

上海农作物种质资源数据库的设计和构建

杨 华¹, 王国军¹, 杨少友², 林 田¹, 魏仕伟¹, 蔡丽娜¹, 王 飞¹
刘鸿艳¹, 罗利军¹, 龙 萍¹

(¹上海市农业生物基因中心, 上海 201106; ²上海协晨信息科技有限公司, 上海 201803)

摘要: 随着农作物种质资源数量的不断增加, 种质资源交流的日益频繁以及种质资源利用的逐步深入, 农作物种质资源数据库的建设对农作物的应用研究和基础性研究至关重要。上海市农业生物基因中心借助于本中心保存的近 21 万份种质资源的信息, 开发了一套上海农作物种质资源数据库, 本文主要阐述了该数据库设计和构建的目的、原则和数据库的架构、运行环境以及主要功能模块。上海农作物种质资源数据库采用了 B/S 结构, 平台系统利用多层逻辑架构实现了对农作物种质资源数据信息的收集、存储、查询、统计分析等功能。外部用户通过浏览器即可访问平台系统, 浏览查询种质资源信息。内部管理员可使用平台的受限功能, 需要认证成功后才可使用后台管理功能, 实现对整个平台数据的管理。注册用户还可通过基因资源数据库提交引种申请, 管理员结合管理信息系统实现对外供种全流程的网络化, 借此可提高工作效率, 增加工作流程的透明度。

关键词: 种质资源; 数据库; 功能模块

Design and Construction of Shanghai Crop Germplasm Resource Database

YANG Hua¹, WANG Guo-jun¹, YANG Shao-you², LIN Tian¹, WEI Shi-wei¹, CAI Li-na¹,
WANG Fei¹, LIU Hong-yan¹, LUO Li-jun¹, LONG Ping¹

(¹Shanghai Agrobiological Gene center, Shanghai 201106;

²Shanghai Xie Chen Mdt InfoTech Ltd, Shanghai 201803)

Abstract: With the increase of crop germplasm resources, the increasingly frequent exchange of germplasm resources and the gradual deepening of the utilization of germplasm resources, the construction of crop germplasm resources database is very important for the application research and basic research of crops. Shanghai Agricultural Biological Gene Center has developed a set of Shanghai crop germplasm resources database, based on the information of nearly 210000 crop germplasm preserved by the Center. This paper mainly describes the purpose and principle of the database design and construction, architecture of the database, running environment and main function modules. The database of Shanghai Crop Germplasm Resources inadopts B/S structure, and the platform system realizes the collection, storage, query, statistical analysis and other functions of crop germplasm resources data using multi-layer logical structure. External users can access the platform system through browser, browse and query germplasm resources information. The internal administrator can use the limited function of the platform and manage the whole platform data, when the authentication is successful. Registered users can also submit the introduction application through the gene resource database, and administrators can carry out network management for the whole process of external supply. This can improve work efficiency and increase the

收稿日期: 2019-06-04 修回日期: 2019-07-03 网络出版日期: 2019-09-09

URL: <http://doi.org/10.13430/j.cnki.jpgr.20190604001>

第一作者研究方向为种质资源的收集、保存和鉴定, E-mail: yh@sagc.org.cn; 王国军为共同第一作者

通信作者: 龙萍, 研究方向为种质资源数据信息, E-mail: lp@sagc.org.cn

基金项目: 上海市农作物种质资源共享服务平台 (18DZ2293700); 国家农作物种质资源共享服务平台 (上海) (NICGR2018-21)

Foundation project: Shared Platform of Crop Germplasm Resources in Shanghai (18DZ2293700), Sharing Platform for National Crop Germplasm Resources (Shanghai) (NICGR2018-21)

transparency of workflow.

