Data for Social Good: A Case Study of Building an Effective Public-Private Partnership on Domestic Violence Prevention

Yu-Hsiu Wang[a], Ya-Yun Chen[b], Sue-Chuan Chen[c], Chia-Kai Liu[d], T. C. Hsieh[d*]

[a] Institute for Information Industry, Taipei, Taiwan

[b] National Yang-Ming University, Institute of brain science, Taipei, Taiwan

[c] Taipei City Center for Prevention of Domestic Violence and Sexual Assault, Taipei, Taiwan

[d] DSP, Inc., Taipei, Taiwan

[*]Corresponding author: johnson@dsp.im

Abstract. Most common collaborations for domestic violence (DV) prevention in Taiwan are between academic researchers and advocacy groups or governments. However, the limitation of these two types of collaborations often fail to produce effective prevention strategies. In the Data for Social Good project, we performed data-driven research to improve DV prevention strategies using a public-private partnership approach in cooperation with social workers from Taipei City government and a team of voluntary fellows from different professional backgrounds. Fellows were divided into two group to build two prevention tools: a community-based interactive DV risk map and an individual-based repeat victimization prediction model. The risk map visualized complex information in a more comprehensible and accessible manner, facilitating the understanding for its intended users to formulate the specific prevention strategies. On the other hand, the prediction model served as a useful tool for front-line social workers to identify the repeat victimization risk levels of new cases without having to understand complex data. Through this project of public-private collaboration, participants of different roles were motivated by different incentives and rewarded with satisfaction and self-growth that encouraged them to invest even more efforts into the project. This project has attracted a large amount of public attention and initiates a trend of applying data science onto DV prevention among other cities.

Keywords: domestic violence prevention, public-private partnership, stakeholder management

1. INTRODUCTION

There are many forms of collaborations between organizations on improving domestic violence (DV) prevention programs around the world. Academy-advocate collaboration is one of the most common forms of collaboration known for good quality of data collection and analysis process, but is often criticized for the unequal power relationship between researchers and advocates. In addition, the focuses of this type of collaborated project were more often on how the project results can benefit academic research instead of making sustainable impacts in the field [1].

In Taiwan, another prevalent form of collaboration is for government to outsource prevention projects to private sector agencies. However, this type of collaboration often ends up with private sector agencies paying only minimum efforts to complete the projects by meet the lowest standard of evaluation criteria, and thus tends to have little impacts on improving the prevention programs. On the other hand, existing studies for Taiwanese prevention programs were mostly qualitative instead of quantitative, despite the rich data from DV case reports generated from DV case management approach. Even choices between using qualitative or quantitative methodologies are under heated debates because of the pros and cons on each side [2].

To solve the above-mentioned problems in the existing DV prevention programs collaborations, our project D4SG Initiative (Data for Social Good) aimed to improve prevention strategies using a matching mechanism that encouraged self-motivated collaboration from both sides of the public-private partnership. This project integrated quantitative methodologies using information technology and machine learning conducted by interdisciplinary voluntary fellows from the private sectors and qualitative methodologies by a continuous consultation that include the DV case management experiences from front-line social workers in adjusting the quantitative model building process [3].

2. THE D4SG INITIATIVE

To understand or solve complex social problems requires close collaboration among multi-sector stakeholders. However, these stakeholders are often divided by their own data silos. The D4SG (Data for Social Good) Initiative adopted a platform strategy to connect and empower change agents across these data silos.

2.1 The beginning

In 2015, the D4SG Initiative was founded by three organizations:

- DSP, Inc., a social enterprise providing professional data as a service (DaaS) solutions for government and enterprises.

- Frontier Foundation, an organization working on creating digital opportunities for NGOs and helping them reduce the digital divide through training, website development and other services since 1993.
- Information Technology Software Academy (ITSA), a research and education program funded by the Ministry of Education that promotes the information technology expertise of students from Department of Information Science.

