

Monograph Output of American University Presses, 2009-2013

A Report Prepared for the Andrew W. Mellon Foundation

February 10, 2017

Prepared by:

Joseph J. Esposito

Karen Barch

Table of Contents

I. Summary	3
II. Introduction	3
III. Methodology	8
IV. Phase One: The Pilot Project	13
V. Phase Two: The Core Database	20
VI. Sidebar: Output for All Titles (not just monographs)	42
VII. Phase Three: An Experiment with OCLC	44
VIII. Discussion	47

I. Summary

This is a report on a linked series of projects to study the output of American university presses in the time period 2009-2013. A large amount of data was collected from the presses themselves; it was then aggregated and analyzed. This analysis yielded reports for the internal use of the Andrew W. Mellon Foundation ("Mellon"), the sponsor of this research. Subsequently we prepared individual reports for the participating presses. We also experimented on a project with OCLC¹ in which we attempted to map (with only partial results) the database of press titles to the holdings in WorldCat. This report (the "public report") represents an edited version of the various private reports submitted to Mellon and the presses. The primary thrust of the project can be found in **Section V: Phase Two: The Core Database.**

II. Introduction

The genesis of this project came about from conversations in 2013-2014 within the academy concerning the condition and future of university press publishing and, in particular, the academic monograph, which is an essential component of scholarly communications and which also plays a role in the credentialing of scholars. There was at that time only anecdotal information on how many monographs university presses published and, hence, no obvious way to measure the size and scope of university presses and the certification system they help to support ("Do university presses publish too many books? Too few? Do they support some fields more than others?"). In order to get some data to assist in other explorations into the university press world, Mellon asked us to tabulate the output of university presses; that tabulation lies at the center of this report.

This report represents our attempt to answer Mellon's question, a three-part project involving American university press book publishing, which we conducted, with some

¹ For information on OCLC, see <https://www.oclc.org/en/home.html?redirect=true>.

breaks along the way, over three years. The three parts will be taken up in turn in the body of this report; they consist of a pilot project, a study of the output of American university press monographs over the period 2009-2013 (the core of the project), and an experiment in which we (with the assistance of personnel from Mellon and OCLC) attempted to map, with mixed results, the information we gathered in Phase Two onto WorldCat, an OCLC service.

It is worth bearing in mind that university presses publish many things, and monographs are but one slice of their overall production, which could include journals, trade books, tests, regional titles (e.g., a tourist's guide to the region where a press is located), and classroom texts. Mellon specifically asked us to determine the number of scholarly monographs in the humanities published each year by American university presses. This immediately raises the question of the definition of a monograph. In our survey of the presses, which were asked to fill out a large template, we used the following instruction:

For a definition of monograph please use John Thompson's from Books in the Digital Age:² "books which are written by scholars and researchers and which are intended primarily for other scholars and researchers" (p. 103). To Thompson's definition we add the following: for purposes of this study, a monograph can have more than one author, but please exclude as monographs books that are collections of essays, even if the essays are all by a single author.

Another question is whether or not an edition is original. For example, if a press publishes a monograph in a cloth binding one year and a paperback (or an ebook) of the very same book a year later, does that count as one monograph or two? We thus introduced the concept of original works, designated in the survey as "primary" works (primary because most original monographs eventually spawned other editions).

² John Thompson, *Books in the Digital Age: The Transformation of Academic and Higher Education Publishing in Britain and the United States* (Cambridge, UK: Polity Press, 2005).

The survey we conducted had a potential sample of 106 presses, 65 of which successfully filled out our template and submitted the data to us. The template the presses used also allowed us to capture some other useful data about the presses--such things, for example, as pricing data, the number of titles published that are not monographs, etc.

The reporting presses published a total of 58,555 books over the five-year period, for an average of 11,711 books per year.³ The number of these books that were described as primary (that is, original) monographs was 14,619 for an average of 2,924 per year. And the number that was described as primary monographs in the humanities was 10,689, an average of 2,138 per year.

Since not every press filled out the template, we extrapolated what the five-year and one-year results would have been if we had had 100% participation for the presses (exclusive of Oxford, Cambridge, and the Associate Members of the Association of American University Presses, henceforth "AAUP").⁴ The extrapolated figures are:

- Total books: 76,000 or 15,000/year
- Primary monographs: 19,000 or 4,000/year
- Humanities primary monographs: 15,000 or 3,000/year

In the body of this report we introduce greater granularity to this analysis. We have divided the press community into six groups (defined below) and provided summaries for each group. We have also extracted data for four subject areas (art, history, literary criticism, and philosophy) in order to assess output at a finer level.

While it is not possible to know if there is a consensus view, we were surprised that the

³ These figures do not include Oxford University Press and Cambridge University Press. We also do not include Associate members of the AAUP, but confined ourselves to full AAUP members.

⁴ The methodology used to derive these figures is described below in the **Methodology** section.

monograph output in the humanities for American presses (exclusive of Cambridge, Oxford, and the Associates) is as small as it is. This suggests to us that programs designed to intervene financially on behalf of monographs will be challenged by the fact that monograph publishing resides in a broader business context, making it difficult to untangle such things as the allocation of overhead.

We do not believe that there is sufficient evidence for the five-year span of this study (which may not be long enough to extrapolate trends) to assert that the level of humanities monograph output has decreased. Our view is that it probably has not, but to prove this point one way or the other would require a study over a longer timeframe. A potential project for 2019 would be to add five years of data to this study (2014-2018), which would enable more meaningful trend analysis.

One item that came up early in our work with the participating presses was that of confidentiality. Many presses expressed discomfort in disclosing so much information about their operations. We suspect as well that since we were collecting information on sales and pricing among other things, some presses were being careful not to leave a suspicion that they were colluding with other presses. Therefore all the information provided to us has been held in confidence, shared only with Mellon.

We thus prepared private reports to Mellon on the data. The first of these was on the pilot project, the second on the rollout of the full survey. We also drafted another private memorandum on the experiment we conducted with OCLC. Internally, we have referred to the report you are reading now as “the public report.” The public report eliminates certain fields of information and any connection between the data and a specific university press. Neither the private reports nor the underlying dataset will be made public.

A note on the presentation of the data. The data falls into three categories: the raw data, the summarized data, and the data that is inserted inline in this report. The first two categories are private, the third public. The raw data is huge and difficult to work

with--it comes to about 1.1 million Excel spreadsheet cells. We made it available to Mellon, but to no one else. The summary data is actually larger--about 3.3 million cells (this is what we mean when we refer to the "scrubbed data")—because we cleaned it up or normalized it to make it more useful. But it is intelligible to anyone with access to the private reports who takes the time to study the column and row heads of the spreadsheets. We made some of the summary data available to Mellon and will shortly be sending it to the individual presses that participated in the study, but it is not inserted inline here.

Which brings us to the OCLC experiment. The background on this experiment derives from research performed by Rick Anderson and Dean Blobaum and published on the *Scholarly Kitchen*.⁵ Rick and Dean sought to determine how important library sales are for university press monographs. They thus took a sampling of titles from The University of Chicago Press, where Dean is employed, and looked up the ISBNs on WorldCat, using WorldCat to identify library holdings. Alan Harvey of Stanford University Press saw this blog post and realized that it would be possible to get a fairly comprehensive picture of the library distribution of university press titles by mapping the data from our project onto WorldCat. We attempted this, but ran into the problem that OCLC does not identify books as publishers do (OCLC is far less reliant on ISBNs). Thus the mapping to WorldCat was only partly successful.

The OCLC experiment raises the question of what other uses the data can be put to. One use was made by Paul Courant and Terry Geitgey, who investigated “free riders” in university press publishing.⁶ (A free rider in this context means an author whose parent institution does not have its own university press.) We suspect, though, that the most practical use will be made by the participating presses, to which we will be distributing individualized reports on their own programs and how their programs compare to those

⁵ <https://scholarlykitchen.sspnet.org/2014/06/23/how-important-are-library-sales-to-the-university-press-one-case-study/>.

⁶ Courant and Geitgey, “Preliminary Examination of “Free Riding” in US Monograph Publication,” *The Journal of Electronic Publishing*, [Volume 19, Issue 1](#): Summer 2016; DOI: <http://dx.doi.org/10.3998/3336451.0019.101>

of university presses of similar size and to the university press community overall. Those individual reports will remain private.

For readers who wish to get right to the primary findings of this study, please go directly to **Section V. Phase Two: The Core Database.**

III. Methodology

As noted above, this project fell into three phases. We developed our methodology over a period of time in collaboration with Mellon and in conversations with members of the press community. Some key decisions concerning scoping that we made along the way:

- We restricted our research to members of the AAUP.⁷ Thus we did not approach university presses that are not members but we also did not pursue other academic publishers, including the many commercial publishers working in this area.
- We studied only American presses, a decision made to limit the amount of data we would have to gather and analyze. It would be desirable to have a supplementary study of international presses, and particularly of Canadian presses, which are highly visible in the American academic publishing scene.
- Although we had the cooperation of the university presses of Oxford and Cambridge, both of which have offices in New York (and hence could be called American university presses), we do not include their data in this report, a matter we touch on below.
- As this project was initiated soon after the Great Recession, we chose to work with a multi-year span, reaching back to 2009, a year before the impact of the recession made itself felt in the lists of the publishers. Thus our five years of data was intended to allow us to normalize press output.

⁷ We wish to thank the AAUP staff for their unstinting support.

Our goals for this study were to:

- Amass a sufficient amount of data to enable us to make some generalizations about university press book publishing;
- Organize that data in an intelligible and useful way;
- Create reports from this data for Mellon's internal use;
- (Under Mellon's guidance) create individualized reports for each of the participating presses;
- Experiment with mapping the press data onto WorldCat;
- Create this public report.

Early on we realized that we would have to manage a large amount of data, which led us to begin with a pilot project to test our assumptions and, not incidentally, to develop a workflow that was reasonably efficient. We discuss the pilot project in some detail in the following section (**Phase One**).

For the central phase of the project (that is, **Phase Two**) we solicited data from all members of the AAUP except for international members and introductory members. (See below for the section on the template that we sent out.) That is, of the AAUP's total membership of 131, we sought information from 106 of them. We received data from 68, for a response rate of 64%. The response rate was materially improved after Mellon sent out a letter to presses that had not responded to us.

Virtually all of the presses had some difficulty in providing the data. After some struggling, we chose not to use the data from three presses as it did not conform to the template we had provided even after we contacted these presses and attempted to help them with the task.

As for the quality of the underlying data, many of the presses experienced difficulties in getting the filled-out template to us. Getting the data from the presses took much more time than we had anticipated, something that should be borne in mind if this project is to

be repeated in the future (it would be a very big project to try to do this on an annual basis). We simply don't know if the project is replicable because the presses had so much trouble with it the first time around; in some instances, they resisted doing the work. Note that the issue here is not the complexity of the software; no one was being asked to learn how to use a new application. We worked entirely with Microsoft Excel, an application that all the presses use routinely. The problem was that for many of the presses, the requested data was simply not maintained in a useful form.

For all these qualifications, we think the data we have gathered permits us to make some reasonable, if not perfect, estimates of press activity.⁸ We spent a good deal of time scrubbing the data to make it useful. We also are confident in our ability to extrapolate data for presses that did not submit anything to us; this confidence arises because of the high rate of response.

