Data Article

An integrated dataset providing a landscape of the Italian Universities

Renato Bruni [a], Cinzia Daraio [a], Simone Di Leo [a,*], Omid Ghamiloo [a], Gianpiero Bianchi [b]

[a] Department of Computer Control and Management Engineering, "Sapienza" University of Rome, Rome, Italy
[b] Dir. for Data Collection, Italian National Institute of Statistics "Istat", Rome, Italy

ARTICLE INFO

ABSTRACT

We provide data describing all the largest Italian Universities from several perspectives, including scientific research, administrative and economic point of view. In particular, associated with each University, we have the following data.

(a) The list of the most representative research keywords (single words and bigrams), automatically extracted from titles, abstracts and other possible metadata of all the research publications available for that University in Scopus database at October 2022.

(b) The Extended_name of the University, Status, University_Type, State_status, number of Managerial and Administrative Staff, Teaching Staff and Researchers, Phd Diplomas, Phd Enrollments, Enrolled Undergraduates, Enrolled Graduates, Graduates, Master I Lv Graduates, Enrolled Master's Degree I Lv, Master II Lv, Graduates Enrolled, Master II Lv, Graduates Specialistic Schools and Enrolled Specialistic Schools were extracted from USTAT database for the years 2016–2018.

(c) The data of Educational Income, Income from Commissioned Research and Technology Transfer, Income from Research with competitive funding, Own Income, Contributions from others (private), Contributions from others (public), Contributions from universities, Contributions from the European Union and the Rest of the World, Contributions

* Corresponding author.
  E-mail address: dileo@diag.uniroma1.it (S. Di Leo).
  Social media: @simonedileo95 (S. Di Leo).

from other local governments, Contributions Regions and Autonomous Provinces, MIUR and other central government grants, Operating Costs, Current Management Costs, Managerial and administrative personnel costs, Research and teaching staff cost, Cost of Lecturers and Researchers, Cost Scientific Collaborators, Cost of Contract Teachers, Cost of Language Experts, Other research and teaching personnel costs, Personnel Costs, Scientific equipment, Concessions, licenses and trademarks, Patent rights are extracted from the unique University Balance Sheet of each university for the years 2016–2018.

These data were of difficult availability; they have been extracted from several heterogeneous sources and have been automatically checked, cleaned from errors, integrated, and missing values have been imputed as much as possible. However, due to large missing portions in the sources, they still contain several missing parts.

Nonetheless, they represent a powerful snapshot of the Italian Universities, and can be of interest to researchers for many analyses of the Italian academic world. All the sources of the openly available data are provided.

Specifications Table

Subject:	Management Information Systems
Specific subject area:	Data from several sources describing Italian Universities from the scientific research, administrative and economic point of view. Can be used for several analyses of the Italian academic world.
Type of data:	Tables in the form of MS Excel spreadsheets.
How the data were acquired:	The data were obtained from 5 different sources.
	- The list of the most representative research keywords of each University have been automatically extracted from titles, abstracts and other possible metadata of all the research publications available for that University in Scopus database at October 2022. The process was based on web scraping, lemmatization, stop-word elimination and identification of relevant words based on frequency and word embedding techniques. After this, we selected the top 30 single words and the top 40 bigrams.
	- The Extended_name, Status, University_Type, State_status, number of Managerial and Administrative Staff, Teaching Staff and Researchers, Phd Diplomas, Phd Enrollments, Enrolled Undergraduates, Enrolled Graduates, Graduates, Master I Lv Graduates, Enrolled Master's Degree I Lv, Master II Lv, Graduates Enrolled, Master II Lv, Graduates Specialistic Schools and Enrolled Specialistic Schools were extracted from USTAT database (the official database of the Italian Ministry of University) for the years 2016–2018.
	- The number of active spin-offs were extracted from Spin-Off Italia (database of spin off and public research). The data are available for the year 2018 only. The data were accessed on the 5th of November 2020, before it became accessible only upon registration.
	- The data of educational income, Income from Commissioned Research and Technology Transfer, Income from Research with competitive funding, Own Income, Contributions from others (private), Contributions from others (public), Contributions from universities, Contributions from the European Union and the Rest of the World, Contributions from other local governments, Contributions from Regions and Autonomous Provinces, MIUR and other central government grants, Operating Costs, Current Management Costs, Managerial and administrative personnel costs, Research and teaching staff cost, Cost of Lecturers and Researchers, Cost Scientific Collaborators, Cost of Contract Teachers, Cost of Language Experts, Other research and teaching personnel costs, Personnel Costs, Scientific equipment, Concessions, licenses and trademarks, Patent rights income were extracted manually from the unique University Balance Sheet of each university for the years 2016–2018.

