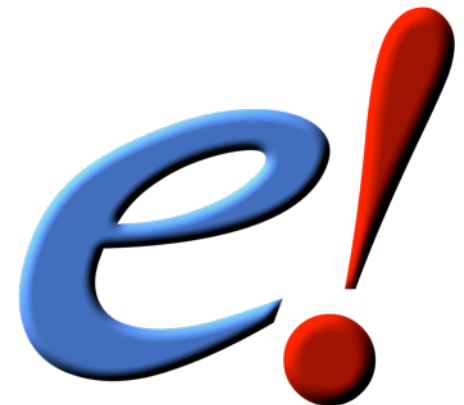


Automated Ensembl Software Environment

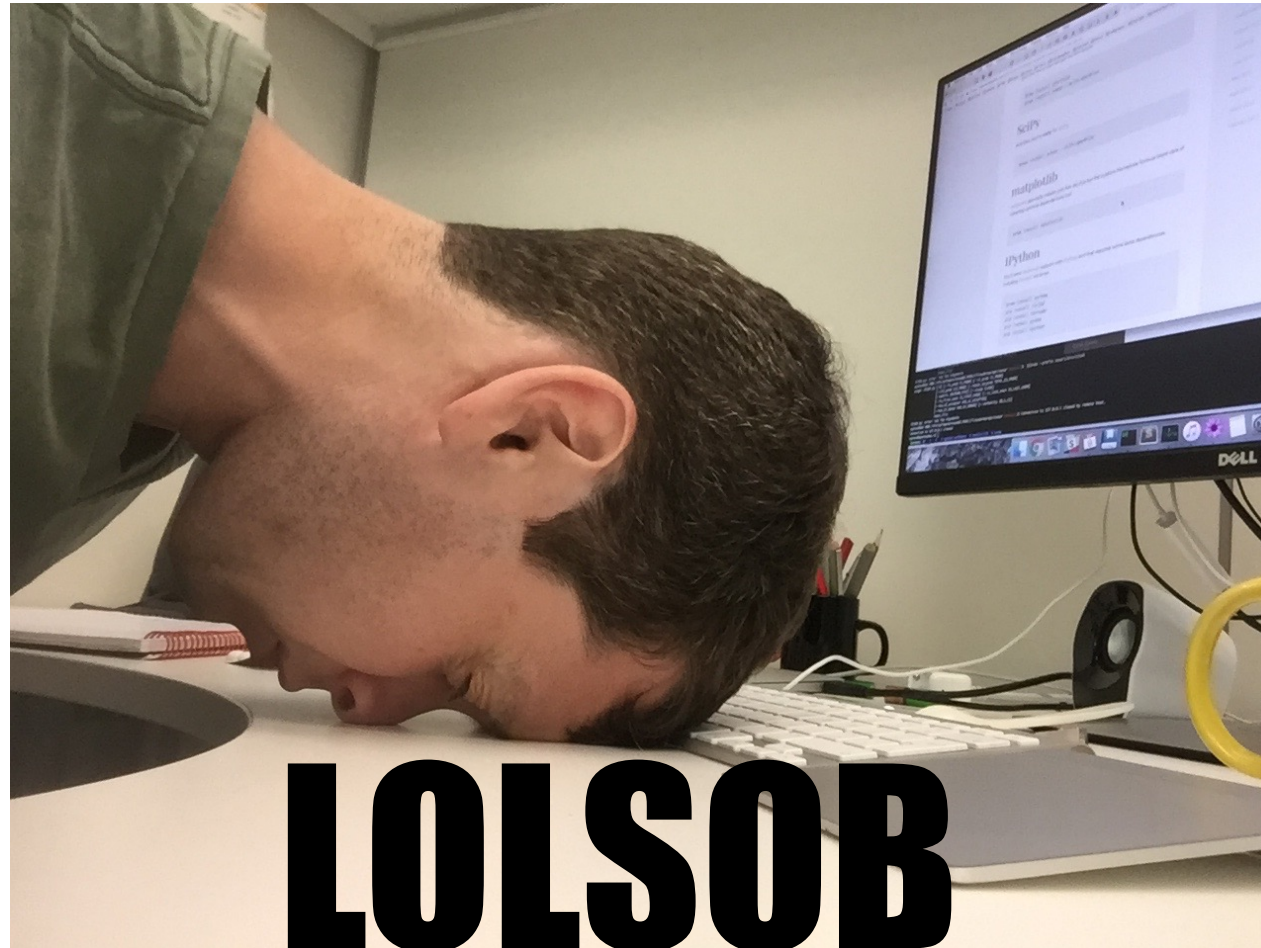
Linuxbrew, Ansible and Shell Scripts Oh
My!

