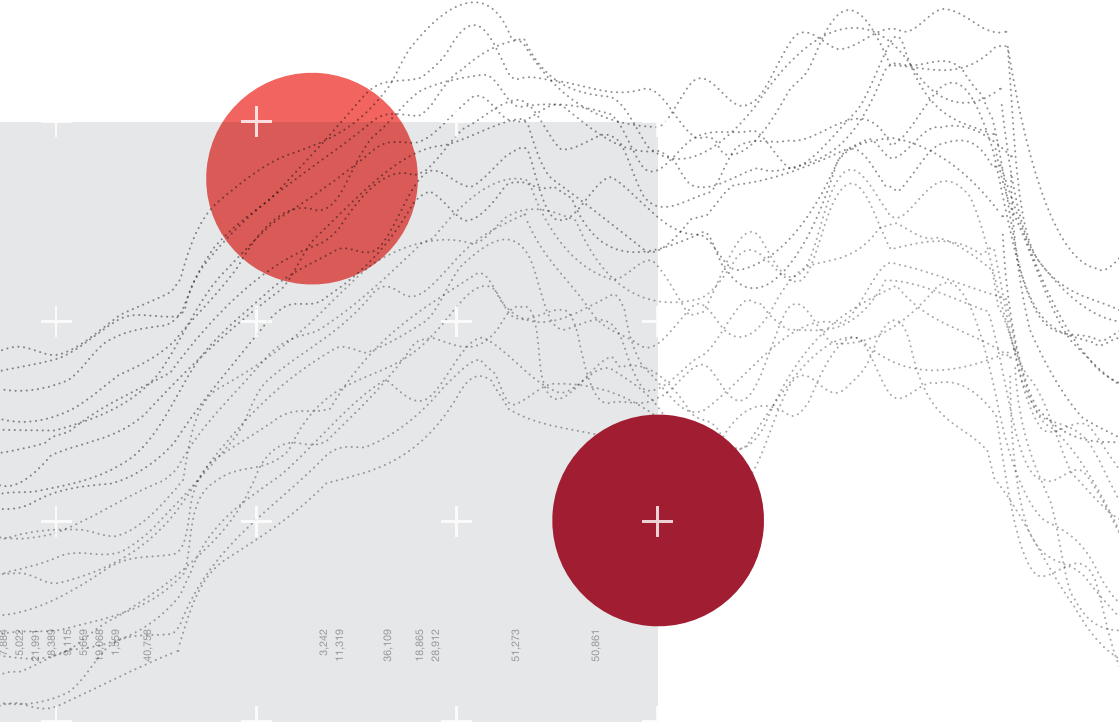
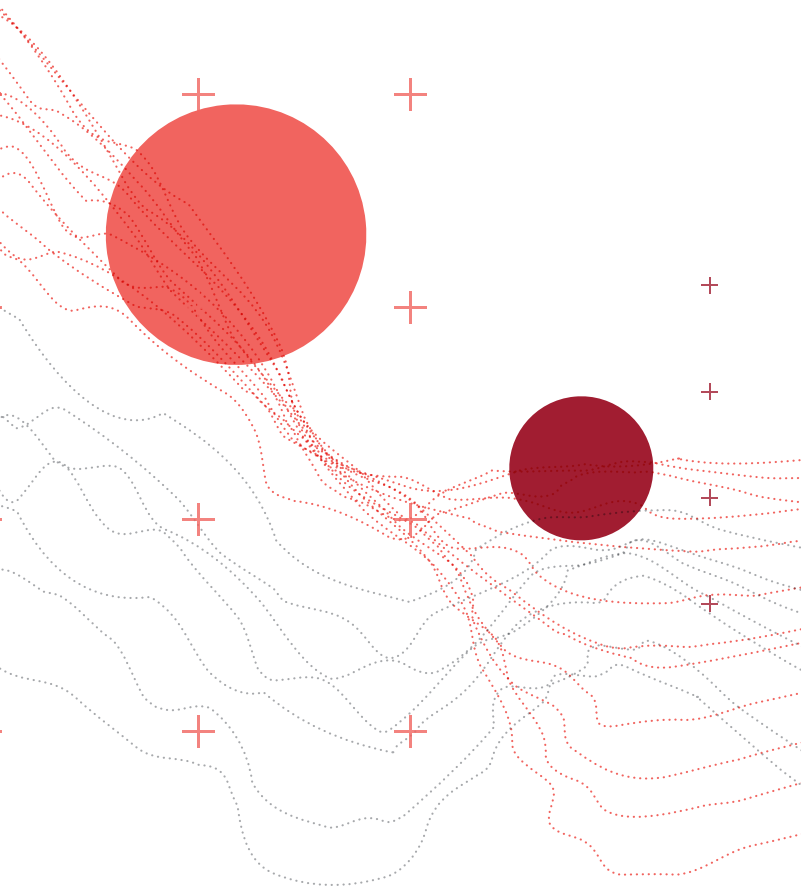




Guidelines for producing gender analysis from innovation and IP data





Why analyze gender in innovation and IP data?

