The mythos surrounding public access to information is powerful: if only we can provide the masses with enough quality information, the story goes, we can improve democracy, the economy, and the human condition. More information, provided across all the typical barriers that divide people – class, gender, race, nationality, etc. – will equalize educational opportunities, and thereby increase equality in general. It will allow the lower classes to pull themselves up by their bootstraps; it will open windows into new realms of knowledge for those interested in exploring them; it will promote the establishment of an informed populace, capable of full participation in discourse and civil life in a democratic society. It’s an inspiring dream, and one in whose service countless institutions and initiatives have been constructed – despite the general lack of consensus about the validity of its underlying causal claims (see, e.g., Levy, 2000; Raber, 1997).

This dissertation will not directly address the causal validity of the information-utopian mythos either – but it will attempt to cast more empirical light on the particular sorts of institutions that have been built in its image. In particular, this dissertation will examine two sets of phenomena from two very different eras: first, the nineteenth-century American public library movement, and second, the current push toward large-scale (even universal), public-access book digitization, by organizations like Google and the Internet Archive. These two phenomena, I contend, have employed closely parallel strategies in pursuit of the information-utopia: that is, both have striven to provide self-directed access to as much quality information as possible, to as many people as possible, free of charge. Yet, on the level of the individual case, the ways in which this mission has been interpreted, and the structures and policies that those interpretations have produced, vary considerably. This dissertation will be the first study to systematically compare these two phenomena, and will specifically focus on these processes of design negotiation within two cases on each side: Google Books and the Open Content Alliance on the one hand, and the Boston Public Library (BPL) and the Carnegie Library program on the other. For each case, the dissertation will examine (1) how (and by whom) the task of democratizing access to information was interpreted and defined, and (2) how those interpretations shaped the designs of the structures and services offered up for public use. This proposal will offer a brief description of the rationale for this comparative study, the specific research questions posed, the theoretical framework guiding those questions, and the methodologies planned to address them.

**Context for and Significance of the Study**
The term “large-scale digitization initiative” was coined in 2006 (Lynch), to describe two then-recently-launched efforts to digitize all the books in the world: Google Books and the Open Content Alliance (OCA). Google Books had been announced in 2004, as a two-pronged effort: on one hand, Google would work with publishers to provide full-text online searching and previews of their books; on the other, they would partner with major research libraries
to work toward the goal of scanning every book ever published (Google, 2004; Taycher, 2010). Books in the public domain would be available to read online or download in their entirety, free of charge; books in copyright or of uncertain rights status would be full-text searchable, but not readable in full online (Google, 2011a, 2011b). In essence, this is still what Google is doing now.

The Open Content Alliance was announced a year after Google Books, as a strategic counterpoint to what many saw as Google’s closed, centralized, and perhaps corporatist approach. The OCA sought to take a more transparent, collaborative, consortial approach to book scanning, led by the Internet Archive (Leetaru, 2008). The OCA has also expressed dreams of digitizing all the books in the world (e.g., Kahle, 2007), but has taken a more deliberate, legally conservative, image-quality-focused approach to achieving that aim (Kahle, 2005, 2007). In recent years, the Open Library (http://openlibrary.org) has become the public face of the OCA, where actual access to books is offered, and the term “OCA” as such has fallen somewhat into disuse; nonetheless, in this proposal I will continue to refer to the project by its original, more encompassing (if somewhat deprecated) title.

