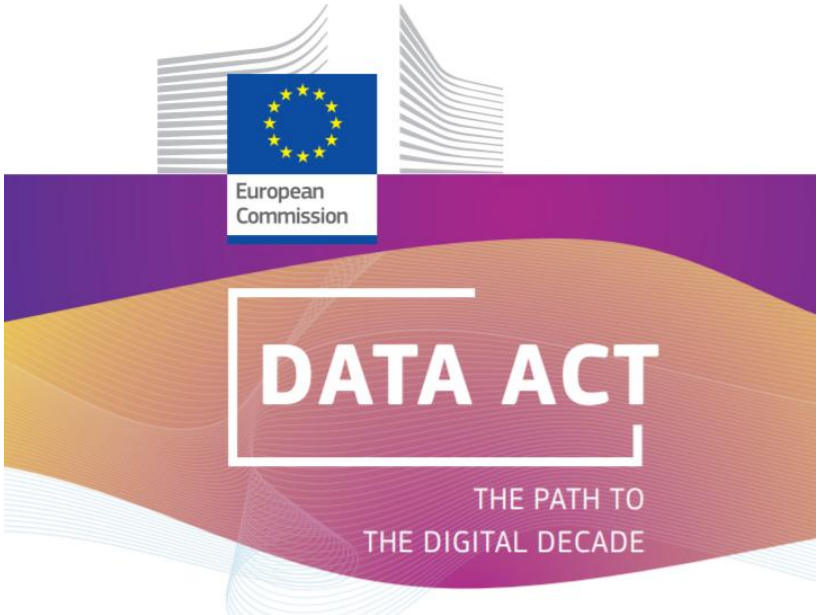
A perspective on the use of personal data stores in data spaces to support Generative AI

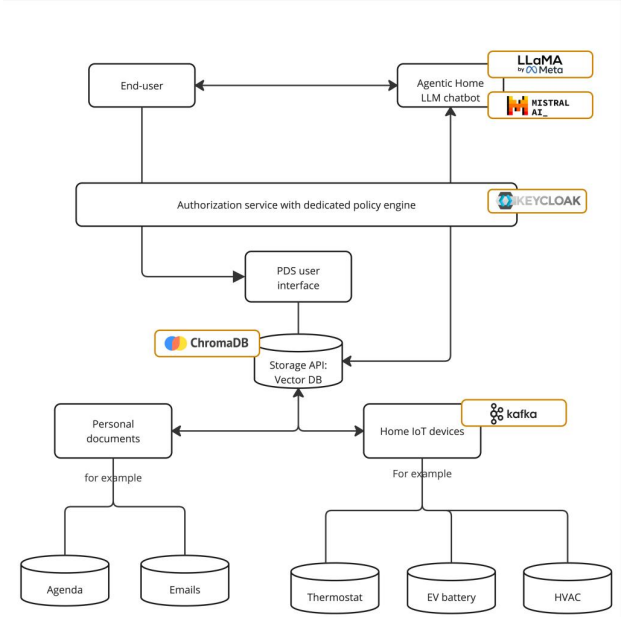
Tanguy Coenen, Data Week Luxemburg, 10/12/2024

What will this talk be about?

2 parts…



An analysis on connected product personal data sharing in the Data Act and its relevance to data spaces and genAI



A technical architecture to link personal data stores to genAI

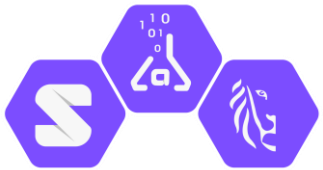
imec

Why talk about this?

1. The EU data strategy is aiming for a **level playing field** in the data economy
2. The **Data Act** is a key piece of legislation in this respect
3. There is still some unclarity on how to **support** the Data Act from a **technical** point of view
4. The EU is struggling to find its competitive position in terms **the AI industry**
5. The killer app for data spaces could well be genAI
6. Applying **GenAI at the edge** offers a possible way forward, combining European strengths, building on **new legislation**

Part 1: the data act and personal data stores

A multidisciplinary (legal, business, computer science) analysis of the Data Act regarding personal data sharing in connected products and a way forward to comply through Personal Data Stores



