DATA+AI SUMMIT 2022

Embedding Privacy by Design Into Data Infrastructure

Through Open-Source Extensive Tooling

ORGANIZED BY databricks

Cillian Kieran
Founder & CEO, Ethyca

Today

What is Privacy-as-code
Overview of Fides
Use case 1: checking policies in CI
Use case 2: automating data rights requests

Privacy is broken.

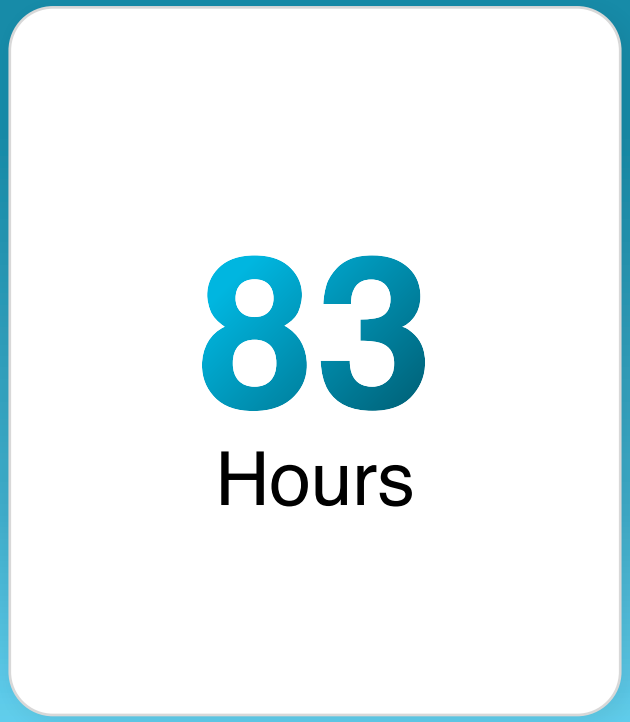
Privacy-as-code is the solution.


