

Report on the International Workshop on Contributorship and Scholarly Attribution

Held Wednesday May 16, 2012, hosted by the Institute for Quantitative Social Science (IQSS) at Harvard University

Report written by Irene Hames, Editorial Advisor and Consultant, with input and some textual contributions from the Workshop Planning Committee and breakout group facilitators

Workshop Planning Committee

- **Liz Allen** Senior Evaluation Adviser, Wellcome Trust
- **Micah Altman** Director of Research, MIT Libraries
- **Amy Brand** Assistant Provost for Faculty Appointments, Harvard University
- **Merce Crosas** Director of Product Development, IQSS/Harvard
- **Diane Scott-Lichter** Publisher and Head of Publishing Division, American Association for Cancer Research
- **Mike Taylor** Principal Investigator, Elsevier Labs

Details of the workshop, agenda and videos of the presentations can be found at http://projects.iq.harvard.edu/attribution_workshop

Contacts: Irene Hames for the report (irene.hames@gmail.com, @irenehames) and Amy Brand for the workshop and next steps (amy_brand@harvard.edu)

Cite this work as:

IWCSA Report (2012). *Report on the International Workshop on Contributorship and Scholarly Attribution, May 16, 2012.* Harvard University and the Wellcome Trust.
http://projects.iq.harvard.edu/attribution_workshop

Published 18 September 2012 under the Creative Commons Attribution CC-BY licence
<http://creativecommons.org/licenses/by/3.0>

Contents

Executive summary	3
Background	3
Problems identified	3
Possible solutions to the problems identified	5
Conclusions and next steps	7
The Workshop	9
Participants and sponsorship	9
Workshop goals	9
Workshop format	9
Introduction Amy Brand	10
Presentations	10
Keynote: The Collective Author, Peter Galison	10
Issues of authorship practice and policy in experimental science research labs, Cassandra Extavour	11
Contribution and attribution in research: funder perspectives, Liz Allen	12
Authorship: a (medical) journal's perspective, Christine Laine	12
Authorship: a provostial perspective, Judith Singer	13
PLOS: big science and medicine and the attribution of authorship, Ginny Barbour	14
Why are the attribution and citation of scientific data important? Christine Borgman	14
FundRef: a new project from a publisher–funding agency collaboration, Fred Dylla	15
Contributor roles in VIVO, Brian Lowe	16
Exploratory research	16
Breakout sessions	17
Breakout Group 1: A taxonomy for roles and contributions in scholarly publications	17
Breakout Group 2: Implementation considerations	19
Breakout Group 3: Social considerations, barriers and incentives	20
Breakout Group 4: The potential for new metrics	21
Breakout Group 5: Citation/attribution for data and non-traditional forms and units of scholarship	22
Summary and further questions for moving the debate forward Clifford Lynch	23
Conclusions and next steps	24
Appendix 1 Workshop participants	26
Appendix 2 Research details Micah Altman and Mike Taylor	28

Executive summary

Background

There are no universal conventions that determine which scholarly contributions qualify individuals as authors or that unambiguously convey authorship precedence. As a consequence, formal attribution of authorship can obfuscate the contributions of those involved in collaborative research and writing endeavors. Publication credit can be misunderstood, and can be misapportioned by traditional impact measures. Because the allocation of credit for research and discovery has such a huge impact on careers and funding, there are increasing concerns about these issues. Contribution opacity also hinders appropriate accountability and responsibility for the integrity of the work being reported. For all these reasons, there is growing interest amongst researchers, funding agencies, academic institutions, editors, publishers and others in increasing the transparency of research contributions, and in more granular tracking of attribution and associated credit.

On May 16, 2012, individuals from a wide range of sectors with interests in scholarly publishing gathered together at an invitation-only workshop at Harvard University to discuss contributorship and attribution, and to explore alternative models. The engaging presentations, vigorous debate from varied perspectives, and productive breakout group discussions combined to produce dynamic interactions and recommendations on how attribution and citation practices could be improved.

The discussions that were generated through this workshop illuminate the ways in which scholarly communication is changing. Scientific communication is becoming more diverse and new channels are being created that make use of semantic technology. Scholars and researchers are creating and disseminating works that do not appear in traditional peer-reviewed journals but that warrant intellectual or technical credit. As the diversity of these works increases and they become more important, the limitations of traditional practices related to authorship are becoming increasingly obvious. New attribution and citation practices and standards are required.

Problems identified

Varied authorship conventions across disciplines

Authorship conventions vary across disciplines, and author position on publications has specific and different meanings in different fields. This makes it difficult not only to assess individual contributions and evaluate their importance in the work being reported, but also to make comparisons across fields. Judith Singer, Senior Vice Provost for Faculty Development and Diversity at Harvard University, made the important point that in interdisciplinary collaborations a person's name could end up in different positions in different publications despite what they did. As interdisciplinary collaborations grow, the problem will get worse. There was general agreement during the workshop that the 'secret code' used in author lists to signal level and type of contribution is 'broken' and new approaches are needed.

Increasing number of authors on articles

The number of authors on articles is increasing. At the upper end, as in high-energy physics, there can be thousands of collaborators and so thousands of names listed as authors on articles. Knowing

who did what and how much they contributed to the effort is becoming impossible. Peter Galison, University Professor at Harvard, described how physicists have had to grapple with problems of credit and attribution over the decades since the 1940s when collaborations involved just a handful of researchers. The 'collective authorship' of today requires complex internal regulation and governance to address the issues that come with it, including balancing credit between different types of contributions and appropriate allocation of accountability and responsibility. The problem isn't restricted to physics. Liz Allen, Senior Evaluation Adviser at the Wellcome Trust, reported that the number of authors on Wellcome Trust-associated genetics papers had increased nearly threefold between 2006 and 2010 (from 10.21 to 28.82, with a range 1-505 for the latter). Author 'inflation' will almost certainly increase as the number of 'big' science consortium projects grows.

Inadequate definitions of authorship

With the increasing complexity and scale of research collaborations it is becoming increasingly difficult to agree on what constitutes 'authorship'. Both Ginny Barbour, Chief Editor of *PLOS Medicine*, and Christine Laine, Editor of *Annals of Internal Medicine*, considered that the ICMJE (International Committee of Medical Journal Editors) authorship criteria, which have been adopted by many journals, are outdated. Barbour commented, though, that they are the best we currently have. New robust definitions are needed that work across disciplines, better capture specific individual contributions at different levels, convey credit more transparently, ensure appropriate accountability for the integrity of the work being reported, and help eliminate gaming and exploitation. Ginny Barbour went so far as to say that the current meaning of authorship is worthless and the time has come to "throw out the term 'author'".

Inability to identify individual contributions

For most scholarly publications all that is available for attribution is a list of authors, hence it is difficult to know what each person contributed and the extent of their input. Some journals require disclosure of individual contributions at article submission, but the practice is patchy, there is no standardization, and reporting reliability can be affected by contribution disclosure form format. As well as preventing accurate attribution and appropriate allocation of credit, the lack of transparency of contribution obscures accountability. The problem of identifying individual contributions in publications makes it difficult for hiring, promotion, and tenure committees to make well-founded decisions. Researchers aren't able to identify and so direct questions to the most appropriate person in an authorship list, for example the person who was responsible for a specific part of the work, or find potential collaborators as readily as they could if more detailed information were provided.

Damaging effect of authorship disputes

Authorship disputes can have serious repercussions for all those involved. Publication of work can be delayed, sometimes for a long time, career progression and publication of related work can be held back, reputations can be damaged, and relationships can break down, to the detriment of both individuals and science more widely. Some disputes stem from the lack of adequate definitions of authorship. But there is also concern that many research groups and collaborations lack systematic and coordinated processes to deal with authorship, and so problems, inequalities, and disputes can arise. Authorship disputes are regularly seen by journal editors. It is not, however, their responsibility to resolve these, or to decide which individuals should be included as authors. Those responsibilities lie with the researchers themselves.

Current metrics are inadequate to capture and include new forms of scholarship and effort

Scholars and researchers are now creating and disseminating works outside the traditional scholarly literature, for example datasets, software and blogs, which represent significant contributions and merit attribution and credit. The importance of capturing and including outcomes beyond journal articles and acknowledging non-traditional scholarly roles such as data acquisition and curation is becoming recognized. Current metrics are inadequate, and new approaches are needed to

accommodate the evolution of scholarly communication and better represent the varied intellectual and technical output of individuals.

Inability of funders to track the outputs of their funding

The lack of any systematic way of identifying funding sources or those funded by specific organizations from the author lists and acknowledgements in publications has made it very difficult for funders to identify the publications that have arisen from their funding. This long-standing problem may soon be resolved. Fred Dylla, Executive Director and CEO of the American Institute of Physics, described 'FundRef' (<http://www.crossref.org/fundref/>), a new initiative from CrossRef launched just before the workshop which will standardize the collection and display of funding-agency information in scholarly publications. A pilot is being developed and refined, and a public demo is scheduled for later this year.

