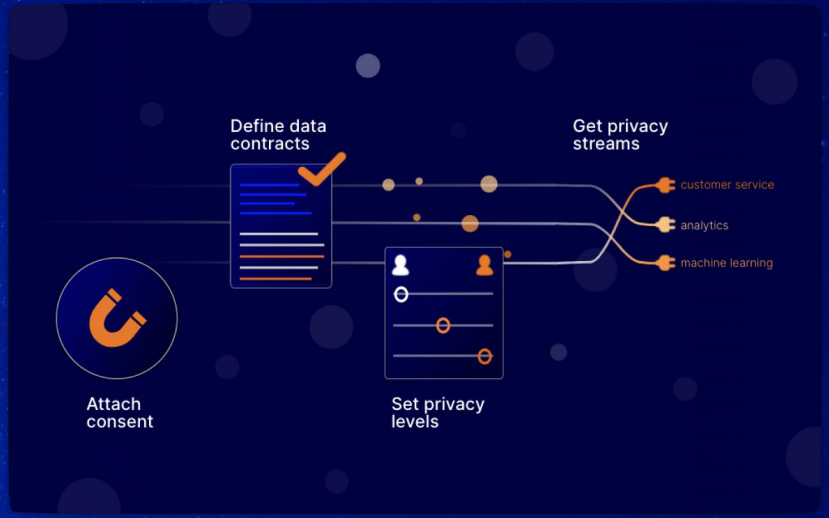
STRM PRIVACY

Data in, privacy out

Data streams for the privacy age

@robintrietsch & @pimnauts



Data Council

Today

```
sections:
- "Learning how to build for privacy"
- "An architecture to shift left with data privacy"
- "(our) solution!"
```

STRM PRIVACY

May 25th 2018

STRM PRIVACY

The Cost of Privacy

Coordination, regulation limit velocity, innovation

Retaining data utility...

Data teams carry ultimate responsibility

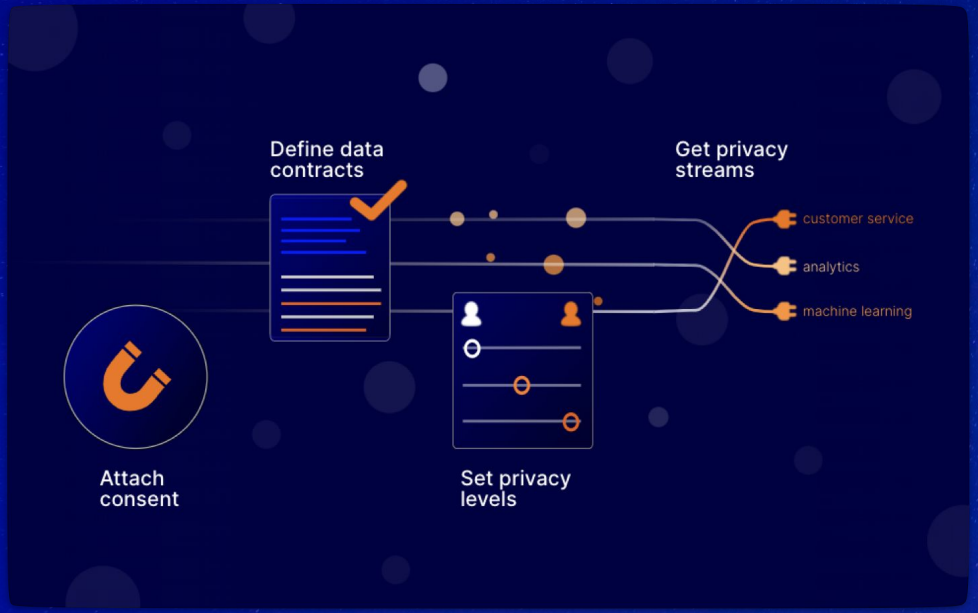
Towards privacy by design for data

Beyond privacy as an access or data sharing problem

- Explicit beats implicit data structure (data contracts)
- Shift left to encode privacy at data collection
- Privacy-transform before any ETL/ELT, retain utility.
- Purpose-binded interfaces (privacy streams)
- Make it interoperable && fast