PIPEDA
GDPR
CDPA
CPRA
CPA
APPI
PIPL
PDPB
LGPD
APA
POP
PDP

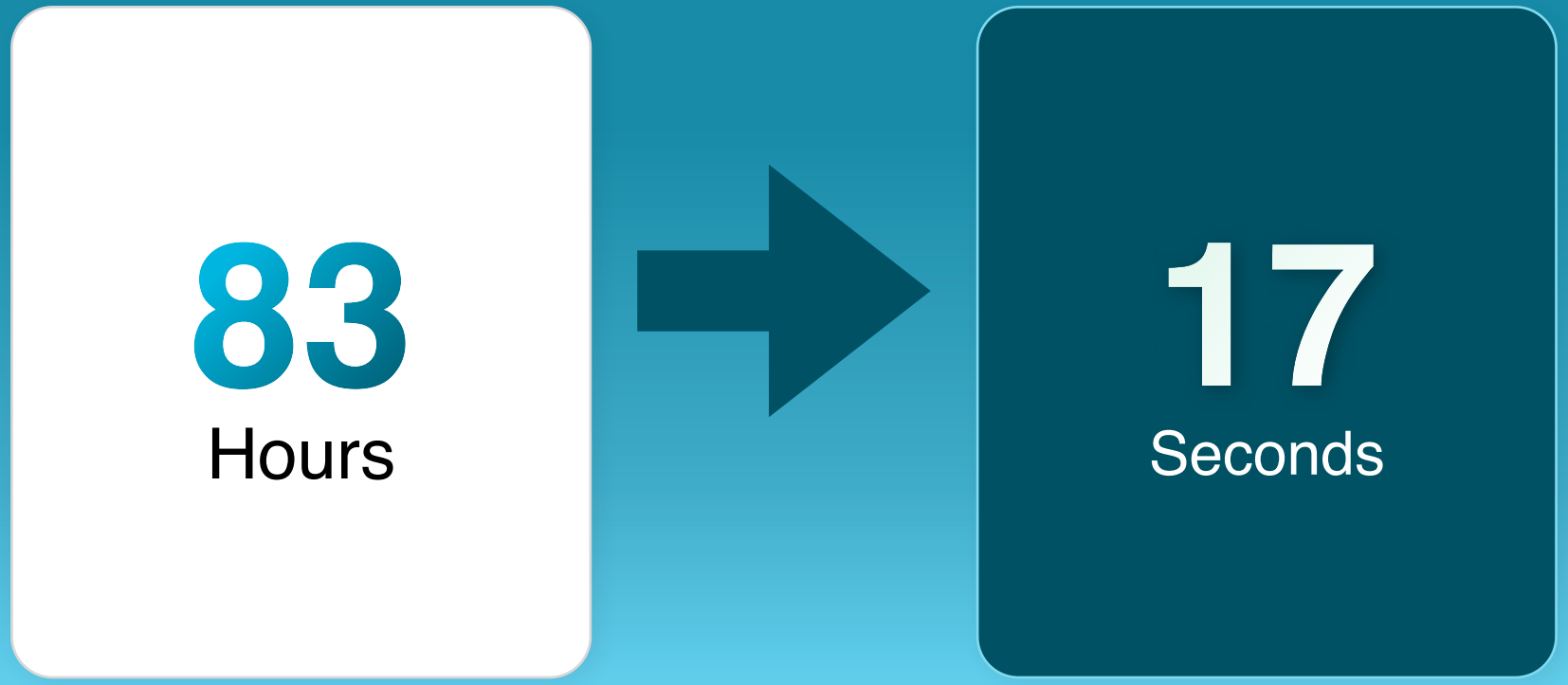
9 out of 10 of companies

have no automated data inventory or data privacy orchestration capabilities

The average privacy request takes



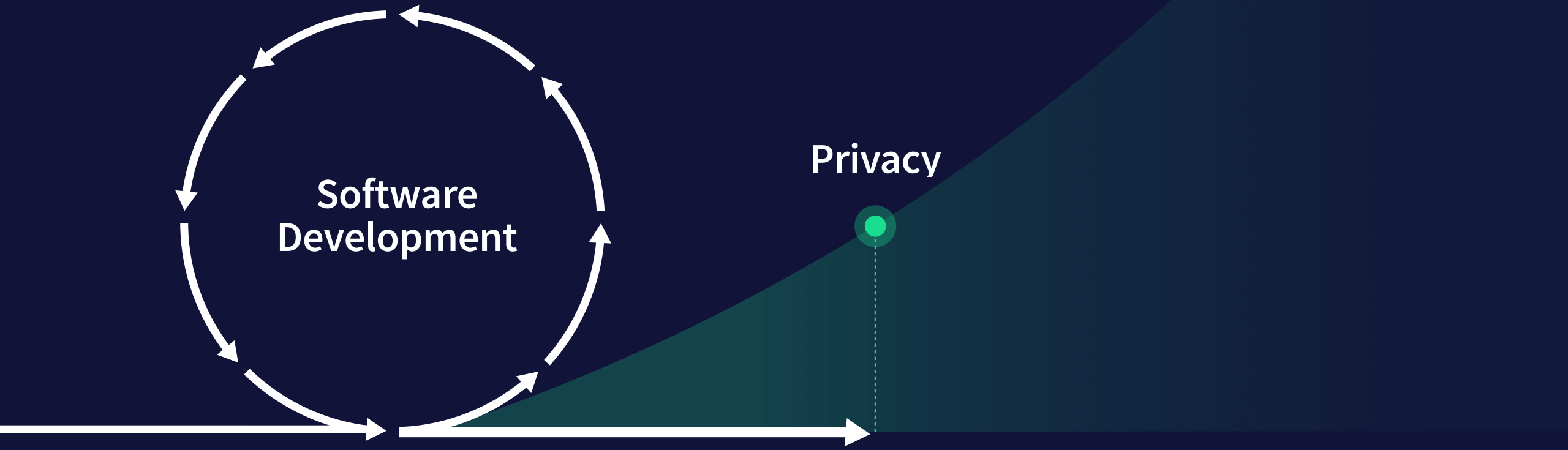
Developer tools that make privacy effortless



Privacy is a fundamental
Human Right.

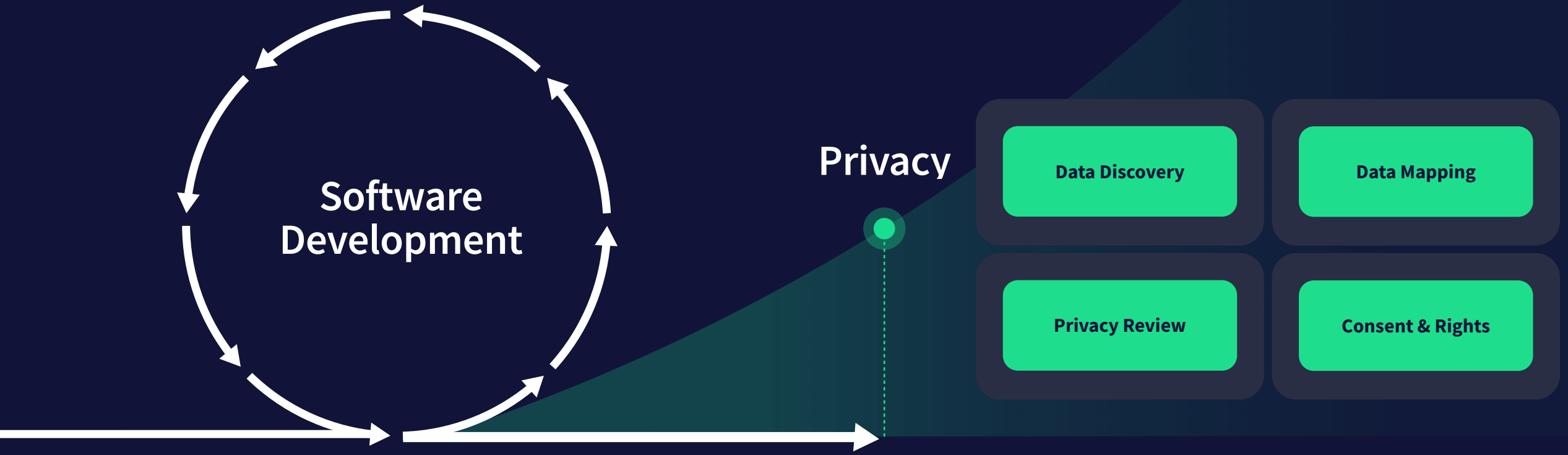
Privacy is really complicated.

Privacy today is considered after software is shipped, leading to pain for developers and lawyers.



Privacy today is considered after software is shipped, leading to pain for developers ~~and~~ lawyers.

from



These laws create tremendous complexity for engineers.

As engineers; what are we trying to accomplish?

CONTEXT

Describe the types of data we're using and what we're using them for.

