全球专利统计数据库（PATSTAT）研究述评*

张静[1]，杨冠灿[1]，刘会景[2]
（1. 中国科学技术信息研究所，北京 100038；2. 北京万方数据股份有限公司，北京 100038）

摘要：从数据库的构成、信息组织架构、技术唯一标识符的设立等方面对PATSTAT数据库进行研究，基于此分析其特点，并从功能实现的角度将其与Orbit、Thomson Innovation等主流商业专利数据库进行比较，总结PATSTAT数据库在数据集成方面的经验。

关键词：PATSTAT数据库；专利数据；数据集成；语义异构
中图分类号：G255
DOI：10.3772/j.issn.1673-2286.2015.12.011

长期以来，专利数据被广泛应用于科技评价活动中。专利数据由于涵盖的信息全面、规范且易于使用，受到了理论界与学术界的广泛青睐[1]。具体而言，专利数据包括了经济、法律以及技术信息，其数据的丰富程度往往是其他技术信息数据来源所不具备的[2]；在有关国际组织的共同努力下，围绕专利数据陆续出台了一系列的标准，使得专利数据较其他数据更为规范与准确[3]。近年来，在数据提供商以及各国专利管理机构的共同努力下，一大批专利数据库被开发出来，既包括免费的，也包括商用的，使得人们可以便利地使用专利数据资源[4]。

全球专利统计数据库（Worldwide Patent Statistical Database，以下简称PATSTAT）是由欧洲专利局创建的以欧洲专利局专利文献主数据库（EPO Master Documentation Database，DOCDB）为主要数据源的快照数据库，收录了全球100多个国家或组织的专利信息，其内容涵盖专利题录数据、引文数据以及专利家族链接。PATSTAT旨在为研究者提供可完全运行于个人电脑的面向统计分析的专利数据库。PATSTAT自2007年向公众发布以来，由于其面向统计分析、数据遵循统一规范、数据开放等特点，在学界得到广泛应用。

1 PATSTAT相关背景

PATSTAT数据构建的最初设想来源于OECD的专利统计专题工作组（Task Force），其成员包括OECD、EPO、JPO、USPTO、WIPO等机构，该组织通过加强成员间的协作关系来促进专利统计质量的提升，尤其是在如下三个方面：①促进专利指标的丰富与标准化；②使专利分析人员能充分获取专利数据；③为基于专利数据的统计决策工作做出贡献。鉴于现有专利数据库之间差异很大，存在数据质量不一，数据加工过程不透明，数据检索、导入、导出受限等问题，不能满足专利数据分析与研究的需求，在OECD的专利统计专题工作组的倡导下，由EPO创建了全新的面向统计决策的PATSTAT。

上述机构的共同努力使得PATSTAT数据库在统计决策的理论准备方面较为坚实。最为显著的成果是2009年OECD出版的《专利统计手册》[1]，汇集专利统计分析领域的研究成果，为使用专利数据来测度科技创新活动提供了指导准则，对专利统计分析的内容、相应指标以及使用的范围进行了系统的阐述，成为了PATSTAT数据库建设的重要理论来源。其次，在工作组的领导下，一批针对专利统计决策分析的研究报告也相继产生，这些报告通过梳理、验证相关理论，构建数据集市和统计测量指标为PATSTAT数据库的不断提升提供有力的支持（见表1）。

2 PATSTAT数据库构成与信息组织

2.1 PATSTAT数据库的构成

PATSTAT数据库由四部分构成：PATSTAT源数

* 本研究得到国家科技支撑计划课题“专利信息资源整合与加工关键技术与规范研究”（编号：2013BAH21B01）和国家自然科学基金青年基金“基于指数随机图模型的专利引用关系形成影响因素及机理研究”（编号：71403256）资助。

