



GlobusWorld 2024

Research Automation with Globus Flows



Ada Nikolaidis
ada@globus.org



Automation with Globus Flows



- **Intro**

- What is it?
- Key concepts
- How to use flows

- **Dive Deeper**

- Automating Instrument Data at Scale (tomorrow)
- Q&A
- Office Hours



What is Globus Flows?

Flows Service



- A platform for managed, secure, reliable task orchestration
- Create, share, and run “flows”

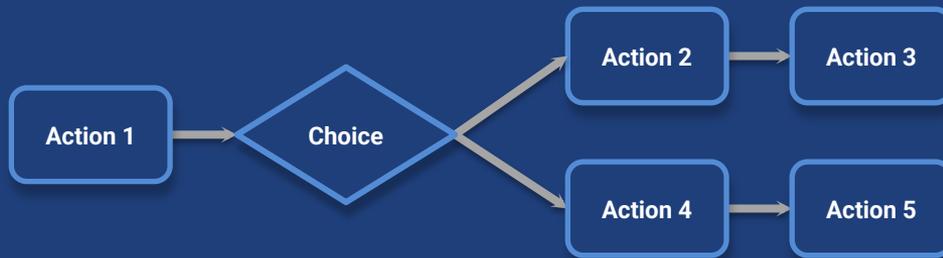


What's a Flow?



A sequence of steps...

- **Hosted**
- **Reusable**
- **Flexible**
- **Shareable**



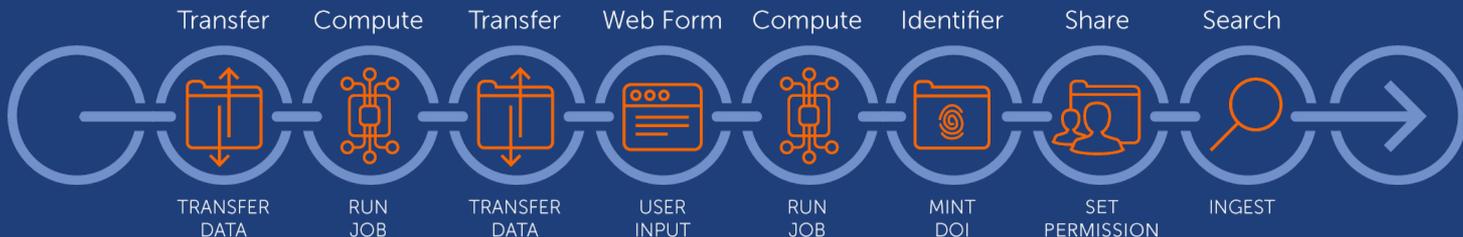


How Do You Use a Flow?



Each time a flow is started, it's called a *run*

- Start from Web App, CLI, Python SDK, API
- Provide input
- Performs a series of actions with that input
- Manage the run (evaluate, share)



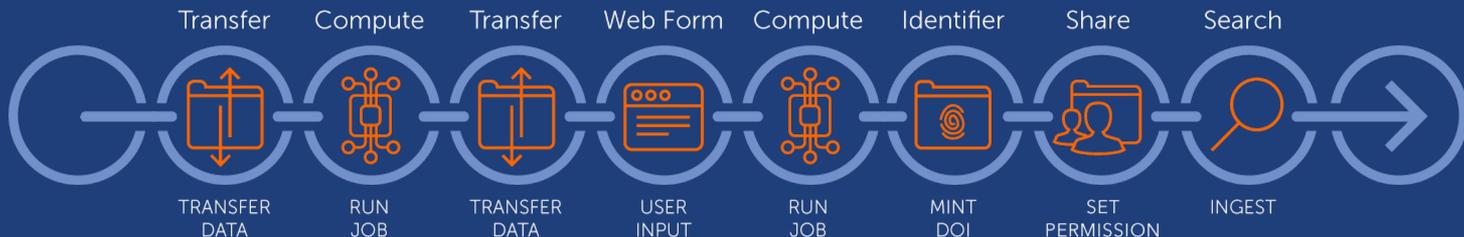


What's an Action?



An *action* is a special type of state

- **AWS Step Functions**
- **“States Language”**
- **Basic states like “Wait”, “Pass”, “Choice”**



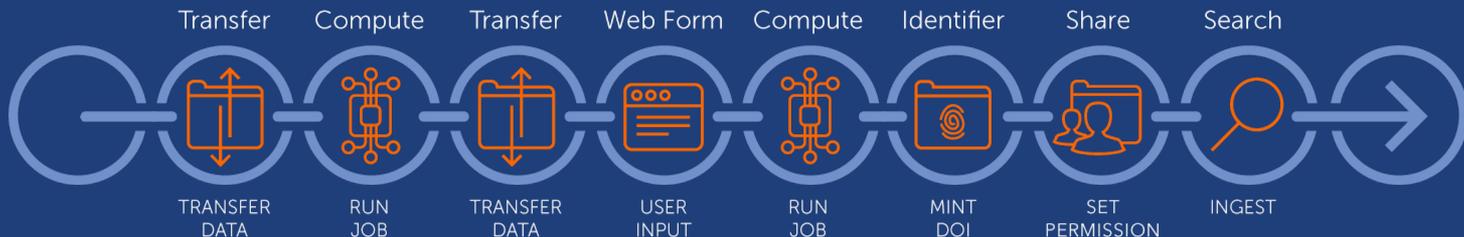


What's an Action?



