

EthicShare Planning Phase Final Report

Submitted by Jeffrey Kahn, PhD, MPH

Prepared by Cecily Marcus, EthicShare Project Director, with contributions from all
EthicShare Partners

July 2007

EthicShare Planning Phase Final Report.....	1
I. Executive Summary.....	3
II. EthicShare Partners.....	5
1. University of Minnesota's Center for Bioethics and University Libraries	5
2. University of Virginia's Institute for Practical Ethics and Public Life.....	6
3. Indiana University – Bloomington’s Poynter Center for the Study of Ethics and American Life	6
4. Indiana University-Purdue University Indianapolis (IUPUI), Center for Bioethics	6
5. Georgetown University, Kennedy Institute of Ethics (KIE) & Library of Information Science (LIS).....	6
III. Content and Collection Development	6
1. Content Acquisition: Georgetown’s Kennedy Institute of Ethics.....	7
a. Georgetown KIE Resources:.....	8
2. Content Analysis	11
a. Subject Area: Public Health Ethics.....	13
b. Subject Area: Bioethics and Religion (search terms: “organ donation” and “Christianity” or “religion”)	13
c. Subject Area: Human Embryonic Stem Cell Research (search terms: various).....	14
3. Digitization Projects and Ingest	15
4. Intellectual Property Issues	16
IV. Specification of Community and Organizational Requirements	18
1. Site Visits.....	19
2. Survey Results.....	22
V. EthicShare Technology Requirements and Functionality	28
1. Selecting a Technology Framework.....	28
2. EthicShare Prototype	32
3. Iterative Feedback for Prototype Functionality	33
4. EthicShare Project Trajectories.....	33
VI. Conclusions from Planning Grant Activities.....	36
VII. Framework for EthicShare’s Proof of Concept.....	37
Appendices	40

I. Executive Summary

Following a 2004 Scholarly Communications Institute hosted by the University of Virginia, faculty and librarians from the group of participating institutions continued to engage the issues surrounding new genre for discourse and exchange. This dialogue resulted in a proposal prepared and submitted to the Council on Library and Information Resources (CLIR) to support a planning effort with a goal of developing what has become EthicShare, a sustainable online environment for the practical ethics community. The effort would assess the requirements to build community, fuel scholarship, and stimulate engagement.

With the University of Minnesota taking the lead, the EthicShare partners—Georgetown University; Indiana University-Bloomington; Indiana University-Purdue University, Indianapolis; and the University of Virginia—envisioned a multi-phase effort, beginning with a foundational planning grant to specify the requirements for such an environment. Funding received from CLIR (with support from the Andrew W. Mellon Foundation) in late 2006 enabled an initial planning phase, the results of which are reported here.

The planning project focused on bioethics—a field of inquiry focusing on the moral dimensions of health care and science—because it enjoys a long history of interdisciplinary scholarship from the humanities, law, and social sciences in addition to the contributions from science and medicine. More than that, however, bioethics provides a scholarly arena and an initial community in which to assess the technical and content requirements, community needs, as well as the governance issues of an online scholarly environment that aims to engage scholars in new and innovative ways. Ultimately, the partners hope to expand EthicShare to serve and support practical ethics scholarship, broadly defined. EthicShare’s planning phase aimed to accomplish three main deliverables:

- Specification of target online content for bioethics
- Specification of desired technology infrastructure to support a discipline-tailored discovery and access environment, with critical functions of inquiry, exchange, and analysis
- Specification of organizational requirements for the EthicShare community, including optimal governance structure, scope, protocols for contribution and exchange, as well as a model for sustained support and development

The story of literature acquisition in the field of bioethics begins with the work of Georgetown University's Kennedy Institute of Ethics and the National Reference Center for Bioethics Literature. The Georgetown KIE data offer significant opportunity for EthicShare's initial collection development, and require substantial effort and investment to migrate and transform the data and re-invent acquisition mechanisms to sustain future growth. The planning focused on increasing the usability and accessibility of Georgetown's path-breaking collection and indexing work, with a goal of addressing the conversion of existing Georgetown data sets into standard formats that allow full text linking (via OpenURL) and other key functionalities. This work is the first step of populating the underlying EthicShare database with high quality and broad ranging resources. Additionally, the EthicShare partners have an opportunity to design new semi-automated and automated ways to create and ingest data records. The goal is to engage in quality collection development, building on the KIE resources, but with attention to reducing labor-intensive efforts and speeding the access to relevant material. Georgetown's KIE reputation, and the quality of its work, position EthicShare very well as an essential discovery environment for scholars in bioethics, and as a foundation for innovative community building.

To compare the target content for EthicShare against the work already done by Georgetown's KIE, EthicShare partners analyzed key collections relevant to bioethics, including digital collections, and performed searches on major databases that support bioethics research to determine the scope, range, and variety of bioethics literature and resources. We investigated intellectual property issues, and in site visits, met with bioethics scholars to discuss the potential of a community-sustained repository for bioethics research, its content needs, and possible user features for discovery, access, and collaboration. We also assessed the research needs and attitudes of bioethics scholars by conducting a survey of bioethics scholars at all EthicShare institutions, and of directors of bioethics research centers across the country.

These efforts informed the planning of EthicShare's technology infrastructure, platform, and design. Over the course of five months, the EthicShare technology team established working principles of development, which include a commitment to open source technologies. Based on the findings of collection analyses, user assessments, site visits, and intellectual property concerns, the technology team assessed options and built a working prototype of EthicShare that was shared with all partners at a planning meeting held at the University of Minnesota in May 2007. The technology team collected feedback on the prototype at this gathering, and from a brief survey of EthicShare partners once they were able to explore the online prototype on their own.

Our major findings were the following:

1. There is a demonstrated need for a robust discovery environment that serves bioethics scholars and that employs easy-to-use interactive features to facilitate interdisciplinary and multi-disciplinary scholarship, collaborative research, and community building and

involvement. While access to content is a core functionality, the planning addressed the necessary balance between developing a repository and mechanisms to federate distributed repositories and publisher services to create a powerful gateway to high quality content.

2. It is clear that the trajectory towards a community-sustained environment requires ample effort by and commitment from the partner institutions, as well as hybrid models of stewardship and contribution between professional and scholarly contributors.

3. There is a need for new models of indexing, classification, and ingest of content that rely more heavily on efficient and flexible semi-automated processes, and less on expensive and labor-intensive efforts.

4. The EthicShare planning phase was successful in identifying the content and collection requirements of an effective discovery and access environment, as well as the needs and preferences of the site's targeted user community.

II. EthicShare Partners

As a fully collaborative project, each partnering institution undertook specific projects to assess the content, technology, and community requirements for EthicShare. Each institution submitted a report documenting the process, findings, and conclusions of the given activities. The content of the reports is incorporated in this final report of the planning project. The activities of each partner were as follows:

1. University of Minnesota's Center for Bioethics and University Libraries

Project Management: Jeffrey Kahn, Kate McCready, Cecily Marcus, and John Butler¹

- a. Platform Analysis: John Butler, Kate McCready, Bill Tantzen, and Chad Fennell
- b. Content Analysis of: Human Embryonic Stem Cell Research and Archival/Core Documents (offline or hard to access commission reports, out of print books, etc.): Kate McCready, Lindsay Reif, and Bart Moffatt
- c. Community Requirements Assessment: Kate McCready, John Riedl, John Butler
- d. Intellectual Property Requirements: Kate McCready, John Butler, Cecily Marcus, with Kenny Crews (IUPUI), consulting.
- e. Governance and sustainability: Jeffrey Kahn, Wendy Pradt Lougee, John Riedl, Cecily Marcus

¹ Due to Kate McCready's maternity leave in April 2007, Cecily Marcus assumed the role of Project Director.

2. University of Virginia's Institute for Practical Ethics and Public Life

- a. Content Analysis: Methodology in Biomedical Ethics and Ethics of Public Health: Jim Childress and Priya Curtis

3. Indiana University – Bloomington's Poynter Center for the Study of Ethics and American Life

- a. Content Analysis: Inquiry into Religion and Medical Ethics: Richard Miller and Karen Boeyink

4. Indiana University-Purdue University Indianapolis (IUPUI), Center for Bioethics

- a. Methodology of Digitization Projects: Identifying Standards and Processes: Eric Meslin, Amy Hatfield and Gabriel Maddox

5. Georgetown University, Kennedy Institute of Ethics (KIE) & Library of Information Science (LIS)

- a. Content Analysis: Assistance Formulating Search Strategies: LeRoy Walters, Doris Goldstein, Joy Kahn and Laura Bishop
- b. Content Analysis: Scope report on LIS databases include major subject areas, numbers of documents, and document types: LeRoy Walters, Doris Goldstein, Joy Kahn and Laura Bishop

Details of our planning efforts follow in the areas of content development and acquisition, digitization needs and best practices, intellectual property issues, community and organizational requirements, and technology development and feedback follow below.

III. Content and Collection Development

During the planning phase of EthicShare (March-July 2007), we identified and documented the specifications for a community-sustained online environment through collaboration with institutional partners and scholars in the bioethics community. Our efforts focused on three fronts: content, technology, and community. The highly respected but underutilized Georgetown KIE data offer significant opportunity for EthicShare's initial collection development, but require substantial effort and investment to migrate and transform the data and reconceive acquisition and indexing mechanisms to sustain future growth. The task of increasing the usability and accessibility of Georgetown's collection and indexing work involves converting existing Georgetown data into standard formats that allow full text linking and other key functionalities.

To identify scholarly databases that provide access to content relevant to bioethics beyond Georgetown KIE's data sets, we also analyzed the holdings of principal databases and collections that serve bioethics scholars. We also documented digitization best practices, surveyed current digitization projects already underway at EthicShare partner institutions, and identified ways that EthicShare can collect and catalog digital materials not readily found in traditional databases. Lastly, with the assistance of copyright expert Kenny Crews, we investigated ongoing copyright issues that EthicShare will need to address as the project moves forward. Details of each effort follow below.

