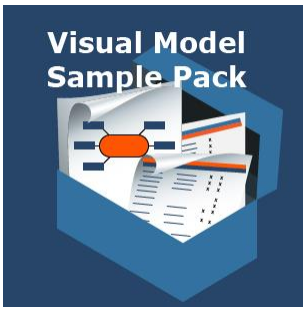


DATA FLOW DIAGRAM

VISUAL MODEL OVERVIEW

A Data Flow Diagram illustrates how information flows through, into, and out of a system.

This Data Flow Diagram was created to illustrate a data transformation process that collected job posting information from 5 job boards, consolidated the job postings together into one file, and output that file to several job aggregators. (Indeed.com is an example of a job board aggregator that still exists in 2013.)

***BABOK® Guide* Connection**

Technique 9.6
Data Flow Diagram

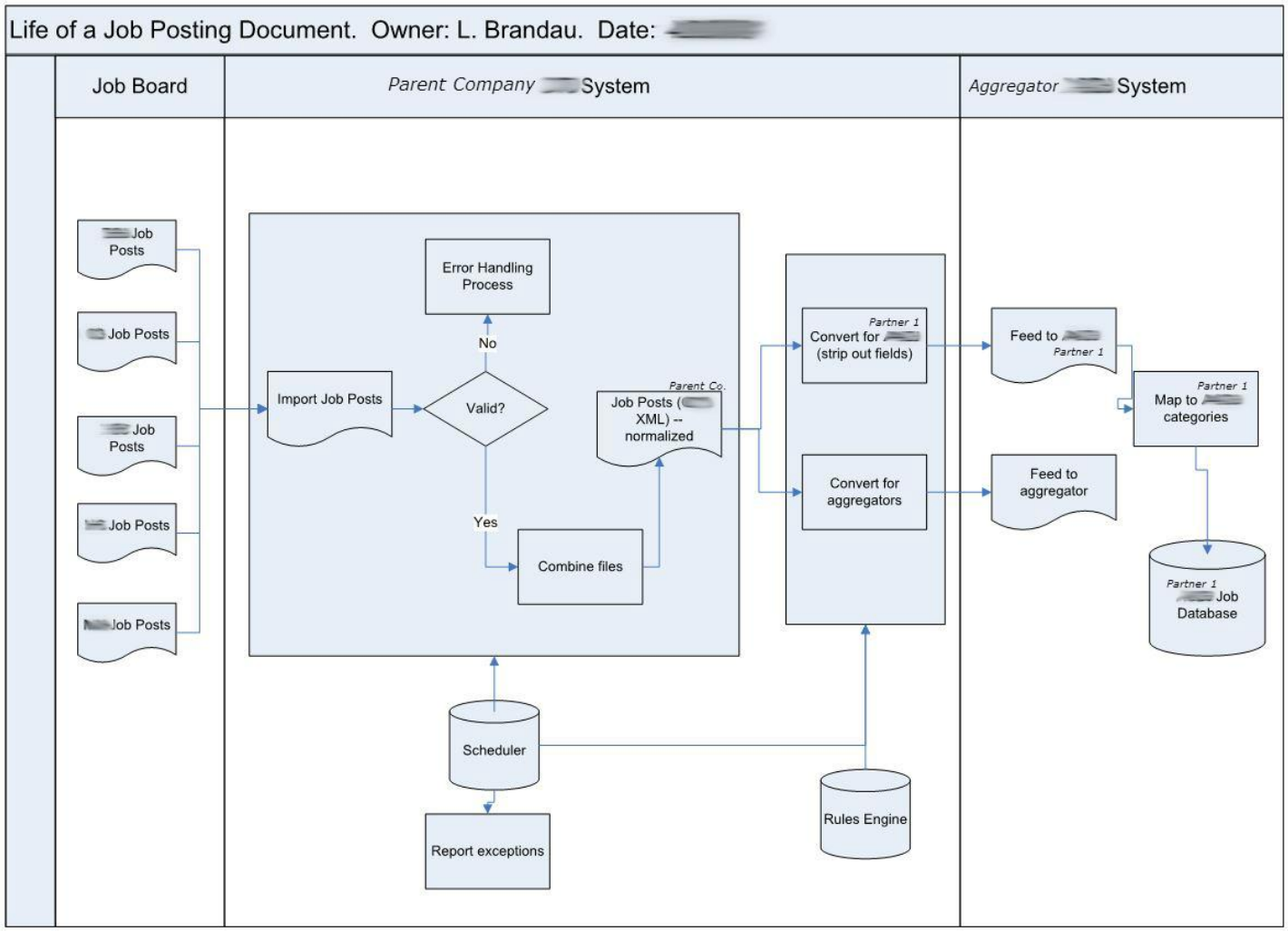
To show how information is input, processed, stored, and output from a system.

