Information & Management 47 (2010) 237–245

Contents lists available at ScienceDirect

Information & Management

journal homepage: www.elsevier.com/locate/im

Governing the data commons: Policy, practice, and the advancement of science

Joshua B. Fisher *, Louise Fortmann

Department of Environmental Science, Policy & Management, University of California at Berkeley, 137 Mulford Hall, Berkeley, CA 94720-3114, USA

ARTICLE INFO

Article history:
Received 2 October 2007
Received in revised form 9 October 2008
Accepted 1 April 2009
Available online 21 April 2010

Keywords:
Common pool resources
Common property
Commons
Data sharing
Design principles
FLUXNET
Intellectual property
Remote sensing

ABSTRACT

Property rights of shared scientific data are often analyzed by applying formal intellectual property law. However, the scales at which these laws apply are not necessarily relevant to data sharing practice among individuals (e.g., in virtual, social communities). Rather, the data sharing communities and their members often form their own policies, practices and norms governing data sharing. Using common property theory, our research objectives were to determine: (1) parallels between data sharing communities and natural resource-sharing communities; (2) factors that lead to successful data sharing; (3) circumstances under which rules, laws and policies govern data sharing. We used cases from two emerging data sharing communities in the environmental sciences—a micrometeorology community (FLUXNET) and a satellite remote sensing community. Cases of data sharing without conflict or irresolvable disputes had more principles in common with a successfully managed natural resource commons than did those characterized by conflict. Successful data sharing requires that biophysical scientists be more attentive to the social nature of data sharing.

1. Introduction

The expense and complexity of data required by researchers working across multiple disciplines at increasing spatial scales have often exceeded the means of a single scientist or institution. Thus, data sharing in distributed collections and virtual communities has become increasingly common. While the literature supports the benefits of data sharing, scientists have given relatively little attention to the property aspects of shared data outside of website statements or institutional guidelines. The literature on data sharing has focused mostly on its technical infrastructure rather than end-users. Our focus was on the property issues that arise when data are shared. We asked how data sharing communities govern themselves.

This is an issue with practical implications. One concept on data sharing in the field of biotechnology demonstrates the idea of an "anticommons" in which too many patents (over-privatization) have blocked advancements. Another potential problem may occur when cooperation or collective action ceases before a product achieves its full potential. For scientific data, misuse (use without permission and/or poor analysis) by users and mistrust by data producers can damage collaboration and interfere with scientific advancement. Understanding how rights to the shared data commons are effectively governed may help avoid adverse effects.

The governance of rights to shared data occurs at three levels. At the *macro*-level, rights to data are governed by international treaties and national laws. At the *meso*-level, rights may be affected by policies of formal and informal associations of data producers and users or institutions with which they are affiliated. At the *micro*-level, rights may be the result of negotiations among individual producers and users.

The research framework for this paper was based on both common property theory of common pool resources (CPR) and intellectual property rights (IPR) theory of intellectual property. Common property theory provides a framework for analyzing how groups develop, implement and sustain collective management regimes for *natural resources* they use in common. Those resources include fisheries, groundwater, and forests, as well as large-scale resources such as the seas, space, and the atmosphere (i.e., the "global commons"). Certain rules or design principles tend to be strongly present in successfully shared natural resource commons and less so in CPR that are in conflict [1]. We define "successful" as data sharing that takes place without conflict or unresolved disagreements. In short, common property theory aims to describe potentially universal principles for sustainable resource-sharing communities.

IPR describes different types of exclusive powers (government-granted) of control and use over *intellectual resources* (non-physical and non-exhaustible). Intellectual resources include inventions (patents), creative works (copyrights), identifying insignia (trademarks), and some forms of confidential information (trade secrets). IPR are created through statutory law and treaties. In short, IPR are

* Corresponding author. Fax: +1 818 354 9476.
E-mail address: joshbfisher@gmail.com (J.B. Fisher).

0378-7206/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.im.2010.04.001

policies, rules, laws and property relations. However, IP scholars are generally concerned with formal entitlements and do not consider how the rules affect behavior.

Data is a form of intellectual property, and the properties of shared data have been compared to CPR [2]. Open source and free software communities have been considered in the context of common property theory [3]. Questions of the public domain have been addressed in the IP literature but without reference to common property theory. Scholarly communication has been compared to an information commons. Our study was oriented towards the governance of data sharing.

We assessed practices at three levels of scientific data sharing, applying CPR theory's design principles for sustainable common natural resources [1] to practices governing commonly shared data. We were particularly interested in the emergence of rules of data sharing written or unwritten, formal or informal, by explicit or implicit understanding and the effect such rules may have on interpersonal conflict. Because individual outcomes affect the pool of data available to the community, we examined individual relationships between data users and data producers. What rights do data producers have over their data? What rights do data users have over the data analysis, manipulation, distribution, modification, and its publication? Do the rights provide enough protection from misuse or data ownership and provide reward (such as monetary, prestige, and publication) resulting in the continuation of data production? What are the sanctions against those who break the rules? Questions about property rights in shared scientific data are located at the intersection of IP theory and common property theory (Fig. 1).

