

# Functionality and Design of EIGE's Database on Gender Statistics

Online Discussion Report



ISBN 978-92-9218-551-0

doi: 10.2839/15378

© European Institute for Gender Equality, 2014

Reproduction is authorised provided the source is acknowledged.

This publication summarises the views expressed by 21 people participating in an online discussion held on 15 July 2014 through the EuroGender online platform of the European Institute for Gender Equality (EIGE). The findings, interpretations, and conclusions are entirely those of the authors and should not be attributed in any manner to EIGE. Neither EIGE nor any person acting on its behalf may be held responsible for the content of the information contained in this publication.

# Contents

Introduction	4
Organisation	4
Participation	6
Background	6
Summary	7
Conclusions and follow-up	13
Annex I: Resources	14
Annex II: Transcript	14
Annex III: List of participants	15



# Introduction

The online discussion on “Functionality and design of EIGE’s database on gender statistics” took place on the 15<sup>th</sup> of July 2014. It was organised to support the EIGE staff in the development of a new database, which will be a centralized resource on gender statistics giving access to a variety of quantitative and qualitative data resources in the European Union. The new database identifies, produces and disseminates statistics that reflect the realities of the lives of women and men and policy issues relating to gender equality.

The main purpose of the online discussion was to receive feedback from potential users of the database and other stakeholders concerning their needs, in particular with respect to the planned content, structure and metadata of the new research tool.

Invitations were sent to a restricted group of selected stakeholders coming from a diverse professional background and working with different institutions such as national statistics offices and/or national machineries for gender equality.

This report of the online discussion aims to enable dissemination of the insights and learning from the debate. It will serve as a resource to EIGE in the development of the future database, which is expected to be operational by the end of 2015.

It also reflects the results of the 1<sup>st</sup> public consultation launched by EIGE about its project on the development of the database.

This report summarises the key points made and themes explored during the debates. Inevitably, it is impossible to reflect in full the variety of ideas and perspectives brought forward.

# Organisation

The online discussion was organised in three sessions. Each session explored a different area of work that could be progressed by EIGE and was facilitated around a set of general questions:

## Session 1: Content and sources of information for the database

This session explored the following questions:

- *Which other sources (than presented during the debate) would you suggest to be considered by EIGE in addition to the suggested ones?*
- *Eventually, the database will have to integrate national and regional data (e.g. data on violence against women, which is very scarce at EU level). Which relevant data sources exist at national/regional level that could be considered for EIGE's database?*
- *Should the database contain only comparable and harmonised data at the EU level or should it also consider non-harmonised data by documenting it in the metadata reference?*
- *How to overcome the problem of making available harmonised data at the EU level with other data, which is not harmonised and comparable at the same level and does not follow the same quality standards for data collection and dissemination?*



## Session 2: Structure of the data and search/extraction options

This session explored the following questions:

- *Of particular interest to EIGE, at this stage, are users' opinions about the proposed browsing tree structure and users' ideas for keywords that should be enabled for search queries. Do you have any proposals in this respect?*
- *The Gender Equality Index structure provides a quite intuitive way for users to access gender equality data. Should it include only variables used in the Index or should it also consider other related variables to the particular domain?*
- *What categories should be included under 'Intersecting inequalities'? Current sub-domains include intersecting inequalities with sub-domains age, citizenship, disability, ethnicity, marital status, religion and sexual orientation. How to deal with age categories under this level? In the future project of the interface maybe it should be foreseen a search tool which will allow looking more into specific categories such as age.*
- *Should 'Indirect violence' be kept under the GBV structure providing access to negative stereotypes, norms and attitudes or should it be deleted?*
- *Does the proposed structure provide an intuitive logic to users or should it be improved? Which reference structure should be applied for its improvement?*

## Session 3: Gender statistics data and metadata

This session explored the following questions:

- *Taking into account the possibility for further breakdowns at the micro data level, which dimensions could be considered for disaggregation? For example, how important is it to present data disaggregated by household type? Which dimensions would be relevant on employment status?*
- *Taken that statistics metadata provide descriptions of data characteristics and support users in the interpretation and analysis of data, which type of information will be most required in the metadata description according to you?*



# Participation

Twenty-one people actively participated in the online discussion and a larger number of 51 people attended. Over one hundred contributions to the various debates were given. The participants were drawn from ten different Member States (United Kingdom, Spain, Ireland, Lithuania, Belgium, Latvia, Romania, Netherlands, Italy and Portugal) and were either individual researchers or experts or representatives of NGOs, local authorities, women's organisations or foundations.