1. Content Acquisition: Georgetown's Kennedy Institute of Ethics

A major element of EthicShare content analysis work was to determine how to maximize the unsurpassed but underutilized collection and indexing work undertaken over three decades by Georgetown University's Kennedy Institute of Ethics (KIE) and its associated National Reference Center in Bioethics Literature (NRCBL). Georgetown's efforts provide the opportunity for a rich core of materials to populate EthicShare in its initial stages. The system developed by KIE seeks to provide comprehensive, cross-disciplinary coverage of substantive English-language materials published since 1973 that discuss ethical and related legal or public policy aspects of the topics and subtopics relevant to bioethics. Citations are drawn from the literatures of the health sciences, the social sciences, law, philosophy, and religion, as well as from the popular media. Georgetown's work is far broader in its coverage than other databases and KIE's reputation for high standards and selectivity has made it a trusted imprimatur of quality among bioethics scholars. Georgetown KIE's reputation, and the quality of its work, positions EthicShare very well as an essential discovery environment for scholars in bioethics, and as a foundation for innovative community building.

Georgetown's citation databases incorporate a variety of publication types, including journal articles; books and chapters within books; newspaper articles; legal documents; government, advisory committee, and task force reports; audiovisual materials; and web-based publications. The monitoring processes used by KIE and NRCBL staff has been international in scope, foreign-language materials have been added to the system since the beginning, and German-language materials indexed since 2005. Primary legal sources include law review articles, court decisions, and government reports (primary sources of relevant legal materials). When laws and bills are newsworthy or potentially trend setting, Georgetown indexes them as well. Georgetown citations also include selected news items from major journals, news magazines, and other popular press sources that together make up an archival record of public concern about bioethical issues. The KIE/NRCBL processes are comprehensive and valued, relying on a well-developed thesaurus and indexing procedures supplemented by classification and indexing routines, carried out by the significant commitment of professional staff.

Most of the indexing and collection work undertaken by Georgetown/KIE has been supported by outside sources such as the National Library of Medicine, the National

Human Genome Research Institute, the Kennedy Foundation, and other public and private sources, but the citations produced by KIE are owned by KIE and can be shared publicly and as a source of EthicShare repository data. At present, KIE databases include abstracts only from selected journals from which specific permission was granted. Abstracts are currently distributed via the annual Bibliography of Bioethics and, in recent years, via the ETHXWeb and GenETHX databases. NLM continues to fund some information activities of the NRCBR and the ETHXWeb database. NLM also indexes several bioethics journals using MeSH headings, but has eliminated coverage of many non-clinical journals that were selectively indexed in BIOETHICSLINE®.

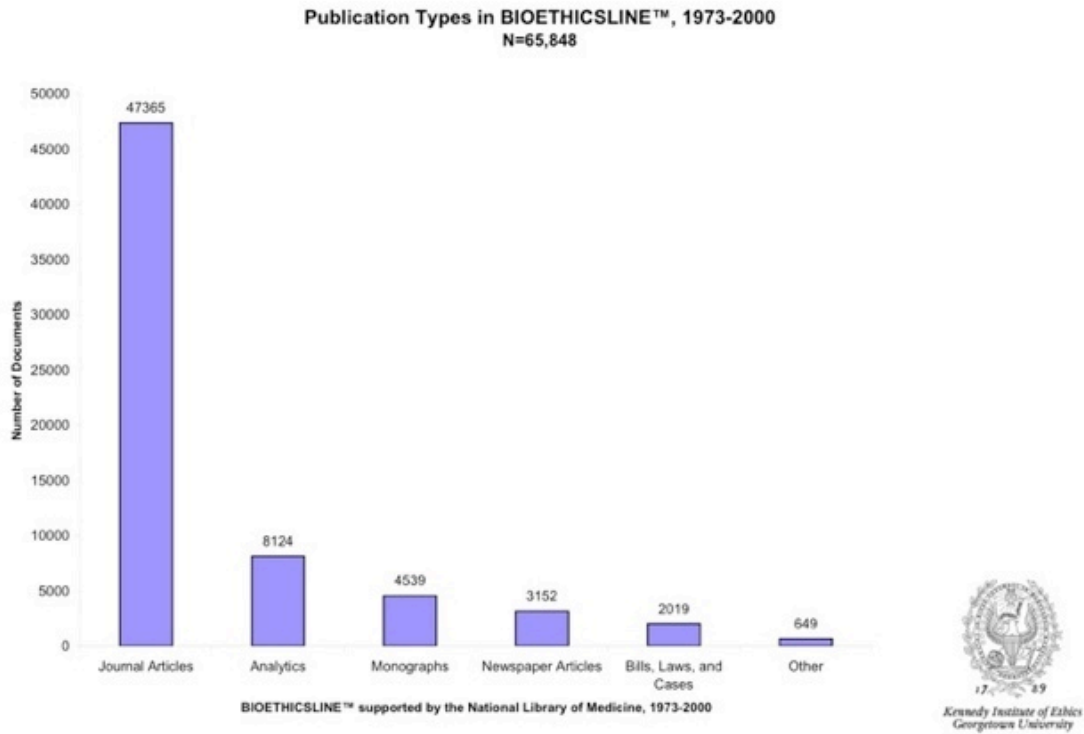
a. Georgetown KIE Resources:

The main bibliographic data sets created by Georgetown include:

- **1973 to 2000: BIOETHICSLINE® Database:** Over 65,00 citations form the "best of the literature"— highly selected materials in the field of bioethics including books, book chapters, law reviews, newspaper articles, journal articles. With modifications and some loss of usefulness, these records were incorporated into either MEDLINE or the NLM Catalog beginning in 2001, in keeping with their respective publication type.
- **1988 to Date: ETHXWeb:** This database holds approximately 230,000 items physically held in the KIE collection. The database records include Georgetown's classification scheme, but no indexing terms. In addition, over 50,000 records for items not held in KIE are included.
- **2000 - 1/2007: PubMed and LocatorPlus:** These records represent those contributed by KIE staff between 2000-2007. The records have some indexing, but only use Medical Subject Headings (MeSH) with some "Other Terms" supplied by KIE staff. PubMed contains journal literature only while LocatorPlus includes books, reports and more, but is not as comprehensive as the former BIOETHICSLINE®.
- **1970 to Date: GenETHX:** National Human Genome Research Institute has funded the KIE's information services in genetics (since 1995). The GenETHX database, a subset of ETHXWeb, comprises new records as well as those added to the library from 1988 to the present.

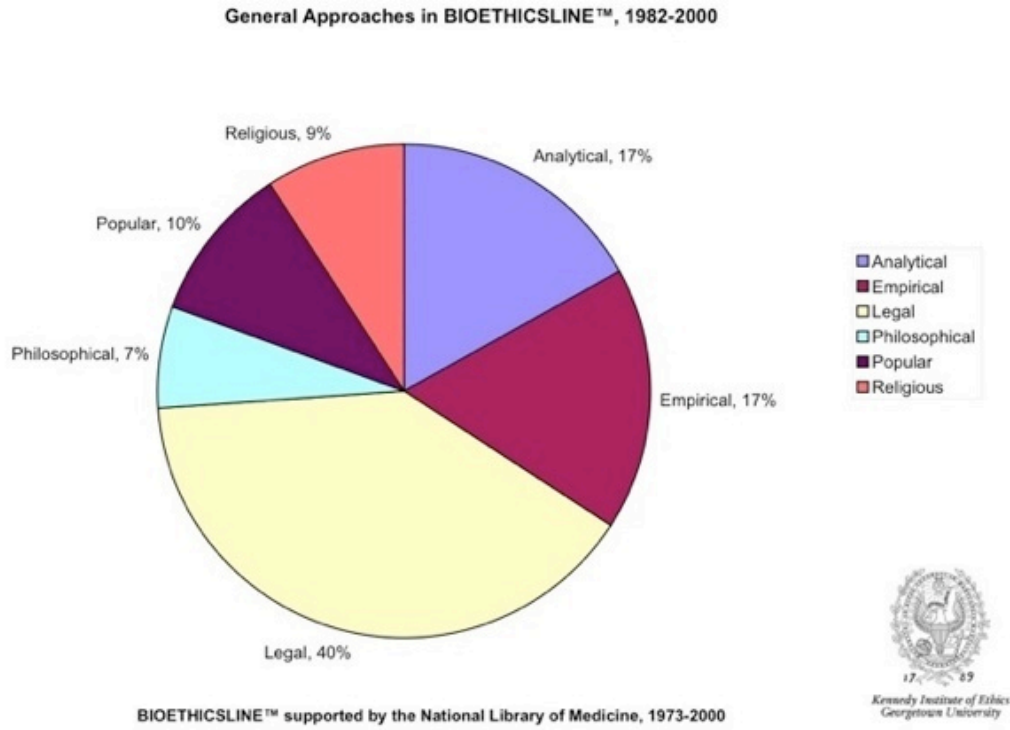
In the specific case of BIOETHICSLINE® (1975-2000), the collection of over 60,000 citations includes over 40,000 journal articles, over 5,000 monographs, about 3,000 newspaper articles, and approximately 2,000 bills, laws, and legal cases See Figure 1.

Figure 1: Publication Types in BIOETHICSLINE®



Broad subject categories in BIOETHICSLINE® (identified between 1982 and 2000) included a broad range of areas including bioethics and professional ethics; war and human rights; death and dying; genetics, reproduction, and abortion; health care and public health; the professional/patient relationship; biomedical and behavioral research; and mental health and behavioral control. See Figure 2.

