



古生物化石标本元数据标准*

徐洪河** 聂婷 郭文 陈焱森 袁文伟

中国科学院南京地质古生物研究所和生物演化与环境卓越创新中心, 现代古生物学和地层学国家重点实验室, 南京 210008

提要 基于古生物学研究现状和各大科研机构对化石标本的收藏情况, 提出古生物化石标本元数据标准。古生物化石标本元数据标准包含十项元数据元素以及一项元数据实体, 即, 标题、数据标识符、许可证标识符、关键词、描述、化石标本参数、数据链接地址、创建者、创建时间和访问限制, 以及作为元数据实体的数据提交机构。十项元数据元素之一的化石标本参数, 又可分为古生物化石标本的实物信息、古生物系统分类信息、地层信息和高维信息。建立古生物化石标本元数据标准旨在完善古生物化石标本数据的规范化与标准化、利于化石标本数据的整理、汇交, 跨平台操作、检索、共享和使用, 促进数据驱动的古生物学与地层学新范式研究变革。

关键词 化石标本 古生物学 数据驱动 元数据

中文引用 徐洪河, 聂婷, 郭文, 陈焱森, 袁文伟, 2022. 古生物化石标本元数据标准. 古生物学报, 61(2): 280–290. DOI: 10.19800/j.cnki.aps.2022007

英文引用 Xu Hong-he, Nie Ting, Guo Wen, Chen Yan-sen, Yuan Wen-wei, 2022. Metadata standard on palaeontological specimens. Acta Palaeontologica Sinica, 61(2): 280–290. DOI: 10.19800/j.cnki.aps.2022007

Metadata standard on palaeontological specimens

XU Hong-he, NIE Ting, GUO Wen, CHEN Yan-sen, YUAN Wen-wei

State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology and Center for Excellence in Life and Palaeoenvironment, Chinese Academy of Sciences, Nanjing 210008, China

Abstract The metadata standard on fossil specimens is proposed based on the palaeontological research status and the fossil collections of several institutions. This metadata standard includes ten core elements and one metadata entity. The ten core elements contain Title, Data Identifier, License URI, Keyword, Description, Fossil Specimen Parameter, Online Address, Creator, Create Date, and Access Constraints. The metadata entity is Point Of Contact. Fossil Specimen Parameter includes physical information, systematic palaeontology information, stratigraphic information, and high dimensional information of fossil specimens. The establishment of metadata standard on the palaeontological specimens facilitates the development toward the standardization of fossil specimen data. The standard will benefit sorting, collecting, interoperating, retrieving, searching, sharing, and using fossil data, and promote the new paradigm shift toward data-driven research in paleontology and stratigraphy.

Key words fossil specimens, palaeontology, data-driven, metadata

收稿日期: 2022-01-27; 改回日期: 2022-03-25; 录用日期: 2022-04-27

* 中国科学院网络安全和信息化专项——古生物学与地层学科学数据库建设项目(CAS-WX2021SF-0205)资助。

** 通讯作者: 徐洪河, 研究员, e-mail: hhxu@nigpas.ac.cn

1 前 言

古生物化石标本是古生物学与地层学研究中最最为重要的研究对象,是生命演化的实证,也是世界自然遗产中的重要资源,对于自然资源普查、地质公园建设以及科学传播都是必不可少的(吴淦国等, 2017)。数据驱动的新研究范式正在古生物学与地层学研究中越发普遍(Alroy *et al.*, 2001; Hammer and Harper, 2006; Rong *et al.*, 2007; Guo, 2017),深刻影响着人们对地球与生命演化的认知(Wang *et al.*, 2021)。

古生物化石标本的数据化与数据库建设是新科研范式在古生物学与地层学研究中应用的基础。随着这门学科研究程度的逐渐加深,世界各地的研究机构都积累了丰富的地质标本收藏,也纷纷开发了化石标本相关的数据库(杨眉等, 2017; Wang *et al.*, 2021)。比如,英国自然历史博物馆(<https://www.nhm.ac.uk/our-science/collections/palaeontology-collections.html>)、美国国立自然历史博物馆(<https://collections.nmnh.si.edu/search/>)、美国耶鲁皮博迪博物馆古脊椎动物部门(<https://peabody.yale.edu/explore/collections/vertebrate-paleontology/>)、中国国家岩矿化石标本资源共享平台(<http://www.nimrf.net.cn/>)等(吴志远, 2015; 谢园等, 2017)。这些数据库都提供了与化石标本(和其他地质标本)相关的查询与检索功能。另外,古生物学与地层学研究领域也有专门的数据库,为开展古生物学定量分析与研究提供数据服务,比如,基于化石产出记录的Paleobiology Database (PBDB; <https://paleobiodb.org/#/>) (Peters and McClennen, 2016)和基于地层剖面的Geobiodiversity Database (GBDB; <http://geobiodiversity.com/>) (Xu *et al.*, 2020)、基于古脊椎动物化石标本的DeepBone Database (<http://deepbone.org/home>) (Pan and Zhu, 2019)、深时数字地球计划和南京大学合作的OneStratigraphy (<http://onestratigraphy.ddeworld.org/>)、世界有孔虫数据库(World Foraminifera Database; <https://www.marinespecies.org/foraminifera/index.php>)等。通过上述的数据平台可以一定程度上检索古生物分类、名称、产地、地质时代等方面的信息,也在一定程度上为新科研范式的古生物学研究提供了数据支持,但它

们却未能全面涵盖古生物化石标本的多类型和多元化的信息。这主要是由于目前缺乏关于古生物化石标本元数据的标准与规范。

元数据(metadata)是关于数据的数据,其涵盖数据的背景信息以及对数据的解释,是汇集、查询数据以及理解数据内容、格式、结构和可用性必不可少的内容(Recknagel and Michener, 2018)。元数据不仅标识和描述信息对象,还记录信息对象的表现形式、功能、应用、与其他信息对象的关系,以及信息对象管理历史(Baca, 2008)。古生物化石标本的元数据是指解释化石标本数据的数据,可提供化石标本数据的背景信息,以利于数据生产者有效管理和维护化石标本数据,便于对化石标本数据的检索、分析与使用,也是与其他相关数据库间开展交互关联的信息基础(樊隽轩等, 2009)。