表1 OECD的专利统计专题工作组为PATSTAT提供的理论及实践支持

序号	研究报告名称	理论价值	数据集
1	测量专利质量：技术与经济价值指标（Measuring Patent Quality: Indicators of Technological and Economic Value[5]）	梳理了国际通用的专利质量评价指标	形成了一套专利质量评价指标体系
2	专利权人、申请人地址信息数据（The OECD REGPAT Database: A Presentation[6]）	对来源EPO和WIPO专利中的发明人、申请人地址信息进行清洗	形成了OECD REGPAT数据库
3	专利家族：定义、范围；三方专利家族（Insight into Different Types of Patent Families[7]; Triadic Patent Families Methodology[8]）	梳理了专利家族的定义、范围以及界定了三方专利家族范围	OECD 三方专利家族数据库
4	统一专利权人名称（OECD Harmonised Applicants' Names database）	以商业数据库中的人名信息为基础构建词表，对专利申请人名称进行清洗	OECD HAN 数据库
5	分析欧洲和国际专利引文（Analysing European and International Patent Citations: A set of EPO Patent Database Building Blocks[8]）	探索建立能够协调多源引文数据的引文数据集市	OECD 引文数据库

据（PATSTAT raw data）、PATSTAT法律事件数据（Legal event data for PATSTAT）、PATSTAT的专利登记信息数据（EP register data for PATSTAT）以及PATSTAT在线数据库（PATSTAT Online Extension）[9]。

PATSTAT源数据是PATSTAT的核心，通常所指PATSTAT数据库也是指的这一部分，该数据主要是从EPO的主著录项数据库（也称DOCDB数据库）中获取的，主要包含的是与专利有关的著录项信息。PATSTAT法律事件数据则主要来源于EPO的全球法律状态数据库（也称INPADOC数据库），主要包含的是专利生命周期过程中发生的法律事件信息，例如审查、续费、失效、权属转移、PCT进入国家阶段以及异议和诉讼信息等。目前，PATSTAT数据加工团队已经设计了整体的专利数据框架，将上述两部分信息进行了有效的组织，可供研究人员进行综合分析，但PATSTAT法律事件数据的数据还需要单独收取费用。

PATSTAT专利登记信息数据是EPO于2013年4月发布的一款新产品，该数据的主要来源为欧专局的专利登记信息数据库（EP register data），包含了在EPO登记的专利（或者通过PCT流程进入到EPO的专利）的著录项信息、法律事件信息以及流程信息，在数据的准确性、详细程度上面具有优势。目前，该部分数据还处于快熟发展期，现有数据库是独立的，并没有整合到之前提及的PATSTAT源数据与法律事件数据的框架中来。

PATSTAT在线数据库是PATSTAT数据库（源数据和法律事件数据）的在线版，EPO通过一个基于SQL检索的平台将PATSTAT数据进行展现。用户可以直接利用SQL检索式对PATSTAT数据进行检索，同时，PATSTAT在线数据库还包含了一些源数据和法律事件数据中不包含的数据，这些数据多是需要通过对来源PATSTAT数据进行初步计算获得的数据（见图1）。

