公共部门开源软件项目——开发是如何组织的?

Johan Linåker 瑞典RISE研究院 瑞典 j
ohan.linaker@ri.se

Björn Lundell
舍夫德大学
瑞典
bjorn.lundell@his.se

Francisco Servant
胡安·卡洛斯国王大学
西班牙
francisco.servant@urjc.es

Jonas Gamalielsson
舍夫德大学
瑞典
jonas.gamalielsson@his.se

Sachiko Muto
瑞典RISE研究院
瑞典
sachiko.muto@ri.se

Gregorio Robles
胡安·卡洛斯国王大学
西班牙
grex@gsyc.urjc.es

arXiv:2304.06093v1 [cs.SE] 12 Apr 2023

摘要

背景／语境：开源软件（OSS）最初是志愿者社区的一项努力，但其实践如今已远远超出最初的应用场景。例如，开源软件在工业领域的战略性应用目前正不断增长，涵盖能源、汽车和医疗卫生等多个垂直行业。然而，公共部门在采用方面却相对滞后，尽管提升互操作性、透明度和数字主权等在公共部门环境中尤为显著的益处已得到指出。公共部门组织（PSO）若寻求参与开源软件相关工作，就会面临挑战：它们往往缺乏必要的技术能力，同时还受到公共采购法规和实践的约束与影响。

目标/宗旨：我们旨在阐明公共部门开源软件项目，即由公共部门组织发起、开发和治理的项目，是如何开发和组织的。基于公共部门组织面临的挑战，我们推测，这类项目的开发组织方式在很大程度上与普遍采用的集市模式（由 Eric Raymond [38] 推广）并不一致；该模式意味着开发是在更大的社区中协作开展的。

方法：我们计划将公共部门开源软件项目与此前报道的一组集市式开源软件项目案例研究进行比较和对照，其中包括 Mockus 等人对 Apache Web 服务器和 Mozilla浏览器开源软件项目的报告 [32], ，以及针对 FreeBSD、JBossAS、JOnAS 和 Apache Geronimo 开源软件项目开展的复现研究 [11, 29]。为确保结果具有可比性，我们将对有目的抽样选取的部分公共部门开源软件项目，复现 Mockus 等人采用的方法论。我们将通过挖掘相关软件仓库，以定量方式识别并描述这些项目；并通过访谈相关组织的成员，对其进行定性调查。

允许为个人或课堂用途免费制作本作品全部或部分内容的数字副本或纸质副本，前提是不得为营利或商业利益而制作或分发副本，且副本首页必须包含本声明和完整引文。必须尊重本作品中归他人而非 ACM 所有的组件的版权。注明出处后可以摘录。若要以其他方式复制、重新发布、发布到服务器或分发到列表，必须事先获得具体许可和/或支付费用。请通过 permissions@acm.org 申请许可。*MSR'23，2023*年*5*月*15*日至*16*日，澳大利亚墨尔本© 2023 美国计算机协会。ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00
https://doi.org/XXXXXXX.XXXXXXX

CCS概念

• **软件及其工程** → **开源模式**；• **以人为中心的计算** → **开源软件**。

关键词

公共部门开源软件、公共部门组织、公共行政部门、集市模式、大教堂模式

ACM 引用格式：
Johan Linåker、Björn Lundell、Francisco Servant、Jonas Gamalielsson、Sachiko Muto 和 Gregorio Robles。2023 年。公共部门开源软件项目——开发是如何组织的？载于论文集：请务必填写版权确认邮件中的正确会议名称（*MSR'23*）。ACM，纽约州纽约市，7 页。https://doi.org/XXXXXXX.XXXXXXX

1 引言

开源软件（OSS）自二十多年前诞生以来（尽管作为一种现象出现得更早），已从一种由社区推动、旨在解决自身迫切需求的方式，发展成为业界围绕通用功能开展协作的成熟模式 [40]。在各个行业领域，竞争者正汇集资源并借助社区解决共同问题，从而让创新能够聚焦于技术栈的上层。由此，开源软件如今已广泛存在于绝大多数公司的内部代码库、产品和服务中，也构成了我们共同数字基础设施的重要部分 [19, 44, 45]。公共部门组织（PSO）同样可以借助开源软件项目和协作，在开发政府服务时更有效地利用资源；然而，开源软件在公共部门的应用并未像许多人预期的那样形成发展势头 [21]。尽管地方、区域、国家乃至超国家层级的许多行政机构已开展大量倡议，将其软件项目作为开源软件开发并从中受益，但迄今为止，这些倡议的影响仍然有限。

在本研究中，我们希望探究这一背景，以及我们所称的公共部门开源软件项目，即那些由一个（或多个）公共部门组织通过内部资源或委托资源发起、开发并治理的项目。具体而言，我们旨在探究，与受到更广泛研究的社区驱动型和行业驱动型开源软件项目相比，这类项目的开发和社会结构有何不同 [6]。

