The Data Consortium for Media and Communications Policy: Toward a Federal Data Agenda for Communications Policymaking

Philip M. Napoli
Donald McGannon Communication Research Center
Fordham University
New York, NY
pnapoli@fordham.edu

&

Joe Karaganis
Social Science Research Council
New York, NY
karaganis@ssrc.org

With Contributions from
Angela Campbell, Institute for Public Representation, Georgetown University
Peter DiCola, Future of Music Coalition
Danilo Yanich, University of Delaware

This report is an outgrowth of an October, 2006 meeting sponsored by the Social Science Research Council and hosted by the American University School of Law

Meeting Participants
Angela Campbell, Institute for Public Representation, Georgetown University
Drew Clark, Center for Public Integrity
Mark Cooper, Consumer Federation of America/Donald McGannon Communication Research Center
Peter DiCola, Future of Music Coalition
Jack Fan, American University School of Law
Harold Feld, Media Access Project
Peter Jaszi, American University School of Law
Guinevere Jobson, American University School of Law
Joe Karaganis, Social Science Research Council
Rik Panganiban, Social Science Research Council
Gregory Rose, Independent Consultant
Andrew Schwartzman, Media Access Project
Jim Snider, New America Foundation
Gloria Tristani, Benton Foundation
Danilo Yanich, University of Delaware
About the Data Consortium

The Data Consortium for Media and Communications Policy is a Social Science Research Council-led initiative that seeks to improve access to datasets used in media and communications policymaking. The consortium includes educational and non-profit partners, from university-based schools, departments, and libraries, to public interest groups and professional associations.

High barriers to access to data are a recipe for poor public policy. Notably, they make independent policy analysis and evaluation of policy outcomes much more difficult. As this situation becomes the norm, media policymaking moves away from basic principles of public accountability.

The consortium is a vehicle for expressing the data-related concerns and collective bargaining power of scholarly and public-interest communities in this area. The consortium is organized around a few core objectives:

- Improve educational and other non-profit access to commercially-produced datasets, especially through cooperation with commercial data providers.
- Facilitate projects that address persistent ‘data gaps’ in our understanding of media and communications policy.
- Expand researcher engagement with datasets, in part by collecting and disseminating information about the uses and terms of access of different datasets.
- Advocate for the principle that public policy should be based on publicly-available data.

We believe that the interests of researchers, public-interest groups, policymakers, industry actors and commercial data providers intersect around these simple goals.

Membership in the Consortium is open to university-based schools, departments, libraries, and individual researchers; public-interest groups; professional associations; and other actors interested in facilitating a rich and more robust understanding of our changing public sphere.

Members include:

The Institute of Communications Research, University of Illinois
The Global Media Research Center, Southern Illinois University
The Annenberg School for Communication, University of Pennsylvania
The Department of Communication, University of Massachusetts at Amherst
The Center for Communication and Civic Engagement, University of Washington
The Department of Communication Studies, University of Michigan
The Annenberg Center for Communication, University of Southern California
The Department of Communication, University of Wisconsin – Milwaukee
The Donald McGannon Communication Research Center, Fordham University


Introduction

Policy debates and decision making at the Federal Communications Commission (FCC) increasingly turn on quantitative data analyses. In such an environment, questions of data quality and access to data are central to assessing the integrity of the policymaking process. These questions extend to the protocols and conventions surrounding how data are gathered (and by whom), the quantity and scope of available data, and the accessibility of data for policy analysis.

This paper describes significant deficiencies in all of these regards in communications policymaking. Our hope is to focus the attention of the policymaking and policy advocacy communities on data quality and access issues, and to begin laying out—for continued discussion, elaboration, and feedback—the contours of an improved federal data agenda for communications policymaking. This data agenda is intended to identify key substantive and procedural changes that should be undertaken at the federal level to improve the quality and accessibility of important categories of data used in communications policymaking and policy analysis. It focuses on the FCC because of its preeminence on many issues of media and communications policy, but extends into the wider ecology of institutions with responsibilities for understanding, regulating, and preserving a democratic, participatory public sphere.

This paper is divided into three sections:

- A description of contemporary shortcomings in the quality and scope of the data available to communications policymakers and policy analysts. This account is placed within the broader historical context of the steadily diminishing federal role in the gathering, aggregating, and disseminating of relevant data.
- An account of the closely-related issue of access to data, which plays an increasing role in ensuring accountability in policymaking and quality control in policy analysis. This section focuses on the impediments facing researchers and policymakers in accessing data for use in
policy analysis. It draws particular attention to the implications of the privatization of many key areas of data gathering, as well as to the political and economic obstacles to greater data access.

- The beginnings of a federal data agenda for communications policymaking, addressing questions of quality, scope, and access. The proposals put forth in this section are tentative and incomplete, as they are meant to serve as the baseline for a broader and deeper discussion of the components of a more robust federal data agenda in this area.