国外已经建立了适用不同领域的元数据标准。都柏林核心(Dublin Core)由联机计算机图书馆中心(Online Computer library Center)和美国国家超级计算应用中心(National Center for Supercomputing Applications)联合提出,起初应用于图书馆信息管理,后来被多个国家采纳并应用于生态学与环境科学(Recknagel and Michener, 2018)等多个领域中,2003年都柏林核心的元素集作为国际标准(ISO15836-2003)发布。达尔文核心(Darwin Core)是用于记录生物群及其信息的词汇和术语标准,主要应用于规范统一现代生物多样性数据,被全球生物多样性信息网络(Global Biodiversity Information Facility, 简称GBIF; <https://www.gbif.org/>)等多个平台采用(潘照晖、朱敏, 2020)。数字地理空间元数据标准(Content Standard for Digital Geospatial Metadata)由美国联邦地理数据委员会元数据工作组制定,是为数字化地理空间数据归档提供术语和定义的通用集合。ISO19115是国际标准化组织地理信息技术委员会制定和发布的《地理信息元数据》国际标准,为许多国家或组织制定和实施地理信息元数据提供参考(樊隽轩等, 2009)。上述元数据标准中仅达尔文核心中提及了古生物学扩展(Paleontology Extension)内容,其与化石标本的元数据有一定的相关性(潘照晖等, 2020),但该扩展内容中仅提供了地层单位相关元素的描述。

关于科学数据的标准,我国在生态学研究方面建立了《生态科学数据元数据》国家标准,并发布了《科技平台资源核心元数据》国家标准。目前还没有确立或发布古生物化石标本元数据的相关标准。

此前,仅河北省岩矿化石标本资源数据库制定了岩矿化石元数据标准(王颖,2009),其参照《国家科技资源平台岩矿化石资源共性描述规范》,制定了标本资源的描述标准、分级归类编码标准,但并没有对岩矿或化石标本元数据标准进行说明,即,未说明元数据实体或元素集。2014年,中华人民共和国国家质量监督检验检疫总局和中国国家标准化委员会发布了《科技平台资源核心元数据》(GB/T30523-2014),该国家标准可适用各类科技平台,适用范围广。该标准也是本文中古生物化石标本的元数据标准的指导性文件。

为了能将标准化、规范化的古生物化石标本数据提供给全球科学共同体,进一步促进古生物学与地层学新范式的科学研究工作,十分有必要在元数据的汇交、共享和使用方面确立权威性的规范与行业标准。本文基于古生物学的研究现状、中国科学院南京地质古生物研究所的化石标本收藏经验,以及中国科学院科学数据中心的建设目标,尝试性地制定了古生物化石标本元数据标准,希望以此进一步规范、整合古生物化石标本数据,使化石标本元数据符合国际通用的科学数据管理中的FAIR原则,即,可查找(Findable)、可获取(Accessible)、可操作(Interoperable)和可复用(Reusable)(邱春艳,2021),助力新范式的基础科学研究。本标准的制定引用了国家标准《GB/T30523-2014 科技平台资源核心元数据》和行业标准《WH/T 46-2012 图像数据加工规范》。凡是标注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版均不适用于本文件,然而,鼓励根据本文件达成协议的各方研究机构使用这些文件的最新版本(白殿一等,2020)。

2 古生物化石标本元数据标准

化石是地质历史时期岩层中保存的生物遗体、遗迹以及其他能指示生命存在的地质记录,化石

的实物标本由科研人员采集、整理、鉴定和保存,用于研究地球科学和生命演化。本文所涉及的化石具体包括:古脊椎动物化石、古无脊椎动物化石、古植物化石以及微体化石(有孔虫、放射虫、牙形类、介形类和孢粉化石)(何明跃等,2015;吴淦国等,2017)。本文涉及的古生物化石标本类别并未完全涵盖古生物研究中的各个门类,实际研究中涉及到的其他类型可通过拓展本标准内容添加,本标准的拓展原则严格参照《GB/T30523-2014 科技平台资源核心元数据》。

元数据标准是用于记录编目信息或结构化描述类记录的标签或编码系统(Xie and Matusiak, 2016)。元数据标准分为数据结构标准、数据值标准、数据内容标准和数据格式(技术交流)标准(Baca, 2008)。本文提出的元数据标准属于数据结构标准,数据结构标准是建立数据库的基础,据此标准,其他方面的内容也可以得到相应规范性补充。建立元数据标准是为了规范数据词汇和明确术语的定义,确保这些术语在同一学科或跨学科应用过程中的一致性。建立行业内一致接受的元数据标准可以简化建立综合的元数据的过程,也可保障元数据长期使用和共享(Recknagel and Michener, 2018)。选择最适合的元数据标准并创建一致的元数据,不仅可以形成特定资源数据的统一描述,还可映射依据其他标准建立的元数据,进一步实现互操作的目标(Baca, 2008)。

元数据标准建立包括定义元数据元素集(或描述符)和元数据实体,以实现元数据的获取、编辑和发表。1995年在美国俄亥俄州都柏林镇召开的元数据研讨首次提出了都柏林核心元数据元素集(Dublin Core Metadata Element Set),用于描述网络资源的元数据元素集,其中就包含数据主题、标题、作者、出版者、其他责任者、日期、对象类型、格式、标识符、关联、来源、语种和覆盖范围共13种元素(或描述符)(网址: <https://www.dublincore.org/specifications/dublin-core/dces/>)。都柏林核心后来为各个行业元数据标准的建立提供了翔实的参照。核心元数据是元数据最小集合,其中包含的元数据元素是元数据的基本单元,是用于解释和标识数据的基本单元。

