



**Food and Agriculture
Organization of the
United Nations**



The International Treaty
**ON PLANT GENETIC RESOURCES
FOR FOOD AND AGRICULTURE**

Item 16.1 of the Provisional Agenda

EIGHTH SESSION OF THE GOVERNING BODY

Rome, 11 – 16 November 2019

**Compilation of information on “digital sequence information” with respect
to plant genetic resources for food and agriculture**

Executive Summary

This Addendum contains the submissions on “digital sequence information” that the Secretariat received after 17 May 2019. All submissions are made available on-line at the following URL:
<http://www.fao.org/plant-treaty/overview/mypow/dsi/en/>.

*This document can be accessed using the Quick Response Code on this page;
an FAO initiative to minimize its environmental impact and promote greener communications.
Other documents can be consulted at [http://www.fao.org/plant-treaty/meetings/meetings-
detail/en/c/1111365/](http://www.fao.org/plant-treaty/meetings/meetings-detail/en/c/1111365/)*



nb636

Submission by Poland

Our contribution to each category of the requested information is as follows:

1. Terminology used in this area

In our opinion, the term 'Digital Sequence Information' is not commonly used or understood by any stakeholder of genetic resources. It is imprecise and unintuitive. We are convinced that it should be replaced by 'Genetic Sequence Data' that is much better recognized and does not cause any scientific controversy.

We do not agree to treat 'DSI' as genetic resources as has been proposed. 'DSI' is neither genetic resource nor genetic material because it is not physical material. In our opinion, this is also not in accordance with Article 2 of the Convention on Biological Diversity, Article 2 of the International Treaty on Plant Genetic Resources for Food and Agriculture and also with Article 2 of the Nagoya Protocol on Access to Genetic Resources and Benefit-Sharing.

2. Actors involved with DSI on PGRFA

- Researchers in the fields of biodiversity, ecology, genetics, genomics, molecular biology and bioinformatics.
- Institutions such as gene banks, research institutes, universities.
- Breeders and breeding companies.
- DNA sequencing international consortia

3. The types and extent of uses of DSI on PGRFA, such as: characterization; breeding and genetic improvement; conservation; identification of PGRFA

Generating and using of 'DSI' is crucial for:

- efficient gene bank management by identification of duplicates, development of core collections
- diversity studies for conservation,
- verification of genetic integrity after long-term storage and regeneration
- molecular taxonomy and identification
- breeding
- comparative genomics for identification genes responsible for important traits

4. The relevance of DSI on PGRFA for food security and nutrition

DSI is a versatile tool for worldwide research aimed at minimizing or excluding food safety problems and sustainable agriculture. It is therefore essential that they should be maintained in an open access for all interested stakeholders in their use. DSI is an effective tool for both *ex* and *in situ* conservation of PGRFA. Systematic genotyping allows gene banks to monitor genetic variability and integrity, identify duplicates, create core collections, taxonomic identification, establish kinship, phylogenetic relations and so on. Omic technologies support crop research on the genetic background of resistance to diseases

Republic of Korea's submission of views and information on Digital Sequence Information to FAO

1. Terminology used

There is international controversy over the scope and definition of the term, 'Digital Sequence Information (DSI)', which partly arise from ambiguity of the term itself. DSI is not generally accepted term in scientific community, but rather a placeholder term. However if it would cause confusion, it is best to be replaced with more accurate, appropriate term for the sake of communication and discussion.

What is being referred to as DSI, as far as how we understand, is sequential order of genetic material (DNA or RNA polymer), not some sequence of numbers or events. So it should best be phrased 'genetic sequence' to clarify the meaning. It should not necessarily be in digital format, so the phrase 'digital' shall be omitted. Also, we suggest using 'data' instead of 'information' as 'information' often refer to broader meaning, thus increasing ambiguity of the term. Sequential order of genetic material, such as ATGCCTCCAGT, clearly fits with definition of 'data'. In conclusion, Republic of Korea suggests using alternative term, 'Genetic Sequence Data(GSD)', for further discussion on the related issues in ITPGRFA. We will use the term GSD for the rest of this document.

