

We cover all four datasets:

1. **arXiv 1912.05530 (Ontology Paper)**
2. **MHP Anxiety/Stress/Depression Dataset (Figshare)**
3. **Kaggle Sentiment Analysis for Mental Health**
4. **PMC Mental Health Studies (PubMed Central article)**

And for **each person**, we explain exactly how to leverage them.



Dataset Usage Guide – Per Team Member

✓ Team Member 1 – Ontology & Knowledge Graph

Engineer

This role builds the ontology, base KG, and dataset-aligned ontology annotations.

Dataset	How Person 1 Uses It
arXiv 1912.05530 (Ontology Paper)	Fundamental reference for ontology structure. Helps identify how mental health concepts and causal relationships are formally modeled in research. Useful for designing classes (Emotion, Symptom, Trigger, Risk patterns), properties, and causal relations. Use citations in your ontology documentation and report.
MHP Anxiety/Stress/Depression (Figshare)	Validates that your ontology captures real, structured mental health constructs like GAD-7, PHQ-9, and PSS-10 components. Person 1 can map scores → ontology concepts. Also helps define severity levels and duration properties in the ontology (e.g., thresholds for risk states).
Kaggle Sentiment Analysis	Use labeled text categories to ensure emotion and symptom classes in the ontology align with real expressive language. For example, the dataset may expose additional symptom phrases that should be included as synonyms or lexical variants in ontology annotations.
PMC Mental Health Article	Use this literature to add detailed definitions, citations, and causal relations in the ontology. For example, if PMC text discusses "sleep disruption → academic performance decline → anxiety," this can be structured as a relation in the KG.

Summary Tasks for Person 1:

- Refine class hierarchy using paper insights
 - Add causal properties informed by research
 - Annotate ontology with dataset-derived synonyms and causal strength hints
 - Document dataset → ontology alignments
-

Team Member 2 – NLP Extraction & Ontology

Mapping Engineer

This role builds the NLP pipeline for extracting emotions, symptoms, triggers, and mapping them to ontology URIs.

Dataset	How Person 2 Uses It
arXiv 1912.05530	 Not a text dataset, but the ontology design strategies in the paper help Person 2 choose which <i>features</i> and <i>semantic labels</i> the NLP should extract. It clarifies the types of concepts the extractor should focus on.
MHP Anxiety/Stress/Depression (Figshare)	 Use survey text and responses to validate extraction patterns. For example, students' answers like "I couldn't sleep" → should map to Insomnia. You can generate sentences from real survey data to test extraction accuracy.
Kaggle Sentiment Analysis	 Directly useful for training/validating concept extraction accuracy. You can use the labeled mental health categories to test your classifier's precision/recall or choose keywords + semantic similarity thresholds. This acts as a <i>development validation set</i> .
PMC Mental Health Article	 Helps expand your extraction lexicon with real clinical/psychological vocabulary (e.g., "anhedonia," "cognitive distortions"), which improves semantic mapping. You can also collect real symptom phrase patterns from these articles.

Summary Tasks for Person 2:

- Build and validate NLP extraction using labeled datasets
 - Create test sentences from real data
 - Expand lexicons using PMC research terms
 - Document confidence score heuristics
-

Team Member 3 – Reasoning & Explanation Engine

Engineer

This role builds the symbolic reasoning engine, causal explanation generator, ranking & confidence, and safety escalation logic.

Dataset	How Person 3 Uses It	
arXiv 1912.05530	Provides methodological insights into how ontologies support reasoning, especially causal links. Useful in the documentation and justification of your reasoning design, especially in the report.	
MHP Anxiety/Stress/Depression (Figshare)	Use this dataset to calculate conditional likelihoods or simple statistical associations between symptoms and risk states (e.g., P(anxiety)) Store these as :causalStrength annotations in the KG, and use them to rank explanation paths.	insomnia + stress)). Store these as :causalStrength annotations in the KG, and use them to rank explanation paths.
Kaggle Sentiment Analysis	Use it to evaluate extraction → reasoning consistency. If you know a sentence is labeled "anxiety," check that your rules infer anxiety risk. This can be part of your reasoning validation and explanation testing.	
PMC Mental Health Article	Use specific relations from PMC (e.g., "chronic stress → sleep disruption → impaired cognition") as rule templates or explanation examples. This can strengthen your SWRL rules and help build causal chains used in the explanation generator.	

Summary Tasks for Person 3:

- Annotate KG with causal strengths from MHP dataset
- Validate reasoning using Kaggle text
- Strengthen SWRL rule templates using clinical relations from PMC
- Document how dataset evidence influences explanation ranking

📌 How Each Dataset Helps the Project As a Whole

Dataset	Role	Primary Contribution
arXiv 1912.05530	1,2,3	Ontology design foundation + reasoning framing
MHP Anxiety/Stress/Depression	1,2,3	Ground truth validation, rule support, ranking signals
Kaggle Sentiment Analysis	2,3	NLP reliability evaluation + reasoning consistency checks
PMC Mental Health Studies	1,2,3	Rich vocabulary + causal relations for ontology & rules

Important Clarification (Team Alignment)

- None of these datasets drive your core symbolic reasoning.

Your **SWRL + ontology + causal rules** remain the authority.

- Datasets are used for validation, lexicon enrichment, annotation, and explanation quality.

They help you ensure the system behaves sensibly with real text and evidence.

- Statistical info from datasets is used only for ranking explanations (NOT for inferring risk states).

This protects the semantic purity of KRR design.

Example Paths of Usage (Concrete)

Person 1 Example

- From PMC article: "Sleep disturbance increases anxiety risk"
→ Encode as :SleepDisturbance leadsTo :AnxietyRisk

Person 2 Example

- From Kaggle: "I feel hopeless and can't focus"
→ Extract "hopelessness" & "lack of focus"
→ Map to ontology URLs with confidence values

Person 3 Example

- From Figshare:
GAD-7 high scores co-occur with stress + insomnia
→ Add :causalStrength 0.78 annotation
→ Use to rank explanation paths
-

Final Note for Each Team Member

Person 1

Your objective is to **capture the world of mental health as structured, causal, and explainable knowledge** – these datasets help you validate that world.

Person 2

Your task is to **translate user text into ontology concepts** – these datasets give you *real sentences and labels* to test and refine that mapping.

Person 3

Your job is to **reason and explain** – these datasets help you measure explanation quality, rank explanations, and align your system with real associations.