本文参考都柏林核心元数据元素集和《科技平台资源核心元数据》国家标准(GB/T30523-2014)建立化石标本的元数据标准,并参考都柏林核心元数据术语建立化石标本元数据中涉及的术语描述。化石标本的核心元数据包括10个元数据元素(化石标本元数据元素集Fossil Specimens Metadata Element Set)和1个元数据实体,均为必选项。10项元数据元素包括标题、数据标识符、许可证标识符、关键词、描述、化石标本参数、数据链接地址、创建者、创建时间和访问限制。元数据实体包括提交机构,元数据实体中包含子元素(下一级元数据元素),分别是提交机构名称、提交机构通信地址、提交机构邮政编码、提交机构联系电话、提交机构电子信箱(表1)。本文从定义、英文名称、短名、数据类型、值域、约束/条件和注解七个方面描述各核心元数据元素,为了保障古生物化石标本元数据更好的通用性,描述方法以及核心元数据元素的英文名称均参考了《科技平台资源核心元数据》(GB/T30523-2014)的内容。

3 古生物化石标本参数术语

上述十项元数据元素之一的化石标本参数,又可分为古生物化石标本的实物信息、古生物系统分类信息、地层信息和高维信息(图1)。其中古生物化石标本的实物信息包含:是否公开发表、参考文献、项目名称、是否模式标本、标本号前缀、标本登记号、野外编号、发表图编号和标本收藏单位等术语。古生物系统分类信息包含:大类/俗名、门、纲、目、科、属、种、修订意见、鉴定人、鉴定/修订时间、命名人和定种年份等术语。地层信息包含:地质时代(宙、代、纪、世、期)、地质时代起始值、地质时代终止值、地质时代中值、岩石地层单位、岩性、剖面名称、产地(国家、省、市、县、村)、经度、纬度、生物带、沉积相和古环境等术语。高维信息分为图像信息和三维模型信息,包括提供者、采集设备名称与型号、采集日期、图像格式、扫描类型、分辨率、数据来源、数据描述、图像文件名称等术语。

以下是对化石标本参数术语的介绍,部分术语

的描述方式参考了伊利诺伊州古生物资料数字化工作组(workshop on paleontology collection digitization) 2005年提交的达尔文核心数据标准古生物扩展部分(Darwin Core Paleontology Extension) (古生物元素定义的网址: <https://github.com/tdwg/wiki-archive/tree/master/twiki/data/DarwinCore/PaleontologyElement.txt>),还包括何心一等(1993)、吴淦国等(2017)、Salvador (1994)的专著内容,《图像数据加工规范》(WH/T46-2012)、邵梦媛等(2017)和杨眉等(2017)文献。

3.1 化石标本的实物信息

是否公开发表Published (Yes/No, 简称Y/N): 化石标本是否已在文章中发表。该项数据为必选项,即,数据库中必备的化石标本信息。

参考文献Reference: 化石标本首次发表时的文献资料,应提供URL或DOI或完整的引文,完整引文包括作者名、出版时间、论文题目或专著名、刊物名或出版者、卷(期)号、页码。化石标本已经公开发表之后的必选项。

项目名称Funding project: 获取化石标本受资助的项目。该项数据为可选项,即,数据库中选填的关于化石标本的信息。在地质资源调查过程中,有的化石标本在采集过程受到某项目资助,但是未公开发表,本选项适合这种情况。

是否模式标本Type specimens (Y/N): 判定是否为模式标本。必选项。

标本号前缀Specimens prefix: 化石标本登记号的英文前缀,通常是馆藏或研究单位名称的英文缩写词,不同的化石收藏单位往往有不一样的化石标本号前缀。必选项。

标本登记号Specimens number: 标本馆藏编号的数字部分,由保存单位或系统统一编号编写。必选项。

野外编号Field number: 化石标本在野外采集时的编号。可选项。

发表图编号Published plate/figure number: 化石标本在已发表的文章或专著中的图版编号和图编号。可选项。

标本收藏单位Collection housed institute: 化石标本保存单位的名称。必选项。

表 1 古生物化石标本核心元数据元素
Table 1 Core metadata elements of alaeontological specimens

元数据基本单元	定义	英文名称	短名	数据类型	值域	约束/条件	注解
元素	标题	Title	title	字符串	自由文本	必选(M)	最大出现次数为 1
	数据标识符	DataIdentifier	dataId	字符串	自由文本, 参见 RFC 2396	必选(M)	最大出现次数为 1
	许可证标识符	LicenseURI	licURI	字符串	自由文本, 参见 RFC 2396	必选(M)	最大出现次数为 1
	关键词	Keyword	keyword	字符串	自由文本	必选(M)	最大出现次数为 N
	描述	Description	descrip	字符串	自由文本	必选(M)	最大出现次数为 1
	化石标本参数	FossilSpecimensParameter	fosSpePara	复合型	自由文本	必选(M)	最大出现次数为 N
	数据链接地址	OnlineAddress	onlAdd	字符串	自由文本, 参见 RFC 2396	必选(M)	最大出现次数为 N
	创建者	Creator	creator	字符串	自由文本	必选(M)	最大出现次数为 1
	创建时间	CreateDate	creDate	字符串	自由文本, 参见 ISO 8601:2004	必选(M)	最大出现次数为 1
	访问限制	AccessConstraints	accConsts	字符串	公开级为 1, 限制级为 2	必选(M)	最大出现次数为 1
实体	提交机构	PointOfContact	idPoC	复合型	自由文本	必选(M)	最大出现次数为 1
	提交机构名称	OrganizationName	orgName	字符串	自由文本	必选(M)	最大出现次数为 1
	提交机构通信地址	Address	cntAdd	字符串	自由文本	必选(M)	最大出现次数为 1
	提交机构邮政编码	PostalCode	postCode	字符串	自由文本	必选(M)	最大出现次数为 1
	提交机构联系电话	Phone	cntPhone	字符串	自由文本	必选(M)	最大出现次数为 N
	提交机构电子信箱	ElectronicMailAddress	cntEmail	字符串	自由文本	必选(M)	最大出现次数为 N
	提交机构名称	提交化石标本核心元数据的组织机构名称					
	提交机构通信地址	与提交机构联系的通信地址					
	提交机构邮政编码	与提交机构通信地址相对应的邮政编码					
	提交机构联系电话	提交机构联系人的联系电话					
实体子元素	提交机构电子信箱	提交机构联系人的电子邮件地址					
	提交机构名称	提交化石标本核心元数据的单位或机构					
	提交机构通信地址	(应列出提交机构名称、提交机构通信地址、提交机构邮政编码、提交机构联系电话和提交机构电子信箱)					
	提交机构名称	提交化石标本核心元数据的组织机构名称					
	提交机构通信地址	与提交机构联系的通信地址					
	提交机构邮政编码	与提交机构通信地址相对应的邮政编码					
	提交机构联系电话	提交机构联系人的联系电话					
	提交机构电子信箱	提交机构联系人的电子邮件地址					
	提交机构名称	提交化石标本核心元数据的单位或机构					
	提交机构通信地址	(应列出提交机构名称、提交机构通信地址、提交机构邮政编码、提交机构联系电话和提交机构电子信箱)					