We should also point out that GSD is not equivalent to genetic material. Article 2 of CBD defines genetic material as 'any material of plant, animal, microbial or other origin containing functional units of heredity', which is clearly not the case of GSD. It is mere intangible data of nucleotide sequence order. Also, since CBD defines genetic resource as 'genetic material of actual or potential use', GSD is not genetic resource as well.

2. Actors involved

The key actors regarding this issue are obviously researchers. Vast majority of GSD is generated and used by researchers and it has become quintessential part of biological studies. It should be noted that there are serious concern among researchers on this issue. GSD is so widely and frequently used by researchers worldwide that any restriction on using and sharing of GSD or even possibility of getting into legal affairs will lead to catastrophic consequence in the field of science. Scientists also stress that studies using GSD expand our understanding on biological diversity and it provides helpful tools, such as DNA markers, for conservation and sustainable use of biological diversity.

Other important players are online database providers, such as academic institutions, journals, governmental agencies. Most of such databases are freely accessible, benefitting researchers all over the world. However, should ABS measures be placed on GSD, many of such open DBs are likely to be closed in order to avoid legal trouble. Such will lead to even greater knowledge gap across the globe and hinder capacity building of researchers in developing countries.

3. Types and extent of uses of GSD on PGRFA and its relevance to food security and nutrition

GSD offers important tools for identification, breeding and conservation of plant genetic resource for food and agriculture. Aforementioned DNA markers, for example, are used to identify species and varieties, predict traits, increase breeding efficiency and select conservation specimen pool to maximize genetic diversity. GSD is also extensively used in the field of fundamental science including taxonomic, molecular and ecological studies. GSD is key to share and verify scientific discovery as well, so many journals require submission of GSD analyzed in the study.

GSD used for crop breeding help enhance food productivity, thus making it more accessible, available and affordable. Also it can be used to select traits for drought, pest and other resistance or higher nutritional value, hence assist stable food production and nutrition. So, GSD contribute to all four pillars of food security and nutrition which are key missions of FAO.

Submission of views and other information on “Digital sequence information”

Terminology used

The term Digital Sequence Information has been widely used by Parties to the Treaty and CBD to describe certain data generated from activities aimed at understanding the molecular genetics of organisms. As such it is widely acknowledged that it has limitations (‘information’ is too broad, ‘sequence’ is very context dependent and whilst ‘digital’ is intended to distinguish such data from the physical resource, it too could have a number of interpretations; for example would digital images be included?). Furthermore, the ITPGRFA’s Global Information System already uses the phrase ‘Digital Object Identifier’ to catalogue physical PGRFA and facilitate access and benefit sharing. The CBD has stated that DSI should be used as a ‘placeholder’ phrase until broad agreement is reached on a more accurate term and associated definition (see decision COP13/16 and COP14/20). The WHO has adopted ‘Genetic Sequence Data’ as an alternative to DSI and some Parties to the ITPGRFA have expressed a preference for this phrase and associated acronym. The UK would be open to adopting such a phrase but recognises the importance of a science-based process to inform any multi-lateral decision.

Actors involved with DSI on PGRFA;

All universities, institutes, companies and other organisations investigating the genetics of plants in the UK probably use DSI to a greater or lesser extent. Below are listed a few key users of DSI in the UK, with a particular focus on genebanks and genetic improvement. The crop Genetic Improvement Networks (GINs) are significant users of DSI, for the identification of new sources of resistance to pests and diseases as well as new sources of tolerance to abiotic stress (eg drought, water logging, nutrient deficiency etc)

- The Pea Genebank and Pulse Crop Genetic Improvement Network (PCGIN) at the John Innes Centre
- The Vegetable Genebank and Vegetable Genetic Improvement Network (VeGIN) at Warwick University
- The Oilseed Rape Genetic Improvement Network (OREGIN) at the University of York
- The Wheat Genetic Improvement Network (WGIN) at Rothamsted Research
- The National Fruit Collection at Brogdale
- The Genetic Resources Unit at John Innes Centre
- The Earlham Institute
- National Institute for Agricultural Botany
- The Millenium Seed Bank at Kew (Wakehurst Place)

