

A hand holding a pen is pointing at a document with a grid pattern. The background is a dark red color with a subtle grid pattern.

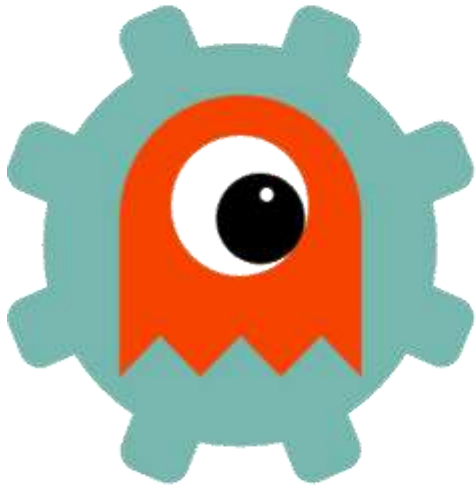
A DATA PRO

WE MAKE SENSE OF INFORMATION

AI-driven web content
enrichment

www.adata.pro

Project CASPAR



CASPAR2

The Combined Automated Semantic
Processing Array



This project has received funding from the Eurostars-2 joint programme with co-funding from the European Union Horizon 2020 research and innovation programme



Ίδρυμα
Πρώτησης
Έρευνας



CASPAR PLATFORM

COMBINED AUTOMATED SEMANTIC PROCESSING ARRAY

CASPAR is a platform utilizing the potential of Artificial Neural Networks (ANNs) for text and image analytics, which will be available to non-tech users. We will offer manually trained ANNs capable of processing content in multiple domains and not requiring structured data or ontologies. The SaaS solution will uniquely combine text and image content analytics technologies, via ANNs and other semantic methods.

