Making the Case: Using a Living and Historic Cultural Universe Big Data System as a Knowledge-Based Foundation for Sustaining and Recapitalizing the Cultural Sector

Louise K. Stevens

ArtsMarket, Inc., Bozeman, Montana

Published online: 13 Dec 2014.

To cite this article: Louise K. Stevens (2014) Making the Case: Using a Living and Historic Cultural Universe Big Data System as a Knowledge-Based Foundation for Sustaining and Recapitalizing the Cultural Sector, The Journal of Arts Management, Law, and Society, 44:4, 208-217, DOI: 10.1080/10632921.2014.960629

Please scroll down for article
Making the Case: Using a Living and Historic Cultural Universe Big Data System as a Knowledge-Based Foundation for Sustaining and Recapitalizing the Cultural Sector

Louise K. Stevens
ArtsMarket, Inc., Bozeman, Montana

The cultural sector needs information to support recapitalization efforts. However the majority of existing studies all point to a knowledge gap within the sector: how to evaluate the amount of capital that can be raised. Through big data, we now have the ability to understand the “universe” of support to the cultural sector. In a test employing over two million data cells from 2007 onward for cultural organizations in Charlotte, North Carolina, the author has found evident patterns and predictors for sector support as a foundation for recapitalization, making the case for big data.

Keywords: big data, cultural universe, fundraising, market analysis, recapitalization, united arts fund

The cultural sector has a need for information to support the recapitalization efforts that have, in recent years, vexed institutional funders and stymied cultural nonprofit organizations of all sizes. Several major research studies have made excellent strides by recommending steps to address planning for capitalization, but the studies all point to a knowledge gap within the sector: how to evaluate just how much capital can be raised. Among these, the TDC study Getting Beyond Breakeven, a Review of Capitalization Needs of Philadelphia-area Arts and Culture Organizations (Nelson 2009), which was conducted immediately prior to and during the recession, offers a particularly illustrative four-point rating scale that showed that even in the relatively “good” times of 2007–2008, seventy-seven percent of Philadelphia’s cultural nonprofits needed additional capital that ranged from simple operating rainy-day reserves to facility-related issues and endowment expansion. The study raised some interesting issues about the lack of critical knowledge on which to ground recapitalization efforts:

1. That despite strong financial literacy, as well as strong interest in building capacity and in strategic planning, nonprofit cultural organizations overwhelmingly posted weak financial positions—that is, even good internal planning wasn’t enough to leverage recapitalization, suggesting fundamental uncertainty that campaigns would succeed.

Address correspondence to Louise K. Stevens, 662 Coffee Creek Road, Bozeman, MT 59715. E-mail: lstevens@artsmarket.com

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/vjam.
2. That highly regarded existing benchmarking tools such as the Cultural Data Project may lead organizations to inappropriately benchmark themselves against equally uncertain peers, and thus pull back from recapitalization, again suggesting that there is a vacuum of other needed data to inform recapitalization.

According to Nelson (2009), “This type of (peer-to-peer) comparison may have unanticipated consequences for the field. Although organizations know that they need a higher degree of capitalization to reach their goals, they are often not realistic in projecting the size of the necessary resources and moreover they don’t size the market to see if strategies to garner these resources are feasible.” As truly valuable as the Cultural Data Project and other benchmarking databases such as the Urban Institute Fundraising Effectiveness Survey are, there is a missing element that is critical to serious recapitalization efforts: sizing up the market. While a feasibility study can help predict gift size from known donors to an individual institution, it cannot typically show the market of donors and donor prospects for the sector as a whole. For cultural organizations within any community, this missing piece of information spells frustration as the recession fades away, leaving behind lost investment capital deepened by double-digit annual dips into existing reserve funds, less earned income, and often many deferred capital needs.

Thanks to the arrival of big data, the legacy of missing market data may finally disappear. Big data has the ability to transform the way researchers understand the “universe” of support to the cultural sector. In general, the field has until now relied upon case studies and sample-based studies to glean information about the size and nature of the sector donor market, and how the donor market relates to the participation market. Even the largest and most continuous sampling ventures, such as the National Endowment for the Arts’ multi-decade Study of Public Participation in the Arts, or the Urban Institute Fundraising Effectiveness Survey, leave the user asking for more.