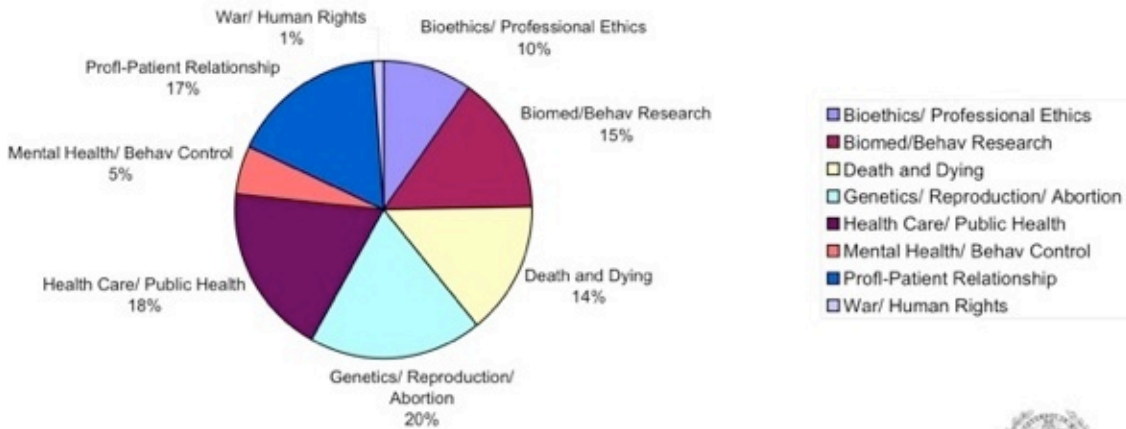
Figure 2: Broad Subject Categories in BIOETHICSLINE®:



The literature catalogued in BIOETHICSLINE® between 1982 and 2000 covers a wide variety of approaches, including legal (40%), empirical (17%), analytical (17%), religious (9%), popular (7%), and philosophical (7%).

Figure 3: General Approaches in BIOETHICSLINE®

Broad Subject Categories in BIOETHICSLINE™, 1982-2000



BIOETHICSLINE™ supported by the National Library of Medicine, 1973-2000



Georgetown/KIE staff has worked over many years to hone the criteria for selection of bioethics-related citations for inclusion in its databases. Since bioethics is a field of study concerned with questions that arise in medicine and health care, in the professional patient relationship, and in biomedical research, it embraces the traditional concerns of medical ethics, focusing on the rights and duties of health professionals and patients, as well as contemporary concerns about the investigator-subject relationship and the social impact of biomedical, behavioral, and genetic research and technology. A third dimension of bioethics is the quest to develop reasonable public policy guidelines for both the delivery of health care and the conduct of research.

Georgetown KIE actively collects and classifies over 60 major topics and sub-topics of bioethics literature. These topics range from professional ethics, to genetics, to reproduction, to human rights. A full list of the Georgetown KIE scope can be found in Appendix A.

2. Content Analysis

In order to identify gaps in existing open repositories or publisher resources and to formulate methods for building EthicShare’s initial core collection beyond Georgetown KIE’s collection, the EthicShare team surveyed the literature available in areas such as stem cell research, organ donation and religion, bioethics methodologies, and public

health ethics. The intention of the collection development exercises was to assess the breadth and depth of the field by looking at small sub-sets of bioethics literature and to be able to compare effectively the holdings of Georgetown KIE's recognized collection and associated databases against other repositories used by bioethics scholars.

Staff members at the University of Minnesota, Indiana University-Bloomington, and the University of Virginia ran repeated search queries on a variety of databases and search engines to determine the number and scope of retrieved items, the percentage of licensed content, the typical availability of full text, and the types of documents represented (books, chapters, journal articles, news stories, reports, etc.). Databases searched included:

- MEDLINE/PubMed
- National Library of Medicine (NLM) Catalog
- Local University catalogs
- WorldCat
- Lexis Nexis Academic
- Lexis Nexis Congressional
- Lexis Nexis Government Periodicals
- Philosopher's Index
- BIOSIS Previews
- Web of Science
- ETHXWeb
- Amazon.com

The findings of the content analyses are undoubtedly indicative of the strategies and approaches adopted by individual researchers who carried out the respective analyses. Consequently, these results say less about the definitive number of sources for a particular sub-set of the field of bioethics than about the range of sources, types of sources, and full-text availability across the field generally. Spanning the humanities, social sciences, and sciences, and including law, religion, philosophy, medicine, public health, politics, and social policy, the corpus of bioethics literature is at once highly distributed and focused. No one database can claim comprehensiveness, but the focused and selected approach of the Georgetown-created datasets comes closest.

With the exception of the resources created by Georgetown University, most databases that serve bioethics scholars are insufficient when considered alone since they largely exclude literature from fields such as philosophy, religion, and law. Additionally, most databases are limited to a few types of literature (journal articles or books), and often exclude literature from the popular press, multi-media sources, government documents, legal cases and analyses, and other types of literature.

At the same time, search engines like Google Scholar are not selective enough to deliver a manageable number of high quality results. Through content analyses of select bioethics subjects, we have determined that available databases such as

MEDLINE/PubMed and NLM Catalog do not have the desired depth and breadth of resources related to bioethics and practical ethics broadly. Though full text accessibility is commonly available for journal articles, other formats (such as commission reports, government documents, and book chapters) are less available in full text online, and are infrequently indexed by most databases.

A brief synopsis of each subject area, as well as some general observations about the scope of bioethics literature surveyed, follows.

a. Subject Area: Public Health Ethics

Public health ethics has only recently been identified as a specific area of applied or practical ethics within the field of bioethics, and there is considerable debate about its boundaries. In a landmark definition, an Institute of Medicine report in 1988 stated: “Public health is what we, as a society, do collectively to assure the conditions in which people can be healthy.” The “ethics” in public health ethics is often closer to social, political, and legal philosophy than to moral philosophy narrowly construed. In other terms, it involves “social ethics” and “political ethics” as well as “ethics and public policy.”

For the purposes of collection analysis for EthicShare, EthicShare project staff chose public health ethics because it is still relatively a new sub-field and its interdisciplinary nature requires a broad approach. Searching across databases was particularly useful because search results included literature on more specific ethical debates within public health relating to issues such as vaccination, quarantine, and drug allocation strategies.

Of the databases searched, the relevant holdings were predominately made up of books and articles, the majority of which were articles. In nearly all cases, the content is licensed, and full text is available in most cases, except for instances where the document type is a commission report. In the area of public health ethics, databases such as WorldCat, Google Scholar, and ETHXWeb yielded the highest number of search results, numbering in the hundreds. Other databases had many fewer returns, numbering in the tens (MEDLINE, PubMed, Web of Science, Philosopher’s Index. This suggests that ETHXWeb is well suited as a source of citation data for EthicShare, since its holdings are broader than any other single source.

b. Subject Area: Bioethics and Religion (search terms: “organ donation” and “Christianity” or “religion”)

The sub-field of religion and bioethics is vast and interdisciplinary, drawing on scholarship from the fields of medicine, law, philosophy, religious studies, and literary criticism. The results yielded from broad searching across a number of databases and websites were so great that EthicShare staff narrowed the scope to include only organ donation and Christianity. When searching the topic of organ donation and Christianity, the NLM catalog returned the most results, but all other databases yielded fewer than 80

returns. Results returned by Google Scholar and PubMed, more than 50% of which were articles. In the case of Google Scholar, 15% of returns were books, 85% articles. ETHXWeb produced very few returns with the narrow query of organ donation/Christianity, but was useful for a broader query of organ donation/religion.

Because the databases being searched returned relatively small numbers of sources, EthicShare staff broadened the scope of databases and found that a number of other sources are extremely useful for research in the area of organ donation and religion, and collect materials not readily found in more mainstream databases. Some examples include:

- www.nccbuscc.org – U.S. Conference of Catholic Bishops
- www.all.org/aba - American Bioethics Advisory Commission
- www.vatican.va
- www.flacathconf.org – Florida Catholic Conference
- www.llu.edu/bioethics - Loma Linda University
- www.linacre.org
- www.parkridgecenter.org
- www.electronicchurch.org – various denominations
- www.cec-kek.org – Conference of European Churches
- www.episcopalbookstore.org
- www.bioethics.gov
- www.elca.org – Evangelical Lutheran Church in America
- www.lahey.org – Lahey Clinic Medical Ethics Journal
- www.ingentaconnect.com
- www.wcc-coe.org/wcc - World Council of Churches

c. Subject Area: Human Embryonic Stem Cell Research (search terms: various)

Research involving human embryonic stem cells occupies a significant segment of bioethics scholarship, and the literature on this topic is vast and interdisciplinary. There are numerous government documents and commission reports dedicated to this issue. Additionally, a significant portion of the academic literature in bioethics takes up the debate on this issue, but the debate is not confined to the academic specialty of bioethics since scientific and legal journals also address these issues. There is also substantial discussion of ethical issues related to human embryonic stem cells in the popular media, with the political debate over funding stem cell research stimulating opinion pieces and feature articles in the news.

The search criteria used for to analyze the literature related to stem cell research were based on the topic of stem cells, and focused on the ethical issues of specifically human embryonic stem cells. The ethical debate surrounding stem cell research is relatively concise, but the large volume of existing literature spans disciplines and media. Google Scholar produced such a large return (over 3000) that search strategies had to be modified

to provide a more usable sample. Databases such as ETHXWeb, MEDLINE and Web of Science also produced significant sets of results (1500-2000), though popular press sources were limited, accessibly most often through Lexis-Nexis. Government documents were most often found in specific databases such as Lexis-Nexis Gov and Congressional. Generally, search results returned article and book sources.

3. Digitization Projects and Ingest

A critical component of the EthicShare project was to assess the feasibility of digitizing relevant, significant print materials and making them discoverable and accessible through the EthicShare database. Towards this end, the EthicShare team at IUPUI documented the definitions, processes, standards, and policies for metadata creation and recommended a digitization workflow process that draw on well-established protocols to assure interoperability and the preservation of data over time. These standards will be shared with all repository administrators who would intend to have their locally hosted content indexed in EthicShare. The most important requirement for partnering repositories will be the existence of metadata that can be brokered through the Open Archives Initiative Protocol for Metadata Harvesting (v.2.x).

