International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 11 | November - 2022 Impact Factor: 7.185 ISSN: 2582-3930

INTELLIGENT GENE BANK MANAGEMENT SYSTEM

Dr.C.K.Gomathy, Assistant Professor, Department of CSE, SCSVMV Deemed to be University, Kanchipuram

Ms.Vangeti Bhavana, Ms.Vummadisetty Sri Jahnavi, Ms.T S Yamini, Ms.Yellamaraju Reddy Priya

UG Scholars, B.E Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa Maha Vidyalaya, Enathur, Kancheepuram, Tamilnadu, India

ABSTRACT:

The composition aims to present the GenBank System for managing the gene bank in the Institute for Plant Genetic coffers in the city of Sadovo (IPGR). The functionality and armature of the system are described. Some aspects of the intelligence of the armature are considered, videlicet functional sidekicks as part of the armature and an ontology for storehouse and processing of data on factory inheritable coffers.

***Keywords*:** GenBank, Gene Bank Ontology, Operative assistance, biological databases, Nucleotide sequence database, Bio informatics.

I. INTRODUCTION:

Bulgaria is one among the richest countries in plant species. Many of those species are disappearing for various reasons – climate change, artificial modification, or adaptations thanks to different climate conditions. One among our main tasks is to preserve and protect these species over time. an outsized number of scientific organizations and institutes study and store plant genetic resources in gene banks under special conditions. one among them is the Institute for Plant Genetic Resources in the town of Sadovo (IPGR), where the most important gene bank in Bulgaria is located. It keeps seeds for over 60,000 species. The processing of data for storage and management of plant genetic resources is performed by using a system developed and designed for the needs of the institute, which is named the GenBank Management System.one among the main tasks of the system is to create descriptions of plant genetic resources in the form of passport data according to the EURISCO standard. Another important mission is to record information about the experiments administered on plant genetic resources and the results obtained therefrom, again in compliance with the EURISCO standard and particularly the part describing the experiments and their results.

The International Plant Genetic Resources Treaty and therefore the Nagoya Protocol, Japan, require that data on the varied genetic resources be available to all stakeholders – institutes, universities, farmers, et al. . This obliged the GenBank System to supply a module, through which these stakeholders could access information on plant genetic resources. For this purpose, the GenBank Ontology was developed and now it's available for free to all users

II. HISTORY OF GENBANK:

GenBank was created in 1979 at the Los Alamos National Laboratory and was called the Los Alamos Sequence Database. it had been renamed GenBank in 1982 and became a public database. During 1989 to 1992, GenBank transitioned to the newly created NCBI, a division of the National Library of drugs (NLM), located on the campus of the US National Institutes of Health (NIH) in Bethesda, MD. GenBank is made and distributed by the NCBI. NCBI began accepting direct submissions to GenBank in 1993.

Since its creation, GenBank has grown at an exponential rate, doubling in size every 18 months. The EMBL was founded in July 1974 on the thought of an intergovernmental treaty of nine European countries plus Israel. it's grown in membership since then; Luxembourg became the 20 th member in 2007, and Australia joined as an associate member in 2008. The EMBL is found in Heidelberg, Germany. An outstation of EMBL is that the ecu Bioinformatics Institute (EBI), located at Hinton, near Cambridge, UK. The EMBL database as a central depository of nucleotide sequence was created in 1981 and was mentioned because the EMBL Data Library. The EMBL Data Library moved to the EBI in 1993, and have become the precursor to the current EMBL-Bank, which is additionally maintained at the EBI. The expression “EMBL-Bank” isn't frequently used. within the literature, the EMBL-Bank is usually referred to as EMBL nucleotide sequence database or EMBL database. during this book, the expression EMBL-Bank are visiting be frequently used. The EMBL-Bank is now a component of the European Nucleotide Archive (ENA), which consists of three main databases: the Sequence Read Archive (SRA), the Trace Archive (these are discussed later), and thus the EMBL-Bank. The ENA is developed and maintained at the EMBL-EBI under the guidance of the International Nucleotide Sequence Database Consortium (INSDC; discussed below). The EMBL-EBI home page is http://www.ebi.ac.uk/. Various databases and tools maintained by EMBL-EBI and made freely available to be used are often accessed using EMBL Services.

