

· 科学论坛 ·

科学研究中原始数据的记录与保存

方玉东^{1*} 陈越¹ 常宏建² 陈克勋³

(1 国家自然科学基金委员会 纪检监察审计局, 北京 100085; 2 山东财经大学 工商管理学院, 济南 250014; 3 山东理工大学 化学工程学院, 淄博 255049)

[摘要] 原始数据的记录与保存,是科学实验的重要环节,是形成科研成果的重要基础。本研究调查了57个课题组的实验原始数据的记录和保存情况,分析发现原始数据的记录和保存存在数据所有权认识模糊、缺乏对研究生的指导、缺乏制度和规范、缺乏监管等问题。提出从规范原始数据记录、科学保存原始资料提高数据使用效率、明确数据记录和保存各行为主体的权利和义务、加强教育和监管四个方面加强和完善原始数据保存的建议。

[关键词] 科学研究,原始数据,记录,保存

在万方期刊论文数据库中,1990年至2014年间,在“题名和关键词”下搜索“原始数据”,共搜索到315篇论文,在“关键词”下搜索“原始数据”,共搜索到201篇论文。分析发现,这些论文主要是研究信息科学领域计算数据的模拟等算法问题,未见针对科学实验过程中原始数据的记录与保存进行研究的学术论文。调查表明,很多科研领域缺乏原始数据记录和保存的标准,只能通过“自我摸索-不断犯错-纠错”的方式来学习数据记录和保存^[1]。原始数据是学术写作的基础,科学家据此可以撰写报告、论文、项目和专利申请。如何科学地记录和保存原始数据是摆在所有科研人员面前的一个重要现实问题。本文就原始数据的记录与保存问题谈一下粗浅的看法。

1 科学研究中的原始数据概述

1.1 什么是原始数据

数据,百度搜索引擎释义为进行各种统计、计算、科学研究或技术设计等所依据的数值。数据有多种形式,实验数据记录本里保存的手写或者打印的测量、观察、计算、解释和结论的记录和报告,我们称之为“无形数据”,或文本数据。另外,还存在“有形数据”,指实验中直接测量得到的数据经过分析产生的图形、图表等,以及其它的有形体现,用来描述

材料,比如细胞、生物标本、图片等^[1]。科学研究中原始数据一般是指未经技术手段处理的,在观察、实验或调研过程中获得的纸质或以计算机为载体形成的文字、图像、声音等数据。

1.2 原始数据的记录和保存

记录原始数据,通常采用的方法是纸质数据记录本、电子记录和数据库保存。由于这些方法各有利弊,它们分别适用于不同科研数据的保存。

(1) 数据记录本是记录科学研究过程的、连续编码的装订本。它是一份记录研究人员实验、观察以及对物理现象最大程度的理解过程中发生的思考和行动的书面材料。纸质记录保存是首选的原始数据记录方法。该方法传统、方便、及时,是绝大多数实验室采用的一种记录原始数据的方法。

(2) 电子记录方式。个人电脑的广泛应用使得电子实验室数据记录本逐渐得到广泛应用,一些科学家使用软件系统记录数据,比如文字信息处理程序、电子数据表程序等。采用电子数据表中的运算工具、搜索功能能够更为方便的处理数据,扫描仪、画图程序和电子数据输出工具的出现,逐步推动了电子记录和保存技术的应用,部分代替了传统的纸质数据记录本^[1]。

(3) 数据库保存。在生命科学研究领域每天都产生海量的DNA、RNA和蛋白质的数据,这类原始

* fangyd@nsfc.gov.cn

本文于2014年4月17日收到。

数据的保存方式一般有数据库保存和电子记录保存。因为不少期刊规定必须把预发表文章中提到的相关数据储存信息明确标明,以利于其他科研人员的使用、查找,所以这些数据一般被保存于学术界认可的数据库中,例如 GENBANK、DDBJ 等。这种保存方式是科研人员接受并认可的一种方式,而且能够实现数据共享,但数据库能够保存的数据类型是有限的。

2 原始数据记录与保存现状

为了初步了解我国科学家科研活动中原始数据记录和保存情况,为规范原始数据记录和保存提供合理建议,本研究调查了 57 个课题组的实验记录和保存现状。

2.1 基本情况

本研究共调查了 57 个课题组,分别属于地球科学、食品科学与技术、林学、农学 4 个学科。具体学科分布如下:

表 1 调查课题组所属学科分布

学科	地球科学	食品科学与技术	林学	农学
数量	21	2	3	31

2.2 关于数据的所有权的认识

调查表明,作为大部分原始数据产生的第一责任人(研究生),44 人(77.2%)希望离开目前单位时能够带走记录或副本,55 人(96.5%)认为实验记录对今后的工作学习会有帮助。这说明大家都意识到了原始数据的重要性,但是也凸显了研究人员对科研数据所有权的模糊认识。

