

Keeping Count: Archiving Women’s Hockey Analytics for Accessibility

Carleen Markey¹[0000–0003–0629–8840], Nix Dube, Alyssa Longmuir, Mike Murphy, An Nguyen, Jessika Ross, and Anna Severn

Carnegie Mellon University, Pittsburgh, PA 15213, USA carleenmarkey@gmail.com

Abstract. Women’s hockey analytics has historically lacked a centralized repository for research, data, and other projects, despite other areas of hockey analytics having such central resources. In this paper, we attempt to fill in this missing piece of women’s hockey analytics by holding an archiving event in which volunteers methodologically gathered as many details on past women’s hockey research and data as possible. Each piece of research and data was then turned into an entry on MetaHockey according to standardized instructions. This event resulted in almost one hundred new women’s hockey focused entries on MetaHockey, whose characteristics largely align with trends in men’s hockey analytics. Examining these entries also empirically reveals an exponentially increase trend in women’s hockey analytics entries year-over-year, demonstrating that both continuing to archive works and taking advantage of this new research in the private and public spheres is conducive to the growth of the field.

Keywords: Women’s Hockey · Hockey Analytics · Archiving.

1 Introduction

Women’s hockey has gained considerable momentum in recent years internationally, and the men’s hockey analytics movement has followed a similar trajectory in the same timeframe. The women’s hockey analytics research and data sources have grown exponentially thanks to these explosions in popularity of the two central components. However, this community has historically lacked easily accessible centralized sources of data and projects, raising the barrier of entry of an already niche subject. Additionally, events focusing on women’s hockey analytics have begun to occur, such as the Big Data Cup [15] and WHKYHAC [8]. Building off of previous works is a crucial aspect to both events and to the progress of this field of research, growing the need for an easily accessed archive of projects.

There have been several websites that have attempted to create such a centralized resource. Even-Strength [1], TheirHockeyCounts [14], CWHL Tracker [2], and pick224 [9] have centralized summary statistics, respectively for advanced PHF statistics, SDHL & NCAA D1/D3 counting statistics, and various leagues’ & international competitions’ counting statistics. While being quite

thorough in their compilation of current women's hockey statistics, they do not document the research and projects being developed in this space.

MetaHockey [4], a site for archiving men's hockey analytics projects, was founded to fill this void in hockey analytics in general. Until October 2021 however, the website contained only seven entries and data sources related to women's hockey [5]. As documented in the WHKYHAC presentation "Contextualizing Historical Data and Current Projects in Women's Hockey" in July 2021, well over ten times that many public women's hockey analytics projects have been created since 2015 alone [10]. Additionally, the authors of this paper, as major participants in the women's hockey analytics research, determined that MetaHockey's organizational system in its current form, does not adequately serve the women's hockey analytics researchers' current needs of an archival website. Among other things, the website is hard to navigate when trying to find code repositories in data, the tagging system does not differentiate between various women's leagues as it does with men's leagues, and documented advances in women's hockey analytics often do not take the form of formal books or articles, which are the two categories available for publications submitted to MetaHockey. Published advances in women's hockey often take the form of twitter threads or Tableau-based tools, which do not fall under either of these categories.

In this project, we take the first step towards fully satisfying the need for easy access to historical women's hockey projects and data sources, as well as continuing MetaHockey's original purpose of serving all sides of hockey analytics. We do this by compiling a detailed list of as many women's hockey analytics projects as possible that were publicly accessible as of October 2021 and adding them to MetaHockey's article repository, with permission and help of MetaHockey site owners and editors. Modifying the MetaHockey website itself to serve users better is left to a future project.

2 Methodology

To proceed with adding women's hockey analytics works to MetaHockey, we followed the methods below in designing the archiving process, designing the archiving materials, putting on the archiving event, and uploading everything to MetaHockey.

2.1 Designing the Archiving Process

Since the authors observed that there is no common publication spot for women's hockey analytics works except for Twitter threads, the most effective way of obtaining the maximum amount of publications, events, and resources was first creating a collaborative list of the people who have created women's hockey analytics projects and compiled data sources, a list of known websites of compiled data sources, a list of direct data sources, and a list of events featuring women's hockey, such as conferences. This is the "To Archive" document [3]. Then,

specific publications, events, and data sources that would become MetaHockey entries would be gathered by searching Twitter and Google for each person/data source/event on the four lists, and creating entries in a Google Sheet ("Whokey MetaHockey Entries") for the publications, events, and data sources found to be related to them [16]. The "Whokey MetaHockey Entries" would then be copied into the Google Sheets-based MetaHockey back-end to get all the entries onto MetaHockey.

