



INSPIRE

D3.2 GEP Prevalence Monitoring Indicator Framework v2

Andrea Löther (GESIS)
Mazlum Karataş (GESIS)
Lena Weber (GESIS)

31 March 2025



**Funded by
the European Union**

Project no.	101058537
Project acronym:	INSPIRE
Project title:	INSPIRE: Centre of Excellence on Inclusive Gender Equality in Research & Innovation: Creating Knowledge & Engaging in Collaborative Action
Call:	HORIZON-WIDERA-2021-ERA-01
Start date of project:	01.10.2022
Duration:	48 months
Deliverable title:	D3.2 GEP Prevalence Monitoring Indicator framework v2
Due date of deliverable:	31.03.2025
Actual date of submission:	31.03.2025
Deliverable Lead Partner:	GESIS
Dissemination level:	Public

Authors & contributors name and organisation	Role
Andrea Löther, GESIS Mazlum Karataş, GESIS Lena Weber, GESIS	Lead author(s)
Hara Stefanou, Innosystems Vaios Kouvelis, Innosystems	Contributor(s)
Marta Warat, UJ	Quality assurance editor



The work contained in this document is subjected to a Creative Commons license (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Disclaimer

While INSPIRE is funded by the European Union, views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the European Research Executive Agency (REA) can be held responsible for them.

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101058537.

Consortium

FUOC	Fundació per a la Universitat Oberta de Catalunya, Spain
JR	Joanneum Research Forschungsgesellschaft MbH, Austria
SDU	Syddansk Universitet, Denmark
UJ	Uniwersytet Jagiellonski, Poland
Notus	Notus, Spain
FLACSO	Facultad Latinoamericana De Ciencias Sociales, Argentina
EM	Europa Media Szolgaltato Non Profitkozhasznu Kft, Hungary
Portia	Portia GmbH, Germany
SRU	Stichting Radboud Universiteit
Fraunhofer	Fraunhofer Gesellschaft zur Forderung der Angewandten Forschung EV, Germany
UH	Universiteit Hasselt, Belgium
ZRC SAZU	Znanstvenoraziskovalni Center Slovenske Akademije Znanosti In Umetnosti, Slovenia
GESIS	GESIS-Leibniz-Institut für Sozialwissenschaften EV, Germany
INNO	Innosystems Symvouleutikes Ypiresies Kai Efarmoges Pliroforikis Ypsilis Technologias Monoprosopi Idiotiki Kefalaiouschiki Etaireia, Greece

Version	Date	Summary of changes	Revised by
01	27.02.2025	First internal revision	Lena Weber
02	11.03.2025	Second internal revision	Andrea Löther
03	11.03.2025	Changes and suggestions are implemented	Mazlum Karataş
04	11.03.2025	Submitted to quality assurance editor	Marta Warat
05	25.03.2025	Suggestions from quality assurance editor are implemented	Andrea Löther
06	26.03.2025	Suggestions from quality assurance editor are implemented	Mazlum Karataş
07	26.03.2025	Third internal revision	Lena Weber
08	27.03.2025	Submitted to the project coordinator	Mazlum Karataş

Executive Summary

A reliable method for systematically monitoring Gender Equality Plans (GEPs) at a supra-institutional level is unavailable. This is the reason why this report summarises the development of a methodology to monitor GEP across Europe in two different ways to figure out which could be the most appropriate in the future or how to combine both methods. The first methodology is *non-reactive* (web scraping and automated text analysis), and the second methodology is a European-wide online *survey* (reactive method). At the end of the report, we compare both methodologies in terms of their advantages and disadvantages for the European-wide monitoring of GEP.

In order to develop the methodologies, a pilot study was initially conducted to test the methods on a smaller scale, encompassing 83 organisations from Germany, Greece, Ireland, and Estonia. Following that, we conducted a full sample survey of approximately 7,000 research-performing (RPOs) and research-funding organisations (RFOs) in all European member states and six associated countries. After removing incomplete data, the survey was sent to 4,571 organizations, and web scraping was conducted for approximately 2000 more cases. The field period of the Europe-wide full sample online survey was in the summer of 2024.

Firstly, the report outlines the theoretical basis of the INSPIRE indicators and their development process for monitoring purposes. The INSPIRE indicators were created based on the T.2.1 Data Monitoring Report, and feedback was received from four different focus groups, which were conducted with a total of 28 participants from all over Europe. The INSPIRE indicators encompass the four areas regarding the prevalence, characteristics, implementation, and impact. While developing INSPIRE indicators, we also consider intersectional and inclusive perspectives.

Second, the nonreactive methods and the survey methodology for data collection will be explained. On the one hand, we have combined web scraping with automated text analysis as a non-reactive method. SerpAPI, a web scraping tool, was utilized to collect data and develop a particular INSPIRE scraper based on Google's crawling database. INSPIRE scraper detected that 41% of 6,571 organisations have a GEP. Furthermore, the INSPIRE Web scraper detects more than the targeted GEPs and also downloads unspecified PDFs, so the total number of PDFs downloaded is 7,018. Therefore, an intermediary classification process is required to clean up the data corpus. We filtered the PDFs using metadata analysis, resulting in a total of 1,564 PDFs. We then removed corrupt GEPs and non-English GEPs so that 1,135 GEPs were available for text analysis. On the other hand, a more conventional online survey was sent to RPOs and RFOs via the online survey platform UNIPARK for data collection. We received 283 responses to the online survey, for a response (completion rate) of 6%. The results of the online survey suffer from nonresponse bias and show 95% of the GEP prevalence rate.

We argue that web scraping and online surveys are complementary rather than rivals in the case of collecting data regarding GEPs. Both approaches demonstrate advantages and disadvantages over one another. Online surveys offer the advantage of obtaining high-quality targeted data in real-time, enabling many observations and facilitating efficient data collection.

However, online surveys face a low response rate and difficulties in acquiring participants' email addresses. In contrast, web scraping offers excellent potential for collecting massive amounts of data without needing any contact details. However, collecting data via web scraping faces challenges, such as selecting appropriate tools for scraping and developing the algorithm to collect the data. Web scraping offers extensive data collection capabilities, but sorting and selecting the data, inconsistencies in metadata, the need for high computational capacity and the absence of standardised terminology for file descriptions pose considerable challenges.

Results from **D.3.3. Report on GEP Monitoring Survey + Webcrawl Result** indicated that using web scraping data serves as a corrective measure for survey data. Conversely, survey data can inform web scraping about gaps that lead to false negatives. This can include the use of diversity and equity concepts instead of gender equality. Consequently, reactive and non-reactive methods mutually reinforce and reveal the biases that lead to overestimation or underestimation. Only by combining survey data, web scraping, and automated text analysis can a complete dataset and a comprehensive picture of the prevalence, characteristics, implementation, and impact of GEPs be obtained.

Table of Contents

List of Acronyms.....	9
List of Figures.....	11
List of Tables.....	11
1 Introduction.....	13
2 Research Design.....	14
2.1 Objectives, Sampling and Implementation of the Pilot and Full Sample Study	14
2.2 Sampling (Full Study).....	16
2.2.1 Sample of Research-Performing Organisations	16
2.2.2 Sample of Research Funding Organisations	18
2.2.3 Sampling Realisation for the Full Sample	19
3 Indicator Development	21
3.1 Political Relevance and Theoretical Framework.....	21
3.2 Methodological Approach of Indicator Development: Process and Quality Check ..	23
3.3 Developed Indicators	24
3.3.1 Prevalence.....	24
3.3.2 Characteristics	26
3.3.3 Implementation	29
3.3.4 Self-Assessment and Impact	37
3.4 Adapting the Indicators for Reactive and Non-Reactive Methods	41
4 Non-Reactive Methods for the Full Sample.....	46
4.1 Web Crawling versus Web Scraping	47
4.1.1 Web Crawling	47
4.1.2 Web Scraping	48
4.2 Detecting The GEPs: The INSPIRE's Web Scraper for the Full Sample	49
4.2.1 Tool Selection For Web-Scraping.....	51
4.2.2 Defining Search Terms	58
4.2.3 Methodology to Reveal the Indicator "Prevalence".....	60
4.2.4 Methodology to Download the GEPs.....	65
4.3 Monitoring Characteristics, Implementation and Impact of GEPs: Classifying PDFs and Text Analysis	69
4.3.1 Text Analysis as a Methodological Approach	69
4.3.2 Classification of PDFs.....	70

4.3.3	Developing the Algorithm for the Text Analysis	78
4.4	Limitations of Web Scraping	88
5	Reactive Method: Full Sample Survey	89
5.1	Questionnaire Design	89
5.2	Translation of the Questionnaire	90
5.3	Email Extraction for the INSPIRE Survey	91
5.4	Implementation of the Survey	93
5.5	Data Cleaning and Appending: Pilot and Full Survey	95
5.6	Response Rate	97
6	Results on Prevalence of GEPs: Reactive vs. Non-Reactive	99
7	Comparison of the Two Approaches: Reactive vs. Non-Reactive Methods	101
8	Input from the Co-Creation Workshop Data Monitoring	104
8.1.1	The Purpose of and Data Access for Monitoring	104
8.1.2	Prevalence of GEP in Private Companies	105
8.1.3	Intersectionality Approach of GEP	105
8.1.4	Improvements for the CORDIS Database for Better Data Quality in the Sampling 105	
9	References	107
10	Appendix	115
10.1	List of Indicators	115
10.2	Web Scraping: Search Terms in the Full Study	119
10.3	Search Terms for Meta-Analysis in the Full Study	126
10.4	Text Analysis: Search Terms	131
10.5	Survey: Questionnaire (English)	134
10.6	Survey: Invitation and Reminder to the Online Survey	153
10.6.1	Invitation	153
10.6.2	The First Reminder	154
10.6.3	The Second Reminder	155
10.7	Country Clusters	156
10.8	Technical Implementation of INSPIRE Scraper	157

List of Acronyms

AC	Associated Countries
API	Application Programming Interface
BART	Bidirectional and Auto-Regressive Transformers
BERT	Bidirectional Encoder Representations from Transformers
BFS	Breadth First Search
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans Apart
CoP	Community of Practice
CORDIS	Community Research and Development Information Service
CSS	Cascading Style Sheets
CSV	Comma Separated Values
CTAM	Computational Text Analysis Methods
D	Deliverable
DBS	Depth First Search
ETER	European Tertiary Education Register
GEP	Gender Equality Plan
GPT	Generative Pre-trained Transformer
GPU	Graphical Processing Unit
HEI	Higher Education Institutions
HTML	Hypertext Markup Language

HTTP	Hypertext Transfer Protocol
IP	Internet Protocol
IR	Information Retrieval
JSON	JavaScript Object Notation
KSH	Knowledge Support Hub
LLM	Large Language Model
MS	Member State
OCR	Optical Character Recognition
PDF	Portable Document Format
PRC	Private Companies which perform research
QA	Question Answering
REC	Public Financed Research Institutions
RFO	Research-Funding Organisation
RPO	Research Performing Organisation
ROR	Research Organisation Registry
URL	Uniform Resource Locator
WP	Work Package
XLSX	Microsoft Excel Spreadsheet

List of Figures

Figure 1	Research design	14
Figure 2	Organisation type by country for the pilot sample	15
Figure 3	INSPIRE web scraping and its application to domains	50
Figure 4	Web scraping process.....	51
Figure 5	GEP prevalence from the incremental steps of the INSPIRE methodology (October – November 2023).....	62
Figure 6	Text analysis process.....	78
Figure 7	Response rate by country	97
Figure 8	Results of GEP prevalence by organization type based on web scraping	100
Figure 9	Results of GEP prevalence by organization type based on online survey	101

List of Tables

Table 1	Implementation of the pilot and full sample study	16
Table 2	Sample and selection process by type of organisation	17
Table 3	Sample by country and type of organisation for the full sample study	18
Table 4	The result of the manual assessment of “not working” URLs.....	19
Table 5	Missing URLs by activity type.....	20
Table 6	Publications: Indicators on the prevalence of GEPs	25
Table 7	Publications: Indicators of leadership commitment and ownership.....	26
Table 8	Publications: Indicators of measures in different areas.....	28
Table 9	Publications: Indicators for monitoring the realisation of GEPs.....	30
Table 10	Publications: Indicators of financial resources	31
Table 11	Publications: Indicators of dedicated staff	32
Table 12	Publication: Indicators of organisational change	33
Table 13	Publications: Indicators of stakeholder involvement.....	34
Table 14	Publication: Indicators for the policy of monitoring and/or controlling gender equality measures and GEPs.....	35
Table 15	Publication: Indicator about data collection and publication.....	36
Table 16	Publications: Indicators on gender diversity and intersectional perspectives in monitoring and data collection.....	37
Table 17	Publications: Indicators on the understanding of impact.....	38
Table 18	Publications: Indicators of data and methods to assess the impact.....	38
Table 19	Publications: Indicators on the representation of women.....	39



Table 20	Publications: Indicators on impact self-assessment.....	41
Table 21	Operationalising the indicators for reactive and non-reactive methods.....	42
Table 22	Comparison table of OpenSearchServer, Scrapy, and SerpApi	56
Table 23	GEP prevalence with OpenServer vs SerpApi (October 2023).....	57
Table 24	A comparison: web scraper vs manual detection.....	63
Table 25	Performance metrics of INSPIRE's scraper for prevalence of GEPs	64
Table 26	PDFs extracted by each step of the 4-step approach.	66
Table 27	PDF extraction results (January 2024).....	68
Table 28	Comparison of LLMs	72
Table 29	Evaluation metrics of the BART and the GPT 3.5	75
Table 30	Performance metrics of the English-only approach.....	77
Table 31	Performance metrics of English plus local languages approach.....	77
Table 32	Information extraction for prevalence indicator: Horizon Europe	82
Table 33	Information extraction for characteristics indicator: Time frame	82
Table 34	Information extraction for characteristics indicators: Areas of Activity	83
Table 35	Information extraction for characteristics indicators: inclusive GEPs	84
Table 36	Information extraction for characteristics indicator: different inequalities	85
Table 37	Information extraction for implementation indicator: dedicated staff	86
Table 38	Information extraction for implementation indicators: Quality Assurance.....	86
Table 39	Information extraction for impact indicator: description of the impact	87
Table 40	Changed items in the full survey compared to the pilot survey	90
Table 41	Translation process	90
Table 42	The timeline of the pilot and full sample surveys.....	94
Table 43	Status statistics of participants	95
Table 44	Response rate by country cluster	98
Table 45	Response rate by availability of the questionnaire in the national language	98
Table 46	Response rate by type or organisation.....	99
Table 47	Search terms used in web-scraping.	119



1 Introduction

The European Commission calls for cultural and institutional change to foster gender equality in science and research. For this purpose, it is encouraged that research-performing (RPOs) and funding organisations (RFOs) implement gender equality plans (Council of the European Union 2015; European Research Area and Innovation Committee 2015; Cheveigné et al. 2017; European Commission 2020). In the last years, the European Commission has expanded the concept of gender equality "by opening policy to intersections with other social categories" (European Commission 2020, 16) and calls for approaches to inclusive gender equality and inclusive gender action plans (European Commission 2022). Recently, having a gender equality plan (GEP) became an eligibility criterion for getting funding in Horizon Europe (European Commission 2021a).

Considering the relevance of GEPs, monitoring their prevalence among research-performing and funding organisations, and assessing their implementation and impact are crucial in assessing changes towards gender equality in European research. The ongoing review and monitoring of GEPs enables adjustments and improvements (European Commission 2021a). An efficient monitoring approach includes – but is not limited to – the accessibility and availability of data and the incorporation of appropriate indicators (Löther, Karataş, and Weber 2023).

Thus, INSPIRE's research programme seeks to develop a robust and efficient methodology to monitor inclusive GEPs across Europe. This includes developing relevant indicators for monitoring inclusive GEP, establishing monitoring instruments through reactive and non-reactive methods and comparing the advantages and challenges of both approaches.

This paper presents the indicators on the one hand and the methodologies and methodological results of the full sample study on the other hand. This paper also reports the changes from the pilot study to the full sample study. The first version of this study, **D.3.1 Report on Prevalence Monitoring Indicator Framework**, which outlines the methodology and results of the pilot study, was published in May 2024 (see Löther, Karataş, and Weber 2024a). The results of the European-wide online survey and web scraping are published in the **D.3.3 Report on GEP Monitoring Survey + Webcrawl Results** in November 2024 (see Löther, Karataş, and Weber 2024b). This paper presents the full sample study consisting firstly of an online survey for the reactive methods and secondly of a combination of web-scraping and text analysis for the non-reactive methods. In addition, this paper incorporates the inputs from the **Co-Creation Workshop Data Monitoring** that took place between 20 and 21 November 2024 in Cologne, Germany.

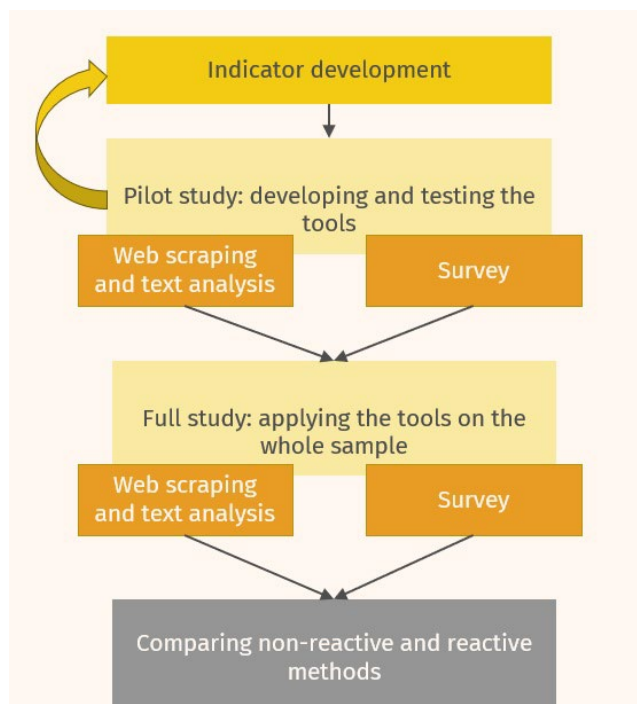
After presenting the research design (2), we explain the indicator development and the indicators (3). Afterwards, we display how we implemented the indicators for non-reactive (4) and reactive methods (5) and present in detail the developed instruments. Finally, we discuss the methodological findings by comparing the two approaches and propose recommendations for the future.



2 Research Design

Figure 1 presents the study's research design. First, we conducted a pilot study for both methodologies. Second, we performed quality assessments and made necessary changes in the methodology. Finally, we conducted the full sample study, which included a Europe-wide online survey and web scraping. The pilot study aimed to check and design the web scraping tools and the questionnaire and to test the field access. The sample for the pilot study – research-performing organisations in Germany, Greece, Ireland, and Estonia – was the same for the web scraping and survey. The sample for the full sample study consists of 27 EU AC and 6 MC (Bosnia-Herzegovina, Israel, Norway, Serbia, Switzerland, and the UK).

Figure 1 Research design



The present paper covers the indicator development, the pilot study and the full sample study. This paper also outlines the methodological refinements and modifications from the pilot study to the full sample study. Deliverable 3.1 (see Löther, Karataş, and Weber 2024a) provided a methodological comparison of the online survey and web scraping in the pilot study, while Deliverable 3.3 (see Löther, Karataş, and Weber 2024b) presented the results of the online survey and web scraping for the full sample study.

2.1 Objectives, Sampling and Implementation of the Pilot and Full Sample Study

The pilot study aimed to develop and test the tools for web scraping, text analysis, and the survey questionnaire. Applying the web scraping and the text analysis on a small sample and comparing the results with a manually controlled corpus allows us to calculate quality indicators. Furthermore, the pilot sample serves as a preliminary test to assess the questionnaire for the full sample.



We constructed the sample for the pilot studies stepwise. In the first step, we selected four countries: Germany, Greece, Estonia, and Ireland. One selection criterion was differences and peculiarities in the linguistic context and grammatical rules. We aimed to assess linguistic challenges by choosing different countries with diverse language backgrounds. For this reason, we chose Germany and Ireland because we are familiar with these languages, whereas we do not know the language of the two other countries. This enabled us to discover when we need cultural expertise to construct instruments. Another criterion was the expected prevalence level of GEPs, ranging from high rates in Germany to lower rates in Estonia.

In the second step, we selected higher education institutions (HEIs) and research-performing organisations (RPOs) from the mentioned countries. We selected institutions by conducting web searches and using the Community Research and Development Information Service (CORDIS)¹ database containing EU research initiatives under the Horizon 2020 (2014–2020) programme. During the selection process, we prioritised institutions with GEPs and those with URLs in the CORDIS database. We also aimed to have an even distribution in the sample concerning the type of organisation in each country. A piloting dataset with web domains has been crafted, featuring four countries (Greece, Ireland, Germany, and Estonia) and approximately 20 organisations per country, spanning academia, research institutions, and private companies. Figure 2 presents the pilot sample of the 83 organisations categorised by country and type of organisation.

Figure 2 Organisation type by country for the pilot sample

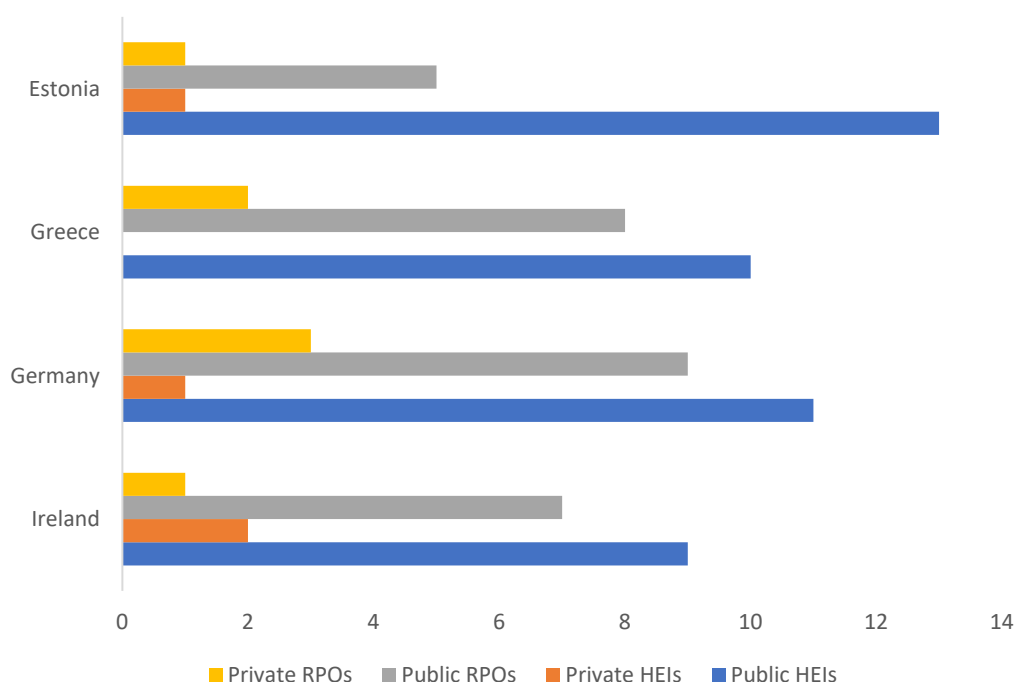


Table 1 gives an overview of the pilot study's implementation.

¹ [CORDIS - EU research projects under Horizon 2020 \(2014-2020\) - Data Europa EU.](#)



Table 1 Implementation of the pilot and full sample study

Tasks	Period
Implementation of web scraping for the full sample	07/2024 – 10/2024
Implementation of European-wide online survey	07/2024 – 08/2024
Developing the web scraper	08/2023 – 01/2024
Developing tools to download and classify GEPs as pdf-file	01/2024 – 08/2024
Developing the algorithm/tool for the text analysis	12/2023 – 03/2024
Developing the survey questionnaire	10/2023 – 02/2024
Pretest of the survey	02/2024 – 03/2024

2.2 Sampling (Full Study)

To apply and test the indicators for both methodological approaches, we use the same sample of research-performing and research-funding organisations in the European Union member states and six states associated with the European Research Area. Initially, these states comprised Israel, Montenegro, North Macedonia, Norway, Switzerland and the UK. The consortium meeting decided to integrate Bosnia-Herzegovina and Serbia instead of Montenegro and North Macedonia because of INSPIRE case studies (WP 3, Task 3.3.) in these countries.

2.2.1 Sample of Research-Performing Organisations

After checking some other sources for research-performing organisations – especially the European Tertiary Education Register (ETER)² – we decided to use the CORDIS database on projects funded in Horizon 2020³ for the following reasons:

- Includes higher education institutions, publicly financed research institutions and private companies, thus a unique source for all organisation types,
- High number of organisations (41,406 for all countries and 37,866 for the countries of the sample)
- Closed program (no further changes in the sample)
- A GEP was not an eligible criterion in Horizon 2020. Thus, there are more differences between having and not having a GEP.

After reducing the project list to a list of organisations, we cleaned the list by:

- Eliminating duplicates

² <https://www.eter-project.com/>

³ CORDIS - EU research projects under Horizon 2020 (2014-2020), URL: <http://data.europa.eu/88u/dataset/cordisH2020projects>, DOI: [10.2906/112117098108/12](https://doi.org/10.2906/112117098108/12). Date of data extraction: 6.7.2023.



- Eliminating departments of universities

Furthermore, we reduced the list to higher education institutions, research organisations and private companies, eliminating organisations of the categories "public bodies" and "other". After cleaning, the list includes 30,545 research-performing organisations from the selected countries (see Table 2).

Table 2 Sample and selection process by type of organisation

	Total number of organisations in CORDIS		Organisations in CORDIS with URL		Sample	
	N	%	N	% of total	N	%
Higher education	1,486	4.9%	1,023	68.8%	1,486	21.5%
Private company	26,268	86.0%	4,245	16.2%	2,627	38.0%
Research organisation	2,791	9.1%	1,462	52.4%	2,791	40.4%
Sum	30,545	100%	6,730	22,0%	6,904	100%

The numbers include only INSPIRE countries.

The web scraping and the survey (search for email addresses) need the website addresses. The CORDIS database includes a column "organisationURL", but the database provides the URL of only 22% of the organisations (see Table 2). In particular, many entries from private companies lack a website address.

Due to financial restrictions (extracting the email addresses), including all organisations (N=30,545) in the sample was impossible. Furthermore, the organisations are unevenly distributed among the organisation types, with a high proportion of private companies. We decided to reduce the sample size but keep all higher education institutions and all research organisations in the sample. A proportional reduction of all organisation types would result in a deficient number of these organisation types. Assuming a response rate of 20-30% would make a valuable analysis impossible. Furthermore, we expect a lower rate of gender equality plans among private companies and, thus, less material for the text analysis.

We decided to include 10% of the private companies in the sample according to the following selection criteria:

- The CORDIS list displays the organisation's website address for practical reasons and to reduce the number of URLs searched.
- The total sum of all Horizon 2020 projects of an individual organisation as a proxy for the size, significance and level of the company's research investments.

The sample consists of 6,904 research-performing organisations. Table 2 shows the distribution among the organisation types. Table 3 displays the sample by country and organisation type.



Table 3 Sample by country and type of organisation for the full sample study

	Higher education	Private company	Research organisation	Total	% of sample
AT	42	82	123	247	3,6%
BA	9	0	10	19	0,3%
BE	28	115	133	276	4,0%
BG	35	5	62	102	1,5%
CH	26	84	58	168	2,4%
CY	10	16	16	42	0,6%
CZ	24	25	82	131	1,9%
DE	212	433	358	1.003	14,5%
DK	15	54	30	99	1,4%
EE	7	8	9	24	0,3%
EL	36	83	63	182	2,6%
ES	88	282	355	725	10,5%
FI	37	64	30	131	1,9%
FR	205	312	195	712	10,3%
HR	16	5	42	63	0,9%
HU	32	30	46	108	1,6%
IE	24	41	24	89	1,3%
IL	24	25	16	65	0,9%
IT	110	269	299	678	9,8%
LT	10	7	19	36	0,5%
LU	1	6	8	15	0,2%
LV	16	3	17	36	0,5%
MT	3	0	5	8	0,1%
NL	49	206	113	368	5,3%
NO	24	41	66	131	1,9%
PL	88	23	115	226	3,3%
PT	44	57	88	189	2,7%
RO	42	17	81	140	2,0%
RS	9	4	31	44	0,6%
SE	37	85	52	174	2,5%
SI	13	26	51	90	1,3%
SK	22	7	31	60	0,9%
UK	148	212	163	523	7,6%
Sum	1,486	2,627	2,791	6,904	100%

2.2.2 Sample of Research Funding Organisations

After checking different sources to create the list of research funding organisations (f.eg. GENDER-NET plus, <https://gender-net-plus.eu/>; EU-Project: Grant allocation disparities (GRANteD), <https://www.granted-project.eu/>; Global Research Council,



<https://globalresearchcouncil.org/>), we decided to use Science Europe⁴ and She Figures 2021 (European Commission 2021b). Science Europe represents major public organisations that fund or perform research in Europe. 37 of the 40 members in 29 European countries are research funding organisations, two of which are both research funding and performing organisations. To complete the sample, we extracted research funding organisations from countries not represented in Science Europe through lists in She Figures (European Commission 2021b, 328–37). The final sample consists of 43 research funding organisations from all INSPIRE countries.

2.2.3 Sampling Realisation for the Full Sample

The next step has been to verify the displayed URLs and find the correct ones. This was necessary to prepare the sample for the web scraping data collection and to be able to extract the email addresses for the invitation letters to the survey. For this reason, we developed a script that uses Python's⁵ request⁶ library to check whether or not the provided URLs work. First, the script checks the status of web pages. Afterwards, the script adds two columns to the current data frame using Python's pandas⁷ library, which reflects the status of the given URL and the final URL. The script can handle basic redirections, such as server codes 301 and 302. If there is a redirection, the last URL column will include the redirected URL.

The script reports that **3992** of 6947 URLs are "working", whereas 2955 URLs are "not working". It should be noted that the "not working" status does not necessarily imply that URLs are not working. This status presents the outcome of the script, which itself possesses inherent constraints and boundaries. In particular, the script may fail to retrieve the information if the webpage contains i) JavaScript-rendered content, ii) Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA)⁸ or anti-bot measures, and iii) complex redirections or session handling. Subsequently, we manually assessed 2955 URLs with "not working" status and assigned them a correct status. The result of the manual assessment is shown in Table 4. Upon manually examining the URLs, we determined that 15% of the URLs initially classified as "not working" were functional when accessed through a web browser. Thus, we indicated redirected URLs. In addition, we manually recovered some URLs while performing the manual assessment, representing 1% of "not working" URLs.

Table 4 The result of the manual assessment of "not working" URLs

Manual Status	Frequency	Percentage	Cumulative
Working	1873	63.38	63.38
Not Working	553	18.71	82.10

⁴ <https://scienceeurope.org/about-us/members/>, Date of data extraction: 21.11.2023.

⁵ Python is considered a high-level, general-purpose programming language. Its design philosophy emphasizes code readability through extensive indentation. See <https://www.python.org/>

⁶ <https://pypi.org/project/requests/>

⁷ <https://pandas.pydata.org/>

⁸ CAPTCHA is considered a type of a security measure and stands for Completely Automated Public Turing test to tell Computers and Humans Apart.



Directed	450	15.23	97.33
No URL	48	1.62	98.95
Manually Recovered	31	1.05	100.00
Total	2955	100.00	

After manually assessing the URLs, we detected that 601 (553 + 48) URLs were missing. 48 organisations had no URL at all, while 553 of them had the status "Not working." The majority of the missing URLs belong to private companies (58%). Table 5 demonstrates missing URLs by activity type.

Table 5 Missing URLs by activity type

Activity type	Frequency	Percentage	Cumulative
PRC	348	57.90	57.90
REC	187	31.11	89.02
HEI	65	10.82	99.83
RFO	1	0.17	100.00
Total	601	100.00	

With: PRC = private companies performing research, REC = public financed research organisations, HEI = higher education institutions, RFO = research-funding organisations.

We relied on automated methods to retrieve the missing URLs, as manually finding them would have been time-consuming. We particularly utilised *The Research Organization Registry's* (ROR)⁹ Application Programming Interface (API) to obtain Uniform Resource Locators (URLs) regarding research organisations, higher education institutions, and research-performing organisations. Unfortunately, private companies are not included in the ROR database. Hence, we used the CUFinder¹⁰ API, a paid service, to extract URLs regarding private companies. Overall, 465 of the 601 missing URLs were successfully recovered; the remaining 136 are still missing and are excluded from the sample.

Nonetheless, we discovered some duplicate URLs within the countries even though we initially removed repetitions when compiling the INSPIRE sample. There are two main reasons for these duplicates. Firstly, certain organisations share a domain with their parent or partner organisation, which means they operate using the same domain. Secondly, big multinational companies have subsidiaries with different names but in the same domain in the database. We identified **240** repeated domains, also removed from the INSPIRE sample, bringing the total number of organizations down to **6,571**.

⁹ <https://ror.org/>

¹⁰ <https://cufinder.io/>



3 Indicator Development

3.1 Political Relevance and Theoretical Framework

To foster cultural and institutional change in research and innovation towards gender equality through GEPs, quality assurance of their implementation and impact is necessary. Monitoring and evaluation aim to assure the quality of the implementation and realisation of GEPs and gender equality policies. Whereas monitoring "is the ongoing process of systematically collecting data on an outcome" and "is a means for measuring progress", evaluation refers "to the systematic assessment of an initiative, its design, implementation and results" and asks for effectiveness and efficiency (Aldercotte 2018, 4). INSPIRE focuses on monitoring and seeks to develop a solid set of indicators that allow monitoring of GEPs, primarily in RPOs, but also in RFOs.

INSPIRE seeks to monitor the GEPs of many RPOs on a European-wide level (27 member states + Bosnia-Herzegovina, Norway, Israel, Switzerland, Serbia and the UK), instead of the effect of a GEP on gender inequalities in a single institution. The analysis will consider national contexts but do not seek to compare countries. Thus, the INSPIRE indicators facilitate GEP monitoring on a supra-organisational, not on an institutional level.