2.3 数据记录本的使用

57 个课题组中,有 38 个(66.7%)课题组使用本研究机构统一的专用数据记录本。54 个(94.7%)课题组表示所使用的数据记录本纸张完整,没有缺页或挖补。

2.4 数据的记录

(1) 关于数据记录的真实性。51 个(89.5%)课题组认为能够真实、及时、准确、完整地记录实验数据,不存在漏记和随意涂改,不存在伪造、编造数据。41 个(71.9%)课题组在修改、增减或删除数据时有较为规范的程序。

(2) 关于数据的完整性。19 个(33.3%)课题组认为能够较为全面的记录实验内容,包括实验名称、目的、时间、方案、过程、观察指标、结果等。

(3) 关于实验记录的规范性。46 个(80.7%)课

题组在实验记录方面较为规范,能够使用专业术语及规范的计量单位。49 个(86%)课题组实验数据的书写用字规范,字迹工整,没有使用铅笔等容易涂改的笔。但是,也有 17 个(29.8%)课题组表示实验记录中存在无所适从的现象,没有可以参照的规范,想怎么记就怎么记。

(4) 关于有形数据。38 个(66.7%)课题组在有形数据记录和保存方面较为规范,能够将有关图表和数据资料按顺序粘贴在数据记录本的相应位置上,并作规范的标注;不宜粘贴的,进行编号、标注,整理装订成册,以备核查。

2.5 数据的保存

57 个课题组均使用了纸质数据记录本,55 个(96.5%)课题组能够妥善保存实验数据记录本,使其免受水浸、墨污、卷边,保持整洁、完好、无破损、不丢失。

3 数据保存和记录中存在的问题分析

通过上述调查,发现在数据记录和保存中主要存在以下问题:

3.1 数据所有权认识模糊

96.5%的受调查人员表示实验记录对今后的工作和学习有益,所以 77.2%的人表示离开岗位后想带走实验记录或副本,对实验数据的所有权认识模糊。一般来讲,受资助项目所收集的科研数据归所在大学或科研机构所有,参加该项目研究的负责人、研究生以及其他工作人员所使用的数据手册,都是该大学或科研机构的资产。研究人员在完成项目、离开实验室从事新的工作时,不允许带走这些原始数据手册。但是如果需要安全备份、为撰写文章或报告提供信息,则可以带走原始数据的复制件。拿走数据复制件应该征得项目负责人的同意^[1]。

3.2 缺乏制度和规范

对于原始数据的记录与保存不同单位有不同的管理方式,据笔者调查了解有的学校和研究部门有统一的要求规定,有的以院系为单位统一要求管理,有的以学科为单位统一管理,甚至有的只是实验室内部统一管理,更有甚者从来没有其他人关心过原始数据,只是由研究者本人使用并自我管理,这样就给科学研究工作带来了极大的不安全和不稳定因素。本次调查发现,近 20%的课题组实验记录没有使用规范的专业术语,没有采用国际标准计量单位,这意味着一旦从事实验的工作人员离开工作岗位,这些实验数据的有效识别、使用将存在很大的困难。

14%的课题组实验数据的书写用字不规范,字迹不工整,有时使用铅笔等容易涂改的笔。29.8%的课题组实验记录过程中没有固定规范,随意性很大。28.1%的课题组存在随意删除、增减数据或修改过程不规范的现象。

3.3 缺乏对研究生的指导

由于缺乏数据记录规范和保存标准,研究生在数据记录和保存方面很少受到指导,36.8%的课题组表示想让实验数据记录更加规范,但是不知道如何做起。38.6%的受调查人员表示从来没有人告知该怎样记录实验数据。

3.4 缺乏监管

科学实验过程中没有监管人员,造成数据记录和保存存在很大的随意性。29.8%的受调查人员表示实验过程中数据记录随意性很大,想怎么记就怎么记,自己认为比较重要的就记录下来,认为不重要的就不做记录。36.8%的课题组表示不了解相关管理规定和规范。

4 改善原始数据记录与保存的建议

4.1 规范原始数据记录

无论采用哪种既定方式记录原始数据,都应该确保原始数据的“原汁原味”。即保证记录具有如下特性:原始性,不允许有任何改动;及时性,对数据的记录一定要及时,不能事后凭回忆记录,以免出现差错;完整性,包括实验材料、方法、过程、结果等不能有疏漏,即使实验失败没有结果也应该有相应的记录;系统性,连续记录实验结果,不能间断;客观性和真实性,不带有任何人为主观因素。