The original mission of D4SG was simply to help public service providers use their own data to improve operational efficiency. However, after studying the data-driven collaborative consulting methods pioneered by DataKind and Data Science for Social Good Fellowship in the University of Chicago, we decided to take a similar approach of launching a short-termed, project-based data science fellowship program.

Since this kind of program was unprecedented in Taiwan, many people were not familiar with the concept of using data science for social good. In addition, most government agencies (GOVs) and nonprofit organizations (NPOs) in Taiwan were not early adopters of technology or innovative approaches. Therefore, we organized a two-day and a four-day hackathon as test drives for NPOs and GOVs respectively. After understanding the needs of the multi-sector stakeholders, we launched the D4SG fellowship program seeking to provide data-driven solutions to important social problems.

2.2 Project implementation

Based on our previous experiences from data hackathons, we divided the participants of a data science project into three roles: proposal organization (professional field), voluntary fellows (data analysis volunteers), and mentor (with data behind the project execution experience). Each of the three coordinating organization took its role in the matching mechanism that paired organization and voluntary fellows. Frontier foundation network recruited NPO and GOV organizations for project proposal while ITSA center recruited voluntary data fellows and DSP, Inc. provided mentorship and technological support.

This was a program without monetary compensation. Each year, there were two three-month cohorts of D4SG Fellowship programs in summer and winter with the rest of the months allocated for business logistics and promotional activities. We organized a review committee and selected organizations to be the official partner proposal organizations based on their data maturity and social influences. We then matched voluntary fellows equipped with skills that met the requirements for solving the problems faced by proposal organizations. Take the 2016 summer cohort for example, we selected three proposal organizations with topics on domestic violence prevention, fire risk map and solar power in farmland and twenty voluntary fellows, 40 % of which were students and 60% were professionals under five years of working experiences with background from computer science, information management, statistics and financial management.

For the proposal organizations, DSP sent professional consultants to help them focus their project goal on generating data-driven solutions and provided data preprocessing suggestions, including transforming operational problems into machine learning solutions, understanding data collection process and the required databases as well as dealing with data privacy, etc.

Before the program officially started, proposal organizations presented the important problems they faced, explained the importance of the problems, listed available data, and described their expected results to inform data fellows why it was important for them to join the projects and what the required skills were.

Since most participants of D4SG have their original full-time occupation, we held our meetings on week nights or weekends for once per week in DSP's office or in proposal organizations' meeting rooms. Meeting sessions were held by mentors to supervise project progress, give suggestions for issues such as setting project requirements and solve data analysis problems, and plan weekly work schedule.

Because of confidential agreements, each project was conducted independently with exchanges of ideas in the mid-term gathering and final project result release presentation. The final project result presentation was open to public, and DSP would issue a project report that was open to public after reviewed by partner organizations.

3. CASE STUDY: THE DV RISK MANAGEMENT PROJECT

The project studied in this paper was proposed by Taipei City Center for Prevention of Domestic Violence and Sexual Assault (TPDVPC), which faced the problem of the insufficient supply of front-line social workers in contrast with the large increase of reported domestic violence cases over the past decade. After several meetings and series of discussion, all the parties agreed that the key factor that caused the failure of curbing the rampant DV cases in Taiwanese society lied in the lack of differentiated prevention strategies targeting different neighborhoods and the inability to identify key variables to prevent repeat victimization from happening. Thus, the team of interdisciplinary voluntary fellows decided to help TPDVPC improve DV prevention strategies on both the community-based level with a DV risk map and the individual level with a repeat victimization prediction model.

3.1 Roles and collaboration methods

After verifying the requirement of TPDVPC, the voluntary fellows in the project were assigned to two groups depending on their interest and capabilities. One group was responsible for the production of the interactive DV risk map. They provided expertise on the extract-transform-load (ETL), descriptive statistics and interactive webpage erection. The other group developed the repeat victimization prediction model, with research skills in surveying key risk factors from existing literature, data concatenation and prediction model building. The social workers from TPDVPC provided fellows with domain knowledge and assistance of data correction. Mentors from the DSP helped with clarifying the requirements from TPDVPC and offered technical supports to voluntary fellows.

