
ANNOTATION GUIDELINE FOR CLIRENER

1 Introduction

This guideline is used for named entity recognition (NER) annotation of scientific texts related to climate change.

Named-entity recognition (NER) (also known as (named) entity identification, entity chunking, and entity extraction) is a subtask of information extraction that seeks to **locate and classify named entities mentioned in unstructured text into pre-defined categories** such as person names (PER), organizations (ORG), locations (LOC), geopolitical entities (GPE), vehicles (VEH), medical codes, time expressions, quantities, monetary values, percentages, etc¹. Most research on NER systems has involved taking an unannotated block of text and highlighting parts of the text that correspond to a certain category:

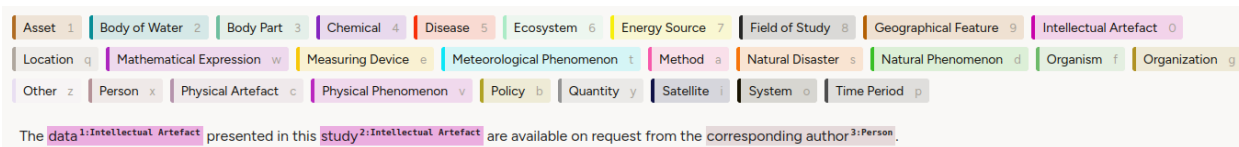


Figure 1: Named entity recognition (NER) example.

In Figure 1, both "data" and "study" are an *Intellectual Artefact* as both are man-made results from scientific (intellectual) work, while "corresponding author" is hopefully a *Person*! In a similar manner, we plan to create a dataset of these annotations for the climate change research domain.

The name of the final product will be the **Climate Research Named Entity Recognition (CliReNER) Dataset**. The CliReNER Dataset consists of **28 named entity categories**: *Asset, Body Part, Body of Water, Chemical, Disease, Ecosystem, Energy Source, Field of Study, Geographical Feature, Intellectual Artefact, Location, Mathematical Expression, Measuring Device, Meteorological Phenomenon, Method, Natural Disaster, Natural Phenomenon, Organism, Organization, Other, Person, Physical Artefact, Physical Phenomenon, Policy or Objective, Quantity, Satellite, System and Time Period*.

2 First Steps

This is a set of steps a **reader/annotator must follow** when using this document:

1. Locate Table 2 (pg. 3) and carefully read the definitions of each entity category, listed in alphabetical order. You will be provided with focused sentences for a subset of all entity categories but are required to **annotate each category that appears in the text**.
2. Revisit the list of assigned entity categories (e.g. Group 4: *Body Part, Chemical, Disease* and *Organism*) and **carefully read each category's detailed directives** (Section 4, pg. 4)².
3. Open the given link (e.g. <http://12.345.678.910:8080/projects/11/data?tab=12>) to access the Label Studio interface. **Familiarise yourself with the interface** by following the guidelines in Section 6 (pg. 33).

¹Definition based on Wikipedia: https://en.wikipedia.org/wiki/Named-entity_recognition.

²Each category's detailed directive page is listed next to the type in Table 2.

4. Before annotation, it is recommended to **read the useful tips on annotation** in Section 7 (pg. 35).
5. Proceed to the sentence example pre-annotated by a predictive model. This model highlights potential entities and assigns the default label - *Other*.
6. Check each pre-annotated example by **answering the following questions**:
 - (a) Does the highlighted textual span fully cover the entity referred to in the sentence?
 - (b) What is the entity category that should be assigned to the textual span?
 - (c) Are there any noun phrases/entities that the model has missed?
7. If the definition is insufficient to answer questions (a) and (b)³, **consult the most likely candidate category and its common ambiguities** (Table 1, pg. 2), and refer to Section 4 (pg. 4).
8. If there are doubts regarding entity boundaries in relation to questions (a) and (c), **consult the Additional Modelling Rules** (Section 5, pg. 30), which include additional case-specific rules and examples.

3 Definitions

Labels	Ambiguities	Labels	Ambiguities
Asset	Field of Study, Organism, Body of Water, Intellectual Artefact	Method	Intellectual Artefact, Measuring Device, Mathematical Expression, Field of Study
Body of Water	Chemical, Ecosystem, System, Geographical Feature	Natural Disaster	Natural Phenomenon, Meteorological Phenomenon
Body Part	Organism	Natural Phenomenon	Meteorological Phenomenon, Physical Phenomenon, Natural Disaster, System
Chemical	Body of Water, System, Physical Phenomenon, Intellectual Artefact	Organism	Ecosystem, Asset, Body Part, Disease
Disease	Organism	Organization	Person, Field of Study, Location, Organism
Ecosystem	Organism, Geographical Feature, Body of Water, System	Person	Organization, Organism
Energy Source	Meteorological Phenomenon, Physical Phenomenon, Quantity	Physical Artefact	
Field of Study	Asset, Intellectual Artefact, Method, Organization	Physical Phenomenon	Meteorological Phenomenon, Natural Phenomenon, Quantity, System
Geographical Feature	Location, Ecosystem, Body of Water, System	Policy	System
Intellectual Artefact	Method, Field of Study, Asset, Geographical Feature	Quantity	Mathematical Expression, Physical Phenomenon, Time Period, Meteorological Phenomenon
Location	Geographical Feature, Organization, Measuring Device, Body of Water	Satellite	Measuring Device
Mathematical Expression	Method, Quantity, Physical Phenomenon	System	Meteorological Phenomenon, Body of Water, Geographical Feature, Policy
Measuring Device	Satellite, Method, System, Location	Time Period	Meteorological Phenomenon, Quantity, Physical Phenomenon, Method
Meteorological Phenomenon	Physical Phenomenon, System, Natural Phenomenon, Natural Disaster		

Table 1: Frequent ambiguities.

³For example, if there is still some doubt about which category it belongs to or the textual span that should be highlighted is unclear.

Type	Definition
Asset (pg. 4)	An Asset is an object or service of value to humans that can get destroyed or diminished by climate disasters/hazards. Key categories are health, buildings, infrastructure, and crops or livestock.
Body of Water (pg. 5)	A Body of Water is a distinct volume or mass of water, whether naturally occurring (like an ocean or river) or contained within a man-made system (like a reservoir or effluent flow). This includes specific named bodies, general types, and scientifically defined water masses.
Body Part (pg. 5)	A Body Part is a structural component of a living organism (plant, animal, or human), which is not the whole organism but a distinct anatomical or morphological part.
Chemical (pg. 6)	A Chemical is a substance with a distinct molecular composition, including elements, compounds, ions, biological molecules, and mixtures. It also encompasses classes of substances, reagents, and terms describing their chemical state or form.
Disease (pg. 7)	A Disease is an abnormal condition, disorder, or pathological state affecting an organism. This includes specific diseases, syndromes, symptoms, signs, injuries, toxicities, and observed pathological changes.
Ecosystem (pg. 8)	An Ecosystem is a community of interacting organisms and their physical environment, or a term used to describe such a community. This includes specific named ecosystems, general types of habitats, and collective terms for the biological components that define them.
Energy Source (pg. 9)	An Energy Source is a substance, material, natural phenomenon, or engineered system that provides or stores usable energy. This includes fuels, renewable resources, electricity, heat, and energy storage technologies.
Field of Study (pg. 10)	A Field of Study is a branch of knowledge, a specific academic or scientific discipline, or a theoretical domain of inquiry.
Geographical Feature (pg. 11)	A Geographical Feature is a natural or man-made feature of the Earth's solid surface or subsurface, including landforms, topographical structures, and defined regions of land.
Intellectual Artefact (pg. 12)	An Intellectual Artefact is a man-made product, typically of an informational, intellectual, or representational nature, resulting from scientific research, data collection, or modeling.
Location (pg. 13)	A Location is a point, area, or region in physical space, defined by natural, administrative, cultural, or geopolitical boundaries or functions. It may occur on Earth or elsewhere in the universe.
Mathematical Expression (pg. 14)	A Mathematical Expression is a term, symbol, equation, or concept related to mathematics, statistics, or formal modeling. This includes formulas, variables, statistical metrics, mathematical operations, and descriptions of quantitative relationships or trends.
Measuring Device (pg. 15)	A Measuring Device is an instrument, apparatus, technology, or system used to obtain quantitative measurements of a physical, biological, or chemical property.
Meteorological Phenomenon (pg. 16)	A Meteorological Phenomenon is a weather- or climate-related event, process, or pattern occurring in the atmosphere and its interaction with the Earth's surface, particularly oceans. This includes specific weather events, large-scale climate oscillations, atmospheric circulation patterns, and long-term trends.
Method (pg. 17)	A Method is a systematic procedure, technique, model, experiment, or approach used to conduct research, perform an analysis, collect data, or achieve a specific practical or scientific outcome.
Natural Disaster (pg. 18)	A Natural Disaster (or Hazard) is a large-scale adverse event, arising from natural, environmental, or climatic processes, that has the potential to cause significant harm or disruption to ecosystems, the environment, or human well-being.
Natural Phenomenon (pg. 19)	A Natural Phenomenon is an observable event or process that occurs in the natural world, driven by biological, physical, or chemical principles. It encompasses interactions, cycles, transformations, and large-scale environmental changes.
Organism (pg. 20)	An Organism is an individual living being (animal, plant, fungus, microbe) or a term that refers to a species, a taxonomic group, or a collection of individuals.
Organization (pg. 21)	An Organization is a formally constituted group, such as a company, institution, government agency, laboratory, or association, with a specific, collective purpose.
Other (pg. 22)	A Other is a miscellaneous category for named entities that do not fit into any of the more specific, defined types. It often includes abstract concepts, generic objects, textual references, and descriptive attributes that have been isolated as entities.
Person (pg. 22)	A Person is an individual or a group of human beings, identified by name, profession, or through a personal pronoun.
Physical Artefact (pg. 23)	A Physical Artefact (or Object) is a tangible, material entity with physical form. This includes both man-made products, such as tools, machines, vehicles, or buildings, and natural objects, such as soil, rocks, leaves, plant samples, powders, or other collected specimens.
Physical Phenomenon (pg. 24)	A Physical Phenomenon is an observable event or property that is not a substance, but rather a manifestation of energy, force, temperature, or state. It encompasses a wide range of processes from molecular interactions to large-scale environmental dynamics.
Policy or Objective (pg. 25)	A Policy or Objective is a defined goal, aim, or challenge that guides actions, or a framework of principles and rules to achieve those goals. This class encompasses high-level aims, specific targets, strategic plans, and guiding principles for management and governance.
Quantity (pg. 26)	A Quantity is a measurement, count, property, or value that is quantifiable. This includes numerical values with units, descriptive terms of amount or degree, statistical measures, and abstract quantifiable properties.
Satellite (pg. 27)	A Satellite is a man-made object, spacecraft, or instrument platform placed into orbit around a celestial body, primarily for purposes like remote sensing, data collection, or communication.
System (pg. 28)	A System is a set of interacting or interdependent components that form an integrated whole. This includes large-scale natural systems, engineered constructs, and abstract or theoretical frameworks that describe a set of interactions.
Time Period (pg. 29)	A Time Period is a specific point in time, a duration, or a recurring interval. This includes dates, years, seasons, epochs, and terms describing frequency or temporal extent.

Table 2: Entity type definitions.

4 Detailed Directives

4.1 Asset

An **Asset** is an object or service of value to humans that can get destroyed or diminished by climate disasters/hazards. Key categories are health, buildings, infrastructure, and crops or livestock.

4.1.1 Asset - Positive

- It is a general term for a resource that has value to humans (e.g., *resource, natural resources, materials*).
- It is a specific natural resource essential for life or economic activity (e.g., *water resources, groundwater resources, food resources*).
- It is an abstract concept representing human well-being, safety, or prosperity (e.g., *human health, public health, mental health, livelihood, welfare, life*).
- It is a fundamental service or infrastructure system that supports a community (e.g., *public drinking water supply, energy infrastructure*).
- It refers to agricultural or biological resources cultivated by humans (e.g., *crops, livestock, fisheries*).