The debate was conducted in English.

# Background

According to EIGE's mid-term programme 2013-2015, the Institute is expected to establish a reliable source of information and data on gender equality in the European Union. This includes the objectives of coordinating, centralising and disseminating research data and statistics on gender equality in Europe.

Under this goal, EIGE has been working with several sources of quantitative and qualitative data within different projects.

Aiming to consolidate, organise and integrate the results of such work, the Institute has been establishing a centralised source of information on gender statistics for EIGE's stakeholders and the general public – EIGE's database on gender statistics.

As a centralised resource on gender statistics, the database aims to identify, produce and disseminate statistics that reflect the realities of the lives of women and men and policy issues relating to gender equality.

Aiming to adequately reflect differences and inequalities in the situation of women and men in all areas of life, gender statistics cut across the traditional fields of statistics. They are defined by the sum of the following characteristics:

- data collected and presented disaggregated by sex as a primary and overall classification;
- data reflecting gender equality issues;
- data based on concepts and definitions that adequately reflect the diversity of women and men and capture all aspects of their lives;
- data collection methods taking into account stereotypes and social and cultural factors that may induce gender biases.

EIGE's database on gender statistics should become a centralised resource for users to obtain statistics, data and metadata according to different search criteria within the perspective of EU policy areas on gender equality and identified keywords.



# Summary

## Session One: Content and sources of information for the database

The database should be a harmonised and centralised source of reliable and comparable data. The following international sources are being considered by EIGE at present:

- Eurostat;
- Eurofound;
- European Union Agency for Fundamental Rights (FRA);
- DG Justice (European Commission's database on Women and Men in Decision Making);
- Other relevant agencies and DGs of the European Commission;
- European Migration Network;
- International Labour Organisation (ILO);
- Relevant agencies of the United Nations system (e.g. UNECE Gender Statistics Database, UNODC, UNSD Statistics Databases);
- World Bank;
- Organisation for Economic Co-operation and Development (OECD);
- Organisation for Security and Co-operation in Europe (OSCE);
- Council of Europe (CoE);
- Other international and civil society organisations;
- Any other research and intergovernmental or supranational organisations working on statistics relevant from a gender perspective within the European Union Member States.

Because of the different ways the data will need to be processed, three types of data will be distinguished:

- data provided as macrodata by the immediate source and used in the database unchanged (such as the employment rates);
- data computed by EIGE from other data provided as macrodata by the immediate source (such as the gender pay gap on monthly wages, computed from levels of monthly wages that are obtained as country-level aggregates from Eurostat);
- data computed by EIGE from microdata (such as average flexibility of working hours for men and women in each country, computed from the micro-level dataset of the EWCS, provided by Eurofound).



### Contributions and key issues

During the online discussion some clarification questions had been answered concerning the contents and sources of information of the new database. Project manager Maris Goldmanis explained that the database is primarily based on reliable and comparable sex-disaggregated data at EU-level from 2005 onwards, covering the 28 EU Member States and countries under the Instrument for Pre-Accession Assistance - IPA (where data is available).

Datasets are broken down by employment status. For some data (especially in the area of gender-based violence), national sources are also considered, but in general data from such sources have the problem of not being directly comparable across countries due to different methodologies (lack of harmonization).

There was an overall agreement to consider regional and national data whenever available, despite the challenge of comparability and harmonization of data (including differences in indicators and language issues). In reaction to a comment of Mrs. Goldmanis, Jeroen Decuyper stated that 'If EIGE wants to become the "one stop shop" for gender equality data, it could consider presenting data from national sources, so that people looking for data on a specific country (without wanting to compare it) can find it in EIGE's database. Of course there will then be a need to indicate that it is based on national data and that the calculating method might not be the same for all countries'.