	- Finally, Institution Name, English Institution Name, Institution AcronymInstitution Category – English, Foundation year, University hospital, Region of establishment (NUTS 2), Region of establishment (NUTS 3), Name of the city, Geographic coordinates – latitude, Geographic coordinates – longitude, Postcode were extracted from ETER (the European tertiary education register).
Data format:	Raw, Imputed and Analyzed
Description of data collection:	The data were extracted from the source databases and processed with several software systems in order to integrate them, remove errors or inconsistencies, impute missing values, identify the desired information.
	In particular, the most representative research keywords of each Italian university have been automatically extracted by using text mining procedures from titles and abstracts of all the research publications available for that university in Scopus database at October 2022. Other university data have been extracted from: USTAT database (the official database of the Italian Ministry of University); Spin-Off Italia (database of spin off and public research); ETER (the European tertiary education register). Moreover, accounting data were extracted manually from the unique University Balance Sheet of each university.
	We also provide in this article a table for the integration of the data obtained from ETER and USTAT through a comprehensive mapping of available institutions. This allows to update and/or expand the dataset more easily.
Data source location:	Primary data sources:
	1) Scopus website: https://www.scopus.com/
	2) ETER Project website: https://www.eter-project.com/#/home
	3) USTAT website: http://ustat.miur.it/opendata/
	4) Portal containing the unique University Balance Sheets:
	https://coep.miur.it/index.php?action=secured-docs
	5) Spin-off Italia: https://www.spinoffitalia.it/
Data accessibility	Mendley Dataset [1], DOI: 10.17632/pycv47nk3p

Value of the Data

- These data are the only available integrated description of the landscape of Italian Universities from the scientific research, administrative and economic point of view.
- These data were obtained by applying several processing steps, some of which quite computationally demanding, in order to integrate them, remove errors or inconsistencies (see also [2]), impute missing values, identify the desired information (for details on the text mining techniques see [3,4]).
- Researchers interested in analysing the Italian academic landscape and the research conducted in Italy can benefit from these datasets. Moreover, they can be of interest to administrators and policymakers working in higher education and to researchers in the sociology of education, education policy, public administration, economics and education statistics.
- The case of the Italian university system is interesting because the system has been greatly affected by several processes of change over the past four decades: the reforms of the 1990s and early 2000s, and the shrinking in the budgets, have stimulated competition for enrolment [5]. Since 1990, numerous new universities have been created, further contributing to the intensification of competition.
- The Italian University system, despite having low tuition, is characterized by low enrolment rates and even lower completion rates than those in other European and American countries. The advent of evaluation deeply affected the system [5,6]. However the evaluation should not be limited to a single dimension (teaching or research) but should consider the multidimensional nature of the activities carried out, since research and teaching are interconnected [7].

1. Objective

The main objective behind the creation of this dataset is to provide an overview of all available data sources on the Italian university system and to offer an integrated database to be able to perform quantitative analyses on the Italian University system. To study this system and de-

velop clustering and comparison analyses of the different universities, which are characterized by considerable heterogeneity [8], a considerable amount of information is needed, regarding the *resources* used and the *outputs* produced by both teaching and research. It is also important to fully consider the contexts in which the different universities are located.

The dataset provided here allows for an overview of the teaching and research activities carried out by Italian universities, and it also describes the resources and infrastructure available to the universities in their context.

2. Data Description

The excel file with the data is divided into several sheets. Each sheet contains information about Italian universities. The "ETER INFORMATION" sheet contains information of geographical description and general information of Italian universities reported in ETER. The "ECONOMIC INFORMATION" sheet contains information on budget data. The "NUMERIC INFORMATION" sheet contains information related to enrollment and number of staff. The "USTAT INFORMATION" sheet contains description information on Italian institutions extracted from the USTAT database. The enrollment and number of staff data refer to the period 2016–2018. Within the file are shown in red and marked with the value of cell "m" the cases of missing value. The "RELEVANT_SINGLEWORDS" sheet and the "RELEVANT_BIGRAMS" sheet contain information respectively on the 30 most relevant single words and the 40 most relevant bigrams describing the reseach activity of Italian universities. Table 1 shows the description of the location of each variable in the various sheets of the file.