Andy Yates - GTI

Ensembl Data Review 12th Sept 2016



Or How to Avoid Repeated



Skip to the End

```
$ ssh ebi-006
```

```
$ source /nfs/software/ensembl/RHEL7/envs/basic.sh
```

Gives you Perl 5.14.2, bwa, 2 versions of Exonerate, blast, kent libs, htlib, and more!

<https://www.ebi.ac.uk/seqdb/confluence/display/ENS/EBI+RHEL7+Farm+Setup>

The Idea of Automated Deployment

- We collected a list of everyone's dependencies (over 50 programs)
- Alessandro worked tirelessly to build a software environment at EBI
 1. Download program X
 2. make && make install
 3. Fix it (edit code or new dependency)
 4. make && make install
 5. See step 1



How Can Code fail to compile?

1. GCC decided to remove standard imports in v4.3
 - Error function “atoi” not found in this scope
 - `#include <cstring.h>`
2. Some software was written for older dialects of C
 - `gcc -std=c89` for bwa
3. Programs do not declare their dependencies
 - Why yes I do need libpng ... thought I said
4. It's Oracle. Oh perl `Makefile.PL` `-l` is the right way to compile this is it? I didn't know

All of that leads to lots of ...



RedHat 7 .. We've got to do it again!



Project Aims

- One command to deploy, update and redeploy our software environment
- Code as Configuration and better dependency management
- Keep the record of dependencies up to date

Death to copy and paste commands from confluence

Installing on the Sanger farm

To be run on farm3, from the ensembl-rest repository.
This presumes most libraries are already available.

These commands were run in a shell

```
wget -O- http://cpanmin.us > stdout
perl stdout -I ~/perl5 App::cpanminus local::lib
echo `perl -I ~/perl5/lib/perl5 -Mlocal::lib`

setenv PERL_LOCAL_LIB_ROOT "$PERL_LOCAL_LIB_ROOT:$HOME/perl5" setenv PERL_MB_OPT "--install_base $HOME/perl5" setenv F

setenv PERL5LIB "$HOME/perl5/lib/perl5:$PERL5LIB"
setenv PATH "$HOME/perl5/bin:$PATH"

cpanm Catalyst
cpanm Catalyst::Devel
cpanm --installdeps .
perl Makefile.PL
```

Installing on a Ubuntu server

The following lists all libraries that have been required in various installations.

Depending on the version of your distribution, some libraries might already be available.

Ubuntu libraries

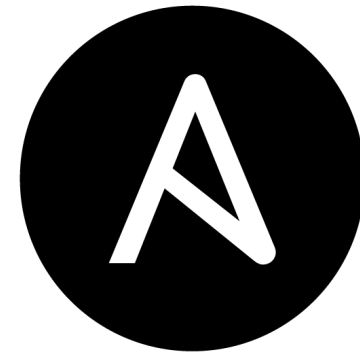
```
sudo apt-get clean
sudo apt-get update
sudo apt-get install -y make
sudo apt-get install -y git
sudo apt-get install -y build-essential
sudo apt-get install -y libxml2 libxml2-dev zlib1g zlib1g-dev libexpat1 libexpat1-dev
sudo apt-get install -y mysql-client mysql-server libmysqlclient-dev
sudo apt-get install -y libcatalyst-action-rest-perl libcatalyst-devel-perl libcatalyst-modules-perl libcatalyst-perl
sudo apt-get install -y libchi-driver-memcached-perl libplack-perl libsafe-isa-perl libmoosex-daemonize-perl
sudo apt-get install -y libclass-dbi-sweet-perl libio-all-lwp-perl libfile-path-tiny-perl libfile-remove-perl
sudo apt-get install -y libfile-sharedir-install-perl libtest-checkdeps-perl libtest-deep-perl libtest-fatal-perl
sudo apt-get install -y libtest-json-perl libtest-trap-perl libtest-xml-simple-perl libtest-xpath-perl libpar-dist-perl
```

The New Tools



- **Linuxbrew**
- Non-root dependency management
- Download a file, get its dependencies, patch/edit, run make, install
- Customisable
- **homebrew/science** lots of **Bioinformatics apps/libs**

- Fancy shell scripts in YAML/Python
- Server management software
- Make servers/directories into what you expect
- Reusable components/plays
- Our environment bootstrapper