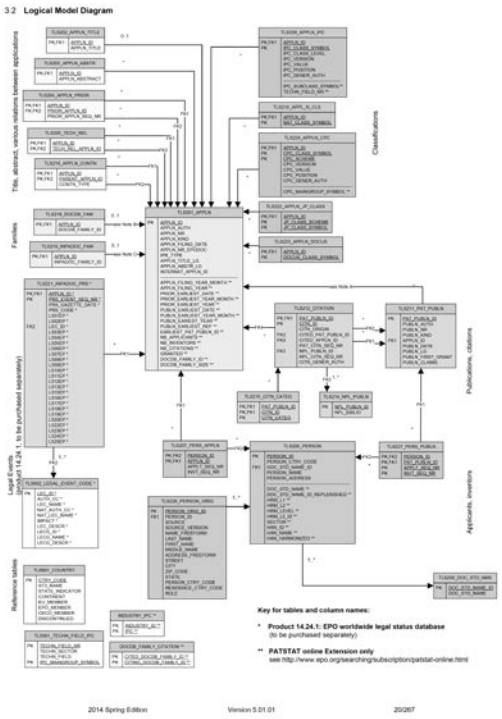

图1.a 源数据及法律事件数据的物理数据模型

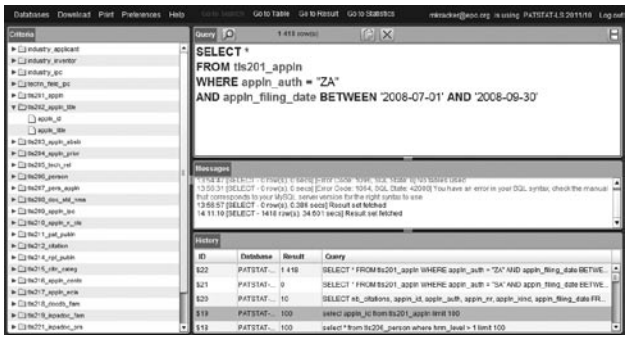

图1.b PATSTAT在线数据检索与分析平台

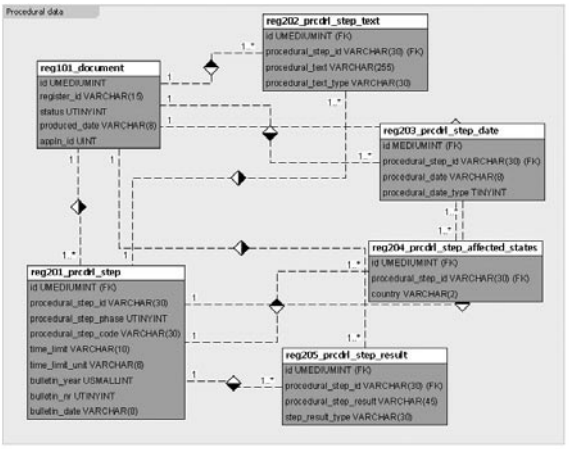

图1.c 登记信息数据的物理数据模型（部分）

图1 PATSTAT信息组织图

2.2 PATSTAT的信息组织架构

PATSTAT数据范围十分广泛，大致可分为六类：号码信息（专利申请号码、专利公开公告号码、专利优先权号码、专利家族号码、参考专利号码），技术信息（主要涉及的是专利著录项信息中的技术相关信息，包括技术分类信息、标题信息、摘要信息等），法律信息（与专利相关的信息，包括授权、缴费、转移、延续、撤销等），“人”的信息（专利相关人员的名称信息，包括申请人、发明人、专利权人、审查员、代理机构的名称信息等），时间信息（与专利生命周期阶段相关的时间信息，包括申请日、授权日、失效日等），地址信息（与专利申请、授权相关的国家以及“人”的地址信息，包括专利申请所在地、专利权人地址等）。这六大类的信息之间存在一定的差异性，即便是同一类别信息之间也存在较大差异。例如，在号码信息下，申请号、公开公告号、优先权号以及家族号就存在较大的差异，如何对上述信息进行科学的组织是PATSTAT需要解决的首要问题。

PATSTAT数据库采用了以专利申请为中心的信息组织方式。从PATSTAT的信息组织模型框架（图2）中可见，专利申请信息在整个关系型数据库中居于核心位置，所有的相关信息都与专利申请信息进行关联。本文认为：PATSTAT之所以建立以专利申请为中心的信息组织方式，主要是基于如下考虑：①从整个专利生命周期视角来考虑，专利申请是整个专利活动的逻辑起点，因此，以专利申请为中心组织整个专利信息，就是从源头上抓住了专利信息；②从信息组织视角来考虑，专利的唯一性原则规定一项专利仅具有一个专利号码信息[10]，而其他信息如公开公告号码、家族号码都有可能存在一对多的状况，因此，以专利申请信息作为信息组织方式也会避免因为号码信息不唯一所带来的对专利信息的歧义；③相对于以家族为中心的组织方式而言，以申请为中心的专利信息组织方法能更便捷地与技术信息、法律状态信息、过程信息相联系，更适合于多维度的专利统计分析。