The IUPUI participants also outlined a recommended work flow process that includes the standards projects might follow for scanning processes, how best to produce automated or manual metadata description and image treatments, and how to preserve and disseminate digital content and metadata.

Current Digitization Activities by EthicShare Partners

Digitization activities are already underway at EthicShare partner institutions Georgetown and IUPUI.

At Georgetown, KIE staff is in the process of digitizing:

- Hastings Center Journals – authorized to digitize all back issues (2 year embargo); approximately 6 years have been done. Files are accessible via ETHXWeb
- IRB newsletter – similar authorization to digitize back volumes with 2 year embargo
- *Scope Notes Series* – This series published by the National Reference Center for Bioethics Research to provide an overview to major issues in bioethics are all available online in full text; genetics *Scope Notes* are updated regularly; others appear as most recently-published version.
- Syllabi project – 185 of 600 course syllabi are available full text; digitizing is supported by NLM contract
- Office of Human Research Protection (OHRP) project – National Commission Reports – this one-year project has been completed; Federal Register materials are accessible via ETHXWeb, a KIE home page for the project, and at the OHRP web site. OHRP plans to post the National Commission reports soon.

At IUPUI's Center for Bioethics Digital Library (BEDL), activities include:

- National Bioethics Advisory Commission Reports
- Central Indiana Bioethics Portal
- International Bioethics Reports and Recommendations
- National Commission for the Protection of Human Subjects in Biomedical and Behavioral Research
- Indiana Eugenics
- Center for Bioethics reports

The Future of Digitization for EthicShare

In researching potential bioethics digitization projects, we found that many materials useful to bioethics scholars are already available in digital form. These include public domain documents such as the current President's Council on Bioethics, various previous presidential commissions and the National Bioethics Advisory Commission. However, many commission reports, government documents, and archival or historical materials are dispersed at a variety of websites and are not generally findable through traditional databases and search engines. The distribution of many resources relevant to bioethics research and scholarship is perhaps a more pressing concern to the scope of EthicShare than the lack of digital resources, and as such, there is a need to develop innovative ways to federate, describe, and provide access to key but obscure digital resources.

At the same time, a number of commission reports published in the 1970s, as well as a significant number of international commissions and reports are currently unavailable online. Out of print and rare books in bioethics are difficult to access, and archival materials such as older meeting notes and ancillary papers of the relevant President's Councils and Commissions are also currently unavailable online.

Current digitization projects already underway at EthicShare partner institutions (Georgetown, IUPUI, and potentially through the planned Google Books program at Indiana University and the University of Minnesota) make it unnecessary for EthicShare to engage in concentrated digitization projects. In looking forward, EthicShare would focus its efforts on aggregating description and access to relevant digital content hosted in distributed repositories, especially those of partners active in digitization projects related to the field of interest. EthicShare will promulgate best practices, and assist in the implementation of OAI mechanisms at partner sites to facilitate metadata harvesting and aggregating in the EthicShare database.

4. Intellectual Property Issues

To investigate issues related to intellectual property right and licenses, EthicShare staff worked with Professor Kenneth Crews, JD, PhD, MLS, outgoing director of IUPUI's

Copyright Management Center (CMC) and future director of Columbia University Libraries Copyright Advisory Office. Professor Crews participated in the May 10 planning meeting held at the University of Minnesota, and served as a consultant on the project generally.

In Crews's estimation, EthicShare's main intellectual property concerns and activities have to do with outlining a clear set of practices that will govern how content is put into the site (content acquisition), what is done with the content once it is there (content access), and the effective management of new materials created for the site (content creation). At all stages, EthicShare must work closely with the host university's legal counsel. An outline of key intellectual property issues of each area of activity follows below.

1. Content ownership and acquisition

- Some EthicShare original work (software, programming code) is protected by copyright. EthicShare is committed to the use of open source software and the open distribution of any software created for EthicShare.
- Comments posted on EthicShare are copyrighted by the authors, and arrangement between the authors and host may be helpful for any redistribution of the content.
- The host institution of the EthicShare project is dedicated to open source and open access principles for software, coding, and appropriate content.
- Main third-party content sources (including citation databases, full text content, digitized content, bibliographies, or selections of citations on a specific topic) are protected by copyright and must be handled accordingly.
- Public domain content, most U.S. government documents, and citations are not subject to copyright restrictions. However, license agreements may prohibit large-scale downloading and public redistribution of content accessed through such agreements. Case-by-case review of public domain status and uses allowed via licensing arrangements is needed.
- Populating EthicShare with Georgetown KIE data gives EthicShare access to a significant amount of data through a single agreement.² By working with Georgetown data, EthicShare can set a good-faith precedent that could bode well in subsequent negotiations with other data providers.
- In reviewing EthicShare institutions' individual license agreements with content providers such as Lexis Nexis, Religion ATLA, WorldCat, and Philosopher's Index, we found that many licenses prohibit public redistribution of the content. As EthicShare proceeds, we should work closely with legal counsel at the host university to discuss any possible ramifications of ingesting citation data.
- EthicShare can work with publishers to seek any necessary licenses for using proprietary datasets.

² Some abstracts attached to Georgetown citations may contain abstracts that are protected under copyright law.

- EthicShare will promote clear standards of use and practice in accordance with intellectual property law.
2. Content Access by users
 - As a discovery and access environment, EthicShare will connect users to copy of full-text content as authorized through their affiliate institution. To do so, EthicShare will leverage the OpenURL protocol, institutionally managed link resolvers, and a global OpenURL resolver registry to connect users seamlessly to appropriate licensed copy.
 - Content that is in public domain or authored by the U.S. government, and some archival content, can be legitimately accessed in full text. All other content will have to proceed according to the license agreements of each individual institution.
 - There is a significant amount of ambiguity in many license agreements and terms of third-party content. EthicShare will have to work with the host university's legal counsel on a case-by-case basis to assess risk factors.
 3. Content Creation (by users and others)
 - Community contribution to the EthicShare database (submitting articles, links, citation, comments, etc.) poses risks of copyright infringement only if a user contributes works for which he or she is not the legal owner. In such cases, a "take down" policy would allow for the expedient removal of any questionable content so that it may be thoroughly reviewed by EthicShare's legal counsel.
 - According to Crews, EthicShare and the university hosting the site may have some protection from possible legal suits as long as EthicShare acts expediently to remove content when necessary.

The main management procedures EthicShare faces in the realm of intellectual property and copyright law include determining the source of content for EthicShare's repository, determining the processes for implementation and removal of content when necessary, and calculating risk based on fair use precedents and principles of open access and advancing scholarship. Given the serious nature of these issues, we will consult widely with relevant experts to determine a protocol for best practices, including the respective general counsels of the EthicShare partners. Professor Crews has proposed to offer guidance with respect to these issues as EthicShare proceeds.

IV. Specification of Community and Organizational Requirements

To better understand the needs of the scholarly community, EthicShare staff captured attitudes, needs, and research challenges of a selected segment of the bioethics community through site visits at all EthicShare partner institutions as well as a paper or online survey of over 90 bioethics faculty members, research associates, postdocs, and

graduate students. The objective of the site visits and survey was to engage in an iterative process to gauge bioethics scholars' attitudes about existing content, systems for information retrieval, social networking features, community participation within online environments, and to elicit overall comments on EthicShare project goals. Details of the site visit and survey findings follow below.

1. Site Visits

Hosted and planned by lead faculty at each participant site (Indiana University—Bloomington's Richard Miller, IUPUI's Eric Meslin, UVA's Jim Childress, Georgetown University's LeRoy Walters, and Jeffrey Kahn at the University of Minnesota), visits to partner institutions by University of Minnesota project staff during the month of February provided invaluable input from partners and the scholars and librarians with whom they work. While each of the five visits was slightly different in terms of agenda and activities, the goals were the same—that is, to introduce EthicShare to scholars in the field of Bioethics and discuss the following:

- User features – what features do they/their colleagues want out of an online community service?
- Content requirements – what type of content is most important? core documents; historical documents; secondary literature; books; recommendations from colleagues?
- Ethics community involvement –what role will scholars play in creating, participating in, and sustaining a community site?

Other issues discussed included methodology and data analysis updates, institutional inventories of bioethics collections, and the involvement of graduate students and librarians in the project. The visits gave partners the opportunity to meet with UM project staff and discuss the particularities of their institutions and areas of expertise.

At each visit, University of Minnesota (UMN) project staff met with a group of faculty members, research associates, postdoctoral fellows, and graduate students (between 7-17 individuals) and gave an overview of the project. A review of online sites with specific features was presented and used to engage the group in discussion of various search functions, user tools, social navigation and networking functionality (e.g., Citeulike.com, Connotea.com). Full text access through OpenURL functionality was also discussed, as well as specific features that facilitate discovery and sharing. After the presentation and a discussion, scholars filled out a paper survey about site features, content needs, and user preferences.

Key findings include the ideas that EthicShare should:

- Allow users to easily identify high quality materials
- Give users comprehensive, full text access to all material types
- Provide access to materials in all related academic fields
- Maintain space for community discussion and community-building
- Allow users to maintain private workspaces within the site.

Other critical issues from the site visits are outlined below.