4.1.2 Asset - Negative

- It refers to the physical feature or body containing the resource, not the resource itself → tag as **Geographical Feature, Body of Water, or Ecosystem** (e.g., *aquifer* is the feature, *groundwater resource* is the asset).
- It refers to the plan or strategy for managing the asset → tag as **Policy or Objective** (e.g., *water resource management* is the policy, *water resource* is the asset).
- It refers to a specific, individual living being rather than its collective value as a resource → tag as **Organism** (e.g., *trees* are organisms, but the *urban forest* as a provider of services could be an asset).
- It refers to a man-made object outside the context of its vulnerability to hazards → tag as **Physical Artefact** (e.g., a *tractor* is an artefact, but *agricultural machinery* as a whole could be considered an asset).

4.2 Body of Water

A **Body of Water** is a distinct volume or mass of water, whether naturally occurring (like an ocean or river) or contained within a man-made system (like a reservoir or effluent flow). This includes specific named bodies, general types, and scientifically defined water masses.

4.2.1 Body of Water - Positive

- It is a specific, named natural body of water (e.g., *Weddell Sea, St. John's River, Pacific*).
- It is a large, defined region within a larger body of water (e.g., *North Pacific, abyssal ocean*).
- It is a scientifically defined water mass with distinct properties (e.g., *AABW* for Antarctic bottom water, *DSW* for dense shelf water).
- It is a man-made or engineered body or flow of water (e.g., *agricultural reservoirs, effluent, inflow, wetlands* when used in the context of a water treatment system).
- It is a generic term referring to water as a physical volume or entity (e.g., *the water, surface waters, ocean*).

4.2.2 Body of Water - Negative

- It refers to the surrounding landform or the integrated biological system rather than the water itself → tag as **Geographical Feature** or **Ecosystem**. For example, in *the wetland ecosystem is diverse, wetland ecosystem* is an **Ecosystem**. However, in *water flows through the wetland, wetland* is treated as a **Body of Water**.
- It refers to a feature that bounds or contains the water → tag as **Geographical Feature** (e.g., *coastline, river basin, continental shelf*).
- It refers to water in a gaseous state or as precipitation in the atmosphere → tag as **Meteorological Phenomenon** (e.g., *water vapor, rain, clouds*).
- It refers to the abstract chemical substance (H₂O) and not a physical volume of it.

4.3 Body Part

A **Body Part** is a structural component of a living organism (plant, animal, or human), which is not the whole organism but a distinct anatomical or morphological part.

4.3.1 Body Part - Positive

- It is a **physical component** of an **organism**.
- It is an **anatomical structure** (e.g., *roots, leaves, stems, trunk, branches, wings, fins, eyes, lungs*).
- It functions as part of the whole organism, not independently.

4.3.2 Body Part - Negative

- It refers to the whole organism → tag as **Organism** (e.g., *tree, fish*).
- It refers to a collective feature → tag as **Ecosystem** (e.g., *vegetation*).
- It refers to a product derived from a body part → tag as **Physical Artefact** or **Other** depending on context (e.g., *timber, meat*).
- It refers to a substance extracted for analysis or a prepared sample of the part → tag as **Chemical** or **Physical Artefact**. (e.g., *serum* → Chemical, *blood sample* → Physical Artefact).

4.4 Chemical

A **Chemical** is a substance with a distinct molecular composition, including elements, compounds, ions, biological molecules, and mixtures. It also encompasses classes of substances, reagents, and terms describing their chemical state or form.

4.4.1 Chemical - Positive

- It is a specific chemical compound, element, or isotope, referred to by its name, formula, or abbreviation (e.g., *uric acid*, *H₂O*, *phenobarbital*, *PB*, *N*, *¹³⁷Cs*).
- It is a specific type of biological molecule or enzyme (e.g., *alanine aminotransferase*, *cytochrome P450*).
- It is a broad class or category of chemical substances (e.g., *greenhouse gases*, *nutrients*, *anti-epileptic drugs*, *enzymes*).
- It is a complex mixture or biological product treated as a single substance (e.g., *excreta*, *human waste*, *digestate*, *PB serum*).
- It is a reagent, solution, or mixture prepared for a scientific process (e.g., *standard solutions*, *enzyme reaction mixture*, *influent*).
- It is a term describing a specific chemical state, form, or ion of a substance (e.g., *monobasic urate form*, *protonated states*).
- It is a material, compound, or treated substance, including mixtures or composites whose identity is defined by its composition or treatment process (e.g., *fungus-treated materials*, *biodegraded wood*, *epoxy resin composite*, *activated carbon*, *biochar*, *soil sample*).

4.4.2 Chemical - Negative

- It refers to the disease or condition caused by a chemical → tag as **Disease** (e.g., *hepatotoxicity* is a Disease caused by the Chemical *PB*).
- It refers to a biological process involving chemicals → tag as **Natural Phenomenon** or **Method** (e.g., *metabolism*, *photosynthesis*).
- It refers to the organism that produces the substance → tag as **Organism**.
- It refers to an energy form rather than a substance → tag as **Energy Source** or **Physical Phenomenon** (e.g., *radiation*, but the source *radionuclide* is a Chemical).
- It refers to a purely physical property rather than the substance itself (e.g., *solubility*, *concentration*).
- It refers to a substance being discussed primarily for its role as a fuel or energy feedstock → tag as **Energy Source** (e.g., *Using ethanol as a fuel source...*).
- It refers to a physical specimen that has been collected for general analysis, rather than a substance defined by its chemical properties or use as a reagent → tag as **Physical Artefact** (e.g., *blood sample*, *tissue sample*).
- It refers to a complex biological tissue or organ system, not a refined substance or mixture derived from it → tag as **Body Part** (e.g., *blood* is a Body Part, but *serum* is a Chemical).

4.5 Disease

A **Disease** is an abnormal condition, disorder, or pathological state affecting an organism. This includes specific diseases, syndromes, symptoms, signs, injuries, toxicities, and observed pathological changes.

4.5.1 Disease - Positive

- It is a specific, named disease, syndrome, or disorder (e.g., *gout, epilepsy, chronic epileptic syndrome*).
- It is an observable symptom, clinical sign, or adverse effect that characterizes a state of ill health (e.g., *seizure, ataxia, anorexia, sedation*).
- It is a pathological process, injury, toxicity, or structural abnormality at the cellular or organ level (e.g., *PB hepatotoxicity, hippocampal atrophy, liver injury, lesions*).
- It is a broad category or general term for illness (e.g., *neurological disease, clinical disease, infectious diseases*).

4.5.2 Disease - Negative

- It refers to the causative agent of the disease → tag as **Organism** or **Chemical** (e.g., *fungus pathogens, a specific toxin*).
- It refers to the unaffected body part where the disease is located → tag as **Body Part** (e.g., *liver* in the case of *liver injury*).
- It refers to a medical treatment or drug used to manage the disease → tag as **Chemical** (e.g., *phenobarbital*).
- It refers to a diagnostic tool or procedure → tag as **Measuring Device** or **Method** (e.g., *EEG, MRI scan*).

4.6 Ecosystem

An **Ecosystem** is a community of interacting organisms and their physical environment, or a term used to describe such a community. This includes specific named ecosystems, general types of habitats, and collective terms for the biological components that define them.

4.6.1 Ecosystem - Positive

- It is a specific, named area defined by its ecological community, whether natural or man-made (e.g., *Orlando Easterly Wetlands (OEW)*, *cloud forests*).
- It is a general type or category of habitat or biome (e.g., *forests*, *wetland*, *grassland*, *cattle pasture*, *tropical rain forests*).
- It is a collective term for the plant or animal life that constitutes a community (e.g., *vegetation*, *plant communities*, *submerged aquatic vegetation (SAV)*, *trophic networks*).
- It is a distinct structural or habitat component within a larger ecosystem (e.g., *cypress dome*, *mixed marsh*, *rookery*).
- It is a large-scale or abstract ecological system (e.g., *biosphere*, *Genus-Physiognomy-Ecosystem (GPE)*).

4.6.2 Ecosystem - Negative

- It refers to an individual organism rather than a community → tag as **Organism** (e.g., *tree* is an Organism, but a *forest* is an Ecosystem).
- It refers to a purely geopolitical or administrative area → tag as **Location** (e.g., *Florida*, *northern Peru*).
- It refers to a non-living landform without emphasizing the biological community → tag as **Geographical Feature** (e.g., *mountain range*, *valley*).
- It refers to a body of water as a feature itself, rather than a habitat defined by its interaction with land and life → tag as **Body of Water** (e.g., *river*, *ocean*).

4.7 Energy Source

An **Energy Source** is a substance, material, natural phenomenon, or engineered system that provides or stores usable energy. This includes fuels, renewable resources, electricity, heat, and energy storage technologies.

4.7.1 Energy Source - Positive

- It is a chemical substance or material explicitly discussed in the context of its use as a fuel or feedstock for energy production (e.g., *ethanol for fuel, biomass, feedstock*).
- It is a material or phenomenon that produces or stores energy (e.g., *fossil fuels, biomass, solar radiation, wind, hydropower*).
- It can be a primary energy source (*solar, wind, biomass, fossil fuels*) or a derived energy carrier (*ethanol, hydrogen*).
- It may appear in terms like *RES (renewable energy sources)* or *energy mix* when clearly referring to the category of sources.

4.7.2 Energy Source - Negative

- It refers to a device or system that uses energy → tag as **Measuring Device, System, or Physical Artefact** depending on context.
- It refers to general energy concepts without specifying a source → tag as **Quantity** (e.g., *energy consumption*).
- It refers to a byproduct or waste that does not serve as an energy provider → tag as **Other** (e.g., *ash from burning biomass*).

4.8 Field of Study

A **Field of Study** is a branch of knowledge, a specific academic or scientific discipline, or a theoretical domain of inquiry.

4.8.1 Field of Study - Positive

- It is a recognized academic discipline (e.g., *physics, human physiology, climatologies*).
- It is a specific sub-field or area of research within a broader discipline (e.g., *phenology*).
- It is a broad category of disciplines (e.g., *historical humanistic*).
- It is a theoretical concept that defines a domain of inquiry (e.g., *niche space*).
- It can be an adjectival form of a discipline when used to describe a theme or category (e.g., *economic themes*).

4.8.2 Field of Study - Negative

- It refers to a specific research project or paper, rather than the discipline itself → tag as **Intellectual Artefact** (e.g., *study, findings*).
- It refers to a method of organization or a classification scheme used within a field → tag as **Method** (e.g., *hierarchical levels*).
- It refers to the practitioners of the field → tag as **Person** (e.g., *climatologists study the Field of Study climatology*).
- It refers to the object being studied, rather than the study of it → tag as **Meteorological Phenomenon** (e.g., *climate*).
- It refers to a term like *phenology* or *physiognomy* when used as a descriptive category or classification level, rather than as the discipline itself → tag as **Other** or **Quantity** depending on context (e.g., *in "classification was based on physiognomy," physiognomy is a category/quantity*).

4.9 Geographical Feature

A **Geographical Feature** is a natural or man-made feature of the Earth's solid surface or subsurface, including landforms, topographical structures, and defined regions of land.

4.9.1 Geographical Feature - Positive

- It is a specific natural landform or a region defined by its topography (e.g., *Antarctic continental shelf*, *Guadalfeo river basin*, *elevations*, *berms*).
- It is a subsurface formation (e.g., *shallow aquifers*).
- It is a general term for a type or expanse of land (e.g., *landmass*, *large landscapes*, *wet prairie*).
- It is a defined region or path, which can be natural or engineered (e.g., *East Asian monsoon domain*, *north flow path*, "northern wet" region).

4.9.2 Geographical Feature - Negative

- It refers to a biological community or habitat, where the focus is on the living organisms and their interactions → tag as **Ecosystem** (e.g., in "the wetland provides habitat," "wetland" is used as a place/feature; in "the wetland ecosystem is fragile," "wetland ecosystem" is an Ecosystem).
- It refers to a large, distinct body of water itself, not the surrounding land → tag as **Body of Water** (e.g., 'Pacific Ocean', 'Yangtze River'). Note that features like 'river basin' or 'continental shelf' are Geographical Features.
- It refers to a geopolitical entity like a country, city, or state → tag as **Location** (e.g., "Spain," "Antarctica"). Named natural features like mountain ranges or specific basins are tagged as Geographical Feature.
- It refers to the entire planet → tag as **Location** (e.g., "Earth").

4.10 Intellectual Artefact

An **Intellectual Artefact** is a man-made product, typically of an informational, intellectual, or representational nature, resulting from scientific research, data collection, or modeling.