### Data Quality in Communications Policymaking

Federal data gathering efforts in the communications arena have been in decline since the 1980s (see Napoli & Seaton, in press), through a mix of passive and active measures that treated reporting requirements as a regulatory burden on industry. This decline occurred as the media and telecommunications fields entered a period of dramatic transformation, during which policymakers and the courts demanded new kinds of data in order to understand the fields in which they worked. These demands were not unique to the communications arena, but rather were part of a larger evolution of evidentiary standards for policymaking, which increasingly privileged the collection and analysis of quantitative data (see Napoli, 2005b).

Throughout this period, the FCC did very little to evaluate what kinds of data were needed to support its policymaking role, or how to obtain them. It came to rely heavily on the datasets developed by commercial providers for their media company clients and the investment community, and to neglect its own substantial data collection capabilities and stated responsibilities. Over time, this created problems in both the scope and quality of policy inputs—scope insofar as commercially collected data were expensive to access and not always structured in ways that illuminated public policy questions, and quality insofar as the FCC’s own data collection mechanisms were no longer consistently maintained or enforced. The discussion that follows highlights some key policymaking areas where problems of the quality, scope, and reliability have undermined the ability of the FCC to make informed policy decisions.
Diminished FCC Data Collection and the Limitations of Commercial Data for Public Policy

As the media and communications landscape changed in the past two decades, the FCC consistently diminished its capacity to collect the data that would illuminate these developments. Repeatedly, it scaled back and then neglected to enforce its own modest reporting requirements for licensees. It halted the gathering of financial statements from broadcasters (Webster, 1990), ceased gathering cable system subscriber data (Dunbar, 2003), and reduced requirements for performance data in connection with the license renewal process (Federal Communications Commission, 1981). As Napoli and Seaton (in press) document, large parts of this data gathering responsibility have been shifted to the commercial sector. This shift raises concerns not only regarding access to and scrutiny of the data (due to the typically high costs and restrictive access provisions associated with commercial databases), but also regarding the applicability of data to policymaking concerns. On many points, commercial data collection and public policy needs fail to align. Commercial data is structured around the financial, investment, and marketing needs of media corporations and investors — the data providers’ primary clients. The FCC, in contrast, answers to a more complex concept of the public interest, which balances economic efficiency with concerns for equity, diversity, and constitutional rights. Data on economically marginalized populations is a frequent break point between the two. For instance, examination of the most widely-used, commercially-available database of broadcast and newspaper market, ownership, and financial data has revealed significant omissions in minority-targeted and foreign-language media outlets (Lloyd, et al., 2005). Intentionally or not, this lack of rigor tracks perceptions about the low value of such media markets.

The dependence of the FCC on commercial providers sets up a frustrating dynamic for researchers. Journalist John Dunbar has documented how efforts to obtain and analyze comprehensive data on the ownership of the media outlets from the FCC—a topic that that falls directly under its regulatory purview—are redirected to commercial data providers, who impose their own, often onerous, terms of access (Dunbar, 2003). Even when access is granted, Dunbar and others have found that the
commercial orientation of these databases makes them ill-suited to many important public policy questions, such as the number and location of duopolies, cross-media ownership patterns, or clustering of radio station ownership. Although such information could easily be collected by providers in the course of other data gathering, it serves no compelling commercial purpose, and the FCC exercises no say in what data are collected or how they are structured.

One of the FCC’s key areas of responsibility is the regulation of the ownership of a wide array of media and telecommunications entities. As all stakeholders would agree, changes in existing ownership policies need to be predicated on a thorough and detailed understanding of the current state of affairs. It has become clear, however, that neither the FCC nor the commercial providers possess adequate data to support informed policymaking in this area. As Hesmond and Pratt have observed, commercial data sources “are functional for market making; but not for an understanding that will provide an evidence base for policy making or intellectual inquiry” (Hesmondhalgh & Pratt, 2005).

**Minority Ownership**

Minority ownership is an area where the research needs of public policy have recurrently diverged from those of commercial data collection. This is particularly troubling given that diversity has been, and continues to be, one of the FCC’s core policy principles (see Napoli, 2001): the Commission is obligated in its decision making to preserve and promote diversity in the media. Because data on minority ownership is of high policy value but little commercial value, federal data gathering has traditionally been the sole source of information. Despite dramatic consolidation of the industry, changes of ownership, and the rise of new minority media, this monitoring role has been significantly reduced in recent years. The National Telecommunications and Information Administration initiated an effort to assess minority ownership in the 1990s, but issued its last report on the subject in 2000 (NTIA, 2000). The FCC continues to gather data in this area, but recent scrutiny of its data has revealed serious problems. Turner and Cooper (2006), in their effort to assess the current state of minority and female television station ownership in the United States, find that the data that the FCC requires all full-power
commercial broadcast stations to submit annually (Form 323) are reported only partially and—even then—erratically. Specifically, summary reports of the data gathered by the Commission only list each minority or female-owned station’s Form 323 response. No aggregation of the data obtained from the stations is conducted for these reports; nor are the responses provided by stations not owned by women or minorities included in the summary reports. Without access to the full range of Form 323 responses, it is difficult for researchers to conduct rigorous analyses, as information from the entire population of stations is not available for analysis. Perhaps more significantly, Turner and Cooper (2006) find substantial omissions in the data, with some of the nation’s largest minority radio and television ownership groups missing from the summary reports; and with some station owners disappearing from the summary reports for years at a time, despite continuity of ownership during the time periods examined.

