

Ensembl migration

The great (WTSI) escape

Alessandro Vullo
Ensembl Data Review
Oct 17 2016

EMBL-EBI



EMBL – European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton, Cambridge, CB10 1SD, UK



EMBL-EBI



Objective(s)

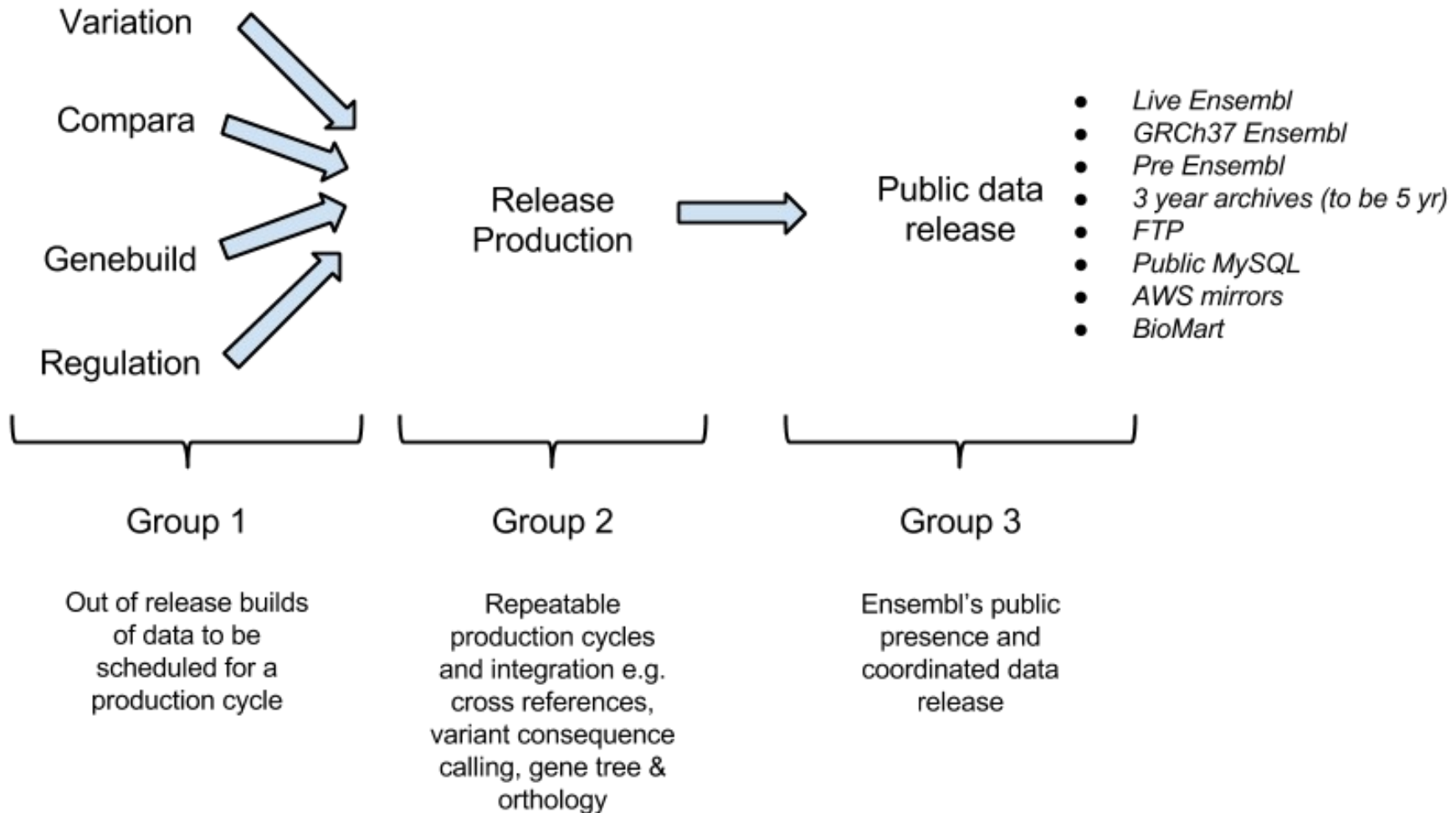
```
assert (e_sanger_footprint ==  
0) if date =~ /Mar 2017/
```

