DATA

Concerning bioconcepts related to plant health, two extensive literature reviews were considered to retrieve text section for this Challenge.

As training data for this Challenge, mentions of specific bioconcepts within 220 text sections were annotated. Participants are provided the 220 text sections as text files (*plh_trainingset_text.txt*) and JSON formatted annotation files (*plh_trainingset_annotations.json*).

As validation data for this Challenge, mentions of specific bioconcepts within 87 text sections were annotated. Participants are provided the 87 text sections as text files (*plh_validationset_text.txt*) and JSON formatted annotation files (*plh_validationset_annotations.json*).

For the test data set, participants are provided ~128 text sections as text files (*plh_testset_text.txt*) to test their NER approach for identifying bioconcepts related to plant health. Participants must submit an JSON-formatted annotation file (*.json) for each of the text files.

PLH: XYLELLA

An extensive literature search was conducted to review new scientific articles published from 1 November 2013 to 20 November 2015, with particular focus on the new findings of host plants species of the *Xylella fastidiosa*.

Total number of references	Training set	Validation set	Test set
220	112	44	64

PLH: Robinia pest list

An extensive literature research was conducted in a specific database platform, ISI Web of Science.

The search had the objective of retrieving all the available papers concerning the pests and pathogens associated to a specific plant genus (gen. *Robinia*)."

Total number of references	Training set	Training set	Test set
215	108	43	64

BIOCONCEPTS

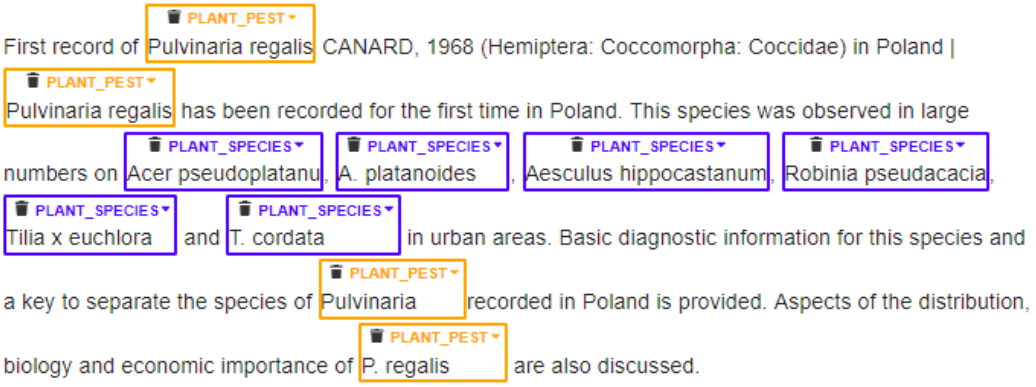
This Challenge is aimed at the annotation of the following the specific types of information, aka Bioconcepts.

Bioconcepts	**Description**
Plant_pest	Any species, strain or biotype of plant, animal or pathogenic agent injurious to plants or plant products. (FAO (Food and Agriculture Organization of the United Nations), 2017d. ISPM (International standards for phytosanitary measures) No. 5. Glossary of phytosanitary terms. 38 pp. Available online: http://www.fao.org/fileadmin/user_upload/faoterm/PDF/ISPM_05_2016_

	En_2017-05-25_PostCPM12_InkAm.pdf)
Plant_species	The susceptible host species (or associated species).
Plant_disease_commname	The pest related disease and/or injury common name

ANNOTATON INSTRUCTIONS

- Annotations are defined as the longest contiguous text that describes the item of interest, including abbreviation definitions.
- Mentions generally do not cross sentence boundaries.
- If an item appears more than once in the text, all instances are annotated, including the use of abbreviations. See example below[i]:



- Concerning Plant_species, both scientific and common names for species are annotated. This includes all species listed within the text, taken separately.

[i] [i] Images illustrating examples were obtained by using LightTag.io annotator tool.