# Violent Event Annotation Guidelines

## Training Data Collection Manual for Event Extraction System

**Version:** 1.0  
**Date:** October 2025  
**Project:** Knowledge Discovery from Free Text: Extraction of Violent Events  
**Prepared by:** Binalfew Kassa Mekonnen  
**Institution:** Addis Ababa University

## Table of Contents

1. Introduction and Overview
2. Annotation Workflow
3. Before You Start: Preparation
4. Entity-Level Annotation
5. Event-Level Annotation
6. Taxonomy Classification
7. Quality Control Procedures
8. Annotation Format Specifications
9. Fully Annotated Examples
10. Common Mistakes and How to Avoid Them
11. Training Exercises
12. Troubleshooting Guide
13. Appendices

## 1. Introduction and Overview

### 1.1 Purpose of This Manual

This manual provides detailed instructions for annotating news articles about violent events in Africa. Your annotations will create the training data that teaches the machine learning system to:

1. **Extract** violent events from news text
2. **Identify** the 5W1H attributes (Who, What, Whom, Where, When, How)
3. **Classify** events into hierarchical taxonomy categories

**Your work is critical** - the quality of your annotations directly determines the accuracy of the final system.

### 1.2 What You Will Annotate

For each news article, you will:

✅ **Identify** all violent events described  
✅ **Mark** entities (actors, victims, locations, dates)  
✅ **Extract** 5W1H attributes for each event  
✅ **Classify** each event using the taxonomy  
✅ **Document** relationships between events  
✅ **Note** confidence and ambiguities

### 1.3 Required Materials

Before starting annotation:

* This Annotation Guidelines document
* Violent Event Taxonomy document (reference)
* Annotation template (Excel/Google Sheets)
* News articles to annotate (provided)
* Access to Google Maps (for geocoding)
* Access to calendar (for date normalization)

### 1.4 Time Expectations

**Training Phase:** 4-6 hours (reading guidelines, practicing examples)  
**Qualification Test:** 1-2 hours  
**Per Article Annotation:** 15-45 minutes (depending on complexity)

**Target:** 8-12 articles per day once proficient

### 1.5 Annotation Philosophy

**Key Principles:**

1. **Annotate what the text says** - Not what you know from other sources
2. **Be consistent** - Use the same rules for all articles
3. **Document uncertainty** - Mark low confidence when unclear
4. **Ask for help** - Better to ask than guess incorrectly
5. **Quality over speed** - Accuracy is more important than quantity

## 2. Annotation Workflow

### 2.1 Overview of the Process

READ ARTICLE  
 ↓  
IDENTIFY EVENTS (How many violent events are described?)  
 ↓  
FOR EACH EVENT:  
 ↓  
1. MARK ENTITIES  
 - Actors (Who)  
 - Victims (Whom)  
 - Locations (Where)  
 - Dates/Times (When)  
 - Weapons/Methods (How)  
 ↓  
2. EXTRACT EVENT TRIGGER (violence-indicating word/phrase)  
 ↓  
3. FILL 5W1H TEMPLATE  
 - Who: [Actor]  
 - What: [Event type]  
 - Whom: [Victim]  
 - Where: [Location]  
 - When: [Date/Time]  
 - How: [Method/Weapon]  
 ↓  
4. CLASSIFY EVENT  
 - Level 1 category  
 - Level 2 category  
 - Level 3 category  
 - Level 4 category (if applicable)  
 ↓  
5. MARK RELATIONSHIPS (if applicable)  
 - Causality  
 - Retaliation  
 - Coordination  
 ↓  
6. QUALITY CHECK  
 - Review all fields  
 - Check consistency  
 - Note confidence  
 ↓  
NEXT EVENT

### 2.2 Detailed Step-by-Step Workflow

**STEP 1: Initial Read (2-3 minutes)** - Read the entire article once without annotating - Get overall understanding - Identify approximately how many events are described - Note the article’s publication date

**STEP 2: Event Identification (2-5 minutes)** - Re-read and identify distinct violent events - Mark event boundaries (which sentences describe which event) - Give each event a temporary ID (Event\_1, Event\_2, etc.) - Determine if article describes one event or multiple events

**STEP 3: Entity Annotation (5-10 minutes per event)** - For each event, mark all relevant entities - Highlight actor spans in text - Highlight victim spans - Highlight location mentions - Highlight temporal expressions - Highlight weapon/method mentions

**STEP 4: Attribute Extraction (3-5 minutes per event)** - Fill in 5W1H template - Extract specific values for each attribute - Normalize dates to absolute format - Geocode locations - Handle missing information

**STEP 5: Taxonomy Classification (2-3 minutes per event)** - Apply taxonomy to classify event - Start with Level 1 (broad category) - Proceed through Levels 2, 3, 4 - Document reasoning if ambiguous - Apply multi-labels if necessary

**STEP 6: Relationship Annotation (1-2 minutes per event)** - Identify relationships between events in same article - Mark temporal sequences - Mark causality if stated - Mark retaliation if stated - Mark coordination if stated

**STEP 7: Quality Review (2-3 minutes)** - Review all annotations for this article - Check for consistency - Verify all required fields completed - Note confidence scores - Flag for adjudication if very uncertain

**STEP 8: Save and Move On** - Save annotations in proper format - Move to next article

## 3. Before You Start: Preparation

### 3.1 Read These Documents First

**Priority 1 (Must Read):** 1. This Annotation Guidelines document (entire) 2. Violent Event Taxonomy document (Sections 1-4, 7)

**Priority 2 (Reference):** 3. Taxonomy Section 5 (Decision Rules) - Keep open while annotating 4. Taxonomy Section 6 (Examples) - Review similar events

### 3.2 Understand Key Concepts

**What is an “event”?** - A distinct violent incident with specific actors, victims, location, and time - Has clear beginning and end - One event = one combination of Who-What-Whom-Where-When-How

**Examples:** - ✅ “Gunmen attacked a village in Mali on Tuesday, killing 12 civilians” = **ONE EVENT** - ✅ “Rebels clashed with army in Gao on Monday; a separate attack occurred in Kidal on Tuesday” = **TWO EVENTS** (different times, different locations) - ✅ “Boko Haram bombed a market, then gunmen stormed the fleeing crowd” = **TWO EVENTS** (different methods, sequential actions)

**What is an “entity”?** - A real-world object mentioned in text: person, organization, location, date, weapon - Can be named (“Al-Shabaab”) or common noun (“gunmen”) - Same entity can be mentioned multiple times with different words (coreference)

**What is “coreference”?** - When different words refer to the same entity - Example: “Al-Shabaab attacked the base. The militants killed 10 soldiers. They then fled.” - “Al-Shabaab” = “The militants” = “They” (all refer to same actor)

### 3.3 Set Up Your Workspace

**Physical Setup:** - Quiet environment with minimal distractions - Computer with large screen (or dual monitors if possible) - Stable internet connection - Comfortable seating for extended work

**Digital Setup:** - Open this guideline document (for reference) - Open taxonomy document (for reference) - Open annotation template - Open article to annotate - Open Google Maps in separate tab - Have notepad for questions/notes

### 3.4 Understand Annotation Format

**We use a structured template with these main sections:**

ARTICLE INFORMATION  
- Article ID  
- Article URL  
- Publication Date  
- News Source  
- Annotator Name  
- Annotation Date  
  
EVENT RECORDS (One row per event)  
- Event ID  
- Event Trigger (word/phrase)  
- Sentence Numbers  
- Who (Actor)  
- What (Event Type)  
- Whom (Victim)  
- Where (Location)  
- When (Date/Time)  
- How (Method/Weapon)  
- Taxonomy L1  
- Taxonomy L2  
- Taxonomy L3  
- Taxonomy L4  
- Confidence Score  
- Notes  
  
ENTITY ANNOTATIONS (Separate section)  
- Entity Text  
- Entity Type  
- Character Span  
- Normalized Form  
- Coreference Group ID  
  
RELATIONSHIP ANNOTATIONS  
- Event 1 ID  
- Relationship Type  
- Event 2 ID

*(Detailed template specifications in Section 8)*

## 4. Entity-Level Annotation

### 4.1 Actor (Who) Annotation

**Definition:** The person, group, or organization that perpetrates the violence.

#### 4.1.1 What to Mark as Actor

**Include:** ✅ Named organizations: “Boko Haram,” “Al-Shabaab,” “M23 rebels”  
✅ Descriptive references: “armed men,” “gunmen,” “militants,” “insurgents”  
✅ State forces: “police,” “military,” “army,” “security forces”  
✅ Specific individuals: “suicide bomber,” “the assailant”  
✅ Ethnic/communal groups: “Fulani herders,” “ethnic militia”

**Exclude:** ❌ Inanimate objects: “the bomb,” “the explosion” (these are methods, not actors)  
❌ Passive constructions without clear actor: “12 were killed” (mark as Unknown Actor)

#### 4.1.2 How to Mark Actors

**Format:** Highlight the complete noun phrase

**Examples:**

✅ Good: “[Armed militants believed to be linked to JNIM] attacked…”  
❌ Too narrow: “[militants] attacked…”  
✅ Good: “[The rebel group’s fighters] stormed…”  
✅ Good: “[A suicide bomber] detonated…”

**Special Cases:**

**Pronouns and Coreference:**

Text: "Al-Shabaab attacked the base. The militants killed 10 soldiers.   
 They then fled."  
  