2. Theoretical approaches: common property and intellectual property rights

2.1. Common property

A shared resource system, or CPR, has been defined [1] as a system that is large, making it costly (but not impossible) to exclude potential beneficiaries from obtaining benefits from its use. A CPR within the context of a legal regime of shared rights is referred to as common property. For the purposes of analyzing the complexity and variability of knowledge or information resources, a commons is defined as a resource shared by a group of people attempting to solve social problems. Once a dataset has been released into cyberspace, it is very costly (if not impossible) to control its use. Shared data in a digital form is thus a global resource.

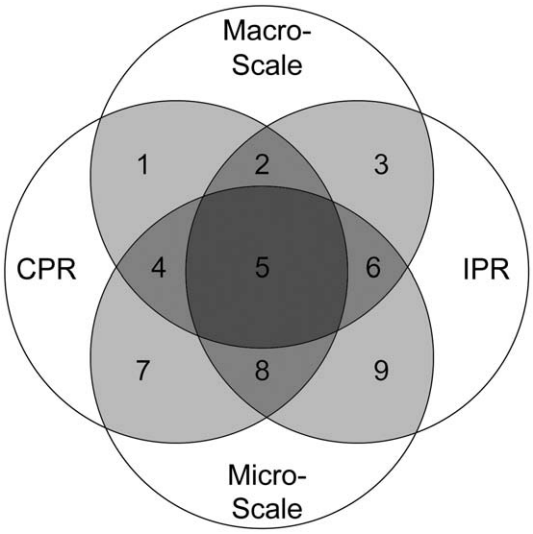

Fig. 1. The theoretical framework of data sharing that links CPR and IPR across the levels. Key:**(1).** CPR at the macro-level (global commons);**(2).** CPR and IPR at the macro-level (global commons with laws and policies);**(3).** IPR at the macro-level (international laws and policies);**(4).** CPR at the meso-level (institutionally-shared);**(5).** CPR and IPR at the meso-level (institutionally-shared IP with policies);**(6).** IPR at the meso-level (institutional policies, rules, and contracts);**(7).** CPR at the micro-level (individually-shared);**(8).** CPR and IPR at the micro-level (individually-shared with rules, understandings and contracts);**(9).** IPR at the micro-level (individual contracts).

The data producers and users differ from most commons communities in two respects. The group is scattered around the globe rather than being contiguous geographically and thus regularly involved in face to face interaction. Thus, the governance of the shared data commons is complicated by the difficulties of communication among individuals and groups with differing power, languages, cultures, and knowledge of their rights and responsibilities in data sharing. Second, a producer provides a particular dataset. While the data sharing community may have a set of appropriation rules, a user's access ultimately depends on negotiation with its producer.

Traditional analysis has shown that a sustainable commons often requires small, relatively homogeneous groups [4]. The question therefore is not whether a data sharing global community of disparate people, cultures, and institutions represents a small and homogeneous group, but whether the definition of "small" and "homogeneous" has changed in cyberspace. It may be faster to send an email to a hundred people on the internet than to say "hello" to ten people in a firm. People in the global commons all share the same transport on the information superhighway.

While shared data meets the definition of CPR, it differs from natural resource CPR because it is not subtractable: the use does not diminish the quality or value. The possibility that the supply of a particular dataset or datasets will be exhausted is not at issue in the management of the shared data commons. Indeed, as long as data users acknowledge the source and as long as their analysis does not reflect poorly on the quality of its data, users enhance the quality of the resource.

The real stakes are threefold. *First*, professionally valuable credit resulting from the analysis of shared data is limited. Publishing data without crediting the producer constitutes "free riding": using a resource without contributing to its production and management or taking more than permitted. *Second*, poor analysis reported in papers with fatal flaws, poor assumptions, etc. can reflect on both the data user and producer because the reader may not be able to identify what was the reason for the bad paper. Such an activity constitutes resource degradation in the natural resource commons. Data producers attempt to ensure that they are credited when a data set they produced has been used and thus to control publication rights. *Third*, although the data are not exhaustible, the analysis might be. Specifically, users should not publish the same data twice to answer the same question twice, thus prior use conditions future use.

CPR design principles are a useful tool for analyzing governance of property. They are conditions associated with the success of voluntary collective management of natural resources by groups of users in accomplishing their outcome. They have been used to analyze common property systems as well as large-scale resources in a global commons. The importance of design principles in successful management has been shown to be highly dependent on the situation.

Each of the these design principles can be applied to shared data:

1. *Clearly defined boundaries* specify who has a right to use a dataset and describe the dataset itself. Datasets should be easily identifiable. Thus poaching can be detected. However, like mobile natural resources, they move or are moved through a territory (cyberspace), which is not easily bounded. The users of