

收稿日期: 2017-07-25
改回日期: 2017-08-08

项目: 中国地质调查局“全国地质资料汇聚与数据整理”项目(1212010040001500018)资助。

doi: 10.12029/gc2017Z101

论文引用格式: 李晨阳, 吴轩, 齐钊宇, 郭慧, 贾丽琼, 商云涛, 孔昭煜, 高学正, 李晓蕾. 2017. 全球地质数据出版模式探索与研究[J]. 中国地质, 44(S1): 1-7.

全球地质数据出版模式探索与研究

李晨阳^{1,2} 吴轩^{1,2} 齐钊宇^{1,2} 郭慧^{1,2} 贾丽琼^{1,2}
商云涛^{1,2} 孔昭煜^{1,2} 高学正^{1,2} 李晓蕾^{1,2}

(1. 中国地质调查局发展研究中心, 北京 100037; 2. 全国地质资料馆, 北京 100037)

摘要: 科学数据是科技创新的基础, 也是人类宝贵的科技信息资源。兼顾保护数据知识产权和推动数据广泛共享是科学数据界长期存在的难题。数据出版(元数据、实体数据、数据论文关联出版)是解决这个难题的有效机制。“全球地质数据出版系统”(中英文)将实现元数据、实体数据、数据论文关联一体出版, 通过互联网实现科学传播和公益性共享机制, 在保护数据知识产权和促进数据共享方面起着重要作用。该系统将推动全世界地质学家共享科学数据, 为地质领域科技创新提供数据基础。本文基于数据出版的概念, 全面介绍了全球地质数据出版, 并探讨其数据共享的意义价值。

关键词: 地质数据; 出版与共享; 模式; 意义; 全球

1 引言

大数据时代, 新一轮的信息技术革命开始启动。作为信息社会新阶段的标志, 大数据正在成为 21 世纪推动社会创新的新动力(曹凌, 2013)。同样, 大数据引发的变革, 推动了信息发布与共享, 也造就了开放数据。开放数据是指公共机构产生、收集或支付的所有信息, 包括地理信息数据、统计资料、气象资料, 由政府资助的研究项目的数据, 以及数字图书, 这些公共数据可以随时访问和咨询, 也可以重新再利用(European Commission, 2012)。

地质领域的资料与数据开放共享一直是相关专家关注的问题(中国科学院地学部, 1996)。中国地质行业长期以来积攒了大量的地质资料和科学数据, 然而缺少相关的共享机制与共享方式, 严重制约着地质领域的科学数据共享(刘闯, 2004)。要解决这个问题, 首先要将科学数据成果的出版纳入科学成果评价体系中。只有这样, 数据作者才能主动、积极和迫切地参与到数据共享的进程中, 才能解决国家科研经费在数据产出方面的成果滞留、遗失和流失的现象(刘闯, 2014)。由此, 中国数据出版应运而生。

长期以来, 中国科学院计算机网络信息中心和依托在该中心的国科联国际科技数据委员会中国委员会及秘书处关注科学数据开放共享问题。他们注意到了数据出版这种最新的数据共享模式, 并于 2016 年创立了中国首个专门的数据期刊《中国科学数据》(郭华东, 2016)。2017 年, 由中国科学院地理科学与资源研究所、中国地理学会联合创办的双

第一作者简介: 李晨阳, 男, 1964 年生, 教授级高级工程师, 从事地质资料数据管理工作; E-mail: chenyang@mail.cgs.gov.cn。

通讯作者简介: 吴轩, 男, 1980 年生, 博士生, 高级工程师, 从事地质资料数据管理工作; E-mail: wxuan@mail.cgs.gov.cn。

语种学术期刊《全球变化数据学报》正式发刊, 响应了大数据时代全球变化科学领域及其相关的地球科学与资源环境科学等众多领域广大科研人员的迫切要求(葛全胜, 2017)。

截至目前, 全国地质资料馆馆藏资料 14 万余档, 单套数据占用容量 170 TB, 每年新进馆资料 5000 余档约 30 万件(齐钊宇等, 2017), 自 2005 年后汇交的地质资料全部包含矢量化的数据。馆藏地质资料中地质大调查成果资料档数仅占 1/14, 但其数据量却占馆藏资料数据总量的 45%, 并且社会化服务占比较高, 所涵盖的地质专业内容具有较高的深度和广度。在 14 万余档馆藏中, 10.2 万档是经数字化扫描形成的电子资料, 其中, 区域地质、区域物化探、区域水工环以及矿产地等数据在地质调查项目的支撑下已经完成了矢量化建库工作。截至目前, 全国地质资料馆共有 30 多个全国性数据库, 3000 多个区域与专题性数据库及数万个点源性数据库。如此海量的数据库(集), 多年来社会化服务却受困于保密、服务模式与对应技术策略和知识产权保护等问题, 因此全国地质资料馆经过多年政策法规解读与技术研究, 形成了一套地质数据出版模式, 建立了全球第一个地质数据出版平台。

全球地质数据出版是基于全球产生的地质资料与科学数据, 建立科学数据出版的政策与制度, 吸纳国际开放数据的理念与方法, 联合 DOI 注册机构, 通过互联网平台发布集创新性、先进性于一体的地质科学数据。本文就地质科学数据出版的概念、出版模式与意义论述如下。

2 全球地质数据出版

2.1 全球地质数据出版定位与内容

全球地质数据出版基于创新的开放数据服务, 定位于建立国际化服务的数据共享平台。全球地质数据出版将开放数据授权, 提供相对完整且真实的数据, 用户可通过公开渠道自由、免费的下载电子化数据。

