



**KEEP  
CALM  
AND  
CURATE**



The Scientific Data Consulting Company  
[www.RanchoBioSciences.com](http://www.RanchoBioSciences.com)

Rancho & Pfizer working  
together for the tranSMART  
Community



**tranSMART**  
v1.2

PM: Tatiana Khasanova  
Admin: Julie Bryant  
April, 2015

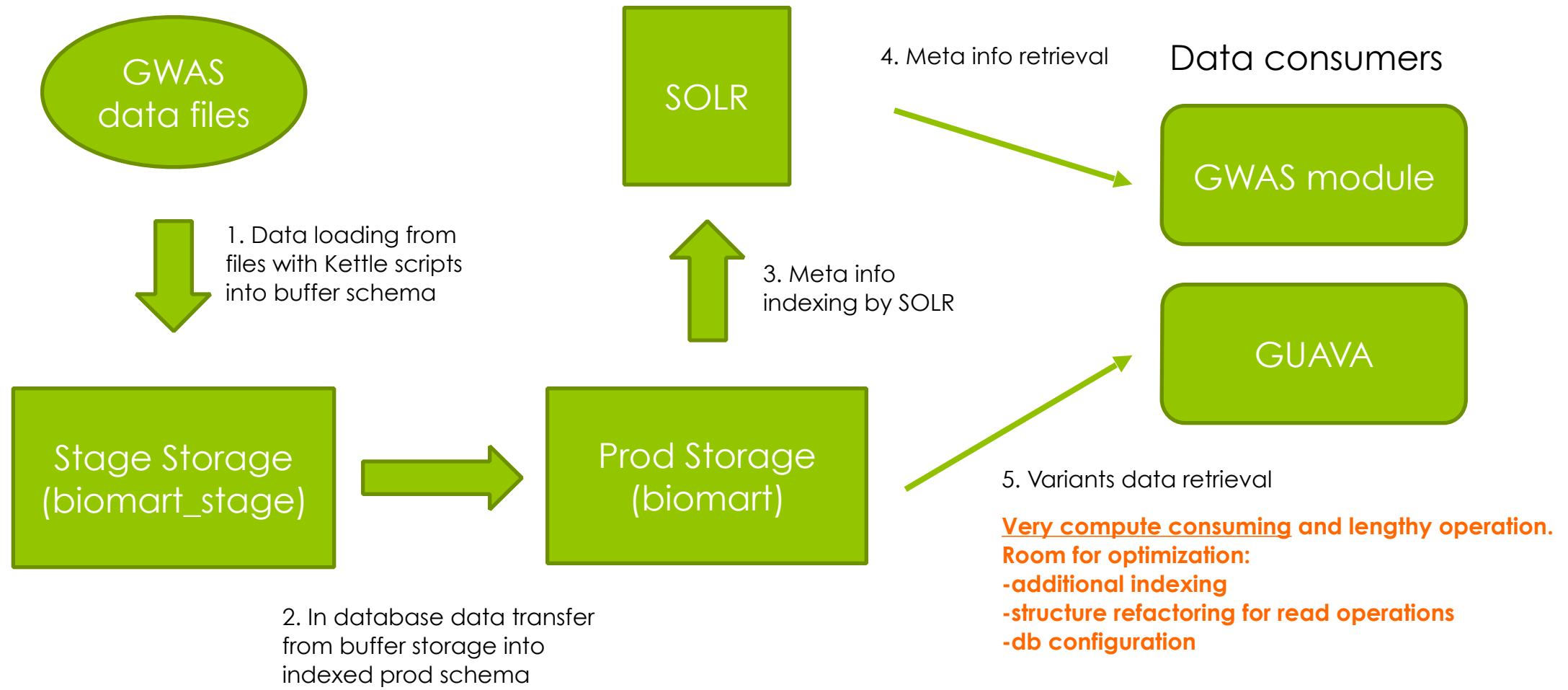
## Customer wanted multiple data types in tranSMART

- Customer chose to have Pfizer GWAS data in tranSMART 1.2
- Rancho could not get the script to work and called Pfizer
- Pfizer very responsive
- We worked together for a solution
- New scripts, documentation in Github

# Tested GWAS package for Oracle

- Data dumps for annotation tables
- Migration SQL scripts for stored procedures
  - Pfizer rolled out scheme of DB but not migration script so we could not apply to existing DB
  - Rancho created migration script to make it work easier-call scripts sequentially
- Indexes for better performance in GWAVA
  - Uses GWAVA to read GWAS data, was slow-made changes to increase speed significantly
- Pfizer Updated Kettle scripts
  - Rancho loaded in Github with package
- Configurations in tranSMART
  - Tweaks to improve for environment with SOLR server-full text search server-grabs any content PDF word etc can configure to search everywhere. Rancho grabs GWAS info meta data. DB working with raw text. Search fast.

