Annotation Guideline

STAGE ONE: Named Entity Recognition

1. Introduction

Purpose of the Manual:
This manual provides detailed instructions for annotating climate-related text or terms extracted from scientific literature. It aims to ensure consistency and accuracy in labelling climate entities, data, and models.
Intended Audience:
The guidelines are designed for annotators, including researchers, climate analysts, scientists, and students, who are familiar with climate science terminology and concepts.
Scope of Annotations:
The annotations focus on specific climate entities, including but not limited to:

- **Earth Systems**: Land, ocean, atmosphere, and biosphere entities.
- **Climate Data**: Specific datasets and measurements.
- **Climate Models**: Global and regional climate models.

2. Definitions and Examples of Key Climate Entities

2.1 Earth Systems
Land:
Refers to a specific region or unit of land that can be described and modeled geographically within the framework of a climate model. **Examples**:

- Continents/Regions: Africa, Ethiopia, United Kingdom (UK), high/mid-latitudes, tropics (tropical regions).
- Land Features: Groundwater, river flow, runoff, streamflow, land cover, land use.
- Specific Landmarks: Amazon Rainforest, Himalayas, United States Midwest (Corn Belt), Antarctica.

Atmosphere:
Refers to the layer of gases surrounding the Earth, which plays a vital role in shaping climate and weather patterns and can be modeled geographically within the framework of a climate model.
Examples:

- Atmospheric Layers: Troposphere, mesosphere.
- Climate Phenomena: Temperature, precipitation, wind, evapotranspiration, clouds.
- Weather Systems: Hadley Cells, Ferrel Cells, Trade Winds, Jet Streams, Monsoons, Intertropical Convergence Zone (ITCZ), El Niño-Southern Oscillation (ENSO), Tornadoes, Thunderstorms.

Oceans:
Refers to the large bodies of saltwater that cover about 71% of the Earth's surface and can be modeled geographically within the framework of a climate model. **Examples**:

- Oceans/Seas: Pacific Ocean, Indian Ocean, Atlantic Ocean.
- Oceanic Features: Gulf Stream, Kuroshio Current, Thermohaline Circulation.
- Climate-Related Ocean Phenomena: Ocean acidification, marine heatwaves, coral reefs, upwelling zones, sea ice, continental shelves.

2.2 Climate Data
Refers to detailed, quantitative measurements or simulations of variables that describe various components of the Earth's climate system. **Examples**:

- Datasets: CRU (Climate Research Unit), GPCC (Global Precipitation Climatology Centre), ERA5 (ECMWF Reanalysis 5th Generation).
- Climate Indices: HadCRUT, MERRA-2, GSMP3.

2.3 Climate Models
Refers to computational models used to simulate the Earth's climate system. **Examples**:
2.4 Global Climate Models (GCMs): CCSM4, CNRM-CM5, HadGEM2-ES.
2.5 Regional Climate Models (RCMs): MICRO, ACCESS-ESM1.5.

3. Key Tags or Labels

Guidelines for Tagging:

- Ensure the correct spelling and usage of tags. For example, use "Variables" consistently, not "Variable>" or other variations.
- Review definitions carefully and apply tags or values strictly based on the provided examples and their accurate definitions.
- If uncertain about the definition of an entity, verify its classification (e.g., variable, teleconnection) before tagging.

