Webinar
Annotate data in the EUDAT CDI

Yann Le Franc - e-Science Data Factory, Paris, France

March 16, 2017

Version **2017-1**

EUDAT receives funding from the European Union's Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065

About esciencedata FACTORY

- **Helping scientists to generate FAIR data**
  - Consulting (Data Management Planning, …)
  - Custom software development:
    - Creating user friendly data management tools for scientists
    - Integrating semantic web and Linked Data in scientific tools
  - Knowledge modeling
  - Research & Innovation (validated by French Ministry of Research)
  - Data curation and publication

Interested in working with us?
contact@esciencefactory.com

Outline of the webinar

- Short introduction about annotations - Q&A
- Demo session - Q&A
- Open discussion - Q&A
- Conclusions

What do we mean by annotation?

- By definition, an annotation is “a note added to a text, book, drawing, etc., as a comment or an explanation” (from Merriam Webster).
- In our context, it is an assertion we want to make about a digital resource i.e. a text file, an image, a recording, a movie,... .

The added-value of annotations

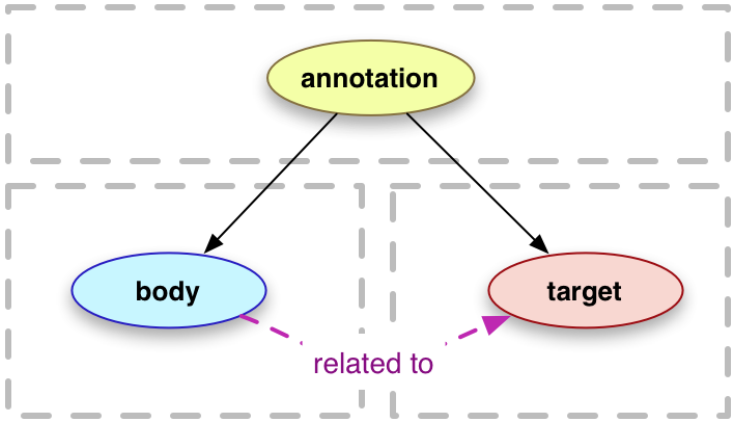
- Enriching digital content with your personal keyword without modifying the data record
- Structure data differently using annotations
- Support data curation before and after publication
- Create aggregated datasets from multi-scale or multi-domain datasets.

B2NOTE: the data annotation service

- Pilot version released: http://b2note.bsc.es
- Three main types of annotations:
  - **Semantic Annotation** of the data in the EUDAT CDI with domain specific ontologies
  - **Free-text keywords**
  - **Comments**
- Based on the W3C Web Annotation Data Model
- Using JSON-LD/RDF format
- Integrated with B2SHAREv2

Web Annotation Data Model

- Use W3C Web Annotation data model – (https://www.w3.org/TR/annotation-model/)



- Serialized in JSON-LD (https://www.w3.org/TR/json-ld/) = JSON based representation of RDF graphs

B2NOTE Pilot service

- **Crowdsourcing annotator**
  - All annotation are public
  - Private annotation in the next release
- **Easy-to-use**
  - Auto-completion with terms from domain specific controlled vocabularies
  - Intuitive User Interface
- **Easily create new datasets selected based on annotations**
- **Easy integration approach based Widget/iframe approach**
  - Integrate with EUDAT services (B2SHARE,…)
  - Integrate with community web UI
- **Easy to deploy**
  - Store triples as JSON-LD in MongoDB backend
  - Uses Django as CMS

Outline of the demo

- Creating an annotation
- View and access your annotation
- Edit your annotation
- Searching for annotated datasets
- Export aggregated dataset
- Export annotations