KU LEUVEN CiTiP
CENTRE FOR IT & IP LAW

IDLab
INTERNET & DATA LAB

imec

Flanders AI research

Flanders
State of the Art

Special thanks to

- Beatriz Goncalves Crisostomo Esteves, UGent
- Michiel Fierens, KULeuven
- Esther De Loof, UGent
- Danica D'hont, UGent

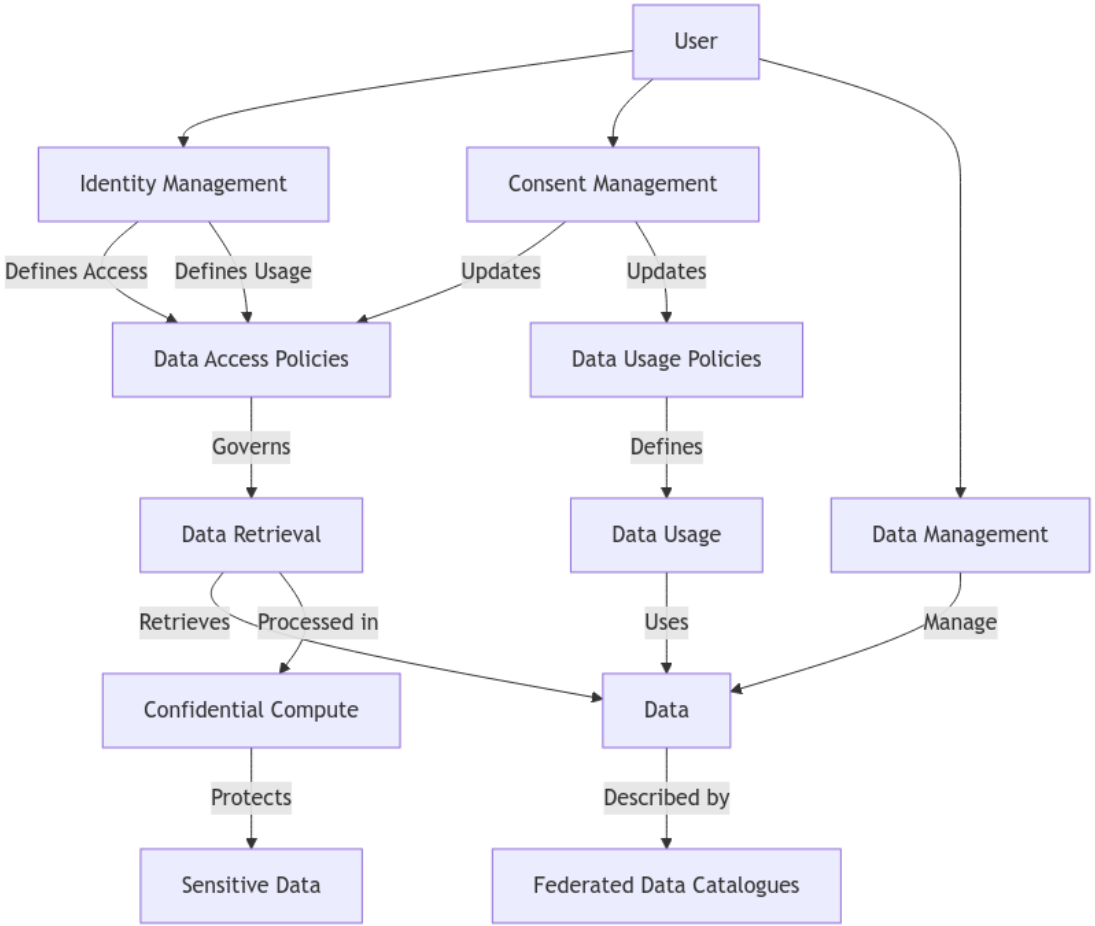
imec

Data Act requirements related to data sharing to personal data in connected products