Name ambiguity leads to misattribution of credit and accountability

Authors' names do not unambiguously identify individuals. This problem is becoming more acute as researchers from nations where a small number of names are shared between very large numbers of people become more active in research publication. Accurate identification of researchers is important, not only so they can receive appropriate credit, but also for accountability. It was noted that the introduction of ORCID (Open Researcher and Contributor ID, <http://about.orcid.org/>) should go a long way to resolving this issue. The registry of unique identifiers will go live in October and one of the recommendations from the workshop was that all researchers should get their ORCID ID and start using it for all their research-related and intellectual outputs, not just for traditional journal articles.

Aggregation of attribution information from a large number of sources

As more and increasingly diverse forms of scholarly communication are produced from a potentially very wide range of sources, attribution information issues will become more complex. Various questions will need to be addressed: Where and by whom will attribution information be gathered? Where will it be stored? Should there be a global identification service? Is a central discovery hub for researcher–work information desirable, and if it is, is it feasible?

Possible solutions to the problems identified

Move to a more descriptive, codified and standardized system of contributor roles and contribution types

A common taxonomy for contributor roles and contribution types in scholarly publications was recommended. This would help resolve a number of the problems identified – different authorship conventions across disciplines, difficulties evaluating individual contributions in multi-author scholarly efforts as author numbers increase, inadequate authorship definitions, and sparse and non-standardized contributorship schemes. Standard terms for roles and contribution types that work across disciplines and allow documentation of a whole range of individual contributions need to be developed. They should also be fit for reuse by a range of diverse third parties. The issue of whether curation and management of standard terms should be done by a trusted neutral organization needs to be considered.

To achieve wide adoption, the new system would need to be viewed as fair, be able to accommodate evolving models of scholarly communication, and be easy to use. Embedding capture of contributorship information into existing workflows and doing this at submission stage, where other metadata and information are being requested, would minimize effort for users. A contribution ontology with three high-level headings was suggested: (i) conceptual and intellectual; (ii) technical and experimental; and (iii) organizational and communication. It was suggested that the

concept of 'authorship/author' be replaced with 'contributorship/contributor type'. Clifford Lynch, Executive Director of the Coalition for Networked Information, cautioned that because authorship is a legal concept as well as an intellectual one, for example in copyright, the implications of making such a move need to be considered and addressed. Lynch also made the very valid point that for any large cultural change to be successful it needs to be evidence-based rather than hope-based, and there was now the opportunity to conduct experiments.

Measures to avoid or minimize authorship disputes

Cassandra Extavour, Associate Professor of Organismic and Evolutionary Biology at Harvard, and Christine Laine, Editor of *Annals of Internal Medicine*, provided similar constructive advice for avoiding authorship problems and disputes. Adoption of this by research groups and those involved in collaborative projects would help avoid misunderstandings and disputes, and so the potential negative consequences that go with them. The following guidelines could apply to both authorship and contributorship:

- (i) Have a clear authorship/contributorship policy.
- (ii) Discuss and document individual contributor roles and provisional authorship early on, ideally at the start of the project before work begins.
- (iii) Review contributions as the work progresses, revise roles and authorship accordingly until journal submission.
- (iv) Keep a descriptive authorship contribution list.
- (v) Document the reasons for author/contributor additions and deletions, and get agreement for changes from all individuals.
- (vi) Make sure all authors/contributors see and approve the final manuscript.

Introduction of new metrics

Rather than attempting to represent an individual's output in a single measure, it was felt that a 'portfolio' approach would allow a researcher's intellectual and technical contributions to be seen in fuller dimensionality. This will become increasingly important as the range of scholarly contributions that warrant attribution and credit broadens.

The following were considered important for evaluation in a perfect world: the diversity of contributions should be recognized; influence and wider impact should be evaluated; evaluations should be appropriate and fit for purpose; teamwork and leadership should be evaluated and rewarded; individuals should be able to provide input on which aspects of their work are representative of their contribution; and it should be possible to trace the chain of personal influence. Various challenges to the introduction of any new system were identified: the discrepancy between aspects that are important and things that can be measured; the need for agreement on what is valued in people; the problem that not all aspects can be quantified, or even identified; the lack of clarity around reasons for assessment and appropriate metrics: different sets of values may be needed for different roles, institutions, career stages, etc; and predictivity is difficult, and some characteristics are valued only retrospectively.

The need to be provocative in creating new metrics was put forward. For example to determine (possibly through surveys of users of attribution metrics) the behaviors it would be beneficial for authors to adopt, and then provide high-profile incentives, such as prestigious fellowships, to motivate those behaviors.

Brian Lowe from Cornell University Library demonstrated with the VIVO project (<http://vivoweb.org/>) that a rich and complex research and scholarship environment can be represented semantically through sharing of linked open data, and include the many roles and processes involved in research and the roles held by individuals over the course of their professional lives.

Development of data citation standards

Christine Borgman, Professor and Presidential Chair in Information Studies at UCLA, stressed that as data become more valuable the concept of authorship in the data world needs to be understood and accommodated. A number of organizations are working on developing data attribution and citation practices and standards. The workshop identified databases and data as being particularly important as new units of scholarship warranting attribution, and also data citation as ripe for standardization. It was felt a priority to work with existing efforts to identify emerging standards and to draft proposals to NISO (National Information Standards Organization; <http://www.niso.org/>) to act as a convenor for stakeholder communities. The NAS Board on Research Data and Information (BRDI; <http://sites.nationalacademies.org/PGA/brdi/index.htm>), CODATA-ICTSI Task Group on data citation (see <http://www.codata.org/taskgroups/TGdatacitation/index.html>), and ORCID were identified as leaders in this area (the last as leaders in author identifiers, which should be integrated into data citation).

Conclusions and next steps

There was general agreement that it had been an excellent meeting and a much-needed look at an important issue. The presence of representatives from varied sectors with interests in scholarly contributorship and attribution meant the discussion was wide-ranging, and diverse perspectives could be taken into account. The problems with current attribution practices identified by the various speakers, who included researchers, scholars, funders, editors, publishers and senior university administrators, resonated with the audience. It was clear that the existence of different conventions for author position in publications across different disciplines is problematic and makes evaluation and decision making difficult in hiring, promotion, and funding processes. The increasing number of authors on publications, up to thousands in some fields, also makes it very difficult to know who did what and so for their contribution to be evaluated. It was felt that the use of author name lists to signal level and type of contribution is 'broken', and that new approaches are required to describe and present contributor roles and contributions. A common taxonomy of standard terms for roles and contribution types that works across disciplines and that can accommodate new forms of scholarship and scholarly communication is needed.

Two new initiatives, ORCID and FundRef, will undoubtedly have great positive impacts on scholarly publication and the evaluation of individual contributions to research efforts. ORCID will, through author name disambiguation, enable not only more accurate attribution in traditional research articles, but also attribution of credit in the new forms of scholarly communication that are arising. FundRef will provide a standard way of reporting funding sources for published scholarly research, allowing funders and others to systematically track the outcomes of the research and individuals they support, something that is currently very difficult to do. Both initiatives will also aid transparency, and so accountability.

The existing metrics for scholarly evaluation were felt to be inadequate and outdated. New metrics are needed, and a 'portfolio' approach to represent an individual's output. As the diversity of scholarly output increases it will become ever more difficult to capture attribution in a single measure. The increasing importance of data and other non-traditional forms of scholarship was recognized, and that improved and new attribution and citation practice standards that encompass them need to be developed. Data and databases were identified as being particularly important, and an effort on this front timely as there is an emerging infrastructure, standards, and community that could form the basis for forward progress.

The breakout sessions were very productive, and the groups came up with lists of recommendations for further actions in the areas they'd been assigned. It was decided that a number would be taken forward and coordinated by the workshop Planning Committee as the next steps in the road map toward improved and enriched scholarly attribution and contributorship practices.

Next steps

Contributorship definition and recoding pilot project A pilot project will be scoped involving publishers, funders, and researchers to review current contributor role vocabularies in use, followed by the design of a baseline cross-disciplinary taxonomy (keeping it contained to roughly 10 roles); this will probably initially focus on biomedical research, given much work has already been undertaken in this arena. Once a prototype has been developed, its usability and value will be tested against existing author annotation systems with a number of recently published articles. The output of this process will then be circulated for reaction and feedback to academic administrators, funders, researchers, publishers, and other relevant groups.

Acknowledgements text-mining research project Work will be undertaken to further analyze article acknowledgement text, to identify contributor roles that do not rise to the level of authorship in the current models being used across the publishing world. Data providers such as Elsevier and Thomson-Reuters could assist with this by making their data available.