科学数据出版是科研人员与数据工作者按照规范的质量管理和控制流程, 以数据论文的方式, 通过互联网公开发布其观察、实验、计算分析等科研过程所产生的原始数据, 或通过已有的数据进行系统化地收集、整理和再加工后形成的数据产品, 使得其他使用者能便捷地发现、获取、理解和再分析利用, 并且可以在科研论文及相关科研成果中引用(黎建辉等, 2015)。

全球地质数据出版主要包括数据论文出版和实体数据出版两个方面。数据论文出版包括描述说明实体数据的数据论文和实体数据元数据信息两个方面。实体数据包括地理信息、地质图和数据库等常见共享数据, 还包括地质工作过程中形成的文献、档案记录、数据表格和其他多媒体、各类以数据为中心的应用、数据库接口服务和专题服务等多种数据类型。

2.2 全球地质数据作者投稿

全球地质数据投稿包括作者承诺和投稿内容两个部分。

2.2.1 作者承诺

申请全球地质数据出版的作者需作出以下承诺:

(1) 每一位数据作者需确认对数据集(库)具有自主知识产权;所投稿件的所有作者,均系本稿件的创造者,对所投稿件内容的正确性、真实性和可靠性等承担责任。一经签署协议,对稿件的署名及署名的顺序将不再变动。作者同意将数据著作权中编辑权、不同介质复制权、依注册中数据公开范围内的数据散发权、网络传播权、多语种翻译权、印刷权和上述产权的转让权与数据出版者共同拥有;该数据集(库)出版权由数据出版者单独所有(刘闯,2014)。

(2) 作者需要保证数据所有权清晰,数据所有权单位清晰无异议。

(3) 保证科学数据的真实性及原创性。作者保证科学数据真实,未作假;所投稿件属作者本人的创造性劳动成果,凡引用了他人的论述、数据、结果等处,均已列出对应的参考文献。

(4) 未一稿多投。一个数据集(库)仅能出版一次,更新数据库以不同版本出版除外。

(5) 作者单位或稿件所述课题,对稿件的发表无保密要求。若有保密要求,经本单位审核并加盖单位公章,同意稿件公开发表。

2.2.2 作者投稿内容

全球地质数据出版作者可选择数据论文与实体数据同时出版,或仅出版实体数据。作者需准备实体数据、元数据两项或实体数据、元数据和数据论文三项进行投稿。

(1) 元数据

元数据是全球地质数据出版注册 DOI 的关键信息,作者需将元数据表格中各项信息逐一填写。

(2) 数据论文

数据论文是一篇集数据集(库)说明与论证该数据集创新性和可靠性于一体的说明论述性文章。因此,数据论文既具有说明文的特点,也具有研究论文的特征。

(3) 实体数据

实体数据是数据出版的核心内容。对于全球地质数据出版来说,数据包含整个地球科学领域专业的科学数据,呈现明显的多样性。实体数据是全球地质数据出版的精髓,是地质科技创新的基础。

2.3 数据出版 DOI

DOI 是美国出版协会于 1998 年提出的数字对象唯一标识符,用于标识网络环境下的任何数字化对象,以便有效管理数字出版物,保护数字出版物的知识产权(尚利娜等,2015;毛军等,2005;张晓林,2001,2002;Paskin,2005)。2013 年,科学研究数据全球联盟(RDA)与世界数据系统(WDS)联合成立数据出版共识组(Data Publishing Interesting Group)。在这些国际组织世界性的大动作驱动下,在知识产权清晰的基础上,全球性科学研究数据大汇集正在涌动并开始形成可持续潮流。以美国主导的全球生物多样性基础设施(GBIF)为例(Penev L,2011),目前已经有 3 亿多数据集(库)业已出版(获得 DOI 注册)(Brase et al.,2006;Florian et al.,2012)、300 多个数据出版中心纳入到该共享网。在德国,仅马普学会 80 个研究所目前已经有 100 多万数据集(库)注册了 DOI/Handel(Lawrence et al.,2011)。

全球地质数据出版采用网刊和纸刊两种方式同时出版，并为实体数据和数据论文分别注册 DOI。出版的实体数据将在清华大学图书馆 (10.23650) 进行 DOI 注册，清华大学图书馆是 DataCite 的注册会员，被授权成为 DOI 注册机构 (RA)；数据论文 DOI 在中文 DOI(10.12063) 进行注册，中文 DOI 是由 IDF 正式授权的 DOI 注册机构 (RA)。

2.4 数据出版流程

全球地质数据出版具体分为实体数据提交、编辑初审、实体数据发布、永久存储、数据引用和影响评价 6 个基本环节。

(1) 实体数据提交。数据作者将整理好的实体数据和数据论文投稿至《全球地质数据》。

(2) 编辑初审。编辑对作者提交的数据论文和数据集进行初审，初审内容包括保密审查、公开性、误差范围等内容。初审不通过时，直接反馈作者；初审通过后，进行同行专家评议。

(3) 同行评议。同行评议是评审的核心部分，主要由相关领域专家对数据完整性、科学性、严谨性、应用价值等方面进行审核评议。

(4) 数据发布。将返回的同行专家评审意见报送责任编辑，责任编辑决定是否发布该数据集和数据论文。为可发布的数据集和数据论文提供数字对象唯一标识符 (DOI)，并在数据平台进行发布。

(5) 永久存储。为了便于用户访问和追溯，建立能够长期稳定运行的数据中心，为可发布的数据集提供永久访问地址。

(6) 数据引用和影响评价。为实现数据与相关论文、相关数据的关联，为可发布的数据集提供数据引用格式，利用 DOI 构建数据引用机制；将数据访问量、下载量及引用量作为共享成效和影响力的评价指标，并定期形成电子简报。

