

FAQs about Peanut Genome Project Policies & Procedures on Data Sharing

Question 1: I agree with the bylaws in principle, that genome sequence information should be kept free of IP. I would prefer that the same be true also for DNA markers. Are they covered?

Response: DNA markers are covered by reference in **Section 3.02 Data Sources**. “Research Data” embodies the recorded factual material commonly accepted in the scientific community that is relevant to the generation, validation or defense of research findings; and in **Section 3.05 Terms & Conditions**. The User shall not claim legal ownership over the information and data found in the data base nor seek intellectual property protection under any form over these information, data and data base. For clarity, the user agrees not to claim any of the sequences disclosed in these databases in any patent application.

Question 2: If I read the document properly, the sequence data would be public, but breeders could use this, say by marker-assisted selection, to develop varieties that would have IP, such as PVP which all public sector peanut breeders in the US are using as far as I'm aware. Is this reading correct?

Response: As stated in **Section 3.05 Terms & Conditions**..., the foregoing shall not prevent the User from releasing, reproducing or seeking intellectual property protection on improved seeds or plants that may be developed using the information for purposes of making such seeds or plants available to farmers for cultivation.

Question 3: The document is broader than sequences alone, and states that other genomic data is to be deposited with LIS.

Question 3(a): What other data would be deposited there?

Response: As stated in **Section 3.08 Further Information**, a physical archive of PGP data is under development at NCGR. Ideas on database content are welcome and will be considered by the EC. Please review current LIS content for typical genomic information in the databases for other species.

Question 3(b): Would there be an option to deposit data from other projects - for example ARS awards?

Response: As stated in **Section 3.06 Acknowledgments and Further Disclaimers**, data owners give NCGR permission to hold copies of data/image(s) within the PGP archive. NCGR holds datasets provided by a variety of public and private bodies, and individual researchers. The data/image owner or copyright holder will be acknowledged as the source of any data that is available for distribution. Data owners retain the copyright of the original data or images at all times. The agreement will help protect the copyright of the data and image owners.

Question 3(c): Would there be a fee involved to do so?

Response: As stated in **Section 3.07 Charges**, PGP data and image resources are made available free of charge to registered individuals on-line for use in not-for-profit decision making, research, education and other public-benefit purposes. However, **Section 3.03 Data Use** states that all data and image resources will be fully interactive with the Legume Information System (LIS) for comparative and association genomic analyses. Pursuant to **Section 3.07** users with access to data and resources housed in LIS may be assessed a nominal charge for any analysis, query or collation of data that requires NCGR staff time. So, there should be no charge for deposition of data in LIS; but, a charge may be warranted in certain cases such as when NCGR staff time is expended to translate data into a syntax that is interactive with LIS programs.

Question 3(d): Would there be a distinction between data generated from grants that are from public institutions (PF, TPPB, ARS) and awards from private companies?

Response: As stated in **Section 3.01 Purpose**, a published Data Sharing & Use Policy contributes to more transparent working relations and sets criteria for managing permission to use and distribute data. Such policy assures providers that data is managed responsibly, and ensures data users acknowledge the data source and the conditions under which the data is made available.

Question 3(e): Would the agreement preclude one from developing genomic data under funding from a private company, that would not be made public?

Response: As stated in **Section 3.02 Data Sources**, Research Data does not include: 1) preliminary analyses; 2) drafts of scientific papers; 3) plans for future research; 4) peer reviews; 5) communications with colleagues; 6) physical objects (for example, laboratory samples, audio tapes, video tapes); 7) trade secrets; 8) commercial information; 9) materials necessary to be held as confidential by a researcher until

publication in a peer-reviewed journal; 10) information which is protected under the law (for example, intellectual property); 11) personnel and medical files and similar files, the disclosure of which would constitute unwarranted invasion of personal privacy; 12) information that could be used to identify a particular person in a research study; or 13) data that are already available to the public domain through an archive or other source. In general, this policy should be interpreted on a case by case basis. Pending additional information certain points can be made in regard to the proposed question: 1) the PGP database will not contain trade secrets; 2) a trade secret is a form of intellectual property protection, but there is no disclosure and no restriction on others who may make the same discovery; 3) Users of the database agree not to claim any of the sequences disclosed in the database in any patent application. So, the agreement does not preclude the development of privately-held information as long as no other form of protection is sought over Research Data that exists in the database.

Question 3(f): As I read it, the agreement explicitly grants the right to develop new varieties and obtain IP over these. Is this correct?