描述开源软件项目社会结构的一种常见方式，是通过所谓的“洋葱” [9] 模型：主要贡献者是项目的核心（“核心”开发者），周围 是一层新的偶尔贡献者。在外层，我们可以找到最终用户。个人距离中心越远，其对项目的投入和影响就越小。在大型开源软件项目中，就个人数量而言，“洋葱”的外层比内层多出几个数量级。这一外围群体通常由开源软件项目的用户和非正式贡献者组成。

成功发展出活跃而充满活力的社区的项目，通常采用集市模式，正如 Eric Raymond 在其著名的《大教堂与集市》一文中所提出的那样 [38], : 开发和沟通流程轻量，管理方式灵活，并向所有人开放，欢迎参与贡献。相比之下，大教堂模式遵循商业开发组织（以及其他开源软件项目）中常见的传统模式；在开源软件项目中，这意味着规划和开发由一位或少数几位成员独立负责，与社区相隔离 [5]。

学者指出，某个特定开源软件项目的开发模式可能会随着时间推移，从以大教堂模式为特征的开发模式转变为集市模式 [5]。此外，同一批学者认为，能够吸引关注和外部贡献者的开源软件项目，其开发模式通常会从大教堂模式转变为集市模式。鉴于此，我们认为，洋葱模型仅适用于采用集市模式或正向集市模式转变的开源软件项目。

在本研究中，我们的总体研究目标是探究公共部门开源软件项目如何开发和组织，并进一步考察研究结果是否可以用集市模式或大教堂模式来描述，还是需要采用新的比喻 [5, 38], 。我们推测，在公共部门围绕开源软件开展倡议二十多年后，公共部门开源软件项目的开发在很大程度上采用了与普遍接受的集市模式不一致的组织方式。进一步而言，我们假设，与最初由 Mockus 等人研究的集市式开源软件项目 [38] 相比，洋葱模型对公共部门开源软件项目 [36]的适用程度存在根本差异 [32]。

这会带来一些影响，因为这些项目无法享受由庞大开发者社区带来的集市模式相关优势。另一个影响可能是，针对社区驱动和行业驱动项目的研究与实践并不适用于公共部门开源软件项目。如果确实如此，就可能需要探索并评估其他方法（管理、技术和流程方面）。据我们所知，目前尚无任何科学研究探讨过这一推测。

我们拟通过将公共部门开源软件项目的特征，与一组此前报告的、按照集市模式开展开发的开源软件项目（下文简称集市式开源软件项目）进行比较和对照，来弥补这一空白。我们特别关注 Mockus 等人的开创性研究；他们研究了 Apache Web 服务器 [31] 和 Mozilla浏览器开源软件项目 [32]。该研究后来首先在由社区驱动的 FreeBSD 开源软件项目上进行了复现（且大部分研究结果得到验证） [11], 随后又在由行业驱动的 JBossAS、JOnAS 和 Apache Geronimo 开源软件项目上进行了复现 [29]。为确保结果具有可比性，我们将复现方法论，并从 Mockus 等人的原始研究中选取一组相关的研究问题 [32]。

接下来，我们将进一步阐述本研究的背景。随后，我们将详细介绍假设和研究问题，以及它们的理论依据。接着，我们将说明整体研究的研究设计，以及如何逐一回答各项研究问题。

2 背景与相关工作

据报道，欧洲与开源软件相关的机构能力规模，与开源软件所创造的价值规模相比明显偏小 [3]。造成这种情-况的因素有多个，包括经济增长、创新和竞争的潜力 [3, 15, 34]。研究还表明，开源软件能够带来在公共部门环境中尤为显著的益处，其中包括改善互操作性 [28], 透明度 [12], 和数字主权 [35]。

另一方面，诸如缺乏软件开发方面的技术能力和胜任力 [4], , 以及与公共采购法规和采购实践相关的问题，都会抑制其采用，而这些问题会影响公共部门组织采购开源软件的条件 [27]。开源软件的一般知识也是一个令人担忧的问题，这既存在于公共部门组织内部，也存在于其供应商中，例如，如何成功创建和组织开源软件社区 [1],, 以及如何在管理层面和开发者层面采用开放协作的原则 [13], [25]。

现有研究报告了多个采用开源软件的案例 [14, 17, 46], , 探讨了政府采用开源软件的益处 [18,21], , 以及如何利用开源软件技术重构公共部门 [16] 并开发新的电子政务服务 [20]。也有研究探讨了采用和发布开源软件相关的风险与关键因素 [23, 24, 42]。然而，尽管公共部门组织如何开展和组织开发与维护工作，以及公共部门组织如何（以及应当如何）参与开源软件项目这一议题早已受到研究社区的关注 [26], 相关研究仍不够充分。一个值得注意的例外是，有研究考察了 X-Road 项目的组织方式。该项目由北欧互操作解决方案研究所治理，由爱沙尼亚和芬兰发起 [39]。