- maintenance of public face is of paramount importance: public web sites (HX/AWS), FTP/MySQL access, RT/mailing lists
- production processes are important as well

The Plan (Summary)

- Migrate 3 groups in turn
 - (1) public archive services (archives, FTP/MySQL)/web presence (HX)
 - (2) release production processes
 - (3) processes outside of release cycle

Ensembl Flow



Sanger Footprint

- ~330TB storage (production, web live/archive servers)
 - DDN SAN 103TB
 - Nexsan 228TB
- 125 servers: ~1300 Cores/~3.2TB RAM
 - 29 DB production
 - 33 DB web & public services
 - 63 web sites (incl. archives)/search/tools
- Farm processing
 - ~500 CPU yrs (2015)
 - ~45 TB clustered filesystem (lustre)
 - ~200TB NFS storage areas (warehouse)

3, 2, 1 go!

- September 2015: commence
- <February 2016: Ok-ish, slow pace
 - initial discussions with EBI/Sanger groups
 - AV takes on PM role (Jan)
 - identify tech requirements in more details
 - benchmarking Sanger machines: mainly storage
 - DB meetings (move data across, back ups, MySQL versions)
- Feb 2016: shutdown in HX, bang!

Hinxton Data Centre Shutdown

- **26-30 Aug 2016**, power cycle (every 5 yrs), electrical maintenance
- Impact: internal/external services HX hosted
- Major consequences for e!

Oh no, suicidal!

- Some production/web db servers on *potentially* faulty IBM blades (50% chance to recover)
- No strategy other than Sanger/IBM try to fix them
- Sanger suggests to migrate away from this HW ASAP!
- Pbs:
 - managing release
 - guarantee public services

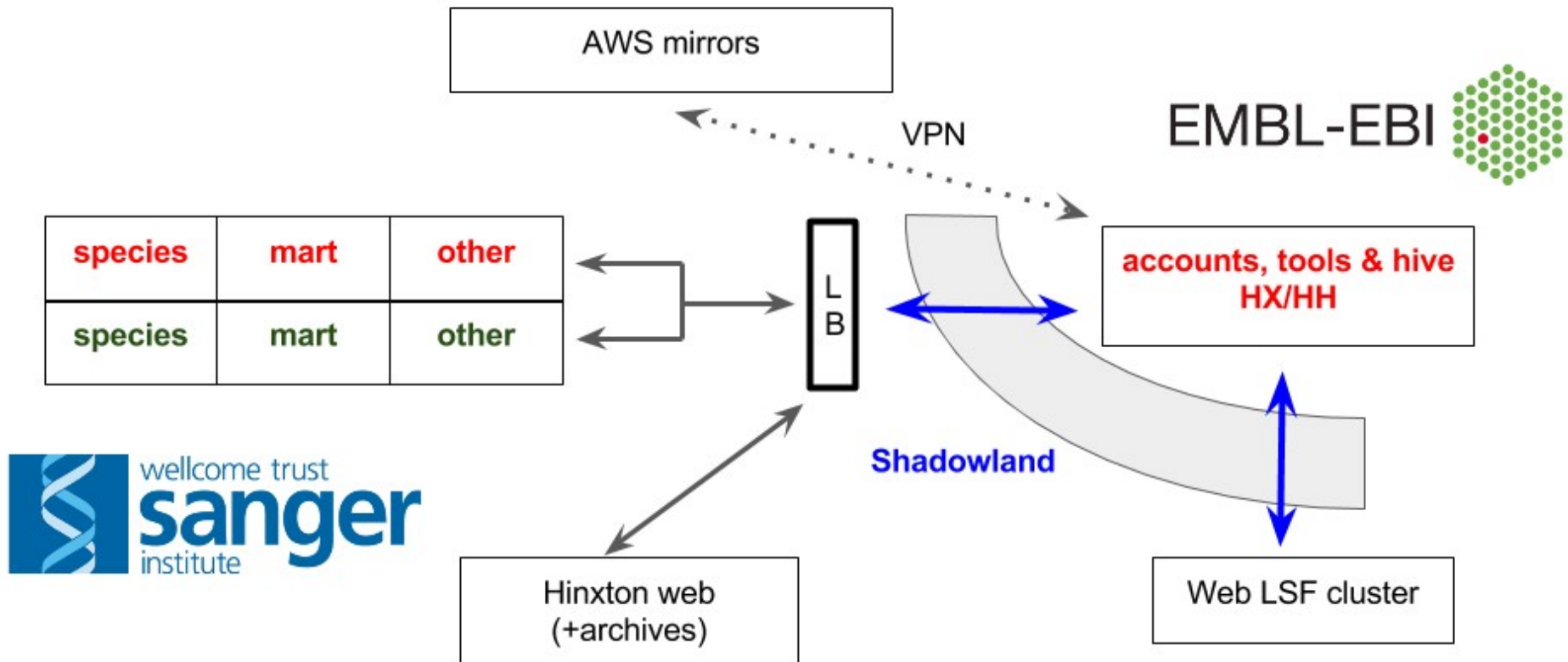
We need HW soon!

