UNDERSTANDING THE CHANGING MARKET VALUE OF A CONSERVATOR: DATAMINING THE DISTLIST ARCHIVES

Matt Cushman, Andrew W. Mellon Fellow in Paintings Conservation, Worcester Art Museum

What Do We Know About Our Market Value?

The in-practice market value of a professional encompasses countless tangible and and intangible qualities that one brings to a job, and although there can be great and varied benefits to working as a conservator, the most easily quantifiable compensation for one's contribution is salary. Conservators are in murky water with respect to compensation information in that there is no annual salary survey for conservation professionals. While the salary surveys administered in the past have produced highly detailed snapshots of a profession at a given time, they have fallen short when it comes to identifying long-term trends.

This presentation touches upon a facet of ongoing independent research aiming to leverage readily available public data so that conservators may be more aware of our successes and shortcomings in terms of encouraging long-term economic sustainability, higher standards for professional satisfaction and happiness, and greater diversity of background, thought, and experience. In particular, a segment of data culled from communications stored within the Conservation DistList Archives is presented, highlighting a somewhat alarming trend identified during this study.

Datamining the DistList

One of the greatest resources for information about the concerns, queries, and professional goings-on of conservators is the Conservation DistList and its online archive. Since its inception in 1987, over 4500 positions have been advertised on the DistList. While one could attempt to obtain accurate salary histories from institutions and individuals, it is far more convenient and less intrusive to pull such data from these job advertisements.