Key words: germplasm resource; database; functional module

农作物种质资源是农业科技原始创新、现代农业发展的物质基础,是保障粮食安全、建设生态文明、支撑农业可持续发展的战略性资源^[1]。我国保存农作物种质资源 48 万余份,这些资源是我国农业发展的基础^[2]。科研人员对作物种质资源的研究积累了大量作物种质资源信息。我国从 20 世纪 80 年代开始建立作物种质资源信息系统,已经建成了集数据信息、工具和服务为一体的信息系统,整合了全国的作物种质资源信息并开放共享^[3]。

上海市农业生物基因中心(以下简称中心)是上海市农业基因资源保存与服务中心,致力于基因资源的收集保存、评价创新和分发利用^[4]。中心自 2002 年成立以来,累计收集保存各类农作物种质资源近 21 万份,包括主要的粮食和经济作物,如水稻、小麦、玉米、大麦、油菜、蔬菜和花卉以及微生物资源和遗传工程材料,尤其是在水稻功能基因资源的收集保存方面已形成特色,并向 100 余家单位和个人提供了 8 万余份次各类种质资源。

中心根据自身特点,将种质资源信息数字化、网络化管理^[5],建成上海农作物种质资源数据库,并根据实际需求,不断改进。目前该数据库已实现对农作物种质资源数据信息的收集、存储、查询、统计分析等功能。注册用户还可以通过基因资源数据库提交引种申请,管理员结合上海农作物种质资源库管理信息系统^[6],实现对外供种全流程的网络化,提高了工作流程的透明度和工作效率,减少了相关工作人员的工作量^[7-9]。

1 上海农作物种质资源数据库建设目的及原则

1.1 上海农作物种质资源数据库建设目的

上海农作物种质资源数据库建设基于 5 个目的:(1)保存种质资源性状数据。农作物种质资源类型丰富,性状多样,造成了性状字段属性的复杂性,需要通过数据库来实现数据的安全保存。(2)实现数据检索。用户可根据目标性状,在数据库中检索出需要的种质资源。(3)指导种质库工作。通过数据库对农作物种质资源信息进行有序管理。种质库管理者可根据所保存资源信息缺失情况,及时安排资源的繁殖鉴定工作。(4)获取实物资源。用户可通过登录数据库,进行目标资源的检索,并发送种质申请单。(5)采集利

用信息。通过资源利用者的反馈以及其他方式,采集利用信息,掌握资源利用方向和成效^[10-13]。

1.2 上海农作物种质资源数据库建设原则

上海农作物种质资源数据库从内部管理上需做到规范、安全、可扩充,从外部使用来说,需做到功能简洁、服务高效。因此在设计该数据库时,遵循以下 5 个原则:(1)规范化原则。数据库涉及的相关农作物种质资源性状基本数据,均严格按照国家农作物种质资源性状描述规范,做到与国家信息系统的一致性。对于国家农作物种质资源描述规范不能满足的性状,另外增加字段进行描述。(2)安全可靠原则。数据的网络化对数据的安全性有更高的要求。通过数据的异地备份,保障数据的安全性。(3)功能可扩充原则。优异资源展示需求会随着工作发展,有不同的需求,因此设计时做到优异资源展示可扩充性。(4)简单实用性原则。平台用户涵盖全国科研院所、高校、公司,甚至是农民个体,因此要求数据库界面友好简洁,便于用户的查询和申请服务。(5)服务快速响应原则。用户提出申请后,管理者第一时间得到申请信息,并作出响应。

2 上海农作物种质资源数据库的设计和运行环境

2.1 总体设计

上海农作物种质资源数据库采用 B/S 结构,外部用户通过浏览器即可访问平台系统,浏览查询种质资源信息。平台系统采用多层逻辑架构,由展示层、业务逻辑层、基础组件层、数据层构成,具体逻辑架构如图 1 所示。

展示层:该层主要分为普通外部用户可直接使用的种质资源信息展示和内部管理员使用的后台信息管理两大部分。普通外部用户无须注册即可直接浏览查询平台中的种质资源等信息;内部管理员可以使用平台的受限功能,需要认证成功后才能使用后台管理功能,管理整个平台的数据。

业务逻辑层:该层实现平台的业务功能,业务逻辑层是展示层和数据层之间的桥梁。业务逻辑层将用户提交的数据经过逻辑处理后,更新到数据层存储,或者根据用户的请求,从数据层获取数据,经过逻辑处理后,返回给展示层显示给用户。

