THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

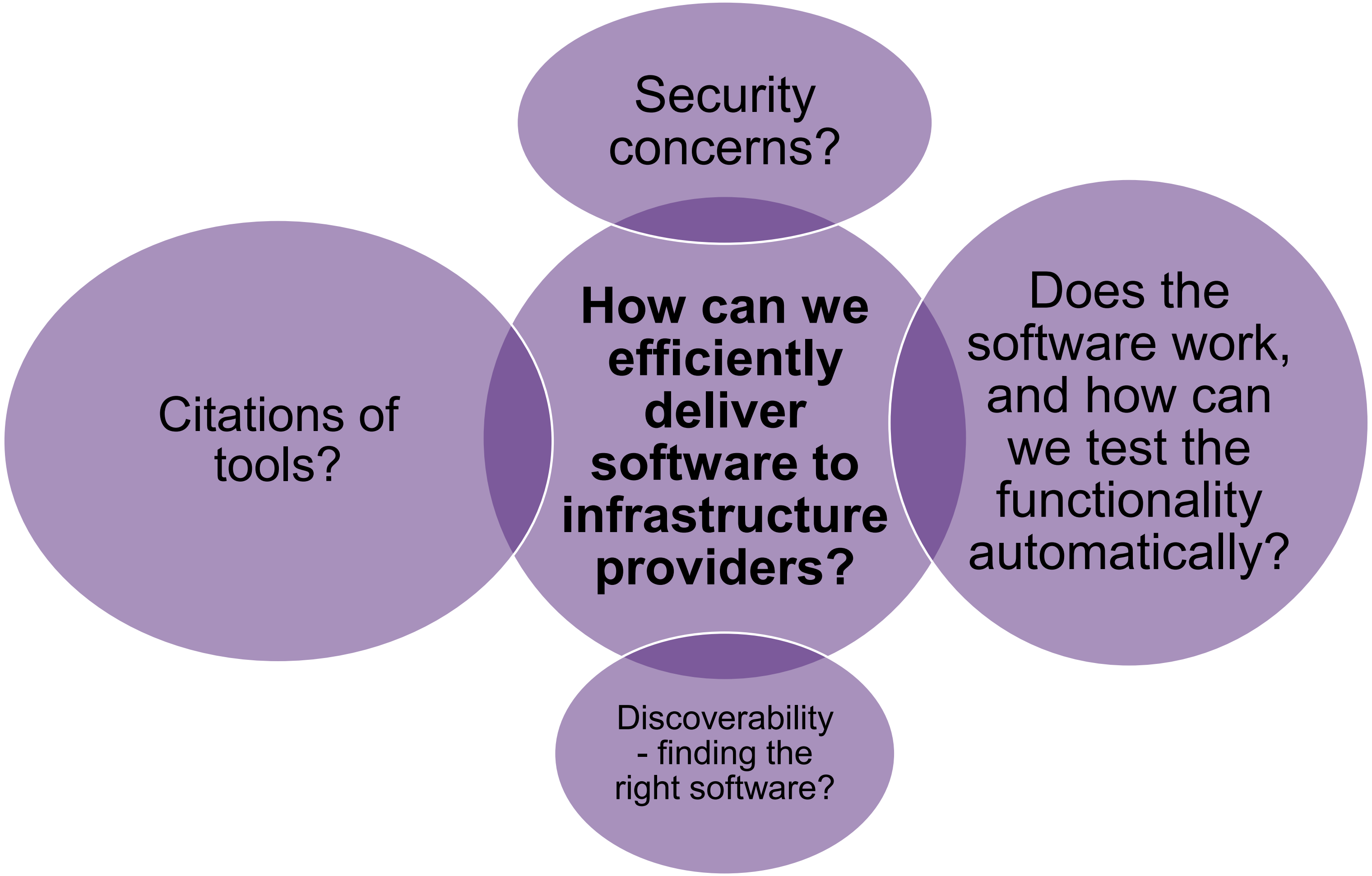
SCIDIR

A **Sci**entific Software **Di**stribution **R**epository

Steffen Bollmann[1], Aswin Narayanan[2], Sarah Beecroft[3], Greg Darcy[5], Nigel Ward[6], Peter Marendy[4]

1) University of Queensland
2) National Imaging Facility
3) Pawsey Supercomputing Research Centre
4) The Queensland Cyber Infrastructure Foundation (QCIF)
5) AARNet
6) Australian BioCommons

What is the problem?



Security concerns?

Citations of tools?

How can we efficiently deliver software to infrastructure providers?

Does the software work, and how can we test the functionality automatically?

Discoverability - finding the right software?

History

- This proposal aims at bringing together various recent developments in different communities

Australian Biocommons Bring-Your-Own-Data Expansion Project

- Aimed at providing shared reference datasets and containers across different infrastructures

AEDAPT NeuroDesk platform project

- Neuroimaging containers automatically built via Github actions and distributed via CVMFS + Desktop environment for easy access

European Environment for Scientific Software Installations (EESSI)

- software layer with EasyBuild, Lmod and archspec; Gentoo Prefix compatibility layer; distribution via CVMFS

EESSI and DRA of Canada

- EESSI plans to build all applications available in easybuild and distribute via CMVFS
- DRA provides CVMFS servers with various software optimized for HPCs

- what they don’t do:
  - provide software containers or workflows in containers (e.g. MRIQC)
  - less focused on reproducibility aspects (focus is on performance, no container archival, no citation of containers, no DOI for containers)

