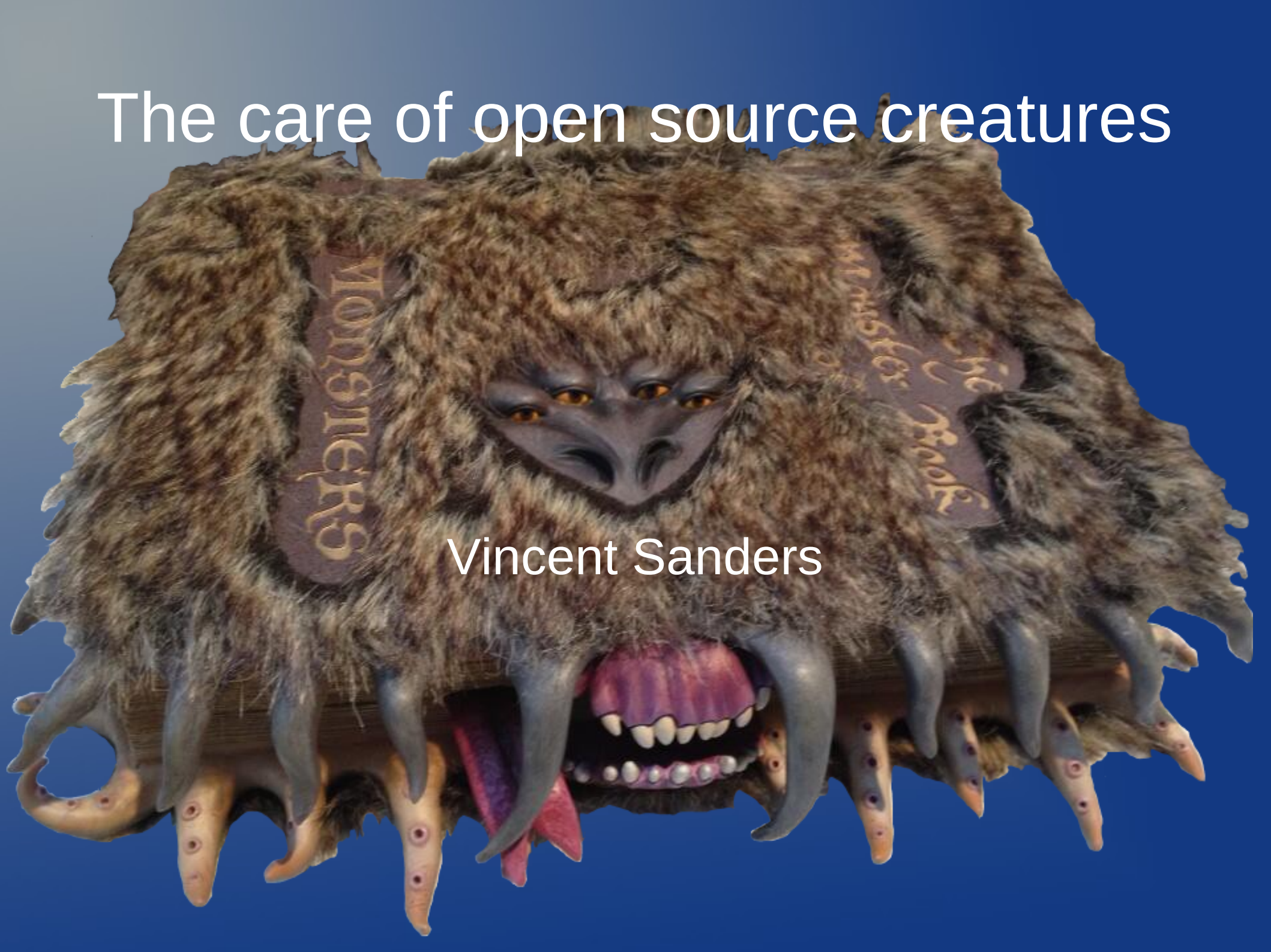
The care of open source creatures



Vincent Sanders

What am I on about?

An examination of:

- What a services a project ought to have
- What options exist to fulfil those requirements
- A practical look at some implementations.

Open Source Life Cycle

- Planning
- Implementing
- Building
- Quality assurance
- Releasing

Planning

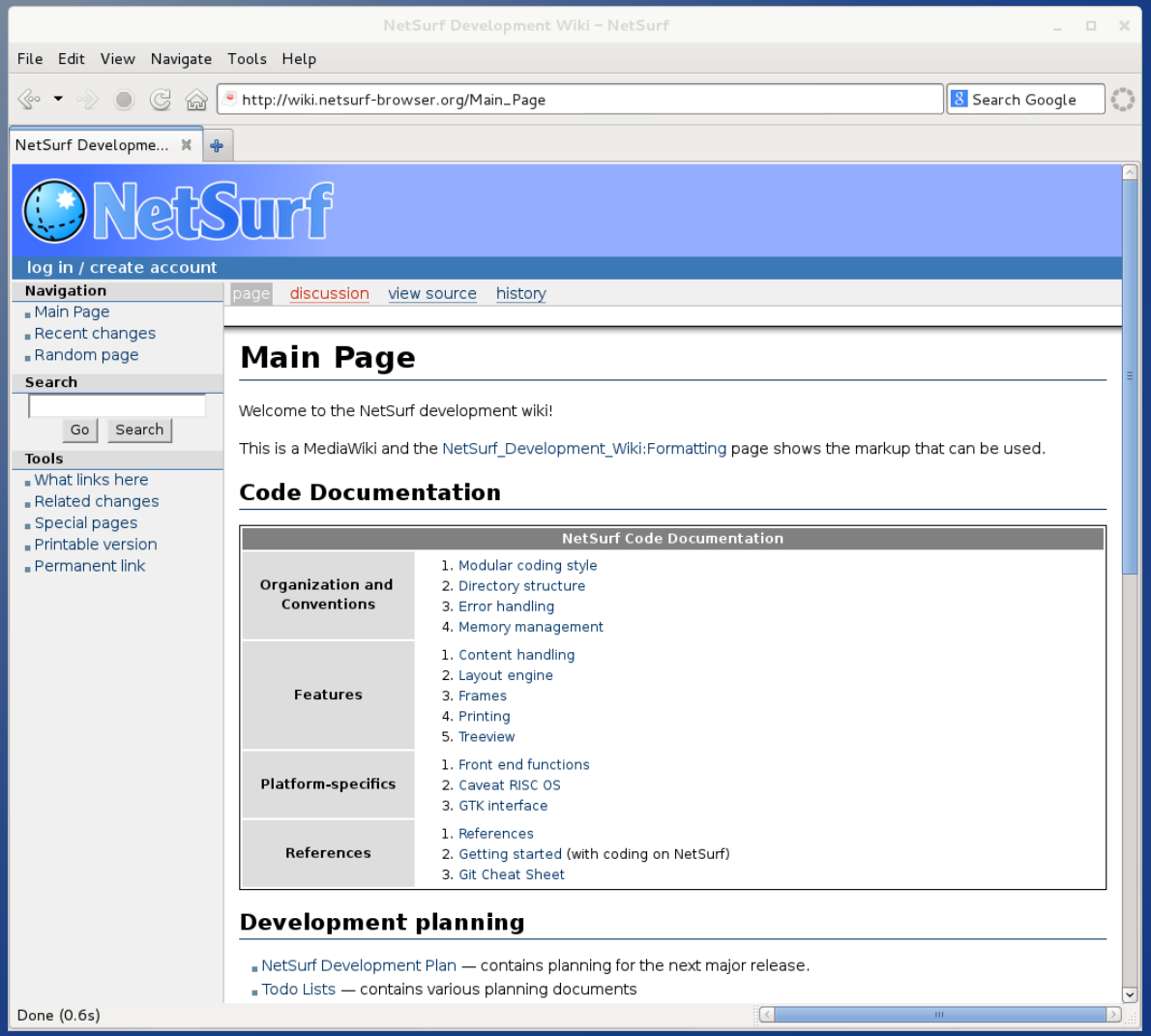
- Planning is usually a social activity
- Important to keep track of decisions
- Communication tools developers actually use
- Flexibility to achieve releases

Planning Infrastructure

- IRC – creating channels on free networks like oftc or freenode is easy
- Email lists are less popular but easy way to communicate with lots of people
- Forums are easy to setup but can degenerate quickly
- Communication with users can occur here to get an idea of what they say they want.