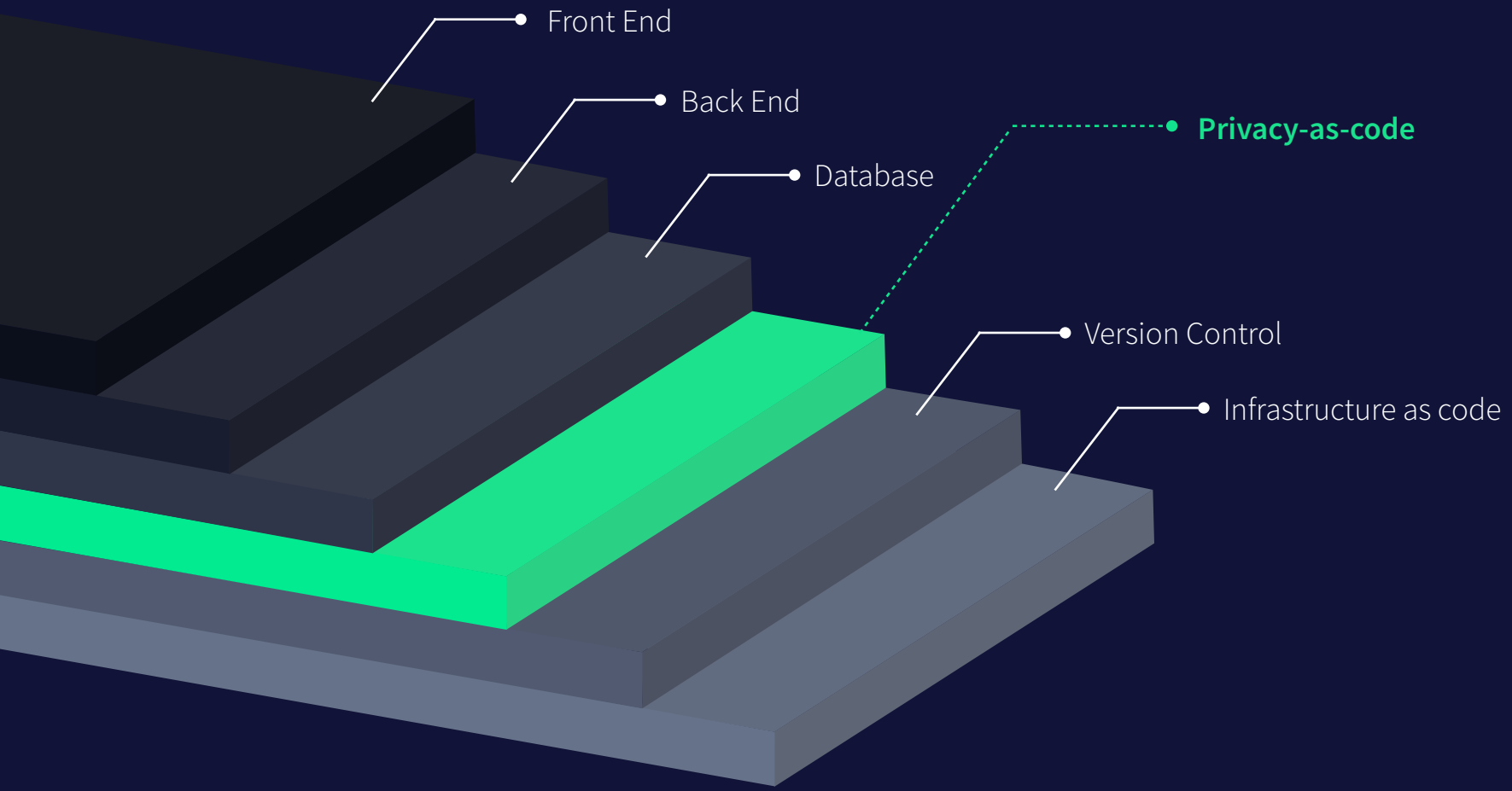
RISK

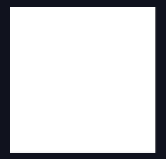
Ensure we're doing anything risky or dangerous with that data.

RIGHTS

Easily manage user rights for access and deletion of data.

Privacy-as-code are tools that make it effortless for developers to implement Privacy by Design



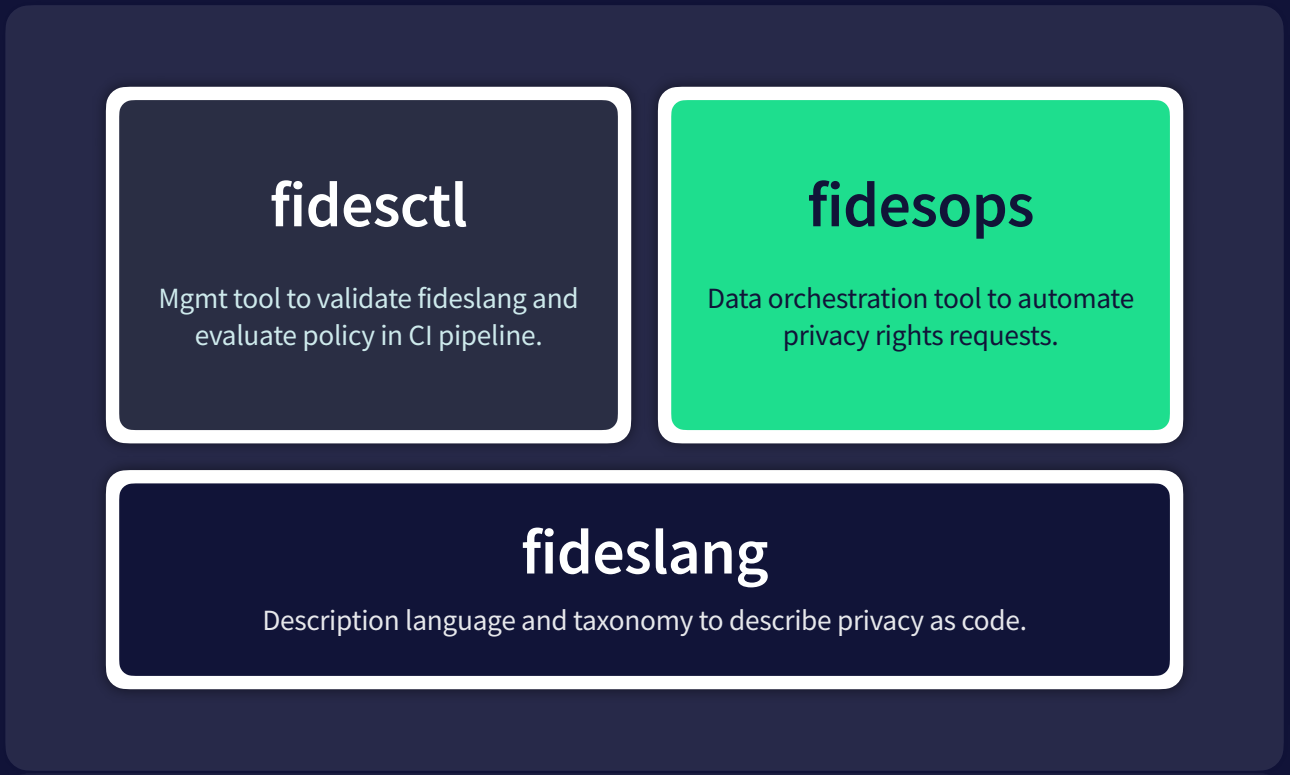


fides

License Apache 2 License CC BY 4.0

Open-source developer tools for Privacy by Design in any tech stack.

