The Generation Challenge Programme Model-Driven Architecture: Scientific Domain Model and Ontology

Genevieve Mae Aquino[1,2], Jeffrey Detras[1], Rosemary Shrestha[2], Martin Senger[1⋆], Kevin Manansala[1], Elizabeth Arnaud[3], Reinhard Simon[4], Jayashree Balaji[5], Thomas Hazekamp[6], Guy Davenport[2], Graham McLaren[1], Theo van Hintum[7], and Richard Bruskiewich1[1⋆⋆]

[1] Crop Research Informatics Laboratory - International Rice Research Institute (IRRI), Philippines
[2] Crop Research Informatics Laboratory - Centro Internacional de Mejoramiento de Maz y Trigo (CIMMYT), Mexico
[3] Bioversity International, Montpellier, France
[4] Centro Internacional de la Papa (CIP), Peru
[5] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India
[6] Bioversity International, Rome, Italy
[7] Centre for Genetic Resources, Wageningen Universiteit en Researchcentrum (WUR), The Netherlands

The Generation Challenge Programme (GCP; `http://www.generationcp.org`) is a globally-distributed crop research consortium directed toward crop improvement through the application of comparative biology and genetic resources characterization to plant breeding. To achieve interoperability and integration within the GCP across diverse data types published by project-distributed data sources and consumed by end-user data analysis tools, the GCP adopted the development paradigm of a "model-driven architecture". At the core of the GCP architecture is a scientific domain model, heavily parameterized with GCP-indexed ontology. The GCP scientific domain model is an object model that encapsulates key crop science concepts and is documented using Unified Modeling Language. The GCP-indexed ontology reuses established international standards where available, converts other publicly-available controlled vocabulary into formally-managed ontology, and develops novel ontology if no public vocabularies yet exist. The GCP architecture is being translated into various open source object-oriented software libraries and into data types for various public semantic web protocols. The GCP model-driven architecture is specified as a common semantic framework for the implementation of interoperability and integration of diverse crop data sets and analysis tools. The architecture and associated open source software are accessible at `http://pantheon.generationcp.org`

Key words: crop informatics, agriculture, ontology, domain model

---

⋆ Presenting author.
⋆⋆ Corresponding author: r.bruskiewich@cgiar.org