Open source ecosystems need equitable credit across contributions

To the Editor — Collaborative and creative communities are more equitable when all contributions to a project are acknowledged. Equitable communities are, in turn, more transparent, more accessible to newcomers, and more encouraging of innovation — hence we should foster these communities, starting with proper attribution of credit. However, to date, no standard and comprehensive contribution acknowledgement system exists in open source, not just for software development but for the broader ecosystems of conferences, organization and outreach efforts, and technical knowledge. Furthermore, both closed and open source projects are built on a complex web of open source dependencies, and we lack a nuanced understanding of who creates and maintains these projects[1]. As a result, large sums and efforts go to open source software projects without knowing whom the investments support and where they have impact[2].

Academia faces a similar recognition problem. Attribution is often collapsed to 'authorship', yet increases in the size and complexity of scientific teams are colliding with this narrow definition of contribution[3]. Focusing only on authorship hides much of the work necessary to publish research[4]. Since this hidden work is performed disproportionately by people from underrepresented communities, the full picture of who is doing work is not accurately represented[5]. Fortunately, new models of recognition are gaining widespread adoption. One example is the CRediT framework, a taxonomy created for 'contributorship, not authorship', to more fully represent the roles that people play in creating research outputs[3]. Contributor roles are categorized by tasks and stages of the research process[6], allowing multiple people to perform the same role or the same person to perform multiple roles. Since its inception, the CRediT taxonomy has been widely adopted (by 33 major publishers so far) in part because it can not only contribute to equity within a project, but also potentially provide the major benefit of standardizing credit across projects and communities[7].

Open source ecosystems need to follow suit and adopt a standard taxonomy of contributor roles. As in academic research, modern open source is a highly collaborative and complex task environment, where overly broad or poorly defined roles easily obfuscate the work of many[2]. For example, a broadly defined role of 'code contributor' fails to distinguish between specific tasks such as adding features, fixing bugs or taking other actions that directly edit the source code. Likewise, other substantial contributions, such as organizing meetings, providing outreach or performing other activities that leave no visible trace within the code, are often neglected. Indeed, some important contributions occur entirely outside of common open source development platforms such as GitHub and often go unrecognized[8]. To succeed, a taxonomy of roles should be simple, comprehensive, use clearly defined non-overlapping categories, represent different types of contributions equally, and must avoid favoring specific platforms.

Existing efforts to recognize contributions to open source are laudable, but gaps remain. For instance, many open source projects include ad hoc attribution lists like 'credit files', without consistent attribution categories. Some approaches propose taxonomies (for example, All Contributors), but do not include clear guidelines around how to apply the proposed taxonomy to various contributions or may miss entire categories altogether. Any confusion around the interpretation of the taxonomy, where different communities could interpret the categories as they wish, negates the benefit of standardizing credit across projects and communities.

Meanwhile, platform-specific metrics like GitHub's 'contributor count' are some of the most visible contribution indicators but are limited in scope and do not generalize to contributions outside the platform. Data-driven efforts that extract contributions automatically from source code, version control records, or other platform data are further limited to only those activities explicitly recorded within the data (see, for example, octohatrack[8] and name-your-contributors). Despite its importance for open source, efforts to recognize contributions more broadly have yet to be widely adopted.

What obstacles have prevented the adoption of a broader, standard recognition model? If CRediT can teach us anything, it is that standards should emerge from the community, undergo many iterations and rounds of feedback, and receive buy-in from major relevant institutions and involved parties. The CRediT taxonomy resulted from a long categorization effort and is a prime example of a working contributor taxonomy. Research leveraging CRediT data is only ramping up[7], yet adoption of the framework continues to rise through community support. A successful taxonomy for open source should develop through a similar community peer review. Just as academic institutions and publishers are embracing the CRediT model, open source contributions need the same attention.

A standard taxonomy of recognized contributions will benefit all levels of open source. Contributors will gain credits beyond the code, providing a clearer signaling of their work for the community. Projects will be able to measure their growth and evaluate their culture. Structural biases will be brought to light, helping to foster more equitable open source communities. Everyone will better understand the interconnected structure of skills, projects and contributions across the broader ecosystems of open source. ❐

Amanda Casari [iD][1], Katie McLaughlin [iD][2], Milo Z. Trujillo [iD][3], Jean-Gabriel Young [iD][3,4], James P. Bagrow [iD][3,5] and Laurent Hébert-Dufresne [iD][3,4]✉

[1]Open Source Programs Office, Google, Kirkland, WA, USA. [2]Open Source Programs Office, Google, Sydney, New South Wales, Australia. [3]Vermont Complex Systems Center, University of Vermont, Burlington, VT, USA. [4]Department of Computer Science, University of Vermont, Burlington, VT, USA. [5]Department of Mathematics and Statistics, University of Vermont, Burlington, VT, USA.
✉e-mail: laurent.hebert-dufresne@uvm.edu

References
1. Eghbal, N. *Roads and Bridges: The Unseen Labor Behind Our Digital Infrastructure* (Ford Foundation, 2016).
2. Eghbal, N. *Working in Public: The Making and Maintenance of Open Source Software* (Stripe Press, 2019).
3. Brand, A., Allen, L., Altman, M., Hlava, M. & Scott, J. *Learned Publ.* **28**, 151–155 (2015).
4. Holcombe, A. O. *Publications* **7**, 48 (2019).
5. Larivière, V. et al. *Soc. Stud. Sci.* **46**, 417–435 (2016).
6. Holcombe, A. O. *Nature* **571**, 147–148 (2019).
7. Allen, L., O'Connell, A. & Kiermer, V. *Learned Publ.* **32**, 71–74 (2019).
8. McLaughlin, K. *Model View Culture* https://modelviewculture.com/pieces/acknowledging-non-coding-contributions (2016).