Annotation:  
- First mention: [Al-Shabaab] = Actor 1  
- Second mention: [The militants] = Actor 1 (coreference)  
- Third mention: [They] = Actor 1 (coreference)  
  
In template, record:  
- Actor: "Al-Shabaab"  
- All mentions: "Al-Shabaab, The militants, They"  
- Coreference Group: ACTOR\_1

**Multiple Actors in One Event:**

Text: "Ethnic militias and armed bandits jointly raided the village"  
  
Annotation:  
- Actor 1: [Ethnic militias]  
- Actor 2: [armed bandits]  
- Record both in template: "Ethnic militias; armed bandits"

**Unclear Actors:**

Text: "Gunmen attacked the market"  
  
Annotation:  
- Actor: [Gunmen]  
- Actor Type: Unknown/Unidentified  
- Note: "No further identification provided in article"

#### 4.1.3 Actor Normalization

After marking actor spans, normalize to standard form:

**Rules:** 1. Use full organization name if known: “Boko Haram” not “BH” 2. Expand acronyms on first mention: “Al-Qaeda in the Islamic Maghreb (AQIM)” 3. Use consistent naming: Always “Al-Shabaab” not sometimes “Shabaab” or “Al Shabab” 4. For descriptive terms, keep as-is: “gunmen,” “armed men,” “militants” 5. For state forces, specify country if clear: “Nigerian military” not just “military”

**Normalization Examples:**

| Original Text | Normalized Form |
| --- | --- |
| “The militants” | “Militants” |
| “AQIM fighters” | “Al-Qaeda in the Islamic Maghreb (AQIM)” |
| “Troops” | “[Country] military” (if country clear) |
| “Unknown assailants” | “Unknown attackers” |
| “Fulani herdsmen” | “Fulani herders” |

### 4.2 Victim (Whom) Annotation

**Definition:** The person(s), group(s), or entities that suffer from the violence.

#### 4.2.1 What to Mark as Victim

**Include:** ✅ Specific individuals: “the mayor,” “aid workers,” “journalist John Smith”  
✅ Groups: “civilians,” “protesters,” “worshippers,” “students”  
✅ Organizations: “the UN compound,” “the hospital”  
✅ Demographic descriptions: “women and children,” “displaced persons”  
✅ Numbers/casualties: “12 people,” “dozens of civilians”  
✅ Infrastructure as victim (when violence targets it): “the power plant,” “the bridge”

#### 4.2.2 How to Mark Victims

**Format:** Highlight complete noun phrase including numbers/descriptions

**Examples:**

✅ Good: “[at least 32 civilians including women and children]”  
✅ Good: “[the presidential palace]”  
✅ Good: “[government soldiers]”  
✅ Good: “[protesters]”

**Special Cases:**

**Casualty Numbers:**

Text: "The attack killed 15 civilians and injured 30 others."  
  
Annotation:  
- Victim: "15 civilians (killed); 30 civilians (injured)"  
- Deaths: 15  
- Injuries: 30  
- Victim Type: Civilian

**Mixed Victim Types:**

Text: "The bombing killed 5 soldiers and 8 civilians"  
  
Annotation:  
- Victims: "5 soldiers; 8 civilians"  
- Deaths: 13 total (5 combatants, 8 civilians)  
- Victim Type: Mixed (combatant and civilian)

**Infrastructure:**

Text: "Insurgents destroyed the water treatment plant"  
  
Annotation:  
- Victim/Target: "water treatment plant"  
- Victim Type: Infrastructure  
- Casualties: None mentioned  
- Impact: Infrastructure damage

**Implied Victims:**

Text: "The suicide bomber detonated at a crowded market"  
  
Annotation:  
- Victim: "Market-goers" or "People at market"  
- If casualties mentioned later: Update with specific numbers  
- If no casualties mentioned: Mark as "Potential victims (casualties unknown)"

#### 4.2.3 Victim Normalization

**Categories for Victim Type:** - Civilian (unarmed, non-combatant) - Combatant (military, police, armed group members) - Mixed (both civilian and combatant) - Infrastructure - Unknown

**Casualty Recording:** - Deaths: Number killed - Injuries: Number wounded/injured - Kidnapped/Abducted: Number taken - Displaced: Number forced to flee - Unknown: “Casualties reported but number unclear”

### 4.3 Location (Where) Annotation

**Definition:** The geographic location where the violence occurred.

#### 4.3.1 What to Mark as Location

**Include:** ✅ Countries: “Nigeria,” “Somalia,” “Democratic Republic of Congo”  
✅ Regions/States: “Borno State,” “Oromia region”  
✅ Cities/Towns: “Mogadishu,” “Gao,” “Maiduguri”  
✅ Villages: “Giwa village,” “Tiloa”  
✅ Specific sites: “the marketplace,” “military base near Kidal”  
✅ Geographic features: “along the Niger River,” “in the Sahel region”

**Exclude:** ❌ Vague references: “in the area,” “in the region” (unless no better location given)

#### 4.3.2 Location Granularity Levels

Annotate at the **most specific level** mentioned:

**Hierarchy:** 1. **Specific site** (most specific): “Bama market,” “military base” 2. **Village/neighborhood**: “Giwa village,” “Bakasi neighborhood” 3. **City/town**: “Maiduguri,” “Gao” 4. **District/county**: “Bama district” 5. **State/province/region**: “Borno State,” “Gao region” 6. **Country**: “Nigeria,” “Mali” 7. **Sub-region** (least specific): “Lake Chad Basin,” “Sahel region”

**Example:**

Text: "Armed militants attacked a market in Bama, Borno State, Nigeria"  
  
Annotation:  
- Location (most specific): "Bama market"  
- City: "Bama"  
- State/Province: "Borno State"  
- Country: "Nigeria"  
- Coordinates: [Use Google Maps to find Bama coordinates]

#### 4.3.3 Geocoding Locations

For each location, provide coordinates when possible:

**Process:** 1. Search location in Google Maps 2. Right-click location → “What’s here?” 3. Copy latitude, longitude 4. Record in format: “Lat, Long”

**Example:** - Location: “Maiduguri, Borno State, Nigeria” - Coordinates: “11.8464, 13.0784”

**If location cannot be found:** - Mark as: “Unable to geocode” - Provide best approximation: “Borno State, Nigeria (approximate)”

#### 4.3.4 Ambiguous Locations

**Multiple locations mentioned:**

Text: "The group has carried out attacks in Mali, Burkina Faso, and Niger"  
  
Annotation:  
- If describing single event: Choose the specific location if later clarified  
- If describing multiple events: This indicates multiple events, annotate separately

**Unclear location:**

Text: "The attack occurred in the north"  
  
Annotation:  
- Location: "[Country] northern region" (if country is clear from context)  
- Note: "Specific location not provided"  
- Confidence: Low

### 4.4 Date/Time (When) Annotation

**Definition:** The temporal reference indicating when the violence occurred.

#### 4.4.1 Types of Temporal Expressions

**Absolute dates:** - “January 15, 2024” - “Monday, March 3” - “15 March”

**Relative dates:** - “yesterday” - “last week” - “on Tuesday” (relative to article publication date) - “three days ago”

**Time periods:** - “early morning” - “overnight” - “during the night”

**Vague temporal references:** - “recently” - “in recent weeks” - “earlier this month”

#### 4.4.2 Date Normalization

**Convert all dates to: YYYY-MM-DD format**

**Process:** 1. Check article publication date (provided with article) 2. Calculate absolute date from relative reference 3. Record in standard format

**Examples:**

| Text Expression | Article Date | Normalized Date |
| --- | --- | --- |
| “yesterday” | 2024-03-15 | 2024-03-14 |
| “last Tuesday” | 2024-03-15 (Friday) | 2024-03-12 |
| “three days ago” | 2024-03-15 | 2024-03-12 |
| “on Monday” | 2024-03-15 (Friday) | 2024-03-11 |
| “January 15” | 2024-03-15 | 2024-01-15 |
| “15 Jan” | 2024-03-15 | 2024-01-15 |

**Uncertainty handling:**

| Text Expression | Normalized Date | Confidence |
| --- | --- | --- |
| “recently” | [Article date - 7 days] | Low |
| “earlier this month” | [First of current month] | Low |
| “last week” | [Article date - 7 days] | Medium |
| “in early March” | 2024-03-05 (approximate) | Medium |

#### 4.4.3 Time of Day

If mentioned, record time:

**Categories:** - Early morning (00:00-06:00) - Morning (06:00-12:00) - Afternoon (12:00-18:00) - Evening (18:00-21:00) - Night (21:00-00:00) - Specific time if given: “08:30 AM”

### 4.5 Weapon/Method (How) Annotation

**Definition:** The instruments, weapons, or methods used to perpetrate violence.

#### 4.5.1 What to Mark as Weapon/Method

**Include:** ✅ Firearms: “guns,” “rifles,” “AK-47s,” “automatic weapons”  
✅ Explosives: “bomb,” “IED,” “grenade,” “explosives,” “suicide vest”  
✅ Vehicles: “vehicle-borne explosives,” “car bomb,” “truck”  
✅ Edged weapons: “knives,” “machetes”  
✅ Fire: “burned,” “set ablaze,” “arson”  
✅ Tactics: “ambush,” “raid,” “coordinated attack”  
✅ Indirect methods: “shelling,” “airstrike,” “mortar fire”

#### 4.5.2 Weapon/Method Categories

**Primary Categories:**

1. **Firearms**
   * Small arms (pistols, rifles)
   * Automatic weapons
   * Sniper fire
   * Unspecified guns
2. **Explosives**
   * IED (Improvised Explosive Device)
   * Suicide bomb/vest
   * Car bomb/VBIED
   * Grenade
   * Landmine
   * Unspecified explosives
3. **Vehicles**
   * As weapon (ramming)
   * As explosive delivery
4. **Edged Weapons**
   * Knives
   * Machetes
   * Swords
5. **Fire/Arson**
   * Building burning
   * Village burning
6. **Heavy Weapons**
   * Artillery
   * Mortars
   * Rocket launchers
7. **Other/Multiple**
   * Combination of methods
   * Unspecified

#### 4.5.3 Tactical Methods

Also record tactical approach:

* Ambush
* Raid
* Assault/frontal attack
* Siege
* Hit-and-run
* Coordinated multi-site attack
* Targeted killing
* Mass shooting/rampage

**Example:**

Text: "Militants ambushed the convoy with IEDs and automatic weapons"  
  
Annotation:  
- Weapon: "IEDs; automatic weapons"  
- Tactic: "Ambush"  
- Method Category: "Explosives; Firearms"

## 5. Event-Level Annotation

### 5.1 Identifying Event Boundaries

**Key Question:** How many distinct violent events does this article describe?

#### 5.1.1 One Event vs. Multiple Events

**Single Event Indicators:** - Same actors - Same location - Same time - Same victims - Continuous action

**Multiple Events Indicators:** - Different times (“on Monday… then on Wednesday…”) - Different locations (“in Maiduguri… separately in Kano…”) - Different actors perpetrating separate actions - Clearly distinct incidents

**Examples:**

**Example 1: ONE EVENT**

"Armed militants attacked a village in Mali on Tuesday morning. The attackers   
killed 12 civilians and burned several houses before fleeing. Witnesses said   
the assault lasted about two hours."  
  