Response: As stated in the preamble for **Article III. Data Sharing & Use Policy**, the PGC has the explicit goal of enabling breeders, geneticists, molecular biologists and other researchers to accelerate the pace of enhancing productivity, crop protection and product quality/safety of the cultivated peanut. PGP investigators, collaborators and associates are encouraged to share data, inventions and other resources, and to make products from the project widely available and useful in a timely manner

Question 4(a): Making data public, the question here is timing. If I read it correctly, the data would need to be made public, but not before publication. This is an issue because putting on a website is considered prior publication by some journals, and would preclude publication in journals. I also expect that universities may consider breeding line data confidential.

Response: Not necessarily. As stated in **Section 3.04 Data Sharing**, The PGC is committed to promoting the use of the data held within the archive. PGC associates believe the subject data resources should be available for use in not-for-profit decision making, research, education and other public-benefit purposes. However, when warranted PGC, as a responsible data custodian, may restrict access to all or part of some data resources. When restrictions are applied the justification and rationale behind the decision will be documented and made available. Examples of reasons why PGC might restrict data access are listed below.

- If the release of certain data is likely to increase the risk of environmental damage or put particularly sensitive species at risk.
- If the release of data is likely to jeopardize the supply of data and collection of future data resources. NCGR is a data custodian and does not own the data resources held in LIS.
- If the data provider has requested that certain data or information be withheld when, for example, data may be commercially sensitive or where the data is under preparation prior to publication. The restriction may be permanent or temporary depending on the restrictions agreed with the data provider.

Question 4(b): At an APRES meeting a few years ago, the PVP office in Beltsville told the peanut breeders that prior publication of breeding data would be cause to deny PVP on the variety.

Response: Obtaining intellectual property rights requires public disclosure. For PVP, full disclosure includes the parentage and breeding methods. In addition, the variety must be novel based on distinctness from all previously existing varieties. Once a complete application is filed in the PVP Office, the application is assigned to an examiner. The examiner conducts a literature search of the crop and gathers descriptive information on varieties from grow-out trials, release notices, seed catalogs, PVP applications, and other published sources. The examiner then uses the appropriate database to determine the novelty of the application variety. This policy has been in effect for decades. Therefore publication would benefit a PVP application by helping document novelty and ownership of the material, which would be extremely important if two different breeders applied for PVP on sister lines from the same parentage.

Pardee WD. 2007. Protecting New Plant Varieties through PVP: Practical Suggestions from a Plant Breeder for Plant Breeders. In *Intellectual Property Management in Health and Agricultural Innovation: A Handbook of Best Practices* (eds. A Krattiger, RT Mahoney, L Nelsen, et al.). MIHR: Oxford, U.K., and PIPRA: Davis, U.S.A. Available online at www.ipHandbook.org

Question 5: What sort of signatures would be needed for the IAA? I think that most public or private sector scientists could not sign it on their own. This would be a legal document and so would also need signatures from higher administrations, and so this would need to be examined by the legal offices of all the institutions involved, as in fact all grant proposals and MTAs are.

Response: Research Data developed by a PGC member that is contributed to the Peanut Genome Project remains the intellectual property of that member and is owned by the member's sponsoring institution. The IAA is a written contract or end-user license agreement (EULA) under which the owners of such intellectual property collectively agree to convey permission through the PGC to a licensee that allows access and use of Research Data in the PGP database. EULAs help protect both provider and user; and have become a standard feature of daily business transactions. Typical examples include rental car agreements or software license agreements that when approved or accepted grant the end user access to the product or service in question. Likewise, the purpose of the IAA is to protect the rights of those who own the information that is deposited in the database. The IAA may be signed by the User or the User's sponsoring organization. The IAA is a moral and ethical covenant between professional colleagues or institutions who own, share, and use the PGP database. Hence, the IAA is an instrument based on trust, and is enforceable through precedents established by the Human Genome Project and by peer conventions within the global genomics research community.

Question 6: If marker data were made available on specific populations, would it be mandatory to release the specific breeding lines or populations?

Response: As stated in **Section 3.02 Data Sources**, Research Data does not include physical objects such as laboratory samples (or breeding lines or populations for example). However, disclosure of the parentage of the breeding lines or populations would be necessary in most cases to ensure the optimal utility of the genetic markers or validation of QTL. The release of such material is left to the discretion of the developer, the owner and acceptance by peers. Use and distribution of such material should be governed by established mechanisms such as Material Transfer Agreements between appropriate parties.