Planning Infrastructure

- A wiki is good for longer term info



Implementing

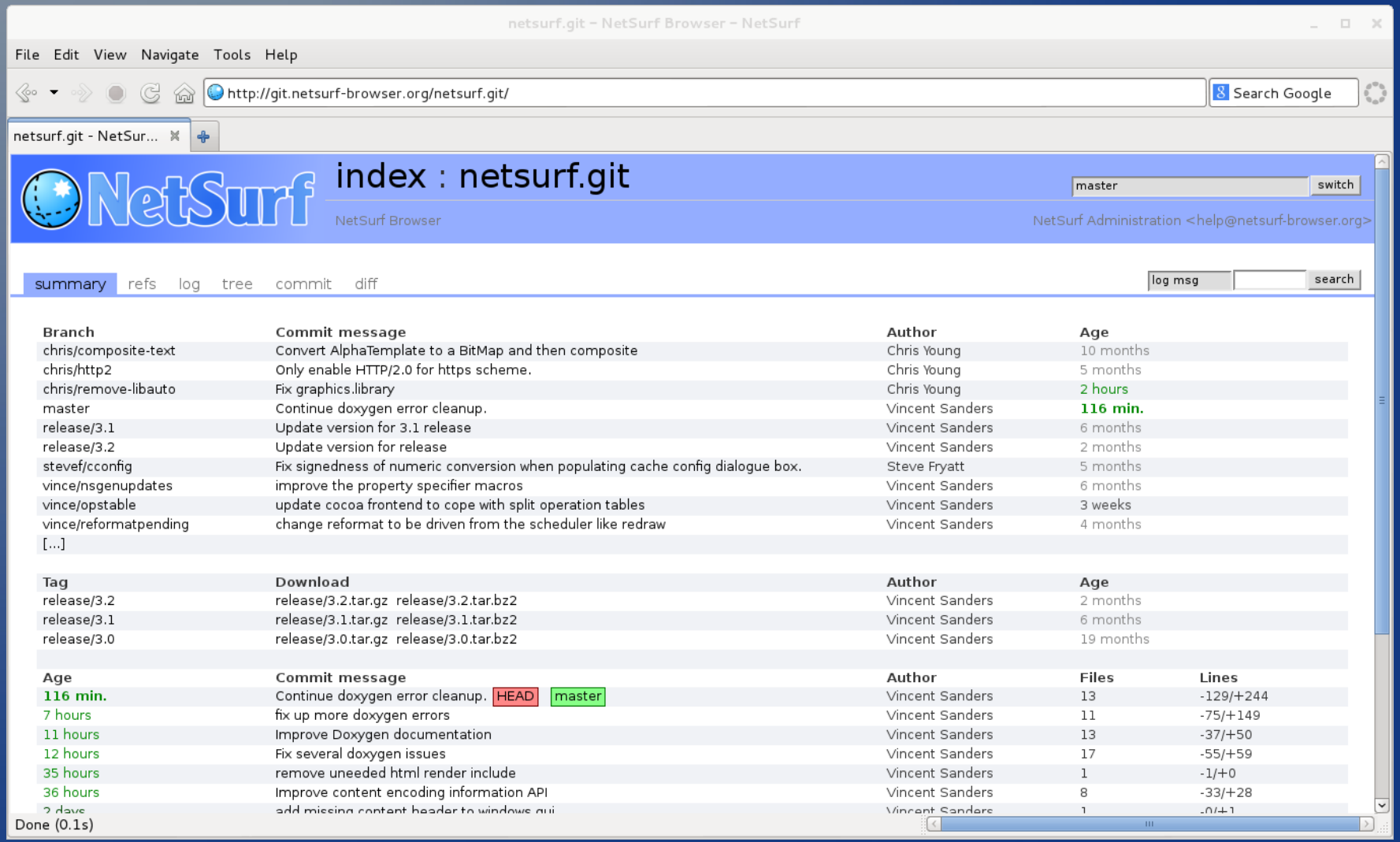
- Code style
- Code documentation

Implementing source control

- Source control is mandatory
- GIT won the argument
- Have a merge policy
- Have a review policy

Implementing source control

- Gitano and cgit are great



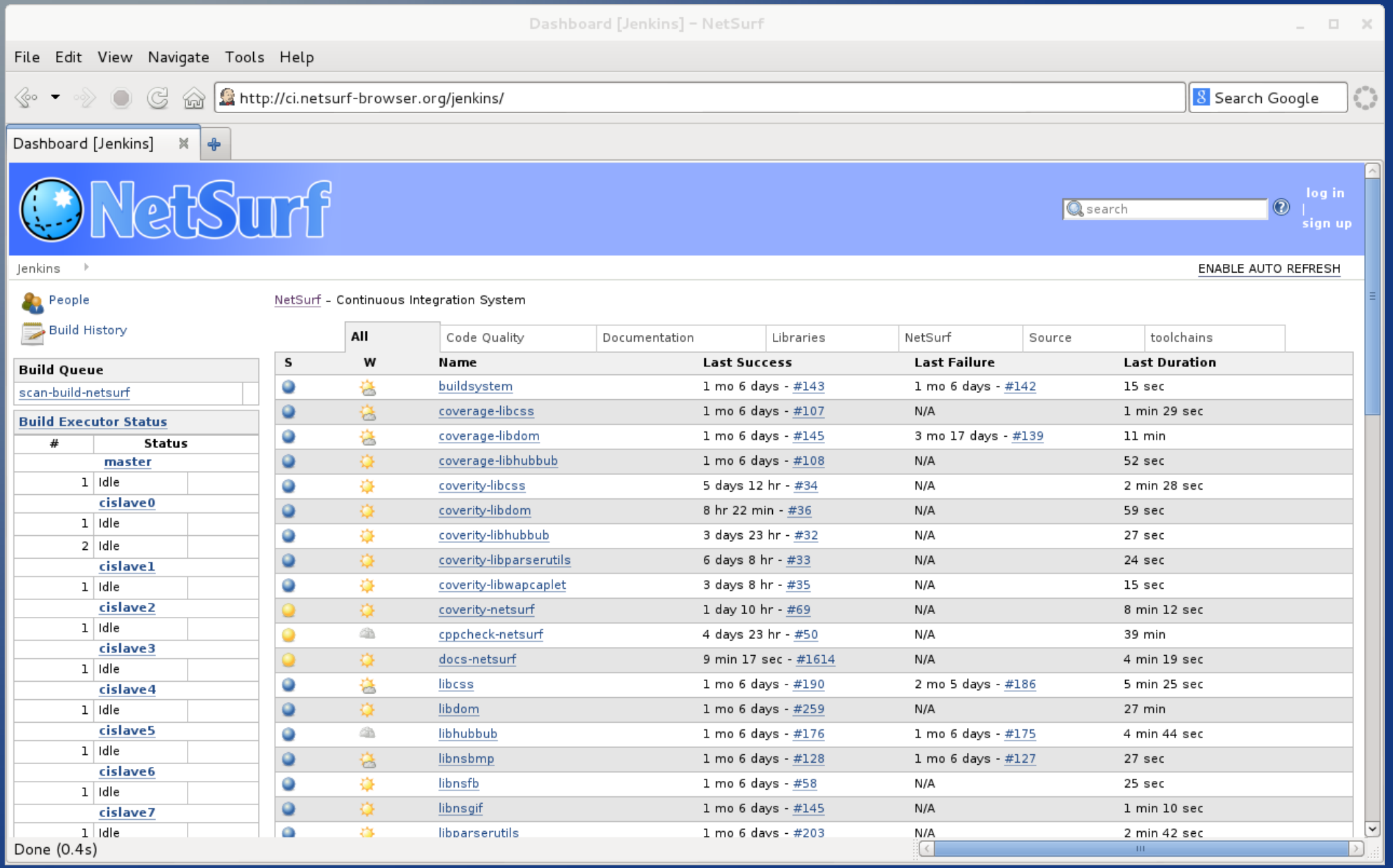
Building

- Master branch should always build
- Getting the software built should be easy
- Build process should be documented
- Continuous integration

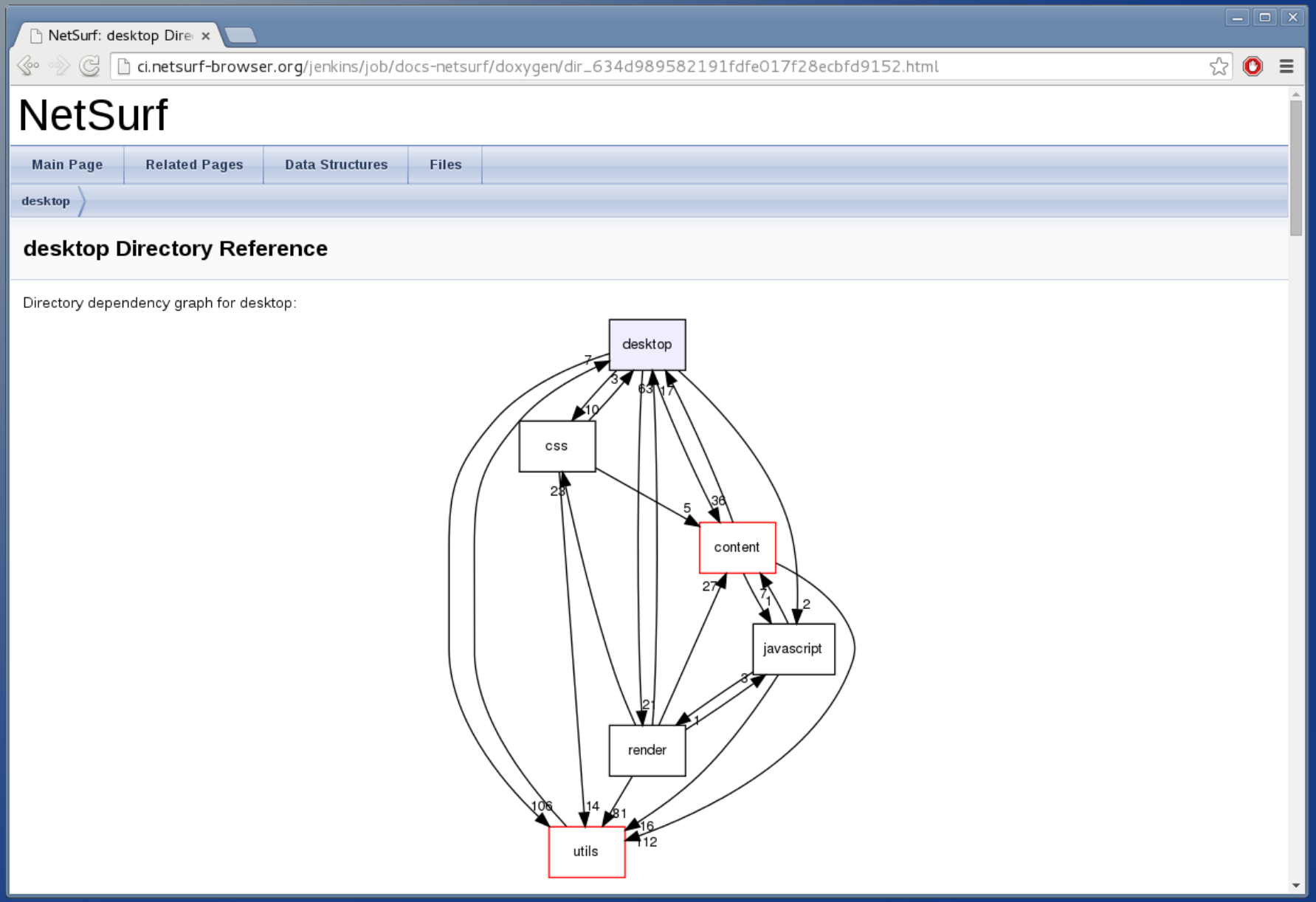
Building with Jenkins

- Jenkins is a CI tool
- Jobs can be triggered by GIT changes
- Jobs can be periodic
- Dependences between modules
- Good mechanisms for feedback