STRM PRIVACY

Data in, privacy out

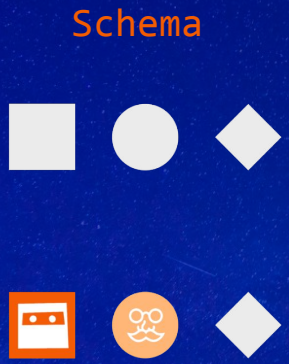


1. Explicit data contracts
2. “Encode” privacy
3. Privacy-transform before anything else
4. Provide privacy streams.

Explicit data structure

Data shape
(fields)

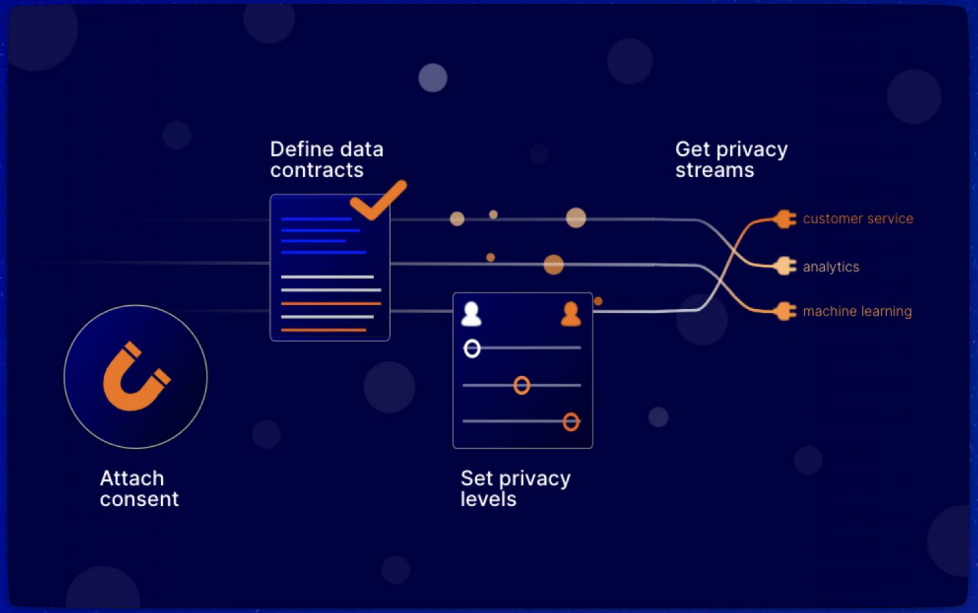
Privacy implications




```
~
%
{
  "ref": {
    "handle": "strmprivacy",
    "name": "sample",
    "version": "1.0.0"
  },
  "schemaRef": {
    "handle": "strmprivacy",
    "name": "webshop-clickstream",
    "version": "4.0.4"
  },
  "state": "ACTIVE",
  "isPublic": true,
  "keyField": "sessionId",
  "piiFields": {
    "previousPage": 1,
    "sessionId": 2,
    "elementName": 3
  },
  "validations": [
    {
      "field": "sessionId",
      "type": "regex",
      "value": "^[0-5]{4}-[5-9]{4}$"
    }
  ],
  "metadata": {
    "title": "STRM privacy Webshop Clickstream"
  }
}
```