EUDAT

TIME FOR A DEMO

https://b2note.bsc.es

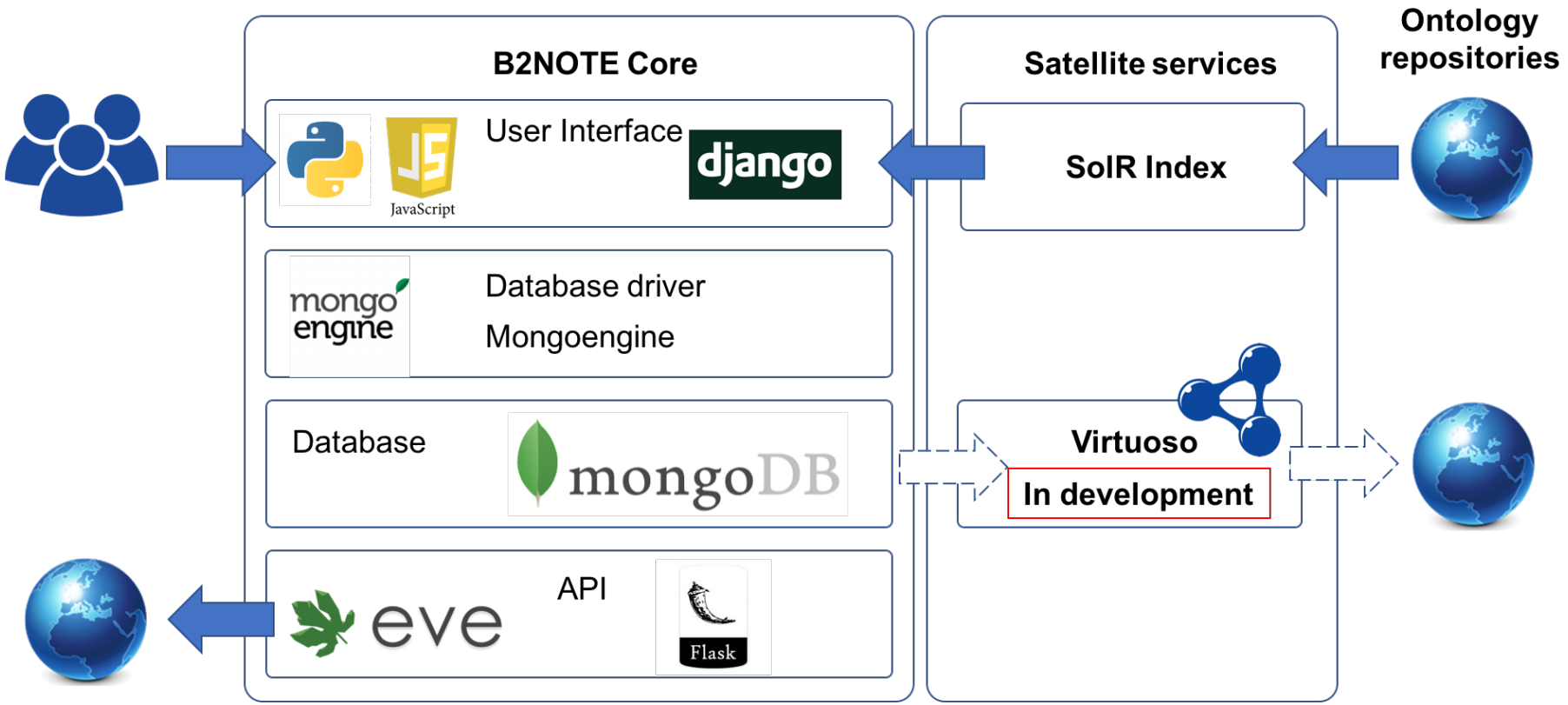
Where to provide feedbacks on B2NOTE?

- Within the service: button "Let us know what you think"
  - Evaluation
  - Request for additional feature
  - Bug report
- By email: ylefranc@esciencefactory.com