This is a time- and labor-intensive process, and the authors recognized the expedited need for the completion of this project by the beginning of the Big Data Cup in spring 2022. As a result, a call was put out for volunteers to help with the searching for and creating MetaHockey entries, and a date was set for an event in which some of the authors would be available over Zoom to help with both [13].

An additional note on this method: simply searching something like "women's hockey analytics" or "women's hockey data" in Twitter's or another website's search engine would have not returned the maximal results for archival entries, as creators tend to title their projects and datasets with the relevant league and area of study/statistics, as seen in entries 714-812 of now-archived women's hockey analytics projects [4].

2.2 Designing the Archiving Materials

Once the three lists were compiled and the overall methodology distilled, an instructional guide, "How to Archive", was designed for volunteer archivists to use for each entry type [11]. The first three pages outline exactly how to go about gathering entry details and adding them into the central archival spreadsheet for each list [16]. The first page of the guide is shown in Fig. 1 and was designed to be used to search for entries using the "people" list from the "To Archive" document.

The second and third pages of this guide are similar to the first in general flow, but with specific modifications for collecting entry data using the "events" list and the "websites"/"direct data sources" lists respectively.

The fourth page, "Creating Entries", continues the workflow from pages 1-3, and outlines how to format the details for each possible entry into the MetaHockey specific format and enter it to the "Whokey MetaHockey Entries" sheet. This fourth page can be viewed in Fig. 2. It is important to note here that volunteer archivists chose the keywords for each entry, as they were women's hockey analytics researchers and therefore familiar with the source material or had help from members like this.

The fifth and final page contains an appendix of common terms used in the "How to Archive" document and instructions on how to select keywords from a suggested list. Keywords not on this list were also able to be added manually for an entry.