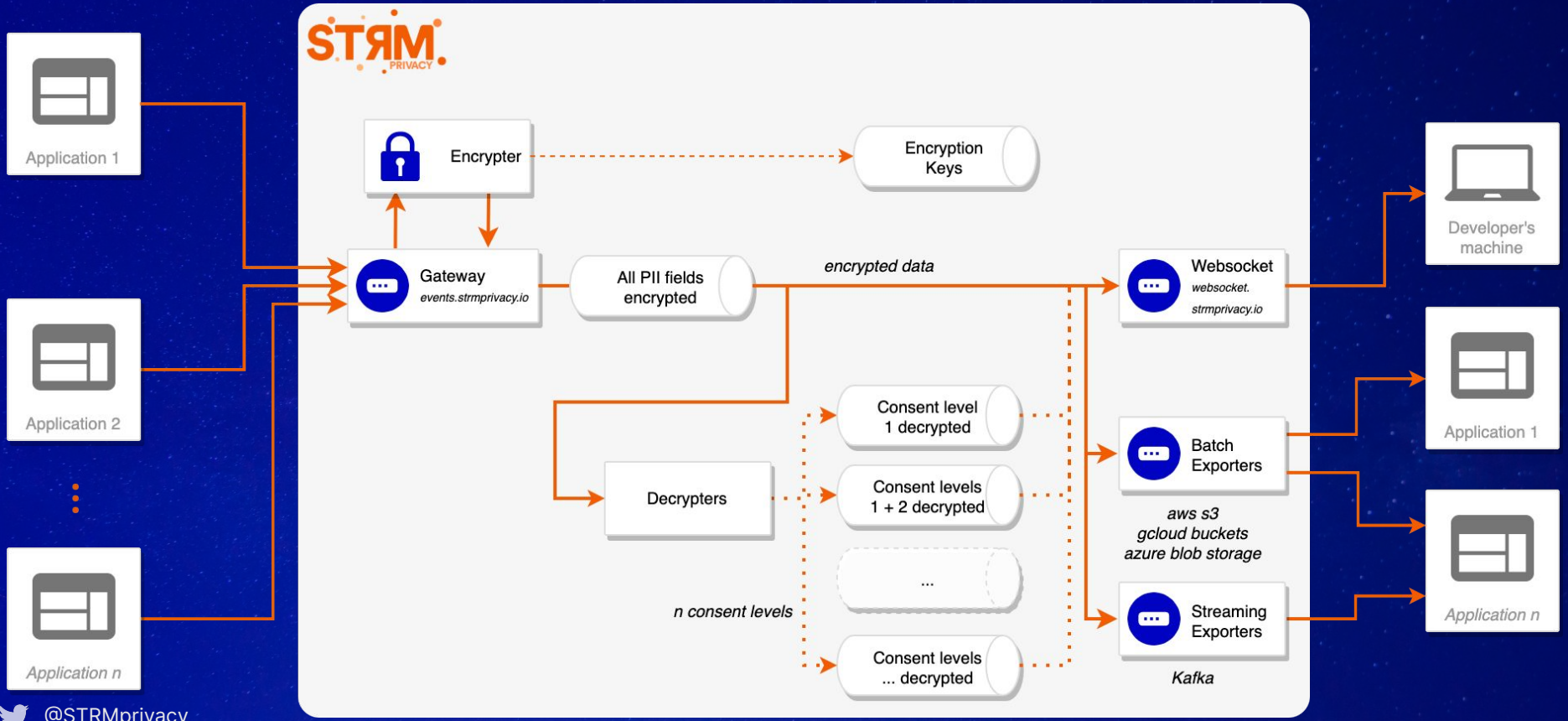
STRM PRIVACY

Data in, privacy out



1. Explicit data contracts
2. “Encode” privacy
3. Privacy-transform before anything else
4. Provide privacy streams.

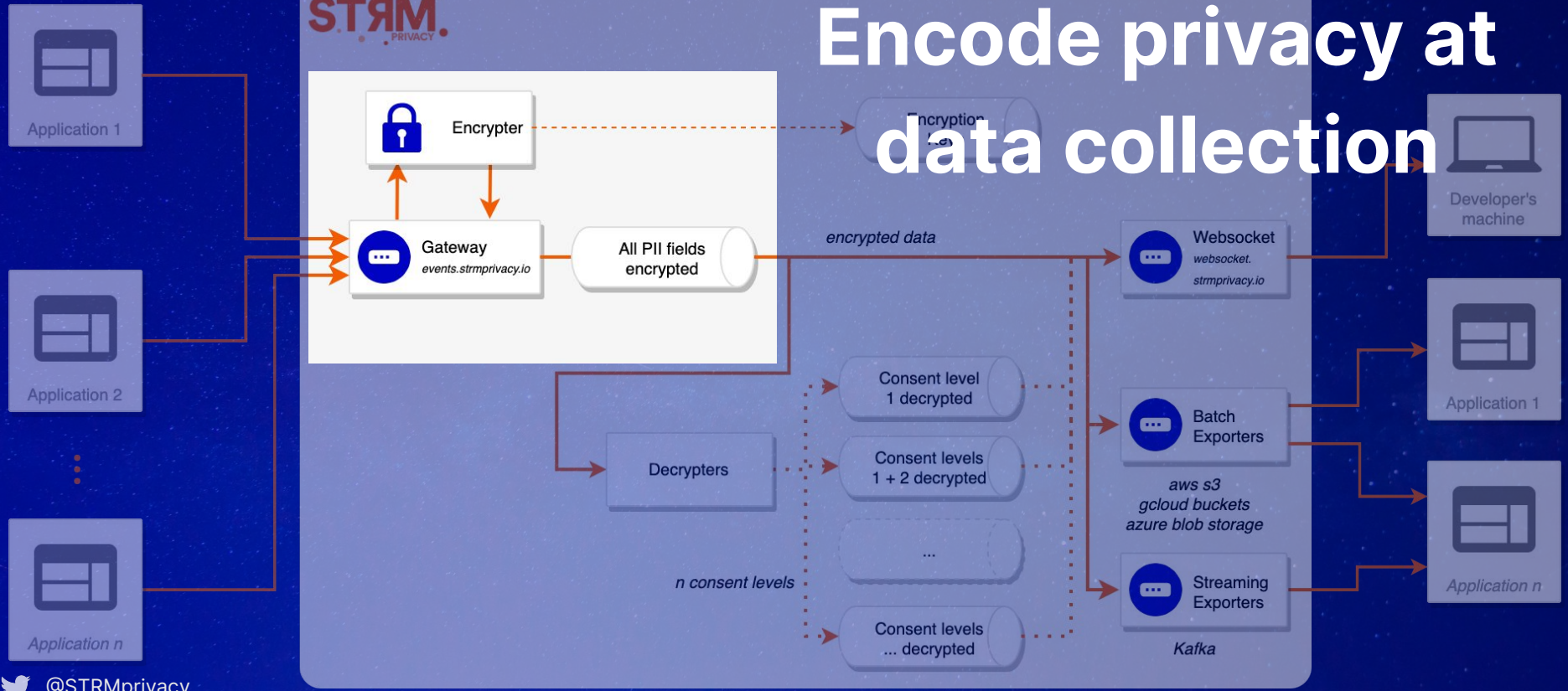
STRM PRIVACY



@STRMprivacy

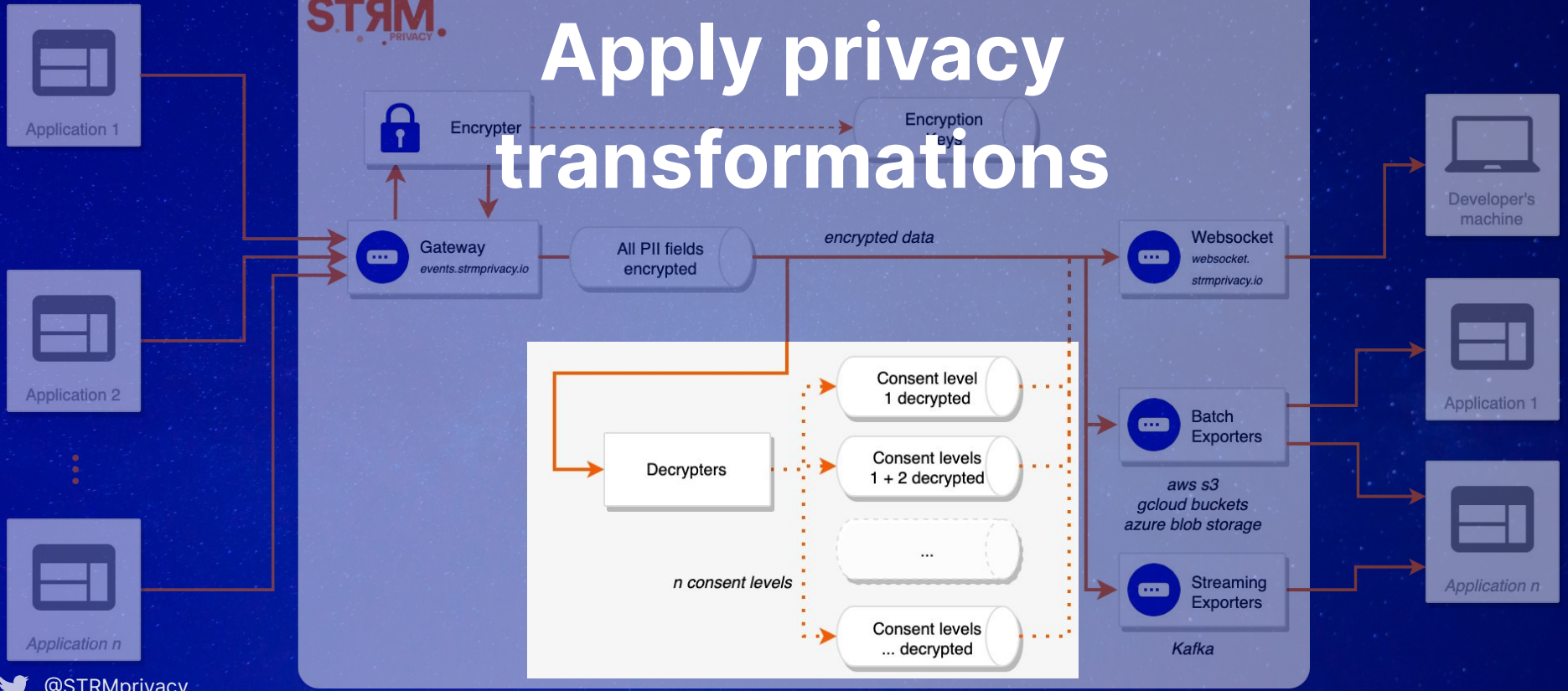