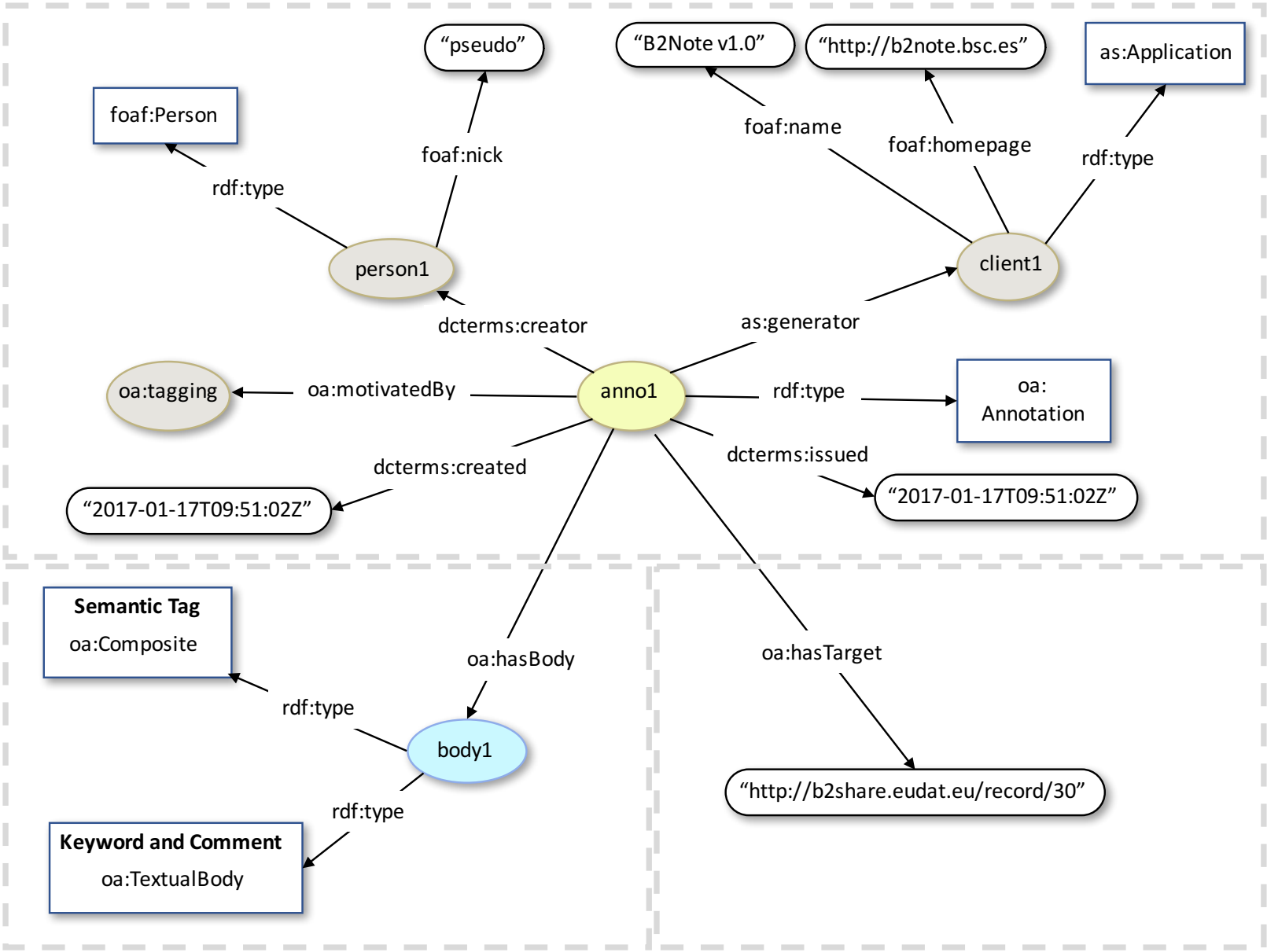
Additional topics to discuss

- Service architecture
- Annotation Data Model
- Using your ontology for annotating files
- Querying annotations as RDF
- API