A N S I B L E

Part 1: Linuxbrew

- **Formula:** a Ruby file for creating a dependency
- **Bottles:** Precompiled software and can be “poured” into a directory
- **Keg:** A versioned install of a program
- **Cellar:** Where the kegs are held
- **Tap:** a repository of formulas normally held in GitHub
 - <http://github.com/Ensembl/homebrew-ensembl>
 - <http://github.com/Ensembl/homebrew-icc>
 - <http://github.com/Ensembl/homebrew-moonshine>
- **Moonshine:** A formula for closed source software (my own term)

Biobambam – In A Linuxbrew Formula

```
class Biobambam < Formula
  desc 'Tools for BAM processing'
  homepage 'https://github.com/gt1/biobambam'
  url 'https://github.com/gt1/biobambam/archive/0.0.191-
release-20150401083643.tar.gz'
  sha256 '1f5be1c5b5c703c0b03969af7d8fa85da280044f131757c76b66e9021b67ae18'
  version '0.0.191'

  depends_on 'ensembl/ensembl/libmaus'

  def install
    libmaus = Formula["ensembl/ensembl/libmaus"].opt_prefix
    system './configure', "--prefix=#{prefix}", "--with-libmaus=#{libmaus}"
    system 'make'
    system 'make install'
  end
end
```

State of our Linuxbrew

- We have our own tap ensembl/ensembl
 - Versioned and custom builds of 72 programs
 - All verified as “working”
 - Some data file formulas as well (Panther for Interproscan)
- Linuxbrew for icc and iccmpi compiled software because icc plus Linuxbrew/gcc hates me
 - Uses formulas from ensembl/ensembl-icc
- We have ensembl/moonshine
 - For closed source license programs like signalp and genscan

ensembl/homebrew-cask

A way of creating meta-formulas for listing broad sets of dependencies like our perl libs or what's needed for assembly mapping

```
#Template generated on 2016-08-19 14:27:03 +0100
#Do not edit this file. Modify the relevant config and regenerate
require 'digest'
class PerlClibs < Formula
  url 'file://'+File.expand_path(__FILE__)
  desc 'Dependencies for the PerlClibs formula'
  sha256 Digest::SHA256.file(File.expand_path(__FILE__)).hexdigest
  version '2'

  depends_on 'mysql-connector-c'
  depends_on 'ensembl/ensembl/kent'
  depends_on 'ensembl/ensembl/htslib'
  depends_on 'ensembl/ensembl/postgres-client'
  # SNIP but there's more here
end
```


That's generated from

```
"perl-clibs":  
  version: 2  
  deps:  
    - "mysql-connector-c"  
    - "ensembl/ensembl/kent"  
    - "ensembl/ensembl/htslib"  
    - "ensembl/ensembl/postgres-client"
```

Custom code available from [ensembl/homebrew-cask](https://github.com/ensembl/homebrew-cask)

Part 2: The *envs

- A suite of libraries for managing programming language versions and installs
- pyenv, plenv, rbenv, jenv
 - All work by creating “shims” which route to an executable
 - Versions can be set globally, in the shell (via ENV variable) or in a local directory
 - **plenv local 5.14.4**
- plenv is a better perlbrew as testified by Prem
- pyenv also has support for virtualenvs