STRM PRIVACY

Encode privacy at data collection



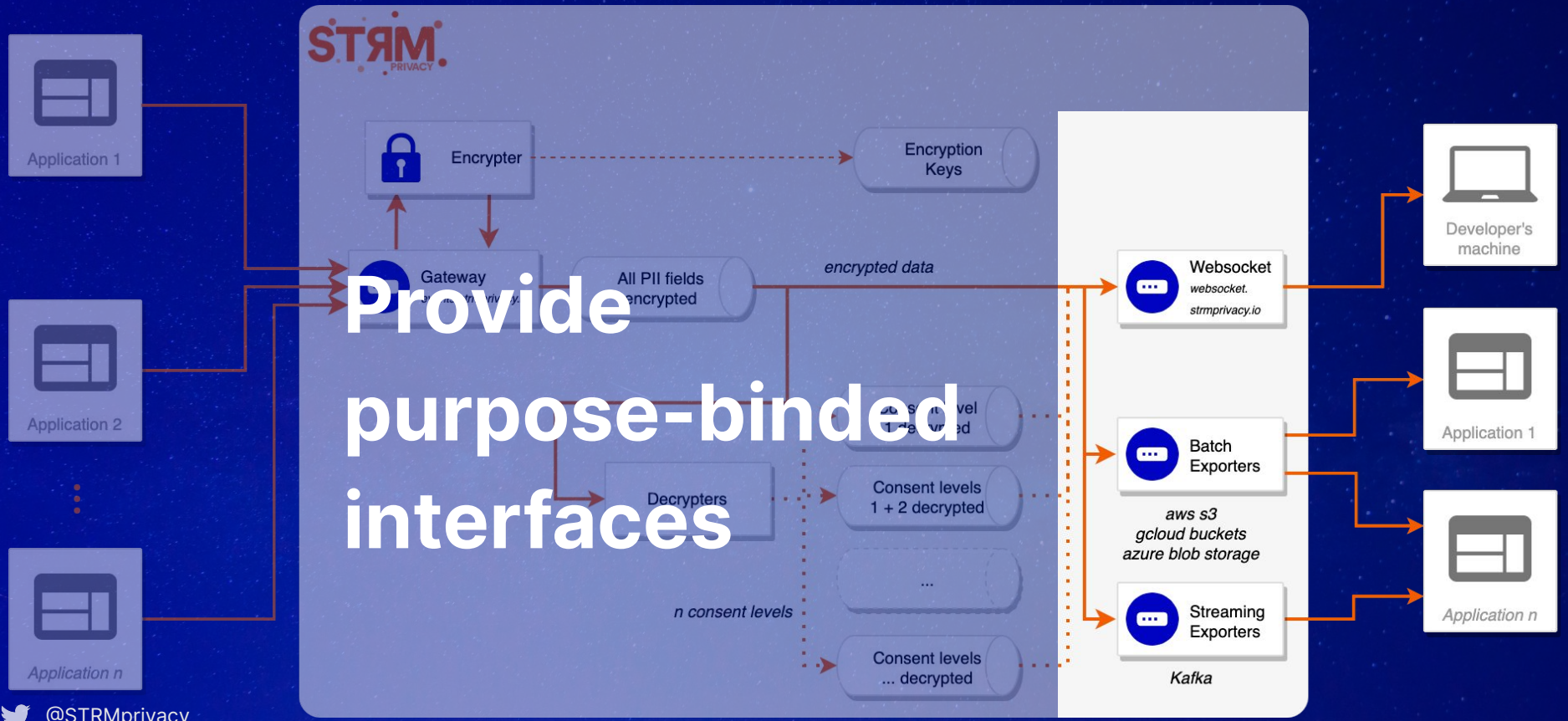
STRM PRIVACY

Apply privacy transformations



@STRMprivacy

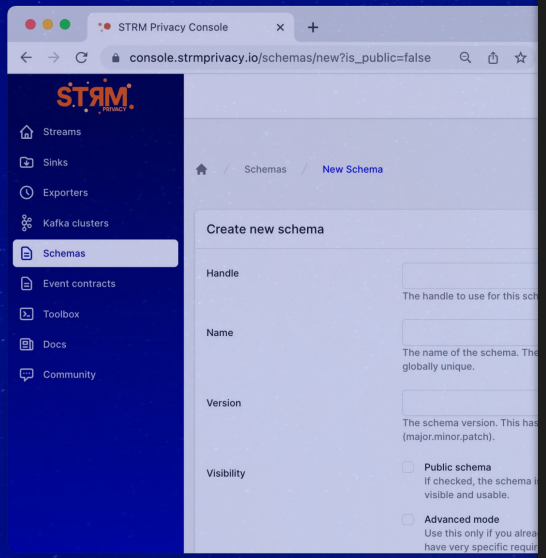
STRM PRIVACY

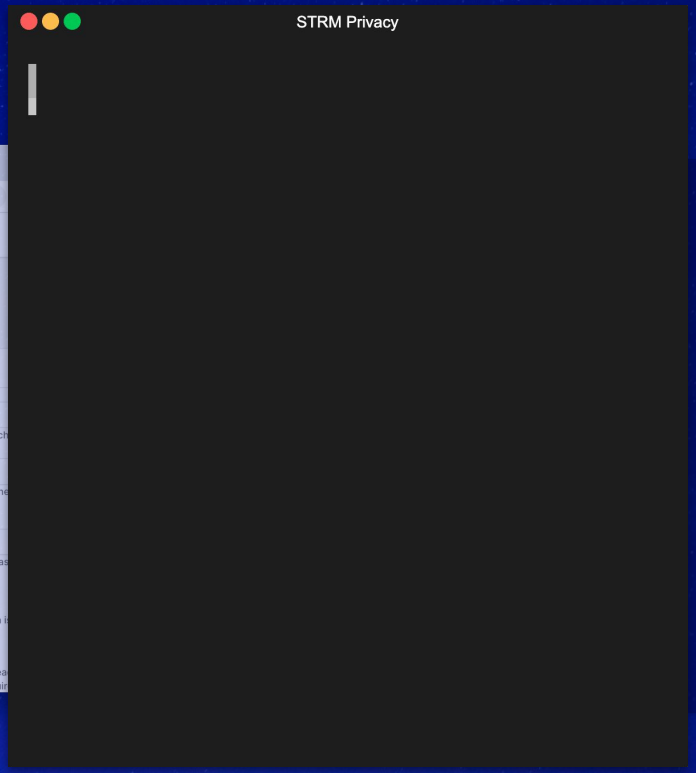


Provide purpose-binded interfaces

@STRMprivacy

