OPEN ACCESS at journalijcir.com

Research Article

ISSN: 2395 -5775

HERBARIUM DATA MANAGEMENT SYSTEM (HDMS): SOFTWARE FOR PROVIDING WEB BASED METHODS FOR VALUE ADDITION IN HERBARIUM SPECIMEN

Uma Singh., Suchit A. John*., JyotiVerma., Shukla B.M and Devranjan and Sudershan Kumar

Abstract

Plant collections and preservation of herbarium specimen in a herbarium is a major activity of taxonomic research. HDMS software is developed using PHP for digitization of herbarium specimen data containing label information, plant name, collectors, locality, date of collections and accession number etc. HDMS is unique and has provision to create data value additions to herbarium label data and these include recent taxonomic nomenclature, geographical data, including GPS data, distribution maps, plant descriptions including scan images of specimens, images & video of live plant, genomic information etc. HDMS has been used to digitize 100000 collections of national herbarium of CSIR- NBRI-LWG.

Key Words: Biodiversity informatics, Herbarium, Databases, Digitization, Software.

INTRODUCTION

Plant collection is a major activity of taxonomists which they have done in the past and they do even today. Plants specimens are made after plant collection and are key to taxonomic and anatomical work, and collecting gene sequences obtained from specimens is central to present-day biological research. Type specimens are the most important herbarium specimen in any herbarium collection. It is that specimen of plant upon which the botanist who names the species and publishes its description (Bridson& Forman, 1998). Herbarium specimens are used in several ways viz. the first, herbaria are essential archives for documenting biodiversity; and hence are use only source which provide frequency and occurrence of a plant species (Joppa *et al*., 2011), the second, herbaria provides information on environmental change. Flowering plants are usually collected when they are in flowering because flowers help in identification of the species. For example if a flower is collected from a locality in a month, and a herbarium sheet of the same species, also in flower in the same area, but was collected 100 years ago in a month, then the data analysis may result as an evidence for impact of climate change (Primack *et al*., 2004) and the third, entomologists used herbarium specimens to find out when a particular species first invaded an area (Zangerl & Berenbaum, 2005). Recently, herbaria are also under focus among molecular biologists. Many plant specimens contain DNA that may be used in genetic studies. About 200 years old herbarium specimens have provided DNA, which could be sequenced (Andreasen *et al*., 2009). In addition to above, now a days, researchers are systematically preserving plant samples for use in sequencing.

Digitization of herbarium collections: It is becoming more common for both sequence data and organism specimen data to be stored electronically and accessed via the Internet which allows for interesting comparative work that would have been difficult, if not impossible, in the past. Having data accessible online means that it is available not only to researchers but to students and general public as well at global level. Many resources are available for dealing with gene sequence data; however, the focus here is on herbaria, collections of preserved plant specimens, which were first created in the 16th century, often became less important in the 20th century, and are regaining important , in part because of efforts to digitize these collections. On global level, the Global Biodiversity Information Facility (GBIF) has created a portal where almost 782 million records about species of all kinds are available electronically (GBIF, 2016).

A perusal of study of GBIF and other specimen databases reveal very large amount of data from India India contributed by outside herbaria. In addition to global databases a few Indian Herbaria have digitized their specimens for example database of IIM herbarium contain data on herbarium specimens label similary herbarium of TBGRI has digitized label data only for few sample specimens and so far value additions have not been done in India herbaria. In the present paper a new software tool for digitization of herbarium specimens has been developed and used for creation of 100000 specimens of National Herbarium of the National Botanical Research Institute and the HDMS software is unique and

[1]Department of Biological Sciences, SHUATS, Allahabad-211007
[2]Software Engineer, Department of Computer Science & Engineering, Indian Institute of Technology (IIT), Kanpur
[3]Bioinformatics Division, CSIR-National Botanical Research Institute Lucknow-226001.
Correspondence and Reprint Requests: Suchit A. John

Received: April,20 2017 | **Accepted:** May,18 2017 | **Published Online:** June,28 2017

Conflict of interest: None declared | **Source of funding:** Nil

allows value additions to the herbarium specimens data and hence immensely useful. HDMS and its usage in value additions are described in the present paper.