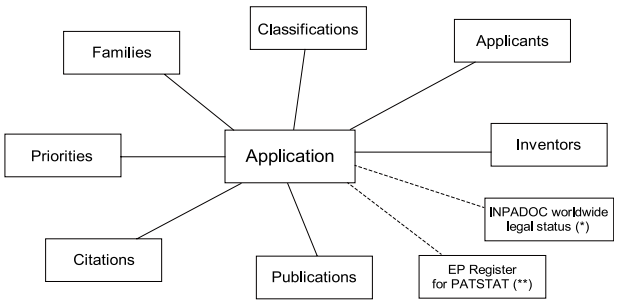

图2 PATSTAT的信息组织模型框架

2.3 PATSTAT的技术标识符的设立

为了确保专利数据库符合参照完整性约束以及提高检索效率的需求，数据库建设过程中通常会统一赋予一个自增的虚拟唯一标识符。从技术的角度而言，这种建立唯一标识符的做法一方面满足了数据的完整性约束要求，确保了数据的唯一性；另外，由于虚拟标识符能够避免在专利申请识别过程中同时需要识别三个实体（专利申请号码通常由国别代码、申请号、辅助属性构成）的弊端，也会降低数据库的存储规模，使数据库具有更高的检索效率。PATSTAT数据库在设立技术标识符的过程中，选择采用申请号作为设立虚拟技术标识符的依据，将全部的数据整合到统一的规则之下，为后续物理建模奠定基础。通过设立专利标识符，PATSTAT将不同来源的专利信息有机地关联起来，而

设立的依据是以专利申请为整个专利生命周期的逻辑起点，通过数据表的关联关系将整个专利生命周期全过程中的信息贯穿起来。

单纯的采用自增的技术标识符往往使得处于不同时点的不同版本之间的PATSTAT数据缺乏关联，同时，PATSTAT数据与其他EPO专利数据源之间也缺乏了对应的关联关系。针对这个问题，PATSTAT选择以DOCDB的技术标识符作为其技术标识符的主要来源（情形1）。当数据出现由于优先权号码缺失导致的技术标识符不一致（情形2），由于引用的专利公开公告号缺失导致的技术标识符不一致（情形3），由于引用的专利申请号缺失所导致的技术标识符不一致（情形4）这三种情形时，才采用自增的技术标识符。上述做法既保证了整个数据仓库符合完整性约束，也保证了EPO的各数据库之间，以及不同时点的数据仓库之间能够实现数据直接关联。

如前所述，为了实现参照完整性约束的要求，PATSTAT在设立技术标识符时需要同时考虑四种不同的情形，因此，PATSTAT设计了一套技术标识符分配方法，其特点在于：能够同时兼容两种技术标识符（相对固定技术标识符以及自增的技术标识符），同时通过号码分配可以将属于不同情形的技术标识符区分开来，使研究人员能仅通过观察技术标识符就能辨识专利申请号所对应的情形。从表2中，我们可以观察到PATSTAT数据库为4种不同情形的专利申请号分配了不同的技术标识符。

表2 PATSTAT数据库不同范围技术标识符范围表

PATSTAT版本	范围1：对应DOCDB数据的技术标识符（APPLN_ID）	PATSTAT自建的技术标识符（APPLN_ID）		
		范围2：因优先权号码信息缺失而增加的技术标识符	范围3：因引用的专利公开公告号缺失而增加的技术标识符	范围4：因引用的专利申请号缺失而增加的技术标识符
2014 Spring	69410835	900000001- 907140127	930000001- 931724340	960000001- 960013546
2013 Oct	67766435	900000001- 907099488	930000001- 931714237	960000001- 960014115

3 PATSTAT数据库特点分析

由于PATSTAT较为详细公开了其数据库设计思路、过程以及元数据，我们可以从数据库的设计方面对PATSTAT的特点进行评价。主要包括如下6个方面：

（1）面向统计决策分析。OECD的专利统计专题工作组通过系统理论研究如编制《专利统计手册》[1]、建立统计数据子集、开展专利统计决策年会、搭建专利统计研讨平台等方式，为PATSTAT奠定了良好的基础。

（2）数据涵盖范围广泛。PATSTAT数据库主要集成了DOCDB数据库、INPADOC数据库（专利家族及法律状态）以及EPR数据库（专利注册信息）三大数据源，同时，在人名信息（专利权人、发明人）、地址信息以及技术分类等信息上集成了多个其他数据源，使得PATSTAT数据库包含了著录项信息、法律状态信息、过程信息等全方面的信息，地域范围包括了90多个多家的7000万条专利信息，更新频率为每年两次[11]。