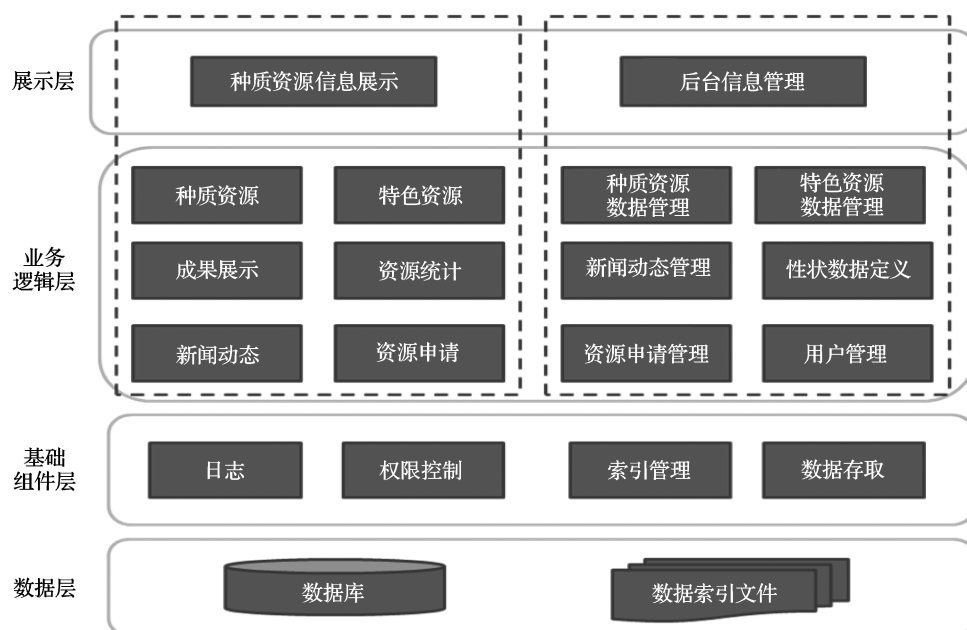


图 1 农作物种质资源数据库分层图

Fig.1 Hierarchical map of crop germplasm resource database

基础组件层: 该层主要提供通用的组件功能, 供业务逻辑层使用。

数据层: 主要用来管理数据的存取, 该层包括数据库和数据索引文件两大部分。数据库主要用于数据的存储和基础数据的查找, 数据索引文件主要用于存储不同类型种质资源的性状描述数据的索引, 便于根据性状描述字段进行数据查找。

2.2 关键设计

本系统在设计上的关键问题是, 由于农作物种质资源类型丰富, 而不同类型种质资源的性状描述字段又有所不同, 按照常规关系型数据库的设计方法, 需要为每种类型的种质资源单独创建一张表, 用于存储该类型种质资源的性状描述数据。常规关系型数据库的设计, 一方面需要预先创建数量众多的、对应于每种类型种质资源的数据库表, 以及开发对应的保存和查询检索功能, 工作量巨大。另一方面如果系统开发完成后, 后续需要在现有类型种质资源上增减调整性状描述数据字段, 或者增加新类型种质资源, 都需要开发人员介入, 手动增加或调整数据库表结构以及对应的功能代码, 缺少足够的灵活性和可扩展性。

针对上述问题的一种解决方案是为每种类型种质资源动态创建数据库表结构, 即在增加新类型的种质资源时, 系统动态创建该类型种质资源的数据库表, 或者在调整种质资源性状描述字段时, 系统动

态调整数据库表结构。但是该方法存在修改表结构时可能会对已保存数据造成损坏的风险, 并且当数据库表中数据量较大时, 调整表结构也可能需要较长时间。

鉴于上述问题, 我们在系统设计中采用了另外一种解决方案。该解决方案是在关系型数据库中使用一张表来保存每种种质资源性状数据定义的元数据信息, 在另外一张表的一个字段中保存种质资源的 JSON 格式的非结构化性状数据, 这样就可以将所有类型种质资源的性状描述数据都统一保存在数据库中。同时, 为了解决非结构化保存的 JSON 格式性状数据不方便通过数据库进行检索查询的问题, 系统通过 Lucene 全文检索引擎构建性状描述字段的索引和原始数据, 当用户进行查询时, 根据数据库表中存储的性状描述定义元数据获取可进行查询的字段, 用户选定查询字段并提供搜索值后, 系统通过 Lucene 全文检索引擎查询索引库, 获取符合条件的种质资源数据, 返回给用户 (图 2)。这样同时保证了数据的安全性、灵活性和可扩展性。