需要指出的是，对于考虑与开源软件项目进行战略参与（超越开源软件的“临时性”采用和使用）的私营组织和公共部门组织而言，有若干挑战需要应对。此前的研究已对组织如何参与不同开源软件项目的若干策略进行了概念化 [28]。在商业（公司）环境中考虑这些策略，与缺乏短期商业目标的公共部门环境相比，意味着条件大不相同。例如，公共部门组织可以发起公共采购项目，采用并部署外部开源软件项目提供的开源软件，供内部使用 [43]。此外，公共部门组织可以选择与对其组织特别重要的特定开源软件项目进行战略参与，例如提交缺陷报告

或寻求在本组织与外部开源软件项目之间建立长期的共生关系 [47]。公共部门组织可以参与某种协会（例如丹麦的 OS2[1]），借此让一群公共部门组织携手合作，这与为特定开源软件项目的治理而成立基金会的方式类似，例如 Eclipse基金会和文档基金会。

3 假设与研究问题

根据我们的推测，并结合此前对 X-Road项目 [39], 的研究，我们预计公共部门开源软件项目由一组掌握经济、决策和战略权力的用户构成。与集市式开源软件项目不同，这些用户（至少在大多数开源软件项目中）不会主动参与开发，而是通过公共采购框架，将开发工作委托给一组开发者（主要是承包商）。除提供技术意见以影响开源软件项目的规划和方向外，这些开发者的权力有限。此外，开发者也可能包括开源软件项目的其他利益相关者和用户，但他们影响项目规划和方向的能力可能有限。因此，我们预计，规划和沟通将自上而下地从公共部门组织中的决策者传达给开发者和其他用户，并采用开放与封闭相结合的沟通渠道。

基于这一背景，我们参照 Mockus 等人 [32] 提出了一系列假设，以便比较和对照本研究所识别的公共部门开源软件项目与 Mockus 等人 [32]及后续复现研究 [11, 29]报告的集市式开源软件项目。值得注意的是，我们排除了一部分原始假设，因为这些假设着重强调开源软件项目相较于商业（闭源）软件在缺陷密度和发布速度方面的优势。

H1a 开源软件开发将拥有一组控制代码库的核心开发者，并由他们创建约 80% 或更多的新功能。如果这一核心群体仅通过非正式、临时的方式协调工作，则人数不会超过 10–15 人。

H1b 开源软件开发中约 95% 或更多的工作将由开源软件项目用户（即公共部门组织）委托的开发者完成。

H2a 无论受委托开发人员的数量多少，项目都会采用非正式的、临时的安排之外的其他机制来协调工作。这些机制可能包括以下一项或多项：明确的开发流程、个人或团队的代码所有权，以及强制检查。H2b 项目由一组用户（即公共部门组织中的决策者）自上而下规划，他们委托开发，并通过开放和封闭的沟通渠道进行沟通。H3 在成功的开源软件开发项目中，报告问题并以其他方式参与规划活动及开源软件项目相关沟通的群体，其规模将比用户群体（即委托开发的公共部门组织）大一个数量级。

[1]关于丹麦 OS2——公共数字化网络：https://os2.eu/english

H4a 拥有强大用户群体的开源软件开发项目，未能吸引大量普通用户参与，因此复用有限甚至不存在；同时，由于缺乏用于发现和修复缺陷的资源，质量也会下降。

H4b 项目的社区将主要由受委托的承包商，以及正在使用或有意使用该开源软件的公共部门组织组成。

为检验这些假设，我们采用 Mockus 等人提出的相同研究问题， [32], 但针对我们分析的公共部门开源软件项目进行研究：

RQ1 开发所识别出的公共部门开源软件项目采用了什么流程？– 对应假设 H1-4 RQ2 有多少人编写了公共部门开源软件项目的功能代码？有多少人报告了问题？有多少人修复了缺陷？他们的所属机构和角色是什么？– 对应假设 H1、H3-4 RQ3 这些职能是否由不同群体的个人承担，即个人是否主要担任单一角色？是否有大量个人较为均衡地参与这些活动，还是只有少数个人承担了大部分工作？– 对应假设 H3-4 RQ4 代码贡献者在代码中负责哪些部分？是否在文件或模块层面严格执行代码所有权？– 对应假设 H3-4 在第 4.3 节中，我们将说明如何回答这些研究问题并探讨相关假设。

4 方法论与实施计划

下文将详细介绍我们的研究设计，包括案例选择标准、其背后的理由，以及定量和定性方法的具体细节。

4.1 总体研究设计

为了使研究结果能够与 Mockus 等人的研究发现进行比较，我们计划尽可能按照他们的方法设计我们的研究方法。需要指出的是，Mockus 等人选择了两个开源软件项目——Apache 和 Mozilla。这些项目具有某些特征，基本上就是拥有规模庞大的开发者（和用户）社区。换言之，他们并非随机抽取所选开源软件项目，因为其中绝大多数规模较小、影响有限 [22, 33], ；相反，他们选择的是规模较大且成熟的项目，这些项目拥有活跃且充满活力的开发者社区（符合集市模式 [5, 38]）。因此，我们首先需要找出由公共部门组织推动、开发活动较为活跃的开源软件项目。