A set of tools built to help engineers and data teams ship privacy-compliant systems effortlessly.

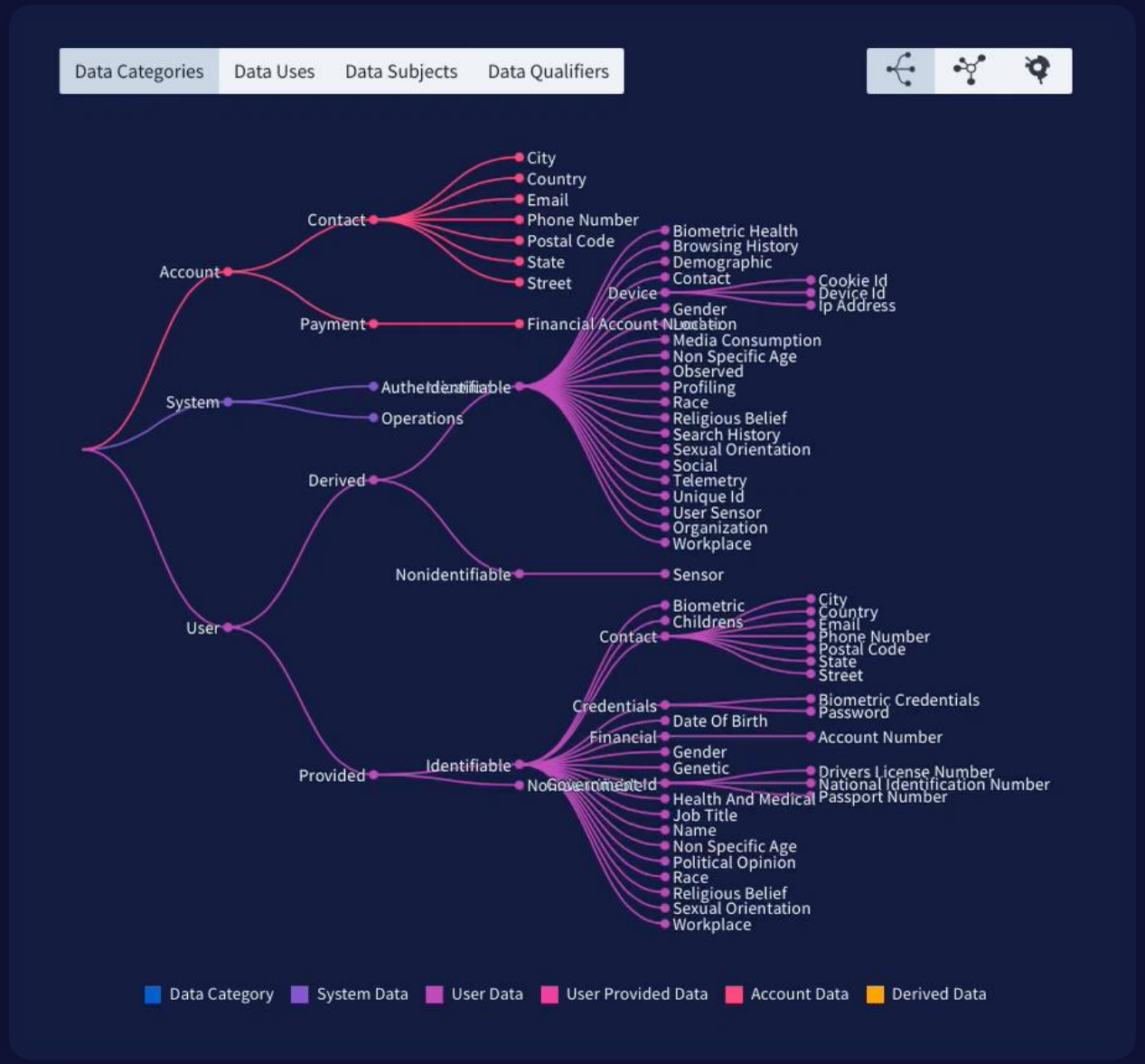

fidesctl
Mgmt tool to validate fideslang and evaluate policy in CI pipeline.
fidesops
Data orchestration tool to automate privacy rights requests.
fideslang
Description language and taxonomy to describe privacy as code.

fides lang

A description language that supports GDPR, CCPA, LGPD and ISO 19944.

In other words:
easily describe personal data in all systems.

Explorer fid.es/taxonomy



Fides Declarations

Light-weight declarative language
Dot notation (mostly)
YAML in your projects (inline declarations coming soon)
System operations data

```
system.operations
```

User provided email address

```
user.provided.identifiable.contact.email
```

Fides Primitives

Organizations
Systems
Datasets
Policies

Organizations

1. Represents all or any part of an organization.
2. Establishes the root of the resource hierarchy.
3. Organizations are unique, i.e. you cannot reference other organization scopes.

Fides Primitives

- Organizations
  - Systems
- Datasets
- Policies

Systems

1. Represents the privacy properties of a single project, services, codebase or application.
2. Describes the categories of data being processed and use of the data in the system.

Fides Primitives

Organizations
Systems
    # Datasets
Policies

Datasets

1. Represent any location data is stored; databases, data warehouses or other stores.
2. You can declare individual fields of data and describe the types of data they are storing.

Fides Primitives

Organizations
Systems
Datasets
  # Policies

Policies

1. Represents a set of rules that a system must adhere to — your privacy policy as code.
2. Fidesctl evaluates these policies against system/dataset declarations for compliance.

