

Exploiting the opportunities of the maturing digital age: the first twenty years of the Scholarly Communications Program of the Andrew W. Mellon Foundation.

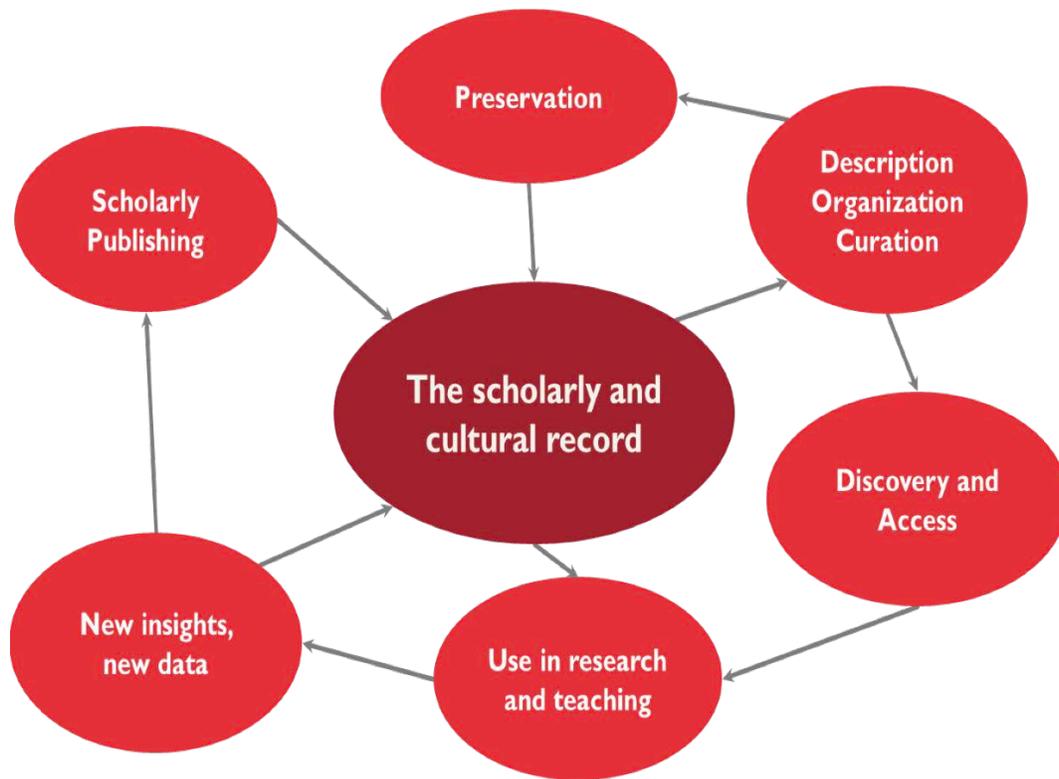
By Michael A. Keller, University Librarian, Stanford University, © April 2019.

Introduction

For the past 20 years, the Andrew W. Mellon's Scholarly Communications Program has seized the advantage of on-rushing developments to significantly structure, manage, and sustain innovation in arts and humanities research and teaching. In this time, the Scholarly Communications Program, under the leadership of a single director, has supported and stimulated projects exploiting digital technologies and network effects in the areas of academic publishing, preservation of resources for scholarly practices, and library services. Many of these projects continue to benefit scholars and students, many have been superseded or suspended, and some have been bell-weather for new pathways for research and teaching in the humanities. Given the hundreds of projects funded through the Scholarly Communications Program, amounting to hundreds of millions of dollars, a deep recitation and detailed history of the investments made is not attempted in this brief presentation. Rather, a few examples of the sorts of projects and their ramifications as well as some observations on the investment principles and analytical processes of the Program's staff are presented.

Proposals based on active imaginations matching needs to opportunities, and prescient choices by the Scholarly Communications Program (hereafter “the Program”), have led to remarkable changes in the expansion of research resources and methodologies available to humanist scholars. The role of the Program staff in stimulating follow-on projects to earlier ones has been important. Inherent in the consideration of initial proposals, the Program’s peer reviewers’ suggestions, and the important sharpening of descriptions, methods, and finances has been the critical judgment of the Program’s leadership, advisors, and staff. Although there has not been anything like a grand strategy for applying the rising tide of technology to all the scholarly boats over these years, the mix of proposals from communities of practice in the target areas of publishing, preservation, and library services combined with broad awareness of needs and aspirations in the Mellon Foundation has produced and augmented flows of new possibilities for scholarly practice and wider appreciation of those possibilities for the current generation and their successors. A few examples of the vital interaction of players and their proposals, communities of practice, and the Foundation’s leadership in the Scholarly Communications Program will make clear this remarkable productivity amid the chaos of these early decades of the digital age. This author’s involvement with many of these projects has provided an exemplary education regarding the nuance and complexities of a successful grant program.