Part 3: Perl Libs

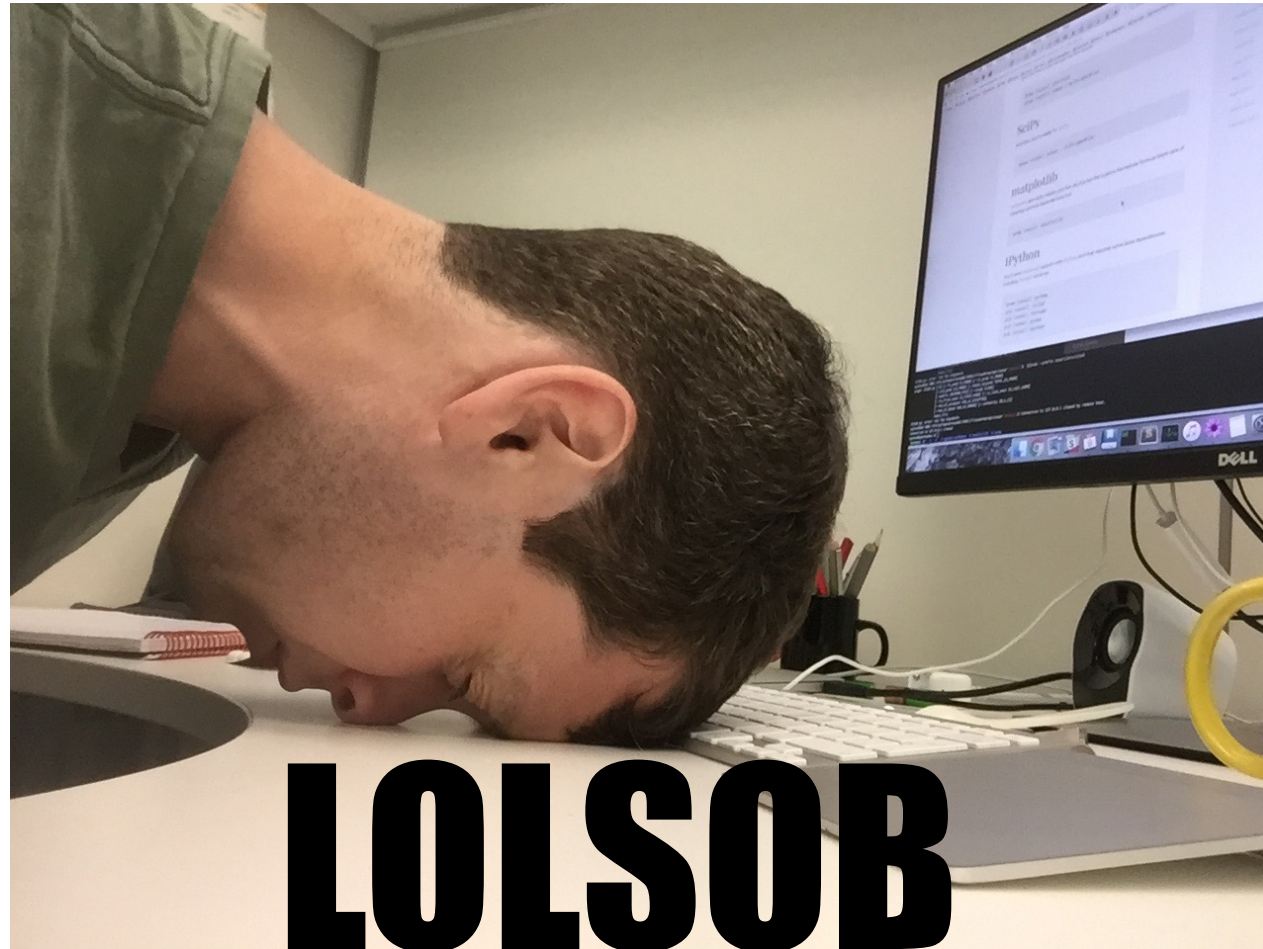
cpanfile for core

```
requires 'DBI';  
requires 'DBD::mysql';  
requires 'HTTP::Tiny';  
requires 'IO::Compress::Gzip';  
requires 'URI::Escape';
```

Most Ensembl repositories have files of their dependencies called cpanfiles. However most do not use them for dependency management