Analysis:   
- Same actors (militants)  
- Same time (Tuesday morning)  
- Same location (one village)  
- Continuous action (attack, killing, burning as part of single assault)  
→ ONE EVENT

**Example 2: TWO EVENTS**

"Rebels clashed with government forces in Gao on Monday, leaving 8 soldiers dead.   
In a separate incident on Tuesday, suspected militants bombed a market in Kidal,   
killing 15 civilians."  
  
Analysis:  
- Event 1: Gao, Monday, rebels vs. soldiers (armed clash)  
- Event 2: Kidal, Tuesday, militants vs. civilians (bombing)  
- Different locations, different times, different nature  
→ TWO EVENTS

**Example 3: TWO EVENTS (Sequential)**

"Boko Haram militants first bombed a crowded market in Maiduguri, then opened   
fire on fleeing survivors. At least 40 people were killed."  
  
Analysis:  
- Event 1: Market bombing (explosive attack)  
- Event 2: Shooting of survivors (armed assault)  
- Same actors, same general location, same time period  
- But different methods and arguably different actions/phases  
→ TWO EVENTS (or can be ONE COMPLEX EVENT - use judgment)  
  
\*\*Decision:\*\* Mark as TWO EVENTS but note they are related/sequential

#### 5.1.2 Complex Scenarios

**Scenario: Attack with Multiple Components**

Text: "The assault began with a suicide car bomb at the entrance, followed by   
gunmen storming the compound and taking hostages. Security forces responded,   
leading to a 3-hour firefight."  
  
How to annotate:  
- Event 1: Suicide car bombing (Terrorism > Bombing > Suicide/Car Bomb)  
- Event 2: Armed assault and hostage-taking (Terrorism > Armed Assault + Kidnapping)  
- Event 3: Armed clash between attackers and security forces (Political Violence > Armed Clash)  
  
Relationship: Events 1, 2, 3 are coordinated/sequential parts of complex attack

**Scenario: Revenge Cycle**

Text: "Following last week's attack by Group A on Village X, Group B retaliated   
yesterday by attacking Village Y."  
  
How to annotate:  
- Event 1: Group A attacks Village X (use date from "last week")  
- Event 2: Group B attacks Village Y (use date from "yesterday")  
  
Relationship: Event 2 is retaliation for Event 1

### 5.2 Event Trigger Identification

**Definition:** The word or phrase that explicitly indicates violent action.

**Trigger Types:**

**Verbs:** - Action verbs: “killed,” “attacked,” “shot,” “bombed,” “kidnapped,” “burned” - Violence verbs: “murdered,” “assassinated,” “massacred,” “executed” - Conflict verbs: “clashed,” “fought,” “raided,” “stormed”

**Nouns:** - “attack,” “bombing,” “shooting,” “massacre,” “raid,” “ambush,” “killing”

**How to Mark:**

Identify the **primary** word/phrase that signals the violent event:

Text: "Armed militants attacked a military base, killing 15 soldiers."  
  
Event Trigger: [attacked]  
(Alternative trigger in same sentence: [killing] - but "attacked" is primary action)

Text: "A suicide bomber detonated explosives at a market."  
  
Event Trigger: [detonated]  
(Alternative: "suicide bomber" but "detonated" is the action verb)

Text: "The massacre in Giwa village left 40 dead."  
  
Event Trigger: [massacre]  
(Noun form - the violence indicator)

### 5.3 Extracting 5W1H Attributes

For each identified event, complete the 5W1H template:

#### Template Format:

EVENT ID: [Unique identifier, e.g., Article01\_Event1]  
  
WHO (Actor):  
- Text span: [exact text from article]  
- Normalized: [standardized actor name]  
- Actor type: [State/Non-state armed group/Criminal/Communal/Unknown]  
  
WHAT (Event Type):  
- Event trigger: [violence-indicating word/phrase]  
- Event description: [brief summary in your own words]  
- Preliminary type: [initial classification guess]  
  
WHOM (Victim):  
- Text span: [exact text from article]  
- Normalized: [standardized victim description]  
- Victim type: [Civilian/Combatant/Mixed/Infrastructure/Unknown]  
- Deaths: [number or "unknown"]  
- Injuries: [number or "unknown"]  
  
WHERE (Location):  
- Most specific location: [village, city, site]  
- Administrative divisions: [district, state, country]  
- Coordinates: [latitude, longitude]  
- Confidence: [High/Medium/Low]  
  
WHEN (Date/Time):  
- Text expression: [exact temporal reference from article]  
- Normalized date: [YYYY-MM-DD]  
- Time of day: [if mentioned]  
- Confidence: [High/Medium/Low]  
  
HOW (Method/Weapon):  
- Weapons: [firearms, explosives, etc.]  
- Tactics: [ambush, raid, assault, etc.]  
- Additional details: [any relevant specifics]

#### 5.3.1 Handling Missing Information

**If information is not mentioned in the article:**

* ✅ Mark as “Not mentioned” or “Unknown”
* ✅ Do NOT infer from other sources
* ✅ Do NOT use your background knowledge
* ❌ Do NOT leave blank - explicitly mark as unknown

**Examples:**

"Gunmen killed 5 people in Nigeria."  
  
Complete as:  
- Who: "Gunmen" (Normalized: "Unknown gunmen")  
- Whom: "5 people"   
- Where: "Nigeria" (no more specific location given)  
- When: Unknown (must be inferred from article date - use article date with low confidence)  
- How: Unknown (no weapon/method specified - "shooting" can be inferred from "gunmen" but mark as assumed)

#### 5.3.2 Handling Ambiguous Information

**If information is unclear or contradictory:**

✅ Note both possibilities  
✅ Mark confidence as Low  
✅ Document the ambiguity in Notes field  
✅ Flag for adjudication if very unclear

**Example:**

"The attack, which may have been carried out by Al-Shabaab or local militia,   
killed between 10 and 20 people."  
  
Annotate:  
- Who: "Al-Shabaab or local militia (uncertain)"  
- Confidence: Low  
- Deaths: "10-20 (range reported)"  
- Notes: "Actor identity unclear in article; casualty number is range"

### 5.4 Coreference Resolution

**Definition:** Identifying when different expressions refer to the same entity.

#### Common Coreference Patterns:

**Pattern 1: Name → Pronoun**

"Al-Shabaab attacked the base. They killed 10 soldiers."  
- "Al-Shabaab" = "They"

**Pattern 2: Name → Description**

"Boko Haram bombed a market. The militants then fled."  
- "Boko Haram" = "The militants"

**Pattern 3: Name → Acronym**

"The Lord's Resistance Army raided villages. The LRA has been active..."  
- "The Lord's Resistance Army" = "The LRA"

**Pattern 4: General → Specific**

"Armed men attacked the village. Witnesses identified the attackers as Fulani herders."  
- "Armed men" = "the attackers" = "Fulani herders"

#### How to Record Coreference:

**Option 1: In entity annotation section**

Entity 1:  
- Mention 1: "Al-Shabaab"  
- Mention 2: "The militants"  
- Mention 3: "They"  
- Coreference Group ID: ACTOR\_1  
  
Entity 2:  
- Mention 1: "the base"  
- Mention 2: "the military compound"  
- Mention 3: "it"  
- Coreference Group ID: LOCATION\_1

**Option 2: In notes**

Notes: "Al-Shabaab" = "The militants" = "They" (coreference)

## 6. Taxonomy Classification

### 6.1 Classification Process

**Step-by-Step:**

**STEP 1: Identify Actor Type** - State forces? → Consider State Violence or Political Violence - Non-state armed group? → Likely Political or Criminal Violence - Communal groups? → Likely Communal Violence

**STEP 2: Identify Motivation** - Political change/control? → Political Violence - Economic/profit? → Criminal Violence - Identity/resources? → Communal Violence

**STEP 3: Identify Target** - Government/military? → Political Violence - Civilians (by state)? → State Violence Against Civilians - Symbolic/terror target? → Terrorism - Other community? → Communal Violence

**STEP 4: Apply Hierarchy** - Choose Level 1 category - Choose Level 2 sub-category - Choose Level 3 specific type - Choose Level 4 detailed subtype (if applicable)

### 6.2 Using the Taxonomy Reference

**While classifying, have the Taxonomy document open to:**

* Section 4: Detailed Category Definitions
* Section 5: Decision Rules
* Section 6: Examples

**Classification Workflow:**

1. Read event description
2. Check Section 5 (Decision Rules) for similar scenarios
3. Navigate to appropriate Level 1 category in Section 4
4. Read Level 2 definitions under that category
5. Select best-matching Level 2
6. Read Level 3 definitions under that Level 2
7. Select best-matching Level 3
8. Check if Level 4 applies
9. Verify with examples in Section 6
10. Document reasoning if ambiguous

### 6.3 Classification Examples

**Example 1: Market Bombing**

Text: "A suicide bomber detonated at a crowded market in Maiduguri, killing 32 civilians.   
Boko Haram claimed responsibility."  
  
