CHAPTER 6

Fostering Open Data on Women in STEM across Latin America: The ELLAS Network

Cristiano Maciel, Indira R. Guzman, Karen da Silva Figueiredo Medeiros Ribeiro, Franci Suni-Lopez, Luciana B. Frigo, Luciana Salgado, Rita Cristina Galarraga Berardi, Sílvia Amélia Bim, Boris Branisa, Elizabeth Jimenez and Luana Bulgarelli Mendes

Introduction

In Latin America, investment in gender data across socio-economic indicators has lagged over the past five years (World Bank Group, 2024). The lack of open and reliable data on women in Science, Technology, Engineering and Mathematics (STEM) and their career trajectories poses a significant barrier to global efforts to close the gender gap in science (UNESCO, 2020; UNESCO, 2024). More data across intersectional lines is needed to track progress and address the challenges women and other underrepresented groups face in STEM fields.

Latin America faces a distinct mix of economic inequality, political instability and deep-rooted gender norms that affect women's participation in science and technology fields (García-Holgado & García-Peñalvo, 2022); without context-specific data, the development and implementation of effective, evidence-based policies that promote gender equality and leadership in STEM are hindered.

In response to this challenge, in 2022, a consortium of researchers from Bolivia, Brazil and Peru created the Equality in Leadership for Latin America STEM (ELLAS) research network to expand the field of studies in Latin America.

The ELLAS Network[1] is an interdisciplinary team that includes professionals from eight universities, spanning fields such as computer science, economics, psychology, social sciences and education. ELLAS launched the "Latin American Open Data for Gender Equality Policies Focusing on Leadership in STEM" initiative (Maciel et al., 2023) intending to gather and analyse data on women in STEM across Bolivia, Brazil and Peru, and create an open data platform (ODP) where this information could be shared and accessed by all.

An ODP is a digital infrastructure that provides public access to a wide range of datasets, offering tools for searching, filtering and visualising data (Keserű & Chan, 2015). By blending their diverse expertise, cultural perspectives and backgrounds, the ELLAS interdisciplinary team aims to offer context-specific insights into research and the ODP development processes.

This chapter highlights the process leading to the development of the ODP and the overview of the collection and structuring of datasets that feed it. Our contribution explores several key dimensions of the ELLAS Network research and its connection to open data. After a brief introduction, we provide a comprehensive overview of the practical aspects of open data, describing how the ELLAS research team collects and structures this information, while highlighting the central research questions it addresses. We then present the ELLAS platform, showcasing its functionalities and illustrating how the ODP serves as a valuable tool for a wide range of societal applications. Lastly, we offer recommendations based on our findings and outline potential directions for future work.

The open data approach for Latin America

The existing data about women in STEM in Latin America is often fragmented, dispersed across open government and non-governmental platforms, or embedded in unstructured formats such as academic papers, grey literature and social media posts. The ELLAS Network positions itself within this context. Our qualitative applied research (Babbie, 2020) proposes an open data approach for Latin America (starting with Bolivia, Brazil and Peru) with a dual focus: first, to explore, extract and structure pre-existing data on these countries about Latin American women in STEM careers as open data; and second, to generate, structure and make available *additional* open data, collected through a comprehensive survey targeted at the three countries' audiences, as detailed in Davila et al. (2024).

1 https://ellas.ufmt.br/

Gathering pre-existing data

Existing data was gathered and organised by the ELLAS research team from various primary and secondary data sources. In the context of ELLAS ODP, 'primary data' refers to unstructured data obtained from scientific literature texts, such as academic papers, grey literature such as official reports, textual content from social media pages and data from a survey created and applied in the three countries. 'Secondary data' refers to pre-existing structured data collected by someone else for a different purpose, usually available in spreadsheets on official websites (Berardi et al., 2024). In our research, secondary data were obtained from publicly accessible spreadsheets and datasets.

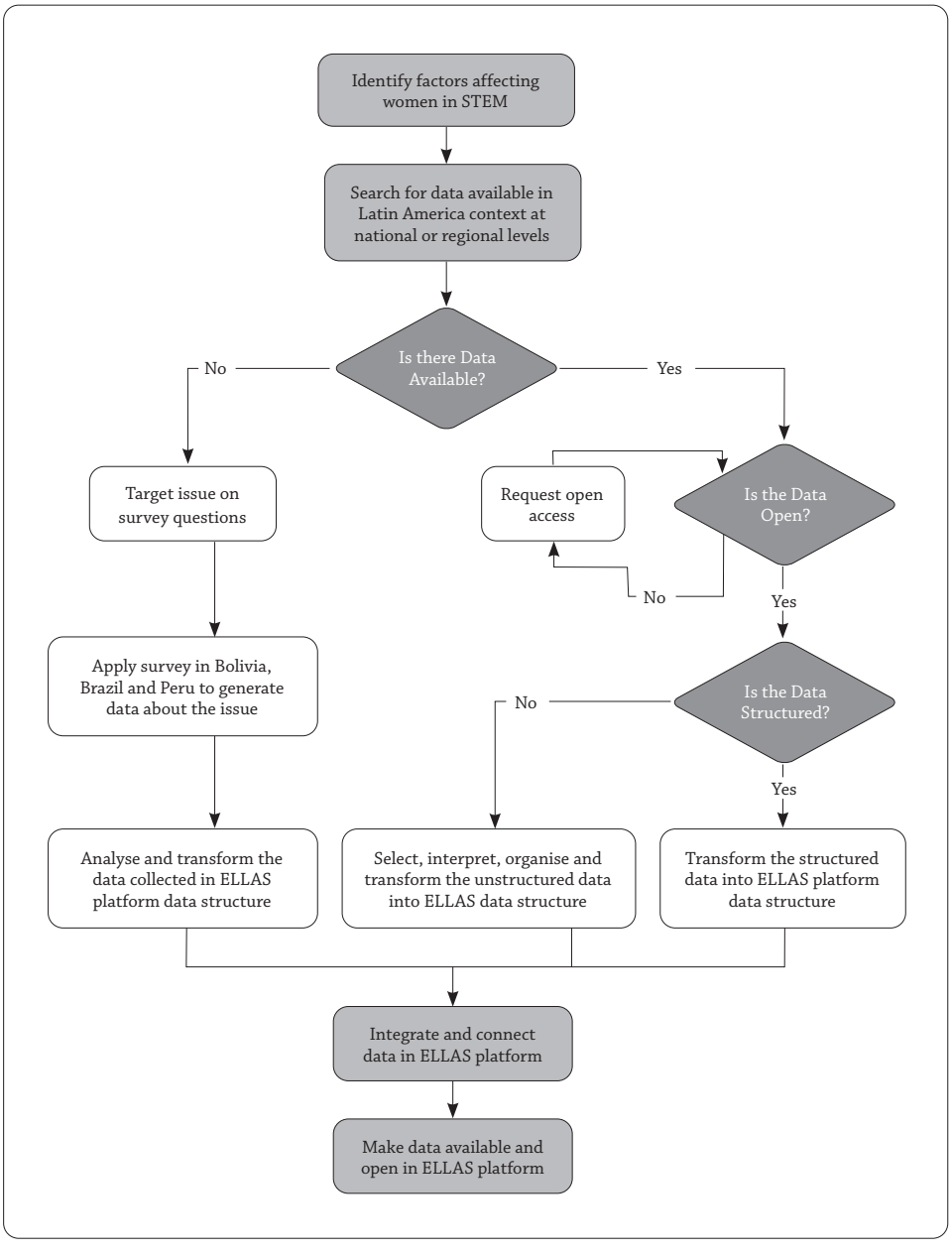
Figure 1 presents a flowchart outlining the process our research team has followed to obtain and integrate data for the ODP. The process began with the identification of factors affecting women in STEM in the Latin American context. Once each factor was identified, the next step was to search for existing data related to these issues at national or regional levels within Latin America. If data were available, open and structured (i.e. secondary data), they were directly translated or converted into structured formats for the ODP. If data were available and open, but unstructured (i.e. primary data), they were analysed and then transformed into structured data for the ODP. In cases where data were available but not in open formats, our researchers actively engaged with data custodians to advocate for open access. When data were unavailable, relevant information was incorporated into our survey instrument (Davila et al., 2024). This survey was administered across the three countries to generate new data, which were subsequently analysed and formatted for inclusion in the ODP.