The excel file contains also a sheet named "ID_MAP" with the mapping between the unique identifiers of ETER, SCOPUS, USTAT and economic data collected. The following Table 2 reports the description of each collected variable and the date when the data were retrieved.

The excel file also contains a "DATA_SUMMARY" sheet, with the organization of the collected data (in practice the same information shown in the above Tables 1 and 2).

3. Experimental Design, Materials and Methods

As reported in the Specifications table, the data were retrieved from 5 different sources.

- The list of the most representative research keywords of the 78 largest Italian Universities have been automatically extracted by means of text mining operations on the bibliographic information of all the research publications available for that University in Scopus database at October 2022. After an initial web scraping operation to obtain the corpus of text, we performed stop-word elimination to remove all irrelevant or generic words. After this, we identified the most relevant terms by using the word embedding algorithm Word2vec [9]. This technique represents each distinct word with a particular list of numbers called its vector. The vectors are generated in such a way that the cosine similarity between the vectors indicates the level of semantic similarity between the corresponding words. In particular, we applied this for single words, and also for couples of words to obtain the relevant bigrams (couples of words generally used together). For single words, we also performed Lemmatization with Part-Of-Speech recognition. This means that, for each word, we remove the inflectional ending to identify its basic lemma, see also [3]. This allows to recognize together the different inflected forms of a word (e.g., plurals of nouns, tenses of the verbs, etc.). For bigrams, we did not apply lemmatization since this would destroy the specific inflectional ending with which the bigram is used. Finally, we selected the top 30 most relevant single words and the top 40 most relevant bigrams for each university. Note that these data were not already present in Scopus in this form.
- The data regarding teaching and administrative staff, enrolled students, graduates and similar were extracted from USTAT database (the official database of the Italian Ministry

Table 1
Structure of excel file including sheet indentification for each indicator present.

Variable	Data Source	Excel Spreadsheet
ETER ID	**ETER**	ETER_INFORMATION
Institution Name	ETER	ETER_INFORMATION
English Institution Name	ETER	ETER_INFORMATION
Institution Acronym	ETER	ETER_INFORMATION
Institution Category - English	ETER	ETER_INFORMATION
Foundation year	ETER	ETER_INFORMATION
University hospital	ETER	ETER_INFORMATION
Region of establishment (NUTS 2)	ETER	ETER_INFORMATION
Region of establishment (NUTS 3)	ETER	ETER_INFORMATION
Name of the city	ETER	ETER_INFORMATION
Geographic coordinates - latitude	ETER	ETER_INFORMATION
Geographic coordinates - longitude	ETER	ETER_INFORMATION
Postcode	ETER	ETER_INFORMATION
BA_ID	**Personal elaboration**	
Educational income	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Income from Commissioned Research and Technology Transfer	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Income from Research with competitive funding	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Own Income	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Contributions from others (private)	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Contributions from others (public)	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Contributions from universities	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Contributions from the European Union and the Rest of the World.	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Contributions from other local governments	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Contributions Regions and Autonomous Provinces	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
MIUR and other central government grants	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Operating Costs	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Current Management Costs	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Managerial and administrative personnel costs	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Research and teaching staff cost	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Cost of Lecturers and Researchers	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Cost Scientific Collaborators	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Cost of Contract Teachers	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Cost of Language Experts	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Other research and teaching personnel costs	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Personnel Costs	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Scientific equipment	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
Concessions, licenses and trademarks	Unique University Balance Sheet Website	ECONOMIC_INFORMATION

Table 1 (*continued*)