在目前暂未出台权威性的操作指南之前,实现原始数据记录的规范性,可以参照国家药品监督管理局《药品研究实验记录暂行规定》^[2]中关于“实验记录的内容”的条款中的相关规定,实验记录的内容通常应包括实验名称、实验目的、实验设计或方案、实验时间、实验材料、实验方法、实验过程、观察指标、实验结果和结果分析等内容。相关机构可以参阅该规定制定符合本研究领域的记录内容,制订相关规章制度,不断完善实验记录内容。归纳起来,首先要明确研究课题的名称,每次实验时间和记录时间。做了哪些实验?使用的实验材料及仪器有哪些?实验方法是什么?有没有参考文献?参考文献的实验方法有没有改进?如何改进的?实验过程中的现象是什么?实验结果如何?本次实验结果与预期结果有无差异?与已发表类似结果有无差异?分

析原因并注明参考文献。记录原始数据时,应该保证记录时间与实验时间的一致性,记录人与实验人的一致性。

4.2 科学保存原始资料,提高数据使用效率

研究数据保存的最终目的是为了查阅和二次利用,因此,应科学地保存原始资料,以提高数据使用效率。

(1)对原始数据进行统筹管理。目前科研单位保存原始数据主要由实验室自身保存,这限制了对数据的交流和使用效率。科研单位可以统筹规划,发挥其他部门如图书馆的作用,对原始数据按照保密级别、研究领域和创建时间分别保存,对于保存期限也可以分别对待。如果能够把与某些课题相关的研究成果如论文、专利等共同保存,则可以加大对科研数据进行再利用,也能大幅度提高科研效率。

(2)改进纸质数据和记录方法。调查发现实验记录多数都会采用纸质记录,但是有近30%的被调查者没有统一的实验记录本,研究生个人自行选择一个笔记本作为实验记录本,对于纸张和写字笔也没有严格要求。因为实验记录需要长期保存并可能多次使用,所以需要采用正规的实验记录本,并且由管理部门统一制定、编号、发放并回收存档。要求记录本耐翻、耐磨,对书写用笔必须使用稳定性好、不易褪色的签字笔和碳素墨水笔书写,不允许有涂改的痕迹。要及时记录数据,不要依靠事后回忆。另外,要注意在纸质文字记录和图片记录相结合,即把相关图片粘贴在文字记录的相应位置,因为如果两者分开保存,往往容易丢失或找不到相对应图片。有的图片能够通过实验补回来,有的图片是补不回来的。

(3)采用电子记录和保存。电子记录和保存数据记录方便、信息储量大、运算功能强,尤其是在图形数据和图像数据的记录和保存方面,具有纸质保存无可比拟的优势。可以从以下几个方面采用电子记录和保存数据:一是可以利用商业化的数据库软件系统作为电子数据记录本。有些数据库可以随意输入大量图表和电子数据,能够记录和保存大量没有结构化的信息,可以用来存储科研数据;二是可以利用一些程序作为一般性的电子记录本。有些商业化程序设计的目的是为了为了满足记录和信息收集的需要,其方法与科研数据记录方法一致,可以进一步开发,用来记录可保存科研数据;三是利用专用的电脑软件。这些专门软件通常都具有文字处理系统、电子数据表程序、画图和数据库程序的功能,可以用来满足科研数据的记录和保存;四是加强数据

安全保护。电子记录容易复制,且容易丢失、被篡改,所以应加强数据安全保护措施;五是加强电子记录保存的软件监督。考虑到知识产权问题,电子数据记录如何见证、如何标注时间和日期是面临的难题,需要加强电子记录保存的软件监督系统的开发和利用^[1]。

4.3 明确数据记录和保存各行为主体的权利和义务

要想保障科学研究中原始资料的真实性、可靠性,并能够实现资源共享,必须由科研相关的各主体人员明确各自的责任与义务,认清自己的定位,并相互配合。王聪和陈越在《科学研究数据保存政策中行为主体的定位》一文中明确阐述了各行为主体的权利与责任^[3]:作为科技资源的分配部门,科研项目资助机构应督促项目依托单位建立与研究数据保存相关的政策,制定研究数据保存最低年限,对于某些特别的数据,建议研究数据保存地点,并在必要时检查所资助项目产生的研究数据;大学与研究机构是研究数据的产生地,也是一部分研究数据的保存地,是研究数据的所有者,应该为研究数据的保存提供设施或保存地点,同时规范研究数据转移;项目负责人对数据的记录、保存和向其他人提供数据的活动负直接责任,同时承担监督研究数据保存的责任;研究人员应遵守相关部门和所在机构关于原始数据保存的有关规定,在数据记录、保存过程中遵守相关规范,以保证研究数据的真实性与可理解性。

虽然各行为主体各负其责,但是项目资助机构、项目依托单位履行职责和义务时困难更大,目前还是一种松散式管理和制度化监督。直接参与科学研究的人员即课题组成员加强对于相关规章制度的学习,明确责任和义务,自觉遵守规章制度,才能够更大程度上保证研究数据的真实性和可靠性。