Deploying Jenkins



Deploying Jenkins



Metrics

The NetSurf Open Sou ×

https://www.openhub.net/p/netsurf

... has had 10,854 commits made by 36 contributors
representing 229,165 lines of code

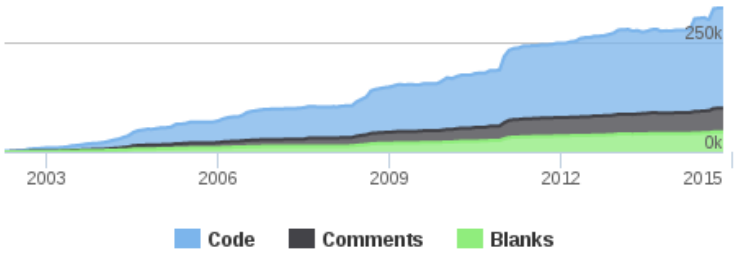
... is mostly written in C
with an average number of source code comments

... has a well established, mature codebase
maintained by a large development team
with stable Y-O-Y commits

... took an estimated 60 years of effort (COCOMO model)
starting with its first commit in April, 2002
ending with its most recent commit 2 days ago

C	86%	XML	6%
12 Other	8%		

Lines of Code



Activity

30 Day Summary
Oct 8 2014 — Nov 7 2014

233 Commits
5 Contributors

12 Month Summary
Nov 7 2013 — Nov 7 2014

1218 Commits
Down -226 (15%) from previous 12 months

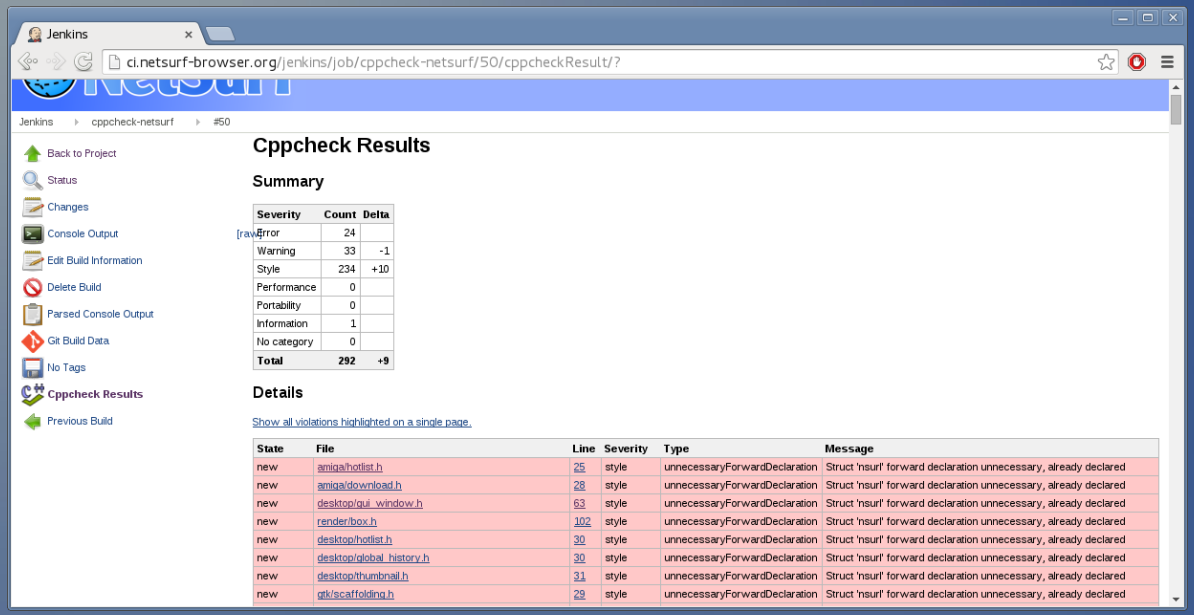
11 Contributors
Down -1 (8%) from previous 12 months

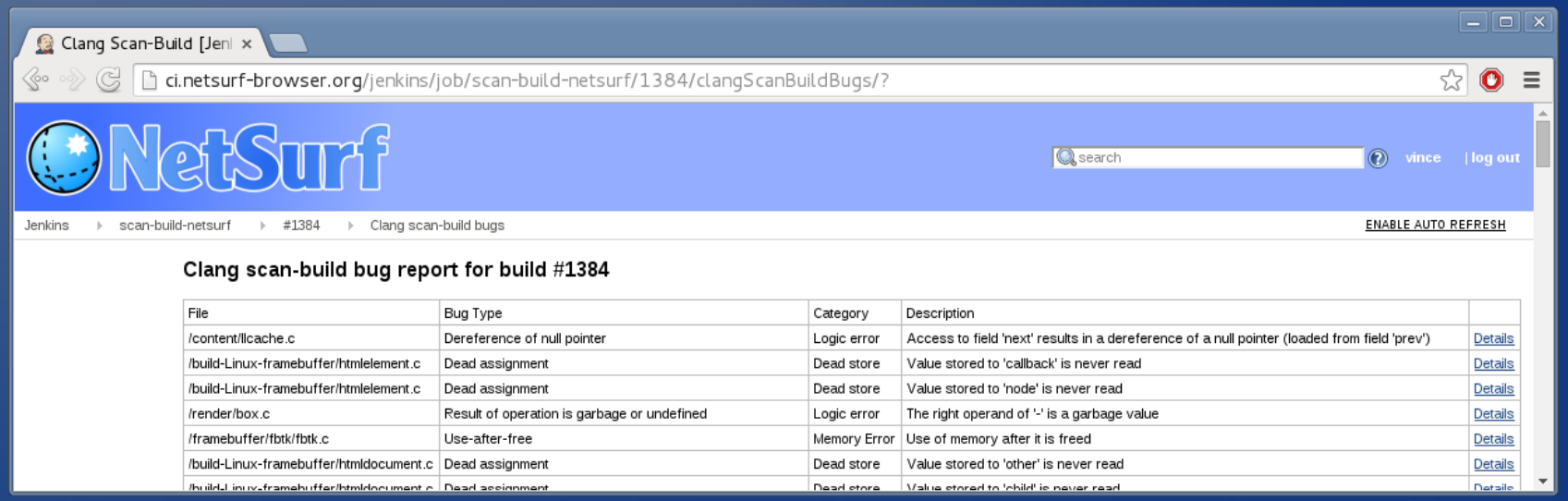
Commits per Month

Quality Assurance

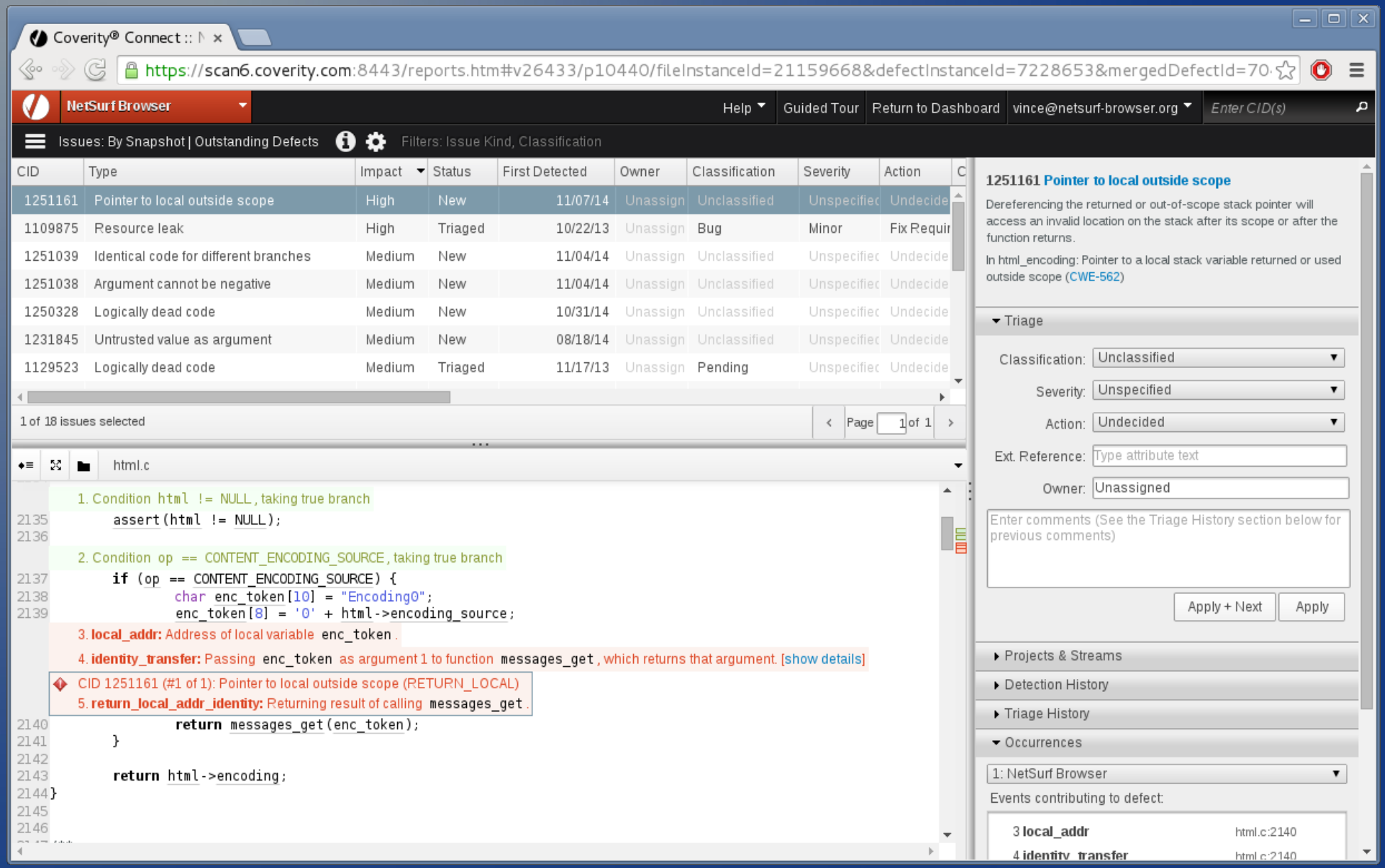
- Static analysis
- Unit testing
- System testing
- Issue tracking
- Metrics