For purposes of this report, we have analyzed the data first by breaking it into six groups: the four groups of the AAUP classification, Oxbridge, and the Associate Members. We considered putting the Associates into the AAUP's four categories, but it became apparent that the Associates work with a different paradigm from the full members of the AAUP. The Associates are a mixed lot: some are research organizations, some cultural institutions, some professional societies. The typical university press, on the other hand, is a unit of a parent institution and solicits manuscripts from scholars everywhere. Interestingly, rarely does a university press have more than 10% of its total authors affiliated with the parent institution.

Now we turn to the template.

⁸ The difficulty in getting good data from the presses has made us wonder about other sources of information on university presses. For example, the AAUP gathers financial data from the presses every year and then publishes an aggregated report, but as this report is based on data that individual presses provide, how reliable is it? Similarly, how trustworthy are the entries in the AAUP *Directory*, in which the total number of publications for each press is listed every year?

To collect the information from the presses, we distributed a template created in Excel. The spreadsheet had multiple tabs. One was a set of instructions; a second presented a sample to show the presses how to put in their data; and another was a list of BISAC (Book Industry Standards and Communications) codes, which were used to identify each book's subject category. The final tab was for the data itself. That data was broken into 17 columns, which are listed below.⁹

A. Title. The book's title was inserted into this column.

B. ISBN. This is the standard identifier for books. Note that while some presses assign ISBNs carefully by format (e.g., an ISBN for a cloth edition, a different ISBN for a digital edition), some presses use ISBNs to encompass multiple formats.

C. Format. Is the book an ebook, a cloth edition, a paperback?

D. Price. The suggested retail price.

E. Monograph. Is this book a monograph? We asked for a yes or no answer. A monograph was defined as a book by a scholar for other scholars. We specifically excluded from this definition collections of essays, even if they were all by the same author. (The lack of consensus on the definition of a monograph was the single biggest problem uncovered in the pilot project.)

F. Primary. This is a yes or no question. Is the book an original copyright? A book that was first published in, say, 2010 might be the primary edition, but the paperback reprint published in 2012 is not primary.

G. Date published. The date for that edition.

⁹ The letters of the subheads in this section correspond to the actual column heads of the template. That template is not part of this public report, but was made available to all of the participating presses.

H. Principal author, last name. Self-evident.

I. Principal author, first name. Self-evident.

J. Advance. This is a yes or no question. Did the publisher pay the author an advance against royalties?

K. Subvention. A yes or no question. Did the publisher receive a subvention for this title? A typical source for a subvention is the author's institution. Many presses found it difficult to answer this easily.

L. BISAC code #1. The primary code assigned to a title. Sometimes books have more than one code (hence columns M and N below).

M. BISAC code #2. Many presses did not provide or provided incomplete data for BISAC codes BISAC code #2 and #3. For the purposes of identifying primary humanities monographs, only BISAC #1 was used.

N. BISAC code #3. As above.

O. Discount code. The discount code indicates the percentage that a wholesaler or retailer subtracts from the suggested list price when paying the publisher's bill. Thus a book with a list price of \$30 might have a code that provides a discount of (say) 35% to a retailer.

P. Total net sales units, print, 2009-2013. Some presses had difficulty providing us with these figures and many presses provided sales dollars instead of sales units. When sales dollars were provided, a standard discount of 30% was applied to estimate the sales in units. Note also that title-by-title comparisons here are not meaningful because books with an earlier publication would have had more time to perform in the marketplace.

Q. Total net sales, digital, 2009-2013. More presses had trouble getting this number, as sometimes their print and digital sales are reported together.

Setting up the template in this way enabled us to drill into the data and get to the results we sought:

1. What was the total number of books your press published in a given year?
2. How many of these books were primary, that is, original copyrights?
3. How many of these books were monographs?
4. How many of the primary monographs were in the humanities? We derived this figure by applying the BISAC codes against #3.

More information on methodology appears in the following section on the pilot project.

IV. Phase One: The Pilot Project

This section summarizes our findings from the pilot project. The pilot was undertaken, of course, to hone our methodology and to help forecast the time and resources that would have to be put into the project.

The decision to work with a pilot was an easy one. This was a complex project with a lot of moving pieces. We also were concerned from the outset with the sheer size of the database and the number of people we would have to interact with (approximately 100 publishers) all at the same time. On the basis of our experience, we are very happy to have begun with the pilot. We encountered a number of issues concerning data collection and definitions of terms (especially for the term “monograph”) that became very time-consuming.

Methodologically, our first step was to design a template using Microsoft Excel with column heads that defined the information that should be put into each cell of the

template. The template covered 19 fields of data, from the obvious (the name of the author) to the more technical. For example, we requested that the publishers identify the subject of each book, as represented by BISAC.¹⁰ BISAC is an industry standard for designating book topics. In addition to the template, the spreadsheet that was sent to the participating publishers included row-by-row instructions and a mock-up of a filled-out template (where we made up the information). None of the presses had any problems working with Excel for the data-gathering, which was an important consideration for the rollout.

The template went out to eight presses. These were selected along four axes. First, we sought two presses from each of the four categories of the AAUP's taxonomy, which is organized by sales volume (the largest presses--Harvard, Princeton, Chicago, California, Yale, etc.--are Group 4, the smallest Group 1). Second, we looked for presses where we had some familiarity from prior or current projects as we believed that this would enable an easier flow of information. Third, we sought presses where we thought we would uncover special information--for example, presses with a distinctive publishing list. Finally, we included two presses where we had no previous connections. We don't want to pretend that this fourth category somehow made the pilot "scientific," but we did view this category as our control.

Although we took a cross-section of the press community, we do not claim that these presses were representative of the university press world as a whole. The aim of the pilot was to test our methodology, not to come up with a short cut to get at the total figures for all the presses. There are simply too many variations from one press to another to do that. A representative sample would have to be much larger. The purpose of using this cross-section was not to say, "We can multiply the figures these presses gave us by 10 or 12 and get the total output of the press community," but to say, "This selection of presses defines the nature of the problems that have to be addressed in order to poll the rest of the presses efficiently."

¹⁰ The Book Industry Study Group has a good FAQ on BISAC: <http://bisg.org/page/BISACFAQ>.

All eight presses were eager to help with a project that bore Mellon's imprimatur,¹¹ but some noted that not all the fields of information were readily available. Indeed, it is clear that each press (by their report) had to devote something in the neighborhood of three person-days to fill out the template, and even then some of the fields were likely to be left blank.¹² This is not surprising: even some of the largest commercial publishers don't have comprehensive databases for all their titles. Systems are built to solve specific problems, not to facilitate research. Thus it is easy for a publisher to provide the discount code for every book because that information is regularly requested by distribution partners. On the other hand, the only way to determine whether or not a book is a translation, for example, is to go to each editorial file, one at a time, and look for this information manually. Were presses to be told that there could be funding available in the future provided that a press kept good records that covered a great many fields, those databases would be put together relatively quickly--because, in effect, they would view the funding agency as a "customer" to which they would be "selling" their internal data. The real lesson from the pilot was that we had to ask the questions carefully so that the presses could provide the information without being chased by us.

The data we received came with the caveat that, except for Mellon, it should only be shared in the aggregate. Thus Mellon could choose to share this report with the AAUP, for example, but the raw data could not be shared. In particular, information about discount codes and pricing could not be shared as this could be construed to touch on antitrust issues. (We are not lawyers, but publishers in general are wary of any public discussion of pricing.)

As for the data itself, we requested and received it from eight presses. We asked for five years of publishing history. We used five years to make sure that we were not

¹¹ No doubt this cooperative spirit owed something to the fact that we handpicked the presses for the pilot. For the full rollout to the entire press community, the degree of cooperation fell off sharply.

¹² A particularly tough field for the presses was to determine the author's institutional affiliation for any particular book. We dropped this field in the full rollout to all the presses.

recording special events in any one year.¹³ We asked the presses to list *all* of their books published during this period. We then asked them to identify all of their *primary* titles, where *primary* was defined as follows: "The primary publication is the first publication of a specific title." A primary publication, in other words, is where a press makes its principal investment in a book. A secondary publication--a paperback edition or an ebook version--requires less investment. We also asked the presses, using their internal notion of what constitutes a monograph, to identify which books were monographs and which were not.

The pilot also captured the sales figures from the presses over the five-year period of the study. We were not sure how to use this information, however, as it is inherently incomplete.

University press books sell over a long period of time. We know from the work Joe did with Kizer Walker and Terry Ehling on patron-driven acquisitions that books in the Cornell library generally have a circulation life of about 12 years.¹⁴ It seems reasonable to assume that book sales also extend over a comparably long period of time, though we do not have information on the "half-life" of university press books. One factor that extends the life of a book is the eventual adoption for classroom use. A book could have been published in 2010, sell a few hundred copies in cloth and perhaps as a Kindle edition, and then was reissued as a paperback for, say, \$20, in 2014. The paperback may trickle out at a rate of 50--or 500--copies a year for many years. It's difficult to know at the time of first publication which books will enter the curriculum (if editors knew this, university presses would be sitting atop a great deal of surplus cash), though presses guess about this all the time. Taking a five-year window, however, is simply insufficient to gauge how well particular books sell overall.

¹³ One press presented a figure for 2013 that was less than half that of 2012. We believe that this reflects records that were not fully up to date when the staff worked on the template.

¹⁴ See "PDA and the University Press" on the *Scholarly Kitchen*: <https://scholarlykitchen.sspnet.org/2012/10/16/pda-and-the-university-press-final-study-now-available>.

A second problem is that the window is five years only for those books published in the first year of the study. As the publication year moves forward one year at a time, the window closes a bit. This makes the sales figures for books published after the first year of the study even less useful for extrapolations.

A refinement of the point in the preceding paragraph is that books are not all published on the same date, and certainly not all on January 1 of each year. So one book may have five years of sales data because it was published early in the year, another book may have been published in the same year but in December, meaning that its sales figures are for slightly more than four years. The way around this (though it's not clear to what purpose this information would be used) is to assume that half the books were published in the first half of the year, half in the second. Thus for books published in the first year of the study, we have 4 ½ years of sales data, for books published in the second year, we have 3 ½ years of data, and so on.

Another complicating issue regarding sales figures is that some books are published in multiple formats, though not all at the same time. Those formats are likely to be cloth (most press books have cloth editions), paperback, and a version of ebooks.¹⁵

Overall, the pilot made us wary of the sales information we gathered. Although we went on in the full rollout (**Phase Two**) to collect sales information, we think the figures understate actual sales performance because of the short time period for which they were recorded.

Throughout the pilot we looked for ways to streamline the process, both for the presses that had to provide the data and for ourselves, as the possibility of getting buried under a mountain of data was always before us. So, for example, we tried to come up with a method by which we could identify which books were truly scholarly monographs in the

¹⁵ We did not isolate data for “shorts” in this study. “Shorts” are books that are not very long, typically between the length of an article and a “full” book. Many presses are experimenting with shorts, both as original publications and as abridged digital reprints.

humanities and which were not and hoped to make this determination by using two mechanical determinants: the discount codes associated with each title and BISAC codes. The discount code identifies the percentage of the suggested retail price that someone pays for a particular book. An individual buying a book directly from a press's Web site, for example, may pay full list price and have a discount of 0%. A library wholesaler may receive a discount of 35%. The list of the various discounts goes on and on; it is a complicated, even arcane aspect of the book business. To describe the situation in somewhat oversimplified terms, academic publishers typically have three broad categories for discounts: a discount for course adoptions (around 20% off of list price), a discount for specialist or professional books (generally around 35% off of list price), and a so-called long or trade discount for books that are sold in bookstores, including online bookstores (between 40% and 50%). We thought that the presses might all use the specialist discount for monographs, and were this true, we would then have a fast and easy way to say what is a monograph and what is not.