Data act requirement categories	Requirements
Data Accessibility for Users	Direct and Easy Access
	Commonly Used and Machine-Readable Format
	Metadata for Interpretation
	Simple Request Mechanism
	Verification of legitimate users
Information Transparency for Users	Pre-Contractual Information
	Changes to Data Access
User Rights Regarding Data	Right to Use and Share
	Right to Rectification and Erasure
	No Unfair Contractual Terms
Fair and Transparent Data Sharing	Transparency in Data Sharing Arrangements
	Prohibition on Unfair Commercial Use of Data
	Reasonable Compensation
	Procedural not substantive requirement
Interoperability	Open Standards and Common Specifications
	Interoperability of Data Processing Services
	Support for Smart Contracts
Protection of Legitimate Interests	Potected data of the data user
	Development of a competing product
	Security Requirements
	Protection of Trade Secrets

imec

Essential concepts of Personal Data Stores to address the Data Act



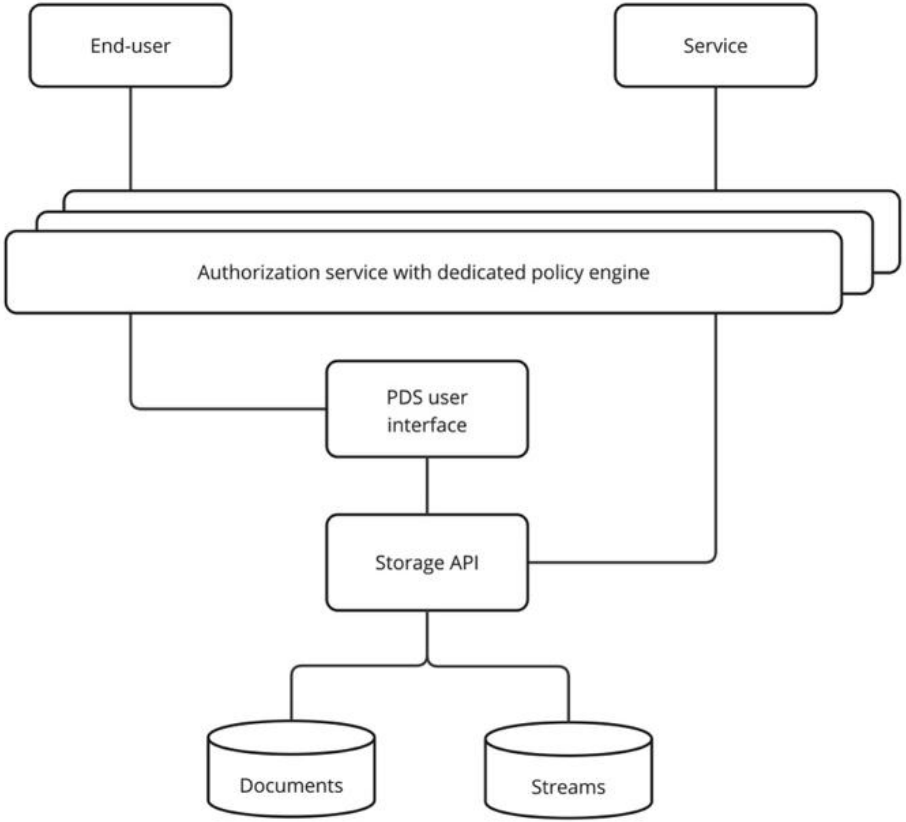
imec

Requirements-technology matrix

Data act requirement categories	Requirements	Identity Management	Consent Management	Data Access Policies	Data Usage Policies	Federated Data Catalogues	Confidential Compute	Linked Data technology
Data Accessibility for Users	Direct and Easy Access	DA1		DA2		DA3		DA4
	Commonly Used and Machine-Readable Format	MR1	MR2	MR3	MR4	MR5		MR6
	Metadata for Interpretation					MD1		MD2
	Simple Request Mechanism	RM1	RM2	RM3	RM4			RM5
	Verification of legitimate users	LU1						
Information Transparency for Users	Pre-Contractual Information			PCI1	PCI2			
	Changes to Data Access			CDA1				
User Rights Regarding Data	Right to Use and Share		RUS1	RUS2	RUS3			
	Right to Rectification and Erasure		RRE1		RRE2			
	No Unfair Contractual Terms				NUC1			
Fair and Transparent Data Sharing	Transparency in Data Sharing Arrangements		TDA1	TDA2	TDA3			TDA4
	Prohibition on Unfair Commercial Use of Data	PUC1			PUC2			
	Reasonable Compensation							
	Procedural not substantive requirement							
Interoperability	Open Standards and Common Specifications	OS1	OS2	OS3	OS4	OS5		OS6
	Interoperability of Data Processing Services							IDS1
	Support for Smart Contracts							
Protection of Legitimate Interests	Potected data of the data user	PD1	PD2	PD3	PD4			
	Development of a competing product							
	Security Requirements	SR1		SR2	SR3		SR4	
	Protection of Trade Secrets						TS1	