（3）专利数据的深层次集成。对于异构、复杂、多源专利数据进行深层次的语义集成是PATSTAT数据库的主要进展之一。具体体现在：专利家族与优先权信息的集成，专利摘要与标题信息的集成，发明人信息与地址信息的集成，以及专利权人信息的集成等方面。这一系列的数据集成使得PATSTAT更加适合全球范围的专利统计分析工作。

（4）体现数据仓库特征。通过物化集中方式将多源异构数据集成到一个统一的系统之下，使数据库具有了表达更加复杂的查询、执行更加复杂的数据转化的能力；数据快照则对于动态信息（法律状态信息、专利权人信息、专利家族信息等）能够进行高效的数据精简，虽然，这也需要以牺牲一定的数据更新效率为代价。

（5）资源共享与协同创新。EPO将PATSTAT视为其加工的专利数据产品中的标杆，为了扩大该数据的影响，推动全球范围内的专利数据分析质量，PATSTAT研发团队也适时地将其在数据库设计与开发过程中的一些核心数据文档分享出来，供数据分析人员参考，这些内容包括了设计基本原则、元数据信息以及具体实施规则与代码等。同时，由于PATSTAT数据的这种公开性，使得它目前已经逐步成为了一个专利数据深加工、清洗与协同的平台[12]。

（6）数据处理过程的公开化。PATSTAT数据库团

队分享了其在数据库设计与开发过程中的一些核心数据文档，包括Patstat Data Catalog[13]，该文档包括了专利数据设计基本原则、元数据信息、数据来源范围、数据指标代码以及业务规则等内容，这些核心文档的公布使得整个PATSTAT的数据库成为了专利分析领域第一个透明的、可复制、可追踪的数据资源，更多负责的数据加工、分析工作可以得以开展。

尽管，PATSTAT数据库存在上面诸多优点，也存在一定的局限性，主要体现为两点：①从数据源来看，主要采集自官方数据（如DOCDB、INPADOC、EP Register数据），上述数据主要是服务于审查员工作流程的，那些对于审查员工作流程起到关键作用的数据，如优先权、引文数据的质量就相对较高，而对于审查员工作流程影响较小的数据，如发明人、申请人地址信息，数据加工的质量就相对较低；②PATSTAT数据的地域倾向性较为突出，即EPO来源的专利数据质量较高，而来源于其他区域的专利数据质量就相对差一些。

4 PATSTAT与主流商业专利数据库的比较

数据库设计最终是为了通过功能实现服务的，通过与市场上其他领先的商业数据库（Orbit数据库、Thomson Innovation数据库）进行比较，可以发现PATSTAT数据库在数据库功能实现上的优势(见表4)。

表3 PATSTAT与主流商业专利数据库的比较

比较项目	Orbit数据库	TI数据库	PATSTAT数据库
提供者	Questel-Orbit	Thomson Reuters	EPO
数据服务类型	商业（付费）数据库	商业（付费）数据库	公共（付费）数据库
是否包含全文	包含多个国家的全文信息（如US, EP, WO）	包含多个国家的全文信息（如US, EP, WO）	不含全文
著录信息来源	DOCDB	DOCDB以及DWPI数据	DOCDB
法律信息来源	INPADOC法律状态、EP的登记信息、日本法律状态信息、美国诉讼数据以及部分国家地区的登记信息（EP外）	INPADOC法律状态、EP的登记信息、美国授权转让、持续期、诉讼、授权后信息；JP专利审查、注册、注册后法律状态、审判、上诉程序信息	INPADOC法律状态、EP的登记信息
专利家族信息	包括PlusPat、FamPat以及扩展专利家族	包括德温特专利家族以及扩展专利家族	包括扩展专利家族及DOCDB简单专利家族
专利权人清洗	利用PatentRatings公司数据进行清洗	自建专利权人清洗规范，辅助以1790 Analytics公司数据	以DOCDB标准化专利权人信息为基础，整合HAN、APE-INV数据
数据更新频率	大部分的全文信息是每周更新，部分国家数据为按月更新	大多数数据是每周更新，专利权人数据是按月更新，DWPI是每三天更新	一年更新两次（春季与秋季）
遗失数据补充	不做数据补充	著录信息不做数据补充，但在核心专利本文插图中的缺失信息已经被补充进来；DWPI增值数据对数据进行了更正	对于来源INPADOC、DOCDB专利文献中的引用、申请号、优先权号遗失信息的进行了补充
错误信息更正	不做更正	DWPI增值数据对数据进行了更正	不做更正，但可以利用多个数据集进行补充择优处理