Major Needs for EthicShare Community Site

- a) Comprehensive access to all material types: full text access to journal/book literature, reports, non-English language materials, audio/visual materials, case studies and legal resources, and conference proceedings and abstracts
- b) Access to materials in all related academic fields: Bioethics, Medicine, Science, Social Sciences, and Humanities—Philosophy, Medicine, Anthropology, Religious Studies, and History
- c) Ability to identify quality resources: recruit experts in the field, present peer-reviewed sources, and display a filtered/sorted hierarchy
- d) Community space: finding collaborative partners, keeping up with new developments in the field, upcoming conferences, and calls for papers
- e) User ability to define audience: make it possible for users to focus their research community, communicate to a specific group, make connections, and provide opportunities for participation
- f) Private work space and collaborative work spaces: project-based tools and services

Desirable Features of EthicShare Community Site

- a) Search/display features: single point of access, new content alerts, categorization by format/discipline/topic, and more
- b) Private space + public space + collaborative space: private workbook, project management features, collaborative writing tools
- c) Evaluation/ratings/rankings/reviews/commentary: accountability is valued, filters between expert and community ratings, ability to make annotations
- d) Usage tracking: collaborative filtering (e.g. Amazon), administrative tracking number of downloads, comments, ratings, and reviews by users in order to represent high usage, use patterns, etc.
- e) Community building features: announcements, message boards, grants, directories
- f) Open courseware: sharing syllabi, lectures
- g) Evaluation/ratings/rankings/reviews/commentary; some comments from potential users:
 - i) Concern over who are the “experts.” Who reviews? What are the incentives?
 - ii) Rating systems are “overrated.” “I wouldn’t trust it ... It’s entertainment and [possibly] damaging.” “Ratings war doesn’t add value.”
- h) Community-sustained approaches; some comments from potential users:
 - i) Adding resources - “Paid staff should do this; this is not a good use of [scholars’] time”
 - ii) Lot of places for people to go....now, one more?

- iii) Comments and ratings by users/scholars more important than those made by paid staff
- iv) Identity of commenter/evaluator is important for making contributions meaningful and trustworthy
- i) Interface design & content
 - i) Interface must be easy to use

General Commentary

- a) Tenure and promotion considerations: some discussion focused on the general usefulness/risks for scholars, especially junior faculty, of a community-based site. Some scholars are hopeful that promotion and tenure pressures may be aided by the system's ability to enable junior faculty to find out about projects. Others wanted to know how and if their institutions would value participation, and count contributions when considering promotion and tenure cases. The esteemed value of editorial boards and the process of peer review were viewed as models for facilitating site ratings, content quality, and other community-added features.
- b) Sustainability/community involvement considerations: though some scholars do not expect to devote significant amounts of time to adding content to a community site, they say they might ask graduate students to do so. Issues of accountability (of quality, reputability) are important to scholars, and they acknowledge the importance of achieving a good balance of usefulness and participation. One scholar comments, "Community is built through trust and cooperation." Scholars also debated the factors that would "entice the community" to participate, and suggested that partnerships with professional and academic associations may be key to sustaining the site.

Partnerships and Sustainability

- a) Developing partnerships: how best to demonstrate value of collaborative efforts that go beyond individual institutions, how to effectively leverage associations, societies, and organizations?
- b) Resources for educators and wider community: potential for EthicShare to be a valued resource for K-12 educators, as well as lawyers/doctors interested in the field of bioethics through quality resources including syllabi and course materials.
- c) Publication distribution and sharing: creating EthicShare as source for pre-prints, white papers, and other open-access publishing models that allow additions to original materials).
- d) Use existing standardization of terminology: importance of leveraging existing standards, e.g., Bioethics Thesaurus (KIE), Proper Names and Organizations (KIE), as well as inter-operating with existing technology architectures.

2. Survey Results

An online and paper survey was distributed to bioethics scholars at Indiana University-Bloomington, IUPUI, Georgetown University, the University of Virginia, and the University of Minnesota (See Appendix 1). Thirty-six participants who attended the site visits completed the paper survey. An online survey, sent to directors of bioethics centers across the country, and to graduate students and postdoctoral fellows at EthicShare partner institutions, was completed by 68 individuals. For comparison purposes, most of our data analysis focused on the differences between responses by graduate students and postdocs (34 responses) versus the responses by faculty members (64 responses). Our sample of respondents was small but representative, and the similarity of many answers (often despite age, institution, or professional status) suggested that there are general conclusions that can be drawn from the responses we did receive.

The average age of faculty member respondents was 51.6 years old and the average number of years working in the field of bioethics was 19.3 years. The average age of graduate students and postdocs was 33.4 years, and the average years working the field of bioethics was 4.75 years. For comparison purposes, the six graduate students who completed the original paper survey were included in the sample of graduate students, (all versions of the survey were identical), creating a sample that included 7 MS students, 17 PhD students, and 4 postdoctoral fellows.

Key findings are presented below.

Materials Used

When asked what types of bioethics materials are important to discover in a community site, all respondents clearly identified “high quality” materials as the most important component. 100% of respondents marked this as either “very important” or “important.” Journal articles were the most important document type among all respondents. Commission reports, often mentioned in site visits as hard-to-find-but-important research resources, rated higher in importance among faculty than among grad students/postdocs (88% of faculty rated commission reports “very important” or “important,” compared to 82% of grad students/postdocs).

Online access to translations, interviews, websites, blogs, and other materials that have not yet been digitized is slightly more important to graduate students and postdocs than to faculty.

Other types of bioethics materials that scholars would expect to find at a community site include legal case materials, links to other bioethics online resources, course syllabi, position papers on issues relevant to bioethics, and links to media reports (television, radio, newspaper) that feature bioethics scholars and scholarship.

At the launch of a EthicShare beta repository, the core documents and collections that bioethics scholars are most interested in include government and commission reports (US and international); access to ETHXWeb resources, journal articles from core publications

such as *Hasting Center Report*, *Journal of Medicine and Philosophy*, and *Kennedy Institute of Ethics Journal*; and case key case studies, among other things.

When asked what non-digital resources should be made accessible in digital form on the site, scholars mentioned older journal articles, books and textbooks relevant to bioethics as well as reference materials, non-digital governmental and commission reports, and other “hard to find” core resources.

In general, scholars are looking for broad access to quality materials regardless of document type or source, though traditional scholarly sources (book chapters and peer-reviewed journal articles) remain paramount.

Resource Description and Retrieval

When asked to rate the usefulness of data fields contained in information about resources (such as document type, searchable full text, a list of and links to cited references, bibliographic citation with abstract and full text, or subject terms assigned by an expert) all respondents cited bibliographic citation with abstract and full text as most important (91% faculty and 96% grad student/postdoc). Searchable full text is also very important (75% faculty; 89% grad/postdoc). Searching by document type is more important to graduate students and postdocs than faculty: (82% grad/postdocs and faculty 75%). Displays of the number of cited references of a resource rated lowest among all respondents: 39% faculty and 43% grad/postdoc.

Site Scope

Respondents indicated a number of desirable features:: find only high quality materials, find and report news in the field, access resources selected by experts, find all relevant materials on a topic, read comments from other users, get feedback, or write comments:

- Finding all materials rates higher among grad students/post docs than faculty: 86% grad students-post docs; 68% faculty.
- Graduate students and post docs are slightly more interested in reading comments than faculty: 50% grad students-post docs; 45% faculty.
- Graduate students and post docs are slightly less interested in writing comments than faculty: 25% grad students-post docs; 29% faculty.

The online resources that bioethics scholars currently rely most on include PubMed, ETHXWeb, the KIE website, Medline (via Ovid), and Google Scholar.

Site Functionality

Respondents were asked to rate the features that would be most attractive to them at the site, including the ability to search by topic/subject, the ability to search within the full text, access to full text, ability to create styled bibliographies, as well as sort results, get updates, keep work private, collect and organize resources, share work, and add, review, and discuss resources.

Searching by topic/subject was rated most important by all respondents, and being able to access and search within the full text was also seen as critical by graduate students, postdocs, and faculty.

Graduate students and postdocs found it more important to sort resources by usage/impact than faculty: 71% faculty; 86% grad students/postdocs.

Other features/content that would be useful to scholars in a community site included:

- Discussion space with shared access to collected resources
- Innovative search options
- Links to and resources from community groups outside of bioethics
- Links to cited references
- Video and multi-media resources

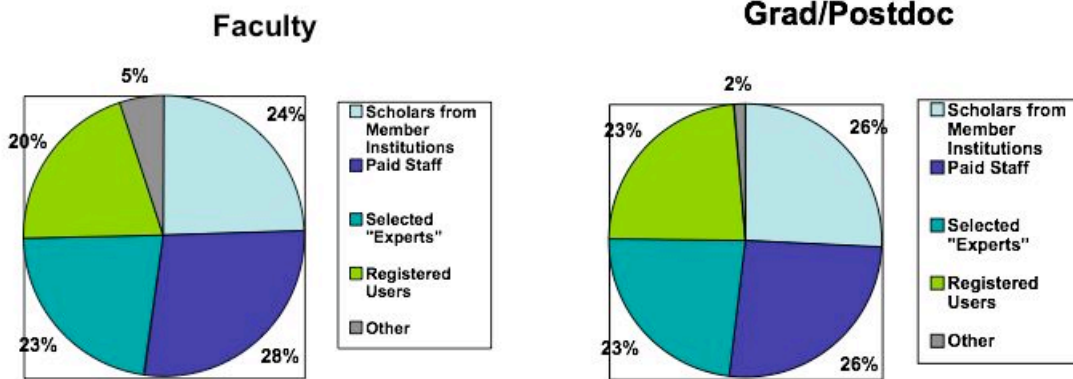
When combining “important” and “very important” ratings, “Social” features rated higher among grad students/post docs than among faculty.

