Debugging Open Source

Dave Hoese

Ugh. Why?!?!?

About Me

- Software Developer at Space Science and Engineering Center (SSEC)
- Open Source Maintainer ← Disclaimer: BIAS!
  - Satpy
  - Pyresample
  - Vispy
  - Donfig
  - Geoxarray
- Open Source Contributor
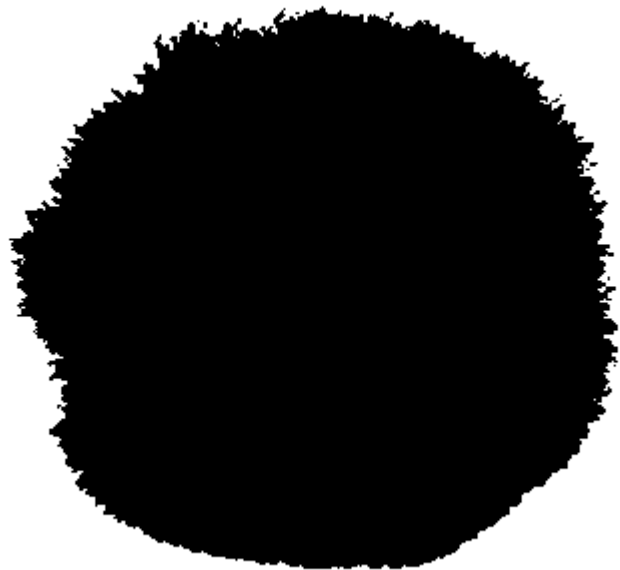
The black hole of debugging
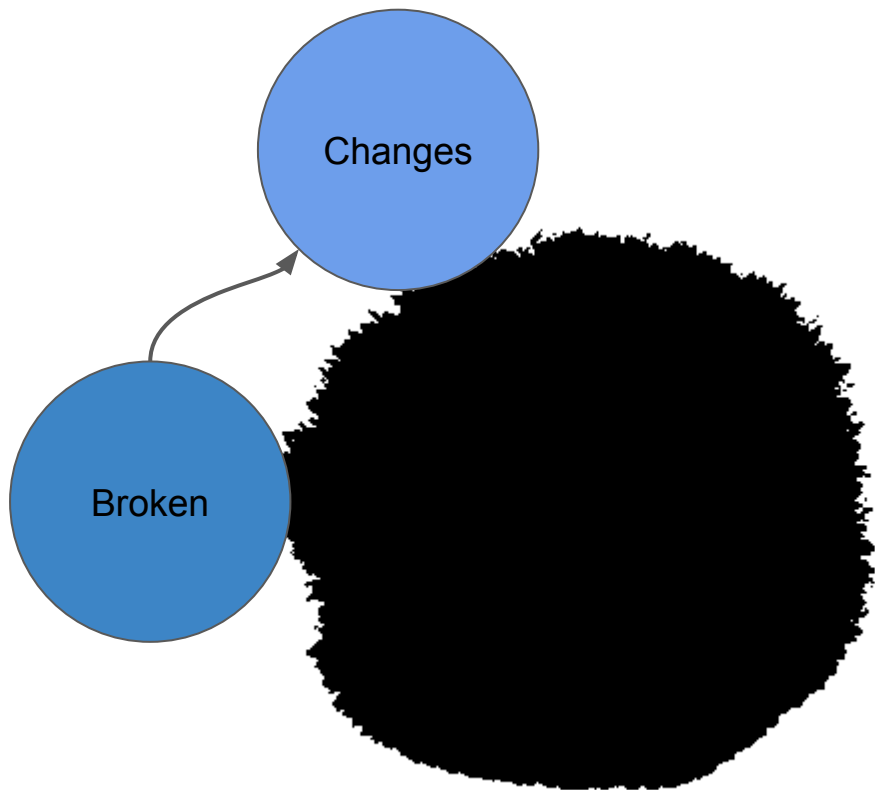
The black hole of debugging