Variable	Data Source	Excel Spreadsheet
Patent rights	Unique University Balance Sheet Website	ECONOMIC_INFORMATION
USTAT_ID	**USTAT**	
Number Of Managerial And Administrative Staff	USTAT	NUMERIC_INFORMATION
Number Of Teaching Staff And Researchers	USTAT	NUMERIC_INFORMATION
Number Of Departments	CercaUniversita	NUMERIC_INFORMATION
Active Spin Offs	Spin off Italia	NUMERIC_INFORMATION
Phd Diplomas	USTAT	NUMERIC_INFORMATION
Phd Enrollments.	USTAT	NUMERIC_INFORMATION
Enrolled Undergraduates	USTAT	NUMERIC_INFORMATION
Enrolled Graduates	USTAT	NUMERIC_INFORMATION
Graduates	USTAT	NUMERIC_INFORMATION
Master I Lv Graduates	USTAT	NUMERIC_INFORMATION
Enrolled Master's Degree I Lv	USTAT	NUMERIC_INFORMATION
Master II Lv Graduates	USTAT	NUMERIC_INFORMATION
Enrolled Master II Lv	USTAT	NUMERIC_INFORMATION
Graduates Specialistic Schools	USTAT	NUMERIC_INFORMATION
Enrolled Specialistic Schools	USTAT	NUMERIC_INFORMATION
Extended_name	USTAT	USTAT_INFORMATION
Status	USTAT	USTAT_INFORMATION
University_Type	USTAT	USTAT_INFORMATION
State_status	USTAT	USTAT_INFORMATION
ID_MAP	**Personal elaboration**	**ID_MAP**
affiliationID	SCOPUS	RELEVANT_SINGLEWORDS
Relevant_SingleWords	**Personal elaboration**	RELEVANT_SINGLEWORDS
affiliationID	SCOPUS	RELEVANT_BIGRAMS
Relevant_Bigrams	**Personal elaboration**	RELEVANT_BIGRAMS

- of University) for the 78 largest Italian Universities for the years 2016–2018 by accessing the USTAT website.
- The number of active spin-offs were extracted from Spin-Off Italia (database of spin off and public research). The data are available for the year 2018 only. The data were accessed on the 5th of November 2020, before it became accessible only upon registration.
- The data regarding income, funding, contributions and costs were extracted manually from the unique University Balance Sheet of each university for the years 2016–2018.
- Finally, general information like Institution Name, Category, Geographic data and similar were extracted from ETER (the European tertiary education register, see also [10]) by accessing ETER website.

Some of the collected data refer to a limited period (years 2018 or 2016–2018), thus not allowing longitudinal analyses, because Italian universities changed their accounting system in 2016 (with Law 240 of 2010, [11]), making it impossible to compare the collected budget items with the previous years. Previously, data for individual universities were collected in the institutional platform called *Bilanci Atenei* (https://ba.miur.it/), which no longer collects data from 2016 onwards. The data collected from budgets stops at present in 2018. This is a limitation that will be overcome in the future whenever more data will be made available.

To the best of our knowledge, there are not similar comprehensive datasets for other countries as the one we provide in this work for Italy. However, there are some websites of national official bodies that report, in part, some of the information collected in our dataset. For the European universities we suggest the European Tertiary Education Register (ETER, https://www.eter-project.com/data-for-download-and-visualisations/database/). Inside the file are the ETER_ID codes to link our dataset with European data. In ETER it is possible to extend the search to other European universities in addition to Italian universities. For data from British universities, we suggest the Higher Education Statistics Agency (HESA) dataset. Several datasets are available on their site for economic/financial and personnel/student data

Table 2
Description of collected variables and date of access.

