OPEN DATA PLATFORM TO PROMOTE GENDER EQUALITY POLICIES IN STEM

Cristiano Maciel, Universidade Federal de Mato Grosso, Brazil, cristianomaciel@ufmt.br
Indira R. Guzman, California State Polytechnic University Pomona, USA, irguzman@cpp.edu
Rita Berardi, Universidade Tecnológica Federal do Paraná, ritaberardi@utfpr.edu.br
Boris Branisa Caballero, Universidad Católica Boliviana San Pablo, Bolivia, bbranisa@ucb.edu.bo
Elizabeth Jimenez Universidad Mayor de San Andrés, Bolivia, ejimenez@cides.edu.bo
Luciana B. Frigo, Universidade Federal de Santa Catarina, Brazil, luciana.frigo@ufsc.br
Luciana Salgado, Universidade Federal Fluminense, Brazil, luciana@ic.uff.br
Nadia Rodriguez-Rodriguez, Universidad de Lima, Perú, nrodrigu@ulima.edu.pe
Silvia Amélia Bim, Universidade Tecnológica Federal do Paraná, Brazil, sabim@utfpr.edu.br
Patricia Cabero Tapia, Universidad Católica Boliviana San Pablo, Bolivia, cabero.patty@gmail.com

ABSTRACT

Women are still underrepresented in Science, Technology, Engineering, and Mathematics (STEM). In the last decade, government and academic institutions have developed initiatives and policies to increase the participation of women in STEM. However, they are rarely published in an open and structured way. This research proposal presents a plan to develop an open data platform of existing policies and initiatives that promote gender equality in STEM in three countries in South America, Bolivia, Brazil, and Peru with an emphasis in technology areas. This project, funded by the Canadian government, has three phases from 2022 to 2025.

Keywords: Women in STEM, Open Data, Latin America, Bolivia, Brazil, Peru, Gender Gap, Women in Leadership, Gender Equality.

INTRODUCTION

The 2030 Agenda for Sustainable Development, adopted by all United Nations Member States in 2015, identified 17 Sustainable Development Goals (SDGs) to end poverty. Goal #5 is framed as "Achieve gender equality and empower all women and girls" (https://sdgs.un.org/goals) emphasizing that promoting gender equality and empowerment is integral to all dimensions of inclusive and sustainable global development of all countries. The problem is that women continue to be underrepresented in Science, Technology, Engineering, and Mathematics (STEM). Only 35% of STEM students in higher education globally are women, and only 3% of female students in higher education choose information and communication technologies (ICT) studies (UNESCO report, 2017). In addition, women hold few leadership positions in this field. Having female leaders in STEM is not only critical to the career advancement of women, but it also has the potential to increase diversity, equity and inclusion.

Reflecting on the status of science, technology, and innovation in Latin America and the Caribbean, it is clear that this region's economy is not well prepared to face the challenges of the knowledge society (Branisa. Cabero & Guzman, 2021). Women in this region are underrepresented in STEM and ICT. According to Basco et al. (2021) women represent only 35% of the workforce in STEM in Latin America. Furthermore, the proportion of women who take on leadership roles is even smaller (Siemiatycki, 2019; Jaccheri, Pereira & Fast, 2020). According to a study by the Inter-American Development Bank (IDB) found, only 15% of management positions are held by women and only 14% of companies are owned by women (Basco, et al., 2021). In STEM areas, the trend is similar. Women in STEM must overcome cultural and societal challenges to achieve leadership positions (Branisa, Kasen & Ziegles, 2013). For

example, only 9% to 27% of members of academy of sciences in Latin American and the Caribbean countries are women (Rodriguez-Hertz, 2018). More work needs to be done to increase representation of women in leadership positions in STEM (Guzman et al. 2020).

In the last decade, government, and academic institutions all over the world have developed several initiatives and policies to increase the participation of women in STEM (Wang & Degol, 2017; Henderson et al., 2011, Cheryan et al., 2017; Maciel et al. 2018). Studies about the gender imbalance are relevant because they find ways to promote social inclusion (Trauth and Connolly, 2021) and help academics and practitioners understand what possible changes can be made to increase diversity in the IT field. Su et al. (2015) and Mitchneck, Smith & Latimer (2016) point out the need for administrative actions/strategies for gender diversity within academia/university, as women also constitute a small population of academic administrators, especially in academic chairs in STEM fields. Research indicates that academic chairs play a critical role in gender diversity efforts (Hurtado & DeAngelo, 2009).