Understanding how women and men can access and use the intellectual property (IP) system equally is key to ensuring that their creativity translates into economic, social and cultural development.

It can be difficult for international and national IP offices to find out the gender of inventors and creators from IP and other innovation data. This makes it harder to create the IP indicators and economic analysis that are needed to inform gender policies.

This short guide summarizes best practice for producing innovation and IP gender indicators. It is hoped that this will be adopted by governments and researchers willing to conduct studies and design policies that will contribute to achieving a more even gender balance.

Roadmap for producing innovation and IP gender indicators

A Collecting innovation and IP gender data

There are different ways to collect innovation and IP data broken down by gender. This guide looks at their advantages and limitations. The two main ways to get innovation and IP gender data are as follows:

- **Direct collection of gender data**, for example during the IP application process.
- **Indirect gender attribution** to the existing innovation or IP data. This can be done by using secondary sources, semantic rules or gender-name dictionaries.

B Using WIPO's gender-name dictionary

The World Intellectual Property Organization (WIPO) and other national IP offices attribute gender to their IP collections based on gender-name dictionaries, such as WIPO's World Gender-Name Dictionary.

C Choosing which gender-sensitive indicator to use

Many useful indicators can be produced by using IP data broken down by gender, though these have different meanings and uses.



A Collecting innovation and IP gender data

Direct collection of gender data

The simplest approach is to request gender information when collecting innovation or IP data. For example, in the case of patent data, you can add fields for inventors and applicants to check on the patent application form:

Box No. II	APPLICANT <input type="checkbox"/> This person is also inventor	Gender: <input type="checkbox"/> F <input type="checkbox"/> M <input type="checkbox"/> Other
-------------------	--	--

This method has the **advantage** of capturing gender-related information **directly from the source**. Self-declaration also has the advantage of allowing users to identify themselves based on their preferred categories.

The **disadvantage** is that it **cannot be applied retroactively**, which can affect historical analysis. It can also be **resource intensive**, as it can take substantial time and money to implement a system change in an IP application form and for users to adapt to it. In addition, it may be **difficult to require that applicants disclose gender information**, due to privacy regulations. But a non-mandatory field could affect the statistical coverage of the exercise.

.....

Smart tip

Whenever possible, collect the gender of all the creators (e.g., inventors, designers, authors) for each IP application.

Gender disparities can occur within creative teams. You can only compute team indicators if you know the gender of all creators.

Indirect attribution

Using secondary sources

One way to attribute gender to existing innovation and IP data is to make use of secondary sources with gender information. Typical secondary sources containing comprehensive coverage and gender markers include national registers of individuals or national employment surveys.

These sources can be linked with the IP data, if individuals are clearly identified in each data set. For example, individual applicants, inventors and creators can be matched to these sources using social security numbers or national ID numbers.

IP data		National employment survey		
Soc. sec. no.	Inventor name	Soc. sec. no.	Employee name	Gender
11111190-2	John Smith	11111190-2	Smith, John	M
11111198-2	John S. Smith	11111191-2	Adams, John	M
11111192-1	Mary Smith	11111192-1	Smith, Mary S.	F
11111192-1	Mary S. Smith	11111193-1	Adams, Mary	F

The main advantages and disadvantages of this method relate to the quality of the secondary source. In other words, the gender attribution will be as good as the reliability of the secondary source.

Prioritize sources that give comprehensive coverage of individuals and that update the information regularly. A further advantage of this method is that it has a low implementation cost.

Smart tip

Note that some national ID numbers reserve a digit for gender information. In this case, you do not need a secondary source to attribute gender.

For instance, in Republic of Korea's ID numbers (RRN), the seventh digit is reserved for gender:

641226-4191838 = F
641226-3191835 = M

data

19,136
8,768
21,474

3,220
16,115
10,003
1,908
1,245
11,481
7,644
610

Patent Families

107
12
198
18,820
21,88
165
37,791
17,999
7,837
4
7,906
1,731
878

1,007
72
33,977
987
6,218
42,989
14,778

14,253
2,108
6,535
18,957
13,716
6,526

489
976

3,186
2,583
1,245
6,051

41,533
53,794

Using semantic rules

Alternatively, you can use **language customs and semantics** about individuals' names to attribute gender.