因此，为了实现我们的研究目标并回答研究问题，我们采用目的性抽样 [37] 并希望报告六个公共部门开源软件项目的案例研究。样本量与 Miller [30] 的做法一致，其依据是每个案例都会为我们计划开展的调查与比较增加复杂性。我们认为，案例研究方法论是最合适的方法，因为 Benbasat 等人认为：“*[一个]* 案例研究在自然环境中考察某一现象，并采用多种数据收集方法。

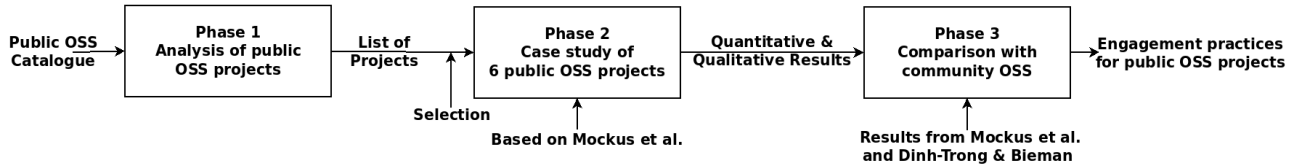

图1：执行计划概览。

从少数实体收集信息。该现象的边界在研究开始时并不明确，也未使用实验控制或操纵" [2]。因此，与Mockus 等人一样，本研究将采用混合方法 [8],，结合定性和定量数据来源。这将基于对公开可用来源的分析，并通过获取公共部门开源软件项目相关利益相关者的反馈，为这些项目提供更完整的视角。

因此，我们将把研究设计分为三个阶段（见图1）：

(1) 在第一阶段，我们将识别由公共部门组织提供的开源软件目录。这些目录是公共部门组织创建协作空间的成果。其中列出了为地方、区域、国家，有时还包括国际层面的公共部门组织开发以及由其开发的软件，并提供源代码和文档的链接。在某些情况下，这些目录中也可以找到通用软件，例如 LibreOffice；尽管公共部门组织广泛使用这类软件，但它们不符合由公共部门组织推动的开源软件项目定义。我们已经识别出约六个此类目录，包括意大利、法国、美国等地公共部门组织所使用的目录。我们将分析这些目录中的所有项目（总计超过 2 万个项目），纳入由公共部门组织推动的项目，并根据相关特征对其进行分类，例如协作者数量（以提交者和缺陷报告者计）以及活动量（以代码提交、缺陷报告和拉取请求计）。这一阶段的产出将是一份按照社区规模和活跃度排序的项目列表。此外，我们还将补充推动该项目的公共部门组织的信息，例如其国籍和政府层级（国际、国家、区域、地方）等其他相关方面。

(2) 在第二阶段，我们计划从第一阶段得出的名单中通过目的性抽样 [37] 选取六个项目。我们的目的不仅是选择规模（以开发者和代码提交数量衡量）最大的六个项目，也希望选出的项目具有多样性（涉及国籍、层级，以及是否为内部开发）。我们将通过挖掘相关软件仓库，对这些项目进行定量调查。我们会联系各项目的代表，询问他们是否愿意对我们的调查提供反馈，从而帮助我们对定量分析的研究发现进行三角验证。因此，项目是否积极回应，将是决定其能否入选并接受进一步研究的一个重要因素。

(3) 在第三阶段，我们将对公共部门开源软件项目样本的观察结果进行定性比较，并将其与 Mockus 等人报告的 Apache Web 服务器和 Mozilla 浏览器开源软件项目的集市模式案例，以及 [32],经复现的 FreeBSD [11],、J BossAS、JOnAS 和 Apache Geronimo 项目案例 [29]进行对照。作为比较公共部门开源软件项目与集市式开源软件项目的参照框架，我们将采用洋葱模型 [36],；该模型传统上用于说明项目中活跃且具有影响力的开发者核心与影响力较弱的用户外围之间的关系 [10]。

下面我们将更详细地阐述方法论及其各个阶段。

4.2 数据集与案例选择

截至目前，我们已发现多个不同国家由公共部门组织使用、为其开发的开源软件项目目录，包括意大利[10]、法国、加拿大、美国、瑞典、丹麦、芬兰、德国以及荷兰。

。图2展示了意大利门户网站的截图。这些目录中收录的项目将作为我们第一阶段的输入。这些目录是通过个人关系网络和在线搜索发现的，应视为潜在相关公共部门开源软件项目的样本，而非完整的数据集。

本研究旨在确定六个特定的公共部门开源软件项目。在选择过程中，我们将采用四项纳入标准（IC）来确定这六个项目：

IC1 首先，作为软件项目成果提供的开源软件，并且在调查时间范围内（2023 年期间）至少被一个公共部门组织使用过的软件。该标准包括作为公共采购项目启动的软件项目，而不论开源软件向公共部门组织提供（由商业供应商提供）是否通过该开源软件项目的公共平台进行。不过，满足该标准的前提是，公共平台上存在一个开源软件项目，即使没有任何公共部门组织的代表直接参与该软件项目的开发和维护（该项目可能最初是作为一个