...So what's an *action*?

- Operations you can perform on other resources and services
- *Managed* interactions
 - Authenticate, authorize, validate, store, auto-retry





Globus-Provided Actions



transfer 	delete 	mkdir 	ls 
stat 	identifier 	ingest 	delete 
notify 	input 	ACLs 	compute 



Examples of Flows



- **Two Stage Globus Transfer**
 - Transfer to temp location, then transfer to final destination
- **Move (Copy and Delete)**
 - Transfer to a destination and delete files on the source
- **Transfer and Set Permissions (coming soon)**
 - Transfer to a destination and share with users and groups



Demo: Running flows





What happens when you
run a flow?



Is this user allowed to run this flow?

Check input using flow's input schema

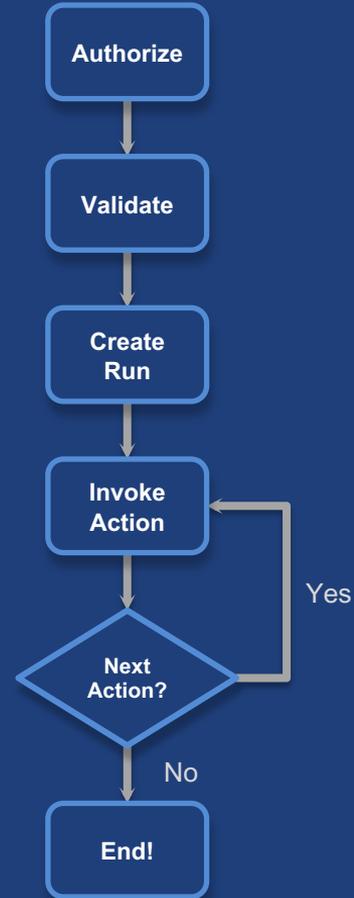
Register the run with the specified settings

Start the first (or next) action

...wait for the action to finish...

Is there another action after this?

The run has finished





Inactive Runs



Most often: A mapped collection

Notification: Runner receives an email

To grant permission: “resume”

Tip #1: Resume with the same application that started the run

Tip #2: Use guest collections when possible



Definition

Operations

Transformations

Conditions

Input Schema

Data requirements

Three key benefits:

1. Prevent errors
2. Guided input
3. Flow validation



Anatomy of an Flow Definition



```
1  {
2    "StartAt" "MyTransfer",
3    "Comment" "A simple file transfer flow using the Globus Transfer Action",
4    "States": {
5      "MyTransfer" {
6        "Type": "Action",
7        "ActionUrl": "https://transfer.actions.globus.org/transfer",
8        "Parameters": {
9          "source_endpoint": "6c54cade-bde5-45c1-bdea-f4bd71dba2cc",
10         "destination_endpoint": "31ce9ba0-176d-45a5-add3-f37d233ba47d",
11         "DATA": [
12           {
13             "source_path.$": "$.source_path",
14             "destination_path.$": "$.destination_path"
15           }
16         ]
17       },
18       "ResultPath": "$.MyTransferResult",
19       "WaitTime": 300,
20       "End": true
21     }
22   }
23 }
```



Anatomy of an Action



```
1  {
2    "StartAt": "MyTransfer",
3    "Comment": "A simple file transfer flow using the Globus Transfer Action",
4    "States": {
5      "MyTransfer": {
6        "Type": "Action",
7        "ActionUrl": "https://transfer.actions.globus.org/transfer",
8        "Parameters": {
9          "source_endpoint": "6c54cade-bde5-45c1-bdea-f4bd71dba2cc",
10         "destination_endpoint": "31ce9ba0-176d-45a5-add3-f37d233ba47d",
11         "DATA": [
12           {
13             "source_path.$": "$.source_path",
14             "destination_path.$": "$.destination_path"
15           }
16         ]
17       },
18       "ResultPath": "$.MyTransferResult",
19       "WaitTime": 300,
20       "End": true
21     }
22   }
23 }
```



Flows at Scale



Command Line Interface

Ad hoc scripting and integration



Python SDK

Applications, Services, Portals



New: Enhanced Validation



Enhanced Validation (Beta)

- **Available in: Globus CLI, Python SDK, API**
- **Catches errors that were previously impossible:**
 - Invalid references to input
 - Invalid parameters to actions
 - Conflicts in state type



Concepts

Flow

Run

Action

Action Provider



Q&A



Thank You!



Don't Forget...

- **Advanced session: *tomorrow***
 - Automating Instrument Data at Scale
- **Office Hours: *today and tomorrow***