

# PCOG Meeting 2017

---



**Sarah Peter**

University of Luxembourg (UL), Luxembourg





---

# Summary

1 About me

2 Selected projects



# Summary

## 1 About me

## 2 Selected projects



## Who I am

- from Saarland, Germany
- Master in Bioinformatics from Saarland University
- worked 4 years as bioinformatician and data manager @ Max-Planck-Institute for Immunology and Epigenetics, Freiburg, Germany
- joined LCSB BioCore and HPC Team in 2015



# What I do

- HPC

- ↪ system administration
- ↪ user support
- ↪ training
- ↪ software installation (EasyBuild, resif)
- ↪ link to LCSB

- LCSB

- ↪ system administration
- ↪ R<sup>3</sup> (Reproducible Research Results)
- ↪ training
- ↪ NGS data processing & management (EpiPGX, CoGIE)
- ↪ software (license) management



# Summary

1 About me

2 Selected projects



# Revolutionary EasyBuild-based Software Installation Framework (RESIF)

- Automatic Management of **software sets**
- Fully automates software builds and supports all available toolchains
- Clean (hierarchical) modules layout to facilitate its usage
- **Versioning** of software set builds
- Easyconfig files from multiple sources
- Define options and software in easy to read **yaml** files
- Targeted at ULHPC sysadmins use case to build many softwares in one go



## Variant processing pipeline

- over 3000 whole-exome-sequencing samples from EpiPGX and CoGIE projects
- **goal:** efficient processing on the cluster
- no time to develop custom pipeline, but ready-made ones do not support OAR
  - ↳ run only on one node
- need to use specific tools, even if they are badly coded or inefficient
- last step requires combining up to all the samples for one analysis
  - ↳ needs a lot of memory
  - ↳ issues with max number of open file handles
  - ↳ tool is written in Java. . .





## Set up physical servers

- **Certon** boxes
  - ↪ old storage servers from Certon (now bankrupt)
  - ↪ were set up as a distributed storage (GlusterFS), but never worked properly
  - ↪ dedicated sub-network behind a management machine
- **Dragen**
  - ↪ FPGA for next-generation sequencing data processing
  - ↪ comes in a dedicated server (Supermicro)
  - ↪ with 3 TB SSD storage
- **Tasks:**
  - ↪ racking, cableing
  - ↪ network configuration
  - ↪ RAID configuration
  - ↪ OS installation
  - ↪ enrolling with config and user management (puppet & freIPA)
  - ↪ setting up NFS server or attach NFS storage



## Alleviating the storage problem

- Watch storage usage, predict demands
- Proactively extend storage
- On new projects estimate storage demands in the beginning and reserve money to pay for it
- **Compress:**
  - ↪ LCSB bought a license for PetaGene compression suite
  - ↪ Can compress .bam and fastq.gz by at least a factor of 2

## Questions?

<http://hpc.uni.lu>

### Sarah Peter

University of Luxembourg, Belval Campus:  
Ketterhill, 4th floor  
6, avenue du Swing  
L-4367 Belvaux  
*mail:* sarah.peter@uni.lu



1 About me

2 Selected projects