Types and extent of uses of DSI on PGRFA

In pre-breeding research, large scale manipulations of DSI data are undertaken using bioinformatics programmes. A multitude of genomics technologies have been developed which use DSI for understanding gene function and genetically characterising individual plants and populations towards the genetic improvement of

crop species (eg TILLING, Genome Wide Association Studies, Genotyping By Sequencing, Marker Assisted Selection, QTL mapping, KASP markers, multiplex PCR, qPCR, associative transcriptomics, exome/promotome capture, AgRenSeq and so on). These methodologies are routinely applied for capturing the genetic potential of the UK's genebanks and are fundamental to the UK's crop Genetic Improvement Networks (GINs)

In the UK, commercial breeders use DSI to assist the generation of higher yielding varieties. Publically funded research in the UK (such as the GINs) however is directed at producing crops which are resilient to food security threats such as pests, diseases and the weather. In the case of pulse crops, research is also aimed at improving seed quality, in an attempt to encourage greater adoption by growers within the arable rotation. The maintenance of the pre-breeding pipeline for these resilience traits is essential for the long term security, productivity and sustainability of the arable sector. In acknowledgement of this fact commercial breeders are key stakeholders working closely with the publically funded GINs to ensure end-user needs are met.

There is widespread recognition within the international research community that it would be useful to work towards harmonised bioinformatics systems, interoperable platforms and shared ontologies for the storage and sharing of DSI. Examples of initiatives that seek to achieve this include BrAPI, Germinate 3, DivSeek, Crop Trust, SeedStor, CIAT etc.

Whilst some developing countries may not have the resources available to generate and apply DSI for the conservation and improvement of native crop species, the UK, and other Parties to the Treaty, have substantial overseas development programmes (for example the UK's Global Challenges Research fund and the Darwin initiative) which foster the collaborations necessary to help in this area. The sharing of DSI is a vital component of these activities for the co-development of genetic maps and markers and the publication of scientific papers leading to advances in crop conservation and improvement.

Relevance of DSI on PGRFA for food security and nutrition

The storage, sharing, characterisation and analysis of Digital Sequence Information (DSI) is a critical element of pre-breeding activities leading to the genetic improvement of crop species. Significant R&D investments are made worldwide in this area, to develop crops which are resistant to pests and diseases, more efficient at using nutrients, and better able to withstand extreme weather. This capability is becoming increasingly important in order to develop crops that continue to survive and perform in the increasingly unpredictable climatic extremes that the world is experiencing as a result of climate change.

Furthermore, part II of the Treaty outlines a number of specific activities in which DSI can play an important role. For example Article 5.1 paragraph a) mentions taking inventory and assessing the status and degree of variation associated with PGR, whilst paragraphs e) and f) describe the need for characterization, evaluation, and

monitoring the genetic integrity of collections. DSI can be employed to assist with these pursuits.

Recently three of the UK GINs (PCGIN, OREGIN and VeGIN) have begun to consider whether there may also be significant genetic potential within their germplasm collections to address issues related to nutritional density, commodity processing and shelf life of fresh produce.

NCP GB8-016 MYPoW/DSI - Submission of views and other information on “Digital Sequence Information”

Contribution by the Government of the Netherlands

Terminology used in this area

- At this stage ‘DSI’, nor digital sequence data can be defined or framed.
- The term ‘DSI’ is used as a place holder, and reference should be made to the ongoing CBD process, under which a dedicated AHTEG will work on clarifying the concept, scope and possible appropriate terminology. Nevertheless, developments are at such a rapid pace any attempt for reaching agreement on a definition will most likely be outdated before any agreement is reached.
- In our view more precise terminology should be based on terms commonly used in the scientific community in the context of genetic research. “Genetic Sequence Data” (GSD) for instance is a commonly used term, also in the context of the WHO PIP Framework. ‘DSI’ and GSD are not equivalent to a genetic resource, and access to ‘DSI’ is not equivalent to access to genetic resources from which it is generated. ‘DSI’ can be generated, following access to particular genetic resources, and therefore conditions to get access to genetic resources should also include permission to generate and to further use ‘DSI’ or any information related to the genetic resource accessed.
- Access to publicly available ‘DSI’ should not be restricted, and making ‘DSI’ publicly available should be further promoted. Open access to publicly available ‘DSI’ is crucial for research and innovation, and will directly contribute to the conservation and sustainable use of PGRFA.
- ‘DSI’ is often ‘re-used’ and transferred to other users, even more frequently than genetic resources. This particularly applies to ‘DSI’ in the public domain and ‘DSI’ deposited in public databases. This will make tracking and tracing very complicated. In addition, individual sequences are not unique, and its uniqueness cannot be proven.