*content = anything whacky related

DATA COLLECTING - PEOPLE

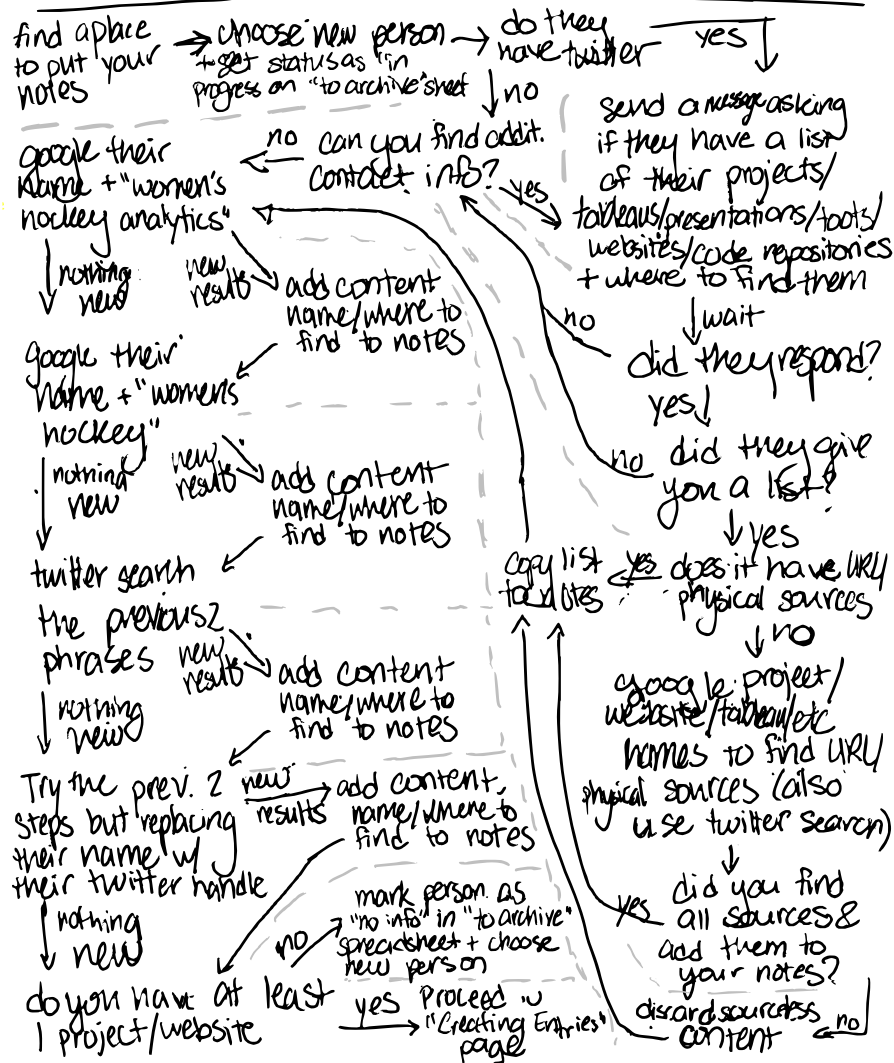


Fig. 1. The first page of the "How to Archive" instructional guide designed to guide a layperson through how to start with a name from the people section of the "To Archive" document, search for the publications, data sources, and events they have been involved with and create detailed MetaHockey entries from those search results.

CREATING ENTRIES

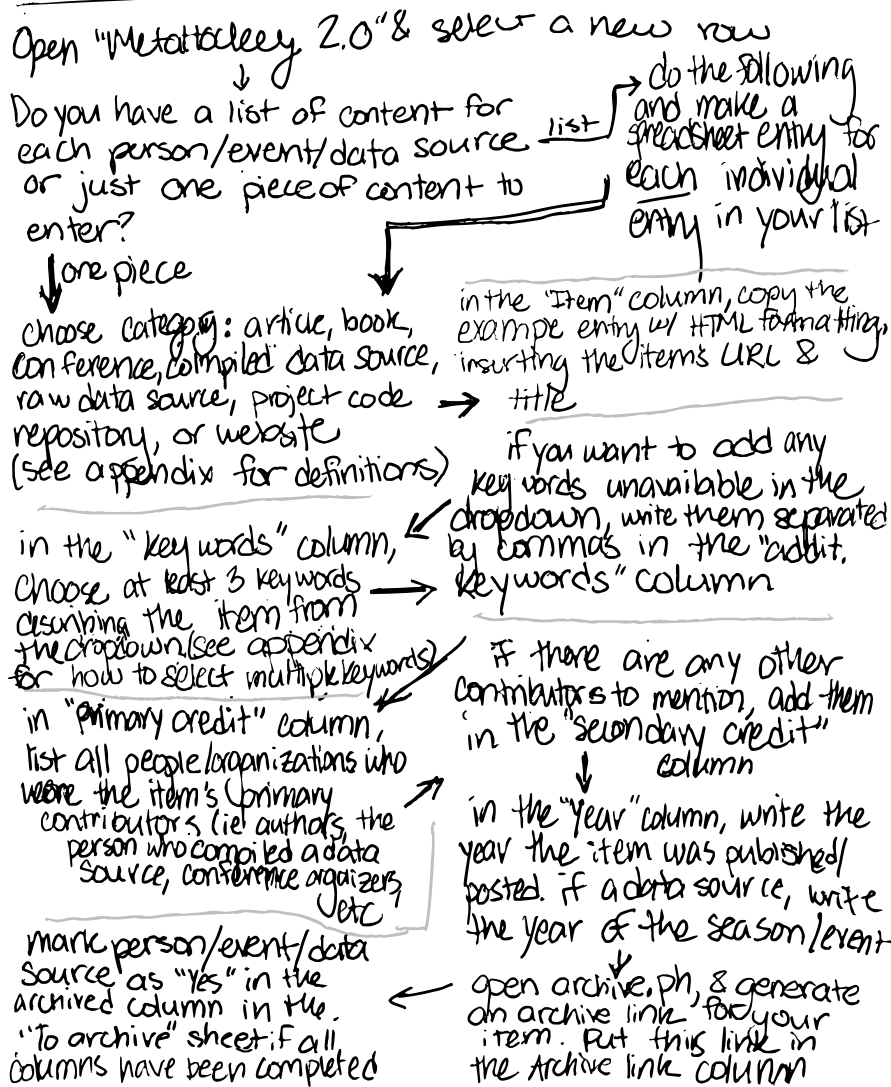


Fig. 2. The fourth page of the "How to Archive" instructional guide designed to guide a layperson through how to format entries into acceptable MetaHockey format and enter it into the "Whokey MetaHockey Entries".

2.3 The Archiving Event and MetaHockey Upload

The archiving event that used this process and these materials occurred on Oct. 23rd, 2021 from 5-8pm EST, with an option for volunteer archivists to keep adding to the "Whokey MetaHockey Entries". Once the event was over, duplicate entries were removed and chosen keywords were checked to be accurate in the "Whokey MetaHockey Entries" Sheet.

