International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I12P202

Volume 70 Issue 12, 13-19, December 2022

Original Article

Using Data Analytics to Monitor Gender Equality in Higher Education Institutions

Silvia Gaftandzhieva[1], Rositsa Doneva[2], Marieta Atanasova[3], Milen Bliznakov[4]

[1,2,3,4]*University of Plovdiv "Paisii Hilendarski", Plovdiv, Bulgaria*

[1]*Corresponding Author : sissiy88@uni-plovdiv.bg*

Received: 23 October 2022 Revised: 10 December 2022 Accepted: 13 December 2022 Published: 24 December 2022

Abstract *- Nowadays, many policies have encouraged organizations to develop Gender Equality Plans (GEPs) and monitor their implementation. The paper presents a prototype of a data analytics tool for gender equality monitoring GEAnalyst designed for the needs of higher education institutions (HEIs) from the perspective of different stakeholder groups. The tool allows them to understand key gaps between women and men within the HEI and its activities, take data-informed decisions to ensure equal access to education and career development, increase the sense of equality among the university community, set priorities, and adjust these priorities as the situation evolves.*

Keywords *- Data analytics, Software tools, Data collection, Monitoring, Higher education institutions.*

1. Introduction

Nowadays, many policies have stimulated organizations from all sectors to develop and implement Gender Equality Plans (GEPs). Organizations are not only strongly encouraged by European policies from all sectors to develop and implement Gender Equality Plans (GEPs), but a Plan like that has become a suitability criterion for higher education institutions (HEIs), all public bodies, and research organizations from the EU member states and accompanying countries who want to participate in the Horizon Europe Framework Programme for Research and Innovation 2021-2027.

Organizations must collect and publish sex/gender-disaggregated data on personnel (and students, where relevant) and carry out annual reporting based on indicators to be eligible for Horizon Europe [1]. The most pertinent indicators should consider by Organizations on how to select, how to collect and to analyse the data, including resources to do so, and should ensure that data is published and monitored on an annual basis.

As a result, the following analysis and the collection of sex-disaggregated data require human resources participation and manually perusing endless data streams. On the other hand, the presented information does not provide information about the organization's current state and is up-to-date at the time of its analysis. It raises the question of whether the analysis and collection of such data can be computerized and thus not only reduce the amount of physical work and the number of people involved and provide current information at any time. In this regard, data analysts face the challenge of leveraging technology to eliminate bias and promote gender equality. The United Nations [2] also encourage data scientists to collect and analyse gender data to measure the progress towards achieving the Sustainable Development Goals (SDGs) for 2030.

Platforms like Gallup Analytics (https://www.gallup.com), Pipeline (https://www.pipelineequity.com/), Hirevue (https://www.hirevue.com), Pymetrics (https://www.pymetrics.ai), Textio Hire (https://textio.com), WEPs Tool (https://weps-gapanalysis.org/), Equialo (https://www.equilo.io/) allows organizations to use data analysis technologies to close the gender gap in the workplace. Organizations from all business areas can use data analytics to illuminate an existing gender gap, monitor trends and discrepancies between the treatment of female and male staff, recruit a diverse workforce, predict whether certain groups of people are more likely to resign than others, and eliminate bias [3-5].

HEIs use administrative staff, students and PhD students, and software systems to automate their activities, which store a large quantity of data on faculty that they could extract and analyse from data analytics tools. Therefore, HEIs have the data sets to benefit from targeted data analytics, which increases interest in analysis and data mining. HEIs are searching for software tools so that they will be allowed to extract data from information systems and convert them into knowledge. As a result, it helps them make important decisions regarding workflows and various processes and improves process management. Data analytics has the potential to act positively on all the main areas of

importance for HEIs, including monitoring gender equality. Data analysis can bring accountability and transparency in the management of the education sector [23] and assessment of institutional performance and progress, as well as help HEI leadership predict future performance and identify potential issues [25].

Automatic extraction, analysis and classification of data are allowed by data analytic tools that support governing bodies of HEIs to make informed decisions. Data analytics tools have been used by many HEIs worldwide [8-17] to improve the quality of learning, track data for institution's work, student retention and improve student results, deliver early interventions and immediate feedback, support decision-making, and make significant progress in improving the university processes, generate reports for internal and external quality assurance, etc. They pay less attention to the possibility of using data analytics tools to monitor gender equality in HEIs. On the other hand, platforms that allow organizations to use data analysis technologies to close the gender gap in the workplace are unsuitable for HEIs because they do not fully reflect the specifics of educational institutions.