This model was created to communicate with the technical team and all business stakeholders about how the job feed aggregation process would happen. It was included in the Scope Document for the project.

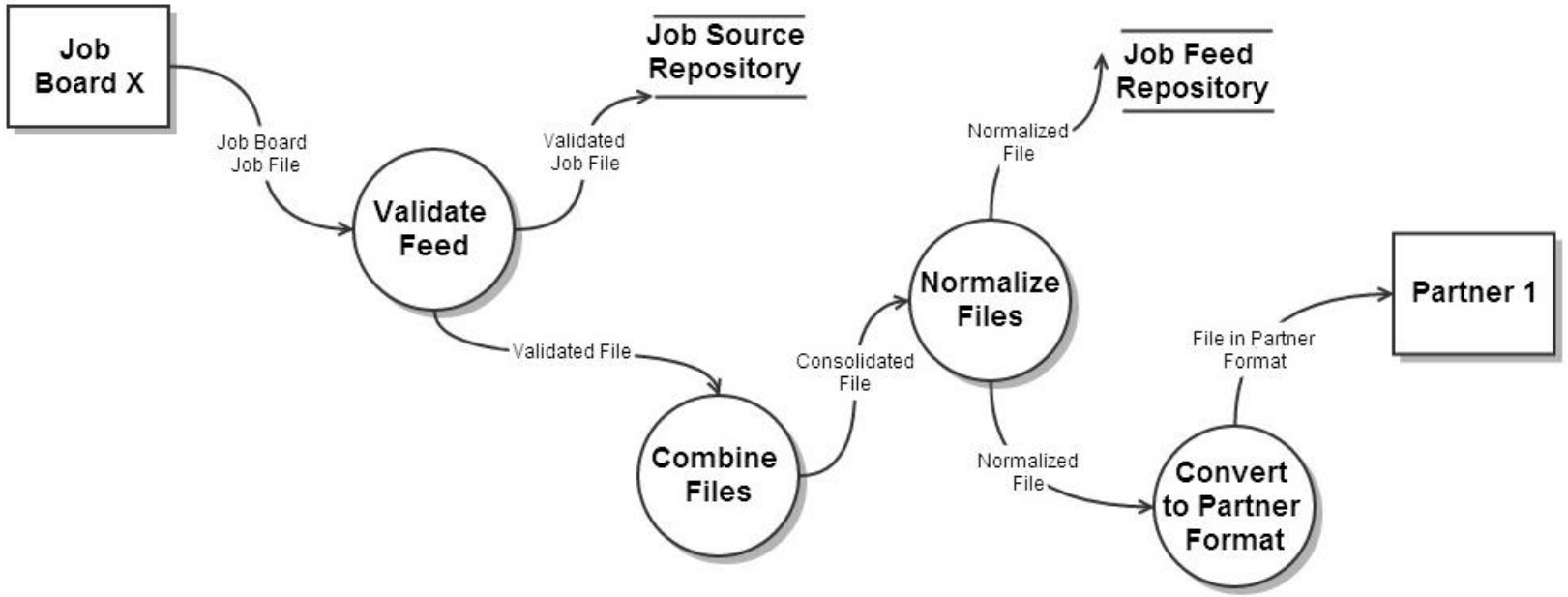
This particular example is a non-standard example of how to put together a Data Flow Diagram. While it realizes the underlying goals of a data flow diagram, it uses notation from a Workflow Diagram rather than one of the standard data flow diagram notations. This is because I was not familiar with the standard notations at the time and the model worked as it was. For the purposes of illustrating how this diagram could have been captured in the standard notations referenced in the *BABOK® Guide*, updated formal samples are included following the original diagram.