After the collection and organisation of all the data, a process of integration and connection was executed to transform the data into linked open data in the format of the ELLAS ODP (further described in the next main section). Linked open data is based on a set of best practices for publishing and connecting data in knowledge graphs on the web. These best practices were introduced by Berners-Lee (2006) in his linked data architecture and became known as the Linked Data Principles, adopted by W3C as linked data publishing standards.

The flowchart in Figure 1 represents a systematic approach to handling gender data in STEM in Latin America. Ultimately, the ODP's goal is to make all data related to women in STEM in Latin America accessible, integrated and usable through the ELLAS platform.

With this in mind, data identification and gathering of pre-existing data began in 2022, when we defined the key concepts we would like to address

Figure 1: Process of systematic data collection and transformation for the ODP



with the data: policies, initiatives, contextual factors and contextual indicators related to women's careers in STEM, directly or indirectly (see Figure 2

for definitions and examples). We believe that those concepts are crucial for revealing the structural and cultural barriers faced by women considering entering or already working in STEM, and as such, are necessary for identifying opportunities to promote gender equity. In addition, Latin America faces unique challenges related to social, economic and gender disparities, making it essential to analyse these variables to inform effective public policies and develop strategies that foster inclusion and female leadership in STEM fields.

Figure 2: Key concept definitions and data samples by country

	Policies	Initiatives	Contextual factors	Contextual indicators
Definition	Official guidelines, regulations, and frameworks established by governments, organisations or institutions to promote gender equality directly or indirectly	Targeted actions or programmes designed to promote gender equality at all levels	Specific social conditions that shape the environments and may act as barriers and enablers to gender equality	Measurable variables that provide insight into the contextual conditions impacting gender equality directly or indirectly
	Examples			
BOLIVIA	National Plan for Equal Opportunities "Women building the New Bolivia to Live Well"	Mujeres TICs BO facebook.com/MujeresTICsBO	Lack of role models (Rosales Rodríguez, 2020)	Women represented 30% of those employed in senior and middle management in 2022 (World Bank Group, 2024)
BRAZIL	CNPq / MCTI / MMulheres Call No. 31/2023, Girls in Exact Sciences, Engineering and Computing	Programa Meninas Digitais https://meninas.sbc.org.br/	Gender pay gap (Lucena et al., 2024)	11.6% of professionals are Black women in the ICT sector in Brazil (Brasscom, 2022)
PERU	Supreme Decree No. 008-2019-MIMP, National Gender Equality Policy (PNIG)	Peruvians in STEM linkedin.com/company/peruvians-in-stem/	Participation in gender equity initiatives (Tull et al., 2018)	47% of STEM graduates are female (UNDP, 2024)

Policies refer to the formal guidelines, regulations and frameworks established at different levels by governments, organisations or institutions to promote gender equality. In the context of our study, these policies may include laws, strategic plans, specific funding, affirmative action programmes and calls for gender quotas aimed at reducing gender disparities related to STEM or women's leadership. In the countries investigated, national-level policies are still limited,

but organisational and institutional policies have emerged (Guzman et al., 2023a; Guzman et al., 2024).

Initiatives include targeted, organised actions, projects or programmes designed to promote gender equality in STEM at different levels. These initiatives can include mentorship programmes, scholarships, training workshops or networks that aim to empower women, enhance their skills or increase their visibility and participation in STEM careers. They are mostly connected to professional networks and academic institutions rather than government bodies.

Contextual factors refer to the specific social, cultural, economic and institutional conditions that shape the environment in which women live and pursue careers and may act as barriers or enablers to gender equality in STEM. These factors include societal attitudes towards gender roles, the availability and quality of education and professional development opportunities, economic disparities and workplace behaviour. In Latin America, these factors vary widely between countries and regions because of the cultural differences that influence women's opportunities and challenges in their lives and STEM careers. A study for Bolivia, for instance, found that traditional gender roles, family influence and discrimination in the labour market discourage women from pursuing STEM careers, even when cognitive ability is not seen as a limitation (Branisa et al., 2020). Understanding the level of influence of these contextual factors is crucial for developing effective strategies and policies to support women in STEM, as they provide insight into the underlying barriers and enablers of gender equity in these professions. In our research, we categorised the contextual factors we identified in our literature search into individual, interpersonal, academic, work-related, family-related, socio-economic, social and historical domains (Davila et al., 2024; Drummond et al., 2023).

Context indicators are measurable variables that provide quantitative insight into the conditions impacting gender equality in those countries. These indicators could be directly or indirectly related to STEM, women's careers or leadership, such as data on gender gaps in STEM education enrolment, employment rates, wage disparities or representation in leadership roles. Context indicators help quantify the influence of various contextual factors and track progress towards gender equality in STEM.