These facts motivate the development of a tool for monitoring gender equality in HEIs. The development of a data analytics tool that will extract and analyse sex-disaggregated data about academic and non-academic staff, PhD students and students will help different stakeholders (e.g. top and middle management, responsible bodies for gender equality, etc.) to make data-driven decisions in order to improve the university environment referring to equality between women and men.

A prototype of a data analytics tool for gender equality monitoring designed is being presented in this paper for the needs of top and middle management and gender equality-responsible bodies. Key gaps between women and men within the HEI and its activities are being understood by this tool. The tool helps managers to make data-informed decisions to ensure equal access to education and career growth for women and men, increase the sense of equality in the university, set priorities and adjust these priorities when the situation evolves.

2. Materials and Methods

Based on the literature review in the field [1, 18-19], the Plovdiv University (PU) GEP in force [24], and available data in potential data sources, a model with a set of indicators is being proposed to serve as a business logic basis of the developed data analytics tool (see Section 3).

The set goal of the tool determines the selection of indicators to collect as much relevant data as possible to enable scrutiny of the differences between men and women in different roles and at different levels within the university's core activities. After the indicators are selected, further breaking down the collected sex-disaggregated data into other categories is considered to study the differences between women and men in the different units of the university. Stakeholders are allowed to examine data on the interrelationship of gender with other characteristics that can highlight specific areas requiring attention.

Gender equality models for monitoring define what type of data they should collect from the institutional information infrastructure that stakeholders can use to monitor gender equality at the university. The model consists of measurable indicators allowing the relevant stakeholder to track sex-disaggregated data for candidate PhD students and PhD students, students, students, and members of the faculty staff for different purposes, e.g. monitoring, analysis, forecast, intervention, recommendations, etc., but finally to improve the university environment in terms of equality between women and men.

They have built the model for gender equality monitoring as hierarchies of indicators of different levels. Indicators from Level 1 represent the object to which the collected and aggregated data relate – candidate PhD students, PhD students, candidate students, students, and faculty staff members. These indicators group together a set of Level 2 indicators that allow the relevant stakeholder to track specific data for this object. Each Level 2 indicator comprises several measurable attributes whose values will be extracted from the university information systems. Table 1 presents the proposed model and its indicators for Levels 1 and 2.

3. Results and Discussion: Data Analytics Tool

Based on the model proposed (Section 2), a corresponding data analytics tool for gender equality monitoring called GEAnalyst is created and implemented.

Through an analytical study of software solutions for analysing, extracting, and visualizing data from different information sources, some technologies and tools for software development were selected. The GEAnalyst has been developed by combining current software solutions, such as *JasperReport Server*, *Jaspersoft ETL* and *JasperSoft Studio* tools (developed by TIBCO Software, https://www.tibco.com/) and the client application.

The *JasperSoft Studio* delivers a variety of instruments to design report templates which are to be filled out with data retrieved from a range of data sources (relational databases, big data sources, or other types of database systems). Together with JasperReport Server, effective report publishing workflows could be created.

There are some options that JasperReport Server offers, such as accessing data collections with a different type of organization (incl. custom - DB, XML, CSV, Hibernate, POJO), organizing structured repositories, and making use of them as data sources for JasperSoft Studio needs when generating and storing reports. Besides, it could present them in the preferable form by the user. The server combines a variety of tools for incorporation with a number of software applications through shared web services.