While those initiatives are important, it is very difficult to find these studies and their results in a systematic way. Researchers often collect and analyze data, but generally, when they publish articles, they do not ensure that the data is available for public use and reusability. In addition, each research study has a different context and level of analysis: project, university, community, and country. Each context may present different meanings for the collected data. Data related to STEM are usually not structured, not well documented, and difficult to find. Research results are rarely published in an open and structured way. One key challenge is hence the lack of open data to support the research of gender equality policies in STEM and cross-cultural analysis. It is necessary to map out factors related to the lack of gender equity and then provide data-based STEM analysis. Developing a structured open infrastructure would increase transparency and easy access to existing policies and initiatives and hence, promote gender equality.

The goal of this research project is to systematically approach the question of female representation and leadership in STEM reviewing the literature with an emphasis in three countries in South America (Bolivia, Brazil and Peru), to map the initiatives and policies implemented so far in each country; and map context aspects that influence women's careers in STEM.

The research questions addressed in this study are:

- RQ1: What are the challenges that women in STEM have to overcome to achieve leadership positions?
- RQ2: What are the policies that could support the career growth of women in STEM?

To answer these questions, we plan to gather comparable primary and secondary data in the three countries, and analyze the data to inform and generate policy recommendations for public and private institutions. We will design a specific ontology for the data (ensuring that it is well structured, easy to access, and reuse) and will create an open data platform, including applications to facilitate and promote its use by anyone interested in the topic. We intend to have products in different formats and in different languages (Portuguese, Spanish and English), ensuring that some of the documents (e.g., policy briefs) use plain language to spread the results to a broader audience and foster debate and discussion.

We use theory of change (Rogers, 2014; Reinholz & Andrews, 2020) to help structure the design and management process of this research project. Theory of change is "theory-driven evaluation aimed to move beyond a simplistic input-output notion of evaluation and instead required that program designers explicitly state how they expected a program to work, thereby making their implicit assumptions explicit"

(Reinholz & Andrews, 2020). Figure 1 shows the inputs, activities, outputs, short/mid/long term outcomes, and impacts envisioned for this project.

