



Generation Challenge Programme Making Data Accessible (SP4)

Tom Hazekamp (Bioversity International) & Guy Davenport (CIMMYT)

The research projects within the Generation Challenge Programme (GCP) generate a large amount of data. The GCP is committed to making these data fully accessible to any authorized user in such a way that they can be fully interpreted and easily converted to any other format. The Bioinformatics and Crop Information Systems subprogramme has developed **DATA TEMPLATES** and a **CENTRAL REGISTRY** to help GCP scientists make their data accessible to others.

How to MAKE YOUR DATA ACCESSIBLE

The data produced are generally suitable for use within the research group which generated them, but they usually need some processing to make them suitable for use by others.

1. Format data using templates

Data templates have been developed for various data types by the Data Templates Project. The objective is to provide simple templates for the temporary storing or distributing of the different data sets that are being produced within SP1, SP2 and SP3, for which there is no current provision in public or institutional databases. These templates contain documentation on how they should be used. The completed data sets should contain the necessary information to be stand-alone and should be simple enough to be understood.

2. Validate data

Software has been developed to validate data which have been formatted using the data templates. The data are validated for the existence of data in mandatory fields, appropriate use of codified data, etc. The stand-alone version of the software can be downloaded from the Data Template section of the GCP Bioinformatics Portal. A web-based version of the validation tool is under development.

Contact *Guy Davenport* for information on Data Templates & Validation Tools

3. Register dataset

The GCP Central Registry provides an on-line central facility where data can be made accessible to authorized users. To make your data available through the Central Registry you need to register the dataset first. During registration you will be asked to provide metadata describing the dataset (e.g. dataset name, originating project, type of dataset etc).

NB. You can register datasets before the data actually exists!

4. Submit dataset (or related resources)

Once the metadata for the dataset has been registered, you can attach resources to it – these might be data files, reports, images or web services. These resources will then become accessible on-line to authorized users.

Contact *Tom Hazekamp* for information on the Central Registry

	SP1	SP2	SP3	SP4	TOTAL
Registered datasets	50	40	11	1	102
Uploaded files	27	11	1	0	39
Datasets with error	0	0	0	1	1