Table 1. Monitoring for gender equality

Level 1	Level 2
1. Candidate students	1.1. Number of both genders (women and men) amongst applicant students per study programme 1.2. Number of both genders (women and men) amongst applicant students per professional field 1.3. Number of both genders (women and men) amongst enrolled students per study programme 1.4. Number of both genders (women and men) amongst enrolled students per professional field 1.5. Number of both genders (women and men) amongst enrolled students per faculty programme
2. Students	2.1. Number of both genders (women and men) amongst students per faculty 2.2. Number of both genders (women and men) amongst students per study programme 2.3. Number of both genders (women and men) amongst students per professional field 2.4. Number of both genders (women and men) amongst dropped out students per faculty 2.5. Number of both genders (women and men) amongst dropped out students per study programme 2.6. Number of both genders (women and men) amongst dropped out students per professional field 2.7. Number of both genders (women and men) amongst graduate students per faculty 2.8. Number of both genders (women and men) amongst graduate students per study programme 2.9. Number of both genders (women and men) amongst graduate students per professional field 2.10. Number of both genders (women and men) at the end of the academic year per faculty 2.11. Average grade of both genders (women and men) at the end of the academic year per study programme 2.12. Average success of both genders (women and men) in graduation per study programme 2.13. Maximum success of both genders (women and men) in graduation per study programme 2.14. Minimum success of both genders (women and men) in graduation per study programme 2.15. Number of students (women and men) who participate in research projects per faculty
3. Candidate PhD Students	3.1. Number of both genders (women and men) - candidate PhD students per study programme 3.2. Number of both genders (women and men) among enrolled PhD students per study programme
4. PhD students	4.1. Number of both genders (women and men) - PhD students per study programme 4.2. Number of both genders (women and men) who dropped out PhD students per study programme 4.3. Number of both genders (women and men) amongst defended doctoral students per faculty 4.4. Number of both genders (women and men) amongst defended doctoral students per study programme 4.5. Number of both genders (women and men) amongst defended doctoral students per professional field 4.6. Number of both genders (women and men) amongst PhD students who participate in research projects per faculty
5. Faculty staff	5.1. Number of both genders (women and men) in an academic position per faculty 5.2. Number of both genders (women and men) in an academic position per professional field 5.3. Number of both genders (women and men) amongst academic staff with scientific degrees per faculty 5.4. Number of both genders (women and men) amongst academic staff with scientific degrees per professional field 5.5. Number of both genders (women and men) amongst academic staff who participate in research projects per faculty 5.6. Number of both genders (women and men) amongst academic staff who participate in research projects per professional field 5.7. Number of both genders (women and men) amongst project managers per faculty 5.8. Number of both genders (women and men) amongst project managers per professional field

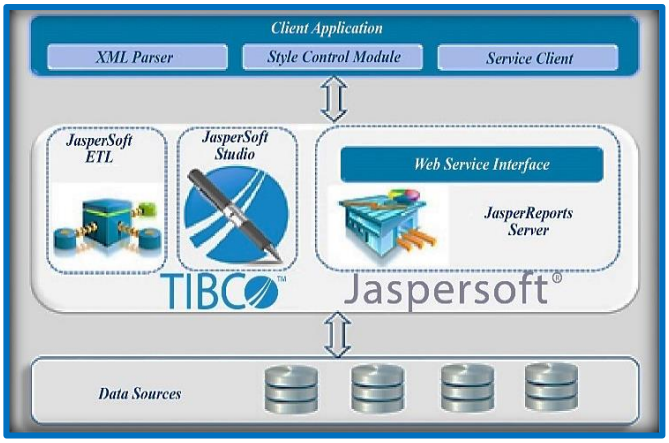

Fig. 1 GEAnalyst Architecture

Jaspersoft ETL (extract, transform, and load) is an open-source tool for data integration. This tool is applied to extract transactional system data to create a joined data warehouse or data mart for analysis and reporting. Some of its applications are graphic design, schedules, and executing data movements and transformations for business intelligence projects, such as loading an Operational Data Store (ODS), Data Mart, or Data Warehouse.

The architecture of the GEAnalyst (see Figure 1) demonstrates the average type of 3-tier architecture with recognized three layers - Presentation, Application and Data layers.

The Client application allows users to ask for a report by a chosen template and view the result of the request (visualized report). The Client application (using XML Parser and Style Control Module functionalities) allows users, through predefined conditions to adjust some view attributes like colour, font size, etc., to visualize the report in the web browser in a user-friendly way.

The report templates design tool JasperSoft Studio has implemented the core functionality of the Application Layer of GEAnalyst and its business logic.

Displaying the developed model (see Section 2) is a key component of this functionality and acquisition of values for the model's indicators of different levels from systems of the university information infrastructure.