MATERIAL AND METHODS

Herbarium specimen contains information on its label on parameters including Family,Genus, Species, Author, Date of collection, Collector, Field No., Notes. In the present paper value additions to label data is done with respect to various aspects as prescribed in the Table1.

Table 1 Herbarium label data, descriptors and other parameters of value additions included in the database

S.No.	Data items	Description
1	**Label Data**	
1.1	Name of the plant	Family name Genus Species Author
1.5	Plant collector	Name of plant collector
1.6	Locality	Locality
1.7	Date of collection	Date
1.8	Determinative	Name of determinative
2.	**Value additions**	
2.1	Vernacular Name	Common name, Language
2.2	Life form	Tree, shrub, subshrub, herbs, and vines
2.3	Life span	Annual, biennial, and perennial
2.4	Geographical Distribution	Country, State, District, Location
2.5	GIS Information	Longitude, Degree, Minute, Second
2.6	GPS Data	Map
2.7	Descriptors	Root, Stem, Leaf, Flower, Inflorescence, Fruit, Seed
2.8	Flowering Time	Flowering Time
2.9	Images	Scan image, Still Image (.jpeg), Video (mp4)

Data Dictionary: Data Dictionary is created for recording data on the important parameters covered in present paper are presented in Table 2. Data dictionary includes details of data items required for digitization of herbarium specimen data and it enables users to understand the definitions and scope of various data items.

Data structure: Data structure were developed and are presented in Table 2.

Maria DB has been used to create database. Maria DB (My SQL replacement) database has several features which make it popular worldwide and these include a) It is an open source, flexible and supports Linux, UNIX & Windows operating system, b) it is reliable & guarantees around the clock uptime, c) It offers one of the most powerful transactional database engines, d) It provides high performance, e) It is suitable for high traffic websites, f) It offers exceptional security features that ensure absolute data protection i.e. only authorized users can enter data into databases server, g) it offers comprehensive support for stored procedure, triggers, functions, views, cursors, SQL and more, h) It has self-managements features like automatic space expansion; auto-restart & dynamic configuration changes etc. and these features provide quick start capability.

Table 2 Data Structure

Field	Length
1. NOMENCLATURE	
1.1 Accepted Name	
Domain	Char 30
Author	Char 30
Super Kingdom	Char 30
Author	Char 30
Kingdom	Char 30
Author	Char 30
Sub Kingdom	Char 30
Author	Char 30
Phylum	Char 30
Author	Char 30
Super Class	Char 30
Author	Char 30
Class	Char 30
Author	Char 30
Sub- Class	Char 30
Author	Char 30
Super Order	Char 30
Author	Char 30
Order	Char 30
Author	Char 30
Sub-Order	Char 30
Author	Char 30
Super Family	Char 30
Author	Char 30
Family	Char 30
Author	Char 30
Sub-Family	Char 30
Author	Char 30
Genus	Char 30
Author	Char 30
Species	Char 30
Author	Char 30
Subspecies	Char 30
Author	Char 40
Variety	Char 30
Author	Char 40
1.2 Synonyms Name	
Domain	Char 30
Author	Char 30
Super Kingdom	Char 30
Author	Char 30
Kingdom	Char 30
Author	Char 30
Sub Kingdom	Char 30
Author	Char 30
Phylum	Char 30
Author	Char 30
Super Class	Char 30
Author	Char 30
Class	Char 30
Author	Char 30
Sub- Class	Char 30
Author	Char 30
Super Order	Char 30
Author	Char 30
Order	Char 30
Author	Char 30
Sub-Order	Char 30
Author	Char 30
Super Family	Char 30
Author	Char 30
Family	Char 30
Author	Char 30
Sub-Family	Char 30
Author	Char 30
Genus	Char 30
Author	Char 30
Species	Char 30
Author	Char 30