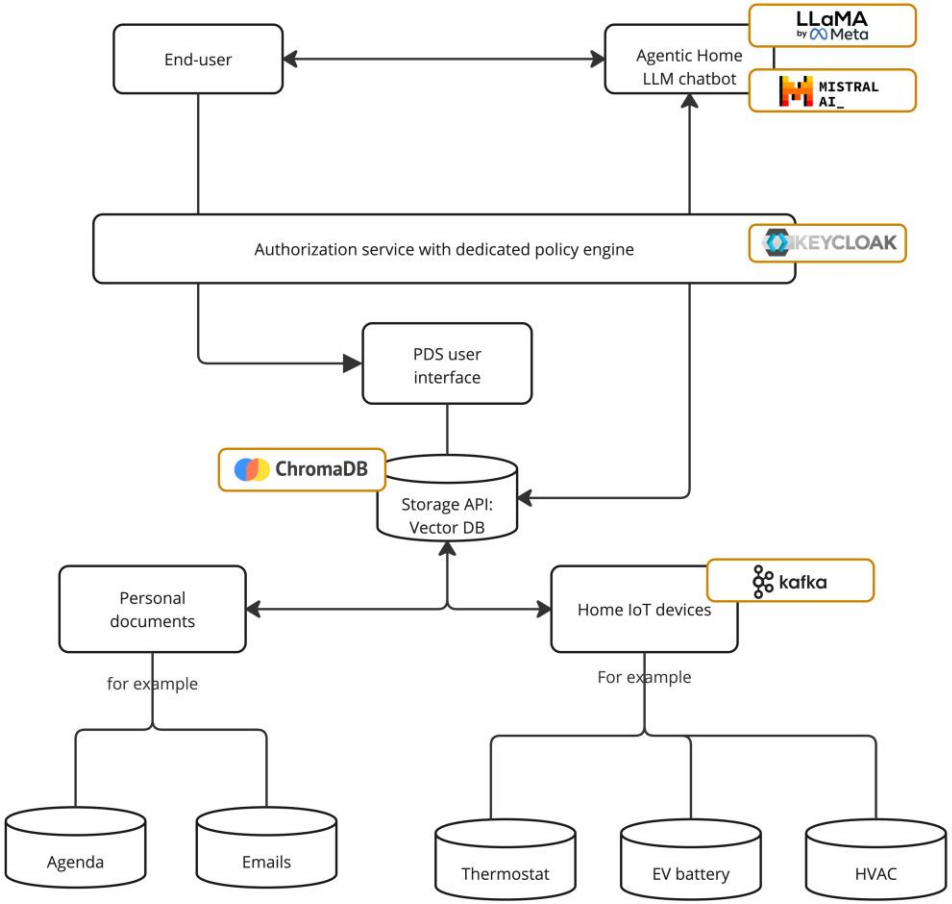
imec

Architecture



Part 2: connecting PDS to (edge) genAI

GenAI agents
RAG
PDS


End-user
Agentic Home LLM chatbot
LLaMA by Meta
MISTRAL AI_
Authorization service with dedicated policy engine
KEYCLOAK
PDS user interface
ChromaDB
Storage API: Vector DB
Personal documents
Home IoT devices
kafka
for example
For example
Agenda
Emails
Thermostat
EV battery
HVAC

Stan

the personal home assistant

Adam Grant

Hi Stan, it's starting to get chilly outside, what's the temperature in our kitchen like?

The current temperature in your kitchen is 20.4 degrees Celsius.

Can I help you with something related to the temperature, such as adjusting the thermostat or finding a cozy spot in the house?