Unfortunately, we could not detect any pattern in the discount codes and whether or not a press deemed a particular book to be a monograph. Some titles marked as monographs bore specialist discounts, some "long" discounts. This meant that for the rollout we still needed a way to identify monographs.

The point of using BISAC was not to say whether a book was or was not a monograph but to place it by domain. This is workable, but not elegant. All the presses have some familiarity with BISAC; the question is which codes (hence subject categories) should be included in the program.

Ultimately we worked with Mellon to come up with a list of fields. We then mapped those codes to the list of the primary monographs. Note the implications of this: we moved from the full list of titles and then got increasingly granular. Thus within the set of all titles we then moved to all *primary* titles. Primary titles were then subdivided into *monographs* and non-monographs. Finally we derived the numbers for *primary monographs*. After Mellon identified the subject areas it was interested in (using the

BISAC alphanumeric system), we were then able to derive *Mellon-approved fields for primary monographs*, which would be the smallest list of titles of all. The caveat here is that the BISAC information the presses provided is messy, so that final list of titles is less than elegant.

Aside from the sheer effort that went into gathering the data, the item that gave the presses the most difficulty was coming up with a definition of a monograph. For the pilot template we tried to have the presses use their internal definition of a monograph and then, title by title, to indicate whether a particular book was a monograph. The feedback we got on this suggests that there is no consistent definition. For the project to roll out, we needed to provide a working definition.

So what is a monograph? We don't think anyone is happy with the literal meaning of a book by a single author. What about the highly specialized title that is written by two authors? Should that not be considered a monograph? Nor do all the presses think of the monograph as something highly specialized (some of the books checked off as monographs bore trade discounts, suggesting that the presses were seeking to put these titles into bookstores).¹⁶

Ultimately we decided for the rollout to use the amended definition of John Thompson, which appears in the Introduction of this report.

As for the figures themselves, they represent only a small sample of the university community (and were derived solely for the pilot), but for anyone who is interested, here they are. The eight presses under examination for this pilot reported that over a five-year period, they cumulatively published a total of 9,786 works. That's an average of 1,957 each year. For those titles defined as "primary" (which we defined as follows: "The primary publication is the first publication of a specific title"), the 5-year figure is 5,334, and the yearly average is 1,067. For those titles described as both primary and

¹⁶ When we finally got to the full project rollout, one press director argued that novels should be classified as monographs.

as monographs, the five-year total was 2,505 books, for an average of 501 per year. Since we could not extrapolate from these figures to provide a reasonable estimate for the totality of press output (that is, across all the members of the AAUP), we needed to roll out our survey to all the presses. The results from that rollout appear in **Section V: Phase Two** below.

V. Phase Two: The Core Database

In this section we will summarize the core findings of the project, where we rolled out our questionnaire/template to the entire American press community. Once again we will be suppressing information about individual presses, but will present information about groups of presses (e.g., all the Group 4 presses) and the AAUP press community as a whole. We will be making distinctions between presses of different size and programs in different subject areas.

A. Focus: Group 4 Presses

As noted above, the AAUP divides presses into four categories based on size. The largest presses (excluding Cambridge and Oxford) are Group 4. The members of this group are:¹⁷

- California
- Chicago
- Columbia
- Harvard
- Johns Hopkins

¹⁷ Note that Duke does not appear here, though it may be a larger press than Columbia (we don't have consolidated sales figures for the presses). We believe this is because Duke's book program is relatively small; the strength of Duke is in journals. Duke is a Group 3 press.

- MIT
- Princeton
- Teacher's College
- Yale

We want to drill into this group at a fine level before turning to the other groups, which we review at a higher level. The reason that we are subjecting Group 4 to this closer analysis is simply these presses' size.

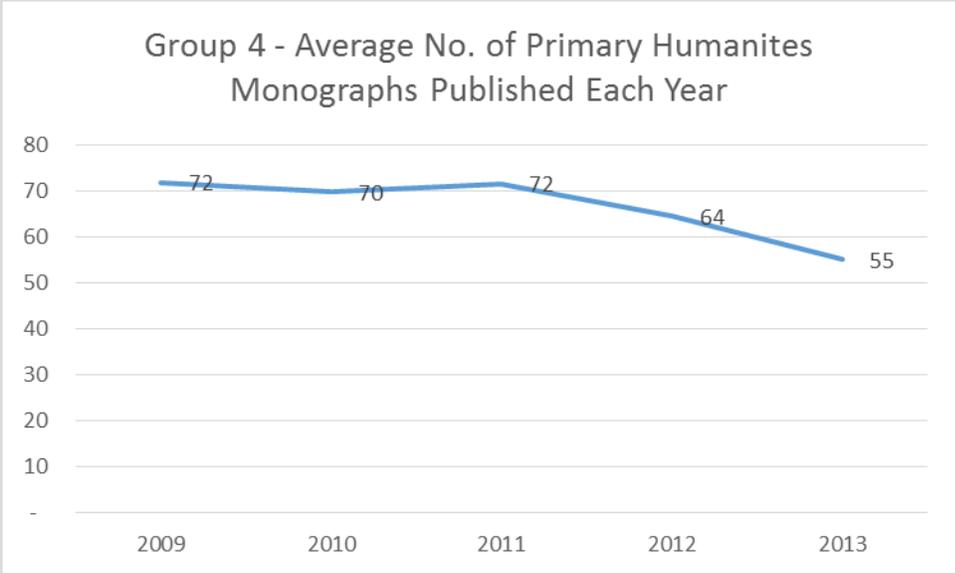
We had 100% participation from this group--nine presses out of nine. Over the five-year span of the survey, this group published a total of 23,167 books. Of this figure, however, the number of books designated as primary monographs was a much lower number, 4,922. When the filter included only original copyrights (or "primary" titles) *in the humanities*, the number dropped further, to 2,992 titles (see chart below).



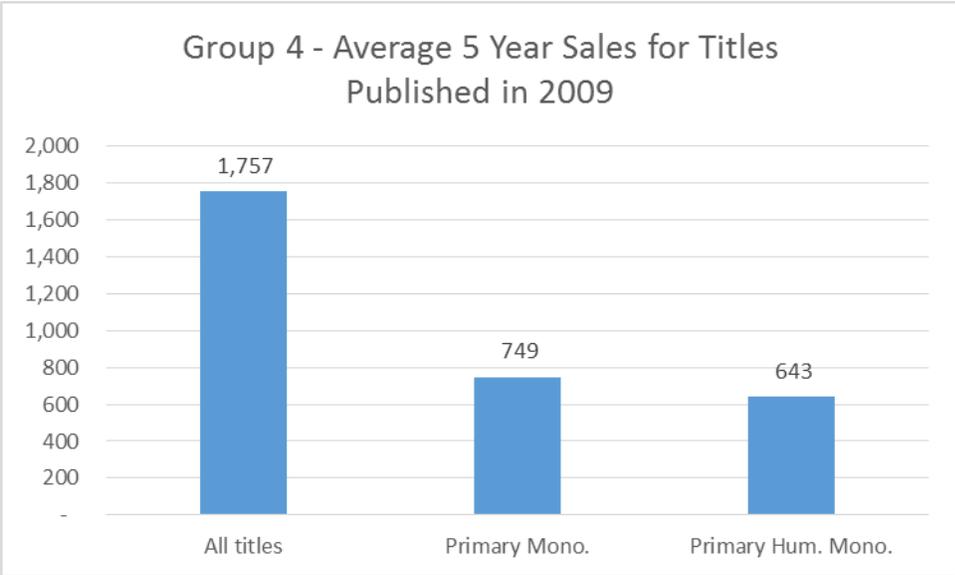
This figure surprised us--we had thought that the humanities primary monograph would be considerably greater than this--and we wonder if this number will come as a surprise to other observers. For Group 4 presses original monographs in the humanities averaged only 13% of total press output over five years.

Can we make some determinations about trends? We have to be careful, as the performance of just one or two presses could skew the results. One of the Group 4 presses, for example, does no monograph publishing at all in the humanities; another has a list heavy with science titles. And then there is the matter of the five-year term of this survey, which is not long enough to be confident about the evolving shape of university press publishing. We don't see any opportunity to get ten years' data from the presses, though, so five years is the best we are likely to have to work with.

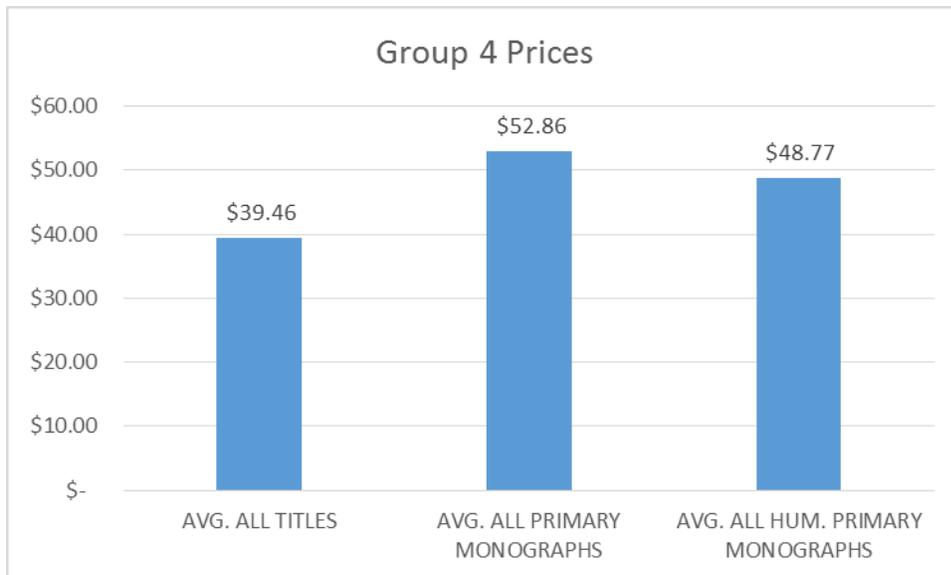
In 2009 each of the Group 4 presses averaged 72 primary monographs in the humanities, a figure that dropped to 55 in 2013. We don't think the evidence (only five years of history) supports the idea that monograph publishing for this group is declining, but it is clear that it is not growing. If the figures in three years continued to show a downward trend, we would feel more comfortable in asserting that monograph publishing is in decline among the larger presses.



The sales figures (in units) for monographs for the Group 4 presses pretty much confirms the figures that are tossed around casually within the press community. For books first published in 2009 (original monographs in the humanities), the average sale in units over five years was 643. It is not meaningful to assess sales for books first published after 2009 within the design of this study because books published later did not have enough time to demonstrate their performance in the marketplace.



On pricing the results for the Group 4 presses pretty much confirmed our assumptions. Humanities monographs are generally priced higher than the rest of a press's list: an average of \$49 for primary humanities monographs in comparison to \$39 for all books. This is likely because monographs are published with a greater emphasis on institutional markets than the many trade titles the Group 4 presses publish.



We asked the presses for information on subventions and advances. Not surprisingly, the results for these presses were widely distributed. Group 4 press averaged subventions on 219 books over the five-year period. During the same period, the average number of advances was 458. (We do not know the size of either the subventions or the advances.)