The moderator, Dorothee Fischer, also pointed out that there is sometimes a confusion of terminology in gender statistics. The term "gender" has often been wrongly used in association with data. "Gender disaggregation" or "data disaggregated by gender" are incorrect terms. Gender statistics are disaggregated by sex, an individual-level characteristic commonly recorded in censuses, surveys and administrative sources, and not by gender, a social concept relevant at the level of a population group. Ligia Nobrega added that in the development of the database the focus is laid on certain areas of concern where women and men may not enjoy the same opportunities or status (such as work status in the labour market or higher education, for example) or where women's and men's lives may be affected in different ways (such as maternal mortality, domestic violence, or occupational injuries, for example). Specific population groups where gender inequality is likely to be present or more pronounced are also of relevance, as long as statistical information is available.

### Interesting content and sources of information for the new database that were suggested during the discussion:

- statistical divisions of government agencies;
- the World Health Organization;
- data from research related to gender equality issues sponsored by the framework programmes (FP7/Horizon 2020);
- in the Netherlands, the Sexual Health Survey of Rutgers WPF is a relevant database for Fertility and Reproductive Rights, including a separate survey for the 15-25 age group. Not harmonized but very sophisticated; the text is in Dutch;
- regional statistics, for example the Basque institute for Statistics which has a specific thematic area relating to gender equality on its website (<http://en.eustat.es/>);
- Interpol and Europol might be appropriate partners for cross-border perpetrators and co-ordination with cyber security realms;
- UNICEF studies providing some databases on gender and education;
- data on social and economic variables, otherwise it is difficult to contextualize the results;
- the Rape Crisis Network Ireland database.

As regards the duration of the data, EIGE will update the content of the database on an ongoing basis, whenever new sources of information are available.

Jacinta Arcadia von Prussia also suggested including qualitative data in the database, e.g. to open access to stories from victims of gender violence. However, qualitative information will be taken into consideration only if relevant for the metadata purposes. At first EIGE will consider data available from official statistics, which follow the quality criteria of the European Statistical System. Other available data will also be included, but it will be documented and reported in a transparent way.





### Interesting examples/references of other databases:

- the World Economic Forum Gender Gaps report:  
<http://www.weforum.org/issues/global-gender-gap>
- survey conducted in Romania:  
<http://www.fundatia.ro/en/node/669>
- in the Netherlands, the Statistical Bureau offers the possibility for 'do-it-yourself' statistics (this could be a reference for the future interface):  
<http://www.cbs.nl/nl-NL/menu/cijfers/statline/zelf-tabellen-maken/default.htm>
- "data for the evaluation of the European semester process from a gender equality perspective" (2012) DG for Internal Policies, European Parliament:  
[http://www.europarl.europa.eu/RegData/etudes/etudes/join/2012/462432/IPOL-FEMM\\_ET%282012%29462432\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/etudes/join/2012/462432/IPOL-FEMM_ET%282012%29462432_EN.pdf)
- the dedicated section on gender statistics of the UNECE database:  
<http://w3.unece.org/pxweb/?lang=1>
- the Eurostat database does not have a dedicated gender section, but its functionality is closest to what EIGE has in mind: an expandable browsing tree; free form search; highly customizable pivot tables; a variety of export options: [http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search\\_database](http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database)
- the database on Women and Men in Decision-Making:  
[http://ec.europa.eu/justice/gender-equality/gender-decision-making/database/index\\_en.htm](http://ec.europa.eu/justice/gender-equality/gender-decision-making/database/index_en.htm)
- and EIGE's own Women and Men in the EU database will be integrated in the EIGE database:  
<http://eige.europa.eu/content/women-and-men-in-the-eu-facts-and-figures>
- the EIGE database will have a macro-level (i.e. country), aggregates from a variety of micro-levels (i.e., individual or household) sources, such as the ESS – European Social Survey:  
<http://nesstar.ess.nsd.uib.no/webview/>
- the GGP - Generations and Gender Programme - data user space:  
<http://www.ggp-i.org/data/data-access.html>
- and the FRA - European Union Agency for Fundamental Rights - study on gender based violence:  
<http://fra.europa.eu/en/publication/2014/vaw-survey-main-results>.

