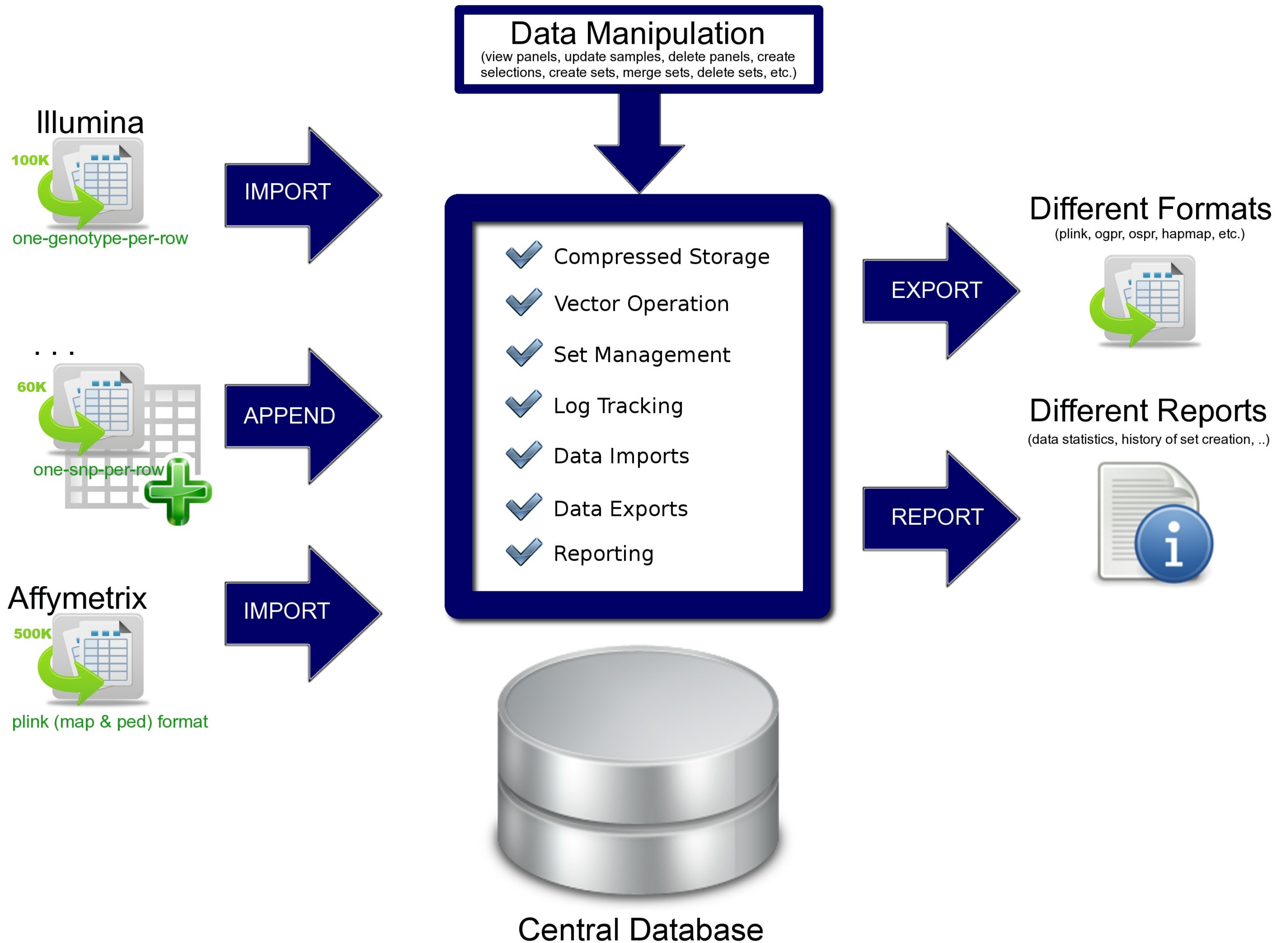


# SNPpit – Efficient Data Management for High Density Genotyping

Eildert Groeneveld and Cong VC Truong  
 Institute of Farm Animal Genetics, Mariensee, FLI, Germany  
 eildert.groeneveld@fli.bund.de

FRIEDRICH-LOEFFLER-INSTITUT  
**FLI**  
 Bundesforschungsinstitut für Tiergesundheit  
 Federal Research Institute for Animal Health



## IMPLEMENTATION

- Linux, Perl and PostgreSQL
- Central Database
- Command Line Interface
- Efficient Storage
- Set based Operations
- Optimal Algorithms using less RAM
- Independent Modules
- Fast Data Retrieval

## PERFORMANCE

- Running on an Intel Core i5 Laptop
- Data: 4 million SNPs \* 270 animals
- Format: Plink; File size: 4.4 GB
- Import time: 30 minutes
- Export time: 10 minutes
- Disk storage: 700 MB
- 500 GB: 700 billion genotypes (500,000 SNPs \* 1.4 million animals)