Static analysis





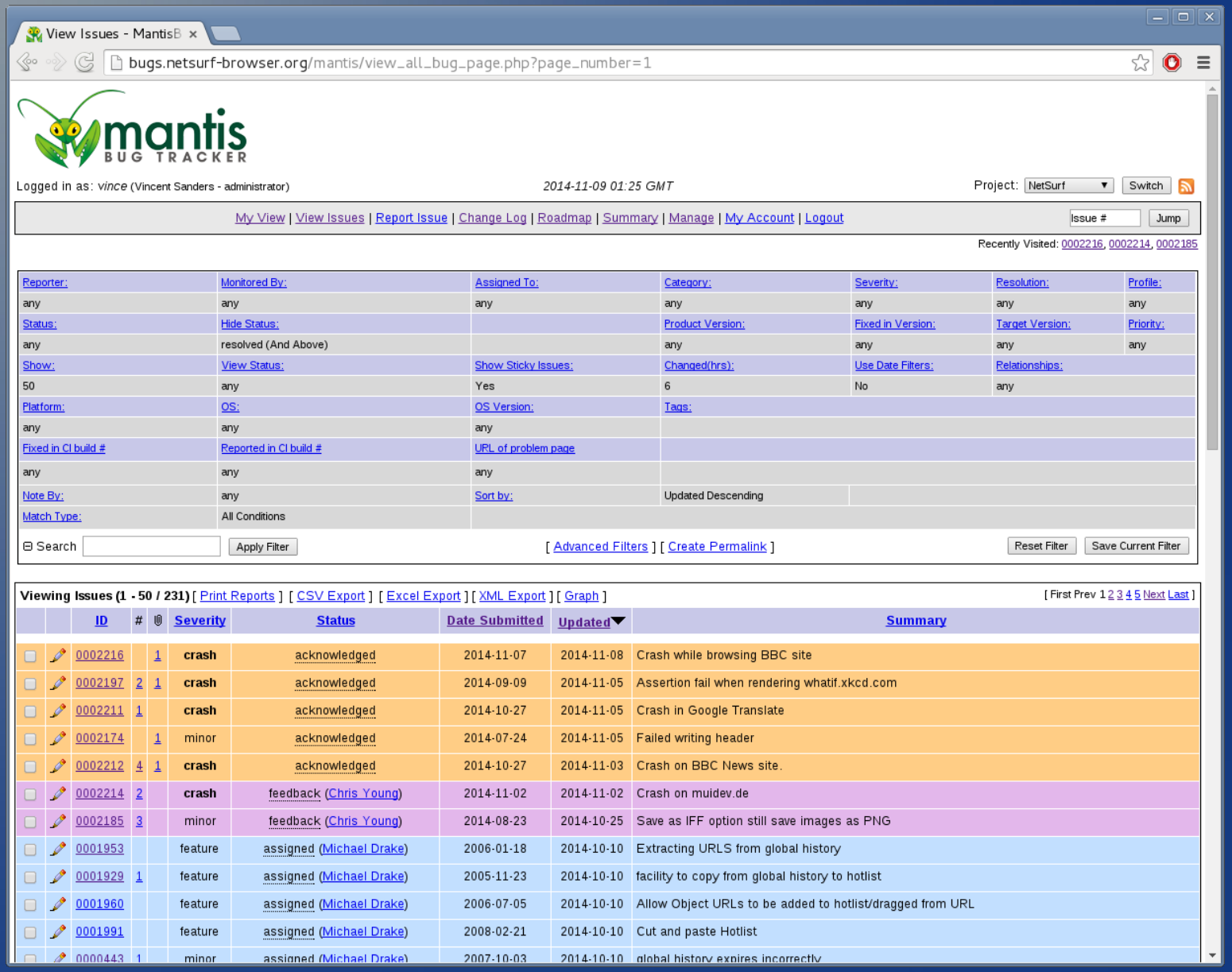
Static Analysis



Issue tracking

- All issue tracking systems are not ideal
- Go with the system that the fewest number of developers dislike
- Remember users have to report issues with it.
- The issue tracker needs a maintainer to be useful
- Double edged sword.

Mantis


View Issues - MantisB
bugs.netsurf-browser.org/mantis/view_all_bug_page.php?page_number=1
mantis BUG TRACKER
Logged in as: vince (Vincent Sanders - administrator)
2014-11-09 01:25 GMT
Project: NetSurf Switch
My View | View Issues | Report Issue | Change Log | Roadmap | Summary | Manage | My Account | Logout
Issue # Jump
Recently Visited: 0002216, 0002214, 0002185
Reporter: any | Monitored By: any | Assigned To: any | Category: any | Severity: any | Resolution: any | Profile: any
Status: any | Hide Status: resolved (And Above) | Product Version: any | Fixed in Version: any | Target Version: any | Priority: any
Show: 50 | View Status: any | Show Sticky Issues: Yes | Changed(hrs): 6 | Use Date Filters: No | Relationships: any
Platform: any | OS: any | OS Version: any | Tags: any
Fixed in CI build # any | Reported in CI build # any | URL of problem page any
Note By: any | Sort by: Updated Descending
Match Type: All Conditions
Search Apply Filter [Advanced Filters] [Create Permalink] Reset Filter Save Current Filter
Viewing Issues (1 - 50 / 231) [Print Reports] [CSV Export] [Excel Export] [XML Export] [Graph]
[First Prev 1 2 3 4 5 Next Last]
ID # Severity Status Date Submitted Updated Summary
0002216 1 crash acknowledged 2014-11-07 2014-11-08 Crash while browsing BBC site
0002197 2 1 crash acknowledged 2014-09-09 2014-11-05 Assertion fail when rendering whatif.xkcd.com
0002211 1 crash acknowledged 2014-10-27 2014-11-05 Crash in Google Translate
0002174 1 minor acknowledged 2014-07-24 2014-11-05 Failed writing header
0002212 4 1 crash acknowledged 2014-10-27 2014-11-03 Crash on BBC News site.
0002214 2 crash feedback (Chris Young) 2014-11-02 2014-11-02 Crash on muidev.de
0002185 3 minor feedback (Chris Young) 2014-08-23 2014-10-25 Save as IFF option still save images as PNG
0001953 feature assigned (Michael Drake) 2006-01-18 2014-10-10 Extracting URLS from global history
0001929 1 feature assigned (Michael Drake) 2005-11-23 2014-10-10 facility to copy from global history to hotlist
0001960 feature assigned (Michael Drake) 2006-07-05 2014-10-10 Allow Object URLs to be added to hotlist/dragged from URL
0001991 feature assigned (Michael Drake) 2008-02-21 2014-10-10 Cut and paste Hotlist
0000443 1 minor assigned (Michael Drake) 2007-10-03 2014-10-10 global history expires incorrectly

Releasing

- All components of a project come together
- Tested build possibly with known issues
- Unreleased software does not exist
- The easier they are to make the more you do

Practical Releasing

- Create CI jobs triggered from a git tag
- Use git sub modules to create a unified source
- Use the CI system to perform build from generated source in known build environment.

Wrapping up

- These are all the parts an open source creature needs to thrive
- Just because a project has these components does not mean it will survive
- The outcome should justify the effort

Any Questions?

Book image from GrrlScientist on flikr

What am I on about?

An examination of:

- What a services a project ought to have
- What options exist to fulfil those requirements
- A practical look at some implementations.

