

Shadow Libraries and Media Access

(aka *Media Piracy in Emerging Economies 2.0*)

As some of you know, the *Media Piracy in Emerging Economies* project (MPEE 1.0) documented the ways in which developing markets for media goods remain 'stuck' in a bad equilibrium, marked by high prices, low sales, and massive illicit access. It argued that, in the absence of efforts to address legal price and distribution issues, the escalating commitments to enforcement that we documented throughout the developing world simply reinforce this dynamic. It also argued that, in the absence of adequate capacities for due process, expanded enforcement tilts toward arbitrary, extrajudicial punishment, primarily in the form of raids and confiscations. Little of this, we argue, has any impact on the actual prevalence of pirated goods, or on consumer practices, which have continued to evolve in ways that ignore both copyright and enforcement threats.

As most or all of you know, this is a recipe for trouble. The current stakeholder politics ensure that enforcement measures fail up, leading to more pervasive, intrusive, and occasionally draconian approaches to enforcing copyright. The falling costs of technology will heighten this contradiction in the next years. The big question, looking forward, is what an eventual realignment will look like. The narrower question, which we propose to explore in this project, is what the evidentiary basis for that realignment will look like.

Our work is based on the premise that charting a more productive path on this front means understanding more about the concrete, practical side of copyright, media access, and enforcement. Although policy can influence behavior, successful policy will be adapted to reality and not struggle against it at all costs. The rapid democratization of digital technologies and, consequently, of copy culture is the unavoidable basis of that adaptation.

As presently conceived, the project has a primary and a secondary research focus: First, it is intended to explore the changing 'ecology of access' to educational and scholarly materials at the university level, particularly in the context of the emergence of large 'shadow libraries' of pirated materials and cheap means of distributing and reading them. Second, it will explore the wider changes in access to digital media and cultural goods in which these transformations are situated. For brevity's sake, we will call these two parts of the project 'Shadow Libraries' and 'Mapping Media Access,' respectively.

Like MPEE 1.0, the new project will combine legal and policy research, political-economic approaches to the affected institutional and industrial sectors, and a mix of interview and focus group methods engaging students, institutional and industry leaders, law enforcement and policymakers. Distinct from the MPEE 1.0 project, it will be built around a narrower research template that will permit stronger qualitative and quantitative comparison across countries and income levels.

At this stage, we have secured IDRC commitments for a 2-year, 3+ country study of shadow libraries, shading into wider questions of media access. We are in discussion with other partners about a complementary survey of digital media access, covering at least the US, but hopefully extending to Brazil, India, and South Africa and potentially one or more countries in Europe. Because the two phases of the work have the potential to be strongly complementary, we are hoping to secure funding for both the sociological and survey components in at least the four countries listed above.

What does this mean for March 3? We are asking for your help in the preliminary design of both phases of the work—at the meeting, but also before and after as the research plans mature. The most ambitious version of this project will require a research model that incorporates both the survey instrument and focus group/interview protocols, and that prioritizes comparability across countries and methods. We also want to build connections to complementary initiatives while avoiding duplicating existing work, so help in identifying people, organizations, and researchers in this area would be great.

We have a remarkable opportunity to shed light on the next turbulent phase of digital culture. What questions should we ask? How should we ask them?

The Growth of Shadow Libraries

Our project starts with the observation that the longstanding crisis of the developing-world library is coming to an end, but not in the way most observers anticipated. Resource scarcity, limited holdings, and poor infrastructure remain the norm. Debates over access to materials continue to revolve around decades-long fights over Xeroxing. But the combination of cheap computers (and soon, readers) and growing 'shadow libraries' of pirated scholarly material is flanking these battles, and transforming access to scholarly materials in the process. As in other fields, the expansion of legal access to materials—including by Google Books, JSTOR, and open access initiatives—will operate in a context shaped by broad de facto access through piracy.