[2]https://developers.italia.it/en/software [3] https://code.gouv.fr/ [4] https://code.open.canada.ca/en/index.html[5]https://code.gov/agencies [6] https://www.offentligkod.se[7] https://os2.eu/produkter[8] http://www.opencode.fi/ 9[10] https://developer.overheid.nl/

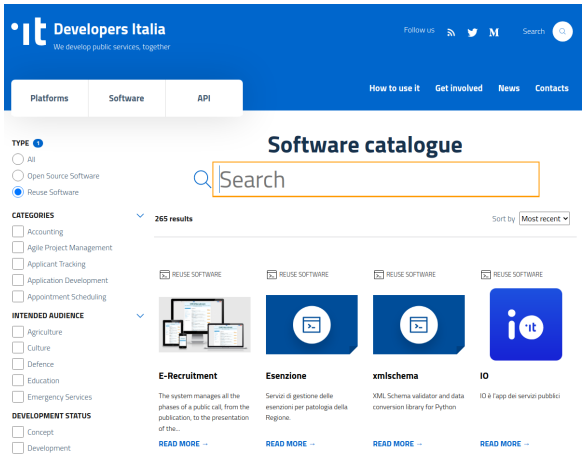

图 2：developers.italia.it 公共部门开源软件网站目录的快照，来自 数字化转型部（意大利数字化转型部）。根据知识共享署名 4.0 许可提供，详见 https://developers.italia.it/en/legal-notice。

作为采购合同一部分在公共平台上提供的开源软件项目）。IC2 第二，过去一年中在公共开发平台上有过活动的开源软件项目。此标准旨在排除托管该项目的公共开发平台上处于不活跃状态的开源软件项目。IC3 第三，开源软件项目的完整源代码及相关开发信息均公开可用，使人们能够检查源代码，并通过使用开源软件许可证许可的开发工具，从该开源软件项目创建可运行的开源软件实例。尽管公共部门组织可能会在缺乏开源软件许可证许可的开发工具的情况下，提供采用开源软件许可证的源代码，但这类项目对其他公共部门组织的相关性有限。因此，这类项目不符合此项纳入标准。IC4 第四，公共部门组织可以（并且已经）获得与该项目相关的商业支持和服务。此标准假定至少有一个公共部门组织已与提供商就该开源软件项目签订商业合同。

IC1-2 将通过本研究第一阶段所述的挖掘过程予以应用，而 IC3-4 将在第二阶段确定六个待调查项目时进行人工验证（见图 1）。

筛选过程的结果将是一组具有某些特征（如上文所述）的项目，我们将对其进行更深入的分析。因此，我们需要访问它们的源代码仓库（如今通常是Git仓库）和缺陷跟踪系统。在大多数情况下，根据对已确定目录的初步检查，我们发现这些项目位于GitHub或GitLab仓库中。具体而言，对于这组项目，我们将遵循下一小节所述的研究设计。

4.3 变量

下面，我们列出为回答研究问题而将要分析的变量。我们旨在采用与 Mockus 等人相同的方法（因此收集相同的变量），以便在第三阶段进行比较（见图 1）。 [32],

4.3.1 RQ1：开发过程。为准确描述开源软件开发流程，Mockus 等人的一位作者先为每个项目（即 Apache 和 Mozilla）撰写了开发过程的描述草稿，然后请核心开源软件开发团队的成员进行审阅。描述范围包括以下信息：

- 角色与职责
- 识别待完成的工作
- 分配并执行开发工作
- 预发布测试
- 检查
- 管理版本发布

我们计划根据项目网页和代码库中的证据，自行撰写一份草稿，以复现这一过程，并请项目成员审阅和确认。访谈还将使我们能够提出后续问题，以澄清并深入了解开源软件项目的开发过程。

4.3.2 RQ2：社区规模。为回答这个问题，我们将像 Mockus 等人那样，识别提交过代码的人（区分添加功能者和修复缺陷者），以及向缺陷跟踪系统提交过缺陷报告的人。

4.3.3 RQ3：参与和角色。Mockus 等人绘制了代码变更和缺陷报告的累计占比随贡献者人数变化的曲线，并统计了排名前 15 位开发者（他们称之为核心群体，沿用洋葱模型的说法）的贡献份额。我们还将考察最活跃的开发者的代码提交规模在统计上是否大于其他开发者的提交规模，以及他们在添加新功能方面的参与情况。我们也将衡量最活跃的开发者和其他开发者的任职时长。

对于缺陷报告，我们将调查报告这些缺陷的人员分布，以及总份额中有多少是由最活跃的开发者完成的。

4.3.4 RQ4：代码所有权。为回答这个问题，我们将追踪修改过文件的人员，看看是否只有一位开发者负责这些文件，还是代码所有权由更多开发者共同拥有。

4.4 定性分析

如第1节所述，我们通过访谈各个案例社区中的成员来验证我们的研究。这既涉及RQ1所强调的过程描述，也涉及RQ2- 4所涵盖的总体研究发现。受访者将根据其社会和技术活动以及其（潜在的）组织隶属关系进行抽样。每个项目至少抽取两名受访者，并视情况增加人数。