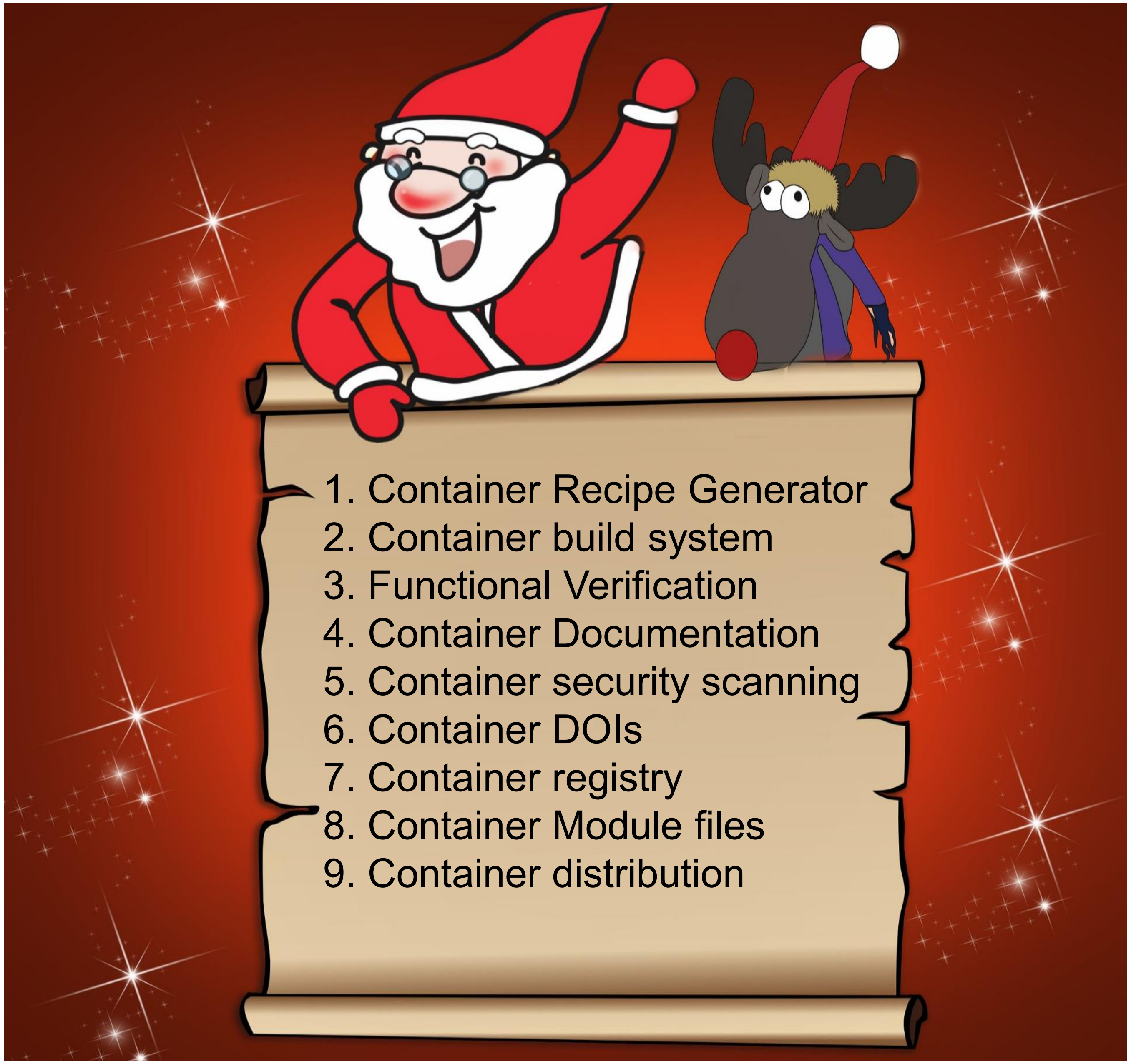
EESSI
EUROPEAN ENVIRONMENT FOR SCIENTIFIC SOFTWARE INSTALLATIONS

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

The gap?

- current software distribution projects don’t focus on containers
- scanning for vulnerabilities or testing for functional correctness currently not in focus
  - -> HPC admins struggle to allow these containers on their platforms, but users are already running it anyway
- software containers are not optimized for different hardware architectures
- no systematic mechanism to capture meta-data, discovery, or citation
- no production CVMFS deployment in Australia

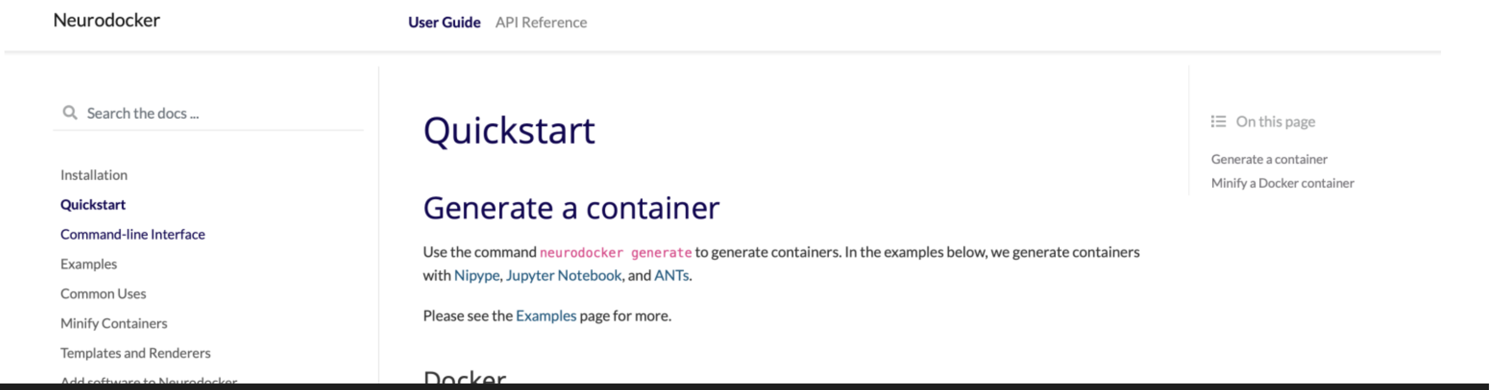
A wish list:



1. Container Recipe Generator
2. Container build system
3. Functional Verification
4. Container Documentation
5. Container security scanning
6. Container DOIs
7. Container registry
8. Container Module files
9. Container distribution

Container Recipe Generator / Template / Renderer

- Writing good docker recipes is tedious
- Domain specific recipe generators like neurodocker are a great help
- other domain agnostic tools
    - HPC Container Maker
    - Spack
    - Easybuild


Neurodocker

User Guide API Reference

Search the docs ...

Installation
Quickstart
Command-line Interface
Examples
Common Uses
Minify Containers
Templates and Renderers

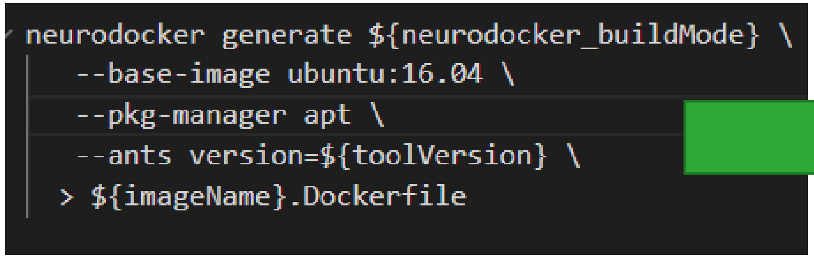
Quickstart

Generate a container

Use the command `neurodocker generate` to generate containers. In the examples below, we generate containers with Nipype, Jupyter Notebook, and ANTs.

Please see the Examples page for more.

