Ensembl Release Coordination

This document describes the tasks that the designated release co-ordinator must carry out over the course of an Ensembl release.

Note that the public release is generally at the beginning of the month, so most of the work for "the March release" is done in February.

Deadlines

The deadlines are counted backwards from the release date so that every phase has a consistent amount of time each month. The "declaration of intentions" is the exception and is always on the 1st of the month.

> *Declaration of intentions*
> 1st - declaration of intentions
>
> *Pre-core handover*
> 3 days before genebuild handover (pseudo-deadline)
> By this time core databases should be ready for stable ID mapping and peptide dumps.
>
> *Genebuild handover to Core/Compara*
> 10 Working days before Core/Compara handover
> Core-like databases, vega should be ready to ecs2d; "ready" meaning: gene build complete stable ID mapping done, protein annotation and mapping done.
>
> *Core/Compara/Variation handover to Mart/Web team*
> 5 Working days before Mart handover
>
> *Mart handover to Web team*
> 2 Working days before release
>
> *Release*
> 1st of the next month or the following Monday if the 1st is on a Friday, Saturday, or Sunday (web team do not release on Fridays).

Note that there is a script at `ensembl-doc/calc_release_deadlines.pl` which will generate the dates for you, given the release date (`--release` argument). It should take account of weekends and Bank Holidays.

Database naming convention

The schema version number changes every month to assist internal version tracking, regardless of whether there are any schema/API changes.
The assembly version number changes if there is a new assembly (or has a letter appended if there is a patch or some other change).
Note that the assembly letter suffix must be added/incremented if there is a change in *any* of the data for a species – i.e. any of core, variation etc.

Database servers

There are several servers used in the release process:

- ecs2:3364 – the staging server, where all the databases are kept and worked on during the release
- ecs2 :3365 – holds a copy of the databases from the currently live release, for scripting against

- ecs3:3307 – holds the actual live databases; do *not* modify these or do anything that will impact performance.

Main steps in release coordination

1. On the 1st of the month previous to release send an email with subject like "December release: declaration of intention", asking people to declare any new databases and expected schema changes. Give people a couple of days to declare, then send a reminder with a summary of what has been declared. Also announce the new version number if it's changed. Also send the summary to the ensembl-dev mailing list (ensembl-dev@ebi.ac.uk) with an appropriate reminder that it's not a guarantee of what will happen.

2. Find out who the release co-ordinators for Compara and Mart (and variation?) are this month.

3. Check how old the `ensembl_go_*` database is, and if it is older than 6 months ask Tony Cox (`avc@sanger.ac.uk`) to generate an updated version. Check how old the `homo_sapiens_disease_*` database is, ask Arek (arek@ebi.ac.uk) if it needs an update in sync with the latest OMIM release.

4. Copy the latest released database versions from ecs3:3307 to ecs2:3365.
   This means you will need to delete the old previous database versions to free disk space.
   Warn people on the ensembl-admin email list about the impending deletion and new copy.
   The script to use is `ensembl/scripts/CopyDBoverServer.pl` scripts, `-h` gives usage information.
   Note:
   - You have to run it as the `mysqlens` user.
   - Must be run on the destination host.
   - It is best to run it on the node that the destination MySQL instance is running on. To do this type
     `caa_stat mysql_3365` (for 3365) or `caa_stat` to list them all

5. Based on the synthesis of what new databases are coming, you can decide which databases will not change, and copy their latest version from ecs3:3307 to ecs2:3364, probably renaming them in the process.
   Warn people on ensembl-admin about the impending deletion, new copy and database renaming.

   Note:
   - use the `drop database` command in a MySQL client, never do a `rm -rf` as `mysqlens` user).
   - Do not delete (without at least checking):
     - databases like `ensembl_%`
     - databases like `%_expression_%` (needed for Mart)
     - `gnf_evoc` (needed for Mart)
   - Check with Compara and Mart people to see if `ensembl_compara_*` and `ensembl_mart_*` can be dropped.

6. Keep in close contact with the persons involved in the production of new data, i.e.

   - Yuan/Daniel for SNP/haplotype/variation,
   - Laura/Felix for worm databases,
   - Val/Steve for mammalian databases,
   - Martin/Vivek/Jan for insect databases,
   - Kerstin/Mario for Zebrafish/Xenopus,
   - Steve Trevanion/Patrick for vega,
   - Cara/Jessica/Abel for compara

   Email them from time to time, to see how things are going.

   Note that, currently, if there are new SNPs, the following needs to be done *in this order*:

- rebuild SNP database
- generate lite databases
- recalculate SNP density features

7. For new genebuilds, various density features and seq_region attributes need to be calculated. To do this the following scripts should be run, all of which are in `ensembl/misc-scripts/density_feature/` and produce several types of density_feature:

   - `gene_density_calc.pl`   (takes about 15 minutes)
   - `percent_gc_calc.pl`   (takes about 12 *hours*)
   - `repeat_coverage_calc.pl`   (takes about 12 *hours*)

   The above three scripts can be run at any time after the database has been received from the genebuilders.

   There are 2 more scripts that need to be run *after* the SNP *and* lite databases have been built:

   `snp_lite_density.pl`   (takes about 35 minutes, creates SNP density features)
   `seq_region_stats.pl`   (takes about 35 minutes, creates `seq_region_attrib`s)

7. On the "genebuild handover" date all databases except Compara are expected to be ready and should be copied over to ecs2:3364.

8. Set up a test web test server on an ecs2 node. Details on how to do this are given in the "Setting up a web server" section at the end of this document. When done, send a mail to ensembl-admin to ask everyone concerned to check the display of "their"data on the web.

9. Ask the release core assistant to run the health check suite and to update schema if needed. Advertise any inconsistency in the databases, and fix them or arrange for someone else to fix them.

