

Research Data

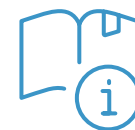
# 数据分享 在中国的挑战和机会

ADVANCING  
**DISCOVERY**

白皮书



# 目录



---

背景 .....	1
数据管理计划 .....	2
数据共享 .....	3
总结和结论 .....	5

该调查是与中国科学院文献情报中心密切合作完成的。感谢中国科学院文献情报中心以及参与该调查的中国科学院研究所的投入和支持！

---

## 作者

Mithu Lucraft, Katie Allin,  
Grace Baynes, Roza Sakellaropoulou

---

2019年1月

---

本白皮书及其原始数据已在Figshare  
存储库中对外公开。

## 获取白皮书:

[https://doi.org/10.6084/m9.  
figshare.7326605](https://doi.org/10.6084/m9.figshare.7326605)

## 获取完整的调查数据集:

[https://doi.org/10.6084/m9.  
figshare.7321604](https://doi.org/10.6084/m9.figshare.7321604)

# 背景

本报告调研了2,000多名在中国工作的研究人员并总结了相关发现，分析了数据共享在中国面临的挑战和机遇。本调研由施普林格·自然（Springer Nature）和中国科学院文献情报中心共同设计和实施完成。完整的数据集可在Figshare<sup>1</sup>上开放获得。

Springer Nature开展了包括本报告在内的一系列全球调研，以更好地了解数据共享所面临的挑战。作为主要的学术出版机构，我们致力于开发出例如数据管理和教育等方面的合作方案。在这项针对中国研究人员的调研之前，我们已经在2017年进行了一项全球调研，在全球范围内共有7,000名研究人员参与<sup>2</sup>，着重探讨在论文投稿时的数据共享。因某些地区和国家的数量不够，特别是日本和中国，所以对研究进行了扩展，以期获得来自日本和中国研究人员的反馈。

## 此次调研的目标是：

- 了解研究人员是否会制定数据管理计划（data management plan），以及他们在定制DMP的过程中遇到的阻碍。
- 了解研究人员在制定这些数据管理计划时所采取的措施。
- 了解研究人员是否正在进行数据的共享，以及数据共享中遇到的阻碍。
- 了解生成的数据类型以及数据的所属权。

该调查得到中国科学院文献情报中心的大力支持。调研以中文和英文两种语言在线进行，向施普林格·自然在中国的邮件注册用户，并通过社交媒体发放了调研邀请。总共收到了有2,202份来自中国研究人员（学术、科学或临床方面的研究人员）的问卷回复，涵盖了主要的研究学科和不同职业阶段。

与大多数基于调查的研究一样，受访者有一定程度的自我选择性，因此，在这种方式下，具有数据管理和数据共享经验的研究人员更有可能来参与此调查，这可能会导致结果具有一定的偏差。

1 Research, Nature; Allin, Katie; Baynes, Grace; Lucraft, Mithu; Penny, Dan; Chong, Steven; et al. (2018): Research data: challenges and opportunities for Chinese researchers-Springer Nature survey data. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.7321604>

2 Stuart, David; Baynes, Grace; Hrynaszkiewicz, Iain; Allin, Katie; Penny, Dan; Lucraft, Mithu; et al. (2018): Whitepaper: Practical challenges for researchers in data sharing. figshare. Paper. <https://doi.org/10.6084/m9.figshare.5975011.v1>

# 数据管理计划

好消息是，93%的中国研究人员表示他们之前已经制定了数据管理计划（DMP），这其中58%的人表示他们为一半或以上的研究项目制定了DMP。与之形成对比的是，2018年的类似研究显示<sup>3</sup>，全球平均有70%的研究人员制定过DMP，并且也显著高于日本的56%<sup>4</sup>。

然而，在中国工作的研究人员中，36%的人很少制定DMP（n = 1,827）。在从未制定过DMP（7%）的研究人员中，50%的人表示他们之前从未听说过DMP，40%的人表示不知道如何制定DMP（n = 126）。69%的中国研究人员“极有可能”或“有可能”在未来两年内制定DMP（n = 1,747），而过去制定过DMP的人表示他们未来更有可能再次制定DMP。这表明，加强有关制定DMP的教育能提升科研工作中经常性制定DMP的研究人员的比例。