Variable	Date accessed	Description
ETER ID	**15/12/2022**	**Eter unique ideintifier for the institution. Can be used also to integrate other data from RISIS 2 Datasets.**
Institution Name	15/12/2022	Name of the institution (Local Language)
English Institution Name	15/12/2022	Name of the institution (English)
Institution Acronym	15/12/2022	Acronym of the institution
Institution Category - English	15/12/2022	Category of the institution: University, Technical University, Online University, Higher school, University for foreigners
Foundation year	15/12/2022	Foundation year of the institution
University hospital	15/12/2022	Value to identify if university has hospital. 0=no hospital, 1=yes
Region of establishment (NUTS 2)	15/12/2022	Region of establishment (NUTS 2)
Region of establishment (NUTS 3)	15/12/2022	Region of establishment (NUTS 3)
Name of the city	15/12/2022	Name of the city
Geographic coordinates - latitude	15/12/2022	Geographic coordinates - latitude
Geographic coordinates - longitude	15/12/2022	Geographic coordinates - longitude
Postcode	15/12/2022	Postcode
BA_ID		**Unique identifier for the institution in this database with available economic data**
Educational income	15/12/2022	Educational income
Income from Commissioned Research and Technology Transfer	15/12/2022	Income from Commissioned Research and Technology Transfer
Income from Research with competitive funding	15/12/2022	Income from Research with competitive funding
Own Income	15/12/2022	Total Income
Contributions from others (private)	15/12/2022	Contributions from others (private)
Contributions from others (public)	15/12/2022	Contributions from others (public)
Contributions from universities	15/12/2022	Contributions from universities
Contributions from the European Union and the Rest of the World.	15/12/2022	Contributions from the European Union and the Rest of the World.
Contributions from other local governments	15/12/2022	Contributions from other local governments
Contributions Regions and Autonomous Provinces	15/12/2022	Contributions Regions and Autonomous Provinces
MIUR and other central government grants	15/12/2022	Miur and other central government contributions
Operating Costs	15/12/2022	Operating Costs
Current Management Costs	15/12/2022	Current Management Costs
Managerial and administrative personnel costs	15/12/2022	Managerial and administrative personnel costs
Research and teaching staff cost	15/12/2022	Research and teaching staff cost (sum of Cost of Teachers and Researchers, Cost Scientific Collaborators, Cost of Contract Teachers, Cost of Language Experts, Other research and teaching personnel costs)
Cost of Lecturers and Researchers	15/12/2022	Cost of Teachers and Researchers
Cost Scientific Collaborators	15/12/2022	Cost Scientific Collaborators
Cost of Contract Teachers	15/12/2022	Cost of Contract Teachers
Cost of Language Experts	15/12/2022	Cost of Language Experts
Other research and teaching personnel costs	15/12/2022	Other research and teaching personnel costs
Personnel Costs	15/12/2022	Personnel Costs (sum of all personell costs)
Scientific equipment	15/12/2022	Scientific equipment (Balance sheet item)
Concessions, licenses and trademarks	15/12/2022	Concessions, licenses and trademarks (Balance sheet item)
Patent rights	15/12/2022	Patent rights (Balance sheet item)
USTAT_ID	**05/11/2020**	**USTAT unique identifier**
Number Of Managerial And Administrative Staff	05/11/2020	Number Of Managerial And Administrative Staff
Number Of Teaching Staff And Researchers	05/11/2020	Number Of Teaching Staff And Researchers
Number Of Departments	05/11/2020	Number Of Departments

Table 2 (*continued*)

Variable	Date accessed	Description
Active Spin Offs	05/11/2020	Active Spin Offs
Phd Diplomas	05/11/2020	Phd Diplomas
Phd Enrollments.	05/11/2020	Phd Enrollments.
Enrolled Undergraduates	05/11/2020	Enrolled Undergraduates
Enrolled Graduates	05/11/2020	Enrolled Graduates
Graduates	05/11/2020	Graduates
Master I Lv Graduates	05/11/2020	Master I Lv Graduates
Enrolled Master's Degree I Lv	05/11/2020	Enrolled Master's Degree I Lv
Master II Lv Graduates	05/11/2020	Master Ii Lv Graduates
Enrolled Master II Lv	05/11/2020	Enrolled Master Ii Lv
Graduates Specialistic Schools	05/11/2020	Graduates Spec Schools
Enrolled Specialistic Schools	05/11/2020	Enrolled Schools Spec
Extended_name	20/12/2022	Extended name of the institution (in Italian)
Status	20/12/2022	Variable to check whether, according to MIUR, the university is still active or not
University_Type	20/12/2022	Variable to identify the type of university according to MIUR
State_status	20/12/2022	Variable to identify whether the university is a state university or not
ID_MAP		**Table to integrate different information from ETER, BA and USTAT. With this mapping, it is possible to further integrate and update the data. Data from USTAT, being provided directly by the Italian Ministry of Education, are updated very frequently.**
affiliationID	october 2022	Scopus affiliation ID
Relevant_SingleWords	4/1/2023	list of the 30 most representative single words
affiliationID	october 2022	Scopus affiliation ID
Relevant_Bigrams	4/1/2023	ist of the 40 most representative bigrams

of universities in Great Britain (https://www.hesa.ac.uk/). For the USA universities we suggest the Integrated Postsecondary Education Data System (IPSED, https://nces.ed.gov/ipeds/). The IPSED database contains information corresponding to that collected in our dataset, such as revenues, costs, assets, personnel, and number of students. For the Australian universities we suggest the official data from the Australian Department of education reporting student and staff (https://www.education.gov.au/higher-education-statistics) and financial data (https://www.education.gov.au/higher-education-publications/finance-publication). For Canadian universities we suggest the Canadian Association of University Business Officers (CAUBO) Financial Information of Universities and Colleges database (FIUC, https://www.caubo.ca/knowledge-center/analytics-and-reports/fiuc-reports/). Their database is a valuable source of information for financial data at Canadian universities and colleges. The FIUC is the only national source for comparable and consistent financial information for Canadian universities.