In their first seven and six years, respectively, these two large-scale digitization initiatives (LSDIs) have already changed the information ecosystem. Between them, they have put more than 17 million books online – more than 15 million at Google Books, and more than 2 million through the OCA (Crawford, 2010; Internet Archive). Yet, within the academy, the only aspect of this transformation that has received a truly extensive amount of scrutiny is its impact on copyright law (e.g., Band, 2006; Bennett, 2007; Grimmelmann, 2009; Romman, 2006; Samuelson, 2010; Travis, 2006). Certain other aspects, such as strategies for building the technological infrastructure to support large digital book collections (e.g., Choudhury et al., 2006; Feng & Manmatha, 2006; Langley & Bloomberg, 2008; Schilit & Kolak, 2008; Vincent, 2008), the problem of providing accurate metadata for such collections (e.g., Duguid, 2007; Jackson, 2008; Lu et al., 2008; Mann, 2008; Mimno & McCallum, 2007; Wang et al., 2008), and the projects’ potential implications for libraries and preservation issues (e.g., Hahn, 2008; Lackie, 2005; Milne, 2006; Nielsen, 2008; Rieger, 2008; Sandler, 2005; Willett, 2009), have also been covered, though to a more modest extent than the legal issues. However, as I have argued elsewhere (Jones, 2010), these bodies of work have remained relatively isolated from one another, and thus provide little in the way of cohesive understanding of what these projects actually are and what they might mean as transformative social phenomena. Returning to the comparison introduced at the outset, then, the research proposed here begins to conceptualize these projects – to understand what they are and what they might mean – by situating them as cases of a broader phenomenon: efforts to democratize information access by providing substantial collections of published media for free public use. In this aspect of their aspirations – and in a surprising number of other ways – LSDIs
strongly echo the precedent of early American public libraries. In both cases, small groups of extremely wealthy individuals and/or institutions threw their financial and political support behind the provision of large quantities of “quality” information to the masses, to be accessed and used in largely self-directed, autonomous ways (Ditzion, 1947; Shera, 1949). And underlying both phenomena runs a current of belief that doing this will improve society – will improve education, equality, democracy – that unproven hypothesis stated at the beginning.

Placing current phenomena in historical context in this way is an essential element in understanding how we have arrived where we are, and to avoiding the repetition of our collective missteps and failures; as Shera notes, just as an individual relies on memories to guide present reasoning, so a society depends on history to shed light on present challenges (1952, pp. 241-242). In the context of LSDIs, the comparison to the early American public library movement will shed light on numerous questions. What sociocultural factors support the initiation of such projects? What assumptions do they reflect and promote? Whose voices are heard (and whose are silent – or silenced) in the process of negotiating the structures of these initiatives? And to what extent have – or can – these systems achieve their universalist, even utopian aspirations? Though this dissertation will not reach conclusions on all of these questions, it will begin to lay a foundation for thinking about them and others.

And furthermore, this comparison will have value for scholarship in library history as well, because the types of questions it raises also remain somewhat underexamined in that field. As Wayne Wiegand has suggested,

On the one hand, [existing work in library history] concentrates too much on the library from the inside out and focuses too much on the institution, the people who practice librarianship within that institution, and the expertise used by the people within the institution itself. On the other hand, it does not concentrate enough on the library from the outside in, nor does it focus sufficiently on people who used (or did not use) the institution, why they used (or did not use) it, and whether the expertise honed within that institution mirrored a particular “language” that was inclusive for some, exclusive for others (2000, p. 21).

These latter questions are precisely the sort being raised by the current research: what is the public library, and how did it come to take that form (or forms)? What sorts of users did those influencing design envision for their new institutions – and what sorts of users did they not envision? As Wiegand notes, there is a real need for this sort of “outside-in” interrogation in the field of library history. This research will act as a bridge between two currently quite dissociated areas of the information field – library history and social analysis of digital libraries and information access efforts – and in so doing will produce new perspectives and understanding on both sides, and, on some level, for thinking more broadly about how the goal of democratizing access to information might best be understood and/or pursued in years to come.

Research Questions
The dissertation proposed here, as noted above, consists of a nested comparative case study analysis of two current
LSDIs and two early American public library systems in their formative years. More specifically, it links these phenomena together as “information democratization efforts” (IDEs): projects where (a) the goal is to provide self-directed access to as many high-quality textual/communicative objects as possible, in a relatively stable setting; (b) the target audience is explicitly everyone, without explicit restrictions based on admission fees, subscriptions, users’ personal characteristics, etc.; and (c) the contents are, at least to a significant extent, free to the user. Initially, I plan to explore the following two questions:

RQ1. Through what social negotiation processes was each project constituted as an “information democratization effort,” in line with the three-part definition above? In particular, how were the key components of the definition conceptualized, and what individuals or groups were most influential in their conceptualization?

