Official NER Annotation Guidelines

Task Type: Span-based Named Entity Recognition (NER)
Tooling: Label Studio (manual annotation)

1. Task Description
Annotators are required to identify and label named entities in short, mixed-domain texts. Each entity is annotated as a text span and assigned a label from a fixed schema. The goal is to produce clean, consistent, machine-readable annotations suitable for training or evaluating NLP models.

2. Entity Schema
Only the following entity types are allowed:

Person - *Named individuals (full proper names only)*
Organization - *Companies, institutions, government bodies, media outlets*
Location - *Geographical places (cities, countries, regions, landmarks)*
Date - *Absolute or relative temporal expressions*

No additional labels may be introduced.If an entity does not clearly fit one of these categories, it must not be labeled.

3. General Annotation Rules
3.1 Span Precision
- Annotate only the exact entity name
- Do not include leading or trailing spaces
- Do not include surrounding descriptive words
    - Correct:
        - Donald Trump
        - University of California, Berkeley
    - Incorrect:
        - President Donald Trump
        - the University of California, Berkeley

3.2 One Entity = One Span
- Multi-word proper names are labeled as a single span
- Do not split entities unless explicitly required by the schema

4. Entity-Specific Guidelines
4.1 Person
Label only explicitly named individuals.
- Include:
    - Barack Obama
    - Bill Gates
- Exclude:
    - Titles without names (the President, the Prime Minister)
    - Job titles (CEO, engineer)

4.2 Organization
Label organizations when they are referenced as entities, not merely as descriptors.
- Include:
    - Google
    - CNN
    - Ministry of Health of Poland
- Exclude:
    - Generic nouns (company, organization, startup)
    - Organization names used only inside titles

4.3 Location

Label geographical locations

Include:

Cities (Paris)

Countries (Albania)

Regions (Asia)

Landmarks (Brandenburger Gate)

Exclude:

Non-geographical institutional references unless clearly locational in context

4.4 Date

Label both absolute and relative temporal expressions, unless explicitly restricted.

Include:

Absolute dates (2019, December)

Relative expressions (yesterday, last Friday, this week, next year)

When relative dates are labeled, annotators may add a comment clarifying the decision.

5. Ambiguity & Context Handling

When an entity could reasonably be interpreted in more than one way:

- Use sentence context to determine the most appropriate label
- Prefer under-labeling over speculative labeling
- Add a short annotation comment if helpful

6. Annotation Comments

Comments are optional and should be added only when useful, such as:

- Explaining exclusions
- Justifying schema decisions
- Clarifying ambiguous cases

Comments should:

- Be concise
- Explain why, not what
- Avoid restating entity names

Example:

"Title excluded; only proper person name labelled."

7. Quality Control Principles

Annotators should aim for:

- Consistent application of rules
- Clean span boundaries
- Minimal but meaningful comments
- Realistic annotation pace reflecting careful human judgment