Application Lifecycle Management – horrid term but in common usage

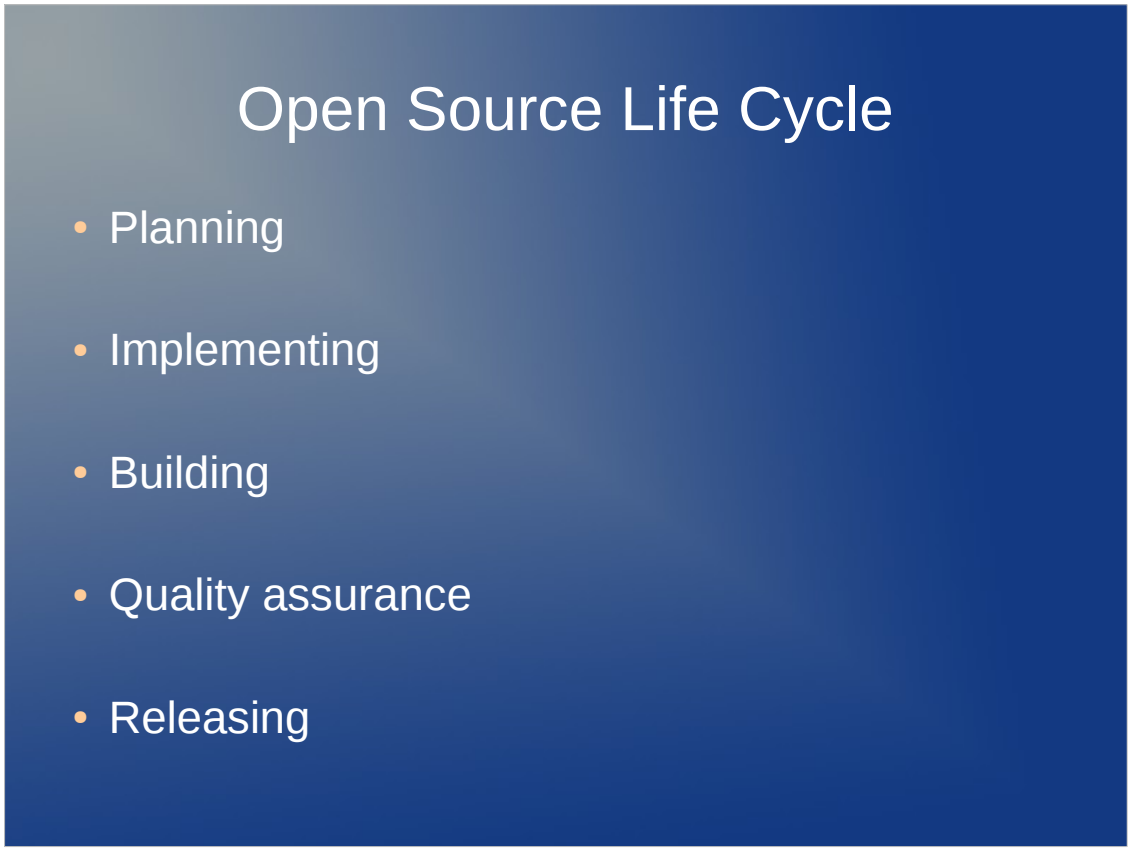
Like Team foundation Server but open source and not crappy

If a project has more than a couple of active developers these are the kind of things that make those people more productive.

The infrastructure should give benefits quickly but be robust enough to grow and adapt

Keep it simple stupid, do not waste more developer time with infrastructure than you gain. This stuff is meant to let you spend more time doing software

Historically sourceforge provided a lot of this, nowadays it is github.



What are the main areas in our lifecycle?

Usually a formal approach has planning and specifications. This is open source, generally we work by consensus and planning is informal at best but it is there. This is not an area technology really helps with and is more a social area. Having said that a wiki is a cheap and easy way to make sure developers ideas do not get lost.

The source code, this is where developers spend most of their effort and scratch their itches, we are generally pretty good at this but it needs managing so revision control systems are needed.

Building the code in all the configurations and environments the project supports can be hard. Continuous integration helps here

QA, Testing, everyone runs their tests all the time before they check in their code, right? Yeah that is what I thought. Having the tests run automatically means you know how healthy the project is

Releasing code is what it is all about. Without this users do not get your software and Debian packages cannot be made. Again CI helps but so does an issue tracker. Especially helpful if the issue tracker allows you to keep track of releases useful for release notes.

Planning

- Planning is usually a social activity
- Important to keep track of decisions
- Communication tools developers actually use
- Flexibility to achieve releases

Most developers do planning in an informal way.

None of these are useful unless developers actually use them, do not implement these unless there is consensus they will be used.

These communication channels are also often where potential new developers join in so its useful to have an easy way to provide answers to all the questions that get repeated a lot (especially gsoc)

No plan survives contact with the enemy, learn to be flexible and ensure your tools are too.

Planning Infrastructure

- IRC – creating channels on free networks like oftc or freenode is easy
- Email lists are less popular but easy way to communicate with lots of people
- Forums are easy to setup but can degenerate quickly
- Communication with users can occur here to get an idea of what they say they want.

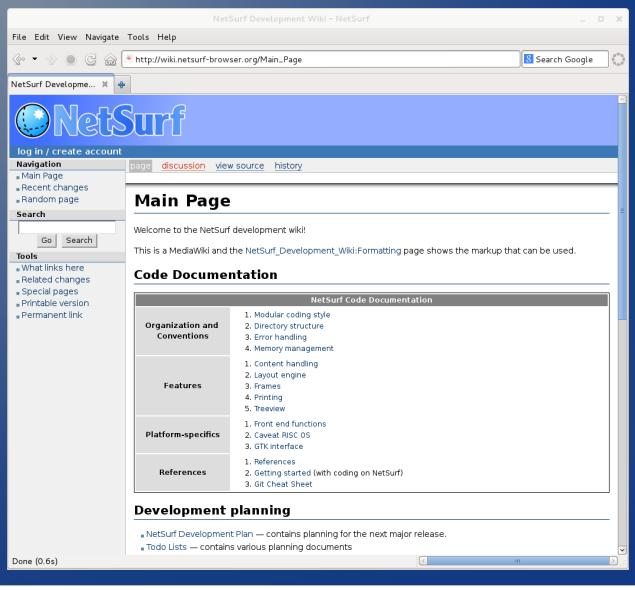
Practically IRC is invaluable for geographically dispersed groups and helps with short term coordination

Email lists or forums are ideal ways to keep in contact with others for with an more permanent record

Forums, especially user forums need ruling with a strong hand to stop them wandering off topic.

Planning Infrastructure

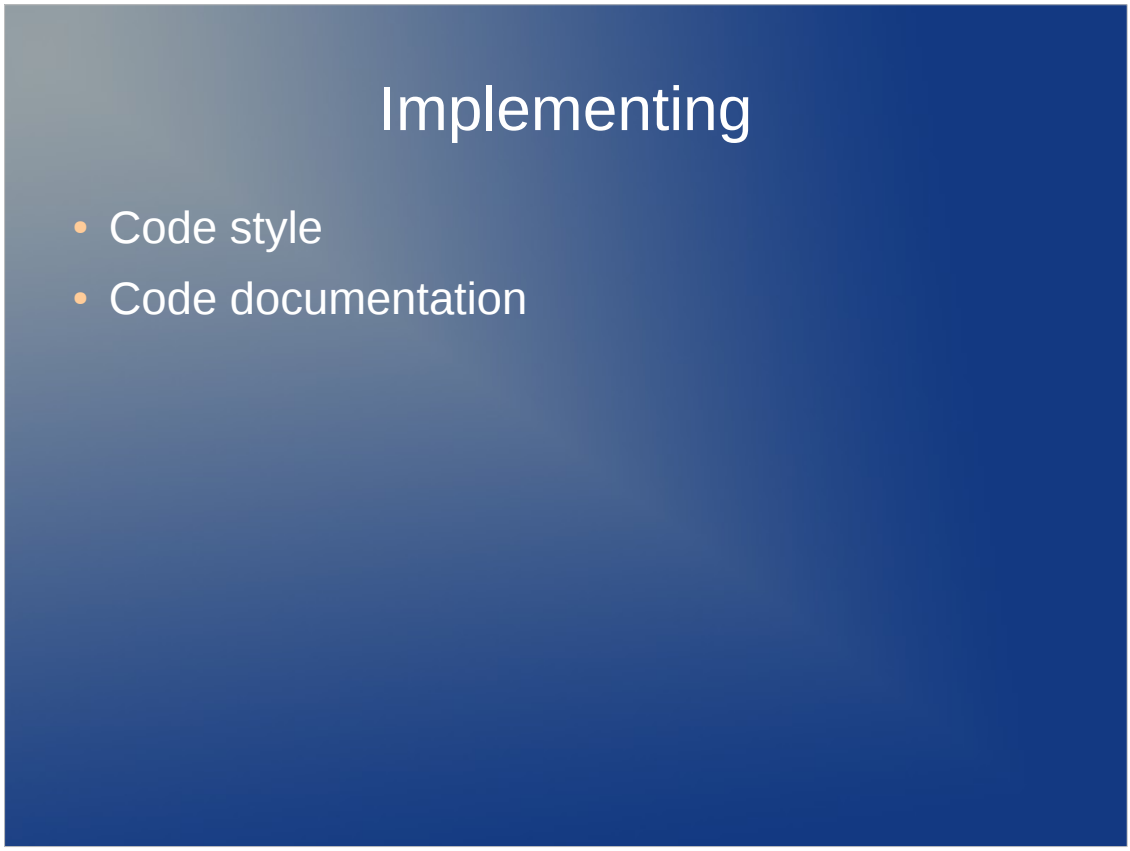
- A wiki is good for longer term info



Wiki is useful for info that is longer term in nature and the easy changeability means you can put info in quickly

Needs gardening if it not to become a spam infested waste of time.