GEPs aim at transformational change toward gender equality in an organisation, and monitoring GEPs aims to check whether transformational change occurs. To monitor change in a structured way, the programme management cycle (Wroblewski 2023, 30; European Commission 2021a, 12; Lipinsky and Schäfer 2016) and the logic chart analysis (Palmén et al. 2019b; Wyatt Knowlton and Phillips 2009) build the framework to structure the monitoring process and the indicators. The programme management or policy cycle describes the implementation of gender equality policies, which includes needs assessment (or gender analysis), planning and decision of the actions, implementation, monitoring and evaluation and adaptations of the measures, by which the cycle starts again. Thus, monitoring is an integral part of gender equality policies. The logic chart analysis provides a framework for evaluating input, output, outcome and impact. Following this distinction, the INSPIRE indicators focus on four areas to monitor the transformational potential of GEPs:

1. Prevalence (as input): Is there a GEP (or equivalent) in the organisation?
2. Characteristics (as output): What are the characteristics of the GEP regarding stakeholders, formal features, areas of activity or inequalities mentioned?
3. Implementation (as outcome and processes): Which measures have been implemented? Are there financial and personal resources for the implementation? Which processes have been introduced?
4. Self-assessment and impact: How does the organisation perceive changes connected to measures and the GEP? How does the organisation analyse changes in gender equality?



Even if monitoring primarily measures output, outcome, and processes, we aspire to gain some glimpse of knowledge about the long-term effects of GEPs (or other equality policies). Sure, for a deeper understanding of the impact of GEPs evaluations would be needed, which regularly rely on sophisticated quantitative and qualitative data and thus are applied on a smaller scale of case numbers. Conventionally, evaluations try to attribute observed effects to the intervention by measuring the counterfactual and searching for causal links (Aldercotte 2018; Gates and Dyson 2017; Kalpazidou Schmidt et al. 2017). It has been also discussed at the co-creation workshop “data monitoring” that it is not feasible to assess impact by monitoring tools and that sophisticated evaluations are necessary for this knowledge.

Despite the differences between monitoring and evaluation, evaluation approaches are valuable in conceptualising monitoring on long-term effects. We rely on “theories of change” and the “impact driver model” because both approaches established a way to get information on impact, not by directly measuring impact but by assessing factors contributing to change. Based on *theories of change*, impact evaluation investigates how “the implementation of the interventions ‘contributed’ to the outcomes and impact of the intervention in combination with a complex array of contextual influential factors” (Palmén and Kalpazidou Schmidt 2019, 7). Using a similar approach, Mergaert, Cacace, and Linková (2022) set up indicators for impact drivers that monitor preconditions for effective change toward gender equality. Their impact driver model combines two approaches: *the institutional capacity model* and *the actor mobilisation model*.

The *institutional capacity model* focuses on an institution's potential and implementation process. Impact drivers for gender mainstreaming are:

- effective leadership
- adequate financial and human resources
- availability of appropriate procedures and processes
- appropriate organisational incentives and accountability structures (Mergaert, Cacace, and Linková 2022, 4–6).

The *actor mobilisation model*, on the other hand, focuses on agency and the activation of internal processes, namely

- transformational agent(s)
- agency dynamics like mobilising stakeholders
- structural features like internal functioning, norms and regulation
- dimensions and areas of sustainable outcomes (Mergaert, Cacace, and Linková 2022, 6–7)

In the combination of both models, the so-called *impact driver model* consists of twelve impact drivers with several indicators each and six stages of institutionalisation. The model envisages monitoring progress in the organisation by self or external assessment.

The indicators are not directly applicable to our supra-organisational GEP monitoring, but the focus on *drivers* and *enablers* for transformational change guides our development and selection of indicators. This approach aligns with the criteria for GEPs set for Horizon Europe, which also focus on several preconditions for change (publication and official endorsement,



dedicated resources, data collection and monitoring, training and recommended thematic blocks) (European Commission 2021a).

3.2 Methodological Approach of Indicator Development: Process and Quality Check

The indicators aim to monitor the prevalence, characteristics, implementation and impact of GEPs European-wide on a supra-institutional level. The indicators represent a conceptual level. They are not associated with a distinct method (survey or non-reactive methods). In another step, we will operationalise the indicators for each instrument (see section 3.4). We describe how to apply the indicators for non-reactive methods (web scraping and automated text analysis) in Chapter 4 and for reactive methods in a European-wide survey in Chapter 5, using the same sample of higher education institutions, research institutions, private research-intensive organisations, and research funding organisations for both methodological approaches (see Chapter 2 on the sample). In this way, the project INSPIRE intends to compare the advantages and disadvantages of each method for monitoring GEPs.

The quality criteria for the development and selection of indicators emerge from the following:

- The subject matter (GEPs): referring to GEPs, not gender inequalities
- The level of analysis: feasible for monitoring a large number of RPOs and RFOs on a European-wide level
- The focus on facilitating factors for impact: providing information on drivers and enablers for transformational change.

A scoping literature review that investigates the state of the art on monitoring gender equality and especially GEPs in research-performing and funding organisations guided the development of the indicators (Löther, Karataş, and Weber 2023). Building on this review, we mapped indicators used or proposed in the literature. By applying the mentioned criteria, we developed a first draft of indicators.

For a quality check and to get feedback, experts on gender equality and monitoring discussed the draft indicators in four online focus groups. Participants of the focus groups came from four geographical areas in Europe (Northern and Western Europe, Eastern Europe, German-speaking countries and Southern Europe). The selection criteria for the participants were expertise in gender equality in science and research, as a researcher or practitioner, and especially in gender monitoring. Furthermore, the participants came from different institutional backgrounds (private research-intensive organisations, higher education institutions, and public research organisations).

The focus groups took place from 26.5.2023 to 14.6.2023, with 28 participants¹¹ in total. The goal has been to have 40 participants. For this purpose, we invited more than 60 people. However, due to internet connectivity issues and time conflicts, especially in Eastern and Southern Europe, some persons who confirmed their participation had to cancel at short notice. We moderated the discussion with a guideline to facilitate comparable results. We recorded and transcribed the discussion for documentation and extracted the main results.

¹¹ Northern and Western Europe: 10; Eastern Europe: 5; German-speaking countries: 8; Southern Europe: 5.



The expert's comments improved the understanding of the indicators' objectives and the impact indicators' comprehensiveness and specified the distinction between the "prevalence" and "characteristics" indicators.

The first version of the indicators, which integrated the results of the focus groups and remarks made by consortium members at the second project meeting (June 20–21, 2023), is published in **D.3.1 Report GEP Prevalence Monitoring Indicator Framework v1** (see Löther, Karataş, and Weber 2024a).

The analysis of the pilot survey led to changes in the questionnaire for the full survey. We revised the indicators based on these changes and the results of the full survey and the full web scraper. The revision also takes the discussion of the co-creation workshop "Data monitoring" (21-22/11/2024) into account.

This report presents the final version of the indicators while the result of the reactive method (i.e., online survey) and non-reactive method (i.e., web scraping and text analysis) is published in **D3.3 Report on GEP Monitoring Survey+Webcrawl Results** (see Löther, Karataş, and Weber 2024b)

3.3 Developed Indicators

A detailed list of indicators can be found in the appendix (see section 10.1).

3.3.1 Prevalence

Prevalence indicators measure whether a GEP is in place in research-performing or funding organisations and inform about the organisational context of these plans.

Indicator 1.1 A gender equality plan or a written and formal institutional strategy that fosters gender equality exists in a research-performing or research-funding organisation.

Different definitions of GEP exist that may hinder the comparability of the data. The European Institute for Gender Equality (2016b) defines a GEP as a set of actions aimed at identifying gender inequalities and bias, designing and implementing measures to correct these, and setting targets and monitoring progress via indicators. In a broader definition, it refers to a planned institutional change approach.

Our study follows the definition given in Horizon Europe, defining a GEP as "a set of commitments and actions that aim to promote gender equality in an organisation through a process of structural change" (European Commission 2021a, 11). The minimum requirements are a public document, dedicated resources, data collection, monitoring, and awareness raising/training on gender equality (European Commission 2021a, 9). Due to practicability, non-reactive methods might have to use a less strict definition. Depending on the country and the organisational context, some organisations might not have a GEP focusing on gender inequalities but a broader equality or diversity plan. Table 6 informs about publications that provide or use indicators on the prevalence of GEPs:



Table 6 Publications: Indicators on the prevalence of GEPs

Publication	Indicator	Used or recommended application
Bührer and Wroblewski (2019, 5)	The existence of an institutional strategy, inter alia, a gender equality plan	Used for a survey of EU-funded researchers (2016)
Research Council of Norway (2016, 16)	Gender equality plans: Do these exist?	Recommendations for transnational indicators
European Commission (2016, 110–11)	The Proportion of research-performing organisations (RPOs) that adopted gender equality plans	Used for an ERA RPOs Survey (2014)
Wroblewski et al. (2015, 64–65)	Share of RPOs with gender equality plans	Used for an ERA RPOs Survey (2013)

Due to the lack of European-wide surveys asking for GEPs, the She Figures used web-scraping since 2018. The indicator displays the proportion of RPOs that have taken measures and actions to promote gender equality (European Commission 2025, 114–20), thus focusing on measures instead of an institutionalised and formalised strategy. Guyan and Douglas Oloyede (2020) and Higher Education Authority (2018) also ask for interventions and initiatives instead of GEPs.

The indicator applies to reactive and non-reactive methods. Nevertheless, survey data are sensitive to a nonresponse bias and overestimating the prevalence, especially when the survey focuses on gender equality. Web scraping data, on the other hand, may underestimate the prevalence in some countries and organisational contexts due to the preference for diversity plans and policies to publish the GEP (see Löther, Karataş, and Weber 2024b, 69–70).

Indicator 1.2 The existence of a GEP in an organisation is linked to national and European initiatives or requirements or third-party funding in terms of:

- **Participation in EU structural change projects, e.g., Horizon 2020, Horizon Europe**
- **Received funding other than EU projects, e.g., from national RFOs**
- **Received non-monetary support**
- **Legal requirements or requirements linked to getting research funding.**

In this indicator, we investigate the link between the prevalence of a GEP and any legal requirements or national factors. Eleven countries require a GEP to be in place, mainly for higher education and research institutions (Wroblewski 2023, 21–22; ERAC Standing Working Group on Gender in Research and Innovation 2021). National initiatives like the Athena SWAN Charter (Barnard 2017) or the German Women's Professorship Program (Biela et al. 2022;



Löther 2019) encourage higher education institutions and research organisations to approve a GEP. On a European level, besides the eligibility criterion in Horizon Europe, the European Commission has supported the implementation of GEPs in thirty structural change projects with over 200 research performing and funding organisations (European Commission 2021b, 168). As context for the prevalence of GEPs, we need to get information about links to national and European initiatives and third-party funding.

All focus groups highlighted external reasons and motivations for setting up a GEP, such as national legal frameworks, Horizon Europe funding criteria and the receipt of funding. To date, the literature does not describe a comparable indicator.

3.3.2 Characteristics

Indicators of the GEP's characteristics refer to their procedural and formal features, including leadership commitment, activities, and relation to other inequalities.

Indicator 2.1 The highest body of an organisation approved and signed the GEP and takes ownership of the GEP.

Leadership commitment is crucial for effective change (Mergaert, Cacace, and Linková 2022; Palmén and Kalpazidou Schmidt 2019). Analysing the ownership of the GEP – who approved and signed it – may serve as a first indication of the commitment. The focus groups emphasised ownership and acceptance of the GEP as crucial. Two publications mention an indicator of leadership commitment:

Table 7 Publications: Indicators of leadership commitment and ownership

Publication	Indicator	Used or recommended application
Mergaert, Cacace, and Linková (2022, 10)	Impact driver “Leadership actively committed to gender equality/gender mainstreaming” with the indicator “There is an explicit and visible commitment of leaders to GM.”	Proposed for assessment
Research Council of Norway (2016, 16)	Gender equality plans: Approved at what level in the organisation?	Recommendations for transnational indicators

Indicator 2.2 Type of Publication of the GEP: The GEP is accessible to people outside the organisation, or is an internal publication which is accessible to all members of the organisation, or is an internal document which is accessible only to the management.

The publication of the GEP is mandatory in the definition of GEPs given by Horizon Europe. It provides transparency and is also an indication of the commitment of the leadership and the organisation. The literature doesn't provide indicators of the GEP's publication.

***Indicator 2.3 Date of the first and current plan's adoption and time frame of the current plan***

The date of the first GEP indicates how long the organisation is engaged in gender equality. The time frame informs whether the GEP is regularly revised and adopted. Furthermore, the date of the first plan and the time frame of the current plan contextualise information about activities and their implementation and impact. We would assume more progress in organisations engaged in gender equality for longer. According to the focus groups, the time frame is a context factor but not a quality criterion.

The literature doesn't provide indicators on this issue.

Indicator 2.4 The GEP addresses measures in each area of activity (grouped according to the specification for Horizon Europe).

Gender equality policy involves measures in different areas. The European Commission (European Commission 2021a, 6) suggests including activities in the following fields: a) actions of awareness-raising and training, b) work-life balance and organisational culture, c) gender balance in leadership and decision-making, d) gender equality in recruitment and career progression, e) measures against gender-based violence, including sexual harassment, and f) integration of the gender dimension into research and teaching content. Feedback from selected experts on the questionnaire recommended further specifying the target groups for the area of activities. Furthermore, data should also measure the quantity of activities in the different areas.

Many publications provide indicators which ask for measures in different areas, typically for monitoring the implementation of measures, not for monitoring the GEP:



Table 8 Publications: Indicators of measures in different areas

Publication	Indicator	Used or recommended application
Doneva, Gaftandzhieva, and Boykova (2022, 3388)	Area of intervention	Proposed GEP implementation reports for individual organisations
Mergaert, Cacace, and Linková (2022, 10)	Impact driver “Coverage of the different dimensions/areas of GE institutional change” with the indicators “comprehensiveness of the GEP/GM work in terms of areas addressed.”	Proposed for assessment
Subdirección General para el Emprendimiento, la Igualdad en la Empresa y la Negociación Colectiva de Mujeres (2021)	Areas for the initial diagnostics: recruitment and selection processes, professional roles, training opportunities, career progression, working conditions, work-life balance, gender-based violence, inclusive language, workplace safety	Proposed indicators for audit and initial diagnostic
Heidler and Reichwein (2018, 7)	Implemented measures by dimensions (staff/early career researchers, work/life balance, research/academic culture, organisational development, gender in research & training, quality assurance)	Used for the analysis of submitted reports
Higher Education Authority (2018, 61–62)	Frequency of recruitment and promotion initiatives (differentiated by type of initiative)	Used for the analysis of Athena SWAN applications and self-audited Irish institutions

Indicator 2.5 The GEP integrates inclusive approaches like:

- **Gender diversity**
- **Intra-categorical differences inside the gender groups (e.g. women with migration background, Black women, etc.)**
- **Intersectionality (on a more conceptual level, not necessarily as a term)**
- **Diversity (inequalities addressed beside each other/"celebrating the differences" approach)**



Recently, the European Commission has expanded the understanding of gender equality "by opening policy to intersections with other social categories" (European Commission 2020, 16) and calls for *inclusive gender equality* approaches and inclusive gender action plans (European Commission 2022). This indicator provides information on whether concepts of intersectionality, inclusion and diversity are included in GEPs. Furthermore, inclusive GEPs also refer to non-binary gender concepts. Thus, the indicators also ask about gender diversity.

Even if several publications on monitoring gender equality interventions discuss intersectional approaches (Guyan and Douglas Oloyede 2020; Mour 2022), we couldn't find indicators of intersectionality and gender diversity which refer to monitoring GEPs.

Indicator 2.6 The GEP addresses different inequalities (race, class/social background, age, etc.).

If the GEPs include an inclusive approach (see Indicator 2.5), this indicator assesses the inequalities addressed in the GEP. As for indicator 2.5, several publications on monitoring gender equality interventions discuss different inequalities (Guyan and Douglas Oloyede 2020; Mour 2022), but we couldn't find indicators which refer to monitoring GEPs.

3.3.3 Implementation

Following the *programme management cycle* (Lipinsky and Schäfer 2016, 2–3), implementation monitoring concerns the realisation or execution of the GEP and, therefore, the outcomes (Douglas Oloyede 2014b). The indicators envisage monitoring the implementation and the prerequisites for a successful implementation, such as financial and personal resources, the involvement of stakeholders, and quality assurance. Implementation indicators monitor, on the one hand, whether the RPOs fulfil the requirements regarding dedicated resources and data collection and monitoring. On the other hand, implementation indicators like financial and personal resources, regulations including gender equality requirements, and stakeholder involvement refer to impact drivers.

Indicator 3.1 Assessment of planned and implemented gender equality measures (by areas of activity):

- ***Level of realisation: planned – started – implemented***
- ***Institutional coverage: implemented across the institution, implemented in some departments, not implemented anywhere.***

Primarily, the indicators on implementation try to get information on whether the organisations realise the measures envisioned in the GEP. The criteria for implementation are the level of realisation on the one hand and institutional coverage on the other hand. Monitoring the implementation in an individual organisation would assess the realisation of each measure mentioned in the GEP. A supra-organisational monitoring can't reach this complexity, but the indicator groups the assessment of implementation by areas of activity (see indicator 2.4). Table 9 displays publications that provide or use indicators for monitoring the realisation of GEPs:



Table 9 Publications: Indicators for monitoring the realisation of GEPs

Publication	Indicator	Used or recommended application
Doneva, Gaftandzhieva, and Boykova (2022, 3388)	Planned indicators; quantitative achievements by measure; qualitative achievement by measure; relevance of the action	Recommended for GEP implementation reports of individual organisations
Mergaert, Cacace, and Linková (2022, 10)	Impact driver “Coverage of the different dimensions/areas of GE institutional change” with the indicator “comprehensiveness and sophistication of the work within addressed areas.”	Proposed for assessment
Heidler and Reichwein (2018, 6–7)	Number of measures per institution by the level of realisation (planned, implemented and established); growth in the implementation of measures by action area	Used for university reports and the analysis of individual reports
Higher Education Authority (2018, 61–62)	Frequency of recruitment and promotion initiatives (differentiated by type of initiative)	Used for the analysis of Athena SWAN applications and self-audited Irish institutions

Indicator 3.2 Financial resources are dedicated to gender equality measures/implementation of GEP.

The implementation of a GEP depends on financial resources. For example, the European Commission states in the eligibility criteria for Horizon Europe that “the GEP has dedicated resources and expertise in gender equality to implement the plan.” The financial volume “should be appropriate to the size and needs of the organisation and its GEP.” (European Commission 2021a, 20). Still, assessing whether the financial resources are sufficient and comparing organisations is challenging. Furthermore, the indicator focuses on internal financial resources because allocating financial resources to gender equality indicates commitment, whereas relying only on external financial resources may disrupt sustainability.

Despite the importance of resources, only a few publications explicitly mention indicators of financial resources:



Table 10 Publications: Indicators of financial resources

Publication	Indicator	Used or recommended application
Mergaert, Cacace, and Linková (2022, 10)	Impact driver “availability of resources” with the indicator “There are funds dedicated to GE”	Proposed for assessment
Nimo (2021, 445)	Rate of variation of the organisation (specific university) budget allocated to the Equality Office; Rate of change in the number of people who have had access to resources on equality	Used for the monitoring of a specific university

Indicator 3.3 A gender equality unit, gender equality committee and/or dedicated staff are in place.

Closely linked to financial resources is dedicated staff. The European Commission proposes “a dedicated gender equality function” as one kind of human resource for implementing a GEP (European Commission 2021a, 20). Dedicated staff ensures expertise in setting up and implementing gender equality policies. Mergaert, Cacace, and Linková (2022, 10) describe two aspects of staff engaged in gender equality as impact drivers: a core team of change agents and the availability of resources, including knowledge and expertise.

The structure of the personal resources – (gender) equality officer or unit, employee representation or part of the management dedicated to gender equality or multiple inequalities – may differ according to legal requirements and the organisation type. Thus, the indicator should grasp the existence of dedicated staff and the organisational varieties. Despite the importance of dedicated staff and expertise, few publications propose indicators for gender equality staff (see Table 11).



Table 11 Publications: Indicators of dedicated staff

Publication	Indicator	Used or recommended application
Mergaert, Cacace, and Linková (2022, 10)	Impact driver “core team of change agents” with four indicators (core team of change agents exists; core team comprises motivated people; core team has a formal mandate and ownership; core team has access to an extended group of change agents); impact driver “availability of resources” with the indicator “internal gender knowledge and expertise are available and used”.	Proposed for assessment
Heidler and Reichwein (2018, 1)	Equal opportunity structures	Used for university reports and the analysis of individual reports

Indicator 3.4 Gender equality is integrated into institutional/internal regulations (appointment regulation, basic rules of the institution, etc.).

Structural and organisational change also involves changes in regulations, which are not primarily concerned with gender equality, like recruitment procedures. Thus, knowledge about regulations, which include gender equality requirements, informs the implementation of organisational change.

Only one publication proposes indicators of organisational change.



Table 12 Publication: Indicators of organisational change

Publication	Indicator	Used or recommended application
Mergaert, Cacace, and Linková (2022, 11)	Impact driver “organisational governance” with the indicators “gender-sensitive routines exist” and “Gender-specific routines exist”.	Proposed for assessment

Indicator 3.5 Internal and external stakeholders are involved in the GEP implementation.

The involvement of internal and external stakeholders, besides the change agents with expertise in gender equality, is essential in gaining support and allies for the GEP (Mergaert, Cacace, and Linková 2022, 6). Among the internal stakeholders, leadership commitment is crucial, as already stated (Research Council of Norway 2016, 10). Thus, the indicator assesses whether internal and external stakeholders are involved in implementing the GEP.

Only one publication proposes indicators of stakeholders.



Table 13 Publications: Indicators of stakeholder involvement

Publication	Indicator	Used or recommended application
Mergaert, Cacace, and Linková (2022, 10)	Impact driver “involvement of internal stakeholders” with five indicators (leadership engagement with the core team of change agents and GE work; variety of internal stakeholders; number of people/size of groups that engage with GE efforts; degree of adherence to GE goals; internal stakeholders start initiatives themselves) and impact driver “involvement of external stakeholders and experts” with four indicators (NGOs/CSOs are involved in the institutional GE work; gender dimension addressed in events; external partnerships with relevant institutions for GE work; external gender expertise available and used)	Proposed for assessment

Indicator 3.6 A policy for monitoring/controlling gender equality measures and the GEP implementation is in place.

In the programme management cycle, monitoring is a critical element in implementing and developing policy initiatives. Monitoring as a part of quality assurance is a prerequisite for a data-driven implementation and further development of the GEP. The European Commission declares data collection and monitoring mandatory for a GEP to comply with the Horizon Europe eligibility criterion (European Commission 2021a, 5).

The indicator generally asks whether a monitoring and/or controlling system for gender equality is in place. A monitoring and controlling system goes beyond data collection because it links data to the objectives of the GEP and measures progress. Systems for quality assurance might also include other approaches like evaluation.

Many publications propose the collection and publication of sex-/gender-disaggregated data, but only a few deal with indicators of a monitoring policy.



Table 14 Publication: Indicators for the policy of monitoring and/or controlling gender equality measures and GEPs

Publication	Indicator	Used or recommended application
Mergaert, Cacace, and Linková (2022, 10)	Impact driver “Transparency and accountability” with “GE is included in unit reports and assessment for internal monitoring” and “GE reporting is done publicly available”; Impact driver “organisational governance” with the indicators “Gender analysis is considered in internal monitoring” and “Gender analysis is considered in internal audits and institutional assessments.”	Proposed for assessment
Research Council of Norway (2016, 16)	What is the system of implementation and monitoring?	Recommendations for transnational indicators

Indicator 3.7 Sex-disaggregated data are collected and published.

Data collection is mandatory for Horizon Europe's eligibility criteria. "To be eligible for Horizon Europe, organisations must collect and publish disaggregated data on the sex and/or gender of personnel (and students, where relevant) and carry out annual reporting based on indicators" (European Commission 2021a, 23). Although Horizon Europe talks about “sex and/or gender”, statistical data usually refer to sex (as “biological and physiological characteristics”)¹², not to gender (as “social attributes and opportunities associated with being female and male and to the relationships between women and men and girls and boys, as well as to the relations between women and those between men”). Thus, the indicator focuses on sex-disaggregated data. The indicator doesn't ask for specific data but, more generally, whether sex-disaggregated data are collected and published.

The literature proposes many quantitative data that rely on collecting sex-disaggregated data (Wroblewski and Eckstein 2018; Nimo 2021), but the publications rarely propose a more general indicator for collecting and publishing data.

¹² EIGE Glossary & Thesaurus, Sex: <https://eige.europa.eu/publications-resources/thesaurus/terms/1048>, gender: <https://eige.europa.eu/publications-resources/thesaurus/terms/1046>.



Table 15 Publication: Indicator about data collection and publication

Publication	Indicator	Used or recommended application
Mergaert, Cacace, and Linková (2022, 10)	Impact driver “Data collection and statistical analysis” with the indicators “institutional gender-disaggregated data are collected” and “institutional gender-disaggregated data and statistics are collected public and accessible”	Proposed for assessment
Higher Education Authority (2018, 46)	A comprehensive gender-disaggregated data collection system will be in place in every HEI.	Recommendations

Indicator 3.8 Gender diversity and intersectional perspectives are integrated into the monitoring or data collection:

- **Inclusion of a broader and non-binary understanding of gender diversity,**
- **Inclusion of other inequalities in conjunction with gender.**

The expanded approach to open gender equality “to intersections with other social categories” (European Commission 2020, 16) and to inclusive gender action plans (European Commission 2022) (see indicators 2.5 and 2.6) also concerns monitoring and data collection as part of inclusive GEPs. Feminist theory and gender research, on the one hand, and legal changes in some European countries, which recognise genders other than men and women, on the other hand, call attention to the fact that gender-disaggregated data must go beyond a binary concept. Thus, the indicator informs whether and how monitoring and data collection integrate a non-binary understanding of gender (Hadler et al. 2022; Lindqvist, Gustafsson Sendén, and Renström 2020).

In addition, integrating intersectional perspectives into monitoring and data collection refers to collecting data on inequalities other than gender, like race, class, and age, and combining different axes of inequalities, e.g., race and gender. The situation of collecting inequality data is very diverse in the European countries: Whereas the Athena SWAN initiative has a strong emphasis and increasing attention to the intersectionality perspectives and data on ethnicity are part of the national statistics in the UK and Ireland, other countries like Spain, France and Germany refrain from collecting such data (Löther, Karataş, and Weber 2023, 37; Claeys-Kulik, Jørgensen, and Stöber 2019; Ovseiko et al. 2019).

We didn’t find any indicator of monitoring the integration of a non-binary gender understanding, and only one proposed an indicator of intersectional approaches in data collection and monitoring.



Table 16 Publications: Indicators on gender diversity and intersectional perspectives in monitoring and data collection

Publication	Indicator	Used or recommended application
Mergaert, Cacace, and Linková (2022, 10)	Impact driver “Data collection and statistical analysis” with the indicator “Intersectional gender-disaggregated data are collected and published.”	Proposed for assessment

3.3.4 Self-Assessment and Impact

Impact refers to the long-term effects of measures and interventions like GEPs. Assessing impacts fully is only possible through the evaluation of GEPs and gender equality policies (Technische Universität Wien 2021; Bühner et al. 2020; Palmén et al. 2019a; Aldercotte 2018; Wroblewski, Kelle, and Reith 2016). The focus groups and the co-creation workshop “Data monitoring” also confirmed that it is impossible to analyse the impact from the outside but rather internally. Using qualitative approaches to capture impact is preferable, as it relies on institutional reflection.

Despite this limitation, the INSPIRE indicators aim to get indications on impact mainly by assessing factors that enable change towards gender equality. Indicators which assess these factors are included in the chapters “Characteristics” and “Implementation”.

Following the allusion to institutional reflection, the following indicators on self-assessment and impact monitor the understanding of the impact and the methods and data used in the RPOs and refer to self-assessment. To obtain quantitative data, the indicators ask for changes in the gender distribution in leadership positions.

Indicator 4.1 The GEP includes a reflection or description of the GEP impact. The organisation reflects its understanding of the GEP impact.

Asking about the impact understanding and methods/data used refers to accountability, transparency, and quality assurance as impact drivers. This indicator informs whether the organisation reflects on the impacts of the GEP.

Douglas Oloyede (2014a) doesn’t provide indicators of the understanding of impact but integrates the reflection on impact in developing qualitative indicators for measuring progress on equality.



Table 17 Publications: Indicators on the understanding of impact

Publication	Indicator	Used or recommended application
Douglas Oloyede (2014a, 10)	What change in factors (experience, attitudes, confidence, behaviour) do you want to see? What would success look like, and how should it be measured?	Questions for developing qualitative indicators

Indicator 4.2 The GEP designates data and methods (e.g. evaluations, surveys or qualitative data) to assess the GEP impact.

The indicator is linked to the previous indicator and refers to accountability, transparency and quality assurance. Indicators 3.6 and 3.7 inform about monitoring systems and data collection, while this indicator provides more precise information on impact assessment methods. We found one publication with such indicators.

Table 18 Publications: Indicators of data and methods to assess the impact

Publication	Indicator	Used or recommended application
Claeys-Kulik, Jørgensen, and Stöber (2019, 36)	How do you measure the impact of your activities on diversity, equality and inclusion? Options: number of students/staff/graduation rate – from underrepresented/disadvantaged backgrounds; success stories	Used in a survey among HEIs

Indicator 4.3 Changes in the participation of women within a fixed period (5-10 years) and comparing RPOs with/without GEP (or gender equality measures) in leading positions

Most gender equality impact indicators presented in the literature rely on data on the representation of women at different career stages. Calculating changes in time and using the difference-in-differences approach (comparing institutions with/without GEPs) are possibilities to causally attribute observed effects to the intervention (implementation of a GEP).

The focus groups proposed integrating several career stages and positions, such as leadership positions, newly recruited women in leadership positions, non-academic positions and administrative staff, senior managerial positions, and decision-making boards. Because the indicators aim to monitor GEPs on a supra-institutional and European-wide level, we decided to restrict the indicator to leadership positions, namely grade A according to the She Figures



(European Commission 2025, 121) and leading researcher (R4) according to the European Framework for Research Careers (European Commission 2011, 2).

Table 19 Publications: Indicators on the representation of women

Publication	Indicator	Used or recommended application
Nimo (2021, 445)	Number of predoctoral contracts, percentage of women graduates, percentage of women in decision-making bodies	Used for monitoring of a specific university
B. Stadler and Wroblewski (2021)	Composition of students, graduates and different groups of employees, gender pay gap among professors, glass ceiling index, presence of women in appointment procedures	Used in gender equality reports of Austrian universities
O'Connor and Irvine (2020)	Gender pay gap, gender profile of the professorate, gender profile of senior management	Used in the evaluation of several gender equality initiatives in Ireland
Löther (2019, 7)	Evaluation of the proportion of women professors at HEIs participating or non-participating in a German gender equality program	Used in the evaluation of the German women professors' programme
Gregory-Smith (2018, 479)	Female employment and female part-time professors; comparison of schools who have applied for Athena SWAN and schools which have not	Used to evaluate the effectiveness of Athena SWAN
Stepan-Norris and Kerrissey (2016, 226)	Percentages of women among faculty, new faculty hires and separations	Used for the evaluation of the ADVANCE Program in one university



Wroblewski and Leitner (2013)	Gender pay gap, gender in recruitment procedures	Used for the analysis of gender equality reports in Austrian universities.
Timmers, Willemsen, and Tijdens (2010)	Changes in the share of women among academic staff, professors, PhD students and students; changes in the glass ceiling index	Used for the evaluation of gender equality policies in Dutch universities

Indicator 4.4 Self-assessment: Rating of the changes in the areas of activity (gender balance, awareness, knowledge about gender (and other) inequalities, promotion and recruitment procedures, work-life-balance) and relevance of GEPs for the achieved changes

Impact monitoring on the level of individual organisations uses surveys among staff and students, but this approach is not feasible for comparing institutions on a supra-organisational level. Besides expert and qualitative assessments (Mergaert, Cacace, and Linková 2022; Douglas Oloyede 2014a), which are not possible for the international monitoring of a large sample, self-assessment of the organisations is another approach, although suffering from biased rating as mentioned in the focus groups.

A GEP and measures, as well as other circumstances inside and outside the organisation, can affect changes in gender equality. Thus, the indicator should also inform about the relevance of the GEP for the achieved changes, which should also be measured through self-assessment.

The “six stages of institutional capacity development” of the impact driver model, labelled as “Starting point, Project, Inception, Growth, Integration, and Institutionalisation” (Mergaert, Cacace, and Linková 2022, 9), could serve as a scale for the self-assessment.

Two publications mention indicators of self-assessment and estimating the relevance of gender equality actions:



Table 20 Publications: Indicators on impact self-assessment

Publication	Indicator	Used or recommended application
Doneva, Gaftandzhieva, and Boykova (2022)	Estimated the relevance of gender equality actions for the achievement of the observed outcome (quantitative and qualitative achievements)	Proposed for GEP implementation reports for individual organisations
Douglas Oloyede (2014a)	Qualitative questions at institutional, departmental and functional levels: How far is equality mainstreamed or embedded? Whether attitudes towards equality have become more positive? A belief that equality is taken into account at the senior strategic level.	Questions for developing qualitative indicators

3.4 Adapting the Indicators for Reactive and Non-Reactive Methods

Based on the conception of indicators for GEP monitoring, we examined how the indicators apply to non-reactive methods (namely web scraping, automated text analysis, and publicly available administrative data) and reactive methods (the survey).