若其回答与反思之间出现不一致，则会进行抽样。

访谈将由两位作者进行，其中一位负责主持访谈，另一位负责记录。观察结果和研究发现将呈现给受访者，随后鼓励受访者根据自身经历进行公开反思并作出回应。访谈将被录音并转录，并根据相应的研究问题进行结构化编码 [41] 。编码数据随后将用于对我们之前的研究发现进行三角验证，并用于我们对各个项目的分析。这些数据还将为我们综合和比较 Mockus 等人所研究的集市式项目、复现研究 [32] 以及本研究所调查的公共部门开源软件项目提供输入 [11, 29], 。

作者将遵守瑞典研究理事会提供的伦理准则 [7]。具体而言，在数据管理方面，受访者的身份信息和所属机构将在访谈转录稿以及任何基于该数据的报告中进行匿名化处理。每位受访者都将获得一份访谈转录稿副本，并获鼓励指出任何误解或敏感信息的提及。拟用于报告的任何引文都将与相关受访者进行核实。本研究报告完成后，所有录音都将被销毁。

5 有效性威胁

由于这是一项复现研究，所采用的方法必须尽可能与原始研究的方法保持一致 [32] ，以便比较研究结果。因此，六位作者分别独立审阅并比较了本研究与原始研究的方法论，并共同讨论了所有可识别的不一致之处。

不过，在设计决策方面确实存在一些差异，例如案例抽样。除了原始研究中的两个案例都拥有规模庞大且活跃的社区这一点外，作者选择这些案例还因为他们在各自的社区中有亲身经历并扎根其中。在本研究中，我们特别关注公共部门开源软件项目的特征，以及这些项目可能与原始研究所考察的集市式开源软件项目有何不同。为避免抽样偏差，我们将挖掘列出不同国家公共部门组织所使用和开发的开源软件项目的软件目录。随后，我们将依据预先设定的纳入标准，对六种不同类型的项目进行抽样。

由于样本有限，我们不声称研究结果具有普遍适用性。相反，我们对所谓的公共部门开源软件项目作出情境化描述，并举例说明这些项目的开发模式（可能）与一般的集市模式开源软件项目以及常见且与之相关的洋葱模型有何不同。通过对这些项目开展定性和定量研究，我们将深入介绍这六个项目的特征，以便对具有相似特征的案例进行轶事式推广。进一步推广是未来研究的一个开放课题。

与原始研究不同，我们不预期自己对待抽样的六个案例有深入了解。因此，为了降低研究偏差和误读观察结果的风险，我们将允许社区代表审阅并批评我们的研究发现（尤其是 RQ1，以及总体上的 RQ2–6）。

另一个威胁涉及本研究的构念效度——我们认为在 Apache Web 服务器和 Mozilla浏览器开源软件项目案例中体现的集市模式，如今是否仍然成立。Mockus 等人的报告 [32] 发表于 2002 年，而实践和流程可能已经演变，因此重新审视这两个社区及其当前的运作方式可能具有参考价值。然而，由于这是一项复现研究，我们只能以 Mockus 等人的报告 [32] 作为基准。此外，我们注意到，自 Mockus 等人的研究 [32] 发表以来，多项研究已对其进行复现和推广（见 [11, 29]），因此我们仍认为该研究有效地呈现了我们所说的集市式项目。此外，我们将把这些对 Mockus 等人研究 [32] 的复现纳入公共部门与集市式开源软件项目之间的比较，以进一步细致地了解集市式项目的特征。

6 启示

在采用开源软件以及过去二十年来行业经历的成熟度演进方面，公共部门一直落后。不过，在利用开源软件促进经济增长 [3, 15, 34], 并实现政策目标方面，仍存在多种机遇，例如与互操作性 [28], 、透明度 [12], 和数字主权 [35]相关的目标。因此，公共部门组织需要考虑如何实现这些目标。

在这方面，公共部门组织可以借鉴业界汲取的经验和成熟的最佳实践。然而，所有公共部门组织运营时都面临许多具体条件，这些条件与公司所处的环境大不相同。因此，借鉴公司环境中的"方法"制定的任何（自上而下的）政策，都可以作为很好的参考，但这些政策需要立足于公共部门组织的实际运营条件。考虑到公共部门组织所面临的特殊情况，例如缺乏技术能力，以及必须遵守公共采购法规和采购实践，这些经验和实践需要在公共部门环境中仔细审视，并据此重新设计。

基于对公共部门开源软件项目的六项案例研究，以及与先前报告的集市式开源软件项目案例的比较，我们旨在设计能够满足这一需求的参与实践。我们希望这些实践能够指导公共部门组织如何发起并参与公共部门开源软件项目，从而实现预期收益。更深入地了解公共部门组织如何才能最好地利用开源软件，有助于改善政府服务，并最终提高公民的生活质量。