Access to sensor data

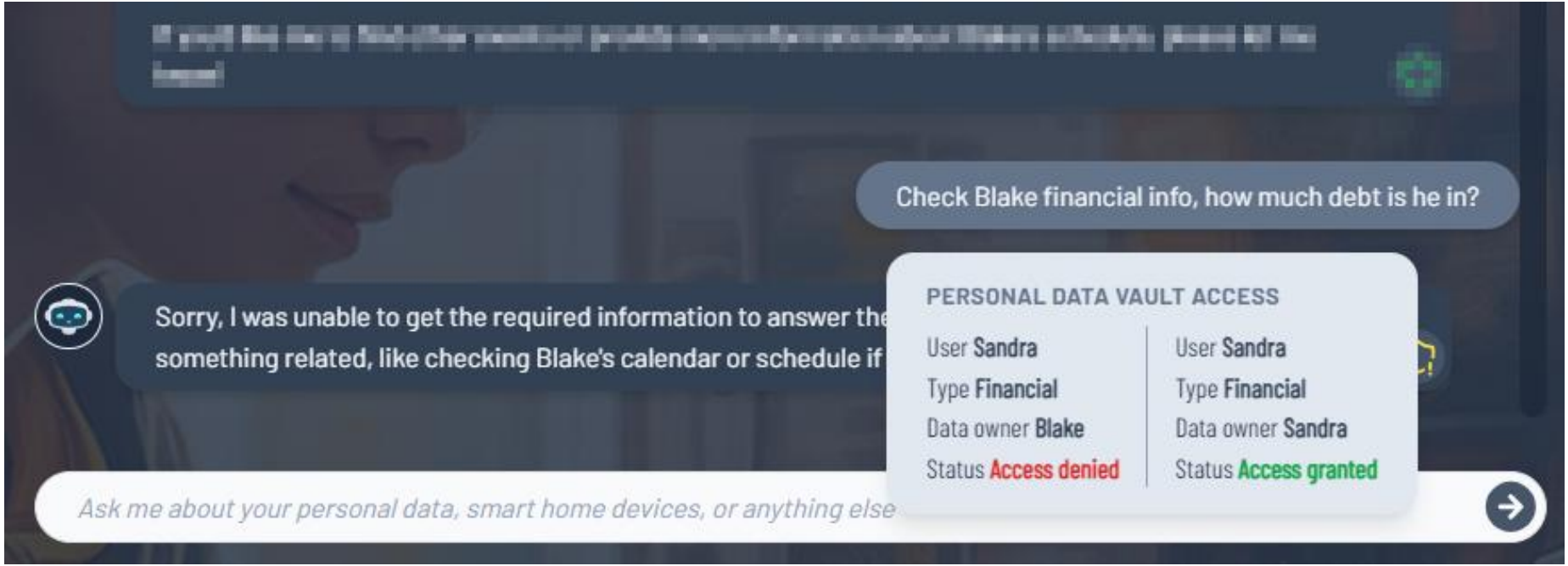
Yes, adjust the thermostat to 22 degrees please

I set the setpoint in the living room to 22 degrees.