2.5 全球地质数据共享服务

全球地质实体数据出版将在线发表在全球地质数据出版系统 <http://dcc.cgs.gov.cn>，数据论文将在线发表在 <http://geodb.cgs.gov.cn>，同时在期刊发表并公开发行（图 1）。用户可免费下载在线的实体数据和数据论文。

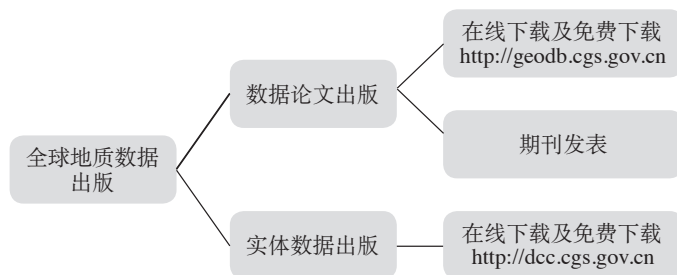


图 1 全球地质数据出版模式

3 地质数据共享政策

参考发展中国家数据共享原则、WDS 数据共享原则、GEO 数据共享原则等国际组织关于科学数据共享的基本原则和“全球变化科学研究数据出版系统”（中英文）制定数据共享政策（刘闯等，2017），全国地质资料馆制定了关于全球地质科学数据管理的有

关规定。“全球地质数据出版系统”(中英文)出版的“数据”包括元数据(中英文)、实体数据(中英文)和通过《全球地质数据》(中英文)发表的数据论文。其共享政策如下:

(1) 经过审核后的全国地质资料馆馆藏地质资料数据和作者提交的地质数据可通过互联网系统免费向全社会开放, 用户免费下载;

(2) 最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源;

(3) 增值服务用户或以任何形式散发和传播(包括通过计算机服务器)“数据”的用户需要与“全球地质数据出版系统”(中英文)编辑部签署书面协议, 获得许可;

(4) 摘取“数据”中的部分记录创作新数据的作者需要遵循 10% 引用原则, 即从本数据集中摘取的数据记录少于新数据集总记录量的 10%, 同时需要对摘取的数据记录标注数据来源。

4 全球地质数据出版意义

(1) 数据出版为数据共享提供强力支撑

数据共享包括两层含义: 一是将保存在科学家个人手中的数据贡献出来, 让更多的人可以使用, 从而扩大数据的使用范围; 二是要保证共享数据的完整性, 通过数据文档、元数据和数据论文等手段对数据进行详细的描述说明, 让已经“共享”的数据可以被更多的人正确使用, 最大程度发挥数据的潜在价值。

为促进科学数据资源的共享和交换, 许多发达国家和国际组织都开展了一系列基于计算机网络的科学数据共享的研究和实践, 目的是将长期积累的科学数据为可持续发展等研究提供数据支撑服务。数据出版是将丰富的公开数据资源开放、为公众提供服务、为国家提供基础资源、实现数据共享的关键环节。

(2) 数据出版为知识产权保护提供平台

科学数据的各个处理阶段都凝聚着科学家的智力劳动, 包括观测仪器布置、数据模拟方法和数据处理方法等。数据(特别是科学数据)具有知识产权已经成为共识。数据知识产权集中体现在数据的版权, 特别是数据的署名权、出版权和编译权。数据出版系统借鉴传统学术出版系统框架, 可有效解决数据版权问题(司莉, 2015)。

通过数据中心出版的数据, 可通过 DOI 注册解决数据署名权及其署名顺序, 通过数据授权解决数据版权问题。数据期刊多采用开源获取(Open Access)的模式, 数据可被无限制地获取和使用, 避免与数据共享产生冲突。

(3) 数据出版促进中国地质数据国际化

在数据出版过程中, 每个数据集的 DOI 及其中英文双语核心元数据、数据论文都可通过 datacite.com 及全国地质资料馆的数据出版平台提供全球范围的查询, 将会大大提高中国地质数据的国际化水平。

(4) 数据出版推动地质资料碎片化、专题化服务

地质资料碎片化服务是馆藏资源服务的最小服务单元, 是有效提升用户体验、构建开放的地质资料信息化建设生态的重要手段。

地质资料专题化服务是对地质资料进行深度加工和整合, 提高地质资料信息的集中度和关联度, 形成分布式、多层次、满足用户不同需求的地质资料信息服务方式。

数据出版将两者进行有效结合, 利用唯一标识颗粒度描述规则实现地质资料统筹

规划、分级管理,以统一标准、信息共享为原则,提供方便、快捷、高效的地质资料服务,逐步延长地质资料社会化服务链。

(5) 数据出版促进原始地质资料汇交和管理

数据共享平台依靠数据出版机制,及时将汇交的数据进行知识产权认证,保护数据工作者的权益,不仅可以提高地质资料汇交人的积极性,还为原始地质资料及时汇交提供强有力的保障。加强地质资料管理,创新汇交机制,维护国家地质数据主权和资源所有权,为地质资料汇交和服务利用形成良性循环。

(6) 促进地质工作者的创新产出

科学数据是科技创新的基础,也是人类宝贵的科技信息资源。地质工作者的创新离不开坚实的数据基础。但以往地质数据获取难度较大,全球地质数据出版将最大程度地提供地质数据,服务于地质工作者,促进地质科技创新。对于地质工作者而言,除报告、专著、学术论文之外,又多了数据出版这样一个新型科技创新成果产出。

5 第一期数据出版

《全球地质数据》第一期数据论文纸刊发表采用《中国地质》(增刊)的方式,实体数据共发布 11 个数据集(库)。每一个数据集(库)均包含数据论文(中英文版)出版和实体数据关联出版。在第一期出版的地质科学数据集(库)中,作者有来自地质学领域的院士及知名专家,有科研院所的科研人员,也有基层地质工作人员。在实体数据方面,有小比例尺全国矢量图件、境外地质成果图件、中国标准分幅地质图库、数据量不足 1 M 的 Excel 格式的数据表格及数据量在 GB 级别的数据集。《全球地质数据》的出版实现了科学数据真正意义上的共享,为地质领域科技创新提供了重要的数据基础。