Classification Process:  
- Actor: Boko Haram (designated terrorist group)  
- Target: Civilians in public place  
- Method: Suicide bombing  
- Motivation: Terror/political-religious  
  
→ Level 1: POLITICAL VIOLENCE  
→ Level 2: Terrorism  
→ Level 3: Bombing/Explosive Attack  
→ Level 4: Suicide Bombing  
  
Final: Political Violence > Terrorism > Bombing/Explosive Attack > Suicide Bombing

**Example 2: Military Clash**

Text: "Rebels clashed with government forces in Gao region, Mali. The battle lasted   
several hours and left 15 soldiers and 8 rebels dead."  
  
Classification Process:  
- Actor: Rebels (armed opposition group)  
- Target: Government military  
- Action: Armed combat/battle  
- Motivation: Political (anti-government insurgency)  
  
→ Level 1: POLITICAL VIOLENCE  
→ Level 2: Rebellion/Armed Insurgency  
→ Level 3: Armed Clash/Battle  
→ Level 4: N/A  
  
Final: Political Violence > Rebellion/Armed Insurgency > Armed Clash/Battle

**Example 3: Farmer-Herder Violence**

Text: "At least 20 people were killed in clashes between Fulani herders and farming   
communities in Benue State over grazing rights and crop damage."  
  
Classification Process:  
- Actor: Fulani herders AND farmers (communal groups)  
- Target: Each other (communal conflict)  
- Motivation: Resource competition (land use)  
- Context: Livelihood-based conflict  
  
→ Level 1: COMMUNAL VIOLENCE  
→ Level 2: Pastoralist-Farmer Clashes  
→ Level 3: Grazing Conflict  
→ Level 4: N/A  
  
Final: Communal Violence > Pastoralist-Farmer Clashes > Grazing Conflict

**Example 4: Police Shooting Protesters**

Text: "Police opened fire on opposition protesters in Harare, killing at least 6   
people and injuring dozens. Protesters were demonstrating against fuel price increases."  
  
Classification Process:  
- Actor: Police (state forces)  
- Target: Protesters (civilians engaged in political activity)  
- Context: Excessive force against largely peaceful protest  
- Ambiguity: Could be State Violence or Political Repression  
  
Decision:  
→ Level 1: POLITICAL VIOLENCE (protesters engaged in political act)  
→ Level 2: Political Repression  
→ Level 3: Violent Suppression of Protests  
→ Level 4: N/A  
  
Alternative classification: STATE VIOLENCE AGAINST CIVILIANS > State Repression of Protests  
  
Decision: Primary = Political Violence > Political Repression > Violent Suppression of Protests  
Secondary = State Violence Against Civilians (multi-label)  
  
Final: Political Violence > Political Repression > Violent Suppression of Protests  
(Note multi-label possibility)

### 6.4 Handling Ambiguous Classifications

**When event could fit multiple categories:**

**Option 1: Choose Primary + Note Alternative**

Primary Classification: [Most prominent/central category]  
Alternative Classifications: [Other valid categories]  
Reasoning: [Why you chose primary]  
Confidence: Medium or Low

**Option 2: Multi-Label (if genuinely ambiguous)**

Classification 1: [First valid category]  
Classification 2: [Second valid category]  
Multi-label: Yes  
Reasoning: [Why both apply]

**When to multi-label:** - Event genuinely spans two Level 1 categories - Cannot determine single best classification from text - Both classifications would be valid interpretations

**Example:**

Text: "Ethnic militia affiliated with ruling party attacks opposition stronghold   
during election period"  
  
This could be:  
- Election Violence (timing and political context)  
- Ethnic/Communal Violence (ethnic militia actor)  
- Political Repression (state-affiliated actor targeting opposition)  
  
Approach:   
Primary: Election Violence > Campaign Violence  
Secondary: Communal Violence > Ethnic Conflict  
Note: "Complex case with political and ethnic dimensions"

### 6.5 Confidence Scoring

**Rate your confidence in the classification:**

**HIGH (0.9-1.0):** - Clear, unambiguous case - All indicators point to same category - Matches taxonomy examples closely - No alternative interpretations

**MEDIUM (0.6-0.8):** - Generally clear but some ambiguity - Most indicators point to chosen category - Minor alternative interpretations exist - Reasonable classification but not certain

**LOW (0.3-0.5):** - Significant ambiguity - Multiple plausible interpretations - Limited information in text - Classification is best guess - Should flag for adjudication

**VERY LOW (0.0-0.2):** - Cannot confidently classify - Insufficient information - Highly ambiguous - MUST flag for adjudication

## 7. Quality Control Procedures

### 7.1 Self-Review Checklist

After completing each article, review using this checklist:

**Article-Level Checks:** - [ ] Article ID recorded correctly - [ ] Article URL/source recorded - [ ] Publication date recorded - [ ] Annotator name and date recorded

**Event Identification:** - [ ] All violent events in article identified - [ ] Event boundaries clear (no merged events) - [ ] No events missed

**Entity Annotation (for each event):** - [ ] Actor identified and marked - [ ] Victim identified and marked - [ ] Location identified and geocoded - [ ] Date normalized correctly - [ ] Weapons/methods identified

**5W1H Completeness:** - [ ] All fields filled (or marked “Unknown”) - [ ] No blank fields - [ ] Coreferences resolved - [ ] Normalization completed

**Taxonomy Classification:** - [ ] Level 1 category assigned - [ ] Level 2 category assigned - [ ] Level 3 category assigned - [ ] Level 4 assigned if applicable - [ ] Confidence score assigned - [ ] Reasoning documented for ambiguous cases

**Relationships:** - [ ] Related events linked - [ ] Relationship types specified

**Notes:** - [ ] Ambiguities documented - [ ] Low confidence cases flagged - [ ] Questions noted for adjudication

### 7.2 Common Mistakes to Avoid

**Mistake 1: Merging Multiple Events** ❌ Wrong: Treating sequential attacks as single event  
✅ Correct: Each distinct incident is separate event

**Mistake 2: Incomplete Coreference** ❌ Wrong: Not linking “The militants” back to “Al-Shabaab”  
✅ Correct: Identify all mentions of same entity

**Mistake 3: Using External Knowledge** ❌ Wrong: “I know Al-Shabaab operates in this area, so they probably did it”  
✅ Correct: Only annotate what the article explicitly states

**Mistake 4: Incorrect Date Normalization** ❌ Wrong: “yesterday” → not converting to actual date  
✅ Correct: “yesterday” + article date → specific date

**Mistake 5: Too Narrow Entity Spans** ❌ Wrong: Marking only “militants” when text says “armed militants linked to AQIM”  
✅ Correct: Mark complete noun phrase

**Mistake 6: Missing Implied Information** ❌ Wrong: Not marking victims when text says “suicide bombing in market” but no casualties mentioned  
✅ Correct: Note “Market-goers (casualties not specified)”

**Mistake 7: Wrong Taxonomy Level** ❌ Wrong: Jumping directly to Level 4 without filling Levels 1-3  
✅ Correct: Complete all hierarchical levels

**Mistake 8: Ignoring Ambiguity** ❌ Wrong: Guessing classification when uncertain  
✅ Correct: Document ambiguity, mark low confidence, note alternatives

### 7.3 Adjudication Process

**When to flag for adjudication:**

✅ Confidence score < 0.5  
✅ Cannot determine event boundaries  
✅ Major ambiguity in actor identification  
✅ Unclear whether one or multiple events  
✅ Cannot confidently classify (multiple equally valid categories)  
✅ Contradictory information in article  
✅ Missing critical information

**How to flag:** - Mark “FLAG FOR ADJUDICATION” in notes - Describe specific issue/question - Note what alternatives you considered - Provide your best attempt but mark as tentative

**Example:**

Notes: "FLAG FOR ADJUDICATION - Unclear if this is one coordinated attack   
or two separate incidents. Article says 'militants attacked the base, then   
separately bombed the market' but timing is ambiguous. Annotated as two events   
but could be single complex attack. Please review."