4.10.1 Intellectual Artefact - Positive

- It is a collection of data, a dataset, or the results of an analysis or simulation (e.g., *data, model results, satellite estimates, bird count data, ensemble averaged results*).
- It is a specific, named dataset or climate model output (e.g., *Chelsa*).
- It is a scientific document, a body of literature, or a specific part of one (e.g., *report, literature, studies*).
- It is a map, image, or other visual representation of data (e.g., *snow albedo map series, satellite images, categorical vegetation maps*).
- It is an abstract product of intellectual work or observation (e.g., *findings, Observations, Knowledge*).
- It is a specific, implemented model (e.g., a software program or dataset) used for simulation (e.g., *the CESM-LME simulation dataset* or *hydrological model* when referring to a specific implementation).

4.10.2 Intellectual Artefact - Negative

- It refers to the tool, instrument, or platform used to generate the data → tag as **Measuring Device** or **Satellite** (e.g., a *lidar* produces *measurements*).
- It refers to the process, technique, or action of collecting or analyzing data → tag as **Method** (e.g., *simulation* is the **Method**, while *simulation results* is an **Intellectual Artefact**).
- It refers to a naturally occurring object in its original state, before being collected or measured → tag as **Organism, Geographical Feature**, etc.
- It refers to the institution or group that creates or provides the artefact → tag as **Organization** (e.g., *Climate Prediction Center* is the **Organization**; its dataset is the **Intellectual Artefact**).
- It refers to a sample that is treated as a substance with specific chemical properties or composition for use in a reaction → tag as **Chemical** (e.g., *standard solution, reagent, serum*).
- It refers to a simple textual pointer or reference → tag as **Other** (e.g., *Figure 5, Table 1*).
- It is a physical sample or specimen that has been collected or prepared for analysis → tag as **Physical Artefact** (e.g., *samples, blank*).

4.11 Location

A **Location** is a point, area, or region in physical space, defined by natural, administrative, cultural, or geopolitical boundaries or functions. It may occur on Earth or elsewhere in the universe.

4.11.1 Location - Positive

- It is a named geopolitical entity (e.g., *Czech Republic, China, Munich, Orange County*).
- It is a large-scale region, continent, or hemisphere (e.g., *North America, Northern Hemisphere, EA* for East Asia).
- It is a specific, defined site or place, which can be natural or man-made (e.g., *park, roost, stations, long-term care facilities, habitat*).
- It is a relative or directional term specifying a part of a larger, often implied, area (e.g., *northern, south, central, extratropical, midlatitude*).
- It is a general term for a place when used in a specific context (e.g., *area, sites, regions, geographical area*).
- It can refer to the entire planet in a geographical or societal context (e.g., *"around the world"*).
- It is a location defined by geographic coordinates or coordinate ranges (e.g., 30° – 45° N, 66° – 82° W).

4.11.2 Location - Negative

- It refers to a type of physical landform rather than a specific place → tag as **Geographical Feature** (e.g., *"basin," "continental shelf," "aquifer"*). While a named feature like "Sierra Nevada" can be a Location, a generic landform type is not.
- It refers to a biological community and its environment → tag as **Ecosystem** (e.g., *"forest," "grassland," "wetland"*). A "park" is a Location because it's a bounded area, whereas a "forest" is an Ecosystem because it's defined by its biological components.
- It refers to the body of water itself, like an ocean, sea, or river → tag as **Body of Water** (e.g., tag *"Pacific Ocean"* as Body of Water, not Location). However, directional descriptors of these areas (like "eastern" in "eastern Pacific") are tagged as Location.
- It refers to the planet as a celestial object in space → tag as **Location** (e.g., *"Earth's orbit"*).
- It refers to a specific environmental layer or substrate that is part of a landscape or ecosystem rather than a discrete location → tag as **Geographical Feature** (e.g., *"forest soil," "mountain slope," "riverbed," "peat layer"*). These describe *physical components* of the environment, not bounded locations.

4.12 Mathematical Expression

A **Mathematical Expression** is a term, symbol, equation, or concept related to mathematics, statistics, or formal modeling. This includes formulas, variables, statistical metrics, mathematical operations, and descriptions of quantitative relationships or trends.

4.12.1 Mathematical Expression - Positive

- It is a formal equation, formula, or a part of one (e.g., $K_{sp}=5120exp1.09I-5700T$, $pH = 9.3$).
- It is a symbol, variable, or acronym representing a mathematical or statistical quantity (e.g., n , R^2 , MAE , K_{sp}).
- It is a named mathematical or statistical operation, model, or concept (e.g., *mathematical model*, *5-yr running mean*, *95% confidence level*, *function*, *variable*).
- It is a term describing a quantitative relationship, trend, or profile in data (e.g., *decreasing trend*, *concentration profile*, *regional trends*, *energy peaks*).
- It is a reference to a specific equation within the text (e.g., *equation (1)*).

4.12.2 Mathematical Expression - Negative

- It refers to a specific measured value or property → tag as **Quantity** (e.g., In "*the concentration is 5 mg/L*," "*5 mg/L*" and "*concentration*" are Quantities.).
- It refers to a textual or documentary element like a figure or table, even if it contains mathematical content → tag as **Intellectual Artefact** or **Other** (e.g., *Fig. 2*, *Table 1*).
- It refers to the general act of analyzing or calculating something → tag as **Method** (e.g., "*analysis*" is a Method, but a specific "*mathematical model*" used in the analysis is a Mathematical Expression).
- It refers to a specific software implementation or dataset of a model → tag as **Intellectual Artefact** (e.g., "*the WiMMed hydrological model*").

4.13 Measuring Device

A **Measuring Device** is an instrument, apparatus, technology, or system used to obtain quantitative measurements of a physical, biological, or chemical property.

4.13.1 Measuring Device - Positive

- It is a specific, named piece of equipment used for measurement (e.g., *Infinite 200 PRO Tecan microplate reader, nephelometer*).
- It is a general type or class of measurement technology, including common acronyms (e.g., *lidar, EEG*).
- It is a term for a measurement process that is used metonymically to refer to the device or technology itself (e.g., *MRI scan*).
- It is a system or collection of instruments that work together to take measurements (e.g., *regional weather stations*).
- It is a generic pronoun or noun referring to a specific device mentioned in the context (e.g., *The instrument*).

4.13.2 Measuring Device - Negative

- It refers to the unit or result of a measurement → tag as **Quantity** (e.g., *510 nm, 24 h*).
- It refers to a software program used to control a device or analyze its data → tag as **Intellectual Artefact** (e.g., *Tecan i-control software*).
- It refers to a general piece of lab equipment not used for quantification → tag as **Physical Artefact** (e.g., *beaker, petri dish*).
- It refers to an abstract protocol or procedure rather than the physical or technological tool → tag as **Method** (e.g., *titration, statistical analysis*).

4.14 Meteorological Phenomenon

A **Meteorological Phenomenon** is a weather- or climate-related event, process, or pattern occurring in the atmosphere and its interaction with the Earth's surface, particularly oceans. This includes specific weather events, large-scale climate oscillations, atmospheric circulation patterns, and long-term trends.

4.14.1 Meteorological Phenomenon - Positive

- It is a specific or general weather event (e.g., *rain, rainfall, snow, hurricane, storm event*).
- It is a large-scale, named climate pattern, circulation, or oscillation, often an acronym (e.g., *East Asian summer monsoon (EASM), Pacific decadal oscillation (PDO), IPO*).
- It is a specific state or phase of a climate pattern (e.g., *cool PDO phase*).
- It is a pattern, anomaly, or variation in a meteorological variable (e.g., *SST anomalies, dipolar patterns, precipitation anomalies*).
- It is an atmospheric pressure system or air circulation pattern (e.g., *anticyclone, cross-equatorial flows, Hadley circulation*).
- It is a long-term process or trend in the climate system (e.g., *climate change, climate variability*).

4.14.2 Meteorological Phenomenon - Negative

- It refers to the overall, long-term state of the atmosphere as an integrated whole → tag as **System** (e.g., *"climate" is a System, but "climate change" is a Meteorological Phenomenon*).
- It refers to the location where the phenomenon occurs → tag as **Location** (e.g., *"North Pacific" is a Location where a "PDO" event occurs*).
- It refers to the data or model that represents the phenomenon → tag as **Intellectual Artefact** (e.g., *"climate model," "weather data"*).
- It refers to the substances involved in the phenomenon → tag as **Chemical** (e.g., *"greenhouse gases," "water vapor"*).
- It refers to an event primarily in terms of its destructive impact on ecosystems or human life → tag as **Natural Disaster** (e.g., *Note: there can be overlap, but hurricane as an atmospheric event is a phenomenon, while its impact could be a disaster*).
- It refers to the measured or quantified amount/value of the process or event → tag as **Quantity** (e.g., *"25 mm of rainfall," "mean annual temperature"*).

4.15 Method

A **Method** is a systematic procedure, technique, model, experiment, or approach used to conduct research, perform an analysis, collect data, or achieve a specific practical or scientific outcome.

4.15.1 Method - Positive

- It is a named scientific, analytical, or medical procedure (e.g., *total Kjeldahl nitrogen method*, *pyrolysis*, *xenotransplantation*, *SID dosing*, *enzymatic assay*).
- It is a general term for a type of scientific action or process (e.g., *analysis*, *sampling*, *monitoring*, *mapping*, *classification*, *quantification*).
- It is an experimental design, simulation, or model used to generate data or test a hypothesis (e.g., *solar-only forcing experiments*, *CTRL experiment*, *hydrological model*).
- It refers to a specific step or component of the research process (e.g., *conceptualization*, *validation*, *formal analysis*, *data curation*).
- It is a specific approach or system for organizing or achieving a goal (e.g., *multidisciplinary approach*, *GPE system*).

4.15.2 Method - Negative

- It refers to the data, results, or product generated by the method → tag as **Intellectual Artefact** (e.g., *"simulation" is a Method, but "simulation results" is an Intellectual Artefact; "analysis" is a Method, but its "findings" are an Intellectual Artefact*).
- It refers to the instrument or device used to perform the method → tag as **Measuring Device** (e.g., *"EEG" is a Measuring Device used for the Method of "monitoring"*).
- It refers to the object or substance being studied or measured → tag as **Organism, Chemical, Ecosystem**, etc. (e.g., *"uric acid" is a Chemical, its "quantification" is a Method*).
- It refers to a natural process occurring on its own, not as a directed technique → tag as **Natural Phenomenon**.
- It refers to a professional role or discipline rather than the actions they take → tag as **Person** or **Field of Study**.

4.16 Natural Disaster

A **Natural Disaster** (or Hazard) is a large-scale adverse event, arising from natural, environmental, or climatic processes, that has the potential to cause significant harm or disruption to ecosystems, the environment, or human well-being.

4.16.1 Natural Disaster - Positive

- It is a specific, destructive natural event (e.g., *volcanic eruptions, drought*).
- It is a process of widespread environmental degradation or damage, especially when the focus is on its adverse outcome (e.g., *eutrophication* leading to fish kills, *pollution, forest loss*).
- It is a major technological or industrial accident with severe environmental consequences (e.g., *nuclear accidents*).
- It is a general term for such catastrophic events (e.g., *global threats*).

4.16.2 Natural Disaster - Negative

- It refers to the polluting substance itself, not the process of pollution → tag as **Chemical** (e.g., *radionuclides, nitrates*).
- It refers to the long-term process driving disasters, not the event itself → tag as **System** or **Natural Phenomenon** (e.g., *climate change*).
- It refers to an outbreak of illness in organisms → tag as **Disease**.
- It refers to the geographical feature from which a disaster originates → tag as **Geographical Feature** (e.g., *"volcano" is the feature, 'volcanic eruption' is the disaster*).

4.17 Natural Phenomenon

A **Natural Phenomenon** is an observable event or process that occurs in the natural world, driven by biological, physical, or chemical principles. It encompasses interactions, cycles, transformations, and large-scale environmental changes.

4.17.1 Natural Phenomenon - Positive

- It is a biological process, state, or behavior (e.g., *growth response, photosynthesis, drug metabolism, rot, disease transmission*).
- It is an ecological interaction or dynamic process, especially when the focus is on the mechanism rather than the destructive outcome (e.g., *eutrophication* as a nutrient cycle process).
- It is a broad physical or environmental process or change (e.g., *climate change, global change, solar cycle, sedimentation*).
- It is a general term for a process or interaction (e.g., *mechanism, interactions, feedbacks, natural variability*).
- It can be a process driven by human influence on the natural world (e.g., *human activity, land use change*).