To collect data related to these concepts, we conducted several systematic reviews of academic literature, grey literature and social media to search the web and gather the available data, such as the Figure 2 examples (Arndt et al, 2023; Drummond et al., 2023; Frigo et al., 2024; Guzman et al., 2023b; Guzman et al. 2024; Ribeiro et al., 2024; Rondon-Pereira, et al., 2024). The data found

in these systematic reviews were structured to compose the ELLAS platform's dataset.

Collection of additional data

In addition to utilising pre-existing data sources, a survey was designed to collect primary data from the populations of these countries; the survey design approach is explained by Davila et al. (2024). The primary goal of this survey was to collect quantitative data on the factors that influence STEM career choices and motivation to lead in STEM in the three countries using the previously conducted systematic literature reviews as a foundation. The survey data offer comparable open data across the three countries, providing essential support to researchers, policymakers and other decision-makers dedicated to closing the gender gap in STEM in Latin America. Moreover, this survey can be replicated by researchers in other regions within and beyond Latin America to identify the contextual factors that hinder or empower women in STEM careers. We made the survey instruments open and accessible via the ELLAS website (ELLAS, n.d.).

The survey instrument was made available in three languages – Spanish, Portuguese and English. We obtained more than 10 000 responses from participants in Bolivia, Brazil and Peru. The primary data collected provide valuable insights into the motivations, challenges and factors that influence women's career decisions and aspirations in STEM (Davila et al., 2024). The responses were meticulously analysed, and in parallel, selected data underwent curation to transform survey data into open data, making it accessible on the ELLAS platform, as shown in Figure 3.

Challenges with data

The ELLAS team used various strategies, incorporating different methods, techniques, procedures and technological tools, to develop the ODP and conduct research to generate data for the ODP. This section outlines the methodological steps and how we overcame challenges to achieve our objectives.

The process of gathering pre-existing data led to several challenges, including inconsistencies in data availability across different countries due to varying levels of transparency and openness in national data policies. For instance, while Brazil's Access to Information Law facilitates public access to government data, Bolivia lacks similar legislation, posing significant barriers to data access (Hildebrand et al., 2024). Additionally, identifying relevant data

sources, determining which courses qualify as STEM and ensuring the data were gender-disaggregated required careful consideration and adaptation of best practices.

In response to these challenges, the ELLAS Network adopted innovative methods, such as incorporating data from social media platforms like Instagram (de Arruda Nunes et al., 2024; Gonçalves et al., 2023), which proved particularly effective in identifying grassroots initiatives that support women in STEM. Traditional data collection methods, such as English-language academic searches, often failed to capture locally relevant information, highlighting the importance of considering local languages and alternative data sources. In the context of Brazil, where only a small percentage of the population has some knowledge of English, leveraging local platforms and languages not only improved data collection but also ensured that the research findings were more relevant and accessible to local stakeholders (Carvalho et al., 2023).

The data organisation of the ELLAS platform involves combining data from different sources that use various vocabularies, different levels of detail and possibly different interpretations. This diversity adds complexity to the integration process, which involves more than just putting the data in one place. We developed methods to organise the data through connections to centralise the data and make interactions with the platform clear and manageable. These connections standardise the data and ensure consistent interpretation by defining the terms used.

In other words, these connections are designed to address ambiguities. For example, suppose we collect data from different sources about initiatives and policies where there are two initiatives with a "Target Audience" label, while there is one policy with an "Audience" label. Although both terms refer to the same concept, they are used differently regarding initiatives versus policies. This makes it hard to integrate the data. In the ELLAS platform, organising and standardising the data ensures that both initiatives and policies are understood as having a "Target Audience" relationship. This approach helps centralise and simplify data interpretation.

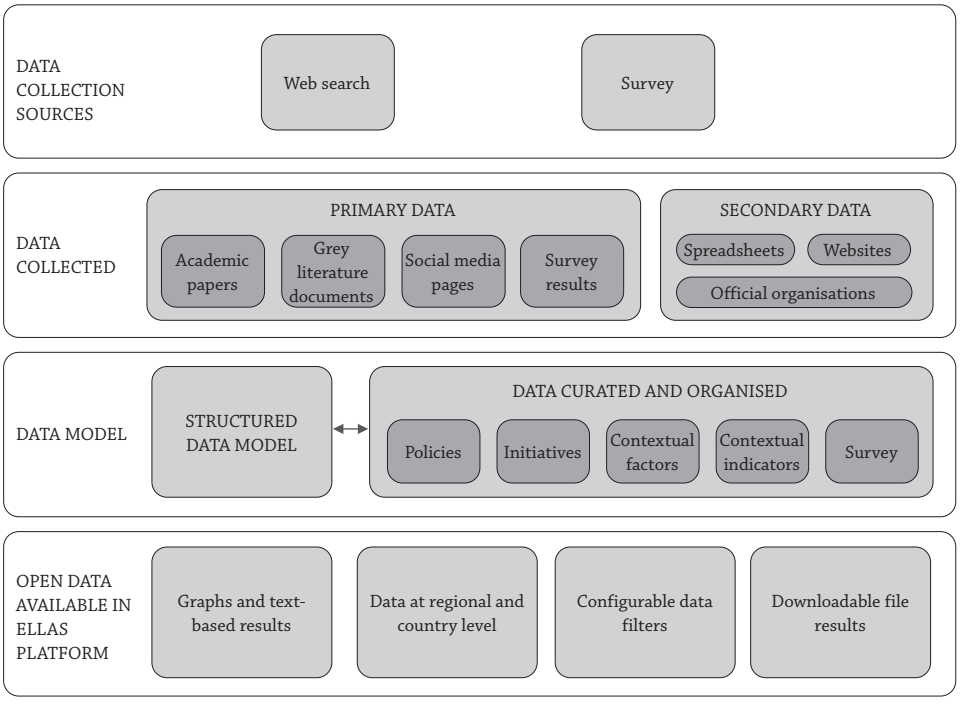
We also connect data across the three languages used on the platform. Our methods handle three languages, so even if the target audience refers to "Undergraduate students", "Estudiantes de pregrado" (in Spanish) or "Estudantes de Graduação" (in Portuguese), the connections we established in our data model recognise them as synonyms. Without proper connections, it might seem like different target audiences; however, it is conceptually the same.

This adaptive approach to data gathering demonstrates the value of using contextually appropriate methods to collect and structure comprehensive and

actionable data. By prioritising local context and leveraging diverse data sources, the ELLAS Network's open data approach addresses current gaps and sets a foundation for more inclusive and informed decision-making. This work can not only support public policymakers in addressing gender disparities in STEM but also contribute to the global call for increased investment in gender-disaggregated data, thereby fostering environments that support women's leadership and career advancement in STEM fields across Latin America.