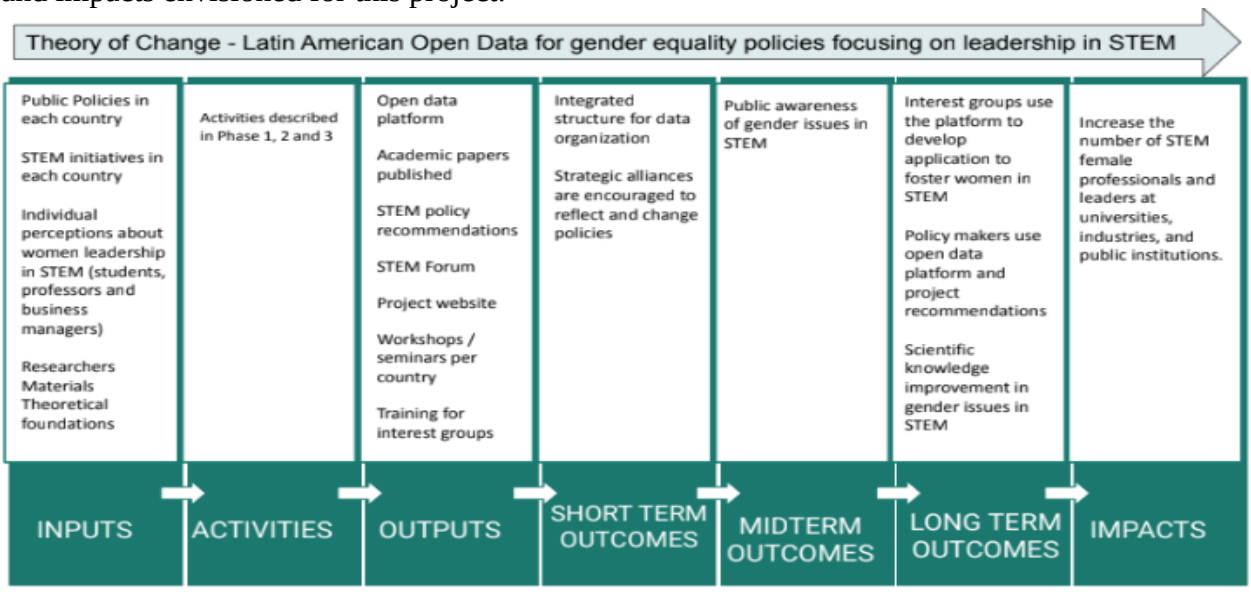

Figure 1. Theory of Change applied to the Research Project

Open Data

Access to data has been generally expensive and restricted based on user's permissions, fees, licenses or policies. Even when datasets have been relatively open and available, they have required special equipment and tools such as software, skills such as statistics, and contextual knowledge concerning a topic, to make sense of them, much of which is beyond the capabilities of the general population (Kitchin, 2014). Because of this, data as well as the information and knowledge derived from them have traditionally been largely "closed" in nature. The open data movement seeks to transform data access, opening data for wider reuse and providing easy-to-use research tools to promote transparency and knowledge sharing (Hyyonen, 2020). The design of an ontology is the first step in the development of an open data infrastructure. The design of Ontology is an interactive process that involves several domain specialists around a subject (Victorino et al., 2018). An ontology is a formal explicit description of concepts in a domain of discourse (classes/concepts), properties of each concept describing various features and attributes of the concept (slots/roles or properties), and restrictions on slots (facets/role restrictions). An ontology together with a set of individual instances of classes constitutes a knowledge base (Noy & McGuinnessl., 2001).

To model these factors and initiatives in an open data infrastructure, we will apply a Semantic Web infrastructure (https://www.w3.org/standards/semanticweb/). These standards allow for linked data to build data stores on the Web, build vocabularies, and write rules for handling data. The Semantic Web Infrastructure builds a technology stack to support open data infrastructures through ontologies that can be used in vertical applications —for example, in education, government, and industry— to improve collaboration, research, and development. It would also be possible to work with inference rules and knowledge graphs with the information available in the data infrastructure. These tools might allow for better policy making and interventions. The proposed open data infrastructure would become a platform to discuss, in an integrated way, how we can reduce the STEM gender gap in these three countries in Latin America. It is thus possible to generate comparable data to inform and generate policy recommendations for public and private institutions. All that will be available on a platform in 3 languages.

The main overall objective is to contribute to the generation and use of cross-country comparable open data in order to assess policies and interventions to reduce the gender gap in STEM, to promote public discussion aimed to increase the number of female leaders at universities, industries, and public institutions, and to foster the development of mobile and web applications based on open data to increase the awareness of the importance of women in STEM.

The Specific Objectives (SO) intrinsically related to the objectives are:
• SO1: To map out the factors, actors, and policies that influence the career development of women in STEM, collect related data, and analyze this data.
• SO2: To build and deploy an open data platform that integrates primary and secondary data about career growth of women in STEM;
• SO3: To promote the use of open data about women in STEM leadership to increase public awareness of gender issues in the field.
• SO4: To provide recommendations for policy makers in Latin America to increase the female representation in STEM with a focus on gender equality and diversity.

Research Plan

This project proposes creating and publishing a linked open cross-cultural data infrastructure that supports research in the STEM field in a comparable and structured way. We will use ontologies to represent different cultural and social aspects involved in collecting data from different countries. The data will be available on a platform. Any country can contribute to the platform with their data and participate in cross-country discussions about the initiatives and effects of gender equality actions. We will apply a semantic approach in conjunction with different countries in Latin America to collect and model the data considering the social and cultural aspects of each unique country. In general, the research will be descriptive and exploratory.

The measurement of the effectiveness of the proposed platform is essential and challenging. We argue to look beyond the quantity of data published and focus on the demand for the data with three key indicators[1]: readiness, implementation, and impact. The data platform usage, ease of getting data, and user's perceived data quality are examples of implementation metrics of an open data access platform. The following are essential metrics concerning impact: policy decisions informed or businesses using the data to make decisions.