Broken → Changes → Communicate

The black hole of debugging

Communicate

Changes

Broken

Find

The black hole of debugging

Broken → Changes → Communicate → Find → Fix

The black hole of debugging

The black hole of debugging

Broken → Changes → Communicate → Find → Fix → Share
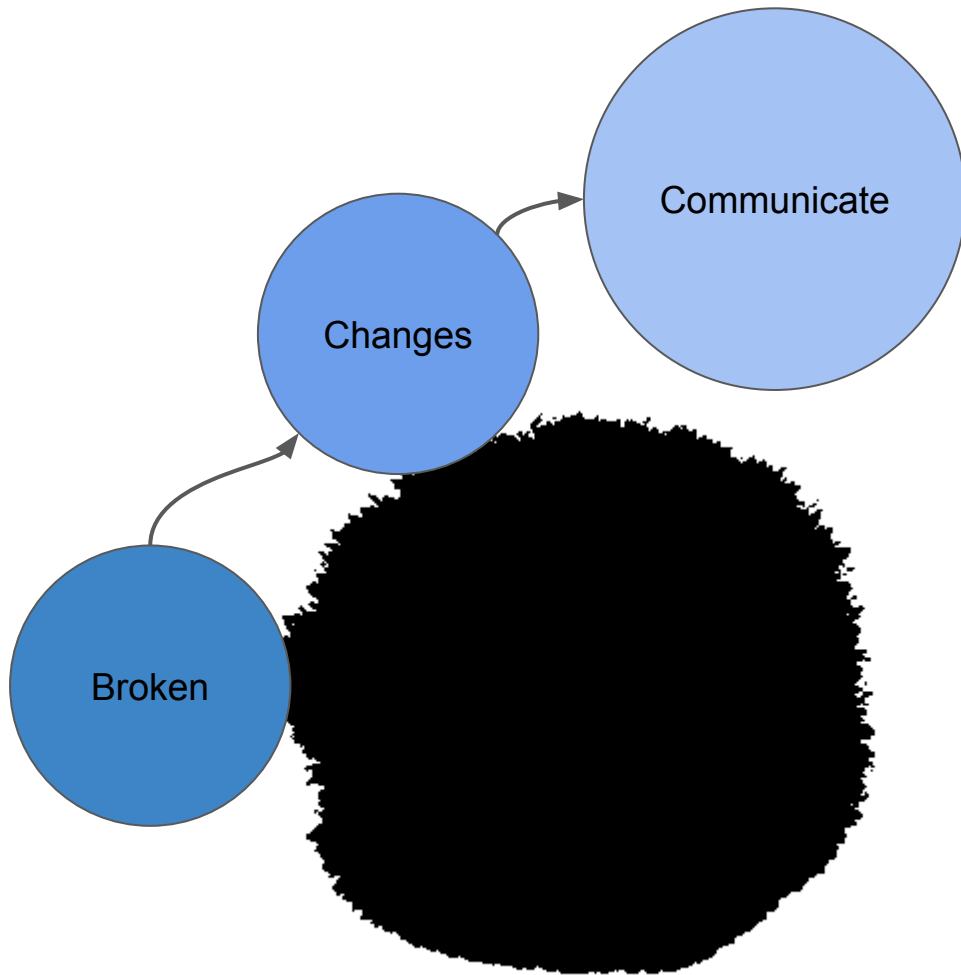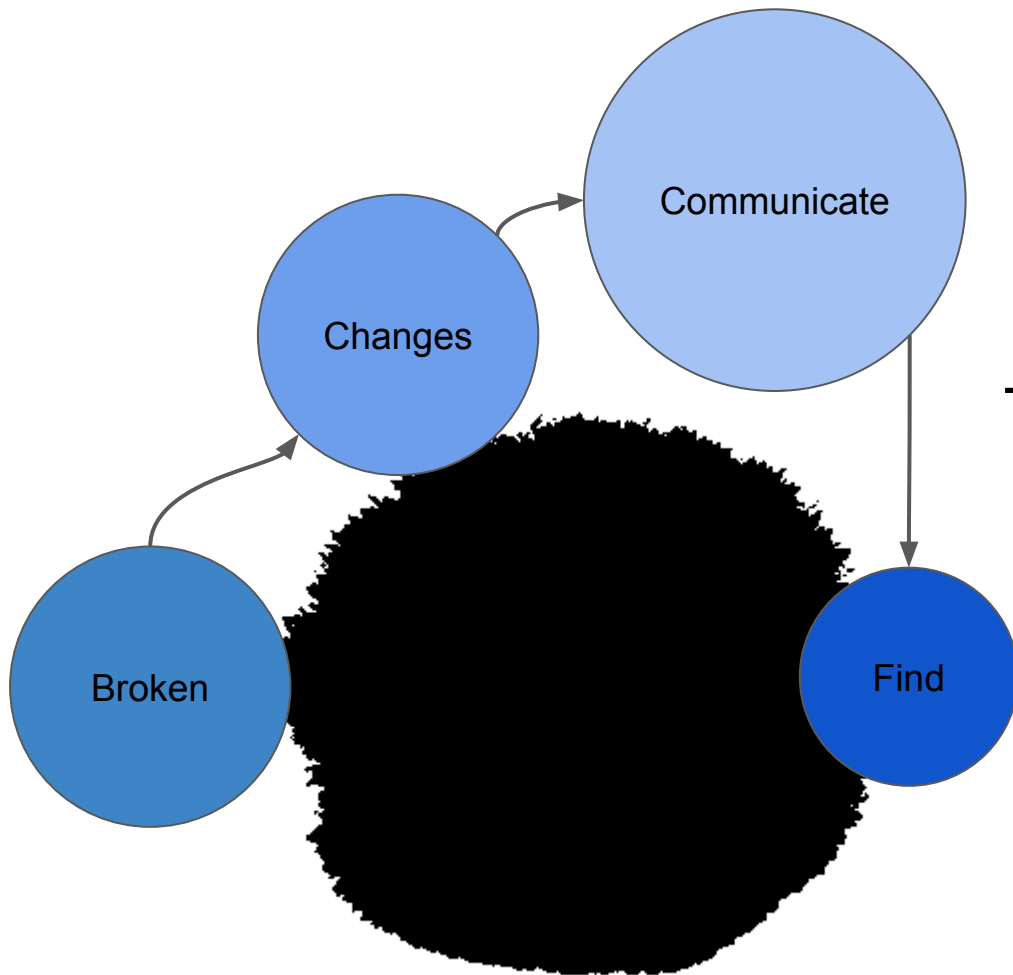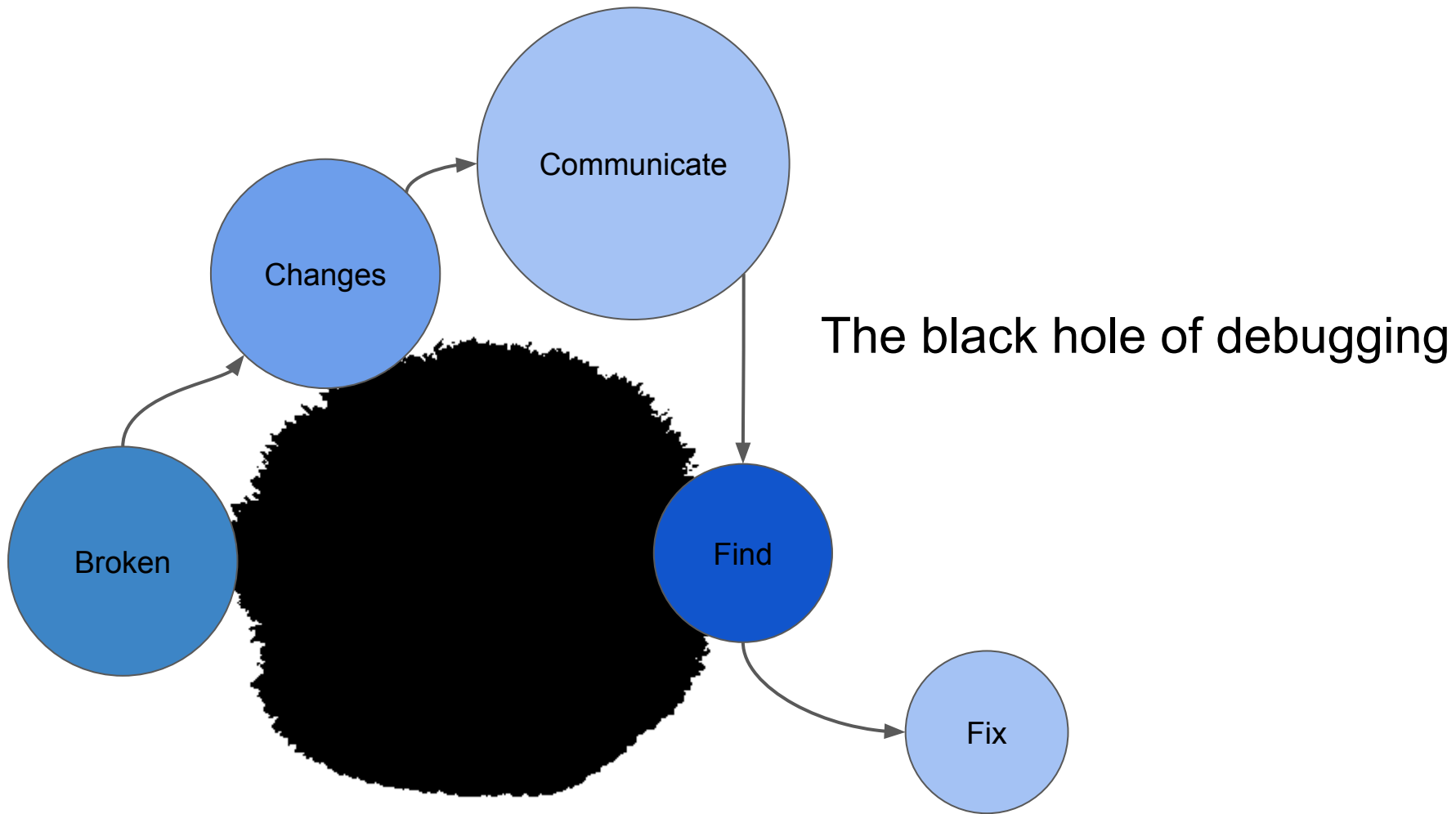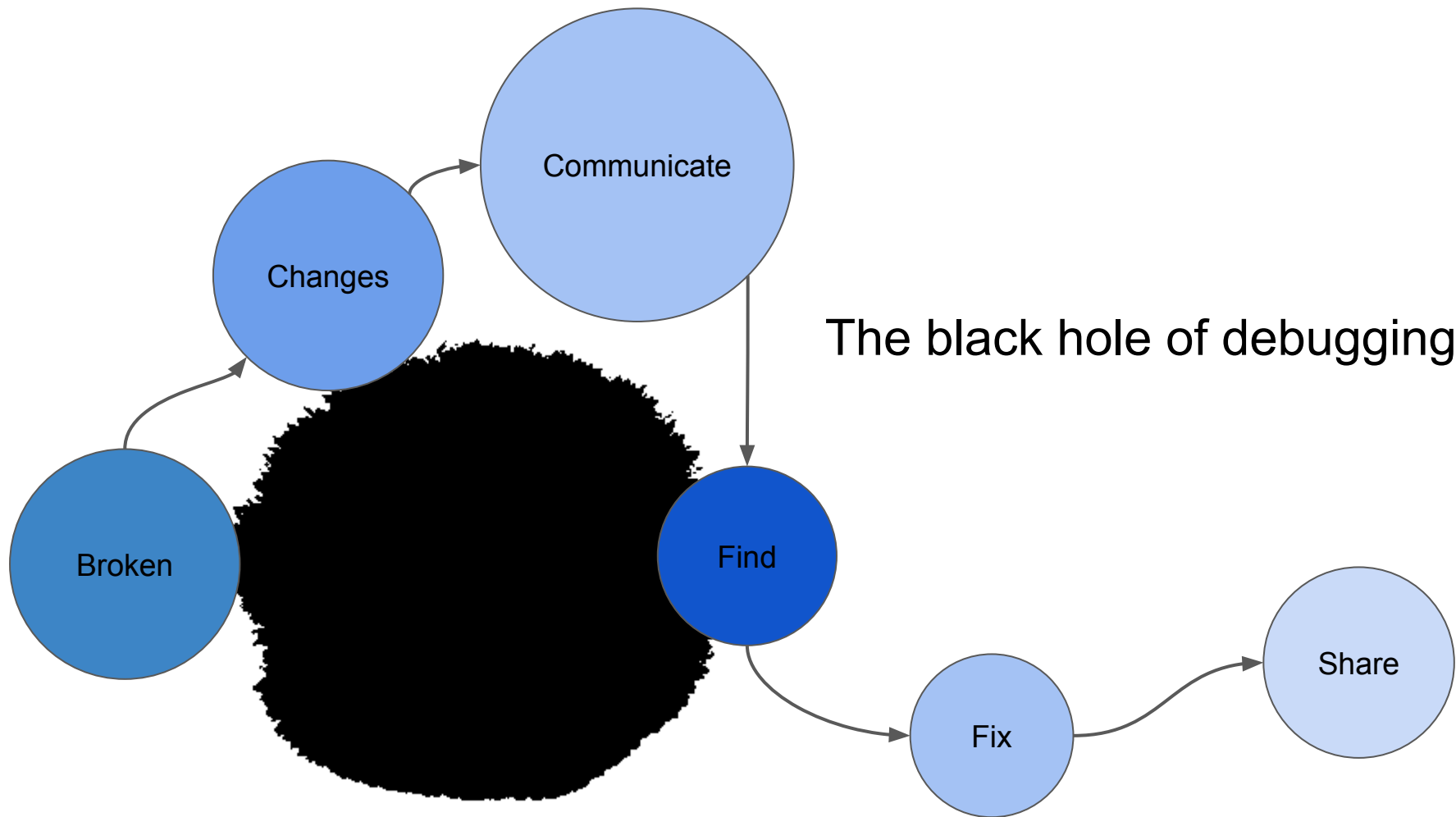
Checklist