Unfortunately at this point, the authors of this paper lost contact with the main editor of the MetaHockey site, and spent the next several months reaching out to various editors of MetaHockey to see if they had site access. The last step of uploading entries onto MetaHockey was finally completed on Feb. 11, 2022 when an editor with site editing access was finally contacted and they agreed to do the current Google Sheet upload, as well as future uploads of entries.

Inevitably with the method used in this paper, publications, data, and events will be missed, since the three lists relies on the authors' memories of such things and ability of volunteers to get accurate search results. Nonetheless, we proceed with this method because the goal is progress, not perfection.

3 Results

As a result of this archiving effort, there are 98 new women's hockey analytics projects, data sets, and research tools on MetaHockey, for a total of 105 entries of the 812 existing MetaHockey entries in the Articles section being works pertaining to women's hockey. Given that this is the first quantitative survey of the field of women's hockey analytics, after qualitative examinations at OTTHAC 2022 [12] and WHKYHAC 2021 [10], it's important to briefly examine these entries statistically. Starting with Table 1, the counts of women's hockey analytics entries are broken down by MetaHockey category label.

Table 1. Popularity of MetaHockey categories among women's hockey entries.

MetaHockey Category Label	# of Entries
Article	33
Files - Raw Data	18
Files - Compiled Data	16
Website (Blog / Tableau Profile / Stats / Etc.)	14
Book	7
Conference	3
Project Code Repository	3

The object of note in this list is the prevalence of files of data, raw and compiled. Several of the book entries are also compiled records of data. The focus on data is likely caused by something the authors are familiar with: the ever-looming possibility of data loss. The authors have heard anecdotes of years

IIHF data being lost to a basement flood, experienced the loss of NWHL/PHF play by play and location data from the league website, and lost access to CWHL statistics when the league ceased operations. It has become a priority of women’s hockey analytics researchers to preserve data whenever possible, as seen with the websites mentioned in the introduction.

The other part of this table that may be surprising to some is the lack of project code repositories. The proposed explanation for this a matter of common practices in the community: project code does not often stand on its own and are often linked within articles to support those projects. Therefore, there is a fundamentally low amount of entries in this category.

Moving past the entry type and onto entry focus, Table 2 is a list of the top 25 keywords associated with entries, excluding the obviously highest use of the women’s hockey tag.

Table 2. Top 25 most popular keywords for women’s hockey entries on MetaHockey.

Keyword	# of Entries
Counting Stats	39
NWHL / PHF	32
CWHL	29
Goalies / Goaltending	23
Big Data Cup	16
NCAA	13
xG	13
Shots / Shooting	8
Passing	8
Central Ontario Women’s Hockey League	7
Western Women’s Hockey League	7
Pre-Shot Movement	7
National Women’s Hockey League (old)	6
PWHPA	5
Olympics	5
Prediction	4
Shot Quality	4
IIHF	4
Play By Play	4
Tracking	4
Advanced Stats	4
SDHL	3
Team	3
World Championships	3
Model(s)	3

The majority of these keywords are linked to either data sources or books, which preserve leagues both defunct and active, as well. The non-data source focused keywords are in line with general trends of hockey analytics study since

2015, namely the focus on xG, shooting, passing, and pre-shot movement. Curiously, goaltending makes a highly ranked appearance on this list. 20 of the 23 entries referencing goalies and/or goaltending can be attributed to one women's hockey analytics researcher who has been preserving goaltending data for the CWHL, NWHL/PHF, and the SDHL since at least 2016 [6].

Lastly, Fig. 3 looks at the number of women's hockey analytics projects (including all projects under all MetaHockey Categories) published in each year since 2014, which is the year of publishing of the oldest project found.