### Session Two: Structure of the data and search/extraction options

There will be two options to access the data: browsing through the **pre-determined tree structure** or by **keyword search**.

Users will be able to browse the database by several pre-determined tree structures. This type of entry will be suitable for the following types of users:

- those who do not have extensive experience in gender statistics and would therefore benefit from a structured guide for the available data;
- those who want to explore the domains of the Index or the areas of the BPfA (Beijing Platform for Action) in their entirety;
- those who are looking for a specific dataset and already know its location in one of the predefined structure.

Another option is being thought for users who have a particular dataset of area of interest in mind but do not know its place in any of the pre-defined structures. Those users might want to take advantage of the **keyword search option**. Such tool should address users' needs on gender statistics by allowing easy access and data extraction according to different criteria.

On the front page of the database, a list of titles of the available tree structures (or "entry points") will be presented. A short description of each tree structure will be presented below the respective title. Clicking on one of the titles will bring up an



expandable tree corresponding to that structure.

These are the five possible entry points to the index structure:

- Thematic areas;
- Policy areas;
- EU strategies;
- Gender Equality Index;
- Beijing Platform for Action;
- Plus an additional section on Gender Based Violence.

Moreover, two ways of accessing the database are foreseen:

- via the browsing tree, where users essentially choose from a predefined, structured menu;
- via keyword/tag search, where users can search for datasets by typing a series of search terms (keywords, tags) into a text box;
- if a particular term (such as “masculinities”) is defined only as a keyword/ tag, but not as an element in the browsing tree, users will not be able to select it as an option from the browsing tree, but will only be able to search for datasets tagged with that term via the search tool.

### Contributions and key issues

There was an overall approval and positive feedback on the possibility of including a keyword search in the database. Or as Jeroen Decuyper phrased it, “this will probably remain the main option for users with little experience in the field and for people who are pressed for time.”

Suggestions for additional key terms included:

- synonyms and related issues, for instance the following terms relating to sexual harassment: harassment, sexual, equality, violence, dignity, honour, European Court of Justice, national legislation, EU directives;
- medically assisted procreation (MAP).

A long discussion focused on the difference between the **thematic areas** and **policy areas**. Some participants were of the opinion that these two categories are somewhat redundant or would create confusion among the users of the database as to where the information is stored. The EIGE team clarified that the label ‘policy areas’ is rooted in the EU policy priorities within the framework of the EU Institutions. It is a particularly relevant entry point, which takes into account EIGE’s main stakeholders who are EU institutions. It is the case of the European Parliament and policy makers in general.

Thematic areas, on the other hand, are based on:

- possible areas of interest on gender equality from the user’s perspective;
- the structured frameworks provided by the data sources where the data are available.

Cristina Castellanos suggested linking the priority areas to the five priorities mentioned in the EU-2020 strategy and relating gender issues to these objectives.

Ana Rita Manca added that it is very relevant to link policy areas with data because this will clearly show statistical gaps: “In other words it will be visible how the policy areas are disproportionately developed in employment and work. I am not saying this is bad, but not at all when it comes to violence for instance. We hope this will contribute in pushing for collecting data in all relevant areas for gender equality.”

In any case, the different entry points often lead to the same type of information and there may even be “a certain degree of overlap” (Maris Goldmanis). This is intentional, so that it allows users with different perspectives/different initial approaches to access all of the data relevant to their topic of interest in one place instead of having to acquire these data



from multiple locations. Thus, for example, "Fertility" appears under "Population and demography", "Health and mortality" and "Reproductive rights", as it seems to be relevant to all three classes of users:

- those who want a general overview of the demographic situation;
- those who want a general overview of health-related data;
- those who are specifically interested in issues of reproductive rights and reproductive health.

