A New VO Publishing Registry Framework at NOIRLab

Robert Nikutta <robert.nikutta@noirlab.edu>

with Chadd Myers, Destry Saul, Igor Suarez Sola

Motivation

- Number of data services, data volume, and complexity growing
- Discoverability suffering disproportionately when not in the VO Registry
- As the years go buy, the "old guard" of engineers is also moving on or retiring
  $\Rightarrow$ *Knowledge drain*
- New devs: often not fluent in legacy languages (e.g., Java)
- Modern S/W engineering today is a totally different game
- Python used much more widely, both in astronomy and industry
- It's much easier to understand a new protocol (e.g., OAI-PMH*) than to master a legacy language
- Much better tooling available for newer tech like Python

(*) Open Archives Initiative Protocol for Metadata Harvesting

Help

- Discussion started at Bologna Interop (May 2023)
- I "complained" that it's opaque for new devs to understand how VO Registry works
  - documentation too high-level
  - some examples don't work
  - the "easy" solution is hacky and manual
  - etc.
- Then we got startup help. Conversations with Registry WG. Renaud, Markus and others organized a telecon, post-Interop.
- Established registry implementers shared their experience and pointers
- Newbies like us learned and asked questions

Thank you!

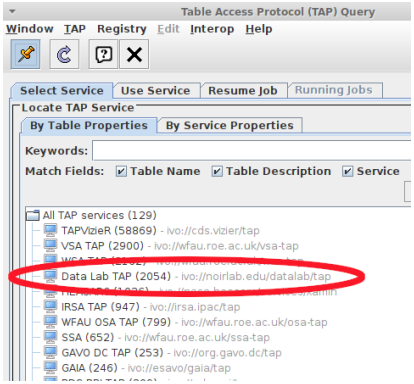
Back at home - Initiate a project

- Began informal conversations with a few devs, including a new and motivated chief S/W person
- 2/3 had not heard of the VO Registry before
- Talked with management (just one level up)
- Sold them on "small effort, big benefits" (maybe, undersold on estimated time a bit)
- We started easy, with just an informal mandate (but with enough personal motivation)
- Met bi-weekly for a while
- *Destry Saul* wrote the registry code (Python), infrastructure code (Terraform), tests, and CI (on Gitlab); got it 90% done
- *Igor Suarez Sola* wrote XML resource creation for some of the Astro Data Lab services and also began CI to automate it

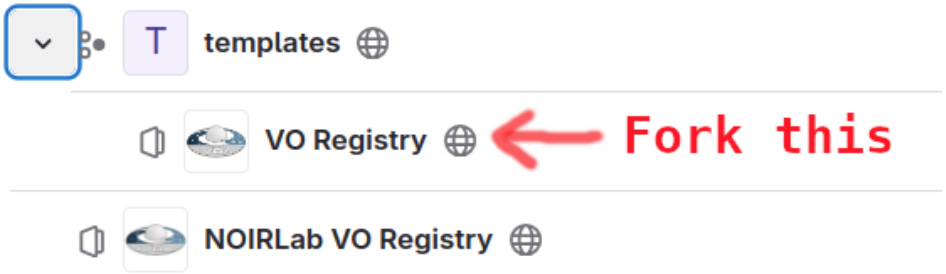
But life isn't kind. . .

- . . . One dev resigned, one moved laterally, our division director left for another job..
- I had to work with the new director, and with a new developer *Chadd Myers*
  - This is probably the hardest part: work to get folks' buy-in, over and over
  - Also requires the will / permission to de-prioritize other projects for a short while
- Chadd learned what was done so far, finished the remaining work, improved the deployment, the CI, and the XML generation, deployed a first version

We are in the VO Registry :-)