Group 4 - Subventions and Advances		
	Total for 5 Years	Average/Press
Subventions	1752	219
Advances	3667	458

How useful is the information about subventions and advances? The subvention figure may be meaningful, as it points to the willingness of an institution to provide support for a faculty member who publishes with a press located at another university; 219 books

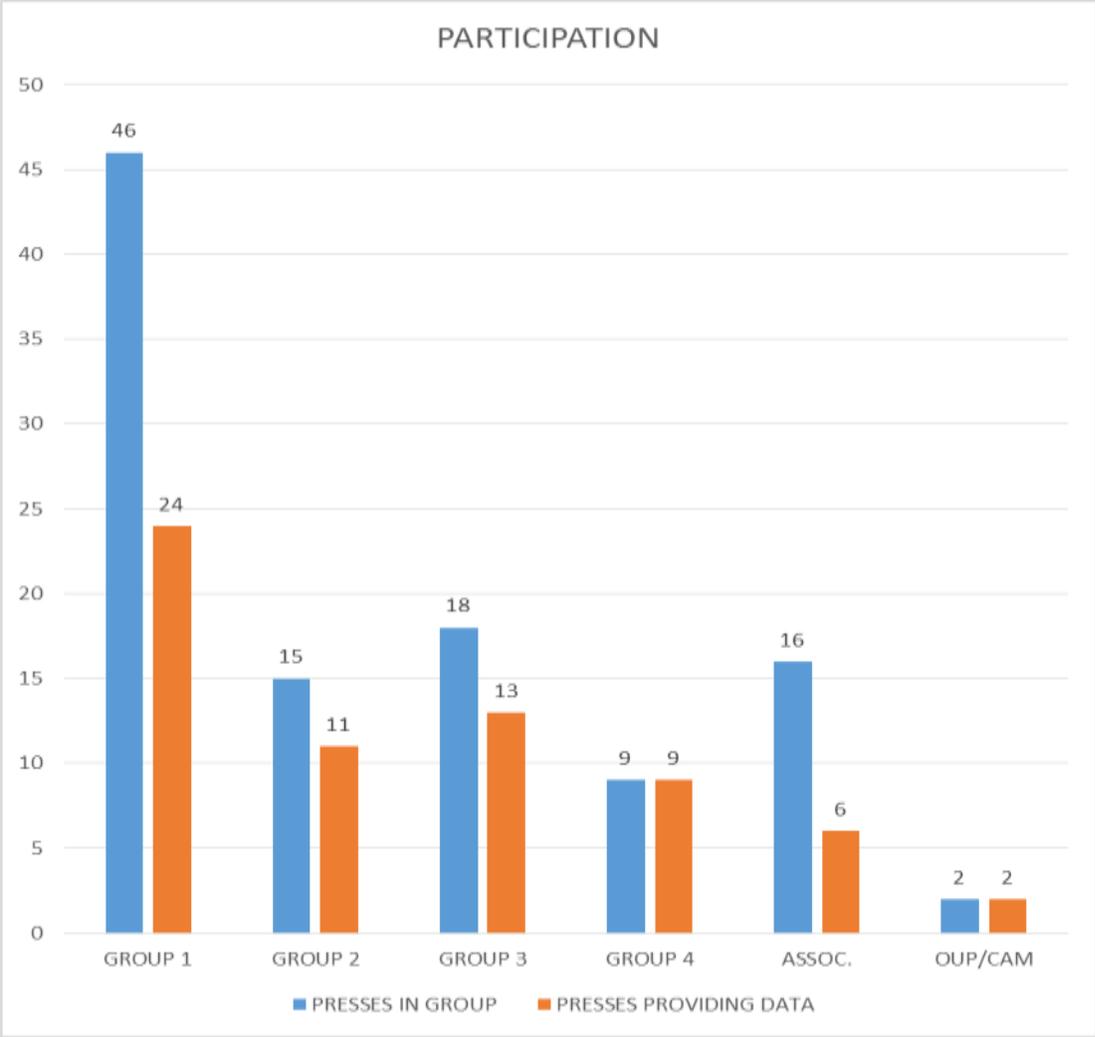
over a five-year period for the average Group 4 press is not a negligible number. The number of advances appears to be less valuable, as advances are skewed toward books with greater market potential in the first place. It is, alas, a defining characteristic of the academic monograph that its market potential is modest.

The information we gathered on discount codes does not seem useful. We originally sought this information because we hoped--incorrectly, it turns out--that we could use the discount codes as an indirect way to identify monographs. The hypothesis was that monographs are largely published for institutional markets and would thus carry a "short" or "professional" discount. The presses, however, declared many titles to be monographs that in fact carried other discounts, including those typically used for trade and course adoption titles. It was for this reason that we opted for a different method to identify monographs (a monograph is a book written by a scholar for other scholars).

B. Summaries of Other Press Categories

We broke the presses into six categories: the four Groups as defined by the AAUP; OUP and Cambridge taken together (aka "Oxbridge"); and the Associate Members of the AAUP. We provided some detailed information on Group 4 presses above. Now let's turn to the other categories.

The participation of the presses varied widely by category. We had 100% participation from Group 4 and also from Oxbridge, but only 52% participation from the Group 1 presses and 38% from Associates. (We understand from the AAUP that this level of participation is about the same as for the AAUP annual survey.) This led us to come up with a way to extrapolate data, which we summarize later in this report. The following graph summarizes participation levels by category:

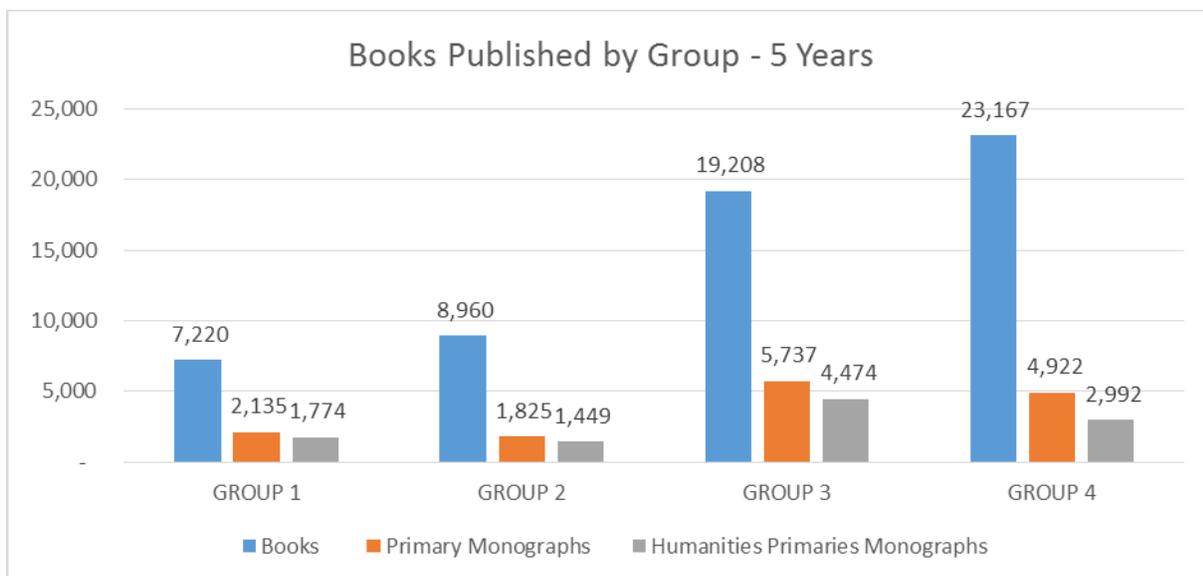


For Group 3, 13 out of 18 presses participated, or 72%. For the five-year period these presses published 4,474 original monographs in the humanities. The number jumps out because it is greater than for the much larger Group 4 presses (2,952). Admittedly the comparisons are not precise. Group 4 only has nine presses, one of which does not publish monographs in the humanities at all. It is nonetheless inescapable that the modest-sized presses of Group 3 play a larger role than one would have expected in monograph publishing. And these figures are before we extrapolate for the non-reporting Group 3 presses.

For Group 2, eleven out of fifteen presses participated, or 79%. The five-year total for

primary monographs in the humanities came to 1,449. Once again, this is not a negligible number.

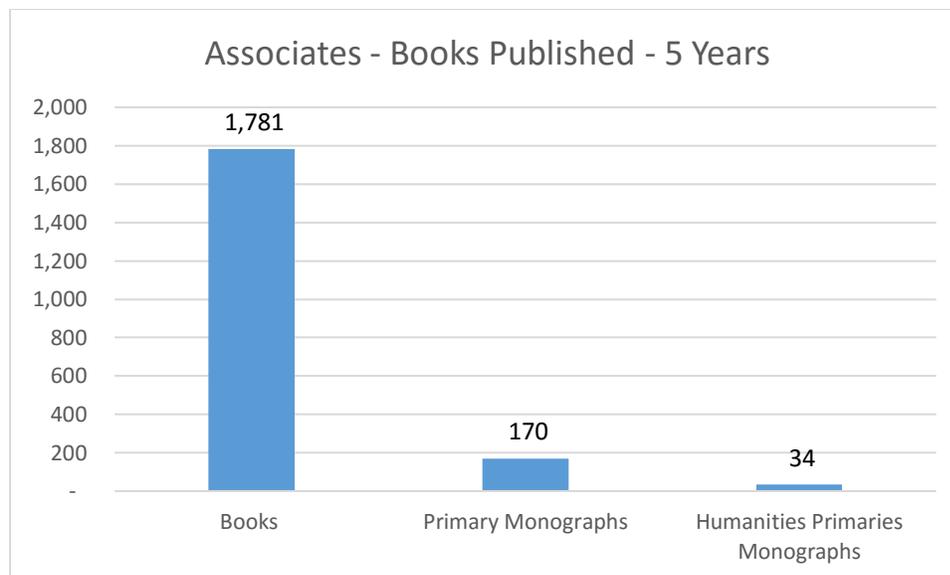
When we get to Group 1, the level of participation flags. Out of 46 total presses in this group, only 24 filled out the template (52%). Of course, these are very tiny organizations, which likely found the task of providing the data to be burdensome. Because of the sheer number of Group 1 presses, however, the aggregate five-year monograph output (primary books in the humanities) came to a substantial 1,774, a bigger number than for Group 2.



What these figures suggest to us is that any program set up to support humanities publishing should be tailored for presses with limited resources. It's important to understand the nature of those limitations. Of course the smaller presses don't have buckets of money, but there is an even bigger limitation in that these presses don't have the management depth to take on complex projects. None of these presses have "benchwarmers" who are in a good position to take on a new project. It would be salutary if programs with these publishers in mind were designed to have little paperwork, for example. Perhaps it would be possible to set up a third-party to handle

some of the administration for any new program.

When we get to the Associate Members of the AAUP, the level of production for primary monographs in the humanities falls off, and the engagement of these presses with this project reflected that. Out of 16 presses in this category, only 37% filled out the template. Over five years, these presses were responsible for only 34 titles defined as primary monographs in the humanities. It should be noted that some of these organizations don't even have book programs. The American Historical Association is perhaps the extreme example. American history is a field rich with publications, but the members of AHA find other venues for the purpose of publishing long-form scholarship, including university presses and even trade houses. One title that immediately comes to mind, Fredrik Logevall's *Embers of War*, a history of the Viet Nam war. Logevall is now on the faculty at the Harvard Kennedy School; we met him when he was at Cornell, overseeing Cornell University Press. His book, published by trade publisher Random House, won a Pulitzer Prize and was a great commercial success.¹⁸



¹⁸ We got to know Fred while working on a project for Cornell University Press. One topic of discussion was for the Press to publish the specialized materials of public intellectuals like himself. Thus Random House would publish commercially strong titles like *Embers of War* and Cornell would publish the papers and notes behind the bestseller. We hope to pursue this area in the future.