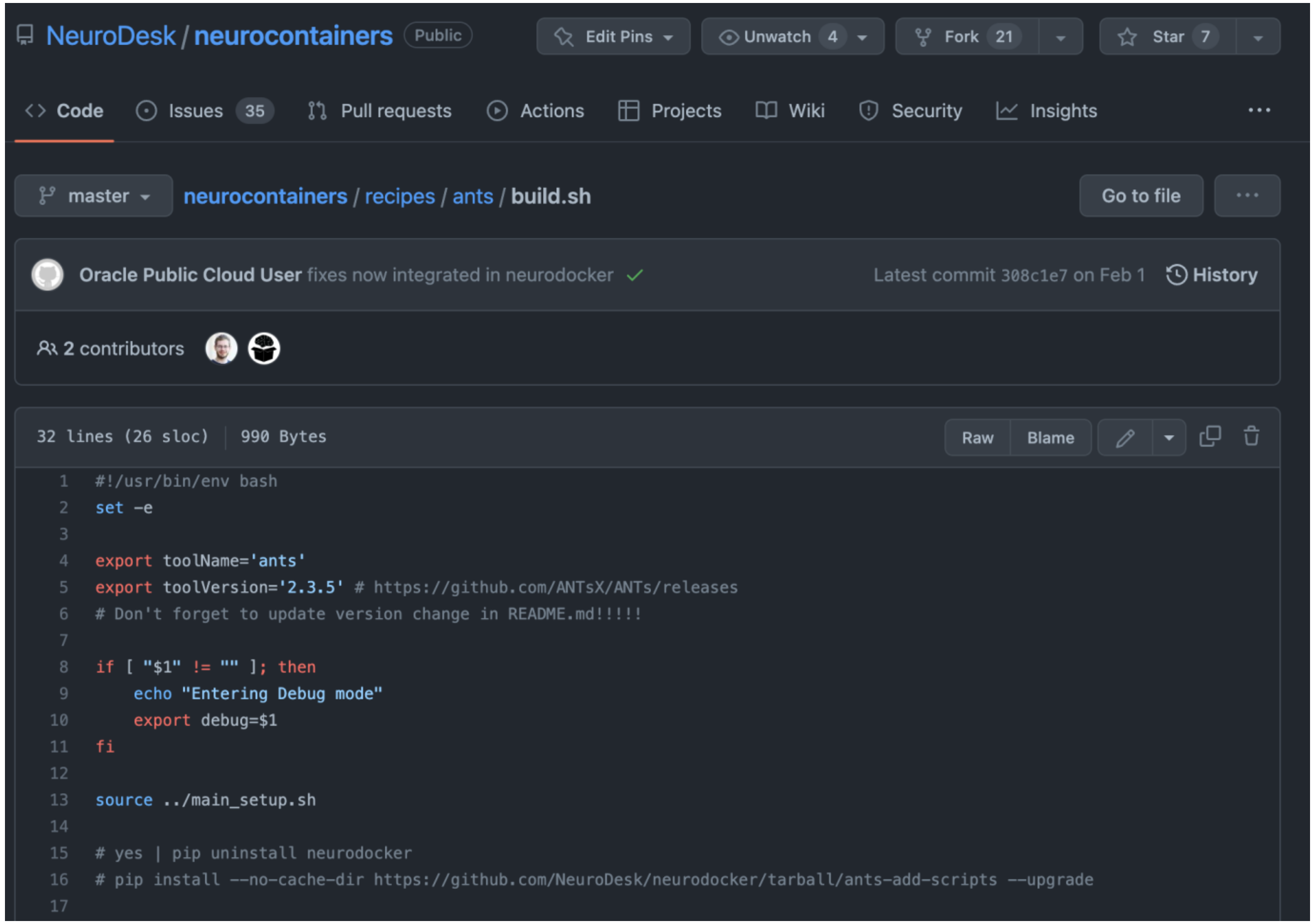
On this page
Generate a container
Minify a Docker container


```
neurodocker generate ${neurodocker_buildMode} \
    --base-image ubuntu:16.04 \
    --pkg-manager apt \
    --ants version=${toolVersion} \
  > ${imageName}.Dockerfile
```


```
FROM ubuntu:16.04
RUN printf '#!/bin/bash\nls -la' > /usr/bin/ll
RUN chmod +x /usr/bin/ll
RUN mkdir /afm01 /afm02 /90days /30days /QRISdata /RDS /data /short /proc_temp /TMPDIR /nvme /local /gpfs1 /working
ENV ANTSPATH="/opt/ants-2.3.4/" \
    PATH="/opt/ants-2.3.4:$PATH"
RUN apt-get update -qq \
    && apt-get install -y -q --no-install-recommends \
           ca-certificates \
           curl \
    && rm -rf /var/lib/apt/lists/* \
    && echo "Downloading ANTs ..." \
    && mkdir -p /opt/ants-2.3.4 \
    && curl -fsSL https://dl.dropbox.com/s/gwf51ykkk5bifyj/ants-Linux-centos6_x86_64-v2.3.4.tar.gz \
    | tar -xz -C /opt/ants-2.3.4 --strip-components 1
ENV DEPLOY_PATH="/opt/ants-2.3.4/"
COPY ["README.md", \
      "/README.md"]
```

Automated Container build

- Github repository that collects Docker recipes from contributors
- Recipes get build via github actions with custom runners + architecture specific builds via custom runners (e.g. ARM, AMD Milan …)
- Periodic rebuilds for new versions?



neurocontainers / recipes / ants / build.sh

Oracle Public Cloud User fixes now integrated in neurodocker ✓ Latest commit 308c1e7 on Feb 1

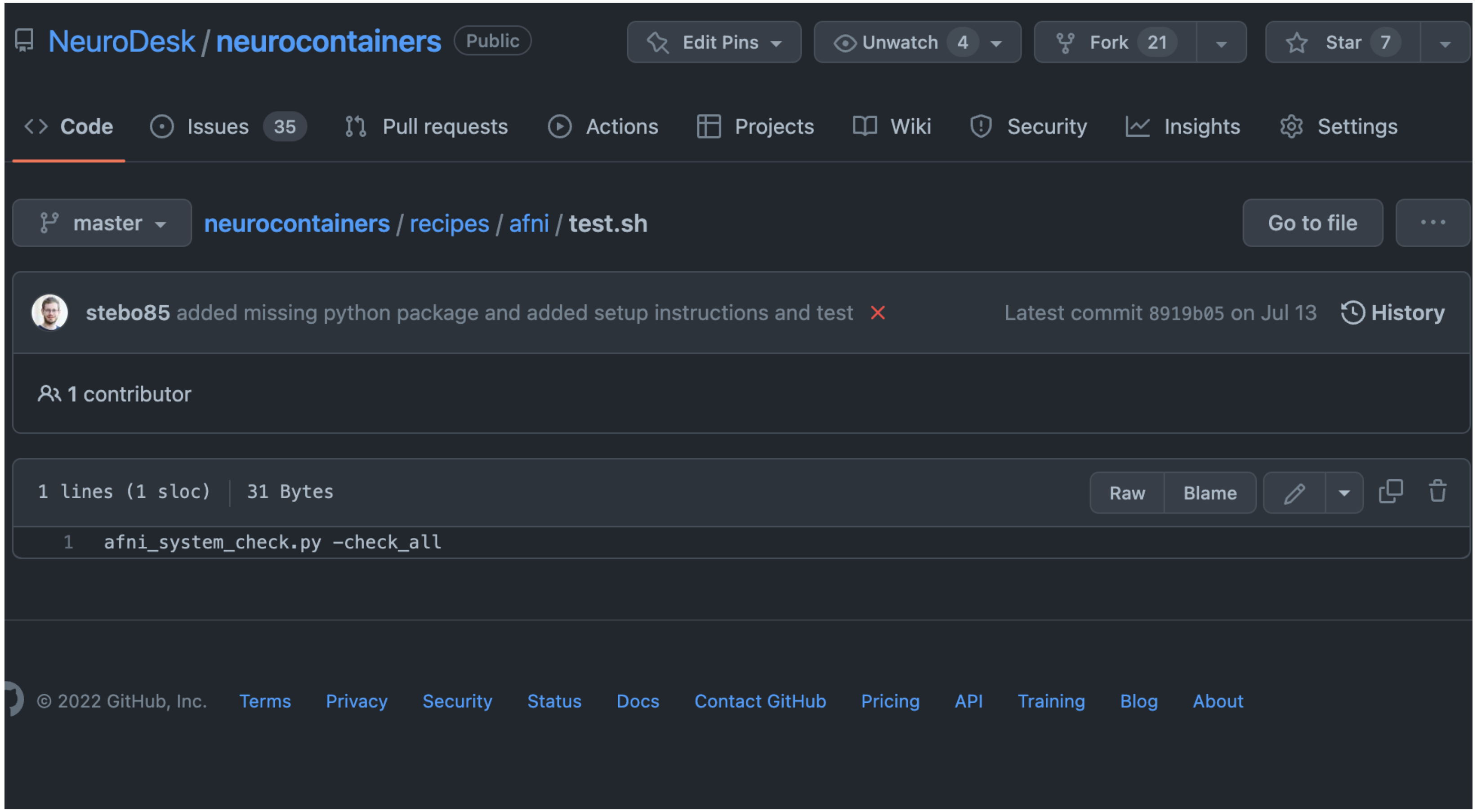
2 contributors

32 lines (26 sloc) | 990 Bytes

```
#!/usr/bin/env bash
set -e

export toolName='ants'
export toolVersion='2.3.5' # https://github.com/ANTsX/ANTs/releases
# Don't forget to update version change in README.md!!!!!

if [ "$1" != "" ]; then
    echo "Entering Debug mode"
    export debug=$1
fi

source ../main_setup.sh

# yes | pip uninstall neurodocker
# pip install --no-cache-dir https://github.com/NeuroDesk/neurodocker/tarball/ants-add-scripts --upgrade

```

Functional Verification

- test script with test data that we can run after the build process?