4.17.2 Natural Phenomenon - Negative

- It refers to a weather or climate pattern in the atmosphere → tag as **Meteorological Phenomenon** (e.g., *monsoon, hurricane, rainfall*).
- It refers to a human-directed, systematic procedure or experiment → tag as **Method** (e.g., *solid-state fermentation* when used as a lab technique, *sampling*). A natural process like "microbial degradation" is a Natural Phenomenon.
- It refers to the destructive outcome of an event rather than the process itself → tag as **Natural Disaster** (e.g., *pollution, forest loss*). Note: there is overlap; "eutrophication" can be either, depending on whether the focus is the process (Natural Phenomenon) or the destructive outcome (Natural Disaster).
- It refers to the integrated system as a whole, not a process within it → tag as **System** (e.g., *climate* is a system, while *climate change* is a phenomenon).
- It refers to the substances involved in the process → tag as **Chemical** (e.g., *carbon dioxide* is a Chemical used in the Natural Phenomenon of *photosynthesis*).

4.18 Organism

An **Organism** is an individual living being (animal, plant, fungus, microbe) or a term that refers to a species, a taxonomic group, or a collection of individuals.

4.18.1 Organism - Positive

- It is a specific species, referred to by its scientific name, common name, or an acronym (e.g., *Pinus sylvestris*, *Scotch pine*, *great egret*, *CSL*).
- It is a general category or type of living thing (e.g., *trees*, *birds*, *microorganisms*, *wildlife*, *aquatic species*).
- It is a collective term for a group of organisms, often of the same or similar types (e.g., *transient bird populations*, *wading birds*).
- It is a pronoun or general noun that refers to a specific organism mentioned in the context (e.g., *the animal*, *it*, *species*).
- It refers to humans or groups of humans as biological subjects (e.g., *human medical patients*, *Human*).

4.18.2 Organism - Negative

- It refers to a structural component of an organism, not the whole organism → tag as **Body Part** (e.g., *hippocampus*, *blood*, *muscle*, *leaves*).
- It refers to a community of different species or the collective plant life in an area that forms a habitat → tag as **Ecosystem** (e.g., An individual *plant* is an Organism, but collective *vegetation* or *plant communities* are Ecosystems).
- It refers to a substance or product derived from an organism → tag as **Chemical** (e.g., *excreta*, *guano*, *uric acid*).

4.19 Organization

An **Organization** is a formally constituted group, such as a company, institution, government agency, laboratory, or association, with a specific, collective purpose.

4.19.1 Organization - Positive

- It is a named government body, academic institution, professional association, or commercial company (e.g., *Environmental Protection Agency, University of Central Florida, American Veterinary Medical Association, Molecular Probes*).
- It is a specific, named facility, plant, or task force (e.g., *Iron Bridge WWTP, International Veterinary Epilepsy Task Force*).
- It is an acronym for a formal organization (e.g., *IPCC, EPA, WWTP*).
- It can be a generic term when it refers to a specific, formal entity within the context (e.g., *reference laboratory* when referring to Idexx Laboratories).
- It can be a publisher or commercial service acting as a proxy for the parent company (e.g., *ScienceDirect*).

4.19.2 Organization - Negative

- It refers to an informal group of people defined by their profession or role → tag as **Person** (e.g., *clinicians, Veterinarians, veterinary staff*).
- It refers to a group of authors in a citation or a pronoun for the authors → tag as **Person** (e.g., *Hahn et al, we*).
- It refers to an abstract project, initiative, or program, not the formal body executing it → tag as **Policy** (e.g., *project, climate initiatives*).
- It refers to a generic place or building, not the entity that occupies it → tag as **Location, Physical Artefact** or **Other** (e.g., *facility*).

4.20 Other

A **Other** is a miscellaneous category for named entities that do not fit into any of the more specific, defined types. It often includes abstract concepts, generic objects, textual references, and descriptive attributes that have been isolated as entities.

4.20.1 Other - Positive

- It is an abstract concept, goal, or state that is not a scientific field or method (e.g., *health, reasons, climate targets, physical dependence, environmental sustainability*).
- It is a textual or documentary reference (e.g., *Figure 5, Table 1, previous studies, final manuscript*).
- It is a generic, non-scientific man-made object that does not fit the informational focus of the **Intellectual Artefact** type (e.g., *street lights, automobile tyres*).
- It is a descriptive attribute, classifier, or characteristic tagged as a standalone entity (e.g., *deciduous, broad-leaved, model-simulated, migratory*).

4.20.2 Other - Negative

- It clearly fits into a more specific category. Always check other entity types first.
- If it is a scientific dataset, report, or prepared sample → tag as **Intellectual Artefact** (e.g., *"data," "blood sample"*).
- If it is a specific procedure or technique → tag as **Method** (e.g., *"medication administration," "forest management"*).
- If it is a diagnosable medical condition → tag as **Disease** (e.g., *"drowning," "reduced appetite"*).
- If it is a person's full name or professional role → tag as **Person** (e.g., *"two anonymous reviewers"*).
- If it refers to a biological community → tag as **Ecosystem** (e.g., *"vegetation types"*).
- If it is a quantifiable value or property → tag as **Quantity** (e.g., *"half-life," "therapeutic range"*).

4.21 Person

A **Person** is an individual or a group of human beings, identified by name, profession, or through a personal pronoun.

4.21.1 Person - Positive

- It is a specific, named individual (e.g., *Wilcox, Brooke Rohman*).
- It is a group of specific individuals, often in a citation format (e.g., *Hahn et al., Mukherjee and Borhad*).
- It is a group of people defined by their profession or role (e.g., *paleoclimatologists, researchers*).
- It is a personal pronoun referring to the authors or researchers involved in the study (e.g., *We*).

4.21.2 Person - Negative

- It refers to a non-human organism → tag as **Organism**.
- It refers to a formal institution or company → tag as **Organization** (e.g., *"University of X," "NASA"*).
- It refers to a concept, method, or object named after a person (an eponym) → tag as **Method** or **Physical Phenomenon** (e.g., *"Arrhenius plots," "Doppler effect"*).

4.22 Physical Artefact

A **Physical Artefact** (or Object) is a tangible, material entity with physical form. This includes both man-made products, such as tools, machines, vehicles, or buildings, and natural objects, such as soil, rocks, leaves, plant samples, powders, or other collected specimens.

4.22.1 Physical Artefact - Positive

- It is a specific man-made structure or facility (e.g., *road, pavements, Hazaribagh tannery*).
- It is a category of man-made equipment or objects (e.g., *machinery, vehicles*).
- It is a processed or manufactured material used for production (e.g., *construction materials, used tyres*).
- It is a component or specific type of a larger man-made structure (e.g., *asphalt pavement*).
- It is a tangible natural object or sample, including soils, rocks, leaves, plant parts, powders, or other collected specimens (e.g., *dried-plant samples, soil samples, fine powder*).

4.22.2 Physical Artefact - Negative

- It refers to the abstract value or service provided by the object, especially in the context of its importance to humans → tag as **Asset** (e.g., *transportation infrastructure* is an asset, a specific *road* is a physical artefact).
- It refers to raw, naturally occurring substances that have not been processed into a product → tag as **Geographical Feature** or **Chemical** (e.g., *rock, crude oil*).
- It refers to the organization that created or operates the artefact → tag as **Organization** (e.g., *the tannery company*).
- It refers to a living organism, even if cultivated by humans → tag as **Organism** (e.g., *trees* in a plantation).

4.23 Physical Phenomenon

A **Physical Phenomenon** is an observable event or property that is not a substance, but rather a manifestation of energy, force, temperature, or state. It encompasses a wide range of processes from molecular interactions to large-scale environmental dynamics.

4.23.1 Physical Phenomenon - Positive

- It is a process involving energy, force, or a change in physical state (e.g., *evaporation, circulation, sedimentation, coastal upwelling, gravity*).
- It is a form of electromagnetic radiation or light (e.g., *artificial light, ultraviolet*).
- It is a property or dynamic related to temperature, heat, or pressure (e.g., *warming, cooling, heat island effect, air temperature field*).
- It is a recurring pattern, cycle, or signal observed in data (e.g., *solar cycle, quasi-11-yr periodicity, seasonal changes*).
- It is a change or variation in a physical property (e.g., *SST variability, contraction, conformational changes*).
- It is a chemical or biochemical process described in physical terms (e.g., *anaerobic digestion (AD), mineralization, protein aggregation*).
- It can be a driver or forcing agent in a system (e.g., *solar forcing, external forcing*).

4.23.2 Physical Phenomenon - Negative

- It refers to a specific weather event or large-scale climate oscillation → tag as **Meteorological Phenomenon** (e.g., Common confusions: *hurricane, monsoon, El Niño*. A general *circulation* is a Physical Phenomenon, but a named atmospheric circulation like *Hadley circulation* is a Meteorological Phenomenon).
- It refers to a biological process driven by life functions → tag as **Natural Phenomenon** (e.g., Common confusions: *metabolism, photosynthesis, growth*. There is overlap; *anaerobic digestion* by microbes can be seen as either).
- It refers to a substance or material → tag as **Chemical** (e.g., *water vapor* is a Chemical, but *evaporation* is a Physical Phenomenon).
- It refers to a quantity or unit of measurement → tag as **Quantity**.
- It refers to a discrete, destructive environmental event → tag as **Natural Disaster** (e.g., *light pollution* is a Physical Phenomenon, but when framed as a large-scale negative impact, it could be a Natural Disaster).

4.24 Policy or Objective

A **Policy or Objective** is a defined goal, aim, or challenge that guides actions, or a framework of principles and rules to achieve those goals. This class encompasses high-level aims, specific targets, strategic plans, and guiding principles for management and governance.

4.24.1 Policy or Objective - Positive

- It is a specific plan, strategy, or structured approach designed to achieve a goal (e.g., *water resource planning, scaled up malaria treatment and control intervention strategies, strategy*).
- It is a framework, principle, or process for managing resources or systems (e.g., *management of water resources, sustainable exploitation, management*).
- It is a high-level goal, aim, or a recognized problem that requires a solution (e.g., *public health challenge, sustainability goals*).
- It refers to a formal commitment or regulatory instrument (e.g., *emission reduction targets*).

4.24.2 Policy or Objective - Negative

- It refers to the specific scientific or technical procedure used to implement or study a policy → tag as **Method** (e.g., *hydrogeological modeling, cost-benefit analysis*).
- It refers to the resource or system being managed, not the management framework itself → tag as **Asset** or **System** (e.g., *urban trees, aquifers*).
- It refers to the institution or body responsible for setting or enacting the policy → tag as **Organization** (e.g., *World Health Organization, government*).
- It refers to the specific report, law, or named agreement that codifies a policy → tag as **Intellectual Artefact** (e.g., *Paris Agreement, The Clean Water Act*).
- It refers to the academic discipline rather than the applied practice of management or planning → tag as **Field of Study** (e.g., *urban planning, policy analysis*).

4.25 Quantity

A **Quantity** is a measurement, count, property, or value that is quantifiable. This includes numerical values with units, descriptive terms of amount or degree, statistical measures, and abstract quantifiable properties.

4.25.1 Quantity - Positive

- It is a specific numerical value, often with units (e.g., *57 million liters, 3.9°C, 1.5 mg/kg, 10 U, 88%*).
- It is a general term for a measured or measurable property (e.g., *concentration, temperature, pH, solubility, dose, salinity*).
- It is a statistical measure, index, or rate (e.g., *mean, average, index, frequency, correlation coefficient, variance*).
- It is a descriptive term of amount, level, or frequency (e.g., *maximum, high, several, twice-daily*).
- It is a term describing spatial extent or distribution (e.g., *distribution, abundance, pixel size, area*).
- It is an acronym for a specific quantitative measure (e.g., *DBH* for diameter at breast height, *TKN* for total Kjeldahl nitrogen).

4.25.2 Quantity - Negative

- It refers to the instrument used to take the measurement → tag as **Measuring Device**.
- It refers to the method of measurement → tag as **Method** (e.g., *"analysis" is a Method, but the resulting "concentration" is a Quantity*).
- It refers to the phenomenon being quantified, not the quantity itself → tag as **Physical Phenomenon** or **Meteorological Phenomenon** (e.g., *"solar activity" is a phenomenon, while its "intensity" is a Quantity*).
- It refers to a substance being measured → tag as **Chemical**.
- It refers to a specific time duration or point in time → tag as **Time Period**.