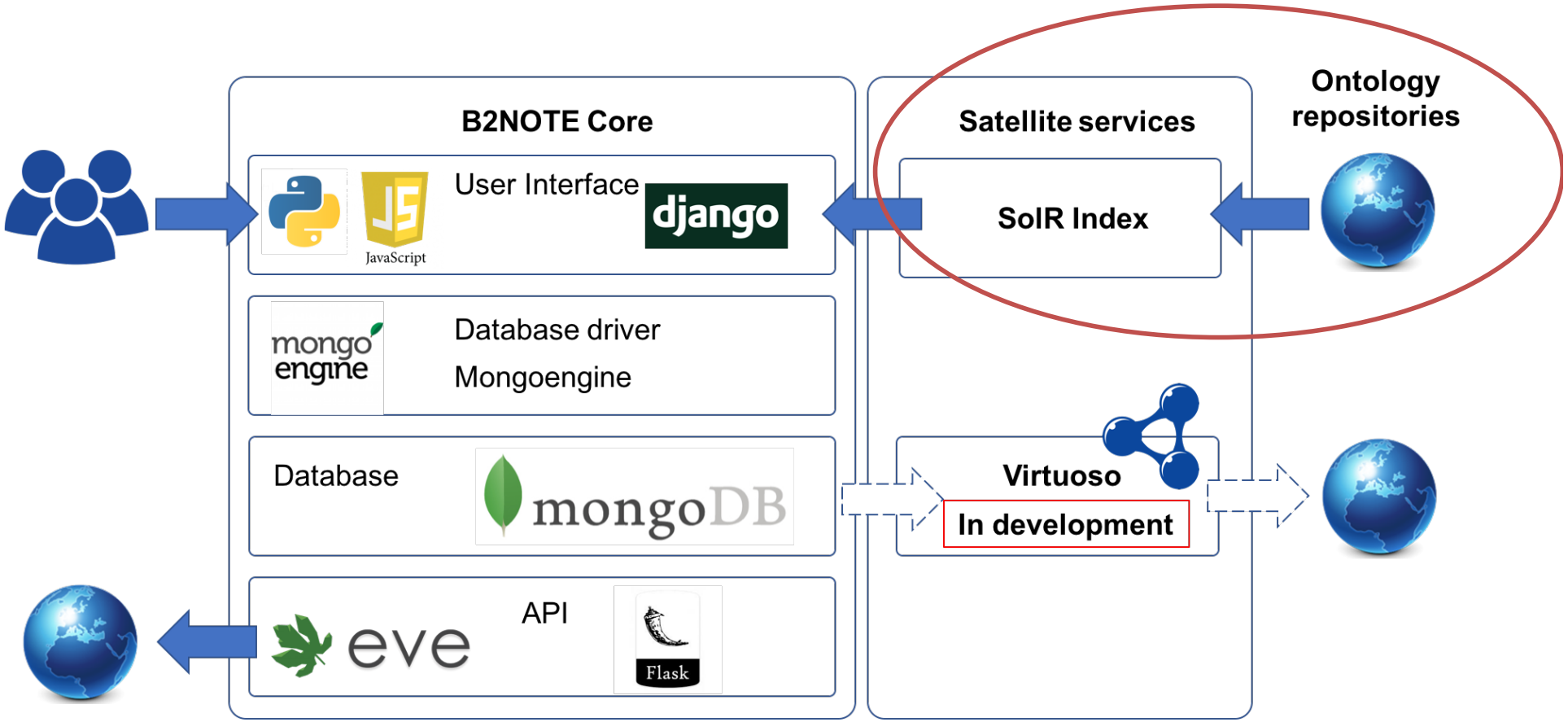
B2NOTE architecture



B2NOTE Annotation Model



Working with ontologies



About the ontology index

- Harvested 1 ontology repository: Bioportal
- 434 ontologies
- More than 5 millions of concepts

→ Problem of interoperability

→ Problem of discoverability

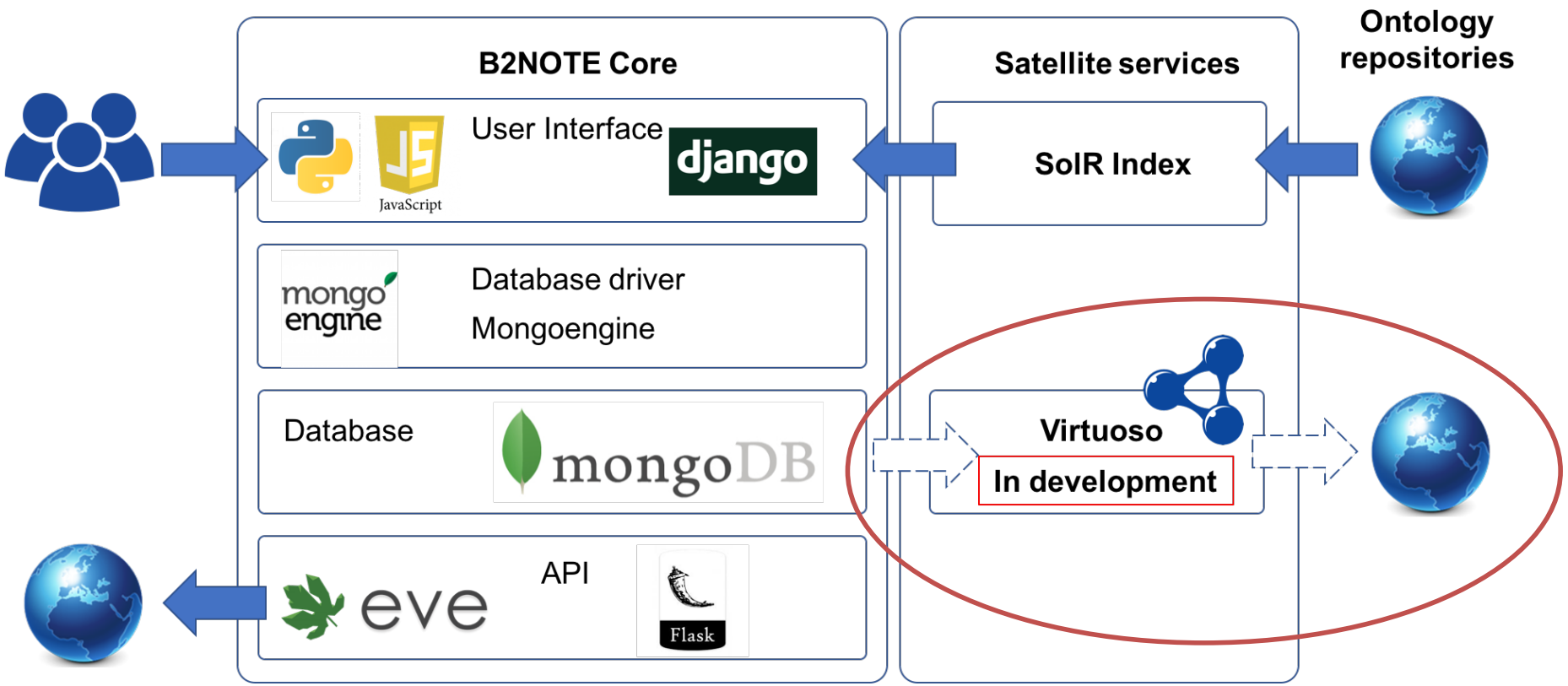
The Ontology Look Up service
Using your own ontology for annotating