Data citation standard Members of the Planning Committee are also members of or closely connected with the key organizations involved in developing data citation practices and standards. In addition, a member of the Planning Committee (Micah Altman) is now taking a leadership role in drafting the upcoming CODATA-ICTSI report on data citation that will be presented at the CODATA assembly in October. Involvement in this report will be continued, and then coordination begun with CODATA and BRDI in initiating a NISO standard's development effort. A group (including Micah Altman and Christine Borgman) will follow up with Todd Carpenter, Executive Director of NISO, to explore the options for creating an officially sanctioned standard for data citation.

~~~~~



## The Workshop

### Participants and sponsorship

A list of the workshop participants is given in Appendix 1. The workshop was sponsored by the Wellcome Trust and the Institute for Quantitative Social Science (IQSS), Harvard University. Kevin Condon from IQSS acted as note taker.

### Workshop goals

This invitation-only workshop brought together members of the academic, publishing, library, technology and funder communities interested in exploring alternative scholarly contributorship and attribution models. There are no universal conventions for agreeing which scholarly contributions qualify individuals as authors or for conveying authorship precedence. As a consequence, authorship tends to obfuscate the contributions of those involved in collaborative research and writing endeavors. Publication credit can be misunderstood, and so misapportioned by traditional impact measures. There is therefore growing interest in increasing the transparency of research contributions, and in more granular tracking of attribution and associated credit. Some journals now require contribution disclosures upon article submission – in either structured or free-text form – but the practice is patchy, there is no standardization and reporting reliability can be affected by disclosure form format.

The objective of the workshop was to explore alternative approaches, and to converge on a road map toward the creation of contributorship and attribution models and technologies that have the potential to:

- facilitate authorship/contributorship disclosure processes and policies
- identify good practices for tracking authorship in portions and versions of work
- minimize authorship disputes
- enable appropriate credit for contributions in multi-authored works – across all aspects of the research being reported (including data curation, statistical analysis, etc.)
- improve automated tracking of funding outcomes and impact
- support new forms of social and research networking
- further developments in data management and nanopublication
- inform the ‘science of science’, e.g. studies of productivity over a career trajectory
- enable new metrics of credit and attribution.

In conjunction with the workshop, the Planning Committee designed and conducted preliminary research to explore attitudes towards scholarly attribution and patterns of attribution for non-authorial contributions.

### Workshop format

There were two main sections to the workshop: a keynote talk plus topic presentations in the morning, and breakout sessions on various aspects of contributorship and attribution in the afternoon. The breakout groups were tasked with producing concise recommendations for taking their areas forward over the next two years.

## Introduction

Amy Brand, Assistant Provost for Faculty Appointments at Harvard University, introduced the day and set the scene by outlining the main issues and problems. The allocation of credit for research and discovery has a huge impact on people's lives, but current citation practices are failing and are outmoded. They are information-poor relative to today's complex and densely populated research culture, authors' names don't unambiguously identify individuals, multi-author citations don't provide rich enough information about who did what in collaborations, and citations can obscure attribution and perpetuate author disputes and misattribution. Credit systems are also being gamed.

Currently all we have in most cases for scholarly publications are author names and their order, and it is difficult to know what each person contributed and the extent of their input. Middle authors particularly miss out, as they are perceived to have contributed less than the first or last authors. Different disciplines have different conventions for name order, making things very confusing. As multi-disciplinary research increases, the situation is likely to become worse. Some name-ordering systems may also have unintended consequences. For example, in economics, where authors are listed alphabetically in publications, it's been shown that those with earlier surname initials are significantly more likely to receive tenure and other professional success.<sup>1</sup>

It is often said that change in scholarly communication will be slow until the people involved in tenure and promotion change their criteria. There are cultural challenges to overcome. A qualitative shift to more descriptive attribution would provide publication records with richer information on what individuals have actually contributed. ORCID (Open Researcher and Contributor ID, <http://about.orcid.org/>) is creating a foundation for more reliable author-centric navigation of the research literature which will enable capture and delivery of more granular contributor data.

The hope is to emerge at end of the day with an outline to create a plan for improved and enriched practices in attribution and citation.

## Presentations

### **Keynote: The Collective Author** Peter Galison, University Professor, Harvard

Peter Galison described the tremendous growth in the size of collaborations in physics over the past 60 years, and how each generation has grappled with the changing scale and the associated issues of credit and attribution. High-energy physics collaborations have grown from a handful of physicists in the 1940s, through dozens in the 1960s, hundreds in the 1980s, to the several thousand scientists who might be involved in experiments today, for example at CERN. Galison introduced the concept of the 'collective author' – a group that functions without the association of the name of a single individual (e.g. in the context of their lab) and often without a lead institution. Complex internal regulation and governance are required, along with novel ways to attribute credit. With collective credit comes collective responsibility. Physicists have been forced to ask very fundamental questions on governance and the representative structure for decision making, and to address the conundrums of collective authorship: how to recognize the prerogative of individuals while preserving coherence of the group; who should speak for the group and who should take responsibility for claims; how to deal with dissent when there is the need to speak with one voice,

---

<sup>1</sup> Einav, L. and Yariv, L. (2006) "What's in a Surname? The Effects of Surname Initials on Academic Success." *Journal of Economic Perspectives*, 20(1): 175–187.

for example if 2500 physicists are involved in a collaboration and 600 don't want to publish; how to balance credit between different contributions.

Career advancement in physics has become harder in the last 10 years because of the large number of authors appearing on publications. As the number of authors on papers has increased, it has become not only more difficult to evaluate individual contributions, but also to make comparisons with other fields in science. The remarkable trajectory of increasing numbers in collaborations in physics has been accompanied by a parallel evolution in the size, scope, and cost of experiments. As much as 10-15 years (and millions of dollars) can be spent on an experiment, so the stakes can be very high as to how governance and attribution decisions are made.

There is a need to better understand attribution practices because of their impact on communities. Galison feels that new systems of evaluation are needed as science moves forward – we can't use "yesterday's evaluations". The focus is already shifting to non-traditional role credit/attribution, and an understanding that one contribution type is not necessarily superior to others. As different disciplines come together, there will be a synthesis of systems, taking into account the different cultures of these disciplines. Studying a diversity of approaches will give more options to develop best practices, but it is likely that one size will never fit all.

**Diverse perspectives on authorship: Scholar, Funder, Publisher, Provost** Moderated by Diane Scott-Lichter, Publisher and Head of Publishing Division, American Association for Cancer Research

**Issues of authorship practice and policy in experimental science research labs** Cassandra Extavour, Associate Professor of Organismic and Evolutionary Biology, Harvard

Cassandra Extavour emphasized that the wants, needs, and goals of scholars at various career stages (undergraduate student, graduate student, postdoc, research assistant, faculty pre- and post-tenure) are different. The importance of authorship and publication is also different at the various stages: for principal investigators (PIs) it is critical, and needs to be of a specific type, demonstrating the ability to lead a productive group; for postdocs it may be important, e.g. if they want an academic position, and needs to be first author; for graduate students it may be important, depending on career goals; for research assistants and undergraduates it's a bonus to be included as an author. Junior members may not realize that it's important for them to be acknowledged, and some may have over-expectations of what it takes to be included as an author. The contributions of undergraduates and research assistants may in some cases be as important as those of postdocs.

Authorship order has specific and different meanings in different fields. In Extavour's own field of evolutionary biology, the first author is recognized as the one who carried out the bulk of the experiments, the last author is the PI (who contributed intellectual content and funding), and the authors in between carried out different amounts of experimental work/analysis. But with large authorship lists it becomes less clear who did what based on position in the list.

In her group, authorship is merited for several types of contribution but must be substantive, and she as the PI decides whether a contribution is substantive. There can be conflicts if there are over- or underestimates of contributions. She handles this by having a clear lab policy on authorship, starting discussions on provisional authorship at the start of projects and revising continually as the work progresses until journal submission. It was disappointing to hear that most of her group have come from labs where these issues aren't discussed. She also always drafts a descriptive authorship contribution list even if not required by journals.

**Contribution and attribution in research: funder perspectives** Liz Allen, Senior Evaluation Adviser, Wellcome Trust

The Wellcome Trust is a UK-based global charitable foundation that spends around GBP650 million each year on research through a number of schemes, supporting more than 3000 researchers in over 30 countries. It needs to account for its spending and decide who to fund in future. Currently, however, it is very difficult to identify individual contributions in publications (which at the moment are used as the main indicator of research outcome) and so answer the question “what did the Wellcome-funded people do?” All funders face similar problems.