6 结束语

科学数据共享是科学界关注的热点问题。全球地质数据涉及专业领域广泛,包括地质学、地球物理学、地球化学、水文、工程、环境、遥感等多个专业,且多领域交叉现象突出。科学数据知识产权保护和数据共享是科学家们长期呼吁和迫切希望解决的问题,也可以说是科学家们的“民生”问题。“全球地质数据出版系统”(中英文)为这个问题解决做出了实践性案例,将实现元数据、实体数据、数据论文关联一体出版,通过互联网实现科学传播和公益性共享机制。

全球地质数据出版虽然不能解决数据的保密问题,但其是较好处理数据保密问题的方法之一。影响地质数据共享的因素很多,对于地质数据而言,保密、服务模式和知识产权保护是重要原因。希望通过地质数据出版工作,在我们已经比较好地处理地质数据保密与公开、服务模式与对应技术策略基础上,能够较好地处理知识产权保护问题。

全球地质数据出版、数字地质资料馆公益平台和“国家地质”虚拟展馆地质成果科普平台,共同构成一个相对完整的地质信息服务体系格局。“全球地质数据出版系统”(中英文)将推动全世界地质学家共享科学数据,为地质领域科技创新提供数据基础。

参考文献

- 曹凌. 2013. 大数据创新: 欧盟开放数据战略研究 [J]. 情报理论与实践, 36(4): 118-122.
European Commission. Digital agenda: Commission's open data strategy, questions & answers [EB/

- OL]. 2012. <http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/891>.
- Florian Quadt, André Düsterhus, Heinke Höck, Michael Lautenschlager, Andreas V. Hense, Andreas N. Hense, Martin Dames, Atarrab. 2012. A Workflow System for the Publication of Environmental Data, Volume 11, 2012. <http://dx.doi.org/10.2481/dsj.012-027>, https://www.jstage.jst.go.jp/article/dsj/11/0/11_012-027/_article.
- 葛全胜. 2017. 全球变化数据学报发刊词 [J]. 全球变化数据学报, 1(1). DOI:10.3974/geodp.2017.01.01.
- 郭华东. 2016. 问渠哪得清如许, 为有源头活水来——《中国科学数据》发刊词 [EB/OL]. <http://www.csdata.org/p/7/12/>. DOI: 10.11922/csdata.0.2016.0014.
- Brase J, Schindler U. 2006. The Publication of Scientific Data by World Data Centers and the National Library of Science and Technology in Germany [J]. Data Science Journal, Volume 5, <http://dx.doi.org/10.2481/dsj.5.205>, https://www.jstage.jst.go.jp/article/dsj/5/0/5_0_205/_article.
- 黎建辉, 吴超, 张丽丽, 李成赞, 胡良霖. 2016. 科学数据出版调查与分析 [J/OL]. 中国科学数据, 1(1). <http://www.csdata.org/paperView?id=9>. DOI: 10.11922/csdata.120.2015.0009.
- Lawrence B, Jones C, Matthews B, Pepler S, Callaghan S. 2011. Citation and Peer Review of Data: Moving Towards Formal Data Publication [J]. The International Journal of Digital Curation 6, 4–37.
- 刘闯, 郭华东, Paul Uhlir, 等. 2017. 发展中国家数据出版基础设施与共享政策研究 [J]. 全球变化数据学报, 1(1):3–11.
- 刘闯. 2014. 论全球变化科学研究数据出版 [J]. 地理学报, 69(增刊):3–11.
- Liu Chuang. 2004. Recent Developments in Environmental Data Access Policies in the P.R. China, Open Access and the Public Domain in Digital Data and Information for Sciences. National Research Council of the National Academies, Washington D.C., USA. 74–78.
- 刘闯. 2004. 我国科学数据共享机制建设研究 [J]. 国土资源信息化, (1): 5–7.
- 毛军, 孟连生, 镇锡惠, 倪金松, 王燕. 2005. 试论我国数字资源唯一标识符发展战略 [J]. 现代图书情报技术, (2):1–4.
- Paskin N. 2005. Digital object identifiers for scientific data [J]. Data Science Journal, 4:12–20.
- Penev L, Mietchen D, Chavan V, Hagedorn G, Remsen D, Smith V, Shotton D. 2011. Pensoft Data Publishing Policies and Guidelines for Biodiversity Data. Pensoft Publishers, http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf.
- 齐钊宇, 孔昭煜, 高学正, 等. 2017. 地质资料数字资源建设现状及发展趋势研究: 以全国地质资料馆为例 [J]. 中国矿业, 26(6):34–38.
- 尚利娜, 牛晓勇. 2015. 我国学术期刊参考文献中 DOI 著录现状分析 [J]. 中国科技期刊研究, 26(5):484–487.
- 司莉, 贾欢, 邢文明. 2015. 科学数据著作权保护问题与对策研究 [J]. 图书与情报, (04):118–122.
- 张晓林. 2002. 开放数字环境下的参考文献链接 [J]. 现代图书情报技术, (1):9–13.
- 张晓林. 2001. 数字对象的唯一标识符技术 [J]. 现代图书情报技术, (3):8–14.
- 中国科学院地学部. 1996. 关于进一步做好我国地球科学、资源与环境科学研究基础资料与数据共享的建议 [J]. 地球科学进展, 11(1):122–123.