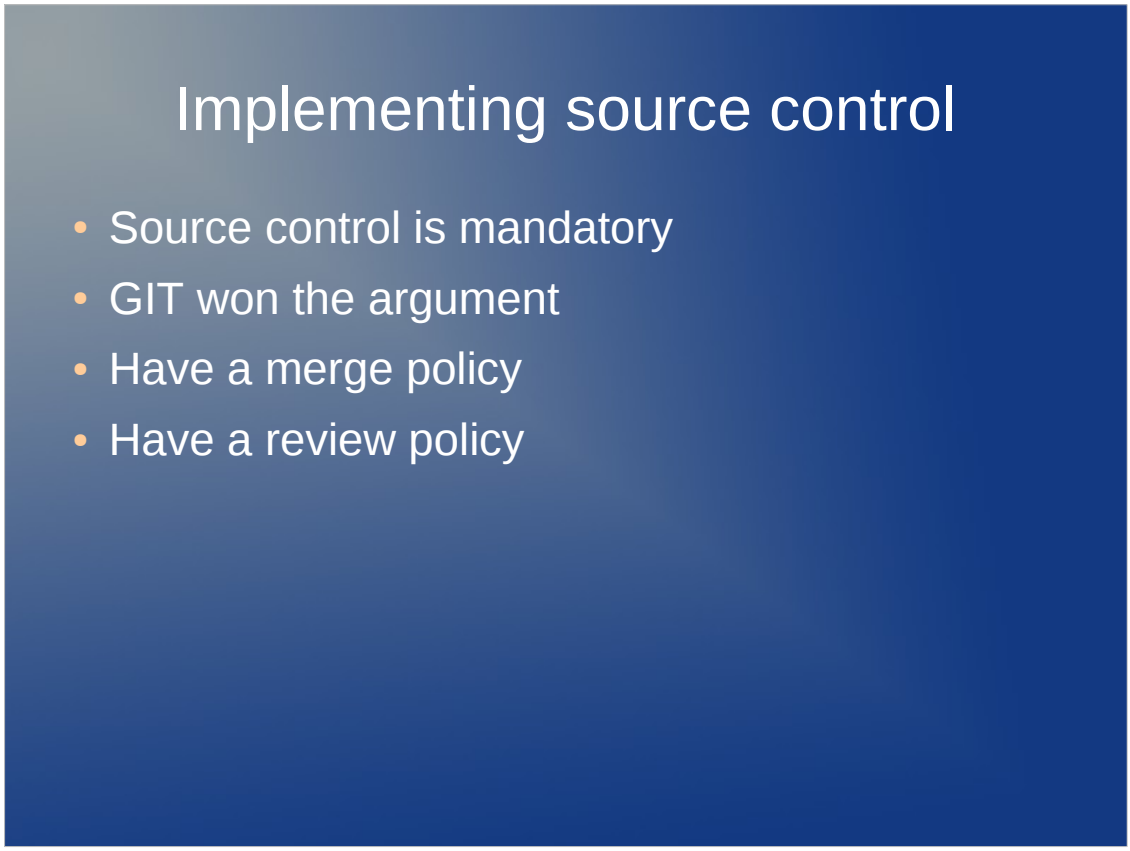
Debian has many to choose from pick one that suits.



This is the bit most developers actually want to do. Personally I love the intellectual rewards

A project should have at least a basic agreement on coding style to stop edit wars breaking out

Basic inline code documentation is useful but it needs maintaining to remain useful



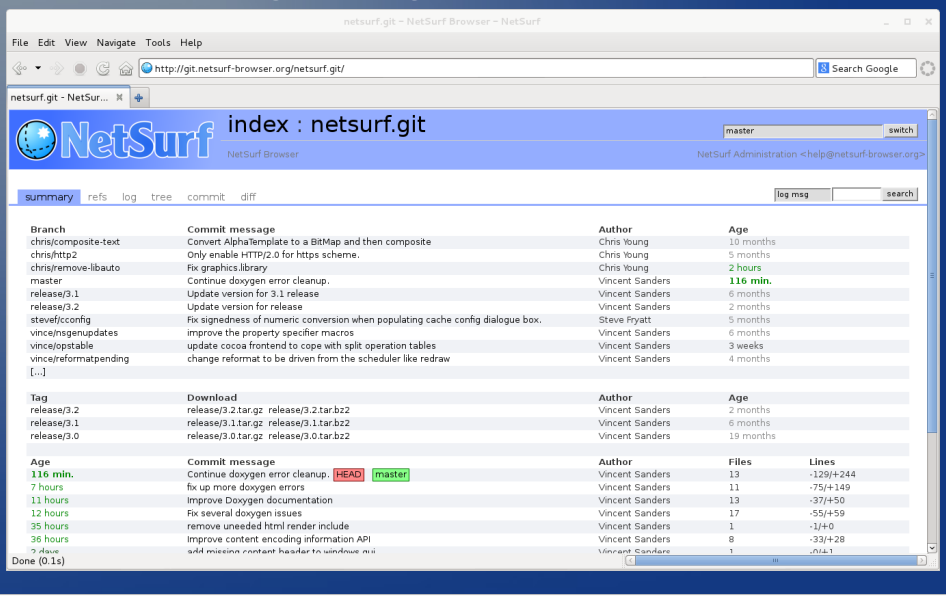
Any project that does not have a easily accessible revision control system is just plain broken.

I have used everything from SCCS through SVN, bzr, perforce and sourcesafe (you may mock me now) and some of those did some things better but GIT won the argument, even emacs is switching.

Make sure you have a sensible merge policy and if you are doing code review ensure the process is clear or it will be ignored.

Implementing source control

- Gitano and cgit are great



Gitano is an excellent git server, encourage Daniel to develop it and get it packaged

Cgit is a great tool especially with a forest of trees.

Failing that there are lots of options or githib is always there if you are willing to use their merge/review model. Heck you even get a basic website interface through github if you want.

Software is no good if developers cannot build it.

Continuous integration is a wonderful tool to ensure the software is always buildable

Especially important if your project has multiple components or architectures.

The more different ways a project can be built to more scope there is that a developer checking in their code will not have tested the alternatives

Building with Jenkins

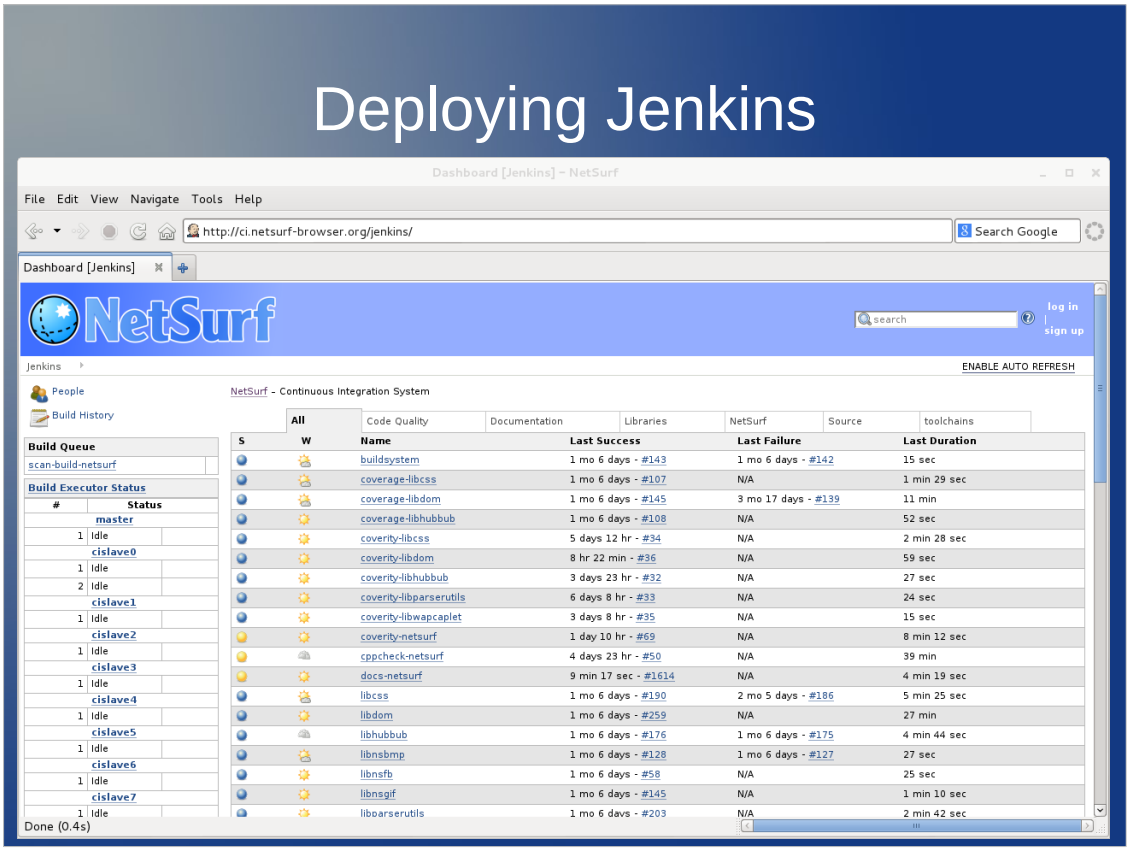
- Jenkins is a CI tool
- Jobs can be triggered by GIT changes
- Jobs can be periodic
- Dependences between modules
- Good mechanisms for feedback

If you are using github their infrastructure integrates travis CI for the rest of us jenkins is probably least bad.

The web based interface and large number of plugins make it easy to deploy.

Start with a small number of jobs and build up means large reward for small initial investment.