There is no systematic way of identifying funding source or those funded by specific organizations from the author listing and acknowledgements in publications. There can even be a disconnect between the information in these two places. The problem of tracking is exacerbated by the increasing numbers of authors listed in publications, particularly in consortium work in general, and in the biosciences in particular. This author ‘inflation’ is likely due to the increase in collaborative ‘big’ science and the stimulus of national research assessment exercises. The average number of authors on Wellcome Trust-associated papers increased between 2006 and 2010 from 10.21 to 28.82 (with a range 1-505 for the latter) for genetics papers, and from 6.28 to 8.32 for all papers. There is a great need for increased transparency so that funders can readily identify the publications and outputs resulting from their funding and who did what.

**Authorship: a (medical) journal’s perspective** Christine Laine, Editor, *Annals of Internal Medicine*

Christine Laine outlined why authorship matters: there are academic, social, and financial implications, readers want to know who did what, and, most importantly, those who are accountable for the integrity of the work can be identified. There are various definitions of authorship, but generally an author is an individual who has made a substantive intellectual contribution. The fundamental principles are: all persons designated as authors should qualify for authorship, all those who qualify should be listed, and each author should have participated sufficiently to take public responsibility for appropriate portions of the content. The key question all authors should ask themselves is: would I be willing to stand by the content of this paper if someone questioned its integrity? Laine feels that the ICMJE authorship criteria,<sup>2</sup> which have been adopted by many journals rather than creating their own, are now outdated.

It is ultimately the role of researchers themselves and not journal editors to decide which individuals have contributed sufficiently to earn the designation ‘author’. Editors cannot police the author or contributor listing for every submission, but may sometimes have suspicions that an author list is incomplete or includes undeserving (guest or gift) authors. Information on and help with authorship problems can be found on COPE’s (the Committee on Publication Ethics) website (<http://publicationethics.org/>). An increasing number of biomedical journals are collecting and publishing information on actual author contributions. Some ask for a ‘guarantor’ to be identified, who takes responsibility for the integrity of work as a whole, from inception to published article. Documenting contributions doesn’t solve everything - authors can mislead or lie - but it helps. At *Annals*, all correspondence goes to all authors, which can also be helpful in identifying misconduct.

Laine proposed a contemporary definition of authorship: (1) authors are individuals identified by the research group to have made substantial contributions to the reported work and who agree to be

---

<sup>2</sup> International Committee of Medical Journal Editors, Uniform Requirements for Manuscripts Submitted to Biomedical Journals, [http://www.icmje.org/ethical\\_1author.html](http://www.icmje.org/ethical_1author.html).

accountable for these contributions; (2) in addition to being accountable for the parts of the work he or she has done, an author should be able to identify which of his/her co-authors are responsible for specific other parts of the work; (3) an author should have confidence in the integrity of the contributions of his/her co-authors; and (4) all authors should review and approve the final manuscript.

Authorship battles undermine confidence in research, and Laine gave some advice on avoiding authorship battles: discuss individual contributor roles and authorship early on, ideally before beginning the work; document discussion consensus; review contributions as the work progresses and modify accordingly; document the rationale for contributor additions/deletions and get consent from all involved; and get all contributors to sign off before manuscript submission.

Laine finished by posing some questions: in scholarly communication, what is the difference (if any) between being an author and being a writer? What level of contribution qualifies a person as an 'author' as opposed to a 'writer' or other type of contributor? Can someone not write a word and still qualify to be an 'author' of a scholarly article? Can an individual draft the entire manuscript and still not qualify to be an 'author' of a scholarly publication?

**Authorship: a provostial perspective (deeply informed by scholar, selection/funding committee, and editorial board perspectives)** Judith Singer, Senior Vice Provost for Faculty Development and Diversity, Harvard University

Judith Singer suggested that there are alignments and misalignments between the many purposes of scholarly publishing and the multiple ways in which scholarly publications are used by individuals occupying different roles to make very different kinds of decisions. Journals and editors evaluate the work in individual papers, but in appointment and funding decisions it is the person and a body of work that are being evaluated, so the stakes are higher. Tenure decisions are very high stake, with long-term financial and staff employment issues, and institutions like Harvard cannot afford to make mistakes. The tendency is therefore to be conservative and this makes authorship issues important. Conventions on author position, however, vary "dizzingly" across fields. In interdisciplinary collaborations, a person's name could therefore end up in different positions in different publications despite what they did.

Singer is interested in predictive validity, taking into account: the sorts of research questions asked by the individual (are they interesting, provocative?); consistency in asking these sorts of questions; the degree to which their work impacts a field (not necessarily based on just citation counts); trans-disciplinary impact; longevity (i.e. are they just a 'one-shot wonder?'); and consistency and independence (are they able to produce results without a collaborator or without the same collaborator?). Harvard uses hard metrics for tenure decisions, but also takes into account qualitative measures (e.g. internal and external letters) and asks expert adjudicators who know the field to look at all the evidence. When asked about the standing of social media and blogging, Singer felt their role is more about getting known, but reputation does matter in funding. Academic communities are reputation driven. She concluded that you have to ask what it is you trying to accomplish. There are many goals, and the same metrics cannot be used for all the goals.

**Related initiatives: context and feasibility** Moderated by Merce Crosas, Director of Product Development, IQSS, Harvard

**PLOS: big science and medicine and the attribution of authorship** Ginny Barbour, Chief Editor, *PLOS Medicine*, Public Library of Science, Chair of COPE

Medical and scientific research is increasingly being done by large groups of collaborators on complex projects, often coming from many different institutions and countries. Ginny Barbour feels that what constitutes ‘authorship’ is becoming harder to agree on and some fundamental issues need to be addressed. Do the current definitions of authorship sufficiently capture the specific contributions of those involved in the research? Are these definitions sufficient to give appropriate academic credit? Do these definitions ensure that authors take appropriate responsibility for the papers on which they are listed as authors?

There is abuse of authorship (e.g. putting a personal stamp on everything even when authorship is not warranted, trading authorship between papers) and this can have a serious impact on research integrity. Editors regularly have to deal with difficult authorship situations, as can be seen from the cases presented at the COPE Forum and detailed on its website<sup>3</sup>. *PLOS Medicine* has very strict authorship/contributorship rules and sends all authors letters asking them to confirm authorship, chasing as necessary, but this is very time consuming. Barbour agrees with previous speaker Christine Laine that the ICMJE authorship guidelines are outdated. They are, however, the best we currently have (although it has been shown how industry uses the guidelines to manipulate authorship<sup>4</sup>).

Barbour thinks the current meaning of authorship is worthless, it is being exploited and gamed and doesn’t encompass the full scholarly spectrum (e.g. blogs, reviews, tweets). We need to rethink authorship in the digital age and not just technically. Robust definitions that are portable across specialties and methods for attribution at different levels need to be developed. The time has come to “throw out the term ‘author’”. Also, article-level metadata allow attribution only at research-article level, and we need to aim to be able to do this at within-article level.

**Why are the attribution and citation of scientific data important? (Introduction to the BRDI-CODATA report)** Christine Borgman, Professor and Presidential Chair in Information Studies, UCLA

Christine Borgman stressed that as data becomes a more valuable object we need to understand what authorship means in the data world. A number of organizations and agencies are working on developing data attribution and citation practices and standards. A joint BRDI-CODATA<sup>5</sup> symposium was held in August 2011<sup>6</sup> on developing and implementing scientific data citation standards and practices. The driving questions it addressed were:

- What are the major technical issues that need to be considered?

---

<sup>3</sup> <http://publicationethics.org/cases>.

<sup>4</sup> Matheson A (2011) How Industry Uses the ICMJE Guidelines to Manipulate Authorship—And How They Should Be Revised. *PLoS Med* 8(8): e1001072. doi:10.1371/journal.pmed.1001072.

<sup>5</sup> BRDI-CODATA: Board on Research Data and Information-Committee on Data for Science and Technology.

<sup>6</sup> Developing Data Attribution and Citation Practices and Standards. US CODATA and the Board on Research Data and Information, in collaboration with CODATA-ICSTI Task Group on Data Citation Standards and Practices ([http://sites.nationalacademies.org/PGA/brdi/PGA\\_064019](http://sites.nationalacademies.org/PGA/brdi/PGA_064019)). (ICSTI: International Council for Scientific and Technical Information)

- What are the major scientific issues that need to be considered? Which are universal for all types of research and which are field- or context-specific?
- What are the major institutional, financial, legal, and socio-cultural issues that need to be considered? Which are universal for all types of research and which are field- or context-specific?
- What is the status of data attribution and citation practices in individual fields in the natural and social (economic and political) sciences in the United States and internationally?
- What are the respective roles and approaches of the main actors in the research enterprise and what are the similarities and differences in disciplines and countries?

The Symposium report is to be published shortly.

Borgman outlined infrastructure considerations for digital objects: social practice (why cite and attribute data? how to cite?), usability, identity, persistence, discoverability (how to identify, locate and retrieve), provenance (how and when data have been modified or transformed, and who has handled it – it is crucial to know this in order to be able to trust the data), relationships, intellectual property (what can I do with it – reuse, reproduce – who owns the rights?), and policy (whose policy and what type when a whole chain of people and organizations have been involved?). She mentioned that some data don't get posted because everyone thinks it is someone else's responsibility. People want credit for what they do, so a mechanism is needed to establish a chain of credit, for the author, depositor, funder and other parties.