Solving the problem of extracting data from information systems requires an in-depth analysis of the PU information infrastructure. The appropriate data sources are projected to be defined by this analysis, what quantity of the stored data and how they can be extracted and analysed for GE monitoring and forming values of the indicators from the model proposed. PU computerizes its learning and administrative processes through software systems that collect data on students and PhD students, faculty and administrative staff and has the data sets which are necessary to benefit from data analytic tools. The analysis, the student information system, the student admission system, the human resource system, and the research reporting system consider powerful data sources of the GEAnalyst. Therefore, all data sources are united through JasperSoft ETL.

The next step in the development process is related to templates of GE monitoring reports for each stakeholder group's necessities which collect appropriate data for the proposed indicators, designed and developed through JasperSoft Studio (see Table 1). The JasperReport Server stores the templates (mentioned above). Besides, all that has an intermediate role between the three architectural layers. The Client Application requests the REST services of the JasperReports Server to run a chosen template and generate a report through the Service Client. The Web Service interface responds to HTTP requests from the client application.

The Data Layer of the GEAnalyst incorporates all the data sources selected from the university information infrastructure (student information system, students' admission system, the research reporting system and human resource system) and JasperReports Server repository. JasperReport Server addresses data sources to retrieve the data needed when creating reports.

GEAnalyst completes the templates with actual data (obtained from information systems or through calculations) and generates reports depending on the user's role (dean, vice-dean, rector, vice-rector, member of GEG). Since indicators from Level 1 and Level 2 are the same for different stakeholder groups, they differentiate in the minor levels, and that is set in the designed templates. A common example is that a dean of a faculty can create reports that provide data only for the faculty s/he heads, and a member of top management can create reports with data for the whole university.

The tool can generate reports automatically based on a fixed schedule or manually by the user when they want to check the present situation in the faculty/university. The generated reports are stored in the repository and can be retrieved by users with access rights.

Tables and diagrams present the results of data processing and allow users to:

- Monitor the ratio of female and male PhD students (candidate students, enrolled PhD students, dropped out PhD students);
- Monitor the ratio of female and male students (candidate students, enrolled students, graduates, dropped out students),
- Monitor the equal access of students, PhD students and employees to education and competitions for academic and administrative positions;

- Monitor the ratio of female and male members of the faculty staff;
- Monitor the ratio of female and male project managers;
- Monitor the ratio of female and male participants in projects (students, PhD students and members of the faculty staff);
- Monitor career growth in terms of gender equality (academic positions and degrees);
- Identify the most desired/undesirable study programmes by female and male future students;
- Identify study programmes with the highest/lowest percentage of female and male graduates;
- Establish a baseline situation in relation to gender equality in PU, against which progress can be monitored and evaluated on an annual basis;
- Carry out an analysis to identify areas of strength and weakness, which will allow better targeting of actions and priorities within the GEP;
- Compare monitoring results from different time periods.

A part of generated report is presented in Figure 2. through the developed tool for *Indicator 5.1. Number of both genders (women and men) in an academic position per faculty* by the university's Vice-rector. The report shows the numbers of both genders (women and men) in academic positions in each faculty. It allows them to monitor the numerical data on the ratio of women and men among academic staff.

Once the data analytics tool creates the reports, the users can analyse them in order to understand key gaps between women and men within the university and its activities. This analysis will help users make data-informed decisions to provide similar career development in the university and access to education. It could also increase the gender balance in the university environment, guide key priorities, and adjust them as the situation evolves. The generated report data (see Figure 2) show that the majority of faculty members are female (e.g. Faculty of Pedagogy) and others are male (e.g. Faculty of Mathematics and Informatics). The last shows that the top management of the faculties where the gender balance is distorted can take measures to stimulate the unrepresented sex and achieve better gender equality (in implementation of Measure 1.7 of PU GEP [21]).

At the end of each year, annual monitoring reports are created for each indicator with summary data for each faculty and the university. These reports can be viewed and downloaded by users with rights through GEAnalyst. Members of GEG can analyse them to recognize whether the activities are impacting and where obstacles persist and compare the results with those from the previous year to show progress or lack thereof. Therefore, the current and last year's results can be presented and the findings to the top and middle management, responsible bodies for gender equality, academic staff and students (where relevant) and other key stakeholders.