✓   \<TODO\>

Background

- Satpy - meteorological satellite processing library
  - Source code hosted on Github
  - Uses Sphinx for documentation rendering
  - Documentation hosted on readthedocs
- Sphinx - documentation tool
  - Converts restructuredtext (.rst) to other formats (HTML, PDF, epub, etc)
  - Basic build is as simple as "make html"
  - Features include reference linking, glossary, indexing, etc
  - Includes interfaces for third-party extensions (API docs from source, link to other libraries, etc)
- ReadTheDocs (readthedocs.io) - documentation hosting
  - Free (for open source) documentation building and hosting
  - Version management
  - GitHub Pull Request integration

Ugh. Why?!?!?

Read the Docs

Sign up · Log in

satpy — AB EN

Versions 13 | Builds 6046

✕ Pull request 2908 / Builds / #27271974

100%

Build failed

Started
6 days ago

Duration
3m 55s

Pull request
#2908 (54c0157c)

⟩_ Output | Raw log | Debug | ✓ Wrap output | View docs

⊗ Build terminated due to excessive memory consumption. — This build was terminated due to excessive memory consumption. Read more about time and memory limits in our documentation.

```
>  git clone --depth 1 https://github.com/pytroll/satpy.git .   0s
>  git fetch origin --force --prune --prune-tags --depth 50 pull/2908/head:external-2908   2s
>  git checkout --force 54c0157c36c99c9d408dd7a75bd46e2c58f7bf37   0s
>  cat .readthedocs.yml   0s
>  git fetch --tags   2s
>  asdf global python mambaforge-4.10.3-10   0s
>  cat doc/rtd_environment.yml   0s
>  mamba env create --quiet --name 2908 --file doc/rtd_environment.yml   164s
>  git update-index --assume-unchanged doc/rtd_environment.yml doc/source/conf.py   0s
>  cd doc/source && if [ "$READTHEDOCS_VERSION" = "latest" ] || [ "$READTHEDOCS_VERSION" = "stable" ]; then python
   generate_area_def_list.py; else touch area_def_list.rst; fi   0s
>  python -m sphinx -T -W --keep-going -b html -d _build/doctrees -D language=en . $READTHEDOCS_OUTPUT/html   42s   Exit: 137
```

Ugh. Why?!?!?

