

PSI Center for Scientific Computing,
Theory and Data

Jobs Release

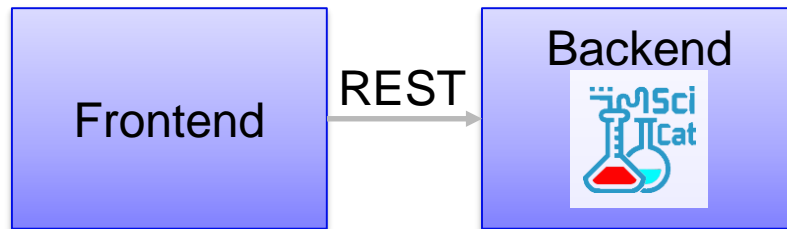
New configuration system for interoperability with external services

Spencer Bliven, Despina Adamopoulou, Sofya Laskina
PSI, 2 July 2024

- Daphne van Dijken (original lead)
- Spencer Bliven (substitute lead)
- Despina Adamopoulou
- Sofya Laskina
- Max Novelli
- Regina Hinzmann
- Nico Hayen
- Carlo Minotti
- Attila Nacsa
- Junjie Quan
- Laura Shemilt
- Marco Leorato

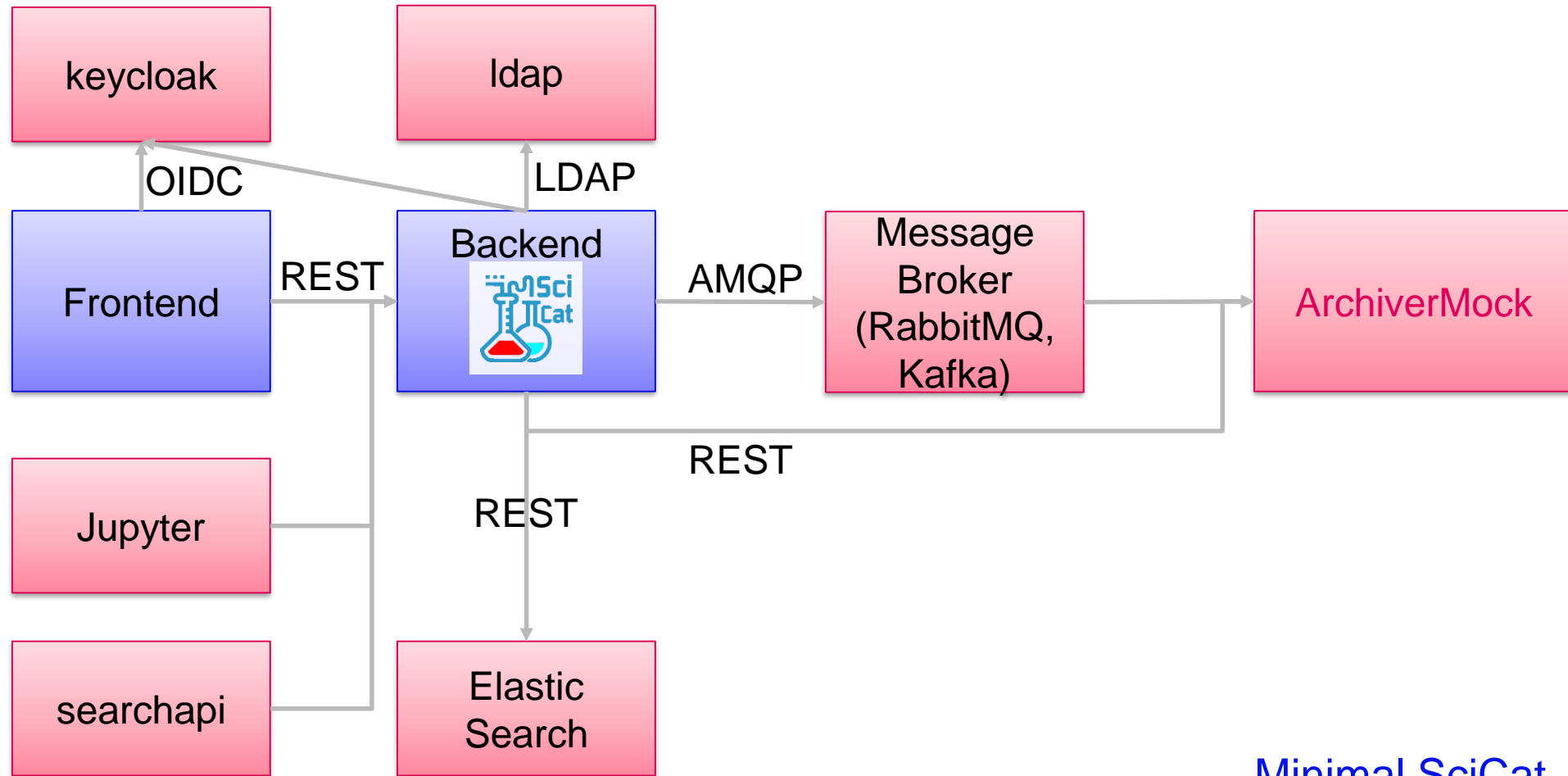
109 commits

```
> git diff --shortstat master release-jobs  
88 files changed, 4026 insertions(+), 3738 deletions(-)
```