- Cope with shutdown, provision in time to advance plan implementation
- Need ~£400K hypervisors/storage
- Have ~£250K (e! overhang budget)
- Will:
 - buy in on e! budget (£150K shortfall)
 - EBI pre-award push shortfall into
 - WT award in Oct charged EMBL pre-award

Strategy (web)

- user/tools/hive DBs migrate to EBI asap
- *shadowland*(?!) connects web/archives to DBs
- EBI-AWS VPN for mirrors
- Sanger blades:
 - web servers OK
 - species/mart/other DBs are suicidal, use Solr servers as fallback

Web & VPN/shadowland



Strategy (production)

- production remains at Sanger
 - e!86 processes finish before (Aug 22)
- staging servers backed up
- teams make sure not to use suicidal HW
- suicidal servers flagged for attention
 - source suitable replacements if pbs

What Happened

- Some VMWare VMs come back with different IPs
- Pb with new Solr servers, java process stopping e!tomcat
- A number of the IBM blades have not come back and some others had trouble with their storage
 - new system board and more memory
 - all blades back and serviceable (Aug 31 evening)

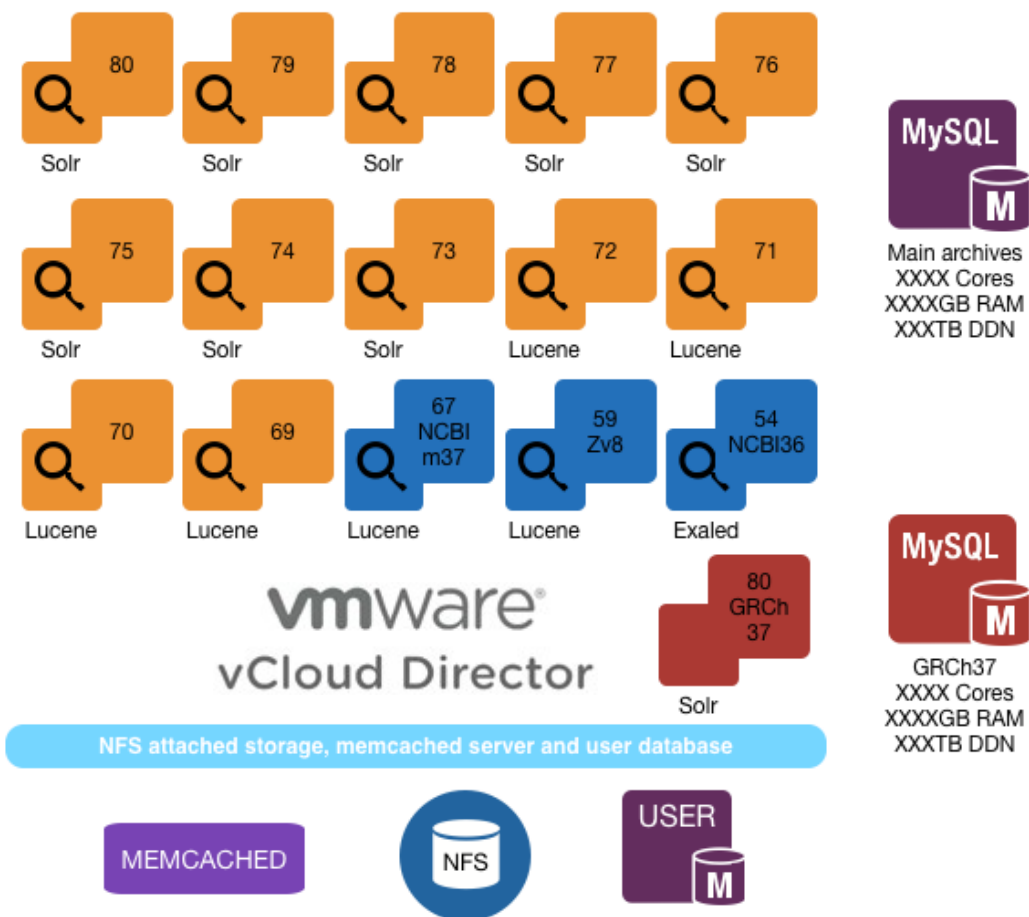
Background

- WP1-2: e! archives
- WP3: e! FTP
- WP4: e! public MySQL
- WP5-6: all content teams
- WP7: live e! websites and services
- WP8: outreach

Archives (WP1-2)

- e! maintains archives (3yrs)
 - past 13 (includes NCBI36 e!54), e.g. may2009.archive.ensembl.org
 - 2 key species archive: NCBI m37 (e67), Zv8 (e59)?
- Objective: **migrate fall 2016**

WTSI set up



Archives Migration Plan

- Host archives on Embassy VMWare
- Spring 2016 - Migrate e!82 archive (WP1)
 - request image
 - requisition DB servers (data hosting, user sessions, mart)/NFS server
 - copy release data from EBI mirror
 - upload WTSI images (web/solr search) – critical

Archives Migration Plan (2)