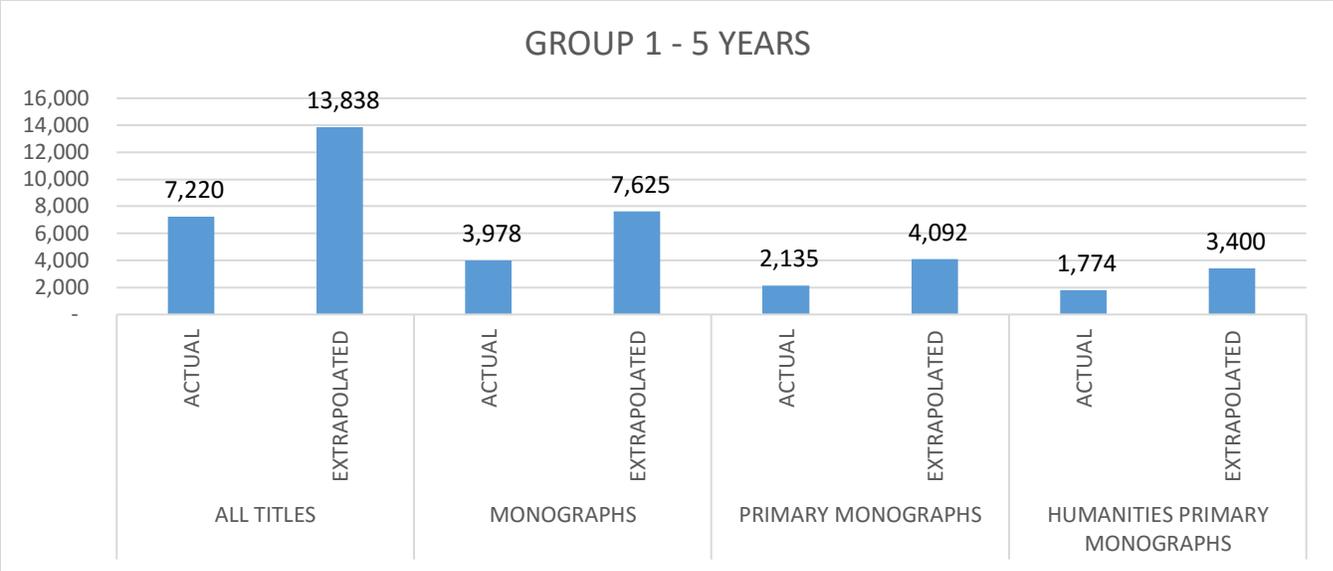
C. Extrapolations

As noted above, we were able to collect enough data to enable us to extrapolate figures for those presses that did not participate.¹⁹ We restricted this to Groups 1-3. This is because we had 100% participation for Group 4; hence no extrapolation is necessary. We excluded the Associate Members, as their programs vary widely, making extrapolation unreliable. And we excluded Oxbridge, as both members of the set participated.

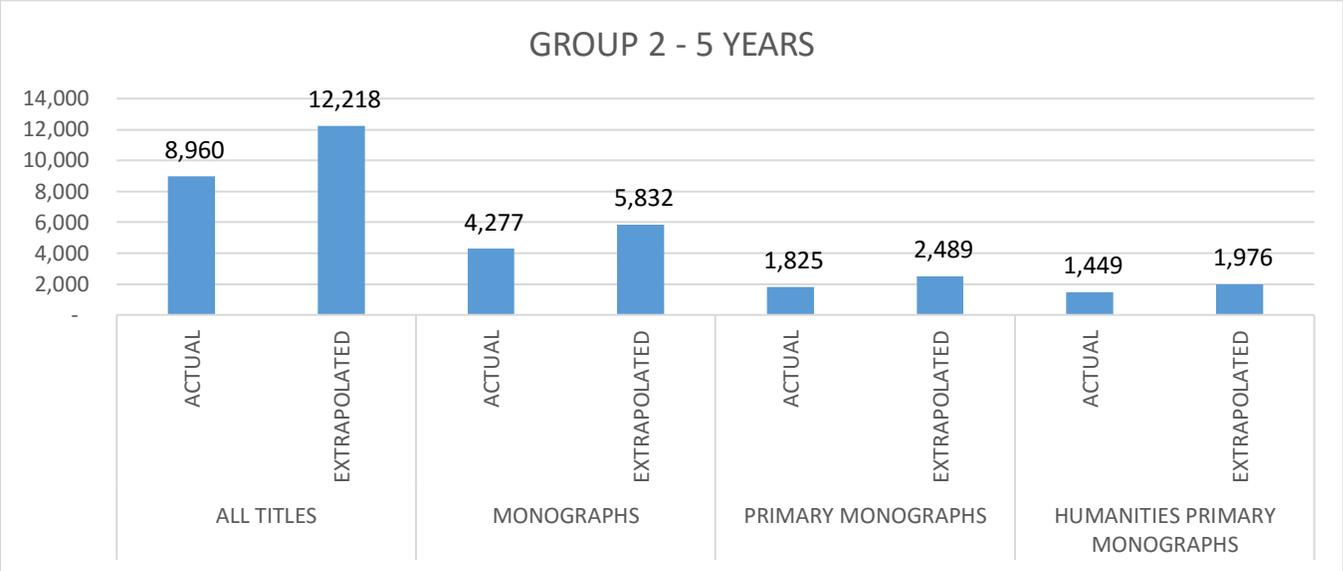
Thus for Groups 1-3 we determined the average number of publications for the presses in each group (that is, the number of primary monographs in the humanities) and then extended those figures to all the presses in the group. We believe that the extrapolations give us a fuller view of total press output.

In the following tables we present the actual values for the participating presses in each group (that is, the sample) and contrast them with the extrapolated or full estimated values. Note that the values for the Group 4 presses are the same for participants and extrapolations (because participation was 100%). First up is Group 1:

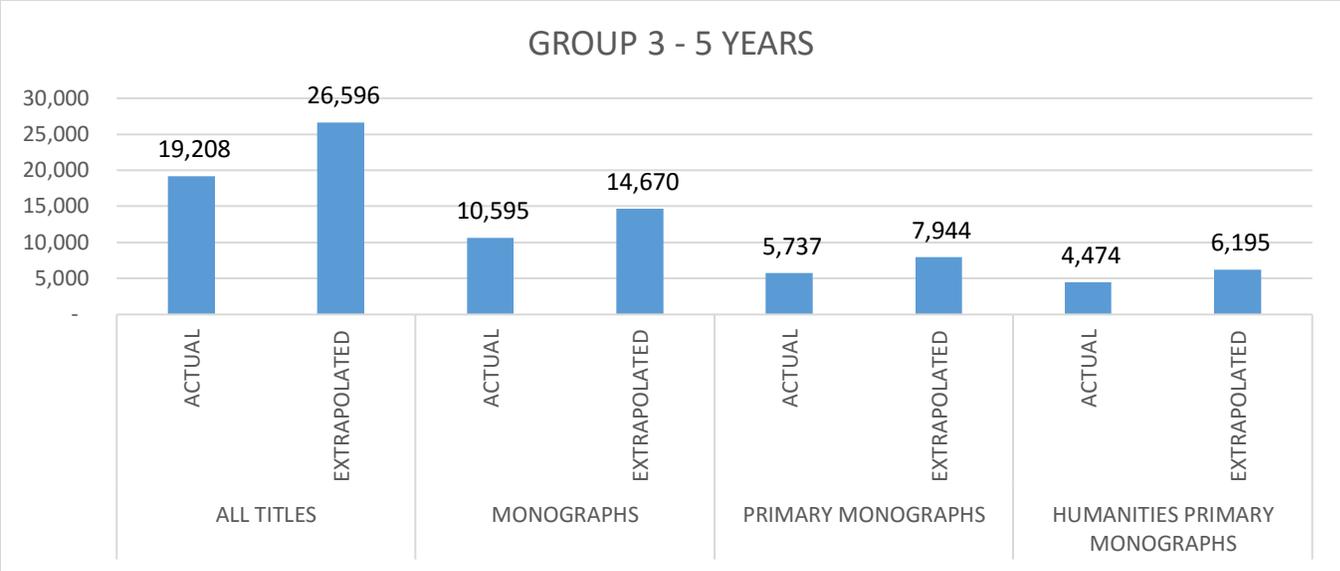
¹⁹ Throughout this section we distinguish between the figures we collected from our sample of the presses (the actual figures) and the estimated total figures, which we derived by extrapolating from the sample.



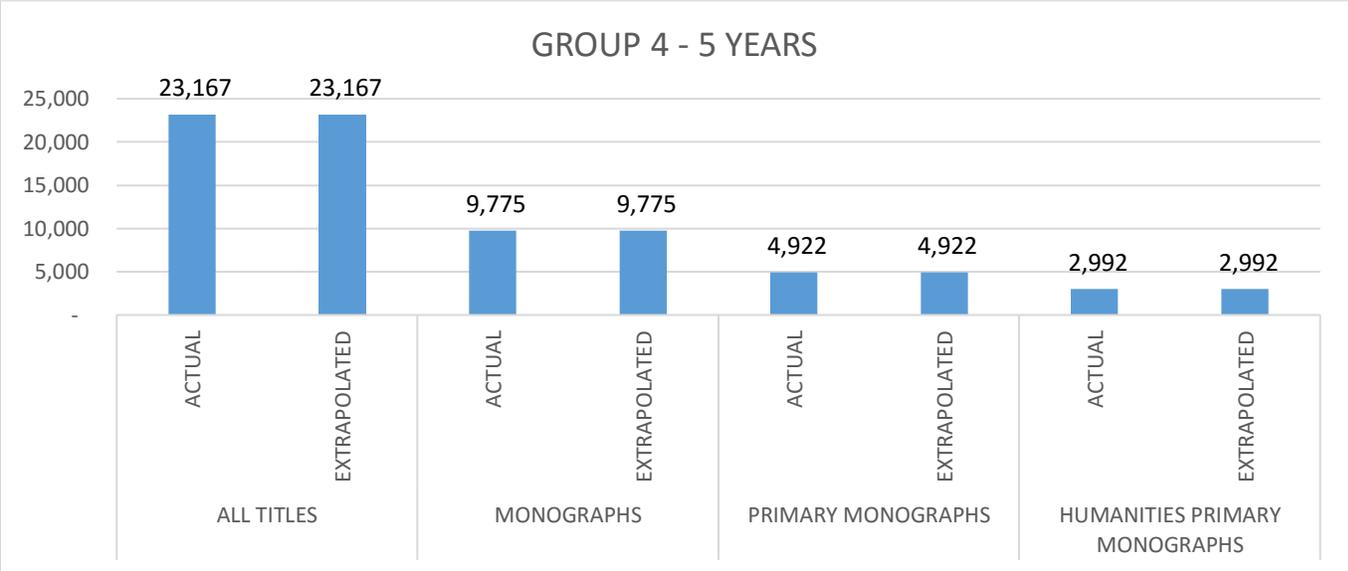
Group 2 had a higher level of participation than Group 1; hence the extrapolations are not proportionately as large:



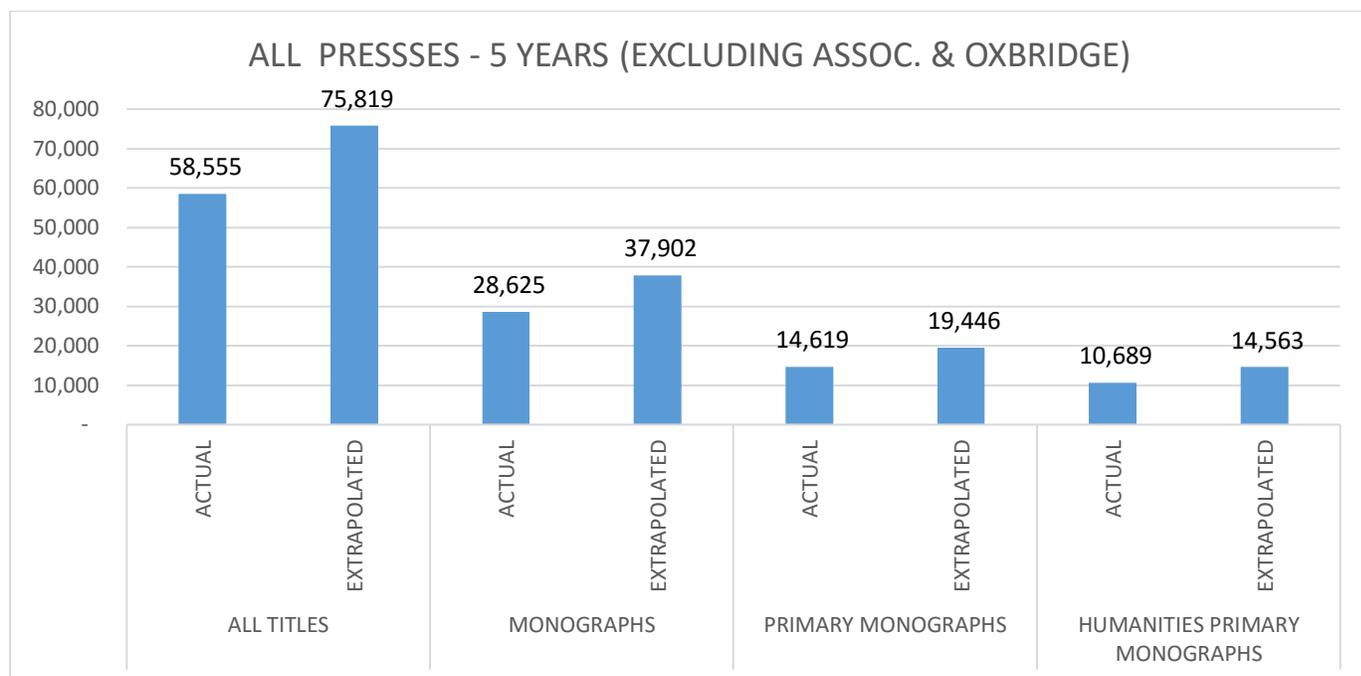
Here is Group 3:



We are showing Group 4 here for the sake of completeness, but the actual and extrapolated values are the same:



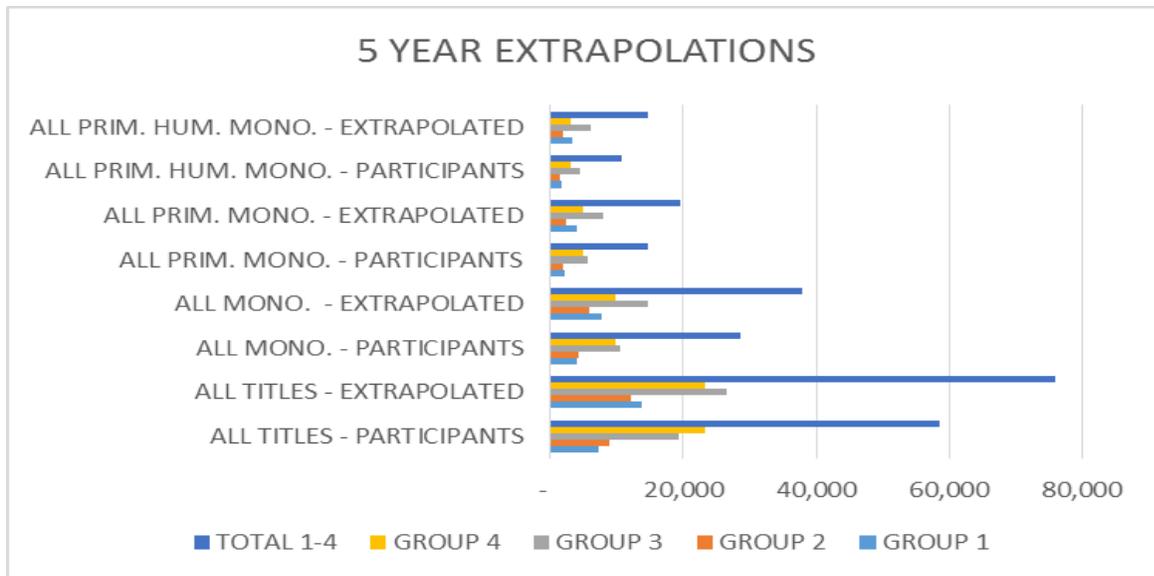
The following bar chart shows the consolidated data for the four AAUP Groups:



Below is a summary expressed in tabular form:

	ALL TITLES - PARTICIPANTS	ALL TITLES - EXTRAPOLATED	ALL MONO. - PARTICIPANTS	ALL MONO. - EXTRAPOLATED	ALL PRIM. MONO. - PARTICIPANTS	ALL PRIM. MONO. - EXTRAPOLATED	ALL PRIM. HUM. MONO. - PARTICIPANTS	ALL PRIM. HUM. MONO. - EXTRAPOLATED
EXTRAPOLATIONS - 5 YEARS								
GROUP 1	7,220	13,838	3,978	7,625	2,135	4,092	1,774	3,400
GROUP 2	8,960	12,218	4,277	5,832	1,825	2,489	1,449	1,976
GROUP 3	19,208	26,596	10,595	14,670	5,737	7,944	4,474	6,195
GROUP 4	23,167	23,167	9,775	9,775	4,922	4,922	2,992	2,992
TOTAL 1-4	58,555	75,819	28,625	37,902	14,619	19,446	10,689	14,563

And finally a summary expressed as a bar chart:



So, to express this in round figures, the American university press community publishes just under 3,000 original monographs in the humanities each year. That figure does not include Oxbridge or the Associates.

D. Oxbridge

We have little to say about Oxford University Press and Cambridge University Press in this public report. Both presses cooperated with our survey, but for a number of reasons we are not including their data here:

- A. OUP and Cambridge are both truly global entities, making comparisons with the American-based presses problematic. Should we include figures for OUP's New York office alone or should we tie in OUP's publications originated in its many offices around the world?
- B. We assured the presses that no information that can be tied back to them would be made public. As a practical matter, the output of Oxbridge is so huge that anyone can identify them in any table or chart even if they are not named.
- C. Finally, Oxbridge's large size skews the results. What do we learn from

comparing Cambridge with California or Chicago except to say that one of these entities is big, big, big?

E. Subject-area Analysis

We looked into the publications by subject area. This was made possible by the use of the BISAC codes. It was not possible to do this for all of the subject areas, as this would literally add thousands of cells to the spreadsheet, so we decided to work with four categories: art, history, philosophy, and literary criticism. These choices were not entirely arbitrary. We selected art because of Mellon’s historical interest in Artstor. History was chosen because of its prominence in university press publishing. Philosophy simply seems to be the bedrock of humanistic study. And literary criticism is a field that is often talked about as a notoriously difficult field to publish in because of market conditions. (Note that these figures do not include Oxbridge.)

The first table is for art:

SUBJECT AREA DATA - PARTICIPANTS - ART	2009	2010	2011	2012	2013	TOTAL	% OF HUM. PRI. MONO.
GROUP 1	1	4	9	5	5	24	1.4%
GROUP 2	11	15	11	17	15	69	4.8%
GROUP 3	15	15	30	24	23	107	2.4%
GROUP 4	21	33	35	35	20	144	4.8%
ASSOC	6	4	9	2	4	25	73.5%
TOTAL	54	71	94	83	67	369	3.4%

The participating presses (that is, not including Oxbridge and without extrapolations) published a total of 369 primary monographs in art over the five-year period. That came to a small percentage (3.4%) of the total output of primary monographs in the humanities (the rightmost column represents the number of humanities primary monographs).

It's a different story when we look at history:

SUBJECT AREA DATA - PARTICIPANTS - HISTORY	2009	2010	2011	2012	2013	TOTAL	% OF HUM. PRI. MONO.
GROUP 1	100	106	99	99	111	515	29.0%
GROUP 2	71	74	82	94	85	406	28.0%
GROUP 3	259	218	202	210	227	1116	24.9%
GROUP 4	216	235	212	181	176	1020	34.1%
ASSOC	0	0	0	0	0	0	0.0%
TOTAL	646	633	595	584	599	3057	28.5%

For history the total output of primary monographs was 3,057 titles, 28.5% of the total output (of primary monographs in the humanities). The sheer amount of history publishing makes us wonder what actions would be supported by this data. Could the American Historical Association be coaxed into providing support for university presses? Would history departments on university campuses be open to providing subsidies to their university presses?

While literary criticism may be under financial pressure, the presses clearly have not given up on this category:

SUBJECT AREA DATA - PARTICIPANTS - LITERARY CRITICISM	2009	2010	2011	2012	2013	TOTAL	% OF HUM. PRI. MONO.
GROUP 1	69	45	59	95	74	342	19.3%
GROUP 2	45	45	53	58	65	266	18.4%
GROUP 3	112	120	91	91	87	501	11.2%
GROUP 4	72	66	74	67	60	339	11.3%
ASSOC	0	0	0	0	0	0	0.0%
TOTAL	298	276	277	311	286	1448	13.5%

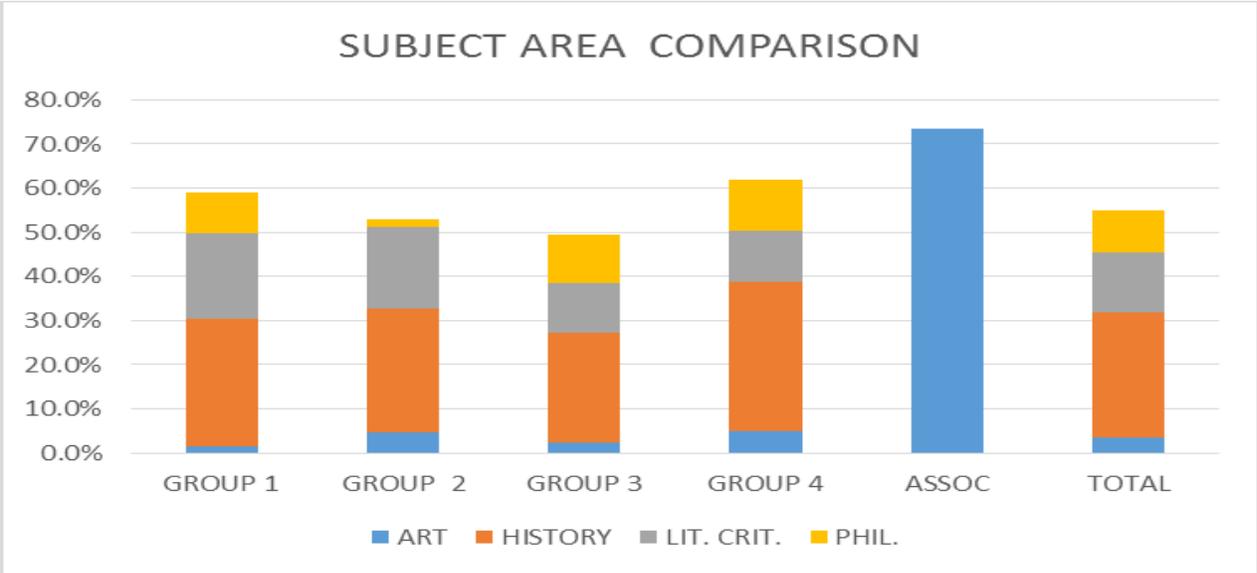
There were 1,448 titles in literary criticism over the five-year period, fully 13.5% of the total output of primary monographs in the humanities.