For example, **name honorific titles** such as “Mrs,” “Ms” or “Mr” in English can refer unambiguously to a given gender. However, titles such as “Dr.” or “Prof.” would require more information for attributing gender.

Similarly, **name structures** can refer to a given gender in some naming conventions. For instance, in Spanish or Portuguese naming convention, names ending in “o” frequently refer to men, and names ending in “a” frequently refer to women.

Unfortunately, many languages do not have clear gender distinctions based on semantic rules, and databases may not collect and store title information.

An advantage of this method is that, as long as naming conventions do not change substantially, it **can be applied retroactively**.

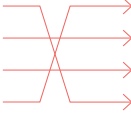
Smart tip

Note that in some languages it is common to mix female and male given names in composed names, like “José María” in Spanish, “Jean-Marie” in French, etc.

When attributing gender in these cases, it is often useful to consider the semantic rules about gender and the position of the given name.

Using gender-name dictionaries

Gender-name dictionaries are lists of names and the gender most commonly associated with them. These lists can be used to attribute gender to the inventors and creators listed in IP data based on their given names.

IP data			Gender-name dictionary	
Inventor surname	Inventor given name		Name	Gender
Smith	Sue		John	M
Smith	George		George	M
Smith	Mary		Mary	F
Smith	John		Sue	F

There are many advantages to using a gender-name dictionary. As long as naming conventions do not change substantially, this method can be easily applied retroactively with a low implementation cost.

A gender-name dictionary may have limitations. Its reliability depends on how adjusted the dictionary is to the languages and ethnicities present in the names of your IP data. In some cases, you will need to accept a certain degree of error. Also, in some languages, many names can be used for either women or men.

Where can I find these dictionaries?

Many name lists exist, and you should find the one that is best suited to your data. WIPO has compiled several of the main public sources to create the World Gender-Name Dictionary. You can find the World Gender-Name Dictionary and links to the underlying public sources on [the WIPO website](#).

How do I use the gender-name dictionary?

In the next section there is a step-by-step guide to using the dictionary to attribute your IP data.

B Using WIPO's World Gender-Name Dictionary

Step 1

Prepare your innovation or IP data

Ideally, IP data will have a field containing the given name for each inventor or creator (and individual applicants if possible).

You can prepare your data as follows:

Separate given names from surnames. If there are multiple given names keep them all, but in separate fields. Remember that the order can determine gender in many languages.

Clean the given names field(s) by removing leading and trailing blanks, double spaces and unnecessary characters (e.g., ' - , ;), and replacing accents and other non-Latin characters (e.g., é→e, ó→o, ß→ss, à→a).

Keep country or language data from the address or country fields.

Step 2

Use the most suitable version of the World Gender-Name Dictionary

Download the latest World Gender-Name Dictionary version from our Resources for Researchers Gender Repository (R4RGender). Check for additional tools there.

Select the World Gender-Name Dictionary **version** according to your country (e.g., France, Chile) or language (e.g., Arabic, Chinese), and apply it to the list of given names in step 1.

Remember, you can **apply more than one version** of the World Gender-Name Dictionary to your data.

Step 3

Assess the attribution results

Check your results! Each data set is different. You can also add new entries to the World Gender-Name Dictionary.

Analyze the coverage of your gender attribution. To what share of records has gender been attributed?

Manually inspect the attribution results. Inspect the 100 most frequent female, male and unattributed given names for (i) wrongly attributed and (ii) unattributed given names that are clearly gender specific in your country.

Create your own gender-name dictionary based on (i) and (ii). Use it to improve the results in step 2.

Send us feedback!

Smart tip

If you cannot find a gender-name dictionary for your country, try to find one from a country that uses the same language. You can also build your own dictionary by manually attributing gender to the most frequent (e.g., top 200) given names in your IP data.



C Choosing which gender-sensitive indicator to use

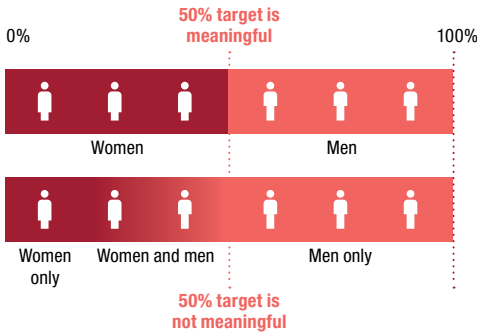
Many useful indicators can be produced using innovation and IP data broken down by gender. However, they have different meanings and uses.