DDBJ has been operational since 1986 and it's maintained at the National Institutes of Genetics at Mishima, Japan. DDBJ is that the sole nucleotide-sequence data bank in Asia. The DDBJ home page is http://www.ddbj.nig.ac.jp/. some recent publications discuss many improvements and added features of DDBJ.

The INSDC (http://www.insdc.org/), a collaborative consortium, was initiated between GenBank, EMBL (ENA), and DDBJ to attach these three databases. This collaboration created the International Nucleotide Sequence Database (INSD). For over 30 years, the INSDC has maintained the first nucleotide-sequence database. The INSDC planning board is composed of members of each of the databases' advisory bodies. The INSDC features a policy of providing free and unrestricted access to all or any or any the available data to scientists worldwide.

III.RELATED WORK:

Many countries in Europe maintain gene banks for plant genetic resources. Specialized environments have been developed for maintenance and access to them. Austria offers its National Inventory, where each user can search for plant genetic resources by Familia, genus, species, or common name. The system outputs all eligible samples; each sample is presented with a common characteristic corresponding to the passport data in line with EURISCO and an image. Belgium provides the LIVCOLL database of the National Botanic Garden of Belgium. All plants can be searched by common name or by botanical name. The database contains over 25,000 taxa of cultivated plants.

Croatia possesses its own database of plant genetic resources PGRD. A search is performed on the main passport data since the samples are divided into groups – vegetables, industrial crops, fruits, medicinal crops, and others.

The National Inventory of Plant Genetic Resources in Germany (PGRDEU) provides access to its gene pool through a system, in which samples can be searched in the various gene banks of different institutes in Germany. The search can be done by basic characteristics of the samples according to EURISCO. Each of the above-mentioned systems has a module for searching for plant genetic resources and as a result the passport data of the samples is supplied.

The main disadvantage of the proposed systems is that users must have knowledge in the field of botany to understand the proposed information and they should be familiar with the names of genetic resources and species. The system provides basic information on genetic resources in accordance with the standard for their passport data. The system proposed in the article tries to solve this problem by developing an ontology containing additional evaluation characteristics for the samples – soil type, fruit characteristics, yield, etc., which will facilitate users.

IV.GENBANK SYSTEM FUNCTIONALITY:

The main goal of the GenBank System is to provide a convenient and easy way to store and manage information about plant genetic resources stored in the gene bank in IRGR, Sadovo. The users of the system are the employees of the institute who maintain and research the resources in the gene bank. In addition, all people external to the institute who are interested in the gene pool of the gene bank are allowed to have access to the information about the stored plant genetic resources.