(责任编辑 王学明)

Received: 25-07-2017
Accepted: 08-08-2017
Fund support: Supported
by China Geological Survey
project "National geological
data convergence and
data collation" funding
(1212010040001500018).

doi: 10.12029/gc2017Z101

Article Citation: Li Chenyang, Wu Xuan, Qi Fanyu, Guo Hui, Jia Liqiong, Shang Yuntao, Kong Zhaoyu, Gao Xuezheng, Li Xiaolei. 2017. Exploration and research on global geological data publishing model [J]. *Geology in China*, 44(S1):1–9.

Exploration and Research on Global Geological Data Publishing Model

LI Chenyang^{1,2}, WU Xuan^{1,2}, QI Fanyu^{1,2}, GUO Hui^{1,2}, JIA Liqiong^{1,2}, SHANG
Yuntao^{1,2}, KONG Zhaoyu^{1,2}, GAO Xuezheng^{1,2}, LI Xiaolei^{1,2}

(1. *Development and Research Center of China Geological Survey, Beijing 100037, China*; 2. *China Geological Survey, Beijing 100037, China*)

Abstract: Scientific data form the basis of scientific and technological innovation and valuable scientific and technological information resources. Protection of intellectual property rights and promotion of broad sharing of data have long been a contradictory problem in the field of scientific data. Data publishing (publishing of integrated metadata, entity data, data papers) is an effective mechanism to solve this problem. By publishing integrated metadata, entity data and data papers, the “Global Geological Data Publishing System” (in Chinese and English) realizes scientific communication and sharing mechanisms through the internet and plays an important role in protecting data intellectual property and promoting data sharing as well. This system will enable geologists worldwide to share scientific data and provide a database for scientific and technological innovation in the geological field. Based on the concept of data publishing, this paper introduces global geological data publishing and discusses the significance of data sharing.

Keywords: geological data; publishing and sharing; mode; significance; global

1 Introduction

A new round of information technology revolution has started in this era of big data. As a mark of the new stage in the information society, big data has become a new drive of the 21st century for social innovation (Cao Ling, 2013). The change triggered by big data has promoted information publishing and sharing and created open data, which means all information produced, collected or paid by public institutions, including geographic information data, statistical data, meteorological data, digital libraries and data of research projects sponsored by the government. These public data can be readily accessed, inquired about and reused (European Commission, 2012).

The opening and sharing of geological information and data has always been a concern of relevant experts (The Earth Sciences Academic Division of Chinese Academy of Sciences, 1996). Abundant geological information and scientific data have been accumulated by the geological community of China over a long period. However, a lack of shared mechanism and methods has severely restricted the share of scientific data in geological fields (Liu Chuang, 2004). To solve this problem, we first need to cover the

About the first author: LI Chenyang, male, born in 1964, senior engineer, mainly engages in geological data management; E-mail: chenyang@mail.cgs.gov.cn.

The corresponding author: WU Xuan, male, born in 1980, doctor candidate, senior engineer, engages in geological data management; E-mail: wxuan@mail.cgs.gov.cn.

publishing of scientific data in the scientific achievements evaluation system. This is the only way to push authors of data to actively and earnestly participate in the process of data sharing and solve the problem of detainment and loss of data produced from projects sponsored by the state scientific research fund (Liu Chuang, 2014). Therefore, data publishing has now begun to emerge in China.

The Computer Network Information Center of the Chinese Academy of Sciences, the CODATA China Committee and the Secretariat of the International Council of Scientific Unions (ICSU) have long been concerned about the opening and sharing of scientific data. They recognized data publishing—the newest data sharing mode, and founded in 2016 *China Scientific Data*, the first special data journal of China (Guo Huadong, 2016). The Institute of Geographic Sciences and Natural Resources Research, CAS and the Geographical Society of China (GSC) jointly founded the bilingual *Journal of Global Change Data and Discovery* in 2017 in response to the urgent demand of scientists and researchers in various fields, including global change sciences and related earth sciences and resources and environment sciences (Ge Quansheng, 2017).

So far, the holdings of National Geological Archives of China (NGAC) exceed 140,000 archives, occupying 17 TB of storage capacity. About 5,000 new archives (300,000 pcs.) are collected by the NGAC each year (Qi Fanyu et al., 2017). All geological data collected after 2005 includes vectorized data. Among the geological data collection, the achievement data of the geological grand survey occupies only one fourteenth of the total number of archives, but 45% of the data size of the total collection. Of the 140,000+ archives, 102,000 are electronic data formed by digital scanning. Under the support of geological survey projects, vectorization database construction has been completed for data of regional geology, regional geophysical and geochemical exploration, regional hydraulic rings and mineral sites. At present, the NGA has more than 30 national databases, 3,000 regional and thematic databases and tens of thousands of point feature databases. However, the public service, with so many databases (datasets), has been hindered by the problem of confidentiality, service mode and corresponding technological strategy and protection of intellectual property rights. Therefore, after conducting interpretation and research on relevant policies and laws for several years, the NGAC successfully created a geological data publishing mode and established the first geological data publishing platform in the world.

Global geological data publishing is a new mode of publishing based on geological information and scientific data produced worldwide. In this mode of publishing, policy and regulations on scientific data publishing are established, the new ideas and methods of international open data are adopted to publish integrated innovative and advanced geological data through the internet in association with the DOI registration agency (RA). This paper explores the concept, mode and significance of geological data publishing.

2 Global Geological Data Publishing

2.1 Positioning and Content of Global Geological Data Publishing

Global geological data publishing is based on a creative open data service and positioned on an establishing international data sharing service platform, which will open the data license and provide relatively complete and true digital data for users to readily download through public channels free of charge.

Scientific data publishing is a concept of publishing whereby scientists, researchers and data personnel publish on the internet, in the form of data papers and in accordance with standard quality management and control processes. This includes the raw data produced in their scientific research processes such as observation, experiment, calculation and analysis or other data products formed as a result of systematic collection, consolidation and processing of existing data. Therefore, other users can conveniently find, acquire, understand, analyse, utilize and cite them in scientific research papers and relevant scientific achievements (Li Jianhui et al., 2015).

