PART
# 01 WHAT IS A DATA ONTOLOGY?

COMMON APPROACH
TO IMPACT MEASUREMENT

## Linking Impact Data: How a data ontology can ease impact data collection and analysis

This series of documents explores the data ontology—a crucial part of the digital infrastructure that will be needed to improve impact measurement in the years ahead.

This document defines what a data ontology is and distinguishes it from vocabularies (sometimes called glossaries) and taxonomies.

The impact investing field has vocabularies and taxonomies. Impact Norms' 5 dimensions of impact are a vocabulary. The IRIS+ impact taxonomy is, as the name states, a taxonomy. The SDGs (goals, targets and indicators) are a taxonomy.

An impact data ontology would connect these taxonomies: linking stakeholder characteristics to SDGs and IRIS metrics. These connections, well structured in data, will provide more context around and more insights into the data that is already being collected. The ontology increases insight through better connections rather than even more data collection.

The proliferation of impact data requires a new, more sophisticated understanding of data itself—its flows between actors; context-sensitivity; and the ways it is or could be informing decisions. Taxonomies provide labels and categories. Data ontology examines the meaningful relationships in information systems. We need both, but the former is not enough any longer. An impact data ontology would facilitate linkages among many standards.

**KEY TAKEAWAY**

Ontologies are better at capturing context. Taxonomy identifies hierarchical relationships within a category; ontologies allow multi-dimensional relationships, connecting many vocabularies and taxonomies.

# How is a data ontology different from a vocabulary or a taxonomy?

Many impact investors are already familiar with the glossaries (called data vocabularies by computer scientists) and taxonomies. Ontologies build on vocabularies and taxonomies to build interconnections and contexts.

### VOCABULARY

The names and terms for things. It ensures that the meaning of data is clear, explicit, and consistent across users.

### TAXONOMY

A set of data classification rules that organizes data by shared characteristics into hierarchies.

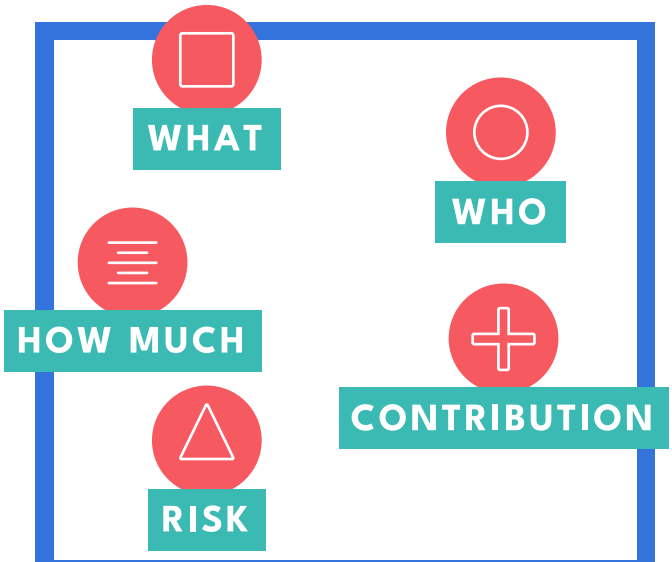
### ONTOLOGY

Concepts and categories that show the properties and relations between data. It connects data taxonomies and vocabularies.

## DATA VOCABULARY

A data vocabulary is a way of making sure that the meaning of the data is clear, explicit, and consistent across the organization (or between organizations). A data vocabulary thus defines and describes the data, and ensures that everyone uses the data in the same way. These can be created within a single organization. They can also be shared standards.

There are many vocabularies in the impact measurement field. Two examples are the **Impact Management Norms' 5 Dimensions of Impact** (Figure 1.1) and **Impact Data Categories**. The 17 SDGs are also a data vocabulary. The defining feature of data vocabularies is that they are just lists; they don't organize the data.