Sign up   Log in

satpy   AB EN ▼

Versions 13   Builds 6046

❌ Pull request 2908 / Builds / #27271974   📅 Started
                                                                6 days ago
                                                             🕐 Duration
                                                 100%   3m 55s
            Build failed   ⑂ Pull request
                                                                #2908 (54c0157c)

>_ Output   🔲 Raw log   🔖 Debug   ☑ Wrap output   📄 View docs

ⓧ Build terminated due to excessive memory consumption. This build was terminated due to excessive memory consumption. Read more about time and memory limits in our documentation.

```
>    git clone --depth 1 https://github.com/pytroll/satpy.git .  0s
>    git fetch origin --force --prune --prune-tags --depth 50 pull/2908/head:external-2908  2s
>    git checkout --force 54c0157c36c99c9d408dd7a75bd46e2c58f7bf37  0s
>    cat .readthedocs.yml  0s
>    git fetch --tags  2s
>    asdf global python mambaforge-4.10.3-10  0s
>    cat doc/rtd_environment.yml  0s
>    mamba env create --quiet --name 2908 --file doc/rtd_environment.yml  164s
>    git update-index --assume-unchanged doc/rtd_environment.yml doc/source/conf.py  0s
>    cd doc/source && if [ "$READTHEDOCS_VERSION" = "latest" ] || [ "$READTHEDOCS_VERSION" = "stable" ]; then python
     generate_area_def_list.py; else touch area_def_list.rst; fi  0s
>    python -m sphinx -T -W --keep-going -b html -d _build/doctrees -D language=en . $READTHEDOCS_OUTPUT/html  42s  Exit 137
```

Checklist

✓   Something's broken

Where do we start?

- Can we reproduce it?
  - Restart job
  - Try locally
- What's changed?
  - "This worked 2 days ago"
  - Did I break this?
  - Did a service (ex. GitHub Actions, AWS, readthedocs) change?
  - Did a dependency change?

Where do we start?

- Can we reproduce it?
  - Restart job…**yep, failed**.
  - Try locally
- What's changed?
  - "This worked 2 days ago"...**yep, yesterday.**
  - Did I break this?...**No?**
  - Did a service (ex. GitHub Actions, AWS, readthedocs) change?...**Unlikely.**
  - Did a dependency change?...**Probably…**

```
mamba env create --quiet --name 2908 --file doc/rtd_environment.yml   164s

Package                          Version  Build          Channel                      Size
_____

  Install:
_____


  + _libgcc_mutex                    0.1  conda_forge    conda-forge/linux-64       Cached

  + _openmp_mutex                    4.5  2_gnu          conda-forge/linux-64        23 KB

  + adwaita-icon-theme              47.0  unix_0         conda-forge/noarch         554 KB

  + affine                         2.4.0  pyhd8ed1ab_1   conda-forge/noarch          19 KB
```

```
+ snappy                    1.2.1   h8bd8927_1     conda-forge/linux-64    42 KB
+ snowballstemmer           2.2.0   pyhd8ed1ab_0   conda-forge/noarch      57 KB
+ snuggs                    1.4.7   pyhd8ed1ab_2   conda-forge/noarch      11 KB
+ sortedcontainers          2.4.0   pyhd8ed1ab_1   conda-forge/noarch      28 KB
+ sphinx                    8.2.0   pyhd8ed1ab_0   conda-forge/noarch       1 MB
+ sphinx_rtd_theme          3.0.1   pyha770c72_0   conda-forge/noarch       4 MB
+ sphinxcontrib-apidoc      0.3.0   py_1           conda-forge/noarch      10 KB
+ sphinxcontrib-applehelp   2.0.0   pyhd8ed1ab_1   conda-forge/noarch      29 KB
+ sphinxcontrib-devhelp     2.0.0   pyhd8ed1ab_1   conda-forge/noarch      24 KB
```

Sphinx 8.2.0

```
pip install Sphinx==8.2.0
```

Released: Feb 18, 2025

Python documentation generator

Project description

pypi v8.2.1   ○ CI passing   docs passing   License BSD 2-Clause

Sphinx makes it easy to create intelligent and beautiful documentation.

Sphinx uses reStructuredText as its markup language, and many of its strengths come from the power and straightforwardness of reStructuredText and its parsing and translating suite, the Docutils.