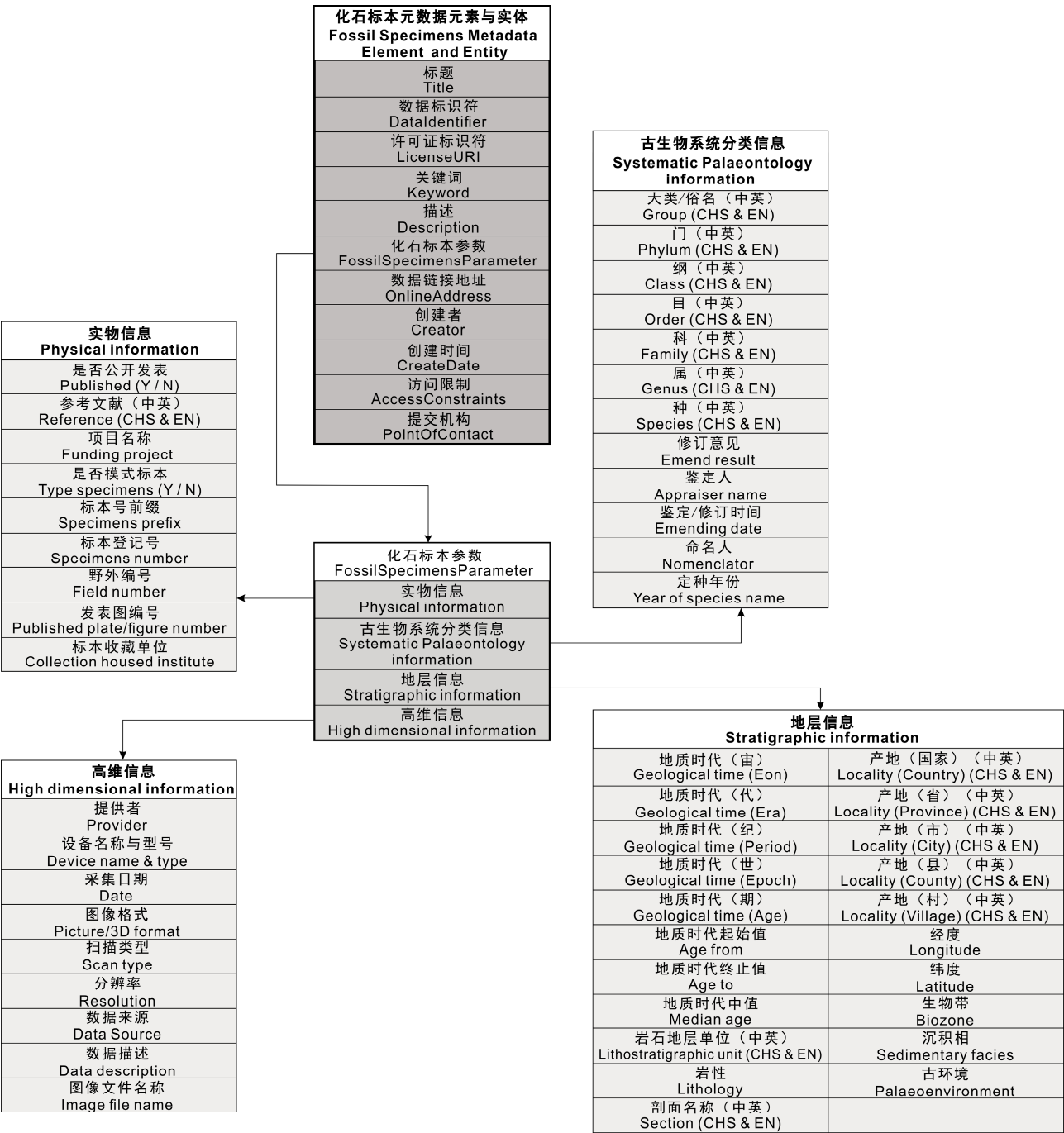


图 1 化石标本元数据结构
Fig. 1 Structure of fossil specimen metadata

3.2 化石标本的古生物系统分类信息

大类/俗名Group: 化石标本的俗称和通俗分类。可选项。

门Phylum: 生物分类法中的第三分类等级, 是位于界和纲之间的生物分类单位。包括中文名、拉丁文名。可选项。

纲Class: 位于门和目之间的生物分类单位。包括中文名、拉丁文名。可选项。

目Order: 位于纲和科之间的生物分类单位。包括中文名、拉丁文名。可选项。

科Family: 位于目和属之间的生物分类单位。包括中文名、拉丁文名。可选项。

属Genus: 是种的综合, 包括若干同源的和形态、构造、生理特征近似的种。本文指化石标本的属名。包括中文名、拉丁文名。必选项。

种Species: 是生物学和古生物学的基本分类单元。本文指化石标本的种名。包括中文名、拉丁文名。可选项。

修订意见Emend result: 已发表的关于化石标本分类位置变动的学术意见。可选项。

鉴定人Appraiser name: 鉴定化石标本分类位置的人的姓名。可选项。

鉴定/修订时间Emending date: 鉴定或修订化石标本分类位置的年份。可选项。

命名人Nomenclator: 化石标本的属种命名者的姓氏。必选项。

定种年份Year of species name: 化石标本种名命名的年份。必选项。

3.3 化石标本的地层信息

地质时代(宙Eon): 地质时代指化石标本目前已知的时代分布范围。宙是比代高一级的地质年代单位。可选项。

地质时代(代Era): 代是与界(年代地层单位)相应的地质年代单位。可选项。

地质时代(纪Period): 纪是与系(年代地层单位)对应的地质年代单位。级别高于世, 低于代。必选项。

地质时代(世Epoch): 世是与统(年代地层单位)对应的地质年代单位。必选项。

地质时代(期Age): 期是与阶(年代地层单位)对应的地质年代单位。可选项。

地质时代起始值Age from: 化石标本对应种分布的地质时代的时间起始数值, 单位为百万年(缩写词: Ma)。可选项。

地质时代终止值Age to: 化石标本对应种分布的地质时代的时间终止数值, 单位为百万年(缩写词: Ma)。可选项。

地质时代中值Median age: 取地质时代起始值和终止值两者的中间值, 单位为百万年(缩写词: Ma), 用于计算古经纬度。可选项。

岩石地层单位Lithostratigraphic unit: 对地球岩石分类时, 根据岩石的特征、属性或地质时代识别为一个单元的岩石体。岩石地层单位包括岩石地层单位中的群、组、段、层。必选项。

岩性Lithology: 地层的岩石组成及岩石特征。包括岩石类型、物质成分、颜色、厚度等特征。必选项。其中岩石类型数据可参考 GB/T 17412.2-1998 沉积岩岩石分类和命名方案国家标准。

剖面名称Section: 地表出露的岩石单位系列的名称, 钻孔岩心等岩石单位可以使用本身名称, 作为虚拟(virtual)或等效的剖面名称。可选项。

产地(国家Country): 化石标本产地所在的国家, 必选项。

产地(省Province): 采集化石标本的省级行政区划, 也包括自治区和直辖市, 必选项。

产地(市City): 采集化石标本的市级行政区划, 也包括地区、自治州、盟, 必选项。

产地(县County): 采集化石标本的县级及同级行政区划, 必选项。

产地(村Village): 采集化石标本的村级行政区划, 可选项。

产地(经度Longitude): 化石标本产地经度数值, 是一个 - 180~180 之间的数字, 负值表示西经的度数, 正值表示东经度数。可选项。