What if it was possible to observe actual participation and actual contribution patterns across an entire geographic universe and over many years? What if, in addition to multi-year sector behavior data, one could use appended demographics to study the nature of all goers and donors? The insights available could make possible a level of analysis and understanding far more locally actionable than sampling and could become the underpinning for sector-wide recapitalization. Even simple but always challenging questions such as how big the sector’s reach really is could be answered as a pathway to advocacy for public support. Other questions that could be answered:

- What share of goers also give, and what share of givers actually go?
- What similarities or differences are there in the universes of all the cultural organizations?
- Are the same individuals and foundations donors to all the organizations, or are there distinct segments of the support base for a cultural system?
- What are the demographic characteristics of goers and donors?

Knowing the answers to these questions is particularly important as the sector rebounds from the multi-year recession and serious losses in both earned and contributed income. Getting to them has been daunting, in part because of sector-wide fear. As Rebecca Thomas (2013) notes, “Many arts leaders live in fear that true openness about their financial state will jeopardize future funding.” Thomas writes of the fear nonprofit leaders have over showing the state of their various financial and capital needs, but her statement also speaks to fear over revealing the state of funding received, in general. Indeed, in over twenty-five years of conducting sector-wide audience studies,
the author has found that fear is a major barrier to many types of research studies. Add donor information to attendance information, and that perceived barrier can become huge.

But done correctly, with privacy guards in place to shield organizations as well as their donors, big data can win out over fear.

Big data are more than what is found on a “big list.” According to Meyer and Schroeder (2012), “Big data are data that are unprecedented in scale and scope in relation to a given phenomenon.” A typical big list that supports arts marketing is compiled of the mailing lists of the participant arts and cultural organizations, which then use the centralized system to swap lists to facilitate direct marketing. In some cities that have developed and currently utilize big lists, the lists have provided an annual census describing participation. Big data goes beyond this to gather every bit of information on each interface every individual in a market area has with the entire sector, accompanied by much additional information that allows for deep analysis of past, current, and future behavior.

Until the arrival of technology capable of managing and viewing big data, it was almost unthinkable to substitute the real thing for sampling. Prior to Excel 2007, for example, an Excel-based data set could not go beyond 256 columns or 65,536 rows. This was barely enough space to assess the universe of cultural participation and support in a small community, and not at all large enough to study multiple years and all relationships, combined with all factors impacting those relationships, in a large metro market. Even multi-layer databases were deficient: as late as the mid-2000s, it was not unusual to run data through the night just to standardize a single organization’s internal databases for analysis. Big data, then, is really in its infancy—not only in collection of data from multiple sources and platforms, but in the ability to standardize, condense, and manipulate it.

Big data answers questions that can’t be addressed with sampling alone. With the full buy-in of the cultural institutions within any geographic area, it is possible to create a historical and living database that not only provides census-type snapshots, but that allows for multi-year trend and predictive analysis.

This was the goal in designing a big data system and using it as the analytics foundation for the Arts and Science Council (Charlotte, NC) as a first step in planning sector-wide recapitalization in the multi-county Charlotte MSA. A total of twenty-eight partner organizations joined into the study. Each provided all of their databases of individual donors, corporate and foundation donors, government grants, as well as members, ticket buyers, subscribers, single ticket buyers, and comp/free ticket recipients, students and amateur participants, and any other attendees, such as special event supporters going back to 2007, organized by year. Capital gifts were separated out from annual gifts, and gifts such as corporate matches to employee gifts or gifts from family funds within a community foundation were also separately indicated for accurate sourcing. In addition, all e-gifts contributed via the Arts and Science Council’s Power 2 Give site were entered, as were subscribers to the Council’s weekly online calendar, CharlotteCultureGuide.com.

With the addition of FY 2012-13 data, there were over two million raw data sets that went into the Cultural Universe Master Database (CUMD). Working with that data, the research team constructed 410 cells of data per each household within the universe of individuals affiliated with the sector, and 168 cells of data per foundation, corporation, or government grant maker affiliated with the sector. To a sample of approximately 100,000 records drawn randomly from the individual attender/donor universe, key demographics were added, including household income and net worth, age, race/ethnicity, as well as presence and ages of children. In addition,
self-reported interest data sets, including arts interest and charitable contribution interest, were added.

There are challenges in standardizing data across organizations. A street address can be recorded in six different ways. Names can be misspelled, resulting in duplication. An organization may not routinely update addresses in its in-house databases, so a donor with a common first and last name who moved a few times could appear in the master database with all prior addresses, suggesting separate individuals rather than one person. Multi-year data inevitably means that more than one person per organization was responsible for the data input, leading to different data in different cells, different spelling of the same person’s name, and the same person showing up with multiple organizational identification numbers. Going beyond name and address standardization is the challenge of extracting data from many different database programs that utilize hidden formulas. A hidden formula can lurk behind what appears as a clean numeral in a cell and wreak havoc on calculations.