Global geological data publishing includes data paper publishing and entity data publishing. The former covers data papers that describe entity data and the metadata information of entity data. The latter includes common shared data such as geographic information, geological maps, databases, documents, records and data sheets formed during geological work processes, other multi-media data, data-centric applications, database interface service and thematic service.

2.2 Author contribution of global geological data

Author contribution of global geological data includes authors' undertakings and content of contribution.

2.2.1 Author undertakings

The author who applies for global geological data publishing shall undertake the following:

(1) Each author of data shall confirm that he/she owns the intellectual property rights to the dataset (database); all the authors of the contributions are the creators of the contribution and are responsible for the correctness, authenticity and reliability of the content of contribution. The authorship and order of authors will not be changed once the publishing agreement is signed. The authors agree to share with the publisher the following copyrights: editorial right, right of reproduction in different media, right of data distribution within registered data publication scopes, network communication right, multilingual translation right, print right and the right to transfer the abovesaid property rights. The right to publish the dataset (database) is exclusively owned by the data publisher (Liu Chuang, 2014).

(2) The author shall guarantee that the ownership of data is clear and the data owner entity is clear and free of disputes.

(3) Guarantee of authenticity and originality. The author shall guarantee that the scientific data is true and not false and the contribution is the creative work achievements of the author. References have been listed for all citations of arguments, data or results of other authors.

The contribution is not made to more than one publisher. One dataset (database), except for the updated version, can be published only once.

Contribution has no confidentiality requirement on the publication of the contribution. In case that confidentiality requirement is involved, the contribution shall be reviewed and approved by the author affiliation for publication by affixing an official seal on it.

2.2.2 Content of contribution

The author may choose to publish data paper(s) and entity data simultaneously or just

publish the latter. The author shall submit both entity data and metadata, or all of the entity data, metadata and data paper.

(1) Metadata

Metadata is the key information for global geological data publishing DOI registration. The author shall fully complete the metadata table item by item.

(2) Data paper

A data paper is an expository and argumentative article that describes the dataset (database) and argues for the creativity and reliability of the dataset. Hence, a data paper has the features of both expository prose and a research paper.

(3) Entity data

Entity data form the core content of data publishing. For global geological data publishing, data must include scientific data in all earth science fields and show obvious diversity. Entity data are the essence of global geological data publishing and the basis for geological innovation.

2.3 Data publishing DOI

DOIs (Digital Object Unique Identifiers) were created by the Association of American Publishers (AAP) in 1998 to identify any digital object in a network environment for effective management of digital publications and protection of the intellectual property rights of digital publications (Shang Lina, 2015; Mao Jun et al., 2005; Zhang Xiaolin, 2001, 2002; Paskin, 2005). In 2013, the RDA and WDS jointly founded the Data Publishing Interest Group. Driven by the worldwide efforts of these international organizations, worldwide collection of global scientific data is emerging and begins to form a sustainable trend on the basis of clear intellectual property rights. For example, the US-led Global Biodiversity Infrastructure (GBIF) (Penev, 2011) has published more than 300 millions datasets (databases) (DOI registered) (Brase et al., 2006; Florian et al., 2012) and over 300 data publishing centers have been incorporated into this data sharing network. In Germany, 80 research institutes of Maxi-Planck-Gesellschaft have more than one million datasets (databases) registered with DOI/Handel (Lawrence et al., 2011).

The Global Geological Data Publishing system publishes data in two forms simultaneously, i.e. online journals and paper journals, and applies for DOI registration for entity data and data papers. Published entity data will be registered with a DOI in the Tsinghua University Library (10.23650), which is a registered member of DataCite and an authorized DOI Registration Agency (RA). Data papers are registered with China DOI (10.12063), which is a DOI RA officially authorized by the IDF.

2.4 Data publishing process

The process of global geological data publishing includes six basic links, i.e. submission of entity data, editor's initial review, entity data publication, permanent preservation, data citation and influence evaluation.

(1) Submission of entity data: the author of the data submits cleansed entity data and data papers to *Global Geological Data*.

(2) Editor's initial review: the editor conducts an initial review on the data paper and datasets submitted by the author, including a review of confidentiality, publicity and error range. If the contribution fails to pass the initial review, the editor will feed it back to the

author directly. Peer review will be carried out after the initial review is passed.

(3) Peer review: as the core part of the review stage, peer reviews are carried out by experts in related fields to review and evaluate the completeness, scientific rigor and application value of the data.

(4) Data publishing: peer review comments are submitted to the responsible editor-in-chief, who will decide whether or not to publish the dataset and data paper. Datasets and data papers that are accepted for publication will be assigned a DOI and published on the data platform.

(5) Permanent preservation: in order to facilitate access and tracing by users and establish a data center that can operate stably for the long term, a permanent access address is provided to the published dataset.

(6) Data citation and influence evaluation: in order to realize the association of data with a relevant paper and data, a data reference format is provided for a published dataset and a data citation mechanism is created with the DOI. The share effect and influence will be evaluated on the basis of data access, downloads and citations and an electronic evaluation report will be produced regularly.