The following table shows the proposed set of activities for each of the three phases of this research project.

Phase	Act.	Activity description	Associated Spec. Obj.
I	1	Make strategic alliances with interest groups	SO1
	2	Mapping of policies in each country	SO1
	3	Mapping Women in STEM Initiatives from stakeholders by country	SO1
	4	Mapping context aspect that influences women in STEM	SO1
	5	Research design for each country for data collection	SO1
	6	Data modeling considering the social and cultural aspects of each country	SO2

[1] https://data.gov.ie/edpelearning/en/module6/#/id/co-01

	7	Design Ontology: metadata & governance.	SO2
	8	Launch Project website	SO1, SO3
	9	Obtain approval of ethical committees for data collection	SO1, SO2
II	10	Primary data collection	SO2
	11	Secondary data retrieval	SO2
	12	Data Analysis & ontology implementation	SO2
	13	Design of ontology final metadata	SO2
	14	Open Data Platform Design: Architecture & Governance	SO1
	15	Beta Platform Deployment	SO2, SO3, SO4
III	16	Training interest groups	SO4
	17	Open Data Platform Design Applications	SO3
	18	Beta Platform Evaluation	SO2, SO3
	19	Open Data Platform Deployment	SO3
	20	Promote adoption for interest groups	SO3
	21	Organize a "STEM Leadership Forum" and the platform launch in each country	SO3
	22	Provide "Women in STEM-Leadership" policy recommendations and policy briefs for public and private institutions	SO4

Table 1. Research Project Phases and Activities

Status of the Project

The main activity in Phase I, year 1 was to officially launch the project. It was an opportunity to gather information of possible stakeholders, including policymakers, the academic community, local and governmental representatives, and initiatives from civil society that would become potential users of the data platform. The project's launching was followed by a series of systematic literature reviews using indexed databases as well as gray literature and the respective initial mapping of initiatives and policies. In addition, we began the research design for primary data collection considering the social and cultural aspects of each country.

Phase II will involve the gathering, organization, and analysis of the data. Socialization of this information during this stage is vital. Socialization will be aimed at ensuring dissemination of the information gathered and trying to analyze and explain the information gathered in each of the contexts where the project takes place. The team will actively participate in academic events and other public spaces where the opportunity to share, dialogue, and debate STEM-related issues arise. Throughout this phase, a series of training workshops will also be organized to ensure that stakeholders and interested community members, in general, can access, explore and use the data and the platform created.

Phase III will focus on the socialization of the open data platform, and the application developed using the data available in the platform and policy recommendations. This will be done both in academic contexts (academic conferences, seminars, and other similar events) and events organized explicitly by STEM leadership for socialization purposes and to reach the public in general. The elaboration of academic papers will be carried out at two levels. On the one hand, a country-specific analysis will be carried out using the data platform and specifically identified themes in the two earlier phases. On the other hand, comparative analysis will also be carried out, gaining from the opportunity to explore and analyze STEM-related issues from a comparative and multidisciplinary perspective.

Recognizing that part of the problem concerning gender gaps in STEM in Latin America is related to the

lack of recent and reliable data, this project aims to contribute to the generation of cross-country comparable data to assess policies and interventions to reduce the gender gap in STEM, especially by increasing the number of female leaders at universities, industries, and public institutions. We plan to map out the factors that influence the career development of women in STEM, as well as document and analyze successful and less successful initiatives and learn from them.

Acknowledgments

The authors would like to thank the International Development Research Centre (IDRC).

REFERENCES

Basco, A.I., Verna, A.B., Monje Silva, A., Barafani, M., Torres, N. & Oueda Cruz, St. (2021). *Una olimpíada Desigual: La equidad de género en las empresas latinoamericanas y del Caribe*. Banco Interamericano de Desarrollo. Nota Técnica IDB-TN-2255.

Branisa, B.; Klasen, S.; Ziegler, M., (2013). "Gender Inequality in Social Institutions and Gendered Development Outcomes," World Development, Elsevier, vol. 45(C), pages 252-268. [Available at https://www.sciencedirect.com/science/article/abs/pii/S0305750X12002975#:~:text=Our%20results%20show%20that%20social,of%20corruption%20in%20a%20country].

