Annex I – THE NATURE OF DATA

I.1 INTRODUCTION

The term "data" is used here in the broadest possible sense. Data are mere basic observations, which are real, not assumed or hypothesized. Data should neither be underrated nor overrated in their importance. They should certainly never be confused with the concept of absolute truth. Data can be thought of as the foundation of an epistemological system that includes information, knowledge, truth, and wisdom. This Annex discusses some of the common classifications of types of data and their treatment; for a more thorough treatment the reader should consult textbooks and other scholarly works on epistemology and statistics.

I.2 BASIC TERMINOLOGY

The word "data" is plural**;** the singular form is "**datum,**" but the terms "**point data**" or "**datapoint**" or "**observation**" are often used with the same meaning, e.g., a single number or observation. The term "**dataset**" is a grouping of related data "**Raw data**" are the initial form of the data, as collected, before any statistical analysis, enumeration, etc. is done. An "**observation period**" is the time during which data are collected. Collected data are considered to be "**samples**" of a larger underlying "**population**" of similar data. "Sample size" is the number of samples in a particular dataset.

I.2.1 Quantitative Data

Data that are related in an ordered or numeric scale. (Also called numeric data.)

- **Rank Data**: The simplest type of quantitative data, needing only to be capable of being sequentially ordered, e.g., one datum is equal or bigger, or better, or more effective, etc., than each other in the set; for convenience, such data may be numerically ranked starting with a value of 1 at one end of the scale. The area of mathematics known as non-parametric statistics is used to describe, summarize, compare, and analyze such data.
- **Measurement Data**: Data collected using a measuring instrument (e.g., a thermometer, force meter, stop watch). Such data are often immediately transduced and digitized for computer analysis.
- **Frequency Data**: Data obtained by counting the number of times an event occurs. If the total number of observed events is known, frequency data may be expressed as a proportion or percentage. If the duration of the observation period is known, such data may be expressed as a rate (e.g., events per hour).
- **Duration Data**: Data obtained by measuring the amount of time that a process was occurring. If the duration of the observation period is known, such data may be expressed as a proportion or percentage.
- **Latency Data**: The time between a designated start of an observation period to the occurrence of a particular event.
- **Interval Data**: The time between two events within an observation period.
- **Dose Response Data**: Data that correlate "dose" to "response." The dose may be any appropriate measure of the amount of non-lethal technology (e.g., joules of energy, watts of electricity, density of a chemical) applied and the response be any conceivable, quantifiable effect of that technology,

including intended and unintended effects. Such data are often illustrated in "dose-response curves" in which dose (low to high) is plotted on the x- axis and response (low to high) is plotted on the y-axis. As the points in such curves usually represent measures of central tendency (e.g., means) of a set of like data, dose-response curves often display an indication of variability or confidence, as well. The medians of such curves, in which 50% of the datapoints fall above and 50% fall below, are often used as a shorthand summary of an effective dose (called "Effective Dose 50", or ED_{50}). Such summary measures may then be compared for intended versus unintended effects to compute a margin of safety or safety factor. While well-established dose-response curves for multiple effects of the same technology is considered ideal, such data currently exist in only a few cases.

- **Threshold Data:** Threshold data are a measure of the minimum dose required to produce some level of an effect in an individual or a population. There are several mathematical approaches to estimating threshold, but the most complete is the use of dose-response data, as described above. Threshold data are usually accompanied by a measure of variability or uncertainty with respect to the population response that is being estimated. The concept of threshold is usually tied to the percentage of a population that displays a particular effect at the threshold dose. For occupational health and safety standards, the threshold dose for an undesired effect may be based on a very small percentage of the population being affected; for such standards, the permissible limit is often much lower than the threshold dose, in order to provide an additional safety factor.
- **Binary Data (or Digital Data):** Binary data are quantitative data represented by a series of the digits "1" and "0," called bits. Ultimately, most quantitative data are converted to binary when processed by a computer. The number of bits used to represent a quantitative datum determines the possible resolution (bit depth) of the measurements; for example, if 8 bits are used, the maximum resolution of the measurement is 2^8, or 256 different levels.

I.2.2 Qualitative Data

Qualitative data include descriptions and classifications of events without reference to any calibrated scale. They are often subjective and/or opinionated. Numbers are arbitrary identifiers of qualitative data and might as well be substituted with words or other symbols. For example, the observation by a battle participant that a particular non-lethal weapon was effective in a particular military operation is an example of qualitative data. There are procedures for maximizing information from qualitative data. Qualitative data are often valued for their usefulness in developing hypotheses that can be tested using experimental, quantitative methods.