产地(纬度Latitude): 化石标本产地纬度数值。是一个 - 90~90 之间的数字, 负值表示南纬的度数, 正值表示北纬度数。可选项。

生物带Biozone: 指生物地层带, 是生物地层单位的统称。常用的生物带包括延限带、间隔带、组合带、富集带和种系带 Salvador (1994)。生物带是基于化石判断地层时代, 并开展地层对比的依据。可选项。

沉积相Sedimentary facies: 沉积环境及在该环境中形成的沉积岩(物) 特征的综合。沉积岩特征则包括岩性特征、古生物特征以及地球化学特征等(何幼斌等, 2007)。可选项。

古环境Palaeoenvironment: 地质历史时期古生物所处的环境。如海水深度、温度、盐度、水体开

放程度、含氧量、水动力条件等。可选项。

3.4 化石标本的高维信息

提供者Provider: 化石标本高维信息(图像或三维模型)提供者的姓名。必选项。

采集设备名称与型号Device name & type: 获取化石标本图像或三维模型数据的设备, 包括设备名称、性能、规格、大小。必选项。

采集日期Date: 获取化石标本图像或三维模型数据的日期。必选项。

图像格式Picture/3D format: 化石标本图像或三维模型文件的存储格式, 指示了高维数据在数据库与网站平台上的解码方式。常见的图像格式如JPG, 常见的三维模型文件格式如OBJ文件等。必选项。化石标本图像需要具有可参照的比例尺。拍摄过程中应适当补光, 避免化石标本投射阴影, 且应避免因环境因素导致化石标本颜色失真。其中微体化石标本图像编码格式为JPG或TIFF, 保留扫描电镜拍摄中设置的比例尺。

扫描类型Scan type: 化石标本图像或三维模型的扫描方式。可选项。

分辨率Resolution: 指示了化石标本图像或三维模型中存储的信息量, 决定图像或三维模型的清晰度。可选项。

数据来源Data source: 化石标本收藏单位或持有者、或化石标本图像或三维模型数据的文献或收藏平台。必选项。

数据描述Data description: 化石标本图像或三维模型数据科学信息。可选项。

图像文件名称Image file name: 化石标本图像或三维模型数据在电脑或存储系统中的文件标识。必选项。

4 应用拓展及案例

数据标准分为数据结构标准、数据值标准、数据内容标准和数据格式(技术交流)标准(Baca, 2008)。数据结构方面的标准化与规范化是元数据

中其他项目标准化的首要基础, 数据结构极大程度上限定甚至决定了数据的值、内容和格式。

本文提出的古生物化石标本元数据标准主要强调了化石标本的数据结构。据此结构, 可以有针对性地采集化石标本数据的相关要素。化石标本的实物信息、古生物系统分类信息和地层信息往往可以提供通过化石标本的标签或已经发表的文献中采集获得, 这些内容中除了描述文本以外, 都可以转变为结构化数据, 其数据值和内容都是易于表格化和规范化的结构性数值或文本, 而结构性的数据格式目前是大型数据最易于开展信息交流的技术, 其数据库开发方面的技术已经相当成熟。

化石标本的高维信息(图像或三维模型)主要是非结构化的数据, 其在采集、检索、查询和共享等方面并不普遍, 一件化石标本往往对应多个图像或三维模型, 其数据值、内容与格式都需要专门进行设计, 以便于数据的安全性、易用性及其与化石标本的其他信息相关联。

在本文的古生物化石标本元数据核心元素中, 首次提出了数据标识符的概念。其目的在于为每一块化石标本匹配唯一的标识, 用户可以通过该标识检索其元数据。效果类似于数字对象标识符(DOI: digital object identifier)的应用情况, 便于在信息时代对化石标本数据的查询、检索与使用。该数据标识符的可行性以及具体的实施仍需要一定的思索与调研, 笔者团队目前正在开展这方面的研究工作。本文以一块笔石化石标本为例(表2, 图2), 说明该元数据标准的应用情况。