Branisa, B.; Cabero, P.; & Guzman, I.R. (2021). The main factors explaining IT Career Choices of Female Students in Bolivia. *AMCIS 2021 Proceedings. 30*. Available at https://aisel.aisnet.org/amcis2021

Chavatzia, T. (2017). *Cracking the code: Girls' and women's education in science, technology, engineering and mathematics (STEM)* (p. 20). Paris, France: UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000253479

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others?. *Psychological bulletin*, *143*(1), 1.

Guzman, I.; Berardi, R.; Maciel, C.; Cabero Tapia, P.; Marin-Raventos, G.; Rodriguez, N.; Rodriguez, M. (2020). Gender Gap in IT in Latin America. [Available at https://aisel.aisnet.org/amcis2020/panels/panels/4/

Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: an analytic review of the literature. *Journal of Research in Science Teaching*, *48*(8), 952–984.

Hurtado, S., & DeAngelo, L. (2009). Keeping senior women at your college. *Academe, 95*(5), 18-20.

Hyvönen, E. et al. (2020). Linked Open Data Infrastructure for Digital Humanities in Finland. *Proceedings of Digital Humanities in Nordic Countries* (DHN 2020).

Jaccheri, L., Pereira, C., & Fast, S. (2020, September). *Gender Issues in Computer Science: Lessons Learnt and Reflections for the Future.* In 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC) (pp. 9-16). IEEE.

Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures & their consequences. SAGE Publications, Limited.

Maciel, C.; Bim, S. A.; da Silva Figueiredo, K. (2018). Digital Girls Program: disseminating computer science to girls in Brazil. Proceedings of the 1st International Workshop on Gender Equality in Software Engineering. ICSE '18. [Available at https://doi.org/10.1145/3195570.3195574]

Mitchneck, B., Smith, J. L., & Latimer, M. (2016). A recipe for change: Creating a more inclusive academy. Science, 352(6282), 148-149.

Noy, N. F., & McGuinness, D. (2001). *Ontology development 101: A guide to creating your first ontology*.

Reinholz, D. L., & Andrews, T. C. (2020). Change theory and theory of change: What's the difference anyway? *International Journal of STEM Education*, 7(1), 1-12. https://doi.org/10.1186/s40594-020-0202-3

Rodriguez Hertz, J. (2018, Oct). Promoting women in science in Latin America and the Caribbean. *Angle Journal.* Retrieved from http://anglejournal.com/article/2018-06-promoting-women-in-science-in-latin-america-and-the-caribbean/

Rogers, P. (2014). Theory of Change, Methodological Briefs: Impact Evaluation 2, UNICEF Office of Research, Florence.

Siemiatycki, M. (2019). The diversity gap in the public–private partnership industry: An examination of women and visible minorities in senior leadership positions. Annals of Public and Cooperative Economics, 90(2), 393-414.

Statista (2021). *Share of women in business leadership roles in Latin American countries 2021.* Retrieved from https://www.statista.com/statistics/1300214/share-women-business-leaders-latam-countries/

Su, X., Johnson, J., & Bozeman, B. (2015). Gender diversity strategy in academic departments: Exploring organizational determinants. *Higher Education, 69*(5), 839-858.

Trauth, E., & Connolly, R. (2021). Investigating the nature of change in factors affecting gender equity in the IT sector: a longitudinal study of women in Ireland. *MIS Quarterly, 45*(4).

Victorino, M., Maristela Terto, d. H., Ishikawa, E., Edgard, C. O., & Chhetri, S. (2018). Transforming Open Data to Linked Open Data Using Ontologies for Information Organization in Big Data Environments of the Brazilian Government: the Brazilian Database Government Open Linked Data – DBgoldbr. *Knowledge Organization: KO, 45(6),* 443-443–466. https://doi.org/10.5771/0943-7444-2018-6-443

Wang M.T.; Degol J.L. (2017). Gender Gap in Science, Technology, Engineering, and Mathematics (STEM): Current Knowledge, Implications for Practice, Policy, and Future Directions. *Educ Psychol Rev.* 2017 Mar;29(1):119-140. doi: 10.1007/s10648-015-9355-x.