New York Philharmonic Attendance Collection

The purpose of this data set is for the collection of musician attendance for the New York Philharmonic between the dates of May 24th, 1872 and April 5th, 1878[1]. The data is organized as a spreadsheet, with the columns representing the date as well as type of performance (either being an extra private rehearsal, private rehearsal, public rehearsal, or concert), and the rows containing a list of the performers names. In each box in which the performers name and performance intersect, there is a symbol, typically being either an X for having attended or an O for being absent. While there is much homogeneity in the manner in which the data is presented across the entire book, in comparing the spreads, there arises several interesting differences. First, is the inclusion of a header in the top left corner of the left page that will indicate the number of the current season and the time period in which that season is taking place. While this is included on most pages, the two or three last pages of a given season will not include this header, seemingly for a crude manner to easily flip through to find the next given season in the book. Another instance of irregular inclusion of information can be seen on the rightmost section of each spread, with two lightly penciled in vertical columns labeled 'amount of time' and '$' (although these two data tables seldom appear, and are only seen in a handful of pages). Continuing on this trend of peculiarity is the symbols used in marking the musicians attendance, which while primarily being the aforementioned X and O, will occasionally use other symbols throughout the progression of the book in spite of no key to deduce their meaning, including 'S', '//', 'ø', 'OX', 'XX', '_', and '*I*X'. This is one small instance of a larger trend in the data set of a reliance on extrinsic markings that, while difficult to make sense of from an outside perspective, clearly serve a valuable purpose in the gathering of attendance data, with another example of this

[1] Attendance Book, 1872-1878, 24 May 1872 - 5 Apr 1878, Folder 498-02-03, Board of Directors Records, New York Philharmonic Leon Levy Digital Archives.
https://archives.nyphil.org/index.php/artifact/c245a42b-2c47-4e28-9a5b-af3714e9e5e9-0.1

being evident in the seemingly random usage of orange pen in certain pages, as shown in example 1. From this, it is applicable to question the capacity digitalization serves in the preservation of data such as this, as well as analysing how much information is sacrificed when viewed through the lens of such a contrasting medium.

In first answering these questions, it seems appropriate to look back and trace the evolution of the collection of musician attendance data for the New York Philharmonic. In doing so, what is immediately apparent is the overlapping of time periods such records share with each other, with books dating between 1871-1875, 1872-1878, and 1875-1880. In analysis of these records, we see a stark contrast in the forms of record keeping of this 1872-1878 book with the others despite serving to collect the same data, indicating that it was an attempt at a new system of organization that was never continued. To best highlight the sheer difference in these books' systems of organization, holding them in juxtaposition we can extract the entirety of the spread in example 2's information[2] in the tiny outlined box of example 3. Yet in tracing the entire available history of musician attendance data collection from 1871 to 1903, the 1872-1876 book remains the sole record to use this far more compact spreadsheet system of organization, but why?

[2] Musician Attendance Book, 1871-1875, 10 Nov 1871 - 24 Apr 1875, Folder 498-02-02, Board of Directors Records, New York Philharmonic Leon Levy Digital Archives.
https://archives.nyphil.org/index.php/artifact/a61cbd21-5c50-43b9-bbd9-36a2806d6694-0.1

In seeking to consolidate this seemingly logical discrepancy, it seems useful to refer to Foucault's work, 'The Order of Things', and in particular, the importance of considering the difference in preconditions that govern a society's understanding of knowledge, or rather, the time periods episteme. It is through an amalgamation of abstract and underlying forces that molds the larger systems of understanding within a society, and ultimately informs how one behaves and interacts with information. Foucault argues that reason itself is derived solely through the comparison of measurement and order. With measurement enabling us to "analyse like things according to the calculable form of identity and difference,"[3] (Pg. 53) and built upon a system of order, which while not contingent on an exterior unit, is built through establishing connections between like things. With this, it is possible to understand how differences in epistemes are achieved through the linking of measurements of the real world to contrive and formulate different systems of order, of which is entirely contingent on the society itself to deduce, as Foucault writes, "without imagination, there would be no resemblance between things." (Pg. 69) Within this framework, it becomes possible to, at least slightly, dissociate from our own contemporary culture in seeking to understand the extent to which different systems of organizing data may serve greater functionality to a given society as determinant of its larger epistemological system.

