

## Keynote Abstract: Too soon? The limitations of AI for event data

**Clionadh Raleigh**

University of Sussex  
5 Stoneham Road, Hove, Sussex Bn3 5hj  
c.raleigh@sussex.ac.uk@abc.org

### *Abstract content*

Not all conflict datasets offer equal levels of coverage, depth, use-ability, and content. A review of the inclusion criteria, methodology, and sourcing of leading publicly available conflict datasets demonstrates that there are significant discrepancies in the output produced by ostensibly similar projects. This keynote will question the presumption of substantial overlap between datasets, and identify a number of important gaps left by deficiencies across core criteria for effective conflict data collection and analysis, including:

**Data Collection and Oversight** : *A rigorous, human coder is the best way to ensure reliable, consistent, and accurate events that are not false positives.* Automated event data projects are still being refined and are not yet at the point where they can be used as accurate representations of reality. It is not appropriate to use these event datasets to present trends, maps, or distributions of violence in a state.

**Inclusion** : *Inclusion criteria should allow for accurate representations of political violence, while being flexible to how political violence has changed.* Who is considered a relevant and legitimate actor in conflict is pre-determined by the mandate of the dataset; the definitions, catchment, and categorization are critical, as they tell a user who and what is likely to be included.

**Coverage and Classification** : *Clear, coherent, and correct classifications are important for users because conflicts are not homogenous: disorder events differ in their frequency, sequences, and intensity.* Event types that reflect the variation of modalities common across conflicts and periods of disorder are basic, central components of insightful and useful analysis.

**Use-ability and Transparency** : *Datasets must be useful and useable if they are to be relied upon for regular analysis, and users should be able to access every detail of how conflict data are coded and collected.* Use-ability is closely tied to straightforward, consistent inclusion criteria and clear methodology.

**Sourcing** : *Extensive sourcing — including from local partners and media in local languages — provides the most thorough and accurate information on political violence and demonstrations, as well as the most accurate presentation of the risks that citizens and civilians experience in their homes and communities.*