Table 21 Operationalising the indicators for reactive and non-reactive methods

Indicator	Non-reactive methods	Reactive methods	Challenges
Indicator 1.1 A gender equality plan or a written and formal institutional strategy that fosters gender equality exists in a research-performing or research-funding organisation.	Web scraping	Survey	Definition of GEP; common and comparable understanding of GEP; different focus on gender equality, equity and/or diversity with differences according to countries and organisation types; unit of the GEP (whole organisation, not departments); publication of the GEP; nonresponse bias
Indicator 1.2 The existence of a GEP in an organisation is linked to national and European initiatives or requirements or third-party funding in terms of: participation in EU-funded structural change projects	Search in CORDIS database	Not necessary when using a list of participating organisations	
Indicator 1.2 The existence of a GEP in an organisation is linked to national and European initiatives or requirements or third-party funding in terms of: Received funding other than EU projects, e.g., from national RFOs	Not possible	Survey	Clear distinction between funding for gender equality measures or setting up a GEP
Indicator 1.2 The existence of a GEP in an organisation is linked to national and European initiatives or requirements or third-party funding in terms of: Received non-monetary support	Not possible	Survey	Clear distinction between funding for gender equality measures or setting up a GEP
Indicator 1.2 The existence of a GEP in an organisation is linked to national and European initiatives or requirements or third-party funding in terms of: Requirements linked to getting European research funding	Web scraping or text analysis	Survey	Survey: answering truthfully
Indicator 1.2 The existence of a GEP in an organisation is linked to national and European initiatives or requirements or third-party funding in terms of: Requirements for research funding (national)	Not possible; using EU reports	Survey	Survey: answering truthfully
Indicator 2.1 The highest body of an organisation approved and signed the GEP and takes ownership of the GEP.	Not possible	Survey	Comparability of boards; defining the highest board



Indicator 2.2 Publication of the GEP: The GEP is accessible to people outside the organisation, an internal publication which is accessible to all members of the organisation, or an internal document which is accessible only to the management.	Not possible	Survey	Publication is one of four criteria for a GEP, according to Horizon Europe.
Indicator 2.3 Date of the first and current plan's adoption	Not possible	Survey	Respondents might not know the first GEP; changes in GEP concepts over time.
Indicator 2.3 Time frame of the current plan	Text analysis	Survey	
Indicator 2.4 The GEP addresses measures in each area of activity (grouped according to the specification for Horizon Europe).	Text analysis	Survey	A clear description of the areas
Indicator 2.5 The GEP integrates inclusive approaches.	Text analysis	Survey	Aspects: gender diversity, intra-categorical differences, intersectionality as a concept, diversity as a concept. Different understanding of the concepts in different countries (gender diversity: references to non-binary gender identities and to gender-mixed teams)
Indicator 2.6 The GEP addresses different inequalities (race, class/social background, age, etc.).	Text analysis	Survey	List of inequalities
Indicator 3.1 Assessment of planned and implemented gender equality measures (by areas of activity)	Not possible	Survey	Common understanding and comparability of the scale "planned", "started", "and implemented"; the status of the implementation depends on the time frame of the GEP
Indicator 3.2 Financial resources are dedicated to gender equality measures/implementation of GEP.	Not possible	Survey	Dedicated resources = one of four criteria for a GEP according to Horizon Europe; amount difficult to determine (no fixed budget, resources in different parts of the organisation)
Indicator 3.3 A gender equality unit, gender equality committee and/or dedicated staff are in place.	Web scraping or text analysis	Survey	Differentiation between staff and internal stakeholders/change agents (voluntary activities); provide exact definitions or list of functions



Indicator 3.4 Gender equality is integrated into institutional/internal regulations (appointment regulation, basic rules of the institution, etc.).	Difficult: web scraping and text analysis of vision, strategy or mission statement	Survey	Provide a list of core regulations
Indicator 3.5 Internal and external stakeholders are involved in the implementation of GEP.	not possible	Survey	Differentiation between the adoption and implementation of the GEP
Indicator 3.6 A policy for monitoring/controlling gender equality measures and the GEP implementation is in place.	Text analysis	Survey	Focus on mentoring or quality assurance (including evaluation)
Indicator 3.7 Sex-disaggregated data are collected and published: collection	Text analysis	Survey	Data collection = one of four criteria for a GEP according to Horizon Europe
Indicator 3.7 Sex-disaggregated data are collected and published: publication	Text analysis	Survey	
Indicator 3.8 Gender diversity and intersectional perspectives are integrated into the monitoring or data collection.	Not possible	Survey	Comparable understanding of gender diversity and intersectional perspectives; clear definition and explanation necessary
Indicator 4.1 The GEP includes a reflection or description of the GEP impact. The organisation reflects its understanding of the impact of GEP.	Text analysis	Survey	Analysis of qualitative data
Indicator 4.2 The GEP designates data and methods (e.g. evaluations, surveys or qualitative data) to assess the GEP impact.	Text analysis	Survey	Provide a list of possible methods
Indicator 4.2 Changes in the participation of women within a set period (5-10 years) and comparing RPOs with/without GEP (or gender equality measures) in leading positions	HEI: ETER database; GOV, BES and RFO: not possible	Survey	Exact definition of career stages; comparability of the career stages; availability of data; willingness to indicate the data in a survey
Indicator 4.3 Rating of the changes in the areas of activity (gender balance, awareness, knowledge about gender (and other) inequalities, promotion and recruitment procedures, work-life) and relevance of GEPs for the achieved changes: assessment	Not possible	Survey	Self-assessment is influenced by the position of the respondent (management, gender equality staff), understanding of the scale, comparability, and lack of a baseline.



Indicator 4.4 Self-assessment: Rating of the changes in the areas of activity (gender balance, awareness, knowledge about gender (and other) inequalities, promotion and recruitment procedures, work-life balance) and relevance of GEPs for the achieved changes: relevance	Not possible	Survey	Self-assessment is influenced by the position of the respondent (management, gender equality staff);
---	--------------	--------	--

4 Non-Reactive Methods for the Full Sample

This chapter will display how we operationalised and applied the developed indicators for GEP monitoring (see Chapter 3) to non-reactive methods (namely web scraping and automated text analysis). INSPIRE developed and piloted a "non-reactive" instrument designed to operate autonomously, navigating through the vast landscape of the internet and extracting pertinent information directly from websites. This innovative approach involves creating and testing a sophisticated automated web-scraping system. It allows for real-time and continuous data gathering, providing a dynamic and up-to-date perspective on the prevalence of GEPs. The web crawler becomes a proactive tool, proactively seeking and extracting relevant data, eliminating the lag associated with reactive survey instruments.

The non-reactive method is an umbrella term which refers to techniques where data is collected without subjects being aware that they are monitored (R. M. Lee 2019; Webb et al. 1999; Janetzko and Kennke 2004). Hence, these methods are also known as *unobtrusive*, *indirect*, *hidden*, *naturalistic*, *noninvasive*, or *non-disruptive*. From an epistemological standpoint, if the researcher does not interact with the subjects during the study, it can be considered non-reactive, i.e., a simple observatory or archive document (Janetzko 2016). The main objective of this strategy is to reduce the influence of observation on the reactions of the individuals or cases who participate in the research. Given that people tend to change their behaviour when they know they are under study, non-reactive approaches allow the capture of more unbiased data (Webb et al. 1966, pr. [1976]). Early literature addresses three non-reactive methods: simple observations, physical traces, and non-reactively gathered archive documents (Webb et al. 1966, pr. [1976]).

For over a decade, novel non-reactive methods such as data gathering from the internet have prevailed, e.g., big data. The tremendous growth of the internet over the last decades has created new possibilities for researchers for non-reactive data collection. This possibility includes using datasets from which information can be gleaned, e.g., via data mining, web scraping, etc. (Leskovec, Rajaraman, and Ullman 2020). Big data represents the new kinds of digital data available and the tools and technologies required to access these data. Big data was claimed to be a paradigm shift in how we comprehend the world (Eynon 2013; Lazer and Radford 2017). Many scholars, including social scientists, are utilising massive quantities of information produced by and about people, things, and their interactions to extend knowledge and respond to research questions. To that end, new technologies (e.g. APIs for web crawling) allow the collection of enormous amounts of data (Olmedilla, Martínez-Torres, and Toral 2016; Eynon 2013).

Recent developments in information technologies have enabled data capture, storage, and handling to be quicker and more effective than traditional methods. Automation is rapidly growing, significantly influencing research (Yarkoni et al. 2021). In that regard, Big Data creates an opportunity in social science from the quantitative side, which can facilitate traditional social science methods where collecting data has always been challenging, time and resource-intensive (Olmedilla, Martínez-Torres, and Toral 2016; Macanovic 2022).

In big data, web scraping and web crawling technologies automate obtaining massive amounts of data from the Internet (Nigam and Biswas 2021). Web scraping and web crawling are related



and essential methods for collecting data from the internet, each serving a distinct role within the broader context of information retrieval (Kulyk 2023).

Before exploring the INSPIRE approach to monitor GEPs via non-reactive methods, some basic information about data extraction from the internet is essential for understanding.

4.1 Web Crawling versus Web Scraping

Although web crawling and web scraping share commonalities in their use of Hypertext Transfer Protocol (HTTP) requests and parsing of Hypertext Markup Language (HTML), their objectives differ fundamentally. Web crawling focuses on systematic exploration and indexing, forming the foundation for search engines (Sheinbaum 2023). Web scraping is centred around targeted data extraction for specific analysis and application purposes (Barton 2023). Together, these techniques complement each other, offering a comprehensive approach to understanding and utilizing the vast resources available on the web.

Understanding the two processes of web scraping and web crawling and how they interact or are used in our INSPIRE project is crucial, especially for tasks involving data collection, analysis, and application.

4.1.1 Web Crawling

Web crawling, often called web indexing, is a robust technique employed to systematically navigate the extensive terrain of the internet and catalogue information from websites. This methodical approach is a foundational process for search engines, facilitating the creation of comprehensive indexes that enable efficient information retrieval (Sharma, Shrivastava, and Singh 2021).

Web crawling involves a crawler, also known as a spider or bot, systematically traversing through web pages by following hyperlinks from one page to another. Unlike web scraping, which focuses on extracting specific data, web crawling is more concerned with indexing and mapping the web structure. The primary objective is to collect metadata and identify relationships between web pages.

The initiation of web crawling starts with a crawler sending HTTP or HTTPS requests to a list of seed URLs. These URLs act as starting points for the crawler to navigate the web. When the crawler visits a webpage, it collects information about the page, including its content, URL, metadata, and any hyperlinks present. The collected data is processed, stored, and organised to create a comprehensive map of interconnected web pages.

One key distinction lies in the breadth of coverage. Web crawling aims to explore many pages and index their content for efficient search functionality. The process involves traversing domains, sub-domains, and various pages within a website, ensuring a thorough examination of the entire web landscape (Sharma, Shrivastava, and Singh 2021).



Web crawlers often prioritize pages based on relevance, popularity, or recency to enhance efficiency. This prioritization ensures that the crawler focuses on indexing the most valuable and up-to-date content. Commonly employed algorithms, such as Breadth-First Search (BFS) or Depth-First Search (DFS), guide the crawler in navigating the web's complex structure.

4.1.2 Web Scraping

"Web-scraping" refers to the automated process of extracting data from webpages, typically using a software program (Luscombe, Dick, and Walby 2022; Noortje and Esther 2012). It is commonly used to gather information from various websites for analysis or research purposes. Using computational methods, web scraping allows data collection activities to be automated with established package examples (McDonnell 2020). Web scraping is also known as web harvesting or web data extraction. It is a systematic approach to gathering valuable data from the web, transforming it from an unstructured format into an organized structure suitable for analysis and application. Web scraping finds extensive use in various domains, including data analysis, research, and competitive intelligence, making it an indispensable tool for companies, researchers, and analysts.

At its core, web scraping involves systematically navigating through the underlying HTML structure of websites to locate and extract specific data. To initiate web scraping, tools send HTTP or HTTPS requests to the targeted website's server, replicating the behaviour of a web browser requesting a webpage and systematically navigating through the underlying HTML structure of websites to locate and extract specific data. Once the HTML content is received, the web scraper parses it. This process involves breaking down the HTML into a tree-like structure, leveraging distinctive HTML tags such as `<div>`, `<p>`, and ``, which organize and present content on the webpage. This parsing process transforms raw HTML into a structured format, facilitating more straightforward navigation and identifying specific elements within the HTML.

The efficacy of web scraping lies in its ability to navigate the structured HTML and pinpoint desired data accurately. XPath and CSS selectors play pivotal roles in this navigation. With its path-like syntax, XPath provides a systematic roadmap for the scraper. Conversely, Cascading Style Sheets (CSS) selectors use patterns for selection, offering flexibility in identifying and isolating elements on a webpage.

These tools act as virtual compasses, guiding the web scraper to specific elements on a webpage for extraction. Be it text, images, links, or other content, the scraper identifies and extracts targeted data using established paths and patterns. This meticulous process ensures precision in data extraction, allowing the scraper to selectively gather information relevant to its intended purpose. Web scraping can automatically parse the HTML of a web page to detect links that point directly to attached Portable Document Format (PDF) files, download these files to the local system, and store them for further processing or analysis. A detailed explanation of the INSPIRE methodology for downloading PDFs can be found in section 4.2.4.



Post-identification and extraction, the next step involves storing the information in a structured format. This structured storage facilitates subsequent analysis, making the extracted data valuable for informed decision-making. Typical formats for storage include Comma-Separated Values (CSV) files, databases, or JavaScript Object Notation (JSON) formats, each chosen based on the specific requirements of the intended analysis.

Websites often include a file called "robots.txt" that provides guidelines for web crawlers and scrapers. This file outlines which sections of the site can be accessed or scraped and which should be avoided. Adhering to these guidelines and respecting a website's terms of service and privacy policies is imperative for ethical and responsible web scraping practices.

Moreover, the dynamic nature of the web demands adaptability from web scrapers. Websites frequently undergo structural changes, and a scraper must be designed to handle these alterations gracefully. Failure to do so can result in broken scraping scripts, underscoring the importance of continuous monitoring and adaptation.

4.2 Detecting The GEPs: The INSPIRE's Web Scraper for the Full Sample

INSPIRE aims to address the existing gap in monitoring the prevalence of GEPs across European countries and sectors. The objective of INSPIRE's web scraper derives from the Prevalence indicator (see section 3.3.1), which aims to measure whether a GEP is in place in research-performing or funding organisations and inform about the context of these plans. To this end, INSPIRE aims to establish a robust, efficient, and scalable methodology for monitoring the prevalence of GEPs across various sectors and regions in Europe, particularly from the challenging-to-reach private sector, that can be reused for future research.

Recent research from the 2018 edition of *She Figures* uses web scraping. *She Figures* measured the proportions of organisations that hold gender equality measures, implementing the web-scraping tool (European Commission 2021b). Their web scraping tool reached an 86% accuracy rate in the trial phase. Nevertheless, this accuracy rate decreased when GEPs' PDFs associated with the webpage were also included in the analysis (European Commission 2021b, 2021c).

Currently, a reliable tool is lacking to track GEPs systematically. INSPIRE's web scraper aims to fill a critical gap in gender equality assessment across diverse sectors in Europe. The INSPIRE web scraper is an extended version of a traditional web scraper. We build upon web crawling that serves as a means of systematic exploration and indexing, ensuring a comprehensive understanding of the web landscape. We use SerpAPI¹³ and specify Google's crawled database to build INSPIRE's scraper. Google regularly crawl the webpages and index them based on their content. This enables INSPIRE to utilize SerpAPI for enhanced web scraping, effectively leveraging data from search engine-crawled sources. We use SerpAPI to enhance our INSPIRE web scraping efforts by incorporating data that search engines have

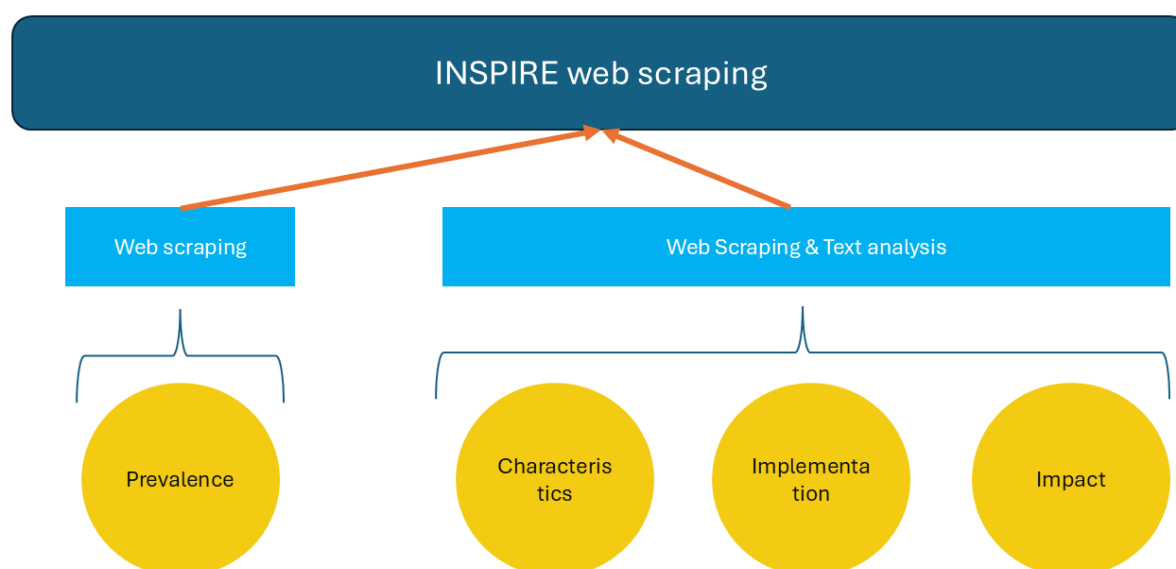
¹³ [SerpApi: Google Search API](#)



already collected. This approach capitalizes on the thorough and continuous web crawling carried out by search engines, enabling us to streamline the extraction of targeted and pertinent information. The web scraper is specifically designed to search for predetermined terms related to GEPs on the identified web domains. By analysing the content of web pages for these specific terms, the web scraper will retrieve relevant search results indicating the prevalence of GEPs within a particular webpage in the domain.

Using SearpAPI for web scraping leverages the autonomy and versatility of existing search engine web crawling, along with focused accuracy, to ensure a complete and detailed understanding of GEP prevalence and to retrieve GEP documents from various websites.

Figure 3 INSPIRE web scraping and its application to domains

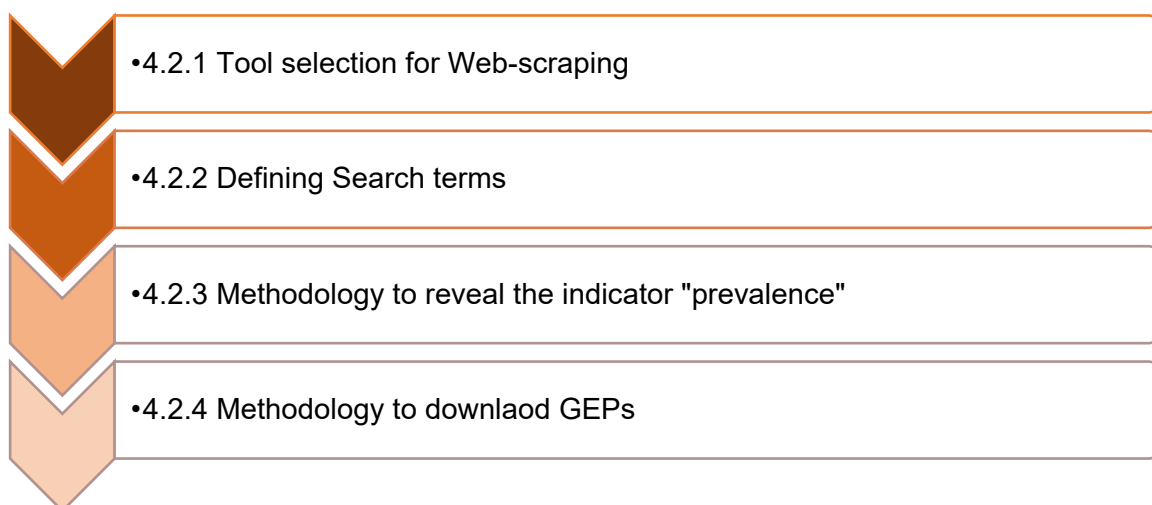


We used web scraping to collect information for our indicators using two different approaches: (i) detecting the prevalence of GEPs, and (ii) downloading GEP documents in PDF format, as shown in Figure 3. These two scripts serve different purposes and are based on different methodologies. Further technical details are provided in the sections 4.2.3 and 4.2.4. Once the PDFs were downloaded, we integrated them into a text analysis to extract relevant information for the indicators. The information extraction methodology for each indicator is explained in detail in section 4.3.

In addition, referring to our methodology as a 'scraper' aligns more accurately with the practices of web scraping. Figure 4 illustrates the web scraping process and the following chapters will explain the development of INSPIRE's scraper steps in depth.



Figure 4 Web scraping process



4.2.1 Tool Selection For Web-Scraping

Numerous libraries, including open-source ones, exist for web scraping and crawling in various programming languages (Glez-Peña et al. 2013). For the INSPIRE project, among many others, we examined three possible tools for constructing the scraper: Scrapy¹⁴, a Python scraping toolkit; OpenSearchServer¹⁵, a Java-based search engine software; and SerpAPI, a real-time API for accessing Google search results. We thoroughly compared these three tools, focusing on critical factors such as performance, features, and ease of use.

Although Scrapy and OpenSearchServer are open-source tools for web scraping while using, we noticed that their scraping rate is slow, making it very difficult to scrape 6,571 URLs for the full sample (see section 2.2). In addition, these tools have a notable risk of the server IP being blocked or receiving abuse reports, necessitating proxy services to mitigate this risk. In contrast, SerpAPI poses no threat of IP blocking and delivers fast, instant outcomes by utilising Google's extensive dataset. Eventually, we chose SerpAPI to develop INSPIRE's scraper since it is easy to use, allowing us fast query results with minimal response time.

Scrapy

Scrapy is a robust open-source web crawling framework for Python that provides comprehensive web scraping and data extraction functionalities. In the realm of continuous text processing, Scrapy offers a range of features that make it a versatile tool for navigating websites, retrieving textual content, and organizing data seamlessly.

Scrapy allows users to define and customize spiders, which are scripts that specify how to navigate a website, follow links, and extract information. This flexibility enables users to tailor their web crawling strategies to the specific structure and requirements of the target websites.

¹⁴ <https://scrapy.org/>

¹⁵ <https://www.opensearchserver.com/>



Scrapy's support for XPath and Cascading Style Sheets (CSS) selectors makes it efficient in extracting continuous text from HTML pages, ensuring accurate textual content retrieval.

The framework's ability to handle asynchronous requests and responses enhances its efficiency in dealing with dynamic content, such as JavaScript-generated pages. This ability ensures that Scrapy can effectively crawl and extract information from modern, interactive websites.

Scrapy supports various storage backends, allowing users to store and organize the scraped data in different formats and databases. Additionally, the framework provides mechanisms for handling common challenges in web crawling, such as handling cookies, redirects, and managing user-agent headers.

Pros:

- **Open Source and Free:** Scrapy is an open-source framework, making it freely accessible to developers and organisations. This affordability factor is particularly advantageous for those operating on tight budgets or looking for cost-effective solutions.
- **Concurrency:** Scrapy supports asynchronous requests and can handle multiple requests concurrently, enhancing performance during web scraping.
- **Compliance with Robots.txt and Noindex Rules:** Scrapy respects the rules outlined in the robots.txt file and adheres to no index directives. This ensures that the scraping activities align with ethical standards and legal guidelines, promoting responsible web scraping practices.

Cons:

- **Slow Scraping:** A notable drawback of Scrapy is its slow scraping speed. This slow processing can hinder the efficiency of data retrieval and impact the overall performance of scraping tasks.
- **Risk of IP Blocking and Abuse Reports:** A notable concern with Scrapy is the associated risk of server IP blocking and the potential receipt of abuse reports. This risk is particularly significant during extensive and prolonged scraping operations, and using proxy services is a must to overcome this issue.

Significant challenges surfaced when we implemented the INSPIRE web scraper while using Scrapy. Performance issues, such as slow scraping and frequent crashes, became apparent early in the implementation phase. For instance, the tool crashed after approximately 20 hours while crawling Greece's URL list without yielding any output. Moreover, a similar scenario occurred while exploring Ireland's URL list, with the tool crashing after around 26 hours and retrieving only a limited set of results. These issues persisted across the testing of multiple countries, URLs, and specific search terms, preventing Scrapy from completing the required



scraping tasks. It became evident that these limitations hindered our ability to achieve the desired functionality and reliability needed for INSPIRE's objectives.

OpenSearchServer

OpenSearchServer is an open-source search engine software that offers robust functionalities in the context of web crawling. It allows users to efficiently index and search through large volumes of data on the web. With its web crawling capabilities, OpenSearchServer can systematically navigate websites, retrieve information, and organize them for effective search operations. The software supports continuous text processing, enabling seamless analysis and indexing of textual content.

OpenSearchServer's web crawling functionality includes trailing websites, following links, and extracting textual data from HTML pages. It can handle dynamic content by executing JavaScript, ensuring it captures information from modern, interactive websites. The software supports customizable configurations for crawling, allowing users to tailor the crawling process to their specific needs.

The continuous text processing aspect of OpenSearchServer involves extracting and analysing textual content from web pages. The software can handle various text formats, ensuring that it can effectively index and search through diverse types of content. It also supports extracting metadata and other relevant information to enhance the search experience.

Pros:

- **Index Creation:** The software creates and utilizes its indexes, forming a crawled site database. This database organizes the indexed information in a structured format, optimizing search performance and retrieval.
- **Indexing Speed:** OpenSearch Server is designed to handle large-scale data and provides good indexing speed for structured and unstructured data.
- **Customization Options:** Users can customize stored values, indexes, scoring mechanisms, and other aspects, tailoring the search engine to specific needs.
- **Parser Capabilities:** The parsers can extract full-text data from various formats, including web pages, PDF files, office documents, and images, enhancing the search engine's versatility.
- **Administrative UI:** The Administrative UI provides a user-friendly interface for managing and configuring the search engine, including the export functionality for query results.
- **API Integration:** The ability to integrate OpenSearchServer through a RESTful API allows for seamless collaboration with other applications and systems.
- **Free to use:** Being open source, OpenSearchServer is accessible to a wide range of users, including developers, businesses, and organisations, without incurring licensing costs.



- **Compliance with Robots.txt and Noindex Rules:** OpenSearchServer respects the rules outlined in the robots.txt file and adheres to no index directives. This ensures that the scraping activities align with ethical standards and legal guidelines, promoting responsible web scraping practices.

Cons:

- **Slow Scraping:** The software may have limitations in terms of scraping speed, potentially resulting in slower data retrieval and indexing processes.
- **Risk of Internet ProtocolIP Blocking:** There is a notable risk of the server IP being blocked or receiving abuse reports, necessitating proxy services to mitigate this risk.
- **Storage Resource Requirements:** OpenSearchServer may require substantial storage resources, with examples indicating that initial indexes for multiple domains can take up significant disk space, leading to potential cost implications.

SerpApi

SerpApi is a third-party API service that extracts Search Engine Results Pages (SERPs) data. It simplifies obtaining information from search engines like Google, Bing, and Yahoo. While SerpApi is not a web crawling framework like Scrapy, it focuses on providing an easy-to-use interface for developers to access search engine results in a structured manner.

With SerpApi, developers can request HTTP API requests, specifying the search query and parameters. The API then returns structured JSON data containing search results, including titles, snippets, URLs, and other relevant information. This approach eliminates the need for complex web scraping and parsing code, as SerpApi handles the intricacies of interacting with search engines.

In terms of continuous text processing, SerpApi allows users to extract and process textual content from search results efficiently. The API response includes the text-based information in the SERPs, enabling users to analyse and utilise the extracted text for their specific needs.

SerpApi abstracts away the complexities of handling various search engine quirks, including changes in HTML structures and anti-scraping measures. It provides a straightforward solution for obtaining search engine data without maintaining and updating complex scraping scripts.

Pros:

- **Uses Google's Database:** SerpApi leverages Google's extensive database for search results, providing users with access to a wealth of data directly from Google's search engine. This ensures the reliability and comprehensiveness of the search results.
- **Instant Query Results:** One of the significant advantages of SerpApi is its capability to deliver instant query results. These results are paginated, allowing users a maximum



of ten results per page, and are ordered based on Google's relevance score. This feature ensures quick and efficient access to up-to-date and relevant information.

- **No Blocking Issues:** Unlike some other scraping solutions, SerpApi boasts a no-blocking characteristic, meaning users are less likely to encounter issues related to IP blocking or restrictions during their scraping activities. This enhances the reliability and continuity of the scraping process.
- **API Integration:** SerpApi supports seamless integration through its API. This allows developers to incorporate SerpApi's functionalities directly into their applications or systems, providing a flexible and efficient solution for various use cases.
- **Response Time:** SerpApi aims to provide fast and reliable responses for search engine results. The API is designed to minimize latency and deliver data quickly.
- **Concurrency:** SerpApi allows for concurrent requests, enabling users to fetch multiple search engine results simultaneously, enhancing performance for large-scale data extraction.
- **Caching:** The API may use caching mechanisms to improve response times further for frequently requested queries.
- **Compliance with Robots.txt and Noindex Rules:** SerpApi respects the rules outlined in the robots.txt file and adheres to no index directives. This ensures that the scraping activities align with ethical standards and legal guidelines, promoting responsible web scraping practices.

Cons:

- **Limited Intervention in Index Schema:** One limitation of SerpApi is the lack of extensive control over the index schema. Users have limited opportunities to intervene and customize the index schema according to their specific requirements. This may be a constraint for those with advanced customization needs.
- **Returns Only Allowed-to-Index Results:** SerpApi returns only results that are allowed to be indexed according to Google's policies. While this aligns with ethical practices, it may limit the types of data available for scraping, especially if the goal is to retrieve results that are not typically indexed.
- **Monthly Subscription Requirement:** To access SerpApi's services, users must subscribe to a monthly plan starting from approximately €50 per month. While this subscription model provides consistent access to the service, it introduces a recurring cost, which may be a consideration for budget-conscious users or those with sporadic scraping needs.

Comparison

Table 22 consolidates the findings of the above analysis and highlights crucial aspects that inform our decision-making process for selecting the most suitable web crawling mechanism for the INSPIRE project.



Table 22 Comparison table of OpenSearchServer, Scrapy, and SerpApi

Feature / Characteristic	OpenSearchServer	Scrapy	SerpAPI
Type	Search Engine Server	Web Crawling Framework	Search Engine API
Open Source	Yes	Yes	No
Language	Java	Python	REST API (various languages)
Scalability	Scalable	Scalable	Highly scalable
Performance	Good	Depends on configuration and hardware	Fast and reliable
Customization	Extensive	Modest	Limited
Web Crawling	Yes	Yes	No (Search API)
Data Source	Full-text search, structured data	Web scraping	Search engine results
Full-Text Extraction	Yes	Yes	Yes
Instant Query Results	No	No	Yes (Paginated)
Risk of IP Blocking	Possible	Possible	No
API Integration	Yes	Yes	Yes
Robots.txt Compliance	Yes	Dependent on Implementation	Yes
Pricing	Free and open-source	Free and open-source	A freemium model with pricing tiers



Intervention in Index Schema	Yes	Limited	Limited
Ease of Use	Moderate	Moderate to Advanced	Easy
Documentation	Comprehensive	Well-documented	Comprehensive

In addition, we have conducted benchmark tests with all tools to assess the prevalence of GEPs in a subset of the piloting dataset. This subset was manually verified to confirm the true GEP prevalence, ensuring a reliable basis for assessing the performance of the tools. This manual verification process refers to assessing whether the institutions in the pilot study (see Chapter 1) have a GEP by checking their web pages manually. The manual verification process enabled us to perform a more comprehensive comparison of the tools, specifically tailored to meet the requirements of the INSPIRE project. We specifically narrowed down our selection to three web scraping tools from many available alternatives: Scrapy, because it was specifically mentioned in the Grant Agreement (GA); OpenSearchServer, because of our team's existing expertise with Java-based solutions; and SerpAPI, which is used by She Figures (see European Commission 2021b, 168–71) and was recommended by other experts during external meetings.

Table 23 GEP prevalence with OpenServer vs SerpApi (October 2023)

	Organisations with GEPs	Detected with OpenSearchServer	Detected with SerpApi
Ireland	12	3	9
Greece	17	4	7
Germany	20	5	7
Estonia	6	3	4

Table 23 illustrates the results of the two web scraping tools, OpenSearchServer and SerpAPI, that use the same search terms. SerpAPI overperforms OpenServer in detecting GEPs in selected institutions for the pilot study. Unfortunately, when we executed Scrapy on our server, the tool crashed after running for a certain period, preventing us from obtaining reliable results. Our server provider also informed us of potential IP blocking or rate-limiting by target websites due to intensive scraping activities, which could explain the instability and crashes experienced during the execution of Scrapy. As a result, after carefully analysing the available features, performance metrics, and ease of integration, we have chosen SerpAPI as the optimal solution for our study. The decision was made with a focus on scalability, ease of implementation, and



the reliable extraction of data, ensuring that our web crawling mechanism aligns perfectly with the objectives of INSPIRE.

4.2.2 Defining Search Terms

The web scraping methodology for detecting Gender Equality Plans (GEPs) was developed through a rigorous and systematic process to ensure accuracy, relevance, and comprehensive coverage. The primary objective was to define a set of **search terms** that would reliably identify the presence of a GEP within an institution's domain while also accounting for linguistic diversity across different countries and regions. The search terms describe the words explicitly searched in a given domain. Determining the appropriate search terms is crucial to the web scraping process since searches examine not only the webpage content but also the metadata¹⁶ of the webpage, including meta keywords, meta descriptions, and meta titles. These elements enhance the precision and scope of searches, ensuring that the most relevant web pages are identified efficiently.

The methodology for determining the final set of search terms followed a structured, iterative approach. We launched our first web scraper approach by doing trials with *She figures'* search terms (European Commission 2021b)¹⁷. This method yielded many hits due to the fact that it aims to detect Gender Equality measures. During the pilot stage of web scraping, we established a core set of terms in English, including phrases such as "**gender equality plan**", "**gender action plan**," "**gender equality action plan**," and "**gender equality charter**". These terms are often used interchangeably and indicate the presence of a gender equality framework. Additionally, we considered country-specific terms that reflect national and institutional frameworks supporting gender equality initiatives, such as "**Athena Swan**" for Ireland, Cyprus and the UK.

We employed the same strategy as in the pilot study to determine the search terms for the full study, as the pilot study yielded statistically robust results for quality assessment (see Löther, Karataş, and Weber 2024a, 57–58). The search terms for the pilot countries are reported in the D.3.1 GEP Prelance Monitoring Indicator Framework v1 (see Löther, Karataş, and Weber 2024a, 97).

To determine the final search terms for the INSPIRE sample, which consists of 27 EU Member States and 6 EU Associated Countries, we followed these steps:

1. **Defining Initial Search Terms in English:** A baseline list of key terms was created, representing commonly used phrases associated with GEPs.
2. **Translating Search Terms into National Languages:** Given the multilingual nature of the dataset, search terms were translated into various languages to ensure comprehensive detection of GEPs beyond English-language content.

¹⁶Metadata on a webpage refers to information about the page's content that is not presented immediately to users while browsing the page.

¹⁷ We are grateful to Quantos SA Statistics and Information Systems, who is responsible for web-scraping of *She figures*, for collaborating by providing us with their search terms for 29 countries.



3. **Expert Verification of Translated Search Terms:** Native language experts validated the translations and ensured contextual accuracy and appropriateness for each country.
4. **Refining Search Terms:** Experts reviewed the translated terms, identifying ambiguities, removing irrelevant terms, and suggesting additional relevant terms.
5. **Finalizing Search Terms:** The finalized list incorporated multiple variations and synonyms, maximizing the likelihood of detecting relevant documents across diverse institutional websites.

One of the key challenges is the validation of search terms across 33 countries, which requires expert evaluation to ensure alignment with each country's specific context. Three sequential processes were taken to validate the search terms: (i) consulting members of the internal team, (ii) consulting members of the INSPIRE project, and (iii) consulting members of the Community of Practices (CoPs) through Knowledge Support Hub's (KSH) network. This multi-step validation process ensured that the search terms were optimized for detecting GEPs across diverse institutional websites. We successfully validated our search terms for almost all languages in the INSPIRE sample, ensuring that the methodology was both comprehensive and adaptable to regional linguistic variations.