The following graphic represents the scope and framework of scholarly communication, starting with source material, then proceeding to the main functions supporting scholarship. Included in the diagram is the essential element of reporting the results of those endeavors for present and future generations of disciplinary specialists and the intermediaries – librarians, archivists, and publishers – who make both source materials and the results of the work of scholars available to current and future generations.



Scholarly Resources

A few examples of projects funded by the Program in the area of medieval studies may suffice to illustrate the extent and significance of that support. From the early years of the first decade of the 21st-century, digitization proposals have led to on-going programs that make otherwise difficult to access manuscript resources available as high resolution digital avatars, often with enhancements -- such as improved descriptions and metadata, appended bibliographies of related material, and searchable graphical elements - that together reduce effort in discovering, studying, collaboration, and citations of source documents.

The Matthew Parker Online Library, aka Parker on the Web, is an example. Matthew Parker (1504-1575), Elizabeth I's first Archbishop of Canterbury, was charged by his queen to complete the conversion of the Church of England from Roman Catholicism to Protestantism. He did so, and that accomplishment, he collected over 530 manuscripts with the intention of showing that the Church of England had a separate lineage from the Church of Rome. Upon his death, that collection was bequeathed to his college, Corpus Christi College at Cambridge, where he had been student, fellow and master, with the proviso that should an annual audit of the collection uncover the loss of three or more manuscripts, the collection would pass to other Cambridge colleges. The fellows of Corpus Christi decided then and held to that decision for over four centuries, severely limiting access to manuscript books in the collection. No audit ever found missing volumes. However, once the Montague Rhodes James catalog of the collection was published in 1912, there were increasing requests for access, most often denied.

Then in 2002, the master and fellows of the college, following a suggestion by an alumnus, by then a successful executive in a Silicon Valley company, decided that the affordances of digitizing the Parker collection for scholars whose access was formerly constrained were highly attractive. Collaborating with the Stanford Libraries and with extensive support from the Program made possible high resolution digitization of the manuscripts, after they were carefully dis-bound, relieving the constraints of too tight binding of most of them in the 19th-century, along with vastly improved descriptions of each manuscript made possible by a team of well-qualified scholars, and, for each manuscript, bibliographies of modern editions, critical commentary, and related scholarly commentary. This concatenation of features was realized in the Parker Online Library (<http://parker.stanford.edu>), now in its second version and now made freely available.

The Parker project was an early manifestation of the provision of high-resolution digital copies of *unica* on the Web and it had, still has, ripples of stimulating consequences for medieval and more general manuscript studies. However, thanks to the forward-thinking leadership of the Program that suggested a structured investigation by the Stanford and Corpus Christi team with medieval manuscript scholars to forecast uses and desired functions of research and scholarly communication postulated on the examples of the digital versions of the Parker manuscripts. Beginning roughly in 2007, conversations, focus groups, and a couple of conferences, with the resulting Shared Canvas data model (<https://iiif.io/model/shared-canvas/1.0/>) becoming visible in

2013, enough conceptualizing had occurred to begin development of the International Image Interoperability Framework (IIIF and see <http://iiif.io>), which with well equipped “viewers” have enabled free access, by streaming from institutional servers to individual scholar’s computers, to thousands of digitized resources, many of the ancient, medieval, and early manuscripts from libraries and museums and many other genres of digitized source materials, such as newspapers, flat artwork, audio and video files, sculpture in 3D, and archival documents, among others.