In Charlotte, where the united arts fund campaign managed by the Arts and Science Council is workplace-based, there were additional challenges. For example, while John Doe may buy tickets and hold museum memberships via his home address, his giving may be recorded only via his work address where he gave his annual campaign gift. In addition, his spouse or partner may be the arts donor in the household, with a different last name and different work address. John Doe may not appear to be very supportive of the arts, if all that is seen is his annual workplace gift tied to his workplace address. But when Jane Doe is found and added, and their shared household address is added, suddenly John and Jane Doe may be among the thousands of high-level, little-known, and under-valued donors we found.

These donors, if found, understood, and cultivated, can play a major role in recapitalizing the sector. There are thousands of John and Jane Does, “lost” donors whose combined household information shows they subscribe to multiple performing arts organizations, have multiple museum memberships, have children who take music lessons at the community music school, and whose memberships and ticket purchases show they have interests in history, in outdoor bird sanctuaries and science museums, in theater, music and dance, in contemporary art and studying pottery. They may alternate their giving across years and organizations: one year, a nature center may get their extra end-of-the-year gift, while the next year it goes to one of the history museums, and the year after to a theatre company. No one thinks they are extraordinary, but once found, they and others like them have the potential to turn the sector around to health. They are the lost investors, the hidden strength that can grow if cultivated in any marketplace. Likewise, there are thousands of small businesses whose multiple small gifts barely make the published gift lists, but are quite significant when their multi-organization support is visible.

As this illustrates, big data can “generate new, precise, and rapid insights into economic, social and political practices and processes. However, it also presents clear methodological, technical, theoretical, and ethical challenges and concerns” (Schroeder and Meyer 2012). Methodologically, drilling the data down to the unique household level is tremendously difficult and time-consuming. There are still thousands of households in the Charlotte CUMD for which we lack enough information to solidly associate the multiple John and Jane Doe work and email addresses to the correct John and Jane Doe street address.

Technical issues are a part of all large databases. Data visualization and predictive analytics require that data move easily from one form to the next, and that data be workable across multiple computers. Data updates can suddenly make a file implode. That last-minute scramble
to associate various data sets can set off a chain reaction that can take considerable time to find and fix. However, software programs such as Alteryx make possible what was only imaginable a few years ago.

But by far the largest issues, as Schroeder and Meyer (2012) point out, are the ethical issues and concerns. It is essential to protect the privacy of the attender or donor, and to protect the private data of the participant organizations. The twenty-eight organizations who joined together as partners in the creation of the Cultural Universe Master Database (CUMD) agreed that the data would only be visible by donor name to their independent, outside research and database development team, far removed from Charlotte, transferred from the partners using security measures. In addition, the partners have been working through a confidentiality protocol in which they have established mutual levels of confidentiality, from public—general output statistics and predictive findings—to extremely confidential. Moving through each higher confidentiality door requires greater privacy commitment from the partners involved.

Is all the time and related investment worth it? Does big data enable the field to learn what it otherwise could not? We believe it does, and in so doing, it makes possible that missing ability to size the market (Nelson 2009). As Randy Cohen and Margaret Jane Wyzsomirski (2002) noted over a decade ago in their work *National and Local Profiles of Cultural Support*, “comprehensive data on funding sources improves the quality of community planning, grant making and cultural policy development efforts, and enables changes in the cultural community to be tracked over time.” They stressed reliable data, collected consistently over time, to begin to counter the lack of longitudinal funding data about the sector.

**EARLY FINDINGS**

Indeed, even in the early drill-downs using the Cultural Universe Master Database, a great deal has been found that bodes well for big data’s use as reliable and consistent background informing for cultural planning and policy making. To demonstrate, a few of the early findings are shared here.

**Portals**

There are distinct portals to cultural support. One of the most surprising—and important for every city that is fortunate to have a united arts fund (UAF)—is that the CUMD found UAF giving is a portal to the sector as a whole, not a capstone that follows after involvement. In fact, workplace giving pre-dates participation among young adults 18–25. If these young donors can become not just engaged, but transformed into a group that demands product (Zakaris and Lowell 2008), their giving can lead to a lifetime of attendance as well as increased financial support. Lost data—they are only known by their workplace address, and only in part through email—lead to lost affiliation with the rest of the sector beyond the UAF. Without home addresses that can be tracked through National Change of Address (NCOA), partner organizations do not know who these donors are or where to find them.