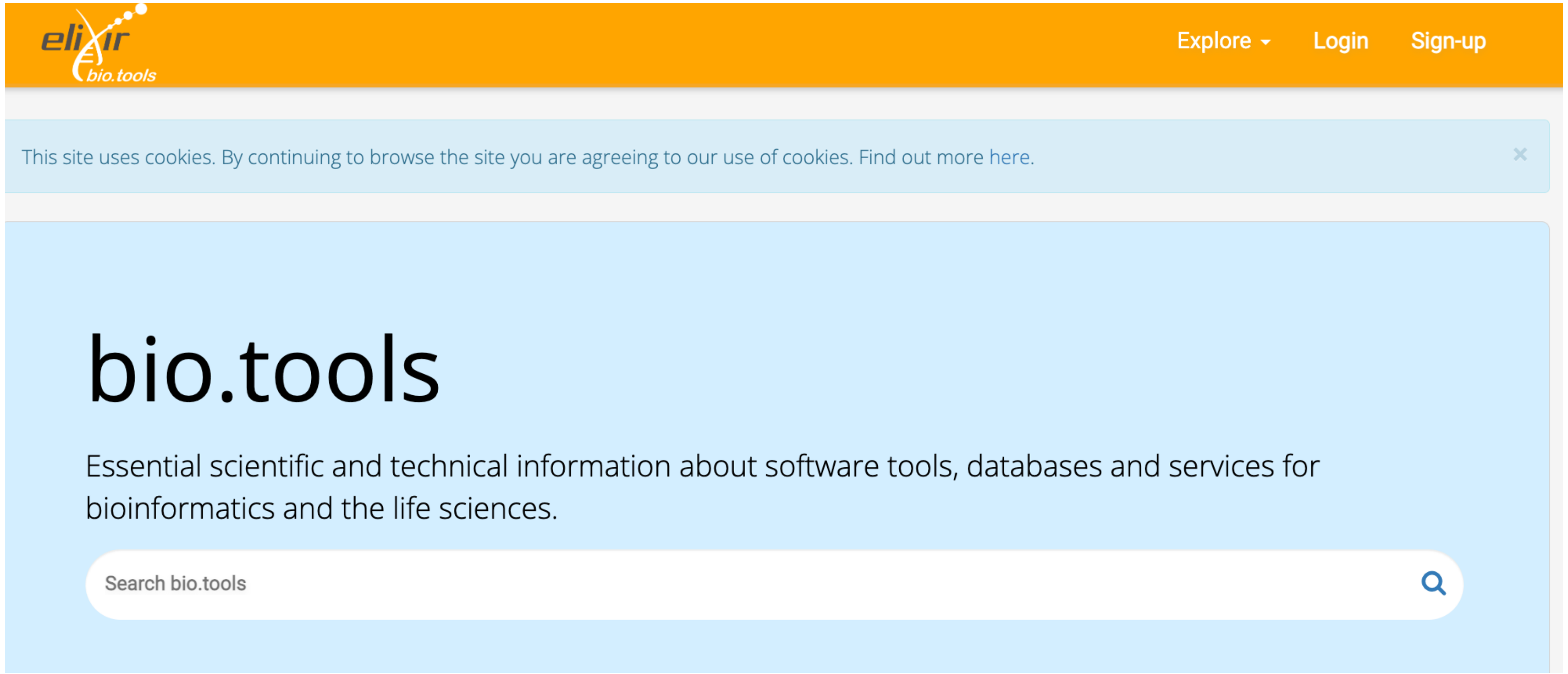
stebo85 added missing python package and added setup instructions and test ✕

```
1 lines (1 sloc) | 31 Bytes
```

```
afni_system_check.py -check_all
```

Container Documentation

- EDAM / Bio.Tools ?
- Voting for new containers and discovering of what’s there -> github issues ?
- how to find people who volunteer to build containers and test them?
- how to handle different hardware architectures / MPI versions ?



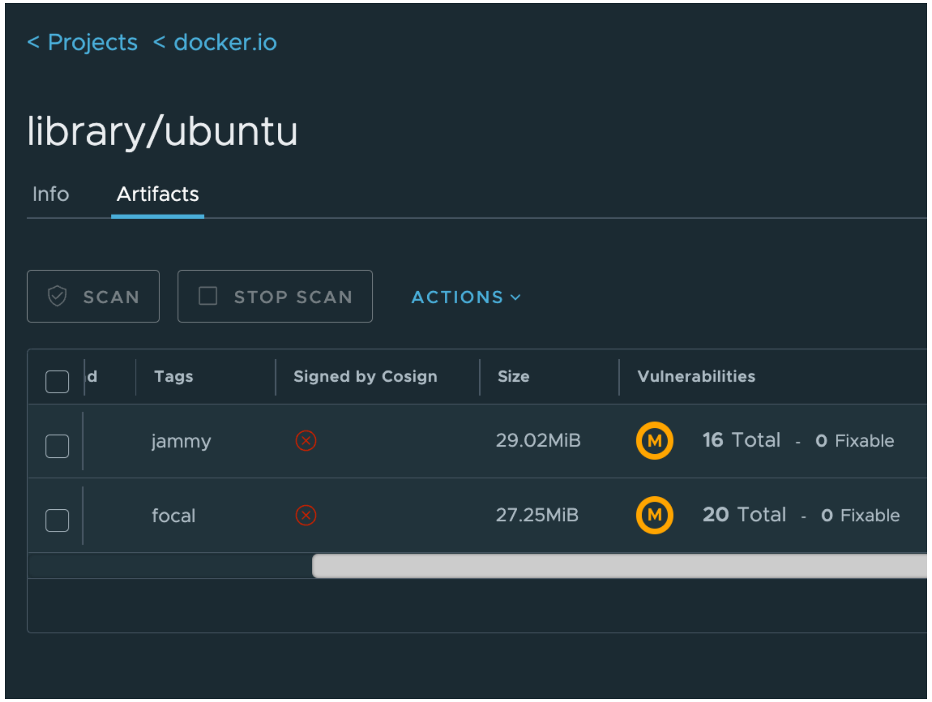
This site uses cookies. By continuing to browse the site you are agreeing to our use of cookies. Find out more here.

bio.tools

Essential scientific and technical information about software tools, databases and services for bioinformatics and the life sciences.

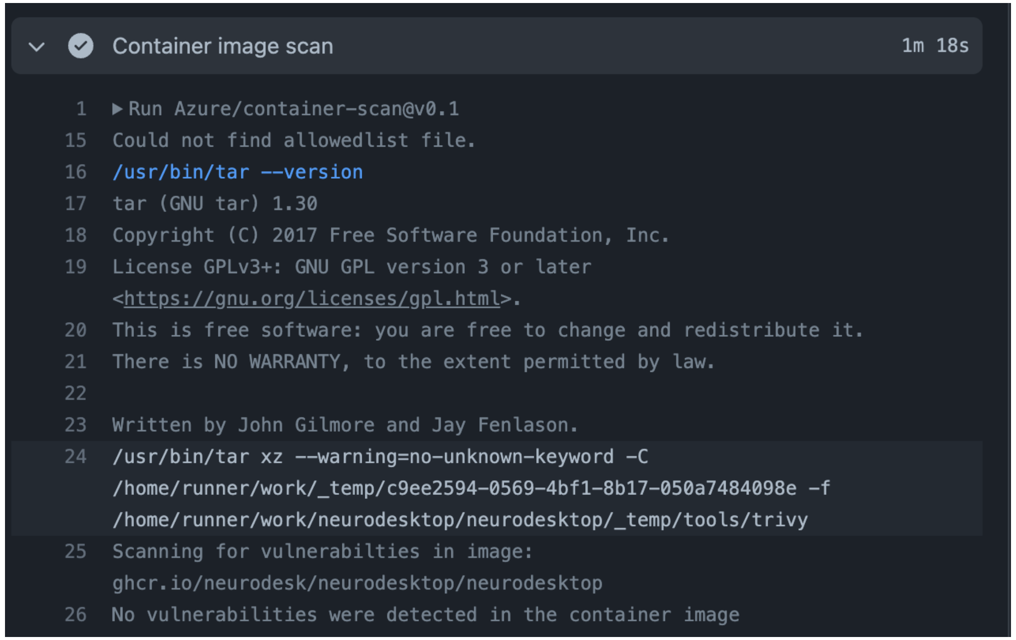
Container Security Scanning

- Scan after build using github action azure container scan
- Scan during storage in registry (e.g. Harbor)
- Regular scans on distribution infrastructure?