Tag	Definition and examples
Variable	represents a specific measurable element or attribute of the climate system that is studied or monitored (e.g., cloud cover, temperature (i.e., surface air, ocean, or groundwater), precipitation, wind speed, vapor pressure, geopotential height, humidity (relative, specific) etc.
Project	refers to a coordinated effort or initiative aimed at investigating specific aspects of climate. Projects often involve multiple stakeholders and produce datasets, models, or assessments (e.g., Coupled Model Intercomparison Project Phase 6 (CMIP6))
Location	refers to the geographic region or coordinates being studied or monitored. This can be global, regional, or local. Examples includes West Africa, Central Africa, East Africa, or Southern Africa; tropics or polar regions; high or mid latitudes regions, specific sites (such as the Amazon, Congo Rainforest or Sahara Desert etc).
Model	refers to computational tool used to simulate and predict climate processes and interactions in the Earth system (e.g., HadGEM3, WRF etc)
Provider	refers to the organization or agency responsible for creating, maintaining, or distributing climate data or tools (e.g., NASA (e.g., GISS for climate models, MERRA datasets); ECMWF (e.g., ERA5 reanalysis datasets); NOAA (e.g., NCEP datasets and climate services).
Instrument	refers to the device or tool used to measure climate variables. Instruments can be ground-based, airborne, or spaceborne. Examples includes Radiosondes (balloons for atmospheric measurements); Satellites (e.g., MODIS, GOES, or Sentinel); Rain gauges and anemometers for ground-level data.
Event	An event is an occurrence or phenomenon in the Earth's system that varies in temporal scale, ranging from short-term weather events lasting minutes to days to long-term climate events spanning decades or more. Examples include remote teleconnection such as ENSO, IOD, etc, droughts, floods, etc
Weather event	Weather events are meteorological occurrences that impact Earth's atmosphere and surface over short timescales (hours to days). Common Weather Events; Rainfall (e.g., Drizzle, showers, or steady rain), Snowfall (e.g., Light , or heavy); Thunderstorms (e.g., storms with lightning, thunder, heavy rain, and hail), Wind Events (e.g., breezes, gusts, and strong winds), Cloud Cover (e.g., Clear skies, partly cloudy, overcast), Temperature Changes (Heatwaves or cold snaps), Fog and Mist, Frost, Dew etc.

Natural Hazard	Natural hazards are phenomena with the potential to cause significant harm to life, property, and the environment. Teleconnection refers to large-scale patterns of climate variability that link weather and climate phenomena across vast geographic areas, influencing atmospheric conditions over long distances. Typical examples of hazards can be broadly classified into geophysical (e.g., earthquakes, volcanic eruptions, tsunamis, landslides), meteorological (e.g., cyclones or hurricanes or typhons, tornadoes, heatwaves), hydrological (e.g., floods, flash floods, drought, avalanches), biological (pandemics, plagues, animal borne diseases), and climatological (e.g., wildfires, frost, cold wave) categories.
Ocean circulation	Ocean circulation is the large-scale movement of water masses in the Earth's oceans, driven by wind, density differences, and the Coriolis effect, regulating Earth's climate. Key examples of ocean circulation, categorized into surface currents (Gulf Stream, Kuroshio Current, California Current, Canary Current, Equatorial Currents), deep ocean currents (North Atlantic Deep Water (NADW), Antarctic Bottom Water (AABW), Mediterranean Outflow Water, Indian Ocean Overturning), Global Ocean Circulation Systems (the Global Conveyor Belt, the Atlantic Meridional Overturning Circulation (AMOC).
Teleconnection	Teleconnection is a large-scale patterns of climate variability that link weather and climate phenomena across vast distances. Examples includes El Niño-Southern Oscillation (ENSO; (El Niño or La Niña), North Atlantic Oscillation (NAO), Arctic Oscillation (AO), Pacific Decadal Oscillation (PDO), Indian Ocean Dipole (IOD), Madden-Julian Oscillation (MJO), Atlantic Multi-Decadal Oscillation (AMO), Southern Annular Mode (SAM), Rossby Waves, Walker Circulation, Monsoonal Systems (i.e., Asian Monsoon and West African Monsoon)