2.3 数据库运行环境

本数据库运行在 WINDOWS SERVER 2008 R2 操作系统上, 采用 JAVA 8 开发, 开发工具 JDK+Eclipse, 运行在 TOMCAT 8.0 下, 数据库采用 MySQL 5.6 版本。

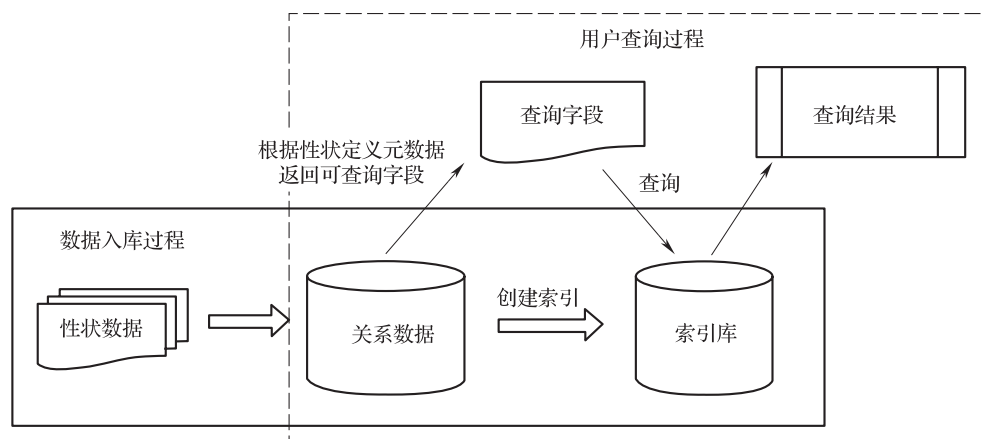


图2 农作物种质资源数据库主要功能

Fig.2 Main function map of crop germplasm resource database

3 数据库的主要功能模块

3.1 外部数据服务

普通用户通过登录上海市农作物种质资源数据库网址 <http://seed.sagc.org.cn/> 进入数据库首页

(图3),并可选择进入最新动态、种质资源、特色资源、成果展示、资源统计、服务指南等几个模块进行浏览,注册用户可以进行资源申请模块申请实物资源。



图3 农作物种质资源数据库工作界面图

Fig.3 Working interface diagram of crop germplasm resource database

3.1.1 新闻动态 用户可以查看到农作物种质资源数据库中最新发布的工作进展、相关新闻、通知等信息。

3.1.2 种质资源 本数据库根据农作物种质资源的类型划分为粮食作物、经济作物、蔬菜、果树、牧草和其他类型。其中粮食作物中分为禾谷类、豆类、根茎

类;经济作物中分为纤维类、油料类;蔬菜作物中分为白菜类、芥菜类、甘蓝类、瓜类、茄果类、根(茎)菜类、绿叶菜类、水生蔬菜类、葱韭类。目前,本数据库保存的种质资源主要是以上三大类。

用户进入需要查询的类别后,选择需要查询的种类,利用默认检索条件进行检索,也可以根据自己

的需要选择检索条件(图4)。用户可以查看某种类别下平台保存的所有种质资源条目以及种质资源的详细信息,并且能够根据常用的性状描述字段快速

查询符合条件的种质资源。本数据库支持用户通过高级查询功能在所有性状描述字段中挑选查询条件进行组合查询种质资源。

The figure displays the search interface of the crop germplasm resources database. At the top, there is a navigation bar with links like '首页' (Home), '最新动态' (Latest News), '种质资源' (Germplasm Resources), '特色资源' (Special Resources), '成果展示' (Achievements), '资源统计' (Resource Statistics), '服务指南' (Service Guide), and '关于我们' (About Us). Below the navigation bar, there is a search bar with a dropdown menu and a search button. The main search area contains several input fields for different search criteria: '种质类型' (Germplasm Type), '库编号' (Library Number), '种质名称' (Germplasm Name), '统一编号' (Unified Number), '田间编号' (Field Number), '种质外文名' (Germplasm English Name), '科名' (Family Name), and '属名' (Genus Name). There are also checkboxes for '更多查询条件' (More Search Conditions) and '查询结果显示字段设置' (Search Results Display Field Settings). At the bottom, there is a '查询' (Search) button.