RQ2. How can the processes and definitions uncovered in RQ1 be seen to have influenced particular design features or policy elements in each case?

These questions emerge in part from the theoretical framework for the study – a combination of structuration (Giddens, 1984) and Social Construction of Technologies (SCOT) (Bijker et al., 1987), discussed in the next section. Accordingly, both questions provide an avenue to explore the power dynamics and mediation processes at play in the construction of information democratization efforts, as well as recursive relationships between these social shaping processes, the information technologies they produce, and the influence of those technologies and the information they carry upon the societies in which they are implemented. And although this dissertation will focus on these projects’ design and definition processes, and not on their usage, it will undoubtedly provide insights valuable to future research assessing use of these systems in particular social contexts and by particular user groups.

Theoretical Framework

The combination of structuration and SCOT theories was selected for this study for two main reasons: (1) their capacity for bridging the social-technical (and functionalist-interpretative) divide, and (2) the provision of a nuanced account of power dynamics (mainly by structuration theory). I will briefly discuss each of these factors before introducing the study’s methodology.

Both physical libraries and web-based book digitization efforts fuse particular kinds of social assumptions, aims, and usage patterns with particular kinds of technical, physical and regulatory structures – and in both, these social and technical aspects are essentially non-dissociable, if one wishes to emerge with a thorough understanding of either as a social phenomenon. The early histories of public libraries and digital libraries, respectively, provide strong illustrations of this point: in the early years of public libraries, the leadership focused on the social aims of the library, leaving concern with useful architecture – something extremely important to attaining those goals – until much later (Dana, 1897; Garrison, 1979; van Slyck, 1991). By contrast, in the early years of digital libraries,
designers initially focused almost exclusively on the technical aspects of information provision, with little thought given to the social assumptions and implications embedded in those systems (Levy, 2000, p. 25). To avoid such blind spots, the current research strives to understand these projects as sociotechnical phenomena, and seeks an epistemological middle way between functionalism and interpretivism, where both human agency and existing social structures can play a role. This orientation is strongly supported by both structuration theory, which conceives of agency and structure as an integrated duality specifically in order to push past the functionalist-interpretative divide (Giddens, 1984, pp. 25, 162), and SCOT, which proposes a recursive relationship between human agents and the technologies they create, in a sense arguably epistemologically parallel to Giddens, but more tailored to the specific context of technology and design (Bijker & Law, 1992a; Pinch & Bijker, 1984).

The second reason for selecting this combination of theories – and structuration in particular – is the need for a nuanced account of power within the study. In comprehending the negotiation processes behind a complex technological system like a physical or digital library, power relations are extremely important, both because, as Giddens suggests, there “is no more elemental concept than that of power” (1984, p. 283), and because of the importance of dealing with the differential and often mediated influence of different groups on the design of these systems. Different actors relevant to the design of information democratization efforts will stand at different levels of remove from the design process, and will be differentially effective in promoting changes to those designs. Structuration provides a nuanced account of power capable of dealing with these dynamics, particularly through its concept of “rules” and “resources” – elements of social reality that emanate from existing social structures, which simultaneously constrain and sanction actions within those structures, and often serve to privilege particular groups or individuals over others in particular ways (1984, p. 25). The structurational concept of power as exercised through rules and resources will provide a useful scaffold upon which to construct an understanding of the multi-layered design negotiation processes under consideration here.

Methodology
As noted earlier, this dissertation will employ a nested comparative case study approach. The case study method was chosen partially for its capacity for addressing complex questions of process and trajectory, where the researcher has little control over events (Yin, 2003, pp. 5-9), and partially based on the precedent for this type of approach in other research assuming a broadly sociotechnical theoretical perspective (e.g., Bijker et al., 1987; Bijker & Law, 1992b; Fox, 1996; LaFollette & Stine, 1991). A comparative, rather than single-case approach was selected both to increase the potential for drawing more powerful logical inferences about information democratization efforts as phenomena (Small, 2009, p. 23; Yin, 2003, p. 53), and, perhaps moreso, to highlight the non-inevitability of any particular form.
of information democratization effort – not all digitization initiatives must look like Google’s, any more than all libraries must resemble Boston Public, and comparisons highlight this fact. Still, the term “comparative case study” only goes so far in articulating a strategy for research; it can be used to describe a whole host of approaches sharing an orientation toward the “case” as a unit (Schutt, 2009). Thus, the precise case selection, data collection, and data analysis methods for this dissertation are discussed in the sections that follow.