THE MODEL

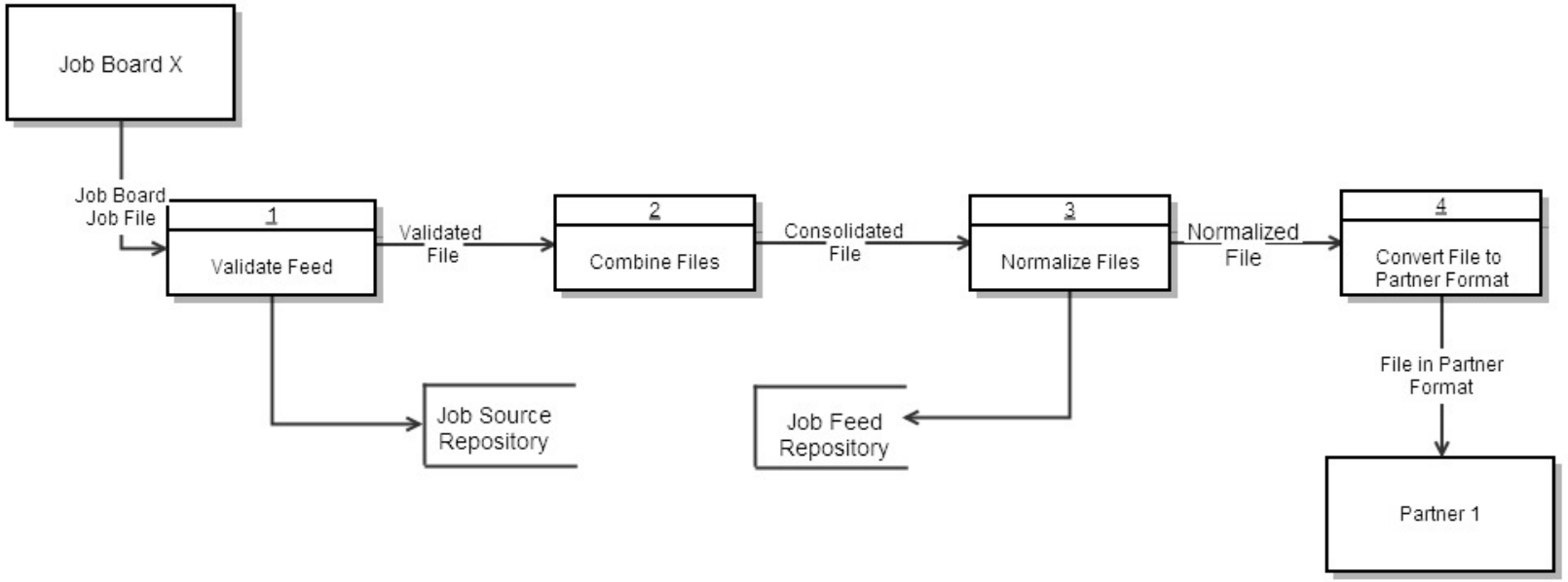
Original Model – Data Flow as a Workflow Diagram



Revised Model – Data Flow Diagram (Yourdon Notation)



Revised Model 2 – Data Flow Diagram (Gane-Sarson Notation)



DETAILED DESCRIPTION

Swimlanes

Swimlanes, the vertical bars breaking the flow chart into logical areas, are used to represent systems. In this case, the Parent Company system is the primary system. The other systems are external systems.