科研处或其他研究人员的支持能最好地帮助研究人员实现其DMP的目标。有将近一半的受访者（48%）表示，他们所在机构的科研处可以最好地支持他们实现未来数据管理计划中所列出的目标，其次为同一研究组的其他人员（37%），第三为导师（36%）。

DMP最常见的内容涉及数据存储（71%）和数据收集（62%）。只有21%的研究人员将数据共享包括了进来（n = 1,490）。

制定DMP的主要原因是为了有良好的数据实践：67%的受访者表示“是为了确保高效和有效的数据管理”，58%的人表示DMP是“做研究时的良好实践”（n = 1,645）。受访者表示，资助者或机构的要求并不是很强的驱动因素（分别为18%和8%）。这与日本类似研究的结果以及与全球数据具有很大的差异，在后两者的结果中，机构和资助者的要求更多地成为了制定DMP的驱动因素。



**93%的研究人员之前制定过数据管理计划（DMP）。其中：**

**58%** 的研究人员为他们一半或以上的研究项目制定了DMP

**36%** 只在很少情况下制定DMP

**在从未制定过DMP的研究人员中：**

**50%** 的人从未听说过DMP

**40%** 的人不知道如何制定DMP  
(n=126)

<sup>3</sup> Digital Science; Hahnel, Mark; Fane, Briony; Treadway, Jon; Baynes, Grace; Wilkinson, Ross; et al. (2018): The State of Open Data Report 2018. figshare. <https://doi.org/10.6084/m9.figshare.7195058.v2>

<sup>4</sup> Research, Nature; Allin, Katie; Baynes, Grace; Lucraft, Mithu; Penny, Dan; Chong, Steven; et al. (2018): Research data: challenges and opportunities for Japanese researchers- Springer Nature survey data. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.6328952.v1>



**DMP最常见的是**

- 数据存储（71%）
- 数据收集（62%）

而只有**21%**的研究人员在DMP中包括了数据共享。

(n=1490)



**69%** 的研究人员表示极有可能或有可能在未来两年内建立DMP。

(n=1747)

# 数据共享



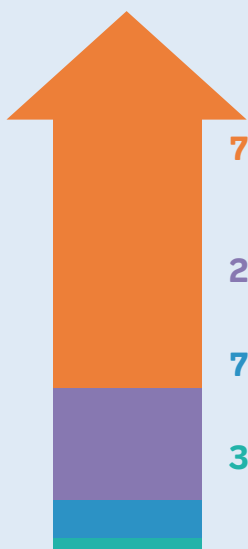
对于回答问卷的大多数中国研究人员来说，数据共享非常重要。79%的受访者认为其数据的可发现性比较重要（给出6分以上，满分11分），这一问题的平均得分为7.99，与全球平均分7.28相当。只有7%的受访者从未以私人或公开的形式共享过数据。70%的受访者以私人或公开的形式都进行过数据共享，这与71%的全球结果相似，并且远高于日本35%的研究人员曾私人共享数据，以及59%的人曾私人或公开共享数据的调查结果（n=975）。在中国有20%的研究人员仅私下分享，3%仅以公开的方式共享（n=1645）。

**79%的受访者认为其数据的可发现性对他们来说是比较重要的。**

促使研究人员共享数据的两大主要原因是“促进他们所属领域的研究发展（46%）”和“提升他们工作的可见度”（44%）（n=1,579）。35%的受访者认为，研究人员没有共享数据的主要原因是期刊没有这方面的要求（n=108），而研究人员在分享数据时所面临的两大担忧为“担心我的数据被滥用”（48%）和“不清楚版权和许可方面的问题”（32%）（n=1,560）。类似的研究发现，在日本（49%）和全球（38%）范围内，最大的担忧来自对数据滥用的考虑。在全球范围内的第二大担忧是“没有合理的荣誉或承认”（n=1356），而在中国，仅有17%的研究人员选择此项。提高期刊要求，以及多在版权和许可方面进行沟通应该能鼓励中国的研究人员更多地进行数据共享。