- Space to collect work (86% grad-post doc; 71% faculty)
- Get updates via email/RSS about new content (64% grad-post doc; 55% faculty)
- Ability to review a resources (54% grad-post doc; 45% faculty)
- Community discussion space (54% grad-post doc; 41% faculty)

Community Maintenance

When asked who should be responsible for adding to the core content of EthicShare, many scholars agreed that paid staff members should bear the largest responsibility in comparison to selected experts, scholars from member institutions, and registered EthicShare users. At the same time, graduate students and postdocs surveyed responded that registered users, experts, and scholars from member institutions should have nearly as much responsibility as paid staff for building up EthicShare’s collection. In general, graduate students, postdocs, and faculty agreed that the responsibility of adding content to EthicShare should be distributed almost equally among paid staff, registered used, scholars from member institutions, and selected exerts in the field of bioethics:

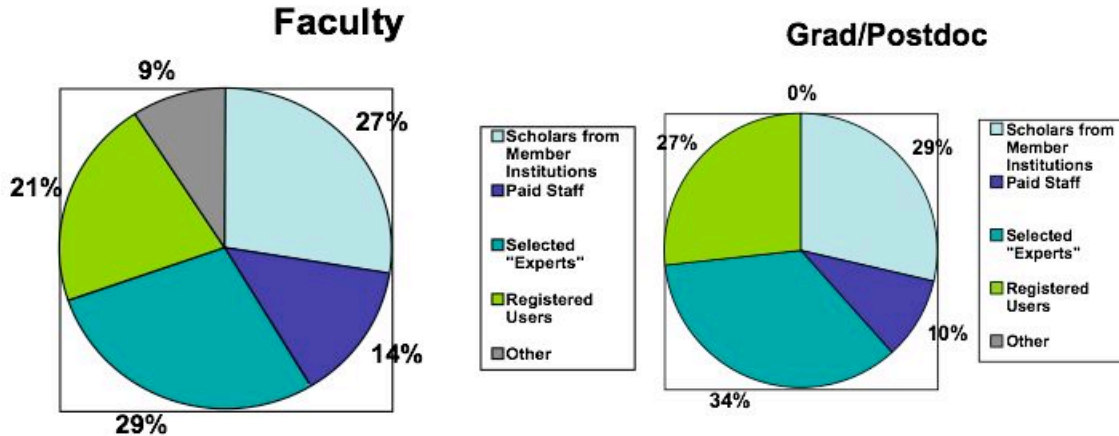
Who Should Add Resources to the Site?



When determining the quality of resources through ratings and comments, scholars trust their peers the most. Graduate students and post docs are slightly more interested in reading comments than faculty, but they are slightly less interested in writing comments than their faculty counterparts. Nearly all survey respondents want to depend most on registered users and selected experts to evaluate and comment on the quality of resources. The importance of an editorial board and other governance structures to ensure community engagement and quality was mentioned repeatedly during site visits. Again, though, agreement across respondents was high, and the distribution of responsibility for adding ratings was nearly equal.

On the question of comments, most respondents agreed that comments posted by paid staff would be less valuable than comments from other scholars and site users.

Who Should Add Commentary to the Site?



Features that would facilitate participation in a community site are:

- Ease and intuitiveness of use; uncluttered and not overpopulated with features that you could get lost in (and lose track of time)
- Community discussion: Ability to get feedback on your own work and to comment on others; discourse among participants/users; ability to connect with and "meet" other scholars- through message boards and online discussions; ability to connect with and "meet" other scholars- through message boards and online discussions; trusted, known users; reciprocity (you contribute and you use).
- Comprehensiveness of resources/ materials included; availability of resources not easily accessible otherwise; search capacity that enabled users to identify more materials than they would be able to find through normal web searches and databases.
- Increased full text access

Survey and Site Visit Findings

While our sample was small, the results are suggestive of some interesting possible cohort differences as well as differentiation of potential roles and responsibilities among categories of users.

1. Community Maintenance—Adding Resources, Commentary Resources
 Survey and site visit findings suggest that EthicShare users may not readily take on the role of populating the site with content. What mechanisms or incentives will facilitate participation of the user community? Part of EthicShare’s process should be to assess the

successes and/or limitations of social features in a scholarly environment, and should work to develop an iterative methodology for engaging community participation

2. Search and Access

Our findings suggest that many bioethics scholars want similar things at a research site (high quality resources, breadth of coverage, the ability to share resources, and the ability to work privately), but interdisciplinary breadth of the fields of practical ethics make it essential that EthicShare provide a wide array of resources in full text. What is the balance between comprehensive access and vetted selection? If EthicShare is not comprehensive, its value-add is a combination of factors:

- Reputable resources and standards for selection
- User features that facilitate and enhance discovery, access, and sharing
- Flexible design that promotes ongoing collection development in innovative ways

Quality control is important to users: mechanisms to review content (e.g., an editorial board model), combined with mechanisms to track user behavior and interests (with attention to privacy), is one potential model.

3. Social tools in scholarly environment

EthicShare will need to balance privacy concerns with user features that allow scholars to work individually, with collaborative partners, and with the larger EthicShare community. The social aspects of EthicShare, though in some cases new to many bioethics scholars, offer possible benefits to the production of scholarship and the development of community, as well as to the creation of new forms of scholarly discourse.

Though there is some expressed ambivalence and even negativity about some social features, scholars did note the potential value of rankings, commenting, recommendations, etc., if they are framed by a focused and meaningful context. Scholars also noted that the identity of the individual providing comments, ratings, and so on, is an important aspect of the value of social input.

Part of the work of EthicShare will be to assess the value of such tools in a scholarly environment and for a community of scholars who are somewhat unaccustomed to social features.

Overall, our findings suggest that mechanisms for long-term maintenance of EthicShare content needs to be built around flexible, low-barrier ingest and evaluation methods that allow for membership contributions and some editorial control and management of content.

V. EthicShare Technology Requirements and Functionality

How can technology help to advance a field of knowledge through strategies that stimulate and sustain the contributions of the scholarly community? For EthicShare technology development, the response to this question began by seeking a deeper understanding of bioethics and practical ethics researchers and the culture, behaviors, and work processes that surround their scholarly practices. Assessment data collected through the project’s focus groups, surveys and other discussions involving scholars in the “EthicShare community” was studied and mapped to emerging models in the development and support of scholarly communities in the online environment. Committed to an iterative design process, we identified and prioritized functional requirements, selected a core platform, and then developed and shared an online prototype with EthicShare partners to gather initial feedback for further refinement and specification.

The EthicShare prototype (<http://ethicshare.cs.umn.edu/>) provides the foundation for a cohesive online environment for thorough information discovery and access, personal and group information management, information and opinion sharing, and scholarly discourse. Tailoring the right mix of content, tools, and communication pathways for this particular community was and continues to be the design priority. We also set as a design goal a flexible infrastructure that could be extended for additional communities of interest. We considered it highly valuable to build on a core technology, sufficiently abstracted and modular, which would facilitate a high level of agile “tuning” at the discipline or community-level for future opportunities.

1. Selecting a Technology Framework

The EthicShare technology team sought an existing core technology framework that 1) satisfied the following principles: open source, open architecture, modular, scaleable, standards-based, use of universally well-understood components that could support rapid development iterations; 2) demonstrated orientation towards support of online communities and that could be customized to the specific needs of the EthicShare community, as well as be extended in future iterations; and 3) provided momentum that we could build on and contribute to, and that demonstrated a developers’ orientation to collaborative information sharing systems and strategies, and a strong community base. In several past projects UMN computer scientist and EthicShare team member John Riedl and his team have explored community maintained repositories under two classes of technology platform:

- **Special-Purpose:** MovieLens is a special-purpose web site for personalized movie predictions. We have explored a range of community maintenance possibilities within MovieLens. Users are enthusiastic about participating in the support of their community, but each new affordance requires special-purpose coding.
- **Wiki-Based:** The WikiLens project (www.wikilens.org) is exploring a "small-

world" recommender site in which all content is provided and maintained by the community. Similarly, in Wikipedia we developed intelligent task routing for editing tasks. Users liked the flexibility and power of the Wiki, which allowed nearly every aspect of the system to be user-editable. However, as developers we found the lack of structured data awkward. Many of our extensions were to support reusable structured data so users' efforts could be focused towards shared goals.

The EthicShare technology team also considered repository-class architectures including:

DSpace:

- Principles (open source, open architecture, modular, scaleable, standards-based, etc.): moderately strong, with distinct drawbacks for modularity and scalability with present release (v.1.4)
- Community interface: weak; end-user interface is generally lacking here
- Momentum: moderately strong for institutional repository aims; for other intents, the rigidity of the architecture has stifled creative development paths

Fedora:

- Principles (open source, open architecture, modular, scaleable, standards-based, etc.): moderately strong with good modularity, but has a track record in the field of being difficult to implement and iterate
- Community interface: quite possible, but scarce evidence of this in practice; mostly meant to be a powerful framework digital object repository
- Momentum: little but gaining; adoption has been very slow and limited to high-end digital library installations.

For the purposes of EthicShare, the technology team chose to use an approach based on a Drupal content management platform because the best available content management systems already include substantial support for structured data and support a wide variety of user-editable items, including wiki-like fields of arbitrary size within structured data. We chose Drupal primarily because of three key advantages:

- 1) A large user base: Drupal is one of the most widely deployed of the content management platforms.
- 2) An effective plug-in architecture: Drupal has a rich ecosystem of plug-in developers, fulfilling nearly every content management need.
- 3) A quality code-base: we investigated the code quality of several of the open source content management platforms, and felt that Drupal's code would be the easiest to adapt to our needs, and the easiest to distribute to future users.

Further, Drupal is an open source framework with an orientation towards support of online communities and collaboration, and it has a thriving developer community. Drupal was designed to provide “a solid base to extend and implement custom content management solutions.” The

success of that philosophy is seen in the enormous number of third party functional modules have been designed for Drupal's core engine.³

The Drupal framework provides a lightweight core that can be readily extended through custom modules. The architecture is generously multi-platform⁴, and uses common web scripting and database tools.⁵ Drupal supports established and emerging standards, with targeted standards including XHTML and CSS. Strict coding standards keep Drupal's data, logic and presentation separate from each other. As a result, all of Drupal's output is completely controlled by the application's presentation layer, known as the 'theme'. This helps to encourage accessibility assurance, as Drupal is Section 508 and WCAG Priority 1, 2, 3 compliant. The source code of Drupal is freely available under the terms of the GNU General Public License (GPL).