Features

- **Output formats**: HTML, PDF, plain text, EPUB, TeX, manual pages, and more

Verified details ✔
These details have been verified by PyPI

Reproduce locally

- Checkout Satpy source
- Create environment with new sphinx (8.2.0)
- Use "/usr/bin/time -v <command>" for time and memory information

GNU "time" - /usr/bin/time -v (8.2.0)

```
Build finished. The HTML pages are in build/html.
    Command being timed: "make html"
    User time (seconds): 84.49
    System time (seconds): 4.46
    Percent of CPU this job got: 100%
    Elapsed (wall clock) time (h:mm:ss or m:ss): 1:28.61
    Average shared text size (kbytes): 0
    Average unshared data size (kbytes): 0
    Average stack size (kbytes): 0
    Average total size (kbytes): 0
    Maximum resident set size (kbytes): 8653184
    Average resident set size (kbytes): 0
```

Checklist

✓   Something's broken

✓   Reproducible

Prove it!

- Create environment with old sphinx (8.1.3)
- Use "/usr/bin/time -v <command>" for time and memory information

GNU "time" - /usr/bin/time -v (8.1.3)

```
Build finished. The HTML pages are in build/html.
    Command being timed: "make html"
    User time (seconds): 75.01
    System time (seconds): 1.38
    Percent of CPU this job got: 100%
    Elapsed (wall clock) time (h:mm:ss or m:ss): 1:15.76
    Average shared text size (kbytes): 0
    Average unshared data size (kbytes): 0
    Average stack size (kbytes): 0
    Average total size (kbytes): 0
    Maximum resident set size (kbytes): 2579220
    Average resident set size (kbytes): 0
```

Checklist

✓ Something's broken

✓ Reproducible

✓ Sphinx did it!

What changed…in sphinx?

Changelog:

https://www.sphinx-doc.org/en/master/changes/8.2.html#release-8-2-0-released-feb-18-2025

Things to look for:

- case specific (memory related)
- new features
- backwards incompatible changes
- suspicious bug fixes

What changed…in sphinx?

Changelog:

https://www.sphinx-doc.org/en/master/changes/8.2.html#release-8-2-0-released-feb-18-2025

Things to look for:

- case specific (memory related)
- new features
- backwards incompatible changes
- suspicious bug fixes
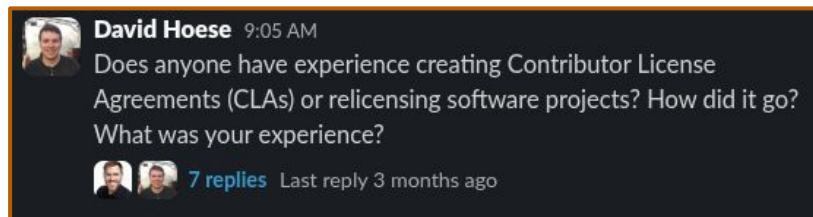
What did I find? NOTHING

Communicate - Is anyone else having this problem?

- Check existing bug reports/issues
- Reach out to colleagues/friends/MadPy
- Mailing lists, discord, discourse, slack, social media, stackoverflow
- File a new issue

Filing a bug report - Best Practice

- Follow issue templates

Create new issue

Bug report · Choose a different template

Add a title *

Title

Describe the bug *

A clear and concise description of what the bug is, including the expected behaviour and what has gone wrong. Please include screenshots, if applicable.

| Write | Preview |

Leave a comment

📎 Paste, drop, or click to add files

How to Reproduce *

Please provide steps to reproduce this bug, with the smallest possible set of source files. For normal bugs this should ideally be one `index.rst` file, and for `sphinx.ext.autodoc` bugs, this should ideally be a single `index.rst` file, and a single example Python module.

| Write | Preview |

```
Minimal method (you can also paste the contents of ``index.rst`` and
``conf.py`` into this report):
```bash
$ echo "Content demonstrating the bug..." > index.rst
$ echo "" > conf.py
$ sphinx-build -M html . _build
```

📎 Paste, drop, or click to add files

Environment info

Environment Information *

Install the latest Sphinx `pip install -U "sphinx>=5.3"` then run `sphinx-build --bug-report` or `python -m sphinx --bug-report` . and paste the output here.

Assignees

No one assigned

Labels

type:bug

Type

No type

Projects

No projects

Milestone

No milestone

ⓘ Remember, contributions to this repository should follow its contributing guidelines.

Filing a bug report - Best Practice

- Follow issue templates
- Simplify - Minimal Verifiable Complete Example (MVCE)
- Provide context - OS, Environment, and any other details

Filing a bug report - Best Practice

- Follow issue templates
- Simplify - Minimal Verifiable Complete Example (MVCE)
- Provide context - OS, Environment, and any other details
- Tone, Language, Phrasing

Filing a bug report - Best Practice

- Follow issue templates
- Simplify - Minimal Verifiable Complete Example (MVCE)
- Provide context - OS, Environment, and any other details
- Tone, Language, Phrasing
- Refine the details and assist in finding the problem
- https://github.com/sphinx-doc/sphinx/issues/13370

Possible outcomes of filing an issue

- Maintainer or contributor fixes it for you
- Won't fix
- Silence
- You're responsible
- All of the above

Checklist

✓   Something's broken

✓   Reproducible

✓   Sphinx did it!

✓   Maintainers have been informed

Finding the problem

- Release notes
- Print debugging
- Interactive debugger (pdb, ipdb, IDE, etc)
- git blame/git log
- git bisect
- strace (segmentation faults or other C-level failures on linux)
- CI "debug mode"
- other?

Narrow it down to a single "thing" (commit or pull request) if possible.

git bisect

- Identify a broken commit (bad)
- Identify a working commit (good)
- Optionally but recommended:
  Create a script to automatically test commits in between
- Fast binary search

git bisect - the script (check_satpy_docs.sh)