4.4 加强教育和监督管理

(1) 加强科研人员尤其是研究生的科研诚信教育。在实际工作中,实验记录的重要性及其作用、规范的实验记录、实验记录内容等问题没有受到研究人员的足够重视,尤其是刚刚涉足科研的研究生,但是国内外大多数科研单位产生研究数据的直接人恰恰是研究生。所以,加强对于研究生的教育和监督,培养良好的科研诚信态度是保证原始数据真实性、可靠性的根本所在。教育是根本,负责任的科研行为教育亟待纳入大学和研究生教育课程^[4]。从研究

生入校就应该对他们教育和引导。首先,教育研究生在思想上树立正确的科学理念,懂得认真做好科学研究包括每一个实验每一次实验记录,懂得原始资料是科学研究的基础,是科学研究的生命线;其次,以其他研究人员的行为和方式去引导刚入校的研究生,让他们意识到不良行为后果的严重性,懂得自觉约束自己,自觉做好实验和记录。研究生是产生原始数据的第一人,导师及其他科研人员不可能每天关注每个学生的实验记录,只有把科研诚信作为一个习惯,把原始记录作为一种使命,把认真实验作为一种责任,才能自觉地做好原始资料记录。

(2) 加强对科研项目负责人的监督,使其切实负起相关责任。科研项目的负责人必须具有高度的责任心,必须高度重视原始数据的记录与保存,不断加强自身科研业务管理的水平。督促项目负责人对原始数据的记录加强管理监督,对新加入科研队伍的人员及时教育。统一发放指定的实验记录本,可以采取定期检查、定期汇报的方式,或者每个研究项目指定专门人员对原始数据的记录进行监管。

(3) 科研管理部门和科研机构应加强科研工作的监管。2009年科技部、教育部等10部门发布的《关于加强我国科研诚信建设的意见》及2007年中国科协发布的《科技工作者科学道德规范(试行)》等规定中都把原始数据记录篡改等作为科研不端行为,但一直没有引起科研工作者的重视。各级项目管理单位要把对原始数据的监督管理作为一项常规工作,不能只是在出现产权纠纷时,出现造假和篡改数据的事情后通过查询原始数据解决问题,需要从根本上杜绝这些问题的发生。所以必须健全原始数据管理制度,并逐步引导科研人员能够自觉学习和遵守相关规章制度,适当的时候可以给予严厉的处罚,把对原始资料记录的考核纳入到科研课题考核中去,在进行科学考评的基础上建立奖惩制度。

参 考 文 献

- [1] Francis L. Macrina 著,何鸣鸿,陈越等译. 科研诚信——负责任的科研行为教程与案例. 北京: 高等教育出版社, 2011.
- [2] 国家药品监督管理局. 药品研究实验记录暂行规定. <http://www.100md.com/Html/Dir0/15/09/47.htm>, 2000-1-3.
- [3] 王聪,陈越. 科学研究数据保存政策中行为主体的定位. 中国科学基金, 2012, 26(4): 217-222.
- [4] 罗晖,陈宜瑜. 负责任的科研行为有赖教育. 科技日报, 2011-8-30.

Record and Storage of Original Data in Science Research

FANG Yudong¹ CHEN Yue¹ CHANG Hongjian² CHEN Kexun³

(1 National Natural Science Foundation of China, Beijing 100085; 2 Shandong University of Finance and Economics, Jinan 250014;
3 Shandong University of Technology, Zibo 255049)

Abstract The original data is the core of scientific research, and it is an important foundation for research. We carried out a questionnaire survey of 57 research groups about the status and problems exist in the record and storage of original data. We analyzed the results of the questionnaire survey carefully and found that there exist lots of problems such as unclear recognition of the ownership of original data, the absence of guidelines for graduate students, and the lack of standards, regulations and supervisions. Rationalization proposals about how to ensure the authenticity of the original data, how to improve the efficiency use of the original data, and how to clarify the rights and obligations of individuals involved in data record and storage were propounded. The emphasis of strengthening the training and supervision of data record and storage was proposed as well.

Key words scientific research, original data, record, storage

· 资料信息 ·

2014年度国家自然科学基金项目申请数量前20位的依托单位:生命科学部

序	单位名称	项数	序	单位名称	项数
1	中国农业大学	329	11	中国热带农业科学院	222
2	浙江大学	326	12	内蒙古农业大学	216
3	西北农林科技大学	325	13	西南大学	216
4	华中农业大学	323	14	石河子大学	197
5	南京农业大学	313	15	复旦大学	196
6	中国科学院上海生命科学研究院	303	16	福建农林大学	186
7	四川农业大学	250	17	北京大学	182
8	上海交通大学	241	18	江苏省农业科学院	164
9	华南农业大学	229	19	河南农业大学	159
10	东北农业大学	226	20	中国科学院植物研究所	158

(张丽萍、李东 供稿)