At the same time as the EthicShare project was underway, the University of Minnesota purchased an enterprise Content Management System (CMS) for use throughout the University environment. This process afforded the Libraries a rare opportunity to compare its existing CMS solutions with the large "enterprise class" solutions (IBM, Stellent, Vignette, etc). In the end, the Libraries Drupal installation proved to be more flexible as a locally managed product than the centrally located enterprise system that was selected for the University system, particularly where integration, customization, system features and available extensions were concerned. Considering how favorably it compared against large industry solutions, the Libraries' confidence in the Drupal framework as a viable long-term platform for delivering web content grew considerably.

With its focus on building an online scholarly community, the EthicShare project cleaves even more closely to Drupal's core architectural philosophy, as it is heavily focused upon online community development. So far, Drupal is proving to be very well aligned with the principles and direction of EthicShare. With considered flexibility and choice, the technology team can build around the Drupal core, develop and share modules, and pursue new directions when necessary without having to sacrifice investments already made in the Drupal technology.

From a strategic perspective, Drupal offers the EthicShare project a ready-made core technology framework for the delivery of critical services, one that has been both thoroughly vetted by the demands of a large open source community and has been shaped by core developers to specifically address the needs of online communities and collaborations.

³ Over 1,000 such modules have been made available to the Drupal community at:
<http://drupal.org/project/Modules>

⁴ Drupal was designed from the start to be multi-platform and can be run on Apache or IIS, Unix (Linux, BSD, Solaris), Windows, Mac OS X.

⁵ Components of the Drupal platform are Apache Web Server, MySQL or PostgreSQL (database), and PHP (scripting). EthicShare has extended the application to integrate with other significant components such as Apache Solr.

Drupal's technical philosophy also supports EthicShare's need for long-term flexibility and customization. These principles include an emphasis on modularity and extensibility; quality coding standards; low resource demands for operation; open source; ease of use; and collaboration⁶.

Drupal's own claims are not without a growing body of evidence to support them. The IBM Internet Technology Group, for example, authored a long series of articles on the construction of an online community site through the use of Drupal⁷. They chose Drupal, in part, because they "felt that Drupal provided the right combination of framework and flexibility to break out of the framework when needed to get the job done."⁸ Having completed the exercise, the IBM team concluded that Drupal was well suited to their purposes:

*Drupal has held up well. When we needed new functions, we could usually find an existing module within the contributions. If not, we were able to quickly build our own custom module to extend the functions of our system. This extensibility, found in many open source CMSs, is critical for addressing new problems as they arise.*⁹

Within the context of EthicShare prototype development, Drupal has fulfilled many of the initial requirements (group collaborative spaces, end user repository contribution mechanisms, integration with external search utilities, user networking features, etc.) without requiring a great deal of customization. Where customization was required, Drupal provided a path to do so. We believe that the flexibility we have seen thus far will continue to reveal itself as the project evolves.

With its large body of developers and users, Drupal also presents an opportunity for EthicShare to contribute a model for online scholarly community development as well as a platform to a wide community of developers with implications well beyond the field of ethics. More concretely the EthicShare project would offer features back to Drupal in an effort to encourage developers external to the project to get involved with intention of refining the code and opening up opportunities for other organizations to implement similar projects.

Ultimately, EthicShare's participants recognize the possibility of platform shifts as emerging technologies reveal themselves in the years to come. However, Drupal appears to be well poised to assume a long-term, highly functional role as a provider of community site services to the EthicShare project. It possesses a critical mass of devoted followers; a newly formalized bureaucracy for long-term planning and support; a growing number of high visibility community site projects in its portfolio; and a core

⁶ <http://drupal.org/principles>

⁷

http://www.ibm.com/developerworks/ibm/osource/implement.html?S_TACT=105AGX46&S_CMP=LP

⁸ <http://www-128.ibm.com/developerworks/ibm/library/i-osource1/index.html?ca=drs->

⁹ <http://www.ibm.com/developerworks/ibm/library/i-osource15/>

philosophy that blends well with EthicShare needs. In summary, we found the selection and implementation of Drupal to be an effective and strategic match for EthicShare needs.

2. EthicShare Prototype

Core functionality of the EthicShare prototype includes a robust search engine against a richly indexed bibliographic data set of 50,000+ records¹⁰, integration of OpenURL linking and full-text services (and use of the OCLC OpenURL resolving registry), document repository deposit and retrieval, customized view of the Google Co-op web portal technology, and an integrated suite of personalization, current awareness, end-user contribution, and social networking options tailored to the needs and interests of the EthicShare research community.

Initial social networking features incorporated into the prototype include the ability to establish interest groups and associate them with specific citations or citation collections, annotation, social bookmarking, navigation tracking and sharing, citation and document posting, and blogging. The integration of an end-user tagging feature, which has been identified as potentially high value to this community, is under investigation. Functional requirements have been (and will continue to be) driven by community assessment data and highly iterative design-release-test cycles involving participating scholars. Core functional requirement priorities were based on survey and site visit findings, as well as the input of UMN EthicShare team members and other EthicShare partners. In order to adequately consider all research and development opportunities, EthicShare partners agreed that the technology team should pursue promising features despite lower ratings by bioethics scholars surveyed. The difficulty ratings were determined by the technical team, and denote the difficulty to implement technically. (See *Appendix C: Core Functional Requirements for EthicShare Prototype Development: Priority and Implementation Difficulty Ratings*)

For screenshots and brief descriptions of services available through the user interface, see *Appendix D: EthicShare Prototype Views*.

The EthicShare prototype extends the core Drupal technology with Solr, an open source indexing engine server based on the Lucene Java search technology. The project has extended Drupal's *biblio* module to direct OpenURL requests through OCLC's link resolver registry to accommodate a user's authorized access to his/her own institution's licensed content. The prototype also leverages Drupal's robust personalization environment, which supports individualized views of both content and presentation.

Other Drupal modules, of which many have a strong orientation to social networking services, have been tailored for the prototype. Solr provides high-performance, scalable indexing with desirable features, such as faceted browsing (i.e., one-click filtering). This technology core holds

¹⁰ For the prototype, topically relevant bibliographic records were extracted from the publicly available PubMed database, processed in BibTeX format, indexed by Solr, and ingested into the Drupal EthicShare instance. An additional smaller set of records was harvested from the DSpace-supported Bioethics Digital Library (BEDL) at IUPUI using the OAI metadata harvesting protocol.

tremendous potential not just for exploratory prototypes, but also for richly functional production-level services. For a comprehensive listing of modules integrated into the prototype, to dates, see *Appendix E: Drupal Modules Configured in EthicShare Prototype*.

3. Iterative Feedback for Prototype Functionality

During a May planning meeting, most EthicShare partners, along with Richard Lucier (UVA Scholarly Communication Institute) and Kenny Crews (IUPUI), met and discussed all aspects of the EthicShare project, including the prototype.

After a presentation of the prototype, EthicShare partners had a lengthy discussion of its features, content, and design. Features that enabled sharing resources, collaborating with colleagues, and tagging were especially interesting. Participants were also interested in pursuing innovative ways to blend controlled vocabularies/thesauri terms with user-generated headings. Ratings the quality of a resource was less interesting to those present at the meeting.

Participants agreed that the design of the site was an important aspect of its effectiveness and appeal as a community resource, both technologically (it should be flexible and intuitive) and philosophically (all communications should be as transparent as possible, and users should be able how their actions are used, recorded, or shared). As a repository and resource suited to the bioethics academic community, EthicShare should be a focused research site that promotes exchange and highlights cutting edge research. EthicShare should maximize technology and open-access principles to make the repository as accessible and open as possible.

Other early feedback from EthicShare partners who explored the prototype online and responded to a brief site survey suggest that the most interesting features of the site were:

- Literature search limits (searching by document type, date, journal title, author, etc.)
- Group discussions and collaborative research capacity
- Google Co-Op: Customized Google search shaped by EthicShare users
- Site-specific bookmarks
- Events/announcement posting

Features that were least interesting to EthicShare partners were:

- Maintaining a personal blog on the site

Gathering and responding to user and community feedback is an important dimension of EthicShare's technological development. The technology team will depend on usability testing, tracked user behaviors, and other assessment tools to further hone EthicShare site features, design, content, and functionalities.

4. EthicShare Project Trajectories

The implementation, release, and review of the EthicShare prototype have helped us to identify

possible future areas of development to be pursued in the future implementation phases of the project. These include ingest of published data; systematic harvesting and data acquisition; data description and indexing; faceted navigation; identity management and OpenURL; social navigation and community engagement; privacy; community-driven design, and technology review.

Ingest of published data: EthicShare's distributed information corpus is of critical value to bioethics scholars and a catalyst for strengthening the scholarly community. Depth, currency, and ongoing relevance are key attributes that bioethics scholars value, and a successful online community must exploit multiple inputs to achieve an effective selection and acquisition of bibliographic and full-text data. This direction of development forces the project to move towards innovative approaches to optimize and balance the various methods by which to enter information into the corpus: community-contributed data, professionally contributed records (e.g., Kennedy Institute for Ethics staff), and algorithmically driven approaches to systematically acquiring content, whether through harvesting, batch loading, or feeds (discussed below). There are significant metadata issues related to normalization, remediation, and de-duplication, as well as intellectual property issues (including copyright and licensing) that need to be investigated and settled. Determining the requirements of scale and dimensionality for the information base are also key questions as the more systematic indexing techniques under consideration are strengthened through the power of content aggregation and engagement of larger numbers of active users. Finally, how "collection development" policies and guidelines are governed and then managed in a community-stewarded environment needs to be determined.

Systematic acquisition and harvesting of data by community and crawlers: The potential is great for record harvesting, crawlers, feeds, and other modes of automated or semi-automated ingest, and as community-based maintenance techniques, to contribute to the long-term growth of the EthicShare repository. To enable community maintenance, computer tools will be developed to support the community in developing information structures that help other members of the community find the information they seek. These tools will include mechanisms for structuring, indexing, and discussing parts of the repository.

