> [T]hose ESRC grant applicants who plan to generate data are responsible for preparing and submitting data management and sharing plans for their research projects as an integral part of the application. It is then a responsibility of the award holder to incorporate data management and sharing as an indivisible part of the research project to increase the potential for data to be shared. We require that the data must be made available for preparation for [reuse] and/or archiving with the ESRC data service providers within three months of the end of the award otherwise we will withhold the final payment.

3. Knowledge Access

Sharing knowledge about scientific discoveries is a foundation of modern science, but workshop participants noted that funding agencies need to understand that knowledge, and therefore knowledge sharing, should be broadly defined to encompass both data and code. They also noted that knowledge sharing takes many forms and should be encouraged, including traditional academic journal publishing as well as other mechanisms such as discussion forums, recommendations, wikis, file-sharing sites, blogs, and microblogs (Trasande and Hannay 2010). However, substantial barriers to knowledge access persist despite mandates to promote sharing. For example, in spite of the embrace of Open Access publishing, the voluntary adoption rate by scientists has been low (around 15%–20%). Mandates have increased these numbers to around 70% for NIH-funded research and in institutions, such as Southampton or CERN, that have adopted these policies. Nevertheless, this means that even with mandatory participation, some 30% of research is not openly available (Fenner 2010).

3.1. Social Issues

Workshop participants agreed that attribution for new forms of scientific activity was critical to promoting knowledge access. Researchers will provide access to their work if they are given credit for their labor. Attribution for scholarly work requires the ability to uniquely identify both specific contributors to research and specific scientific contributions (Fenner 2010). Participants felt strongly that an author-identification system that transcends institutional, disciplinary, and national boundaries would help create a “clear and unambiguous scholarly record” of research activities associated with an individual and help provide unambiguous attribution for researcher contributions, whether they appear as publications, patents, or data sets (Office of Science and Technology Policy 2009; National Science Foundation 2011). An author-identification system would also allow for “microattribution” for research contributions not associated with a peer-reviewed journal publication (Credit Where Credit Is Due 2009). In the current system, a significant portion of scientific work remains unrecognized because there are no formal methods for providing attribution for this labor

(Pfeiffenberger 2010; Elias 2010; Lambe 2010; Neylon 2010b). Examples include the work of students, research assistants, and other non-author collaborators or the participants in large-scale scientific infrastructure projects. This renders much of the labor that makes science possible "invisible"—"the visible manifestations of science conceal an intricate social network of relationships, trust and perceived authority, underlying how science gets funded, how scientists decide to collaborate, and how new knowledge gets validated" (Lambe 2010).

With an international, unambiguous, and comprehensive attribution framework, data could be collected on the full range of research labor and outputs (Lane 2010), enabling a "wide range of derived metrics and rankings" (Trasande and Hannay 2010) that could be used to better understand the complexity of scientific labor and research. The Knowledge Organisation System described in Section 2.2 may help make "visible" the scientific labor previously left "invisible" by connecting formally recognized scientific outputs and metrics to the informal labor and social networks that support the generation, dissemination, and reuse of scientific knowledge.

Authentication, verification, quality assurance and control, and privacy provisions are critical to the success of a persistent author-identification system (Neylon 2010b). The systems must be able to handle duplication and redundancies, and should "not be affected by name changes, cultural difference in name order, inconsistent first-name abbreviations or the use of different alphabets" (Credit Where Credit Is Due 2009). There is a financial incentive for this as well: "In the current world ill-considered, non-transparent, and irreproducible metric systems will almost inevitably lead to legal claims" (Neylon 2010b).

At the heart of resolving this issue are establishing and authenticating unique researcher identifiers. To avoid misidentifications, access to individual researcher identifiers should be under the control of individual researchers (Neylon 2010b), and researchers should be required to authenticate their biographical and professional information (Trasande and Hannay 2010). A centralized database designed to populate grant and job applications, bio-sketches, or reports, which can otherwise be onerous and repetitive tasks (Evans 2010), would likely be an incentive for researchers to keep this information current. The Lattes Platform, a research database, adopts a related approach, requiring users to register before applying for government funding.