To enhance adaptability to diverse linguistic environments, the finalized list of search terms incorporated multiple alphabets, including Latin, Greek, Hebrew, and Cyrillic. Given that different alphabets pose unique challenges in search algorithms, special attention was given to ensuring accurate character encoding and query handling. Additionally, SerpAPI, the search tool used for web scraping, was tested to confirm its capability to process different alphabets effectively. The full list of search terms for the full sample countries is provided in the annexe, section 10.2.

To maximise detection rates while maintaining search efficiency, both English- and local-language searches were prioritised where applicable. This means that in the GEP prevalence script, we used both English and national search terms to identify whether organisations had a GEP, whereas in the GEP document download script, only English search terms were used for reasons of efficiency. In the GEP prevalence script, only English is used if a country's national language is already English. In Ireland and Malta, searches were conducted exclusively in English. In Luxembourg, searches were conducted in German along with English, as it was deemed the most appropriate language for detecting GEPs in that context. In Belgium and Switzerland, search terms were used in the respective languages based on regional linguistic distinctions. For example, Dutch and French were used together with English for Belgium, while German, Italian, and French together with English were used for Switzerland.

By integrating language-specific and multi-alphabet searches, the web scraping methodology facilitated a comprehensive mapping of GEPs across institutions, providing a valuable tool for assessing the implementation of gender equality policies on a broad scale.



4.2.3 Methodology to Reveal the Indicator "Prevalence"

Following the definition of search terms, the INSPIRE web scraper was deployed to measure the prevalence of GEPs across institutional websites. The objective was to identify and quantify the presence of publicly available GEP documents, ensuring a comprehensive and accurate representation of institutional compliance with gender equality mandates.

Initially, we implemented the scraper using the national languages of the respective countries (e.g., employing Greek search terms for Greek organisations). However, although the scraper retrieved ten results per search, the search did not yield the anticipated level of accuracy in detecting GEP prevalence. In response to these findings, we refined our methodology by integrating English search terms across all countries except for those where English is already the official language. This adjustment enabled a thorough cross-linguistic analysis, allowing for better detection across multilingual datasets.

This iterative process was carefully structured to optimize search effectiveness and minimize false positives. The methodology followed a sequential four-step approach designed to refine search results progressively and ensure a structured data retrieval process. The web scraper was configured to extract up to ten results per query, repeating the process at each step until either the limit was reached or all steps were completed.

The 4-step approach

The evolution of our methodology led to the development of a refined and detailed 4-step approach, aiming to enhance precision in the search process:

1. **Initial Search with English Key Term:** Conduct an initial search using a single key term, specifically "gender equality plan" in English. If the search yields no results or fewer than 10, proceed to the next step. We decided to limit our selection to the first 10 search results to strike a balance between accurately targeting the GEPs and the performance of the scraping process. Retrieving more than 10 results per step would significantly slow down the scraping, especially as Google displays 10 search results per page by default. This choice ensured both operational efficiency and effective targeting of relevant content.
2. **Local Language Search with Key Term Translation:** If the initial English search is unproductive, search using the translated equivalent of the key term in the local language. For example, if the project is focused on a Greek organisation, the search would utilize the translated term "gender equality plan" in Greek. If the results are insufficient (less than 10), proceed to the next step.
3. **Extended Search with Additional English Terms:** Expand the search by incorporating multiple terms in English to ensure a more comprehensive exploration. This step is an alternative in cases where the initial strict search did not reveal adequate results or none. If the outcome is still unsatisfactory (less than ten results), move on to the final step.



4. **Extended Search with Additional Local Language Terms:** Similar to step 3, extend the search by introducing various terms in the local language. We designed this step to capture nuanced aspects of gender equality plans that the earlier searches may not adequately represent. If the script does not retrieve any result in all four steps, the organisation is assumed not to have a GEP.

This systematic 4-step approach was implemented to refine the search process, ensuring that targeted and alternative methods are employed to reveal the prevalence of GEPs within specific domains.

While we experimented with executing the 4-step approach in the order 1-3-2-4, we observed worse performance in identifying GEP prevalence, leading us to decide to abandon this particular sequence. Our commitment to an iterative approach allows us to adapt and refine our methods for optimal outcomes. This adaptability, demonstrated in our methodology, underscores its applicability beyond the pilot dataset, enabling us to navigate the complexities of identifying GEP prevalence across diverse linguistic and organisational landscapes.

We released INSPIRE's scraper script for GEP prevalence as an open-source tool on a public GitHub repository. This initiative was aimed at promoting transparency, reproducibility, and collaborative improvements to the methodology. The repository can be accessed via the following link: <https://github.com/InspireQualityeu/gep-scraper> (the script can be found in /webcrawler/crawler-prevalence.php, while technical details can be found in the README.md file and in annexe section 10.8).

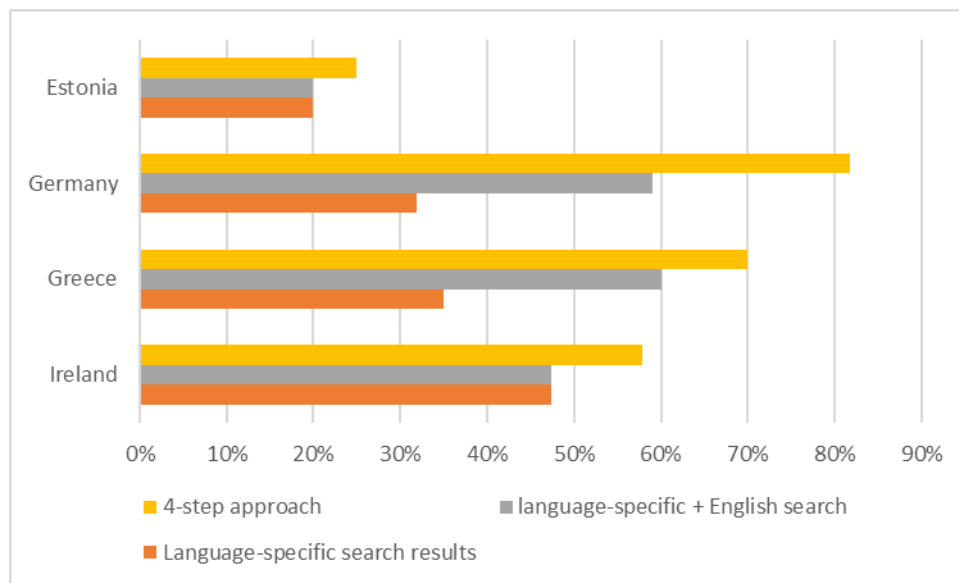
Results of the 4-step approach

Hereby, we present a comprehensive analysis of the GEP prevalence using the three distinct approaches: language-specific searches, language-specific combined with English searches, and the refined 4-step approach (see Figure 5). We examine the effectiveness of these methodologies in extracting relevant information from the pilot dataset of the four European countries: Ireland, Greece, Germany, and Estonia. For each country, we manually detected:

- i. The number of language-specific search results
- ii. Results from language-specific combined with English searches
- iii. Results obtained through a refined 4-step approach revealed the existence of a GEP within an organisation.



Figure 5 GEP prevalence from the incremental steps of the INSPIRE methodology (October – November 2023).



The percentage values represent the proportion of relevant results among the total number of organisations.

Key Findings:

- **Language-Specific Searches:** The 4-step approach consistently outperforms traditional language-specific searches in all countries.
- **Language-Specific + English Searches:** Greece and Germany significantly improve information retrieval when combining language-specific and English searches, emphasizing the importance of multilingual queries.
- **4-Step Approach:** The 4-step approach yields the highest percentage of relevant results across all countries, showcasing its effectiveness in extracting targeted information.

This comparative analysis underscores the importance of a tailored approach to language-specific information retrieval. The 4-step methodology is a robust strategy, while the combination of language-specific and English searches shows promise, particularly in multilingual contexts. As information retrieval methods evolve, understanding these nuances becomes crucial for optimizing search outcomes in diverse linguistic landscapes.

Quality assessment of the INSPIRE scraper for the Indicator "Prevalence"

We assessed the quality of INSPIRE's web scraper in gathering data on the "prevalence" indicator in four pilot European countries: Ireland, Greece, Germany, and Estonia. The quality assessment involved comparing manually detected GEPs with those obtained through automated scraping to gain insights into auto-scraping efficiency for this indicator. We first aimed to manually determine whether 83 organisations in our pilot sample (see section 2.1)



had a GEP. Then, we performed 4-step web scraping on the same 83 organisations. Finally, we obtained results from both approaches to compare their results and assess the success of INSPIRE's web scraper. Manual detection revealed that 55 out of 83 organisations have a GEP, accounting for 66%. We assessed the success of INSPIRE's scraper using four measures from the evaluation metric: *accuracy*, *sensitivity*, *specificity*, and *precision*. These measures are widely used in information retrieval, machine learning, data mining, and medical diagnostics (Das et al. 2019; Hossin and Sulaiman 2015; Powers 2020; Wang 2022). Based on the evaluation metrics used, we determined the success of INSPIRE's scraper.

Table 24 A comparison: web scraper vs manual detection

		Truth (as assessed by us manually)	
		Positive	Negative
Scraping result	Positive	$N_{tp} = 55$	$N_{fp} = 7$
	Negative	$N_{fn} = 0$	$N_{tn} = 21$

- N_{tp} refers to the number of true positives.
- N_{fp} refers to the number of false positives.
- N_{tn} refers to the number of true negatives.
- N_{fn} refers to the number of false negatives.

- i. The first measure of evaluation metrics is *accuracy*, which quantifies the overall accuracy of a model or test, computed as the ratio of correct results (including true positives and true negatives) to all cases assessed. Accuracy is a straightforward and easy-to-understand performance measure, giving a rapid overview of how well a model performs for both positive and negative categories.

$$\text{Accuracy} = \frac{N_{tp} + N_{tn}}{N_{tp} + N_{tn} + N_{fp} + N_{fn}}$$

- ii. The second measure of evaluation metrics is *sensitivity*, which determines the likelihood that the organisation has a GEP and how effectively the scraper can detect this.

$$\text{Sensitivity} = \frac{N_{tp}}{N_{tp} + N_{fn}}$$



- iii. The third measure of evaluation metrics is *specificity*, which assesses the likelihood of the scraper detecting an absence of a GEP if the organisation does not have one.

Specificity = $\frac{N_{tn}}{N_{tn}+N_{fp}}$

- iv. The fourth measure of evaluation metrics is *precision*, which indicates the proportion of correctly identified positive cases out of all those flagged as positive by the scraping process.

Precision = $\frac{N_{tp}}{N_{tp}+N_{fp}}$

Table 25 Performance metrics of INSPIRE’s scraper for prevalence of GEPs

Measurement	Accuracy	Sensitivity	Specificity	Precision
Statistics	92%	100%	75%	89%

As Table 25 shows, INSPIRE’s scraper exhibits high accuracy with an overall accuracy rating of 0.92. This figure shows a high level of correctness in the scraper's operations. With a precision rate of 0.89, the scraper is quite effective at ensuring that 89% of the retrieved results it identifies as GEPs are actually relevant, demonstrating its ability to sift through large volumes of online data to extract relevant documents. The specificity score is measured at 0.75, showing that the scraper effectively filters out 75% of non-GEP content, thus eliminating significant irrelevant data. Notably, INSPIRE’s scraper achieved a sensitivity (or recall) rate of 1.00, meaning it successfully identified all actual GEPs in the dataset without missing any relevant documents. These metrics collectively illustrate the robustness and effectiveness of the scraper in identifying GEPs, highlighting its value as an essential tool for researchers and policymakers focusing on gender equality initiatives.

Operationalisation of methodology to reveal the indicator "prevalence"

After obtaining a reliable statistical measure of quality in the pilot study, we applied the methodology to the full dataset of 6,571 web pages to determine the "prevalence" indicator (see also Löther, Karataş, and Weber 2024b).

During implementation, an important adjustment was made: Search terms were systematically split into multiple queries for specific languages due to Google Search's limitations of 32 words or 2,048 characters per query. This limitation necessitated strategic adjustments, particularly for languages with longer word constructions or more complex grammatical structures. The languages affected by this constraint included Bosnian, Bulgarian, French, Czech, Croatian,



Polish, Portuguese, Serbian, and Slovenian. In these cases, search queries were split into smaller batches to ensure that all relevant terms could be applied while staying within the imposed limits. This adjustment was essential to accommodate the extensive range of search terms required to capture the diverse terminological expressions of GEPs across different linguistic and cultural contexts. While this approach increased the number of individual search queries, it enabled a more granular and precise exploration of GEP prevalence. These refinements collectively aimed to enhance data accuracy, reduce false positives, and ensure a robust and comprehensive assessment of GEP prevalence across institutional websites.

4.2.4 Methodology to Download the GEPs

The initial phase of our methodology involved the utilisation of a script designed to measure the prevalence of GEPs from websites within the piloting dataset. To analyse the GEP structure and content, we recognised the need to enhance our approach by incorporating a more intricate layer – retrieving associated documents containing detailed GEP information. Given that the text analysis tool can analyse PDFs, we targeted our refinement process to PDF only, leaving aside the rest of the types of files that could also include detailed GEP information, i.e. HTML or .doc files.

This advanced script goes beyond web content extraction and is now able to download PDF documents that match identified GEPs. This innovative script harnesses the power of Google search, using the 'filetype:pdf' filter and GEP-related terms to accurately target and download relevant PDF documents.

The script initiates the process by conducting targeted *Google* searches employing the specified file type filter, narrowing the search results to PDF documents. Simultaneously, the script integrates the GEP-related terms similarly to the prevalence script to ensure a refined focus on documents pertinent to GEPs. This strategic combination streamlines the search process and enhances the accuracy of identifying and retrieving GEP-related PDFs.

The iterative nature of the script ensures a meticulous verification of GEP presence on websites before proceeding to the PDF download phase. This two-step process enhances the reliability of the data collected, as it confirms the alignment between the identified GEP terms on websites and the contents of the downloaded PDFs.

By harnessing the capabilities of Google searches with tailored filters, our script enhances the efficiency and precision of data extraction, aligning with the evolving landscape of web-based information retrieval. This refined approach signifies a technological leap in our methodology and underscores our commitment to capturing a comprehensive and nuanced understanding of GEPs within the INSPIRE project's scope.

As part of our ongoing optimisation efforts, we fine-tuned the 4-step approach by strategically altering the order of search terms in steps 3 and 4. Recognising the dynamic nature of web searches, this adjustment aimed to enhance the precision and effectiveness of the INSPIRE web scraper. By experimenting with different arrangements of search terms, we sought to uncover patterns that could yield more accurate and relevant results. Indeed, the modification



in the order of German search terms yielded a noteworthy enhancement in the success rate for German scraping, resulting in a substantial improvement of 20%. This iterative process reflects our commitment to continually improving the methodology's adaptability to diverse linguistic and organisational landscapes.

Another significant fine-tuning involved broadening the search scope by incorporating Google suggestions without using quotation marks. By omitting quotation marks, we aimed to increase the flexibility of the search process, enabling the scraper to adapt to variations in language usage and capture a more comprehensive set of relevant results. This fine-tuning proved particularly effective in countries with more complex grammar and syntax in their native language, ultimately improving performance. This refinement aligns with our commitment to staying responsive to evolving search dynamics and maximising data extraction accuracy in the INSPIRE project.

We released INSPIRE's scraper script for downloading the GEPs as an open-source tool on a public GitHub repository. This initiative aimed to promote transparency, reproducibility, and collaborative improvements to the methodology. The repository can be accessed via the link: <https://github.com/InspireQualityeu/gep-scraper> (the script can be found in /webcrawler/crawler.php, while technical details can be found in the README.md file and in annexe section 10.8).

Result of the 4-step approach for downloading GEPs

We conducted a comprehensive evaluation of the performance of each step within the 4-step approach for the pilot study, which was composed of 83 organisations from Ireland, Greece, Germany and Estonia. The detailed breakdown in Table 26 offers a nuanced understanding of the number of identified GEPs at each stage of the data extraction process for every country in our analysis.

Table 26 PDFs extracted by each step of the 4-step approach.

	Step 1	Step 2	Step 3	Step 4
Ireland	5	N/A	6	N/A
Greece	6	6	1	1
Germany	11	3	0	5
Estonia	2	0	0	1
Contribution	44%	17%	13%	13%



Table 26 shows the contribution of each step to the 4-step approach, i.e., how many PDFs have been extracted from each step.

Key findings:

- **Variability in results:** There is notable variability in the performance of each step across countries.
- **Local Language Search Impact (steps 2 and 4):** The local language search (simple and extended) demonstrates a significant impact on results, particularly in Greece (via step 2) and Germany (via step 4), where there is a notable increase in the GEP pdf inclusion.
- **Dependency on Language and Cultural Context:** The results underscore the importance of tailoring the methodology to each country's linguistic and cultural contexts, as observed in the impact of local language searches.

These findings demonstrate that the effectiveness of the four-step information retrieval process varies across the examined European countries. All four steps are required depending on the case, emphasizing the need for a tailored and adaptable approach when extracting data related to GEPs across diverse linguistic and organisational landscapes.

Quality check for downloading the GEPs

We comprehensively analyse the PDF retrieval success rates across the four piloting European countries: Ireland, Greece, Germany, and Estonia. The data includes manually detected PDFs of GEPs and those obtained through web scraping, providing insights into the efficiency of web scraping in capturing relevant documents. The success rate is defined by whether web-scraped PDFs contain real GEPs.

Key Findings:

- **High Success Rates:** Ireland and Greece showcase high success rates of 92% and 94%, respectively, indicating the effectiveness of automated scraping in capturing the majority of relevant PDFs. Germany demonstrates an 86% success rate, suggesting that while auto-scraping is effective, additional factors may influence document availability.
- **Challenges in Estonia:** Estonia experiences a lower success rate of 60%, highlighting potential challenges in automated scraping. This is attributed to the non-indexing of PDFs in Google, revealing a specific hurdle in retrieving relevant information through automated means.

This analysis provides valuable insights into the success rates of automated scraping in capturing relevant PDFs across different European countries. While Ireland and Greece demonstrate robust performance, Germany presents a slightly lower success rate, and Estonia faces challenges, indicating the importance of considering regional variations in information retrieval methodologies.



To check the quality of results and assess the algorithm's performance, we use the same four measures from the evaluation metric Table 27: accuracy, sensitivity, specificity, and precision. Note that the web scraping algorithm returns multiple PDFs per organization and cannot distinguish which one is the GEP. When calculating the statistics, we take into account whether a GEP was downloaded or not. Therefore, the calculated statistics are low. In particular, the accuracy rate is 17 percent, indicating a low overall success rate of the scraper in downloading GEPs. This is mainly due to the presence of more than one document per institution. We introduced an intermediate step—PDF classification—and, based on the different approaches adopted, calculated quality statistics. The PDF classification process is explained in detail in section 4.3.2.

Table 27 *PDF extraction results (January 2024)*

Measurement	Accuracy	Sensitivity	Specificity	Precision
Statistics	17%	59%	10%	10%

Operationalisation of methodology to download GEPs

A crucial aspect of operationalizing the methodology was streamlining the document retrieval process to enhance computational efficiency. By concentrating exclusively on English-language documents, we facilitated quicker processing times and optimized resource utilization. This strategic decision enabled the management of a large dataset comprising 6,571 institutional domains while maintaining a practical balance between accuracy and computational feasibility. The focus on English PDFs also significantly simplified subsequent tasks, such as metadata extraction, filename parsing, and content analysis, ensuring that the retrieved documents could be systematically processed and categorized with minimal additional effort.

Expanding the linguistic scope to include multiple languages would have introduced substantial complexities, particularly in terms of query formulation, document classification, and multilingual text processing. The increased operational burden and computational demands associated with such an expansion would have significantly lengthened the analysis phase. To assess the feasibility of incorporating multilingual searches, we conducted experimental searches using native languages for institutions in Germany, Spain, and France. However, these tests did not yield substantial improvements in retrieval rates. Furthermore, the additional processing requirements imposed by multi-language queries resulted in a considerable increase in computational load and resource consumption. Given the lack of significant added value relative to the increased cost and complexity, we determined that expanding the linguistic scope would not represent a cost-effective enhancement to the methodology.

As a result, English-language documents were given precedence in the final implementation of the scraper. This approach ensured that the methodology remained scalable, efficient, and capable of delivering reliable results without overburdening computational systems. The



refined scraping methodology thus effectively balanced accuracy, efficiency, and feasibility, reinforcing the robustness of GEP document retrieval across institutional websites.

All in all, INSPIRE's scraper downloaded 7,018 PDFs from 6,571 web pages.

4.3 Monitoring Characteristics, Implementation and Impact of GEPs: Classifying PDFs and Text Analysis

We aimed to investigate GEPs beyond the prevalence indicator (see section 3.3). Our analysis aims to retrieve information for the INSPIRE indicators regarding the characteristics, implementation and impact of the extracted GEPs. Text analysis is used in two different manners: first, to **classify GEPs** among the retrieved PDFs, and second, **to extract information** from GEPs to derive INSPIRE indicators. An intermediate step, the classification of PDFs, is necessary because the web scraping carried out in the pilot study produced a considerable corpus of data with some inappropriate files. Table 27 demonstrates that the accuracy of the extracted PDFs is relatively low and there are more than one PDF per institution. Many false positives make identifying the true GEP for each institution challenging. Using web scraping alone does not suffice to obtain one accurate PDF per institution. For this reason, classification process have been developed to get one PDF (i.e., GEP) per institution. During the pilot phase, we used two different classification methods: zero-shot classification using LLMs and metadata analysis. After a thorough comparison of the two approaches, we chose metadata analysis for efficiency reasons. Section 4.3.2 explains the classification issue in more detail and presents quality statistics obtained during the pilot study. Finally, a rule-based approach was used to extract information for the INSPIRE indicators. The tools used, the algorithm development process, the quality statistics from the pilot study and the results of the full study are presented in detail in Section 4.3.3.

4.3.1 Text Analysis as a Methodological Approach

Recent sociological approaches have already used well-established methodological practices for text analysis, such as quantitative text analysis (e.g., dictionary-based procedures) and qualitative content analysis (Macanovic 2022; Spörlein and Schlueter 2021). The field of Natural Language Processing and machine learning have brought up new opportunities to automate text analysis methods in social sciences due to proven efficiency concerning time and cost.

Computational text analysis methods (CTAM) are an umbrella term for many approaches to analysing digital data. These methods differ in the way they involve techniques for extracting specific information using keywords or formatting guidelines and advanced software solutions such as BART, GPT, and other large language models (Baden et al. 2022; Mazel 2023). Indeed, some of these techniques, such as Information Retrieval (IR), have already been used in computer science. The information retrieval process entails retrieving unstructured content; usually text-based documents, that matches the information needs from extensive collections often kept on computers (Manning, Raghavan, and Schütze 2012). Unstructured text mainly presents challenges for researchers to analyse. It is defined by the absence of a predictable



framework or order, the use of natural language, which may be imprecise and complex, and variances in context, style, and syntax (Inmon and Nesavich 2007; Manning, Raghavan, and Schütze 2012). Recently developed tools now allow for content extraction from PDFs with significantly reduced data loss (Meuschke et al. 2023).

4.3.2 Classification of PDFs

With a dataset comprising 6,571 institutional domains and a substantial volume of downloaded PDFs (7018), including a significant number of false positives, identifying the true GEP for each institution is challenging. This challenge is compounded by the fact that many of the institutions' websites do not follow a standardized structure, which complicates the process of retrieving the correct document. Each institution's website is unique in layout and design, which means the location and presentation of the target GEP document can vary drastically.

Additionally, the metadata associated with these documents is often insufficiently detailed to support accurate classification. Metadata, in many cases, is sparse or inconsistent, further hindering the ability to automatically identify the right GEP document for each institution. On top of this, the presence of multiple languages across the websites introduces further complications, as the data has not been standardized or pre-processed to account for linguistic diversity. This makes the classification task even more difficult, as the same document could be presented in different languages or formats across different institutions, and without a standardized system, automated classification processes struggle to make accurate decisions.

To address these challenges, a robust and multi-faceted methodology was essential to accurately identify the relevant GEP for each institution. Simply relying on web scraping alone was insufficient, as web scraping tends to gather as many documents as possible that meet the predefined criteria without taking into account whether these documents are truly the relevant GEP for that institution. This issue of false positives, or irrelevant documents, is a significant barrier in efficiently retrieving the correct data.

In response to this, we introduced an intermediate automatic text analysis algorithm and a classification process designed to refine the search and help pinpoint one specific PDF—namely, the GEP document—per institution before proceeding to the text analysis phase. This provided a necessary step between raw data collection and final document classification. In particular, we tested two approaches to classify the documents: a) Zero-shot classification and b) Metadata analysis.

Zero-shot Classification with BART and GPT

Text classification can be done through different approaches, such as supervised machine learning and Large Language Models (LLMs). A recent study showed that LLMs outperformed traditional machine-learning models for text classification tasks (Chae and Davidson 2023). Advanced LLMs have a high amount of representational richness and the ability to generalise, allowing them to effectively complete new tasks without additional training. This capacity is usually known as zero-shot learning (Brown et al. 2020; Chae and Davidson 2023; Socher et al. 2013). Zero-shot learning is advantageous over supervised machine learning models, which



do not require training data. Developing a training dataset and testing a classifier can be costly and time-consuming. Yet, advanced LLMs can be modified with a few lines of code to achieve good performance with a few high-quality instances (Do, Ollion, and Shen 2022; Chae and Davidson 2023). Among many others, some of the models have proven a great success for zero-shot classification, namely, Generative Pretrained Transformer (GPT) (Brown et al. 2020; Ouyang et al. 2022), Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2018) and Bidirectional and Auto-Regressive Transformers (BART) (Lewis et al. 2019).

We performed some predictions with these models' variations; namely, we used GPT-3.5 turbo instruct and the BART Facebook/bart-large-mnli model¹⁸. Although GPT-4 is more capable than the GPT-3.5 turbo instruct model, it has a chat endpoint, which means GPT-4 can not provide predicted probabilities for categories. Predicted probabilities linked to the labels enable adjustment of thresholds for predicted labels and facilitate selection among the predicted labels for each organisation. We compared the two large language models (candidate models), which is displayed in Table 28.

¹⁸ See [facebook/bart-large-mnli](https://huggingface.co/facebook/bart-large-mnli) · Hugging Face



Table 28 Comparison of LLMs

	GPT - 3.5 turbo instruct	BART – facebook/bart-large-mnli
Resources	Uses OpenAI server. No need for additional computational sources.	Source intensive. Needs powerful computational sources.
Cost	Cost for API 0,003 Euro per 1K token for GPT 3.5 Turbo instruct model	Using BART is free because It is an open-source project.
Time	Since its server is used, it has been pretty fast. Also, GPT turbo models are optimised for their speed. It takes 3-4 minutes to classify ~700 rows (first 200 words of PDFs).	The model is cumbersome. Although Google Colab is used, it took ~120 minutes to classify ~700 rows (first 200 words of PDFs).
Language/s	It is trained in many languages. Its capacity in English is much better. It is strong in understanding other widely spoken languages such as Spanish, French, and German. However, it has an understanding of many European languages. There is no published list for this.	Only trained in English; therefore, the text should be translated into English for other languages. This requires using other LLMs, which are source-intensive as well.
Prompting	It is highly flexible for prompting. It is designed for prompting requests. The result of classification can be improved via efficient prompting.	It is not naturally designed for prompting. However, prompting can be adopted.
Understanding context	GPT is designed for chat interaction and prediction of the next words. However, it has an excellent understanding of the context.	BART is naturally designed to understand the content of the text.
Limit	The GPT 3.5 Turbo instruct model currently has a 4096 token limit per request. Which means we cannot send longer texts to make predictions. 4096 tokens are equivalent to ~3000 words.	512 tokens per request. Equivalent to 450 words.

The GPT 3.5 turbo instruct model is preferable due to its speed and less need for computational capacity. However, this model can only be used when purchasing tokens from Open AI.

We evaluated the performance of the BART and GPT 3.5 models. For both models, processing the entire document for analysis requires high computational and time resources considering the final study, which aims to contain 6,571 institutions. Therefore, we concentrated on processing only some portion of each retrieved document to decrease the required computational resources. Our first approach was to process each PDF's first 1-2 pages.



Usually, each document has a complete title referring to "gender equality plans". However, this approach did not provide the target information as some documents contain an opening page or an initial page, which is not a text and is not entitled to a "gender equality plan". We then changed our approach to retrieving the first 100 words of each PDF. This strategy produced more robust results, indicating that the information collected for each PDF is more consistent across the sample.

For each model, an intermediate step, preprocessing the text data, was necessary and carried out differently due to the models' different natures. Zero-shot classification using the BART model is performed in Google Colab¹⁹. For the **BART** model, the following preprocessing steps were performed:

- i. Reading each PDF and extracting the first 100 words.
- ii. The text is preprocessed, and duplicate rows are eliminated.
- iii. The first 100 words were translated into English via LLMs.
- iv. We created defined candidate labels based on existing PDFs.
- v. We asked the machine to classify the PDFs based on the content of the 100 words of each PDF using the BART model.
- vi. The model produces an estimated label and its associated probability.
- vii. Decisions are made based on predicted labels and probabilities.

Where we define candidate labels as follows for the BART model:

- Gender Equality Initiatives
- Educational Program Details
- University Administrative Records
- Research and Development
- Academic Publications and Materials
- Strategic Development Plans
- Legal and Policy Documents
- Student and Faculty Resources

¹⁹Google Colaboratory, is a free cloud-based platform provided by Google.



- Public Health and Nutrition Studies
- International Relations and Agreements

For the GPT model, the preprocessing steps were slightly different. The main difference was that we did not translate the text into English since GPT can understand German, Greek, and Estonian. The preprocessing steps for the **GPT** model are defined as follows:

- i. Reading each PDF and extracting the first 100 words.
- ii. The text is preprocessed, and duplicate rows are eliminated.
- iii. We asked the machine to classify based on the content of the 100 words of each PDF using the GPT 3.5 Instruct model.
- iv. The classification task is carried out via the prompt.
- v. The model produces an estimated binary (Yes/No) category and its associated probability.
- vi. Decisions are made based on predicted labels and probabilities.

The prompt is defined as follows:

“Is the following text part of a Gender Equality (Action) Plan or not? A Gender Equality Plan (GEP) is a set of commitments and actions that aim to promote gender equality in an organisation through a process of structural change, particularly in workplaces, educational institutions, and public bodies. Provide "Yes" or "No" as an answer.”

Table 29 displays a comparison of the BART and GPT models' performance.



Table 29 Evaluation metrics of the BART and the GPT 3.5

Country	Model	Precision	Sensitivity	Accuracy	Specificity
Ireland	BART	75%	67%	64%	60%
Ireland	GPT 3.5	38%	100%	47%	20%
Greece	BART	100%	60%	60%	N/A
Greece	GPT 3.5	100%	33%	33%	N/A
Germany	BART	100%	82%	82%	N/A
Germany	GPT 3.5	71%	100%	72%	17%
Estonia	BART	100%	40%	77%	100%
Estonia	GPT 3.5	100%	29%	62%	100%
Total	BART	94%	62%	71%	80%
Total	GPT 3.5	77%	65%	53%	46%

As Table 29 shows, BART and GPT models work well to distinguish gender equality plans/initiatives from other documents, such as educational programs, administrative records, research and development papers, and academic publications and materials. However, the model cannot predict the targeted GEPs correctly because many documents are similar to GEPs.

One of the assumptions made while evaluating the quality of LLMs is that the presence of departmental GEP is considered a true positive for the organization. Distinguishing between different types of GEPs is still highly challenging for the model. One of our objectives is to optimize the predictions for finding true GEPs per institution in the classification step. Based on our above results, the BART model overperforms the GPT 3.5 model in our specific setting. However, the BART model needs high computational capacity. Changing the prompt of the GPT 3.5 turbo instruct model, defining better candidate labels for the BART model or improving the hardware/software equipment (server capacity) could make the process better.

Metadata analysis

The metadata-based classification approach was implemented through a structured process focused on enhancing document relevance and reducing false positives.



Metadata analysis comprises several steps (Löther, Karataş, and Weber 2024, 68–70). First, we analyzed file indicators, including the filename, the first page of the document, and other embedded metadata (e.g., document title, author information, keywords) to identify documents containing relevant terms. Subsequently, keyword extraction techniques were applied specifically to the first page of each PDF to further assess its relevance. The list of search terms used for the metadata analysis, tailored to each country's specific terminology, can be found in section 10.3. This list of terms played a crucial role in filtering out irrelevant documents. When multiple GEP documents were detected for a single institution, the first document identified was assumed to be the official version.

The approach used several specific steps to refine the classification process and achieve the desired results:

1. **Focus on English PDFs:** By narrowing our focus to English documents, we streamline the subsequent stages, optimizing efficiency and resource utilization.
2. **Filtering PDF filenames based on specific terms:** To further narrow down the pool of documents, we implemented a filtering mechanism based on specific terms within the filenames of downloaded PDFs. The filenames were scrutinized for keywords directly related to GEPs, such as "Gender Equality Plan," "GEP," "Action Plan," "Gender Equality Action Plan," and "GEAP" (both in English and in combinations of English and various local languages). Any files that lacked these crucial terms were excluded from the dataset, ensuring that only PDFs that were likely to contain relevant GEP information were retained.
3. **Keyword search on the first page of PDFs:** Another integral step of our classification process involved a detailed examination of the first page of each downloaded PDF. By employing keyword searches specific to Gender Equality Plans, similar to the terms mentioned above, we can verify the content's relevance and legitimacy. This step is essential for discerning whether the downloaded files genuinely pertained to GEPs, contributing to a higher degree of accuracy in our dataset.
4. **Filtering based on the metadata of the PDFs:** Another step towards classification involves scrutinizing the metadata of each downloaded PDF, such as document title, author information, creation date, and keywords, to identify and retain documents with relevant GEP information.

Table 30 and Table 31 report the quality assessment of the meta-analysis based on the pilot countries.



Table 30 Performance metrics of the English-only approach

	Precision	Sensitivity	Accuracy	Specificity
Ireland	60%	75%	68%	64%
Greece	73%	62%	58%	50%
Germany	75%	50%	52%	57%
Estonia	100%	60%	87%	100%
Total	77%	64%	67%	67%

Table 31 Performance metrics of English plus local languages approach

	Precision	Sensitivity	Accuracy	Specificity
Ireland²⁰	60%	75%	68%	64%
Greece	79%	79%	70%	50%
Germany	61%	73%	58%	36%
Estonia	100%	60%	87%	100%
Total	75%	71.75%	70.75%	62.5%

The classification task was carried out using exclusively English search terms and a mix of English and local language search terms. As Table 30 and Table 31 show, superior outcomes were achieved with the combined approach, and Table 31 displays its performance metrics.