### 7.4 Inter-Annotator Agreement

**If multiple annotators work on same article (for quality checks):**

**Process:** 1. Both annotators independently annotate same article 2. Compare annotations 3. Measure agreement 4. Discuss disagreements 5. Reach consensus 6. Learn from patterns

**What to compare:** - Number of events identified - Entity spans (character-level match) - 5W1H values - Taxonomy classifications - Confidence scores

**Goal:** Agreement > 80% on major decisions (event count, Level 1-2 classification)

## 8. Annotation Format Specifications

### 8.1 Template Structure

**File Format:** Excel (.xlsx) or Google Sheets

**Sheet 1: Article Metadata**

| Field | Type | Required | Example |  
|-------|------|----------|---------|  
| Article\_ID | Text | Yes | ART\_001 |  
| Article\_URL | URL | Yes | https://... |  
| Article\_Title | Text | Yes | "Armed Attack in Mali" |  
| News\_Source | Text | Yes | "Reuters" |  
| Publication\_Date | Date | Yes | 2024-03-15 |  
| Annotator\_Name | Text | Yes | "John Doe" |  
| Annotation\_Date | Date | Yes | 2024-10-20 |  
| Article\_Text | Long text | Yes | [Full article] |  
| Notes | Text | Optional | Any general notes |

**Sheet 2: Event Records**

| Field | Type | Required | Example |  
|-------|------|----------|---------|  
| Event\_ID | Text | Yes | ART\_001\_EVT\_001 |  
| Article\_ID | Text | Yes | ART\_001 |  
| Event\_Trigger\_Text | Text | Yes | "attacked" |  
| Sentence\_Numbers | Text | Yes | "3,4,5" |  
| Actor\_Text | Text | Yes | "Al-Shabaab militants" |  
| Actor\_Normalized | Text | Yes | "Al-Shabaab" |  
| Actor\_Type | Category | Yes | "Non-state armed group" |  
| Actor\_Confidence | Number | Yes | 0.9 |  
| Victim\_Text | Text | Yes | "at least 15 civilians" |  
| Victim\_Normalized | Text | Yes | "15 civilians" |  
| Victim\_Type | Category | Yes | "Civilian" |  
| Deaths | Number | Optional | 15 |  
| Injuries | Number | Optional | 30 |  
| Victim\_Confidence | Number | Yes | 0.8 |  
| Location\_Text | Text | Yes | "Bama market, Borno" |  
| Location\_Normalized | Text | Yes | "Bama, Borno State, Nigeria" |  
| Location\_Specific | Text | Yes | "Bama market" |  
| Location\_City | Text | Yes | "Bama" |  
| Location\_State | Text | Optional | "Borno State" |  
| Location\_Country | Text | Yes | "Nigeria" |  
| Location\_Coordinates | Text | Optional | "11.5204, 13.6891" |  
| Location\_Confidence | Number | Yes | 0.9 |  
| Date\_Text | Text | Yes | "yesterday" |  
| Date\_Normalized | Date | Yes | 2024-03-14 |  
| Time\_of\_Day | Category | Optional | "Morning" |  
| Date\_Confidence | Number | Yes | 0.9 |  
| Weapon\_Text | Text | Yes | "guns and explosives" |  
| Weapon\_Normalized | Text | Yes | "Firearms; Explosives" |  
| Weapon\_Category | Category | Yes | "Multiple" |  
| Tactic | Text | Optional | "Raid" |  
| Taxonomy\_L1 | Category | Yes | "Political Violence" |  
| Taxonomy\_L2 | Category | Yes | "Terrorism" |  
| Taxonomy\_L3 | Category | Yes | "Armed Assault" |  
| Taxonomy\_L4 | Category | Optional | "Armed Raid" |  
| Classification\_Confidence | Number | Yes | 0.85 |  
| Alternative\_Classification | Text | Optional | "Rebellion > Rebel Attack" |  
| Multi\_Label | Boolean | No | FALSE |  
| Event\_Description | Text | Yes | Brief summary |  
| Notes | Text | Optional | Any notes/flags |  
| Flagged\_for\_Review | Boolean | No | FALSE |

**Sheet 3: Entity Annotations (Detailed)**

| Field | Type | Example |  
|-------|------|---------|  
| Entity\_ID | Text | ENT\_001 |  
| Event\_ID | Text | ART\_001\_EVT\_001 |  
| Entity\_Text | Text | "Al-Shabaab militants" |  
| Entity\_Type | Category | "Actor" |  
| Start\_Char | Number | 245 |  
| End\_Char | Number | 265 |  
| Sentence\_Number | Number | 3 |  
| Normalized\_Form | Text | "Al-Shabaab" |  
| Coreference\_Group\_ID | Text | "ACTOR\_1" |  
| All\_Mentions | Text | "Al-Shabaab militants; The attackers; They" |

**Sheet 4: Relationships**

| Field | Type | Example |  
|-------|------|---------|  
| Relationship\_ID | Text | REL\_001 |  
| Event\_1\_ID | Text | ART\_001\_EVT\_001 |  
| Relationship\_Type | Category | "Retaliation" |  
| Event\_2\_ID | Text | ART\_001\_EVT\_002 |  
| Description | Text | "Event 2 is retaliation for Event 1" |  
| Confidence | Number | 0.8 |

### 8.2 Category Options (Controlled Vocabularies)

**Actor\_Type:** - State forces - Non-state armed group - Criminal organization - Communal group - Unknown - Multiple

**Victim\_Type:** - Civilian - Combatant - Mixed (civilian and combatant) - Infrastructure - Unknown

**Weapon\_Category:** - Firearms - Explosives - Edged weapons - Fire/Arson - Heavy weapons - Multiple - Unknown

**Tactic:** - Ambush - Raid - Assault - Siege - Hit-and-run - Coordinated attack - Targeted killing - Mass shooting - Other - Unknown

**Time\_of\_Day:** - Early morning (00:00-06:00) - Morning (06:00-12:00) - Afternoon (12:00-18:00) - Evening (18:00-21:00) - Night (21:00-00:00) - Unknown

**Relationship\_Type:** - Temporal\_Sequence (Event A before Event B) - Causality (Event A caused Event B) - Retaliation (Event B is revenge for Event A) - Coordination (Events are coordinated) - Related\_Context (Events are contextually related)

### 8.3 Confidence Score Scale

**Scale: 0.0 to 1.0**

* **0.9-1.0:** Very high confidence, clear and unambiguous
* **0.7-0.8:** High confidence, minor uncertainty
* **0.5-0.6:** Medium confidence, some ambiguity
* **0.3-0.4:** Low confidence, significant ambiguity
* **0.1-0.2:** Very low confidence, major uncertainty
* **0.0:** Cannot determine, completely unclear

### 8.4 Naming Conventions

**Article IDs:** - Format: ART\_XXX where XXX is 3-digit number - Example: ART\_001, ART\_002, …, ART\_999

**Event IDs:** - Format: ArticleID\_EVT\_YY where YY is 2-digit event number within article - Example: ART\_001\_EVT\_01, ART\_001\_EVT\_02

**Entity IDs:** - Format: EventID\_ENT\_ZZ where ZZ is entity number - Example: ART\_001\_EVT\_01\_ENT\_01

**Coreference Group IDs:** - Format: EventID\_ACTOR\_N or EventID\_VICTIM\_N or EventID\_LOC\_N - Example: ART\_001\_EVT\_01\_ACTOR\_1

## 9. Fully Annotated Examples

### Example 1: Simple Single Event

**Article Text:**

Maiduguri, Nigeria - At least 12 people were killed when a suicide bomber   
detonated explosives at a crowded marketplace in Maiduguri, Borno State on   
Tuesday morning. Witnesses said the bomber was wearing a vest packed with   
explosives. Boko Haram, the militant group that has been waging an insurgency   
in northeastern Nigeria since 2009, claimed responsibility for the attack on   
social media. The marketplace was crowded with shoppers when the explosion   
occurred around 10 AM. At least 30 others were injured in the blast.  
  
- Source: Reuters  
- Publication Date: 2024-03-20

**ANNOTATION:**

**Article Metadata:** - Article\_ID: ART\_001 - Article\_URL: [example URL] - Article\_Title: “Suicide Bomber Kills 12 in Maiduguri Market” - News\_Source: “Reuters” - Publication\_Date: 2024-03-20 - Annotator\_Name: [Your name] - Annotation\_Date: 2024-10-20

**Event Record:**

**Event\_ID:** ART\_001\_EVT\_001

**Event Identification:** - Event\_Trigger\_Text: “detonated” - Sentence\_Numbers: “1, 2, 5” - Event\_Description: “Suicide bombing at crowded marketplace in Maiduguri”

**Actor (Who):** - Actor\_Text: “a suicide bomber” - Actor\_Normalized: “Boko Haram (suicide bomber)” - Actor\_Type: “Non-state armed group” - Actor\_Confidence: 0.95 - Notes: “Boko Haram claimed responsibility (sentence 3)”

**Victim (Whom):** - Victim\_Text: “At least 12 people… at least 30 others” - Victim\_Normalized: “12 civilians (killed); 30 civilians (injured)” - Victim\_Type: “Civilian” - Deaths: 12 - Injuries: 30 - Victim\_Confidence: 0.9 - Notes: “Marketplace shoppers = civilians”

**Location (Where):** - Location\_Text: “crowded marketplace in Maiduguri, Borno State” - Location\_Normalized: “Maiduguri, Borno State, Nigeria” - Location\_Specific: “Marketplace in Maiduguri” - Location\_City: “Maiduguri” - Location\_State: “Borno State” - Location\_Country: “Nigeria” - Location\_Coordinates: “11.8464, 13.0784” (Maiduguri center) - Location\_Confidence: 0.95

**Date/Time (When):** - Date\_Text: “on Tuesday morning” - Date\_Normalized: 2024-03-19 (Tuesday before publication date 2024-03-20) - Time\_of\_Day: “Morning” (article says “around 10 AM”) - Date\_Confidence: 0.95

**Weapon/Method (How):** - Weapon\_Text: “explosives… vest packed with explosives” - Weapon\_Normalized: “Suicide vest explosive” - Weapon\_Category: “Explosives” - Tactic: “Suicide bombing”

**Taxonomy Classification:** - Taxonomy\_L1: “Political Violence” - Taxonomy\_L2: “Terrorism” - Taxonomy\_L3: “Bombing/Explosive Attack” - Taxonomy\_L4: “Suicide Bombing” - Classification\_Confidence: 0.95 - Classification\_Reasoning: “Boko Haram terrorist group, civilian target, suicide bombing method = clear terrorism case”