According to Turner and Cooper (2006), these inaccuracies are most likely an outgrowth of problems with FCC data handling—in this case, the automated process via which data are harvested from the electronic filings. This process appears incapable of accurately capturing the various levels and categories of ownership and the consequent complexity of the information filed by the individual stations (some of which file more than 20 Form 323s due to the complicated ownership of many individual stations). Regardless of the reason, these shortcomings need to be addressed if the FCC wants robust policy analysis and effective policy in this area.

**Employment**

A wide array of FCC regulatory activity depends on the Commission’s ability to accurately gauge employment patterns in the various industry sectors under its regulatory authority. This includes regulations dealing with gender and ethnic diversity hiring practices in the media industries. As in the minority ownership area, this is a field almost entirely dependent on federal data collection (some professional associations also gather data in this area, but seldom are such data sufficiently comprehensive for policy analysis). There are, unfortunately, a number of troubling problems in this area as well.
A first order problem is that the generality of federal data collection on employment and wages conducted by the Bureau of Labor Statistics makes it difficult to answer questions about communications and media industries in particular. This has been a problem in recent work on the impact of regulatory changes on the commercial radio industry, where the data are typically aggregated and reported at the market level, making outlet- or employer-level analyses difficult (see DiCola, 2007). Respondent sample sizes for individual communications industry sectors such as radio, moreover, are often too small to facilitate useful analyses; nor are the data always sufficiently broken down along industry lines.

Federal data gathering on minority employment exhibits similar inadequacies. As the FCC’s Equal Employment Opportunity rules have been scaled back over the years, so too has the Commission’s practice of gathering minority employment data through its licensing procedures (Federal Communications Commission, 2001). Data on the actual race/gender breakdown of employees is no longer required. Instead, the Commission now only asks licensees to provide information about the number of vacancies open and filled during the license period, with some additional information about recruitment and outreach sources utilized. As in other cases, diminished data collection is accompanied by erratic reporting: one recent analysis (Napoli, 2005a), encountered missing data issues in the filings of 204 of 350 audited licensee responses examined for the 2003 audit, and 162 of the 350 audited licensee responses examined for the 2004 audit. In practice, the FCC does not assess compliance with even these stripped-down EEO rules. Along similar lines, in 2004 the FCC proposed to eliminate its 22-year old requirement that common carriers provide annual reports on minority and female employment (Honig, 2003).
Content

Although the FCC has dramatically scaled back the extent to which it directly regulates media content, there has been an increase in attention in recent years to the indirect impact of various regulatory measures on content (see Napoli, 2005b). The FCC has itself conducted a number of studies in this area on, for example, the relationship between ownership, market conditions and the provision of local news and public affairs programming (Spavins, et al., 2002); the relationship between ownership structure and the political orientation of news content (Pritchard, 2002); and the diversity of programming in both radio and television (Einstein, 2002; Williams, Brown, & Alexander, 2002).

Given the growing role of this kind of analysis, the inadequacy of existing federal efforts to gather relevant content data is disconcerting. Any systematic efforts to assess television or radio programming—particularly at the local market level at which most contemporary regulations are directed—faces an immediate problem of scarce and fragmented data sources. The FCC gathers minimal content-related data. Broadcast licensees are required to maintain lists of programming addressing the needs and interests of their communities, but this information is not even required to be submitted to the Commission. Rather, it need only be made available on-site for the public to access. An effort by the Commission in 2000 to expand and standardize this reporting requirement met with substantial industry resistance, and this rulemaking proceeding has since languished (see Federal Communications Commission, 2000). The only detailed content data that licensees are required to submit to the Commission focuses on the three-hour-per-week educational children’s programming requirement. These quarterly Children’s Television Programming Reports are submitted to the Commission (which maintains a publicly accessible on-line database), as well as maintained in the licensee’s public inspection file (see Federal Communications Commission, 2000).

Of course, such logs provide an indirect indicator of the actual media content, at best, and the records themselves usually suffer the flaws associated with unmonitored self-reporting. Researchers seeking direct access to media content (particularly on an outlet-by-outlet basis) have relatively few
archival options. The well-known Vanderbilt Television News Archive provides a reasonably comprehensive historical database of the Big Three broadcast television networks’ nightly newscasts (see Althaus, Edy, & Phalen, 2002; Breeding, 2003), but little else. When researchers try to assess either non-news programming at the national level or any form of programming broadcast at the local level, systematic archival resources are virtually nonexistent (Murphy, 1997).

In 1997, the Library of Congress released the results of a detailed study on the current state of American television and video preservation (Murphy, 1997). This study was the outgrowth of numerous hearings involving a wide range of stakeholders (see Library of Congress, 1996a, 1996b, 1996c). The report paints a very grim picture of television and video preservation in the U.S. as a whole, but reserved its bleakest assessment for the state of preservation for local television newscasts—content that informs a wide range of analyses relied upon by policymakers (see Federal Communications Commission, 2006). A number of recent FCC analyses related to its media ownership rules have focused on local news (e.g., Pritchard, 2002; Spavins, et al., 2002), which is often treated as an indicator of responsiveness to local communities. In addition, challenges of broadcast license renewals have, in many instances, revolved around the analysis of local news programming, dating back to many of the earliest license challenges (see, e.g., Mills, 2004). Yet there never has been any meaningful federal effort to archive the content that grounds analytical inputs in these policy areas.