**FundRef: a new project from a publisher–funding agency collaboration to standardize the collection and display of funding agency information in scholarly publications** Fred Dylla, Executive Director and CEO, American Institute of Physics

Funding agencies and the research communities supported by funding agencies have a responsibility to report the results of the supported research to various constituencies, including the public. But presently, Fred Dylla explained, funders and grant numbers are not always included in the resulting scholarly publications. Where publishers do collect this information it is presented in non-standard locations in the publications, making it difficult to find. If this information can be collected, displayed and archived as an agreed-upon component of the standard metadata associated with a scholarly publication, an important service will be provided to the authors, funding agencies and readers who access these publications. A new initiative is aiming to do this. CrossRef is managing a pilot project, called 'FundRef' (<http://www.crossref.org/fundref/>), to develop and test a protocol for collecting funding-agency information (funder name and grant number) from authors at the time of manuscript submission to journals. Seven scholarly publishers and four funding agencies are participating in the design and implementation of the pilot project.

FundRef's goals are to create a standard nomenclature for the capture, display and use of funding-agency metadata, and to demonstrate an industry-wide methodology for connecting scholarly publications to research funders. A taxonomy of funder names will be tested, using the registry of 4000 funder names that Elsevier is making available as the pilot taxonomy. Funders will help maintain the registry of funder names and instruct awardees to provide correct information to publishers. CrossRef is enabling publishers to deposit funder and grant number metadata as part of the regular CrossRef deposit, so that will be available along with the DOI and bibliographic data. Funders will use the data to report the published results of research awards. The DOIs could be used to trace funding, but many agencies don't have a place for the DOI to resolve to. The pilot will test the feasibility of the workflow and integration with the funder name registry, and an API to query the CrossRef data using funder names/grant numbers/publication metadata will be developed. The

project was officially launched in March, pilot development and refinement will take place April-August, with subgroups working in parallel, and a public demo will be available in the Fall.

**Contributor roles in VIVO** Brian Lowe, VIVO Semantic Lead Developer, Cornell University Library

VIVO (not an acronym; <http://vivoweb.org/>) is a project that enables international networking of researchers across all disciplines through the sharing of linked open semantic data. It originated at Cornell University, to provide a single point of entry for discovery of the research and scholarship environment there, but has, with NIH funding, expanded to other universities nationally. All publicly available information on the people, departments, facilities, and other resources that collectively make up the research and scholarship environment are brought together in a single site. Brian Lowe described how researcher roles and collaborations are represented, and what makes VIVO work. Critically, all the data are represented semantically, using WC3 semantic web technology. No traditional relational databases are used.

As the project progressed it was realized that a greater complexity was needed to describe what was required. All the roles and processes involved in research, and the roles held by people in different time spans over the course of their professional lives had to be included. How roles relate to processes had to be configured. The outcomes of all the processes had to be identified and the appropriate level of granularity for reporting. The importance of including, and capturing, researcher outcomes beyond journal articles was recognized, for example datasets and nanopublications, that currently don't appear in traditional publication records. If items can be captured and modelled earlier in the process than the point of publication, it will be possible to end up with more detailed information than just the list of authors at the end of the process.

## Exploratory research

In conjunction with the workshop, the Planning Committee designed and conducted preliminary research to explore attitudes towards scholarly attribution and patterns of attribution for non-authorial contributions. Mike Taylor, Principal Investigator, Elsevier Labs, presented a preliminary analysis at the workshop, which was subsequently revised for this report.

To explore attitudes, a survey was developed and fielded. Respondents to the survey generally approached authorship from a holistic perspective, emphasizing overall responsibility for the project, its operation, and results. They also recognized most other research activities as meriting authorship in most cases, with the exceptions of potentially non-intellectual contributions such as the contribution of reagents, funding, and general administrative help.

To explore patterns of attribution related to non-authorial contributions, a text analysis of the 'Acknowledgements' section of a large sample of scientific publications was performed. For comparison, the frequency of these terms across papers indexed in CiteSeer was also examined. In both the Elsevier and CiteSeer sample, financial support and data were frequently acknowledged, and manuscript, technical and statistical assistance were also regularly acknowledged in both. A conjecture is that authors tend to acknowledge substantial effort and resources that are necessary for the research but not considered a substantial intellectual contribution in that instance. This interpretation is consistent with the survey results.

Details of the survey and an analysis of the results to date are given in Appendix 2.



## Breakout sessions

### Breakout Group 1: A taxonomy for roles and contributions in scholarly publications

*Facilitated by David Shotton, Emeritus Fellow, University of Oxford, and Margie Hlava, Founder/CEO, Access Innovations* (Session participants: Ginny Barbour, Amy Brand, Irene Hames, Veronique Kiermer, Christine Laine, Brian Lowe, Clifford Lynch, August Muench, Diane Scott-Lichter, Liz Williams)

#### (1) Scope

The group set out to consider how publications currently capture contributor role distinctions, what works and what doesn't in indicating level of effort or different types of contribution, and whether there are any existing models that should be considered. They also asked whether it is possible to arrive at a taxonomy of common contributor role definitions that works across disciplines. David Shotton's SCoRO (Scholarly Contributions and Roles Ontology, <http://purl.org/spar/scoro/>) model, prepared for the workshop, was used as a springboard for the discussions.

#### (2) Issues and questions

##### What would the new taxonomy be used for?

- *Credit for tenure, promotion, and grant funding* Author roles and contributions need to be clarified because, at present, interpretation of the meaning of author order is complex and very field-specific. The meeting as a whole described this system as 'broken'.
- *Communicating with peers* Attribution of specific work to individuals will enable effective peer-to-peer search, for example for potential collaborators, and direct contact with the relevant people.
- *Accountability and responsibility* There will be transparency, of contribution, funding and conflicts of interests.

##### Requirements of the taxonomy

- Needs to be generic and work across all disciplines, allow documentation of a whole range of individual contributions, be extensible and permit entry of details of contributions.
- Needs to be fair. Avoiding, for example, assigning percentage of total credit for a paper to individual authors as this could be used to promote purely bibliometric evaluations of individuals' worth.
- Must be easy for everyone to use and understand.
- Drivers are required as change won't happen unless funders and leading universities, in addition to publishers, support it. Any change must also be acceptable to those who generate citation metrics.

##### Features of the ontology

A high-level outline with three headings focused on scholarly publication and the nature of the contributions was proposed:

- *Conceptual and intellectual* Formulation of the research questions; design of the experiments; interpretation of the results; intellectual and moral responsibility for the integrity of the paper, as a whole and for individual contributions.

- *Technical and experimental* Implementation of the investigation and undertaking of experiments; obtaining specimens and subjects; acquisition and processing of data; analysis of results.
- *Organizational and communication* Writing the project proposal and obtaining funding; ensuring compliance, e.g. ethics committees approval, informed consent; writing up results into a paper; illustration of the paper – selection and use of data; reporting.

### **(3) Proposed actions**

- *Establish a Working Group* A core group of activists is needed to drive this initiative forward, before involving various relevant organizations.
- *Begin building a suitable taxonomy of contribution types* A survey of current practices is needed and a review of classifications in current use.
- *Prepare a grant proposal* Sources of funding need to be identified and an application submitted to cover building of the taxonomy, research on acceptability, implementation, and a feasibility study. Funding for the latter will be needed because journals/publishers will require resources to incorporate the project into their workflows, especially if a large-scale effort is needed.
- *Express revised classification in a number of formats* For example, as a simple taxonomy and an OWL ontology.
- *Provide draft taxonomy for comments to all relevant groups* For example, publishers, funders, researchers, scholarly publishing associations and providers of services, standards organizations, individuals responsible for hiring, tenure, and promotion decisions, citation metrics groups.
- *Revise taxonomy to incorporate feedback*
- *Establish pilot project for feasibility study with retrospective analysis* Possible duration 18 months (12 months for implementation + 6 months for evaluation). Target and invite a small number of journals to take part (broad subject coverage is needed since different disciplines currently use different methods). Each early implementer exemplar journal retrospectively encodes all or selected papers for one year of publication using the new contributor annotation system. Adjustments are made as needed to the system and scalability determined. The results are made available to the other constituent communities to evaluate and report on. Implementation is revised as required to accommodate feedback and ensure wide uptake.
- *Abolish 'authorship/author' and replace with 'contributorship/contributor type'* 'Authorship is broken' has become somewhat of a marketing phrase to indicate that the problem is in need of serious discussion and decisions need to be made for a better representation of the current types of attribution for research in the public record. Many at the meeting felt that this was rather an extreme statement and it isn't authorship per se that is broken, but rather that it is the interpretation of contribution, and hence assignment of credit, based on author order that is. The suggestion from the group was that 'authorship/author' be replaced with 'contributorship/contributor type'.
