Cheat sheet on how to calculate realized genetic gain

Giovanny Covarrubias-Pazaran
Breeding scheme optimization lead, Excellence in Breeding Platform

This cheat sheet provides a step-by-step approach to calculate the realized genetic gain. In essence we need to generate phenotypic data, split it based on targets, and analyze it in the proper way.

	Analysis of historical data	Analysis of era trial data
D A T A G E N E R A T I O N	Approach your database manager.	Go to your germplasm bank manager.
	Gather the data records from a selected pipeline we are interested in.	Gather the list of materials for the pipeline you are interested in.
	Split the data by market target (TPE + product features) (e.g. Drought conditions environments late maturity).	Split the list of materials by market target (TPE + product features) (e.g. Drought conditions environments late maturity).
	Keep the split data for a time period of interest (i.e. last 10 years).	Keep the list of materials for a time period of interest (e.g. last 10 years).
	Keep the split data for a given germplasm stage(s) of interest (e.g. PYT & AYT or OnFarm).	Keep the list of materials for a given germplasm stage(s) of interest (e.g. PYT & AYT or OnFarm).
	Keep the split data only for the trait(s) of interest (e.g. yield).	Decide the traits to phenotypes in the experiment.
	Hand the batches of data to your Biometrician.	Hand the list of materials to your Biometrician to run an experimental design.
		Plant the experiment and harvest the phenotype data.
A N A L Y S I S	Ask your Biometrician to fit a multi environment trial (MET) analysis with a mixed model accounting for nuisance (spatial, environment, year) and genotype (genotype, genotype by year, genotype by location, genotype by year by location) factors. Extract adjusted means for all genotypes.	
	Ask your Biometrician to merge the year of origin of the genotypes to their adjusted means, and ask him/her to fit a linear model of the form *adjusted.mean~year.origin*	
	Take the slope of the regression and present as genetic gain to stakeholders (e.g. donors and other organizations).	

To calculate the realized genetic gain once the data its in proper shape the Excellence in Breeding team has made scripts available. They can be found at
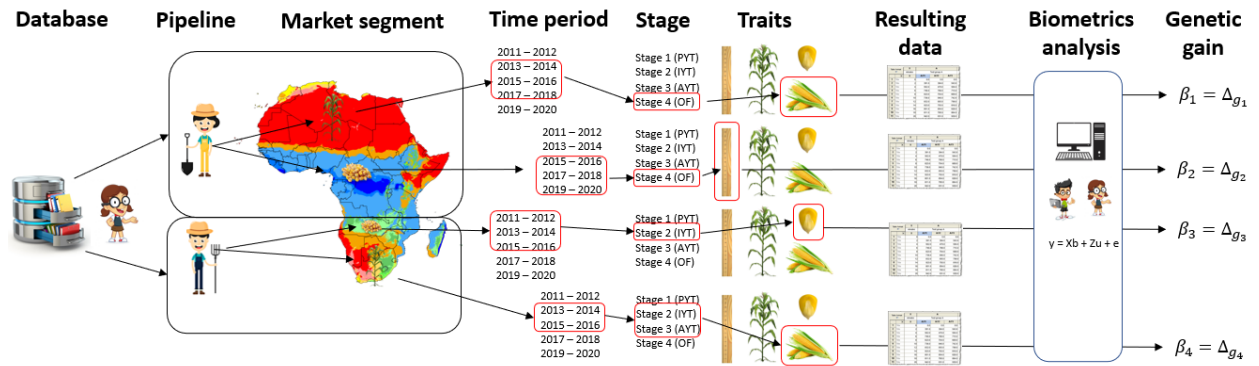https://gitlab.com/excellenceinbreeding/module2

Figure 1. Graphical representation of major steps to calculate realized genetic gain using historical data. A critical part is to split the data in a meaningful way to interpret properly the resulting value of genetic gain.