I.3 DATA QUALITY – THE COLLECTION OF DATA

All data are not collected equal. Particular data may be described with various qualifying terms, such as good, bad, reliable, biased, etc. The principles of data quality may be derived from common usage. Good data are **not** data that agree with the experimenter's cherished theory. They are merely data that have been collected in a reliable, reproducible manner, with opportunities for error and bias minimized as much as possible. Any data can be biased or outright falsified. However, as the quality of the data increases, the potential for falsification and bias becomes easier to detect, control, and avoid. Hearsay and other verbal report, given long after an incident and by an observer with a clear vested interest in the use of the data, has the highest potential for bias and inaccuracy. Standardized procedures for data collection (e.g., entry forms, time limits, structured questionnaires, written classification criteria, multiple observers, etc.) can increase the quality of data. Poor measurement devices (e.g., biased, unreliable, or uncalibrated) can result in both increased variability and systematic error. The gold standard for high quality data is that obtained from a well-designed, controlled experiment using a calibrated measuring instrument and skilled observers who have little knowledge of the

hypothesis being tested or of which treatments are administered to which groups (often called a "blind" observer). The insistence on such procedures does not impugn the honesty or trustworthiness of the observer; research has shown that experimental bias can occur unconsciously in even the most honourable and well-intentioned observers. Even qualitative data may be judged according to the known accuracy and expertise of the observer. Thus a qualitative observation from an experienced and trusted individual may be more relevant and valuable than any number of datapoints or amount of statistical analysis. However, data quality should not be equated to data relevance, as rarefied experimental data may or may not provide valid predictions of effects in the real world.

I.4 STATISTICAL ANALYSIS OF DATA

Raw data provide little information without further processing. Statistics is the discipline that uses mathematics and probability theory to process raw data into various types of information. Statistics is both a well established and evolving discipline and only the barest introduction is attempted here.

- **Descriptive Statistics**: Descriptions associated with central tendency (mean, mode, medium) or variability (e.g., range, standard deviation) of a dataset.
- **Estimating Populations Parameters**: Statistical procedures used to estimate the parameters (e.g., central tendency, confidence levels) of an underlying population based on the data obtained from samples from that population.
- **Hypothesis Testing**: Statistical procedures used to draw conclusions from data, for example to decide if two or more sets of data have been drawn from the same or different populations, i.e., to estimate if they are statistically significantly different from one another?
- **Statistical Significance**: Using calculations of size, central tendency, and variability of the samples, and assumptions regarding independence or dependence of the samples, the predicted direction of the difference, the sampling method used, and the distribution of the underlying populations, measures of statistical significance estimate the probability that observed data and datasets are related by chance or otherwise. The most common question is whether two or more datasets come from the same or different underlying populations. An example question for non-lethal weapons is whether or not a particular exposure, e.g., to a putative acoustic weapon, has had an effect on task performance. The experimenter collects quantitative data under the conditions of exposure and no exposure and compares the results using an appropriate statistical test. The result of the test indicates the probability (p) that the samples differ by chance. By convention, the probability level of 0.05 is accepted as being "statistically significant." The label "$p < 0.05$" indicates that "the probability that the apparent relationships between these datasets will appear by chance is less than .05 (i.e., one time in 20). It is important that interpreters of statistical results realize that if a probability level of 0.05 is used as the criterion, results labelled as being "statistically significant" will have a purely chance relationship 5% of the time.
- **Using Appropriate Statistics**: Statistical tests make assumptions regarding the mathematical nature of the sampled data and the underlying population from which the sampled data are collected. These include assumptions about the type of data (see discussion above in 1.2.1) and the distribution of data within the underlying population. The statistical approach selected must match the nature of the data being analyzed. There is no more common error in data analysis than the inappropriate use of parametric statistics on non-parametric data, with the use of the mean as a measure of the central tendency of ranked data being an egregious example.

I.5 CHALLENGES FOR NLW HUMAN EFFECTS DATA COLLECTION

Researchers on the human effects of NLW face a difficult challenge in that they need to collect data on human subjects, yet current regulations for human experimentation do not adequately address the types of experiments they need to perform. Most regulations for human use are oriented toward improving medical treatment or elucidating fundamental knowledge about human physiology or behaviour. Measuring the risks or effectiveness of a weapon system, albeit a non-lethal weapons, is not a familiar justification for permitting human experimentation. Existing regulation, ethical considerations, risk-benefit calculations, available alternatives, and the need for reliable applicable data all affect this complicated issue, which can only be noted in this paper.

I.6 DATA AND MATHEMATICAL MODELS: GARBAGE IN – GARBAGE OUT

Raw data, analysed data, and the relationships supported by the results of statistical comparison are often used to create mathematical models that describe these relationships in a form that can be readily graphed, simulated, and more easily understood; such models are called "**descriptive models."** When such models are used to predict relationships and outcomes that have not actually been used to create the model, they become **"predictive models."** Accurate predictive models are extremely useful, however their accuracy must be validated by additional data collection before they can be relied upon. An issue that sometimes arises in non-lethal weapon testing is the claim or assumption that a descriptive model is also an accurate predictive model, without the validating step of demonstrating that the model can accurately predict new data. Models are often based on very limited observations made in highly controlled, unrealistic situations using numerous untested assumptions. Worse is when models are initially based on poor quality data analysed within inappropriate statistics. The most excellent model cannot make up for bad data. Models can be used to guide the collection of additional data and new data can be used to improve the models; however, models do not eliminate the need to collect new data.