参考文献

[1] Jono Bacon. 2012. 社区的艺术：构建参与的新时代. O'Reilly Media, Inc. [2] Izak Benbasat, David K Goldstein, and Melissa Mead. 1987. 信息系统研究中的案例研究策略. *MIS* 季刊 (1987), 369–386.[3] Knut Blind, Mirko Böhm, Paula Grzegorzewska, Andrew Katz, Sachiko Muto, Sivan Pätsch, and Torben Schubert. 2021. 开源软件和硬件对欧盟经济技术独立性、竞争力和创新的影响. 最终研究报告。欧盟委员会，布鲁塞尔，*doi* 10 (2021), 430161. [4] Markus Borg, Thomas Olsson, Ulrik Franke, and Saïd Assar. 2018. 瑞典政府机构的数字化——透过软件开发普查的视角. In *2018* 年 *IEEE/ACM* 第 *40* 届国际软件工程会议：社会中的软件工程（*ICSE-SEIS*）. IEEE, 37–46.

[5] Andrea Capiluppi 和 Martin Michlmayr。2007。《从大教堂模式到集市模式：志愿者社区项目生命周期的实证研究》。载于 IFIP 开源系统国际会议。施普林格，31–44。[6] Eugenio Capra 和 Anthony I Wasserman。2008。《开源项目管理风格评估框架》。载于《开源开发、社区与质量》，Barbara Russo、Ernesto Damiani、Scott Hissam、Björn Lundell 和 Giancarlo Succi（编）。施普林格美国，波士顿，马萨诸塞州，1–14。[7] 瑞典研究理事会。2017。《良好科研规范》。瑞典研究理事会。[8] John W Creswell、Vicki L Plano Clark、Michelle L Gutmann 和 William E Hanson。2003。《高级混合方法研究设计》。载于《社会与行为研究中的混合方法手册》209，240（2003），209–240。[9] Kevin Crowston 和 James Howison。2005。《开源软件开发团队的社会结构》。《First Monday》10，2（2005）。[10] Kevin Crowston、Kangning Wei、Qing Li 和 James Howison。2006。《自由/开放源码软件团队通信中的核心与边缘》。载于《第39届夏威夷国际系统科学会议论文集》（HICSS'06），第6卷。IEEE，118a–118a。[11] Trung T Dinh-Trong 和 James M Bieman。2005。《FreeBSD 项目：开源开发的复制案例研究》。《IEEE 软件工程汇刊》31，6（2005），481–494。[12] OpenForum Europe。2022。《开放战略自主》。可访问：https://openforumeurope.org/publication-open-strategic-autonomy/。最后访问日期：2022年6月29日。[13] Joseph Feller 和 Brian Fitzgerald。2002。《理解开源软件开发》。Addison-Wesley，伦敦。[14] Brian Fitzgerald 和 Tony Kenny。2004。《使用开源软件开发信息系统基础设施》。《IEEE 软件》21，1（2004），50–55。https: //doi.org/ 10 1109 2004 1259216 . /MS. . [15] Shane Greenstein 和 Frank Nagle。2014。《数字暗物质与 Apache 的经济贡献》。《研究政策》43，4（2014），623–631。[16] Antti Hautamäki 和 Kaisa Oksanen。2018。《用于重构公共部门的数字平台》。载于《平台经济中的协作价值共创》。施普林格，91–108。[17] Viktor Hollmann、Hwansoo Lee、Hangjung Zo 和 Andrew P Ciganek。2013。《考察开源软件仓库的成功因素：OSOR.eu 门户案例》。《国际商业信息系统期刊》14，1（2013），1–20。[18] Philip Huysmans、Kris Ven 和 Jan Verelst。2008。《数据密集型公共管理机构不采用 OpenOffice.org 的原因》。《First Monday》13，10（2008），10。[19] 开源促进会。2022。《开源现状调查》。可访问：https://opensource.org/ten-takeaways-from-the-2022-state-of-open-source-survey。最后访问日期：2022年6月28日。[20] Ahto Kalja、Kristiina Kindel、Riina Kivi 和 Tarmo Robal。2007。《电子政府服务：如何开发和管理？》。载于《工程与技术管理》，波特兰国际中心。IEEE，2795–2798。[21] George L Kovács、Sylvester Drozdik、Giancarlo Succi 和 Paolo Zuliani。2004。《面向公共管理的开源软件》。载于《第6届计算机科学与信息技术国际研讨会论文集》。[22] Sandeep Krishnamurthy。2002。《洞穴还是社区？：对100个成熟开源项目的实证考察》。《First Monday》（2002）。[23] Victor Kuechler、Carlos Jensen 和 Deborah Bryant。2013。《美国能源行业采用 FOSS 的误解与障碍》。载于 IFIP 开源系统国际会议。施普林格，232–244。[24] Johan Linåker 和 Björn Regnell。2020。《分享什么、何时分享以及在哪里分享：平衡开源软件贡献的目标与复杂性》。《实证软件工程》25，5（2020），3799–3840。[25] Johan Linåker 和 Per Runeson。2020。《公共部门平台走向开放：通过开放协作开发创建并发展生态系统》。