- Autumn 2016 – Host all archives on Embassy (WP2)
 - transfer/configure all images
 - EBI-Embassy VPN (shared accounts DB)
 - deploy service monitoring infrastructure
 - control procedures/load balancer
 - transfer stats
 - DNS redirect
 - SOP to create archive at EBI

Milestones (70% done)

- Jan '16: **1%**
 - request 82 VMs, set up Embassy space
- May '16: **12%**
 - upload 82 images and data
 - set up main DB/Mart/NFS servers set up
- Jul '16: **25%**
 - infrastructure devs: networking (route trafficking), security
 - shadowland ready
 - upload/testing e54/67 images
- Aug '16: **40%**
 - requisition enough compute/storage: 90GHz/344GB RAM/80TB
 - Sanger-EBI data migration (archive DBs/Mart/NFS)
- Sep '16: **60%**
 - uploaded Sanger archive/search instances (e!73-84)
 - EBI-Embassy data transfer (martdb/nfs)
- Oct '16: **70%**
 - EBI-Embassy data transfer (ensdb)
 - all web machines configured (networks, accounts, passwords)
 - uploaded Sanger e!85 instances

FTP (WP3)

- ~35TB disk: FTP, HTTP(S) and Rsync
- EBI set up
 - 2 LB mirrored servers HH + HX (semi-auto) fail over, 75TB storage
 - extra purchase
- Objective: **migrate fall 2016**

FTP Migration Plan

- Requisition server/storage
- Test FUSE fs access
- Data transfer (Sanger->EBI->HH)
- Release deployment procedures
- Switch DNS

Milestones (80%)

- Sep '15: **0%**
 - pilot study into move
- Jan '16: **1%**
 - request space (not available, need money)
- Apr '16: **5%**
 - data transfer access point/ftp test server available
 - test transfer e!82-83
- July '16: **25%**
 - FTP layout agreement, requisition server/storage
- Aug '16: **40%**
 - internal beta testing available
 - service ready in HH, HX mount point
- Sep '16: **70%**
 - Sanger-EBI data migration (shadowland)
- Oct '16: **80%**
 - moving to HH

Public MySQL (WP4)