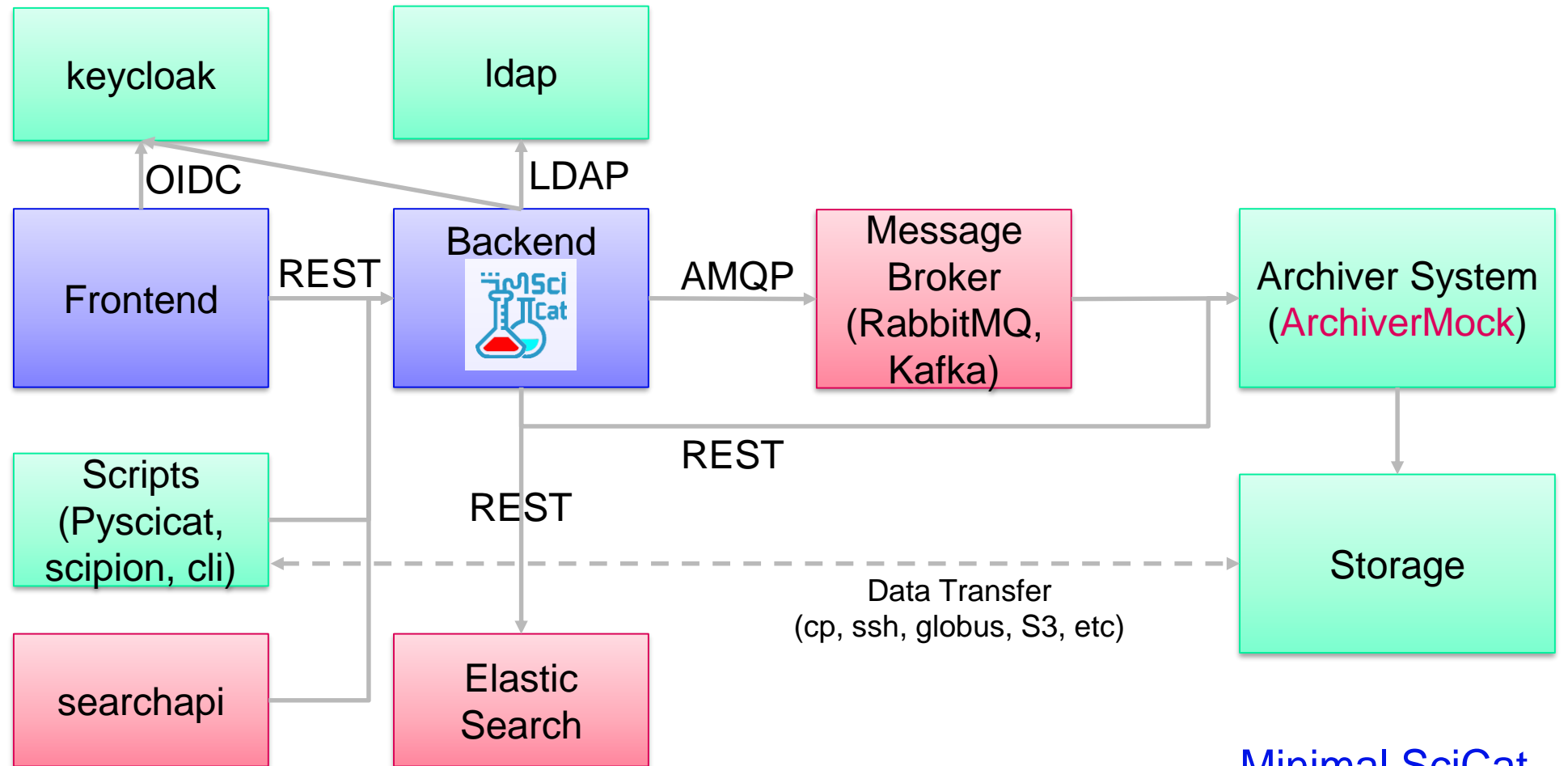
Minimal SciCat

SciCat Ecosystem



Minimal SciCat
SciCatLive

SciCat Ecosystem



Minimal SciCat
Other Containers
Central Services

Jobs in Scicat v3



- Fixed job types:
 - `archive`: Indicates that data should be moved to tape storage
 - `retrieve`: Indicates that data should be retrieved internally from tape storage
 - `public`: retrieve data to public storage location
- Hard-coded actions for events
 - Publish message to a Message Broker
 - Dispatch email following a template
- Limitations
 - Customization requires rebuilding SciCat images with custom js
 - Code spread among many locations, files (and site-specific monkey patches)
 - All dependencies are compiled into the backend, even if unused

User / Jobs / [c03b53c3-0964-4eb9-bd81-aa28e7c1f929](#) /

✉ Email Job Initiator	spencer.bliven@psi.ch
👤 Type	archive
⚙️ Creation Time	2022-04-29 07:29
⚙️ Job Params	{ "username": "oidc.bliven_s" }
✉ Date Of Last Message	2022-04-29 07:29
📁 Dataset List	"20.500.11935/31334ea3-53e1-48e9-8ac7-18f0325f3f82"
📅 Created At	2022-04-29 07:29
📅 Updated At	2022-04-29 07:29