According to the Library of Congress report, “The most devastating losses have already occurred among news film and videotape files of local television stations across the United States” (Murphy, 1997, p. 9). It is estimated that less than 10 percent of local news programming survives. “Even today,” according to the report, “local news tapes are rarely kept more than a week before they are recycled” (Murphy, 1997, p. 9). And yet, according to the report, “Every group that has studied the selection of television for preservation has concluded that all news programs should be retained and preserved as aired” (Murphy, 1997, p. 9).

Of course, as the report notes, the preservation of entertainment programming has improved over the years, due in large part to the tremendous commercial incentives associated with archiving
programming for later monetization through ancillary outlets (cable, syndication, on-demand, DVD, online, etc.). News, in contrast, has a much more limited shelf life and, consequently, more limited long-term revenue prospects. Commercial incentives to systematically preserve news content are largely absent and, as the report notes, “there is no FCC requirement that local newscasts be saved” (Murphy, 1997, p. 56). Scattered local news programming preservation efforts do exist at the state and local level—including some 40 archives housed in places such as state historical societies and universities. These efforts often operate under conditions that do not ensure comprehensive data collection, long-term preservation, or efficient access (Murphy, 1997).

What is particularly troubling about this state of affairs is the extent to which the Federal Communications Commission continues to ask questions for which the data necessary for developing meaningful answers are virtually non-existent. For instance, the FCC’s recently announced slate of studies to be conducted in connection with the current review of media ownership regulations (see Federal Communications Commission, 2006) include analyses that would seem to rely upon data that would be difficult, if not impossible, to obtain with any comprehensiveness, rigor, or assurance of accuracy. One of the Commission’s studies promises to “analyze the effect of ownership structure and robustness . . . on various measures of the quantity and the quality of different types of TV programming, including local news and public affairs, minority programming, children’s programming, family programming, religious programming, and violent and indecent content” (Federal Communications Commission, 2006). Completing a study of such breadth and depth within the allotted four-month window would be miraculous even if the content data were readily available. Given the rudimentary and fragmented state of television archiving, an authoritative study on this subject is basically impossible.

Because of the lack of archival resources on programming, many policy-oriented analyses of media content rely primarily on commercially-produced program schedule or playlist databases. These often make content-based inferences from the databases without ever assessing the underlying content directly (e.g., DiCola, 2006; Hamilton, 2000; Spavins, et al., Williams, Brown, & Alexander, 2002; Yan & Napoli, 2006). While such metadata sources can dramatically reduce the time and labor associated
with investigating certain types of research questions, they provide only limited opportunities for
verifying or challenging the tagging of data on key policy questions, such as distinctions between news
and entertainment, or children’s and adult programming. Such metadata is also generally inadequate for
investigating questions about content quality and substance that increasingly are being asked in the policy
arena (see above).
Access to Data

The previous section focused primarily on the quality and scope of the data currently being gathered to support communications policymaking and policy analysis. This section takes up the closely-related issue of effective access to that data. It is becoming increasingly clear that access to policy-relevant data is too restrictive to serve the needs of the communications policymaking process. Part of this problem relates to the privatization of data collection, in which policy researchers outside the employment of media companies have virtually no market power. Part of the problem relates to the new and largely unrealized opportunities to use World Wide Web to collect and disseminate knowledge.

The Web has greatly impacted the ability of regulatory agencies such as the FCC to gather and disseminate the raw data that factor into their policy analyses. The FCC has not been idle in this area, and currently maintains a number of freely accessible online databases: the above-mentioned collection of Form 398 reports for educational children’s television programming; statistical data on indecency complaints and indecency actions; and data on telephony charges, usage, and service provider performance, among others. For reasons that are not clear, the preponderance of this data relates to telecommunications industries and service providers, rather than to media industries and media outlets.

Despite these positive steps, there are a number of factors that impede access to policy-relevant data, from opaque pricing structures and restrictive licenses for commercial data, to legal barriers to access, to the basic adversarial nature of contemporary communications policymaking. The sections below discuss bottlenecks at the level of the FCC’s discretionary authority and in relation to commercial datasets.

Discretionary Authority

The FCC collects Form 477 (“Local Telephone Competition and Broadband Reporting”) from broadband providers in an effort to gauge the extent of broadband availability across the United States. This information is gathered at the zip code level, which facilitates a wide range of potentially valuable
analyses of broadband deployment that could guide policymaking. Under the terms of the Freedom of Information Act, the non-partisan public interest organization, the Center for Public Integrity, sought access to these data (Center for Public Integrity, 2006a), but was denied by the Commission on the grounds that a) the data fall within certain Freedom of Information Act (FOIA) exemptions (due to the inclusion of “commercially sensitive” data); and b) the Center for Public Integrity failed to present a compelling public interest reason for disclosure of the information (Federal Communications Commission, 2006b). This latter argument is particularly striking in light of the fact that the Center for Public Integrity sought access to the data in part to resolve a discrepancy between FCC and U.S. Government Accountability Office analyses of the data. According to the FCC’s analysis, the data showed that the median number of broadband providers in a zip code was eight, whereas the GAO’s analysis indicated that the median number was only two (Center for Public Integrity, 2006a). Equally troubling was the Commission’s decision to deny access to the entirety of the data, as opposed to only those portions of the data involving commercially sensitive information (see Center for Public Integrity, 2006b). In this instance and in several others, the FCC has chosen to exercise its discretionary authority to block access to data.

