DATA GLOSSARY

Why do we need a data glossary?

The data language can be complex, with numerous terms, many of which can be used interchangeably. These are the key ones we use at FFI with definitions of what we mean when we say them.

This glossary has been developed to assist those engaging in data activities including the data literacy course. Here is a link to our MEL Glossary

COMMON DATA TERMS

Accuracy	Data accuracy refers to the correctness and reliability of data (from Atlan). Accurate data correctly represents the real-world scenario or event it is supposed to depict. It's free from errors, especially those that occur due to incorrect data entry or faulty processes (from Atlan).
Aggregated data	Data clustered to represent large groups. Opposite of disaggregated data.
Algorithm	A well-defined, step-by-step computational procedure that takes an input (or set of inputs) and produces an output (or set of outputs). It's a sequence of instructions designed to perform a specific task or solve a specific problem (Cormen et al., 2009).
Application	A type of software that assists people in performing a particular activity on their device.
Attribute field	An attribute field is a piece of data or information associated with a particular object or record contained in a database. They are used to store characteristics or properties associated with the object or record and can be utilized to filter, search, or sort data (from Arena Solutions). With GIS data, this nonspatial information about a geographic feature in a GIS, usually stored in a table and linked to the feature by a unique identifier (from arcgis.com).
Big data	Extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions.
Categorical data	Data recorded in intervals or by groups. There might not be a logical order to the categories. For example, gender, age group or species group.
Causation	The relationship between two variables where a change in one variable directly results in a change in the other variable.
Consistency	Data consistency refers to the accuracy, reliability, and uniformity of data within a system or database, ensuring that information remains the same across different instances or datasets.
Continuous data	Data measured along a scale. For example, height, weight or temperature.

Correlation	The process of establishing a relationship or connection between two or more - variables. In a data context, correlation refers to a statistical measure that describes the extent to which two variables change together.
Dashboard	A visual interface that aggregates and displays data in a cohesive and easily interpretable manner. Dashboards often provide real-time insights and are designed to aid decision-making by presenting key performance indicators (KPIs), metrics, and other vital data points in graphical formats such as charts, graphs, and tables.
Data	Collection of related facts, usually organized in a particular format such as a table or database and gathered for a particular purpose (Tanner et al., 2020)
Data analysis	Examining and interpreting data to draw conclusions about the information with the goal of looking for patterns, and themes and making sense of and summarising the data.
Data cleaning	Process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.
Data compliance	Data compliance refers to the practice of ensuring that data handling, storage, and management align with applicable laws, regulations, standards, and internal policies. This means that organizations must manage personal, financial, and other sensitive data in ways that meet the legal requirements for privacy, security, and integrity.
Data democratisation	Data democratization seeks to make data more accessible to non-technical users, in part, by making the tools that access the data easier to use. This includes tools that do not require advanced technical skill or deep understanding of data analytics to use (IBM, 2023).
Data governance	Collection of processes, roles, policies, standards, and metrics that ensure the effective and efficient use of information in enabling an organisation to achieve its goals.
Data interoperability	Data interoperability is the ability of different systems, software applications, and organizations to exchange, interpret, and use data effectively and seamlessly. It ensures that data can flow across various platforms without compatibility issues, allowing diverse systems to work together and understand shared information.
Data lifecycle	The steps that data moves through during its life.

Data management	Process to ensure that diverse data sets can be efficiently collected, integrated/processed, labelled/stored, and then easily retrieved through time by people who want to use them.
Data outputs	Data output is the process of presenting, displaying, or transmitting processed information from a system, software application, or database to an end-user or another system. This output can take various forms, such as visual displays on screens, printed reports, exported files, graphs, and data feeds, depending on the intended use and audience.
Data practices	Includes design of methods, planning and data management, collection, storage, processing, and visualisation and analysis of information for interpretation and use.
Data privacy	The ability of a person to determine for themselves when, how, and to what extent personal information about them is shared with or communicated to others.
Data processing	The systematic transformation, organization, and analysis of raw data into a meaningful and useful format. This stage involves cleaning, organising, and transforming raw data into a usable format.
Data registry	Collates important information about the datasets held by a project.
Data repository	
Data security	Is the process of safeguarding digital information throughout its entire life cycle to protect it from corruption, theft, or unauthorized access
Data sharing	The practice of making data available to others, either within an organization or publicly.
Data storage	Data storage is the process of saving digital information in a structured format that allows for retrieval, management, and modification. It can involve various media, such as hard drives, cloud services, or databases, and is essential for preserving data for future use, analysis, or compliance purposes.
Data validation	Data validation is the process of ensuring that data entered into a system or database meets predefined accuracy, quality, and format criteria. It checks for correctness, completeness, and reasonableness to reduce errors and improve data reliability.
Data verification	The process of ensuring that data is accurate, complete, and consistent. It is a critical step in data quality management and involves comparing data against a known and trusted source to check for errors and inconsistencies.
Data visualization	The graphical representation of information and data. By using visual elements like charts, graphs and maps, data visualization tools provide an accessible way to see and understand trends, outliers and patterns in data.
Database	A large amount of information stored in a computer system in such a way that it can be easily looked at or changed.

