/web/20170813062942/http://theonlineadvertisingguide.com/glossary/non-human-traffic/>) from the original on August 13, 2017. Retrieved April 24, 2019.
6. "BEHAVIORAL PROFILING AND PERSONALIZATION: CUSTOMER EXPERIENCE FIRST" (<https://www.selligent.com/blog/inspiration/behavioral-profiling-and-personalization-customer-experience-first>). *selligent.com*. Archived (<https://web.archive.org/web/20190424083739/https://www.selligent.com/blog/inspiration/behavioral-profiling-and-personalization-customer-experience-first>) from the original on April 24, 2019. Retrieved April 24, 2019.
7. "What is Targeting – Meaning" (<https://sendpulse.com/support/glossary/targeting>). *selligent.com*. Archived (<https://web.archive.org/web/20190424083726/http://sendpulse.com/support/glossary/targeting>) from the original on April 24, 2019. Retrieved April 24, 2019.
8. "What is Data Stream and how to use it" (<http://www.onaudience.com/resources/what-is-data-stream-and-how-to-use-it/>). *OnAudience.com*. Archived (<https://web.archive.org/web/20190424083725/http://www.onaudience.com/resources/what-is-data-stream-and-how-to-use-it/>) from the original on April 24, 2019. Retrieved April 24, 2019.
9. "The 6 types of user segmentation and what they mean for your product" (<https://uxdesign.cc/how-to-think-segmentation-from-day-1-f714df093ccb>). *uxdesign.cc*. June 12, 2018.
10. "What is hybrid data management" (<https://www.ibm.com/analytics/data-management/resources/what-is-hybrid-data-management/>). *ibm.com*. January 2, 2018. Archived (<https://web.archive.org/web/20190424083730/https://www.ibm.com/analytics/data-management/resources/what-is-hybrid-data-management/>) from the original on April 24, 2019. Retrieved April 24, 2019.
11. "What is personally identifiable information (PII)? How to protect it under GDPR" (<https://www.csoonline.com/article/3215864/how-to-protect-personally-identifiable-information-pii-under-gdpr.html>). *csoonline.com*. Archived (<https://web.archive.org/web/20190424083727/https://www.csoonline.com/article/3215864/how-to-protect-personally-identifiable-information-pii-under-gdpr.html>) from the original on April 24, 2019. Retrieved April 24, 2019.

---

Retrieved from "[https://en.wikipedia.org/w/index.php?title=Data\\_stream&oldid=1091042824](https://en.wikipedia.org/w/index.php?title=Data_stream&oldid=1091042824)"

---

**This page was last edited on 1 June 2022, at 21:15 (UTC).**

Text is available under the Creative Commons Attribution-ShareAlike License 3.0; additional terms

may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.