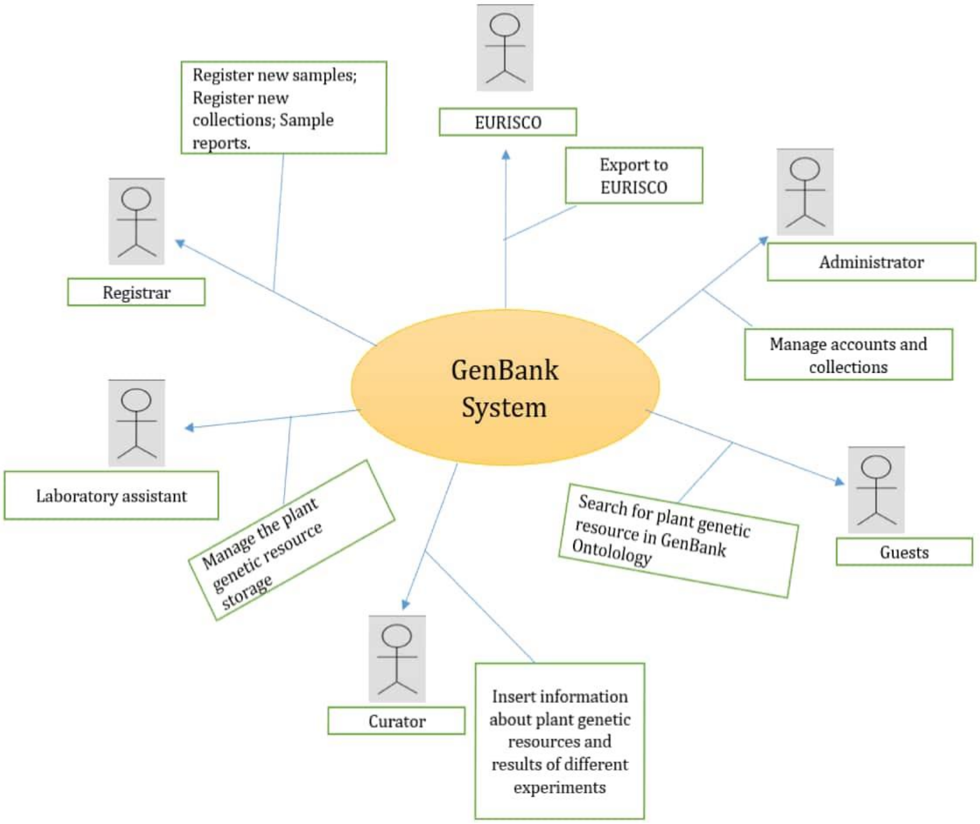

Fig. 1. Context diagram of the GenBank System.

The main task of the employees in the gene bank in IPGR is to register all plant genetic resources received in the gene bank and to manage their storage by using the GenBank System. These plant genetic resources are registered in the gene bank as samples. The sample presents an RGR of a certain genus, species, and reproduction by year. The characteristics of the samples provide the so-called passport data of the sample set in the EURISCO standard. Each sample is enlisted in the gene bank in different collections:

- **Storage collections**, which offer short-term, medium-term, and long-term storage. Each type of storage has its own characteristics including storage temperature and shelf life;

- **Curator's collections** – these are collections in which there are only samples of short-term storage in order to conduct various types of experiments on them. These experiments are carried out by curators; each curator has a collection of specific specimens, for which he/she is responsible. The workflow in the gene bank is performed by the following users of the system: Registrar, Administrator, Laboratory assistant, Curator, EURISCO system, and Guest. Curators conduct experiments with the samples included in their collections and they record and analyze the results. When a new sample arrives at the gene bank, the Registrar registers the sample in the system with its passport data and enrolls it in a specific storage collection as well as in a curator's collection. The laboratory assistants are employees who are responsible for the collections for storage of the samples in the gene bank. For a certain period of time, the samples from the medium and long-term collections are subjected to experiments related to proving the quality of the seeds. In case of poor quality indicators, the main goal of the laboratory assistant is to make sure that the sample is replaced by a new one. This becomes possible by the fact that part of the seeds are used for reproduction. The GenBank System is responsible for informing the laboratory assistant about the expiration of the storage period for each of the samples and notifying him/her about the performance of experiments related to establishing the quality of the stored seeds. All data on the shelf life, seed quality, experiments performed, and their results are stored and analyzed by the laboratory assistant in the system. A requirement for each gene bank is to register its available samples in the EURISCO system. The samples are registered with their passport data and this is the reason why the GenBank System offers data export for each of the samples in the format required by the EURISCO system. External users of the GenBank System are defined as Guests. They have access to the plant genetic resources information only through the ontology developed in the system. In this way, the database will remain available only to the employees of the institute and the ontology will contain data, to which all interested parties are granted access by the institute. The possibility for semantic processing of the ontology and the inclusion of additional information in it, which is understandable for everyone, makes the dissemination of information about the plant genetic resources affordable for everyone.