Actors involved with DSI on PGRFA

- A large diversity of actors is involved in both commercial and non-commercial use, including research, breeding, governmental institutions and civil society organizations. Access to ‘DSI’ should not be subject to administrative measures that will limit research, innovation and knowledge development, or restrict exchange of information. Restricting access will also have negative effects on smaller actors, who directly rely on public data and analysis tools.

The types and extent of uses of DSI on PGRFA, such as: characterization; breeding and genetic improvement; conservation; identification of PGRFA

- The use of ‘DSI’ on PGRFA ranges from characterization and identification of PGRFA to conservation, genetic improvement and innovation. Genetic improvement and development of new varieties is based on the effective use of both genomic and phenotypic data. Biological research benefits from open access to ‘DSI’. Restrictions to access will have negative consequences, and for specific research areas it may be even impossible to continue the work.

The relevance of DSI on PGRFA for food security and nutrition

- The use of ‘DSI’ on PGRFA is already fully integrated in research, innovation and breeding processes.
- Access to ‘DSI’ and related technologies is crucial for any stakeholder and country, in order to reach long term food security objectives, to be able to adapt to climate change, to deal with human health issues, and to contribute to the conservation and sustainable use of biodiversity.

**Submission of views and information on “Digital Sequence Information” in
Plant Genetic Resources for Food and Agriculture
October 8, 2019**

In preparation for the 8th Session of its Governing Body, the International Treaty for Plant Genetic Resources in Food and Agriculture (Plant Treaty) has asked stakeholders for information on digital sequence information (DSI), including: (1) terminology used in this area; and (2) types and extent of uses of DSI on PGRFA in breeding and genetic improvement, identification, characterization and conservation of plant genetic resources for food and agriculture (PGRFA). DivSeek International Network Inc (DIN) is pleased to offer its views on these issues for consideration by the Governing Body of the Plant Treaty.

The DivSeek International Network (DIN) is a global, community-driven initiative that aims to harness the genetic potential of crop diversity using science and technology. DIN facilitates networking, sharing of community standards and best practices in order to enable breeders and researchers to mobilize plant genetic variation to accelerate the rate of crop improvement and furnish food and agricultural products to the growing human population. DIN has 67 Member institutions from 28 different countries, as well as four Observers who participate in DIN meetings but do not have voting rights in the organization (<http://www.divseek.org/partners>).

1. Terminology Used in this Area

DIN suggests use of the term “genetic sequence data” rather than “digital sequence information.”

The “data” generated by DNA sequencing refers to a collection of raw observations about the physical order of the four bases, thymine (T), adenine (A), cytosine (C), and guanine (G) in a molecule of DNA or RNA. When the data are processed, organized, and interpreted in a given context, they become meaningful and useful to users and are referred to as “information.”

“Genetic sequence data” about PGR is abundant and can be accumulated, stored, processed, organized, shared, utilized and interpreted to generate many different kinds of “information.”

2. Uses of Genetic Sequence Data

a. Breeding and genetic improvement.

Advances in whole genome sequencing have transformed modern breeding in the last decade by allowing researchers to more easily quantify diversity, identify genetic variants contributing to agronomically important traits, and develop genomic prediction models.

