DATA

Concerning bioconcepts related to animal health, an extensive literature review were considered to retrieve text section for this Challenge.

As training data for this Challenge, mentions of specific bioconcepts within 228 text sections were annotated. Participants are provided the 228 text sections as text files (*ahaw_trainingset_text.txt*) and JSON formatted annotation files (*ahaw_trainingset_annotations.json*).

As validation data for this Challenge, mentions of specific bioconcepts within 53 text sections were annotated. Participants are provided the 53 text sections as text files (*ahaw_validationset_text.txt*) and JSON formatted annotation files (*ahaw_validationset_annotations.json*).

For the test data set, participants are provided 71 text sections as text files (*ahaw_testset_text.txt*) to test their NER approach for identifying bioconcepts related to animal health. Participants must submit an JSON-formatted annotation file (*.json) for each of the text files.

AHAW: DACRAH

An extensive literature search was performed to assess the worldwide occurrence of the vector borne diseases (VBDs) and to study the prevalence, incidence or occurrence of pathogens or a previous exposure to pathogens in an area. (DOI: 10.2903/j.efsa.2017.4793)

Total number of references	Training set	Validation set	Test set
352	228	53	71

BIOCONCEPTS

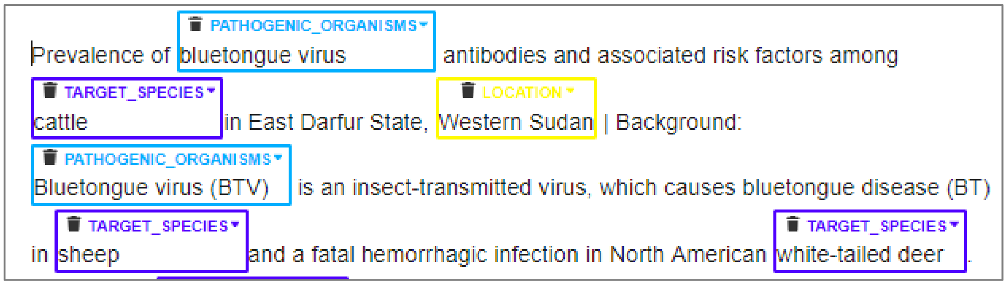
This Challenge is aimed at the annotation of the following the specific types of information, aka Bioconcepts.

Bioconcepts	**Description**
Location	Geographic location where the study was conducted
Pathogenic_organisms	Anything that can produce disease in the target species
AnMethod	Analytical method used to test samples
Prevalence	Calculated prevalence or number of tested samples and number of positive samples
Target_species	The susceptible host species (the species in which the pathogen replicates)
Year	The year during which the study took place or started

ANNOTATON INSTRUCTIONS

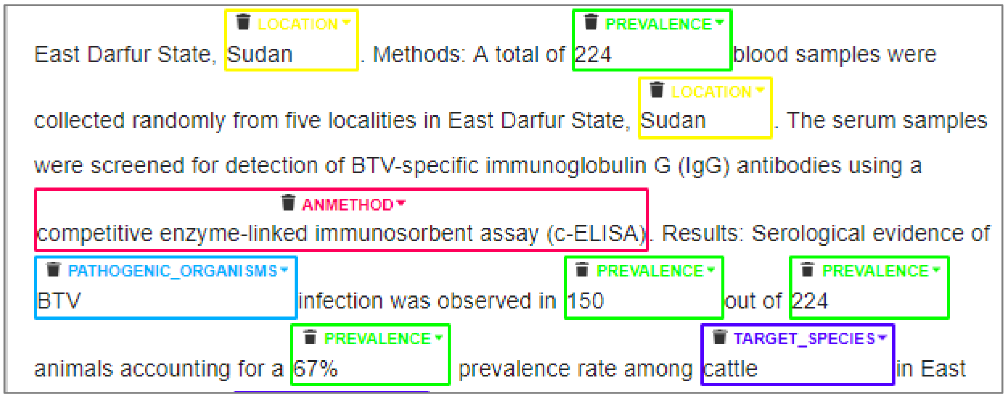
- Annotations are defined as the longest contiguous text that describes the item of interest, including abbreviation definitions.
- Mentions generally do not cross sentence boundaries.

- In the examples below[i], the abbreviation '(BTV)' should be included in the annotation.



Prevalence of bluetongue virus antibodies and associated risk factors among cattle in East Darfur State, Western Sudan | Background: Bluetongue virus (BTV) is an insect-transmitted virus, which causes bluetongue disease (BT) in sheep and a fatal hemorrhagic infection in North American white-tailed deer .

- If an item (often the case for Target_species) appears more than once in the text, all instances are annotated, including the use of abbreviations.
- Concerning Target_species, both scientific and common names for species are annotated. This includes all species listed within the text, including those not used in the experiment itself.
- Concerning Prevalence, all the text that refer and describe element that characterize prevalence are annotated. In the example below, the text referring to prevalence (“67%”) is annotated, together with the text referring to element related to prevalence such as the number of infected “150” and the total number of samples tested “224”.



East Darfur State, Sudan . Methods: A total of 224 blood samples were collected randomly from five localities in East Darfur State, Sudan . The serum samples were screened for detection of BTV-specific immunoglobulin G (IgG) antibodies using a competitive enzyme-linked immunosorbent assay (c-ELISA). Results: Serological evidence of BTV infection was observed in 150 out of 224 animals accounting for a 67% prevalence rate among cattle in East

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