Into theUnknowns:
Free Culture, Online Cooperative Systems Research, and Beyond

Aaron Shaw
ashaw@cyber.law.harvard.edu
Berkman Center for Internet and Society

The challenges of analyzing free culture and the digital commons as a whole derive in part from the from the characteristics of the Internet itself. Due to the difficulties of discovering or sampling Internet-based phenomena, the full extent of cooperative and commons-based endeavors online is not only an *unknown* at this time, but may in fact be *unknowable*. Insofar as this workshop offers an opportunity to reflect on the state of knowledge about free culture and the digital commons, this last point about our collective *unknowledge* is worth dwelling on a bit. The achievements and efforts represented among the attendees of this workshop notwithstanding, my sense is that there is an awful lot that we do not know about existing free culture and commons-based practices, never mind the potential for such practices to transform creative work, cultural production, and social exchange more generally. As a result, researchers like ourselves must confront an array of methodological, theoretical, and empirical challenges that many of our colleagues and peers do not. This short essay provides a brief overview of the Online Cooperation Research project at the Berkman Center for Internet and Society in the context of these challenges and free culture research more generally.

In the past twenty years, the Internet has given rise to a thriving ecosystem of networked communities that create public goods through the collaboration of thousands of individuals. Massive, distributed systems of social interaction, these communities challenge dominant theories of human organization and production. Why does this sort of distributed cooperative system work so well? If existing mechanisms for the production and distribution of intangible assets have failed to adapt to the digitally networked environment, can online cooperative systems provide new models for the provision of knowledge-based resources?

Through the Online Cooperation Research (OCR) project we aim to contribute to this emerging field of research through three distinct interventions: (1) the largest-scale observation survey of cooperative systems online to date; (2) a series of exploratory and analytical case studies that extend Yochai Benkler's theoretical framework of the design levers for cooperative systems and commons-based production; (3) the technical development of open platforms for distributed and collaborative research. All three of these interventions have proceeded in parallel as distinct "phases" of our research agenda, illuminating each other and providing a broad empirical foundation upon which to base descriptive and analytical claims about online cooperative systems. More details about each of the components of the project will be forthcoming as we continue the project and write up the results.

A brief description of the sampling methods we are using illustrates how the OCR responds to some of the earlier concerns I raised about the state of free culture and online cooperation research. One of the problems in generating a sample of Online Cooperative Systems is the lack of an authoritative set of definitions or lists with which to find cases. We aim to overcome these limitations through a method we call "Collective Intelligence Sampling." This method leverages our existing knowledge of the field together with computational link-mining and text analysis in order to discover a large sample of sites that meet the criteria for inclusion in our study. In effect, it entails using the contents of Wikipedia as a sampling frame for broadly defined or hidden populations of websites and then retroactively assessing the bias of the sampling frame. As we are operationalizing it, Collective

Intelligence Sampling involves the following steps:

(1) Identify a valid set of "seed" terms or cases that capture a range of concepts and examples of online cooperative human systems.

(2) Exhaustively crawl the English language Wikipedia using an algorithm that scores the semantic relatedness of every article in relation to the terms of our "seed" set.

(3) Use the results of the algorithm's crawl to discover the set of Wikipedia articles that meet a minimum threshold for relatedness.

(4) Review these highly related results to identify and validate the cases of online cooperative human systems that meet the criteria for inclusion in our study.

(5) Assess the external validity of the English language Wikipedia through a series of diagnostic tests looking at a series of discrete domains for which more exhaustive data exists.

Clearly, this sampling procedure entails a number of practical compromises and decisions that merit further discussion in another venue. Nevertheless, I contend that the choices we have made offer an intriguing model for subsequent work in this field.

Going back to my earlier claims about unknowledge, the OCR project has approached the challenges of studying online cooperation and commons-based production by bracketing the "unknown unknowns" about our topic, and then using "known knowns" to inform the design of research methods that pursue "known unknowns." In this (Rumsfeldian) sense, the project looks to expand existing comparative knowledge of cooperation and commons-based activities online. The insights we generate will help to inform both the subsequent design of cooperative systems as well as the application of effective policy and regulations to this dynamic space of innovation. In this way, the OCR complements a number of projects and research agendas represented at the Free Culture Research Workshop by expanding the scope of analysis beyond well-known cases of commons-based peer production online. Lastly, through the creation of free and open research platforms and new methodologies, we believe that the OCR will facilitate the future work of scholars in this area.