Examples of vocabularies in impact measurement include Impact Management Norms' 5 dimensions of impact (pictured above*), the 17 goals of the UN SDGs, and the IRIS+ metrics.

**Impact Norms 5 Dimensions of Impact, from Impact Frontiers*

*Figure 1.1*

# DATA TAXONOMY

A data taxonomy is a set of data classification rules that organizes data by shared characteristics into hierarchies. Taxonomies are useful because they allow for the quick identification of relevant information. They help put related items together.

Organizations use data taxonomies to organize impact data into categories that make sense for them. The IRIS+ System by GIIN, for example (see Figure 1.3), identifies a taxonomy of impact categories, themes, goals and metrics.

Biological classifications offer a familiar example of the difference between a vocabulary and a taxonomy. Biology names animals. This is the vocabulary as shown in Figure 1.2. Biology also classifies organisms into a taxonomy that includes kingdom, phylum and class. In our example, the vocabulary names the animals in the forest and the taxonomy helps to communicate a hierarchical relationship among them based on common characteristics. (We should note that some data scientists sometimes argue that the biology taxonomy is not, strictly speaking, a true taxonomy, but the argument is technical. We have chosen to use this example because it is one that readers will be familiar with and because it lets us put cute forest animals in a document about data ontologies.)

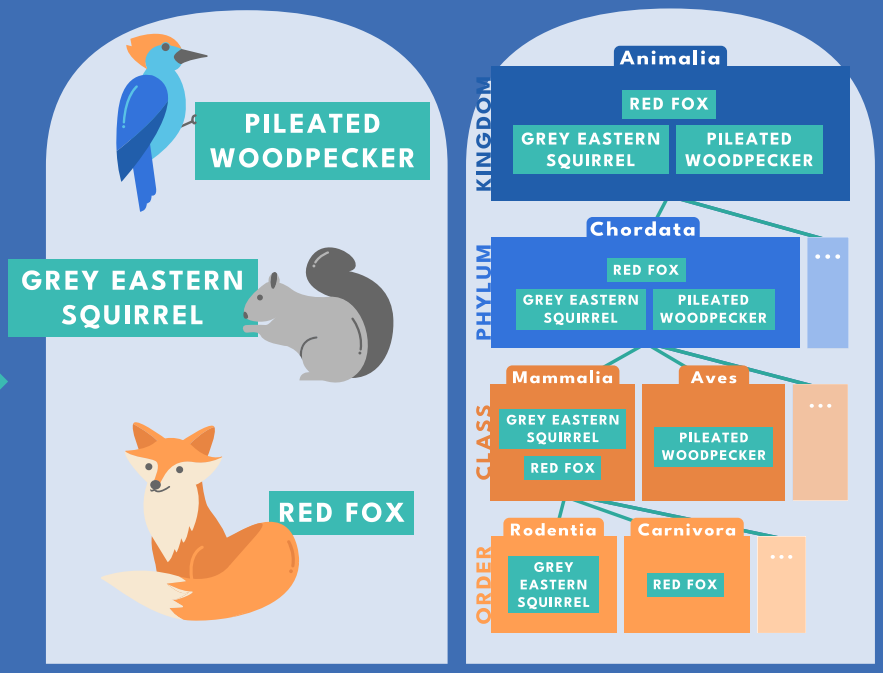



*Figure 1.2*

Examples of data taxonomies:
*IRIS taxonomy*[1]*:*


*Figure 1.3*

Data taxonomies are not effective at connecting other taxonomies. A taxonomy is unable to represent the connections between the IRIS+ taxonomy and the Impact Management Norms data vocabulary, even though **many IRIS+ metrics have been linked to the Impact Norms**. That linkage is the role of a data ontology.

[1] *https://iris.thegiin.org/document/iris-thematic-taxonomy/*

# DATA ONTOLOGY

A data ontology is a richer representation of data. It connects data taxonomies and vocabularies.