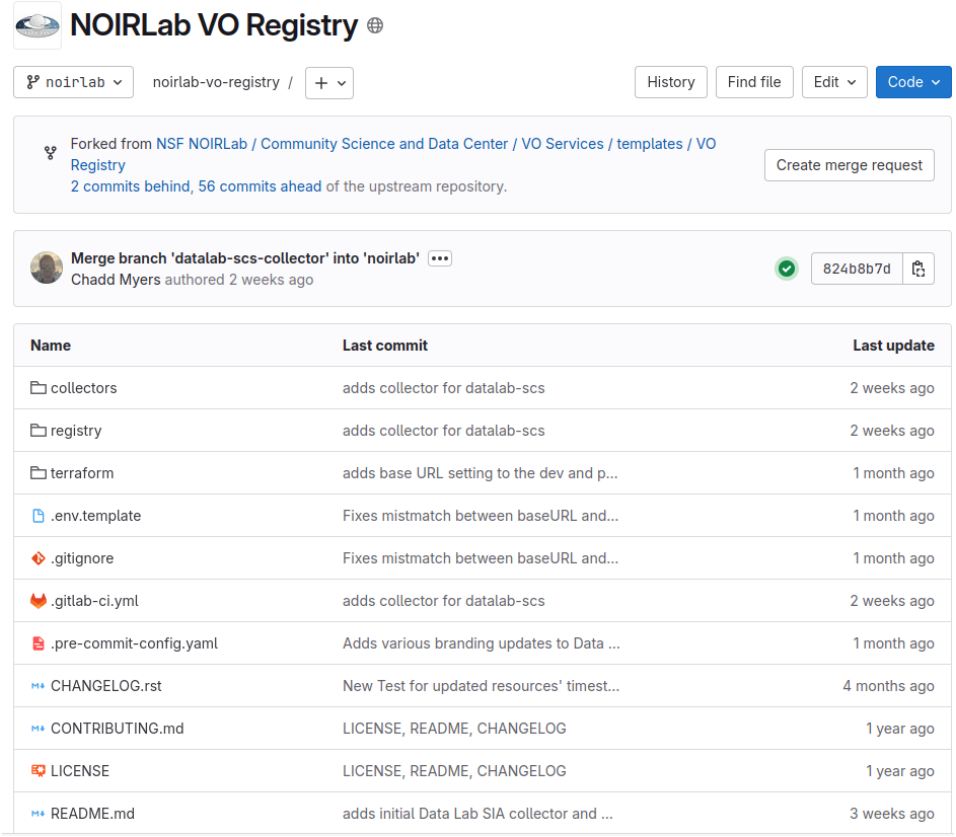


The implementation - Template and Forks



https://gitlab.com/nsf-noirlab/csdc/vo-services

The implementation - Code organization



NOIRLab VO Registry

noirlab | noirlab-vo-registry / | History | Find file | Edit | Code

Forked from NSF NOIRLab / Community Science and Data Center / VO Services / templates / VO Registry
2 commits behind, 56 commits ahead of the upstream repository.

Create merge request

Merge branch 'datalab-scs-collector' into 'noirlab'
Chadd Myers authored 2 weeks ago

824b8b7d

Name	Last commit	Last update
collectors	adds collector for datalab-scs	2 weeks ago
registry	adds collector for datalab-scs	2 weeks ago
terraform	adds base URL setting to the dev and p...	1 month ago
.env.template	Fixes mistmatch between baseURL and...	1 month ago
.gitignore	Fixes mistmatch between baseURL and...	1 month ago
.gitlab-ci.yml	adds collector for datalab-scs	2 weeks ago
.pre-commit-config.yaml	Adds various branding updates to Data ...	1 month ago
CHANGELOG.rst	New Test for updated resources' timest...	4 months ago
CONTRIBUTING.md	LICENSE, README, CHANGELOG	1 year ago
LICENSE	LICENSE, README, CHANGELOG	1 year ago
README.md	adds initial Data Lab SIA collector and ...	3 weeks ago

The implementation - Registry

- Written in Python, uses FastAPI (good tooling available)
- It's a publishing registry only for now (implements OAI-PMH), not a searchable registry. $\Rightarrow$ Meant to be harvested.
- Implements verbs:
  GetRecord, Identify, ListIdentifiers, ListMetadataFormats, ListRecords, ListSets
- System requirements: Docker Engine, Make (only for local development)
- Documentation has steps to configure, deploy, operate
- Tests: unit & end-to-end
- VOResources are stored in static XML files in `vo_registry/static/resources/`
- The two required VOResources (Registry and Authority) are included (edit their content).

The implementation - XML record generation

- One of many ways:
  - collectors
  - custom
  - manual
  - etc.
- We use "collectors" which are simple workflow-triggered functions.
- For SIA & SCS we call a small internal discovery service:

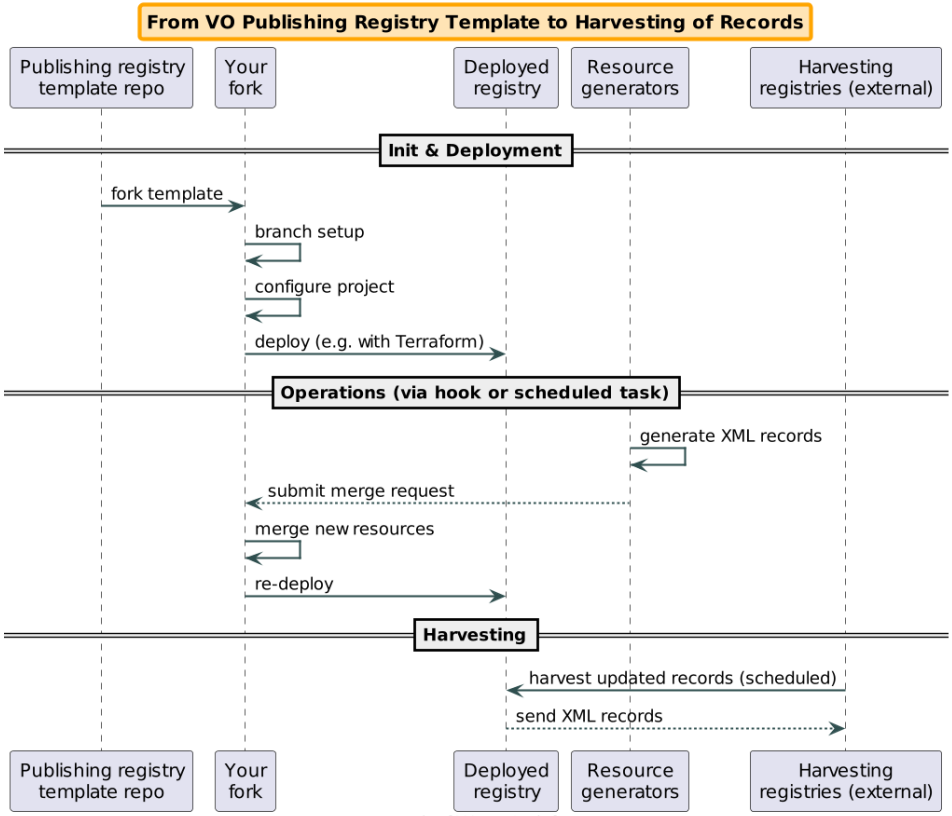
  "What SIA services do you have?"

  (For TAP we still need to write one.)
- Response converted to XML records, written to static files.
- Merge Request made automatically to our fork repo $\Rightarrow$ Human approves $\Rightarrow$ Re-deployed.

The implementation - Infrastructure

- Uses Poetry / pyproject.toml for package config
- Terraform is set up, just add your config params
  - But it's optional; you can deploy whichever way you like
- We deploy at GCP, on a cheap on-demand instance
- Easy to adapt to other providers, or deploy on-prem
- CI (we use GitLab), set up a few env variables

From template to harvesting



Currently exposed services

In first iteration: Astro Data Lab science platform services

TAP: AllWISE, Buzzard (sim), CatWISE2020, WISE2020, *DAD*, *DECaPS*, *DELVE*, *DES*, *DESI EDR*, Gaia, *GNIRS-DQS*, *GOGREEN*, Hipparcos, *DESI Imaging Legacy Surveys (LS)*, TRILEGAL / LSST-Sim, *NOIRLab Source Catalog (NSC)*, PG-IR, PHAT, SDSS, Siena Galaxy Atlas, SkyMapper, *SMASH*, *SPARCL*, S-PLUS, 2MASS, Tycho, UKIDSS, UnWISE, USNO, VHS

SIA (raw+calibrated+stacks): *DES*, *DECaPS*, *DeMCELS*, *GOGREEN*, *LS*, *NSC*, *SMASH*, S-PLUS, *NOIRLab Archive*

SCS: AllWISE, *DECaPS*, *DELVE*, *DES*, Gaia, *GOGREEN*, Hipparcos, *LS*, *NSC*, PHAT, SDSS, SkyMapper, *SMASH*, *SPARCL*, S-PLUS, 2MASS, Tycho, UnWISE, UKIDSS, USNO, VHS

Next steps / possible directions

- Add coverage information
- Integrate other NOIRLab projects with VO services, including:
  - NOIRLab Astro Data Archive
  - ANTARES alert broker (TAP)
  - Gemini Archive
  - Rubin
  - future US-ELTP
- Some conversations have started, some have yet to be initiated.

Thank you!

- **VO Registry template (fork this)**

  https://gitlab.com/nsf-noirlab/csdc/vo-services/templates/vo-registry

  MRs are welcome! Ideas for improvement too. Please get in touch with us.

  Talk to us also if you want to talk about how to deploy your own fork.

- **Contact**

  Robert Nikutta (project): <robert.nikutta@noirlab.edu>

  Chadd Myers (technical): <chadd.myers@noirlab.edu>