> ***Exemplar*: Open Researcher and Contributor ID (ORCID)**
> Proposed in 2009, ORCID is a system of unique alphanumeric strings for each researcher. It is backed by 23 organizations, including Thomson Reuters, the British Library, and the Wellcome Trust (Credit Where Credit Is Due 2009). ORCID intends to create a central registry of unique identifiers linked to other author schemes (http://www.orcid.org/).

For a system based on unique researcher identifiers, all parties must trust that the identifier system is reliable, authentic, and immutable. Accordingly, the development of systems that are unable to generate a significant level of trust is "likely to limit and fragment any effort to coordinate, federate, or integrate differing identity solutions in the research space. Therefore interoperability of any developed system with the wider web must be a prime consideration" (Neylon 2010b). One possible method for establishing trust and ensuring proper attribution is encouraging the use of an identification system like ORCID (see Exemplar), and establishing publishing practices that make the "creation and capture [of unique identifiers] an integral part of the editorial process" (Trasande and Hannay 2010).

3.2. Technical Issues

The development of a persistent, trusted, ubiquitous, and interoperable centralized repository for housing the unique researcher identifiers may provide a "trusted broker" for promoting knowledge access and attribution (Trasande and Hannay 2010; Neylon 2010b). Currently, a number of identification tools exist or are under development, including ORCID, Vivo, Lattes Platform, Public Library of Science (PLoS), and PubMedCentral. For example, ORCID proposes to create a "central registry of unique identifiers for individual researchers and an open and transparent linking mechanism between ORCID and other current author ID schemes" (http://www.orcid.org/). Lattes, a fully developed researcher database that allows for verification (Aragão 2010; Lane 2010) has now been adopted in 17 countries in Latin America, Europe, and Africa [Aragão].

3.3. Role of Funding Agencies

Funding agencies worldwide can play a critical role in encouraging knowledge access and the implementation of an identification system to facilitate attribution. Agencies are uniquely positioned to require data and code sharing in publicly funded work, and they support the infrastructure and tools for data and code sharing. Participants felt that funding agencies should embrace the creation of identification systems and ensure their adoption by requiring registration as a prerequisite to applying for agency funding (Trasande and Hannay 2010). Participants also thought that agencies could support a research library coalition that would provide an international open-standard data set for bibliometric information for all published work worldwide (Conlon 2010).

4. Conclusions and Next Steps

At the conclusion of the workshop, participants agreed on a set of attributes of the "ideal" attribution landscape 5 years into the future [Greer]. It would include a framework of openness and international standards for data and knowledge; reliable and unique identifiers for each researcher, organization, publication, and the

relationship to each other; a link between all publications and their appropriate data; continuous investment for data preservation and access; and formal and informal training of students, researchers, and personnel at funding agencies.

In white papers submitted before the workshop, and during presentations and in discussions at the workshop itself, participants identified a set of actions that would achieve this vision:

(1) Establish a system of persistent identifiers for both researchers and their outputs. The following specific suggestions were made about the characteristics of such a system:

- Create taxonomies of scientific data—Enable the cataloguing, tagging, and parsing of data sets for automated recall.

- Create incentives—Encourage and offer incentives to researchers to routinely use the standardized identification schemes to annotate their data, a process that will be aided by the further development of software tools. Provide researchers with incentives to encourage them to make data sets available to the wider research community through the development and use of attribution systems. Help ensure that data sets are linked to subsequent publications and other research outputs, further aiding attribution and the reproducibility of research. Publish data and code to facilitate assessment and certification of quality and allow data sets to become part of the citable "scientific record."

- Create independent standards—Establish federally funded platforms for data and code sharing that are independent of institutions and individual researchers, and use standards of unique identification for citation and version control.

- Create a legal framework—Develop an Open Research License (ORL) to resolve conflicts between reproducibility and copyright law.

- Create a registration mechanism—Encourage the development, implementation, and use of standardized identification systems to facilitate attribution by requiring system registration as a prerequisite to applying for agency funding.

(2) Develop national and international pilot projects that compare different technical solutions for developing and maintaining open data platforms, fostering the replication of scientific research, and ensuring attribution for the intellectual contributions of researchers.

(3) Foster formal and informal training to ensure that open data and knowledge systems are maintained.

Workshop participants agreed that engaging in these efforts will provide opportunities to work across counterpart funding agencies to encourage international cooperation and the dissemination of knowledge and data.