Orbit数据库由于有了FAMPAT和PLUSPAT对家族数据的精确定义，使得系统能够在更深层次上对数据实现整合，例如专利家族的引用关系、专利家族之间的关系等。另外，该数据库在全文信息、插图信息以及法律诉讼数据方面的集成都使其能够较好地满足专利审查员的专利审查工作。然而，Orbit数据库也存在一些不足，如没有采取数据补充措施，专利权人清洗方面的效果并不好，检索式并不够灵活等等。

TI数据库包含目前最广泛的专利信息，有最为专业的表格检索、专利号检索、专家检索，及时的更新频率和数据覆盖范围。另外，其独有的DWPI增值专利信息也使得该数据集成为专利技术研发人员和专利分析专

家的有力助手。然而，Derwent封闭的专利家族定义如同黑箱，难以在大数据的范围下复用；另外，TI的数据虽包含了最为广泛的信息，但似乎没有很好的组织，内部缺乏简明的逻辑性，一般用户是难以理解其内在逻辑的，软件使用的门槛较高。

PATSTAT的优势在于其是从统计决策视角对DOCDB、INPADOC等数据集进行的一次深层次的加工。PATSTAT在统计决策方面具有强大的理论支撑，这是其他数据集所不具备的。其次，由于EPO采取了开放的措施，分享了很多数据库设计、操作规范，使得整个数据操作过程是透明的（这是其他商业数据不具备的）。另外，该数据还具有数据全面性（包括著录项、法律状态、登记信息）、数据精简性（不包括全文、插图等信息）、数据的统一性、灵活性、易操作性等特点。由于直接提供数据，可以利用SQL直接检索，是较为有效的检索方式。缺点在于：仅限于统计决策用途，不包含全文、说明书、插图信息等；数据检索、操作方式较为专业，无法为一般用户所使用。

5 结语

PATSTAT作为专利信息服务领域较杰出的产品，其在专利数据加工、集成、设计方面的方法和经验具有独到之处，非常值得学习和借鉴。具体到专利数据集成主要体现在如下四点。

（1）数据集成思路：PATSTAT从统计决策支持的应用场景出发，综合考虑数据集成的数据范围、查询效率、系统架构等问题，将技术创新理论发展与专利数据加工实践有机结合，形成了有鲜明特征的专利统计数据库。具体而言，PATSTAT积极吸纳技术创新理论发展的最新成果，如将《OECD专利统计手册》作为其理论依据[1]；吸纳OECD在专利引文[14]、专利家族[15]、专利权人清洗[16]、专利地理信息[17]、专利质量评价[5]方面的最新成果等，形成了一系列的专利数据集[18]；定期召开国际专利统计年会等。

（2）数据易用性：PATSTAT第一次提供了一个线下的全球专利数据库。为了方便统计分析人员的使用，数据库的数据存储方式采用CSV格式，保证了数据能够便捷的在各种数据库工具上使用；另外，EPO也通过详细的操作手册指导，帮助用户利用SQL数据库直接对全球专利数据进行查询、操作和分析。

（3）数据异构集成技巧：PATSTAT数据范围是全球专利数据，这是一项异常艰巨的任务。不同国家的专利数据在收录范围、数据内容、数据质量、数据结构、语言等方面存在巨大差异，对于这样异质数据的集成不仅需要对各国的专利数据资源有详细的了解，同时要有专业的数据加工团队支持。PATSTAT数据库公开了其在数据库设计与开发过程中的一些核心数据文档，以便于数据分析人员参考；同时，这些设计文档中涉及的数据库设计基本原则、元数据信息以及具体实施规则与代码，都能够对未来专利数据的集成提供较好的帮助。

（4）开放数据：在PATSTAT数据库的研发、更新的过程中，开放思想体现的较为明显。首先，在数据库研发过程中，各国专利局（OECD、WIPO、USPTO等）都积极参与到了其数据库的设计过程；其次，在数据库初步建设完成之后，EPO面向研究人员公布了其核心的数据文档，使广大科研人员能够以其数据库为基础，更新、改进、修正具体的数据集；另外，PATSTAT也广泛吸收新的研究成果，并不断利用这些成果来改进其数据库。因此，通过这一系列的数据开发分享措施，很好地实现了PATSTAT数据库的生态自循环。

参考文献

[1] OECD. OECD Patent Statistics Manual[M]. ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2009.