Operationalisation of PDF classification for the full sample

As shown in the quality assessment of the alternative classification methods in the previous section, the BART model outperformed GPT 3.5 and metadata analysis in precision (94%), accuracy (71%) and specificity (80%), while metadata analysis outperformed BART and GPT

²⁰ For Ireland, only English search terms used.



3.5 in sensitivity (71.75%). In addition, metadata analysis outperformed GPT 3.5 in all metrics except precision.

The metadata analysis method proved to be more resource-efficient while maintaining high classification accuracy, and it was ultimately adopted. The main reason is the availability of resources, as Large Language Models (LLMs) demand either a graphics processing unit (GPU) or a token-wise subscription to use their API models. In addition, optimizing the LLM through prompt engineering would take considerable time. All in all, there are no significant differences between the quality assessments of zero-shot classification and metadata analysis. However, with the appropriate resources, LLM has the potential to achieve much higher statistics.

We created a country-specific search terms for metadata analysis. The search terms are created on the basis of a manual assessment of the GEPs for 33 sample countries for the full study. The full list of search terms can be found in annexe (section 10.3).

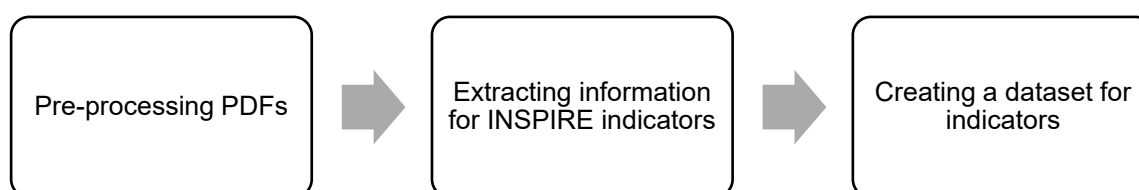
INSPIRE's scraper yielded 7,018 PDFs for the full sample study. After conducting metadata analysis, the number was decreased to 1,564. Thus, 1,564 are to be preprocessed and analysed.

4.3.3 Developing the Algorithm for the Text Analysis

The full study aims to explore a text analysis method that proves to be the best automated approach to retrieve information for INSPIRE indicators in the full sample. Our text analysis process has changed slightly from the pilot study (text analyses process of the pilot study see Löther, Karataş, and Weber 2024a, 70). The main change is that the translation of non-English PDFs into English has been omitted. This is because the INSPIRE scraper for downloading the GEP is only performed in English for efficiency reasons. (see section 4.2.4).

Our text analysis for the full sample comprises three steps, as demonstrated in Figure 6.

Figure 6 Text analysis process



Preprocessing PDFs

Processing text for analysis using the Portable Document Format (PDF) is not straightforward. Tables in PDFs can be challenging to work with due to the embedded tabular structure (Rastan, Paik, and Shepherd 2019), text alignment, and formatting details. Moreover, PDF



files can also contain images, footnotes, and figures, which makes it difficult for software or other assistive technologies to process information. Given that downloaded GEPs in PDF form do not share a uniform structure, extracting the texts for analysing purposes poses challenges. To overcome this challenge, we use a **Python**²¹ library called "**pdfplumber**"²² to extract the text from GEPs in PDF form. However, pdfplumber works very well with machine-generated PDFs rather than scanned PDFs. However, we discovered that some of the PDFs are scanned images rather than computer-generated documents. To overcome this issue, we included a drawback function in the script to convert scanned images (OCR-processed) to .txt format using the "**tesseract-ocr**"²³ library (Smith 2007) if the "**pdfplumber**" fails. The script is designed with a robust implementation capable of processing both text-based PDFs and scanned PDFs containing image-based text. However, we have identified certain corrupted PDFs that could not be converted to text.

The GEPs were stored in separate subfolders for each organization within each country in the full sample, resulting in a total of 33 country-specific folders. The "**os**"²⁴ library in Python allowed us to navigate through subfolders and extract GEPs by specifying their parent directory while retaining information about the corresponding organizations within each country.

Furthermore, the dataset we are working with is accessed via Python, demonstrating the programming language's flexibility and ease of use when dealing with multiple data formats. However, the existing format of our data poses substantial challenges to efficient operation. The lack of structure, in particular, limits our ability to apply analytical functions directly, change the information, or even execute fundamental data exploration tasks with the speed and efficiency we require. In Python, structured data formats, such as data frames, provide various benefits, including faster data access, more straightforward data cleaning and transformation processes, and the ability to apply complicated analytical algorithms with relatively simple code. Therefore, we created a data frame using Python's "**pandas**"²⁵ library. Reading PDF documents presents significant challenges and requires a substantial amount of time. As the final goal of text processing, we saved each PDF as a plain text file (.txt) to determine the optimal method for extracting information in the following phase because Python is substantially faster at processing basic text files.

We noticed that some of the PDFs retrieved by the INSPIRE scraper were not in English, despite the fact that the search query for downloading PDFs was in English only. To filter out non-English GEPs, we used Python's "**langdetect**"²⁶ library to detect the GEPs' language automatically. *Langdetec* library was developed by Shuyo. The library predicted the language

²¹ Python is a popular interpreted programming language that emphasizes readability, flexibility, and simplicity. It has since become prevalent in many domains, including artificial intelligence, scientific computing, data analysis, and web development.

²² See [pdfplumber · PyPI](#)

²³ <https://github.com/tesseract-ocr/tesseract>

²⁴ OS module provides portable operating system-dependent functionality (see [os — Miscellaneous operating system interfaces — Python 3.12.1 documentation](#))

²⁵ Pandas is an effective open-source tool for data analysis and manipulation (see [pandas - Python Data Analysis Library \(pydata.org\)](#))

²⁶ See [langdetect · PyPI](#)



of the GEPs with a 100% success rate in the pilot study. Indeed, Langdetec Library has been proven to have a more than 99% success rate in 49 languages (Shuyo 2010).

We conducted a thorough comparative analysis of available automated translation tools for translating GEPs, highlighting their benefits and limitations in the **D.3.1 Report GEP Prevalence Monitoring Indicator Framework v1** (see Löther, Karataş, and Weber 2024a, 72–74).

The preprocessing steps are outlined below:

- Python receives PDF files from the specified directory.
- Extracts the text from each page for each PDF.
- Concatenates text from every page for each PDF.
- Creates a data frame containing PDF names, PDF texts, organization URLs, language of the text, and belonging country.
- Filters out non-English GEPs
- Saves each PDF as a plain text file (.txt).

Extracting Information for INSPIRE Indicators

After transforming all GEPs from PDF to plain text format, we extracted information for the INSPIRE indicators (see section 3.3) by applying text analysis techniques. Our information extraction strategy for text analysis relies on searching for specific terms or patterns within the GEPs. **Regex**²⁷, a Python library, is employed for text analysis.

A regular expression (often abbreviated as regex or regexp) is a sequence of characters that defines a search pattern. It is mainly used for string matching and manipulation. Regular expressions are a powerful tool in text processing (Mitkov 2022). In our situation, we implemented regex in two distinct manners: matching strings and patterns. To do string matching, we defined specific search terms for our indicators. String matching can be approached in different ways. Standard string matching focuses on identifying a particular pattern within a larger text regardless of its context or word boundaries. On the other hand, standalone string matching requires that the pattern be recognized as an independent word and separated by word boundaries from other text elements, ensuring more contextual sensitivity in search results. For instance, consider a text saying, "Recently, age management gained popularity," simple string matching for "age" would find matches within "age" and "management", focusing only on the sequence of letters without considering context or word boundaries. Whereas standalone string matching would only match with the word "age". Both standard and standalone string matches are used to extract information for the INSPIRE indicators, depending on the case.

²⁷ [re — Regular expression operations — Python 3.12.2 documentation](#)



Afterwards, we developed a function for each indicator, which follows the subsequent reasoning:

- i. The function transforms specific words or phrases into regular expression patterns, normal or standalone string-matching.
- ii. It changes the entire text to lowercase to enable case-insensitive searching.
- iii. The lowered text is then scanned for instances of each regular expression pattern.
- iv. The identified words or phrases are collected in a set to detect any matches.
- v. If at least one match is found, it returns 'Yes'; otherwise, it returns 'No'.

Finally, each function is applied to the GEPs within a pandas DataFrame in Python, leading to systematic processing of the DataFrame on a row-by-row basis. In other words, GEPs are the input for the created functions. By processing the DataFrame row by row, each function can efficiently analyse and extract information for INSPIRE indicators.

Quality assessment and results of INSPIRE indicators

The tables below demonstrate the INSPIRE indicators, the methods and tools employed, their respective success rates, and the results. Quality assessment based on 44 GEPs of the pilot study. Note that the quality assessment (i.e., success rate) was only performed for the pilot study, as it would be very time-consuming to perform on a large volume of documents. After achieving good statistics in the pilot study, we implemented the same information extraction script for the full sample study. We analysed 1,135 GEPs in the full sample study. The last columns in the tables show the mean of each indicator based on these GEPs. The specified search terms for each indicator are available in the annexe (see section 10.4). For a more detailed analysis, including cross-tabulations by the cluster of countries and by type of organization, please refer to the publication **D 3.3 Report on GEP Monitoring Survey + Webcrawl Results** (see Löther, Karataş, and Weber 2024b).



Prevalence

Table 32 Information extraction for prevalence indicator: Horizon Europe

Indicator	Tool-method used	Success rate in %	Result in % (Mean)
Indicator 1.2 The existence of a GEP in an organisation is linked to national and European initiatives or requirements or third-party funding in terms of: Requirements linked to getting European research funding (HORIZON Europe)	Regex - Normal match string	100	29

Characteristics

Table 33 Information extraction for characteristics indicator: Time frame

Indicator	Tool-method used	Success rate in %	Result in % (Mean)
Indicator 2.3 The time frame of the current plan	Regex - Pattern match	75	N/A ²⁸

²⁸ The variable is a string variable, thus its mean cannot be computed. It is formatted as a time frame, such as "### - ###".



Areas of Activity (grouped according to the specification for Horizon Europe) addressed in the GEP

Table 34 Information extraction for characteristics indicators: Areas of Activity

Indicator	Tool-method used	Success rate in %	Result in % (Mean)
Indicator 2.4 Areas of activity - Actions of awareness-raising and training addressed in the GEP	Regex -Normal string match	100	71
Indicator 2.4 Areas of activity - Work-life balance and organisational culture addressed in the GEP	Regex -Normal string match	100	80
Indicator 2.4 Areas of activity - Gender balance in leadership and decision-making addressed in the GEP	Regex -Normal string match	100	82
Indicator 2.4 Areas of activity - Gender equality in recruitment and career progression addressed in the GEP	Regex -Normal string match	100	75
Indicator 2.4 Areas of activity - Measures against gender-based violence, including sexual harassment, addressed in the GEP	Regex –Normal string match	100	65
Indicator 2.4 Areas of activity - Integration of the gender dimension into research and teaching content addressed in the GEP	Regex -Normal string match	93	58



The GEP integrates inclusive approaches

Table 35 Information extraction for characteristics indicators: inclusive GEPs

Indicator	Tool-method used	Success rate in %	Result in % (Mean)
Indicator 2.5 Inclusive approaches The GEP integrates intersectionality	Regex -Normal string match	100	15
Indicator 2.5 Inclusive approaches The GEP integrates diversity	Regex -Standalone string match	100	72
Indicator 2.5 Inclusive approaches The GEP integrates gender diversity	Regex -Standalone string match	100	41



The GEP addresses different inequalities

Table 36 Information extraction for characteristics indicator: different inequalities

Indicator	Tool-method used	Success rate in %	Result in % (Mean)
Indicator 2.6 Addressing different inequalities The GEP addresses race	Regex - Standalone string match	98	41
Indicator 2.6 Addressing different inequalities The GEP addresses nationality	Regex - Standalone string match	100	18
Indicator 2.6 Addressing different inequalities The GEP addresses religion	Regex - Normal string match	98	33
Indicator 2.6 Addressing different inequalities The GEP addresses class/social background	Regex - Standalone string match	93	15
Indicator 2.6 Addressing different inequalities The GEP addresses age	Regex - Standalone string match	95	49
Indicator 2.6 Addressing different inequalities The GEP addresses sexual orientation	Regex - Standalone string match	93	30
Indicator 2.6 Addressing different inequalities The GEP addresses disability	Regex - Normal string match	98	36



Implementation

Financial and personal resources, commitment

Table 37 Information extraction for implementation indicator: dedicated staff

Indicator	Tool-method used	Success rate in %	Result in % (Mean)
Indicator 3.3 A gender equality unit, gender equality committee and/or dedicated staff are in place	Regex- Normal match string	95	46

Quality Assurance

Table 38 Information extraction for implementation indicators: Quality Assurance

Indicator	Tool-method used	Success rate in %	Result in % (Mean)
Indicator 3.6 A policy for monitoring/controlling gender equality measures and the GEP implementation is in place	Regex - Standalone string match	91	82
Indicator 3.7 Sex-disaggregated data are collected and published: collection	Regex - Normal Match string	86	54
Indicator 3.7 Sex-disaggregated data are collected and published: publication	Regex - Normal match string	89	44



Self-assessment and impact

Table 39 Information extraction for impact indicator: description of the impact

Indicator	Tool-method used	Success rate in %	Result in % (Mean)
Indicator 4.1 The GEP includes a reflection or description of the GEP's impact. The organisation reflects its understanding of the impact of GEP.	Regex - Standalone string match	66	78

A pandas DataFrame containing the relevant indicator and its corresponding output is generated upon executing the Python script. The success of the information retrieval script was subsequently evaluated by comparing it to the manually assessed GEPs during the pilot phase. Overall, the performance of the information retrieval script is generally high, with the lowest success rate belonging to the impact indicator at 66%. Although regex is typically effective, we have outlined possible factors that may contribute to its lower success rate. These include:

- **Contextual issues:** The text contains search terms that are used in a different context than intended. In other words, the search term appears but does not refer to the indicator. We observed that the search term can be found in a footnote or reference section with no relevance to the indicator in certain instances.
- **Textual issues:** Some words could be corrupted during the text extraction process, or there might be hyphens within a word that the script may not be able to identify.
- **Translation issues:** We manually evaluated the GEPs from Germany in their original language and then translated them into English for text analysis before running the script. As a result, there could be potential translation discrepancies.

Using regular expressions (regex) for information retrieval has numerous benefits. Regex offers a versatile and potent approach to searching, matching, and modifying text, enabling accurate pattern matching and extracting specific data from extensive textual collections. While regex can process large amounts of text data as anticipated in the full sample of INSPIRE, it struggles with interpreting natural language variations and uncertainties. This limitation can result in incorrect identifications and omissions in information retrieval processes since regex lacks the ability to comprehend the context of the text beyond extracting predefined information.

The developed script for the information extraction has also been shared on the GitHub repository under the same username (<https://github.com/InspireQualityeu>) that was created for the INSPIRE project. The repository associated with the information extraction for the INSPIRE



indicators is called “gep-information-retrieval” and can be reached via the following URL: <https://github.com/InspireQualityeu/gep-information-retrieval>.

Realisation of text analysis

All in all, our web scraping application for the full sample study yielded 7018 PDFs. After conducting metadata analysis, the number was decreased to 1564. We managed to convert 1530 PDFs to plain text format out of 1564 PDFs. Thirty-four PDFs were corrupted and could not be processed and were therefore excluded from the analysis. Finally, we deleted non-English GEPs using “**langdetect**” library, yielding **1135** GEPs in .txt format for the analysis. The result of the text analysis is published in the **D.3.3 Report on GEP Monitoring Survey + Webcrawl Result** (see Löther, Karataş, and Weber 2024b).

4.4 Limitations of Web Scraping

While the web scraping methodology provided a systematic approach to GEP detection, several inherent limitations were identified. One primary constraint was the dependency on predefined search terms, which restricted the ability to detect GEPs labelled under alternative terminologies such as “Gender Equality and Diversity Plan” or “Equity, Diversity and Inclusion Plan”. As a result, the scraping process often did not detect GEPs using different language or terminology. This limitation reduced the methodology's comprehensiveness, as it could not adapt to the evolving ways in which institutions could label or categorize their GEPs.

Furthermore, the methodology was constrained by the availability of online-published GEPs on the organisation's website. Specifically, web scraping can only extract data from documents published on the organisation's webpage. In these cases, web scraping was unable to detect and capture the relevant documents, as it relied solely on accessible, publicly published content. Institutions that either did not post their GEPs or kept them behind closed portals created a significant gap in the data collected through this method, contributing to some gaps in the dataset.

Another significant challenge was posed by web security measures implemented by institutional websites. Anti-bot mechanisms, including CAPTCHA verification, robots.txt restrictions, and Cloudflare protection, can hinder automated data extraction, reducing accessibility to certain documents. These security measures can reduce the accessibility of certain documents and, as a result, may lead to the incomplete extraction of the necessary GEP data.



5 Reactive Method: Full Sample Survey

A survey can be proactive or reactive, seeking information directly from participants. We consider the INSPIRE-survey reactive because we asked RPOs and RFOs whether they have implemented a GEP and which measures it encompasses. Based on the indicators for GEP monitoring (see section 3.3), this chapter reports on the application of the indicators to the pilot survey and the full sample survey and highlights the differences between them. In addition, this chapter describes the technical implementation of the full survey and the data cleaning process and provides statistics on the response rate for both surveys (for the result of the survey, see Löther, Karataş, and Weber 2024b).

5.1 Questionnaire Design

After adapting the indicators to the web scraping and text analysis (see sections 4.2 and 4.3), we developed the items for the survey. Items that had already been developed in the literature were not deemed to be suitable for the purpose of our survey. We thus formulated the items following the conceptual ideas of our indicators. The INSPIRE questionnaire for the full survey consists of several steps, and these steps ensure the quality of the questionnaire:

- Internal revision of the first draft,
- External feedback through INSPIRE consortium members and eight German gender equality officers,
- Testing the questionnaire in a pilot survey (81 RPOs from four countries),
- Revision of the questionnaire by integrating participants' feedback and detecting inconsistencies in the analysis of the pilot data

The pilot survey was particularly helpful for designing the full sample survey since we implemented two open-ended questions for the quality check to help us improve the full survey (issues with the survey's language; feedback on the questions).

The analysis of the survey results informed the revision of the questionnaire according to the following aspects:

- Detecting misunderstandings through inconsistent answers.
- Detecting misunderstandings through answers in the text field “other”.
- Detecting problems in analysing and interpreting the results due to item or question inconsistencies.
- Integrating items which stem from the text fields “other”.
- Exploring answers to the feedback questions (survey language, feedback on the questions).

Table 40 displays the number of items changed after the revision of the pilot questionnaire. In total, 66% of the survey items asked in the pilot remained for the full sample survey.



Table 40 Changed items in the full survey compared to the pilot survey

Amount of changes	N	%
No changes	31	66.0%
Minor changes (small reformulation to be more precise)	4	8.5%
Major changes (additional categories, comprehensive reformulation, merging or splitting items)	12	25.5%
Total	47	100.0%

5.2 Translation of the Questionnaire

The survey was displayed in English and German during the pilot phase. We compiled the pilot survey in English and translated it into German. For this purpose, we tested two translation tools for translating into German. The first tool is a LLM called m2m_100_418M²⁹ (A. Fan et al. 2021); the second tool is deepL.com³⁰, a commercial translation service. DeepL.com provided significantly better translation quality than the *LLM m2m*. Therefore, we prioritize using deepL.com for the full sample survey.

The full survey was made available in five languages: English, French, German, Polish and Spanish. Financial resources restricted the translation to four languages. The selection criterion for these languages was the number of organisations in which these languages are spoken. Further, they represent each country cluster, and the selection has been confirmed by the INSPIRE consortium. Table 41 displays the translation process of each language for the full survey.

Table 41 Translation process

Language	Translation process
German	Translation software (deepL), revision by German native speakers (project team member)
French	External professional translation, revision by project team members fluently speaking French
Polish	External professional translation, revision by Polish native speakers (INSPIRE Consortium)
Spanish	Translation software (deepL), revision by Spanish native speakers (INSPIRE Consortium)

²⁹ Many to many 100 languages with 418M parameter

³⁰ see <https://www.deepl.com/translator>



In the same manner, described above, we translated the invitation letter and two reminders into four languages (see the annexe, section 10.6.1 for the invitation letter for the full survey and sections 10.6.2 and 10.6.3 for the reminders).

5.3 Email Extraction for the INSPIRE Survey

This section outlines the structured process we followed to identify key personnel responsible for GEP implementation, extract their email addresses and support the efficient and targeted distribution of the INSPIRE survey. Our approach involved an extensive search across 6,571 websites belonging to HEI, REC, PRC and RFO. A hierarchical ranking system was implemented to prioritize roles, favouring Gender Equality Officers, followed by Rectors/Presidents/CEOs, Heads of Administration/HR Department and Heads of Communication Department.

Step 1: Automated Web Scraping and Data Extraction

The process began with an automated web scraping phase, where custom Python scripts were deployed to systematically search through institutional websites for contact pages related to GEPs or roles tied to gender equality. We used a variety of specific search terms to locate the most pertinent pages, including:

"gender equality", "diversity", "inclusion", "rector", "president", "CEO", "head of administration", "head of HR", "head of communication"

Once relevant pages were identified, regular expressions were used to extract any email addresses associated with these pages. Where available, the script was designed to capture additional details, such as the names and roles of individuals, which were then structured into a database.

In cases where no emails were identified, an alternative approach was employed using Google Search, executing the following query:

"{org_name} {keywords} email -filetype:pdf"

The goal was to find emails on a webpage on the organization's website that referenced PDFs containing GEP-related documents. We hypothesized that these PDFs might include the contact details of personnel responsible for the document, allowing us to obtain the necessary contact information indirectly.

If this method was unsuccessful, we employed an advanced AI-driven approach using ChatGPT with the following prompt:

"Find email addresses of key personnel (gender equality officer, diversity officer, inclusion officer, rector, president, CEO, head of administration, head of HR, head of communication) for the organization: {org_name}."



The results were added to a .csv file, including all sets of names, emails, and roles that were identified for each institution or organization.

Step 2: Manual Quality Check

After the automated extraction process, a thorough manual review was conducted to ensure the accuracy and relevance of the collected data. The review process was crucial, as missing or incomplete information was prevalent in a significant portion of the extracted data. The manual check involved the following:

- **Deleting invalid email addresses:** Email addresses deemed invalid, such as those tied to certified electronic mail systems (e.g., Italy's PEC addresses), were removed from the dataset to avoid sending the survey to non-functional or inappropriate contacts.
- **Verifying missing or incorrect information:** Any missing or incorrect email addresses were manually retrieved or verified by visiting the official websites of the respective organizations to find the correct information.

In cases where a Gender Equality Officer could not be identified during the automated extraction phase, the manual review process followed the hierarchical order of importance. The next role in the hierarchy—such as a Rector, President, or CEO—was sought out and added. If no specific role was found, we resorted to collecting general administrative contacts, ensuring that the survey reached the most appropriate individuals, even in the absence of a dedicated GEP role.

Step 3: Data Compilation and Reporting

Once the contact information had been thoroughly reviewed and validated, the final dataset was compiled. The data was organized and structured into .xlsx format, incorporating the initial website information and appending each entry with the following relevant fields:

- Contact Names
- Contact Emails
- Contact Roles

This dataset formed the foundation for the distribution of the INSPIRE survey, enabling targeted and efficient outreach to the key personnel responsible for GEPs across the institutions. This dataset was prepared for distribution to support the administration of the GEP survey. INSPIRE survey is sent out to **4,571** organisations, which is **2000** less than the web scraping sample. This was because the email extraction process was challenging due to the uniform structure of each web page, which made automated approaches to retrieving email addresses less effective. In addition, the presence of different languages posed further



obstacles. Finally, the organisations that received invitations to the pilot study were excluded from this email-extracting process.

5.4 Implementation of the Survey

We used Unipark as the software tool for the survey. We used the questionnaire of the the pilot survey into which we implemented changes to the questionnaire (see minor and major changes see Table 40, p. 90). Subsequently, we modified the German version and integrated French, Spanish, and Polish translations into the full survey, utilising Unipark's "export/import" feature in the language editor, allowing for a straightforward and error-robust implementation. The first question in the survey is language selection, which asks for the participants' preferred language; then, the rest of the survey is displayed in the selected language. Furthermore, the second essential question in the survey is about consenting to completing the survey and storing the participants' data. If participants do not agree, the survey ends, and they can no longer access the survey.

INSPIRE survey includes various types of questions such as open-ended questions (text input), single choice (radio buttons), multiple choice (checkboxes), dropdown menu (single choice), matrix (Likert scale - single choice per row), and slider question. In addition, before sending out both the pilot and full survey, we performed several quality checks to ensure that everything worked as intended. These quality checks include i) validating the filter status in Unipark, ii) creating test data in Unipark, and iii) having internal team members test the survey.

Using filters helps us improve data quality by creating questionnaires tailored to specific groups. For example, the INSPIRE survey is designed regardless of the status of GEPs. This means that completing the survey does not require having a GEP. Nonetheless, we designed some questions for institutions with and without a GEP. Furthermore, we offered a "return" (back) option so that participants could change their answers if they wished. We enabled the option "back button data correction". Thus, Unipark resets the data and automatically saves the last response entered. The survey also contains mandatory questions, which implies that participants cannot proceed without filling them out.

Participants could take a break and complete the survey at any time until the survey's end date. When participants wished to continue after closing the browser, the last page they sent was displayed. Participants could also forward the invitation email to the appropriate bodies within their organisation. Whoever possessed the URL had access to the survey. Thus, multiple persons could complete the survey.

We prepared an invitation letter and two reminders in English, German, Spanish, French, and Polish for the full survey. We communicated specifically in German for German and Austrian organisations, French for French organisations, Spanish for Spanish organisations, and Polish for Polish organisations. We sent invitations and reminders in English for the remaining organisations in the INSPIRE sample. We included six members of the CEWS team at GESIS in participant lists of the full survey and assigned different languages in order to determine if the invitation letters and reminders worked effectively in the assigned languages. We correctly



received invitation letters and reminders in assigned languages. Initially, we planned a field phase of four weeks for both the pilot and full survey. However, the full survey was extended by two more weeks to mitigate the risk of low responses owing to the summer holiday. Table 42 displays the timeline of the surveys.

Table 42 The timeline of the pilot and full sample surveys

Survey type	Invitation letter/Starting date	The first reminder	The second reminder	End date of the survey
Pilot survey	28.02.2024	11.03.2024	21.03.2024	27.03.2024
Full survey	04.07.2024	22.07.2024	12.08.2024	19.08.2024 ³¹

The survey has been sent via Unipark. Disposition codes³² are assigned by Unipark for participants based on their status. Unipark assigns the status, but we (as admin) can also modify it. Table 43 shows the status statistics of participants. When participants have not used the link yet, their status is assigned as "invited" (disposition code:12) . When at least 30 minutes of inactivity was observed, their status becomes "suspended" (disposition code:22); if participants successfully complete the survey with or without break³³ their status become "completed" (disposition code:31 and 32). The status "screen out" is assigned when participants are excluded from the survey. Participants might be excluded from the survey in two ways: they refuse informed consent, or they have no knowledge of gender equality issues at their institution (disposition codes 37 and 38). The status "custom unavailable" is a customised status (disposition code 18) akin to "cannot be reached" that can be further differentiated. We assigned "custom unavailable status" to organisations who explicitly emailed us and stated they did not want to participate in the survey. The status "email could not be delivered" (disposition code 14) relates to situations in which the mail server refuses to accept an email repeatedly or cannot connect to the mail server repeatedly. Unipark indicates that four attempts are made per default. The status "can not be reached" refers to a situation in which participants can not be reached. This status can be assigned manually. However, we have not assigned any of the participants. Possible reasons for "cannot be reached" include³⁴, but are not limited to: i) invalid email addresses, ii) full mailboxes, iii) address not found, and

³¹ The table shows the dates on which emails were sent. Due to the high volume of email requests overwhelming Unipark's system, the actual delivery of the invitation letter or reminders may take one day longer than indicated in the table.

³² For all disposition code in Unipark, please see <https://qbdocs.atlassian.net/wiki/spaces/DOC/pages/1441563458/Disposition+Codes>

³³ Unipark has different disposition codes completing with break and without break. For convenience we merged two cases.

³⁴ For more detail see <https://qbdocs.atlassian.net/wiki/spaces/DOC/pages/4095541263/Good+Mail+Bad+Mail+How+to+Avoid+Invitations+Ending+up+in+the+Spam+Folder>



iv) authentication issues. All in all, Unipark reports the survey was not received by 222 ("email could not be delivered" + "can not be reached", which makes up 4.8%) participants.

Table 43 Status statistics of participants

Survey type	Total sample	Email could not be delivered	Can not be reached	Custom unavailable	Invited	Screen Out	Suspended	Completed
Pilot survey	83	1 (1.2%)	0 (0%)	1 (1.2%)	58 (69.9%)	0 (0%)	8 (9.6%)	15 (18.1%)
Full survey	4571	49 (1.1%)	172 (3.8%)	3 (0.1%)	3743 (81.7%)	8 (0.2%)	336 (7.3%)	268 (5.9%)
Total = Pilot + Full	4654	50 (1.1%)	172 (3.7%)	4 (0.1%)	3793 (81.5)	8 (0.2%)	344 (7.4%)	283 (6.1%)

Nevertheless, it may also occur that our survey invitation could end up in the spam bin due to a variety of circumstances, including i) non-personalised communications and ii) the inclusion of a link to the survey. In several situations, from French organisations, we received some responses suggesting that we needed to click on the provided link to confirm that our survey invitation was not spam. Their emails indicate that this is a security measure because we are emailing participants for the first time. For all cases, we clicked the suggested link to prove to them that the invitation mail was not spam.

5.5 Data Cleaning and Appending: Pilot and Full Survey

The logic for data cleaning for the full and pilot surveys was mainly comparable. The distinction between them in terms of data cleaning lies in questionnaire elements such as inclusion, exclusion, and modification of survey items, including item choices. The codebooks for both surveys differ because the items are not identical, and we prepared two separate scripts for data cleaning. Nevertheless, 66% of the items remain the same in the pilot and full survey, as Table 40 shows. It is necessary to have exactly the same variable names to append two datasets to have a unique dataset. We defined the variable names identically for the same variables even though they had distinct names in their respective codebooks.

We exported the datasets into Excel files (.xlsx) as well as their respective codebook files (.pdf) containing variable names and their corresponding items. We used Stata for the data-cleaning



process. Thus, we imported Excel files into Stata. We start the cleaning process by eliminating the observation that is not needed. We eliminated testers; we had one tester (internal team member) in the pilot sample and seven testers (internal team members) in the full sample. Testers were introduced to check surveys during the course of the surveys. We also eliminated observations who did not complete the survey based on their disposition code. In particular, we removed the following cases from the samples: "invited", "email could not be delivered", "cannot be reached", and "custom unavailable". In this case, participants who rejected the informed consent and those who had no knowledge about gender equality plans were also eliminated from the full sample. We only kept observations in which their status was either "completed" or "completed after the break". We also noted that two individuals' behaviour in the survey was ambiguous, which means they saw and answered both sets of questions designed for organisations with and without GEPs. In fact, they should have seen only either set of questions designed for organisations with or without a GEP. Hence, we also eliminated these two participants from the full sample. Consequently, the number of observations was reduced to 15 for the pilot and 266 for the full sample.

Cleaning the data entails transforming different types of questions into a more generic way to prepare for the analysis. It also necessitates correct labelling of the variables, answer items and corresponding questions. The fundamental distinction is among question types. First, we categorised questions into closed-ended questions (which accept values based on established options) and open-ended questions. We encoded closed-ended questions as bytes or float and open-ended questions as strings. Byte refers to a numeric number in binary (0 or 1) or small categorical data, whereas string indicates a textual response. Float could hold much higher numbers than byte, and was automatically assigned by STATA when a new variable was created. We set up numerical values as bytes, we built labels for both questions and options, which we then allocated to the variables. Unipark also makes distinctions about the type of variables when it comes to reporting the missing values. Following the Uniparkn guidelines, we coded missing values for different types of variables in the dataset. Finally, we removed project variables that Unipark produces that are unnecessary, such as "duration", "last-page", "dispcode" etc.

We appended two datasets to create a unique sample of the dataset. Sixty-six per cent of the data from the pilot and the full sample are present in the unique dataset. The dataset does not include the remaining variables. The dataset, pilot, and full sample must all have the exact same variables in order to move further with the append. In Stata, the "append" command essentially stacks them on top of one another. There are 281 observations in the appended sample; 15 were gathered from the pilot sample and 266 from the full sample.

Initially, the data cleaning process was performed using STATA; then, the cleaning script was translated into the open-source software R³⁵. Data analysis was performed using STATA, Excel, SPSS³⁶, and Python. We also translated some of the data analysis scripts into R. Both scripts, the data cleaning script (<https://github.com/InspireQualityeu/gep-survey-data->

³⁵ <https://www.r-project.org/>

³⁶ <https://www.ibm.com/de-de/products/spss-statistics>



[cleaning](#)) and the data analysis (<https://github.com/InspireQualityeu/gep-survey-data-analysis>) script, are released as R code in the GitHub open source repository.

