-->

Data Flow Diagrams

Last Modified on 06/01/2026 11:08 am EDT

What is a Data Flow Diagram?

A data flow diagram illustrates how data moves throughout an application, service, or business process. It provides a visual representation of how information enters, traverses, is stored within, and exits a system. A data flow diagram helps answer key questions such as:

- How does data enter the system?
- Where does data travel within the system?
- How is data processed or transformed?
- Where is data stored?
- How does data leave the system?
- What security controls protect the data throughout its lifecycle?

> Wharton Computing staff have access to Draw.io in Confluence which is a useful tool to create these diagrams!

Key Components of a Data Flow Diagram

A comprehensive data flow diagram should identify:

- Systems, applications, services, and infrastructure components that process, transmit, or store institutional data
- All entry points where data enters the environment
- All exit points where data leaves the environment
- Internal data flows between systems, services, and subsystems
- Trust boundaries and segmented network zones (such as firewalls or restricted environments)

Data Flow Requirements

Arrows should clearly indicate the direction of data flow. Each flow should include relevant information about the data being transmitted, including:

- Data classification level (highlight regulated or sensitive data where applicable)
- Transport protocol and transmission method
- Encryption used for data in transit
- Encryption used for data at rest
- Authentication mechanisms
- Authorization and access controls
- Storage location and storage type

Network segmentation boundaries, including firewalls or isolated environments, should also be clearly represented.

Process for Creating a Data Flow Diagram

Step 1: Create a Context Diagram

Begin with a high-level context diagram that illustrates the overall system and its interactions with users, external services, and connected systems.

Anyone reviewing the diagram should be able to quickly understand:

- The major system components
- Where data originates
- How data moves through the environment
- Where data is stored
- Where data exits the system

The context diagram serves as the foundation for more detailed diagrams.

Step 2: Identify Trust Boundaries and Firewalls

Clearly indicate network segmentation and trust boundaries, including:

- Firewalls
- Restricted network zones
- Third-party environments
- Cloud and on-premise boundaries

Dashed lines or labeled boundary markers may be used to visually distinguish these segments.

Step 3: Document System Processes

Add technical processes that occur within each system or entity, including:

- Data processing activities
- Application services
- Authentication services
- API communications
- Scheduled jobs or automated workflows

This step helps illustrate how data is transformed or handled throughout the system lifecycle.

Step 4: Label Data Flows

Each data flow arrow should include sufficient detail to support security and risk analysis.

Recommended information tags include:

- Data classification
- Encryption in transit
- Encryption at rest

- Service or protocol used
- Authentication method
- Authorization or access control model

Sensitive or regulated data should be clearly identified to support compliance and security review activities.