Example archive job in v3

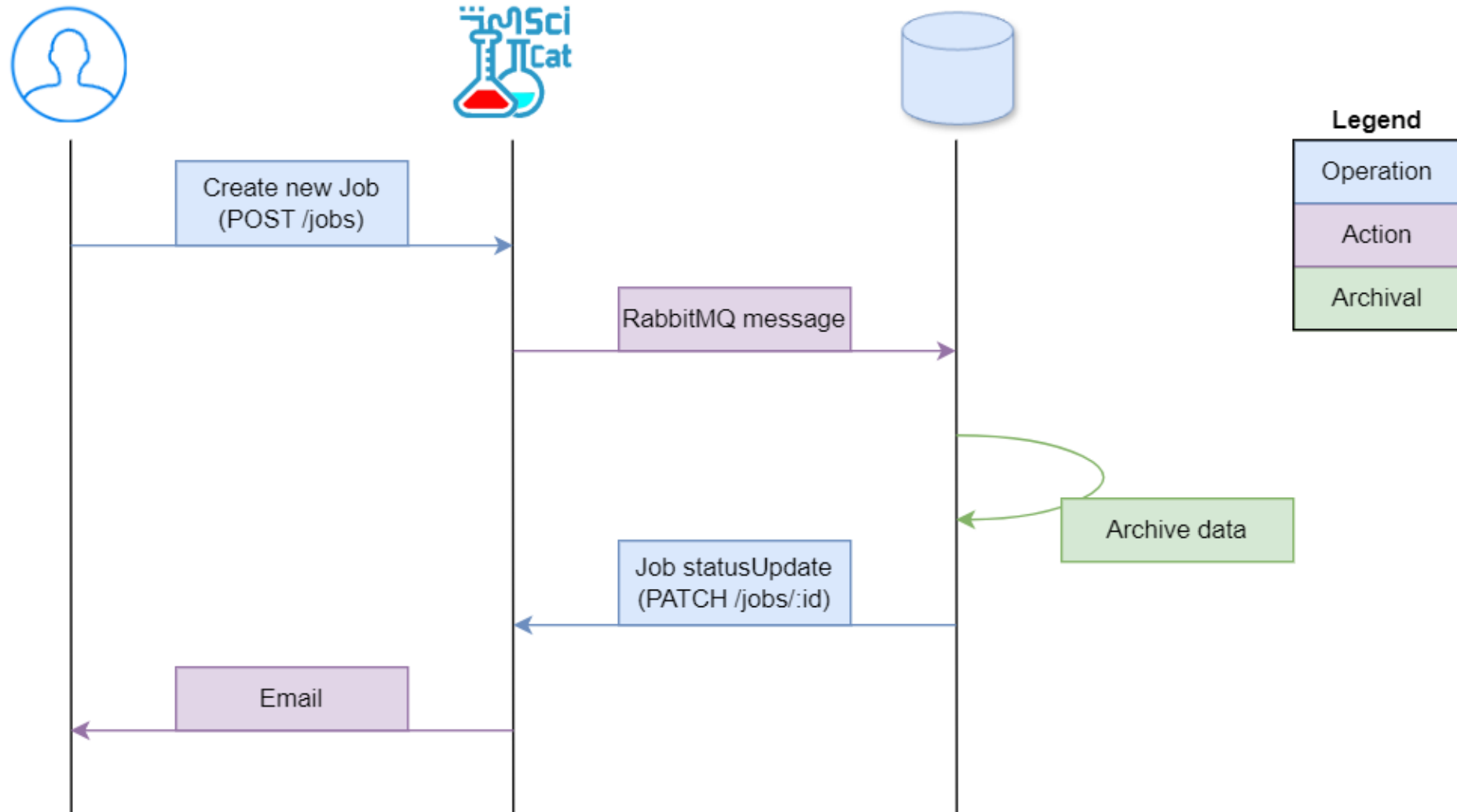
Goals for jobs in v4.0



In the old backend, jobs were hardcoded, so not flexible to edit. Instead, we will now have a JSON that defines the job configuration:

- Configure jobs without code changes
- Refactor code to be modular and easy to add functionalities
- Enable the use of pre-built containers without modification
- Better testing, particularly integration tests
- (Phase 2) Move RabbitMQ/Kafka dependencies from the backend to independent plugins so facilities can only install what they use

Actions overview



- `jobType` (string): archive, retrieve, public, or custom types
- `jobOperations` (enum): create, statusUpdate
- `auth` (enum or string): authorization policy (eg #all), specific user, specific group
- `JobActions` (class): send email, call URL, post RabbitMQ message, etc
- `jobParams` (object): datasetIds, action-specific configuration parameters
- `ownership` (string): ownerUser, ownerGroup

Input fields for the job to be created

```
{
  "type": "all_access",
  "jobParams": {
    "datasetIds": [
      "datasets1",
      "dataset2"
    ]
  },
  "ownerUser": "user1",
  "ownerGroup": "group",
  "contactEmail": "user1@psi.ch"
}
```

Documentation coming before official release.