Manage + integrate



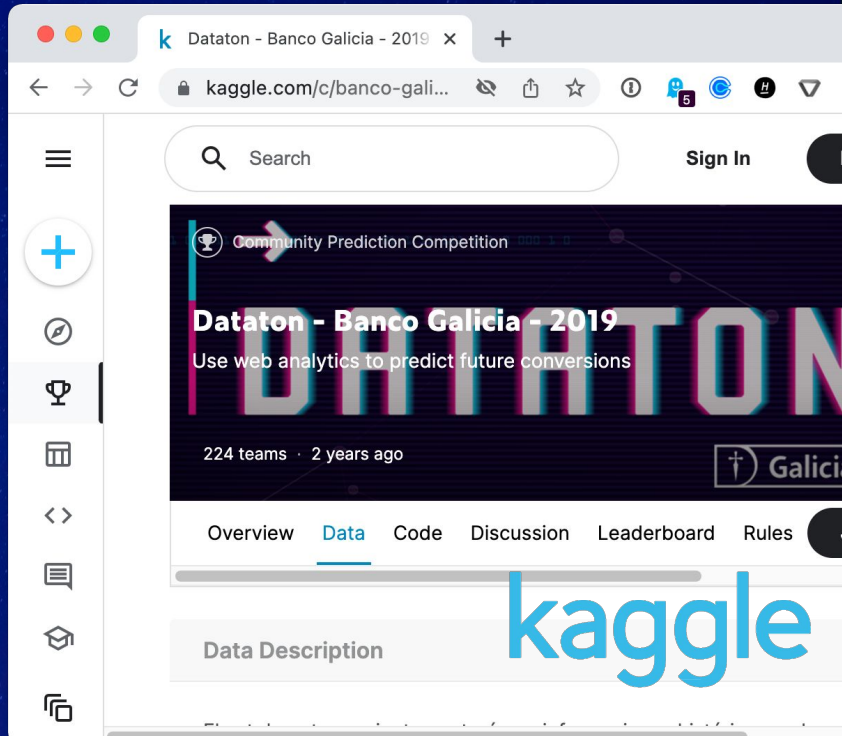


@STRMprivacy

STRM PRIVACY

Demo

Visualize on real-time privacy streams



STRM PRIVACY

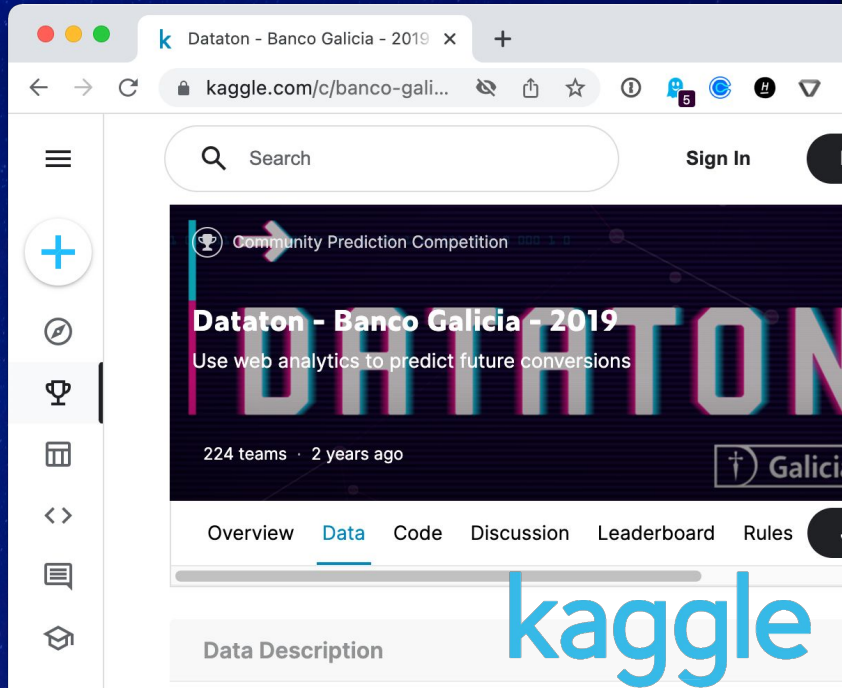
Banco Galicia clickstream

Clickstream Session Data

1. Create a data contract
2. Data replay to simulate a stream