The scholarly functions arising from the insights and desiderata of those scholars consulted from 2007 and more specialists joining the constant dialog between them and the information technologists contributing their programming and organizational skills to new developments are varied. The practice of listening to scholars as they specify functions they desire to use in their research and teaching, as in the case of Shared Canvas, has led to the development of several “viewers”, among them Mirador (<http://projectmirador.org>). Mirador is an open-source, web-based multiple frame viewer that supports zoom-pan-rotate functionality, ability to display and compare simple images.¹ Re-assembling manuscripts and even manuscript pages that were disassembled in prior times, annotating pages and sections of pages or aspects of works of art (both flat and 3D) for further contemplation or sharing with colleagues, comparison of numerous manuscripts and other pages, making extracts for research

¹ For a presentation demonstrating the power of the combination of IIIF and the Mirador “viewer”, please see: <http://web.stanford.edu/group/dmstech/cgi-bin/wordpress/tag/mirador/>, a presentation by Benjamin Albritton.

and communications purposes, and enabling transcription and translation virtually layered on a digital page are among the functions now supported and heavily used.

More IIF and Mirador developments for images arising from STEM, arts, humanities scholarship are underway, through the volunteer efforts of dozens of volunteers. From a growing community now over 300 cultural heritage institutions worldwide, perhaps up to 2 billion images of tens of thousands of source materials are available without paywalls for research and teaching. That these remarkable attainments occurred and are still occurring is the result of the foresight and gentle influence of the Program's leadership as well as the funding the Program supplied.²

Related projects funded and shaped by the Program, such as E-Codices, the Virtual Manuscript Library of Switzerland, the Roman de la Rose project at Johns Hopkins, the papyri and cuneiform projects, and the Electronic Enlightenment among many others are changing humanities scholarship and indirectly benefitting archival and artefactual practices, including exhibitions and other services to the public.

Publishing humanities and arts monographs

² An excellent summary of the Parker, Roman de la Rose, and E-Codices projects and the "shared canvas" notions is presented in: Waters, Donald J., "Digital humanities and the changing ecology of scholarly communications" in **International Journal of Humanities and Arts Computing**, v. 7, March 2013: 17-19.

The domination of a few for profit publishers whose foci are all disciplines in the global academy of higher education and research has resulted over the past 30+ years in reduced spending by university and college libraries on publications in the arts, humanities, and social sciences. Particularly ill affected are most university presses in North American and the British Commonwealth countries. Most university libraries spend 80% of their allocations for acquisition of products of for profit publishers, most of which focus principally on STEM disciplines' journals, driving out spending on humanities publications. All but a few university presses have experienced dramatic reductions in sales to university libraries as a result.

In recent years, the Program has invested in numerous projects, some of them consortial, to make more efficient through shared resources and easily adapted new practices humanities monograph publishing. Other projects benefitting from the Program's attention are intending to exploit some of the "natural" characteristics of the Web to make more common updates to e-articles and e-books after initial publication and to "publish" peer reviewed interactive scholarly works, manifestations often by teams of scholars and their information technologist partners that cannot appear as the ribbon of texts that have typified in printed and e-forms written communication for millennia. The sum of the Program's investments intend to make more viable in business senses university presses, not for profit agencies of over 100 universities in North America alone, in the work of publishing important scholarship that for profit publishers might never undertake on the basis of limited marketplaces and thus limited

or non-existent potential for return on investments, e.g. profits. Examples of these projects are the Stanford Press' Interactive Scholarly Works program (see <https://www.sup.org/digital/>) and the University of Michigan Press (digital culture program, see <http://www.digitalculture.org> and its EBook Collection, see <https://www.press.umich.edu/librarians>).

There is a related "free-rider" problem arising from the fact that all appointments, promotions, and tenure decisions depend upon documented contributions to the commonwealth of knowledge by aspirant scholars. With only about 100 university presses available to serve humanities scholars with peer reviewed publications programs employed among perhaps 1,000 institutions of higher education, there are dependents on a relatively small number of presses that receive little or no contributions to their fiscal stability beyond publishable monographs, each one a small cost center. The Mellon Foundation has been aware of and sensitive to this problem, attempting in these grants to university presses to begin to address that problem, but it persists. More attention is needed here, not necessarily by the Mellon Foundation's Scholarly Communication Program.

