

FAIRagro use case

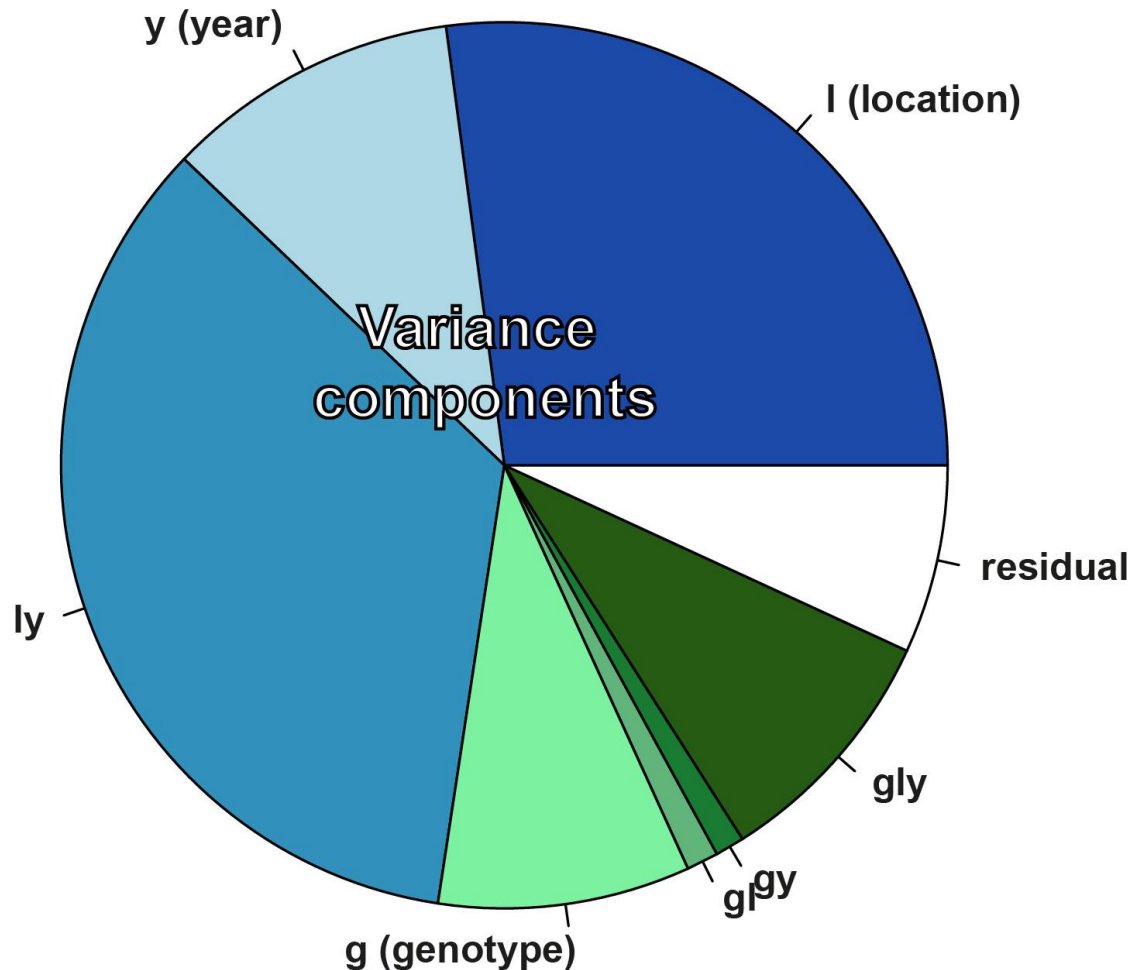
„Exploiting genotype \times location \times year \times management interactions for sustainable crop production“

Prof. Dr. Jochen C. Reif

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK);

reif@ipk-gatersleben.de

Photo: Petair / Fotolia



Agronomy and plant genetics are rarely considered in an integrated fashion. This is mainly due to the difficulties of having integrated curated data available.

=> Can we close the gap between yields measured in experimental field trials and on-farm yields?

Scales

The use case bridges the scales “plant” - “field” - “farm”

Agricultural scientific discipline

The use case integrates partners from following agricultural scientific disciplines: Plant breeding, Plant nutrition, Plant pathology, Agronomy, Bioinformatics, and Biostatistics/Modelling.

The main objective in this use case is to **enable, develop, and validate knowledge-based prediction models considering genotype × location × year × management interactions** using **wheat and maize** as model crop. To this end, data on environmental parameters, weather data, trial design, genetics of varieties, and their yield and quality parameters under agricultural production conditions will be curated, harmonized, stored and made available.

The data will be provided by public real-world laboratories focusing on experimental field stations. The development of the knowledge-based prediction models considering genotype × location × year × management interactions can build upon ongoing research conducted by the participating partners. Thus, within 5 years these prediction models can be developed.

What already exists?

- First generation of algorithms available for predicting genotype × location × year × management interactions (e.g., <https://doi.org/10.1038/s41467-020-18480-y>);
- Partially curated data on agronomic traits (<https://doi.org/10.1038/sdata.2018.278>) and SNP profiles (<https://doi.org/10.5447/IPK/2018/9>) of comprehensive populations.
SNP <https://doi.org/10.1038/s41588-018-0266-x>;
- Data on environments is available in other domains (e.g., <https://lpdaacsvc.cr.usgs.gov/appears/>; <https://doi.org/10.1038/s41597-019-0146-y>).
Detailed data on crop management is often not recorded in plant genetic studies.
- Repository for SNP profiles (e.g. <https://www.ebi.ac.uk/eva/>) and agronomic data. Often not linked with detailed characterization data of locations and management.
SNP: EBI EVA Agronomic: [e!DAL-PGP](#)
- Experimental meta data are recorded in a non-public LIMS/ELN database

What is realized in the use case (outcomes)?

- Interoperability between platforms providing integrated and curated data allowing to study the potential of genotype × location × year × management interactions.
- Advanced algorithms for predicting genotype × location × year × management interactions.
- Connectivity between predictions and on-farm yield.

Generation and curation of phenotypic data, genomic data, and metadata for classical plant breeding trials:
IPK, JKI, UHOH, ...

Detailed description of environments: Thünen Institut, ...

On-farm data: Technische Hochschule Ostwestfalen-Lippe, ...

Increase of data providers desirable. In particular, official variety testing data or data from regional trials are very interesting, but access is challenging due to legal aspects.

Genomic data => interoperability between different platforms (e.g. between sequence data and SNP arrays; often associated with legal issues such as knowledge about array design)

Phenotypic data => Hard to work with standardized trait measurements. Detailed metadata is required and case-by-case decision on how to integrate heterogenous data.

Envirotyping data => Interoperability to phenotypic data is often missing.