Case Selection
The cases for this study were selected primarily based on their perceived historical significance as embodiments of the information democratizing phenomena they represent: LSDIs and early American public libraries, respectively.

On the LSDI side, Google Books was chosen for its remarkable scale and pace of digitization – an order of magnitude larger and faster than any other digitization project that has ever been undertaken (Internet Archive, 2010; Project Gutenberg, 2010; Singer, 2011) – as well as for its status as the most well-known mass book-scanning project (St. Clair, 2008), its demonstration of a particular model for such a project (corporate, centralized, private), and the potential of the project to reshape not only access to books but also U.S. copyright law as it applies to them (e.g., Samuelson, 2010). The Open Content Alliance, in turn, was chosen both because it is the second-largest book-digitization project in history and because it provides an analytically interesting contrast to Google Books in terms of openness, imaging standards, and legal strategy (Kahle, 2007; Leetaru, 2008).

On the public library side, the Boston Public Library (BPL) and the Carnegie Library program were chosen for the well-established magnitude of each system’s influence on the public library movement, albeit via different mechanisms. The Boston Public Library exerted its influence mainly through emulation of its example: it was the first major urban public library in the United States, and its early trustees created many systems for library administration, funding, and access which were later widely emulated in other library systems (Ditzion, 1947, p. 13; Shera, 1949, pp. 179-181). The Carnegie Libraries, by contrast, were more influential due to their geographic ubiquity and relative uniformity than emulation of them as role models (though that also happened). Within the United States, Carnegie donations funded 1,679 library buildings (Bobinski, 1969), reflecting relatively uniform and centrally-planned policies for local tax support, priorities for building design, and general philosophy of library service (van Slyck, 1991, pp. 369, 376-381). The Carnegie case, however, for the purposes of this dissertation, will not encompass the program as a whole. Instead, it will focus on (a) the elements of the program that were centrally planned by the Carnegie Corporation, and (b) the design negotiation processes underlying two of the 44 Carnegie library donations in the state of Washington – one urban (in Seattle), and one rural (e.g. Prosser or Aberdeen), to provide insight into differences in process related to the size of the municipalities and their respective grants.
Data Collection

Investigating the research questions for these four cases will entail two distinct sets of data collection strategies: one for the public libraries, one for the digitization initiatives.

For the public libraries, the processes of design negotiation relevant to this study occurred long ago: for the Boson Public Library, roughly 1842-1860; for the Carnegie Libraries, roughly 1889-1910. Thus, data collection for these cases will be primarily archival in nature, piecing together correspondence among trustees and administrators, architectural documentation, contemporary media accounts of public reaction and experiences, secondary source accounts, and other historical evidence to construct a picture of the processes involved in the practical negotiation of the fundamental concepts that guided the libraries’ structural and policy development. Many of these documents, including the Boston Public Library’s Trustees’ meeting minutes from 1852-1860, proposed and implemented rules and regulations for the early BPL, and all of the Carnegie Corporation’s correspondence regarding the 44 Washington Carnegie Libraries as well as the 21 library donations made prior to 1899, have already been gathered on visits to the Boston Public Library and the Carnegie Corporation (CCNY) archive at Columbia University in June 2011. Two supplementary archival forays are also tentatively planned for the fall: one to retrieve earlier correspondence dealing with the Carnegie Library program from the Library of Congress’s collection of Andrew Carnegie’s papers, and one to retrieve additional documents related the two Washington Carnegie Libraries selected from local collections in Seattle and elsewhere in the state of Washington.