Fides Taxonomy

- # Data Categories
- # Data Subjects
- # Data Uses
- # Data Qualifiers

Data Categories

"What" type of data am I processing

```
user.provided.identifiable.contact
```

Fides Taxonomy

Data Categories
    # Data Subjects
Data Uses
Data Qualifiers

Data Subjects

"Who" is the owner of the data

```
- user
- customer
- patient
```

Fides Taxonomy

Data Categories
Data Subjects
    # Data Uses
Data Qualifiers

Data Uses

*"**How**"* is the data being used

```
personalize.system
```

Fides Taxonomy

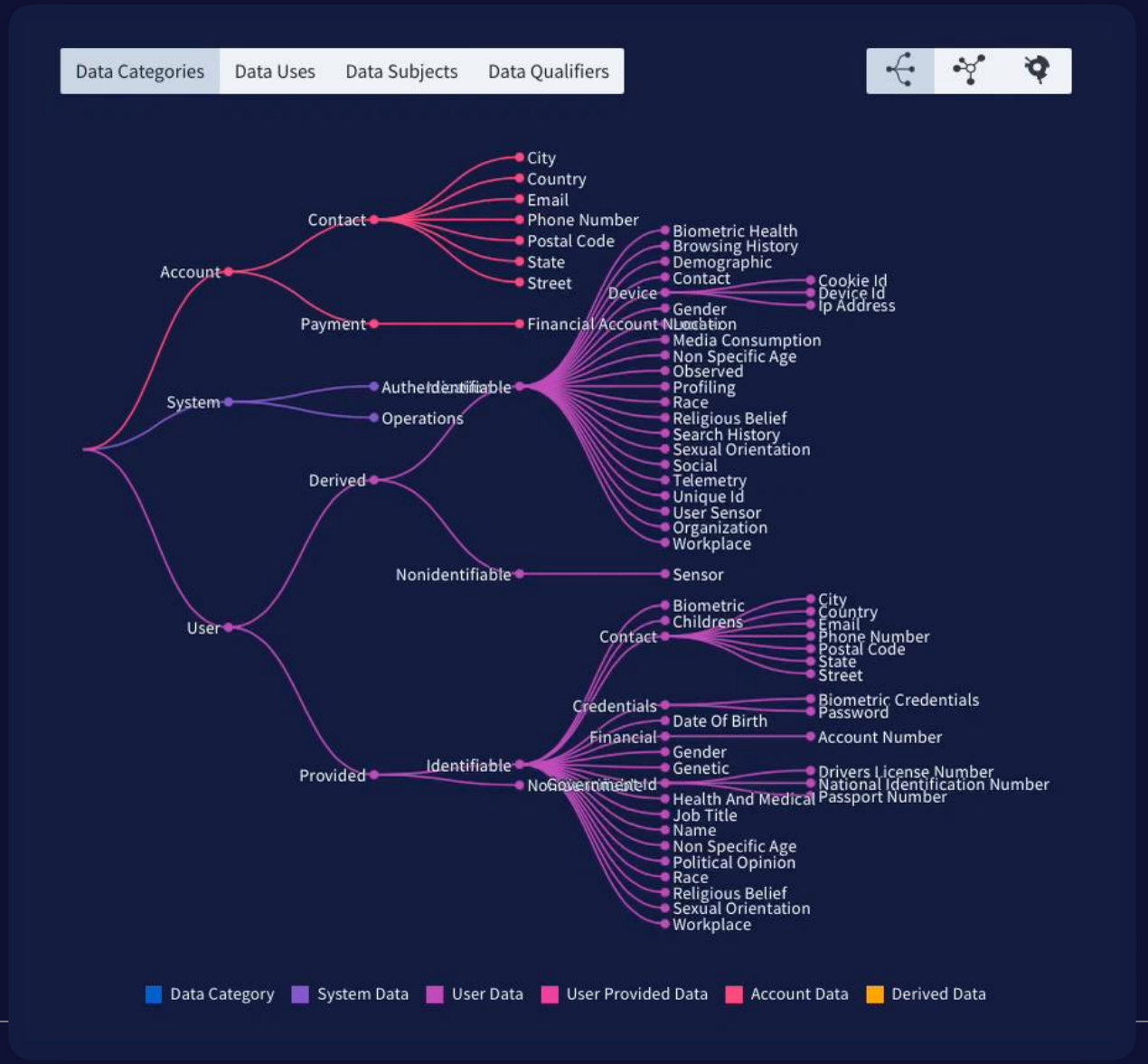
- # Data Categories
- # Data Subjects
- # Data Uses
  - # Data Qualifiers

Data Qualifier

How re-identifiable a person is

```
aggregated.anonymized
```

fid.es/taxonomy



Using Fides, you can describe…

- What type of data your application processes (`data_category`)
- How your system uses that data (`data_use`)
- What policies you want your system to adhere to
- And more!

fidesctl

Developer tools
Pushing Privacy Left
Compliance checks in CI/CD pipelines

fidesops

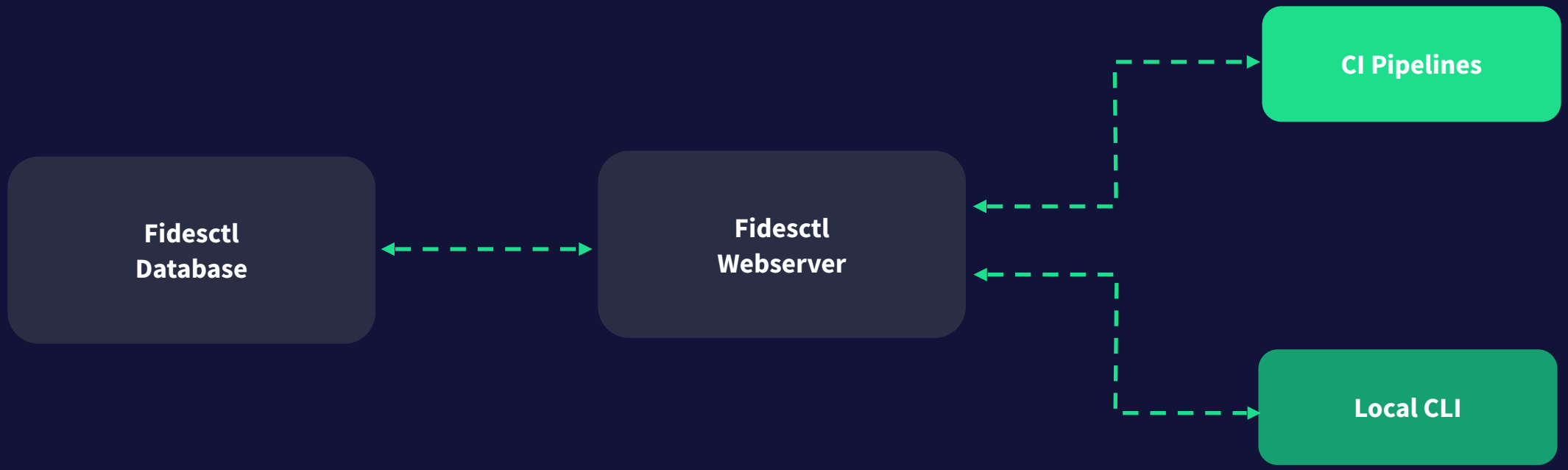
Runtime Application
Manage and automate Data Subject Requests to comply with GDPR et al.