PLOVDIV UNIVERSITY PAISII HILENDARSKI

Gender Equality: Data analytic tool

VISUAL EDITOR

Repository | Settings | Login: Vice-rector

STORAGE FOR PU "PAISII HILENDARSKI"

Monitoring and analysis of numerical data on the ratio of women and men in academic positions (by units and positions held)

Indicator: Numbers of women and men in academic position

Faculty	Other		Assistant		Assist. Prof.		Assoc. Prof.		Professor		Total	
	M	F	M	F	M	F	M	F	M	F	M	F
Faculty of Biology	0	0	0	0	14	17	9	13	5	7	28	37
Faculty of Pedagogy	4	5	2	15	6	23	8	14	7	10	27	67
Faculty of Economics and Social Sciences	1	0	5	2	8	6	8	14	8	4	30	26
Faculty of Mathematics and Informatics	1	0	10	7	17	14	12	15	12	4	52	40
Physico-technological faculty	0	1	9	2	7	7	10	12	2	4	28	26
Faculty of Philology	1	5	7	19	6	24	9	26	3	9	26	83
Faculty of History and Philosophy	0	0	3	2	4	10	14	17	6	3	27	32
Faculty of Chemistry	0	0	0	4	5	12	7	11	3	1	15	28
Law faculty	0	0	1	1	5	6	12	10	8	8	26	25

Fig. 2 Generated report

4. Conclusion

The developed application is the first of its kind that allows the monitoring of gender equality and career development in HEIs. GEAnalyst retrieves and summarizes data that enable gender equality monitoring without human resources participation and manually perusing endless data streams. The tool helps HEIs managers to make data-informed decisions to ensure equal access to education and career growth for women and men, increase the sense of equality in the university, set priorities and adjust these priorities when the situation evolves.

GEAnalyst provides real-time testing at PU. Various stakeholder groups will use the tool to generate reports needed for the status-quo assessment of the university and the monitoring procedure for PU GEP [21].

Based on the feedback received from the ongoing evaluation of the tool by various stakeholder representatives, the final version of the device will be developed, which will ensure requirements for users' privacy and data protection [22].

In future, the tool functionality will be expanded to allow data extraction for other quantitative indicators (e.g. number of women and men in decision-making positions, number of publications of women and men, average numbers of years needed for women and men to make career growth, numbers of women and men applying for/taking parental leave etc.) as well as automated analysis of the results of surveys conducted among academic staff and students.

Regardless of the software systems, each HEI can adapt the developed data analytics tool for its needs. In order to achieve this, HEI needs to identify data analytics purposes and map its context.

Funding Statement

The paper is partly supported by the projects "Supporting and Implementing Plans for Gender Equality in Academia and Research" (Grant Agreement No 824544) and "Application of big data analysis methods in higher education". This paper reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

References

[1] European Commission, "*Horizon Europe Guidance on Gender Equality Plans*," Publications Office of the European Union, Luxembourg, 2021. *Crossref*, http://doi.org/10.2777/876509

[2] C. Lopes, and S. Bailur, "*Gender Equality and Big Data: Making Gender Data Visible*," 2019.

[3] Beatriz Sanz Sáiz, "Five Ways Data Analytics Can Help Close the Gender Gap," 2019. [Online]. Available: https://www.ey.com/en_bg/digital/five-ways-data-analytics-can-help-close-the-gender-gap

[4] Kayla Matthews, "*How HR is Using Data Science and Analytics to Close the Gender Gap*," 2020. [Online]. Available: https://www.kdnuggets.com/2020/01/hr-data-science-analytics-gender-gap.html

[5] Vandana Joshi, "Gender Discrimination at Work Place: A Significant Barrier for Women Empowerment," *SSRG International Journal of Humanities and Social Science*, vol. 6, no. 1, pp. 12-15, 2019. *Crossref*, https://doi.org/10.14445/23942703/IJHSS-V6I1P103

[6] G. Muneeswari et al., "Urban Computing: Recent Developments and Analytics Techniques in Big Data," *International Journal of Engineering Trends and Technology*, vol. 70, no. 7, pp. 158-168, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I7P217

[7] Surabhi Singh, Lata Bajpai Singh, and Anjali Rai, "Investigating the Analytics for Workforce Automation," *SSRG International Journal of Economics and Management Studies*, vol. 9, no. 5, pp. 23-31, 2022. *Crossref*, https://doi.org/10.14445/23939125/IJEMS-V9I5P103

[8] Bart Rienties et al., "A Review of Ten Years of Implementation and Research in Aligning Learning Design with Learning Analytics at the Open University UK," *Interaction Design and Architecture(s)*, vol. 33, pp. 134–154, 2017.