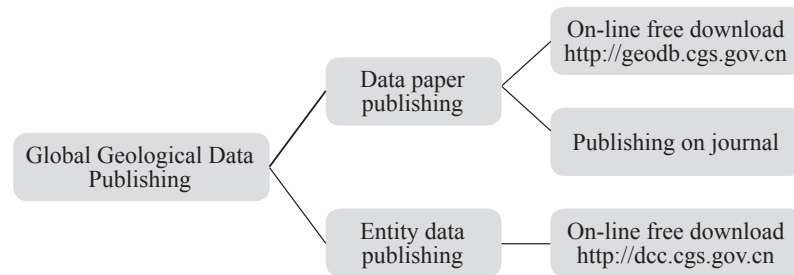


Fig.1 Model of Global Geological Data Publishing

2.5 Global geological data sharing service

Global geological entity data will be published at <http://dcc.cgs.gov.cn> (Global Geological Data Publishing System), and data papers will be published at <http://geodb.cgs.gov.cn>. They are synchronously published in the journal and distributed openly (Fig. 1). The users may download the online entity data and data paper free of charge.

3 Geological Data Sharing Policy

A data sharing policy was established in reference to the data sharing principles of developing countries, the basic principles of scientific data sharing established by international organizations, including WDS and GEO and the Global Change Research Data Publishing and Repository (in Chinese and English) (Liu Chuang et al., 2017). The National Geological Archives established relevant regulations on the management of global geological data. The data published by the Global Geological Data Publishing System (in Chinese and English) includes metadata (in Chinese and English), entity data (in Chinese and English) and the data papers published in *Global Geological Data* (in Chinese and English). The sharing policy is as follows:

(1) The geological data held by the National Geological Archives, upon review and approval, and the geological data submitted by authors are open to the public for download

by end users free of charge on the internet.

(2) When the data are cited, the end users shall indicate the sources of data in the reference or at other appropriate locations in the reference format.

(3) Users of the value-added service and users who distribute and communicate the data in any form (including through computer servers) shall sign an agreement in writing with the Editorial Office of the Global Geological Data Publishing System (in Chinese and English) to obtain a license.

(4) Authors who extract a part of the data to create new data shall follow the principle of 10% citation, i.e., the data record extracted from this dataset shall be less than 10% of the total records of the new dataset, and the source of extracted data records shall be indicated.

4 Significance of Global Geological Data Publishing

(1) Data publishing provides powerful support.

Data sharing has two layers of meaning: first, it means that the data in the hands of individual scientists are contributed and made available to more people so as to enlarge the scope of data application; second, it means that the integrity of shared data shall be ensured by detailed description of data through a data file, metadata and data paper, so that the shared data can be properly used by more people and the potential values of the data are utilized to the maximum extent.

In order to promote sharing and exchange of scientific data resources and provide data support services to research on sustainable development with scientific data accumulated over a long period, many developed countries and international organizations have conducted research and practice of scientific data sharing based on the internet. Data publishing is the key link to share and open abundant data resources to the public.

(2) Data publishing provides a platform for protection of intellectual property rights.

Each processing stage of scientific data is condensed with the intellectual labor of scientists, including setting up of observation apparatus, data analogue and data processing methods, etc.. Intellectual property rights to data (especially scientific data) have been commonly recognized. Data intellectual property rights are mainly represented by copyrights, especially right of authorship, right of publication and right of compiling. The data publishing system can effectively solve the problem of copyrights by using the frame of traditional academic publishing, (SI Li, 2015).

For data published by the Data Center, authorship of data and order of authors are protected through DOI registration, and data copyrights are protected by data licenses. Data journals mainly use the open access mode to allow unlimited access and use of data and avoid conflict with data sharing.

(3) Data publishing promotes internationalization of geological data from China

During the process of data publishing, the DOI of each dataset, and the Chinese and English bilingual core metadata and data papers can be queried worldwide through <https://datacite.com> and the data publishing platform of the National Geological Archives. This will greatly improve the internationalization level of geological data from China.

(4) Data publishing promotes the geological data fragmentation and thematic service.

The geological data fragmentation service, as the smallest service unit of the collection resources service, is an important means to effectively improve users' experiences and construct open geological data informatization ecology.

The geological data thematic service is a kind of information service mode whereby geological data are further processed and integrated to improve the data concentration and association level, and form distributed and multi-level geological data to meet the different demands of users.

By effectively integrating these two types of service and realizing uniform planning and hierarchical management of geological data through the unique identifier granularity description rule, data publishing provides a convenient, fast and efficient geological data service and gradually extends the public service chain of geological data on the principle of uniform standard and information sharing.

(5) Data publishing promotes submission and management of original geological data.

With the data publishing mechanism, the data sharing platform efficiently certifies the intellectual property rights to submitted data to protect the rights and interests of data authors. It not only improves the activity of submitters of geological data but also provides a strong guarantee for timely submission of original geological data. Data publishing can strengthen the management of geological data, offer innovations to the submission mechanism, safeguard the national sovereignty and ownership of geological data resources and create an authoritative cycle for submission, service and application of geological data.

(6) Data publishing promotes the creative output of geologists.

Scientific data is the basis for scientific and technological innovation and the precious scientific information resources of researchers worldwide. It facilitates the innovation of geologists, which cannot be made without a solid data foundation. However, it was quite difficult to obtain geological data in the past. The Global Geological Data Publishing system will provide geological data to the maximum extent, serve geologists and encourage innovation in geology. For geologists, data publishing is a new type of scientific output in addition to reports, books and academic papers.

5 The First Issue *Global Geological Data*

The first issue of *Global Geological Data* published data papers in the form of a paper journal called *Geology in China* (Supplement) and 11 datasets (databases). Each of the datasets (databases) includes publication of a data paper (in Chinese and English) and entity data. Among the authors of the first issue are academicians and established experts in geology, and researchers from institutes and geological personnel from units on a basic level. The published entity data includes small scale national vectorized maps, foreign geological maps, standard geological maps of China, data sheets in the Excel format with data sizes ranging from less than 1 MB to the GB level. The publishing of *Global Geological Data* realized sharing of scientific data in its real sense and provided an important data foundation for scientific and technological innovation in geology.