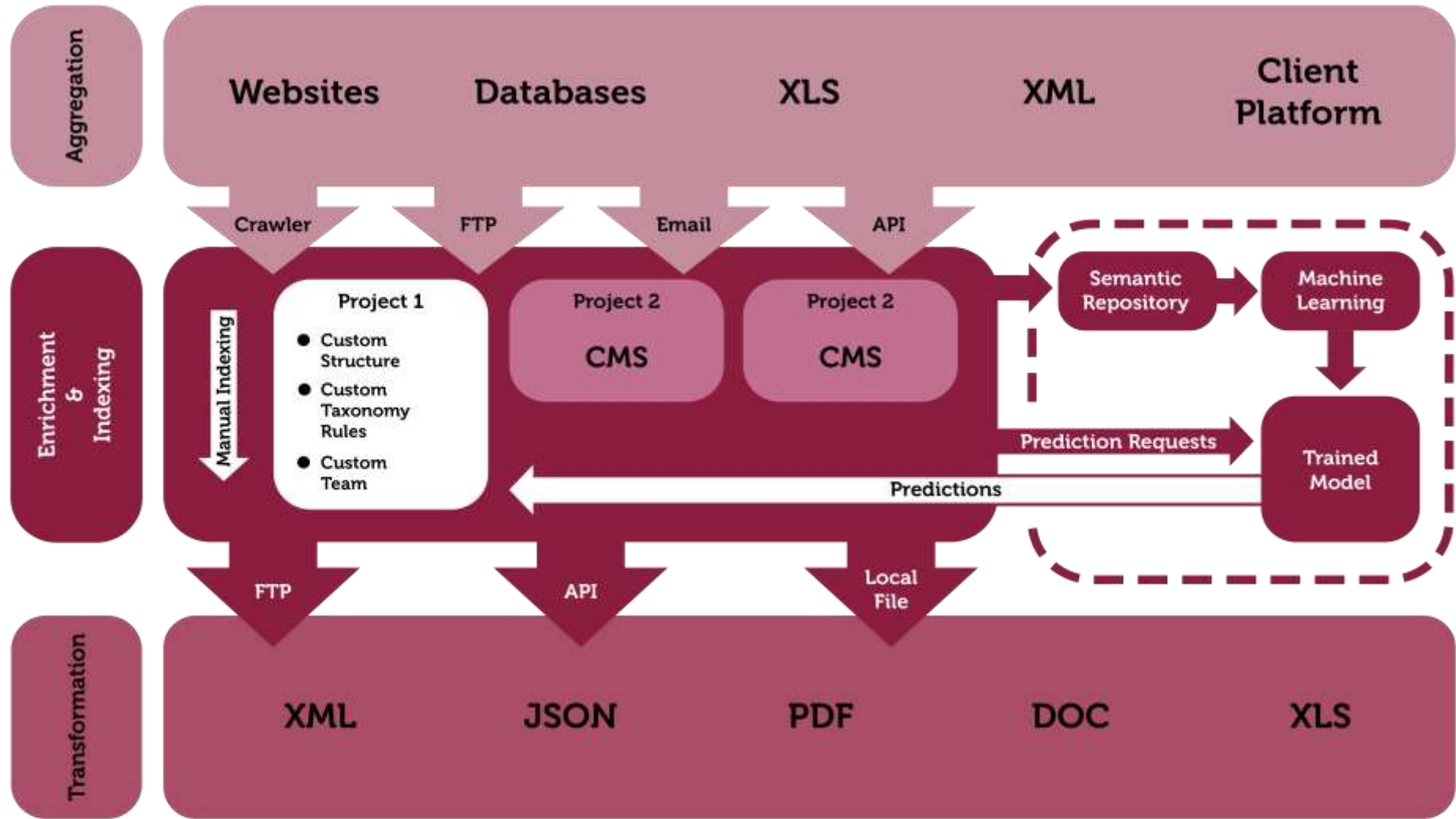
AUTOMATION

IT skills

- Elasticsearch/Kibana
- XML/JSON
- Python/TensorFlow
- JAVA/Weka/DL4J
- PostgreSQL
- Rest API
- ANTLR
- XPATH/Regex
- RDF – OWL

What we do

1. Data aggregation
2. Metadata enrichment
3. Data transformation



DATA AGGREGATION

Crawler architecture

- Source management
- Crawler training
- Content selection & extraction
- Alerting system
- Search indexing
- Publishing & access



DATA AGGREGATION

Supported sources

- Online media
- Government websites
- Company websites
- Forums
- Job advertising websites
- Blogs
- Facebook
- Twitter

Highly specialized sources

- Stock exchange listings
- Financial information
- Shareholder reports
- Sanction lists

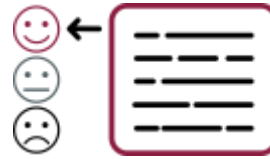
TAILOR-MADE TEXT MINING

- **Classification**
- **Named entity recognition**
- **Document similarity & clustering**
- **Use case: Media listening services**
- **Use case: Knowledge extraction/ representation solution**
- **Use case: Smart data management system**