**Third Party Research and Data**

The privatization of data collection has placed control over access to data firmly in the hands of commercial data providers. Researchers interested in audience ratings or industry structure have to negotiate with providers for licenses to use portions of their datasets. Commercial providers have extensive discretionary authority over who they license to and on what terms. There are no rules, norms, or ‘markets’ that shape pricing. These lines of authority become less clear in the case of data utilized in studies submitted to FCC proceedings, where access to data is an obvious condition of accountability in policymaking. This intersection between public policy and private inputs is poorly defined and increasingly contested. In 2002, the FCC took a partial step toward transparency by making the underlying data for the studies it commissioned in the context of its 2002-2003 media ownership
proceedings (mostly) available on-line for download and (re)analysis (see Federal Communications Commission, 2002a; see also Napoli & Seaton, in press). The commercial, proprietary status of some of the data used in these studies created complications, however, and some of the data were made available only on an isolated terminal physically located at the FCC. When dealing with large datasets, such compromises make it virtually impossible to do more than spot check the data, making them of little or no value to researchers.

Because the 2002-2003 studies were commissioned by the FCC, the public status of the underlying data was relatively clear. Most of the research utilized by the FCC, however, is submitted by third parties—researchers, advocacy groups, media companies, professional associations, and so on—in the course of comment periods during FCC proceedings. Although the submission process for comments makes these studies ‘public’ and accessible online, the data underlying submitted studies has not been subject to disclosure requirements. This creates obvious impediments to scrutiny of the studies, both by other stakeholders in the proceedings and by the FCC itself. In the case of independent research that relies heavily upon commercial data sources, the researchers themselves are usually contractually prohibited from disseminating or disclosing the data.

This double-bind is coming under increasing pressure from stakeholders and advocates involved in FCC policy processes. In 2006, EchoStar Satellite, a Direct Broadcast Satellite provider, argued in court that the Administrative Procedures Act required that any data relied upon by the FCC (or any other federal agency) in its decision-making must be made available in the public record (EchoStar Satellite v. Federal Communications Commission, 2006). In this case, the company was seeking access to broadcast signal strength data that the FCC used in its determination of broadcast signal transmission rights under the Satellite Home Viewer Act of 1998. This data was not submitted in raw form, but rather was analyzed in an engineering report submitted to the FCC by the National Association of Broadcasters (NAB) and the Association for Maximum Service Television (AMTS). Unfortunately, the U.S. Court of Appeals for the D.C. Circuit did not address EchoStar’s argument regarding its right to access the data, on the grounds that EchoStar did not request the data until after the Commission had issued its final decision.
(EchoStar Satellite v. Federal Communications Commission, 2006). The FCC, on the other hand, argued that EchoStar was not entitled access to the data because the Commission “had not relied upon them when it issued its final rule. Rather, the Commission based its analysis upon the description, methodology, and results of the study contained in the public comments” (EchoStar Satellite v. Federal Communications Commission, 2006, p. 20). The FCC’s perspective on this issue begs the question of whether relying upon a study utilizing a particular data set is different from relying upon the data analyzed within that study. Superficially, this would appear to be shaky grounds on which to deny access to data—if indeed access to policy data is required by statute.

A related issue arose recently in connection with the FCC’s ongoing media ownership proceeding. The Smaller Market Broadcasters Coalition (2006) filed a comment arguing that it was entitled to see data underlying a study cited in other comments filed by Consumers Union, the Consumer Federation of America, Free Press, and the Office of Communication of the United Church of Christ. The study in question was an academic research paper presented at a recent conference by two University of Michigan scholars (Yan & Park, 2006). The Small Market Broadcasters Coalition (2006) took issue with some of the findings in the paper involving the relationship between duopoly ownership of television outlets and the provision of local news and public affairs programming. They argued that the study “should not be given any consideration until the underlying data are placed on the record and the public has had an opportunity to evaluate those data and comment on the Study” (p. 2).

EchoStar and the Small Market Broadcaster Coalition identify the same need for access to underlying data, but differ significantly in the boundaries they propose. Echostar requested access to the data underlying a study conducted and submitted by the NAB and its partners. The Small Market Broadcaster Coalition requested access to the data underlying an independent study cited by Consumers Union and its partners—a study they neither conducted nor funded, and which was not submitted to the FCC. Consumers Union and partners do not have access to Yan and Park’s data. Moreover, Yan and Park would likely be in violation of their commercial data license if they it chose to release it. The Small Market Broadcaster Coalition request reaches beyond the formal comment process and implicates
the practice of citation and independent research in general. The burdens placed on parties filing comments in this scenario would be completely insurmountable (Consumers Union, Consumer Federation of America, & Free Press, 2006) and would ultimately chill the usage of research in policy advocacy.