fidesctl

Use Case 1

Policy Enforcement in Development

fidesctl Components



Getting Started: Configuration

```
[cli]
server_url = “http://fidesctl:8080"

[api]
database_url = “postgresql+psycopg2://postgres:fidesctl@fidesctl-db:5423/fidesctl"
test_database_url = “postgresql+psycopg2://postgres:fidesctl@fidesctl-db:5423/fidesctl_test"

log_destination = ““
log_level = “INFO”
log_serialization = ““
```

Github repo fid.es/ctl

Config docs fid.es/config

Getting Started: Commands

`annotate`	Annotate fidesctl resource types.
`apply`	Validate local manifest files and persist any changes via the API server.
`db`	Database utility commands.
`delete`	Delete a resource on the server.
`evaluate`	Compare your System's Privacy Declarations with your Organization's Policies.
`export`	Export fidesctl resource types
`generate`	Generate fidesctl resource types
`get`	View a resource from the server as a JSON object.
`init`	Initializes a Fidesctl instance, creating the default directory.
`scan`	Scan external resource coverage against fidesctl resources

```
# fidesctl -h
```

fid.es/commands

PIA

An overview of the requirements and methodology

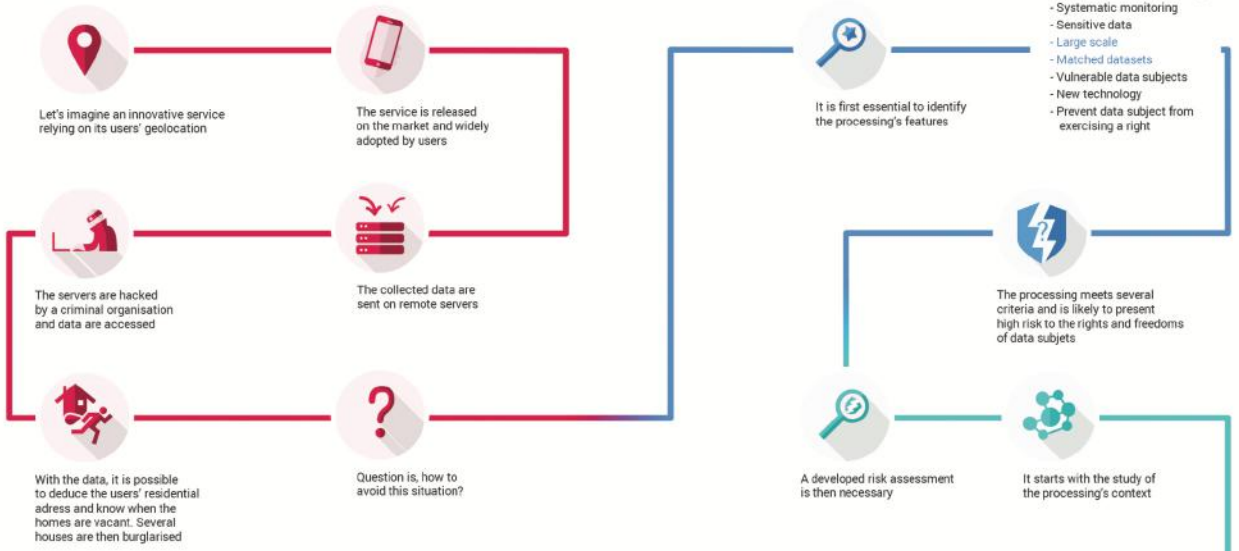
0. Launching a new processing

Every day in the digital realm, numerous services are created.

Those services usually rely on the processing of personal data aiming at fulfilling the needs of organisations or their users.

The supporting assets used to store the data have different levels of vulnerabilities toward feared events such as illegitimate acess, unwanted change, or disappearance of personal data.

Those risks are likely to have significant impacts on the users' privacy.



1. Considering the processing

For the data processor as well as the data subjects, those risks are unwelcome.

Before carrying out a processing, it is essential to analyse it to understand its inherent risks.

Several factors affect the riskiness of a processing, as the kind of data processed.

Generally speaking, if a processing meets two of the criteria listed, then it is likely to present high risks and would require to carry out a privacy impact assessment.

2. Evaluating the privacy risks

The assessment first establishes the context in which the processing is carried out, including its purpose and technical features.