CLASSIFIER

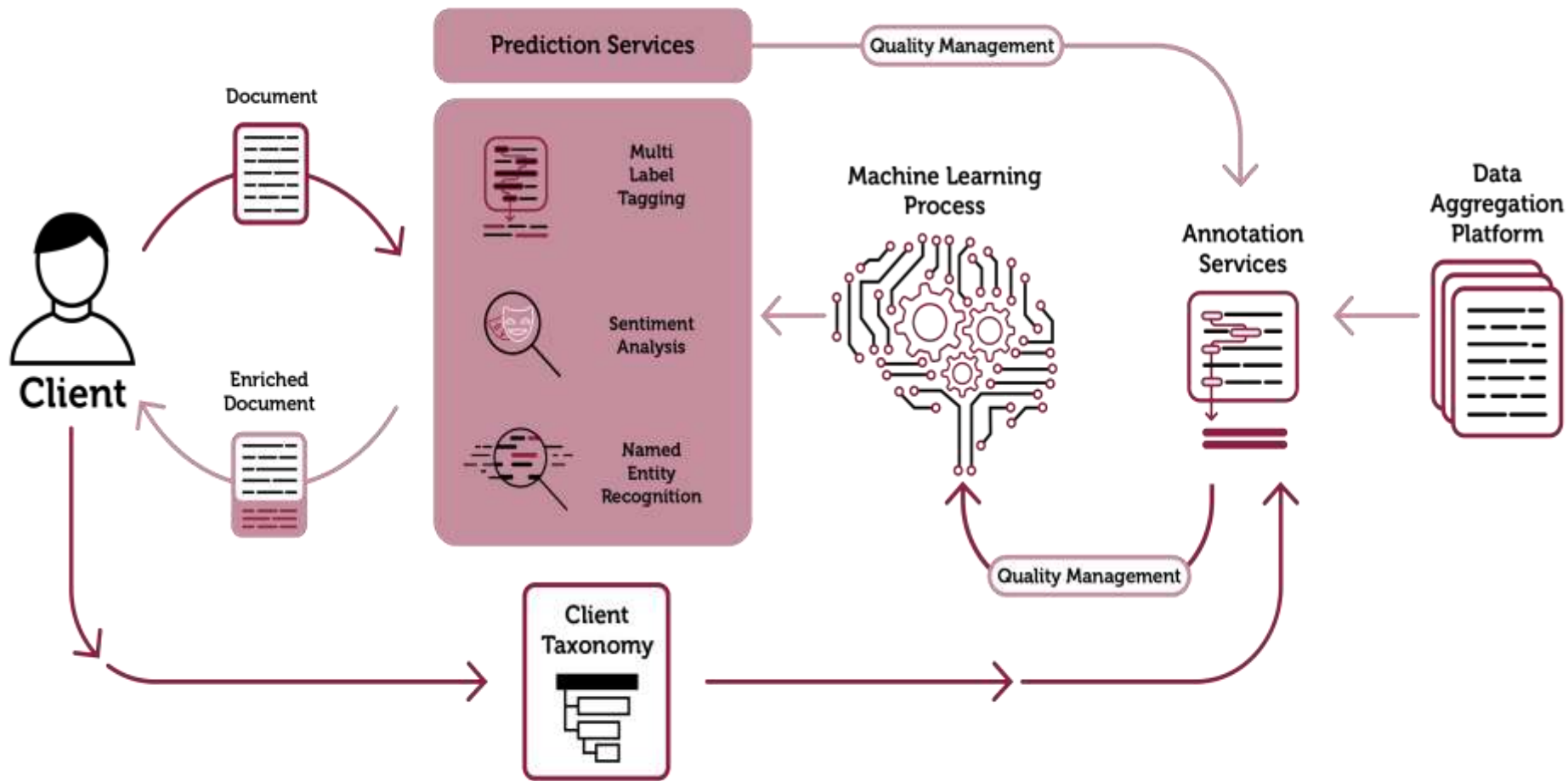


TAXONOMY
CLASSIFIER



SENTIMENT
CLASSIFIER

- **Context-aware** and **domain-specific** classifiers
- **Higher quality**, usually achievable through custom model training
- **Speed, scale** and **efficiency**
- **Tailor-made process**, including trainings, support and consulting



TAXONOMY CLASSIFIER

- **Custom taxonomies** are supported
- **Multi-label tagging technology**, capable of evaluating articles and tagging them to a set of categories
- A **predefined taxonomy** and **some examples** where the taxonomy has already been applied is **all our developers need**

SENTIMENT CLASSIFIER



GENERAL
SENTIMENT



SENTIMENT
PER
ENTITY



SENTIMENT
PER
TOPIC

- **Effective alert system** for PR, communications and risk & compliance professionals

NAMED-ENTITY RECOGNITION

A DATA PRO

live demo

This demo is powered by **Identric**, the innovation hub of A DATA PRO.

But Mr Trump said his former lawyer provided no evidence about alleged collusion between the Trump campaign with Russia during the 2016 US presidential elections. "He didn't lie about one thing. He said, no collusion with the Russian hoax. And I said, I wonder why he didn't lie about that, too, like everything else," Mr Trump said. Michael Cohen, 52, will start a three-year prison term in May for the campaign finance violation of paying hush money to one of Mr Trump's alleged mistresses, tax evasion and lying to Congress.

723 chars left



ANALYZE

But **Mr Trump** said his former lawyer provided no evidence about alleged collusion between the **Trump** campaign with **Russia** during the **2016** US presidential elections. "He didn't lie about one thing. He said, no collusion with the Russian hoax. And I said, I wonder why he didn't lie about that, too, like everything else," Mr **Trump** said. **Michael Cohen**, 52, will start a three-year prison term in **May** for the campaign finance violation of paying hush money to one of Mr **Trump**'s alleged mistresses, tax evasion and lying to **Congress**.