Formal inclusion in a policymaking process, however, provides a potentially more solid point of application for rules on data disclosure. Such rules would have to encompass not just FCC-conducted or commissioned studies, but also outside research submitted in proceedings. In all cases, meaningful access to data would have to include access to any commercial datasets used in the study. Effective data disclosure would likely require two types of action in this context: (1) rulemaking by the FCC to require the disclosure of data submitted in formal policy proceedings, and (2) an accompanying shift in the licensing terms used by data providers to permit disclosure in public policy contexts.

**Content**

The scarcity of data on media content is compounded by problems of accessibility. Efforts by the FCC to require that stations make even their rudimentary, licensee-required data on programming available online (in addition to on-site) have encountered substantial resistance from industry and from within the Commission (see Federal Communications Commission, 2000). Rulemaking on this issue remains unresolved despite the fact that the Commission itself acknowledges that “members of the public have encountered difficulties accessing information under existing procedures” (Federal Communications Commission, 2000, p. 4).

Researchers trying to work directly with media content also confront other problems that go well beyond the FCC’s jurisdiction. Current copyright law presents major obstacles. A recent case study by Ubois (2006) illustrates the difficulties associated with accessing electronic media (i.e., the news footage and television episodes) associated with the well-known Dan Quayle/Murphy Brown controversy that erupted in 1992 following Vice President Quayle’s public criticisms of the television program and its lead character. As Ubois (2006) discovered:
“Reconstruction of the … primary source materials proved effectively impossible, despite extensive and prolonged efforts. The speech by Dan Quayle that initiated the controversy was inaccessible for reasons of copyright, and the owner of the Murphy Brown episodes refused to provide them for educational use. Other news and entertainment footage was difficult to find, expensive, or unavailable” (Ubois, 2006, p. 8).

As in the area of data collection, private control of video archives places enormous discretionary authority in the hands of the private interests—even in cases where research on these materials would clearly fall under ‘fair use’ exceptions to copyright law. Had Ubois possessed a copy of the materials—e.g., from his own recording of the broadcasts—he would been able to make his derivative analyses and assertions of fair use. In the latter case, the burden of challenging a research-related or educational use would fall upon the copyright holder. Without access, the negotiation of those boundaries never takes place. It is worth noting, too, that Ubois was lucky in knowing who held the copyrights on the desired materials. Because there is no registry of copyrights, and because copyrights are both transferable and often multiple with respect to audio/visual works, ownership often is impossible to ascertain (see, e.g., Lessig, 2004).

In a corporate environment marked by the monetization of media archives, media outlets have frequently been uncooperative with researchers in providing access to relevant content data—even in contexts which have nothing to do with copyright concerns (see Murphy, 1997; Ubois, 2006). Issues of costs and labor associated with accessing, reproducing and transferring relevant content arise, as does more difficult-to-document reluctance among media outlets to aid researchers whose work might eventually used against them in adversarial policy proceedings (see McGehee, 2006). The current structure of access provides industry with some veto power over academic research agendas.

This is relevant especially in the context of efforts to independently monitor and analyze media content, which has frequently been used to advocate policies or regulatory actions that run counter to the interests of the regulated industries (e.g., Mills, 2004). Such monitoring efforts encompass a wide range of content areas, including news and public affairs, violent programming, children’s programming, and indecent programming (e.g., Children Now, 2003; Fowler, Goldstein, Hale, & Kaplan, in press; Parents Television Council, 2003).
One of the ironies of this situation is that, while the print media is relatively well-archived and sufficiently accessible to facilitate systematic research (via widely used and accessible, if somewhat expensive, data sources such as LEXIS/NEXIS), the electronic media is nowhere near as accessible. As legal scholar Lawrence Lessig (2004) has noted, “Why is it that the part of our culture that is recorded in the newspapers remains perpetually accessible, while the part that is recorded on videotape is not? How is it that we have created a world where researchers trying to understand the effect of media on nineteenth-century America will have an easier time than researchers trying to understand the effect of media on twentieth-century America” (p. 111).
Toward a Federal Data Agenda for Communications Policymaking

As this paper has illustrated, many of the basic questions that policymakers, the courts, and stakeholders pose in regard to communications policy cannot be answered due to the poor quality, scope, and accessibility of policy-relevant data. The result is a frustrating Catch-22 in which the studies that are conducted are subjected to withering methodological critiques—and thus frequently discredited—while little effort is made either to produce better data or better access to existing datasets (see McGehee, 2006). This situation undermines the extent to which research can effectively inform public policymaking.

Some of these problems are challenging, and would involve legislative efforts or a significant rethinking of the FCC’s research role. Others are very modest and simple to address, provided the FCC and other relevant actors choose to act. This section offers the beginnings of a concrete agenda for change. We recognize that it is neither definitive nor fully inclusive of the areas of communications policymaking where meaningful improvements in federal data gathering policies are needed and possible. We do hope that it can be refined, expanded, and developed with greater specificity in the coming months via comments and feedback from interested stakeholders.