Fig 1. D4SG fellowship in 2016 summer. This photo was taken during a discussion where social workers of TPDVPC set project requirements and provided support of domain knowledge in response to research question asked by voluntary fellows.

3.2 Results

All data used in this project were from the National Domestic Violence, Sexual Assault, Children and Juvenile Protection Information System at the Taiwan Ministry of Health and Welfare (MOHW).

(1) The risk map

In the last stage of the ETL, we converted the addresses into latitude and longitude and transformed them into the district and village levels, which were the third-level and fourth-level administrative unit in Taiwan. Then, we calculated the amount of occurrence in each case type and analyzed the sociodemographic features in an area. Afterwards, the analyzed data were visualized into an interactive DV risk map under website framework.

We sorted the case into four intimate partner violence (IPV) case types, including children/adolescent, elderly, and inter-sibling and others. The summary for all DV case types is shown in Table 1. The gender of victims was significantly biased to female (female to male was 81% to 19%). The result revealed that different case types had distinguished characteristics. The children and adolescent protection cases had the biggest proportion of victims from low- and middle-income families. However the victims from elderly protection and inter-sibling cases tended to with disabilities or mental illness.

Table 1. Data Summary of four general categories

Case type	IPV	children/ adolescent	elderly	inter-sibling and others
Proportion	52%	8%	6%	34%
Gender ratio (Male/Female)	19/81	54/46	37/63	41/59
Low- and middle- income	8.7%	24.2%	9.2%	12.4%
Disabilities and mental illness	2.1%	1.8%	3.9%	4.2%

The map was for community-based level prevention and the intended users of the map were governmental officials, chiefs of villages, and social workers, etc. With this map, policymakers were able to properly respond to the specific DV case types in local neighborhoods under their governance and to formulate effective policies.

(2) The repeat victimization prediction model

The data cleaned from the process of making risk map was then utilized by prediction model group. After taking into consideration of data quality and limitation of available time, the IPV case type was the main source of data used for building repeat victimization prediction model. Both data from the case reports and the Taiwan Intimate Partner Violence Danger Assessment (TIPVDA) form were included in our prediction model. This model was built for front-line social workers to evaluate the repeat victimization risk level of victims when a case was first reported. Preventing repeat violence in the future would reduce the burden of front-line social workers and the needs to deluge in complex data.

The pattern of the cases reported more than twice a year was different from those reported equal to or smaller than twice a year. Thus, cases with the former situation were assigned a value of 1 and latter as a value of 0. The random forest model was then performed to conduct classification to identify the cases' repeat victimization risk by their demographic features, case characteristics, and the answers of TIPVDA. The result showed that our model had a high performance accuracy and sensitivity with the average 96.3% accuracy and 88.9% recall rate. Although the model showed a precision and F1-measure of 48.7% and 62.8%, the low percentages did not influence the evaluation of the model since it came from a high proportion of false positives. In other words, cases with report times less than two times a year were predicted as high repeat risk cases with more than twice reports a year. It is acceptable because social workers and their managers tend to take on a more conservative case management approach in social work practices.

4. DISCUSSION

This project adopted public-private partnership in improving DV strategies with two data analysis tools: the interactive risk map and repeat victimization risk prediction model. The risk map assisted social workers in developing differentiated strategies targeting each village according to the specific case types and the demographic features of victims in that village. The repeat victimization risk prediction model helped social workers identify the risk level of each case during its first report. Over the course of completing this project, the incentives for each participant occupying different roles need to be fulfilled. In addition, each participant obtained different types of satisfaction and achieved self-growth that in return motivated them to contribute even more into this project. The following paragraphs summarize insights gained from this project in terms of the roles of participants, and the last section suggests directions for future studies.

