Data onboarding workflow

Learn Splunk's five-step process to build best-practice data onboarding workflows.

You can build an effective data onboarding workflow by mapping the data onboarding process into five phases.

These guidelines can help you streamline data requests, define the use case, validate data, and properly communicate the availability of new data.

Audience

- Developer
- Knowledge manager
- Program manager
- Project manager

For more about these roles, see Setting roles and responsibilities.

Guidelines for establishing a data onboarding workflow

The recommended data onboarding workflow consists of five steps: request data, define the use case, implement, validate, and communicate.

During the process, document your approach so your community is well informed. This can deflect many questions, establish user expectations, make your users more aware of their responsibilities, and teach your users how they can make effective contributions to the data onboarding process.

Step One: Request data

The data onboarding workflow begins with a request to add data. You can keep it as simple as an email, you can establish a formal process with templates and request requirements, as in the checklist attached to the bottom of this article, or you can even leverage an enterprise change control system.

1. Simplify your data requests. Capture only the essentials. Avoid using Splunk-specific terms, such as index name, field extractions, and so on.

2. Ask for specific known information. Ask for concrete details, such as data source host names and IP addresses, path, location, and access information; retention requirements (how long they need to keep the data); and a brief description of what the data represents. This will help you prioritize the request and define source types.
3. **Estimate data volume.** The requester may know their estimated data volume, but it may be more efficient for you to review the source location and do the math yourself. A near-maximum of the data volume (such as 95th percentile) works well with the Splunk licensing model. Do not take the average or median data volume, since the actual data volume will exceed that threshold 50% of the time. For help estimating data volume, download the [Data Volume Calculator](#) from Splunkbase.

### Step Two: Define the data

Hold a data definition meeting to clarify details of the request. Be thorough during this stage to reduce the chance of miscommunication or misunderstanding and to help the implementation phase go more smoothly.

1. **Get a sample of the data prior to the data definition meeting.** Review the sample data up front to verify:
   - Splunk's ability to access the data
   - Permissions to the data within Splunk
   - Forwarders (if needed)
   - Dependency on a modular input (if needed)
   - Any data retention and storage considerations

2. **Verify the requester's commitment.** If the requester is enthused and prioritizes this meeting with you, you'll know this request is truly important to them and the fruit of your labors will be enjoyed. If the requester does not make this meeting a priority, it may indicate they are not as invested in this use case as they could be.

3. **Define the use case with the requester.** Validate Splunk-relevant details about the information, such as event breaks, timestamps, and other critical source-type elements. Discussing the use case with the requester enables you to uncover searches or dashboards that will be immediately useful to them. The scope should be to assist the requester with their initial search and dashboard setup to get them going, not a commitment to own their use case.

4. **Empower the requester to own the use case.** Make sure the requester has completed the appropriate education path to enable them to own their use case. The requester should be responsible to own further search-time activities. For more information about how to establish education paths, see User and Team Lifecycle.

### Step Three: Implement the use case

Once the data is defined, proceed with technical implementation.

1. **Build out search and reporting artifacts.** Use the information gathered in the define data step. Focus on value-add elements that only you can uniquely provide, such as tags, reports, saved searches, dashboards, forms, field extractions, and any other elements you have uncovered or nice-to-haves submitted by the requester.

2. **Ask for clarifications as needed.** Ask the requester if you need more information about the data, details or objectives of the use case during implementation.
**Step Four: Validate**

After developing the use case artifacts, validate that they achieve the expected results.

1. **Run through the use case in your lab.** Run the artifacts you created through testing in your own lab using sample data relevant to the use case.
2. **Invite the requester to validate the use case.** Have the requester review the results you generated from your tests to make sure the use case meets the requester’s expectations. Make any adjustments needed.

**Step Five: Communicate**

This phase ensures that each diverse data point added to an analytic (or KPI) directly contributes to business value.

1. **Send an announcement about the availability of the new data.** Communicate with the wider user community that the use case is available. This enables other users to consider how these data points might help them.
2. **Help the community understand current and potential use case(s) for the data.** In your announcement, suggest some creative applications of the data. Provide use case information that will help the community understand how this data can support stronger, data-driven decisions.
3. **Include details in the announcement.** Include details in your announcement, such as how to access the data (index, sourcetype, tag name), what the data represents (use information from the data request and/or data definition meeting, what knowledge objects exist for it already (fields, dashboards, saved searches, and so on).

**More resources**

- To learn about getting data into Splunk Enterprise, see [What data can I index?](#) in the *Splunk Enterprise Getting Data In* manual.
- For an introduction to data onboarding best practices, see the .conf2017 session [Data Onboarding: Where do I begin?](#)

[Splunk Data On-Boarding Checklist.pdf](#)