**Other Fields:** - Alternative\_Classification: None - Multi\_Label: FALSE - Flagged\_for\_Review: FALSE

**Entity Annotations (Detailed):**

Entity 1: - Entity\_ID: ART\_001\_EVT\_001\_ENT\_001 - Entity\_Text: “a suicide bomber” - Entity\_Type: “Actor” - Normalized\_Form: “Suicide bomber (Boko Haram)” - Coreference\_Group\_ID: “ART\_001\_EVT\_001\_ACTOR\_1” - All\_Mentions: “a suicide bomber; the bomber”

Entity 2: - Entity\_ID: ART\_001\_EVT\_001\_ENT\_002 - Entity\_Text: “Boko Haram” - Entity\_Type: “Actor\_Organization” - Normalized\_Form: “Boko Haram” - Coreference\_Group\_ID: “ART\_001\_EVT\_001\_ACTOR\_1” - All\_Mentions: “Boko Haram; the militant group”

Entity 3: - Entity\_ID: ART\_001\_EVT\_001\_ENT\_003 - Entity\_Text: “At least 12 people” - Entity\_Type: “Victim” - Normalized\_Form: “12 civilians (killed)” - Coreference\_Group\_ID: “ART\_001\_EVT\_001\_VICTIM\_1”

Entity 4: - Entity\_ID: ART\_001\_EVT\_001\_ENT\_004 - Entity\_Text: “at least 30 others” - Entity\_Type: “Victim” - Normalized\_Form: “30 civilians (injured)” - Coreference\_Group\_ID: “ART\_001\_EVT\_001\_VICTIM\_2”

Entity 5: - Entity\_ID: ART\_001\_EVT\_001\_ENT\_005 - Entity\_Text: “a crowded marketplace in Maiduguri, Borno State” - Entity\_Type: “Location” - Normalized\_Form: “Maiduguri marketplace, Borno State, Nigeria” - Coreference\_Group\_ID: “ART\_001\_EVT\_001\_LOC\_1” - All\_Mentions: “a crowded marketplace in Maiduguri; The marketplace”

### Example 2: Multiple Events with Relationships

**Article Text:**

Bamako, Mali - Malian armed forces clashed with suspected JNIM militants near   
Gao on Monday, leaving at least 8 soldiers and 5 militants dead, according to   
military sources. The fighting erupted when troops on patrol were ambushed by   
the armed group.  
  
In a separate incident on Wednesday, a vehicle carrying UN peacekeepers struck   
an improvised explosive device (IED) on a road between Gao and Kidal. Three   
peacekeepers from Chad were killed and two others were wounded. No group has   
claimed responsibility for the IED attack, but JNIM has previously targeted UN   
forces in the region.  
  
- Source: AFP  
- Publication Date: 2024-03-21

**ANNOTATION:**

**Article Metadata:** - Article\_ID: ART\_002 - Publication\_Date: 2024-03-21 - News\_Source: “AFP”

**EVENT 1:**

**Event\_ID:** ART\_002\_EVT\_001

**Event Identification:** - Event\_Trigger\_Text: “clashed” - Event\_Description: “Armed clash between Malian forces and JNIM militants”

**Actor (Who):** - Actor\_Text: “suspected JNIM militants” - Actor\_Normalized: “JNIM (Jama’at Nasr al-Islam wal Muslimin)” - Actor\_Type: “Non-state armed group” - Actor\_Confidence: 0.85 - Notes: “Described as ‘suspected JNIM’ - not 100% confirmed”

**Victim (Whom):** - Victim\_Text: “at least 8 soldiers and 5 militants” - Victim\_Normalized: “8 Malian soldiers (killed); 5 militants (killed)” - Victim\_Type: “Combatant” (both sides) - Deaths: 13 total (8 government, 5 militants) - Victim\_Confidence: 0.9

**Location (Where):** - Location\_Text: “near Gao” - Location\_Normalized: “Gao region, Mali” - Location\_Specific: “Near Gao city” - Location\_City: “Gao” - Location\_Country: “Mali” - Location\_Coordinates: “16.2710, -0.0445” (Gao coordinates) - Location\_Confidence: 0.8 (exact location unclear - “near Gao”)

**Date/Time (When):** - Date\_Text: “on Monday” - Date\_Normalized: 2024-03-18 (Monday before Thu publication) - Date\_Confidence: 0.9

**Weapon/Method (How):** - Weapon\_Text: Not explicitly stated - Weapon\_Normalized: “Firearms (assumed)” - Weapon\_Category: “Firearms” - Tactic: “Ambush” - Notes: “Article says troops were ‘ambushed’”

**Taxonomy Classification:** - Taxonomy\_L1: “Political Violence” - Taxonomy\_L2: “Rebellion/Armed Insurgency” - Taxonomy\_L3: “Ambush” - Taxonomy\_L4: N/A - Classification\_Confidence: 0.9 - Reasoning: “Armed group ambushing government forces = classic insurgency ambush”

**EVENT 2:**

**Event\_ID:** ART\_002\_EVT\_002

**Event Identification:** - Event\_Trigger\_Text: “struck” - Event\_Description: “IED attack on UN peacekeepers”

**Actor (Who):** - Actor\_Text: “No group claimed responsibility” - Actor\_Normalized: “Unknown (possibly JNIM)” - Actor\_Type: “Unknown” - Actor\_Confidence: 0.3 - Notes: “No claim of responsibility but JNIM suspected based on past patterns”

**Victim (Whom):** - Victim\_Text: “Three peacekeepers from Chad… two others” - Victim\_Normalized: “3 Chadian UN peacekeepers (killed); 2 peacekeepers (injured)” - Victim\_Type: “Combatant” (peacekeepers are armed forces) - Deaths: 3 - Injuries: 2 - Victim\_Confidence: 0.95

**Location (Where):** - Location\_Text: “on a road between Gao and Kidal” - Location\_Normalized: “Road between Gao and Kidal, Mali” - Location\_Specific: “Gao-Kidal road” - Location\_City: N/A (between cities) - Location\_State: “Gao region” - Location\_Country: “Mali” - Location\_Coordinates: “17.0, 0.5” (approximate midpoint) - Location\_Confidence: 0.7

**Date/Time (When):** - Date\_Text: “on Wednesday” - Date\_Normalized: 2024-03-20 (Wednesday before Thu publication) - Date\_Confidence: 0.9

**Weapon/Method (How):** - Weapon\_Text: “improvised explosive device (IED)” - Weapon\_Normalized: “IED (Improvised Explosive Device)” - Weapon\_Category: “Explosives” - Tactic: “Roadside IED”

**Taxonomy Classification:** - Taxonomy\_L1: “Political Violence” - Taxonomy\_L2: “Rebellion/Armed Insurgency” - Taxonomy\_L3: “Ambush” - Taxonomy\_L4: “IED Ambush” - Classification\_Confidence: 0.75 - Reasoning: “IED targeting military forces = insurgent tactic. Could also be Terrorism if goal is to intimidate UN, but given target (military) and method (IED on military route), Rebellion/Ambush is more appropriate.” - Alternative\_Classification: “Terrorism > Bombing > Roadside IED” - Notes: “Borderline case - IED could indicate terrorism, but targeting of military convoy suggests insurgent ambush”

**Relationship:** - Relationship\_ID: REL\_001 - Event\_1\_ID: ART\_002\_EVT\_001 - Relationship\_Type: “Related\_Context” - Event\_2\_ID: ART\_002\_EVT\_002 - Description: “Both events involve suspected JNIM in same region around same time period” - Confidence: 0.6

### Example 3: Complex Event with Ambiguity

**Article Text:**

Mubi, Nigeria - At least 45 people were killed in a revenge attack in Adamawa   
State on Saturday. According to local sources, armed men believed to be Fulani   
herders raided three villages in the Mubi area, burning homes and shooting   
residents. The attack is said to be in retaliation for the killing of several   
Fulani herders by farmers last month following a dispute over grazing rights and   
crop destruction.   
  
Witnesses report that the attackers arrived at dawn and systematically went   
house to house, killing men, women, and children. At least 20 homes were burned   
to the ground. Local police have yet to make any arrests. Community leaders   
called for government intervention to stop the escalating cycle of violence   
between farmers and herders in the region.  
  
- Source: Daily Trust  
- Publication Date: 2024-03-24

**ANNOTATION:**

**Event\_ID:** ART\_003\_EVT\_001

**Event Identification:** - Event\_Trigger\_Text: “raided” - Event\_Description: “Revenge raid by Fulani herders on farming villages” - Sentence\_Numbers: “1,2,3,4”

**Actor (Who):** - Actor\_Text: “armed men believed to be Fulani herders” - Actor\_Normalized: “Fulani herders (suspected)” - Actor\_Type: “Communal group” - Actor\_Confidence: 0.7 - Notes: “‘Believed to be’ indicates uncertainty. Not confirmed by authorities.”