The process depicted in Figure 3 illustrates the structured methodology adopted by the ELLAS research team to systematically gather, integrate and transform diverse data sources into a unified ODP. This systematic approach reinforces the project's commitment to addressing the gender data gap and advocating for open data practices, ultimately facilitating more informed research, policymaking and strategic actions to promote gender equity in STEM fields. The open data available in the ELLAS platform is presented in the form of graphs and text-based results, which means it can compile data at regional and country geographic levels, it has configurable data filters and people can download the results in open formats.

Figure 3: ELLAS open data approach for Latin America



The ELLAS linked open data platform

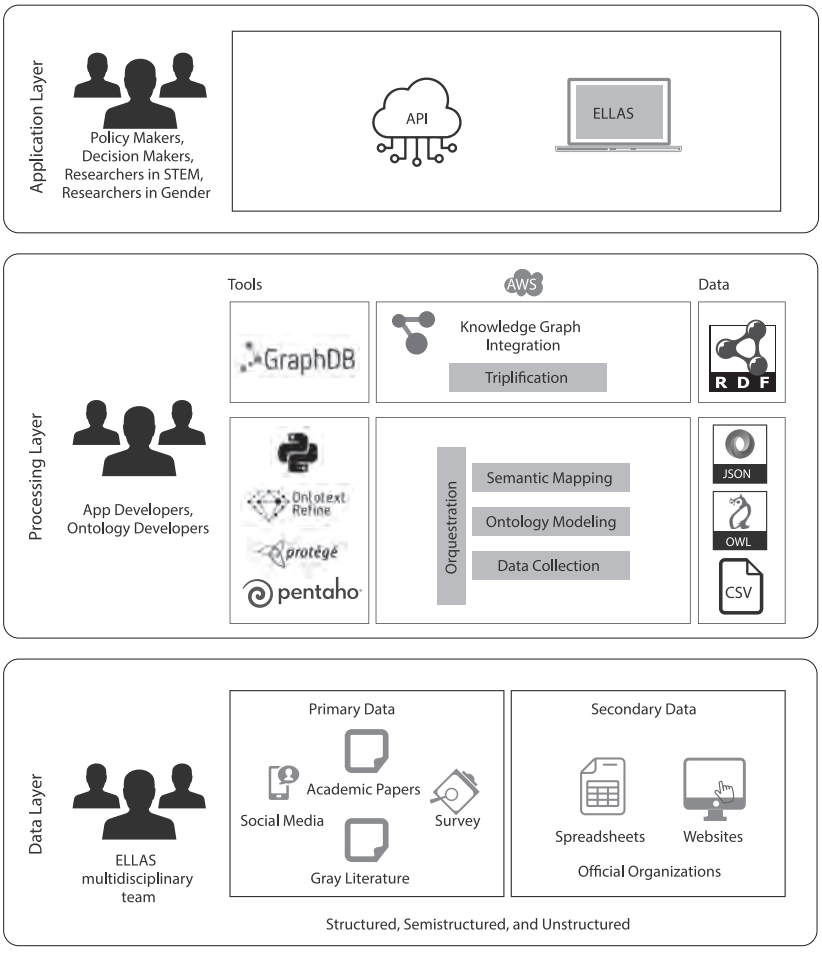
The platform architecture

We start this section by briefly explaining the open data ELLAS platform architecture, which represents the fundamental structure of our software system and is designed to allow the system to grow, providing a better user experience without performance loss. The open data ELLAS platform architecture can be viewed in more technical detail in Maciel et al. (2024). The designed architecture allows for the curation of different data sources, from raw data to linked data, to be consumed by individual users such as researchers, policymakers and decision-makers working on STEM and gender issues. The ELLAS platform has a layered architecture using cloud services throughout all processes, as we can see in Figure 4. This comprises:

- The *data layer* which organises two types of data sources defined in the context of the ELLAS platform as "primary" and "secondary" data, as previously explained. Multidisciplinary and multicultural teams collaborated to curate and prepare the data for integration into the next layer.
- The *processing layer* which collects structured files (with specific comma-separated values – CSV format) for ontology modelling in three languages (Portuguese, English and Spanish), using technologies such as Protégé and OWL. This complex pipeline is managed by Pentaho and Python technologies, involving platform and ontology developers. Finally, it is possible to integrate data via knowledge graphs, scattered across different sources, into linked data according to ontologies created by the ELLAS team.
- The *application layer* which enables users to search, understand and utilise data. It is being designed to ensure a good user experience when interacting with the platform and to help users perform tasks such as searching, selecting and understanding the steps necessary to do so.

The ELLAS website (ELLAS, n.d.), hosted by the Federal University of Mato Grosso (UFMT), provides access to the ODP and additional information about the project. The following section describes the application layer, which is the layer accessed by users.

Figure 4: Open data ELLAS platform architecture



Source: Maciel et al. (2024)

The application layer: Intended use of the ELLAS platform

The ongoing project follows a gender-sensitive design based on feminist human-computer interaction (HCI) (Bardzell, 2010; Bardzell & Bardzell, 2011; Bellini et al., 2022; Drummond et al., 2024), highlighting the importance of including gender considerations from the initial conception of the platform to promote diversity on the social web. Figure 5 shows the interaction design (ID) methodology (Oliveira et al., 2024).

By creating and making available a data platform that can be consulted at no cost, and initially available in three languages (Portuguese, Spanish and English), we believe that we are offering a model of how to develop accessible gender databases and platforms. Additionally, by providing access to data in Portuguese and Spanish, we are minimising language barriers and hopefully increasing opportunities for Brazilian and Latin American women scientists to contribute to research in this field.

We also believe that by interacting with the data available on the platform, people could better understand gender inequalities in leadership positions in technology and engineering contexts. Consequently, with knowledge and better comprehension of inequalities, ELLAS ODP users could promote positive change for women in their professional environments.

Finally, the platform can facilitate collective learning and foster collaborative research environments for anyone with fluency in the available interface languages, digital literacy and internet connectivity. For example, in the area of initiatives, we collected data about each initiative's social media information and the gender of its coordinators. Users can ask: *What is the initiative's social network(s)?* to gain access to the social media information for the initiative. Users can also search for the following: *What is the social gender of the people responsible for the initiatives?*, to find information about the gender of those coordinating the initiatives if it was available in the initiative's public media. This feature allows the platform to identify and highlight women leaders in STEM and to facilitate the creation of new women's networks. However, it is important to clarify that the platform does not provide individuals' direct contact details but offers general information about initiatives. The focus on ethical and privacy considerations ensures that the shared data represent demographic overviews rather than personal details.

