

Open data on higher education and research

Documentation

Ministry of Higher education and Research

2022 Edition, version 1

Introduction

The Ministry of Higher education and Research in Luxembourg awards financial aid to students studying in a higher education institution in the form of grants and loans. There are several eligibility criteria, but in simple terms, any student residing in Luxembourg, or whose parents work in Luxembourg, is eligible (non-residents are eligible since 2013). Each academic year, a document was published by the ministry containing a series of statistics, outlining the number of beneficiaries disaggregated by country of residence, country of study, citizenship, field of study... These documents can be found on the website of the ministry, on following link¹.

Since the academic year 2021/2022, the data behind these graphs is now available as a flat file on the open data portal of Luxembourg. The portal can be accessed here², and the data itself can be downloaded here.

This document provides an overview of the variables included in the data, as well as important remarks to properly interpret the data.

High-level overview of the data

The source of the data is the financial aid data base maintained at the Ministry of Higher education and Research in Luxembourg, which contains information on every single grant application. The data published on the open data portal contains only a subset of the available information, and is aggregated. Data starts at the 2016/2017 academic year until the academic year 2021/2022. New data is released in autumn each year, according to the schedule detailed in the next section. The Excel file also contains two datasets on the BTS programmes. The source for this data are the ministry's administrative data base which is used to manage the programmes.

Publication schedule

The last academic year available in the data is provisional, and is to be considered final after two years. The table below summarises the publication schedule:

Year Y	Year Y+1	Year Y+2
Provisional data for Y-1	Provisional data for Y	Provisional data for Y+1
Final data for Y-2	Final data for Y-1	Final data for Y
Final data for Y-3	Final data for Y-2	Final data for Y-1
...

As such, any result computed using provisional data is subject to change with future revisions of the data. In rare cases, it may happen that data that was considered final may change. This should happen rarely, and

¹<https://mesr.gouvernement.lu/en/publications.html>

²<https://data.public.lu/en/>

the impact of such changes very limited.

List of datasets

The Excel document has 6 sheets, each one containing a flat dataset. Here is a brief description of each dataset:

- **effectifs**: shows the number of students per academic year and other variables;
- **demarches**: shows the number of grant applications per academic year;
- **montants**: shows the amount of grants and loans awarded per academic year;
- **beneficiaires**: shows the number of students that received each type of grant or loan;
- **bts_inscription**: shows the number of students studying in a BTS programme;
- **bts_diplomed**: shows the number of BTS degrees awarded per academic year.

The first 4 datasets are related to financial aid, while the last 2 are not. This means that the population for the first 4 datasets are students that ask for financial aid or in other words, a non-random sample of the whole population of students. However, since the eligibility criteria are broad, it is very likely that this sample covers a very large part of the population, at least for residents. The last 2 datasets are the actual complete population of students studying in a BTS programme.

Most of the data is available per academic year and semester. Some care is required when analyzing these data. Indeed, in some cases computing a total per academic year does not make sense. For example, computing the total, per academic year, of students in the dataset **effectifs** is wrong. This is because most students apply for financial aid in the winter and in the summer semester. Thus, summing them up would result in counting most students twice. This is not an issue for the dataset **demarches**, because that would result in the total number of applications per year. Since most students introduce at least two applications, one for the winter semester and one for the summer semester, this results in a correct figure that can be interpreted.

List of variables

Student-specific variables

- **resident**: a factor with two levels, **resident** or **non-resident**;
- **pays_residence_iso3**: a factor with 12 levels, giving the country of residence of the students: **Autres pays** (other countries), **AUT** (Austria), **BEL** (Belgium), **CHE** (Switzerland), **DEU** (Germany), **ESP** (Spain), **FRA** (France), **ITA** (Italy), **LUX** (Luxembourg), **NLD** (Netherlands), **PRT** (Portugal), and missing values
- **sexe**: a factor with two levels, **M** and **F**;
- **nationalite**: a factor with seven levels, giving citizenship of the students: **Autres** (Other), **BEL** (Belgium), **DEU** (Germany), **FRA** (France), **LUX** (Luxembourg), **PRT** (Portugal)