Subspecies	Char 30
Author	Char 40
Variety	Char 30
Author	Char 40
1.3 Common Name	
Common Name	Char 30
Language	Char 30
2.Geographical Distribution	
2.1 Occurrence	
Country	Char 25
State	Char 25
District	Char 25
Location	Char 25
2.1.2 Altitude	Char 8
2.1.2 Latitude	Text 1
Degree	d 8
Minutes	d 8
Second	d 8
2.1.3 Longitude	Text 1
Degree	d 8
Minutes	d 8
Second	d 8
2.2 G.P.S. Data	
2.3 Map Data	.jpg path
2.4 Flowering Time	
Year	d 8
Month	Char 20
Week	Text 3
2.5 Habit	
Herb	Text 35
Shrub	Text 35
Tree	Text 35
2.6 Habitat	
pH	Text 2
Temperature	d 8
Moisture (Soil)	Text 5
Light Conditions	Text 3
Mineral Nutrition	Text 10
2.7 Soil Type	Char 25
3. DESCRIPTORS	
3.1 Root	
a. Root parts	
b. Root cap	Y/N
c. Root hair	Y/N
d. Secondary root	Y/N
e. Root types	
Adventitious	Y/N
a. Primary	Y/N
b. Secondary	Y/N
c. Root structural types	Char 25
d. Image	.jpg path
3.2Stem	
a. Stem parts	
Bark	Char 25
Pith	Char 25
Wood	Char 25
b. Stem types	Char 40
c. Stem structural types	Char 25
d. Image	.jpg path
3.3 Buds	
a. Buds parts	Char 25
b. Bud types	Char 40
c. Bud structural types	Char 25
d. Image	.jpg path
3.4 Leaf	
a. Leaf parts	Char 50
b. Leaf types	Char 50
c. Leaf structural types	Char 25
d. Petiole and Petiolule structural types	Char 40
e. Stipule and Stipel types	Char 40
d. Image	.jpg path
3.5Inflorescence	
a. Inflorescence parts	Char 25
b. Inflorescence types	Char 50
c. Image	.jpg path
3.6Flower	
a. Flower parts	Char 50
b. Flower types	Char 30
c. Image	.jpg path
3.7Fruit	
a. Fruit parts	Char 30
b. Fruit structural types	Char 50
c. Image	.jpg path
3.8 Seed	
a. Seed parts	Char 25
b. Seed types	Char 40
c. Embryo parts	Char 25
d. Embryo types	Char 25
e. Image	.jpg path
4. IMAGE	
Photo	.jpg path
Video	.dat, .mp3, .avi path
Still Image	.jpg path

PHP has several features which include i) PHP codes are executed faster than ASP because PHP codes runs in its own memory space while ASP codes in use head server, ii) PHP tools are open sources, iii) PHP require Linux Server for hosting while ASP codes needs IIS installed on windows server & thus hosting with PHP is less expensive, iv) PHP had features to connect server databases used in My SQL, however, for using ASP, MS-SQL should be purchased, and v) PHP codes can be kept as encrypted using Ion Cube encryption software.

Herbarium label contains very minimum information and hence provision for value additions to label data was made and these are discussed below:

Nomenclature updating: HDMS software allows creation of data in respect of recent nomenclature of species and the schematic diagram presented in Figure 1 A sample input screen showing data creation of herbarium specimen label data is presented in Figure 2.

Geography: Herbarium label data contains very crude geographic information on label data and it is mostly restricted to the name of locality. HDMS software has provision to update geography data of locality with its corresponding district, state and country. It allows entry of GPS coordinates and schematic diagram is presented in Figure 3 An input screen as generated using HDMS software is presented in Figure 4.