基于古生物化石标本元数据标准, 笔者团队收集、汇交了笔者单位收藏的1550块笔石化石标本的科学数据, 这些化石标本代表了奥陶纪到志留纪地质时期中113种笔石, 据此开发了古生物化石标本的科学数据库和化石标本可视化系统, 用户可以通过该可视化系统的网页端开展古生物化石标本的综合查询, 也可以查看化石标本的高分辨率图像(图3)。

表 2 古生物化石标本核心元数据示例

Table 2 An example of the core metadata of palaeontological fossil specimens

标题	14545 号古生物化石标本数据
数据标识符	FSI.gra.nigp.14545
许可证标识符	open
关键词	标本登记号、属、剖面、地质时代、图像
描述	标本登记号 14545 化石标本数据, 包含该标本的实物信息、古生物信息、地层信息和高维信息
化石标本参数	用途: 研究、科普
提交机构名称	标本登记号 14545
提交机构通信地址	中国科学院南京地质古生物研究所大数据中心
提交机构通信地址	南京市北京东路 39 号
提交机构邮政编码	210008
提交机构联系电话	025-83282230
提交机构电子邮箱	hhxu@nigpas.ac.cn
数据链接地址	http://fsidvis.fossil-ontology.com:8089
创建者	徐洪河
创建时间	Wed 2021-12-29T09:02:50+08:00
访问限制	1

实物信息			
是否公开发表	是		
参考文献	杨达铨, 1964. 浙江安吉下志留统中的几种笔石. 古生物学报. 12: 628–635 (in Chinese with English summary). DOI: 10.19800/j.cnki.aps.1964.04.005		
Reference	Yang Da-quan, 1964. Some Lower Silurian graptolites from Anji, northwestern Zhejiang (Chekiang). Acta Palaeontologica Sinica. 12: 628–635. DOI: 10.19800/j.cnki.aps.1964.04.005		
项目名称	中国科学院先导专项 A 项目 XDA19050101		
是否模式标本	否	标本号前缀	NIGP
标本登记号	14545	野外编号	Ya18
发表图编号	Pl. I Fig.6–11	标本收藏单位	中国科学院南京地质古生物研究所标本馆
古生物系统分类信息			
大类/俗名	Graptolite	门	Hemichordata
纲	Graptolithina	目	Graptoloidea
科	Dimorphograptidae	属	Akidograptus
种	Akidograptus ascensus	修订意见	
鉴定人	杨达铨	鉴定/修订时间	1964
命名者	Davies	定种年份	1929
地层信息			
宙	显生宙	代	古生代
纪	志留纪	世	兰多维列世
期	鲁丹期	地质时代起始值	443.8
地质时代终止值	440.8	地质时代中值	442.3
岩石地层单位	龙马溪组	岩性	页岩
剖面名称	黄墅村	产地(国家)	中国
产地(省)	浙江	产地(市)	湖州
产地(县)	安吉	产地(村)	黄墅
经度	119.676	纬度	30.608
生物带	Akidograptus ascensus 带	沉积相	
古环境			
高维信息			
图像信息			
提供者	徐洪河	设备名称与型号	Nikon D800E
采集日期	2021-10-22		
图像格式	jpg	扫描类型	反射光影拍摄标本
分辨率	300 dpi	数据来源	中国科学院南京地质古生物研究所地层古生物大数据中心
数据描述	Akidograptus ascensus 幼年标本	图像文件名称	14545Akidograptus_ascensus.jpg
三维模型			
提供者		设备名称与型号	
图像格式		扫描类型	
分辨率		数据来源	
数据描述		图像文件名称	

图 2 古生物化石标本参数数据示例

Fig. 2 An example of parameter data of a palaeontological specimen

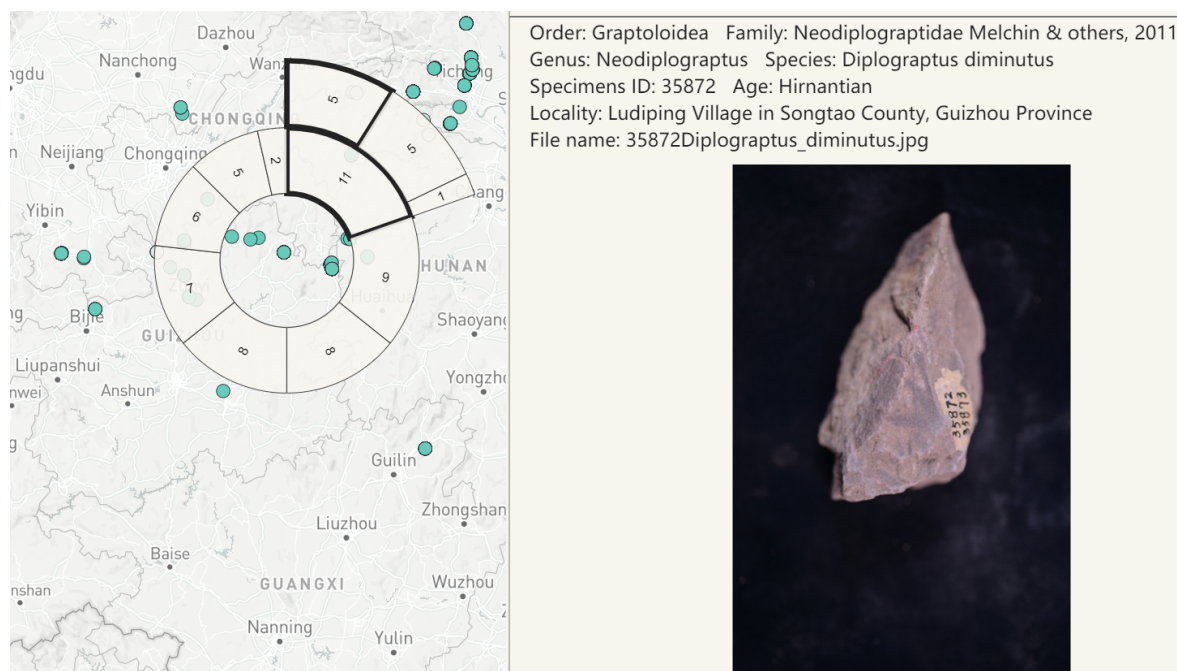


图3 化石标本图图像数据库可视化系统(FSIDvis, Fossil specimen image data visualizer, <http://fsidvis.fossil-ontology.com:8089/>) (徐洪河等, 2021)显示界面。通过该软件可以查询到化石标本的古生物学分类名称、地质时代、化石产地、标本编号、图像等信息