Finally we get to philosophy:

SUBJECT AREA DATA - PARTICIPANTS - PHILOSOPHY	2009	2010	2011	2012	2013	TOTAL	% OF HUM. PRI. MONO.
GROUP 1	31	25	26	41	41	164	9.2%
GROUP 2	3	7	9	3	4	26	1.8%
GROUP 3	95	97	66	123	111	492	11.0%
GROUP 4	89	54	65	72	65	345	11.5%
ASSOC	0	0	0	0	0	0	0.0%
TOTAL	218	183	166	239	221	1027	9.6%

Philosophy accounted for 1,027 titles or 9.6% of total output of primary monographs in the humanities.

Here is another view that compares press output by subject area, which serves to underscore how heavily weighted toward history monograph publishing is. The y-axis of the graph represents the percentage of total output for primary monographs in the humanities. Thus the Group 4 presses, for example, have just over 60% of their output in primary humanities monographs in just four subject areas, with history being the largest by far.



We included the Associates in this chart, as it illustrates how lopsided their output is, a function of the specific missions of these organizations and the level of participation in the survey.

In summary, over the five-year period of this study, the reporting presses exclusive of Oxbridge published a total of 10,723 primary monographs in the humanities, of which 369 (3.4%) were in art, 3,057 (28.5%) were in history, 1,027 (9.6%) were in philosophy, and 1,448 (13.5%) were in literary criticism.

In reviewing these figures, we don't see anything to suggest a sharp decline or uptick in the output for any of these fields. While no one would suggest that university press publishing is a growth business, the chorus of doom that one so often hears seems to be exaggerated.

F. Pricing

While there was some variation for pricing among the different categories of presses, overall we found the pricing to be more consistent than not. (Oxbridge is not included here.)

PRICES	GROUP 1	GROUP 2	GROUP 3	GROUP 4	ASSOC.
AVG. ALL TITLES	\$ 33.70	\$ 37.05	\$ 41.82	\$ 39.46	\$ 40.38
AVG. ALL PRIMARY MONOGRAPHS	\$ 41.78	\$ 47.78	\$ 45.72	\$ 52.86	\$ 54.45
AVG. ALL HUM. PRIMARY MONOGRAPHS	\$ 42.05	\$ 48.15	\$ 45.56	\$ 48.77	\$ 43.72

Pricing touches on the matter of formats and of market segments. For the most part, presses identified the cloth edition as the primary publication. This is not surprising, as cloth editions have been the mainstay of scholarly publishing for years. The prices in this table reflect mostly cloth prices, which is why they are so high. Libraries may be comfortable spending \$45 and more for a book, but few individuals are; and certainly almost no student wants to pay that much. Thus the pattern is to publish a high-priced cloth edition aimed toward the library market, to publish a lower-priced digital edition (often through Amazon or other online vendors) to bring some individuals into the marketplace, and to use paperbacks for classroom use. We don't have data to suggest whether ebooks are taking much market share away from paperbacks in the classroom,

though we suspect that ebooks are indeed making inroads, largely on the strength of lower prices. We have no data whatsoever for the pirated digital editions, which some students (and perhaps some faculty) download over the Internet, but their price is zero, which tends to put downward pressure on the pricing of the other editions.

Thus someone could purchase a monograph in the humanities (but probably not a *primary* monograph in the humanities) for a lower price than shown in this table, but it would be a paperback or ebook reprint of the primary title.

G. Subventions and Advances

Many of the presses were unable or unwilling to provide this data, but the following table shows what information we have:

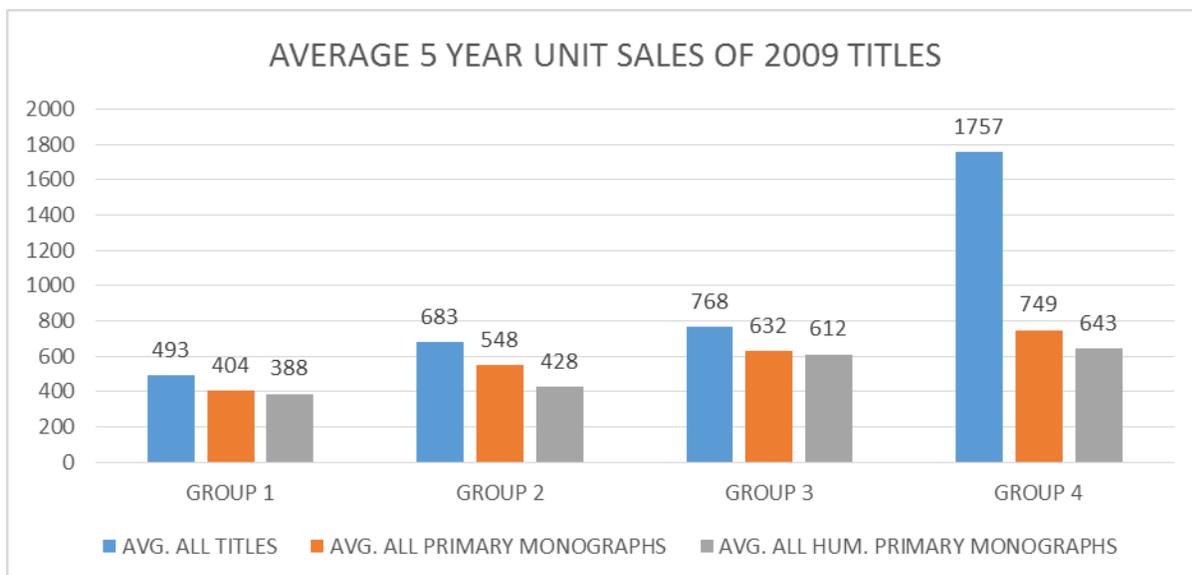
SUBVENTIONS	GROUP 1	GROUP 2	GROUP 3	GROUP 4	ASSOC.	TOTAL
TOTAL FOR 5 YEARS - PARTICIPANTS	964	1129	3727	1752	148	7720

ADVANCES	GROUP 1	GROUP 2	GROUP 3	GROUP 4	ASSOC.	TOTAL
TOTAL FOR 5 YEARS - PARTICIPANTS	536	559	5273	3667	819	10854

As noted earlier, we found the data on subventions to be useful (because it demonstrated an institution's willingness to support its faculty), but the advances to be less useful (because they are usually paid only on titles with stronger commercial potential. Overall the presses reported receiving 7,720 subventions over the five-year period, for an average of 1,544 per year. While we don't know if subventions went to titles outside of the focus of this study (that is, for trade books or monographs in fields outside the humanities), it seems probable that most of the subventions went to monographs because those are the books that struggle in the marketplace.

H. Some Other Fields of Data

Some of the other fields of data provide interesting views into the press community. For example, see this table with unit sales of titles first published in 2009:



We are using the figures for 2009 publications because they permit us to evaluate five years of sales history. What jumps out here is the extent of the Group 4 publishers' non-monograph business. So while primary monographs in the humanities struggle to sell something just short of 600 copies, the total unit sales for all titles (including monographs) is over 1,700.

Also revealing is the information on book formats:

FORMAT	GROUP 1	GROUP 2	GROUP 3	GROUP 4	ASSOC.
AVG. 5 YR. SALES - ALL CLOTH BOOKS	418	757	505	1562	4595
AVG. 5 YR. SALES - ALL CLOTH HUM. PRIMARY MONOGRAPHS	347	475	586	548	496
AVG. SALES - ALL PAPER BOOKS	473	467	725	1280	4302
AVG. SALES - ALL PAPER HUM. PRIMARY MONOGRAPHS	350	446	494	721	721

The average five-year unit sales for cloth book shows a great disparity between the Group 4 presses and the presses in Groups 1-3, but when we look just at cloth titles

that are primary humanities monographs, the disparity disappears. This implies that even the largest presses haven't solved the riddle of how to sell more humanities monographs, though they have had success in publishing in other areas. The figures for the Associate Members suggest that these presses have a locked-in market among their memberships, something that the conventional university presses do not.

For paperbacks the Group 4 presses and the Associates have very strong sales until we qualify this to include only primary titles in the humanities, where the Group 4 presses and the Associates outperform the other categories, but not by a tremendous amount.

We don't think this study told us much that was meaningful about ebook sales. There are a couple reasons for this. First, the study begins with data from 2009, a mere two years after Amazon launched the Kindle (the iPhone launched a few months before the Kindle in 2007). Second, the presses don't report ebook sales uniformly. We know from other work we have done that virtually all new press titles have electronic editions--the presses, in other words, are catching up with the trade--but we don't see what these digital editions tell us about monograph sales.

I. Discussion of Phase Two Findings

To summarize: the participating American university presses in Groups 1-4 published a total of 58,555 books over the five-year period of the study. Extrapolating this figure to give us an estimate that includes the non-participating presses, the total title output comes to approximately 76,000 or just over 15,000 per year. The total output of primary monographs in the humanities over five years for the participating Groups 1-4 presses comes to 10,689, or 2,138 per year. When extrapolated to all Group 1-4 presses, the total number of primary humanities monographs is approximately 15,000 or 3,000/year. Primary monographs in the humanities thus represent 19% of total press output. These figures do not include Oxbridge.

More can be made of this data in three different ways. First, some new questions could

be put to the raw data. This is very time-consuming, however, and cannot be undertaken idly. Second, new questions can be put to the scrubbed data. It is easier to work with the scrubbed data, but undoubtedly there are many questions that we did not anticipate when we scrubbed the data (which would then require us to go back to the raw data). Third, the data can be brought to bear on other data--the holdings of academic libraries, for example, which can be gleaned from WorldCat. And this in fact is what we experimented with and discuss below.

We did not attempt any extrapolations for the international members of AAUP. We don't know enough about many of them to hazard any generalizations. It may be useful to add a study of the Canadian presses, however, as many of them have programs that seem similar to the American presses. See, for example, the book programs for the presses at The University of Toronto and McGill and compare them with the presses at The University of North Carolina and Stanford.

We have come away from this part of the study with this overall impression: the presses are creating more publications than we had believed, but fewer original scholarly monographs, including fewer humanities primary monographs. A figure that has been bandied about in the press community for the number of publications per year is 12,000 or perhaps a bit more. Some people have suggested that that figure is high because it includes trade titles; the figure for monographs was put in the range of 8,000-9,000 per year. But the extrapolated figures we have derived are 15,164 for all titles, 3,889 for primary monographs, and fewer, 2,913, if we focus only on original works in the humanities. These figures are exclusive of Oxbridge and the Associates.

How to reconcile these figures? It is probable that the biggest variance is due to counting each edition (cloth, paperback, digital) rather than just the original work. Eliminating trade, regional, reference, and text titles lowers the figure some more. We then reduced the total number of monographs further by filtering for the humanities.