We collected sex – and gender – as disaggregated data at the country level, allowing for users to make searches by country. For example, by selecting the question *What types of gender policies were implemented in countries in a given year?*, the user can select which country(ies) they want to search.

We also collected data that indicate the educational levels of all initiatives. Consequently, from selecting the question: *What initiatives are being developed at a given school level?*, the user can choose which school level they would like to investigate.

The platform offers the possibility for a user to provide feedback regarding their satisfaction with the data presentation and if they find any technical issues or errors.

Figure 5 shows a storyboard with an example of an interaction scenario. We expect that with these functionalities, the platform can foster a culture of accountability and collaboration with the contribution of the public interacting with it and sharing their data with regular updates.

Throughout the platform's design process, we aimed to ensure that photos, symbols and illustrations represent Latin American women in technology and engineering, mainly women from the countries directly involved in the project: Bolivia, Brazil and Peru. We believe that by doing this, we can recognise and address implicit biases from the evaluation of the platform prototype, thereby contributing to reducing gender stereotypes and promoting inclusion, diversity and equity. For example, the Bolivian team answered that they did not feel represented by the photos used in the first prototype. So, with the help of a Bolivian female student, we searched for other photos that could be more inclusive and then included a photo of a Cholita, representing indigenous women in Bolivia.

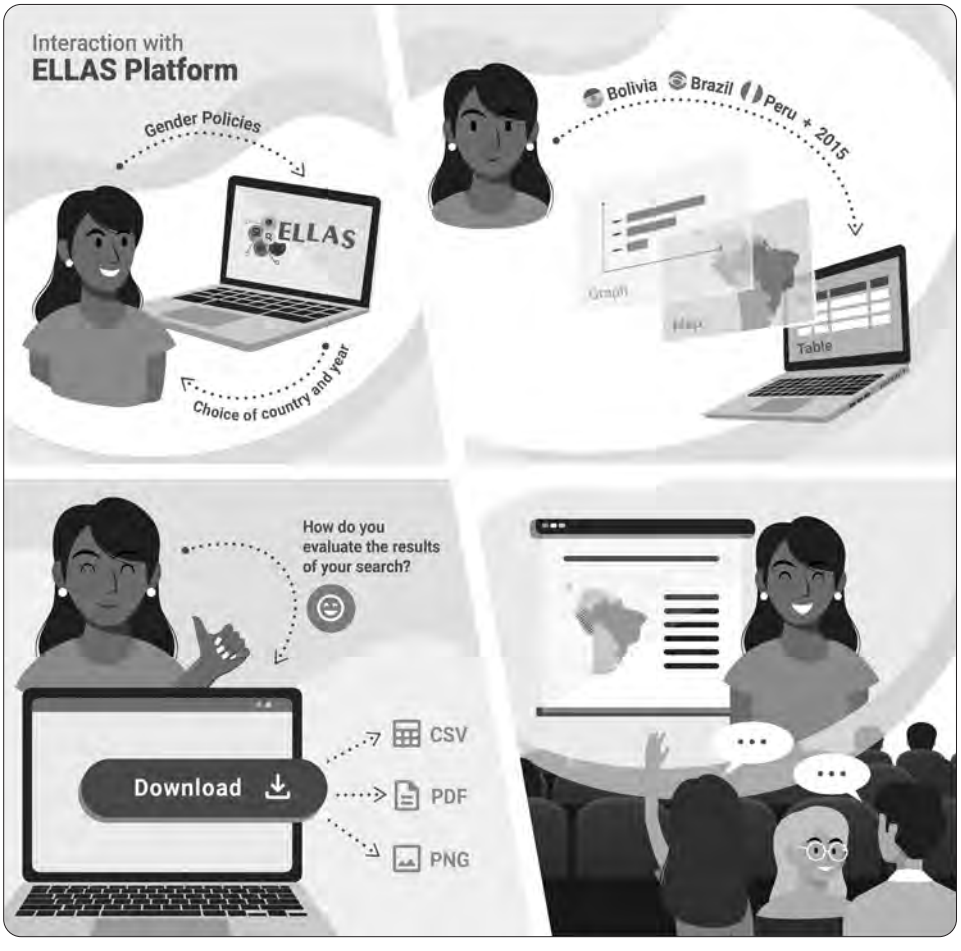
By accessing and using the ELLAS ODP, users can explore data related to initiatives, contextual factors, context indicators, policies and survey data from Bolivia, Brazil and Peru. The ELLAS ODP offers users access to curated datasets that can be downloaded for further exploration and analysis. This organised data can serve as a valuable starting point for further exploration and informed discussion. We hope this resource will be especially useful to decision-makers, researchers and policymakers in the following ways:

- **Data on initiatives could help in the identification of best practices, resource allocation and programme design:** Data on various initiatives from these countries could assist in identifying which programmes and approaches are encouraging women's participation and leadership in STEM. By navigating the data, insights from existing initiatives could guide decision-makers on where to allocate resources to maximise impact. Furthermore, a greater understanding of past initiatives already established in these countries could aid in designing new programmes.
- **Data on contextual factors could help to better understand barriers in order to tailor solutions and measure progress:** Contextual data helps in identifying cultural, social and economic barriers that women face in STEM fields. By understanding the specific context of each country, policies and initiatives could be tailored to address unique local challenges. Also, context indicators such as gender equality indices, education levels and employment data are available on the platform to download and could be used to understand their impact.

Figure 5: Storyboard illustrating an interaction with the ELLAS platform

Interaction scenario example
Patrícia Ribeiro is a student at a university in Latin America who is active in the feminist and student movements. This semester, she is taking an elective course on Public Policy. One of her course assignments involves searching the internet for data on the gender gap in STEM. Patrícia needs to research the gender indices of Latin American countries and universities, the types of gender policies already implemented by countries or universities, and what gender aspects these policies address.

Patrícia recently learned that a data platform made by the ELLAS Network is available, so she decided to explore the data from this source. She searched for policies, refining the search by country and time interval (in years). The results were presented in different formats, such as a map, tables and graphs. Patrícia was very pleased with the results. She downloaded the results to include them in her class presentation. The debate with the class, based on the data collected by Patrícia on the ELLAS platform, was very rich and informative.