Entity Recognition

person

location

date

organization

NAMED-ENTITY RECOGNITION



PER BRAND
OR TRADEMARK

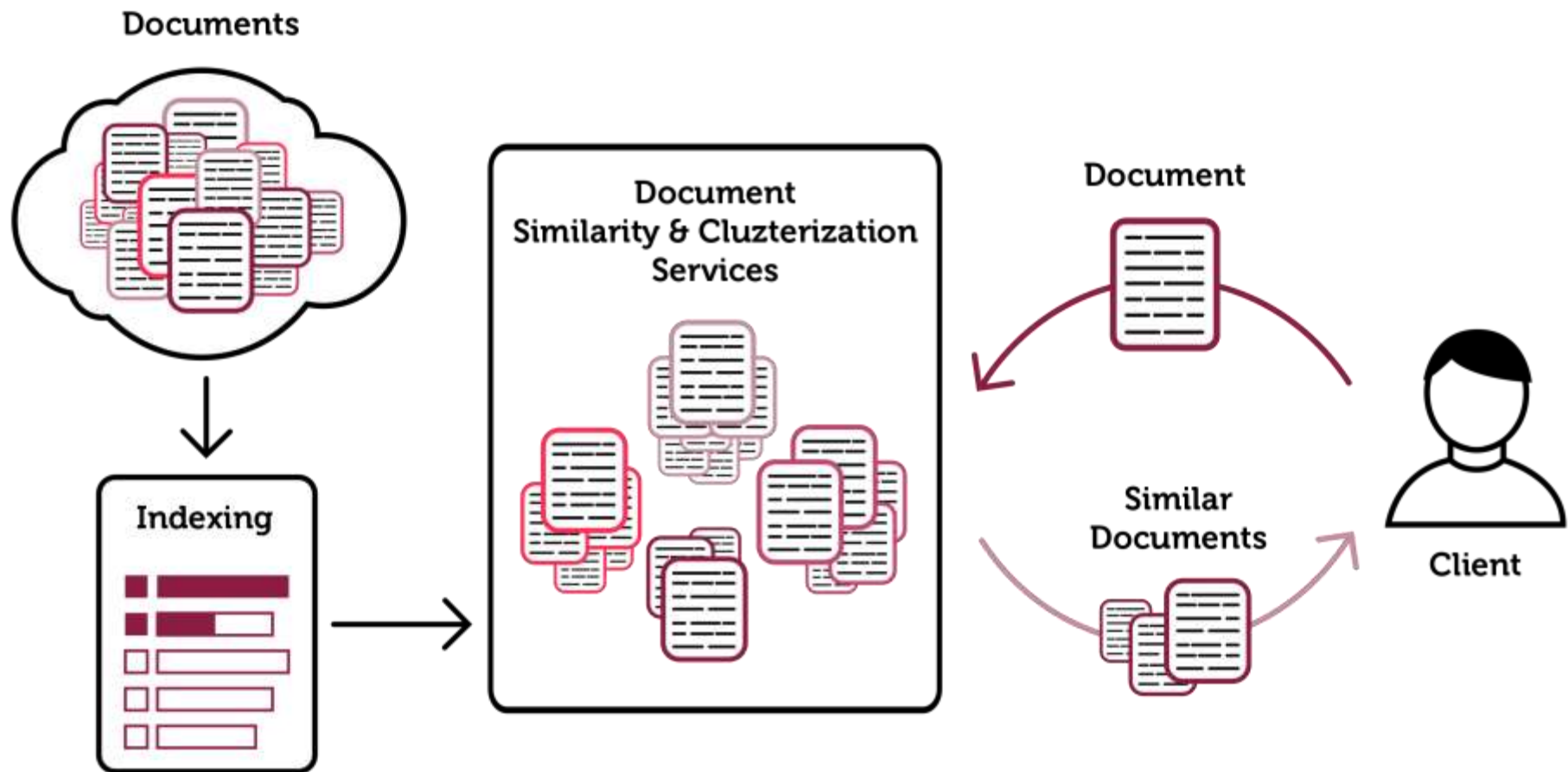


PER
PRODUCT

- **The custom model** can learn fast to identify all textual mentions it has been trained to recognize
- **Less common languages** are supported
- **Integrated** with other services