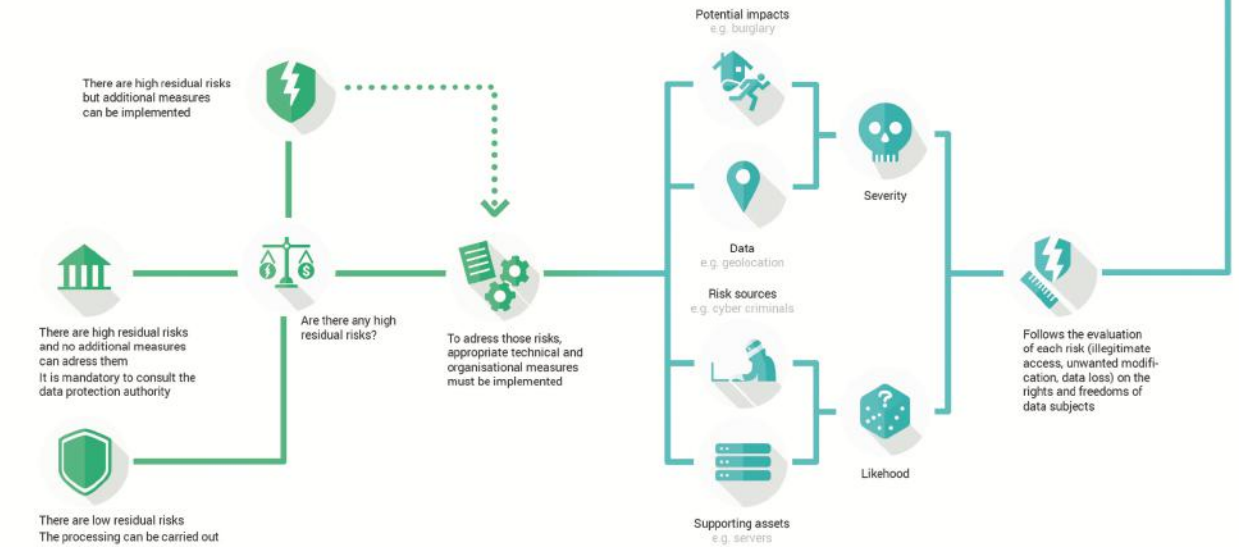
In addition to studying the fundamental principles, made up of the necessity and proportionality of the processing, each risk has to be analysed to evaluate its severity and likehood according to its potential impacts on the rights and freedoms of data subjects, the data processed, the risks sources and the supporting assets.

3. Adressing the risks

Once the risks have been identified, it should be determined if they are acceptable given the existing and planned technical and organisational measures.

If it doesn't seem possible in regard of the foreseen measures, the data protection authority has to be consulted.

In any case, it is mandatory to implement the planned controls before carrying out the processing.



Privacy Dev Tools: Check Policy in CI

What type of data we're processing and storing

What we're doing with that data in our system

Check it against the policies set by our organization

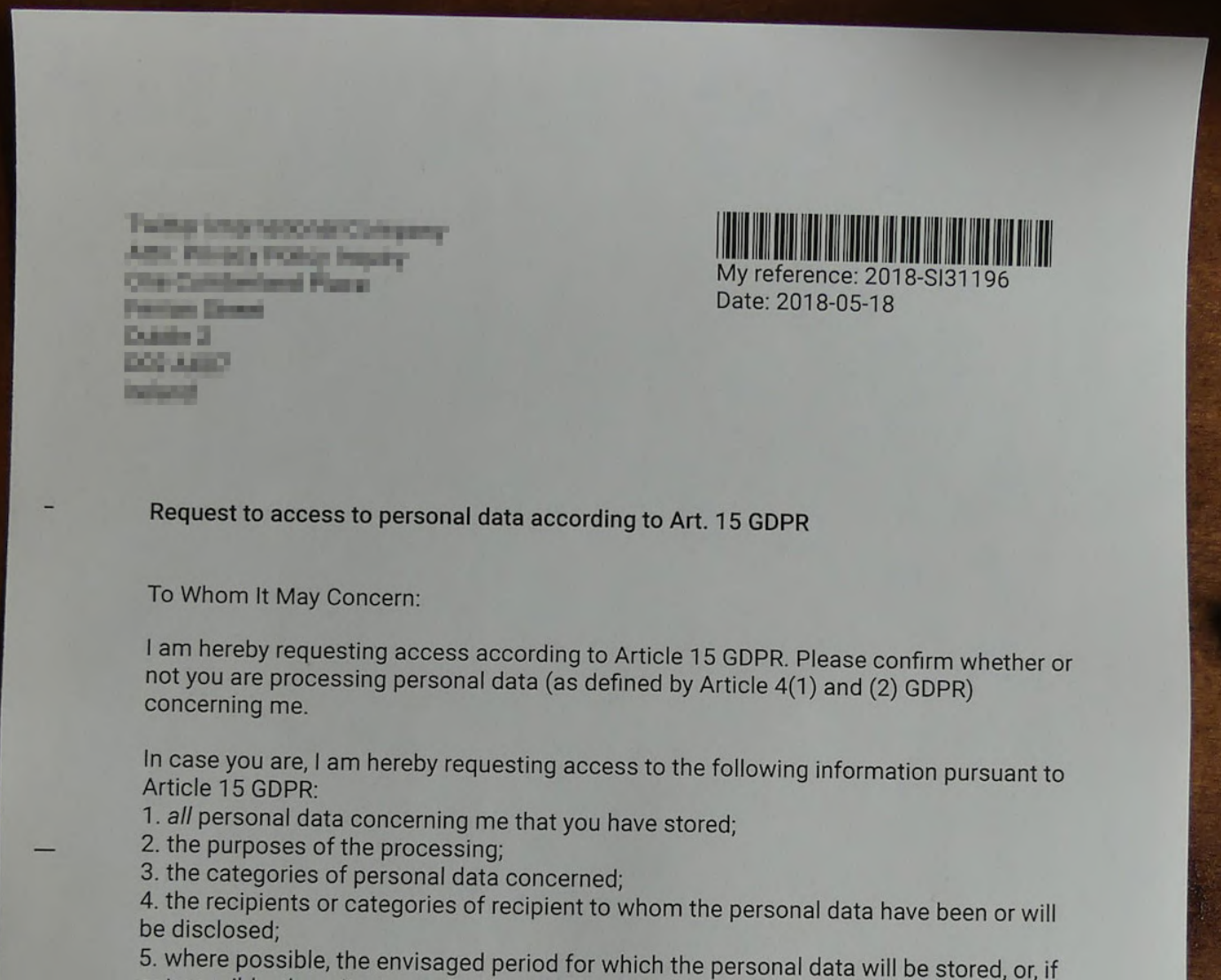
Maintain an audit trail of evaluations

fidesops

Use Case 2

Programmatic Data Rights Requests

Data Subject Access Requests look like this…



[redacted]