4.26 Satellite

A **Satellite** is a man-made object, spacecraft, or instrument platform placed into orbit around a celestial body, primarily for purposes like remote sensing, data collection, or communication.

4.26.1 Satellite - Positive

- It is the specific name of an artificial satellite or a series of satellites (e.g., *Landsat*, *Landsat 8*).
- It is a specific instrument or program that is located on a satellite platform and is used as a proxy for the satellite itself (e.g., *SPOT VEGETATION*).
- It is the general term 'satellite' used in the context of remote sensing or orbital technology.

4.26.2 Satellite - Negative

- It refers to the data or images generated by a satellite → tag as **Intellectual Artefact** (e.g., "*Landsat images*").
- It refers to the technique or process of using a satellite for data collection → tag as **Method** (e.g., "*satellite remote sensing*").
- It refers to a natural satellite, like a moon → tag as **Location**.
- It refers to the agency or organization that built or operates the satellite → tag as **Organization** (e.g., *NASA*, *ESA*).
- It refers to a ground-based technology or device → tag as **Physical Artefact** or **Measuring Device** (e.g., *a receiving station*).

4.27 System

A **System** is a set of interacting or interdependent components that form an integrated whole. This includes large-scale natural systems, engineered constructs, and abstract or theoretical frameworks that describe a set of interactions.

4.27.1 System - Positive

- It is a large-scale, complex natural system involving multiple interacting parts (e.g., *climate, atmosphere, global overturning circulation*).
- It is a persistent meteorological or atmospheric feature composed of interacting elements (e.g., *North Pacific high, mei-yu fronts*).
- It is an engineered or man-made construct designed to perform a specific function (e.g., *engineered wetland, flow train, deep marsh water control structure*).
- It is an abstract or theoretical framework, mechanism, or structure that describes how components interact (e.g., *top-down mechanism, north–south dipolar structure*).
- It is a generic term for an interconnected set of parts (e.g., *Earth systems, subsystems*).

4.27.2 System - Negative

- It refers to a biological community interacting with its physical environment, where the emphasis is on the living organisms → tag as **Ecosystem** (e.g., a *forest* is an Ecosystem, but an *engineered wetland* which emphasizes the design is a System).
- It refers to a single, discrete event or process rather than the larger collection of interacting parts that cause it → tag as **Physical Phenomenon** or **Natural Phenomenon** (e.g., *rainfall* is a phenomenon, while *climate* is the system).
- It refers to a static landform or topographical feature → tag as **Geographical Feature** (e.g., *river basin*).
- It refers to a software model or a dataset representing a system → tag as **Intellectual Artefact** (e.g., *climate model*).

4.28 Time Period

A **Time Period** is a specific point in time, a duration, or a recurring interval. This includes dates, years, seasons, epochs, and terms describing frequency or temporal extent.

4.28.1 Time Period - Positive

- It is a specific date, year, or range of years (e.g., *September 2017*, *1993*, *AD 850–1850*).
- It is a specific season, month, or part of a day (e.g., *summer*, *December*, *night*).
- It is a duration or interval of time (e.g., *seven-month period*, *several decades*, *21 days*, *48 h*).
- It is a named epoch, era, or defined period (e.g., *strong 11-yr solar cycle epoch*, *preindustrial period*).
- It is a term describing a recurring interval or frequency (e.g., *once a day*, *8–16-yr*, *decadal*).
- It is a relative or general temporal reference (e.g., *present*, *past*, *future*, *modern days*).

4.28.2 Time Period - Negative

- It refers to a physical or meteorological phenomenon that has a temporal component but is not the time period itself → tag as **Physical Phenomenon** or **Meteorological Phenomenon** (e.g., Common confusion: *solar cycle* is a Physical Phenomenon, but *11-yr* is the Time Period that describes it).
- It refers to an abstract temporal property rather than a specific duration or point in time → tag as **Quantity** (e.g., in "we measured the duration", *duration* is a Quantity, but in "the duration was 5 hours," *5 hours* is a Time Period. The term *frequency* is also a Quantity).
- It refers to an event or action rather than the time it occurred → tag as **Method** (e.g., *sampling*).

5 Additional Modelling Rules

Each rule should be concise, numbered, and written in a consistent “IF–THEN” or “WHEN–THEN” structure. The intent is to provide precise modeling guidance for recurring or ambiguous annotation cases.

5.0.1 Quantity - Modeling Change or Variation

Condition (WHEN): A term expresses a change, trend, or fluctuation in a measurable property (e.g., *temperature increase, price fluctuations, decline in precipitation*).

Guideline (THEN):

- Annotate only the core measurable property (e.g., *temperature, price, precipitation*) as a **Quantity**.
- The dynamic meaning (e.g., *increase, decline, fluctuate*) will be captured in the relation semantics between entities.
- Do not create compound Quantity tags like “temperature increase” or “price fluctuation.”

Example(s):

"Thanks to **low**^{Quantity} **operating costs**^{Quantity}, **renewable energy sources**^{EnergySource} are little affected by **fossil fuel**^{EnergySource} **price**^{Quantity} fluctuations [27]".

5.0.2 Location – Directional or Regional Descriptors

Condition (WHEN): A Location or large Body of Water is modified by a directional or regional descriptor such as *northern, southern, eastern, western*, or similar (e.g., *northern Peru, western Pacific, eastern Amazon Basin*).

Guideline (THEN):

- Annotate the descriptor (e.g., *northern*) as a standalone **Location** entity.
- Annotate the base place (e.g., *Peru, Pacific*) using its usual category (typically **Location** or **Body of Water**).
- Do not combine them into a compound entity (e.g., avoid “northern Peru” as a single Location span).

Example(s):

" We^{Person} compared **5 years**^{TimePeriod} of **in situ**^{Method} **weather data**^{IntellectualArtefact} from **cloud forests**^{Ecosystem} in **northern**^{Location} **Peru**^{Location},"

5.0.3 Coordinated Modifiers – When Shared-Head Constructions Prevent Full Entity Spans

Condition (WHEN): A phrase contains multiple coordinated descriptors linked by “and” (e.g., *seasonal and annual*) that refer to distinct conceptual entities (e.g., *seasonal mean precipitation* and *annual mean precipitation*), but the sentence structure provides the head noun only once (e.g., *seasonal and annual mean precipitation*). Because the head noun is not repeated, the full conceptual entities cannot be annotated as separate spans.

Guideline (THEN):

- Annotate each descriptor (e.g., *seasonal*, *annual*) as its own entity with its appropriate category (e.g., **Time Period**).
- Annotate the shared head noun (e.g., *mean precipitation*) once, according to its category (e.g., **Quantity**).
- Treat the descriptors as independent entities that semantically modify the shared head, even though the full spans (*seasonal mean precipitation*, *annual mean precipitation*) cannot be annotated.
- If one descriptor cannot be typed confidently, annotate as **Other**.

Example(s):

"Over the coastal^{Location} and developing^{Other} countries^{Location} of the southern^{Location} African region^{Location}, such as Mozambique^{Location}, there will likely be an increase in the frequency^{Quantity} and intensity^{Quantity} of heat waves^{MeteorologicalPhenomenon}, a decrease in seasonal^{TimePeriod} and annual^{TimePeriod} mean precipitation^{Quantity}, ..."

5.0.4 Excluding Value Relations from Quantity Spans – Restriction of quantity spans to value and unit.

Condition (WHEN): A specific numeric value or measurement is preceded by comparative, relational, or approximation modifiers (e.g., *more than*, *less than*, *over*, *approximately*, *up to*, *at least*) that indicate a range, limit, or estimation.

Guideline (THEN): Annotate only the numeric value and its associated unit (if present) as the **Quantity**. Exclude the relational phrases or modifiers from the entity span.

Example(s):

"Currently^{TimePeriod}, the share of renewable energy^{EnergySource} in the total consumption^{Quantity} of the Czech Republic^{Location} is more than 17%^{Quantity}, while the vast majority of this energy^{EnergySource} comes from various forms of biomass^{EnergySource} [70]."

5.0.5 Prepositional Boundaries – Splitting entities at prepositions to ensure granularity.

Condition (WHEN): Two distinct conceptual entities or descriptors are connected by a preposition (e.g., *of*, *in*, *among*, *at*, *for*). The sentence structure creates a compositional phrase (e.g., *Analysis of X*, *Interaction among Y*).

Guideline (THEN):

- Do not annotate the entire compositional phrase as a single entity.
- Annotate the head concept (before the preposition) and the modifying concept (after the preposition) as separate independent entities.
- **Exclude** the preposition itself from all entity spans.

Example(s):

" Quantitative analyses^{Method} of the complex interactions^{NaturalPhenomenon} and feedbacks^{NaturalPhenomenon} among plants^{Organism}, soils^{GeographicalFeature}, climate^{System}, and herbivores^{Organism} at the landscape scale^{Quantity} require ..."

Template

[Rule Title] - short descriptive label.

Condition (IF/WHEN): state the situation or ambiguity.

Guideline (THEN): specify how to annotate.

(Optional) Example(s): demonstrate correct usage.

6 Label Studio Quick Guide

6.1 Basic Information

Upon opening the link, the Label Studio user interface is visible as shown in Figure 2. Here, all tasks for a given annotator are listed in the forms indicated by the red ellipse numbered 2. Each task has its core information, including: *Completed*, *Total annotations per task*, *Total skipped annotations*, *Annotated by*, *Updated at*, *sentence*, *sentence_id* and *paper_id*.

To start the annotation process, press **Label All Tasks** indicated by the red ellipse numbered 1.

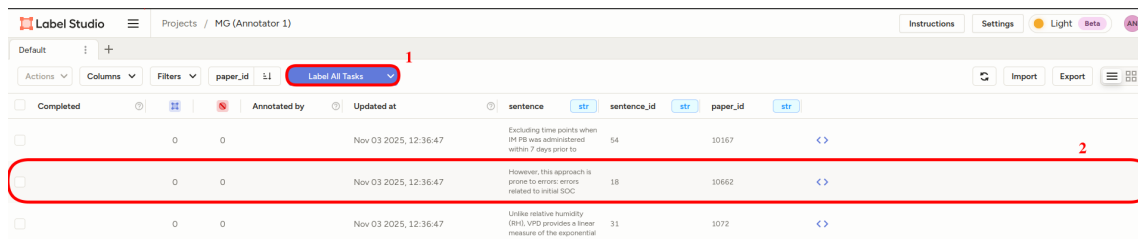


Figure 2: Label Studio interface.

6.2 Deletion and Correction

After pressing **Label All Tasks**, the annotation interface becomes visible, as shown in Figure 3:

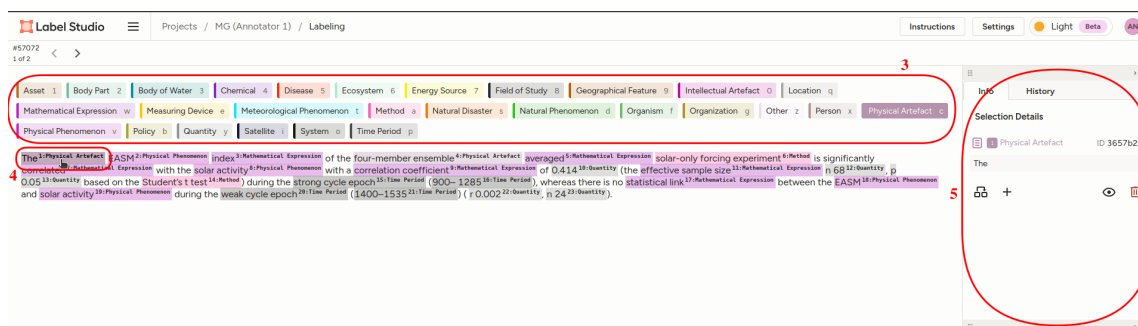


Figure 3: Annotation interface - Deletion.