- Provide access to multi-disciplinary ontological resources (discoverability)
- Register and describe your endpoint/API for harvesting
- Register and describe your ontology:
  - propose a mapping with internal OLS data model
  - Use B2SHARE to publish your ontology

EUDAT Semantic Working Group Workshop

- Barcelona – April 3-4
- "How to improve the discoverability and the interoperability of multi-disciplinary scientific semantic resources?"

Querying the Annotation graph



Querying the Annotation graph

- Triple Store: OpenLink Virtuoso
- Script converting JSON-LD to RDF
- Pending issues:
  - Configuration of the SPARQL endpoint
  - Design of a workflow to update RDF content with new annotations

B2NOTE API

- Built using the Python REST API framework Eve
- Accessing annotation
  - Accessing all annotations: https://b2note.bsc.es/api/annotations
  - Use filters to access specific annotations
  - Use projections to retrieve specific elements of the annotations.

Future work

- Improvement of the User Interface and User functionalities
- Using W3C DCAT model to structure the aggregated datasets
- Improvement auto-complete function
- Integration with other EUDAT services
- Development of production-ready service

Thanks

- Antoine Brémaud, PhD (e-Science Data Factory)
- Pablo Rodenas (Barcelona Supercomputing Center)

Contact Info

- B2NOTE: Yann Le Franc, PhD : ylefranc@esciencefactory.com
- e-Science Data Factory: contact@esciencefactory.com

Q&A
and
Concluding Remarks

EUDAT & RDM Summer School

3-7 July 2017, FORTH, Heraklion, Crete, Greece

eudat.eu/eudat-summer-school

EUDAT receives funding from the European Union's Horizon 2020 programme - DG CONNECT e-Infrastructures. Contract No. 654065

EUDAT

EUDAT
SUMMER SCHOOL
3-7 July 2017 @ Heraklion, Greece

What is the Summer School about?

Focused on **Data Management** and using **EUDAT services**, the EUDAT Summer School aims to introduce early-career researchers to the principles and tools needed for careers in **data intensive science and data management**.

The course will provide attendees with a better understanding of the European e-Infrastructure landscape, the different tools and services offered by them, and how they can be used to improve the quality of your research outputs.

Who should apply?

Early-career researchers working with big data, as well as **researchers** from less data-intensive communities and **data managers**, interested in furthering their careers in the fields of data management, data science or digital preservation.

EUDAT

EUDAT
SUMMER SCHOOL
3-7 July 2017 @ Heraklion, Greece

What is the goal?

Attendees will understand how the international e-infrastructures, which originate in different fields of research, are building blocks to allow a more integrated solution to meet their needs; they are expected to actively explore data services guided by our experts.

The topics covered by the Summer School are:

- The Research Data Lifecycle
- The FAIR Data Concept
- Writing a Data Management Plan
- The EUDAT Service Suite Overview
- High Performance Computing (HPC) Programming Models
- Using the EGI Federated Cloud for Data Analysis
- Linking HPC to Data Management
- Open Data and Cross-disciplinary Research
- Long Term Data Curation

EUDAT

How to apply?

Visit

eudat.eu/eudat-summer-school

for criteria and financial support opportunities

When is the deadline for applying?

Monday 17 April 2017 @ 23:59 CET