Institution-specific variables

- **pays_etablissement_iso3**: a factor with 15 levels, giving the country of study of the students: **Autres pays** (other countries), **AUT** (Austria), **BEL** (Belgium), **CAN** (Canada), **CHE** (Switzerland), **DEU** (Germany), **ESP** (Spain), **FRA** (France), **GBR** (Great-Britain), **IRL** (Ireland), **ITA** (Italy), **LUX** (Luxembourg), **NLD** (Netherlands), **PRT** (Portugal), and **Pays manquant** (missing value, usually because the students did not fill out this part of the form. In most cases, incomplete requests for financial aid are denied.);
- **localite_etablissement**: a factor with many levels, giving the city of study;
- **domaine_formation**: a factor with 18 levels, giving the field of study of the student;
- **diplome**: a factor with 8 levels, giving the degree the student is currently completing

Grant-specific variables

- **annee_aca**: academic year;
- **semestre**: the semester an application was introduced;

- **demande_statut**: a factor with two levels, **REFUSE** (financial aid application rejected) and **VALIDE** (financial aid application accepted);
- **type_bourse**: a factor with 6 levels, giving the type of financial aid received (grant);
- **type_pret**: a factor with 6 levels, giving the type of financial aid received (loan);

Quantitative variables

- **total_etudiants**: Total students in each category;
- **total_beneficiaires**: Total beneficiaries of each type of grant (only relevant for the sheet called **beneficiaires**);
- **total_bourse**: Total amount of grants awarded for all students in this category;
- **total_pret**: Total amount of loans awarded for all students in this category

BTS specific variables

- **annee_aca**: academic year from;
- **semestre**: semester;
- **resident**: a factor with two levels, **resident** or **non-resident**;
- **sexe**: a factor with two levels, **M** and **F**;
- **nom_institution**: a factor with 12 levels indicating the school at which the student is enrolled: **ECG** (École de Commerce et de Gestion), **EHTL** (École d'Hôtellerie et de Tourisme du Luxembourg), **LAM** (Lycée des Arts et Métiers), **LCD** (Lycée Classique de Diekirch), **LGK** (Lycée Guillaume Kroll) **LJBM** (Lycée Josy Barthel Mamer), **LN** (Lycée du Nord), **LTB** (Lycée Technique de Bonnevoie), **LTC** (Lycée Technique du Centre), **LTETT** (Lycée Technique d'Ettelbruck), **LTPS** (Lycée Technique pour Professions de Santé), **MLG** (Maacher Lycée Grevenmacher)
- **domaine_formation**: a factor with 6 levels, giving the field of study of the student;
- **total_etudiants**: Total students in each category;

Important remarks

Breaks in the data

- Starting with the academic year 2021/2022, the amount reported for loans is the actual amount taken up by students. Before this date, it was the amount granted, but students always had the option to not take the loan.
- Starting with the academic year 2021/2022, every application is introduced online, with some rare exceptions, for example, when a student provided wrong information and this must be corrected past the submission deadline. In such cases a paper application is used.

Preparation steps for released

The data released in the open data portal was aggregated and cleaned before release. Here are the various steps that were taken to clean the data:

- Grant applications that were started but never completed were ignored;
- Duplicated lines were removed;
- Missing data was filled: for instance, in cases where the city of study was only given in the winter semester, the same value was used for the summer semester, unless the student also changed university;
- Degree names (and spelling mistakes) were normalized: for instance, “licence”, “bacharelato”, “baachelor” were all collapsed into “Bachelor”;
- Rarely occurring degrees, nationalities, cities of study or fields were collapsed into an “Other” category;
- Country names were replaced by their iso3 codes, and country names occurring rarely were collapsed into an “Other” category;

Several unit tests were written to ensure that these steps did not introduce any errors, however if, despite all the care we put into preparing these data, you find any errors, or have any questions, contact the ministry's statistician Bruno Rodrigues: `bruno [dot] rodrigues [at] mesr [dot] etat [dot] lu`.