```bash
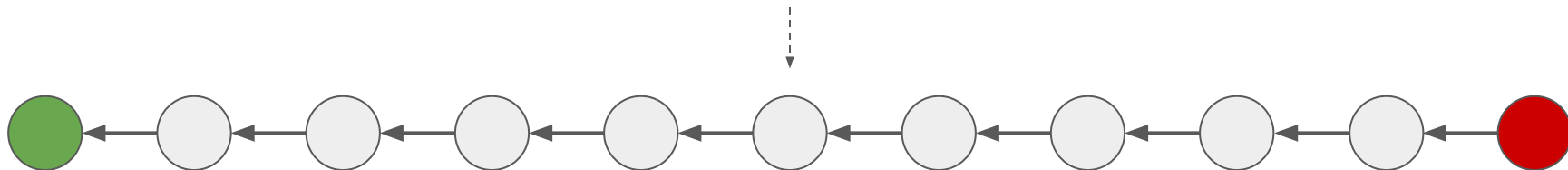#!/usr/bin/env bash

echo "Installing sphinx..."
pip install -e . --no-deps

echo "Building docs..."
cd ~/repos/git/satpy/doc
make clean
max_rss=$(/usr/bin/time -v make html 2>&1 | grep "Maximum resident set" | awk
'{print $6}')
echo "Maximum RSS: $max_rss"
if [[ $max_rss -gt 3000000 ]]; then
    exit 1
else
    exit 0
fi
```

git bisect - run it