- **Data on policies could help policy development, evaluation, advocacy and lobbying:** The ELLAS ODP presents data about STEM-related policies identified across these countries, which could help in evaluating the effectiveness of different policy approaches. Learning from existing policies could aid in the development of new, more effective policies. Additionally, analysing the data on existing policies, alongside survey data on individual perceptions of these policies, could be used as evidence for advocating for changes at both national and regional levels.
- **Data from the survey can help in analysing the impact of the different factors and make comparisons among the three countries:** The survey data could help decision-makers understand the primary factors that encourage or discourage women from pursuing STEM careers in Bolivia, Brazil and Peru. By knowing what drives or hinders career choices, policies and programmes can be developed to address specific barriers and enhance motivations. The comparable data across the three countries can allow policymakers to identify common trends and country-specific challenges, facilitating tailored and more effective interventions. The shared data can foster collaboration between Bolivia, Brazil and Peru, allowing for the exchange of best practices and the development of joint strategies to promote women in STEM. The data may help in crafting contextually appropriate solutions, ensuring that policies and programmes resonate with the unique experiences and challenges faced by women in each country. Additionally, the open nature of the survey instrument allows for replication in other regions, enabling policymakers to extend their understanding of the factors influencing women in STEM beyond these three countries and apply successful strategies in diverse contexts.

The overall benefits of the use of the ELLAS platform by decision-makers and policymakers are the evidence-based decision-making process to craft effective policies, the regional collaboration that comes with understanding the different approaches and outcomes to promote women in STEM in neighbouring countries, and the cultural features of the platform data to ensure that policies are culturally sensitive and more likely to succeed in their specific contexts. By leveraging the ELLAS platform data, decision-makers and policymakers could create more effective, gender-equitable and sustainable strategies to enhance women's participation and leadership in STEM across Bolivia, Brazil, Peru and other Latin American countries with data on the platform. The platform and available data can be accessed on the ELLAS website (ELLAS, n.d.).

Conclusion

In this chapter, we shared the outcomes of our efforts to promote open data for gender equality in STEM fields in Latin America (for additional work on gender equality in Latin America, see chapters 2, 5 and 11). We presented the core concepts driving the ELLAS project – including policies, initiatives, contextual factors and indicators – demonstrating how these frameworks guided the retrieval of existing data from the web and the collection of new data through a survey on women in STEM in Bolivia, Brazil and Peru. We also described how we created a model to structure this data as open data and made it accessible through a platform that we designed and developed called ELLAS ODP.

The ELLAS ODP empowers people to access valuable data on women in STEM in Latin America, enabling research, analysis and evidence-based policy and decision-making to enhance the representation of women in these fields. Furthermore, the structural model created is designed to be connectable, flexible and scalable, allowing others to contribute new data to the platform to expand its coverage across other countries. We recognise the scarcity of data on this topic and its impact on the visibility and career advancement of women in STEM across the region. This is why we believe that advocating for open data is crucial in addressing this issue and that our work is a meaningful step towards supporting gender equality in STEM in Latin America.

By increasing data availability and centralising it in one accessible and contextually sensitive designated place, the ELLAS platform becomes a practical tool that people can rely on and use for various purposes, like research, curiosity, inspiration or in their careers. We hope to spark curiosity, inspire action and encourage more people to explore and utilise these data, contributing to a future where gender equality in STEM is not just an aspiration but a reality.

The ELLAS platform provides a structured and accessible way to collect and report open gender data (Guzman et al., 2020) at various levels, directly contributing to monitoring and promoting transparency and gender equality across Latin America.

As the ELLAS Network, we focus on advancing gender equality in STEM through informed, strategic and actionable measures. We advocate for promoting transparency by making data about women in STEM, particularly in leadership roles, openly available. This transparency is crucial for fostering accountability and driving progress. Additionally, we emphasise the importance of creating a culture where open gender data is utilised for evidence-based policy and decision-making, particularly at local and regional levels where

context-specific solutions are often needed. To facilitate this, we recommend developing user-friendly and culturally sensitive open data and open science platforms and tools that make gender data accessible and actionable. Lastly, we stress the need to support and integrate open data platforms, such as the ELLAS platform, to ensure long-term sustainability and effectiveness. By implementing these recommendations, we aim to create a robust infrastructure that enhances data availability and drives meaningful and sustained improvements in gender equality within STEM fields.

Looking ahead, our future work will focus on broadening the impact and utility of the ELLAS platform. We plan to socialise the use of the open gender data platform with diverse audiences, ensuring that a wide range of stakeholders can leverage its resources. To enhance the reliability and applicability of our data collection efforts, we aim to validate the survey data via statistical analysis to formalise it as a replicable tool for gathering gender data on women's leadership in STEM. We will also develop comprehensive documentation and protocols to guide users in utilising the ELLAS platform and its data for evidence-based policy and decision-making via use cases. Additionally, we are committed to ensuring the scalability of the ELLAS data warehouse, expanding its capacity to include new primary and secondary datasets from other Latin American and Caribbean countries, as well as from other countries in the Global South. To secure the platform's long-term viability, we will build pathways towards sustainability by integrating data translation processes and aligning with established national and international open data standards. Through these efforts, we strive to enhance the platform's impact and contribute to the ongoing advancement of gender equality in STEM across the region.

Previous publications are available for those interested in learning more about the ELLAS Network; these publications provide details about the methods, techniques and tools used in different stages of the research process and development of the platform. For example, Maciel et al. (2024), Berardi et al. (2024) and Lima et al. (2023) elaborate on the platform architecture. The conceptual models used to build the survey are available in Drummond et al. (2023) and Guzman et al. (2023a), while the survey itself can be found in Davila et al. (2024), and the information they collected for the platform in other studies is described by Ribeiro et al. (2024), Rondon Pereira et al. (2024), Guzman et al. (2024) and Frigo et al. (2024).

Acknowledgements

We appreciate the financial support of the International Development Research Centre (IDRC) of Canada and the administrative support of the Fundação de Apoio e Desenvolvimento da Universidade Federal de Mato Grosso (UNISELVA). We would also like to thank all individuals who have contributed to the ELLAS project and research network.

About the Authors

Cristiano Maciel is a Professor at the Institute of Computing, Federal University of Mato Grosso (UFMT), Brazil – ELLAS Project General Coordinator.

Indira R. Guzman is a Professor of Computer Information Systems and Director of the Mitchell Hill Center for Digital Innovation, California State Polytechnic University, Pomona, USA – ELLAS Research Consultant.

Karen da Silva Figueiredo Medeiros Ribeiro is a Professor at the Institute of Computing, Federal University of Mato Grosso (UFMT), Brazil – ELLAS Research Consultant at UFMT.

Franci Suni-Lopez is a Professor at the Laboratorio de Inteligencia Artificial, Universidad de Lima, Perú – ELLAS Research Consultant at ULima.