Genomic Information: HDMS software allows linkage of genomic information including EST, Protein, Nucleotides etc. and hyper link has been provided to web portal https://www.ncbi.nlm.nih.gov//and provides user instantaneously the associated genomic information.

RESULTS

HDMS Software created during the present study and it is capable of storing master data on various field viz. family, genus, species, subspecies, variety, author, common name, district, state and county and these master files were made to store data once and the permission of modification was allowed only to administrator level. Thus this helps in speedily creation of data online.

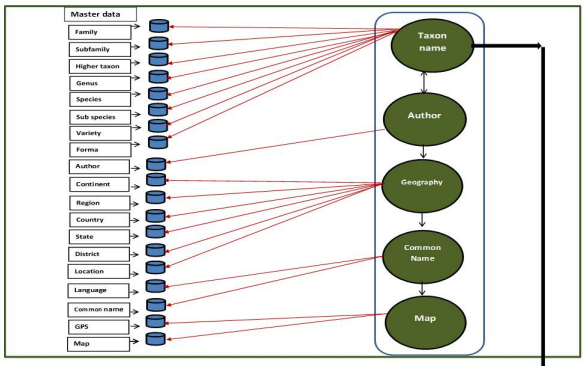

Fig 1: Schematic model show data entry mechanism of label data

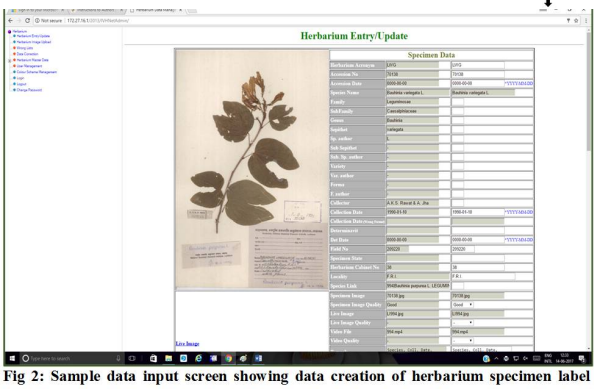

Fig 2: Sample data input screen showing data creation of herbarium specimen label data

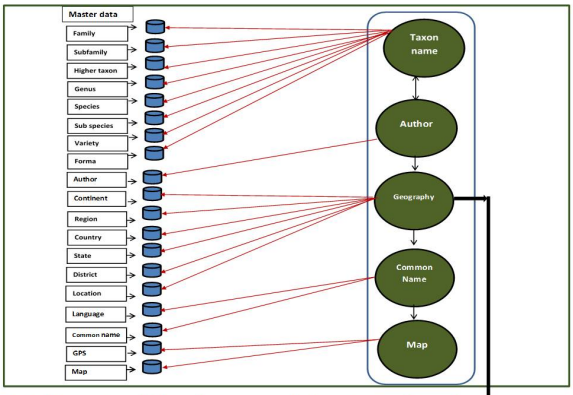

Fig 3: Schematic model show data entry mechanism of nomenclature data

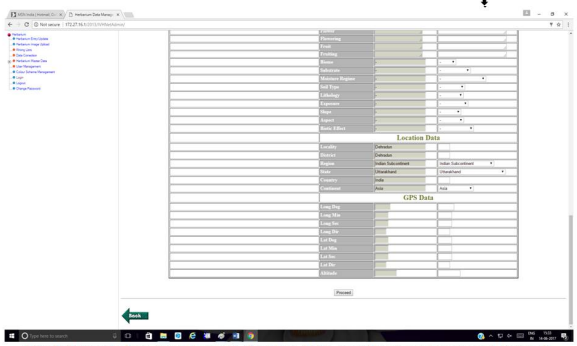

Fig 4: Sample data input screen showing data creation recent nomenclature of herbarium specimen