3. Real-time data visualization

All with real-time anonymization

on (PII) data.



STRM PRIVACY

Customer use cases

- Predict on real-time anonymized streams (search, recsys) < demo
- Operational analytics in sensitive domains (e.g. mobility)
- Low-latency, high-volume external data sharing (e.g. health sensor/IoT)
- Structured privacy (privacy policy 🤝 data processing)

STRM PRIVACY

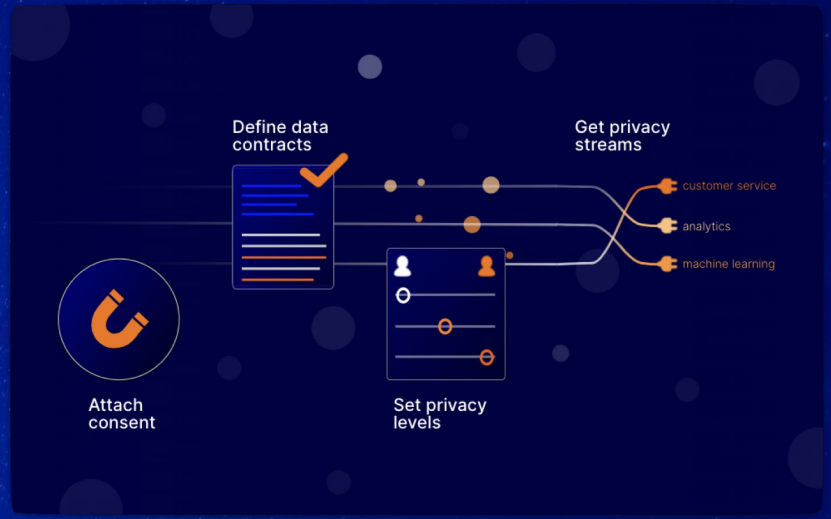
Summary

- Insert data privacy before any downstream usage or storage.
- Encode privacy policies into data through **data contracts**
- Align privacy policy 🤝 data processing in one go.
- Distribute purpose-driven data interfaces through **privacy streams**
- All in real-time if necessary (and batch, too 😉)

STRM PRIVACY

Data in, privacy out

Data streams for the privacy age



strmprivacy.io

/strmprivacy

@STRMprivacy

Data Council

Various icons provided by The Noun Project