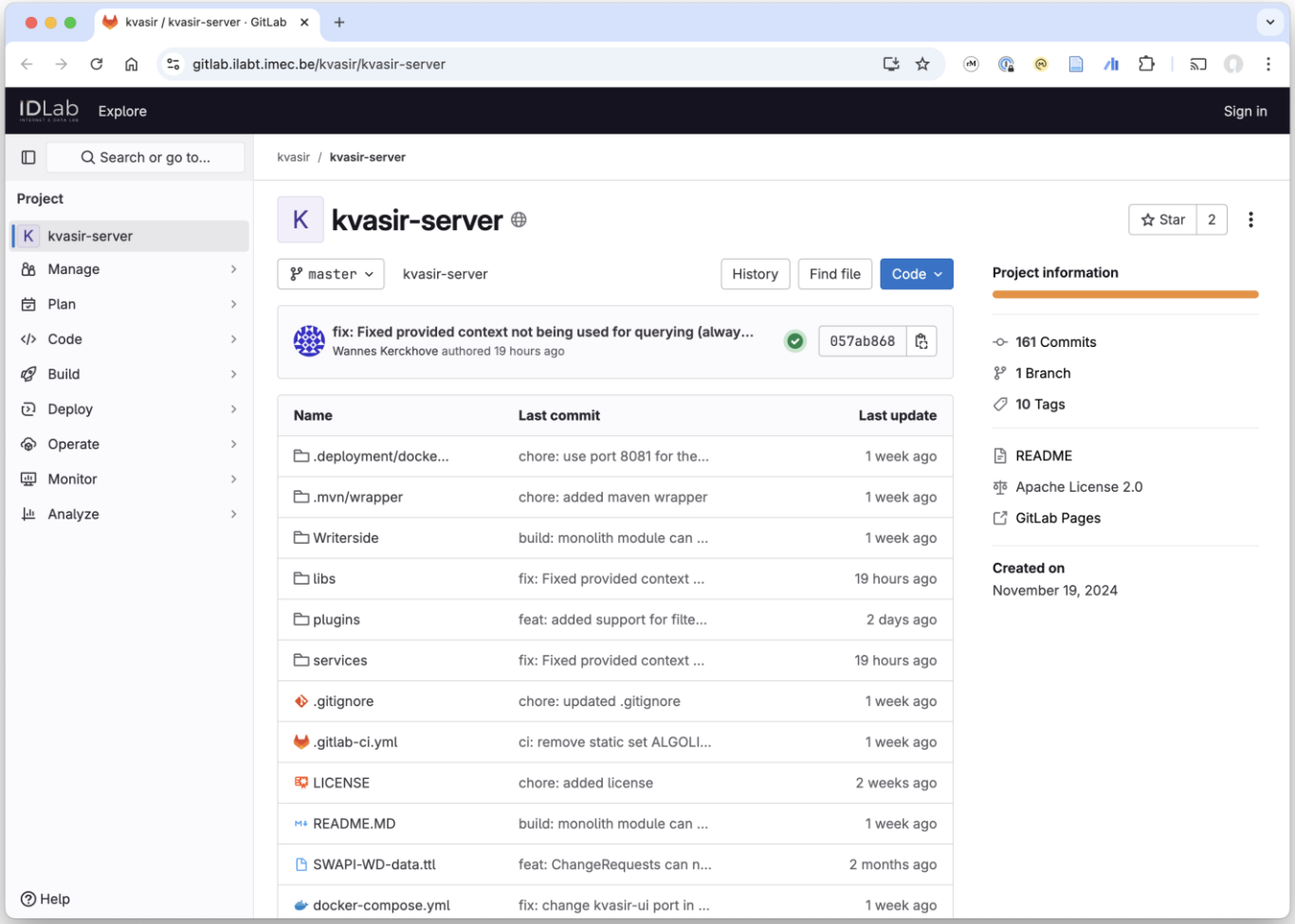
Actuation through method calling