For the digitization initiatives, the processes of design negotiation are young and ongoing, thus allowing for a more interactive set of methods. For these cases, semi-structured interviews with project leaders and designers, exploring their views of the projects’ development, their role within it, and their take on which other individuals or groups played influential parts, will comprise the bulk of the data collected. Interview questions will be developed iteratively based both on secondary source research and on the preliminary findings from early archival visits. On-site interviews will be sought with individuals directly involved in project leadership or design at either the OCA or Google Books headquarters (both in or near San Francisco, CA), on a visit to be made in September 2011; telephone interviews will be sought with individuals unavailable during that visit, and also with dispersed stakeholders such as the directors or digitization managers of partner libraries. Identification of potential informants will follow a snowball sampling methodology, wherein current informants are asked to suggest valuable additional informants; I would estimate that the total number of individuals interviewed will be between 10 and 25. Interviews will be recorded and transcribed to make them accessible for content analysis. Additional data sought will include observations of the institutional contexts in which the projects are managed, secondary accounts, and, wherever
possible, internal documentation of decision-making processes analogous to that sought for the library cases.¹

**Data Analysis**

Data analysis for all cases will occur on an ongoing basis throughout the archival research and interviewing process, using an analytical-inductive approach (Patton, 2002, p. 454), wherein a start list of codes, generated based on theory and secondary source research prior to entry into the field, is used to guide data collection and preliminary analysis. That list will then be iteratively revised based on trends in the data (Miles & Huberman, 1984, pp. 57-60). The data analysis procedures for this study will differ based on what type of data is being analyzed: primary source materials (actual utterances of project leadership, including both written evidence like correspondence/meeting minutes and transcripts of interview audio) will be subjected to formal content analysis, aided by software such as Atlas.ti, focusing on the actual words used, concepts and metaphors employed, etc. My own coding will be selectively checked against additional coding by a Master’s student assistant conducted during Autumn quarter 2011, to improve consistency and reliability (Miles & Huberman, 1984, pp. 60-63).

Secondary sources and non-textual data (e.g. architectural plans, observations of buildings and web interfaces) will be purposively utilized to confirm or contest emergent hypotheses from the primary source data. For example, if an internal memo between the Boston Public Library’s trustees were to make a claim about public perception or acceptance of the new building, that claim might be verified or contested by newspaper or magazine accounts of the time. Or if an interview with an LSDI project leader indicated that a particular feature was implemented (or retracted) in response to public outcry, the blog postings and media coverage from that time might offer alternate perspectives on what happened. Non-textual data will be assessed in a similarly purposive fashion to secondary source texts, though to a slightly different end: that is, to the extent that design processes are explicitly discussed in primary sources, the designs subsequently produced can be referenced in order to literally illustrate the results of those discussions: which design concepts made their way into physical reality, and which did not? And what might this tell us about the priorities and/or relative influence of those responsible for the decision?

At present, the research is in the midst of the data collection stage, with the Proposal Defense having been completed in early May, the coding start list developed in late May, and the first archives visits completed in early June. Interviews with project personnel are planned for September and October, and a second set of archives visits for later in the autumn. Data analysis will begin over the summer, to lay the groundwork for succeeding phases of the research, and a second coder will begin analysis on a subset of documents in September.

¹ Based on my preliminary conversations with representatives from both projects, I would anticipate that such documents will be easier to obtain from the Open Content Alliance than from Google Books, due to legal restrictions. Both projects’ ephemera – though Google’s moreso – fall under a variety of confidentiality agreements which I do not intend to challenge.
Limitations
As with any bounded research project, the methodologies and sources of evidence drawn upon in this dissertation impose certain limitations. In particular, issues related to both bias and scope are foreseeable; here I will briefly discuss how each pertains to this study. First, given the centrality of content analysis of utterances by project leaders and personnel in this research, there is an issue of self-reporting bias in the data, as well as one of investigator bias in the coding. Especially for the LSDI cases, the first-person accounts collected will undoubtedly be colored to some extent by the way in which project leaders would like their projects to be perceived. I hope to counter some of this effect by (a) constructing the interview protocols using methods shown to produce more accurate historical narratives (e.g., Belli, 1998) and (b) triangulating multiple data sources – including multiple types of informant – as noted above. However, it is likely that in some cases the available data will not completely address concerns of bias.