10. Run a healthcheck again just before handover (checking things we had done before plus new things that have been added during the data shakedown) - if possible of course.

11. By the Core/Compara handover date if everything is OK, send an email to hand-over the databases, listing all databases, specifying what new data is coming in and any schema changes. Also mention when is the deadline for Mart database handover.

12. If there have been any schema changes, send the appropriate patch SQL file to the web team so that they can publish it. By convention these are stored in `ensembl/sql` and named e.g. `patch_26_27.sql`

13. Even if by this date, the main part of your work is over, you will still need to keep track of email about database problems, and decide with the web team on how to fix them.

14. By the day of release, organise a post-mortem meeting to discuss problems that arose during the release, propose/ask for possible solutions/improvements for the next release.
    The main people to be there should be Jim, James, Arne, Arek, Abel, Val, Steve, Glenn, Ewan.
    Once you've decided on a mutually convenient time and date, book a room for the post-mortem.
    Send out an agenda beforehand and ask people if there's anything else that needs to be discussed.
    If the release has slipped and the post-mortem ends up being on or before the release day, it's better to re-arrange it for a few days later.

15. Send an email minute to ensembl-admin to summarize what was discussed/decided during the post-mortem.

Setting up a web server

Note that all these steps should be done as the `ensembl` Unix user.

The default port for the web server is 5000, but you can choose whatever you like as long as you announce it on ensembl-admin.

Websites live in `/nfs/acari/websites`, and it's customary to create a directory for each release, e.g. for release 26 use `/nfs/acari/websites/ensembl_26`

Check out the various modules from CVS:

```
su ensembl
mkdir /nfs/acari/websites/ensembl_26
cd /nfs/acari/websites/ensembl_26
cvs -d :ext:ensembl@cvs.sanger.ac.uk:/nfs/ensembl/cvsroot co sanger-web ensembl-api
```

Note that `ensembl-api` and `sanger-web` are CVS aliases which each check out several modules.

You'll need BioPerl; often you can just link to the one in the previous ensembl release webserver directory, but if you need the latest version do the following (full instructions at `http://www.bioperl.org/Core/Latest/index.shtml`):

```
setenv http_proxy "http://wwwcache.sanger.ac.uk:3128"   (to make wget work)
wget http://bioperl.org/DIST/bioperl-1.2.3.tar.gz
gtar xvzf bioperl-1.2.3.tar.gz
rm -f bioperl-1.2.3.tar.gz
ln -s bioperl-1.2.3 bioperl-live
```

Create a directory for the webserver logs:

```
mkdir logs
```

Then create a symbolic link to the `httpd` executable in a `src` directory

```
mkdir src
ln -s /usr/local/ensembl/apache/bin/httpd src/httpd
```

Modify some configuration files (all kept in the `conf/` directory)

```
cd conf
```

Modify each file in turn:

MULTI.ini - set the names of the multi-species databases. Comment out Compara and/or Mart if they're not available yet.

DEFAULTS.ini - update the following variables if not already done:

```
ENSEMBL_HOST            = ecs2
ENSEMBL_HOST_PORT       = 3364
ENSEMBL_WRITE_USER      = ensadmin
ENSEMBL_WRITE_PASS      = xxxx
```

SiteDefs.pm - update the following variables if not already done:

```
$ENSEMBL_PORT           = '5000';
$ENSEMBL_USER           = 'ensembl';
$ENSEMBL_ERRORS_TO      = 'glenn@ebi.ac.uk';   put your email address here!
```

`Homo_sapiens.ini` - update the following variables if not already done:

```
ENSEMBL_GOLDEN_PATH      = NCBI35
ENSEMBL_DB               = homo_sapiens_core_26_35
ENSEMBL_DISEASE          = homo_sapiens_disease_26_35
ENSEMBL_EST              = homo_sapiens_est_26_35
ENSEMBL_ESTGENE          = homo_sapiens_estgene_26_35
# ENSEMBL_FASTA            = homo_sapiens_fasta_26_35
ENSEMBL_HAPLOTYPE        = homo_sapiens_haplotype_26_35
ENSEMBL_LITE             = homo_sapiens_lite_26_35
ENSEMBL_VEGA             = homo_sapiens_vega_26_35
ENSEMBL_SNP              = homo_sapiens_snp_26_35
```

... and so on for each species to be served `Mus_musculus.ini`, `Rattus_norvegicus.ini`, etc.

Note if there are new species you'll need to create a new `.ini` file and add the relevant entry to the `$ENSEMBL_SPECIES_ALIASES` hash in `SiteDefs.pm`. Also there won't be a static index page for it yet so you'll need to point people directly at the mapview page e.g. http://ecs2d:5000/Tetraodon_nigroviridis/mapview

At this stage you can also comment out the `ENSEMBL_FASTA` lines.

As Mart is not used, the following 2 lines in the `perl.startup` script can be commented out:

```
#use MartDefs;
...
# MartDefs->new();
```

Then start the server:

```
./ctrl_scripts/start_server
```

or, if you want to delete `config.packed`

```
./ctrl_scripts/start_server -r
```

This will go through all the configuration files and create the file `conf/config.packed`; this is used to speed up starting the server. If you change a config file, you will need to delete this file before you re-start the httpd, so that the changes will be taken into account.

To stop the server:

```
./ctrl_scripts/stop_server
```

The web server can be accessed on http://ecs2a.internal.sanger.ac.uk:5000/ (or whatever server and port you've chosen). EBI users will need to either set up an SSH tunnel or run a web browser over X.

If you experience problems connecting to ecs2a or whichever machine the web server is running on, try using the generic host (ecs2 or equivalent5) instead.