Fig. 3 Interface of the FSIDvis (Fossil specimen image data visualizer, <http://fsidvis.fossil-ontology.com:8089/>) (Xu *et al.*, 2021), showing information of an individual fossil specimen, including classification, geological age, locality, serial number, and image

致谢 评审专家和学科副主编提出宝贵修改建议, 在此一并感谢!

参考文献 (References)

- 白殿一, 刘慎斋, 王益谊, 杜晓燕, 李佳, 逢征虎, 2020. 标准化文件的起草. 北京: 中国标准出版社, 1-405.
- 诸云强, 孙凯, 杨雅萍, 王筱萱, 乐夏芳, 朱华忠, 宋佳, 何明跃, 陈佑启, 彭克银, 付磊, 杨彦臣, 杨眉, 邹金秋, 陈艳, 李金斌, 刘燕, 雷蕾, 杨杰, 李威蓉, 张金区, 2017. 科技基础性工作数据资料的汇交与整编. 中国科技资源导刊, 49: 12-20.
- 樊隽轩, 迟昭利, 陈峰, 陈清, 2009. 元数据标准及其在古生物数据库中的应用. 地层学杂志, 33: 391-397.
- GB/T 30523-2014 科技平台 资源核心元数据
- 何明跃, 吴淦国, 曹希平, 高星, 杨群, 2015. 岩矿化石标本资源收集整理保存技术规程. 北京: 地质出版社, 1-435.
- 何心一, 徐桂荣, 1993. 古生物学教程. 北京: 地质出版社, 1-255.
- 何幼斌, 王文广, 2007. 沉积岩与沉积相. 北京: 石油工业出版社, 1-378.
- 潘照晖, 朱敏, 2020. 国内外古脊椎动物数据库综述. 高校地质学报, 26: 424-443.
- 邱春艳, 2021. 国内外科学数据相关实践中的元数据研究进展. 情报资料工作, 42: 104-112.
- 邵梦媛, 杨眉, 何明跃, 2017. 古生物群专题数据库的构建与展示. 科研信息化技术与应用, 8: 77-82.
- 王颖, 2009. 河北省岩矿化石数据库的构建与元数据规范研究. 石家庄: 河北师范大学硕士研究生学位论文, 1-58.
- WH/T 46-2012 图像数据加工规范
- 吴淦国, 何明跃, 杨良峰, 杨群, 刘金毅, 朱祥根, 2017. 脊椎动物、植物、牙形类化石及旧石器标本资源描述标准. 北京: 地质出版社, 1-255.
- 吴志远, 2015. 国家岩矿化石标本资源共享平台研究及应用. 中国地质大学(北京), 博士学位论文, 1-121.
- 谢园, 杨眉, 何明跃, 吴志远, 2017. 科学标本资源汇交整编规范与实践. 中国科技资源导刊, 49: 21-29.
- 徐洪河, 牛志彬, 陈焱森, 2021, 软件名称: Fossil specimen image data visualizer 软件 v1.0, 登记号 2021SR1333435, 证书号: 软著登字第 8056061 号.
- 杨达铨, 1964. 浙江安吉下志留统中的几种笔石. 古生物学报, 3: 628-637.
- 杨眉, 何明跃, 施爽, 严薇, 2017. 地学类标本资源的规范数字化及质量控制. 中国科技资源导刊, 49: 45-51.
- Alroy J, Marshall C R, Bambach R K, Bezusko K, Foote M, Fusich F T, Hansen T A, Holland S M, Ivancy L C, Jablonski D, Jacobs D K, Jones D C, Kosnik M A, Lidgard S, Low S, Miller A I, Gottshall P M, Olszewski T D, Patzkowsky M E, Raup D M, Roy K, Sepkoski J Jr, Sommers M G, Wagner P J, Webber A, 2001. Effects of sampling standardization on estimates of Phanerozoic marine diversification, Proceedings of the National Academy of Sciences of the United States of America, 98: 6261-6266. DOI: 10.1073/pnas.111144698.
- Baca M, 2008. Introduction to Metadata. Odyssey Press Inc. 1-81.
- Bai Dian-yi, Liu Shen-zhai, Wang Yi-yi, Du Xiao-yan, Li Jia, Pang Zheng-hu, 2020. Drafting of standardizing documents. Beijing: Standards Press of China. 1-405 (in Chinese).

