



# Data Exchange Solution between PubAnnotation and AlvisAE

Mouhamadou Ba, Robert Bossy and Claire  
Nédellec

*firstname.lastname@inra.fr*



**Bibliome Team from the  
Research Unit 'Applied  
Mathematics and Computer  
Science, from Genomes to  
the Environment' in France**



# Who we are ?

**Bibliome** is NLP research team

- at INRA (National Institute of Agricultural Research)
- as part of Research Unit MaAGE (Mathematics and Computer Science, from Genomes to the Environment)

The team is specialized

- in NLP applied to biology, domain-specific Knowledge Acquisition
- with area of interest in agriculture, food, biology (animal, plants, microbes ...) ecology

**Multidisciplinary skills within the team**

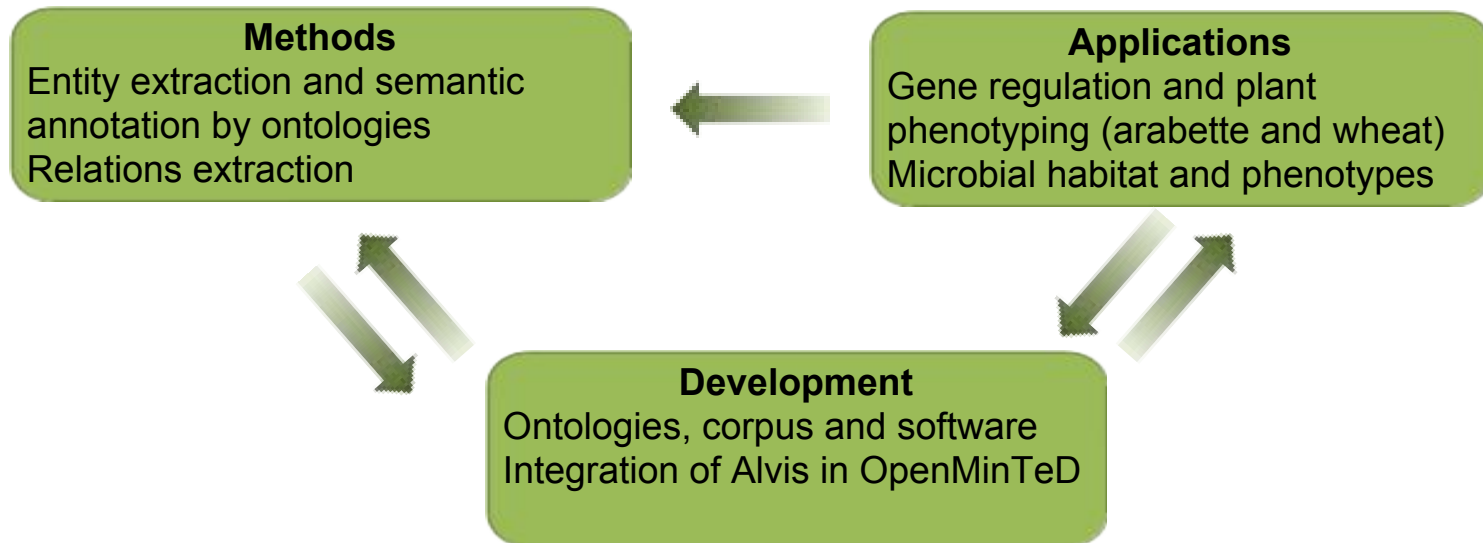
- Biology
- Bioinformatics
- Knowledge Engineering
- Data and Software Engineering
- Machine Learning
- Natural Language Processing

# What we do ?

We develop methods for extraction and formalization of knowledge from textual documents in life sciences

Our research and developments

- are applied to biology
- from end-users needs (biologists)
- are as generic and reusable as possible



## Team achievements

- **AlvisAE**: annotation editor
- AlvisNLP/ML: corpus processing engine
- AlvisIR: semantic search engine framework
- TyDi: terminology and ontology editor

→ <https://github.com/Bibliome>

- We organize BioNLP Shared Tasks (BB and SeeDev tasks)
- We are part of european project OpenMinTeD (text-mining e-infrastructure)

an online annotation editor to perform annotation of textual documents for

annotation : [10645449] Genotypic

Habitat

Genotypic characterization of drug-resistant **Mycobacterium tuberculosis** isolates from **Peru**.

Twenty-nine epidemiological unrelated and mostly multidrug-resistant **Mycobacterium tuberculosis** (**MDR-TB**) strains from **Peruvian patients**

To investigate the molecular genetics of **MDR-TB** strains recovered in a Latin American country.

Antimicrobial agent susceptibility testing, major genetic group designation, IS6110 fingerprinting, spoligotyping, and automated deoxyribonucleic acid sequencing of regions of the katG, rpoB, embB, gyrA, and pncA genes with mutations commonly associated with drug resistance.

Nineteen isolates were found to be multidrug-resistant by susceptibility testing. IS6110 typing showed that virtually

Annotations Text selection

Id	Annotation	KI	Type	Details	Vis
4e02e	[imported] imported from annotation by estelle in campaign 35		Geographica	Peru	
7600a	[imported] imported from annotation by estelle in campaign 35		Habitat	Peruvian patients	

11

ay

ain

on

n for

# AlvisAE

an online annotation editor to perform annotation of textual documents for knowledge extraction

used to organize annotation campaigns with several users in a collaborative way (e.g., used to produce datasets of BioNLP-ST challenges (BB, SeeDev))

## Functionalities

- visualize and annotate entities and relations of textual documents using domain ontologies
- manage the lifecycle of the annotation campaigns (create campaign, add documents, assign users, define a workflow of tasks, do annotation and adjudication)
- Do linguistic processing via AlvisNLP/ML (for example, automatic annotation using NLP/ML modules)

*Papazian, F., Bossy, R., & Nédellec, C. (2012, July). AlvisAE: a collaborative Web text annotation editor for knowledge acquisition. In Proceedings of the Sixth Linguistic Annotation Workshop (pp. 149-152). Association for Computational Linguistics.*

# AlvisAE architecture

Client server platform composed of two main components

## Web Service

- Implement REST operations
- Manage storage/access to the contents (storage of documents and annotations, authentication and authorization of the annotators, workflow enforcement)

## Web Application Client

- Implement the editors (offers web interfaces for users to login, access and visualize documents, edit annotations, etc.)
- Exchange format with the Web Service component is based on JSON

## Goal during BLAH4

Make **AlvisAE** and **PubAnnotation** interoperable by implementing a **data-exchange** solution

Enable users to share annotations from AlvisAE and to access documents from PubAnnotation

Example of usage

- get documents from PubAnnotation to AlvisAE,
- do annotation tasks,
- get the annotated documents back to PubAnnotation

Facilitate the creation and sharing of annotations produced with domain experts



# Plan and Tasks

1. Ensure conversion of contents between PubAnnotation and AlvisAE
  - Define the mappings to/from the JSON format of PubAnnotation
  - Implement a converter
  
2. Define an exchange policy/protocole
  - Define the characteristics of the import/export operations
  - Implement the operations as REST calls (based on existing work)
  
3. Add a trigger interface
  - Define the trigger interface
  - add an interface to AlvisAE

# Materials

## Source code

- AlvisAE is available on github  
<https://github.com/openminted/alvisae>
- the web service component is based on framework lift (old version 2.8, written in scala, <https://liftweb.net/>)
- the web application component is based on GWT (version 2)
- Alv2

## Working deployment

- AlvisAE may be deployed using docker
- <https://github.com/openminted/alvisae/README>



# Thank you !