We have anecdotal accounts of these developments in a number of developing-world universities, but no systematic analysis of them. We have no forward-looking analysis on the possible impact on developing-country publishing industries, existing library infrastructure, or legal strategies for addressing access to educational materials. Building on the MPEE network, we will conduct a study of these developments, grounded in a mixed method approach to the larger ecology of scholarly communication and publishing around major universities. This will include survey and interview methods with students, analysis of publishing markets, discussion with online resource providers (including the operators of several major curated pirate archives), and analysis of university and library strategies around access to materials. It will combine the high level policy and political economic analysis of industry sectors with ground level analysis of what students, faculty, and other participants in the educational sector are actually doing to access the materials they need, and how these practices are evolving in function of expanding access to computers and broadband. In this respect it will intersect, but not overlap, IDRC funded projects on access to knowledge and copyright law. Indeed we expect the projects to be strongly complementary. As in the broader debate around copyright and enforcement, this ground-

level, empirical perspective is largely absent from policy debates at the state and institutional levels. Yet an understanding of these practices must be the basis for any durable, scalable solutions.

We believe the time to map this ecology of access is now, as the digital book phenomenon is shifting into overdrive with the emergence of ubiquitous internet connections and cheap readers. It is almost certain to become an important policy battle in the next years—ending the relative passivity of publishers in the enforcement wars. And it will be of paramount importance to students and universities, whose educational trajectories and decisions about the allocation of scarce resources will be most directly affected.

Mapping Media Access

Although the primary focus of this work will be access to educational materials, the survey and interview method will also permit a much better window on patterns of media consumption and digital media practices than is currently available. Comprehensive empirical data on media consumption in the era of digital convergence is remarkably scarce. Traditional consumption models have broken down as forms of digital access have proliferated. The speed of change also ensures that studies grow stale quickly. Most longitudinal studies address only very broad categories of activity, such as TV, radio, and the Internet (e.g., the Pew Biennial Surveys of Media Consumption in the US). There are a number of consumer surveys of piracy and counterfeiting (notably, BASCAP's 2009 comparative survey), but these tilt strongly toward broader categories of illicit goods and toward determining motives for participation in illicit economies (and by extension, toward determining effective messaging for industry groups). There are a number of surveys of particular media types (e.g., music) or distribution channels (p2p) but nothing that puts the larger picture together.

Today, a reasonably comprehensive study of access to audio-visual media would have to accommodate a very complex ecology, encompassing broadcast and direct-to-home services, the purchase and rental of disks, legal downloads, p2p and file-locker piracy, legal and illegal streaming services, large scale sneaker nets trading in portable hard drives, and a variety of other means of distribution. For texts, the vectors would include book purchases, library use, subscription-based access (e.g., through universities), open access content, and the similarly wide variety of pirate channels. We are aware of no studies that shed light on this digital media ecology or its differential structure in high, medium, and low-income populations. At present, there are no good comprehensive studies of what 'access' means for consumers faced with a spectrum of options—some physical, some digital; some legal, some not.

Among high-income consumers, the sheer breadth of access has become an obstacle to understanding how copyright, piracy, and enforcement function in practice. Moreover, there is growing reason to doubt that consumers themselves can reliably distinguish these channels, as they swim through a sea of media with various (and increasingly irrelevant) provenances. Recorded media, in particular, has become dissociated from the modes of ownership and use that once defined them, such as collecting, and its associations with intentional, managed, legal acquisition. Consumption is changing at a fundamental level, and we have few benchmarks to either describe it or guide policy.

Our survey and interview-based inquiry into how these issues affect access to educational materials will provide a basis for exploring broader questions of access to movies, music, and software—the core categories employed by the stakeholders in the copyright area—and to more general categories of fiction and non-fiction. It will also include questions about what consumers know about the law and enforcement regime that has grown up around them in the past decade—putting the significant industry and government investment in building a “culture of respect” for intellectual property to an all-too-infrequent test.

In the end, the study will be a powerful contribution to debates over IP policy, to conversations about the future of media markets, and to our ongoing efforts to empower developing countries to design policies and programs that enhance access to media. Because much of the enforcement debate, in particular, focuses on comparisons between developed and developing countries, we will be strongly advocating for a still-wider comparative approach to these issues with other funders, including at least one European market.