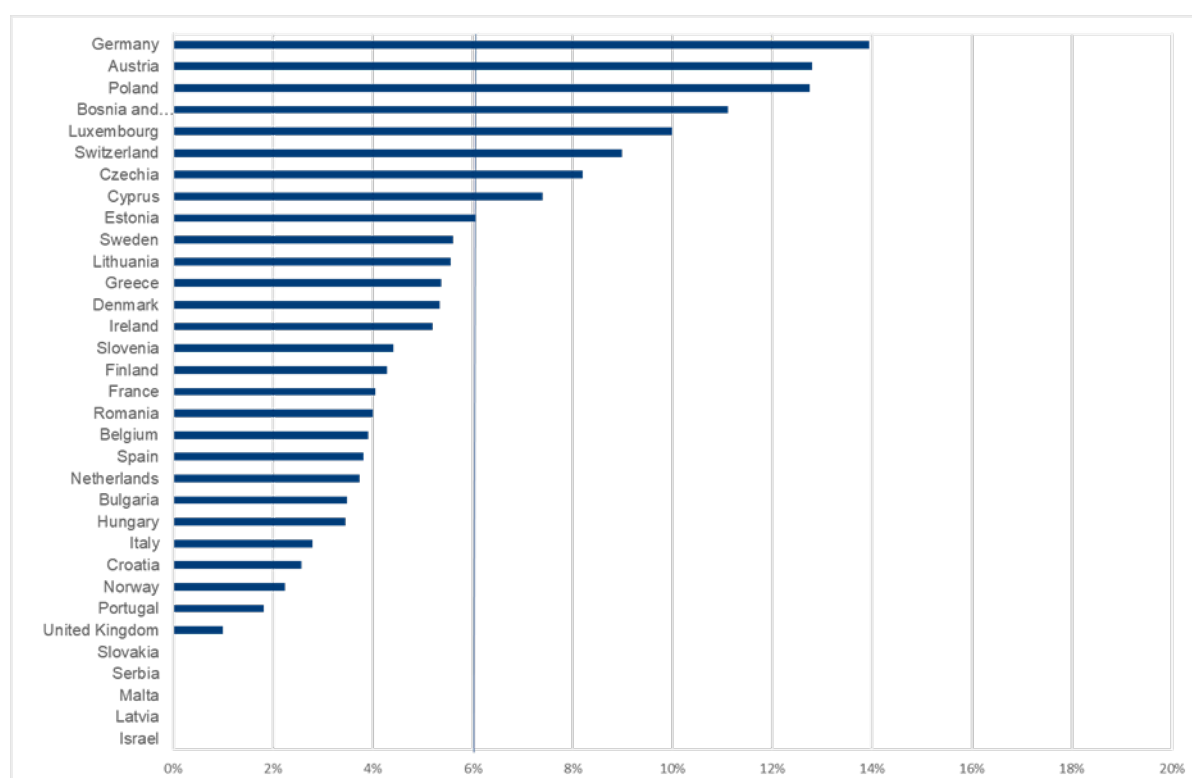
5.6 Response Rate

The final data set of the survey (merged pilot and full survey) includes 281 research-performing and research-funding organisations. The response rate is 6,0% and lower than expected. There are different organisational and technical issues that explain the low response rate:

- Lacking quality of the email addresses (f.ex. info@##) for the invitation; the invitation was not directed to a correct person
- Invitation has been classified as spam by email programs
- The time period: the survey was online during the summer and vacation period (July and August)

These issues are linked to the characteristics of the organisations that answered or did not answer. First, there are significant differences in the response rate between countries, as shown in Figure 7.

Figure 7 Response rate by country



Source: INSPIRE survey Q 2.1.

The response rate is above average for eight countries: Germany, Austria, Poland, Bosnia and Herzegovina (N=2), Luxembourg (N=1), Switzerland, Czechia and Cyprus (N=2); only very



few organisations from the United Kingdom and Portugal answered the survey and no response was received from Israel, Latvia, Malta, Serbia and Slovakia.

Considering low numbers in several countries – both in the survey and the web scraping data – we clustered the countries in four clusters for European Countries – Central West, Central East and Eastern, Southern and Northern/West – and a cluster “not belonging to a cluster” which only includes Israel (see section 10.7) The same country clusters are part of the INSPIRE structure.

Clustering the countries (see Table 44) to regional areas, the response rate was especially high in central western countries and low in southern and northwestern countries. In the cluster “Central East and Eastern”, the response rate varies between higher than the average (Poland and Czechia) and countries with no answers (Latvia, Serbia and Slovakia).

Table 44 Response rate by country cluster

Country cluster	N	Response rate
Central West	161	9.4%
Central East and Eastern	54	6.5%
Southern	43	3.4%
Northern/West	23	2.8%
Not belonging to a cluster	0	0.0%

Table 45 Response rate by availability of the questionnaire in the national language

Language	N	Response rate
German	133	13.3%
Polish	26	12.7%
French	20	4.0%
Spanish	21	3.8%
Not in the national language	73	3.8%
English	8	1.7%



Table 46 Response rate by type or organisation

Type of organisation	N	Response rate
Higher education	141	10.7%
Research organisation	122	6.0%
Private company	7	0.6%
Research-funding organisation	11	26.2%
Total	281	6.0%

Source: INSPIRE survey Q 2.1 and Q 2.2.

The availability of a questionnaire in the national language explains higher response rates in some countries (see **Table 45**). Providing the questionnaire in German and Polish impacts the response rates, but this is not true for French, Spanish and English. Thus, we can explain the high response rate in Poland by the availability of a Polish questionnaire, but language doesn't account for the high response rate in Czechia and the low rate in the United Kingdom.

Furthermore, the response rates differ according to the type of organisation. Whereas more than one-quarter of the research-funding organisations and 10 % of the higher education institutions answered the survey, very few private companies with research activities (N=7) participated in the survey. The quality of the email addresses and classification as spam might impact the response rate of private companies.

Finally, we might also assume a nonresponse bias that organisations not involved in gender equality policies didn't participate in the survey. Nonresponse bias is a sort of selection bias that arises in surveys where people who do not participate differ significantly from those who do (Berg 2005). The results on the prevalence of gender equality plans confirm this assumption. This nonresponse bias also explains the low response rate among private companies, as they are less likely to have a GEP.

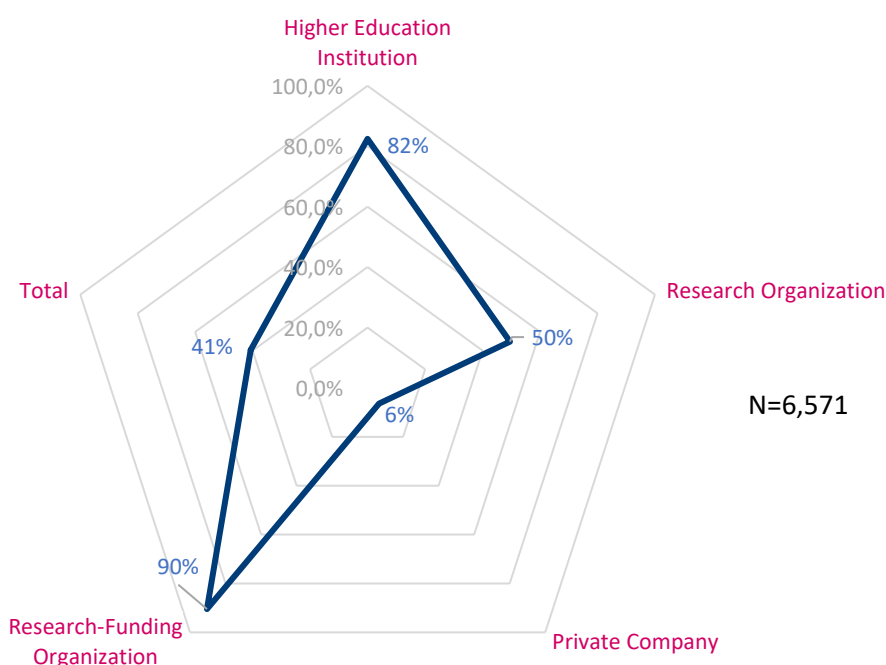
6 Results on Prevalence of GEPs: Reactive vs. Non-Reactive

Provided URLs are available, web scraping makes it possible to assess the prevalence of GEPs across many organizations. Using web scraping data from 6,571 organizations, we found that 41% have a GEP (see Figure 8). Private companies had the lowest GEP prevalence at 6%, followed by research organizations at 50% and higher education institutions at 82%. Research funding organizations had the highest prevalence, with 90% having a GEP (see Löther, Karataş, and Weber 2024b). Due to certain limitations of web scraping, such as (i) anti-bot measures, (ii) strict adherence to predefined search terms, which may prevent the detection of plans other than gender equality plans (GEPs), and (iii) the requirement that GEPs be published on an organization's website, the prevalence of GEPs identified through web



scraping may be underestimated (see section 4.4 and Löther, Karataş, and Weber 2024b, 33–41).

Figure 8 Results of GEP prevalence by organization type based on web scraping



Source: INSPIRE web scraping. N= 6,571

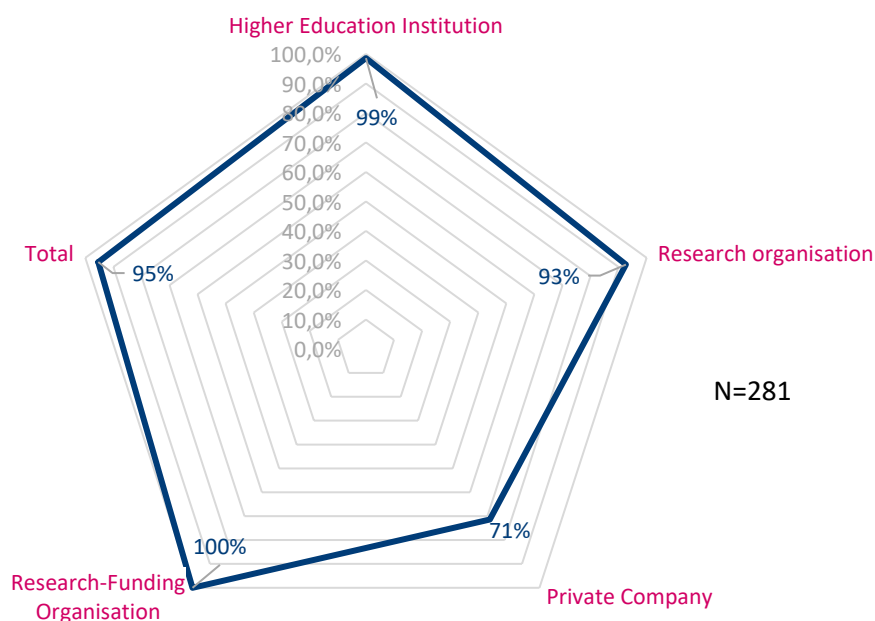
Compared to the web scraping data, the survey data on the prevalence of GEPs has a strong nonresponse bias, as almost all organizations that participated in the survey reported having a GEP, 95% of 281 (see Figure 9).

The web-scraping results confirm this nonresponse bias, as organizations already engaged in gender equality policies were more likely to participate in the survey. As a result, surveys which focus especially on gender equality are unreliable to measure the prevalence of GEPs and lead to biased estimates.

Despite the aforementioned nonresponse bias and the resulting unreliability of survey data on the prevalence of GEPs, the INSPIRE survey does provide useful information regarding the circumstances for adopting a GEP, as well as its characteristics, implementation, and impact (see Löther, Karataş, and Weber 2024b).



Figure 9 Results of GEP prevalence by organization type based on online survey



Source: INSPIRE survey 3.1. N= 281

7 Comparison of the Two Approaches: Reactive vs. Non-Reactive Methods

Our study compares reactive and non-reactive methods to monitor GEPs on a supra-institutional and European-wide level. This chapter presents methodological results of comparing both approaches, building on the experiences in the full sample study, which comprises 6,571 organisations from 33 countries. The two strategies – online survey as a reactive method and web scraping together with automated text analysis as a non-reactive method – present significant advantages and obstacles, each in particular circumstances. We argue that both methods are complementary rather than competing with each other.

Regarding **data quality**, the online survey turned out better than web scraping because of the personalised research design. The survey allows us to collect specific information and design questions that directly address our research objectives, prevalence, characteristics, implementation and impact of GEPs. The survey also allowed us to disaggregate and collect data in a more detailed manner, such as distinguishing between different types of plans (e.g., gender equality plan, gender equality and diversity plan, and diversity, equity, or inclusion plan). In addition, the survey enabled us to collect data on indicators unavailable via web scraping, such as the date a GEP was first adopted and the percentage change in the representation of women in senior research positions. Survey data also informed web scraping



about gaps that lead to false negatives, such as diversity and equity concepts instead of gender equality.

Surveys are typically known for a low response rate. This issue was also the case for the INSPIRE survey. The response rate of the survey is 6 per cent out of 4,571 organisations. The online survey is subject to nonresponse bias, which occurs when individuals who do not respond systematically differ from those who do respond, leading to biased results. In this case, the bias occurs because the survey focused on gender equality and GEPs and thus attracted predominantly respondents from organizations with a GEP (prevalence rate of 95%). The prevalence data derived from online surveys focusing on gender equality are not reliable. Comparing the survey data with web scraping data suggests that percentages are overestimated for issues sensitive to nonresponse bias. Web scraping data confirms the nonresponse bias, indicating a GEP prevalence of 41% among 6,571 organizations. Only 7 (0.6%) private companies responded to the online survey, making it difficult to reach private companies. However, nonresponse bias may have played a role here since private companies are less likely to have a GEP – 6% indicated via web scraping—and EU horizon eligibility criteria do not apply to them.

Web scraping reaches a large and unbiased sample, and the assessment in the pilot study proves the high quality of the developed web scraping tool. In particular, the prevalence data are reliable. The large number of observations allows for differentiation by country. Web scraping data thus serves as a corrective for survey data. Nevertheless, the survey data provide information about the limitations of web scraping data. In particular, web scraping is sensitive to the search terms (gender equality vs diversity, equity and inclusion) and whether a GEP is published on the website. Another disadvantage of web scraping is that it faces difficulties when websites implement anti-bot strategies to prevent web scraping. These limitations can lead to false negative results and underestimating the percentages, especially for private companies.

Both designing an online survey and developing a web scraper are **time-consuming**. Designing a survey involves careful consideration to ensure the questions are unbiased and adequately measure the intended outcome. Deciding on and configuring the web scraper tool, i.e., SerpAPI, was time-consuming since it requires trials with different search terms and languages. Furthermore, our online survey procedure required significant time to retrieve participants' email addresses — of those persons who are responsible for gender equality in the organisations — and in 2,000 cases, we were unable to find the email addresses of responsible persons. Identifying the URLs of the target organizations for web scraping and extracting email addresses proved challenging despite initial assumptions that the process would be straightforward. Since the CORDIS database was not up to date, many URLs were found to be non-functional, and it was necessary to verify and correct the URLs before performing web scraping. However, accurate URLs are essential for both methods.

The **cost** of implementation may differ based on the tools chosen for each approach. In our case, we opted to purchase a paid web scraping tool due to its ability to gather high-quality data. In addition, the expenses associated with classification and translation can be significant.



The questionnaire needed to be at least proofread by native speakers or professional translators. The translation is also essential for the classification and analysis of the "big data" corpus when GEPs are available in different languages. An automated machine translation via LLMs requires special hardware equipment, i.e., GPUs. For efficiency, we limited the full sample study to GEPs available in English. Regarding the INSPIRE survey, we utilised a licenced survey software called Unipark through GESIS. However, other options, such as free web scraping tools and open-source survey software, may be more cost-effective for specific projects.

Ethical considerations are another aspect of data collection. In the survey, participants must typically sign data consent forms in compliance with the General Data Protection Regulation³⁷ (GDPR) enforced by the EU. As of 2022, no laws or rules govern web scraping (Kryukov 2023). Each case of web scraping should be examined contextually, considering criteria such as the nature of the data being scraped, the source of the scraping tools and data protection, and the website's terms of service. Thus, collecting data regarding GEPs with SerpAPI should be aligned with the legal context since we are retrieving the crawled information from Google's database.

The development of new tools in Information Technologies (IT) and Natural Language Processing (NLP) enhances the capabilities of web scraping and text analysis, suggesting a potential shift towards more automated methods in the future. In this study, some LLMs were used and reported during the pilot phase. In particular, GPT (3.5 instruct) and BART models are used for PDF classification, and m2m100b is used to translate GEPs from non-English to English. Nevertheless, there are newly developed LLMs that continuously improve the performance, so this could further improve the performance of zero-shot classification tasks in the future. Currently, several advanced language models are available, in addition to those used in this study, including GPT-4, LLaMA 2³⁸, Flan-T5³⁹, and various BERT variants⁴⁰. Practical concerns include addressing hardware limitations when deploying large language models (LLMs) and analysing a vast quantity of PDFs, which require strategic decisions to optimize the study's methodology and technical setup. With adequate resources, the performance of large language models (LLMs) can be enhanced through fine-tuning. However, this process demands substantial computational power and considerable time. Furthermore, LLMs could also be useful for information retrieval. Currently, our information retrieval strategy is based on predefined search terms or patterns. In contrast, transformer-based LLMs (Question Answering (QA) models in LLMs) can understand the context and provide accurate answers. They are specialised systems that provide accurate answers to user queries based on a given context or corpus of information. Certain variations of BERT and GPT have been adapted for use in developing QA models.

³⁷ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504>

³⁸ <https://github.com/meta-llama/llama>

³⁹ <https://github.com/google-research/FLAN>

⁴⁰ <https://github.com/google-research/bert>



8 Input from the Co-Creation Workshop Data Monitoring

This section incorporates the insights from the **Co-Creation Workshop Data Monitoring** that took place on the 20th and 21st of November 2024 at GESIS in Cologne, Germany. Twenty experts in research, monitoring procedures, gender equality politics and European politics participated in the workshop. The aim of the workshop was to discuss the first results of the INSPIRE survey and web-scraping data and how these insights could be intensified but also incorporated into other established procedures for gender monitoring and the monitoring of gender equality plans in the European context.

8.1.1 The Purpose of and Data Access for Monitoring

The participants discussed the meaning of monitoring. It should be clear to all what is meant by monitoring. The experts pointed out that the purpose of the monitoring of GEPs should also be defined, and the reason for the monitoring should be clearly stated. They also asserted that member states should engage with the monitoring process of GEPs (also different stakeholder groups). Additionally, participants argued that high-quality data is crucial to make evidence-based decisions. Sustainable monitoring should be led at a higher level (i.e., EC) and should be mandatory for all applying research organisations, including private companies. Sustainable procedures, tools, and standards should be highlighted and promoted.

The monitoring of GEP could be improved by integrating also qualitative data and tools that would be able to measure the impact of GEPs, here are research gaps that needs to be investigated. Integrating **qualitative data** was identified by the experts as a crucial aspect. In particular, collecting data via interviews is suggested due to its ability to capture people's experiences, which quantitative data is incapable of. Developing **tools to measure the impact** of GEPs is suggested since there is a lack of tools to measure the impact of GEPs. In addition, progress should be measured against a stated action plan rather than solely on data, recognizing critical evaluation and learning from actions that did not reach the desired outcomes.

Further, the participants collected ideas how to improve the data access for GEP monitoring. First, participants suggested that GEP publication should be required in a **central repository** since obtaining the data via web scraping is highly challenging. It would be optimal if a GEP database of Horizon Europe participants is created at a higher level. In particular, EC could push to collect GEPs as PDFs or links to GEPs at the submission stage. This would facilitate monitoring and focus more on analysing the content of GEPs. Establishing a standardized GEP template could also facilitate its monitoring.

Second, publishing GEP in **English** would also contribute positively to the monitoring process since it would eliminate the translation burden.

Third, participants recognize a **lack of legislation for GEPs** at the national and/or regional level, as discussed by the Genderaction+ project (see Knapieńska and Chrobak-Tatar 2023),



so that the requirements for GEPs differ considerable. Fourth, experts emphasized other projects/tools for **monitoring gender equality**, such as DIMIS (see G. Stadler et al. 2023), GEAM (see Aldercotte et al. 2021), and INTERMAPS⁴¹ and how the combination could be improved of the different tools for monitoring purposes.

Finally the experts suggests, that it should be more explored how the use of LLMs could enhance **content analysis of GEPs**. Despite using GPT and BERT models, other EU-based LLMs like MISTRAL.AI⁴² and ALEPH ALPHA⁴³ have emerged, they are prioritizing regional-level data protection.

8.1.2 Prevalence of GEP in Private Companies

Web-scraping results show that the prevalence of GEPs is quite high in publicly financed organisations (50% for REC, 82% for HEI, 90% for RFO), but low for private companies (low prevalence rate 6%). There should be more political emphasis to increase the engagement in private companies for GEPs. Experts suggest that the eligibility criteria should also apply to private companies, requiring them to have a GEP. For this purpose, the awareness among RPOs on the benefits of GEPs for the organization should be highlighted by politicians. Another suggestion by the experts is to create communities of practices to foster the responsibility for GEPs on the ground.

8.1.3 Intersectionality Approach of GEP

Currently, the eligibility criteria do not require GEPs to incorporate intersectional and inclusive perspectives. Thus, many organizations do not include intersectional considerations in their GEPs. Participants argued that intersectionality should be more explicitly required in GEPs to ensure that intersectional perspectives are embedded in identified actions and aligned with the organizational context. However, participants also discussed a lack of knowledge about intersectionality and how it could be applied in GEP. This could be an additional "burden" for newcomers, and it can be problematic in certain regions.

Participants discussed that GEPs should be designed to address gender while taking into account intersectional dimensions to ensure that they cover "gender and beyond". Monitoring gender equality using disaggregated data - categorized by different dimensions of inequality - allows for a more nuanced understanding of inclusiveness and intersectionality.

8.1.4 Improvements for the CORDIS Database for Better Data Quality in the Sampling

A good sampling database is crucial for effective monitoring. In particular, the experts discussed the need for a good database with a balanced sample. We noticed that the database used to create the sample was not up to date and that there were a lot of non-functional URLs.

⁴¹ <https://www.upf.edu/web/intermaps/the-project>

⁴² <https://mistral.ai/en>

⁴³ <https://aleph-alpha.com/>



Thus, we spent additional time recovering non-working URLs, making the process more time-consuming and challenging. In collaboration with the experts, we propose including a new column in the CORDIS database to display the Research Organisation Registry (ROR) ID. An ROR ID in the database provides several benefits, including globally unique identification, interoperability with other research infrastructure systems, and improved data consistency, particularly in naming organizations. In addition, ROR provides an application programming interface (API) free of charge, which can make the process of sampling a lot faster.

Furthermore, participants also suggested that compliance with the recommendations of the Frascati Manual (Organisation for Economic Co-operation and Development 2015) on the reference population be ensured, and the sampling framework should be aligned with national R&D survey populations, where available.



9 References

- Aldercotte, Amanda. 2018. "Monitoring and Evaluating Impact." Unpublished manuscript, last modified February 01, 2023. <https://www.advance-he.ac.uk/knowledge-hub/monitoring-and-evaluating-impact>.
- Aldercotte, Amanda, Maria Caprile, Kevin Guyan, Memory Malibha-Pinchbeck, Jörg Müller, Rachel Palmen, and Carla Startin. 2021. "ACT - Gender Equality Audit and Monitoring (GEAM) - Version 2."
- Baden, Christian, Christian Pipal, Martijn Schoonvelde, and Mariken A. C. G. van der Velden. 2022. "Three Gaps in Computational Text Analysis Methods for Social Sciences: A Research Agenda." *Communication Methods and Measures* 16 (1): 1–18. <https://doi.org/10.1080/19312458.2021.2015574>.
- Barnard, Sarah. 2017. "The Athena SWAN Charter: Promoting Commitment to Gender Equality in Higher Education Institutions in the UK." In *Gendered Success in Higher Education: Global Perspectives*, edited by Kate White and Pat O'Connor, 155–74. London: Palgrave Macmillan.
- Barton, David. 2023. "Web Crawling Vs. Web Scraping: What's the Difference?" Accessed March 27, 2024. <https://blog.apify.com/web-crawling-vs-web-scraping/>.
- Berg, Nathan. 2005. "Non-Response Bias."
- Biela, Jan, Katharina Warta, Nadia Galati, Simon Zingerle, and Silvie Klein-Franke. 2022. "Evaluation Des Professorinnenprogramms Des Bundes Und Der Länder: Dritte Programmphase Und Gesamtevaluation: Evaluationsbericht." Unpublished manuscript, last modified January 17, 2025. https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/Evaluation_des_Professorinnenprogramms_Bericht_Januar_2022.pdf.
- Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. 2020. "Language Models Are Few-Shot Learners." *Advances in neural information processing systems* 33: 1877–1901.
- Bührer, Susanne, Evanthia Kalpazidou Schmidt, Rachel Palmén, and Sybille Reidl. 2020. "Evaluating Gender Equality Effects in Research and Innovation Systems." *Scientometrics* 2 (4). <https://doi.org/10.1007/s11192-020-03596-1>.
- Bührer, Susanne, and Angela Wroblewski. 2019. "The Practice and Perceptions of RRI-a Gender Perspective." *Evaluation and program planning* 77:101717. <https://doi.org/10.1016/j.evalprogplan.2019.101717>.
- Chae, Youngjin, and Thomas Davidson. 2023. "Large Language Models for Text Classification: From Zero-Shot Learning to Fine-Tuning." <https://osf.io/preprints/socarxiv/sthwk>.



- Cheveigné, Suzanne de, Bente Knoll, Maria Bustelo, Eivind Engebretsen, and Ulf Sandström. 2017. "Interim Evaluation: Gender Equality as a Crosscutting Issue in Horizon 2020." Unpublished manuscript, last modified March 21, 2023. <https://hal.science/hal-02948895/document>.
- Claeys-Kulik, Anna-Lena, Thomas Ekman Jørgensen, and Henriette Stöber. 2019. "Diversity, Equity and Inclusion in European Higher Education Institutions: Results from the INVITED Project." Unpublished manuscript, last modified March 21, 2024. https://eua.eu/downloads/publications/web_diversity%20equity%20and%20inclusion%20in%20european%20higher%20education%20institutions.pdf.
- Council of the European Union. 2015. "Advancing Gender Equality in the European Research Area: Council Conclusions (Adopted on 01.12.2015)." 14846/15. Accessed February 15, 2016. <http://data.consilium.europa.eu/doc/document/ST-14846-2015-INIT/en/pdf>.
- Das, Amita, U. Rajendra Acharya, Soumya S. Panda, and Sukanta Sabut. 2019. "Deep Learning Based Liver Cancer Detection Using Watershed Transform and Gaussian Mixture Model Techniques." *Cognitive Systems Research* 54: 165–75.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805*.
- Do, Salomé, Étienne Ollion, and Rubing Shen. 2022. "The Augmented Social Scientist: Using Sequential Transfer Learning to Annotate Millions of Texts with Human-Level Accuracy." *Sociological Methods & Research*, 00491241221134526.
- Doneva, Rositsa, Silvia Gaftandzhieva, and Kirina Boykova. 2022. "An Approach to Monitoring Gender Equality Plans Implementation." In *EDULEARN22 Proceedings: 14th International Conference on Education and New Learning Technologies*, edited by Louis Gómez Chova, Agustín López Martínez, and Joanna Lees, 3383–91. EDULEARN proceedings (Internet). Valencia: IATED Academy. <https://library.iated.org/publications/EDULEARN22>.
- Douglas Oloyede, Freya. 2014a. "Measuring Progress on Equality: Qualitative Evidence." Unpublished manuscript, last modified February 06, 2023. http://www.ecu.ac.uk/publications/files/measuring-progress-on-equality-qualitative-evidence.pdf/at_download/file.
- Douglas Oloyede, Freya. 2014b. "Measuring Progress on Equality: Qualitative Evidence." Unpublished manuscript, last modified February 06, 2023. http://www.ecu.ac.uk/publications/files/measuring-progress-on-equality-qualitative-evidence.pdf/at_download/file.
- ERAC Standing Working Group on Gender in Research and Innovation. 2021. "Gender Equality Plans as a Catalyst for Change: ERAC 1202/21." Unpublished manuscript, last modified December 09, 2022. <https://data.consilium.europa.eu/doc/document/ST-1202-2021-INIT/en/pdf>.



- European Commission. 2011. "Towards a European Framework for Research Careers." Unpublished manuscript, last modified April 12, 2019. https://cdn5.euraxess.org/sites/default/files/policy_library/towards_a_european_framework_for_research_careers_final.pdf.
- European Commission. 2016. "She Figures Handbook 2015." Unpublished manuscript, last modified August 28, 2024. <https://data.europa.eu/doi/10.2777/402057>.
- European Commission. 2020. "A New ERA for Research and Innovation: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions." Unpublished manuscript, last modified March 22, 2023. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0628>. Com(2020) 628 final.
- European Commission. 2021a. *Horizon Europe Guidance on Gender Equality Plans*. Luxembourg: Publications Office of the European Union. Accessed March 17, 2023. <https://op.europa.eu/en/publication-detail/-/publication/ffcb06c3-200a-11ec-bd8e-01aa75ed71a1/language-en/format-PDF/source-232129669>.
- European Commission. 2021b. "She Figures 2021: Gender in Research and Innovation - Statistics and Indicators." Unpublished manuscript, last modified December 02, 2021. <https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/67d5a207-4da1-11ec-91ac-01aa75ed71a1>.
- European Commission. 2021c. "She Figures Handbook 2021." Unpublished manuscript, last modified December 02, 2021. <https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/058103b5-4da0-11ec-91ac-01aa75ed71a1>.
- European Commission. 2022. "Towards Inclusive Gender Equality in Research and Innovation." https://research-and-innovation.ec.europa.eu/document/download/e5a7cdaf-5b24-40af-9f82-2870285da0aa_en. Factsheet.
- European Commission. 2025. "She Figures Handbook 2024." Unpublished manuscript, last modified February 26, 2025. <https://op.europa.eu/en/publication-detail/-/publication/05d87867-e838-11ef-b5e9-01aa75ed71a1/language-en>.
- European Institute for Gender Equality. 2016a. "Gender Equality in Academia and Research - GEAR Tool: Making a Gender Equality Plan." Unpublished manuscript, last modified September 20, 2018. <http://eige.europa.eu/sites/default/files/mh0716096enn.pdf>.
- European Institute for Gender Equality. 2016b. "Integrating Gender Equality into Academia and Research Organisations: Analytical Paper." Unpublished manuscript, last modified July 26, 2024. <https://eige.europa.eu/publications-resources/publications/integrating-gender-equality-academia-and-research-organisations-analytical-paper>.
- European Research Area and Innovation Committee. 2015. "European Research Area (ERA) Roadmap 2015-2020: ERAC 1208/15." Unpublished manuscript, last modified March 22, 2023. <https://era.gv.at/era/era-2000-2021/era-roadmap/european-era-roadmap-2015-2020/>.



- Eynon, Rebecca. 2013. "The Rise of Big Data: What Does It Mean for Education, Technology, and Media Research?" *Learning, Media and Technology* 38 (3): 237–40. <https://doi.org/10.1080/17439884.2013.771783>.
- Fan, Angela, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines et al. 2021. "Beyond English-Centric Multilingual Machine Translation." *Journal of Machine Learning Research* 22 (107): 1–48.
- Gates, Emily, and Lisa Dyson. 2017. "Implications of the Changing Conversation About Causality for Evaluators." *American Journal of Evaluation* 38 (1): 29–46. <https://doi.org/10.1177/1098214016644068>.
- Glez-Peña, Daniel, Anália Lourenço, Hugo López-Fernández, Miguel Reboiro-Jato, and Florentino Fdez-Riverola. 2013. "Web Scraping Technologies in an API World." *Briefings in bioinformatics* 15 (5): 788–97. <https://doi.org/10.1093/bib/bbt026>.
- Gregory-Smith, Ian. 2018. "Positive Action Towards Gender Equality: Evidence from the Athena SWAN Charter in UK Medical Schools." *British Journal of Industrial Relations* 56 (3): 463–83. <https://doi.org/10.1111/bjir.12252>.
- Guyan, Kevin, and Freya Douglas Oloyede. 2020. "Equality, Diversity and Inclusion in Research and Innovation: UK Review." Unpublished manuscript, last modified December 09, 2022. <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-EDI-EvidenceReviewUK.pdf>.
- Hadler, Patricia, Cornelia E. Neuert, Verena Ortmanns, and Angelika Stiegler. 2022. "Are You...? Asking Questions on Sex with a Third Category in Germany." *Field Methods* 34 (2): 91–107. <https://doi.org/10.1177/1525822X211072326>.
- Heidler, Julia, and Eva Reichwein. 2018. "The DFG Research-Oriented Standards on Gender Equality: Selected Findings from a Study on Implementation and Impact." DFG infobrief 1.18. Unpublished manuscript, last modified February 15, 2019. http://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/infobriefe/ib01_2018_en.pdf.
- Higher Education Authority. 2018. "Gender Action Plan 2018 - 2020: Report of the Gender Equality Taskforce." Unpublished manuscript, last modified January 19, 2023. <https://www.gov.ie/en/publication/3c3c6f-gender-action-plan-2018-2020/>. Accelerating gender equality in Irish higher education institutions.
- Hossin, Mohammad, and Md Nasir Sulaiman. 2015. "A Review on Evaluation Metrics for Data Classification Evaluations." *International journal of data mining & knowledge management process* 5 (2): 1.
- Inmon, William H., and Anthony Nesavich. 2007. *Tapping into Unstructured Data: Integrating Unstructured Data and Textual Analytics into Business Intelligence*. Pearson Education.
- Janetzko, Dietmar. 2016. "Nonreactive Data Collection Online." *The SAGE Handbook of Online Research Methods*, 76.
- Janetzko, Dietmar, and Roman Kennke. 2004. "Workshops During the German Online Research 04." Accessed March 15, 2024. https://www.gor.de/archive/gor04/work_e_2.htm.



- Kalpazidou Schmidt, Evanthia, Susanne Bühner, Martina Schraudner, Sybille Reidl, Jörg Müller, Rachel Palmén, Sigrid Haase et al. 2017. "A Conceptual Evaluation Framework for Promoting Gender Equality in Research and Innovation: Toolbox I - a Synthesis Report." Unpublished manuscript, last modified June 02, 2020. <https://www.efforti.eu/sites/default/files/2018-03/EFFORTI%20D3.3%20FINAL%20report%2027032018.pdf>. EFFORTI - Deliverable 3.3.
- Knapińska, Anna, and Magdalena Chrobak-Tatar. 2023. "Benchmarking Analysis of Monitoring/evaluation of GEPs." Unpublished manuscript, last modified January 25, 2024. https://genderaction.eu/wp-content/uploads/2023/07/GENDERACTIONplus_D6.1_Benchmarking-analysis-of-monitoringevaluation-of-GEPs.pdf. Deliverable 6.1.
- Kryukov, Denis. 2023. "Is Web Scraping Legal?" Accessed March 15, 2024. <https://infatica.io/blog/web-scraping-legality/#:~:text=Yes.,you%27re%20breaking%20a%20law.>
- Kulyk, Oleg. 2023. "Web Scraping Vs Web Crawling - Use Cases and Differences." Accessed March 27, 2024. <https://scrapingant.com/blog/web-scraping-vs-web-crawling>.
- Lazer, David, and Jason Radford. 2017. "Data Ex Machina: Introduction to Big Data." *Annual Review of Sociology*. <https://doi.org/10.1146/annurev-soc-060116-053457>.
- Lee, Raymond M. 2019. "Unobtrusive Methods." In *Handbook of Research Methods in Health Social Sciences*, edited by Pranee Liamputtong, 491–507. Singapore: Springer Singapore.
- Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. 2020. *Mining of Massive Datasets*: Cambridge University Press.
- Lewis, Mike, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. "Bart: Denoising Sequence-to-Sequence Pre-Training for Natural Language Generation, Translation, and Comprehension." *arXiv preprint arXiv:1910.13461*.
- Lindqvist, Anna, Marie Gustafsson Sendén, and Emma Renström. 2020. "What Is Gender, Anyway: A Review of the Options for Operationalising Gender." online first. *Psychology & Sexuality*. DOI: 10.1080/19419899.2020.1729844.
- Lipinsky, Anke, and Maria Schäfer. 2016. "INTEGER Guidelines for the Self-Assessment of Transformational Gender Action Plans Set up in Higher Education and Research Institutions." *Cologne, Germany: GESIS Leibniz Institute for the Social Sciences*. [Google Scholar].
- Löther, Andrea. 2019. "Is It Working? An Impact Evaluation of the German "Women Professors Program"." *Social Sciences* 8 (4): 116. <https://doi.org/10.3390/socsci8040116>.
- Löther, Andrea, Mazlum Karataş, and Lena Weber. 2023. "Literature Review Data Monitoring: Deliverable D2.1.".
- Löther, Andrea, Mazlum Karataş, and Lena Weber. 2024a. "D3.1 GEP Prevalence Monitoring Indicator Framework V1.".