- Berners-Lee T, Fielding R, Masinter L, Uniform Resource Identifiers (URI): Generic Syntax. RFC Editor, 1998.
- Darwin Core Paleontology Element Definitions v1.0. 2005-07-03. <https://github.com/tdwg/wiki-archive/tree/master/twiki/data/DarwinCore/PaleontologyElement.txt>.
- DCMI Metadata Terms: Dublin Core Metadata Element Set, Version1.1. 1995-07. <http://www.dlib.org/dlib/July95/07weibel.html>
- Fan Jun-xuan, Chi Zhao-li, Chen Feng, Chen Qing, 2009. Application and discussion of metadata in Palaeobiology Geobiodiversity Database. *Journal of Stratigraphy*, 33: 391–397 (in Chinese with English abstract).
- GB/T 30523-2014 Science and technology infrastructure – resource core metadata.
- Guo Hua-dong, 2017. Big Earth data: A new frontier in Earth and information sciences, *Big Earth Data*, 1: 4–20. DOI: 10.1080/20964471.2017.1403062.
- Hammer O, Harper D A T, 2006. *Paleontological Data Analysis*, Blackwell Publishing, Oxford. 1–351. DOI: 10.1002/9780470750711.
- He Ming-yue, Wu Gan-guo, Cao Xi-ping, Gao Xing, Yang Qun, 2015. The standard technological guideline of the collection, arrangement and preservation of rock, mineral and fossil specimens resources. Beijing: Geological Publishing House. 1–435 (in Chinese).
- He Xin-yi, Xu Gui-rong, 1993. *Course of Paleontology*. Beijing: Geological Publishing House. 1–255 (in Chinese).
- He You-bin, Wang Wen-guang, 2007. *Sedimentary rocks and sedimentary facies*. Beijing: Petroleum Industry Press. 1–378.
- ISO 8601:2004 Data elements and interchange formats-Information interchange-Representation of dates and times.
- Pan Zhao-hui, Zhu Min, 2019. VPPDB: hosting the data for vertebrate paleontology and paleoanthropology. *Acta Geologica Sinica (English Edition)*, 93 (supp.1): 61–63. DOI: 10.1111/1755-6724.14246.
- Pan Zhao-hui, Zhu Min, 2020. Review of databases in vertebrate paleontology. *Geological Journal of China Universities*, 26: 424–443 (in Chinese with English abstract). DOI: 10.16108/j.issn1006-7493.2020022.
- Peters S E, McClennen M, 2016. The Paleobiology Database application programming interface. *Paleobiology*, 42: 1–7. DOI: 10.1017/pab.2015.39.
- Qiu Chun-yan, 2021. Metadata research progress in scientific data-related practices at home and abroad. *Information and Documentation Services*, 42: 104–112 (in Chinese with English abstract). DOI: 10.12154/j.qbzlgz.2021.05.012.
- Recknagel F, Michener W K, 2018. *Ecological Informatics*. 1–428.
- Rong Jia-yu, Fan Jun-xuan, Miller A I, Li Guo-xiang, 2007. Dynamic patterns of latest Proterozoic-Palaeozoic-early Mesozoic marine biodiversity in South China. *Geological Journal*, 42: 431–454. DOI: 10.1002/gj.1073.
- Salvador A, 1994. *International Stratigraphic Guide*. Boulder, Colorado: Geological Society of America.
- Shao Meng-yuan, Yang Mei, He Ming-yue, 2017. Construction and display of palaeobiota subject database. *e-Science Technology & Application*, 8: 77–82 (in Chinese with English abstract). DOI: 10.11871/j.issn.1674-9480.2017.04.011.
- Wang Cheng-shan, Hazen R M, Cheng Qiu-ming, Stephenson M H, Zhou Cheng-hu, Fox P, Shen Shu-zhong, Oberhänsli R, Hou Zeng-qian, Ma Xiao-gang, Feng Zhi-qiang, Fan Jun-xuan, Ma Chao, Hu Xiu-mian, Luo Bin, Wang Juan-le, Schiffries C M, 2021. The Deep-Time Digital Earth program: data-driven discovery in geosciences. *National Science Review*, 8: nwab027. DOI: 10.1093/nsr/nwab027.
- Wang Yin, 2009. The rock, mineral, fossil database construction and origin-data standard research of Hebei Province. Master Dissertation. 1–58 (in Chinese with English abstract).
- WH/T 46-2012 Image data reformatting and transferring specification and guideline.
- Wu Gan-guo, He Ming-yue, Yang Liang-feng, Yang Qun, Liu Jin-yi, Zhu Xiang-gen, 2017. *The descriptive standard of vertebrate, plant, conodont fossils and Paleolithic specimens resources*. Beijing: Geological Publishing House. 1–255 (in Chinese).
- Wu Zhi-yuan, 2015. Research and application on national infrastructure of mineral rock and fossil specimen resources. China University of Geosciences (Beijing), PhD Dissertation. 1–121 (in Chinese with English abstract).
- Xie I, Matusiak K, 2016. *Discover digital libraries: theory and practice*. Elsevier. 73–88.
- Xie Yuan, Yang Mei, He Ming-yue, Wu Zhi-yuan, 2017. Aggregation standard and experience of scientific specimen resources. *China Science & Technology Resources Review*, 49: 21–29 (in Chinese with English abstract). DOI: 10.3772/j.issn.1674-1544.2017.05.003.
- Xu Hong-he, Niu Zhi-bin, Chen Yan-sen, 2020. A status report on a section-based stratigraphic and palaeontological database—the Geobiodiversity Database. *Earth System Science Data*, 12: 3443–3452. DOI: 10.5194/essd-12-3443-2020.
- Xu Hong-he, Niu Zhi-bin, Chen Yan-sen, 2021. Software title: Fossil specimen image data visualizer v1.0. Register No. 2021SR 1333435. Certificate No. 8056061.
- Yang Da-quan, 1964. Some lower Silurian graptolites from Anji, northwestern Zhejiang (Chekiang). *Acta Palaeontologica Sinica*, 3: 628–637 (in Chinese with English abstract). DOI: 10.19800/j.cnki.aps.1964.04.005.
- Yang Mei, He Ming-yue, Shi Shuang, Yan Wei, 2017. Digitization and quality control of geological specimen resources. *China Science & Technology Resources Review*, 49: 45–51 (in Chinese with English abstract). DOI: 10.3772/j.issn.1674-1544.2017.05.006.
- Zhu Yun-qiang, Sun Kai, Yang Ya-ping, Wang Xiao-xuan, Yue Xia-fang, Zhu Hua-zhong, Song Jia, He Ming-yue, Chen You-qi, Peng Ke-yin, Fu Lei, Yang Yan-chen, Yang Mei, Zou Jin-qiu, Chen Yan, Li Jin-bin, Liu Yan, Lei, Yang Jie, Li Wei-rong, Zhang Jin-qu, 2017. Data resources collection and reorganization for national special program on basic works for science and technology of China. *China Science & Technology Resources Review*, 49: 12–20 (in Chinese with English abstract). DOI: 10.3772/j.issn.1674-1544.2017.05.002.

(责任编辑: 黄 冰)