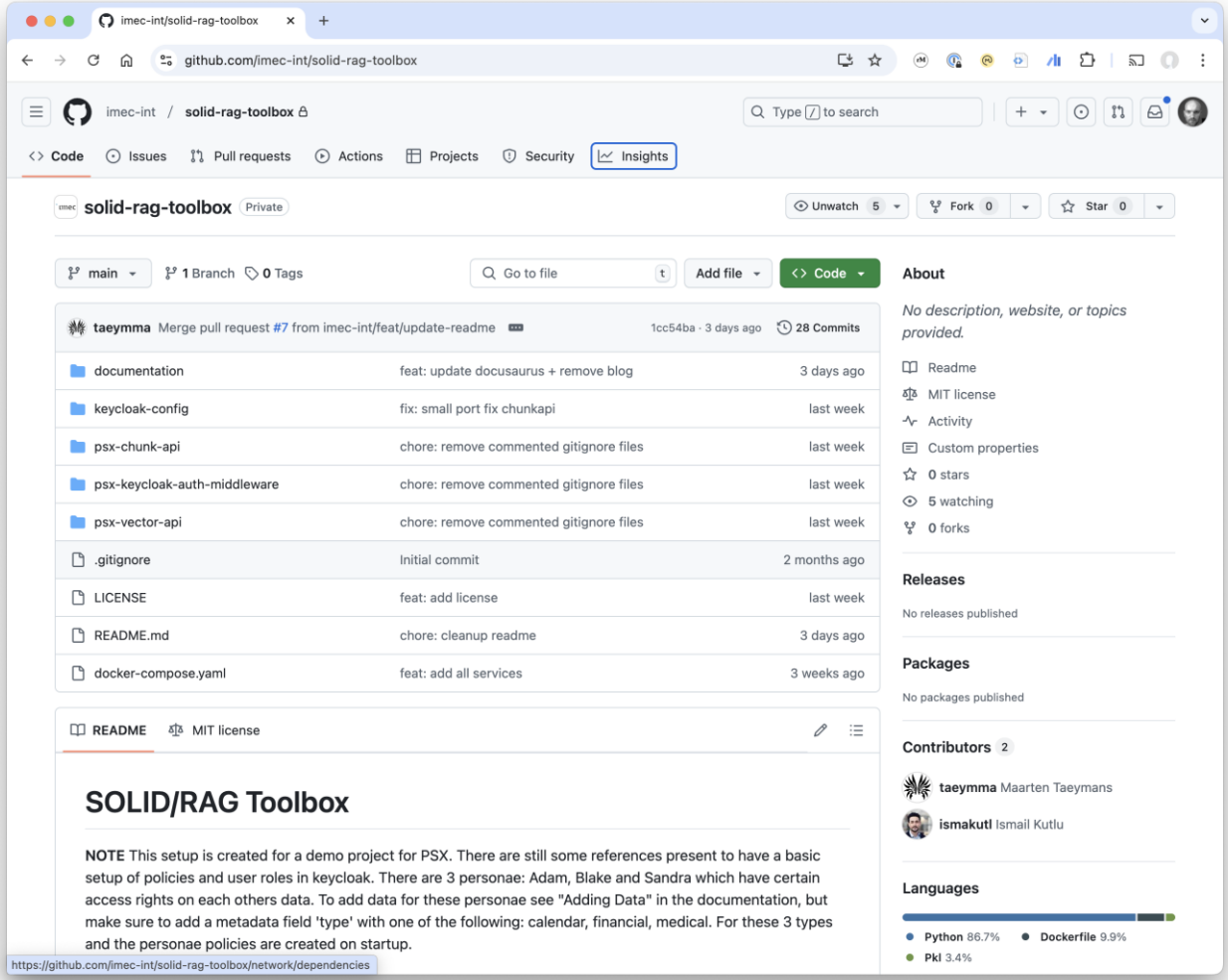
Ask me about your personal data, smart home devices, or anything else