V.GENE BANK SYSTEM ARCHITECTURE:

The architecture of the GenBank System consists of three main layers

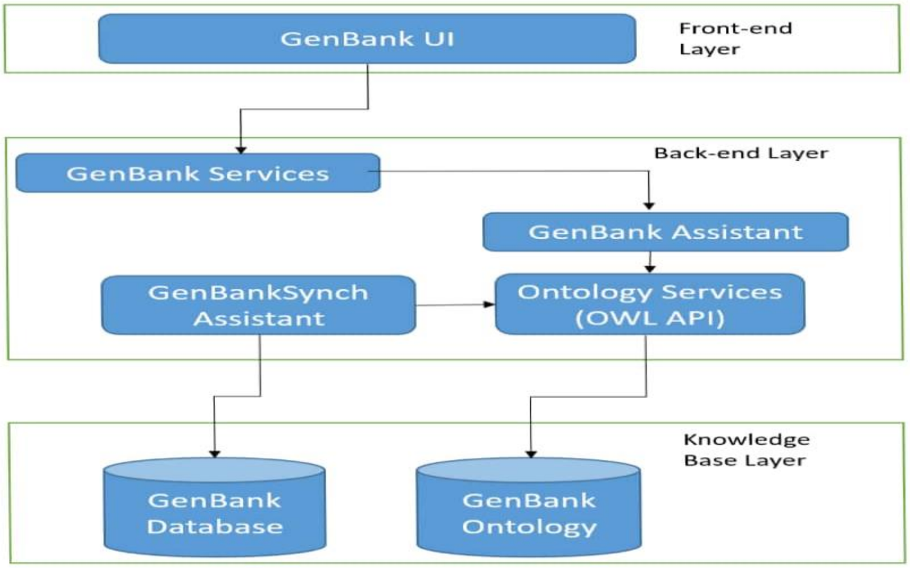

Fig. 2. The GenBank system architecture

The front-end layer is used to visualize information about the plant genetic resources. The interface is organized according to the user access. Information on plant genetic resources is freely available from unregistered users and for this purpose the interface components use the GenBank Ontology.

The levels of access in the architecture are different and they are presented in more detail in

The back-end layer consists of three main components – GenBank-Services, GenBank-Assistant, and GenBank-Synch-Assistant. The GenBank-Services is a component that implements the basic functionalities for storage and processing of samples of the plant genetic resources. The information that is entered and used by employees is recorded in a database – the GenBank Database. The GenBank-Assistant and the GenBank-Synch-Assistant components were developed as operational assistants. The GenBank-Assistant was created to process and provide information on plant genetic resources using the GenBank Ontology for this purpose. Its main functionality is to search for plant genetic resources in the ontology according to specific criteria. The GenBank-Synch-Assistant was developed to synchronize the database and the ontology in the application. For this purpose, this operational assistant monitors the status of the database and when entering a new sample, it makes sure that the information about it is recorded in the ontology. In this way, a correspondence is maintained between the samples registered in the database and those in the ontology. In addition to the state of the database, the GenBank-Synch-Assistant receives messages from the GenBank-Assistant in cases where a sample is not found in the ontology. Upon receiving such a message, it again synchronizes between the database and the ontology, which also ensures consistency between the two sample repositories.

Knowledge base – this layer contains knowledge about the plant genetic resources. Two main components have been developed – GenBank-Database and GenBank-Ontology. The database contains all the information about the plant genetic resources as well as all experiments and results performed on them. Access to the database is limited and external users are not entitled to it. The ontology for plant genetic resources – the GenBank Ontology, is provided for free access. It contains information on all plant genetic resources that can be accessed by any interested party. The specific information, which is the focus of research in the IPGR in Sadovo on plant genetic resources, is not available in the ontology. The GenBank Ontology contains information accessible to everyone and also enables intelligent search by stakeholders. Not every stakeholder has knowledge of the botanical names of the plant species and varieties so the ontology provides a lot of synonyms for their names and many additional features that may be familiar to users. The next section of the article discusses part of the GenBank Ontology.

