

Version française — Instructions d’annotation K2K

1. Contexte

Pour chaque requête (*query*), vous aurez accès à :

- La requête utilisateur.
- La réponse du système.
- La liste des fichiers/datasets retournés (titre, description, URL).
- Éventuellement des éléments de contexte ou d’ontologie.

Votre rôle : évaluer la qualité du résultat sur plusieurs dimensions : **pertinence, clarté, fidélité, utilité**.

2. Champs à remplir

Champ	Valeurs possibles	Règle
ID de la requête	fourni	—
ID annotateur	texte	Votre identifiant
Top dataset pertinent ?	Oui / Non / N/A	Oui si le 1er dataset est pertinent
Précision@5	0..5	Nombre de datasets pertinents dans les 5 premiers
Pertinence	1–5	1 = pas du tout pertinent, 5 = parfaitement pertinent
Clarté	1–5	Facilité de compréhension, structure
Utilité	1–5	Actionnabilité de la réponse
Fidélité	1–5	Appui factuel dans les sources fournies
Comportement de clarification	Oui / Non / N/A	Si info manquante, demande pertinente ?
Faits manquants/incorrects	texte	Résumé court
Type d’erreur majeure	Hallucination / Dataset erroné / Téléchargement manquant / Changement de langue / Peu clair / Autre	—
Confiance	1–5	Votre certitude dans l’annotation
Temps passé (min)	nombre	Optionnel

3. Définitions clés

Pertinence : un dataset est pertinent s'il couvre la variable, la zone et la période demandées.

Clarté : lisibilité, absence d'ambiguïté, structure logique.

Utilité : liens directs, étapes claires, explications exploitables.

Fidélité : exactitude des affirmations, appui dans les sources ou fichiers.

Clarification : détection des manques et demande d'info pertinente.

4. Cas particuliers

- **Langue** : la réponse doit rester dans la langue de la requête.
 - **Correspondance partielle** : noter en commentaire et ajuster la note.
 - **Pas de fichiers trouvés** : pertinence moyenne possible si guidage clair.
 - **URLs payantes/inaccessibles** : considérer comme absence de fichier téléchargeable.
 - **Doublons** : à compter une seule fois dans Precision@5.
-

5. Checklist rapide

- Pertinence de la réponse → oui/non + score
 - Exactitude des datasets (variable / zone / période)
 - Téléchargeable ?
 - Justification dans les sources ?
 - Clarification demandée si nécessaire ?
-

English version — K2K Annotation Guidelines

1. Context

For each *query*, you will see:

- The user query.
- The system's response.
- A list of returned datasets/files (title, description, URL).
- Optional: context chunks or ontology.

Your role: evaluate result quality along **relevance, clarity, faithfulness, usefulness**.

2. Fields to fill

Field	Possible values	Rule
Query ID	provided	—
Rater ID	text	Your identifier
Top dataset relevant?	Yes / No / N/A	Yes if 1st dataset is relevant
Precision@5	0..5	Relevant datasets among top 5
Relevance	1–5	1 = not relevant, 5 = fully relevant
Clarity	1–5	Readability, structure
Usefulness	1–5	Actionability of the answer
Faithfulness	1–5	Claims supported by provided sources
Clarification behavior	Yes / No / N/A	If info missing, was a useful question asked?
Missing/Incorrect facts	text	Short note
Major failure type	Hallucination / Wrong dataset / Missing download / Language mismatch / Unclear / Other	—
Confidence	1–5	Your certainty in the rating
Time spent (min)	number	Optional

3. Key definitions

Relevance: a dataset is relevant if it matches the requested variable, region, and time period.

Clarity: readability, absence of ambiguity, logical structure.

Usefulness: includes direct links, clear next steps, actionable info.

Faithfulness: factual accuracy supported by returned files/sources.

Clarification: detection of missing info and asking for it.

4. Special cases

- **Language:** reply should remain in the query's language.
 - **Partial match:** note in comments, adjust score.
 - **No files found:** mid relevance possible if clear guidance is given.
 - **Paid/inaccessible URLs:** treat as missing downloadable file.
 - **Duplicates:** count once in Precision@5.
-

5. Quick checklist

- Does the answer address the query?
 - Are datasets correct for variable / region / period?
 - Can files be downloaded?
 - Are claims supported by sources?
 - Was a clarification asked when needed?
-