[9] Sadiq Hussain et al., "Prediction Model on Student Performance based on Internal Assessment using Deep Learning," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 8, pp. 4-22, 2019. *Crossref*, https://doi.org/10.3991/ijet.v14i08.10001

[10] Nurbiha A Shukor, and Zaleha Abdullah, "Using Learning Analytics to Improve MOOC Instructional Design," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 24, pp. 6-17, 2019. *Crossref*, https://doi.org/10.3991/ijet.v14i24.12185

[11] Silvia Gaftandzhieva et al., "Mobile Learning Analytics Application: Using Students' Big Data to Improve Student Success," *International Journal on Information Technologies & Security*, vol.10, no. 3, pp. 53-64, 2018.

[12] Billy Tak-ming Wong, and Kam Cheong Li, "A Review of Learning Analytics Intervention in Higher Education (2011–2018)," *Journal of Computers in Education*, vol. 7, pp. 7–28, 2020. *Crossref*, https://doi.org/10.1007/s40692-019-00143-7

[13] Miloslava Cerna, "Modified Recommender System Model for the Utilized e-learning Platform," *Journal of Computers in Education*, vol. 7, no. 1, pp. 105–129, 2020. *Crossref*, https://doi.org/10.1007/s40692-019-00133-9

[14] S. Gaftandzhieva, and R. Doneva, "Data Analytics to Improve and Optimize University Processes," *ICERI21 Proceedings*, pp. 6236-6245, 2021. *Crossref*, http://doi: 10.21125/iceri.2021.1404.

[15] Sadiq Hussain et al., "Regression Analysis of Student Academic Performance Using Deep Learning," *Education and Information Technologies*, vol. 26, no. 1, pp. 783–798, 2021. *Crossref*, https://doi.org/10.1007/s10639-020-10241-0

[16] Silvia Gaftandzhieva, Rositsa Doneva, and Milen Bliznakov, "Internal and External QA in HE: LA Tools and Self-Evaluation Report Preparation," *International Journal of Emerging Technologies in Learning*, vol. 15, no. 16, pp. 191-199, 2020. *Crossref*, https://doi.org/10.3991/ijet.v15i16.14401

[17] Rositsa Doneva et al., "Learning Analytics Software Tool Supporting Decision Making in Higher Education," *International Journal on Information Technologies and Security*, vol. 12, no. 2, pp. 37-46, 2020.

[18] TARGET, "*Taking a Reflexive Approach to Gender Equality for Institutional Transformation*," 2018.

[19] EIGE, "*Gear Tool*," 2016. [Online]. Available: https://eige.europa.eu

[20] Venkateswarlu Pynam, R Roje Spanadna, and Kolli Srikanth, "An Extensive Study of Data Analysis Tools (Rapid Miner, Weka, R Tool, Knime, Orange)," *SSRG International Journal of Computer Science and Engineering*, vol. 5, no. 9, pp. 4-11, 2018. *Crossref*, https://doi.org/10.14445/23488387/IJCSE-V5I9P102

[21] R. Doneva, S. Gaftandzhieva, and K. Boykova, "An Approach to Monitoring Gender Equality Plans Implementation," *EDULEARN22 Proceedings*, pp. 3383-3391, 2022. *Crossref*, https://doi.org/10.21125/edulearn.2022.0829

[22] Radi Petrov Romansky, "A Survey of Digital World Opportunities and Challenges for User's Privacy," *International Journal on Information Technologies and Security*, vol. 9, no. 4, pp. 97-112, 2017.

[23] L. Madamshetty, K. Suresh, and B. Naidu, "*Integrating Big Data in Higher Education: Perspectives and Challenges*," 2020. [Online]. Available: https://www.¬ee.co.za/-article/integrating-big-data-in-higher-education-perspectives-and-challenges.html

[24] Plovdiv University, Paisius of Hilendar, "*Gender Equality Plan (2021-2024)*," 2021. [Online] Available: https://gender-spear.eu/assets/content/¬PU_GEP_EN¬finalS.PDF

[25] Ben Daniel, "Big Data and Analytics in Higher Education: Opportunities and Challenges," *British Journal of Educational Technology*, vol. 46, no. 5, pp. 904-920, 2019. *Crossref*, https://doi.org/10.1111/bjet.12230