Here, all **28 entity categories** are visible (red ellipse number 3). Below this, a **sentence to be annotated** is presented. Each sentence comes with a pre-annotation performed by a prediction model. These model predictions, such as the one denoted by red ellipse number 4, are more often than not incorrect. By pressing the highlighted annotation, for example, "The" indicated by red ellipse number 4, its details appear in the **Selection Details window** (red ellipse number 5). The prediction model assigned to this highlighted part of the text ("The") is *Physical Artefact*. This prediction is incorrect and can be deleted by pressing the **bin icon** in the Selection Details window. **In this way, each highlighted text (annotation) can be deleted.**

Next, as depicted in Figure 4, pressing on the highlighted text "EASM" (red ellipse number 6) shows that it is currently annotated by a prediction model as "Physical Phenomenon" (visible in the Selection

Details window). "EASM", which stands for "East Asian summer monsoon", better fits the *Meteorological Phenomenon* category, as it is a weather- or climate-related event. Thus, by pressing on the desired category - *Meteorological Phenomenon* (red ellipse number 7) - the category of "EASM" is changed (see Figure 5). In this way, each highlighted text (annotation) can be corrected.

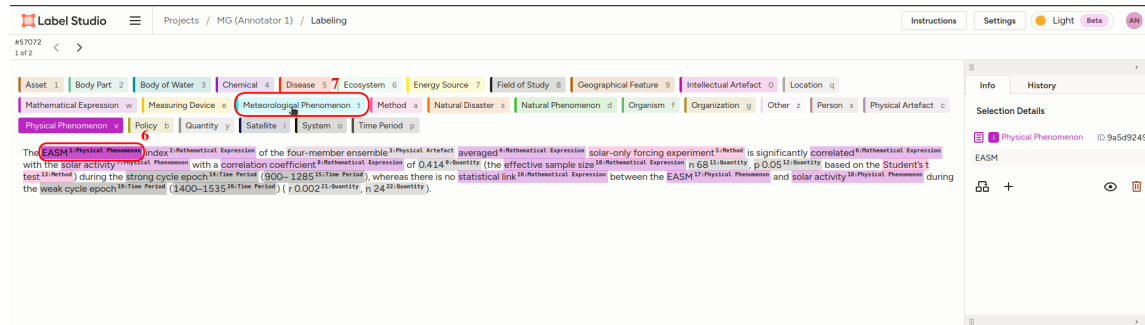


Figure 4: Annotation interface - Correction.

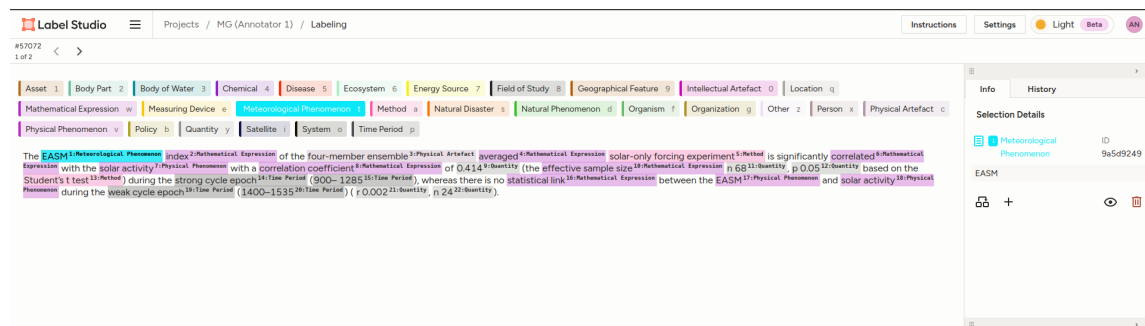


Figure 5: Annotation interface - Correction result.

6.3 Annotation

Finally, on closer reading, it is clear that "EASM" is actually part of a larger contextual entity, specifically, "The EASM index" or simply "EASM index". The "EASM index", like any other index, is a numerical and quantifiable value, and thus fits better in the *Quantity* category. For this reason, both the "EASM" and "index" annotations should be deleted (see Section 6.2). After performing the deletion, as shown in Figure 6, pressing the desired category, *Quantity*, sets it as the active annotation category (red ellipse number 8). Pressing and holding the left mouse button at the beginning of the desired text ("EASM index") and dragging the mouse to the end of the text (movement depicted by red arrow number 9) creates a new annotation:

The *EASM index*^{*Quantity*} of the

In this way, each highlighted text can be annotated, deleting previous annotations if they exist and highlighting the desired text with a new active category.

Once all annotations are correct, press the **Submit/Update** button (Figure 7, red ellipse number 10). In this way, the annotation for a given task (single sentence) is completed and a new one is presented.

Alternatively, view the video at the link.

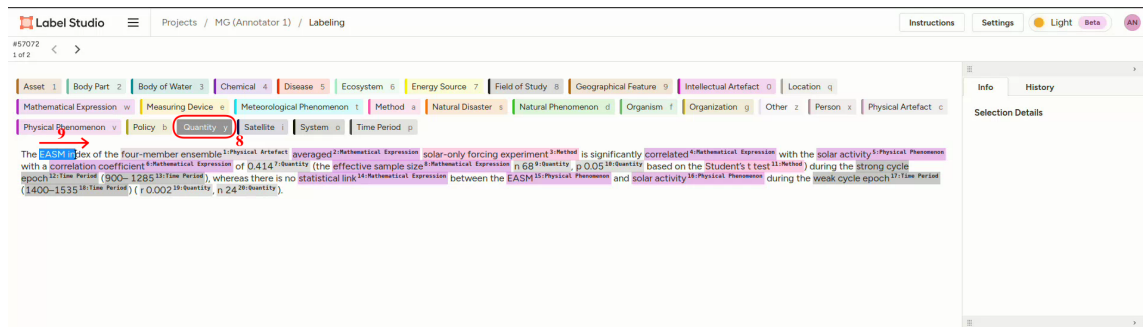


Figure 6: Annotation interface - Annotation.



Figure 7: Annotation interface - Annotation final.

7 Useful Tips

- When annotating, it is useful to familiarise yourself with frequent examples of a given category. For this, see Figures 8 to 13.
- When annotating, it is useful to open the article from which the sentence originates. This can be done by checking the sentence's `paper_id` and comparing it with the `paper_id` in Table 3 (pg. 36). **This is particularly useful for acronyms (e.g. *EASM*).**
- When annotating, it is useful to explore the meaning of a given term on the web (e.g. *snow albedo*). Using the search term "*What is snow albedo?*", one of the results is: "*Snow Albedo refers to the measure of how much sunlight, or solar radiation, a snow-covered surface reflects back into space.*"⁴.

⁴<https://climate.sustainability-directory.com/term/snow-albedo/>

This is particularly useful when deciding on the entity category. Here, *snow albedo* is clearly a Quantity.

paper_id	Title	Link
1072	Spatiotemporal climate variability in the Andes of northern Peru: Evaluation of gridded datasets to describe cloud forest microclimate and local rainfall	Link
2416	Agro-climate changes over Northeast Asia in RCP scenarios simulated by WRF	Link
6641	Assessment of Non-Anthropogenic Addition of Uric Acid to a Water Treatment Wetlands	Link
10061	Effect of Na Doping on the Electrochemical Performance of Li _{1.2} Ni _{0.13} Co _{0.13} Mn _{0.54} O ₂ Cathode for Lithium-Ion Batteries	Link
10167	Clinical Observations Associated with Phenobarbital Serum Monitoring to Manage Epilepsy in a California Sea Lion with Domoic Acid Toxicosis	Link
10662	Design and Implementation of a Smart Lithium-Ion Battery System with Real-Time Fault Diagnosis Capability for Electric Vehicles	Link
10756	Air Pollution Tolerance Index and Heavy Metals Accumulation of Tree Species for Sustainable Environmental Management in Megacity of Lahore	Link
12778	A Correlation-Scale-Threshold Method for Spatial Variability of Rainfall	Link
13006	Assessment of Physicochemical Properties and Comparative Pollution Status of the Dhaleshwari River in Bangladesh	Link
13798	Analytical Study for the Determination of the Energy Use Potential of Forest Dendromass in the Czech Republic	Link
16783	A Green Quality Management Decision Model with Carbon Tax and Capacity Expansion under Activity-Based Costing (ABC)—A Case Study in the Tire Manufacturing Industry	Link
16811	Synthesis and Characterization of Epoxidized Beechwood Pyrolysis Bio-Oil as a Curing Agent of Bio-Based Novolac Resin	Link
20390	Uncertainty Estimation and Evaluation of Shallow Aquifers' Exploitability: The Case Study of the Adige Valley Aquifer (Italy)	Link
21622	Bottled Mineral Waters as Unconventional Sampling in Hydro-Geological Research	Link
22971	Optimization of Solid-State Fermentation of Switchgrass Using White-Rot Fungi for Biofuel Production	Link
27835	Assessing Climate Driven Malaria Variability in Ghana Using a Regional Scale Dynamical Model	Link
28638	No City Left Behind: Building Climate Policy Bridges between the North and South	Link
28901	Snow Surface Albedo Sensitivity to Black Carbon: Radiative Transfer Modelling	Link
30003	Effects of Climate and Drought on Stem Diameter Growth of Urban Tree Species	Link
32771	Decomposition of Used Tyre Rubber by Pyrolysis: Enhancement of the Physical Properties of the Liquid Fraction Using a Hydrogen Stream	Link
33310	Coastal Quarries as Relative Sea-Level Markers: A Methodological Approach Applied in the Apulia Region (Southern Italy)	Link
35800	Estimating the Statistical Significance of Cross-Correlations between Hydroclimatic Processes in the Presence of Long-Range Dependence	Link
36618	Development and Evaluation of Sustainable Bituminous Paver Blocks	Link
39322	Evaluating Possible Changes in Air Temperature and Precipitation Patterns in Mozambique by Comparing Present and Future RegCM4 Simulation	Link
40290	Comparison between Snow Albedo Obtained from Landsat TM, ETM+ Imagery and the SPOT VEGETATION Albedo Product in a Mediterranean Mountainous Site	Link
41282	Circulation of ¹³⁷ Cs in Various Forest Plants in the Chernobyl Exclusion Zone during the Year	Link
45266	Spatial Distribution of Snow Cover in Tibet and Topographic Dependence	Link
46034	Genus-Physiognomy-Ecosystem (GPE) System for Satellite-Based Classification of Plant Communities	Link
48798	Ammonia Cycling and Emerging Inorganic Secondary Aerosols from Arable Agriculture	Link
50441	Food Waste Digestate-Based Biorefinery Approach for Rhamnolipids Production: A Techno-Economic Analysis	Link
53166	Analyzing How Governance of Material Efficiency Affects the Environmental Performance of Product Flows: A Comparison of Product Chain Organization of Swedish and Dutch Metal Packaging Flows	Link
55512	Comparison of Satellite Rainfall Estimates and Rain Gauge Measurements in Italy, and Impact on Landslide Modeling	Link
56596	Analysis of Daytime Cloud Fraction Spatiotemporal Variation over the Arctic from 2000 to 2019 from Multiple Satellite Products	Link
62269	Decadal Variations of the East Asian Summer Monsoon Forced by the 11-Year Insolation Cycle	Link
63539	Recovered frog populations coexist with endemic <i>Batrachochytrium dendrobatidis</i> despite load-dependent mortality	Link
65935	Regional mapping of species-level continuous foliar cover: beyond categorical vegetation mapping	Link
71011	Determination of tropical deforestation rates and related carbon losses from 1990 to 2010	Link
71538	Warmer and browner waters decrease fish biomass production	Link
72508	Substantial contribution of iodine to Arctic ozone destruction	Link
72553	Oxygenation of the Baltoscandian shelf linked to Ordovician biodiversification	Link
75726	Impact of Atmospheric Retrievals on Hurricane Florence/Michael Forecasts in a Regional NWP Model	Link
84006	Ambient Aerosol Hygroscopic Growth From Combined Raman Lidar and HSR	Link
86033	A simple decomposition of European temperature variability capturing the variance from days to a decade	Link
87857	Quasi-biweekly oscillation of the Asian monsoon rainfall in late summer and autumn: different types of structure and propagation	Link
90439	Intrinsic nonlinear dynamics drive single-species systems	Link
99773	Probing both sides of the story	Link
170785	Recent reduced abyssal overturning and ventilation in the Australian Antarctic Basin	Link
170943	Slowdown of Antarctic Bottom Water export driven by climatic wind and sea-ice changes	Link
171290	Mitigating poverty: The patterns of multiple carbon tax and recycling regimes for Peru	Link
171401	Zero carbon homes: Perceptions from the UK construction industry	Link

Table 3: Paper id mapping to research articles.