6 Conclusion

Sharing of scientific data is a hot issue of concern in the scientific community. Global geological data involves a wide range of disciplines, including geology, geophysics, geochemistry, hydrology, engineering, the environment and remote sensing, being often multi-disciplinary. The issue of intellectual property right protection and data sharing is an urgent problem that scientists have been appealing to solve and, so to speak, improve

the “livelihood issue” of scientists. As a practical case to solve this problem, the Global Geological Data Publishing System (in Chinese and English) publishes integrated metadata, entity data and data papers and realizes scientific communication and public sharing mechanisms through the internet.

The Global Geological Data Publishing System is one of the methods that handle data confidentiality, though it cannot completely solve the problem. There are a number of factors that affect the sharing of geological data. Confidentiality, service mode and protection of intellectual property rights are the most important factors. We hope that the issue of intellectual property rights can be properly handled through the geological data publishing system on the basis that we have found ways to properly handle the issue of confidentiality and publishing of geological data, service mode and corresponding technical strategy.

The Global Geological Data Publishing System, the Digital Geological Archives Public Platform and the geological achievements platform in the virtual National Geology exhibition hall consist a relatively complete geological information service system. The Global Geological Data Publishing System (in Chinese and English) will promote geologists worldwide to share scientific data and provide a database for scientific and technological innovation in the geological field.

References

- Cao Ling. 2013. Big data innovation: EU open data strategy research [J]. *Information Studies: Theory & Application*, 36(4): 118–122 (in Chinese).
- European Commission. Digital agenda: commission’s open datastrategy, questions & answers [EB/OL]. 2012. <http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/891>.
- Quadt F, Düsterhus A, Heinke H, Lautenschlager M, Hense A V, Hense A N, Dames M & Atarrabi. 2012. A workflow system for the publication of environmental data, Volume 11, 2012. <http://dx.doi.org/10.2481/dsj.012-027>, https://www.jstage.jst.go.jp/article/dsj/11/0/11_012-027/_article.
- Ge Quansheng. 2017. Journal of global change data & discovery published words [J]. *Journal of Global Change Data & Discovery*, 1(1), DOI:10.3974/geodp.2017.01.01 (in Chinese).
- Guo Huadong. 2016. Ask the way which is clear, as the source of living water to——《CSdata》Published words [EB/OL]. <http://www.csdata.org/p/7/12/>. DOI: 10.11922/csdata.0.2016.0014 (in Chinese).
- Brase J & Schindler U. 2006. The publication of scientific data by world data centers and the national library of science and technology in Germany, *Data Science Journal*, Volume 5, 2006, <http://dx.doi.org/10.2481/dsj.5.205>, https://www.jstage.jst.go.jp/article/dsj/5/0/5_0_205/_article.
- Li Jianhui, Wu Chao, Zhang Lili, Li Chengzan, Hu Lianglin. 2016. Investigation and analysis of scientific data publishing[J/OL]. *Chinese scientific data*. 1(1). <http://www.csdata.org/paperView?id=9>. DOI:10.11922/csdata.120.2015.0009. (in Chinese).
- Lawrence B, Jones C, Matthews B, Pepler S, Callaghan S. 2011. Citation and peer review of data: Moving towards formal data publication[J]. *The International Journal of Digital Curation* 6, 4–37.
- Liu Chuang, Guo Huadong, Uhlir P, et al. 2017. Data sharing principles in developing countries [J]. *Journal of Global Change Data & Discovery*, 1(1): 3–11 (in Chinese).
- Liu Chuang. 2014. Global change research data publishing and repository [J]. *Journal of Geographical Sciences*, 69 (Supp.): 3–11 (in Chinese).

- Liu Chuang. 2004. Recent Developments in Environmental Data Access Policies in the P.R. China, Open Access and the Public Domain in Digital Data and Information for Sciences. National Research Council of the National Academies, Washington D.C., USA. 74–78.
- Liu Chuang. 2004. Research on the construction of scientific data sharing mechanism in China [J]. Land and Resources Information, 2004, (1): 5–7.
- Mao Jun, Meng Liansheng, Zhen Xihui, et al. 2005. Establish a framework of digital resource unique identifier(DrUI) system in China: Strategy and Economics [J]. New Technology of Library and Information Service, (2): 1–4 (in Chinese).
- Paskin N. 2005. Digital object identifiers for scientific data [J]. Data Science Journal, 4: 12–20.
- Penev L, Mietchen D, Chavan V, Hagedorn G, Remsen D, Smith V, Shotton D. 2011. Pensoft data publishing policies and guidelines for biodiversity data. Pensoft Publishers, http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf.
- Qi Fanyu, Kong Zhaoyu, Gao Xuezheng, et al. Research on the present situation and development trend of digital resources of geological data: a case study of NGA [J]. China Mining Magazine, 26(6): 34–38 (in Chinese).
- Shang Lina, Niu Xiaoyong. 2015. An analysis on the current situation of DOI documentation of references in academic journals [J]. Chinese Journal of Scientific and Technical Periodicals, 26(5): 484–487 (in Chinese).
- Si Li, Jia Huan, Xing Wenming. 2015. Study on problems and countermeasures in copyright protection of scientific Data [J]. Library & Information, (04): 118–122 (in Chinese).
- Zhang Xiaolin. 2002. Reference linking in a distributed and open digital [J]. New Technology of Library and Information Service, (1): 9–13 (in Chinese).
- Zhang Xiaolin. 2001. Unique identifiers for digital objects [J]. New Technology of Library and Information Service, (3): 8–14 (in Chinese).
- The Earth Sciences Academic Division of Chinese Academy of Sciences. 1996. Suggestion on further development of earth sciences, resources and environment sciences research basic information and data sharing [J]. Advances in Earth Science, 11(1):122-123.