In brief, to accomplish this feat, the entirety of the DistList Archives was downloaded and merged into a singe plain text file. A parsing script, written in Perl, was run line-by-line over the master text file, pulling out key words, phrases, and numbers of interest -- such as date, job title, number of positions in that title, specialty, years of post-graduate experience required, and maximum and minimum salary, etc. —and inserting that data into a database. Email communications such as those on the DistList are unstructured and messy by nature. To assist in the transformation and clean-up of the raw data, a tool called
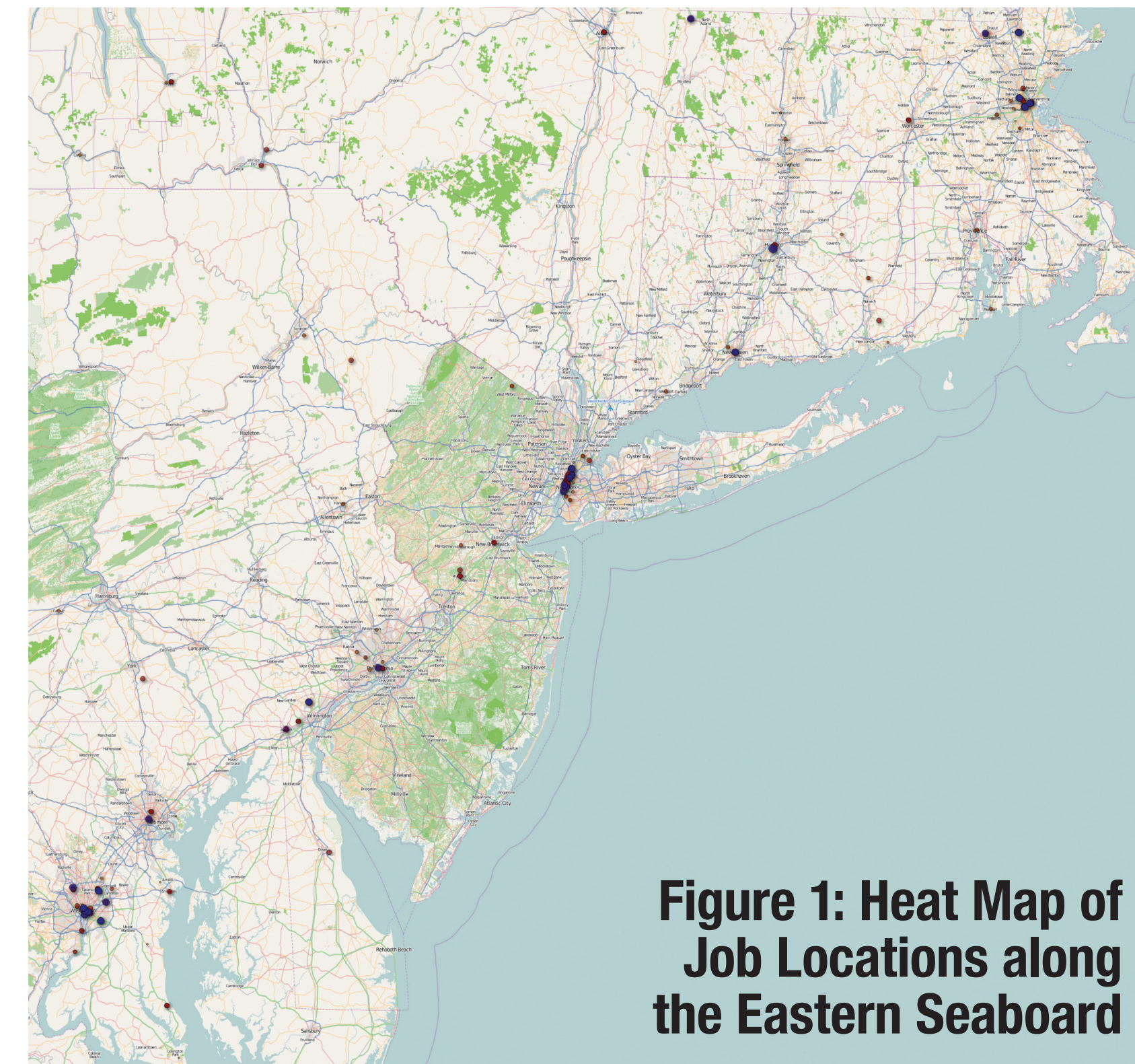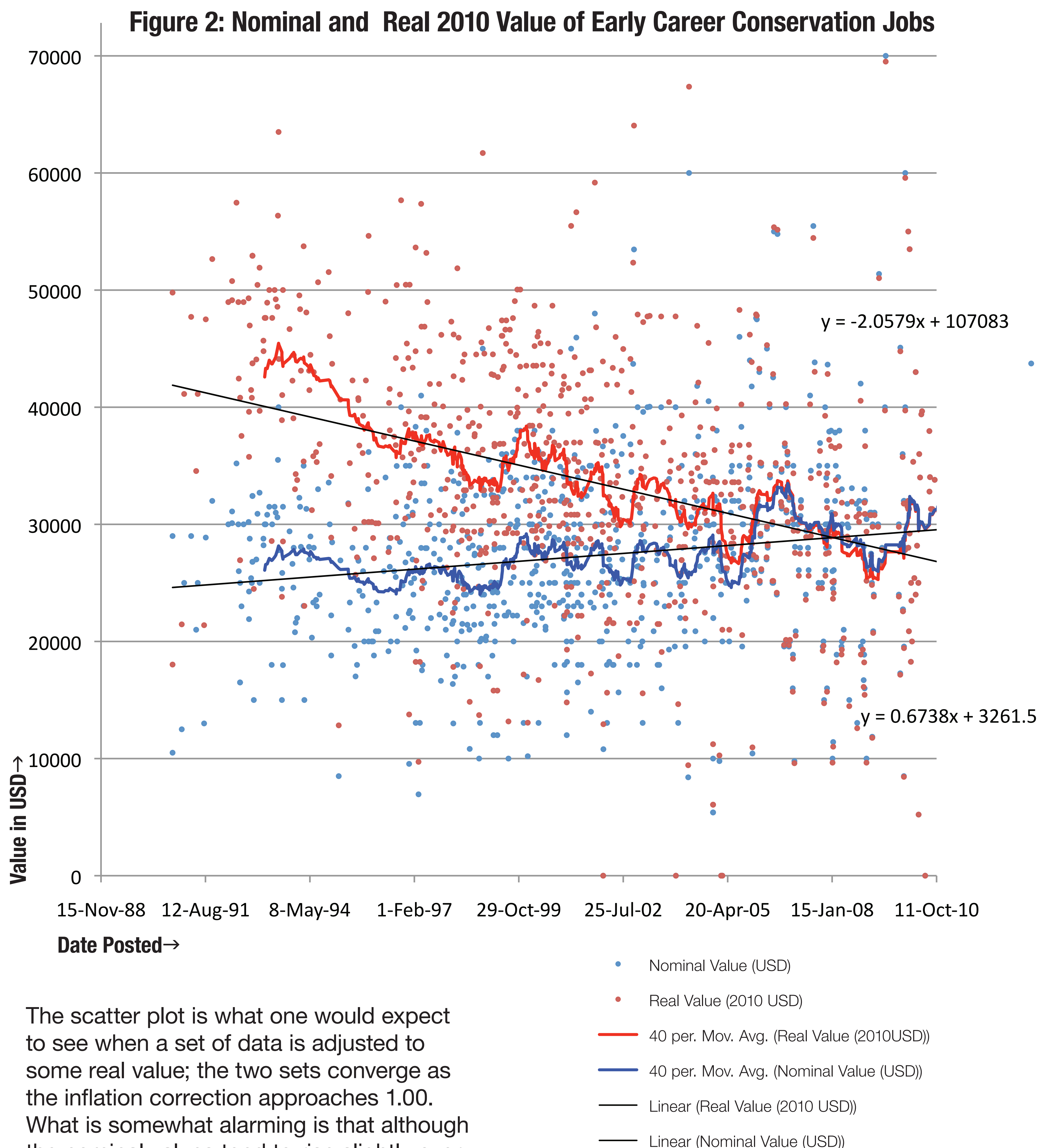