- Genomic sequence data plays an increasingly important role in genetics as it applies to all aspects of crop breeding, including understanding domestication origins and subsequent radiation of crops, varietal identification, pedigree verification, tracking of germplasm, gene discovery, marker assisted selection, and genomic selection.
- Genomic-breeding approaches have been thoroughly integrated into the breeding pipelines for most large commodity crops and are now considered essential for minor crops, as new sequencing platforms have dramatically reduced the costs.
- Genomic sequence data have created new opportunities to dissect the architecture of complex traits and infer the evolutionary history of genes and genomic regions, making it possible to trace the ancestry of specific genes and alleles back to landrace varieties and/or wild crop relatives. Identifying new traits and alleles in exotic germplasm, and introgressing those alleles into modern varieties can help adapt existing crops to new environments and changing climates. It also allows breeders to begin to harness the wealth of untapped variation that exists in wild and exotic gene pools.
- Lastly, genomic data for crops are being leveraged to build quantitative genetic models that can predict field performance based solely on genetic sequence data. Genomic selection modelling is a promising application for increasing breeding efficiency; it facilitates efforts to combine favorable traits into the same genotype more precisely and with fewer selection cycles.

As the prevalence of genomic sequence data has increased, the number, diversity and complexity of genomic and bioinformatic resources has also increased. These computational resources are essential for managing genomic sequence data, enabling researchers to systematically accumulate, store, process, organize, interpret, share, and utilize the data.

b. Identification, characterization and conservation of PGRFA.

Genomic sequence data are being widely generated for the world's gene bank holdings. DIN initiated a survey of gene bank curators from its membership. From this survey and other input from national gene banks, we documented that genomic characterization is moving from individual experiments on select accessions to more systematic screening of diversity across entire collections. There is a broad consensus that the availability of genomic sequence data aids gene bank managers in identifying, organizing and selecting accessions for users, and that these data increase the efficiency of gene bank management and will enhance utilization of collections in the future.

The utilization of genetic sequence data for gene bank curation activities can be organized into three main categories:

- *More accurate estimates of identity and integrity of an accession.* Genomic sequence data can be efficiently used to identify and discriminate individual accessions. Variation at individual nucleotides can be used to assess levels of fine scale diversity, providing curators with information about duplication, contamination and possible gaps in the collection.
- *More precise estimation of the phylogenetic relationships among and genetic history of accessions.* Understanding the genetic relationships among accessions is vital in gene bank management. Structuring the collection among species at the phylogenetic level and then subdividing the accessions into genetic population clusters has become a routine procedure in many global collections. With the advent of a high-density single nucleotide polymorphisms (SNPs) the clarity of these estimations is greatly increased.
- *More efficient development of core sets, custom reference sets and improved access to diversity.* Subsets of large collections can be identified that contain a representative fraction of the overall diversity. Subsets can either use whole genomic data or can be used to target specific user-defined regions of the genome to meet a user's needs.

c. Basic research in the biological, ecological and environmental sciences.

Genomic sequence data are essential to fundamental research in biology, ecology and environmental science, with important applications in agriculture, medicine, natural resources and energy.

Fundamental research areas driven by the widespread availability of genomic sequence data include gene discovery across biological kingdoms, gene regulation (including epigenetics), gene and genome evolution, pan-genome conservation and diversity, cell, tissue and organismal development, the microbiome, ecophysiological modeling, plasticity of populations and ecosystems. These and related areas of research generate fundamental understanding that impacts the well-being of human populations and the value of PGRFA.

3. Additional Considerations Regarding Genetic Sequence Data

DIN acknowledges that different communities have different requirements for genetic sequence data, data analysis and data management tools, and for data utilization and breeding strategies. DIN supports efforts to meet community-specific needs and objectives. In addition, DIN strongly supports the need for ongoing capacity building, knowledge-exchange, and technology-sharing to augment the ability of all parties to interpret and utilize genetic sequence data in their efforts to accelerate the development of resilient and sustainable agricultural practices. The amount of genetic sequence data available for each crop or species varies considerably, and the process by which communities extract

benefit from these data also differs because they have different needs and objectives. For this reason, capacity building must be needs-based and targeted to the community and crop or species at issue.

A central aim of DIN is to facilitate the development of best practices, shared vocabularies, and community standards to enable collaboration, capacity building, and data integration, once the terms and conditions for access and benefit-sharing are agreed upon. In so doing, DIN aims to ensure that genetic sequence data are not only widely accessible, but that the benefits they provide are universally actionable.