与直接的同事和合作者进行数据共享比公开地分享更重要，在他们的回答中，与所在机构的同事（61%）或与知名同行（55%）间的私人共享最常见。有更多的中国研究人员认为，与中国的其他研究人员（64%）分享数据比公开地共享更重要：只有45%的人认为对所有人公开数据很重要。在研究人员中，通过物理驱动器进行一对一的共享仍然非常普遍，49%的人通过USB或闪存盘方式共享，43%的人通过email方式共享，30%的人通过电脑硬盘的方式共享（n=1,441）。在对日本的研究人员进行的类似的调查中，大部分的私人共享是通过电子邮件（65%），USB或闪存驱动器（41%），以及在线文件共享（39%），例如Dropbox或者谷歌。

## 私人间的数据共享比公开的数据共享更为常见：



**70%** 的受访者以私人或公开的形式都进行了数据共享

**20%** 的受访者只进行过私人间的数据共享

**7%** 的受访者从未共享过数据

**3%** 的受访者只进行过公开的数据共享

三种最常见的私人间数据共享的方式为：



USB / 闪存驱动器 (49%)



电子邮件 (43%)



PC硬盘驱动器 (30%)

绝大部分是与直接的同事和合作者进行的数据共享，包括：



与所在机构的同事私下共享数据 (61%)



与知名同行私下共享 (55%)



认为与中国其他研究人员的数据共享很重要 (64%)



认为将数据对所有人公开很重要 (45%)

公开的数据共享的最常见方式是将数据作为期刊文章的补充信息进行提交（42%， $n = 1119$ ），其次是在数据期刊中进行共享（33%）。这个百分比是比较高的，因为目前来说专门的数据期刊相对较少，据Springer Nature估计，在全世界每年发表的180万篇研究论文中，只有不到1%的文章发表在数据期刊上。2018年“开放数据状况报告”称，18%的受访者已经在数据期刊上有过发表经历，这一数据与全球研究成果相比也很高。但我们还需要做更多的工作来检验中国和全球的研究人员对“数据期刊”的理解，以确保我们对这个结果的诠释是正确的。我们还在主题学科方面观察到了差异：有更多的生物科学研究人员将数据存放在特定学科的数据库中（34%），而有更多的工程研究人员则使用代码共享服务（19%）。虽然在中国，对数据库的了解程度和使用频率似乎比日本的更高（42%的中国研究人员知道并且已经在专门的数据库中存放了数据，而在日本，这一比例为30%）；然而，这还是低于全球范围60%的平均比例。

在共享的数据类型和时间点方面，公开的和私人间的数据共享也是不同的。在私人间，不管数据是正面的还是负面的，都平等地共享（84%），然而在公开的数据共享过程中，负面数据的共享比例更低（46%，正面数据的共享比例为65%）。很多受访者表示不会以公开的方式共享涉及未发表结果的数据（75%）以及涉及国家安全（70%）的数据和个人数据（70%）（ $n=1,196$ ）。共享的时间点也各有不同：对于私人间的共享，研究人员在整个研究过程中都会共享数据，而对于公开的数据共享，绝大部分共享发生在发表之后（46%）或在发表之时（33%）（ $n=1479$ ）。

在发表之前和之后，研究人员对数据所有权的感受是不同的（ $n=1,902$ ）。相比于发表之前（79%），在发表之后仍认为其拥有数据的研究人员人数大幅减少（56%）。相比于发表之前（4%），在发表之后有更多的研究人员认为数据由出版机构拥有（43%）。这与全球科研人员对数据所有权的看法相当。