Preservation of the scholarly and cultural record

A direct benefit of the Program's investment in numerous digitization projects, each with features and functions beyond merely providing high resolution images of

scholarly resources, is that of reducing the need for scholars to actually visit and physically touch unique originals. The assertion that 95% of scholars' requirements in the use of originals can be satisfied by good, high resolution digital copies is apparently accurate. Thus the wear on the source manuscripts, newspapers, and the like is reduced, with the desired and hoped for result of prolonging the life of at least the more ancient cultural resources. Nevertheless, for those older materials and the more fragile of modern resources, such as newspapers and books printed on acid laden paper, digitization projects make possible digital preservation of the essential content of such genres. The challenges of digital preservation, the digital archive, are rapid changes in related technologies such as operating systems, applications, hardware, data formats, and single points of failure occurring in dependence upon a single archival "solution" that become apparent only after a failure has occurred.

Complementary investments by the Program in Portico (see <https://www.portico.org>) and LOCKSS (see <https://www.lockss.org/>) could address the single point of failure possibility, but also serve important access functions for legions of scholars and students. In addition, the CLOCKSS (see <https://clockss.org>) alliance of libraries and publishers that makes possible both digital preservation in original forms and formats of scholarly publications and access to those publications in the event of disruptions due to economic, environmental, political, and technological failures through a network of 12 CLOCKSS servers around the world.

The Artstor service is a notable example of the Program's investments that principally provides access to and important functions in support of research and teaching for digitized art resources from over 300 institutions based around the world, replacing numerous art slides collections in colleges and universities, many populated by sub-standard images of art works copied from books. Artstor, also, should be regarded as a preservation service for access over the years and decades to high resolution images fragile artistic works that are subject to the privations of natural and unnatural disasters as well as thefts and damage by miscreants. Artstor's service is a hedge against losses or damage to originals.

More obvious services in support of saving a few originals from collective consortial library collections of published works are the WEST and EAST Storage Trusts (see <https://www.cdlib.org/services/west/> and <https://eastlibraries.org>) both stimulated and supported in development by the Program and both members of the Partnership for Shared Book Collections, itself an extension of the Program's initiatives. While preserving a few copies of original publications, these agencies make possible the more difficult work of coordinated collection development among research libraries, on the shoals of which many well intentioned earlier attempts have foundered.

Regranting Programs

In 2008, with the support of The Andrew W. Mellon Foundation through its Scholarly Communications Program, the Council on Library and Information Resources began investing in innovative and efficient approaches to describing rare collections through *Cataloging Hidden Special Collections and Archives: Building a New Research Environment*. The impetus for this program was a widely shared acknowledgement of the need to rethink cataloging methods toward greater standardization, efficiency, and scholarly impact. The urgency of this need, explored through a decade of research beginning in the 1990s, compelled the Foundation and CLIR to create a national program that would fund the creation of records for unique cultural heritage that would be available through the internet and the Web.

By 2014, when the program announced its final awards of this first regranting program, CLIR had provided 129 cataloging grants totaling over \$27.5 million to a wide variety of institutions in the United States and Canada. These grants have made it possible for scholars, students, and the general public to find and use vast quantities of diverse materials that were not previously discoverable online. Collections exposed through the program include rare books and serials, manuscripts, archives of all kinds, architectural drawings, photographs, artworks, maps and almanacs, textiles, audio and audiovisual recordings, ephemera, and much more.

Cataloging Hidden Special Collections and Archives was conceived as a kind of incubator which, if successful, would contribute to a shared understanding within professional

and academic communities that while all special materials are by nature local, creating standardized descriptions of them that can be accessible anywhere and anytime alongside related but dispersed collections is an exceptionally important goal. Grant recipients have generously shared their experiences, resources, and lessons learned with others through the Hidden Collections Registry, listing over 2,700 entries for collections newly made visible through these CLIR Programs, through program symposia in 2010 and 2015, and various publications, blog entries, and presentations at scholarly conferences.

From 2015 to 2017, CLIR directed a regranting program providing over about \$4 million per year to numerous proposals from a range of institutions in the Digitizing Hidden Collections of Libraries and Archives. The source of the financial support for this effort was again the Scholarly Communications Program of the Mellon Foundation.

From 2015 to 2017, a successor program of CLIR, *Digitizing Hidden Special Collections* has awarded approximately \$4 million annually to institutions holding collections of high value for research, teaching, and learning. *Digitizing Hidden Special Collections and Archives: Enabling New Scholarship through Increasing Access to Unique Materials* is built upon the model of CLIR's *Cataloging Hidden Special Collections and Archives* program (2008-2014).

In January 2018, a new, related program was announced: CLIR's Digitizing Hidden Collections of Libraries and Archives (DHC) initiative based on a generous, \$13.2 million three-year grant from The Andrew W. Mellon Foundation.