**Additional themes for the three levels in the tree structure were suggested:**

- asylum and migration, security and crime, energy;
- add "status attainment" in the work place (by sex and by the same level of education);
- gender pay gap (per country);
- number of specialists in stereotypical feminine domains;
- number of specialists in the stereotypical masculine domains;
- distinctions among the number of hours worked in addition to full/part-time employment;
- parental leave and childcare: include a category on paid and non-paid parental leave, transferable and non-transferable, explicitly stating whether they are calendar or working days, the distribution of total paid and unpaid days by sex in each country;
- "fertility and reproductive rights", "Medically Assisted Procreation (MAP)" in the second level of the thematic area;
- legal status/minority status/disability, e.g. asylum seekers/refugees, people with disabilities in the area of sexual violence/abuse;
- a subcategory for time in informal care for (in law) parents (with an ageing Europe the relevance of this category for gender equality will increase);
- distinction between individual and household earnings and labour or other sources;
- marital property regimes.

In the foreseen structure of the database, current sub-domains include intersecting inequalities with the sub-domains age, citizenship, disability, ethnicity, marital status, religion and sexual orientation. The participants were asked feedback on how to deal with age categories under this level. Cristina Castellanos was wondering whether nationality is embedded in citizenship and she suggested, moreover, to add individual and household incomes, and number of children or childless as well as house ownership.

Also in the Netherlands, it is important with respect to employment status and individual income to compare women with children with women without children in a relevant age category (e.g. 20-50 years). And ethnicity is relevant, but often statistical and technical problems are in the way (Atria Institute on Gender Equality and Women's History).

The Rape Crisis Network in Ireland recalled that in Ireland they use "civil status" as opposed to "marital status", as gay people are not allowed to get married there yet. At the same time they were wondering about the term "family status".

The debate also touched on the question of complexity of the tree structure versus a simple structure where information can easily be found. Jeroen Decuyper suggested including "simple and concise instructions" for the users to avoid confusion. The Rape Crisis Network in Ireland explained in detail their data collection methodology, confirming that they had prepared a comprehensive user guide "which contains explanations and definitions of everything asked on the system".

One open issue is also the way of presenting the BPfA indicators and the Gender Equality Index scores, which will most likely be placed under their respective entry points. Maris Goldmanis underlined that EIGE is "willing to consider also displaying BPfA and Gender Equality Index indicators/scores under the other trees, whenever they can be clearly mapped to a particular section of those trees. The only reason we are currently thinking of not doing so is that this might create too much 'clutter' (too many different types of data presented together)".



### Session Three: Gender Statistics metadata and data

The final session dealt with metadata, which make it possible for users to understand, interpret, evaluate and analyse statistical information.

Two types of metadata can be distinguished: 1) **structural metadata**, which provide a structured description of the way the statistical data and the reference metadata are organised, and 2) **reference metadata**, which provide additional descriptive information on the concepts used, the data collection and generation methods employed, and the quality of the data.

All observations in the EIGE database will be at the country and year level. It follows that all datasets will contain the dimension "**Country**" and the dimension "**Year**". In addition, all datasets, except those that provide direct measures of the relative situation of men and women (such as values of the Gender Equality Index itself, gender pay gap measures, measures of gender gaps in other variables -prepared for the Index- and measures of horizontal segregation in occupations and education) will also provide gender-disaggregated data, i.e., will contain the dimension "Sex".

### Contributions and key issues

The discussion started with some further clarifications on what metadata are.

“Metadata is the information linked to the data, meaning the number we read in the table. It is useful because it tells us about the process of collecting and producing this number. How can one for example read the number 30? It could be Euros, hours or years. Metadata will help us understanding and completing the information on this number. This is not enough because metadata gives information on the periodicity of releasing statistical information, the reference period, the legal framework of the data collected, information on data quality, including comparability, accuracy, relevance and reliability of the data. This information leads us to make a clear distinction between those data produced by official statistics and those that are not. This has an implication from the user point of view and helps to assess the quality of data. Finally it will help data producers to use a shared methodology towards the harmonisation of statistics in the EU.”  
Ana Rita Manca

“Gender statistics require at least two statistical variables cross-tabulated: sex and the main characteristic that is studied, such as educational attainment or labour force participation. A meaningful gender analysis commonly requires disaggregation by sex and other characteristics at the same time.” Ligia Nobrega

Although this was the most abstract of the three discussion topics and there were fewer contributions altogether, it seems that there was an overall consensus among the participants agreeing with the EIGE proposal on metadata.