In our forest animal example, an ontology can represent the taxonomy of kingdom, phylum, class, order, etc., as well as food web relationships, habitat relationships, and parent-child relationships. This is illustrated in Figure 1.4.

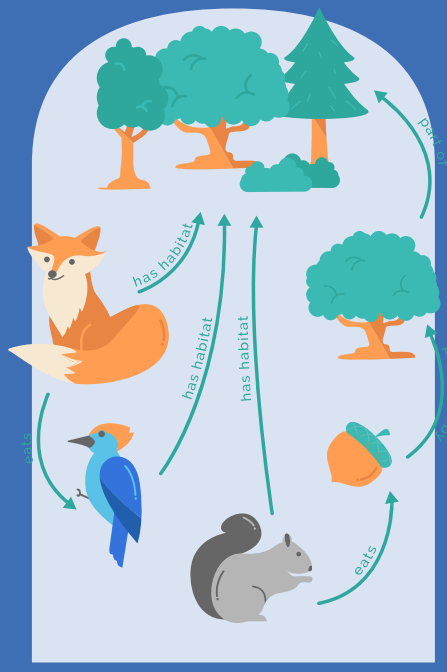



*Figure 1.4*

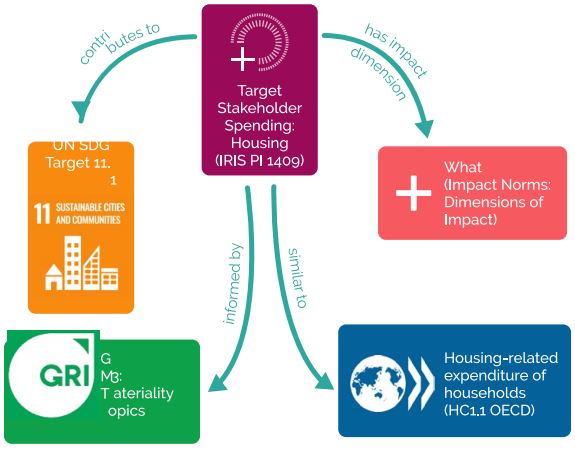



*Figure 1.5*

Figure 1.5 is an ontology that represents the relationship between IRIS+ Metric PI 1409 Target Stakeholder Household spending with other standards. The ontology notes that IRIS Metric PI1409 is part of the "What" dimension of the 5 Dimensions of Impact, and it contributes to UN SDG Target 11.1, which is part of UN SDG Goal 11. It is similar to OECD indicator HC11, "housing-related expenditures of households." In this fictitious example, the organization creating this report selected IRIS+ PI1409 after undertaking a materiality assessment in accordance with GRI 3.

A data ontology explicitly defines the different types of data and the relationships between the different types of data. It names and defines each thing and how those things connect. You can imagine a data ontology like an elaborate spreadsheet template with all the column headings already in place, except that because it is an interconnected web of data, it is much more complex than can be reasonably captured in one spreadsheet.

A data ontology defines these relationships in machine-readable language so that the software platforms know that a metric can be associated with many outcomes, but that a given metric report can only ever be associated with one time period (e.g. a year). Defining these relationships in a common way is essential for the interoperability of data from one software platform to another. It ensures that when an investee sends data that has a metric associated with three outcomes connected to various SDGs and IRIS+ themes and categories, the investors' software platform has a place to store the metric, the outcomes, and the relationships between them. Those many-to-many relationships are very challenging to maintain in spreadsheets. And it is those many-to-many relationships that add context and insight that is analytically useful.

To make the concept of an impact data ontology more concrete, we can use the example of the Common Impact Data Standard. The **Common Impact Data Standard** is an ontology that provides the integrative map enabling linkages and synergy across many standards. It is a shared reference model that structures relationships linking dimensions of existing frameworks into a cohesive lattice.

CONTINUE TO:

PART 02 **THE PROBLEMS THAT AN IMPACT DATA ONTOLOGY CAN SOLVE IF WIDELY ADOPTED**