Data Sources

 The data sources or inputs for the process are represented by the rectangles with the curved bottom in the "Job Board" swimlane. This is typically referred to as the "Document" shape. In Data Flow Diagramming terms, these would be considered external entities. In the Yourdon notation example, external entities are represented by text with a line above and a line below. In the Gane-Sarson example, external entities are represented by rectangles without the right line.

Data Processing

The two blue boxes within the Parent Company System contain a sequence of processes and checks that are performed to manipulate the data. In Data Flow Diagramming terms, these are called Data Processes and Data Flows. Here I used the Activity Box from a Process Flow Diagram to capture these processes and show the flow of data. In the Yourdon notation example, the Data Processes are captured in circles and the Data Flows using curved lines. In the Gane-Sarson example, the Data Processes are captured in numbered boxes and the Data Flows using straight lines.

As is typical, the rules governing each of these data processes was further defined in supporting documentation, which included XML schemas, matrix of fields and rules, and data translation tables. (See *Data Model* for an example of such a matrix.)

Engines Supporting Data Processing

The "Scheduler" and "Rules Engine" data store elements illustrate that the system would be built to store rules in external files and repositories. These rules would

be used to implement the data processing and data flow. This part of the diagram represents the System Architecture more than the Data Flow and so is not included in the Data Flow Diagrams completed using the Yourdon or Gane-Sarson notations.

Data Outputs

The data outputs of this process are represented using the "Document" shape in the Aggregator System swimlane. In Data Flow Diagramming terms, these would be considered external entities, which are discussed above in the Data Inputs section.

The Aggregator System swimlane shows the data processing to happen by one of our partners, in this case Partner 1. This was an important element to show the narrowed scope of the data processing we needed to perform internally before distributing jobs to aggregators. In this case, Partner 1 planned to handle some of the required post-processing to fulfill the terms of our agreement.

TERMINOLOGY

Data Flow Diagram is the formal term used within the *BABOK® Guide* to refer to this type of model. Other common terms include:

- **Data Integration Flow**
- **Data Processing Diagram**

This diagram also represents certain aspects of the following types of diagrams:

- **Relationship Map –** Because each swimlane represents an organization, this model shows one aspect of the relationship between these organizations.
- **State Diagram –** Because each step in this workflow illustrates a state of job posting data, it has some parallels to a state diagram which shows the lifecycle of a data entity or class.
- **System Architecture** – The data store elements represent an aspect of the system architecture. *See System Architecture Diagram.*

- **Workflow Diagram** – The Activity boxes, validation check, and lines are drawn from standard workflow modeling notation. *See Process Flow Diagram.*

POSSIBLE USES

Use this type of model:

- To show how data will flow within a system or among different systems.
- To identify interfaces between systems that require data sharing or processing.
- To identify existing data processes.
- To identify key data processes to be created.

HOW TO CREATE A SIMILAR MODEL

Follow these steps to create a similar model for your project.

1. Identify the known inputs and outputs for the data flow. (Often a data flow sequence starts by understanding how the data exists today and identifying what the desired output is.) Add these to your diagram as external entities.
2. Identify the key data processes that need to happen to transform the input into the output. Add these to your diagram as data processes.
3. Identify the key data stores that will be necessary to capture data in a specific state for a period of time. Add these to your diagram as data stores.
4. Identify aspects of the data process where data is moved from one place to another. Add these to your diagram as data flows.

WATCH OUT FOR

A standard Data Flow Diagram does not show exceptions or every possible scenario. Attempting to do so would complicate the flow. If this information becomes necessary, incorporate elements of a Process Flow Diagram to accompany your Data Flow Diagram.

The Yourdon and Sane-Garson notations are more likely to be new to your stakeholders and therefore more difficult to interpret. Also, the level of abstraction required for this type of diagram is counter-intuitive for typical stakeholders. When using these notations, be sure to explain the notations and level of abstraction to all stakeholders as you review the diagram with them. Consider supplementing with appropriate additional models to drill into the further detail.