图4 农作物种质资源数据库检索页面图

Fig.4 Search page diagram of crop germplasm resources database

3.1.3 特色资源 特色资源模块展示了上海市农业生物基因中心收集的特色资源。该模块采用的是可扩展模式,今后随着特色资源的增加,可以不断添加到该界面进行展示。目前收集了上海审定品种 846 份、上海地方品种 783 份、节水抗旱稻资源 21 份、茶梅资源 214 份、生菜资源 756 份、萱草资源 40 份。

3.1.4 资源统计 用户可以查看本数据库保存的各类种质资源的相关统计信息,目前本库已收集各类资源 22.1 万份,整合数据 800 万条。

3.1.5 资源申请 用户需提交申请,成为注册用户,并登录。注册用户检索的资源可以加入种质资源申请单,在线提出引种申请,种质库管理员接到申请后,开展相关服务流程。注册用户还可通过该模块查看到本次引种申请的审批和出库等进度和以往的服务记录。

3.2 内部数据管理

3.2.1 性状描述元数据定义 该模块可以根据种质资源的不同类别,分别定义不同的性状描述字段集合,从而实现支持不同类别的种质资源使用不同的性状描述信息。性状描述字段可以支持常用的文本、数值、日期、选项、图片等类型字段。

3.2.2 种质资源管理 该模块主要用于管理种质资源数据。不同类别的种质资源使用不同的数据字段,因此该模块对每种种质资源分开进行管理。管理员可以通过 Excel 文件上传,增加或更新种质资源的数据。

3.2.3 特色资源管理 该模块管理前台展示的特色资源数据,管理员也可通过 Excel 文件进行增加或更新。

3.2.4 新闻动态管理 管理员通过该模块可以增加或更新新闻动态信息。

3.2.5 资源申请管理 管理员通过该模块可以审批外部用户提出的引种申请,并且更新该申请的处理进度信息。

4 问题和讨论

随着国家对种质资源的重视,种质资源库中农作物种质资源的数量将越来越多,类型将越来越丰富,相关研究也会越来越深入,为实现种质资源库的有效、准确地管理,信息化、数据化和电子化是数据库管理的必然趋势。为了适应这一趋势,本中心也在这方面做了有效尝试。通过多年实际应用来看主

要存在如下问题:(1)数据信息的标准化需要加强。以前对农作物种质资源的繁种,性状的记载不是按照数据信息化要求进行的,主观性描述较多,数据完整性不够,标准化不强。随着作物表型组学的发展,与表型组学发展相对应的仪器更加高通量化、标准化和准确,这对推动种质资源库相关种质资源的信息保存会起到越来越重要的作用。因此,建议相关部门立项,委托专业人员鉴定性状,按照描述标准采集基础数据,扩充到数据库中。(2)从现有农作物种质资源相关数据库功能来看,主要功能还局限于以种质资源数据的存储和基本性状的检索为主。今后数据库可重点开发数据统计与分析的模块。把这些数据有机整合,从中获取感兴趣的信息。如某高产水稻,通过数据库相关数据的分析,能快速得出其高产类型,主要是穗数型还是大粒型,或是高结实率的品种。