In focusing on the differences in organizational systems used by the New York Philharmonic, being that of the spread-sheet and more basic ledger systems, we are able to better understand how such a seemingly inferior form of data collection may in fact have better suited the needs of the philharmonic at the time. In comparing the historical usage of a ledger and that of a spreadsheet, there is a substantial difference between when the peak popularity of each would occur, with a clear and traceable connection between the height of their usages to the

[3] Foucault, Michel. *The Order of Things*. 2nd ed., Routledge, 2001.

available technologies existent within society at the time. While there is well documented usage of paper spreadsheets with large roots in the purpose of accounting from beyond the 19th century[4], it would appear to remain a relatively obscure system for a purpose such as this, which would evidently change during the early to mid 20th century with the surge in computational advancement to make the spreadsheet a vital and most prominent tool in the organization, storage, and analysis of data, with an increase in its applicability for more widespread purposes. Without this added factor, it is understandable as to why it would exhibit far less prevalence during the late 19th century and why the list-formatted attendance ledger would be considered the easier and more appropriate system to use here. Another important factor to consider is comparing the ease of access/alteration of information with the degree to which you are able to optimize its storage. It seems in the modern age that this tradeoff is largely abated as computer memory storage and compression algorithms continually advance, allowing for a far larger degree of ease in accessing and sorting through inconceivably large data sets. In examining the broader context, it would make sense as to why the complexity in a grid system may have served as an unnecessary, if not counterproductive, addition towards the tracking of attendance. In the modern context, it is far easier to consider the advantages that a more complex spreadsheet system would serve when considering modern computational technology, yet when bringing into consideration the prevalence of nondescript markings outside the confines of the provided framework, one may question its applicability to the fast paced and quickly changing context of live performance.

It is revealing to look at a system of data organization with a completely different context, yet with a somewhat similar purpose, as seen in example 4 of a loading manifest for

[4] Power, D. J., "A Brief History of Spreadsheets", DSSResources.COM, World Wide Web,http://dssresources.com/history/sshistory.html, version 3.6, 08/30/2004.

Operation Neptune in 1944[5]. While this system includes five additional points of data in comparison to the attendance sheet, the most notable difference is the usage of the drop order of the parachuters, having the order of names be a point of data in and of itself. Since the attendance sheet of the philharmonic did not necessitate a particular order of the musicians, it is understandable that having a fixed list may have only negatively affected the system of organization, not allowing the organizers to only list the given musicians of a certain performance, but instead requiring them to actively go through and mark all of those not participating as absent.

From here, leaves the question of digitalization, and while it is seemingly easy to imagine how such a grid like system of organization would fit onto a digital spreadsheet, there is a much larger question over the amount that is both gained and lost in doing so. With ~57 pages of performances containing ~18 performances each, there are ~1026 performances in the entire book, which combined with ~15 performers per performance, leaves ~15,000 points of attendance information. While such information would seem to warrant a binary value of either present or not present, the intricacy of the real world would instead require a possibility of ~6 values of attendance status for all of the symbols included. However, given that the data follows a somewhat clear layout throughout the book, its digitalization to the spreadsheet could be conducted in a rather timely manner. From the digitalization of this information, one could easily have access to a trove of insightful and useful information tracking much more than attendance, but also see analytical information about the regularity of performances, the average size of practices, and so forth. What's more is the difficulty in determining what information, which

[5] "D-Day Jump Manifest Document." Donated by Winfield Wood, Air Mobility Command Museum, 18 Apr. 2015, https://amcmuseum.org/collections/d-day-jump-manifest-document/.

could theoretically be extracted from the book, should in fact be included in its digitalization. Were a performer to be initially included but have their name crossed out, should there be documentation of this? Furthermore, should the side notes be included as well? A particularly interesting example of this can be seen in example 5, with the markings next to three performers reading ‘dead’, ‘absent’, and ‘lost’. In examining the degree to which the data has already been digitized, there has been a large, yet incomplete, number of pages converted to text listing the names of all the performers in a very similar vein to a digital version of example 2, however there is no organization done in a spreadsheet manner to which more thorough analysis of the data could be easily conducted. In implementing a plan to digitize a cohesive and contextualized form of this record, I would argue for the inclusion of footnotes and crossing outs, while still maintaining it in a spreadsheet format. It is clear that while much of the record is centered around a spreadsheet, in best conserving its intent and purpose, such seemingly extraneous additional information is necessary.

{{datasets/schafferexample1}}

{{appendix/schafferexample2}}

{{appendix/schafferexample3}}

{{datasets/es5045_d-day-jump-manifest-document}}

{{appendix/schafferexample5}}