8 Statistics

Top 20 Most Frequent Entities per Label (FDEL_annot)

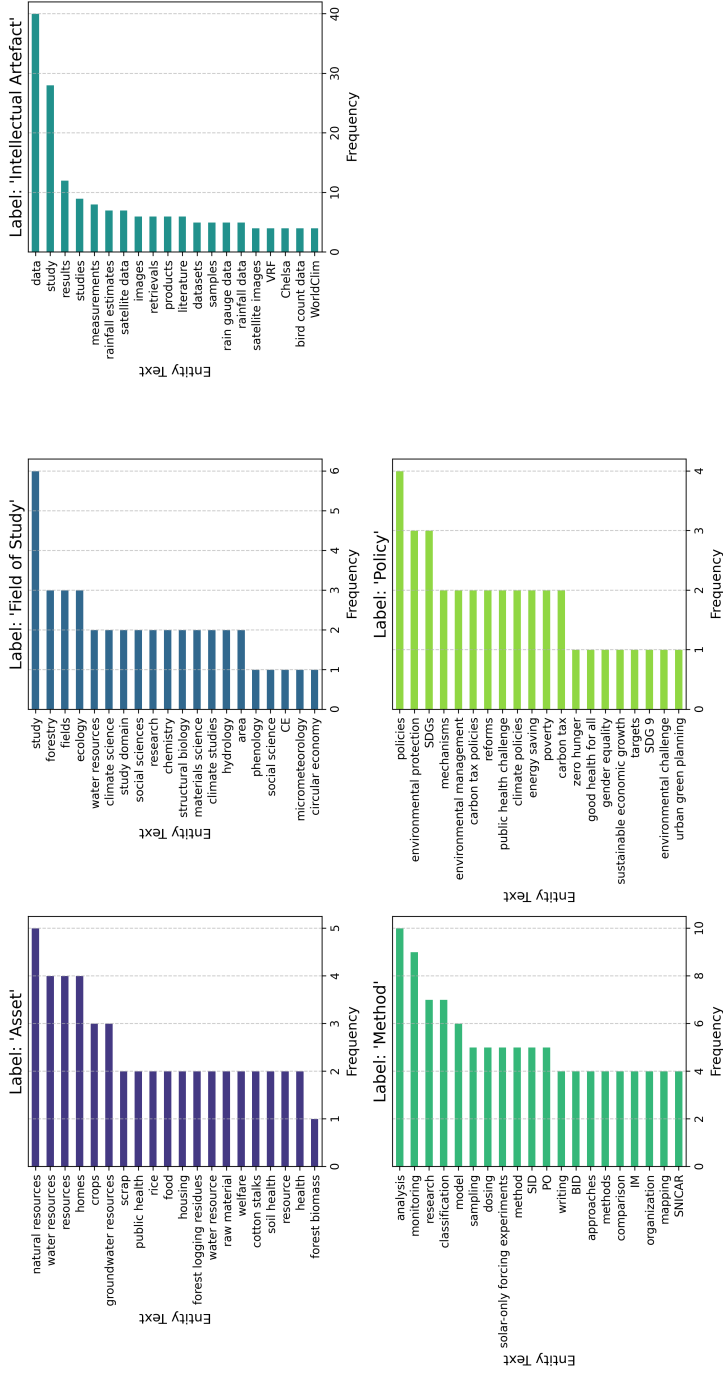


Figure 8: Entity category top 20 - Group 1

Top 20 Most Frequent Entities per Label (FDEL_annot)

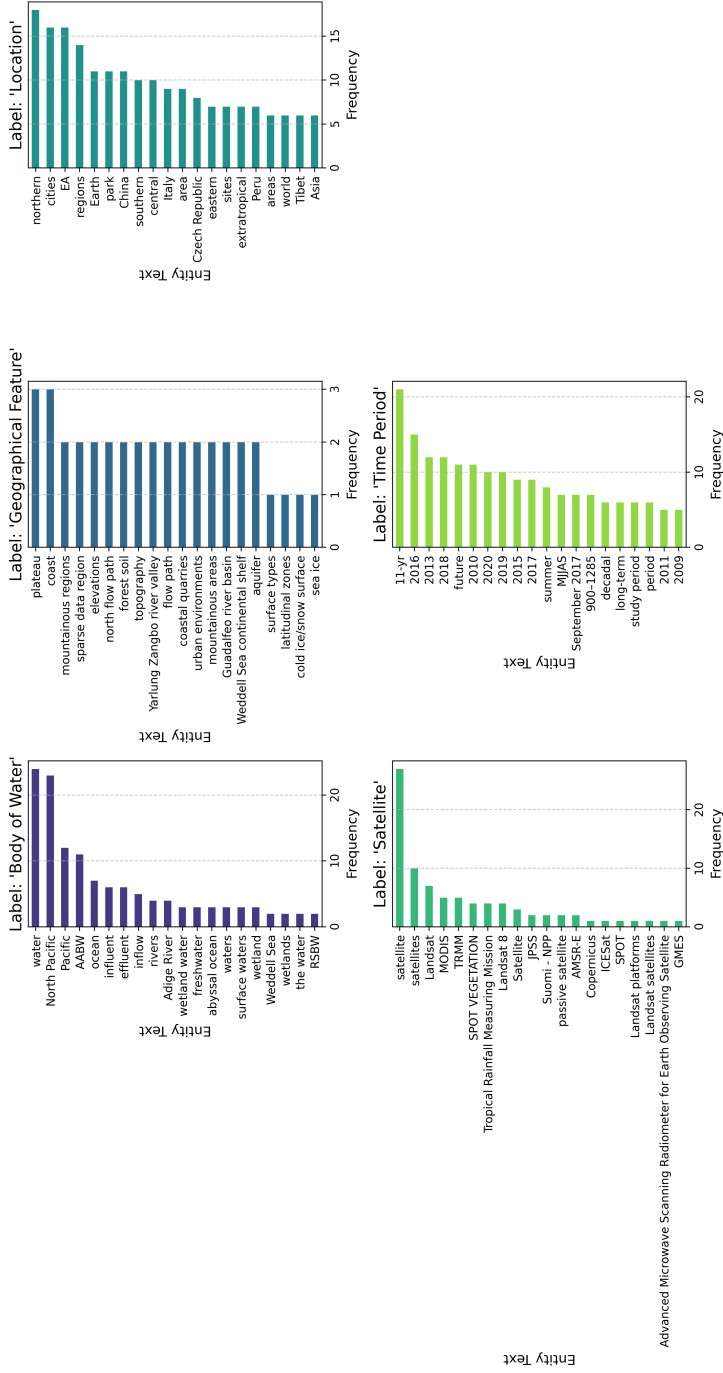


Figure 9: Entity category top 20 - Group 2

Top 20 Most Frequent Entities per Label (FDEL_annot)

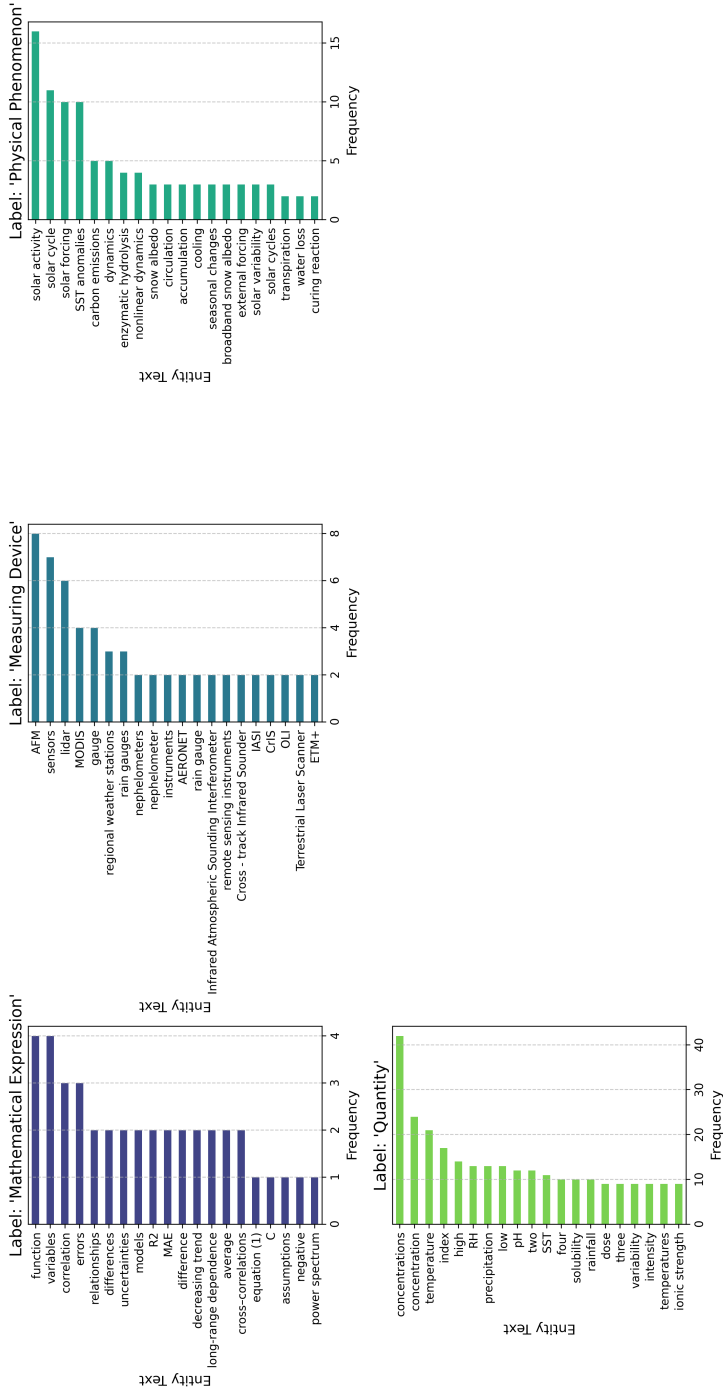


Figure 10: Entity category top 20 - Group 3

Top 20 Most Frequent Entities per Label (FDEL_annot)

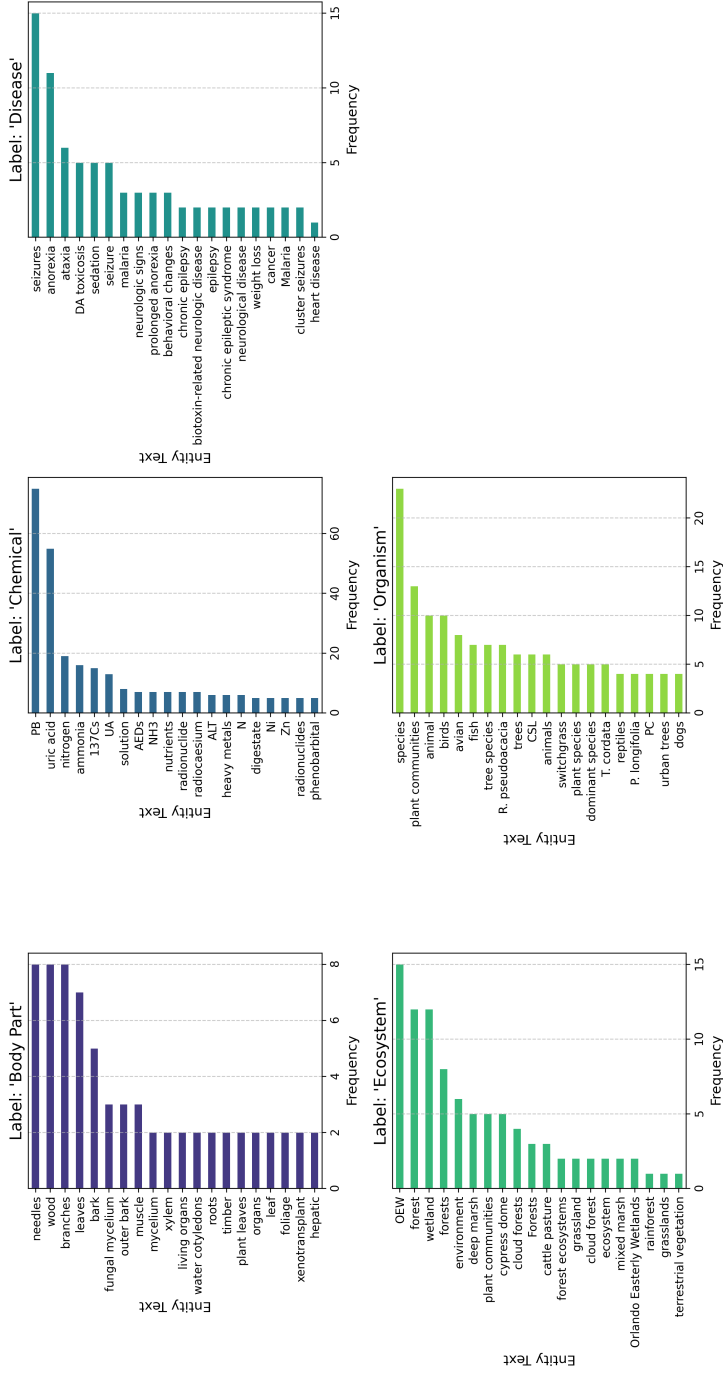


Figure 11: Entity category top 20 - Group 4

Top 20 Most Frequent Entities per Label (FDEL_annot)