Gender benchmarks

Women listed as inventors or creators is the most intuitive indicator, as it provides **clear targets for gender policies**, for example 50 percent women listed as inventors or creators.

The main limitation is that IP applications can have more than one inventor or creator, so the total does not reflect the actual number of IP filings.

The percentage of applications with at least one woman inventor or creator uses the same unit as other typical IP indicators.



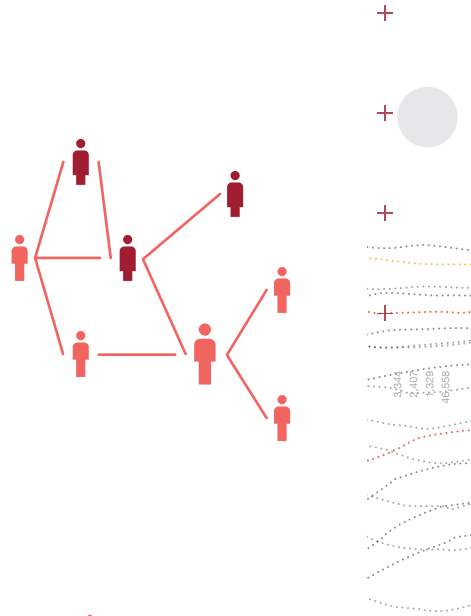
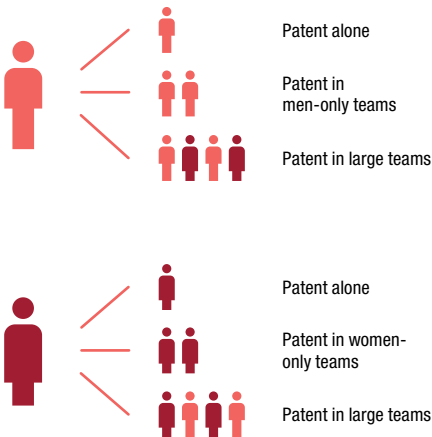
However, it provides no clear gender policy target and **risks understating gender imbalances**.

 **Indicator: percentage of women listed as inventors or creators**

Collaborators and group size

The number of co-inventors and co-creators in the same IP application can provide information on the nature of collaboration and gender distribution.

Collaborations can be illustrated by **network data visualizations**.



 **Indicator: percentage of IP filings with women only or with majority women**

 **Indicator: average team size by gender**

 **Indicator: co-inventor networks by gender**


indicators

Industry and sectors

Many IP classifications (e.g., International Patent Classification, Nice and Locarno) enable analysis of women’s representation in **industries or technological fields**.

Information on the gender gap by sector can come from analysis of the representation of women listed as inventors in patent applications from **companies and universities**.

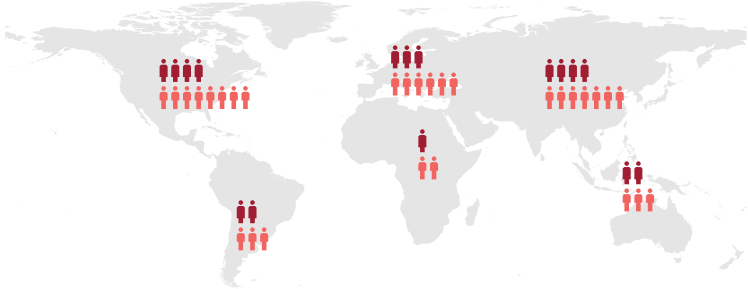
 **Indicator: percentage of women listed as inventors by class**