We can automatically download, give to cpanminus and install all dependencies

Apart from Oracle::DBD



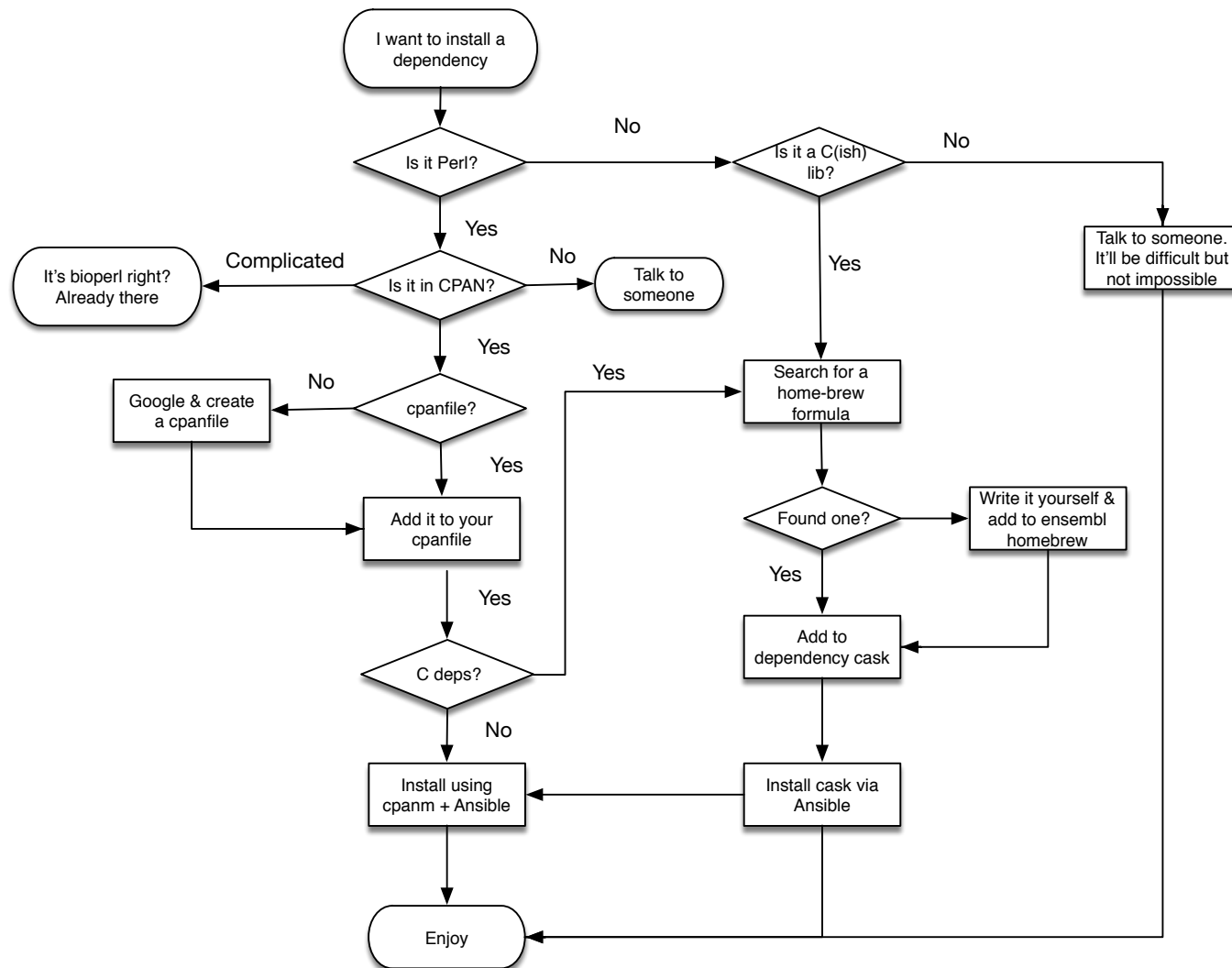
Part 4: Putting It All Together



- Ansible runs tasks in a dependable order
- Tasks can be reused and composed into plays
- One play that can
 - Create the /nfs/software/ensembl/RHELVERSION
 - Install all required Linuxbrews and *envs
 - Run the default Ensembl tasks
 - Install all publically available cpanfiles
 - Create bash environment files

```
ansible-playbook -i localhost, ensembl-software-  
environment/roles/ebi-farm.yml --extra-vars  
"software_home_base_dir=$PWD"
```

What If You Want Software Installed?



Talk to
someone
with ensw
access

The End

```
$ ssh ebi-006
```

```
$ source /nfs/software/ensembl/RHEL7/envs/basic.sh
```

<https://www.ebi.ac.uk/seqdb/confluence/display/ENS/EBI+RHEL7+Farm+Setup>

The Entire Ensembl and Ensembl Genomes Team

Andrew Yates¹, Wasiu Akanni¹, M. Ridwan Amode¹, Daniel Barrell^{1,2}, Konstantinos Billis¹, Denise Carvalho-Silva¹, Carla Cummins¹, Peter Clapham², Stephen Fitzgerald¹, Laurent Gil¹, Carlos García Girón¹, Leo Gordon¹, Thibaut Hourlier¹, Sarah E. Hunt¹, Sophie H. Janacek¹, Nathan Johnson¹, Thomas Juettemann¹, Stephen Keenan¹, Ilias Lavidas¹, Fergal J. Martin¹, Thomas Maurel¹, William McLaren¹, Daniel N. Murphy¹, Rishi Nag¹, Michael Nuhn¹, Anne Parker¹, Mateus Patricio¹, Miguel Pignatelli¹, Matthew Rahtz², Harpreet Singh Riat¹, Daniel Sheppard¹, Kieron Taylor¹, Anja Thormann¹, **Alessandro Vullo¹**, Steven P. Wilder¹, Amonida Zadissa¹, Ewan Birney¹, Jennifer Harrow², Matthieu Muffato¹, Emily Perry¹, Magali Ruffier¹, Giulietta Spudich¹, Stephen J. Trevanion¹, Fiona Cunningham¹, Bronwen L. Aken¹, Daniel R. Zerbino¹ and Paul Flicek^{1,2,*}

¹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK and ²Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

Funding