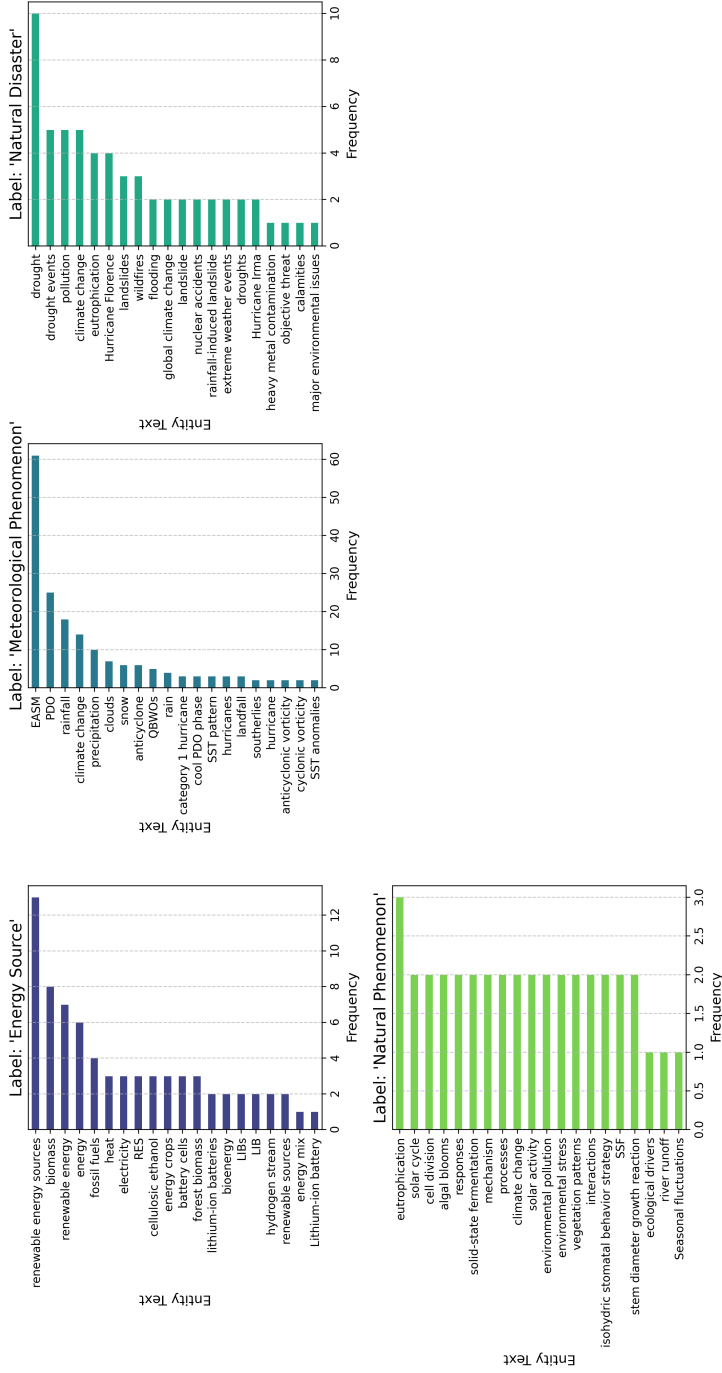


Figure 12: Entity category top 20 - Group 5

Top 20 Most Frequent Entities per Label (FDEL_annot)

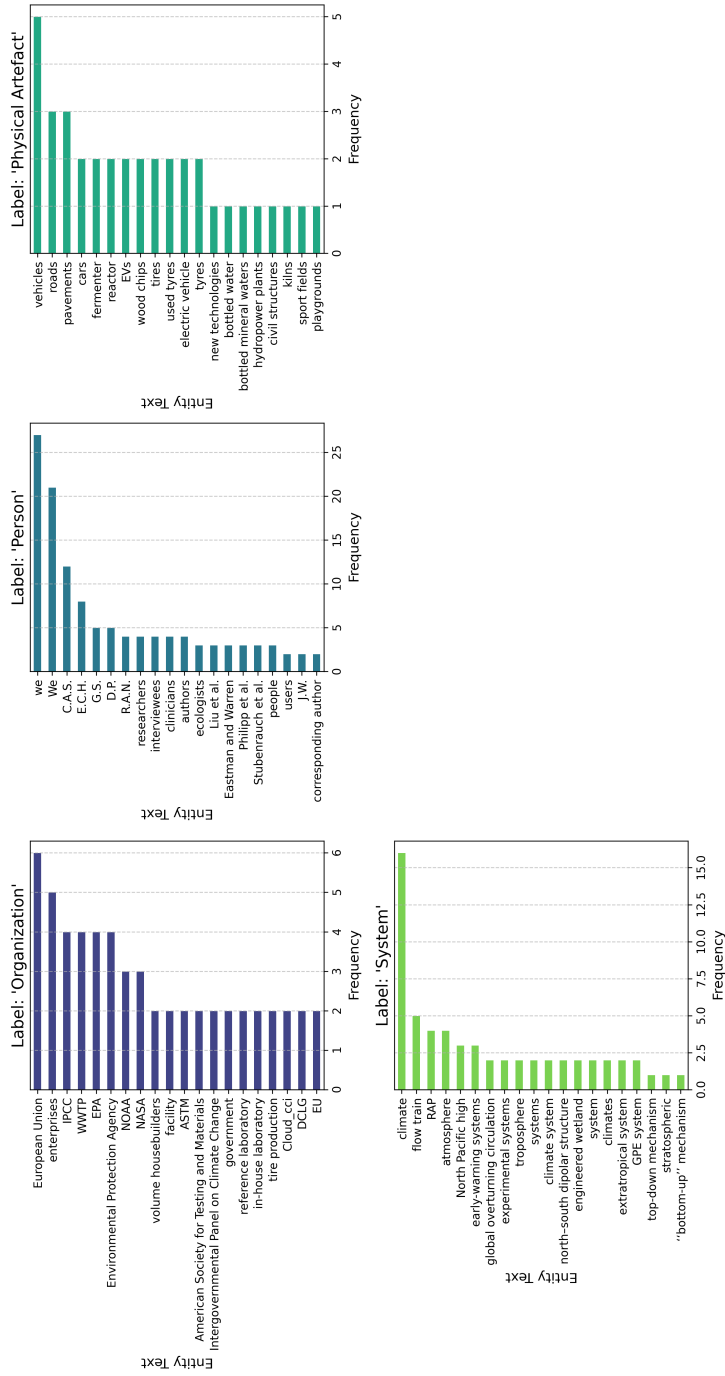


Figure 13: Entity category top 20 - Group 6

9 Model Inputs

An **Asset** is an object or service of value to humans that can get destroyed or diminished by climate disasters/hazards. Key categories are health, buildings, infrastructure, and crops or livestock.

A **Body of Water** is a distinct volume or mass of water, whether naturally occurring (like an ocean or river) or contained within a man-made system (like a reservoir or effluent flow). This includes specific named bodies, general types, and scientifically defined water masses.

A **Body Part** is a structural component of a living organism (plant, animal, or human), which is not the whole organism but a distinct anatomical or morphological part.

A **Chemical** is a substance with a distinct molecular composition, including elements, compounds, ions, biological molecules, and mixtures. It also encompasses classes of substances, reagents, and terms describing their chemical state or form.

A **Disease** is an abnormal condition, disorder, or pathological state affecting an organism. This includes specific diseases, syndromes, symptoms, signs, injuries, toxicities, and observed pathological changes.

An **Ecosystem** is a community of interacting organisms and their physical environment, or a term used to describe such a community. This includes specific named ecosystems, general types of habitats, and collective terms for the biological components that define them.

An **Energy Source** is a substance, material, natural phenomenon, or engineered system that provides or stores usable energy. This includes fuels, renewable resources, electricity, heat, and energy storage technologies.

A **Field of Study** is a branch of knowledge, a specific academic or scientific discipline, or a theoretical domain of inquiry.

A **Geographical Feature** is a natural or man-made feature of the Earth's solid surface or subsurface, including landforms, topographical structures, and defined regions of land.

An **Intellectual Artefact** is a man-made product, typically of an informational, intellectual, or representational nature, resulting from scientific research, data collection, or modeling.

A **Location** is a point, area, or region in physical space, defined by natural, administrative, cultural, or geopolitical boundaries or functions. It may occur on Earth or elsewhere in the universe.

A **Mathematical Expression** is a term, symbol, equation, or concept related to mathematics, statistics, or formal modeling. This includes formulas, variables, statistical metrics, mathematical operations, and descriptions of quantitative relationships or trends.

A **Measuring Device** is an instrument, apparatus, technology, or system used to obtain quantitative measurements of a physical, biological, or chemical property.

A **Meteorological Phenomenon** is a weather- or climate-related event, process, or pattern occurring in the atmosphere and its interaction with the Earth's surface, particularly oceans. This includes specific weather events, large-scale climate oscillations, atmospheric circulation patterns, and long-term trends.

A **Method** is a systematic procedure, technique, model, experiment, or approach used to conduct research, perform an analysis, collect data, or achieve a specific practical or scientific outcome.

A **Natural Disaster** (or Hazard) is a large-scale adverse event, arising from natural, environmental, or climatic processes, that has the potential to cause significant harm or disruption to ecosystems, the environment, or human well-being.

A **Natural Phenomenon** is an observable event or process that occurs in the natural world, driven by biological, physical, or chemical principles. It encompasses interactions, cycles, transformations, and large-scale environmental changes.

An **Organism** is an individual living being (animal, plant, fungus, microbe) or a term that refers to a species, a taxonomic group, or a collection of individuals.

An **Organization** is a formally constituted group, such as a company, institution, government agency, laboratory, or association, with a specific, collective purpose.

A **Other** is a miscellaneous category for named entities that do not fit into any of the more specific, defined types. It often includes abstract concepts, generic objects, textual references, and descriptive attributes that have been isolated as entities.

A **Person** is an individual or a group of human beings, identified by name, profession, or through a personal pronoun.

A **Physical Artefact** (or Object) is a tangible, material entity with physical form. This includes both man-made products, such as tools, machines, vehicles, or buildings, and natural objects, such as soil, rocks, leaves, plant samples, powders, or other collected specimens.

A **Physical Phenomenon** is an observable event or property that is not a substance, but rather a manifestation of energy, force, temperature, or state. It encompasses a wide range of processes from molecular interactions to large-scale environmental dynamics.

A **Policy or Objective** is a defined goal, aim, or challenge that guides actions, or a framework of principles and rules to achieve those goals. This class encompasses high-level aims, specific targets, strategic plans, and guiding principles for management and governance.

A **Quantity** is a measurement, count, property, or value that is quantifiable. This includes numerical values with units, descriptive terms of amount or degree, statistical measures, and abstract quantifiable properties.

A **Satellite** is a man-made object, spacecraft, or instrument platform placed into orbit around a celestial body, primarily for purposes like remote sensing, data collection, or communication.

A **System** is a set of interacting or interdependent components that form an integrated whole. This includes large-scale natural systems, engineered constructs, and abstract or theoretical frameworks that describe a set of interactions.

A **Time Period** is a specific point in time, a duration, or a recurring interval. This includes dates, years, seasons, epochs, and terms describing frequency or temporal extent.