- *Consideration of reservations* There were concerns that abolishing 'author' is not feasible because the concept is too deeply embedded in scholarly culture. Also, if 'authors' are to be replaced by 'contributors', only some of whom may be 'writers', there may be legal complications (this issue was discussed further in Clifford Lynch's summary of the day).

## Breakout Group 2: Implementation considerations

Facilitated by Gudmundur Thorisson, Research Associate, University of Leicester, and Geoffrey Bilder, Director of Strategic Initiatives, CrossRef (Interim Technical Director, ORCID) (Session participants: Alberto Accomazzi, Patricia Brennan, Stephanie Diment, Fred Dylla, Jo McEntyre, Rebecca Rosen, Griffin Weber, David Weinberger)

### (1) Scope

The group focused on two main themes: (1) data gathering and (2) data storage and access. Participants were asked to think about barriers and specific actions that could/should be taken to overcome them.

### (2) Issues and questions

The group came up with a list of important questions that need to be addressed:

- Where will attribution information be gathered?
- Where will the attribution information be stored, and, potentially, later copied to?
- Who will gather the information and how? Will it be mostly manual input, mostly automated, or both?
- What can/should be the role of ORCID as a global identification service and, potentially, as a central discovery hub for researcher ↔ work information?
- What level of detail does a central discovery hub need to cover to be of real use for data aggregation?
- If a central hub is not feasible, or even desired, how do we deal with the problem of usefully aggregating research information from a potentially large number of sources (journals, funders, digital repositories, wikis etc.), and with even knowing about all relevant sources?
- What can/should be the role of ontologies?
- How can we leverage tools familiar to non-tech end users (like Excel spread sheets with embedded ontologies) to gather data in structured form and with sufficient semantics?
- How do we balance a completionist approach (being thorough) with a pragmatist approach ("the stupidest thing that could possibly work")?

### (3) Proposed actions

The group felt it was not possible to come up with specific implementation actions because most strategic goals were yet to be defined. They therefore proposed some more general guidelines and pointers on implementation.

- *Data capture* The focus should be on embedding capture of contributorship information into existing workflows, rather than creating new ones, to minimize the effort for authors as much as possible. Capture should ideally take place at the stage of submission of manuscript/dataset/other work where the person(s) responsible are currently asked for various manuscript metadata, including the list of co-authors/contributors. The group also felt it would be useful, if it were possible, to go back and annotate (possibly via ORCID) publication claims retrospectively. Such self-claims by authors would, however, need to be clearly distinguished from 'authoritative' contributor information provided at submission time.
- *Data structure and semantics* The use of standard terms for roles and contributions is crucial for integration, but adoption uptake will be low unless the popup-menu lists or other user-interface elements presented to end users are kept short, certainly no more than 10 to 15 items long. It should be possible to further qualify broad categories with free-text comments (note: this is basically a shorter version of what the taxonomy group, Group 1, has elaborated in greater detail). There were suggestions to use text-mine-available content to generate and refine suitable terms in semi-automated fashion. The issue of whether

curation and management of standard terms ought to be done via a trusted, neutral organization (similar to how CrossRef is running FundRef) was also raised.

- *Data availability and discovery* Whichever contributorship information is gathered needs to be fit for reuse by a range of diverse third parties. This implies that data must be accessible in structured, machine-readable form and under an appropriate license or public-domain waiver. Data also need to be easily discoverable, either via a central data clearinghouse or, alternatively, via central indexing services (one of which could be ORCID) which contain pointers to where the contributorship information is actually stored.

### **Breakout Group 3: Social considerations, barriers and incentives**

*Facilitated by Cameron Neylon, Biophysicist, Science in the Open, and Anita de Waard, Director of Disruptive Technologies, Elsevier Labs* (Session participants: Liz Allen, Monica Bradford, Michael Fisher, Alyssa Goodman, Elizabeth Knoll, Alexa McCray, Judith Singer, Hideaki Takeda, Simeon Warner)

#### **(1) Scope**

The group considered how, in a perfect world, they would like to be evaluated for their work, the barriers that prevent them and others evaluating their work in this ideal way, general actions that could be taken to overcome these barriers, and specific actions they could take within the scope of their own work to help overcome them. The group also asked some open questions: Is the way people are currently assessed (by small groups of senior, entitled 'judges' sitting in a room) still a valid method of evaluation? Is it valid to evaluate people, rather than 'work', 'group', and so on? Is person-focused evaluation the best way to support science?

#### **(2) Issues and questions**

For 'perfect world' evaluation of work, the following were considered important:

- To have the diversity of contributions (domain, media, output types) recognized.
- For individuals to be able to provide input on which aspects of their work are representative of their contribution.
- To be evaluated on influence and wider impact.
- For evaluations to be appropriate and fit for purpose.
- To be able to trace the chain of personal influence.
- To see teamwork and leadership evaluated and rewarded.

They also identified challenges to a 'perfect world' evaluation:

- The discrepancy between aspects that are important and things that can be measured.
- The need to agree on what we value in people. Valued characteristics the group came up with included: creativity, networking skills, influence on peers in the same field, influence on peers in other fields, translational influence on policy, practice and the public, persuasiveness, good 'taste' in choosing research questions (are they exciting, novel?), knowledge of subject area, craftsmanship, leadership, 'generativeness', originality vs curation, perseverance, productivity.
- There are some issues with using these values:
  - Not all aspects can be quantified, or even identified (even by the person themselves).
  - There is lack of clarity around reasons for assessment and appropriate metrics: different sets of values are needed for different roles, institutions, career stages, etc.
  - Predictivity is difficult – some characteristics are valued purely retrospectively.

### **(3) Proposed actions**

- Expand the list of valued characteristics and see which aspects can be tracked/measured.
- Give everyone their ORCID ID and tell them to start using it. Start to track and collect all personal activities, thoughts, blogs, and so on using this ID.
- Publishers to allow a space in the author's details for the ORCID ID; add microattribution?
- Use the ID in workflow/codeshare systems.
- To improve buy-in with, for example, university provosts and deans, create a few (real) exemplars to show that these values are an improvement on existing measures and can answer questions such as: Who did we miss? How could we have found them?

### **Breakout Group 4: The potential for new metrics**

*Facilitated by Caitlin Trasande, Head of Science Metrics, Digital Science, and Jason Priem, PhD Student, University of North Carolina Chapel Hill (Session participants: Kevin Dolby, Erin Driver-Linn, Glen Ellison, Christopher Erdmann, Julia Hawks, Rebecca Lawrence, Alberto Pepe, Lydia Snover)*

#### **(1) Scope**

The group looked at the current state of metrics and considered what the ideal attribution economy would look like and how to get there.

#### **(2) Issues and questions**

- How do attribution and metrics intersect? Is attribution 'credit'?
- What is the current state of the art?
  - What are the standard creatorship metrics? Mainly: patents (citation from papers and citation from other patents), papers and conference papers (citations from other papers, co-authorship, especially interdisciplinary), and grant funding.
  - What is wrong with these existing metrics? Some suggestions: out of date, lack of high-level overview, lack of dimensionality, lack of aggregation, the portfolio is unbalanced, or we don't have the chance to balance it.
- What should we encourage?
  - What kinds of creatorship deserve to be recognized?
  - What kinds of behaviors do we want to encourage (openness, troubleshooting, etc.) in research?
- How do we encourage it?
  - How might new models of attribution encourage the behaviors we like?
  - What are some of the features of a microattribution economy? Who wins in such an economy? Who loses?
  - What potential new metrics are there? What are their strengths and weaknesses?
- Where do we go from here?
  - How can we make this happen?
  - Which people are most likely and instrumental to do to this?

#### **(3) Proposed actions**

- Be provocative in creating new metrics which utilize author attribution (the group were unanimous on this).
- Survey users of attribution metrics (e.g. deans, provosts, chairs, chancellors) to determine the top desirable behaviors of authors that would make their assessments better. For example, in order to combat uncertainty about what author number 34 on a paper of 50 authors actually

did, have authors micropublish work related to their contribution on an open-access (OA) platform.

- Determine the behaviors you want authors to adopt.
- ‘Characterize don’t measure’. Rather than trying to boil things down to a single number (how many papers, how many co-authors), have a ‘portfolio’ approach to representing an author’s output. This would allow a user of the metric (academic administrator, search committee) to see an author in fuller dimensionality. Parameters could include diversity of output, influence of work, teamwork, leadership.
- Focus on young researchers. They are more malleable and have a long career ahead.
- Provide a high-profile prize incentive (along the lines of the MacArthur Fellowship) to motivate desired behavior. For example, to incentivize micropublishing on OA platforms, target new postdocs and require award recipients to OA micropublish all of their research for the duration of the award. The award would need to have substantial cash funds attached so that these and the prestige of the prize suffice as an incentive to the author. Funding would need to be secured and a selection committee formed. Award multiple scientists so that a cohort can be established and tracked. Create prestige with a strong brand (e.g. The Harvard OA Micropublishing Fellow or the Nature Micropublishing Fellow). Compare awardees’ scientific trajectory on a number of measures with a control group (e.g. anonymized equally-qualifying applicants who were not selected) and assess:
  - Did awardees publish more?
  - Were they more successful at attracting grants?
  - Were they more successful on the job market?
  - Did they establish more collaborations?
  - Were they invited to more talks, conferences, etc?
  - Is their work more interdisciplinary?
  - Did their work receive more citations/downloads/social media attention?

Track award recipients beyond award duration to determine if the desired behavior (e.g. micropublishing) sustainably established itself with that scientist, influenced collaborators, seeded the laboratory culture, and so on.