library/ubuntu

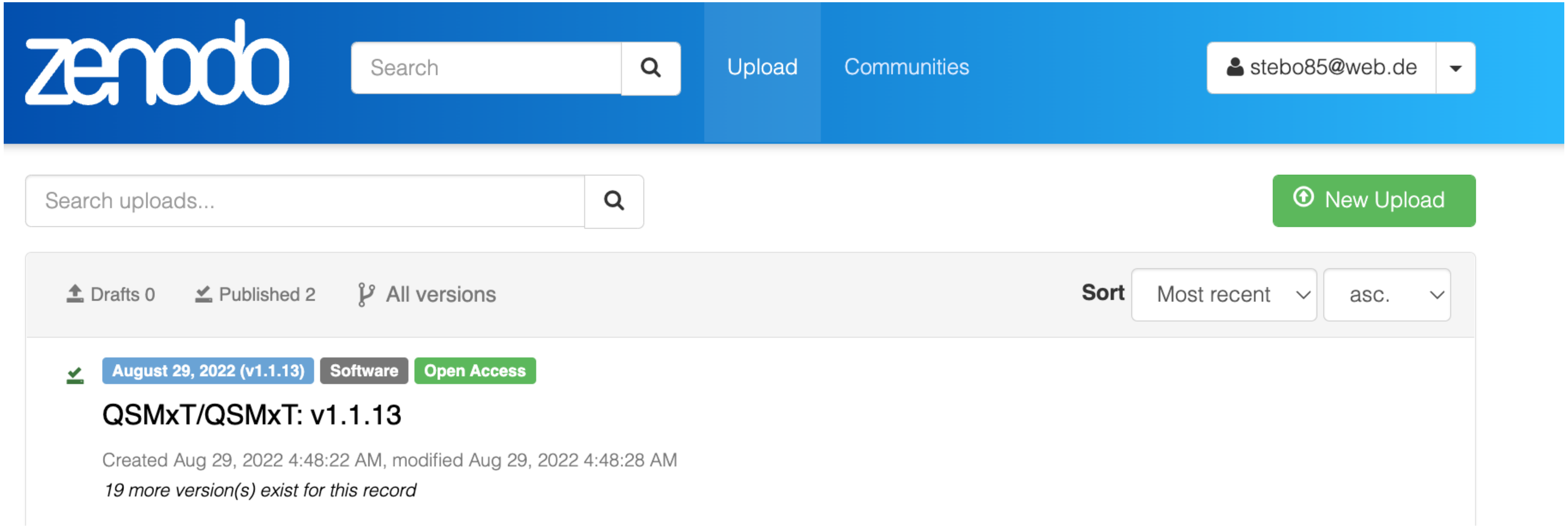
☐	d	Tags	Signed by Cosign	Size	Vulnerabilities
☐		jammy	☒	29.02MiB	M 16 Total - 0 Fixable
☐		focal	☒	27.25MiB	M 20 Total - 0 Fixable



```
▶ Run Azure/container-scan@v0.1
Could not find allowedlist file.
/usr/bin/tar --version
tar (GNU tar) 1.30
Copyright (C) 2017 Free Software Foundation, Inc.
License GPLv3+: GNU GPL version 3 or later
<https://gnu.org/licenses/gpl.html>.
This is free software: you are free to change and redistribute it.
There is NO WARRANTY, to the extent permitted by law.

Written by John Gilmore and Jay Fenlason.
/usr/bin/tar xz --warning=no-unknown-keyword -C
/home/runner/work/_temp/c9ee2594-0569-4bf1-8b17-050a7484098e -f
/home/runner/work/neurodesktop/neurodesktop/_temp/tools/trivy
Scanning for vulnerabilties in image:
ghcr.io/neurodesk/neurodesktop/neurodesktop
No vulnerabilities were detected in the container image
```

Making Containers citable

- Upload Singularity files to Zenodo and create DOI?
- Or can we mint our own DOIs and link these to nectar object storage?



