Why

We want language and notation for recording elements of a set.

Definition

A *dataset* of size n *in* (or *from*) a non-empty set A is an element of the direct product A^n. In other words, it is an n-tuple. In the case that we speak of a dataset, we call its coordinates the records. So, for example, we call the first coordinate the first record and the second coordinate the second record. The ith coordinate is the ith record, for $i = 1, \ldots, n$.

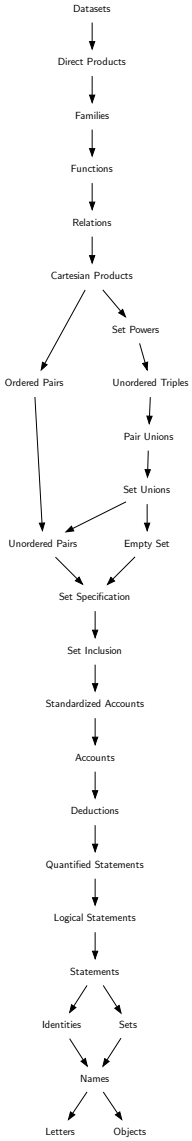
Although the terminology dataset is standard, datasets are not sets. Datasets may contain an element multiple times. Thus we use the single word "dataset," in contrast with the two words "data set," which may together suggest that the object involved was a set.

In the case that the dataset is from a set A which is a product of two sets (say, $\mathcal{U}$ and $\mathcal{V}$, so that $A = \mathcal{U} \times \mathcal{V}$) we say that it is a dataset of *record pairs*. A record pair is sometimes called an *observation*.

Notation

Let A be a non-empty set. Although we usually denote sequences by $(a_1, \ldots, a_n)$, we will denote datasets by $(a^1, \ldots, a^n)$; notably, using a superscript. Often A is itself a set of sequences, and so we will want to refer to the ith component of

a^1 by a_i^1.


Datasets
Direct Products
Families
Functions
Relations
Cartesian Products
Set Powers
Ordered Pairs
Unordered Triples
Pair Unions
Set Unions
Unordered Pairs
Empty Set
Set Specification
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Identities
Sets
Names
Letters
Objects