Like its predecessor program Cataloging Hidden Special Collections and Archives, DHC supports projects that make digitized resources easily discoverable and accessible alongside related materials held within the home institution as well as by other collecting institutions. The program's core values include collaboration, by promoting strategic partnerships rather than duplication of capacity and effort; sustainability, by promoting best practices for ensuring the long-term availability and discoverability of digital files created through digitization; and openness, by ensuring that digitized content will be made available to the public as easily and completely as possible, given ethical and legal constraints. Committing to these values requires technological, cultural, and behavioral strategies to encourage collaboration, sustainability of effort, and content preservation and reusability over time.

Finally, the *Recordings at Risk*, a national regrating program administered by CLIR to support the preservation of rare and unique audio, audiovisual, and other time-based media of high scholarly value through digital reformatting has been generously funded by Scholarly Communications Program since January 2017. The program will run nine competitions from January 2017 to April 2021 and will award a total of \$4.5 million.

Awards will cover costs up to certain limits of preservation reformatting for fragile

and/or obsolete time-based media content by qualified external service providers. Eligible media may include, but are not necessarily limited to, magnetic audio and video tape, grooved discs, wax cylinders, wire recordings, and film (with or without sound).

Recordings at Risk encourages professionals who may be constrained by limited resources and/or technical expertise to take action against the threats of degradation and obsolescence. The program aims to help institutions identify priorities and develop practical strategies for digital reformatting, build relationships with partners, and raise awareness of best practices.³

Such regranting programs sponsored by the Mellon Foundations like those cited above as well as other ones providing support for Ph.D. dissertation research in original sources support and post-doctoral fellowships providing career experiences for potential subject specialists in academic libraries and for those interested in data curation roles in academic libraries have placed 190 Ph.D. candidates in 81 host institutions and 168 Post Docs in the US as well as 22 in Canada and elsewhere. Similar programs have been given Mellon support in the ACLS for digital innovation fellowships and digital extension grants, supporting digitally based research projects in all disciplines of the humanities and humanities-related social sciences. One

³ The paragraphs concerning the CLIR regranting programs have been freely adapted from reports in the CLIR website (see <http://clir.org>).

interpretation of regranteeing programs widen the scope of possible recipients of the Program's funding through a few grants to trusted institutions, whose staff and reviewers are supplementing the staff of the Program.

Tools

The Program has invested in a dozen digital tools or toolkits that were meant to take advantage of the growing capabilities of the World Wide Web in supporting aspects of scholarly communication. Among the dozen or so applications that have been supported, either wholly or in concert with other funders, are several that deserve recognition and demonstrate the range of thinking in this category: Fedora; DSpace; CollectionSpace; ArchivesSpace; Zotero; BitCurator; and Editoria. All of these are Open Source and widely implemented, each with an associated development community.

Fedora is a robust, modular, open source repository system for the management and dissemination of digital content. It is especially suited for digital libraries and archives, both for access and preservation. It is also used to provide specialized access to very large and complex digital collections of historic and cultural materials as well as scientific data. Fedora has a worldwide installed user base that includes academic and cultural heritage organizations, universities, research institutions, university libraries, national libraries, and government agencies. The Fedora community is supported by the stewardship of the DuraSpace organization.

DSpace is a web application devised by the HP Labs and MIT that consists of a database, storage manager and front end web interface. The architecture includes a specific data model with configurable metadata schemas, workflows and browse/search functionality. It is easily implemented and used widely as a digital repository for libraries and archives.

CollectionsSpace is an open-source, web-based collections management software for museums and similar institutions. It is a platform for collections information management that: supports traditional collections management activities; enables the integration of emergent and dynamic new technologies into the information ecologies of museums; and is an effective and affordable alternative to one-off applications developed in-house and proprietary offerings.

ArchivesSpace is a web application for managing archives information. The application is designed to support core functions in archives administration such as accessioning; description and arrangement of processed materials including analog, hybrid, and born-digital content; management of authorities (agents and subjects) and rights; and reference service. The application supports collection management through collection management records, tracking of events, and a growing number of administrative

reports. The application also functions as a metadata authoring tool, enabling the generation of EAD, MARCXML, MODS, Dublin Core, and METS formatted data.