 **Indicator: percentage of women listed as inventors by applicant type**

Geography

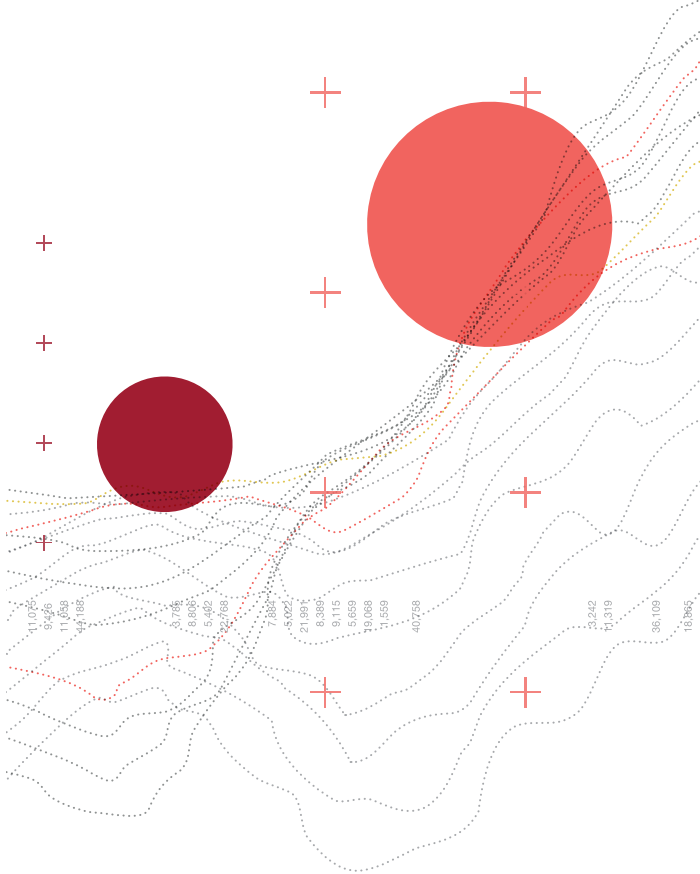
Addresses of inventors or creators enable deeper analysis of gender disparities by country, region or city.

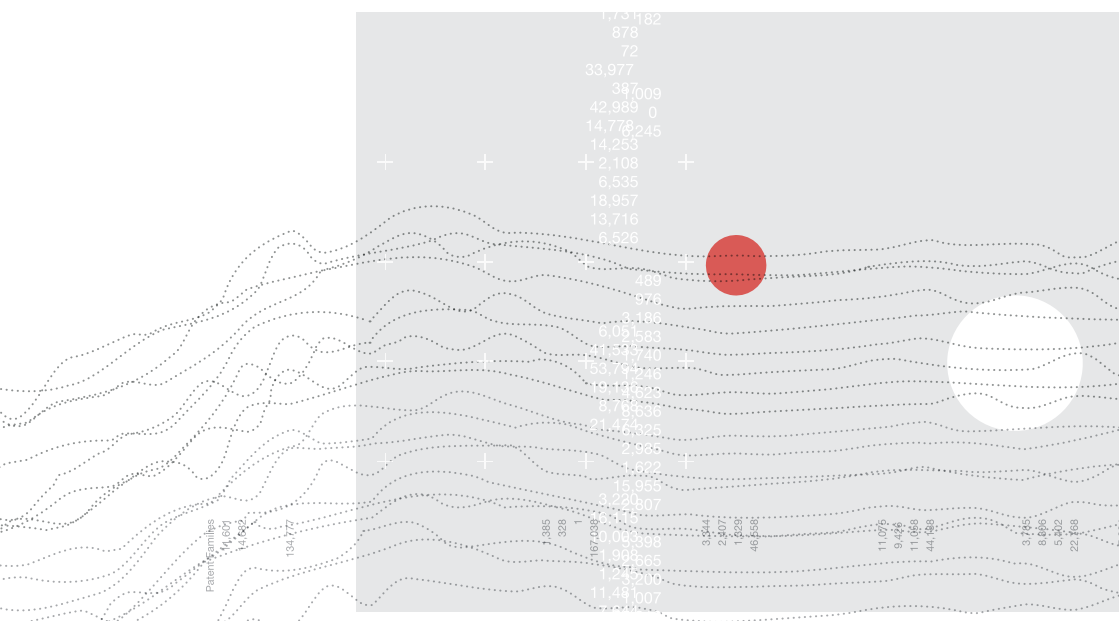
Geographical distribution can be illustrated with **data-generated maps**.



 **Indicator: percentage of women listed as inventors or creators by state or province**

 **Indicator: percentage of women listed as inventors or creators by city**





World Intellectual Property Organization
 34, chemin des Colombettes
 P.O. Box 18
 CH-1211 Geneva 20
 Switzerland

Tel: +41 22 338 91 11
 Fax: +41 22 733 54 28

For contact details of WIPO's
 External Offices visit:
www.wipo.int/about-wipo/en/offices

© WIPO, 2022



Attribution 4.0 International
 (CC BY 4.0)

The CC license does not apply to non-WIPO content in this publication.

WIPO Publication No. 1073E
 DOI 10.34667/tind.45020
 ISBN 978-92-805-3410-8