1. Internal Data Collection:

Management and Compliance with Existing Reporting Requirements

Communications policy analysis would be significantly improved if the FCC did a better job of managing its existing reporting requirements. The collection of such information would seem to be a core part of the FCC’s responsibility to monitor the state of the industries under its watch. Shortcomings in FCC-collected data are a self-inflicted blow to its policymaking capabilities, and cover a wide range of policymaking areas. Specifically, the FCC should:

- Devote more resources to the efficient and reliable gathering and processing of existing data.

Researchers have identified serious problems with the reporting and/or analysis of:
- Form 323 (dealing with ownership)
- Form 398 (dealing with children’s educational programming)
- Form 477 (dealing with broadband deployment)

- Better enforce compliance with reporting requirements by licensees. Improved compliance goes hand in hand with better data processing. More severe sanctions should be considered for organizations that fail to provide complete or accurate information in these regards.

**Expanded Data Collection**

In several areas, the full exercise of the FCC’s regulatory responsibilities would seem to require modest additions to its data collection practices. These could dramatically increase the quality of research inputs into policymaking. Among these:

- High priority should be given to the gathering of comprehensive information on the financial state of the outlets under its jurisdiction. Revenue data, in particular, must be better captured by the Commission, given the policy importance it accords the analysis of competition within individual communications markets. The fact that the Commission does not have comprehensive information on the financial status of individual media outlets and media markets makes it virtually impossible for the agency to apply its own preferred analytical frameworks to policy decisions. Reliance on commercial sources for such data is inadequate because the methods of gathering and reporting of such data are susceptible to marketplace demands (and to shifts in these demands). The FCC used to gather). Revisiting earlier FCC policy requiring financial statements from broadcast licensees (Webster, 1990) would be a good first step. To adequately account for the industries under FCC jurisdiction, such requirements would have to be expanded beyond broadcast licensees.

The regulated industries have often objected that such reporting reveals commercially sensitive information. Such claims need—first and foremost—to be assessed against the availability of
comparable data from commercial sources. There is little logic, for instance, to the argument that station revenue data is commercially sensitive, and therefore protected from federal data gathering efforts, given that such data are readily available (for a substantial price) via a number of commercial data sources.

2. Content Archiving

The archiving of media content is increasingly important to communications policy research and is an area in need of dramatic improvements. The FCC can play a constructive role here, although a more comprehensive solution to archiving and access will almost certainly require action at other levels of government, such as changes to copyright law, or expansion of the mandate of the Library of Congress. As the FCC looks ahead toward the challenges of communications policy in the next decades, it would be well served by beginning such dialogues with other agencies. In the meantime:

Programming and Performance Archiving

It is squarely within the FCC’s authority to require broadcast licensees to provide a tangible, accessible, and reliable record of station programming and performance. A return to the era of detailed program logs would be one possible element of such a shift. Such materials should be available to the public online and, perhaps preferably, be submitted to the FCC for verification and aggregation into a publicly available master data set. The FCC began to move (somewhat tentatively) in this direction in 2000 in its proceeding on the reporting requirements of broadcast licensees (see Federal Communications Commission, 2000), yet this proceeding has languished for six years. The Commission needs to revisit this question of the mechanisms by which such data should be made available to the public.

- An appropriate longer-term goal would be the establishment of a centralized content archive in which all FCC licensees are required to deposit on an annual basis, at minimum, some
representative sample of their content output, in order to facilitate the outlet-level and market-level analyses that are becoming increasingly important in policymaking. The scope of such an enterprise would require further discussion. A great deal of cross-market and longitudinal analyses of programming practices could be accomplished with even a modest randomly-constructed sample of programming. A more ambitious solution would target the complete content archive of all programming—a possibility that is already within technical reach at relatively modest cost.

The regulatory authority to back a more comprehensive archiving agenda is already present, as is the appropriate federal infrastructure for handling a larger archiving enterprise. The American Television and Radio Archives Act (a section of the 1976 Copyright Act) established the American Television and Radio Archives within the Library of Congress, for the purpose of preserving “a permanent record of the television and radio programs which are the heritage of the people of the United States and to provide access to such programs to historians and scholars” (American Television and Radio Archives Act, 1976). While this archive is relatively strong in the areas of prime-time network television programs and PBS content (see Murphy, 1997), it is not a robust archive for the content output of individual radio and television broadcast licensees across the United States; though it certainly has the potential to take on this role. A coordinated effort by the Federal Communications Commission and the Library of Congress could realize the archive’s potential and make an enormous contribution to American culture. At a more technical level, it would lead to a dramatic improvement in the FCC’s ability to answer policy questions about media content.

3. Third Party Data

The FCC needs more robust policies regarding quality assessment and access to data with respect to studies submitted by third parties. It also needs stronger policies regarding the commercial data sources
utilized directly by the Commission. The Data Quality Act (2001) requires government agencies to develop procedures and standards for addressing issues related to the quality of data used in agency decision-making. The FCC responded to the Act by issuing a set of Information Quality Guidelines in 2002, which stated its commitment “to ensuring that all data it disseminates reflect a level of quality commensurate with the nature of the information. Further, the Commission seeks to disseminate all its data as broadly and promptly as possible. This commitment applies to all data and information disseminated by the Commission” (p. 2).