DOCUMENT SIMILARITY & CLUSTERING

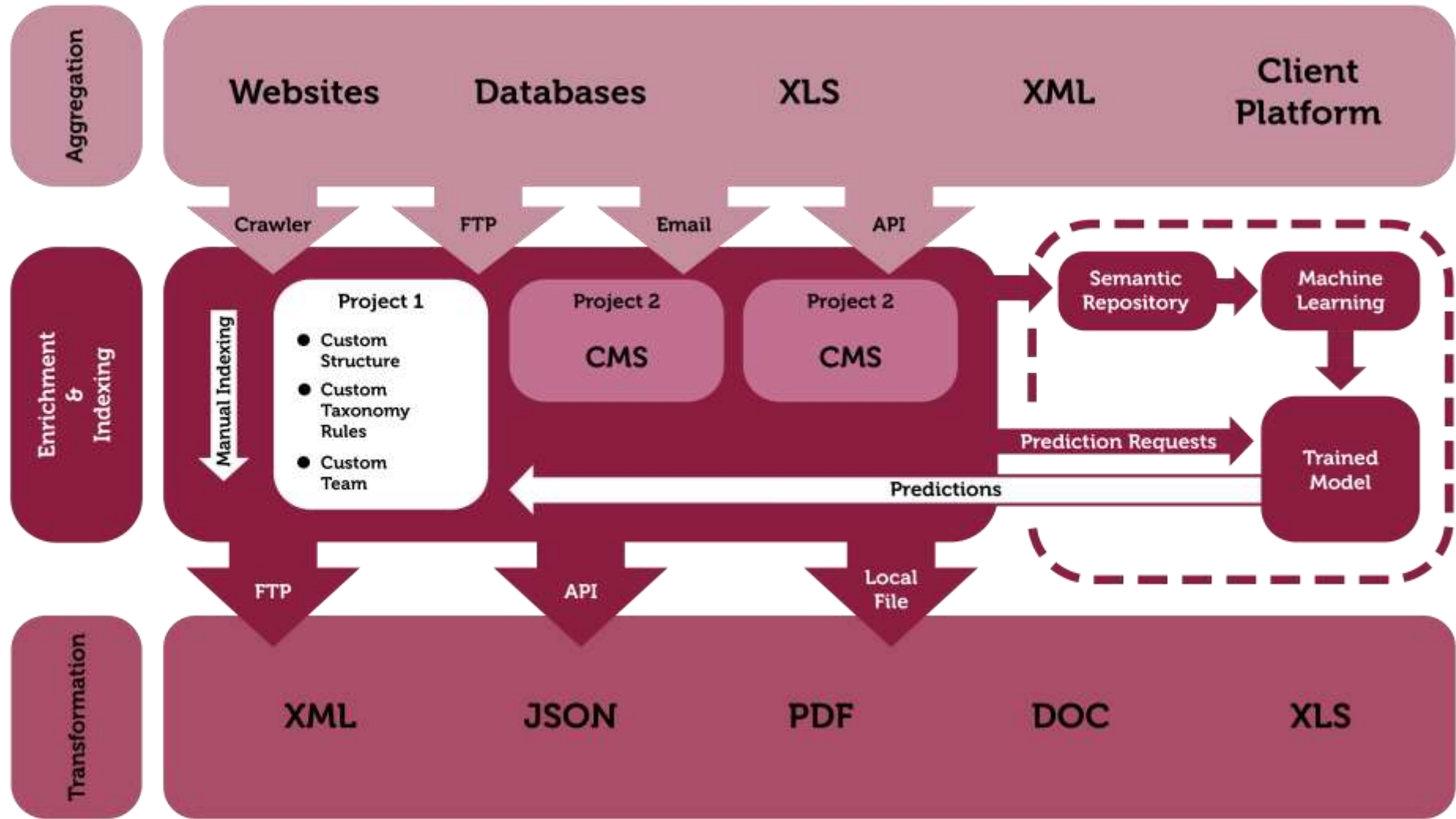
- **Finds and groups similar documents** even in a sea of data
- **30+ languages** are supported
- **No training** necessary



INFRASTRUCTURE

- More than 400 CPUs
- More than 2TB RAM
- 10TB Storage
 - read speed 5 GB/s
 - write speed 2.7 GB/s
- 3 locations
- 1 Gbps links
- Tesla GPUs





Links

<http://e-caspar.com>

<http://www.adata.pro>

<http://www.identrics.net>

A hand holding a pen pointing at a document with a grid pattern. The background is a dark red color with a subtle grid pattern.

A DATA PRO

WE MAKE SENSE OF INFORMATION

THANK YOU

www.adata.pro