**Victim (Whom):** - Victim\_Text: “At least 45 people… men, women, and children” - Victim\_Normalized: “45 civilians (killed) - farmers/farming community” - Victim\_Type: “Civilian” - Deaths: 45 - Injuries: Unknown - Victim\_Confidence: 0.9 - Notes: “Men, women, children indicate civilian casualties”

**Location (Where):** - Location\_Text: “three villages in the Mubi area… Adamawa State” - Location\_Normalized: “Mubi area, Adamawa State, Nigeria” - Location\_Specific: “Three villages near Mubi” - Location\_City: “Mubi” - Location\_State: “Adamawa State” - Location\_Country: “Nigeria” - Location\_Coordinates: “10.2688, 13.2669” (Mubi coordinates) - Location\_Confidence: 0.8 - Notes: “Specific village names not provided”

**Date/Time (When):** - Date\_Text: “on Saturday” - Date\_Normalized: 2024-03-23 (Saturday before Sun publication) - Time\_of\_Day: “Early morning” (article says “at dawn”) - Date\_Confidence: 0.95

**Weapon/Method (How):** - Weapon\_Text: “burning homes and shooting residents” - Weapon\_Normalized: “Firearms; Arson” - Weapon\_Category: “Multiple” - Tactic: “Raid” - Notes: “Systematic house-to-house killing indicates organized raid”

**Taxonomy Classification:** - Taxonomy\_L1: “Communal Violence” - Taxonomy\_L2: “Pastoralist-Farmer Clashes” - Taxonomy\_L3: “Revenge Raid (Pastoralist)” - Taxonomy\_L4: N/A - Classification\_Confidence: 0.7 - Reasoning: “Actor = Fulani herders (pastoralists), Victim = farmers, Context = grazing dispute and explicit revenge motive, Pattern = farmer-herder violence cycle” - Alternative\_Classification: “Communal Violence > Ethnic/Tribal Conflict > Ethnic Revenge Attack” (since Fulani have ethnic dimension) - Multi\_Label: Possible - Notes: “Could be classified under Ethnic Violence since Fulani identity has ethnic component, but pastoralist-farmer framing seems more appropriate given explicit grazing/crop context. Article emphasizes ‘farmers and herders’ rather than ethnic identities.”

**Context/Relationship:** - Notes: “Article references previous triggering event: ‘killing of several Fulani herders by farmers last month.’ This is Event 2 (which occurred outside our annotation timeframe) that this event is retaliating for.” - Implied\_Event: “Farmers killed Fulani herders last month (not fully described)”

**Ambiguities:** - Actor identification: “believed to be Fulani herders” - not 100% certain - Exact location: village names not given - Previous triggering event: described but not detailed

**Flagged Issues:** - Flagged\_for\_Review: TRUE - Review\_Reason: “Actor identification uncertain (‘believed to be’). Also borderline between Pastoralist-Farmer and Ethnic Violence classifications. Seek adjudication on taxonomy choice.”

## 10. Common Mistakes and How to Avoid Them

### Mistake 1: Merging Sequential Events

**WRONG:**

Article describes: "Militants bombed a market on Monday, then attacked a police   
station on Tuesday"  
  
Annotation: ONE EVENT - "Militants carried out attacks"

**CORRECT:**

Annotation: TWO EVENTS  
- Event 1: Market bombing, Monday  
- Event 2: Police station attack, Tuesday  
- Relationship: Sequential actions by same group

**How to Avoid:** - Check if time, location, or target differs - If yes → separate events

### Mistake 2: Including Opinion/Analysis as Facts

**WRONG:**

Article says: "Experts believe the attack was carried out by Al-Shabaab"  
  
Annotation: Actor = "Al-Shabaab" (Confidence: 0.9)

**CORRECT:**

Annotation: Actor = "Unknown (Al-Shabaab suspected by experts)" (Confidence: 0.5)  
Notes: "Attribution is expert speculation, not confirmed"

**How to Avoid:** - Distinguish statements of fact from speculation - Words like “believed,” “suspected,” “reportedly” indicate uncertainty - Lower confidence scores for unconfirmed information

### Mistake 3: Using External Knowledge

**WRONG:**

Article: "Armed men attacked a village in northern Nigeria"  
  
Annotation: Actor = "Boko Haram" (because I know they operate there)

**CORRECT:**

Annotation: Actor = "Armed men (unidentified)" or "Unknown armed group"  
Notes: "No specific identification in article"

**How to Avoid:** - **ONLY annotate what the article explicitly states** - Your background knowledge is irrelevant - If article doesn’t say it, mark as “Unknown”

### Mistake 4: Inconsistent Normalization

**WRONG:**

Event 1: Actor = "AQIM"  
Event 2: Actor = "Al-Qaeda in Islamic Maghreb"  
Event 3: Actor = "Al Qaeda Maghreb"

**CORRECT:**

All events: Actor = "Al-Qaeda in the Islamic Maghreb (AQIM)"  
(Use consistent standard form)

**How to Avoid:** - Establish standard names at start of annotation session - Keep reference list of normalized names - Always use full name with acronym in parentheses

### Mistake 5: Wrong Hierarchy Level

**WRONG:**

Taxonomy:  
- L1: Political Violence  
- L2: [blank]  
- L3: [blank]  
- L4: Suicide Bombing

**CORRECT:**

Taxonomy:  
- L1: Political Violence  
- L2: Terrorism  
- L3: Bombing/Explosive Attack  
- L4: Suicide Bombing

**How to Avoid:** - ALWAYS fill all levels from top to bottom - Cannot skip levels - Refer to taxonomy document for correct hierarchy

### Mistake 6: Ignoring Coreference

**WRONG:**

Sentence 1: "Al-Shabaab attacked the base."  
Sentence 2: "The militants killed 10 soldiers."  
  
Annotation:  
Event 1 Actor: "Al-Shabaab"  
Event 2 Actor: "The militants" (treated as separate entity)

**CORRECT:**

Annotation:  
Event 1 Actor: "Al-Shabaab"  
Coreference: "Al-Shabaab" = "The militants" = same entity  
(One event, not two)

**How to Avoid:** - Always resolve pronouns and descriptions to referents - Track entities throughout article - Mark coreference groups

### Mistake 7: Too Narrow or Too Broad Entity Spans

**WRONG:**

Text: "Armed militants believed to be linked to JNIM"  
Marked span: "militants"

**CORRECT:**

Marked span: "Armed militants believed to be linked to JNIM"  
(Complete noun phrase with all modifiers)

**How to Avoid:** - Include all adjectives, prepositional phrases, relative clauses - Mark the complete descriptive phrase - If unsure, mark broader rather than narrower

### Mistake 8: Not Documenting Ambiguity

**WRONG:**

Ambiguous case → guess classification → mark high confidence

**CORRECT:**

Ambiguous case → document both options → mark low confidence → flag for review

**How to Avoid:** - Be honest about uncertainty - Note alternatives considered - Flag genuinely ambiguous cases - Low confidence is acceptable and valuable

### Mistake 9: Incorrect Date Calculation

**WRONG:**

Article published: Friday, March 22, 2024  
Article says: "yesterday"  
  
Annotated date: March 21, 2024 (Thursday)  
[Correct]  
  
Article says: "last Monday"  
  
Annotated date: March 21, 2024 (Thursday)  
[WRONG - should be March 18]

**CORRECT:**

"last Monday" when article published Friday March 22 = Monday March 18

**How to Avoid:** - Use calendar - Count backwards carefully - “last [day]” = most recent occurrence of that day BEFORE article date

### Mistake 10: Missing Implied Violence

**WRONG:**

Article: "A suicide bomber entered the market"  
Annotation: No casualties marked  
  
[Misses implied violence]

**CORRECT:**

Annotation: "Casualties: Unknown but implied (suicide bombing in market)"  
Victim: "Market-goers (number unknown)"

**How to Avoid:** - Consider context - Suicide bombers detonate (implied action) - Market bombings have victims (implied casualties) - Note “Unknown” rather than leaving blank

## 11. Training Exercises

### Exercise 1: Event Identification

**Instructions:** Read the article and identify how many distinct violent events are described.

**Article:**

Nairobi, Kenya - Armed robbers attacked a bank in downtown Nairobi on Monday,   
killing a security guard and escaping with an undisclosed amount of money.   
Police are searching for the gang.  
  
In a separate incident on the same day, suspected Al-Shabaab militants attacked   
a police station in Garissa County, near the Somali border. At least 3 police   
officers were killed in the assault. The militants also burned a police vehicle   
before fleeing.

**Question:** How many events? What are they?

**ANSWER:**

TWO EVENTS  
  
Event 1: Bank robbery in Nairobi  
- Actor: Armed robbers (criminal)  
- Victim: Security guard (1 killed)  
- Location: Downtown Nairobi  
- Date: Monday  
- Type: Criminal Violence > Armed Robbery > Bank Robbery  
  
Event 2: Attack on police station in Garissa  
- Actor: Al-Shabaab militants  
- Victims: 3 police officers (killed)  
- Location: Garissa County  
- Date: Monday (same day, different location)  
- Type: Political Violence > Terrorism or Rebellion  
  
Reasoning: Different actors, different locations, different motivations,   
different nature of violence = separate events despite same date

### Exercise 2: Taxonomy Classification

**Instructions:** Classify the following events using the full taxonomy (L1, L2, L3, L4).

**Event A:**

"A suicide bomber detonated explosives at a church in Maiduguri during Sunday   
service, killing 18 worshippers. Boko Haram claimed responsibility."

**YOUR CLASSIFICATION:**

L1: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
L2: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
L3: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
L4: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ANSWER:**

L1: Political Violence  
L2: Terrorism  
L3: Bombing/Explosive Attack  
L4: Suicide Bombing  
  
Reasoning: Terrorist group, civilian target (worshippers), suicide bombing method,   
symbolic target (church) = classic terrorism case

**Event B:**

"Fulani herders and farming communities clashed over grazing rights in Benue   
State, killing at least 15 people from both sides. The violence erupted after   
cattle destroyed crops."

**YOUR CLASSIFICATION:**

L1: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
L2: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
L3: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
L4: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ANSWER:**

L1: Communal Violence  
L2: Pastoralist-Farmer Clashes  
L3: Grazing Conflict  
L4: N/A  
  
Reasoning: Herders vs. farmers, resource conflict (land/crops), communal groups   
not state or organized armed groups = pastoralist-farmer violence

### Exercise 3: Coreference Resolution

**Instructions:** Identify all mentions of the same actor and mark coreference.

**Text:**

"Al-Shabaab militants attacked a military base in Somalia on Tuesday.   
The insurgents killed 8 soldiers before government reinforcements arrived.   
They then fled the area. The Islamist group later claimed responsibility   
on social media."