```
$ git bisect start  # <-- 8.1.3
status: waiting for both good and bad commits
$ git bisect good
status: waiting for bad commit, 1 good commit known
$ git checkout v8.2.0
Previous HEAD position was a1510de47 Bump to 8.1.3 final
HEAD is now at 78bb5a184 Bump to 8.2.0 final
$ git bisect bad
Bisecting: 250 revisions left to test after this (roughly 8 steps)
[4d77504a7b0088d491f7e84a38ea534d766379b3] Suggest GitHub Discussions in the issue
chooser
$ git bisect run ./check_satpy_docs.sh
running './check_satpy_docs.sh'
Installing sphinx...
```

git bisect - run it

[fec4d7c2f1f741374b3a6bba16960ab232a8f3dd] Viewcode: Fix issue with import
paths that differ from the directory structure (#13195)
running './check_satpy_docs.sh'
Installing sphinx...
…
Successfully installed Sphinx-8.2.0
Building docs...
…
Maximum RSS: 6116100
fec4d7c2f1f741374b3a6bba16960ab232a8f3dd is the first bad commit
commit fec4d7c2f1f741374b3a6bba16960ab232a8f3dd
Author: ProGamerGov <ProGamerGov@users.noreply.github.com>
Date: Sun Jan 5 03:43:57 2025 -0700

        Viewcode: Fix issue with import paths that differ from the directory
structure (#13195)

git bisect - results

[Commit](#) → [Pull Request](#)

Communicate Again!

- On the commit itself?
- In the pull request?
- In the issue

Checklist

✓   Something's broken

✓   Reproducible

✓   Sphinx did it!

✓   Maintainers have been informed

✓   Found the problem code

Finding the solution

- Understanding the code
    - The commit/pull request and its goal
    - Where that code fits into the bigger picture
- Understanding the tests
    - Where is the code tested? How?
    - Is your use case represented? Should it be?
- How to find the fix?
    - TDD?
    - Your MVCE?
    - Depends

Finding the solution - sphinx viewcode

- Parse API docs and generate source code HTML page
  - https://satpy.readthedocs.io/en/stable/api/satpy.scene.html#satpy.scene.Scene
- API docs refer to importable python object
  - `mylib.mysubpkg.mymod.myobj` → "myobj" in "mylib/mysubpkg/mymod.py"
  - `mylib.mysubpkg.myobj` → "myobj" in "mylib/mysubpkg/__init__.py"
- Understand importlib
- Understand going from "standard" importlib to low-level
- Understand the role of "sys.modules"
- Learned: low-level importlib doesn't use "sys.modules"

Finding the solution - sphinx viewcode

- Parse API docs and generate source code HTML page
  - https://satpy.readthedocs.io/en/stable/api/satpy.scene.html#satpy.scene.Scene
- API docs refer to importable python object
  - `mylib.mysubpkg.mymod.myobj` → "myobj" in "mylib/mysubpkg/mymod.py"
  - `mylib.mysubpkg.myobj` → "myobj" in "mylib/mysubpkg/__init__.py"
- Understand importlib
- Understand going from "standard" importlib to low-level
- Understand the role of "sys.modules"
- Learned: low-level importlib doesn't use "sys.modules"
- viewcode was re-importing modules for every documented API object and keeping them in memory

Finding the solution - sphinx viewcode

- Parse API docs and generate source code HTML page
  - https://satpy.readthedocs.io/en/stable/api/satpy.scene.html#satpy.scene.Scene
- API docs refer to importable python object
  - `mylib.mysubpkg.mymod.myobj` → "myobj" in "mylib/mysubpkg/mymod.py"
  - `mylib.mysubpkg.myobj` → "myobj" in "mylib/mysubpkg/__init__.py"
- Understand importlib
- Understand going from "standard" importlib to low-level
- Understand the role of "sys.modules"
- Learned: low-level importlib doesn't use "sys.modules"
- viewcode was re-importing modules for every documented API object and keeping them in memory
- Solution: Go back to basics

Checklist

✓   Something's broken

✓   Reproducible

✓   Sphinx did it!

✓   Maintainers have been informed

✓   Found the problem code

✓   Found the fix!

Submitting the solution - Best Practice

- Contributing Guide!
  - Usually in a CONTRIBUTING.md file or "Contributor's Guide" in the documentation
  - Might include things like:
    - Coding style
    - Development install
    - Testing
    - pre-commit
    - Commit message format
    - How to submit a pull request
- Create the pull request - pull request templates

Submitting the solution - sphinx viewcode

https://github.com/sphinx-doc/sphinx/pull/13380

Checklist

✓   Something's broken

✓   Reproducible

✓   Sphinx did it!

✓   Maintainers have been informed

✓   Found the problem code

✓   Found the fix!

✓   Submitted the fix

Submitting the solution - Gotchas

- Copyrights and Contributor License Agreements
- Authors and Changelogs
- Pull request will automatically update after new commits
- DON'T USE YOUR "main" BRANCH!

Git branching - Main Merging

Upstream "main":

Fork "main":

Git branching - Main Merging

Upstream "main":

Fork "main":

Git branching - Next contribution

Upstream "main":

Fork "main":

Submitting the solution - one branch per pull request

- Create a new branch
  - Old:
    ```
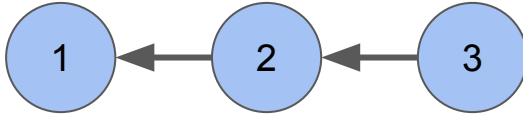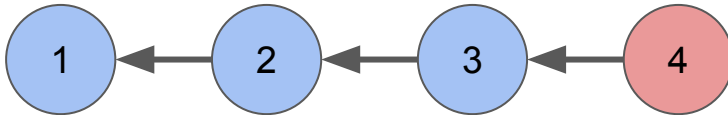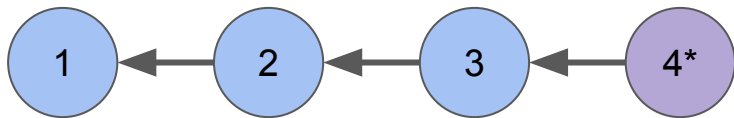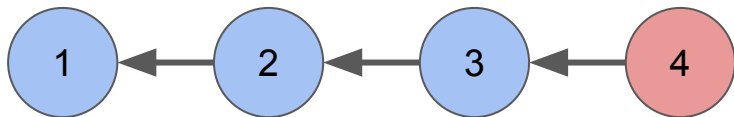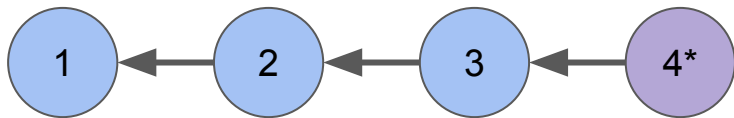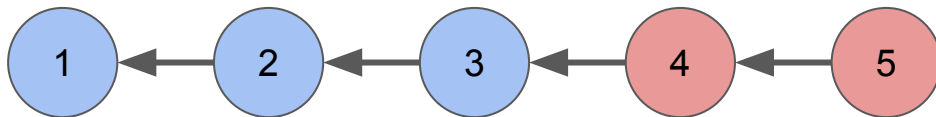    git checkout -b mybranch
    ```
  - New:
    ```
    git switch -c mybranch
    ```
- Delete the branch after merge
  - Github button
  - Command line:
    ```
    # delete remote branch:
    git push -d upstream mybranch
    # delete local branch:
    git branch -d mybranch
    ```

Submitted the solution - Best Practice

- Respond to comments and reviews
- Hope to be merged

Submitted the solution - Best Practice

- Respond to comments and reviews
- Hope to be merged
- Wait for release!

Checklist

✓ Something's broken

✓ Reproducible

✓ Sphinx did it!

✓ Maintainers have been informed

✓ Found the problem code

✓ Found the fix!

✓ Submitted the fix

✓ Fix released!

Questions?

Bonus…

Bonus - 2.6GB is good?

Q: Why, in the working good case, does Satpy's documentation use 2.6GB of memory to generate HTML?

Bonus - 2.6GB is good?

Q: Why, in the working good case, does Satpy's documentation use 2.6GB of memory to generate HTML?

A: Bad test writing