DATAFLOW DIAGRAM: A HIGH-LEVEL MODELING TOOL

Structural part:

(1) *External nodes:*

- These are the data-sources and data-sinks external to the system, and are drawn as boxes.
- An external node has at least an arc to it or from it (or both) from one or more internal (specifically, process) nodes. There are no links joining two external nodes.
- An external node may represent a data-store or a process external to the system being modeled, i.e., the part described by the internal nodes.

(2) *Internal nodes:*

(a) Process nodes;

- These are drawn as circles/ellipses, and each of them has ≥ 1 incoming arcs and ≥ 1 outgoing arcs. Exception: a random-number generator or some initialization process (with no incoming arcs), and a process that removes files or frees up memory (with no outgoing-arcs). These exceptions are not considered here.
- The other end of an arc joining to/from a process can be an internal or external node.

(b) Data-store nodes:

- These are drawn as boxes with open vertical-sides (horizontal-sides) or the sides slightly in-set.
- Each data-store has ≥ 0 incoming-arcs and ≥ 1 outgoing-arcs, but no arcs join two data-stores or a data-store and an external node.

Each internal node (save data-stores and other exceptional process-nodes mentioned above) should be on a path from a source-node to a sink-node. Also, the internal nodes should form a connected graph, when viewed as an undirected graph.

(3) *Labels:*

- Nodes have distinct labels (verb-phrases for process-nodes and noun-phrases for data-stores).
- Each arc (dataflow) is labeled by one or more data items and these labels are distinct from node labels. Two different arcs from a process-node may have common labels, but not so for arcs from two different process-nodes (see next item).
- All labels to and from a data-store are the same, and hence this may be omitted and assumed to be the same as the label of data-store itself. One exception is a data-store name like "d_1 or d_2", in which case a dataflow to the data-store may be labeled any of "d_1", "d_2", and "d_1, d_2"; the dataflows from the data-store are similarly labeled.

Semantic Part:

Process: A major function (it may be implemented by several smaller functions).

Data-store: An interface between asynchronous processes (functions), which do not have a calling-relationship. It includes the case where the data must survive beyond the current execution of the software.

Labels: The set of labels on all arcs from a process-node together gives the outputs of the process, all or only a part of which may be used in another processes and in a data-store. The order in which the outputs are created is not represented in DFD. Similarly for inputs of a process (note that no two arcs to a process have a common label).

A dataflow diagram should have enough details to allow high-level walk-throughs/use-cases. One should avoid cycles as much as possible; if one must have a cycle, then there should be some indication of where the computation begins on the cycle.