On harvesting: Currently there is no centralized repository for capturing the broader bioethics literature, especially that published in the fields of law and philosophy broadly construed. Computer tools will be developed that automatically crawl publishers' web sites to ingest freely usable material, such as the titles and authors of newly published articles. These materials will be made visible to the community through the community maintenance tools, to allow specially designated community members to select which of the new materials belong in the repository, and which are out-of-scope. Rights relationships will be sought with publishers to enable the inclusion of additional materials such as abstracts in the repository, under the argument that inclusion of the richer material will make it more likely that users of EthicShare will discover articles they wish to access from the publishers, increasing the value of the publishers' catalogs. Over time, these crawlers may become outdated with respect to the publishers' web sites. We will explore two techniques to ameliorate this challenge. First, the crawlers will be developed upon a common architecture, and will be made open source, so that members of the community with technical skills can maintain or extend them. Second, special ingest methods will be developed for industry standard catalog export techniques. As EthicShare gains critical mass,

publishers will be encouraged to export their catalog in one of these formats, which can be ingested in sustainable ways.

Description, categorization, and indexing: Economic pressures suggest that the exploration of effective alternatives to professionally conducted description, coding, and classification of bibliographic records (to enhance navigation and access) is necessary. The EthicShare pilot will explore the use of end-user tagging, and will leverage research on computationally derived taxonomy with an eye toward providing tags for users to select in order to help the community develop an effective ontology. A key opportunity is to integrate end-user tagging with taxonomies designed by experts (e.g., the Kennedy Institute's *Bioethics Thesaurus*). Pursuing this approach will enable comparisons of the cost and effectiveness of user-generated tagging vs. staff-generated description. We also hope to understand of the potential complementarity of these approaches in benefiting a scholarly user community. Semantic indexing techniques, which seek to find useful patterns in unstructured data, may also be explored.

Faceted navigation: The prototype revealed the power, flexibility, and performance of integrating the Solr indexing engine and faceted browsing services into the EthicShare navigation scheme. What, then, is the impact of faceted navigation on the discovery practices of scholars? How should facets be identified for a specialized community and how might facets affect the types of data that are leveraged through indexing? For example, might data contributed by users (e.g., a specified individual's tags, ratings, or commentary) be indexed as a navigation option? Also, does faceting lend greater effectiveness in the navigation of multidisciplinary literature? What might this suggest for the amassing of larger data aggregations and the broadening of discovery without the loss of efficiency?

Improving identity management aspects of the OpenURL registry: The EthicShare prototype demonstrates a real-world, large-scale interest in enhancing and improving the current framework of network registries (i.e., in this case the OCLC OpenURL Registry). This will be imperative for multi-institutional collaborations involving institutionally defined authorizations for access to licensed content and other local services (e.g., inter-library loan). Specifically, the project needs to solve the problem of how users at participating *EthicShare* institutions can access full-text resources when not on their campus network, and thus not conducting their access through a recognized and institutionally registered IP address. Further, there needs to be a sensible and useful response from the registry to users who hold no entitlements to access content through an institutional arrangement (e.g., such a request could be resolved through a search in worldcat.org). The project wishes to engage with OCLC in exploring, developing, and testing a scaleable solution to this identity discernment issue. Shibboleth and OpenID are among the avenues to investigate in enabling authorized users full access to the resources to which they're entitled via OpenURL/CrossRef regardless of their location on the network.

Social navigation applications for improved discovery and scholarly community engagement and contribution: Social navigation technologies to provide users with recommendation services help users navigate complex information spaces. A future EthicShare pilot will leverage pioneering expertise residing at the University of Minnesota in this area, especially computer scientist John Riedl's interdisciplinary work on how social web applications

foster community and engagement, drawing on models from the fields of economics and social psychology.

Privacy and disclosure: The opportunities for managing and sharing personal information in a community environment raise numerous policy and implementation questions. Determining policy regarding the point(s) at which informed consent is required. Clarity about when, how, and the extent to which personally held information is disclosed through various functions and services offered is also a significant obligation of the system. Engaging the community and the partnerships' institutions in these discussion and formulation of policy and controls is critical.

Community-driven design model: Developing a community-driven and community-stewarded online environment in a multi-institutional and multi-disciplinary context calls for innovative and engaging processes. Establishing an effective and accepted governance model for technology (and other) decisions is a fundamental element. Communications and participation in the processes that inform decision-making is clearly another aspect. A highly iterative design process is essential to the delivery of meaningful services to individual scholars and the community as a whole. This process must involve a very agile development environment where design iterations become quick responses to user inputs, and where there are continuous feedback channels and assessment efforts. We will employ usability testing, use data analysis, and intelligent feedback mechanisms that capture explicit and implicit user inputs.

Technology review panel: As EthicShare moves to the next stage of planning, an external review panel will be formed to review available data on the prototype and provide input to the design team architectural issues relating to performance, scaling, interoperability, and pertinent trends in the virtual community development arena. Seeing possibilities for converging projects, this panel may provide the opportunity to formalize a developers' interest group in this area of technology development.

VI. Conclusions from Planning Grant Activities

As a result of the efforts in content and collection analysis, user requirements and community assessment, and technology development, and in light of emerging trends online and in scholarly communities, EthicShare partners conclude that:

1. There is a demonstrated need for a robust discovery environment that serves bioethics scholars and that employs easy-to-use interactive features to facilitate interdisciplinary and multi-disciplinary scholarship, collaborative research, and community building and involvement. While access to content is a core functionality, the planning addressed the necessary balance between developing a repository and mechanisms to federate distributed repositories and publisher services to create a powerful gateway to high quality content.

Such a site has tremendous potential to move a field forward and to develop an active and engaged community of scholars who thus far have had limited exposure to newer technologies and web service functionalities.

2. It is clear that the trajectory towards a community-sustained environment requires ample effort by and commitment from the partner institutions, as well as hybrid models of stewardship and contribution between professional and scholarly contributors.

As a whole, bioethics scholars are not yet motivated to take on the responsibility of building, maintaining, and growing an online scholarly environment. In order to engage the bioethics community in the activities of adding to and using the environment, the EthicShare partners would need to promote and demonstrate the benefits of a community-sustained environment. Additional partners might also allow the EthicShare team to broaden the site's potential range. The long-term responsibility for the site may be a combination of institutional support and community participation, and would require commitment from the site's host institution for sustained development, improvement, and longevity.

3. There is a need for new models of indexing, classification, and ingest of content that rely more heavily on efficient and flexible semi-automated processes, and less on the expensive and labor-intensive efforts of library or other staff.

At the same time, the human component of vetting content and quality, as well as interpreting and producing scholarship, is a critical aspect of the scholarly process for bioethics scholars. Can the behaviors of scholars— through tagging, bookmarking, commenting, and sharing—help serve as valuable indices of scholarly content and quality that can be leveraged in an automated way? Further, as we develop systems that can “learn” from user behavior and draw on that behavioral intelligence to create useful services, how do we measure *EthicShare*'s value and effectiveness over time?

4. The EthicShare planning phase was successful in identifying the content and collection requirements of a valuable discovery environment, as well as the needs and preferences of the site's targeted user community.

The planning phase also allowed the EthicShare team to make substantial progress in the technological development, prototyping, and establishment of an iterative design process for the potential site. These efforts lead us to ask new questions about the future of libraries, scholarly communities, and their interactions online. A key question motivating EthicShare is: How do we move away from the classic model of the high cost, labor intensive library (or digital library?) to a more agile and flexible cooperative model in a way that privileges both human craftsmanship and the value of distributed technologies and virtual user behaviors?

VII. Framework for EthicShare's Proof of Concept

The goals of EthicShare go beyond the creation of a repository model and towards a robust discovery environment that models and supports new forms of collaboration. A successful collaboration environment will require a model for sustainability. The

analysis of these issues during the planning phase has led us to the conclusion that sustainability must be viewed as both sustainability of the technology environment (i.e., the repository and associated functions/services) and sustainability of the collaborative activity (i.e., an engaged community with commitment to participate and contribute to the environment). Our iterative implementation of EthicShare will focus on the issue of sustaining and enhancing the collaboration. We expect that the University of Minnesota would ultimately assume responsibility for hosting the technology environment once established.

The University of Minnesota Center for Bioethics, University of Minnesota Libraries, and *EthicShare* partners plan to document the next stages of EthicShare's development in a proposal for future support. A future pilot would include three areas of investment:

1. Building an online discovery environment and community for bioethics scholars. This virtual community centers around a core collection of scholarly materials, governmental and other policy documents, materials from the popular press, and multimedia resources, as well as a powerful index of aggregated citation and metadata that allows EthicShare scholars easy access to non-locally hosted resources. The environment is fueled by open source technological development and robust functionality that will allow scholars to interact with resources and each other in new ways, based on what has been learned and developed during the EthicShare planning phase.

Tools for tagging, bookmarking, commenting on, and sharing formal and in-process resources will be integrated into the site repository. It will also facilitate collaboration among groups of users based on topic interests, projects, joint research and writing, and the like. The repository will initially be populated by existing data sets created by Georgetown KIE.

2. Developing an open architecture technological platform with intent for documented open source release, methods of semi-automated data population and ingest of content, and a model of community and institutional engagement that is sustainable over time, and that can be used by scholars in other disciplinary and interdisciplinary areas.
3. Designing mechanisms to support ongoing development and population of the repository, thus fostering broad bioethics community involvement in the governance and maintenance of the community site. We will also investigate the incentives and motivations that drive community participation, as well as the rules and regulations that engender scholars' confidence and participation over time.

Deliverables include:

- An operational discovery environment for the scholarly community of bioethics, with an initial focus on Georgetown data
- An open source, OpenURL extensible platform maintained by web crawlers and automated indexing that is sustainable over time

- Best practices documentation of all technology, copyright, and digitization efforts
- A model to enable scholars to engage in their scholarly work and, through the use of interactive web features and functionalities, participate in and contribute to a virtual community of bioethics scholars