My reference: 2018-SI31196
Date: 2018-05-18

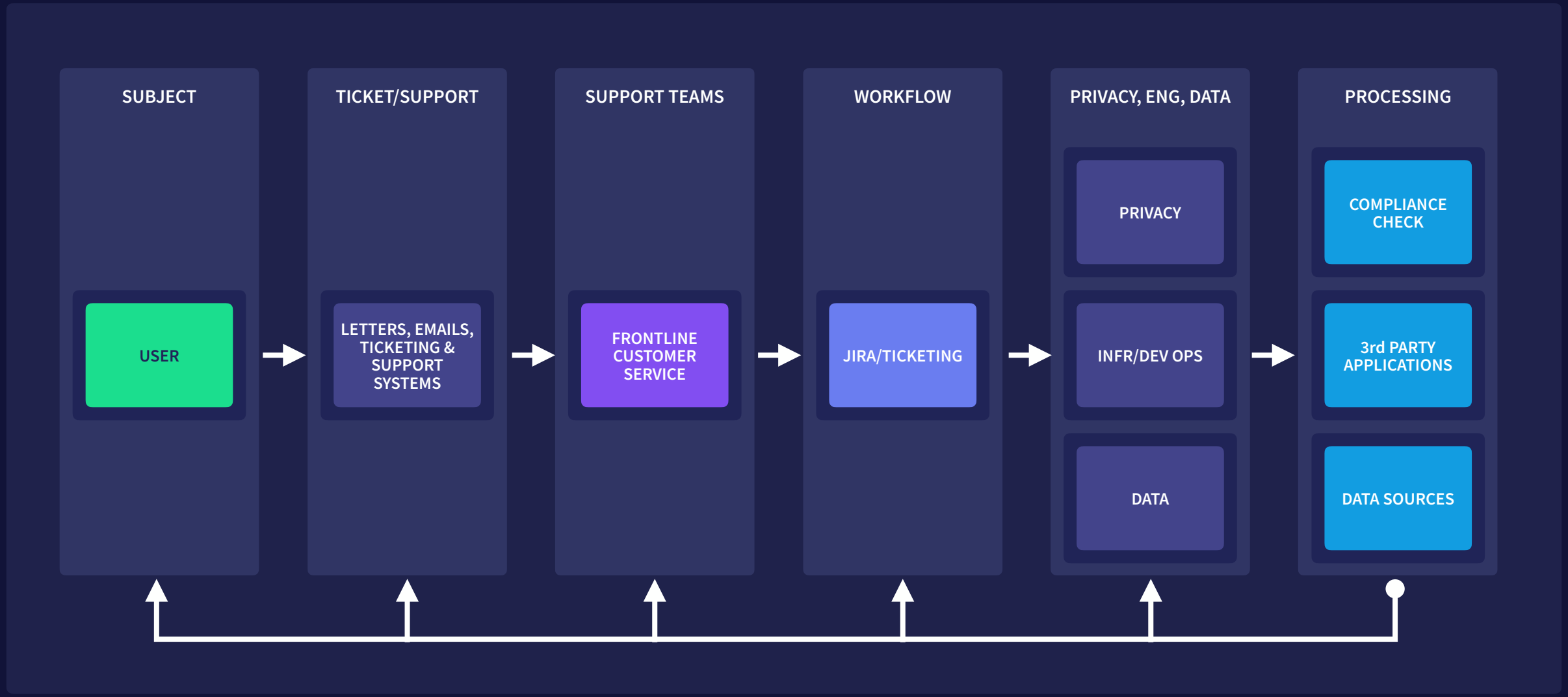
Request to access to personal data according to Art. 15 GDPR

To Whom It May Concern:

I am hereby requesting access according to Article 15 GDPR. Please confirm whether or not you are processing personal data (as defined by Article 4(1) and (2) GDPR) concerning me.

In case you are, I am hereby requesting access to the following information pursuant to Article 15 GDPR:

1. *all* personal data concerning me that you have stored;
2. the purposes of the processing;
3. the categories of personal data concerned;
4. the recipients or categories of recipient to whom the personal data have been or will be disclosed;
5. where possible, the envisaged period for which the personal data will be stored, or, if


SUBJECT
TICKET/SUPPORT
SUPPORT TEAMS
WORKFLOW
PRIVACY, ENG, DATA
PROCESSING
USER
LETTERS, EMAILS, TICKETING & SUPPORT SYSTEMS
FRONTLINE CUSTOMER SERVICE
JIRA/TICKETING
PRIVACY
INFR/DEV OPS
DATA
COMPLIANCE CHECK
3rd PARTY APPLICATIONS
DATA SOURCES

Privacy Dev Tools: Automation Data Request

Construct a request policy

Retrieve data and collate data across systems

Make a development change that affects the data model

Re-run the request to see dev change reflected in model

Privacy Dev Tools: Summary

- # Described our system's personal data characteristics
- # Run evaluations in CI to know we're complying with policies
- # Automated retrieval of user data across systems
- # Automatically update requests against dev changes

Helpers: Simplifying Privacy

Many helpers for simplifying privacy

ML classifier to automate dataset labeling/annotation

UI for better visual management of the data map and taxonomy

Roadmap tackles each major privacy engineering problem

Conclusion

Privacy Engineering Platform

We've covered

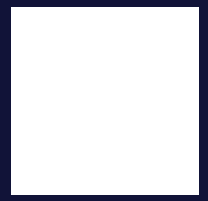
- What is Privacy-as-code
- Overview of Fides
- Use case 1: checking policies in CI
- Use case 2: automating data rights requests

Easy to get your team using Fides

Fides is a **free, open-source** Python project on Github

Up and running in **15 mins with the quick-start** engineering guide

Great documentation and **Fides slack community** to support you



fides

Open-source developer tools for privacy.

Thank you. Comments, Issues, PRs welcome!

Join the community fid.es/join

Download this presentation fid.es/databricks