## **Breakout Group 5: Citation/attribution for data and non-traditional forms and units of scholarship**

*Facilitated by Micah Altman, Director of Research, MIT Libraries, and Tim Clark, Director of Bioinformatics, MGH/HMS (Session participants: Marguerite Avery, Chris Biemesderfer, Christine Borgman, Todd Carpenter, Merce Crosas, Gregg Gordon, Laurel Haak, Mark Hahnel, Robert Hilliker, Emilie Marcus, Craig Van Dyck, Alex Wade)*

### **(1) Scope**

The group focused on how attribution (and citation) processes, policies and practices can be improved with respect to ‘non-traditional’ scholarly works (such as datasets, blogs, and software), and what might ultimately be gained from these improvements. Ideally, what these non-traditional works suggest as practices ought not to be considered as ‘lesser’ alternatives to traditional practices, but should help inform a broad-spectrum unified approach to both traditional and non-traditional works.

### **(2) Issues and questions**

- Scholars and researchers are now creating and disseminating works which do not appear in

traditional peer-reviewed journals but reflect significant contributions and so could (or might?) warrant attribution. The group identified the following: databases (data edits?), software (bug fixes?), blog entries (blog comments?), problem lists (syllabi?), lemmas (proof steps?), protocols (reagents? reagent preparation methods?), peer reviews (peer comments?), visualizations (dynamic processes?), tweets (retweets?), workflows, syllabi, talks (slide decks?), web sites (wiki edits?).

- The group identified databases/data as being particularly important, and software as important (and a subset identified protocols as important for attribution). Data were considered particularly timely, as data citation practices are clearly developing through systems such as *DataCite*, *Dataverse*, and *Pangea*. Thus, there is an emerging infrastructure, standards, and community to form a basis for forward progress.
- In particular, the group identified the area of data citation as ripe for standardization, and felt that it was a priority to work with existing efforts to identify emerging standards on data citation, and draft a proposal to NISO to act as a convenor for stakeholder communities. The NAS Board on Data and Information, CODATA-ICTSI Task Group on data citation, and ORCID were identified as leaders in this area (the last as leaders in author identifiers – which should be integrated into data citation).

### (3) Proposed actions

- Review Thorny Staples project at the Smithsonian.
- Circulate ORCID-EU proposal – how to use ORCIDs in data citations.
- Review forthcoming BRDI/NAS report.
- Coordinate with CODATA-ICTSI.
- Identify existing standards on data citations and approach NISO with a draft proposal to convene stakeholder communities.
- Reach out to funding agencies (e.g. Wellcome, Sloan, Gates) to clarify definition of data, revisit the model of tripartite responsibility.
- Develop proposed standard.
- Proof of concept.
- Pilot – early adopters.
- Disseminate standard.
- Evaluate.

### **Summary and further questions for moving the debate forward** Clifford Lynch, Executive Director, Coalition for Networked Information

Clifford Lynch found it difficult to summarize the day as it had been a very wide-ranging, and provocative, discussion. He would attempt to pull out the main lines of argument and identify the most important questions for moving the debate forward. The system of scholarly communication and evaluation is a complex process and there had been talk during the day of changing things in a short time. He stressed that there was now the opportunity to conduct experiments rather than just writing a manifesto. Any large cultural change needs to be evidence-based rather than hope-based, and adoption would be faster if there was supporting evidence.

There had been a lot of discussion on the notion that ‘authorship is broken’. The increasing numbers of authors on articles – thousands in the case of high-energy physics and ‘big’ science – alters the scale of various problems (what holds for a small number of authors doesn’t extrapolate to big science) and it’s clear that the system of credit, accountability, and understanding of what each person has done is breaking down. The significance of position in author listings as regards level of

participation and contribution varies from discipline to discipline. This makes evaluation and decision making especially problematic for tenure committees, which are made up of individuals from different disciplines. There had been general agreement during the workshop that the 'secret code' used in author lists to signal level and type of contribution is "terribly broken".

A whole new approach is needed, especially as scholarly publishing becomes increasingly electronic and moves to the 'article of the future'. More than one solution may be needed. How do we change traditional journal articles to make evaluations easier, better and more representative? How are the roles of various 'authors' to be codified? Moving to a codified and standardized system where descriptions of what the various authors did will allow greater understanding of their contributions and aid accountability. It will also make it easier for researchers trying to find collaborators to identify the specific individual in a long list of authors who was responsible, for example, for the technique they're interested in. An experiment had been proposed during the workshop (by the taxonomy breakout group, Group 1) to take a small number of journals, implement parallel processes of attribution, using a new codification of contributor roles, and see if this makes a difference.

Scholarly communication is becoming broader than the journal article and we need to accommodate this evolution. An important question involves working out how to evaluate individual contributions in these new forms of scholarly communication. A prominent issue is data, and the efforts being made to develop data attribution and citation practices and standards had been reported. There will be the opportunity to recognize the roles of data acquisition, provision and curation.

An important conversation is also going on about metadata. What do we want to attach to journal articles? Who uses the metadata and for what? Metadata captured during the publication process can be fed into a range of processes. FundRef is a very interesting and welcome new initiative which is aiming to standardize the collection and display of funding information in scholarly publications. This will both help transparency of sponsorship and allow funding organizations to track the outcomes of the research and researchers they have funded.

Lynch put forward a word of caution. It had been suggested that the time had come to "throw out the term 'author'", and that 'author' should be replaced with 'contributor'. What are the implications of moving away from 'authorship' to 'contributorship'? Authorship is both an intellectual and a legal concept. The legal framework of copyright is centered round the concept of authorship, and if articles are changed so as to have only 'contributors', some of whom would be the article's 'writers' and others who would not qualify for authorship by current criteria, then to whom would copyright and legal responsibility apply? This needs to be thought through.

## **Conclusions and next steps**

There was general agreement that it had been an excellent meeting and a much-needed look at an important issue. The presence of representatives from varied sectors with interests in scholarly contributorship and attribution meant the discussion was wide-ranging, and diverse perspectives could be taken into account. The problems with current attribution practices identified by the various speakers, who included researchers, scholars, funders, editors, publishers and senior university administrators, resonated with the audience. It was clear that the existence of different conventions for author position in publications across different disciplines is problematic and makes evaluation and decision making difficult in hiring, promotion, and funding processes. The increasing number of authors on publications, up to thousands in some fields, also makes it very difficult to know who did what and so for their contribution to be evaluated. It was felt that the use of author



name lists to signal level and type of contribution is 'broken', and that new approaches are required to describe and present contributor roles and contributions. A common taxonomy of standard terms for roles and contribution types that works across disciplines and that can accommodate new forms of scholarship and scholarly communication is needed.

Two new initiatives, ORCID and FundRef, will undoubtedly have great positive impacts on scholarly publication and the evaluation of individual contributions to research efforts. ORCID will, through author name disambiguation, enable not only more accurate attribution in traditional research articles, but also attribution of credit in the new forms of scholarly communication that are arising. FundRef will provide a standard way of reporting funding sources for published scholarly research, allowing funders and others to systematically track the outcomes of the research and individuals they support, something that is currently very difficult to do. Both initiatives will also aid transparency, and so accountability.

The existing metrics for scholarly evaluation were felt to be inadequate and outdated. New metrics are needed, and a 'portfolio' approach to represent an individual's output. As the diversity of scholarly output increases it will become ever more difficult to capture attribution in a single measure. The increasing importance of data and other non-traditional forms of scholarship was recognized, and that improved and new attribution and citation practice standards that encompass them need to be developed. Data and databases were identified as being particularly important, and an effort on this front timely as there is an emerging infrastructure, standards, and community that could form the basis for forward progress.

The breakout sessions were very productive, and the groups came up with lists of recommendations for further actions in the areas they'd been assigned. It was decided that a number would be taken forward and coordinated by the workshop Planning Committee as the next steps in the road map toward improved and enriched scholarly attribution and contributorship practices.

### **Next steps**

**Contributorship definition and recoding pilot project** A pilot project will be scoped involving publishers, funders, and researchers to review current contributor role vocabularies in use, followed by the design of a baseline cross-disciplinary taxonomy (keeping it contained to roughly 10 roles); this will probably initially focus on biomedical research, given much work has already been undertaken in this arena. Once a prototype has been developed, its usability and value will be tested against existing author annotation systems with a number of recently published articles. The output of this process will then be circulated for reaction and feedback to academic administrators, funders, researchers, publishers, and other relevant groups.

**Acknowledgements text-mining research project** Work will be undertaken to further analyze article acknowledgement text, to identify contributor roles that do not rise to the level of authorship in the current models being used across the publishing world. Data providers such as Elsevier and Thomson-Reuters could assist with this by making their data available.

**Data citation standard** Members of the Planning Committee are also members of or closely connected with the key organizations involved in developing data citation practices and standards. In addition, a member of the Planning Committee (Micah Altman) is now taking a leadership role in drafting the upcoming CODATA-ICTSI report on data citation that will be presented at the CODATA assembly in October. Involvement in this report will be continued, and then coordination begun with CODATA and BRDI in initiating a NISO standard's development effort. A group (including Micah Altman and Christine Borgman) will follow up with Todd Carpenter, Executive Director of NISO, to explore the options for creating an officially sanctioned standard for data citation.

## Appendix 1 Workshop participants

|                                                                                                           |