Luciana B. Frigo is a Professor at the Knowledge Engineer Department, Federal University of Santa Catarina (UFSC), Brazil – ELLAS Local Manager at UFSC.

Luciana Salgado is a Professor at the Institute of Computing, Fluminense Federal University (UFF), Brazil – ELLAS Local Manager at UFF.

Rita Cristina Galarraga Berardi is a Professor at Federal University of Technology – Paraná (UTFPR), Brazil – ELLAS Local Manager at UTFPR.

Sílvia Amélia Bim is a Professor at the Academic Department of Informatics, Federal University of Technology – Paraná (UTFPR), Brazil – ELLAS Local Leader at UTFPR.

Boris Branisa is a Professor at the School of Production and Competitiveness (ePC), Universidad Católica Boliviana (UCB) "San Pablo", Bolivia – ELLAS Local Manager at UCB.

Elizabeth Jimenez is a Professor at the Centre for Multidisciplinary Development Studies (CIDES), Universidad Mayor de San Andres (UMSA), Bolivia – ELLAS Local Manager at UMSA.

Luana Bulgarelli Mendes is the ELLAS Project Manager, Foundation for Support and Development of the Federal University of Mato Grosso (UNISELVA), Brazil.

References

Arndt, G. J., Gonçalves, M. B., Miguel, R. D. B. P., & Frigo, L. B. (2023). Mulheres em STEM: Produções acadêmicas no contexto brasileiro. *Interfases*, (018), 133–148. https://revistas.ulima.edu.pe/index.php/Interfases/article/view/6615

Babbie, E. R. (2020). *The practice of social research* (15th ed.). Cengage. https://books.google.co.za/books?hl=en&lr=&id=lFvjDwAAQBAJ&oi=fnd&pg=PP1&ots=I4yT4z2UOb&sig=-s0uh_z0KSefdHyjky-VOmXeKonU&redir_esc=y#v=onepage&q&f=false

Bardzell, S. (2010). Feminist HCI: Taking stock and outlining an agenda for design. In *CHI '10: Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1301–1310). Association for Computing Machinery. https://doi.org/10.1145/1753326.1753521

Bardzell, S., & Bardzell, J. (2011). Towards a feminist HCI methodology: Social science, feminism, and HCI. In *CHI '11: Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 675–684). Association of Computer Machinery. https://doi.org/10.1145/1978942.1979041

Bellini, R., Meissner, J., Mitchell Finnigan, S., & Strohmayer, A. (2022). Feminist human-computer interaction: Struggles for past, contemporary and futuristic feminist theories in digital innovation. *Feminist Theory, 23*(2), 143–149. https://doi.org/10.1177/14647001221082291

Berardi, R. C. G., Auceli, P. H. S., Maciel, C., Davila, G., Guzman, I. R., & Mendes, L. (2024). ELLAS architecture and process: Collecting and curating data on women's presence in STEM. *Journal on Interactive Systems*, *15*(1), 530–540. https://journals-sol.sbc.org.br/index.php/jis/article/view/3853

Berners-Lee, T. (2006, July 27). Linked data [Blog post]. *W3C*. https://www.w3.org/DesignIssues/LinkedData.html

Branisa, B., Cabero, P., & Guzmán, I. (2020). ¿P*or qué tan pocas mujeres en Bolivia optan por carreras universitarias relacionadas con STEM (Ciencia, Tecnología, Ingeniería y Matemáticas) y qué podemos hacer para cambiar esta situación?* SDSN Bolivia. https://www.sdsnbolivia.org/wp-content/uploads/2020/10/10-Estudio-Transversal-Por-que-tan-pocas-mujeres-optan-por-carreras-universitarias-relacionadas-con-STEM-en-Bolivia-y-que-podemos-hacer-para-cambiar-esta-situacion.pdf

Brasscom. (2022). Relatório de Diversidade no Setor TIC. https://brasscom.org.br/wp-content/uploads/2024/03/BRI2-2023-013-Diversidade-v18-1.pdf

Carvalho, L. P., Bim, S. A., Santoro, F. M., & Oliveira, J. (2023). A critical analysis on Brazilian computational scientific events linguistic aspects. In *IHC '23: Proceedings of the XXII Brazilian symposium on human factors in computing systems* (pp. 1–11). Association of Computing Machinery. https://doi.org/10.1145/3638067.3638125

Davila, G. A., Guzman, I. R., Rodriguez, N., Frigo, L., Bim, S. A., & Maciel, C. (2024, June). *Development of a survey instrument for assessing the gender gap in STEM* [Paper presentation]. Paper presented at the 19th Iberian Conference on Information Systems and Technologies (CISTI), University of Salamanca, Salamanca, Spain.

de Arruda Nunes, A. C., Moro, F. F., Frigo, L. B., Miguel, R. d. B., & Bim, S. A., (2024). Mapeamento de iniciativas Brasileiras em STEM no Instagram: Uma discussão metodológica. In *XVI Congress of Latin American women in computing 2024 (LAWCC/CLEI)* (pp. 75–86). https://ceur-ws.org/Vol-3872/paper7.pdf

Drummond, B. M., Cardoso de Castro Salgado, L., &Viterbo Filho, J. (2024). From barriers to inclusion: A female-inclusive assessment framework for interface evaluation. In *Anais do XV Workshop sobre aspectos da interação humano-computador na web social* (pp. 37–44). Sociedade Brasileira de Computação. https://sol.sbc.org.br/index.php/waihcws/article/view/32172

Drummond, B., Salgado, L., Avelino, M., Viterbo Filho, J., Ribeiro, K., Cigüeñas, M., Dávila, G., & Branisa, B. (2023). Mapping contextual aspects that influence women in computing in Latin America. *Interfases*, (018), 19–30. https://revistas.ulima.edu.pe/index.php/Interfases/article/view/6610/6663

ELLAS. (n.d.). *Equality in Leadership for Latin America STEM* [Website]. ELLAS. https://ellas.ufmt.br/

Frigo, L. B., Ferreira dos Santos, A., Rodriguez-Rodriguez, N., de Barros Pinto Miguel, R., Bim, S. A., Maciel, C., & Guzman, I. R. (2024). Mapping women STEM initiatives in Latin American Countries: Bolivia, Brazil and Peru. In Á. Rocha, C. Ferrás, J. Hochstetter Diez & M. Diéguez Rebolledo (Eds.), *Information technology and systems. ICITS 2024*, *933*(1), 401–409. Springer.

García-Holgado, A., & García-Peñalvo, F. J. (2022). A model for bridging the gender gap in STEM in higher education institutions. In *Women in STEM in higher education: Good practices of attraction, access and retainment in higher education* (pp. 1–19). Springer.