Zotero is a free and open source application employed by many scholars to manage bibliographic data and related research materials (such as PDF files). Notable features include web browser integration, online syncing, generation of in-text citations, footnotes and bibliographies, as well as integration with some word processors such as Microsoft Word, LibreOffice Writer, and Google Docs. It is produced by the Center for History and New Media at George Mason University

BitCurator intends to support digital forensics in libraries, archives, and museums in order to help ensure the longevity and reliability of the cultural, scientific, and historical record. Functions supported by BitCurator are: acquisition and processing of digital collections; maintaining the original order of digital materials; surveying the extent and composition of digital collections; redacting personally identifiable information; extracting technical and preservation metadata; packaging digital materials for archival storage. BitCurator was developed by the University of Maryland Institute for Technologies in the Humanities (MITH).

Editoria is a digital workflow supporting production of books including styling, copyediting, author review, and proofreading in a browser-based system. It supports collaboration among editors and book production specialists in real time. It is an

automated “type setting” application enabling use of templates according to a publisher’s desired style(s). It outputs files in several formats.

Standards

The Open Archives Initiative and particularly its Protocol for Metadata Harvesting (OAI-PMH) is a protocol funded by the Program in 2002. It is in nearly universal usage in the Open Access publishing movement and is essential in discovering through applications that make use of it. The Open Archives Initiative has developed interoperability standards that aim to facilitate the efficient dissemination of content. The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication. The fundamental technological framework and standards that are developing to support this work are independent of the both the type of content offered and the economic mechanisms surrounding that content. These promise to have much broader relevance in opening up access to a range of digital materials.

Open Archives Initiative’s Object Reuse and Exchange (OAI-ORE) defines standards for the description and exchange of aggregations of Web resources. These aggregations, sometimes called compound digital objects, may combine distributed resources with multiple media types including text, images, data, and video. The goal of these standards is to expose the rich content in these aggregations to applications that

support authoring, deposit, exchange, visualization, reuse, and preservation. Although a motivating use case for the work is the changing nature of scholarship and scholarly communication, and the need for cyberinfrastructure to support that scholarship, the intent of the effort is to develop standards that generalize across all web-based information including the increasing popular social networks of “web 2.0”.⁴

Conclusion

The remarkable history of the first twenty years of the Program’s decisions, threads of development, and stimulation of both initial and succeeding proposals needs to be told from an elevated perspective and respectful of confidentiality. Among the sterling qualities exhibited consistently over these years is the acumen, patience, and creativity of the leadership of the Program. Since 1999, informal conversations, small and large, well-focused conferences convened by the Program’s leadership, preliminary prospectuses submitted by hopeful principal investigators, and thoughtful appreciation of the connections among seemingly disparate developments have resulted in remarkable, if varied, progress. Some proposers have spoken so appreciatively of the chance to have a proposal, an idea even, considered by the Program that the resulting refinement of proposals, described by some as “Waters torture” or “whack-a-mole” or

⁴ The paragraphs quote directly from the OAI webpages; see <https://www.openarchives.org>.

“surgery without anesthesia” have been experienced perhaps not joyfully, but certainly with equanimity.

The future responsible parties of the Scholarly Communications Program of the Andrew W. Mellon Foundation in support of humanities scholarship and teaching along with the related fields of practice in libraries, archives, and to some extent museums, would be wise to continue the practices and the exercise of sometimes risk-taking judgment that has characterized this first and most productive twenty years, almost a generation. Environments involving cloud storage and virtual engines in the clouds, artificial intelligence plus machine learning plus neural networking, and exploitation of linked open data for discovery, inventorying, business functions, and reconciliation of RDF triples vastly expanding imaginations of intellectuals, scholars, and students lie in the immediate future. However, the constant flow of new possibilities, new technologies, new coordination of effort among scholars with ideas of somehow utilizing those possibilities for research and teaching, the community of academic service providers who are also innovators (librarians, archivists, museum professionals) with ideas of their own for clientele writ large, publishers with their own notions about communicating advances, and finally thought and action leaders of scholarly NGOs with horizon-wide vision and foresight into the distant futures will tell the tale for the more distant experiments the Program might support. There is no doubt that the benefits and meaningful support by the Mellon Foundation’s several programs and particularly this one devoted to the broad definition of Scholarly Communication

has conveyed numerous and highly significant benefits to the fields of the humanities cited in this brief presentation.