- Two servers, releases back to 24 (MySQL 4/5.1)
- GRCh37 server: e!79-86
- Objective: **migrate autumn 2016**

Migration Plan

- Requisition HW/storage
- Copy DBs:
 - 1. 24-47
 - 2. 48-86
 - 3. GRCh37 79-86
- Backup strategy/SOP to push DBs to servers per release
- Sanger-EBI switch

Milestones (71%)

- May '16: **5%**
 - HW procurement order
- Jul '16: **15%**
 - warehouse/shadowland available
- Aug '16: **50%**
 - HW ready
 - data transfer (78TB, no e! 86) to warehouse (shadowland)
- Sep '16: **70%**
 - layout agreement (failover of public services on HH, backup, update)
 - HW available
- Oct '16: **71%**
 - 86 DBs copied (www, grch37)
 - preparing env

Content Teams (WP5-6)

- Filesystems/storage:
 - 45TB clustered (HX farm) + 200TB NFS long-term warehouse (lustre backup)
- Compute:
 - 500 CPU yrs (2015) LSF farm
 - run genebuild/production pipelines etc.
- Objective: **run e!88 at EBI**

Migration Plan

- Archive CVS
- Requisition HW/farm compute/storage
- Survive shutdown
- Software env
- Test pipelines@EBI
- Migrate warehouse, old DBs etc.

Milestones (**65%**)

- Feb '16: **1%**
 - clustered storage options/e! shared user
- Apr '16: **5%**
 - software deps/install (RH6: perls, required libraries)
 - EBI cluster update to RH7!
 - Sanger CVS RO
- May '16: **15%**
 - HW purchase
 - EBI farm minicourse
 - CVS repo at EBI (local co/tarball)
- Jun '16: **16%**
 - RH7 software env (Andy) commence
 - Sanger CVS turned off
- Jul '16: **40%**
 - RH7 software area ready
 - farm storage/compute ready (100TB lustre/300TB warehouse)
- Aug '16: **55%**
 - production HW ready
 - HX shutdown
 - pipelines tested (RH6)
- Sep '16: **65%**
 - RH7 software env ready
 - HW available
- Oct '16: ?
 - pipeline tests ongoing (RH7)

Websites (WP7)

- Services:
 - **Core website (incl. Mart)**
 - **REST**
 - **Search**
 - Accounts
 - Tools
 - Sessions
- www, AWS, GRCh37, e!g delivered from HH
- Objectives: **migrate Q1 2017 (e!88)**
 - infrastructure ready end 2016 (e!87)

Services (Core)

- 3 ro nfs volumes (localsw, 100GB pre-cached data, 10 TB datafiles)
- rw nfs for temp files
- rw sessions DB
- load balanced web/database servers
- dev environment
- email server