Figure 1: Heat Map of Job Locations along the Eastern Seaboard

Google Refine was used. Once the data were in a usable form, they were filtered and faceted using Google Refine and finally examined in Microsoft Excel.

A natural language processing and machine learning approach is currently being developed so that a finely tuned pipeline is in place for similar but cleaner analysis of job postings from various sources.

What to do with all of this data?

The data can be sliced and represented any number of ways. For instance, Figure 1 shows a heat map of jobs along the eastern seaboard with the size and color of the dot correlating to the number of years of post-graduate experience required to be considered for the position. The salary data are troublesome in that there are inconsistencies in currency and there can be great variation in salaries between, say, a department head and a post-graduate intern, or even between the head of a division and a senior conservator. For this particular study, only postings for jobs within the United States were considered. Finally, the segment of the data with the greatest percentage of listed salaries and therefore the most likely to be representative of an underlying trend in salaries was considered: that of the early career conservator. Figure 2 shows the nominal listed salary for nearly 900 job postings between 1990 and 2010 as well as those same salaries adjusted for inflation using a modified Consumer Bundle index. The Consumer Bundle was chosen because it is essentially a measure of a unit's purchasing power in terms of goods and services, including items like insurance premiums. A 40-unit running average for both the nominal and real 2010 values is provided to give a sense of local trends over time.

Figure 2: Nominal and Real 2010 Value of Early Career Conservation Jobs

$y = -2.0579x + 107083$

$y = 0.6738x + 3261.5$

Value in USD→

Date Posted→

- Nominal Value (USD)
- Real Value (2010 USD)
- 40 per. Mov. Avg. (Real Value (2010USD))
- 40 per. Mov. Avg. (Nominal Value (USD))
- Linear (Real Value (2010 USD))
- Linear (Nominal Value (USD))

The scatter plot is what one would expect to see when a set of data is adjusted to some real value; the two sets converge as the inflation correction approaches 1.00. What is somewhat alarming is that although the nominal values tend to rise slightly over time (a very rough estimate of approximately $250/year is given by the linear regressions), the real value of the same kinds of positions appears to be decreasing roughly $750/year over that twenty year span. Whether the same can be said for positions requiring greater experience and those outside of the US will be the focus of further study.

The causes for this phenomenon are unclear. Regardless, given the current economic climate and the toll it has taken on nearly all cultural institutions, it would be naive to expect this trend to change overnight. However, it is hoped that, with greater awareness and further curious investigation of data conservators have produced over the decades, these and many other issues can be surely addressed as the field continues to grow and develop.

References

- Google-Refine: code.google.com/p/google-refine
- Lawrence H. Officer & Samuel H. Williamson, "Measures of Worth," MeasuringWorth, 2010.
- OpenHeatMap: www.openheatmap.com