“Yes, I don't know what I can add to your description of two types of metadata, I just agree. It is important that I can check how for instance “self assessed work-life balance” is measured. With employment it is important to know what definition is used. And with the gender gap in employment rates it is important to include the gap in FTE's (and explain how Eurostat computes this variable).” The Atria Institute on Gender Equality and Women's History

“The description on how the data providers compute a certain variable and/or processes statistics information should be included in the metadata description. This will be the case of the example on FTE (full time equivalent).” Ligia Nobrega

“It would also be very useful to know the gap time between the data delivered and the referring time, and the periodicity of the information, and the local/ regional/national coverage, error sample and what kind of categories are representative.” Cristina Castellanos



## Conclusions and follow-up

All in all, there was a very positive reaction on the new EIGE database – some of the participants of the online debate explicitly stated what a useful tool it will be for their work in the future.

An important part of the work has already been developed and the debate helped EIGE to finetune some of the open issues and get direct feedback from users on the planned content and structure of the database.

Thus, the participants made suggestions for possible data sources at the regional, national and international level as well as for keyword search terms and categories for the tree structure.

Some issues will need to be clarified, in particular the planned tree structure (e.g. whether to keep the planned entry points “thematic priorities” and “policy priorities” or whether to fuse them). Some of these open questions will probably only be answered during a test phase of the new database. Cristina Castellanos suggested in this context “to open a table, asking people to give their feedback, especially in domains where data is less available or comparable.”

And as some participants stated, the database is thought to be a dynamic product. It will develop according to the users’ needs and will be progressively adjusted, even once it is online.

Another online debate is planned for this autumn concerning the interface and visual presentation of the database. Certainly this complementary feedback from the users will also be very helpful to shape the future EIGE database on gender equality.



# Annex I: Resources

**The background document made available on EuroGender platform during the discussion**

[http://eurogender.eige.europa.eu/sites/default/files/web-discussions-files/Proposed%20Tree%20Structures%20for%20Online%20Discussion\\_1.pdf](http://eurogender.eige.europa.eu/sites/default/files/web-discussions-files/Proposed%20Tree%20Structures%20for%20Online%20Discussion_1.pdf)

# Annex II: Transcript

**Link to the online discussion transcript**

<http://eurogender.eige.europa.eu/documents/transcripts-online-discussions-eiges-database-gender-statistics>



## Annex III: List of participants

Name	Institution	Country
Alexandrina Satnoianu	EIGE	Lithuania
Alina Žvinklienė	Chief researcher	Lithuania
Ana Sofia Fernandez	Resource & Documentation Centre (RDC), Officer	Lithuania
Anna Rita Manca	EIGE, Statistics Officer	Lithuania
Eelco Wierda	ATRIA – Institute on Gender Equality and Women's History	Netherlands
Christina Andersson	World Peace Foundation, Ambassador coordinator	Belgium
Cristina Castellanos Serrano	Tavistock Institute, Senior Researcher and Consultant	United Kingdom
Dorothee Fischer	Cultura Lavoro Srl	Italy
Emakunde – Basque Women's Institute	Emakunde – Basque Women's Institute	Spain
GLA – Greater London Authority	GLA – Greater London Authority	London
Helena Morais Maceira	EIGE	Lithuania
Jacinta Arcadia Von Prussia	Director, Individual Expert	Belgium
Jeroen Decuyper	Federal Institute for the equality of women and men, Attaché	Belgium
Kristaps Petermanis	EIGE, Administrative Agent	Latvia
Ligia Nobrega	EIGE, Gender Expert	Lithuania
Loredana Ivan	Associate professor	Romania
Maria José Carrilho	Councillor for Gender Equality - Alternate member	Portugal
Maris Goldmanis	Royall Holloway University of London	United Kingdom
Rape Crisis Network Ireland	Rape Crisis Network Ireland, Data and Services Information Manager	Ireland
Therese Murphy	EIGE, Head of Operations	Ireland



## **Contact information**

European Institute for Gender Equality

Gedimino, 16

LT-01103 Vilnius (Lithuania)

Tel +370 5 215 7444

Tel +370 5 215 7400

<http://eige.europa.eu>