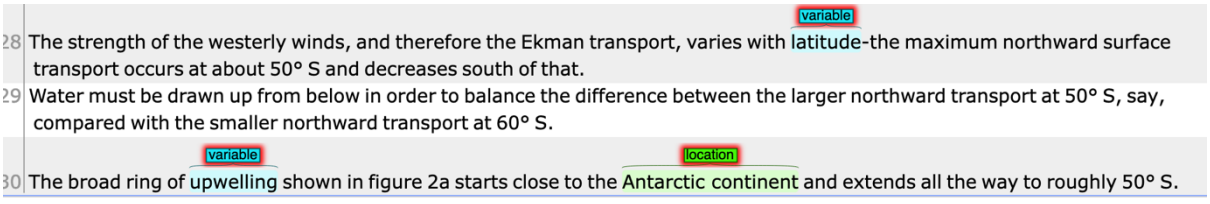
4. Example

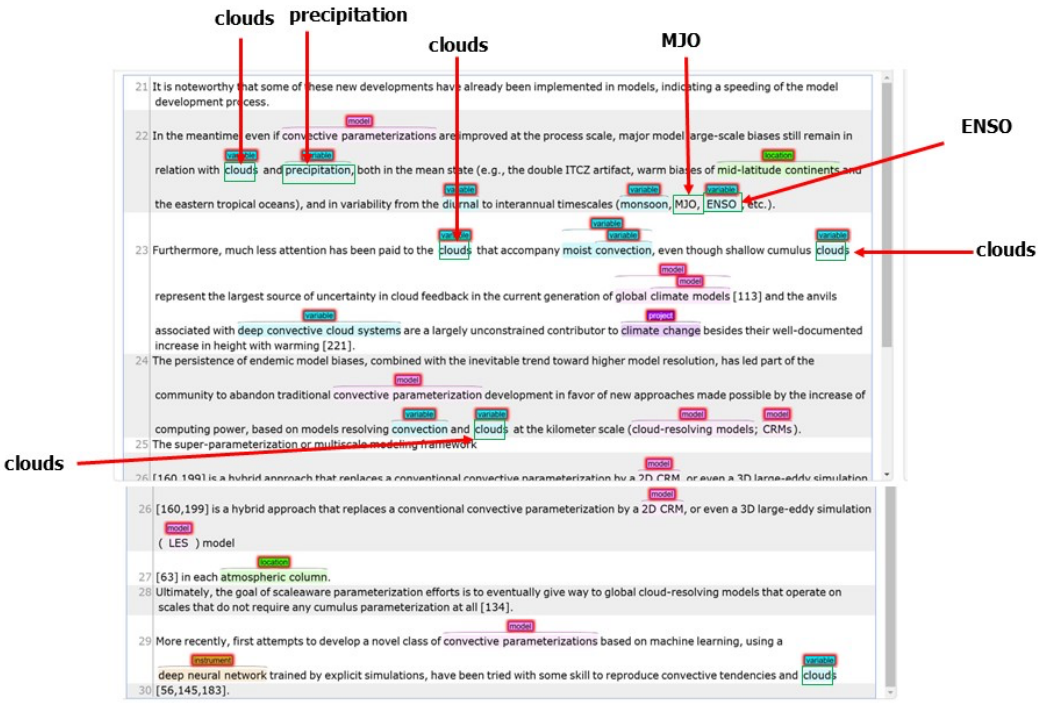
Example: *"This annotation manual aims to provide consistent methods for annotating climate data. Our primary focus is 09bdb7d909ed6615760571a6aa14051133179aee.xmi"*

Task one: see the scientific literature with serial number above.

Role of the annotator: The annotator is expected is to read each sentence carefully. Then, you are required to perform these tasks concurrently.