- Löther, Andrea, Mazlum Karataş, and Lena Weber. 2024b. "D3.3 Report on GEP Monitoring Survey + Webcrawl Result."
- Luscombe, Alex, Kevin Dick, and Kevin Walby. 2022. "Algorithmic Thinking in the Public Interest: Navigating Technical, Legal, and Ethical Hurdles to Web Scraping in the Social Sciences." *Qual Quant* 56 (3): 1023–44. <https://doi.org/10.1007/s11135-021-01164-0>.
- Macanovic, Ana. 2022. "Text Mining for Social Science - the State and the Future of Computational Text Analysis in Sociology." *Social Science Research* 108:102784. <https://doi.org/10.1016/j.ssresearch.2022.102784>.
- Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. 2012. *Introduction to Information Retrieval*: Cambridge University Press.
- Mazel, Adam. 2023. "Digital Scholarship: Computational Text Analysis." Accessed February 14, 2024. <https://libguides.union.edu/digital-scholarship/cta>.
- McDonnell, Diarmuid. 2020. "Web-Scraping for Social Science Research: Websites as a Source of Data." Accessed January 21, 2023. https://dam.ukdataservice.ac.uk/media/622652/web_webscrapingwebsites23apr20.pdf.
- Mergaert, Lut, Marina Cacace, and Marcela Linková. 2022. "Gender Equality Impact Drivers Revisited: Assessing Institutional Capacity in Research and Higher Education Institutions." *Social Sciences* 11 (9): 379. <https://doi.org/10.3390/socsci11090379>.
- Meuschke, Norman, Apurva Jagdale, Timo Spinde, Jelena Mitrović, and Bela Gipp. 2023. "A Benchmark of PDF Information Extraction Tools Using a Multi-Task and Multi-Domain Evaluation Framework for Academic Documents." In *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity*. Vol. 13972, 383–405. LNCS. Cham: Springer Nature Switzerland.
- Mitkov, Ruslan. 2022. *The Oxford Handbook of Computational Linguistics*: Oxford University Press.
- Mour, Chloé. 2022. "Anti-Discrimination Policies in Higher Education Institutions: An Interdisciplinary Scoping Review." LIEPP Working Paper 132 | hal-03625791. <https://hal-sciencespo.archives-ouvertes.fr/hal-03625791/document>.
- Nigam, Harshit, and Prantik Biswas. 2021. "From Web Scraping to Web Crawling." In *Applications of Artificial Intelligence and Machine Learning: Select Proceedings of ICAAAIML 2020*, 97–112.
- Nimo, Silvana María Sánchez. 2021. "The Evaluation of Equality Plans at the University of the Balearic Islands: A Proposal for a Gender-Sensitive Indicator System." *Inv. fem.* 12 (2): 439–48. <https://doi.org/10.5209/infe.72029>.
- Noortje, Marres, and Weltevrede Esther. 2012. "Scraping the Social?" <https://doi.org/10.1080/17530350.2013.772070>.
- O'Connor, Pat, and Gemma Irvine. 2020. "Multi-Level State Interventions and Gender Equality in Higher Education Institutions: The Irish Case." *Administrative Sciences* 10 (4): 98. <https://doi.org/10.3390/admsci10040098>.



- Olmedilla, M., M. R. Martínez-Torres, and S. L. Toral. 2016. "Harvesting Big Data in Social Science: A Methodological Approach for Collecting Online User-Generated Content." *Computer Standards & Interfaces* 46:79–87. <https://doi.org/10.1016/j.csi.2016.02.003>.
- Ouyang, Long, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang et al. 2022. "Training Language Models to Follow Instructions with Human Feedback." *Advances in neural information processing systems* 35: 27730–44.
- Ovseiko, Pavel V., Pololi H. Linda, Edmunds Laurel, Civian Jan, Daly Mary, and Buchan Alastair. 2019. "Creating a More Supportive and Inclusive University Culture: A Mixed-Methods Interdisciplinary Comparative Analysis of Medical and Social Sciences at the University of Oxford." *Interdisciplinary Science Reviews* 44 (2): 166–91.
- Palmén, Rachel, and Evanthia Kalpazidou Schmidt. 2019. "Analysing Facilitating and Hindering Factors for Implementing Gender Equality Interventions in R&I: Structures and Processes." *Evaluation and program planning* 77:101726. <https://doi.org/10.1016/j.evalprogplan.2019.101726>.
- Palmén, Rachel, Evanthia Kalpazidou Schmidt, Clemens Striebing, Sybille Reidl, Susanne Bühner, and Dóra Groó. 2019a. "Measuring Gender in R&I – Theories, Methods, and Experience." *Interdisciplinary Science Reviews* 44 (2): 154–65. <https://doi.org/10.1080/03080188.2019.1603873>.
- Palmén, Rachel, Sybille Reidl, Susanne Bühner, Evanthia Kalpazidou Schmidt, Florian Holzinger, Dora Groo, Dorottya Rigler et al. 2019b. "EFFORTI –Deliverable 4.2 Synthesis Report." Unpublished manuscript, last modified November 15, 2019. https://www.efforti.eu/sites/default/files/2019-11/D4.2_EFFORTI_Comparative_Report_FINAL_September2019.pdf.
- Powers, David M. W. 2020. "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation." *arXiv preprint arXiv:2010.16061*.
- Rastan, Roya, Hye-Young Paik, and John Shepherd. 2019. "TEXUS: A Unified Framework for Extracting and Understanding Tables in PDF Documents." *Information Processing & Management* 56 (3): 895–918. <https://doi.org/10.1016/j.ipm.2019.01.008>.
- Research Council of Norway. 2016. "Synthesis Report on Selected Structural Change Initiatives and Indicators for Monitoring of State-of-Play and Progress." Unpublished manuscript, last modified November 28, 2016. http://www.gender-net.eu/IMG/pdf/Rapport_Gender-net_2-8-v3.pdf.
- Sharma, Arvind K., Vandana Shrivastava, and Harvir Singh. 2021. "Experimental Performance Analysis of Web Crawlers Using Single and Multi-Threaded Web Crawling and Indexing Algorithm for the Application of Smart Web Contents." *Materials Today: Proceedings* 37: 1403–8.
- Sheinbaum, Gil. 2023. "Web Crawling Vs. Web Scraping: Key Differences." Accessed March 27, 2024. <https://nimbleway.com/blog/web-scraping/web-crawling-vs-web-scraping/>.
- Shuyo, Nakatani. 2010. "Language Detection Library for Java." Accessed May 23, 2024. <https://www.slideshare.net/slideshow/language-detection-library-for-java/6014274>.



- Smith, Ray. 2007. "An Overview of the Tesseract OCR Engine." In *ICDAR '07: Proceedings of the Ninth International Conference on Document Analysis and Recognition*, 629–33. Washington, DC, USA: IEEE Computer Society. <https://storage.googleapis.com/pub-tools-public-publication-data/pdf/33418.pdf>.
- Socher, Richard, Milind Ganjoo, Christopher D. Manning, and Andrew Ng. 2013. "Zero-Shot Learning Through Cross-Modal Transfer." *Advances in neural information processing systems* 26.
- Spörlein, Christoph, and Elmar Schlueter. 2021. "Ethnic Insults in YouTube Comments: Social Contagion and Selection Effects During the German "Refugee Crisis"." *European Sociological Review* 37 (3): 411–28. <https://doi.org/10.1093/esr/jcaa053>.
- Stadler, Bettina, and Angela Wroblewski. 2021. "Wissen in Zahlen. Potenziale Von Gender-Monitoring Im Gleichstellungspolitischen Prozess Am Beispiel Österreichischer Universitäten. (German)." *GENDER* 13 (2): 142–58. <https://doi.org/10.3224/gender.v13i2.10>.
- Stadler, Gertraud, Marie Chesaniuk, Stephanie Haering, Julia Roseman, Vera Maren Straßburger, Schraudner Martina, Ahmad Aisha-Nusrat et al. 2023. "Diversified Innovations in the Health Sciences: Proposal for a Diversity Minimal Item Set (DiMIS)." *Sustainable Chemistry and Pharmacy* 33:101072. <https://doi.org/10.1016/j.scp.2023.101072>.
- Stepan-Norris, Judith, and Jasmine Kerrissey. 2016. "Enhancing Gender Equity in Academia: Lessons from the ADVANCE Program." *Sociological Perspectives* 59 (2): 225–45. <https://doi.org/10.1177/0731121415582103>.
- Subdirección General para el Emprendimiento, la Igualdad en la Empresa y la Negociación Colectiva de Mujeres. 2021. "Guía Para La Elaboración de Planes de Igualdad En Las Empresas." https://www.igualdadenaempresa.es/asesoramiento/herramientas-igualdad/docs/Guia_pdi.pdf.
- Technische Universität Wien. 2021. "GEECCO Evaluation & Monitoring Tutorials." Accessed June 15, 2023. <https://www.tuwien.at/en/tu-wien/organisation/zentrale-bereiche/genderkompetenz/gender-in-der-forschung/geschlecht-innovation/geecco-results/evaluation-and-monitoring-tutorials>.
- Timmers, Tanya M., Tineke M. Willemsen, and Kea G. Tijdens. 2010. "Gender Diversity Policies in Universities: A Multi-Perspective Framework of Policy Measures." *High Educ* 59 (6): 719–35. <https://doi.org/10.1007/s10734-009-9276-z>.
- Wang, Lulu. 2022. "Deep Learning Techniques to Diagnose Lung Cancer." *Cancers* 14 (22): 5569.
- Webb, J. Eugene, T. Donald Campbell, D. Richard Schwartz, and Lee Sechrest. 1966, pr. [1976]. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. 12. pr. Rand McNally sociology series. Chicago: Rand McNally.
- Webb, J. Eugene, T. Donald Campbell, D. Richard Schwartz, and Lee Sechrest. 1999. *Unobtrusive Measures*: Sage Publications.



- Wroblewski, Angela. 2023. "First Report on Monitoring ERA Action Implementation at National Level." Unpublished manuscript, last modified March 15, 2024.
https://genderaction.eu/wp-content/uploads/2023/09/GENDERACTIONplus_D5.1_First-report-on-monitoring-ERA-action-implementation-at-national-level.pdf. Deliverable 5.1.
- Wroblewski, Angela, Susanne Bühner, Andrea Leitner, and Cheng Fan. 2015. "Monitoring the Evolution and Benefits of Responsible Research and Innovation (MoRRI): Analytical Report on the Gender Equality Dimension." Unpublished manuscript, last modified March 22, 2018.
https://www.isi.fraunhofer.de/content/dam/isi/dokumente/ccp/2015/morri_analytical_report.pdf.
- Wroblewski, Angela, Udo Kelle, and Florian Reith, eds. 2016. *Gleichstellung Messbar Machen: Grundlagen Und Anwendungen Von Gender- Und Gleichstellungsindikatoren*. Wiesbaden: Springer VS.
- Wroblewski, Angela, and Andrea Leitner. 2013. "Analyse von Gender - Indikatoren: WB - Kennzahl 1.A.5 Gender Pay Gap und Datenbedarfskennzahl 1.3 Geschlechterrepräsentanz im Berufungsverfahren." Unpublished manuscript, last modified January 27, 2023. http://www.equi.at/dateien/IHS_Bericht_Gender_Indikator.pdf.
 Endbericht. Studie im Auftrag des BMWF.
- Wyatt Knowlton, Lisa, and Cynthia C. Phillips. 2009. *The Logic Model Guidebook. Better Strategies for Great Results*. Los Angeles, California: Sage Publications.
- Yarkoni, Tal, Dean Eckles, James Heathers, and Margaret C. Levenstein, Paul E. Smaldino, and Julia Lane. 2021. "Enhancing and Accelerating Social Science via Automation: Challenges and Opportunities." Accessed January 25, 2024.

10 Appendix

10.1 List of Indicators

Prevalence

- Indicator 1.1 A gender equality plan or a written and formal institutional strategy that fosters gender equality) exists in a research-performing or research-funding organisation.
- Indicator 1.2 The existence of a GEP in an organisation is linked to national and European initiatives or requirements or third-party funding in terms of::
- Participation in EU structural change projects, e.g., Horizon 2020, Horizon Europe
 - Received funding other than EU projects, e.g., from national RFOs
 - Received non-monetary support
 - Legal requirements or requirements linked to getting research funding

Characteristics



- Indicator 2.1 Ownership and hierarchical level of the organisation that adopted the GEP: ownership of the GEP
- Indicator 2.2 Publication of the GEP: publication accessible to people outside the organisation, internal publication (accessible to all members of the organisation), internal document (accessible only to the management)
- Indicator 2.3 Date of the first and current plan's adoption and time frame of the current plan
- Indicator 2.4 Areas of activity (grouped according to the specification for Horizon Europe) and target groups addressed in the GEP
- Prevalence: Are there measures in each area of activity
 - Quantity: How many measures in each area of activity
- Indicator 2.5 ***The GEP integrates inclusive approaches like::***
- Gender diversity
 - Intra-categorical differences inside the gender groups (e.g. women with migration background, Black women, etc.)
 - Intersectionality (on a more conceptual level, not necessarily as a term)
 - Diversity (inequalities addressed beside each other/"celebrating the differences" approach)
- Indicator 2.6 The GEP addresses different inequalities (race, class/social background, age, etc.).

Implementation

- Indicator 3.1 Assessment of planned and implemented gender equality measures (by areas of activity)
- Level of realisation: planned – started – implemented
 - Institutional coverage: implemented across the institution, implemented in some departments, not implemented anywhere



- Indicator 3.2 Financial resources are dedicated to gender equality measures/implementation of GEP.
- Indicator 3.3 A gender equality unit, gender equality committee and/or dedicated staff are in place.
- Indicator 3.4 Gender equality is integrated into institutional/internal regulations (appointment regulation, basic rules of the institution, etc.).
- Indicator 3.5 Internal and external stakeholders are involved in the GEP implementation.
- Indicator 3.6 A policy for monitoring/controlling gender equality measures and the GEP implementation is in place.
- Indicator 3.7 Sex-disaggregated data are collected and published.
- Indicator 3.8 Gender diversity and intersectional perspectives are integrated into the monitoring or data collection:
- Inclusion of a broader and non-binary understanding of gender diversity
 - Inclusion of other inequalities in conjunction with gender



Impact

- Indicator 4.1 The GEP includes a reflection or description of the GEP impact. The organisation reflects its understanding of the GEP impact.
- Indicator 4.2 The GEP designates data and methods (e.g. evaluations, surveys or qualitative data) to assess the GEP impact.
- Indicator 4.2 Changes in the participation of women within a set period (5-10 years) and comparing RPOs with/without GEP (or gender equality measures) in leading positions
- Indicator 4.3 Rating of the changes in the areas of activity (gender balance, awareness, knowledge about gender (and other) inequalities, promotion and recruitment procedures, work-life-balance) and relevance of GEPs for the achieved changes



10.2 Web Scraping: Search Terms in the Full Study

Table 47 Search terms used in web-scraping.

Country	Search term 1	Search term 2	Search term 3	Search term 4	Search term 5	Search term 6	Search term 7	Search term 8	Search term 9	Search term 10	Search term 11	Search term 12	Search term 13	Search term 14
Austria	Gleichstellungsplan	Frauenförderungsplan	Aktionsplan AND Gleichstellung	Frauenförderplan	Chancengleichheitsplan									
Belgium - France	Charte pour l'égalité de genre	plan d'action pour l'égalité	plan d'action sur l'égalité	plan égalité hommes-femmes	Plan Egalité de Genre									
Belgium - Dutch	Gendergelijkheidsplan	Gendergelijkheidsplan	Gendergelijkheid en diversiteitsplan	Diversiteit, Inclusie en Gender Equality Plan										
Bosnia and Herzegovina	Gender akcioni plan	Plan rodne ravnopravnosti	Акциони план за родну равноправност	Gender akcijski plan	Гендер акциони план	Akcioni plan za rodnu ravnopravnost	Akcijski plan za ravnopravnost spolova	Акциони план родне равноправности	Povelja o jednakosti žena i muškaraca	Plan za rodnu ravnopravnost	Akcioni plan rodne ravnopravnosti	Rodni akcioni plan	Povelja za ravnopravnost spolova	
Bulgaria	План за равнопоставеност между половете	План за равенство между половете	План на дейности за осигуряване на равнопоставеност	План за насърчаване на равенството между половете	Харта за равнопоставеност между половете	Харта за равенството между половете	План за действие за равнопоставеност	План за действие за равенството между половете	План за действие на полове					



			на половете				между половете							
Croatia	Akcijski plan za rodnu ravnopravnost	Plan za ravnopravnost spolova	Plan rodne ravnopravnosti	Akcijski plan za ravnopravnost spolova	Plan ravnopravnosti spolova	Povelja za ravnopravnost žena i muškaraca	Povelja o ravnopravnosti spolova	Povelja za ravnopravnost spolova	Povelja a ravnopravnosti spolova	Gender akcijski plan				
Cyprus	Σχέδιο Ισότητας των Φύλων	Σχέδιο δράσης για την ισότητα των φύλων	Athena Swan	Σχέδιο δράσης για την Έμφυλη Ισότητα	Χάρτης για την Ισότητα των φύλων									
Czech Republic	Plán pro rovnost žen a mužů	Gender akční plán	Akční plán pro rovnost žen a mužů	Charta rovnosti žen a mužů	Akční plán pro rovné odměňování žen a mužů	Akční plán genderové rovnosti								
Denmark	Ligestillingsplan	Handleplan for ligestilling, diversitet og inklusion	Handleplan for ligestilling og diversitet	Action plan for gender equality	Action plan for equality									



Estonia	Soolise võrdõiguslikkuse kava	soolise võrdõiguslikkuse tegevuskava	soolise võrdõiguslikkuse plaan	võrdse kohtlemise kava	Soolise võrdõiguslikkuse põhimõtted ja tegevuskava									
Finland	Tasa-arvosuunnitelma	Tasa-arvo- ja yhdenvertaisuussuunnitelma	Tasa-arvosuunnitelma	Sukupuolten tasa-arvon edistämisen toimintasuunnitelma										
France	plan d'action pour l'égalité	plan d'action sur l'égalité	Plan d'action égalité femmes hommes	plan d'égalité	plan d'action relatifs à l'égalité	Plan pour l'égalité	Charte pour l'égalité							
Germany	Gleichstellungskonzept	Gleichstellungsplan	Frauenförderplan	Rahmenplan AND Gleichstellung	Rahmenplan AND Chancengleichheit	Chancengleichheitsplan	Gleichstellungsstrategie	Aktionsplan AND Gleichstellung						
Greece	Σχέδιο για την ισότητα των φύλων	Σχέδιο Δράσης για την Ισότητα των Φύλων	Σχέδιο Δράσης για την Έμφυλη Ισότητα	Σχεδίου Δράσης για την Έμφυλη Ισότητα										
Hungary	Nemek közötti egyenlőség terv	Nemi egyenlőségi terv	Nemi esélyegyenlőségi terv	A nemek közötti egyenlőség	Nemek közötti egyenlőség terv									



				cselekvési terve	támogató terv									
Ireland	Gender Equality Plan	Gender Action Plan	Gender Equality Action Plan	Athena SWAN	Gender equality charter									
Israel	תוכנית שוויון מגדרי	תכנית פעולה לשיוויון מגדרי	תוכנית פעולה לשוויון בין המינים	תכנית להטמעת שוויון מגדרי	קידום שיוויון מגדרי									
Italy	Il piano di parità di genere	Piano di azione di genere	Piano d'azione per l'uguaglianza di genere	Carta dell'uguaglianza di genere	Piano per l'uguaglianza di Genere									
Latvia	Dzimumu līdztiesības plāns	Dzimumu līdztiesības rīcības plāns	Dzimumu līdztiesības pamatprin cipi un plāns	Dzimumu līdztiesības īstenošanas plāns	Dzimumu līdztiesības pamatprin cipu īstenošanas plāns	Dzimumu līdztiesības harta								
Lithuania	Lyčių lygybės planas	Lyčių lygybės veiksmų planas	Moterų ir vyrų lygių galimybių veiksmų plano	Lyčių lygybės ir įvairovės politika										



Luxembourg	Gleichstellungsplan	Frauenförderplan	Chancengleichheitsplan	Aktionsplan AND Gleichstellung	Rahmenplan AND Gleichstellung	Rahmenplan AND Chancengleichheit	Gleichstellungskonzept	Gleichstellungsstrategie						
Malta	Gender Equality Plan	Gender Action Plan	Gender Equality Action Plan	Gender equality charter										
Netherlands	Gender gelijks plan	Gendergelijksheidsplan	Gendergelijkheid en diversiteits plan	Diversiteit, Inclusie en Gender Equality Plan										
Norway	Handlingsplan for likestilling	Handlingsplan for kjønns	Handlingsplan for kjønnsmanngfold											
Poland	Plan równości płci	Plan na rzecz równości płci	Plan równości	Plan wdrażania polityki równości płci	Plan równości szans	Plan na rzecz równych szans	Plan na rzecz równości kobiet i mężczyzn	Plan działań na rzecz równości płci	Plan Równości Szans Płci	Program działań na rzecz równości płci	Strategia na rzecz równości płci			
Portugal	Plano para a igualdade de género	Plano de Ação de Género	Carta de Princípios para a Igualdade de Género	Plano de Ação para a Igualdade de Oportunidades	Estratégia para a Igualdade de Género	Plano de Ação AND Igualdade de Género								



Romania	Planul de egalitate de gen	Plan de egalitate de gen	Planul de acțiune pentru egalitatea de gen	Carta egalității de gen										
Serbia	Plan o rodnoj ravnopravnosti	Акциони план за родну равноправност	План за постизање е родне равноправности	Акциони план родне равноправности	План мера за остваривање и унапређење родне равноправности	Plan mera za ostvarivanje i unapređenje rodne ravnopravnosti	Akcioni plan za ostvarivanje rodne ravnopravnosti	План за родну равноправност	Akcioni plan za rodnu ravnopravnost	Plan rodne ravnopravnosti	Родни акциони план	Evropska povelja o rodnoj ravnopravnosti	Европска повеља о родној равноправности	Povelja Atena Svon
Slovakia	Plán rodovej rovnosti	Rodový akčný plán	Akčný plán pre rodovú rovnosť	Charta rodovej rovnosti	Stratégia rodovej rovnosti	Stratégia pre rodovú rovnosť								
Slovenia	Načrt za enakost spolov	Akcijski načrt za uveljavljanje enakosti spolov	Načrt enakosti spolov	Načrt za uveljavljanje enakosti spolov	Akcijski načrt za enakost spolov	Akcijski načrt vzpostaviti ve enakih možnosti glede na spol	Listina o enakosti spolov							
Spain	Plan de igualdad de género	Plan de acción de género	Plan para la igualdad entre mujeres y hombres											



Sweden	jämställdhetsplan	action plan for gender mainstreaming	Handlingsplan för jämställdhetsintegrering	Plan för jämställdhetsintegrering	handlingsplan för jämställdhet	Handlingsplan för jämställdhet mellan könen	Handlingsplan för Lika villkor och jämställdhet	jämställdhetsstadgan						
Switzerland - French	plan d'action pour l'égalité	plan d'action sur l'égalité	Plan d'action égalité femmes hommes	plan d'égalité	plan d'action relatifs à l'égalité	Plan pour l'égalité	Charte pour l'égalité							
Switzerland - German	Gleichstellungsplan	Chancengleichheitsplan	Aktionsplan AND Chancengleichheit	Aktionsplan AND Gleichstellung	Frauenförderplan									
Switzerland - Italian	Il piano di parità di genere	Piano di azione di genere	Piano d'azione per l'uguaglianza di genere	Carta dell'uguaglianza di genere	Piano per l'uguaglianza di Genere									
United Kingdom	Gender Equality Plan	Gender Action Plan	Gender Equality Action Plan	Athena SWAN	Gender equality charter									



10.3 Search Terms for Meta-Analysis in the Full Study

Country	Search term 1	Search term 2	Search term 3	Search term 4	Search term 5	Search term 6	Search term 7	Search term 8	Search term 9	Search term 10	Search term 11
Austria	Gleichstellungsplan	Frauenförderungsplan	Frauenförderungsplan	Chancengleichheitsplan	Aktionsplan	Frauenförderplan	frauenförderplan				
Belgium - France	paep	egalite plan	action egalite	pae	actions egalite	action égalité					
Belgium - Dutch	genderplan	gender plan	gender gelijkheidsplan	plan genre	Gendergelijkheid	Gendergelijkheidsplan					
Bosnia and Herzegovina	Gender akcioni plan	gender ap	plan rodne ravnopravnosti	LGAP	AKCIJSKI PLAN	Akcioni Planovi	rodnu ravnopravnost	rodne ravnopravnosti			
Bulgaria	ravnopostavenost	План за равнопоставеност на половете	План за равнопоставеност между половете	План за равнопоставеност	Plan ravenstvo	ravenstvo	План равнопоставеност				
Croatia	Rodna ravnopravnost	ravnopravnost spolova	ravnopravnost i spolova	rodne ravnopravnosti	Akcijski plan						
Cyprus	ΣΧΕΔΙΟ ΔΡΑΣΗΣ	ΣΧΕΔΙΟ ΙΣΟΤΗΤΑΣ	SDIF	ΣΔΙΦ	ISDIF	ΙΣΔΙΦ	sxedio isotitas filon	ΕΣΔΙΦ	sxedio drasis	ESDIF	ΣΔΕΙ



Czech Republic	Plan rovnosti	Plán genderové rovnosti	Plan genderove rovnosti	AP pro rovnost	Akční plán	Akcni plan	Genderová rovnost	Genderova rovnost	Strategie rovnosti		
Denmark	Ligestillingsplan	Ligestillingshandleplan	handleplan	handlingsplan	ligestilling						
Estonia	soolise vordoiguslikkuse	Soolise võrdõiguslikkuse	SVK								
Finland	Tasa arvosuunnitelma	Tasa arvosuunnitelmaohje	tasa arvosuunnitelminen	yhdenvertaisuussuunnitelma	TASAARVOSUUNNITELMA	tasa arvo suunnitelma					
France	paep	egalite plan	action egalite	pae	actions egalite	action égalité					
Germany	Gleichstellungsrahmenplan	Gleichstellungskonzept	Gleichstellungsplan	Rahmenplan	Chancengleichheitsplan	Aktionsplan	Frauenfoerderplan	frauenförderplan			
Greece	ΣΧΕΔΙΟ ΔΡΑΣΗΣ	ΣΧΕΔΙΟ ΙΣΟΤΗΤΑΣ	SDIF	ΣΔΙΦ	ISDIF	ΙΣΔΙΦ	sxedio isotitas filon	ΕΣΔΙΦ	sxedio drasis	ESDIF	ΣΔΕΙ
Hungary	Nemek közötti egyenlőségi terv	Nemek kozotti egyenlosegi terv	nemek kozotti eselyegyenlosegi	NemekKozottiEselyegyenlosegiTerv	Nemi Egyenlőség i Terv	Nemi Egyenlosegi Terv	nemi eselyegyenlosegi terv				
Ireland	gender equality plan	GEAP	Action Plan	GAP	Gender Equality Action Plan	genderequalityplan					



Israel	תקינה-מגדרית	שוויון מגדרי	תוכנית פעולה בנושא מגדר								
Italy	Piano uguaglianza di genere	Piano per la parità di genere	piano uguaglianza								
Latvia	Dzimumu līdztiesības plans	Dzimumu līdztiesības plana	dzimuma līdztiesības	dzimuma līdztiesības	Dzimumu līdztiesību	Dzimumlīdztiesība					
Lithuania	Lyčių lygybės planas	lyčių lygybės planas	Lyčių lygybės	lyčių lygybės	lygybės ir įvairovės						
Luxembourg	Gleichstellungsrahmenplan	Gleichstellungskonzept	Gleichstellungsplan	Rahmenplan	Chancengleichheitsplan	Aktionsplan	Frauenförderplan	frauenförderplan			
Malta	gender equality plan	GEAP	Action Plan	GAP	Gender Equality Action Plan	genderequality plan					
Netherlands	genderplan	gender plan	gender gelijkheidsplan	plan genre	Gendergelijkheid	Gendergelijkheidsplan					
Norway	handlingsplan	likestilling	Handlingsplan for likestilling								
Poland	plan równości płci	PRP	Plan RP	RP	rownosci szans	Plan Równości Płci	plan równosci	rownych szans			



Portugal	PI Género	PI Genero	Plano Igualdade	plano igualdade genero	PlanoligualdadeGenero	Planoligualdad e	Igualdad de Género	Plano para a Igualdade	Plano		
Romania	Plan egalitate	Planul de Egalitate	Plan pentru egalitatea	Plan de egalitate							
Serbia	Plan rodne ravnopravnosti	akcioni plan RR	rodne ravnopravnosti	Акциони план за родну равноправност	LAP	rodnu ravnopravnost	LAP PR	Plan rodna ravnopravnost	lokalno akcioni plan		
Slovakia	Plán rodovej rovnosti	Plan rodovej rovnosti	rovnosti zien muzov	Akčný plán	Akcny plan	strategia rovnosti					
Slovenia	Načrt za enakost spolov	Nacrt za enakost spolov	Nacrt enakosti spolov	NES	načrt enakosti spolov	Uveljavljanje enakosti spolov					
Spain	Plan de Igualdad	La Igualdad de Género	PLANIGUALDAD	Plan Igualdad							
Sweden	Jamstalldhetssplan	Jamstalldhetslagen	jämställdhetsintegrering	jämställdhetsplan	handlingsplan						
Switzerland - French	paep	egalite plan	action egalite	pae	actions egalite	action égalité					
Switzerland - German	Gleichstellungsplan	Rahmenplan	Chancengleichheitsplan	Aktionsplan	Frauenförderplan	frauenförderplan					

D3.2 GEP Prevalence Monitoring Indicator Framework v2



Switzerland - Italian	Piano uguaglianza di genere	Piano per la parità di genere	piano uguaglianza								
United Kingdom	gender equality plan	GEAP	Action Plan	GAP	Gender Equality Action Plan	genderequalityplan					

10.4 Text Analysis: Search Terms

Prevalence

Indicator 1.2 The existence of a GEP in an organisation is linked to national and European initiatives or requirements or third-party funding in terms of: Requirements linked to getting European research funding (HORIZON Europe)

- Participation in EU structural change projects, e.g., Horizon 2020, Horizon Europe

Search Terms = "Horizon Europe", "Eligibility criterion", "Horizon research funding"

Characteristics

Indicator 2.3 Time frame of the current plan

Search Pattern = `r\s*\d{4}\s*? [\u002D\u0058A\u005BE\u2010\u2011\u2012\u2013`

`\u2014\u2015\u202E3A\u202E3B\uFE58\uFE63\uFF0D]\s*\d{4}\s*?'`

Indicator 2.4 Actions of awareness-raising and training addressed in the GEP

Search Terms = "awareness training", "raise awareness", "increase awareness", "awareness-raising", "unconscious bias", "gender stereotypes", "recruitment training", "selection training", "sensitisation", "sensitivity to gender equality", "capacity-building training", "gender equality training", "training sessions", "training of the recruitment committees", "raising awareness", "training courses", "training events", "awareness and competence development", "awareness of gender issues", "employee awareness", "gender equality awareness", "gender awareness", "unconscious gender bias", "awareness of gender equality"

Indicator 2.4 Work-life balance and organisational culture addressed in the GEP

Search Terms = "Work-life balance", "Caring responsibilities", "Care responsibilities", "Parental leave", "Family leave", "Maternity leave", "Paternity leave", "Career breaks", "Family-related break", "Return4r", "Parents", "Carers", "Flexible working time", "Flexibility of working time", "Working time arrangements", "Remote working", "Hybrid working", "Family audit", "Childcare facilities", "Childcare", "Nursery", "Day care center", "Family support", "Dignity at work", "Inclusive language", "Gender-sensitive language", "Gendered language", "Welcoming", "Workload", "Open work environment", "Inclusive work environment", "Family life", "Child care leave"

Indicator 2.4 Gender balance in leadership and decision-making addressed in the GEP

Search Terms = "Gender balance", "Decision-making body", "Decision-making bodies", "Academic leadership", "Administrative leadership", "Representation", "Leadership position", "Gender quota", "Gender balance", "Election"



procedures", "Leadership competences", "Leadership development", "Decision-making process"

Indicator 2.4 Gender equality in recruitment and career progression addressed in the GEP

Search Terms = "Recruitment procedure", "Recruitment committee", "Recruitment panel", "Selection process", "Selection procedure", "Promotion committee", "Career progression", "Career development", "Evaluation criteria", "Appraisal criteria", "Performance assessment", "Performance appraisal", "Mentoring", "Career ladder", "Career path", "Career advancement", "Recruitment process", "Gender equality in recruitment", "Recruitment transparency"

Indicator 2.4 Measures against gender-based violence, including sexual harassment, addressed in the GEP

Search Terms = "Gender-based violence", "Sexual harassment", "Harassment", "Sexual violence", "Bullying", "Gender-based discrimination", "Sexualized violence", "Sexualised violence"

Indicator 2.4 Integration of the gender dimension into research and teaching content addressed in the GEP

Search Terms = "Gender dimension", "Sex analysis", "Gender analysis", "Gender impact assessment", "Incorporating gender", "Incorporating sex", "Research funding", "Research priorities", "Teaching activities", "Educational activities", "Gender perspectives", "Gender in research", "Gender in teaching"

Indicator 2.5 The GEP integrates intersectionality

Search Terms = "Intersectional", "Intersectionality"

Indicator 2.5 The GEP integrates diversity

Search Terms = "Diversity"

Indicator 2.5 The GEP integrates gender diversity

Search Terms = "Non-binary", "Gender-diverse", "Trans", "Transgender", "Gender diversity", "Gender identity", "All genders", "Other genders"

Indicator 2.6 The GEP addresses race

Search Terms = "Race", "Ethnic background", "Ethnic minorities", "Migrant", "Immigrant", "Ethnic origin", "Ethnicity", "Black", "People of colour", "Black", "People of color", "PoC", "BIPoC", "Indigenous", "Minority ethnic",

Indicator 2.6 The GEP addresses nationality



Search Term = “Nationality”

Indicator 2.6 The GEP addresses religion

Search Terms = “Religion”, “Religious”, “Religions”, “Religiose”, “Religiosity”, “Religiously”, “Muslim”, “Jewish”

Indicator 2.6 The GEP addresses class/social background

Search Terms = “Class”, “Social background”, “First generation”, “Socio-economic status”

Indicator 2.6 The GEP addresses age

Search Terms = “Age”

Indicator 2.6 The GEP addresses sexual orientation

Search Terms = “Sexual orientation”, “Gay”, “Lesbian”, “Bisexual”, “Queer”, “LGBTIQ”, “Sexual identity”, “LGBTI”,

Indicator 2.6 The GEP addresses disability

Search Terms = “Disability”, “Disable”, “Disabled”, “Disabling”, “Disablism”, “Disablist”, “Ableism”, “Physical Impairment”, “Physical limitation”, “Chronic disease”

Implementation

Indicator 3.3 A gender equality unit, gender equality committee and/or dedicated staff are in place. Search Terms = “Gender equality officer”, “Gender equality commissioner”, “Equal opportunities officer”, “Equal opportunity officer”, “Equal opportunity commissioner”, “Gender equality unit”, “Gender equality team”, “Equality officer”, “EDI manager”, “Diversity manager”, “Diversity team”, “EDI committee”, “EDI director”, “Working group”, “GEP working group”, “Gender equality working group”, “Equality committee”, “EDI committee”, “Athena SWAN steering group”, “EDI joint advisory committee”, “Gender equality steering group”, “Senate commission for gender equality”, “Senate commission for equality”, “Senate commission for equal opportunities”

Indicator 3.6 A policy for monitoring/controlling gender equality measures and the GEP implementation is in place

Search Terms = “Monitoring”, “Controlling”, “Evaluation”, “Quality assurance”, “Gender equality analysis”, “Gender equality audit”, “Gender equality assessment”, “Equality analysis”, “Equality audit”, “Equality assessment”, “Self-assessment”, “EDI surveying”, “Equality impact assessment”



Indicator 3.7 Sex-disaggregated data are collected and published: collection

Search Terms = “Data collection”, “Collection of data”, “Sex-disaggregated data”, “Gender statistics”, “Gender data”, “Collecting data”, “Share of women”, “Share of men”, “Proportion of women”, “Proportion of men”, “Regular statistics”

Indicator 3.7 Sex-disaggregated data are collected and published: publication

Search Terms = “Reporting”, “Gender data report”, “Gender equality report”, “Equality monitoring report”, “EDI reporting”, “Progress report”, “Publication of data”, “Gender distribution”, “Distribution of positions”, “Distribution of academic positions”, “Annual report”

Impact

Indicator 4.1 The GEP includes a reflection or description of the GEP impact. The organisation reflects its understanding of the GEP impact.