Configuration



```
{
  "configVersion": "v1.0 2024-03-01 6f3f38",
  "jobs": [
    {
      "jobType": "archive",
      "create": {
        "auth": "#all",
        "actions": [
          {
            "actionType": "log"
          },
          {
            "actionType": "url",
            "url": "http://localhost:3000/api/v3/health?jobid={{id}}",
            "headers": {
              "accept": "application/json"
            }
          },
          {
            "actionType": "rabbitmq",
            "hostname": "rabbitmq",
            "port": 5672,
            "username": "guest",
            "password": "guest",
            "queue": "jobqueue"
          }
        ]
      },
      "statusUpdate": {
        "auth": "archivemanager",
        "actions": [ ... ]
      }
    }
  ],
},
```

- Parse configuration file and validate its schema
- For each job type, register actions per job operation
- Validate actions before starting the job operation
- Perform actions after a successful job operation

Template example



jobConfig.json:

```
{
  "actionType": "email",
  "to": "{{job.contactEmail}}",
  "from": "dacadmail@psi.ch",
  "subject": "[SciCat] Your {{job.type}} job submitted
  successfully",
  "bodyTemplate": "job-template.html"
}
```

job-template.html excerpt:

```
<h2>{{subject}}</h2>
<p><b>Job {{job.pid}} will be performed on the following dataset(s):</b></p>
<table>
  {{#each job.jobParams.datasets}}
  <tr>
    <td><a href="{{@root.config.scicatHref}}/datasets/{{urlEncode pid}}">{{pid}}</a></td>
    <td>{{ownerGroup}}</td>
    <td>{{sourceFolder}}</td>
    <td>{{formatBytes size}}</td>
  </tr>
  {{/each}}
</table>
```

- Most actions accept handlebars templates
- Templates receive the job parameters passed during job creation (such as linked datasets), and additional params passed to statusUpdate calls
- (Email particulars still under development)

Permissions



- Two levels of permissions are required to create a job: endpoint and job instance authorization
- Different identities get different permissions:

Job Create Authorization	Endpoint Authentication Translation	Endpoint Authentication Description	Instance Authentication Translation	Instance Authentication Description
<i>#all</i>	<i>#all</i>	any user can access this endpoint, both anonymous and authenticated	<i>#all</i>	Any user can create this instance of the job
<i>#datasetPublic</i>	<i>#all</i>	any user can access this endpoint, both anonymous and authenticated	<i>#datasetPublic</i>	the job instance will be created only if all the datasets listed are public
<i>#authenticated</i>	<i>#user</i>	any valid users can access the endpoint, independently from their groups	<i>#user</i>	any valid users can create this instance of the job
<i>#datasetAccess</i>	<i>#user</i>	any valid user can access this endpoint, independently from their groups	<i>#datasetAccess</i>	the job instance will be created only if the user has access to all the datasets listed
<i>#datasetOwner</i>	<i>#user</i>	any valid user can access this endpoint, independently from their groups	<i>#datasetOwner</i>	the job instance will be created only if the user is part of all the datasets owner group
@GROUP	GROUP	only users that belongs to the specified group can access the endpoint	GROUP	the job instance will be created only if the user belongs to the group specified
USER	USER	only the specified user can access the endpoint	USER	the job instance can be created only by the user indicated

Permissions



more restrictive →



more restrictive →

- unit tests are written for
 - every user type
 - job configuration
 - (optional) dataset permissions
- tables with tests outcomes available for each of the endpoints
- ≈ 200 tests to test authorization of jobs CRUD operations
- Next step: integration tests with scicatlive

Roadmap



Status	Task
Done	Core Job implementation
Done	Permission Model
Done	Unit tests
Review	RabbitMQ action
In progress	Email action
In progress	URL action
Summer 2024	Early adopters: PSI, RFI, MAXIV
Fall 24	Merge to master

Planned:

- Plugin system ([#862](#))
- More integration tests using scicatlive
- Additional action ideas

Thanks!