Descriptive statistics	Descriptive statistics summarize and describe data – they help you explain what the data says. This can include averages like mean, mode and median. Other descriptive statistics include standard deviation, variance, percentiles, and more.
Digital Object Identifier (DOI)	A DOI name is a digital identifier of an object, any object — physical, digital, or abstract. DOIs solve a common problem: keeping track of things. Things can be matter, material, content, or activities. Designed to be used by humans as well as machines, DOIs identify objects persistently. They allow things to be uniquely identified and accessed reliably. You know what you have, where it is, and others can track it too (DOI Foundation, 2023)
Disaggregated data	Data that is presented in sub-categories (for example by gender or education level). This is opposite to aggregated data which are clustered to represent large groups.
Discrete data	Data that can only take certain fixed values.
Derived data	Data which is generated from raw data.
Evidence	FFI adopts a broad definition of the term evidence, acknowledging that it can range in format, quality and relevance. We recognise that evidence can come from a range of sources from tacit knowledge, generated through experience, values, and observations including local and indigenous knowledge to more formal (empirical) evidence or explicit knowledge that is systematically researched and documented in reports, papers, guidelines.
Geospatial data	
Indicators	A sign, clue or marker that helps answer our monitoring and evaluation questions and indicates if we are on the pathway to progress.
Inferential statistics	Inferential statistics make predictions or suggestions about a bigger population than the data you've captured. They work by using statistical models to test the confidence of our results.
Inputs*	The financial, human, and material resources used for the conservation interventions. (Adapted from OECD definition)
Integrity	Data integrity is about maintaining the consistency, trustworthiness, and reliability of data throughout its lifecycle. It ensures that the data remains unchanged from its source and has not been tampered with or altered without proper authorization (from Atlan).
Metadata	Structured documentation about the dataset, I.e., data that describes other data. "Meta" is used to describe something self-referential.
Method	An information/data gathering or analysis activity, procedure or technique. For example, focus groups, semi-structured interviews, surveys, participatory rapid appraisal, direct species population surveys, and camera trap survey.
Methodology	The overall package of an analytical framework, research design, methods, and an assessment process that links everything together (e.g., SAPA guidelines).

Milestone*	Sometimes also called 'intermediate results' or 'benchmarks', milestones are situations that, when achieved, indicate that the interventions are on the way to reaching the desired target. These are often defined when the overall impact will take a long time to see, but progress towards it can be measured in this way.
Monitoring	The collection of repeated evidence to determine progress over time on use of resources, project implementation, and the success of project interventions in bringing about desired changes.
Occupancy	Proportion of sites "occupied" by a species.
Open data	Data which is publicly available and can be re-used without restrictions
Outputs*	The direct products of activities that have been implemented. They are relevant to the achievement of project outcomes and are typically tangible and easy to measure. For example, if we were to host a training, the output would be the number of persons who would have attended the training.
Parameter	A numerical or other measurable factor forming one of a set that defines a system or sets the conditions of its operation.
Precision	A parameter is a specific value or characteristic that serves as an input to a function, model, or process, helping to define, control, or influence its behavior or outcome. Parameters set boundaries or limits within which the process operates and can be adjusted to alter results.
Qualitative data	Data that tells us about the qualities rather than the number of something. Often, they are represented by text, and sometimes images.
Quantitative data	Data that can be represented as numbers including both continuous data measured along a scale & categorical data recorded in intervals or by groups.
Raw data	Collected data that hasn't been processed or formatted (also known as primary data).
Responsible data	The duty to ensure people's rights to consent, privacy security and ownership around the information processes of collection, analysis, storage, presentation and reuse of data, while respecting the values of transparency and openness. (Responsible Data Forum, working definition, September 2014).
Rigorous	The thoroughness, precision, and strict adherence to established methodologies and standards when collecting, analysing, and interpreting data.
Robust	The ability of a statistical method, model, or system to produce consistent and reliable results despite the presence of anomalies, outliers, or other challenges in the data.
Sample	A smaller set of data that a researcher chooses or selects from a larger population using a pre-defined selection method. These elements are known as sample points, sampling units, or observations.

Sampling strategy	The process of identifying your population and determining how best to select a sample from it (from The Research Competition). The approach used to select a sample representative of a larger population so that generalisations and testing hypotheses are valid, and replicable. There are many sampling strategies to choose between including simple random, semi-random, systematic, targeted, purposive, and stratified sampling.
Sensitive data	Information or data, that if disclosed or accessed without proper authorisation, are likely to cause harm or negative impacts and/or be disadvantageous for affected people, fauna and/or flora.
Software	Programs or operating information used by a computer. Note: Applications are a type of software that assists people in performing a particular activity on their device (e.g., Survey123, WhatsApp).
Statistical model	The use of mathematical models and statistical assumptions to generate sample data and make predictions about the real world. A statistical model is a collection of probability distributions on a set of all possible outcomes of an experiment. (from HEAVY.AI)
Survey	A survey is a method of collecting data or information from a group of individuals or entities, typically to gather insights, opinions, behaviors, or characteristics about a specific topic.
Tool	Specific information/data gathering instrument used within a method. Fauna & Flora consider "data tools" to include applications, programs, software or hardware used within the data lifecycle (for example, Survey123, camera traps, or drone).
Tabular data	Data organised as a table with rows and columns.
Thematic analysis	Thematic analysis is a qualitative data analysis method that involves reading through a data set (such as transcripts from in depth interviews or focus groups), and identifying patterns in meaning across the data to derive themes (from DelveTool.com).
Variable	A value collected or measured within a survey or held within a dataset or database. A variable is any characteristic, number, or quantity that can be measured or counted.