# GWAS data flow



# Results

- Instruction tested on public clear build TM  $\geq$  v1.2.2 (Oracle 11g)
- MAGIC dataset is a sample for fast test of concept
- Approach has been used on two big pharma GWAS datasets (10 GB)
  - Loading time took less than 2 days
- Gives community a sense of how long it should take to compare how they are doing

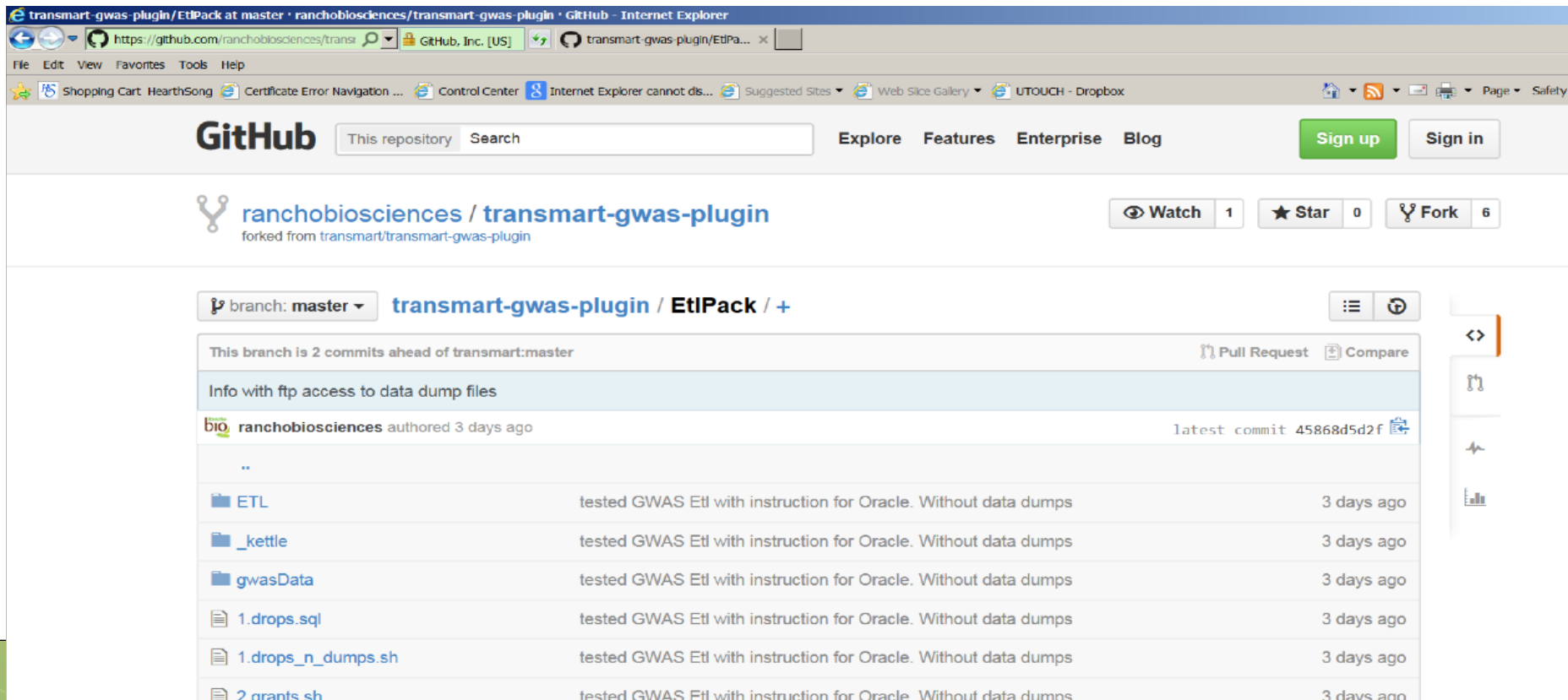
## Future plans

- Test on PostgreSQL version
- Optimize data retrieval operation for GUAVA
- Fix issues with GWAS UI
- All changes posted in Github

# Thank you for your time

GWAS ETL package link:

<https://github.com/ranchobiosciences/transmart-gwas-plugin/tree/master/EtIPack>



The screenshot shows a web browser displaying the GitHub repository page for `ranchobiosciences/transmart-gwas-plugin`. The browser's address bar shows the URL `https://github.com/ranchobiosciences/transmart-gwas-plugin/tree/master/EtIPack`. The page header includes the GitHub logo, a search bar, and navigation links for `Explore`, `Features`, `Enterprise`, and `Blog`. There are `Sign up` and `Sign in` buttons. The repository name `ranchobiosciences / transmart-gwas-plugin` is displayed, along with statistics: `Watch 1`, `Star 0`, and `Fork 6`. The main content area shows the `EtIPack` directory structure, with a commit message: `Info with ftp access to data dump files`. The commit was authored 3 days ago by `ranchobiosciences`. The file list includes:

File Name	Commit Message	Time
ETL	tested GWAS Etl with instruction for Oracle. Without data dumps	3 days ago
_kettle	tested GWAS Etl with instruction for Oracle. Without data dumps	3 days ago
gwasData	tested GWAS Etl with instruction for Oracle. Without data dumps	3 days ago
1.drops.sql	tested GWAS Etl with instruction for Oracle. Without data dumps	3 days ago
1.drops_n_dumps.sh	tested GWAS Etl with instruction for Oracle. Without data dumps	3 days ago
2.grants.sh	tested GWAS Etl with instruction for Oracle. Without data dumps	3 days ago

# Thank you

## **Merck**

Meiping

Scott

## **Pfizer**

Jay

Haiyan

Hugo

## **Rancho**

Alex

Tania