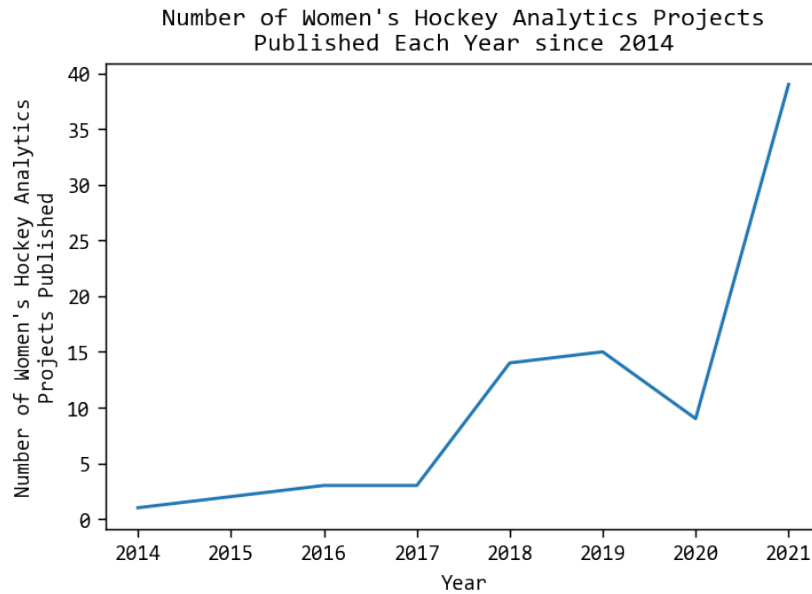


Fig. 3. A chart displaying an increasing trend of women's hockey analytics projects each year, with the exception of 2020. 2020 was when most women's hockey leagues and tournaments were inactive due to the COVID-19 pandemic [7], and therefore there was no new data to work with.

As described in the introduction, women's hockey analytics is on a trajectory of exponential growth. Fig. 3 shows that this is not just conjecture or wishful thinking. Teams, leagues, and researchers would be wise to turn attention to this field of research as women's hockey analytics continues on the rise in the public and private spheres.

4 Conclusion

Women's hockey analytics has a history of projects and data that needed to be centrally archived in order for the community and research to continue to grow. Thanks to a volunteer-based effort, an integral first step has been made towards fulfilling this need. By investigating all available avenues in which projects and data might be found in a procedural manner, details and archived copies of nearly one hundred women's hockey projects, data, and events have made it onto MetaHockey. This set of now-archived research and data reflect the recent priorities of the women's hockey analytics community of data preservation and bringing the field up to speed with men's hockey analytics. It also shows a concrete trend of women's hockey analytics research exponentially growing in the last few years. In the future, we hope to continue archiving women's hockey analytics research in a more periodic manner and hope that as the community gains more momentum, referencing previous works will become more prevalent and will be used to accelerate the public and private development of the field.

Acknowledgements Thanks to Mike Gallimore, Prashanth Iyer, and Mike Pfeil for their permission and help with uploading articles to MetaHockey. We would also like to acknowledge and thank all the volunteers who helped search for and create MetaHockey entries, for all of this would not have been possible without them.

References

1. Matt Barlowe, Jake Flancer, Alyssa Longmuir, and CJ Tuturo. Even-Strength. <https://web.archive.org/web/20210415171201/https://even-strength.com/>
2. Jeff Craig. CWHL Tracker. <https://cwhl-tracker.herokuapp.com/>
3. Nix Dube, Alyssa Longmuir, Mike Murphy, Carleen Markey, An Nguyen, Jessika Ross, and Anna Severn. "To Archive". <https://bit.ly/3KI2kGr>
4. Mike Gallimore and Prashanth Iyer. Metahockey. <https://bit.ly/3LM4ar8>
5. Mike Gallimore and Prashanth Iyer. Metahockey, Dec. 11, 2020. <https://bit.ly/3saOxBE>
6. Jeremy/Giants in the Crease. <https://creasegiants.wixsite.com/index/single-post/2016/08/20/top-goalies-in-the-cwhl-draft>
7. Emily Kaplan. "How the COVID-19 pandemic widened the gender gap in hockey". https://www.espn.com/olympics/hockey/story/_/id/31394538/how-covid-19-pandemic-widened-gender-gap-hockey
8. Alyssa Longmuir, Carleen Markey, and Mike Murphy. WHKYHAC. <https://www.whkyhac.com/>
9. Dave MacPherson. Pick224. pick224.com/women
10. Carleen Markey. "Contextualizing Historical Data and Current Projects in Women's Hockey". <https://youtu.be/WyhSZpKxRFM>
11. Carleen Markey. "How To Archive". <https://bit.ly/3KIgCqc>
12. Carleen Markey. "On the Breakaway: Accelerating the evolution of women's hockey analytics". <https://youtu.be/r-iVM1WACHo>
13. MetaHockey Twitter Account. <https://bit.ly/3OTc185>

14. Mike Murphy. TheirHockeyCounts. <https://theirhockeycounts.com/>
15. Stathletes. Big Data Cup. <https://www.stathletes.com/big-data-cup/>
16. "WHockey MetaHockey Entries". <https://docs.google.com/spreadsheets/d/1J-Lg18bUdwVC71O5g47UTYch2ZxUR-2zUeR226GpDL4/edit?usp=sharing>