Search Terms = “Impact”, “Effect”, “Outcome”, “Output”

10.5 Survey: Questionnaire (English)

Survey to Monitor Gender Equality Plans

Starting page

Welcome to the survey to monitor gender equality plans.

Please select your preferred language.

- ☐ English
- ☐ German – Deutsch
- ☐ Spanish – Español
- ☐ French – Français
- ☐ Polish – Polski

1 Introduction

Thank you for participating in this survey.

The online survey aims to monitor the existence of a Gender Equality Plan or an equivalent, its characteristics, its implementation, and an assessment of its impact. The survey is provided by GESIS - Leibniz-Institute for the Social Sciences as part of the INSPIRE project, funded by the European Union. GESIS is researching how to monitor Gender Equality Plans (GEPs) of Research Performing Organisations (Higher education institutions, public research institutions, private companies) and Research Funding Organisations in the European landscape.



If you don't have any information about gender equality activities in your organisation, please transmit the survey to someone with the knowledge.

Informed consent

We, Lena Weber, Andrea Löther and Mazlum Karataş – researchers at GESIS – hereby inform of the following:

1. Purpose of the study

GESIS – Leibniz-Institute for the Social Sciences undertakes the survey in the context of the INSPIRE project. INSPIRE (<https://inspirequality.eu>) is a project funded by the European Union to build Europe's Centre of Excellence on inclusive gender equality in research and innovation. It brings together cutting-edge knowledge, ambitious policy approaches, and innovative practices to provide a gateway for scholars, equality experts, practitioners, and trainers to connect and share resources, as well as co-create new ones. INSPIRE's ambitious research programme develops new, relevant indicators for developing inclusive Gender Equality Plans (GEP). The project conducts a GEP monitoring survey across Europe to identify the necessary conditions for GEP impact. INSPIRE counts on four Knowledge & Support Hubs (KSHs) led by academics and practitioners throughout Europe to develop cutting-edge knowledge on sustaining change, widening participation, intersectionality and fostering innovation and change in the private sector. These KSHs will provide support to 12 communities of practice to facilitate GEP implementation and foster mutual support for the co-development of innovative practices, customised training and pan-European data collection.

The online-survey is oriented to map GEP prevalence and impact in Research Performing Organisations and Research Funding Organisations in the European Research Area. The survey will cover key process and content features of GEPs as specified in the Eligibility Criteria for Horizon Europe.

2. Who takes part in the study

Participants in the online survey are the organisations in question in the EU 27 as well as Bosnia-Herzegovina, Israel, Norway, Serbia and Switzerland from the associated countries and the UK.

3. Procedure

The survey consists of six parts and will take approximately [15-25 minutes] of your time to complete. We ask you kindly to answer as many questions as you can. Answers are saved automatically when you click "Continue" in the online survey interface. It is possible to interrupt the survey and re-visit your answers later by clicking the same link.. During the survey, you can go back and modify your answer by clicking "Back" button. If you change your mind during the survey and do not wish to participate, you can simply close the browser. Once you have reached the last page and do not wish to make further changes to your answers, we kindly ask you to click "Continue".



4. Confidentiality

When filling out the online survey, we will ask you if you want to receive more information about the INSPIRE project from the INSPIRE consortium members. If so, you will be redirected to the INSPIRE homepage to fill in your contact details.

As the online-survey is hosted by Unipark, GESIS will have access to the answers you provide in the survey. However, before sharing any data with other INSPIRE consortium members, the data file is split so that the main part of the survey is completely anonymous.

The remaining data is analysed anonymously and published only in aggregated form. Organisation names are asked so that collaboration networks can be mapped.

We found your data (name, function and mail-address) via a web search performed by INNO Systems, one of the INSPIRE Consortium members. Your personal data will be treated in accordance with the privacy policy of GESIS, which you can access from the research organisation's website (<https://www.gesis.org/en/institute/data-protection>), and according to the requirements of the European General Data Protection Regulation (GDPR 2016/679) and the German General Data Protection Regulation "GDPR" (2018). You can exercise your rights of accessing, modifying, opposing, and cancelling your data by contacting us (inspire@gesis.org). Your data and data of the submitted online-survey will be stored until the end of the project (30 September 2026). After this period, the data will be blocked until the applicable expiry period has elapsed. You may request to erase your data before the project ends by sending an email to (inspire@gesis.org). In this case your data will be blocked until the applicable expiry period has elapsed.

5. Dissemination of findings

The findings of this study are part of a European project. Findings can be used for conference presentations and publications in academic journals and/or book chapters. INSPIRE strives to make research data FAIR (findable, accessible, interoperable, and reusable) and available to the public and other researchers through open access repositories. All dissemination of research findings will follow the requirements of the EU legal framework for privacy, data protection and security of personal data.

6. Feedback

If you wish so, you can receive an electronic copy of any publications resulting from the research conducted in the INSPIRE project. Please send any requests via email to inspire@gesis.org.

7. Complaints



Any complaints regarding the ethical aspects of this study should be directed to the email of Datenschutz-GESIS@he-c.de at GESIS.

8. Who to contact if you have any questions

If you have any questions about this project, please get in touch with the principal investigator Rachel Palmén via email: rpalmen@uoc.edu.

9. Consent

I consent to participate in this survey.

- ☐ Yes
- ☐ No

Option No – exit

2 Information about the organisation

First, we like to get some information about your organisation.

2.1 Country

List/drop-down menu

2.2 Field of activity of your organisation

- ☐ Higher Education Institution
- ☐ Research organisation
- ☐ Private Company
- ☐ Research Funding Organisation

2.3 Size of your organisation

Please indicate the number of employees (academic, administrative and technical staff).

- ☐ Less than 50
- ☐ 51 - 250
- ☐ 251 - 500
- ☐ 501 – 1000
- ☐ 1001 – 5000
- ☐ More than 5.000

2.4 Name of your organisation

We ask the organisations' names to link the responses with other characteristics. The data is analysed anonymously and published only in aggregated form.

- ☐ The name of organisation:



free text field

- ☐ I don't want to share the name

3. Prevalence

Our following questions concern the existence of a Gender Equality Plan and the motivations for adopting it.

3.1 Does your organisation approve a Gender Equality Plan or another institutional strategy to foster gender equality and/or diversity?

In this survey, we understand a Gender Equality Plan (GEP) as a written document that describes the institutional strategy to foster gender equality in your organisation. It could be named differently in your organisation. The GEP regularly identifies gender inequalities, describes measures to counteract them, sets targets for gender equality, and instruments to monitor the progress. We can also speak of a GEP when it addresses gender among other dimensions of inequality (such as race, class, sexual orientation, age, etc.). In the framework of this survey, we investigate the gender equality plan of the whole organisation, not of departments or single units.

Please indicate the kind of institutional strategy. Several responses are possible.

1. Gender Equality Plan (focusing primarily on gender equality)
2. Gender equality and Diversity plan (including several inequalities but focusing on gender)
3. Diversity, equity or inclusion plan (dealing with several inequalities without highlighting one)
4. No such plan
5. Not known
6. No answer

Explanatory text for:

Filter 0 = 5 or 6

If you don't have any information about gender equality activities in your organisation, please transmit the survey to someone with the knowledge.

3.2 Does the equity or diversity plan integrate targets and measures to foster gender equality?

Filter: 0 = (3)

- ☐ Yes
☐ No



- ☐ Not known
- ☐ No answer

Explanatory text for

Filter: 3.1 = 1 + 2 or 3

If your organisation has adopted a diversity plan or strategy in addition to the gender equality plan, the questions in the survey deal exclusively with the gender equality plan.

3.3 In what language/s is the Gender Equality Plan available in your organisation?

Filter: 0 = 1-3

Please select from the list. Several answers are possible:

- ☐ National language/s
- ☐ English
- ☐ Other language

3.4 Has your organisation received funding or non-monetary support to set up a gender equality plan? your organisation received funding or non-monetary support to set up a gender equality plan?

Filter: 0 = 1-3

Funding for setting up a gender equality plan does not include financing gender equality measures. Non-monetary support does not include support for gender equality measures in general or stakeholders of your organisations engaged in setting up the gender equality plan.

Please select from the list. Several answers are possible:

- ☐ Participation in a structural change project funded by the EU (e.g., Horizon 2020)
- ☐ Funding from a national agency
- ☐ Training or counselling offered nationally or regionally
- ☐ Cooperation in professional networks or networks of universities or research organisations
- ☐ Not known
- ☐ No answer

3.5 HORIZON Europe requires organisations to submit a gender equality plan (GEP) if they apply for research funding. Did this requirement influence your organisation to set up a GEP?

Filter: 3.1 = 1-3



- ☐ Yes – Our organisation set up a GEP because of this requirement.
- ☐ Yes – Our organisation adopted an existing plan to meet the requirements of HORIZON Europe.
- ☐ No – Our organisation set up a GEP before HORIZON Europe started (2021).
- ☐ No – Our organisation set up a GEP recently but regardless of the HORIZON Europe requirement.
- ☐ Not known
- ☐ No answer

3.6 Please indicate the requirements and reasons below the European level that motivated your organisation to set up a gender equality plan.

Filter: 3.1 = 1-3

Please select from the list. Several answers are possible:

- ☐ Legal requirement to have a GEP (national or regional laws)
- ☐ Getting (research) funding demands a gender equality plan
- ☐ National or regional gender equality initiative that demands a gender equality plan, e.g., Athena SWAN (UK, Ireland), the Senior Academic Leadership Initiative (Ireland), Equal opportunities for men and women at universities / gender studies (Switzerland) or the program for women professors (Germany)
- ☐ Voluntary initiatives like diversity or equality audits
- ☐ Willingness of the organisation to foster gender equality
- ☐ Bottom-up initiatives in the organisation to foster gender equality
- ☐ Other, please specify - free text field
- ☐ Not known
- ☐ No answer

3.7 Please specify the legal requirement or the national/regional gender equality initiative.

Filter: 4.6 = 1 and/or 2

Free text field

3.8 What is the most crucial reason for your organisation not having a gender equality plan?

Filter: 0 = 4

Please select from the list. Several answers are possible:

- ☐ Time-consuming



- ☐ No personal resources
- ☐ No acceptance
- ☐ No necessity
- ☐ Other, please specify - free text field

4. Characteristics

In this section, we ask you about the characteristics of the gender equality plan, such as its duration, the involved stakeholders, and the key areas addressed.

4.1 Which bodies of your organisation approved the Gender Equality Plan?

Filter: 3.1 = 1-3

Please select from the list. Several answers are possible:

- ☐ Management board, rectorate, executive committee, directorate
- ☐ Decision-making body of internal members: e.g. Senate
- ☐ Decision-making body of (mostly) external members: e.g. university council, supervisory board
- ☐ Gender equality committee (or equivalent committee)
- ☐ Gender equality officer (or equivalent)
- ☐ Scientific council
- ☐ Members assembly, staff assembly
- ☐ Other, please specify - free text field

4.2 In which way is the Gender Equality Plan published and accessible?

Filter: 3.1 = 1-3

Please select from the list.

- ☐ The document is accessible to people outside the organisation
- ☐ Only internal document and accessible to all members of the organisation
- ☐ Only internal document and accessible only to the management

4.3 Please state the URL via which the Gender Equality Plan is publicly accessible.

Filter: 4.2 = 1

Free text field



4.4 When was the first gender equality plan adopted by your organisation (year)?

Explanation: Please consider the first plan or concept that addresses gender equality (or an equivalent concept such as the advancement of women), regardless of whether this plan meets the current requirements for a gender equality plan or is exactly named so.

Filter: 3.1 = 1-3

In the year: field in the format #####

- ☐ Not known
- ☐ No answer

4.5 What period does the current plan cover?

Filter: 3.1 = 1-3

Year [field in the format #####] - to the year [field in the format #####] The current plan does not have a specific time frame.

- ☐ The current plan does not have a specific time frame.
- ☐ Not known
- ☐ No answer

4.6 A gender equality plan provides for interventions in different areas. Please indicate the number of interventions (activities, policies, regulations etc.) in each area.

Filter: 3.1 = 1-3

[Matrix]

Scale:

- no measures designed
- 1 measure
- 2-5 measures
- 6-10 measures
- more than 10 measures
- not applicable

List:

- Awareness-raising and training
(e.g. workshops and training on gender bias, training for recruitment committees, booklets, films or posters)



- Work-life balance and organisational culture (e.g. child-care facilities for staff and students, dual career policy, network of fathers on the campus, respectful interaction, welcoming culture)
- Gender balance in leadership and decision-making (e.g. quota for decision-making bodies, gender-integrated leadership program)
- Gender equality in recruitment and career progression (e.g. active recruitment, gender equality in appointment procedures, coaching and mentoring programs for women researchers)
- Measures against gender-based violence, including sexual harassment, (e.g. complaints office, guidelines on sexual harassment)
- Integration of the gender dimension into research and teaching content (e.g. counselling for research funding, gender lectureship)

4.7 Which target groups does the Gender Equality Plan address?

Filter: 0 = 1-3

Please select from the list. Several answers are possible:

- ☐ Students
- ☐ PhDs
- ☐ Academic or scientific staff (or employees)
- ☐ Administrative and technical staff (or employees)
- ☐ Leading positions
- ☐ Applicants
- ☐ Reviewers
- ☐ Other, please specify

4.8 Which terms and perspectives about diversity and gender are used in the gender equality plan?

Filter: 3.1 = 1-3

- (1) It mentions differences within gender groups (e.g. Black women or persons with disabilities).
- (2) It provides measures to enhance equal opportunities for disadvantaged groups (e.g. Black women, persons from disadvantaged socioeconomic backgrounds or international students).
- (3) It includes the concept of “intersectionality”.
- (4) It includes the concept of “diversity”.
- (5) It mentions gender diversity (e.g. non-binary, diverse, trans, etc. persons).
- (6) It uses gender-neutral or gender-sensitive language, e.g. with signs to draw attention to gender-sensitive language (like “*”, “_”, “.” or “/”).

Coding for every item: Yes – No – Not known – No answer

4.9 Which inequalities other than gender does the Gender Equality Plan mention?



Filter: 8 = yes on one of 1-4

Please select from the list. Several answers are possible:

- ☐ Race, ethnicity, ethnic minorities, BAME [Black, Asian, Minority Ethnic], BIPOC [Black, Indigenous People of Colour]
- ☐ Nationality
- ☐ Religion
- ☐ Class, socioeconomic status, first-generation students or academics
- ☐ Age
- ☐ Sexual orientation
- ☐ Gender identity
- ☐ Disability/chronic health/mental impairment
- ☐ Care responsibilities
- ☐ Additional inequalities not listed
- [+ free text field]

5. Implementation

The following questions concern the implementation of the Gender Equality Plan, focusing on its realisation, personnel and financial resources and monitoring.

5.1 Are staff, committees or management positions in your organisation specially dedicated to implementing gender equality (or equal opportunities, equality or diversity)?

According to national or organisational requirements or traditions, these positions might have different names, such as “officer for the advancement of women” or “equity unit”. If these positions deal with gender equality in a broad sense, please include them in your answer.

Please select from the list. Several answers are possible:

- ☐ Gender equality [equality and diversity, equal opportunities...] unit (staff dedicated to gender equality)
- ☐ Gender equality officer / equal opportunity officer (member of the organisation charged with supervising gender equality in the organisation)
- ☐ Gender equality [equality and diversity, equal opportunities...] committee
- ☐ Gender equality officers (or equivalent), staff or committees in the departments
- ☐ Member of the top management (president, vice-chancellor, CEO, head of administration, etc.) charged with gender equality [equality and diversity, equal opportunities...] responsibilities
- ☐ Member of the department management or administration charged with gender equality [equality and diversity, equal opportunities...] responsibilities
- ☐ Other, please specify



5.2 Does your organisation dedicate any internal financial resources to implement gender equality measures?

Explanation: Financial resources include resources for staff charged with gender equality. Third-party funding is excluded.

- ☐ Yes
- ☐ No
- ☐ Not known
- ☐ No answer

5.3 Please assess the relation of the dedicated budget and the planned and implemented gender equality measures.

Filter: 5.2= yes

- ☐ Sufficient
- ☐ Almost adequate
- ☐ Hardly adequate
- ☐ Minimal

5.4 Which stakeholders are involved in the implementation of the Gender Equality Plan?

Filter: 3.1 = 1-3

Please select from the list. Several answers are possible:

- ☐ Management board, rectorate, executive committee, directorate
- ☐ Decision-making body of internal members: Senate
- ☐ Decision-making body of external members: university council, advisory board
- ☐ Members assembly, staff assembly
- ☐ All employees
- ☐ Administration, especially heads of the administration
- ☐ Departments, especially heads of the departments
- ☐ Gender equality officer (or equivalent)
- ☐ Gender equality unit
- ☐ Gender equality committee
- ☐ National or regional ministry
- ☐ Consultant, trainer, coaches
- ☐ Scientific council
- ☐ Other, please specify -free text field



5.5 Does your organisation provide training and capacity-building measures for gender equality (or diversity/equity)?

- ☐ Yes
- ☐ No
- ☐ Not known
- ☐ No answer

5.6 Does your organisation monitor gender equality and/or the implementation of the Gender Equality Plan?

- ☐ Yes
- ☐ No
- ☐ Not known
- ☐ No answer

5.7 Does your organisation regularly collect gender- or sex-disaggregated data for the following groups?

Please select from the list. Several answers are possible:

- ☐ Students
- ☐ Exams and graduates (BA, MA and equivalent)
- ☐ PhDs
- ☐ Academic or scientific staff
- ☐ Leading positions
- ☐ Administrative and technical staff
- ☐ Boards, councils and panels
- ☐ Success in acquiring funding
- ☐ Scientific publications
- ☐ Other, please specify -free text field
- ☐ No regular collection of gender- or sex-disaggregated data

5.8 Does your organisation regularly publish gender- or sex-disaggregated data for the following groups?

Please select from the list. Several answers are possible:

- ☐ Students
- ☐ Exams and graduates (BA, MA and equivalent)
- ☐ PhDs



- ☐ Academic or scientific staff
- ☐ Leading positions
- ☐ Administrative and technical staff
- ☐ Boards, councils and panels
- ☐ Principal investigators who successful in acquired funding
- ☐ Authors of scientific publications
- ☐ Other, please specify -free text field
- ☐ No regular publication of gender- or sex-disaggregated data

5.9 Does your organisation collect gender- or sex-disaggregated data about the inequalities listed below?

Please select from the list. Several answers are possible:

- ☐ Race, ethnicity, ethnic minorities, BAME (Black, Asian, Minority Ethnic), BIPOC (Black, Indigenous People of Colour)
- ☐ Nationality
- ☐ Religion
- ☐ Class, socioeconomic background status, first-generation students or academics
- ☐ Age
- ☐ Sexual orientation
- ☐ Gender identity
- ☐ Disability/chronic health/mental impairment
- ☐ Care responsibilities
- ☐ Additional inequalities not listed
[+ free text field]
- ☐ No data collection of gender and other inequalities

5.10 When collecting data on gender, which items are generally used in your organisation?

Filter: 5.6 ≠ no

Please select from the list. Several answers are possible:

- ☐ Women, female
- ☐ Men, male
- ☐ Legal term for other genders (f.ex. "divers" in Germany)
- ☐ Non-binary
- ☐ Trans* /TIN
- ☐ Opportunity for self-identification
- ☐ The variable "No answer"



- ☐ Other, please specify -free text field

5.11 Please estimate the overall implementation status of the measures planned in the gender equality plan.

Explanation: Estimate the status quo regardless of the duration of the plan. We will take the duration into account when analysing the survey data.

Filter: 0 = 1-3

Please note that the sum should add up to 100%.

Fully implemented: ## % of the measures

Have started: ## % of the measures

Not started yet: ## % of the measures

5.12 Which documents/regulations/procedures of your organisation include gender equality issues?

Please indicate from the list which documents (regulations/procedures) are available in your organisation and which contain gender equality issues. Several answers are possible:

Document, regulation or procedure	Exist in the organisation	Includes gender equality issues
Mission statement		
Strategy, university development plan		
University charter		
Quality assurance		
Recruitment or appointment regulations or policies		
Regulations or policies on career development, human resources development		
Internationalisation strategy or guest researcher program		



Regulations for internal research awards or grants		
Regulations for internal research support		
Regulation for internal teaching support		
Guidelines for reviewers and/or applicants		

6. Impact

The last part focuses on the impact of the Gender Equality Plan and the methods used to evaluate its effects.

6.1 Does your organisation assess the impact (the long-term effects) of the Gender Equality Plan and/or gender equality measures?

Filter: 3.1 = 1-3

- ☐ Yes
- ☐ No
- ☐ Not known
- ☐ No answer

6.2 Please name keywords on how your organisation assesses the impact of the Gender Equality Plan.

Explanation: Impact assessment issues can be the awareness of gender bias among staff and decision-making bodies or the gender distribution among students and staff.

Filter: 6.1 = yes

Free text field

6.3 Which approaches does your organisation use to measure the impact?

Filter: 6.1 = yes

Please select from the list. Several answers are possible:



- ☐ Self-evaluation
- ☐ External evaluation
- ☐ External audit
- ☐ Other approaches, please specify -free text field

6.4 Which data and methods does your organisation use to measure the impact?

Filter: 6.1 = yes

Please select from the list. Several answers are possible:

- ☐ Survey on working conditions and/or organisation climate among students
- ☐ Survey on working conditions and/or organisation climate among staff
- ☐ Qualitative interviews or focus groups with students
- ☐ Qualitative interviews or focus groups with academic and/or administrative staff
- ☐ Data analysis of the administrative data about gender distribution among departments
- ☐ Analysis of monitoring data (e.g. mentoring, workshops or other equality measures, complaints about sexual harassment)
- ☐ Other, please specify -free text field

6.5 Please indicate the current percentage of women in your organisation's highest research position (or position with research tasks) and that five years ago.

Explanation: Definition of the highest position: Leading Researcher, researchers leading their research area or field (R4 according to the European Framework for Research Careers); the single highest grade/post at which research is normally conducted (Grade A according to the She Figures), e.g. in universities the highest grades of professors

2022 ## %

2017 ## %

Please describe the position for which you provide the data (salary group, title or other characteristics)

Free text field

6.6 Please rate the extent of changes toward gender equality in the following areas in your organisation since 2017.

Score: -5 +5, with -5: significant negative changes, 0=no changes. +5=significant positive changes



Explanation: Please rate changes toward gender equality in terms of more (or less) awareness of gender bias, better (or worse) possibilities to combine private life, care responsibilities and work, organisational culture that takes diversity into account or a higher gender balance in decision-making bodies.

- Awareness of gender equality and gender bias in academia and science
- Work-life balance and organisational culture
- Gender balance in leadership and decision-making
- Gender equality in recruitment and career progression
- Gender-based violence, including sexual harassment,
- Integration of the gender dimension into research and teaching content

6.7 Please rate how far your organisation has stabilised activities for gender equality in the following areas.

Scale: 0 = starting point: nothing in place, 1 = minimal activities, 2 = inception (first isolated activities), 3 = growth (growing number of activities), 4 = integration (activities integrated into a coordinated concept), 5 = institutionalisation: gender equality policies fully integrated into the organisation

- Actions of awareness-raising and training
(e.g. workshops and training on gender bias, training for recruitment committees, booklets, films or posters)
- Work-life balance and organisational culture
(e.g. child-care facilities for staff and students, dual career policy, network of fathers on the campus)
- Gender balance in leadership and decision-making
(e.g. quota for decision-making bodies, gender-integrated leadership program)
- Gender equality in recruitment and career progression
(e.g. active recruitment, gender equality in appointment procedures, coaching and mentoring programs for women researchers)
- Measures against gender-based violence, including sexual harassment,
(e.g. complaints office, guidelines on sexual harassment)
- Integration of the gender dimension into research and teaching content
(e.g. counselling for research funding, gender lectureship)

6.8 Please rate the relevance of the Gender Equality Plan for the achieved positive changes.

Score: 0-5, 0 = no influence, 5 = high influence

Filter: 0 = 1-3

6.9 You stated that the Gender Equality Plan has low influence on the achieved changes in gender equality. Which policies or instruments had a bigger influence than the Gender Equality Plan?



Filter: 6.8 = 0-2

Free text field

7. Further Questions

Finally, we want to know how you handled the survey.

7.1 Which individuals were involved in the processing of this survey?

Please select from the list. Several answers are possible:

- ☐ Member of the top management (e.g. president, vice-chancellor, rectorate, executive committee)
- ☐ Staff linked to the leadership
- ☐ Gender equality officer (or equivalent)
- ☐ Gender equality unit
- ☐ Administration staff (e.g. human resources department)
- ☐ Other, please specify - *free text field*

7.2 Would you be interested in joining a Community of Practice (CoP) for Research Funding Organisations (RFOs) or for companies (both aim to exchange knowledge on gender and innovation)?

Information about the INSPIRE's Communities of Practice can be found here:
<https://www.inspirequality.eu/support>

- ☐ yes
- ☐ No interest

7.3 Please contact with the following e-mail address to be part of Community of Practice : innovationKSH@inspirequality.eu

Filter: 7.2 = yes

7.4 Do you have any further remarks on gender equality or gender equality plans in your organisation?

Free text field

Final Page 1 => Informed Consent= No

Thank you for taking the time to consider participating in our survey. We sincerely respect your decision to decline after reviewing the informed consent. Your comfort and informed choice are essential to us. If you have any questions or would like to address this topic, please do not hesitate to contact us.

inspire@gesis.org



Final Page 2 => Prevalance= “Not known” or “No answer”

We appreciate your interest in our survey. Apparently, you don't know if the gender equality plan exists in your organisation. If you clicked the answers “not known”/”no answer” by mistake, please contact us directly. Your feedback is valuable, and we like to assist with any clarifications or issues you may have.

inspire@gesis.org

Final Page 3 => The real final page

Thank you for taking the time to complete our survey on gender equality plans. Your contributions are highly beneficial as we strive to monitor gender equality plans and their impact. We appreciate your willingness to share your experiences, which are critical to our commitment to creating a more inclusive and equitable environment. If you have any questions or want to discuss your responses in more detail, please contact us at

inspire@gesis.org.

For more information about Inspire project please click the link :

<https://www.inspirequality.eu/>

10.6 Survey: Invitation and Reminder to the Online Survey

10.6.1 Invitation

Subject: Invitation to INSPIRE's Survey to Monitor Gender Equality Plans

Dear Madam or Sir,

To whom it may concern,

We kindly invite you to participate in a survey to monitor gender equality plans in academic and research-performing and funding organisations. The survey is available in the following languages: **English, Français, Deutsch, Polski** and **Español**.

GESIS – Leibniz-Institute for the Social Sciences undertakes this survey in the context of the INSPIRE project (<https://inspirequality.eu/>). INSPIRE is a project funded by the European Union to build Europe's Centre of Excellence on inclusive gender equality in research and innovation. As part of this project, GESIS investigates how to monitor gender equality plans of research performing and funding organisations in the European landscape via an online survey. The survey asks about the existence of a gender equality plan or an equivalent, its characteristics, its implementation, and an assessment of its impact. We found your data (name, function and mail address) via a web search by INNO Systems, one of the INSPIRE Consortium members.

Participating in the survey, which will take 15-20 minutes, will help get valuable and essential information about implementing gender equality plans in the European Research Area.



To participate, please click on the following link. **The survey is available until 01.08.2024.**

#code_complete#

If you don't have any information about gender equality activities in your organisation, please transmit the survey to someone who knows.

Best regards

Lena Weber, Andrea Löther and Mazlum Karataş

10.6.2 The First Reminder

Subject: Extended deadline and Friendly Reminder: INSPIRE's Survey on Monitoring Gender Equality Plans

Dear Madam or Sir,

To whom it may concern,

The deadline for the INSPIRE survey has been extended. **The survey is available until 19.08.2024.**

This is a friendly reminder for our survey about gender equality plans. Please help us receive more stable evidence about gender equality plans in European research-performing and funding organisations while answering our survey. The survey is available in the following languages: **English, Français, Deutsch, Polski and Español.**

GESIS – Leibniz-Institute for the Social Sciences undertakes this survey in the context of the EU-funded INSPIRE project (<https://inspirequality.eu/>). INSPIRE is a project to build Europe's Centre of Excellence on inclusive gender equality in research and innovation. As part of this project, GESIS is researching how to monitor gender equality plans of research performing and research funding organisations in the European landscape via an online survey. The survey asks about the existence of a gender equality plan or an equivalent, its characteristics, its implementation, and an assessment of its impact. We found your data (name, function and mail address) via a web search by INNO Systems, one of the INSPIRE Consortium members.

Participation in the survey will only take 15-20 minutes of your time.

To participate, please click on the following link.

#code_complete#

If you don't have any information about gender equality activities in your organisation, please transmit the survey to someone who knows.



Best regards

Lena Weber, Andrea Löther and Mazlum Karataş

10.6.3 The Second Reminder

Subject: Friendly Reminder: Only Seven Days Left to Share Your Insights on Monitoring Gender Equality Plans

Dear Madam or Sir,

To whom it may concern,

As a friendly reminder, you have seven days left to participate in our survey about gender equality plans. Please help us receive more stable evidence about gender equality plans in European research-performing and funding organisations while answering our survey. The survey is available in the following languages: **English, Français, Deutsch, Polski** and **Español**.

Participation in the survey will only take 15-20 minutes of your time.

To participate, please click on the following link. The survey is available until **19.08.2024**.

#code_complete#

If you don't have any information about gender equality activities in your organisation, please transmit the survey to someone who knows.

Best regards

Lena Weber, Andrea Löther and Mazlum Karataş



10.7 Country Clusters

Central East and Eastern	Bosnia and Herzegovina
	Bulgaria
	Croatia
	Czechia
	Estonia
	Hungary
	Latvia
	Lithuania
	Poland
	Romania
	Servia
	Slovakia
	Slovenia
Central West	Austria
	Belgium
	France
	Germany
	Luxembourg
	Netherlands
	Switzerland
Northern/West	Denmark
	Finland
	Ireland
	Norway
	Sweden
	United Kingdom
Southern	Cyprus
	Greece
	Italy
	Malta
	Portugal
	Spain
Not belonging to a cluster	Israel



10.8 Technical Implementation of INSPIRE Scraper

This package provides scripts to research organisations if they have a gender equality plan (GEP) or not (Prevalence) using the serpapi.com API.

webcrawler/crawler-prevalence.php : Check whether GEPs are in place.

webcrawler/crawler.php : Extract the more relative PDF GEPs.

Requirements

- Apache or Nginx web server with shell access
- PHP 8.1 or later
- Composer (latest version)
- Required PHP extensions:
 - ext-json
 - ext-mbstring
 - ext-zip
 - ext-gd
 - ext-iconv

Installation

1. Clone the Repository

```
git clone https://github.com/InspireQualityeu/gep-scraper.git
```

```
cd webcrawler
```

2. Install Composer Dependencies

Ensure Composer is installed. If not, download and install it from [Composer's official website](https://getcomposer.org/).

Then, run:



composer install

3. Configure API Key and Variables

Before running the scripts, update the required configuration variables within:

- API Key: Obtain your API key from [SerpAPI](#) and replace the placeholder.
- Countries: Modify the list of countries as needed.
- Search Terms: Update the terms according to your requirements.

4. Run the PHP Scripts

php crawler-prevalence.php

php crawler.php

Troubleshooting

- If composer install fails, ensure PHP and Composer are installed correctly.
- If missing extensions are reported, install them using your system's package manager. For example:
 - Ubuntu/Debian: `sudo apt install php-mbstring php-zip php-gd php-json php-iconv`
 - CentOS/RHEL: `sudo yum install php-mbstring php-zip php-gd php-json php-iconv`
 - Windows: Enable the extensions in php.ini and restart your web server.