Through a CUMD, however, they can be identified and the best organizations—based on predictive modeling—can take on the task of further building their interest in cultural participation. Shared system-wide effort is critical, in that the CUMD identified key donor giving years to
specific partners as 41–65, which we found to be the age group that dominates high-value, multi-
organization donors. Keeping more of the young portal entrants moving through the sector will
lead to larger commitment by more donors, eventually moving from the UAF to the partner
organizations themselves. A surprising number of 18-25 donors are high value donors, as well, so
keeping them engaged will likely lead to larger gifts as their earning capacity grows (Figure 1).

**Churn**

In the past few years, churn has been studied extensively by sector groups such as the American
Symphony Orchestra League, which found that, between 2005-2007, fifty-five percent of ticket
buyers from the previous year did not return and used this as a basis of a lower-churn effort
(Wizenried 2006). The CUMD found that it isn’t just symphony orchestras that suffer churn: it
identified a seventy-eight percent churn in the complete universe of performing arts attenders and
a 133 percent churn in museum members across six years (this includes households that lapse for
a year or two and then return again). Not surprisingly, this becomes compounded when it comes
to donors. The CUMD showed that the average number of years (not just length of consecutive
years) a donor was affiliated with any organization was two years across all twenty-eight partners.
Individuals participate in an organization as a ticket buyer or member for at least one year before
giving, which means that most donors migrate away from an organization after three years, during
which time enormous resources have been used to win them, educate them about the value of the
organization, and keep them. Then, the donors in essence disappear. The organization that invested
in them then loses out after all that early investment in donor cultivation. This local-level, sector-
wide evidence supports the Urban Institute/Association of Fundraising Professionals Survey of
Fundraising Effectiveness 2012 finding that net gain in donors after losses from 2011 to 2012
across all portions of the nonprofit field was 2.6 percent. The Urban Institute (2013) report states
that “increasing gains by 10 percentage points—would double net growth, and reducing (churn)
losses by 10 percentage points would also double the net from 10 percent to 20 percent. And,
a reduction of losses by 20 percentage points would triple the net. It usually costs less to retain
and motivate an existing donor than to attract a new one. For most organizations—and especially
those that are sustaining losses or achieving only modest net gains in gifts and donors—taking positive steps to reduce gift and donor losses is the least expensive strategy for increasing net fundraising gains.”

If partnership in a CUMD can engage partners in thinking about the sector as a whole rather than only about their own organization—which it absolutely can—those donors do not need to be lost at all. For example, predictive analysis can be used to identify what one or two disciplines and organizations donors of various ages are most likely to move to when they leave organization A. This makes it viable for portal organizations to hand off the information they have collected, and the relationship they have built, to the next logical organizations. In this way, a CUMD leads to sector-wide problem solving. Donors could be kept within the sector, their gifts would consistently grow over time, leading to greater efficiency at every stage of donor cultivation. Potentially, it would be possible to envision a sort of cooperative structure in which the organizations that do the early donor training receive a share of revenue back from the eventual organizations further along the food chain.

Under-giving

Many followers of trends in charitable giving have reported under-giving by wealthy individuals. Some have linked the phenomenon to greater expectations among the wealthy that their gifts are leading to specific results, and that such results are often not conveyed clearly enough within the arts and cultural sector. According to Osili et al. (2013), “High net worth donors continue to be impact driven. High net worth donors are consistently motivated to give because they feel moved about how their gift can make a difference, with the highest proportion of donors reporting this as a motivation in both 2009 and 2011 (72.4% and 74%, respectively).”

Based on the CUMD findings, the apparent disconnect between high-net-worth donors’ results-orientation philosophy and how the sector shapes messages deserves additional attention. In the CUMD’s multi-year view, donors with $1 million+ net worth—those most likely to be able to provide large annual gifts and capital campaign gifts—have not given close to their level of capacity, including pre- and post-recession. For example, thirty-six percent of donors with $1 million+ net worth made 2012 gifts to the sector of $1 to $49, while only seventeen percent gave gifts over $1,000 (Figure 2).