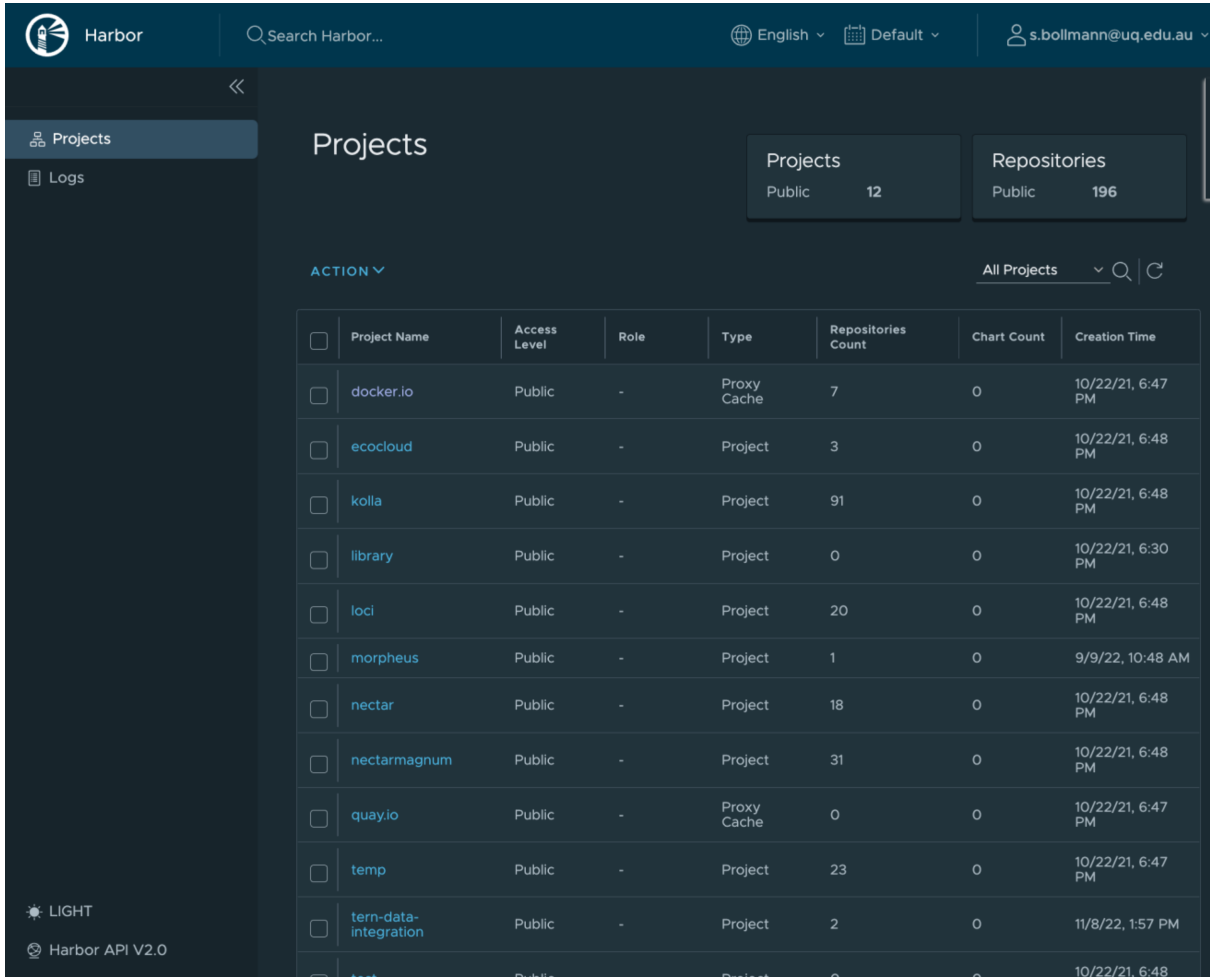
QSMxT/QSMxT: v1.1.13

Created Aug 29, 2022 4:48:22 AM, modified Aug 29, 2022 4:48:28 AM

19 more version(s) exist for this record

Container Registry

https://registry.rc.nectar.org.au/harbor



Projects

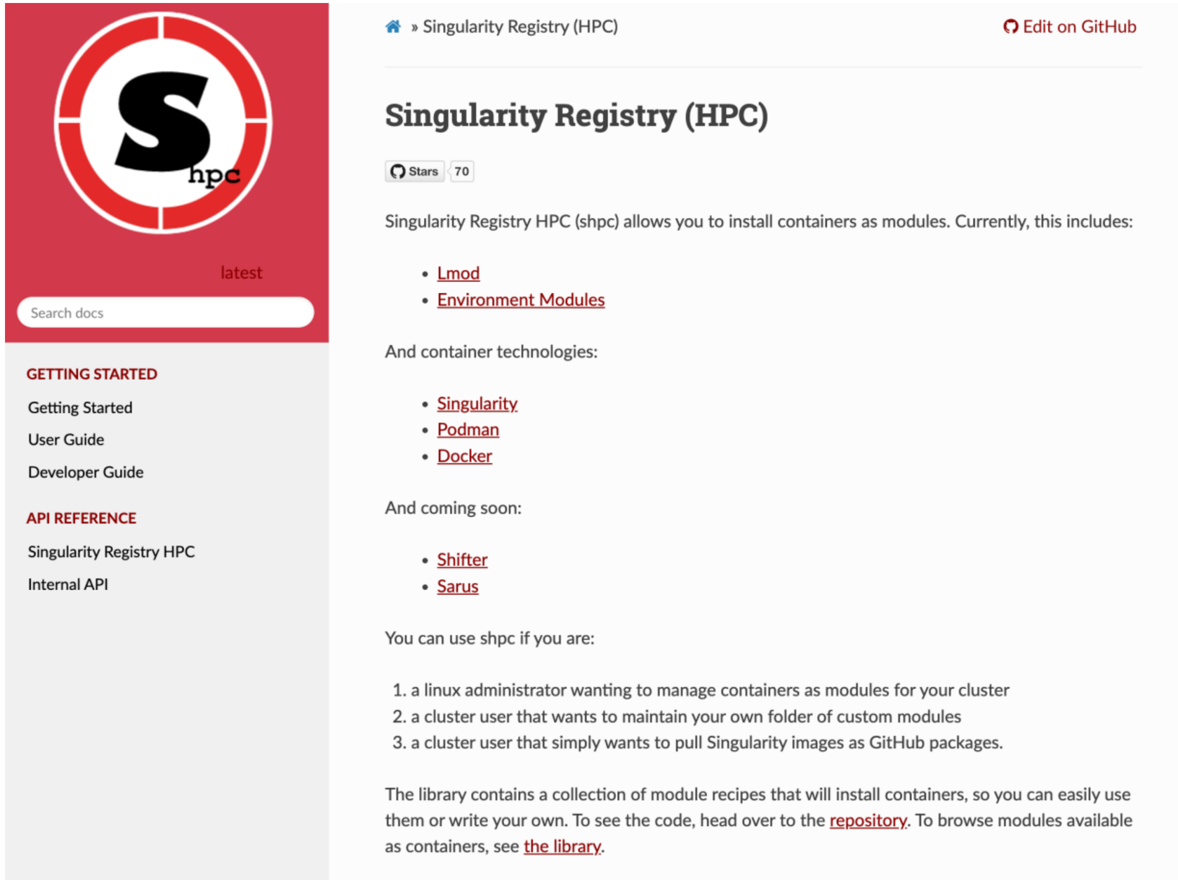
Projects
Public 12

Repositories
Public 196

☐	Project Name	Access Level	Role	Type	Repositories Count	Chart Count	Creation Time
☐	docker.io	Public	-	Proxy Cache	7	0	10/22/21, 6:47 PM
☐	ecocloud	Public	-	Project	3	0	10/22/21, 6:48 PM
☐	kolla	Public	-	Project	91	0	10/22/21, 6:48 PM
☐	library	Public	-	Project	0	0	10/22/21, 6:30 PM
☐	loci	Public	-	Project	20	0	10/22/21, 6:48 PM
☐	morpheus	Public	-	Project	1	0	9/9/22, 10:48 AM
☐	nectar	Public	-	Project	18	0	10/22/21, 6:48 PM
☐	nectarmagnum	Public	-	Project	31	0	10/22/21, 6:48 PM
☐	quay.io	Public	-	Proxy Cache	0	0	10/22/21, 6:47 PM
☐	temp	Public	-	Project	23	0	10/22/21, 6:47 PM
☐	tern-data-integration	Public	-	Project	2	0	11/8/22, 1:57 PM
							10/22/21, 6:48

Container Modules

- integration of containers in Module System in HPCs via Lmod
- SHPC?
    - can SHPC automatically expose individual binaries inside a container? Currenlty every binary needs a separate entry in the yaml?
- How to handle different sites?
    - every site has a subdirectory on distribution system with their own container subset and SHPC config?



Singularity Registry (HPC)

Singularity Registry HPC (shpc) allows you to install containers as modules. Currently, this includes:

- Lmod
- Environment Modules

And container technologies:

- Singularity
- Podman
- Docker

And coming soon:

- Shifter
- Sarus

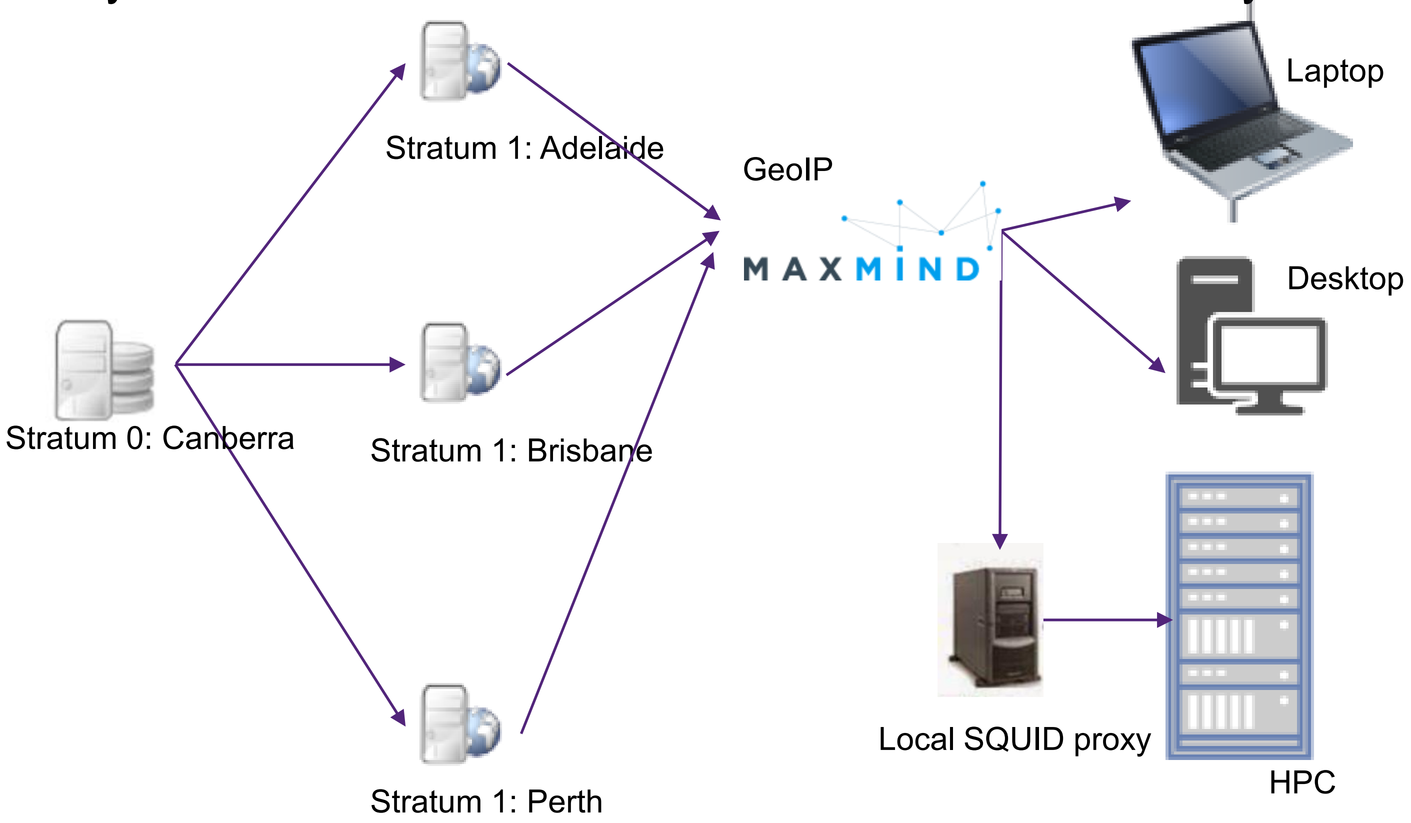
You can use shpc if you are:

1. a linux administrator wanting to manage containers as modules for your cluster
2. a cluster user that wants to maintain your own folder of custom modules
3. a cluster user that simply wants to pull Singularity images as GitHub packages.

The library contains a collection of module recipes that will install containers, so you can easily use them or write your own. To see the code, head over to the repository. To browse modules available as containers, see the library.

Distributing Singularity containers via CVMFS

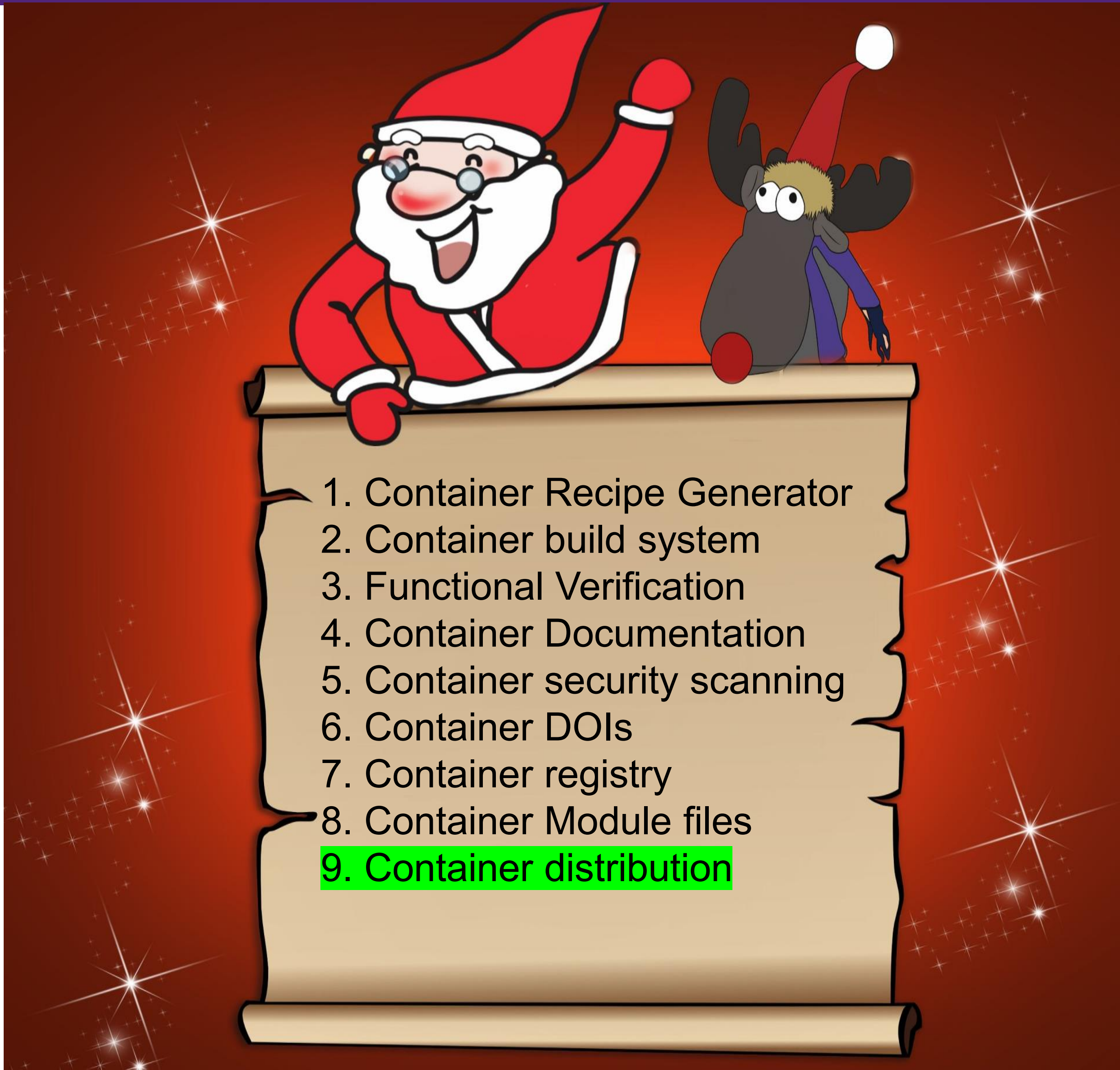
- download and unpack singularity containers to CVMFS storage for distribution and on-demand access and distribute reference data
- every user-site has their own subset of containers that they offer to their users (e.g. architecture optimized)



Proposed first phase

- Network of Stratum 1 servers in Australia hosted by AARNet & ARDC
- These Stratum 1 servers distribute existing community operated Stratum 0 servers, such as
  - EESSI,
  - DRA,
  - Neurodesk,
  - Galaxy Project,
  - AARNet's Biocommons RNAseq repositories,
  - SHPC BioConda containers