[2] Griliches Z. Patent Statistics as Economic Indicators: A Survey[J]. JOURNAL OF ECONOMIC LITERATURE ,1990, 28(4): 1661-1707.

[3] WIPO. Handbook on industrial property information and documentation (WIPO publication) [M]. WIPO,2015.

[4] Albrecht M A, Bosma R, van Dinter T, et al. Quality assurance in the EPO Patent Information Resource[J]. WORLD PATENT INFORMATION. 2010, 32(4): 279-286.

[5] Squicciarini M, Dernis H, Criscuolo C. Measuring Patent Quality. 2013.

[6] StÈphane Maraut, HÈlÈne Dernis, Colin Webb, et al. The OECD REGPAT Datatbase: A Presentation[M]. Paris, 2008.

[7] Martinez C. Insight into Different Types of Patent Families. OECD SCIENCE, TECHNOLOGY AND INDUSTRY WORKING PAPERS[C]. 2010.

[8] Webb C, Dernis H, Harhoff D, et al. Analysing European and

International Patent Citations. OECD SCIENCE, TECHNOLOGY AND INDUSTRY WORKING PAPERS[C]. 2005.

[9] EPO. EPO Worldwide Patent Statistical Database (PATSTAT). 2015.

[10] 中华人民共和国国家知识产权局. 专利文献号标准: ZC 0007-2004[Z]. 中华人民共和国知识产权行业标准. 北京: 2004.

[11] EPO. EPO Worldwide Patent Statistical Database - 2014 Autumn Edition.2014.

[12] Coffano M, Tarasconi G. CRIOS - Patstat Database: Sources, Contents and Access Rules[M]. CENTER FOR RESEARCH ON INNOVATION, ORGANIZATION AND STRATEGY CRIOS, 2014.

[13] EPO. Data Catalog V 5.01 Patstat. 2014.

[14] Webb C, Dernis H, Harhoff D, et al. Analysing European and International Patent Citations: A Set of EPO Patent Database Building Blocks[J]. OECD PUBLISHING ,2005(9):31.

[15] Dernis H, Khan M. Triadic Patent Families Methodology[J]. OECD PUBLISHING, 2004.

[16] Ribeiro S P, Menghinello S, De Backer K. The OECD ORBIS Database: Responding to the Need for Firm-Level Micro-Data in the OECD[J]. OECD PUBLISHING, 2010.

[17] Squicciarini M, Dernis H. A Cross-Country Characterisation of the Patenting Behaviour of Firms based on Matched Firm and Patent Data[J]. OECD PUBLISHING, 2013.

[18] OECD. OECD work on patent statistics. 2015.

作者简介

张静，女，1975年生，博士，副研究员，研究方向：专利数据挖掘、信息分析。
杨冠灿，男，1981年生，博士，助理研究员，研究方向：专利数据、技术竞争情报等。
刘会景，女，1984年生，硕士，专利分析师，研究方向：专利分析、专利数据挖掘。

Review of Worldwide Patent Statistical Database (PATSTAT)

ZHANG Jing[1], YANG GuanCan[1], LIU HuiJing[2]
（1. Institute of Scientific and Technical Information of China, Beijing100038, China;
2. WANFANG Data Co., Ltd, Beijing100038, China）

Abstract: This paper dissects PATSTAT from points of database structure, information organizationarchitecture and technical identifier, then analysis the features of PATSTAT, and compares PATSTAT with Leading Commercial Patent Databases, such as Orbit, Thomson Innovation. As a conclusion, the paper summarizes the patent data integration experiences of PATSTAT database.

Keywords: PATSTAT Database; Patent Data; Data Integration; Semantic Heterogeneity

（收稿日期：2015-07-10）