Fewer Donors

The CUMD shows that cultural sector support is back to pre-recession levels, but with gains made from many fewer donors. While annual operating support from all sources—individuals, foundations, corporations, and government—has exceeded what it was in 2007, this was achieved with a net loss of twenty-two percent in the number of donors to the sector as a whole (2012 compared to 2007). But, because of inflation, just staying even with the $28 million baseline of support from 2007 would require $29.5 million in 2012 (Giving USA 2013) (Figure 3).

This means that the $31 million received in 2011 and again in 2012 represents only a real gain of $3 million against losses realized in 2008 and 2009 of $11 million, for a net loss over the years of $8 million. Such a loss, combined with twenty-two percent fewer donors, shows the dire
need for sector-wide joint effort in recapitalization; otherwise, the sector could take decades to rebound.

Geographic Considerations

By tracking every identifiable household’s interaction and contributions to the sector over six years, the CUMD profiled geographic market penetration. Just as cultural mapping has proven to be a valuable tool in visualizing cultural assets, cultural universe mapping is a valuable tool in visualizing the geographic spread of the support universe for organizations within a single metropolitan area. Figure 4, illustrating households within the southeast US that are identified...
as participants and/or donors to the Charlotte partners, 2007-2012, shows tantalizing opportunity and vexing challenges. The actual market penetration by the partners during 2007-2012 within the Charlotte Metropolitan Statistical Area (MSA) was only twenty-six percent of all households, but this was nearly exactly compensated for by reach outside the thirty-mile radius market, as illustrated by the circle drawn on the map. One could imagine the same approximate equation in any MSA.

But attenders from outside the market area are markedly less likely to be donors than attenders inside the market area for whom support to local organizations has an emotional tug. Eighty-five percent of individual donors who gave $200 or more in any year 2007–2013, and eighty-one percent of all corporate, foundation, and government supporters 2007-2013, are from within the MSA.

This geography of support hints at opportunity in identifying overlapping attenders/donors across markets, and could suggest multi-market approaches to recapitalizing the sector. One could imagine numerous shared high-level donors, for example, between the various key markets in North and South Carolina, and shared strategies to move donors through the regional sector that would enlarge upon the work of the Charlotte partners.

It also may suggest migration out of the market to consume the arts and culture. The 2012 SPPA (NEA) study shows that, based on its sample, forty-nine percent of US households attended at least one visual or performing arts destination or event in that year. In comparison, the total CUMD reached into only twenty-six percent of Charlotte MSA households during the combined 2007-2013 years. There are, of course, many free and un-ticketed events and museums that can account for a larger overall consumption rate that would be captured by the SPPA and not by household-level data. But the CUMD raw data also contain thousands of MSA households that used local Charlotte ticket vendors to purchase event tickets for Atlanta and elsewhere, though not to the level of out-of-market households coming into Charlotte for events. There may also be thousands of MSA households that affiliate with very small, local cultural organizations that weren’t partners. Does this take participation from the evident twenty-six percent to the SPPA’s modeled forty-nine percent? It may take similar CUMD analysis elsewhere to begin to assess this.

Overall, the geographic view of the market is yet one more dynamic that illustrates the challenges in new capitalization. There is no single participation hub, but there is a localized...
sense of giving that would mean a participant from sixty miles away will give in her own community while benefiting from participating in the sector in Charlotte.

**SUMMARY**

After ten months of building and refining, the CUMD is presently in its early months of yielding valuable information to guide policy, planning, advocacy and, above all, sector-wide recapitalization. The research team has begun additional analytics that factor in demographics and donor interests, lifetime giving expectations based on when donors enter the universe, and more. Overall, it shows the important role that technology plays and will continue to play in revealing important new information about the sector, and has paved the way for others to do similar work with big data, offering greater insights about the universe of support for the sector.

Action is the next step, and it will need to continue to be guided by protocols of confidentiality and mutual respect. At present, the CUMD is informing recapitalization planning and advocacy sector-wide. It will soon also be possible to envision sub-groups of the partners working cooperatively to transform attenders to annual donors, and gradually move their databases about them to other partners who will benefit in future years. For now, the Arts and Science Council continues to build the CUMD. The partners have the added benefit of system-wide data management protocols that are strengthening their internal databases pre-handoff to the CUMD and, as a group, the partnership is invested in learning and transforming through a knowledge foundation of big data.

**REFERENCES**


Thomas, Rebecca. 2013. “Capitalization, for Arts Sake!” *Grant Makers in the Arts (GIA) Reader*, Fall.