Empirical Research Methodology

The present proposal builds on the rich multidisciplinary approach to media practices and legal regimes developed in MPEE. We expect to reproduce the mix of legal, political-economic, survey-based, and ethnographic inquiry that guided that project, but with a narrower focus on the ecology of access to educational materials and a stronger research protocol. This inquiry will range from qualitative, small-group and guided interview based inquiry into individual practices of access, to interviews and policy analysis at the institutional level, such as universities, publishers, and (legal and illicit) access providers, to higher level analysis of policymaking—with the precise mix to be refined through discussion with partners and operationalized in function of the strengths of the different teams.

Pending further discussion with Google and potentially other funders, we also plan to conduct a major phone survey of individual practices around these issues. We expect this work to provide a much more robust, comparative, quantitative dimension to the work than was possible in the MPEE project, and plan to closely integrate the survey and qualitative analysis.

Much of the report will rely on an approach comparable to what we've done in the overview sections of MPEE. There are, however, some notable exceptions:

Although we explored relationships with the top P2P sites in the course of our MPEE work, we never succeeded in gaining sufficient access to data anchor a substantial account. In this case, our India and Brazil partners have access to a number of the administrators of these sites who can assist with such an account. The launch of our piracy report will also give us significant credibility in this space, hopefully in ways that can translate into access. At a minimum, this section is conceived as a history of these services and their communities. More ambitiously, we believe we can secure access to traffic logs that will shed light on user activities and demographics. Relatedly, we do not see a strong rationale for data mining of P2P sites in this project. This proved to be a relatively low-value direction of work for us in the MPEE project and we have not, at present, identified a set of questions that would be profitably

engaged with such methods. We would be very very interested if you think otherwise. The time to question these assumptions is now.

Report Structure

The primary output of the empirical research described above will be an 80,000-100,000 word report, similar to but more compact than the gigantic MPEE report, organized along the following lines:

1) Intro/overview sections:

- (a) An account of the global political economy of books and access.
 - i. Relevant regulation and policy fights
 - ii. The copyright battles of the 1960s-1970s.
 - iii. The prior history of the informal library: Institutional and library photocopying practices.
 - iv. Processes of technological diffusion (computers, broadband, readers) and market growth/income issues.
 - v. Publishers in the lobbying and enforcement landscape
- (b) History and analysis of top-level online pirate libraries – Aarg (social sciences), Gigapedia (general library), Karagarga (film), etc.
 - i. Analysis of usage patterns and user geography based on traffic data (very likely Aarg; possibly Karagarga).
 - ii. Analysis of the emergence of cloud-based curation and curatorial communities, massive personal libraries, and the transformation of the collection.
- (c) Examination of existing research on book markets and piracy.
- (d) Synthesis of the disciplinary studies (below)
- (e) Synthesis of the media mapping

2) Three Primary Country Studies: BR, IN, SA; Limited comparison in the US

- (a) The local history of university-publisher battles, fair use, copying, enforcement, legal reform.
 - i. the life of the university library; financing and future plans

- (b) Political economic and pricing analysis of the publishing business
 - i. analysis of shifting business models; politics/economics of shift to open models.
- (c) Histories of access to materials in 3 fields: law, engineering, and media/comm studies.
 - i. Primary investigation to include 2-3 focus groups with students, a student survey, and interviews with faculty, librarians, and other relevant actors.
 - ii. Secondary investigation to include a breakdown of curricula from one or more core classes, including determination of pricing, faculty strategies for selection of materials (open/not open), and actual practices of access among students.
- (d) Use of those interviews and surveys to conduct a broader mapping of media practices beyond books (software and audio-visual media).
- (e) Narrower analysis of the more organized informal libraries or interesting local ecologies that are emerging in schools and departments. This will be depend on initial surveying of the terrain.
 - i. Opportunistic mobilization of other research

3) Conclusion and recommendations.

---- ---- ----

4) Large-scale phone survey component