In order to ameliorate both this and more investigator-based biases, especially deliberate analysis methods will be required, involving both general inter-coder contestation of interpretations, and strategies like Klein & Meyers’s “Principle of Suspicion” (in which socially-created distortions are systematically surfaced) (1999, p. 77).

The second significant foreseeable limitation of the research proposed here relates to scope. That is, as currently formed, the research questions are quite broad. The reason for this, in a nutshell, is that, especially for the historical cases, it was impossible at the outset to predict the type and extent of data that would be retrievable: it was (and still is, to some extent) completely possible that the records retained by the archives visited would address certain pieces of the research questions and not others, with no way of knowing in advance which pieces would be which. Still, I do not anticipate the research questions maintaining their current breadth. Now that some data has been collected, I am beginning the process of iterating and narrowing down the research questions based on the relative prevalence of particular themes within the actual documents gathered. Though this likely means that some of the first data collected will not remain relevant to the dissertation, that scenario seems far preferable to the risk of having too little to report based on too-narrow initial questions, and in any case simply leaves me with a cushion of interesting data to draw upon for further work after the dissertation is completed.

Anticipated Contributions
This research will make three types of contribution to scholarship: descriptive, comparative, and theoretical. First, especially for the digitization cases, it will describe a set of projects that, as noted, have been underassessed to this point, thus laying a foundation for future descriptive and analytical work in this area. It will also describe the BPL and Carnegie Libraries from a novel angle, and will thus also contribute - though perhaps not as radically - to the already substantial body of analysis on the history of those library systems. Second, this comparison is, to my
knowledge, completely novel, at least in the academic literature. Nobody has tried to systematically compare LSDIs to American public libraries. And I suspect there are some powerful insights to be gained from that comparison, regarding first, what elements seem to be general characteristics of this kind of effort, and which might be more idiosyncratic to a particular context; and second, building on that, how these cases, taken together, might inform the way we think about the future of information access – and especially the structures we build to provide it. And finally, it seems likely that this work will lay the groundwork for some extension to the theories discussed earlier, particularly in illustrating theoretical issues specific to information-intensive technologies: by examining the social shaping processes of four initiatives specifically geared toward changing society through the provision of vast quantities of information, this research will help shed light on how the informational content of an information access system might influence the structures that surround its design and use (and vice-versa), and how accounting for that additional level of influence might alter our understanding of the sociotechnical cycle in such cases.

**Conclusion**

The idea that increasing public access to books will improve equality, democracy, and the human condition may not be a proven fact, but it is an extremely compelling theory. And, full disclosure: it’s one that I believe to be true. But the way in which access happens matters. The assumptions ingrained in a context of use – from the height of a bookshelf to the epistemological biases of a genre set – have a meaningful impact on just who actually has effective access to the resources provided – and who gets excluded. If we accept for a moment the utopian hypothesis – that access to information is positive for individuals, and equal access is positive for societies – then this leads us to ask of builders of IDEs like the ones studied here: who is your “everyone”? What is your “everything”? How free is “free”? And building on the answers to those questions, how might those systems be made more inclusive, more extensive – more true to the ultimate aims that underlie this sort of informational utopianism?

These are precisely the kinds of questions being asked by the research proposed here. In their founding era, public libraries were responsible for a massive increase in the amount of high-quality information available to their communities, free of charge; in the current era, projects like Google Books and the Open Content Alliance are striving toward the same basic goal, except that their community is the entire world, and the capacity of their shelves virtually limitless. Their potential significance is profound. These projects could unilaterally (bilateral?) change the entire world’s relationship to books, reading, information in general – indeed, perhaps they already have. Or they could become a transitional form, setting the stage for the next shift in information access to come. And if we are to have a hope of comprehending either of these eventualities, we must begin to understand what these projects are, and what they mean, in the broader context of history.
Works Cited