Gonçalves, M. B., Arndt, G. J., Miguel, R. de B. P., & Frigo, L. B. (2023). "El lugar de una mujer está en la tecnología": Un análisis de perfiles eños de Instagram sobre mujeres en tecnología. *Interfases*, (018), 159–167. https://revistas.ulima.edu.pe/index.php/Interfases/article/view/6617

Guzman, I., Berardi, R., Maciel, C., Cabero Tapia, P., Marin-Raventos, G., Rodriguez, N., & Rodriguez, M. (2020). Gender gap in IT in Latin America. *AMCIS 2020 Proceedings*. 4. https://aisel.aisnet.org/amcis2020/panels/panels/4

Guzman, I. R., Sánchez, A., Gonzales Lopez, R. (2023a). Motivation to lead of IT professionals: The role of IT occupational culture, commitment, and gender equality policies in Latin America. *AMCIS 2023 Proceedings*. 2. https://aisel.aisnet.org/amcis2023/lacais/lacais/2

Guzman, I. R., Viterbo, J., dos Santos Nunes, E. P., Suni-Lopez, F., Jimenez, E., & Casagrande, A. L. (2023b). Gender equality policies in STEM in Latin America – The use of gray and systematic literature searches. *SIGMIS-CPR '23: Proceedings of the 2023 computers and people research conference* (Article 21, pp. 1–2). Association for Computing Machinery. https://doi.org/10.1145/3579168.3632631

Guzman, I. R., Viterbo, J., dos Santos Nunes, E. P., Suni-Lopez, F., Jimenez, E., Casagrande, A. L., & Maciel, C. (2024). Gender equality policies in STEM in Latin America – A systematic literature review. In Á. Rocha, C. Ferrás, J. Hochstetter Diez & M. Diéguez Rebolledo (Eds.), *Information technology and systems. ICITS 2024, 933*(1), 410–419. Springer. https://doi.org/10.1007/978-3-031-54256-5_39

Hildebrand, N. D., Amador, B. O., Maciel, C., & Berardi, R. C. G. (2024, July). A escassez de dados abertos estruturados em Países Latino-Americanos com enfoque de gênero na educação superior. In *Women in information technology* (pp. 161–172). Sociedade Brasileira de Computação (SBC). https://sol.sbc.org.br/index.php/wit/article/view/29561/29364

Keserű, J., & Chan, J. K-S. (2015). *The social impact of open data*. Paper presented at the 3rd International Open Data Conference 2015 (IODC), 28–29 May, Ottawa. https://www.researchgate.net/publication/298646716_The_Social_Impact_of_Open_Data

Lima, P. R. S., Salgado, L. C. de C., & Bim, S. A. (2023). Primer paso en el proyecto para la interacción de una plataforma de datos abiertos con la lente feminista interseccional. *Interfases*, (018), 65–74. https://revistas.ulima.edu.pe/index.php/Interfases/article/view/6605/6667

Lucena, L., Santos, M., Gonçalves, S., & Coutinho, J. (2024). Gênero, parentalidade e remuneração em cargos de desenvolvimento na indústria de software. In *Anais do XVIII Women in information technology* (pp. 24–34). Sociedade Brasileira de Computação (SBC).

Maciel, C., Guzman, I., Berardi, R., Caballero, B. B., Rodriguez, N., Frigo, L., Salgado, L., Rodriguez-Rodriguez, N., Bim, S. A., & Tapia, P. C. (2023). Open data platform to promote gender equality policies in STEM. In *Proceedings of the Western Decision Sciences Institute (WDSI)*. https://ellas.ufmt.br/open-data-platform-to-promote-gender-equality-policies-in-stem/

Maciel, C., Guzman, I. R., Berardi, R. C. G., Rodriguez-Rodriguez, N., Salgado, L., Frigo, L. B., Branisa, B., & Jimenez, E. (2024). An open data platform to advance gender equality in STEM in Latin America. *Communications of the ACM, 67*(8), 90–92. https://doi.org/10.1145/3653294

Oliveira, R., Salgado, L., Bim, S. A. (2024). Desenvolvimento de uma plataforma de dados abertos a partir do design sensível ao gênero. In *Anais do XV workshop sobre aspectos da interação humano-computador para a web social (WAIHCWS)* (pp. 29–36). Sociedade Brasileira de Computação. https://sol.sbc.org.br/index.php/waihcws/article/view/32171

Ribeiro, K., Gomes, J. M., & Aguiar, B. R. (2024). Abertura e disponibilidade de dados de gênero sobre liderança de mulheres Brasileiras em TI. In *AMCIS 2024 Proceedings*. 2. https://aisel.aisnet.org/amcis2024/span_lacais/span_lacais/2/

Rondon Pereira, L. R., Ferreira de Lima, S. D., Maciel, C., & Guzman, I. R. (2024). Equity in focus: A qualitative study of female representation in the context of academic administration and STEM of federal universities in Brazil. *SIGMIS-CPR 2024: Proceedings of ACM computer people research* (pp. 1–8). Association of Computing Machinery. https://doi.org/10.1145/3632634.3655858

Rosales Rodríguez, M. A. (2020). Relación entre la inclusión y el abandono de mujeres jóvenes en carreras y áreas STEM y TIC. *Fides et Ratio-Revista de Difusión Cultural y Científica de la Universidad La Salle en Bolivia, 20*(20), 141–166.

Tull, R., Jangha, S., Medina, Y., Bell, T., & Parker, R. (2018). Sharing peace engineering with US-based minority students, through the UN's Sustainable Development Goals, in Peru. In *2018 World Engineering Education Forum – Global Engineering Deans Council (WEEF-GEDC)* (pp. 1–6). IEEE. https://doi.org/10.1109/WEEF-GEDC.2018.8629764

UNDP (United Nations Development Programme). (2024, May 7). Coded bias: The underrepresentation of women in STEM in Latin America and the Caribbean [Blog post]. *United Nations Development Programme*. https://www.undp.org/latin-america/blog/coded-bias-underrepresentation-women-stem-latin-america-and-caribbean

UNESCO (United Nations Educational, Scientific and Cultural Organization). (2020). *Closing the gender gap in STEM: Drawing more girls and women into Science, Technology, Engineering, and Mathematics* [Programme and meeting document]. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000245717?posInSet=1&queryId=ead18c73-9f22-45aa-a7dc-d908b865ebc4

UNESCO. (2024). *UNESCO Call to action: Closing the gender gap in science* [Programme and meeting document]. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000388641

World Bank Group. (2024). *Gender Data Portal: Gender data availability*. https://genderdata.worldbank.org/en/data-availability