**Question:** List all mentions of the actor and show coreference.

**YOUR ANSWER:**

Mention 1: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
Mention 2: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
Mention 3: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
Mention 4: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
Coreference Group ID: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ANSWER:**

Mention 1: "Al-Shabaab militants"  
Mention 2: "The insurgents"  
Mention 3: "They"  
Mention 4: "The Islamist group"  
  
Coreference Group ID: ACTOR\_1  
  
All refer to same entity: Al-Shabaab  
  
Normalized Actor: "Al-Shabaab"

### Exercise 4: Date Normalization

**Instructions:** Normalize the following temporal expressions to YYYY-MM-DD format.

Article publication date: **Thursday, March 21, 2024**

| Expression | Your Answer | Confidence |
| --- | --- | --- |
| “yesterday” |  |  |
| “on Tuesday” |  |  |
| “last weekend” |  |  |
| “three days ago” |  |  |
| “early this month” |  |  |

**ANSWERS:** | Expression | Normalized Date | Confidence | |————|—————–|————| | “yesterday” | 2024-03-20 (Wednesday) | High | | “on Tuesday” | 2024-03-19 | High | | “last weekend” | 2024-03-16 or 2024-03-17 (Sat/Sun) | Medium | | “three days ago” | 2024-03-18 (Monday) | High | | “early this month” | ~2024-03-05 (approximate) | Low |

### Exercise 5: Complete Annotation

**Instructions:** Fully annotate this short article.

**Article:**

Gao, Mali - Rebels attacked a UN convoy near Gao on Wednesday, killing   
two peacekeepers from Burkina Faso. The convoy was traveling between Gao   
and Kidal when it was ambushed by armed militants using automatic weapons   
and rocket-propelled grenades. A UN spokesperson confirmed the attack and   
said three other peacekeepers were wounded. No group has claimed responsibility.  
  
- Source: BBC  
- Publication Date: 2024-03-22 (Friday)

**YOUR ANNOTATION:**

Event\_ID: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
Actor (Who):  
- Text: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- Normalized: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- Type: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- Confidence: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
Victim (Whom):  
- Text: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- Normalized: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- Deaths: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- Injuries: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
Location (Where):  
- Text: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- Normalized: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- Coordinates: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
Date/Time (When):  
- Text: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- Normalized: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
Weapon/Method (How):  
- Text: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- Tactic: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
Taxonomy:  
- L1: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- L2: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- L3: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- L4: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
- Confidence: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ANSWER KEY:**

Event\_ID: EX\_001\_EVT\_001  
  
Actor (Who):  
- Text: "armed militants"  
- Normalized: "Unknown armed militants"  
- Type: "Non-state armed group (identity unknown)"  
- Confidence: 0.6 (no specific identification)  
  
Victim (Whom):  
- Text: "two peacekeepers from Burkina Faso... three other peacekeepers"  
- Normalized: "2 Burkinabe UN peacekeepers (killed); 3 UN peacekeepers (wounded)"  
- Deaths: 2  
- Injuries: 3  
  
Location (Where):  
- Text: "near Gao... traveling between Gao and Kidal"  
- Normalized: "Road between Gao and Kidal, Gao region, Mali"  
- Coordinates: ~16.5, -0.2 (approximate midpoint)  
- Confidence: 0.7 (general area but not exact location)  
  
Date/Time (When):  
- Text: "on Wednesday"  
- Normalized: 2024-03-20 (Wednesday before Friday publication)  
- Confidence: 0.95  
  
Weapon/Method (How):  
- Text: "automatic weapons and rocket-propelled grenades"  
- Normalized: "Automatic firearms; RPGs"  
- Tactic: "Ambush"  
  
Taxonomy:  
- L1: Political Violence  
- L2: Rebellion/Armed Insurgency  
- L3: Ambush  
- L4: N/A  
- Confidence: 0.8  
- Reasoning: "Armed militants ambushing military convoy (peacekeepers) =   
 insurgent tactic. Could be terrorism but targeting of military forces   
 suggests rebellion/insurgency."

## 12. Troubleshooting Guide

### Problem: Can’t determine if one event or multiple

**Solution:** - Ask: Different time? Different location? Different actors? - If ANY are different → probably separate events - If unsure: default to MORE events (better to split than merge) - Document reasoning in notes

### Problem: Article gives conflicting information

**Example:** “Sources say 10-20 people were killed” or “Al-Shabaab or local militia (unclear)”

**Solution:** - Record the range/ambiguity: “Deaths: 10-20 (range reported)” - Note alternatives: “Actor: Al-Shabaab or local militia (article unclear)” - Mark confidence as LOW - Flag for adjudication

### Problem: Can’t find location on map

**Solution:** - Try alternate spellings - Try searching “[place name] + [country]” - Try searching in local language if known - If truly cannot find: Mark “Unable to geocode” and use approximate region coordinates - Note in comments: “Specific location not found on maps; used regional approximation”

### Problem: Don’t know which taxonomy category

**Solution:** 1. Go to Taxonomy Section 5 (Decision Rules) 2. Find similar scenario 3. Apply decision logic 4. If still unclear: document both options, mark LOW confidence, flag for adjudication 5. It’s OK to be uncertain - documenting uncertainty is valuable

### Problem: Article is in summary/listicle format

**Example:** “Violence this week: Monday - attack in X, Tuesday - bombing in Y, Wednesday - clash in Z”

**Solution:** - Treat each listed incident as separate event - May have limited information for each - Mark confidence as appropriate for available info - Note: “Article format is summary list with limited detail”

### Problem: Article discusses violence generally without specific incidents

**Example:** “Violence has been increasing in the region with frequent attacks”

**Solution:** - This is NOT a specific event - Do NOT annotate - Note: “Article discusses violence trends but describes no specific incident” - Skip this article

### Problem: Annotation template fields don’t fit this event

**Solution:** - Use “Notes” field extensively - Use “Other” or “Multiple” categories when available - Document what doesn’t fit - Suggest template improvements to supervisor - Still complete all required fields as best possible

### Problem: Made error and need to fix

**Solution:** - Review your work before finalizing - If found before submission: simply correct - If found after submission: document correction in separate log - Note: “Corrected [field] from [old value] to [new value] - reason: [explanation]”

### Problem: Running out of time / annotation taking too long

**Solution:** - **Don’t rush** - quality over speed - If taking >45 min per article consistently: discuss with supervisor - May need more training or may be encountering unusually complex articles - Track time: note which articles took longest and why - Efficiency improves with practice

## 13. Appendices

### Appendix A: Quick Reference Checklist

**For Every Article:** - [ ] Read entire article - [ ] Identify number of events - [ ] Annotate Article Metadata - [ ] For each event: - [ ] Mark entities (actors, victims, locations, dates, weapons) - [ ] Extract 5W1H - [ ] Classify using taxonomy - [ ] Rate confidence - [ ] Note relationships - [ ] Quality check - [ ] Save annotations

### Appendix B: Taxonomy Quick Reference

**Level 1 Categories (4):** 1. Political Violence 2. Criminal Violence 3. Communal Violence 4. State Violence Against Civilians

**Most Common Level 2:** - Rebellion/Insurgency - Terrorism - Political Repression - Armed Robbery - Ethnic Conflict - Pastoralist-Farmer Clashes - Extrajudicial Killings

### Appendix C: Contact Information

**For Questions/Issues:** - **Project Lead:** [Your name/email] - **Taxonomy Questions:** [Advisor contact] - **Technical Issues:** [Tech support] - **Adjudication:** [Lead annotator]

**Response Time:** - Urgent issues: Within 24 hours - General questions: Within 48 hours - Adjudication: Within 1 week

### Appendix D: Annotation Progress Tracking

**Daily Log Template:**

Date: \_\_\_\_\_\_\_\_\_\_  
Annotator: \_\_\_\_\_\_\_\_\_\_  
  
Articles Completed: \_\_\_\_\_\_\_\_\_\_  
Events Annotated: \_\_\_\_\_\_\_\_\_\_  
Flagged for Review: \_\_\_\_\_\_\_\_\_\_  
Questions/Issues: \_\_\_\_\_\_\_\_\_\_  
Time Spent: \_\_\_\_\_\_\_\_\_\_ hours  
  
Notes:  
- Challenges encountered:  
- Interesting cases:  
- Lessons learned:

### Appendix E: Glossary of Terms

**Event:** A distinct violent incident with specific actors, victims, location, and time

**Entity:** A real-world object mentioned in text (person, organization, location, date, weapon)

**Coreference:** When different words refer to the same entity

**Trigger:** The word or phrase that indicates violent action

**Normalization:** Converting varied text expressions to standardized forms

**Annotation:** The process of marking and labeling information in text

**Taxonomy:** Hierarchical classification system for organizing categories

**Confidence Score:** Numerical rating (0-1) of certainty in annotation

**Adjudication:** Process of resolving disagreements or unclear cases

**5W1H:** Who, What, Whom, Where, When, How

**END OF ANNOTATION GUIDELINES**

**Document Status:** Version 1.0 - Ready for Annotator Training  
**Last Updated:** October 2025  
**Contact:** [Your contact information]

**Next Steps for Annotators:** 1. Read this entire document (4-6 hours) 2. Complete training exercises (Section 11) 3. Take qualification test 4. Begin annotation with supervision 5. Regular check-ins with project lead

**Good luck with your annotations! Your careful work is essential to the success of this research project.**