VI.CONCLUSION:

The GenBank System was developed for the needs of IPGR in the town of Sadovo for management of the plant genetic resources in the national gene bank. The architecture provides the application of intelligent components to facilitate its maintenance and use. The system is designed to be expanded with new features. The ontology as a component of the architecture enables semantic search, which will assist non-specialist users in the field. The ability to synchronize data automatically between the ontology and the database makes it easier for employees to maintain the system. The creation of the GenBank Ontology is time-consuming and requires specialists in the field of plant genetic resources. Experiments with the GenBank System are in progress and we are trying to offer a convenient and flexible interface to the users to take advantage of it.

VII. REFERENCES:

1. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2010;**38**: D46–D51.
2. Kulikova T, Akhtar R, Aldebert P, Althorpe N, Andersson M, Baldwin A, Bates K, Bhattacharyya S, Bower L, Browne P, et al. EMBL nucleotide sequence database in 2006. Nucleic Acids Res. 2007;**35**: D16–D20.
3. Sugawara H, Ogasawara O, Okubo K, Gojobori T, Tateno Y. DDBJ with new system and face. Nucleic Acids Res. 2008;**36**: D22–D24.
4. Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Edgar R, Federhen S, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2010;**38**: D5–D16.
5. Dr.C K Gomathy et al,Web Service Composition In A Digitalized Health Care Environment For Effective Communications, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, April 2016, ISSN: 2278 – 1323.
6. Vishnupriya C K and et al,Dimensional and Morphologic Variations of palatal Rugae-a hospital based study among Chennai populations, International Journal Of Science Research, ISSN No: 2277-8179 Volume 7, Issue 7,P.No-19-20,July ’2018
7. Dr.C K Gomathy et al, Machine Learning-Based Clinical Decision Support System, International Journal of Scientific Research in Engineering and Management (IJSREM) Volume: 06 Issue: 10 | October - 2022 Impact Factor: 7.185 ISSN: 2582-3930
8. Dr.C K Gomathy et al,A Review On IOT Based Covid-19 Patient Health Monitor In Quarantine, International Research Journal of Engineering and Technology (IRJET),e-ISSN: 2395-0056 Volume: 08 Issue: 09 | Sep 2021 www.irjet.net p-ISSN: 2395-0072`
9. Dr.C K Gomathy, et al, A Medical Information Security Using Cryptosystem For Wireless Sensor Networks, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) E-Issn: 2395-5325 Volume3, Issue 4, P.No-1-5,April ’2017

10. Dr.C K Gomathy and et al, The Parkinson's Disease Detection Using Machine Learning Techniques, International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 10 | Oct 2021, e-ISSN: 2395-0056, p-ISSN: 2395-0072.

11. Dr.C K Gomathy and et.al, The Prediction of Disease Using Machine Learning, International Journal of Scientific Research in Engineering and Management (IJSREM) Volume: 05 Issue: 10 | Oct- 2021, ISSN: 2582-3930.

12. Dr.C K Gomathy and et.al, The Smart Stick Assistant For Visually Challenged People Using Ai Image Recognition, International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 9 | Sep 2021, e-ISSN: 2395-0056, p-ISSN: 2395-0072.

AUTHOR'S PROFILE:

Ms. V. Bhavana, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa MahaVidyalaya Enathur, Kanchipuram, India. Her area of interest: database management.

Ms. V. Sri Jahnavi, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa MahaVidyalaya Enathur, Kanchipuram, India. Her area of interest: Front-end Development

Ms. Y. Reddy Priya, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa MahaVidyalaya Enathur, Kanchipuram, India. Her area of interest: Python

Ms. T.S.Yamini, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa MahaVidyalaya Enathur, Kanchipuram, India. Her area of interest: Python

Dr. C.K. Gomathy is Assistant Professor in Computer Science and Engineering at Sri Chandrasekharendra Saraswathi Viswa MahaVidyalaya, Enathur, Kanchipuram, India. Her area of interest in Software Engineering, Web Services, Knowledge Management