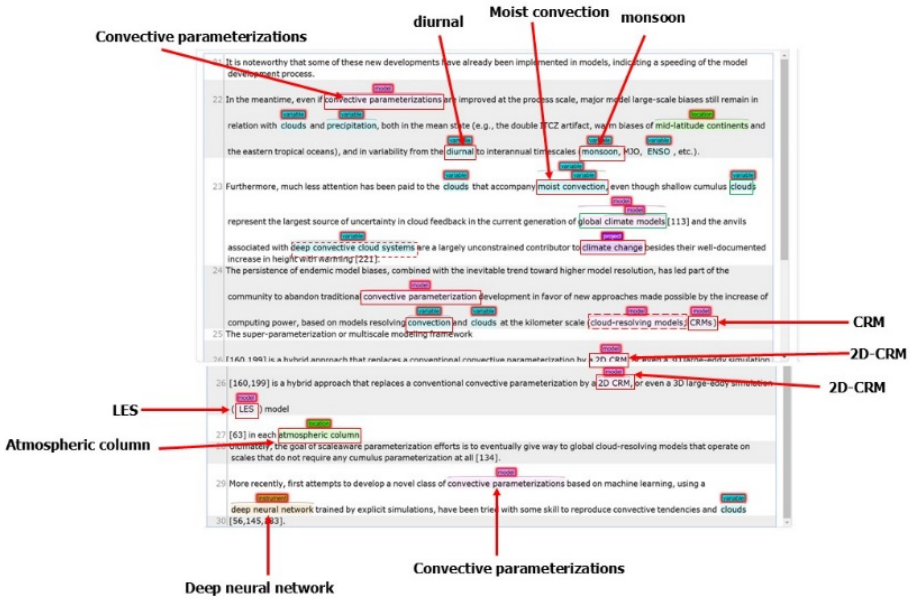
1. Verify specific pre-annotated climate entries of interest in line 22: (E.g., "clouds", "precipitation", "ENSO") and other scientific terms such as "mid-latitude continents". (see details below for more information).
2. Delete pre-annotated test that involves a "process" or "methods", "tools", frameworks, "instrument of measurements", "units of measurement", "temporal, threshold or range of values" (e.g., convective parameterisation, diurnal, monsoon (see details below for more information).
3. Annotate missing but relevant "un-annotated" text of interest (E.g., Westerly Winds) (see details below on how to annotate).

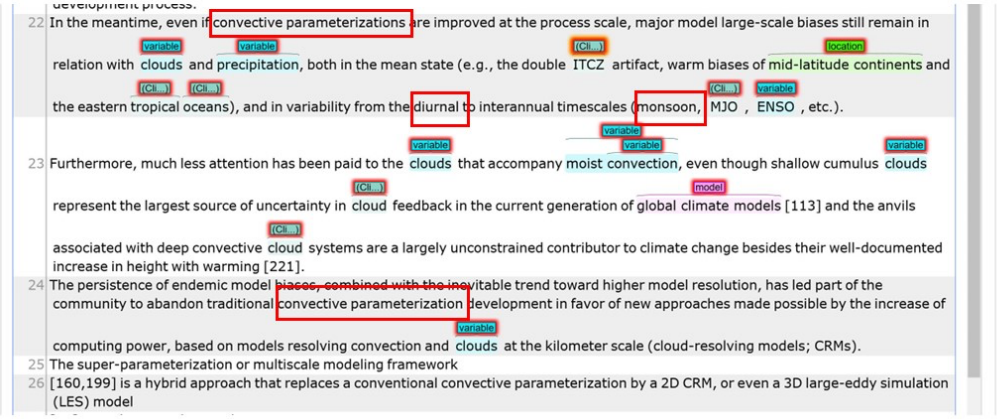




Other Scientific Terms: You may find other climate variables such as temperature, wind speed or wind, sea surface temperature or SST; rainfall, cyclones, aerosols, etc

Delete wrongly pre-annotated climate entities. These may include but not limited to methods, materials, processes, units of measurements, threshold, or range of values, etc
Units of Measurement: (e.g., Celsius for temperature, mm for rainfall, km/h for wind speed).
Thresholds and Ranges: Values or thresholds or ranges. E.g., 10°C for temperature or mm for precipitation."
Standardization: standardizing annotations across climate entities. For example, temperature (delete prefix “minimum or min”, “maximum or max”, “nighttime”, “daytime” for temperature annotations to ensure consistency (e.g. minimum temperature to temperature).
Other Scientific Terms: Phrases that are a scientific term but do not fall into any of the above classes E.g. diurnal, interannual,





STAGE TWO: Entity Linking

1. Tag Selection Guidelines

- **Allowed Tags**: Only the following values should be selected as tags. Do not type any tags manually; only select from the provided list: project, location, model, experiment, platform, instrument, provider, variable, weather event, natural hazard, teleconnection, ocean circulation
- **Spelling and Formatting**:
    - Ensure all tags are in **lowercase**.
    - Do not use uppercase letters or modify the spellings in any way.
    - If you encounter any foreign or unrecognized tags, do not use them.