载于《第16届国际开放协作研讨会论文集》。1–10。[26] Björn Lundell、Morten Amundsen、Rishab Aiyer Ghosh、Jean-Luc Hardy 和 Per-Ola Sjöswärd。2009。《专题讨论：公共部门中的开源》。载于 IFIP 开源系统国际会议。施普林格，368–369。[27] Björn Lundell、Jonas Gamalielsson、Simon Butler、Christoffer Brax、Tomas Persson、Anders Mattsson、Tomas Gustavsson、Jonas Feist 和 Jonas Öberg。2021。《通过采购项目促进开源软件使用：如何避免锁定效应？》。载于 IFIP 开源系统国际会议。施普林格，16–27。[28] Björn Lundell、Jonas Gamalielsson、Stefan Tengblad、Bahram Hooshyar Yousefi、Thomas Fischer、Gert Johansson、Bengt Rodung、Anders Mattsson、Johan Oppmark、Tomas Gustavsson 等。2017。《通过开源应对锁定、互操作性和长期维护挑战：企业如何战略性地使用开源？》。载于 IFIP 开源系统国际会议。施普林格，查姆，80–88。[29] Xiujuan Ma、Minghui Zhou 和 Hong Mei。2010。《开发者如何参与开源项目：基于 JBossAS、JOnAS 和 Apache Geronimo 的复制案例研究》。载于《第1届实证软件工程研究复制国际研讨会论文集》。[30] George A Miller。1956。《神奇的数字七，加减二：我们信息处理能力的某些局限》。《心理学评论》63，2（1956），81。[31] Audris Mockus、Roy T Fielding 和 James Herbsleb。2000。开源软件开发案例研究：Apache 服务器。载于第*22*届国际软件工程会议论文集。263–272。[32] Audris Mockus、Roy T Fielding 和 James D Herbsleb。2002。开源软件开发的两个案例研究：Apache 和 Mozilla。*ACM* 软件工程与方法学汇刊（*TOSEM*）11，3（2002），309–346。[33] Nuthan Munaiah、Steven Kroh、Craig Cabrey 和 Meiyappan Nagappan。2017。为工程化软件项目整理 GitHub。实证软件工程22，6（2017），3219–3253。[34] Frank Nagle。2019。政府技术政策、社会价值与国家竞争力。哈佛商学院战略部门工作论文19-103（2019）。[35] Frank Nagle。2022。强化数字基础设施：自由及开源软件的政策议程。链接：https://www.brookings.edu/research/strengthening-digital-infrastructure-a-policy-agenda-for-free-and-open-source-software/。最后访问于2022年6月29日。.[36] Kumiyo Nakakoji、Yasuhiro Yamamoto、Yoshiyuki Nishinaka、Kouichi Kishida 和 Yunwen Ye。2002。开源软件系统及社区的演化模式。载于软件演化原理国际研讨会论文集（*IWPSE'02*）。ACM，佛罗里达州奥兰多，76–85。[37] Michael Quinn Patton。2014。定性研究与评估方法：理论与实践相结合。Sage 出版社。[38] Eric Raymond。1999。《大教堂与集市》。知识、技术与政策 12，3（1999），23–49。[39] Gregorio Robles、Jonas Gamalielsson 和 Björn Lundell。2019。基于开源软件构建政府3.0解决方案：X-Road案例。载于国际电子政务会议。施普林格，69–81。[40] Gregorio Robles、Igor Steinmacher、Paul Adams 和 Christoph Treude。2019。开源软件的二十年：从怀疑到主流。*IEEE Software* 36，6（2019），12–15。[41] Johnny Saldaña。2021。定性研究人员编码手册。Sage。[42] Thomas P Scanlon。2019。美国国防部开源发展的关键因素。*IEEE Software* 36，6（2019），29–33。[43] Maha Shaikh。2016。在英国公共部门协商采用开源软件。《政府信息季刊》33，1（2016），115–132。[44] 新思科技。2020。开源安全与风险报告。链接：https://www.synopsys.com/blogs/software-security/open-source-trends-ossra-report/。最后访问于2022年6月28日。.[45] Tidelift。2018。专业开源调查。链接：https://blog.tidelift.com/open-source-is-everywhere-survey-results-part-1。最后访问于2022年6月28日。.[46] Kris Ven、Dieter Van Nuffel 和 Jan Verelst。2006。布鲁塞尔公共行政部门引入 OpenOffice.org。载于开源系统，*IFIP* 开源软件工作组*2.13*基金会，*2006*年*6*月*8*日至*10*日，意大利科莫（*IFIP*，第*203*卷），Ernesto Damiani、Brian Fitzgerald、Walt Scacchi、Marco Scotto 和 Giancarlo Succi（编）。施普林格，123–134。https://doi.org/10.1007/0-387-34226-5 12[47] Susan Winter、Nicholas Berente、James Howison 和 Brian Butler。2014。超越组织“容器”：构想21世纪的社会技术工作。《信息与组织》24，4（2014），250–269。