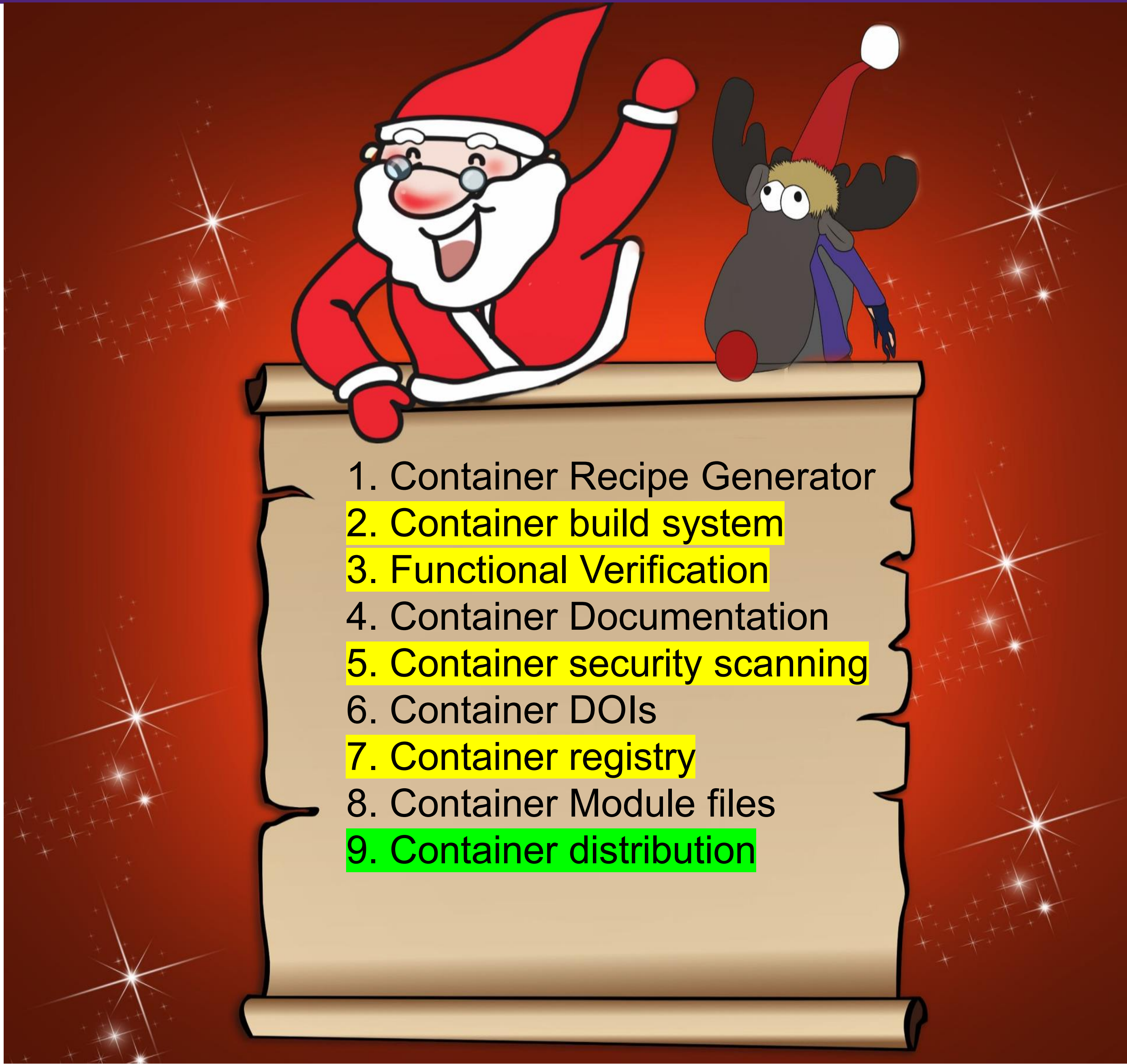
-> Is there demand for a national service like this?



1. Container Recipe Generator
2. Container build system
3. Functional Verification
4. Container Documentation
5. Container security scanning
6. Container DOIs
7. Container registry
8. Container Module files
9. Container distribution

Proposed second phase

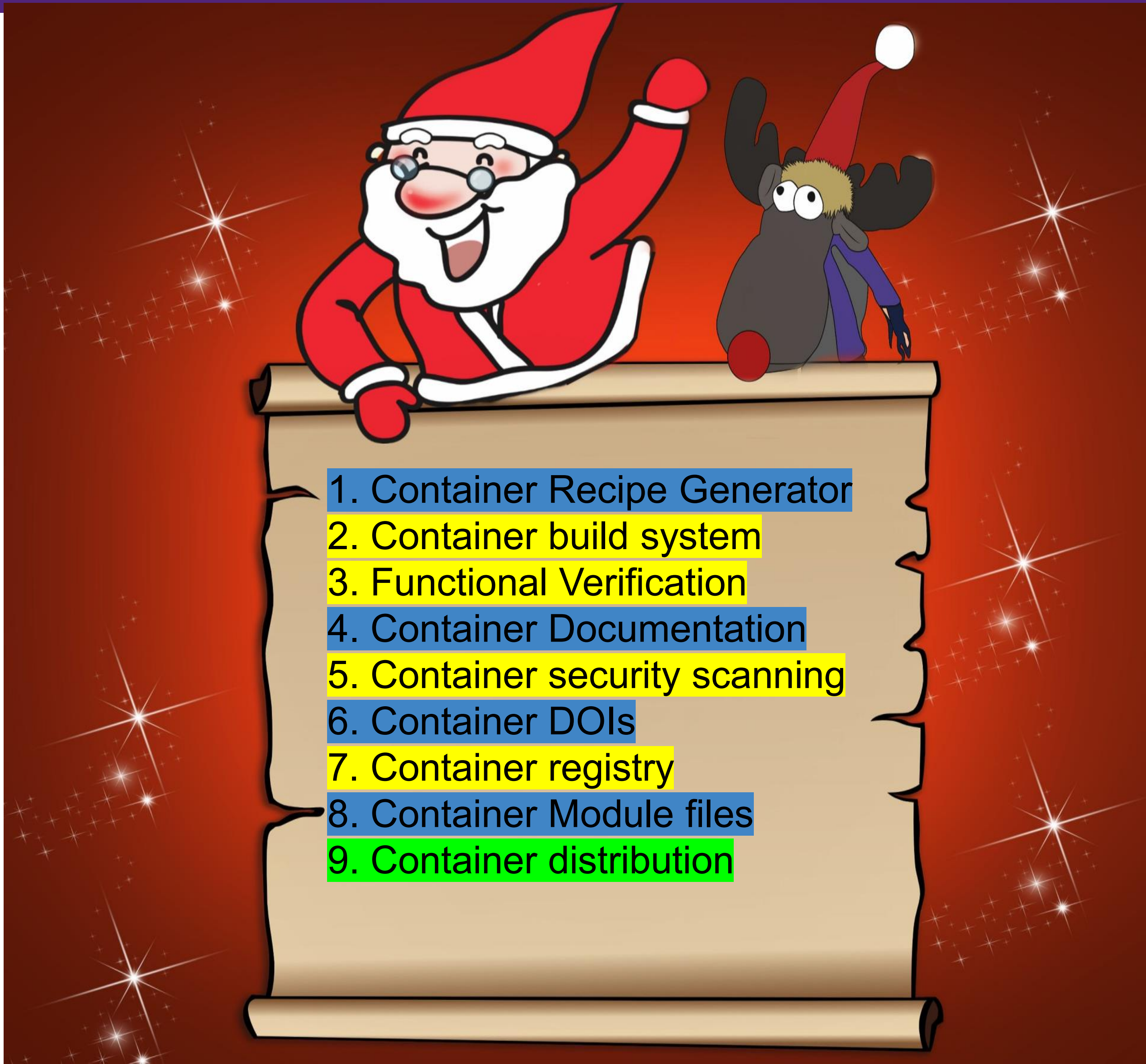
- building up a blueprint for a community to
  - build and test (functional+security) their own containers using
  - Github actions + Nectar Cloud + Nectar Harbour registry + cvmfs stratum 0 setup



1. Container Recipe Generator
2. Container build system
3. Functional Verification
4. Container Documentation
5. Container security scanning
6. Container DOIs
7. Container registry
8. Container Module files
9. Container distribution

Proposed third phase

- making it easier to generate container recipes
- enabling the long-term archival of containers and DOI minting
- building up a tool registry for hosting metadata about containers + discovery
- improve generation of module files for HPC use cases



1. Container Recipe Generator
2. Container build system
3. Functional Verification
4. Container Documentation
5. Container security scanning
6. Container DOIs
7. Container registry
8. Container Module files
9. Container distribution

Governance

- ARDC ARCOS Technical Working group for coordination between communities?
- joint working group between ARDC/AARNet/QCIF/community representatives

Join the discussion, Tuesday 17th, 14.15-15.15

Boulevard B2

Enabling Platforms & Services

Aleem Uddin & Greg D’arcy

BoF Session

Software Containers in Australian Research Sector: ARCOS Community Building and Collaboration on SciDir – a Scientific Software Distribution Project

Aleem Uddin, Greg D’arcy, Steven Manos, Mark Gray, David Powell, Paul Coddington, Jake Yip, Ping Chen

https://bit.ly/scidirpoll

- weighting of priorities/phases?
- anything else that should be on the Wishlist?
- governance model?