One data point that continues to intrigue us. We have been in touch with Palgrave

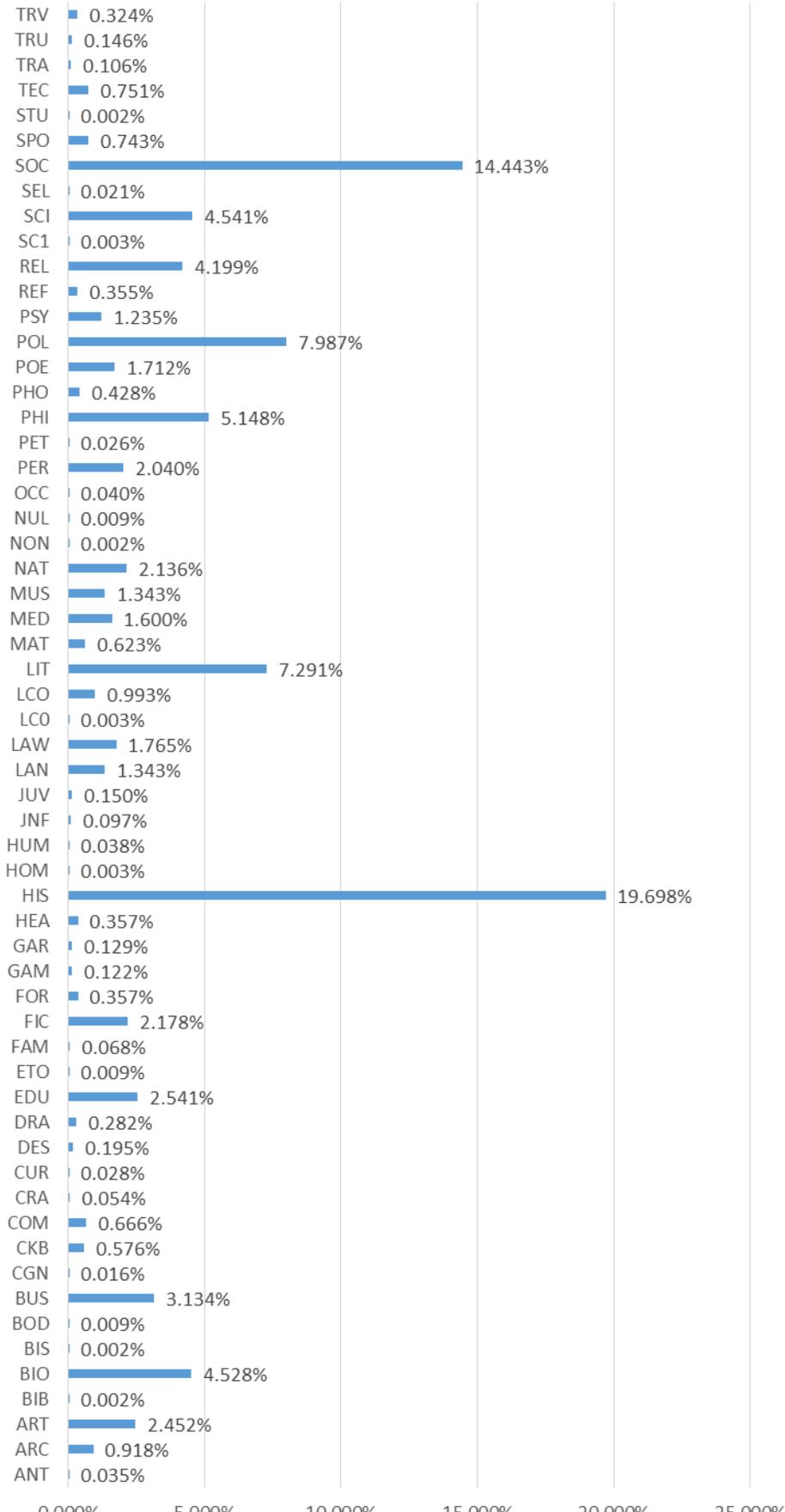
Macmillan and Taylor & Francis (Routledge), both of which have academic book programs that do not look very different from that of many university presses. The sales are almost entirely institutional. These commercial programs are significantly profitable. Is there a crisis in the monograph or is the crisis in the marketing departments of university presses?

VI. Sidebar: Output for All Titles (not just monographs)

Although our emphasis here has been on university press monographs, we wanted to take a look at the presses' overall programs. Monographs, after all, sit among many other types of publications and have the benefit of sharing organizational overhead with non-monograph titles.

The following graph shows the output of all Groups 1-4 presses by BISAC code, that is, by subject matter. The figures show the dominance of just four categories: sociology, literary criticism, political science, and history. Taken together, these four categories comprise approximately half of the total output of the American university press world.

% OF TOTAL BISAC CODES ALL GROUP 1-4 PRESSES



VII. Phase Three: An Experiment with OCLC

As noted in the Introduction, we took the hint from Rick Anderson, Dean Blobaum, and Alan Harvey and put together an experiment with OCLC. The idea was to take all the ISBNs we had gathered (summarized above as Phase Two) and to attempt to map them against the WorldCat database. If this were successful, the presses would then have a vastly expanded amount of information on their operations. For example, for each ISBN a press would be able to see its Dewey Decimal Classification (DDC) and Library of Congress number (LOC). The ISBN would also be linked to library holdings--so for example, a press could determine if that particular book (as identified by the ISBN) was in the library at, say, Dickinson College, the University of Minnesota, and Georgetown University. The promise of the WorldCat goes further than that, however, as WorldCat has international participation from OCLC's member libraries and includes libraries beyond the academy. Few book publishers have anything like a comprehensive view of the distribution of their titles.

On a fairly low level the reports from this database could have a practical impact on press operations. Why are some titles selling well in the UK but not elsewhere outside the U.S.? Is our sales organization properly constructed? Should we be reevaluating our marketing plans? It is hard to predict all the uses that this data could be put to, but it seemed worthwhile to put the data into the presses' hands and let them play with it. Our own view is that WorldCat is a greatly underutilized source of information for publishers.

It should be noted that nothing we were suggesting concerning WorldCat required our involvement if a press were highly motivated without us. Nothing is to stop a publisher from looking up titles, whether by ISBN or author and title, one by one on WorldCat. The problem is scale: how to get information on thousands of ISBNs all at once and to organize the information in a manner that would be useful for business analysis? This is what we were seeking.

We spent a great deal of time working with OCLC to try to normalize the data--that is, to allow our data to map directly onto theirs--but, unfortunately, OCLC could only provide useful data on about two-thirds of the 60,000 titles.

It's worth considering why this should be so. In our first project we had the university presses provide title-by-title information to us by using ISBNs as identifiers. ISBNs are the principal identifiers all book publishers use,²⁰ as the ISBN identifies a salable commodity. OCLC informed us, however, that librarians are often skeptical about ISBNs and typically assign other identifiers, Dewey Decimal Codes and Library of Congress numbers among them. (Of course, we were trying to get to those DDCs and LOCs through the vehicle of the ISBN.) Some cataloguers do not use ISBNs at all; in some instances the ISBNs are linked to other identifiers, including other ISBNs. For example, a publisher might provide the ISBN for a paperback edition of a book, but a cataloguer may link that ISBN to the ISBN for a cloth or digital edition. Other linkages include ISBNs for the same book whose price has changed, a unique ISBN that is connected to a specific distributor (the distributor naturally wants orders to come its way), ISBNs for revised or earlier editions, links to foreign-language versions, and ISBNs for each of several digital formats. On top of this is the distributed nature of the cataloguing operation: with so many hands involved in the process (which today we would call "crowdsourcing"), it is not surprising that some errors would slip in. At bottom these problems point to the basic asymmetry of the use of the ISBN: publishers wish to identify a specific *edition* for sale, while librarians seek to identify a specific *text* to satisfy the needs of its patrons.

OCLC took the ISBNs provided to them and introduced the *workset*: a collection of all linked identifiers. Thus for a specific ISBN, the workset might turn up a number of such linked identifiers, which would certainly be of interest to a librarian. This is not helpful to

²⁰ An emerging exception to this is the new self-publishing market, where authors may publish directly with online retailers, some of which do not use ISBNs. Amazon, for example, has its own identifier system. At this time few academic authors are working with these venues.

university press publishers, however, who will likely look to our ultimate report (private and customized to their individual press) to guide them on the holding of specific editions in specific libraries and categories of libraries. We worked closely with OCLC on the workset data, but finally concluded that it was not useful for our purposes (however useful it might be for other purposes). In the end we had OCLC do a restricted mapping in which they only turned up “direct hits” for the ISBNs we had provided. Of those 60,000 ISBNs, OCLC was able to provide information for about 40,000.

We pondered the possibility of merging the original dataset that we had generated (for 60,000 titles) with the OCLC data in the individualized reports to the participating presses, but decided that it would be misleading to do so. First, the OCLC data was missing one-third of the ISBNs. But more importantly, our data was for the years 2009-2013, whereas the OCLC was a snapshot for one moment in time, June 2015. This would not quite be adding apples to oranges, but it implies greater normalization of the data than is warranted. We thus decided that we would present reports on each dataset separately. There were two exceptions to this. We took from our dataset two fields of information: whether a book was a monograph and whether it was the primary edition. The “monograph” tag was applied by the publishers to the original dataset (“a book by scholars for scholars, but not including collections of essays”). An original (“primary”) edition referred to the first appearance of a title. Thus a title might appear in cloth, paperback, and as an ebook, all of which were designated as monographs, but only the first edition (typically the cloth version) was deemed to be primary or original.

We called the OCLC phase of this project an experiment, and it was. What did we learn from this?

The most important thing is that WorldCat is of enormous potential but of uneven practical utility to publishers at this time. It may be outside OCLC's mission to make WorldCat more useful to publishers, but the database would be much improved if in addition to the records created by library cataloguers, it imported metadata directly from publishers. The publishers' metadata could be the baseline record, and the cataloguers

would then augment the fields of information. This would effectively create an unrivalled source of information about books. Marry ecommerce capabilities to this and OCLC could become a significant online bookseller.

The second lesson is that even the partial utility of WorldCat to publishers has value to individual publishers who can then see a great deal of information about their own publications (provided that the title was found in the OCLC databases). We anticipate that publishers will increasingly find new uses for WorldCat and hope that OCLC will improve the database to accommodate their interest.

The data from the OCLC experiment will not be made public--though, we hasten to reiterate, all of the data is available on WorldCat for anyone who takes the trouble to search for it manually. We have crafted individual reports that are being sent to the participating publishers.

VII. Conclusion

There are many ways this project could be improved, though it is unclear which of these improvements would be worth the effort. Most obviously, it would be wonderful to have data through 2016 (current data ends in 2013). We would also like to see ten years of data in order to study long-term trends. And it would be great to have more participating publishers. On the WorldCat front, we would like to see enriched, mappable data that would make WorldCat a better tool for publishers and not only libraries and library patrons (for whom WorldCat provides a very good service). We believe, though, that it will be hard to do this study again for the simple reason that it is so difficult to get the data from the presses. If the presses could be persuaded to participate once again, we would propose doing this project again in 2019, thus giving us ten years' of data (2009-2018).

Overall, however, we believe that the project answers the question we set out to answer: What is the output of the university press community and of the core

monograph programs in particular? This study shows that the presses have more diverse programs than previously assumed (at least by us) and that monographs are but one component of their total publishing lists.