4.1 Insight for social workers and supervisors in the prevention center

This project has improved the work transition process and data science competency of front-line social workers and higher level supervisors. First of all, the social work discipline tends to have high turnover rate that makes it difficult for the hands-on knowledge and experiences of dealing with DV cases to be passed on inside the organization. Hence, important factors and patterns identified by experienced social workers were not able to be applied to future case management after they leave the organization. However, our prediction model that was built from the knowledge of social workers and was verified using rigorous model building process served as a useful tool for social workers to obtain prediction results by simply inputting case data without having to be equipped with complex data knowledge and case management experiences with them. In addition, through completing the project together and with voluntary fellows explaining how field experiences can be translated into the language of data science, staff in prevention center also built up their data science competency and were able to identify key concepts of data science methodologies, and further pass on their knowledge to fellow social workers, professionals in the similar disciplines, as well as activist and administrative staff in local neighborhoods. Moreover, because our project has successfully aroused public attention and gained recognition in the DV prevention realm in Taiwan, more city governments and prevention centers have requested for similar cooperation opportunities in hope of leveraging data science in improving prevention strategies.

4.2 Insight for fellows

The voluntary fellows of this team were from different professional backgrounds of finance, information management, neural science and social science discipline. For fellows who joined the project out of compassion for improving social goods, it was rewarding to have front-line social workers share their field experiences and to see the positive impact of our project on improving the efficiency of prevention strategies. In addition, for professionals who were interested in data science but were lacking chances to be engaged in a hands-on project, this project was a good opportunity to see the whole implementation process of a data project. The more experienced data science professionals were intrigued by the leading

role they took in making key decisions on the research process and the sense of achievement of completing a not-for-profit data science project while networking with interdisciplinary professionals.

4.3 Insights for future studies

This project not only served as a model study for applying data science on DV prevention, it also prepared the relevant data ready for open to public and for academic research. Similar to previous studies' conclusion that a successful collaboration is dependent upon trust between participants and continuous learning and communication [4], what distinguished this project from academic research was that it was not conducted for the sake of research per se, but to really solve social problems, which also enabled the project to gain full support from relevant stakeholders. For example, when data was incomplete, social workers would voluntarily go through documents to find missing data or even apply for data access from other governmental division such as Department of Education and Department of Health, showing how this form of cooperation can enable dedication to the project from all parties involved.

5. CONCLUSION

This project adopted a public-private partnership in improving DV prevention strategies by leveraging data science. The result of the project exemplifies that unlike in academic and public collaborations, where researchers are most likely dominating the projects, the public-private collaboration approach can better motivate researchers, social workers and social work supervisors to work together equally toward the same project goal.

6. ACKNOWLEDGMENTS

We deeply appreciate the D4SG Initiative (Data for Social Good) for connecting and empowering change agents across these data silos. We would also like to thank the social workers team from TPDVPC for contributing DV domain knowledge and assisting us in data correction.

7. REFERENCES

[1] Williams, L. M. (2004). Researcher-advocate collaborations to end violence against women. Journal of Interpersonal Violence, 19(11), 1350-1357.

[2] SALE, J.E.M., LOHFELD, L. H., & BRAZIL, K. (2002). Revisiting the Quantitative-Qualitative Debate: Implications for Mixed-Methods Research. Quality & Quantity, 36(1), 43–53. http://doi.org/10.1023/A:1014301607592

[3] Chen, Y.-Y., Wang, Y.-H., Hsieh, Y. S., Ke, J. T., Liu, C. K., Chen, S. C., & Hsieh, T. C. (2017). Measuring the unmeasurable - a project of domestic violence risk prediction and management. Data for Good Exchange Conference 2017 (submitted).

[4] Vinton, L., (2003). A model collaborative project toward making domestic violence centers elder ready. Violence Against Women, 12, 1504-1513.