**很多受访者表示，不会以公开的方式共享涉及未发表结果的数据（75%）以及涉及国家安全（70%）的数据和个人数据（70%）（ $n=1,196$ ）。**

# 总结和结论

总的来说，在中国有很多研究人员正在创建DMP（93%），但是制定这项计划的频率差异很大，只有58%的人为其一半或一半以上的研究制定DMP。有效地让人们认识到DMP的价值有可能会使这一数字上升，因为过去创建过DMP的研究人员表示他们更有可能在未来再次创建DMP。中国的DMP应当包括更多有关数据存储和共享计划的细节和信息，要解决这一问题就要提供有关存储库的更多信息并增加对数据存储和共享的支持。

数据共享对研究人员非常重要，79%的研究人员认为数据的可发现性很重要，93%的人以私下或公开的方式分享过他们的数据。与直接的同事和合作者进行的私人间的共享仍然比公开的共享更常见。无论是通过USB或email进行的私人间的共享，还是作为期刊文章的补充信息进行的公开的共享，大部分的数据共享都没有做到最好。缺乏期刊要求，对数据滥用的担忧以及缺乏版权和许可方面的相关知识是研究人员在共享数据方面所面临的主要问题，因此，通过让更多的期刊做出要求以及围绕版权和许可进行更多的沟通有可能会鼓励中国的研究人员增加数据的共享。

期刊出版业有责任鼓励并激励最佳实践，包括鼓励作者将研究数据存入存储库，并将可引用的数据链接以及数据可用性声明作为文章的常规部分。在本次调查中，导致研究人员不进行数据共享的首要原因是期刊没有此要求，事实上，期刊/出版机构的要求排在共享数据激励因素中的前三位。除了期刊的政策要求之外，鼓励发表数据描述集和数据记录方面的文章也值得考虑。此外，数据类文章在提供完善的荣誉机制（即，可引用的出版物）的同时，也使得数据集更易于查找、访问和再次使用。

虽然以补充信息的形式共享数据比完全不共享数据更好，但是这只是个次优解决方案。将数据存储存储在存储库中能使数据更易于查找和访问。我们需要在全球范围内做出努力，让研究人员相信在存储库中共享数据的好处，并简化具体的操作流程。在中国，这对于研究人员来说尤为重要，因为目前看来他们在数据库中储存数据的水平要低于全球同行。2018年3月由国务院办公厅发布的“科学数据管理办法”的文件就要求研究人员在国际期刊上发表文章之前将数据储存在数据库中。

最后，我们建议要着重强调研究数据的最佳实践在文章发表和引用方面所带来的好处，这些最佳实践包括在存储库中共享数据以及在文章中加入数据可用性声明和数据的引用。

## 一些建议：

- 提供有关创建和实施DMP等方面的信息并给予一定的支持，重点放在数据共享上
- 提供一些培训，让研究人员知道DMP的价值
- 进行更多的研究，以更好地了解研究人员对于“数据滥用”的担忧
- 明确有关版权和许可选项方面的信息
- 期刊加强这方面的政策要求，包括鼓励数据存储、数据引用和数据可用性声明
- 以更快、更简便的方式将数据存储存储在存储库中，并清楚地解释将数据储存在社区存储库中的好处
- 鼓励发表数据文章，让作者通过良好的数据实践获得认可
- 宣传数据管理的好处，包括在文章出版和引用方面的优势

# Appendix

## Number of respondents by subject area

Total	100%	2128
Biology	19%	410
Medicine	14%	293
Engineering	13%	272
Earth and Environmental Science	11%	237
Physics	11%	231
Materials Science	10%	204
Chemistry	7%	157
Other (please specify)	6%	128
Social Sciences	6%	118
Arts & Humanities	2%	33
Business/Investment	1%	29
Astronomy and planetary science	1%	16

## Number of respondents by job title

Total	100%	1499
<b>First stage researcher</b> Inc. PHD Student, Post Doc, Technician/Research Assistant, Student	55%	824
<b>Recognised researcher</b> Inc. Research Scientist, Staff Scientist, Associate Professor/Lecturer	25%	367
<b>Established researcher</b> Inc. Senior Scientist, PI	5%	78
<b>Leading researcher</b> Inc. Consulting/Fellow/Attending Physician, Lab Director/Head, Medical Professional, Professor and Research Director/VP of Research	13%	190
<b>Other</b>	3%	40