Deploying Jenkins



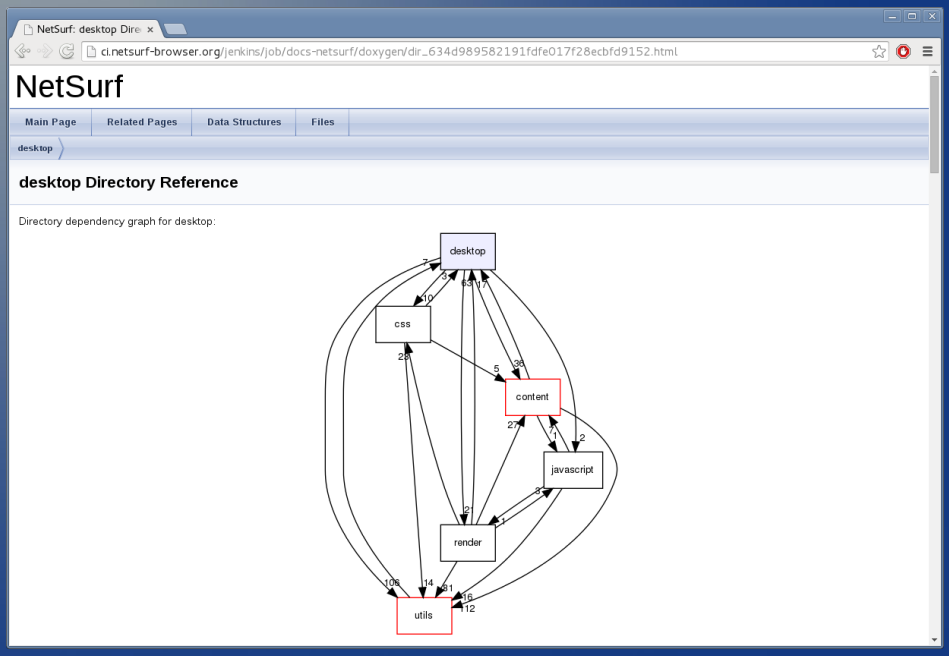
DEMO

I deployed jenkins for the netsurf project more than 18months ago

Started with small number of jobs and now now it build all the libraries and netsurf for 9 OS

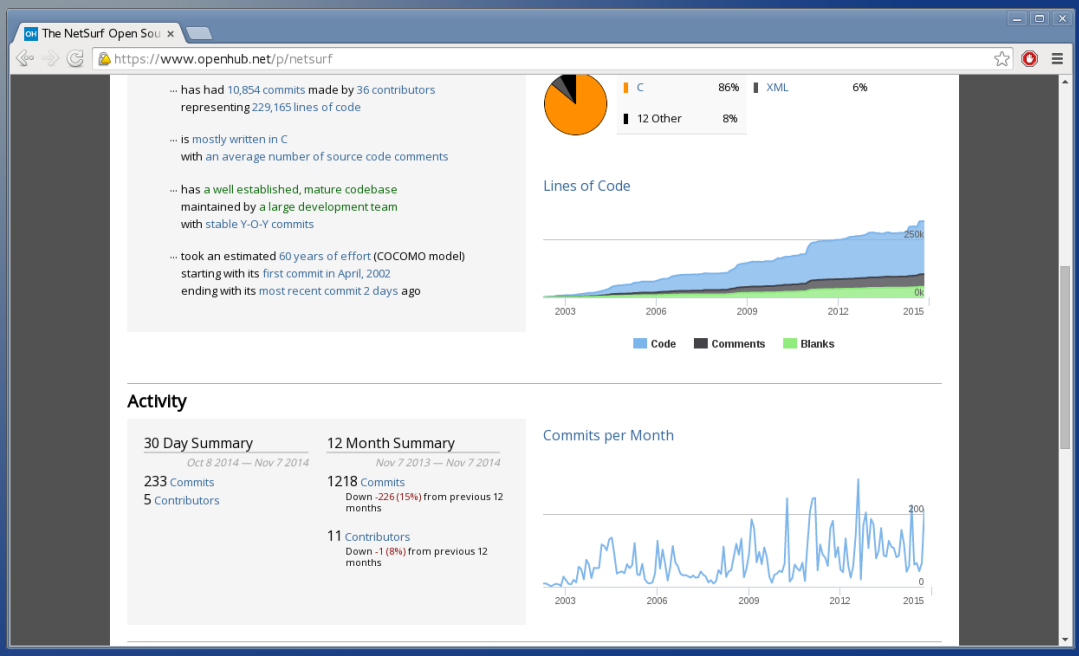
Netsuf job builds multiple configurations (with clang and gcc for nine tookits)

Deploying Jenkins

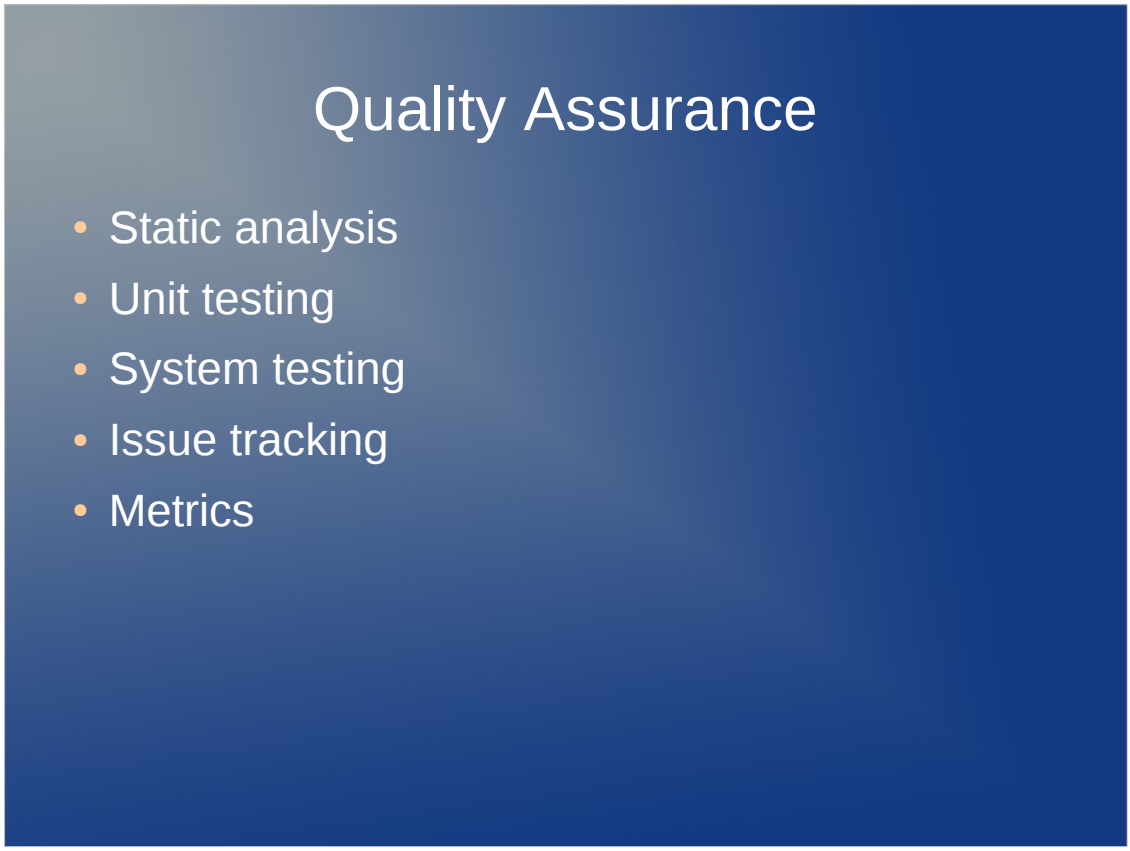


Jobs can be added to ensure things like the code documentation (via doxygen) is updated at the same time as code is built.

Metrics



Can add other jobs like metrics but sites like oholu now black duck open hub provide these and this kind of info is not hugely helpful



Static analysis is a powerful tool that can help find issues before your users do. It is only part of the story but works best when automated and a developer does not have to do anything. No new regressions is helpful target with this.

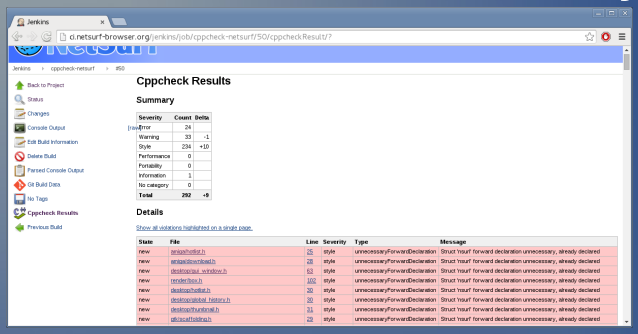
A project should have at least some basic unit testing although this gets missed a lot. Again automation of running the tests is best. Gamification helps

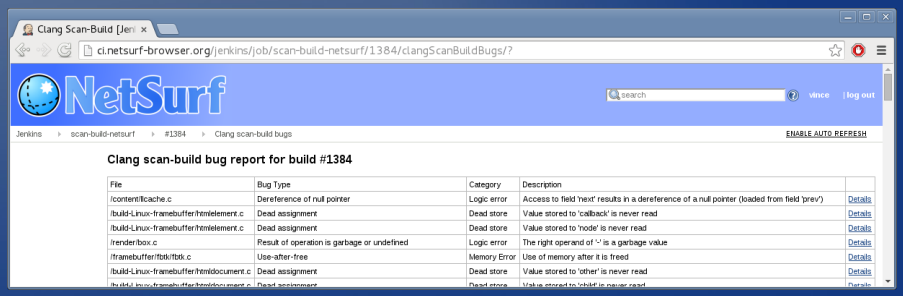
System testing is hard but useful if it can be implemented.

An issue tracker is a useful tool both for keeping track of QA issues and for bugs found by users in releases.