A key feature of the Commission’s interpretation of the Act is that it applies only to “reports prepared for Congress or required by legislation.” This language aligns the Data Quality Act with the Freedom of Information Act (Data Access Act, 1998), which provides the public the right to access data produced with government funding (though even here there are significant caveats; see Hornstein, 2003; Napoli & Seaton, in press). Among the many things that the FCC says the Act does not apply to are:

- public filings, subpoenas, or adjudicative processes; non-scientific/non-statistical general, procedural or organizational information; information that is not initiated or sponsored by the Commission; information that expresses personal opinions rather than formal agency views; information for the primary use of federal employees, contractors, or grantees; responses to requests made under the Freedom of Information Act, the Privacy Act, the Federal Advisory Committee Act, or similar laws; agency correspondence; archival records; trade secrets, intellectual property, confidential data or information; and non-routine or emergency public safety information (Federal Communications Commission, 2002b, p. 8).

This list effectively excludes much of the research and data on which the FCC relies in its policymaking—especially third-party, publicly-filed research. The FCC interpretation keeps the Data Quality Act at a distance from much of the actual process of policymaking—in fact it is not clear that this interpretation would apply even to its recent rounds of commissioned research on media ownership (despite the precedent it set in providing access to the data underlying the first round of studies in 2002). The studies have been commissioned as contracts, not as grants—a technicality that circumvents the strict letter of the Data Quality Act, which was conceived around the model of laboratory science.

The FCC’s position has been challenged by the Center for Regulatory Effectiveness (2006), the primary watchdog organization associated with the Data Quality Act. The CRE recently filed comments
in the FCC’s media ownership proceeding arguing that “All of the data used or relied on by the Commission, whether developed internally, by agency contractors, or by independent third-parties, will need to adhere to applicable Data Quality standards” (p. 2). According to the CRE (2006), “The FCC will need to apply OMB and Commission Data Quality standards to all substantive data submitted by commenters. The Commission is only able to use and rely on third-party information that fully complies with Data Quality standards” (p. 4). Moreover, petitions claiming that information disseminated by the Commission has failed to meet applicable data quality standards “may be filed against FCC-developed information or against FCC information that is based on third-party materials” (Center for Regulatory Effectiveness, 2006, p. 4).

The inclusion of public comments under the Data Quality Act would lead to a more realistic discussion of quality assessment and access to data in FCC policymaking—one that would also likely require a more robust internal process of verification and review of all research used in policymaking decisions (see Gasser, 2003). Because so much of the policy analysis and review process at the FCC is externalized through the comment process, a commitment to data quality would also require a much more substantive discussion about access to the data underlying third-party submissions. Because so much of this data, in turn, is governed by commercial licenses that restrict or forbid secondary dissemination, third-party submitters are often legally enjoined from disclosing their data. This creates a basic obstacle to review, both for external commenters and the FCC itself, and poses a serious challenge to quality and accountability of policymaking at the FCC. If the commission is to continue to rely on external data collection and an adversarial comment process, it should require that:

- Any study submitted by a commenting party to a Commission proceeding be accompanied by the associated underlying data.
- Such data must be made available for reanalysis by other interested parties. This would probably require an accompanying process of rethinking commercial licensing practices, to permit less restrictive terms of disclosure for public policy purposes.
4. Advisory Committee on Data Quality, Integrity, and Access

Because the issue of data and its uses in communications policymaking are complex and evolving, we recommend the creation of a Federal Advisory Committee on Data Quality and Integrity. Such a committee could be comprised of a mix of FCC personnel, industry representatives, academic researchers, and members of the public interest/advocacy community. This committee would be charged with establishing specific baseline standards for the Commission’s data needs, as well as with assessing the quality and integrity of the various data sets relied upon not only by the Commission, but also by the various external stakeholders that submit research to the Commission in individual proceedings. Such a committee could also engage in regular systematic inventories and assessments of the various forms that the FCC solicits from the organizations under its regulatory authority, as well as continue to improve the accessibility of relevant data. The Committee would then make recommendations to the Commission regarding data gathering needs, processes, and access policies. This work would also recuperate the stalled 2000 proceeding on reporting requirements, and create a stronger basis for the FCC to think about the future of the communications arena and its role in ensuring a vibrant and participatory public sphere.

Conclusion

As was noted at the outset, this document is very much a work in progress. More input is sought in regards to additional areas where the federal role in data gathering is in need of improvement, and in regards to specific recommendations for how this role could be enhanced. At present, this document represents a first pass through this very complex and wide-ranging issue. Comments, feedback, additions, and critiques are all enthusiastically solicited as this effort to build a substantive federal data agenda for communications policymaking moves forward.
References

Federal Communications Commission (2002b). In the matter of implementation of guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information pursuant to


Office of Management and Budget (2002). Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by federal agencies. Federal Register 67(36), 8452-8460.


Smaller Market Broadcasters Coalition (2006). In the matter of 2006 quadrennial regulatory review: Coalition request for underlying data, MB Docket No. 06-121.