imec Flanders AI Research Program SolidLab Flanders Universiteit Antwerpen IDLab


Check Blake financial info, how much debt is he in?
Sorry, I was unable to get the required information to answer the
something related, like checking Blake's calendar or schedule if
PERSONAL DATA VAULT ACCESS
User Sandra
Type Financial
Data owner Blake
Status Access denied
User Sandra
Type Financial
Data owner Sandra
Status Access granted
Ask me about your personal data, smart home devices, or anything else

imec

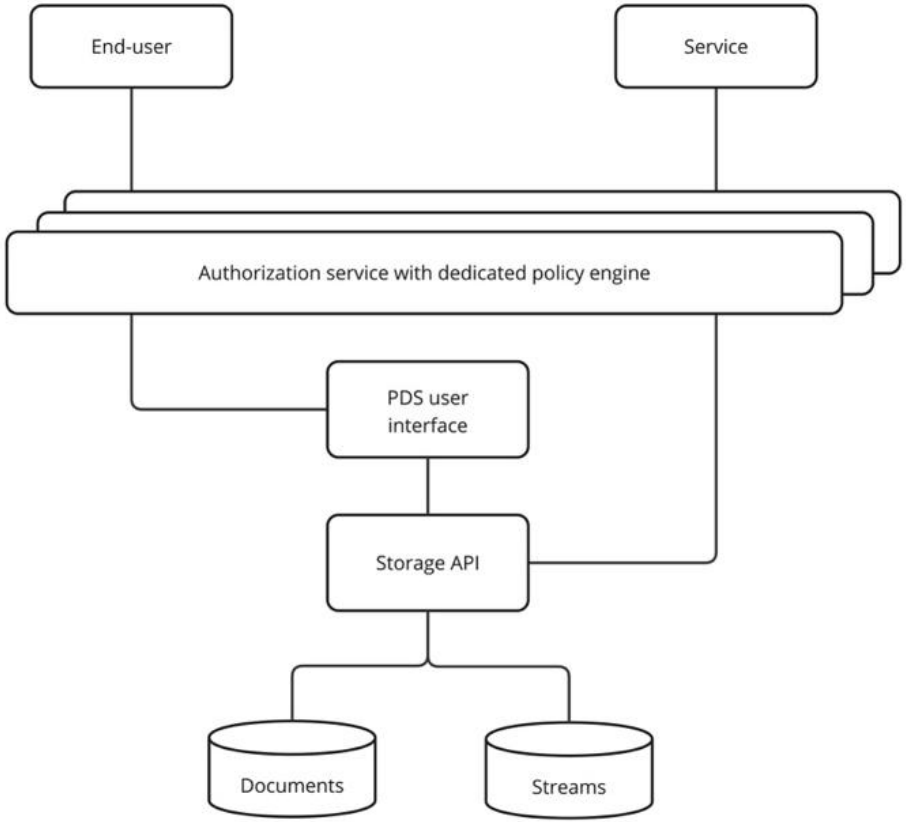

kvasir / kvasir-server · GitLab
gitlab.ilabt.imec.be/kvasir/kvasir-server
IDLab
Explore
Sign in
Search or go to...
kvasir / kvasir-server
Project
kvasir-server
Manage
Plan
Code
Build
Deploy
Operate
Monitor
Analyze
Help
kvasir-server
Star 2
master
kvasir-server
History
Find file
Code
fix: Fixed provided context not being used for querying (alway...
Wannes Kerckhove authored 19 hours ago
057ab868
Name
Last commit
Last update
.deployment/docke...
chore: use port 8081 for the...
1 week ago
.mvn/wrapper
chore: added maven wrapper
1 week ago
Writerside
build: monolith module can ...
1 week ago
libs
fix: Fixed provided context ...
19 hours ago
plugins
feat: added support for filte...
2 days ago
services
fix: Fixed provided context ...
19 hours ago
.gitignore
chore: updated .gitignore
1 week ago
.gitlab-ci.yml
ci: remove static set ALGOLI...
1 week ago
LICENSE
chore: added license
2 weeks ago
README.MD
build: monolith module can ...
1 week ago
SWAPI-WD-data.ttl
feat: ChangeRequests can n...
2 months ago
docker-compose.yml
fix: change kvasir-ui port in ...
1 week ago
Project information
161 Commits
1 Branch
10 Tags
README
Apache License 2.0
GitLab Pages
Created on
November 19, 2024


imec-int/solid-rag-toolbox
github.com/imec-int/solid-rag-toolbox
imec-int / solid-rag-toolbox
Type / to search
Code
Issues
Pull requests
Actions
Projects
Security
Insights
solid-rag-toolbox
Private
Unwatch 5
Fork 0
Star 0
main
1 Branch
0 Tags
Go to file
Add file
Code
taeymma Merge pull request #7 from imec-int/feat/update-readme
1cc54ba · 3 days ago
28 Commits
documentation feat: update docusaurus + remove blog 3 days ago
keycloak-config fix: small port fix chunkapi last week
psx-chunk-api chore: remove commented gitignore files last week
psx-keycloak-auth-middleware chore: remove commented gitignore files last week
psx-vector-api chore: remove commented gitignore files last week
.gitignore Initial commit 2 months ago
LICENSE feat: add license last week
README.md chore: cleanup readme 3 days ago
docker-compose.yaml feat: add all services 3 weeks ago
README
MIT license
SOLID/RAG Toolbox
NOTE This setup is created for a demo project for PSX. There are still some references present to have a basic setup of policies and user roles in keycloak. There are 3 personae: Adam, Blake and Sandra which have certain access rights on each others data. To add data for these personae see "Adding Data" in the documentation, but make sure to add a metadata field 'type' with one of the following: calendar, financial, medical. For these 3 types and the personae policies are created on startup.
About
No description, website, or topics provided.
Readme
MIT license
Activity
Custom properties
0 stars
5 watching
0 forks
Releases
No releases published
Packages
No packages published
Contributors 2
taeymma Maarten Taeymans
ismakutl Ismail Kutlu
Languages
Python 86.7%
Dockerfile 9.9%
Pkl 3.4%
https://github.com/imec-int/solid-rag-toolbox/network/dependencies

Yes, but…

… how is this compatible with a data space?

Data space compatibility



Connector control plane

Connector data plane

Catalogues

Contract/policy negotiation

Transaction logging / payment

imec

Our 2025 ambition is to build an integrated open source personal data store stack that is compliant by design, AI ready and data space compatible.

Reach out if you want to connect!

tanguy.coenen@imec.be

imec
embracing a better life