参考文献

- [1] 中华人民共和国农业农村部,中华人民共和国国家发展和改革委员会,中华人民共和国科学技术部. 全国农作物种质资源保护与利用中长期发展规划(2015-2030年).(2015-04-29)[2019-06-04]. http://www.moa.gov.cn/nybg/b/2015/si/201711/t20171129_6134098.htm
Ministry of Agriculture of the People's Republic of China, National Development and Reform Commission of the People's Republic of China, People's Republic of China Ministry of Science and Technology. National medium-and long-term development plan for the protection and utilization of crop germplasm resources(2015-2030).(2015-04-29)[2019-06-04] http://www.moa.gov.cn/nybg/b/2015/si/201711/t20171129_6134098.htm
- [2] 刘旭,李立会,黎裕,方涛. 作物种质资源研究回顾与发展趋势. 农学学报,2018(1): 1-6
Liu X, Li L H, Li Y, Fang W. Crop germplasm resources: advances and trends. Journal of Agriculture, 2018(1): 1-6
- [3] 方涛,曹永生. 中国作物种质资源信息系统. 科研信息化技术与应用,2012,3(6): 66-73
Fang W, Cao Y S. Chinese crop germplasm resources information system. E-science Technology & Application, 2012,3(6): 66-73
- [4] 杨华,龙萍,夏辉,陈海荣,林田,罗利军. 上海水稻种质资源的研究与利用进展. 植物遗传资源学报,2015,16(2): 245-249
Yang H, Long P, Xia H, Chen H R, Lin T, Luo L J. The research and utilization progress of rice germplasm resource in Shanghai. Journal of Plant Genetic Resources, 2015,16(2): 245-249
- [5] 龙萍,曹开雄. 上海市农业生物基因资源数据库的构建及其应用. 上海农业学报,2004,20(3): 7-10
Long P, Cao K X. Construction and application of agricultural biological gene resource database in Shanghai. Acta Agriculturae Shanghai, 2004,20(3): 7-10
- [6] 杨华,王国军,朱天生,林田,魏仕伟,石群芳,张前荣,王飞,刘鸿艳,罗利军,龙萍. 上海农作物种质资源库管理信息系统建设. 植物遗传资源学报,2019,20(2): 459-465
Yang H, Wang G J, Zhu T S, Lin T, Wei S W, Shi Q F, Zhang Q R, Wang F, Liu H Y, Luo L J, Long P. Construction of Shanghai crop germplasm resources management information system. Journal of Plant Genetic Resources, 2019,20(2): 459-465
- [7] 杨欣,颜伟,许大光,蔡仕宾,王才林. 江苏省农业种质资源库管系统设计与构建. 农业网络信息,2010(11): 44-47
Yang X, Yan W, Xu D G, Cai S B, Wang C L. Design and development of the Jiangsu agricultural germplasm information system. Agriculture Network Information, 2010(11): 44-47
- [8] 杨文娟,高媛,魏鑫,王林海,黎冬华,周璐,游均,张秀荣,于景印,张艳欣. 芝麻种质资源信息数据库的设计与构建. 中国油料作物学报,2018,40(1): 57-63
Yang W J, Gao Y, Wei X, Wang L H, Li D H, Zhou R, You J, Zhang X R, Yu J Y, Zhang Y X. Design and establishment of information database of sesame (*Sesamum indicum* L.) germplasm resources. Chinese Journal of Oil Crop Sciences, 2018,40(1): 57-63
- [9] 叶锡君,孙敬,张天真. 农作物特种遗传资源共享平台的建立. 南京农业大学学报,2011,34(6): 7-12
Ye X J, Sun J, Zhang T Z. Genetic resources sharing platform construction of crops. Journal of Nanjing Agricultural University, 2011,34(6): 7-12
- [10] 曹永生,方涛. 国家农作物种质资源平台的建立和应用. 生物多样性,2010,18(5): 454-460
Cao Y S, Fang W. Establishment and application of national crop germplasm resources infrastructure in China. Biodiversity Science, 2010,18(5): 454-460
- [11] 曹永生,陈育,孔繁胜. 中国作物种质资源信息共享网络的建立. 资源科学,2001,23(1): 46-48
Cao Y S, Chen Y, Kong F S. Establishment of Chinese crop germplasm information sharing network. Resources Science, 2001,23(1): 46-48
- [12] 栾汝朋,孙素芬,张峻峰,于峰. 北京农作物种质资源信息服务平台的设计与构建. 中国农学通报,2010,26(20): 383-387
Luan R P, Sun S F, Zhang J F, Yu F. Design and construction of Beijing crop germplasm information platform. Chinese Agricultural Science Bulletin, 2010,26(20): 383-387
- [13] Kaplan K. International partnership to develop a global plant genebank information system. (2008-02-11)[2019-06-04]. <http://www.ars.usda.gov/is/pr/2008/080211.2.htm>