Those listed above are examples of the databases containing the same type of information as our dataset. Moreover, additional details on the institutions are often available on the national websites of the various national statistical institutes. However, the main limitation of these databases for comparison comes both from the difference in accounting systems in the various countries and the local language in which the data description and collection documents are written. For example, data on costs and revenues per institution are available on the website of the Spanish Ministry of Universities (https://www.universidades.gob.es/catalogo-de-datos/) in their platform "Sistema Integrado de Información Universitaria" (https://www.universidades.gob.es/sistema-integrado-de-informacion-universitaria/) but the detailed documentation is written in Spanish.

Due to the heterogeneity of the sources, a number of procedures had to be developed to link and integrate the data, check their correctness, or at least their plausibility and compatibility, and impute the missing values [2]. Some of these procedures were quite computationally

demanding, too. Note, however, that, due to large missing portions in the sources, they still contain several missing parts. Despite this, the integrated dataset presented in this work constitutes a rare and powerful snapshot of the situation of the Italian Universities. Thus, this dataset can be of interest to researchers interested in analysing the Italian academic landscape and the research conducted in Italy from several points of view. Moreover, it can be of interest to administrators and policymakers working in higher education [12].

Ethics Statements

This work did not involve the use of human subjects or animals.

CRediT Author Statement

Renato Bruni: Conceptualization, Methodology, Writing **Cinzia Daraio**: Conceptualization, Methodology, Writing. **Simone Di Leo**: Data Curation, Investigation, Writing. **Omid Ghamiloo:** Software, Data Curation. **Gianpiero Bianchi**: Software, Writing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Landscape of the Italian Universities (Original data) (Mendeley Data).

Acknowledgments

References

[1] G. Bianchi, R. Bruni, C. Daraio, S. Di Leo, O. Ghamiloo, An integrated dataset providing a landscape of the Italian Universities, Mendeley Data, V1 (2023), doi:10.17632/pycv47nk3p.

[2] R. Bruni, C. Daraio, D. Aureli, Imputation techniques for the reconstruction of missing interconnected data from higher educational institutions, Knowl.-Based Syst. 212 (2021) 106512, doi:10.1016/j.knosys.2020.106512.

[3] R. Bruni, G. Bianchi, Website categorization: a formal approach and robustness analysis in the case of e-commerce detection, Expert Syst. Appl. 142 (2020) 113001.

[4] R. Bruni, G. Bianchi, P. Papa, Hyperparameter black-box optimization to improve the automatic classification of support tickets, Algorithms 16 (1) (2023) 46, doi:10.3390/a16010046.

[5] F. Rossi, Increased competition and diversity in higher education: an empirical analysis of the Italian university system, Higher Educ. Policy 22 (4) (2009) 389–413.

[6] G. Capano, The re-regulation of the Italian university system through quality assurance. A mechanistic perspective, Policy Society 33 (3) (2014) 199–213.

[7] D. Checchi, T. Jappelli, A. Uricchio, Teaching, Research and Academic Careers: An Analysis of the Interrelations and Impacts, Springer, 2022.

[8] R. Bruni, G. Catalano, C. Daraio, M. Gregori, H.F. Moed, Studying the heterogeneity of European higher education institutions, Scientometrics 125 (2) (2020) 1117–1144.

[9] T. Mikolov, Q.V. Le, I. Sutskever: Exploiting similarities among languages for machine translation, in arXiv:1309.4168, 2013.

[10] C. Daraio, R. Bruni, G. Catalano, A. Daraio, G. Matteucci, M. Scannapieco, D. Wagner-Schuster, B. Lepori, European Tertiary Education Register (ETER): evolution of the data quality approach, J. Data Inf. Sci. (2020), doi:10.2478/jdis-2020-0029.
[11] Law no. 240 of 30 December 2010 published in G.U. no. 10 of 14 January 2011 - Ordinary Suppl. no. 11.
[12] G. Rebora, M. Turri, Critical factors in the use of evaluation in Italian universities, Higher Educ. 61 (5) (2011) 531–544.