|-----------------------------------------------------------------------------------------------------------|
| Alberto Accomazzi, Program Manager, Harvard-Smithsonian Center for Astrophysics                           |
| Liz Allen, Senior Evaluation Adviser, Wellcome Trust                                                      |
| Micah Altman, Director of Research, MIT Libraries                                                         |
| Marguerite Avery, Senior Acquisitions Editor, The MIT Press                                               |
| Ginny Barbour, Chief Editor, <i>PLOS Medicine</i> , Public Library of Science; Chair, COPE                |
| Chris Biemesderfer, Director of Publishing, American Astronomical Society                                 |
| Geoffrey Bilder, Director of Strategic Initiatives, CrossRef (Interim Technical Director, ORCID)          |
| Christine Borgman, Professor and Presidential Chair in Information Studies, UCLA                          |
| Monica Bradford, Executive Editor, <i>Science/AAAS</i>                                                    |
| Amy Brand, Assistant Provost for Faculty Appointments, Harvard University                                 |
| Patricia Brennan, Product Manager, Thomson Reuters                                                        |
| Todd Carpenter, Executive Director, National Information Standards Organization                           |
| Tim Clark, Director of Bioinformatics, MGH/HMS                                                            |
| Kevin Condon, QA and Production Support Manager, IQSS                                                     |
| Merce Crosas, Director of Product Development, IQSS/Harvard                                               |
| Anita de Waard, Director of Disruptive Technologies, Elsevier Labs                                        |
| Stephanie Diment, Publisher Nature Life Sciences Research, Nature Publishing Group                        |
| Kevin Dolby, Evaluation Adviser, Wellcome Trust                                                           |
| Erin Driver-Linn, Director of Institutional Research and the Initiative in Learning and Teaching, Harvard |
| H. Frederick Dylla, Executive Director and CEO, American Institute of Physics                             |
| Glen Ellison, Professor of Economics, MIT                                                                 |
| Christopher Erdmann, Head Librarian, Harvard-Smithsonian Center for Astrophysics                          |
| Cassandra Extavour, Associate Professor of Organismic & Evolutionary Biology, Harvard                     |
| Michael Fisher, Executive Editor for Science and Medicine, Harvard University Press                       |
| Peter Galison, University Professor, Harvard                                                              |
| Alyssa Goodman, Professor of Astronomy, Harvard University                                                |
| Gregg Gordon, President & CEO, SSRN                                                                       |
| Laurel Haak, Executive Director, ORCID                                                                    |
| Mark Hahnel, Founder, figshare                                                                            |
| Irene Hames, Editorial Advisor and Consultant                                                             |
| Julia Hawks, VP North America, Symplectic LTD                                                             |
| Robert Hilliker, Digital Repository Manager, Columbia University                                          |
| Marjorie Hlava, President, Access Innovations                                                             |
| Veronique Kiermer, Executive Editor, Nature Publishing Group                                              |
| Elizabeth Knoll, Executive Editor-at-Large, Harvard University Press                                      |
| Christine Laine, Editor, <i>Annals of Internal Medicine</i>                                               |
| Rebecca Lawrence, Director New Product Development, Faculty of 1000                                       |
| Brian Lowe, VIVO Semantic Lead Developer, Cornell University Library                                      |
| Clifford Lynch, Executive Director, Coalition for Networked Information                                   |
| Emilie Marcus, CEO and Editor-in-Chief, Cell Press                                                        |
| Alexa McCray, Associate Professor of Medicine, Harvard Medical School                                     |

|                                                                                                          |
|----------------------------------------------------------------------------------------------------------|
| Jo McEntyre, Head of Literature Services, EMBL-EBI                                                       |
| August Muench, Astrophysicist, Harvard-Smithsonian Center for Astrophysics                               |
| Cameron Neylon, Biophysicist, Science in the Open                                                        |
| Alberto Pepe, Postdoctoral Researcher, Harvard-Smithsonian Center for Astrophysics                       |
| Jason Priem, PhD Student, University of North Carolina Chapel Hill                                       |
| Rebecca Rosen, AAAS Science & Technology Policy Fellow, National Science Foundation                      |
| Diane Scott-Lichter, Publisher and Head of Publishing Division, American Association for Cancer Research |
| David Shotton, Emeritus Fellow, University of Oxford                                                     |
| Judith Singer, Senior Vice Provost, Harvard University                                                   |
| Lydia Snover, Director of Institutional Research, MIT                                                    |
| Hideaki Takeda, Professor, National Institute of Informatics, Japan                                      |
| Mike Taylor, Principal Investigator, Elsevier Labs                                                       |
| Gudmundur Thorisson, Research Associate, University of Leicester                                         |
| Caitlin Trasande, Head of Science Metrics, Digital Science                                               |
| Craig Van Dyck, Vice President Global Content Management, Wiley-Blackwell                                |
| Alex Wade, Director Scholarly Communication, Microsoft Research                                          |
| Simeon Warner, Director of Repository Development, Cornell University Library                            |
| Griffin Weber, Chief Technology Officer, HMS                                                             |
| David Weinberger, Harvard Library Innovation Lab, Harvard Law School                                     |
| Liz Williams, Executive Editor, <i>Journal of Cell Biology</i>                                           |

## **Appendix 2 Research details** Micah Altman, Director of Research, MIT Libraries, and Mike Taylor, Principal Investigator, Elsevier Labs

### *Details of the survey*

The survey into attitudes towards scholarly attribution included the following:

- A set of questions measuring attitudes toward various forms of research activity – and in particular, whether they should be recognized as authorship.
- A set of questions measuring the respondents' direct responsibility for (the same) forms of research activity.
- A set of demographic questions aimed at eliciting contributors' professional role, discipline, and career stage.

The survey was sent out electronically to all invitees to the workshop, and to a pool of 500 active researchers previously recruited by Elsevier Labs for the purpose of responding to market surveys. 67 responses were received from the first pool and 177 from the second pool.

Most (>55%) of respondents strongly agreed that responsibility for the overall results, interpretation of the results, and conception of the project each merited full authorship. Most respondents also agreed or strongly agreed that design of experimental methods, responsibility for day-to-day operations, direct oversight, critical review, statistical analysis, assurance of data accuracy, and performing the literature review also merited full authorship. Most respondents agreed at least slightly that software development activities, preparing graphics, managing the laboratory, contribution of reagents, and funding procurement merited authorship. There was, however, a substantial minority of respondents that opposed or strongly opposed linking authorship to funding, or to contribution of reagents, and most respondents opposed linking authorship with administrative duties.

### *Details of the text analysis*

To explore patterns of attribution related to non-authorial contributions, a text analysis of the 'Acknowledgements' section of a large sample of scientific publications was performed. For the main part of this analysis, Elsevier supplied a sample of 10400 acknowledgement sections. The acknowledgements data was a random selection of articles containing acknowledgements sections from 1061785 articles published across all disciplines on ScienceDirect since 2008.

These extracts were converted to XML form, and parsed using the text-mining package [tm](http://tm.r-forge.r-project.org/) (<http://tm.r-forge.r-project.org/>), available for the statistical programming language, R. The resulting corpus was normalized (for case, whitespace, and punctuation), cleaned (numbers, individual letters, and a conservative list of stop words were removed), and stemmed.

There were 367 (stemmed) terms that each appeared in at least 1% of the documents, 64 terms that appeared in at least 2%, and 35 terms that appeared in at least 5% of documents. (Testing with 10% and 20% thresholds did not reveal additional keywords relevant to contributorship.) Of these keywords, most were clearly semantically unrelated to contributorship or scholarly roles. The remaining, potentially relevant, keywords were confirmed through inspection of a selection of texts in which these terms occurred.

For comparison, the frequency of these terms across papers indexed in CiteSeer was also examined. The AckSeer service was used to measure term frequency. It was found (based on an examination of

queries for the 40 most common English words) that this service indexed approximately 492242 non-empty acknowledgement sections.

Several forms of contributorship were regularly acknowledged ( $\geq 5\%$  of responses in the Elsevier sample):

- contributions of funding, grants and financial support (34% Elsevier, 34% CiteSeer)
- assistance with manuscripts (17% Elsevier, 5% CiteSeer)
- technical and technological assistance (16% Elsevier, 6% CiteSeer)
- contributions of data (12% Elsevier, 13% CiteSeer)
- assistance with animals/veterinary assistance (12% Elsevier, <1% CiteSeer)
- statistical assistance (5% Elsevier, 3% CiteSeer).

Also notable, 'reviewers' and 'referees' were acknowledged in 3% of the Elsevier sample, and in 11% of the CiteSeer sample.

A conjecture is that authors tend to acknowledge substantial effort and resources that are necessary for the research, but not considered a substantial intellectual contribution in that instance. This interpretation is consistent with the survey results.

~~~~~