Migration Plan

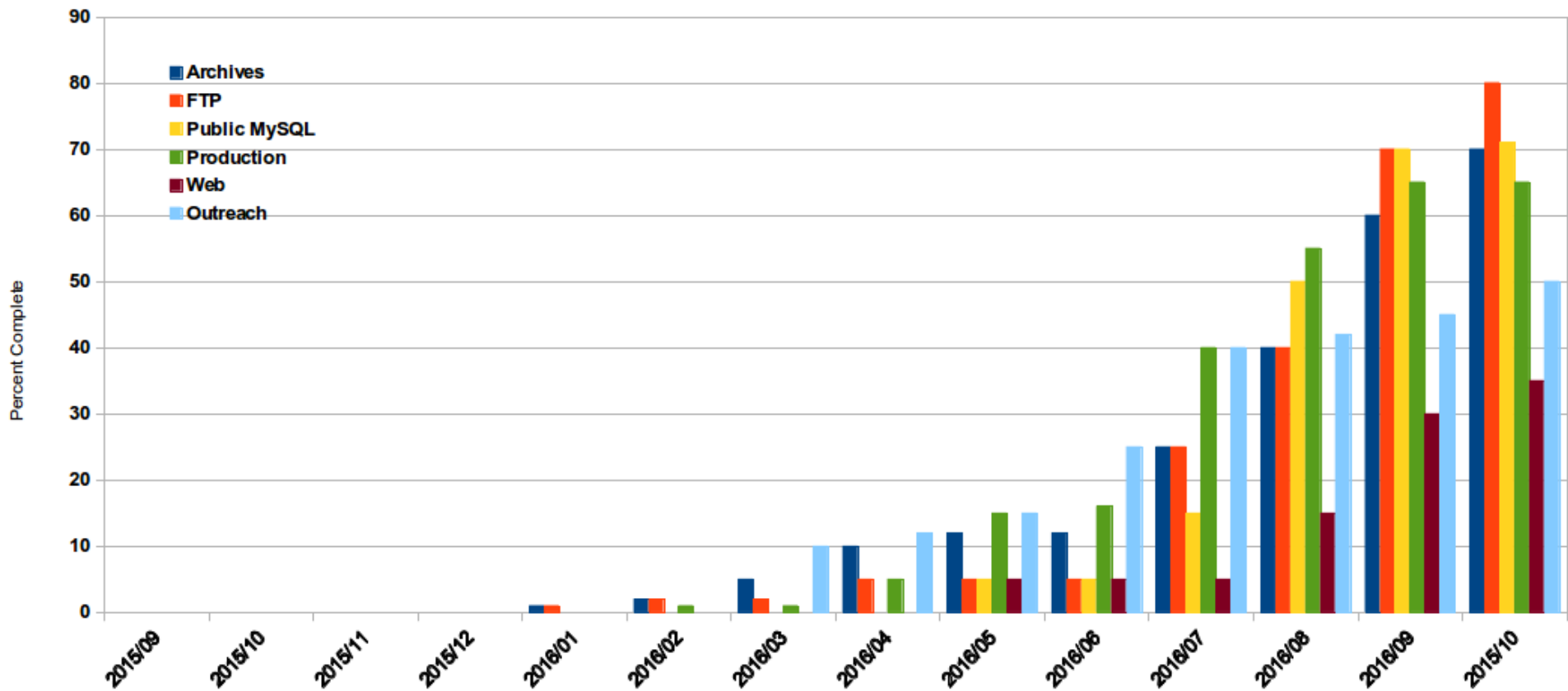
- Requisition HW
- Infrastructure for Tools framework
- Dev areas for web devs
- SOPs
 - test/live sites, search
 - post-release updates (data/code)
 - AWS
- Monitoring/notification/stats
 - define systems, transfer stats from Sanger
- Host in HH/HX
- Transfer domain
- Redirect DNS

Milestones (35%?)

- Mar '16: **1%**
 - domain transferred (no .cn)
- May '16: **5%**
 - HW purchase
- Aug '16: **15%**
 - HW ready
 - accounts/tools/hive DBs at EBI
 - HX shutdown
- Sep '16: **30%**
 - web topology
 - defined dev env
 - dev MySQL server
 - e!86 DBs at EBI
- Oct '16: **35%**
 - request prod DBs, virtual users/project space

Overall: ~60%

- HX shutdown
- Infrastructure:
 - HW resources purchased/partly allocated (budget employed)
 - networking infrastructure to support data transfer
- Public distribution:
 - Archives/FTP/public MySQL on the go (thanks to shadowland)
 - Web infrastructure work started (moved user/tools, web dev/prod envs)
- Production:
 - HW requisitioned (DB servers, Farm, data storage)
 - CVS migrated
 - Long-term (RH7) software environment
- Outreach
 - RT migration to happen next week



e! Migration Project Management

Milestones

- [Archives - WP1/2](#)
- [FTP - WP3](#)
- [Public MySQL - WP4](#)
- [Production \(incl. genebuild\) - WP5/6](#)
- [Web - WP7](#)
- [Outreach - WP8](#)

Activity Logs

- [Archives - WP1/2](#)
- [FTP - WP3](#)
- [Public MySQL - WP4](#)
- [Production \(incl. genebuild\) - WP5/6](#)
- [Web - WP7](#)
- [Outreach - WP8](#)

Acknowledgements

- Mag, Daniel
- Andy, Steve, Dan S²
- Migration champions
- Everybody in e!
- John Barker
- Asier Roa
- Andy Bryant/Cafferky
- Luke Burling
- Paul Bevan