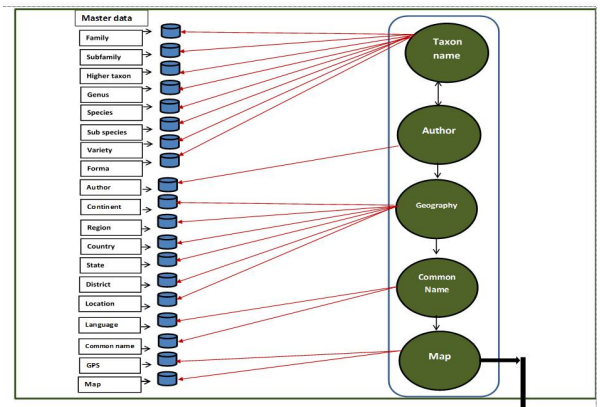

Fig 5: Schematic model show data entry mechanism on geographical data

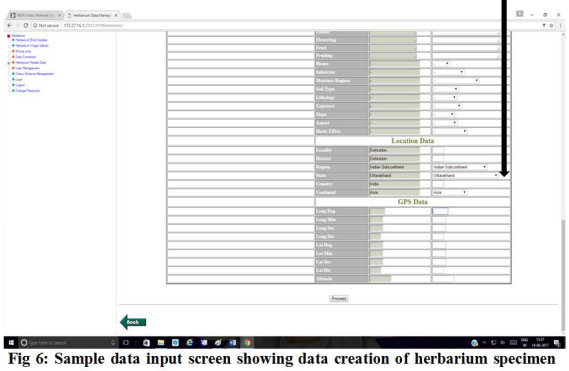

Fig 6: Sample data input screen showing data creation of herbarium specimen geographical data including GPS data

HDMS software allows entry of data available herbarium specimens on taxon name, geography, common name, maps, genomic data, as depicted in schematic diagrams Figure 1 and sample input screen in Figure 2. Herbarium specimen scan images are uploaded in data base and nomenclature directory HDMS upload images and user can use zoom in zoom out feature to view the label data and this ensure authentic data creation

HDMS software has been developed and is capable of creating web enabled database on herbarium specimens of India. It has provision to store information on various aspects, namely, plant name, its locality and data of collection, collector name, field no., field notes etc. HDMS allows value addition to label data with respect to several aspects viz. recent nomenclature, geography, common mane, GPS data, and genomic data and thus this helps in storage of herbarium specimen label data in electronic form. The web enable software can be used by other herbaria of India for digitizing their collection and thus about 2 million collections may be digitized using HDMS.

Acknowledgement

The authors are thankful to Prof. (Dr.) S.K. Barik , Director CSIR-NBRI and Prof. (Dr.) P.W. Ramteke, Dean P.G. studies and HOD, Department of Biological Sciences, SHUATS Allahabad, India.

Reference

1. Diane Bridson and Leonard Forman, (1998) . The Herbarium handbook revised edition. Royal Botanic Gardens, Kew, UK, (1-334) .
2. F.A. Bisby, G.F. Russell and R.J. Pankhurst (1993). Designs for a global plant species information system. Oxford University (1-342)
3. http://www.gbif.org/
4. Kumar S. & Sane, P.V.(2003). Legumes of South Asia-a checklist. Royal Botanic Gardens, Kew, UK, (1-536) .
5. Lucas N. Joppaa.b.c, David L. Robertsb. c, Norman Myersd,1, and Stuart L. Pimme (2011).Biodiversity hotspots house most undiscovered plant species vol. 108 no. 32, 13171-13176.
6. Maura C. Flannery, (2013). Department of Biology, Plant Collections Online: Using Digital Herbaria in Biology Teaching); *Journal of College Biology Teaching*, v39 (3-9).
7. R. Zangerl and May R. Berenbaum(2005). Increase in toxicity of an invasive weed after reassociation with its coevolved herbivore; vol. 102 no. 43

****