Metrics are pretty but do not serve much beyond that

Static analysis

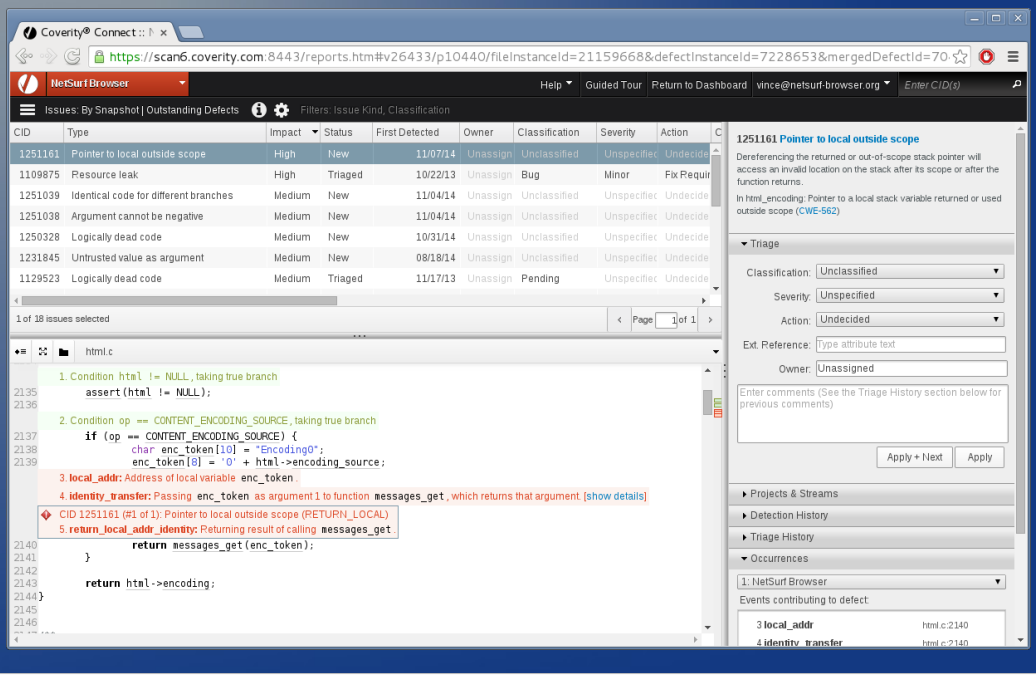




Static analysis is powerful for netsurf we have CI jobs that run scan-build (clang) , cppcheck and the proprietary but free (beer) coverity

The free tools are ok but have a lot of false positives and are difficult to manage.

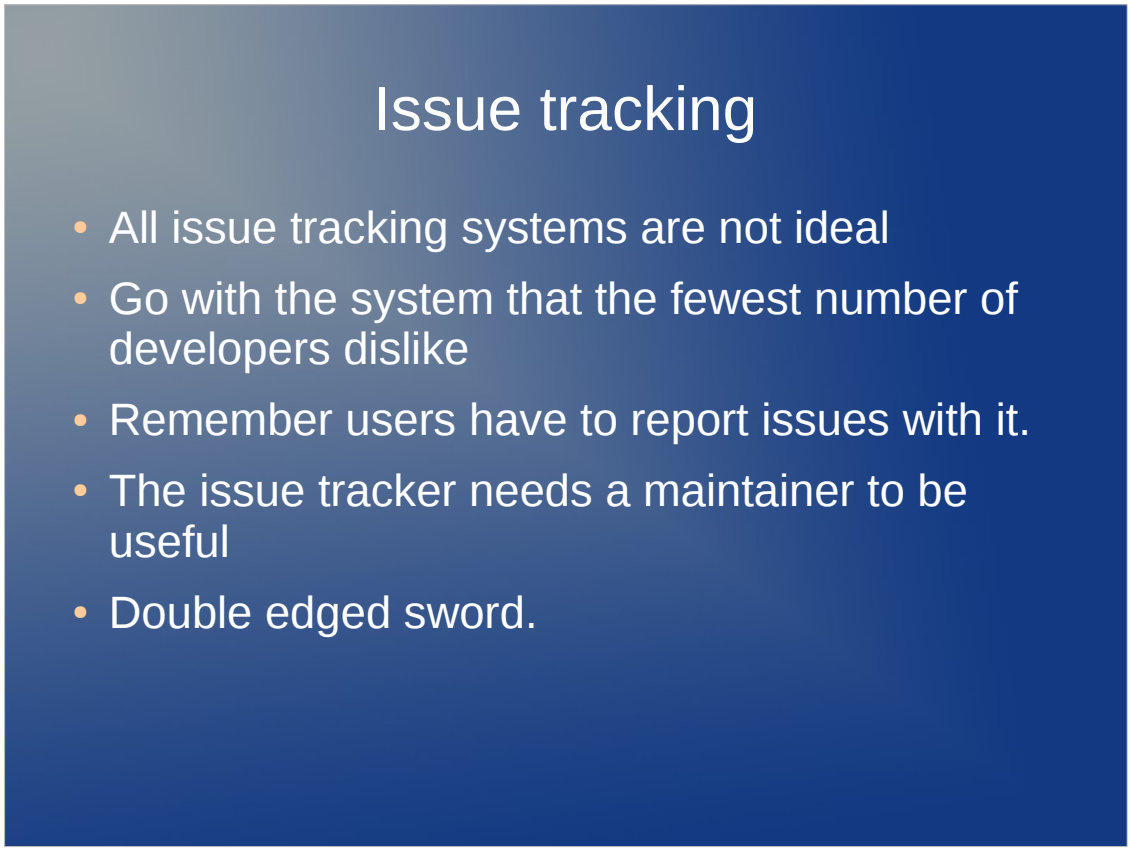
Static Analysis



Coverity is much easier to use and worthwhile looking at fro any open source project as you can use their free scan service.

I wrote a lot about this in a blog post

Error analysis is the sweet spot for improvement



Issue trackers seem to be like mail clients, they all suck, some less than others

If your project went with github they have an integrated solution, if not practical options are basically bugzilla, trac or mantis

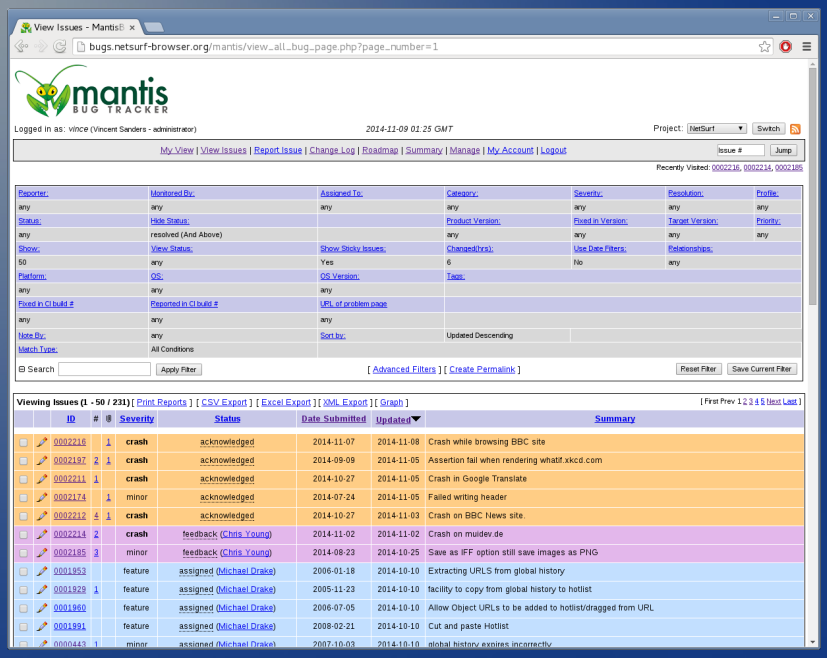
Bugzilla is ubiquitous and hard to admin.

Trac gives you a wiki as well as issue tracker and source viewer but forces a workflow

Mantis is very simple, is usable without javascript/html5 features making it fast to use. The simplicity could be limiting if your project needs more but remember the maintenance burden.

Users need training in whatever tool, its a big investment.

Mantis



Netsurf deployed mantis

Importing the old data from sourceforge took a lot of time and we **never** look at the historical data

I spend an hour a week minimum just ensuring new bugs are acknowledged and basic triage.

We benefit from it by users reporting issues with CI builds quickly

We have a large backlog of unreproducible crash bugs on minority platforms which appear to be caused by external factors. Users feel slighted if you just close them.

Releasing

- All components of a project come together
- Tested build possibly with known issues
- Unreleased software does not exist
- The easier they are to make the more you do

A release is an opportunity for the developers to get all the moving parts of their project in a state it can be used by non-developers

Allows more in depth QA and gives confidence to users they can update.

As far as users are concerned if it is unreleased the software does not exist.

The easier the release process is for developers the less anxiety over a release there is.

One strategy is to use the CI system to publish builds all the time so the difference between a release build and a CI build is negligible aka continuous deployment

Practical Releasing

- Create CI jobs triggered from a git tag
- Use git sub modules to create a unified source
- Use the CI system to perform build from generated source in known build environment.

Release process for sub modules is as simple as

```
git tag -s -m 'Official Release'
release/<version number>

git push --tags
```

Then updating the sub modules in the netsurf-all repo and pushing a signed tag to that

CI will generate the source tarballs and build them without human involvement. Release can be mechanically generated in under 30 minutes. Most time is spent checking we are releasing what we intend to.

Process in wiki

Wrapping up

- These are all the parts an open source creature needs to thrive
- Just because a project has these components does not mean it will survive
- The outcome should justify the effort

The health of open source projects, like any creature, are not solely dependant on the care given to them. If there is no demand for the software then a project will die as developers move away and do something more interesting but without care a project will definitely fail

Developers should always consider the overhead of implementing things and ensure they will get a worthwhile return

I once killed a Tamagotchi in 90 minutes, maybe there are better ways.

Any Questions?