2. Annotation Setup

- Open **two tables** simultaneously:
    1. **Annotation Table**: The document or interface where you are performing the annotations.
    2. **Knowledge Base Table**: A reference table or database containing entity identifiers and their corresponding information.

- Use the knowledge base to search for and verify the correct identifiers for each entity. Make sure to check if the definitions and the path match the semantic meaning.

3. Task Description

- **Objective**: Link each entity in the text to its corresponding identifier in the knowledge base.
- **Steps**:
    1. Identify the entity in the text.
    2. Double check the tag from the allowed list (e.g., location, variable, etc.).
    3. Search the knowledge base to find the correct identifier for the entity.
    4. Link the entity to its identifier in the annotation table.

4. Quality Assurance

- Double-check the spelling and formatting of tags.
- Ensure that all entities are linked to the correct identifiers in the knowledge base.
- If an entity cannot be found in the knowledge base, flag it for review rather than making an assumption.

STAGE THREE: Relationship

1. Relationship Types and Definitions

Below are the relationship types to be annotated, along with their definitions and examples. Ensure that you correctly identify the **source entity** and **target entity** for each relationship.

1. **ComparedTo**
    - **Definition**: The source entity is compared to the target entity.
    - **Example**: A climate model, experiment, or project (source entity) outputs data (target entity).
    - **Template**: [Source Entity] ComparedTo [Target Entity]
2. **RunBy**
    - **Definition**: Experiments or scenarios (source entity) are run by a climate model (target entity).
    - **Example**: An experiment (source entity) is executed by a climate model (target entity).
    - **Template**: [Source Entity] RunBy [Target Entity]
3. **ProvidedBy**
    - **Definition**: A dataset, instrument, or model (source entity) is created or managed by an organization (target entity).
    - **Example**: A dataset (source entity) is provided by a research organization (target entity).
    - **Template**: [Source Entity] ProvidedBy [Target Entity]
4. **ValidatedBy**
    - **Definition**: The accuracy or reliability of model simulations (source entity) is confirmed by datasets or analyses (target entity).
    - **Example**: A climate model simulation (source entity) is validated by observational data (target entity).
    - **Template**: [Source Entity] ValidatedBy [Target Entity]
5. **UsedIn**
    - **Definition**: An entity, such as a model, simulation tool, experiment, or instrument (source entity), is utilized within a project (target entity).
    - **Example**: A climate model (source entity) is used in a research project (target entity).
    - **Template**: [Source Entity] UsedIn [Target Entity]
6. **MeasuredAt**

- **Definition**: A variable or parameter (source entity) is quantified or recorded at a geographic location (target entity).
- **Example**: Temperature data (source entity) is measured at a specific weather station (target entity).
- **Template**: [Source Entity] MeasuredAt [Target Entity]

7. **MountedOn**
   - **Definition**: An instrument or measurement device (source entity) is physically attached or installed on a platform (target entity).
   - **Example**: A weather sensor (source entity) is mounted on a satellite (target entity).
   - **Template**: [Source Entity] MountedOn [Target Entity]
8. **TargetsLocation**
   - **Definition**: An experiment, project, model, weather event, natural hazard, teleconnection, or ocean circulation (source entity) is designed to study, simulate, or focus on a specific geographic location (target entity).
   - **Example**: A climate model (source entity) targets the Amazon Rainforest (target entity).
   - **Template**: [Source Entity] TargetsLocation [Target Entity]

2. Annotation Instructions

1. **Identify Entities**:
   - Clearly identify the **source entity** and **target entity** in the text.
   - Ensure that both entities are correctly tagged (e.g., model, location, variable, etc.) before annotating the relationship.
2. **Select Relationship Type**:
   - Choose the most appropriate relationship type from the list above based on the context.
   - Refer to the definitions and examples to ensure accuracy.
3. **Annotate the Relationship**:
   - Use the provided templates to annotate the relationship between the source and target entities.
   - Double-check that the relationship type aligns with the context of the text.
4. **Verify Consistency**:
   - Ensure that the relationship annotation is consistent with the definitions and examples provided.
   - If unsure, consult the knowledge base or flag the relationship for review.