

**Documentation of statistics for
Research and development in the public sector 2016**

1 Introduction

The purpose of the R&D statistics is, among other things, to analyse the scope of research and experimental development undertaken within the public sector. The results forms part of the measurement of R&D in relation to GDP. The survey is conducted in accordance with OECDs guidelines for R&D described in the Frascati Manual.

On 1st January 2008, Statistics Denmark took over the responsibility of compiling statistics on research, development and innovation from the Danish Centre for Studies in Research and Research Policy.

2 Statistical presentation

The statistics on Research and Development (R&D) cover a large number of indicators describing the resources used with the focus being expenditures and personel (in numbers and in full-time equivalents). The information is distributed by main sectors, types of expenditure, fields of science and type of research.

2.1 Data description

The statistics regarding public research and development (R&D) cover a large number of indicators describing the resources used with the focus being expenditures and personel (in numbers and in full-time equivalents). The information is distributed by main sectors, types of expenditure, fields of science and type of research.

2.2 Classification system

Statistics on R&D in the public sector are published according to a number of various groupings and classifications depending on the type of R&D institution and the R&D personel.

Sector

Eurostat:

- Higher Education (HE)
- Government (GOV)
- Private non-profit(PNP)

Denmark:

- HE: Higher Education (Split in to University Hospitals and Higher Education in published tables)
- GOV: Other government research institutions
- Sector research institutions
- PNP: Private non-profit organisations

In a few supplementary tables there are specific information for each individual university.

Personnel categories

The national category:

- Videnskabeligt personale (VIP), cover the international category

Researchers, which include the subcategories (which are only available in Danish terms):

Universities:

- Professorer
- Lektorer/seniorforskere
- Seniorrådgivere
- Adjunkt/Post doc. Forsker
- Ph.d.- og kandidatstipendiat
- Øvrige forskere

University hospitals:

- Professorer
- Overlæger
- Afdelingslæger
- 1. reservelæger/reservelæger
- Post Docs
- Ph.d.- og kandidatstipendiat
- Øvrige forskere med kandidatgrader
- Kliniske assistenter/forskningsstipendiater
- Forskere med mellemlange videregående uddannelser.

Other government sector:

- Forskere med Ph.D. grad
- Forskere med kandidatgrad
- Øvrige forskere

The national category:

- Teknisk administrativt personale(TAP), cover the international categories
- "Technicians" and "Other Support Staff".

Technicians include the subcategories (which are only available in Danish terms):

Universities:

- Teknisk Personale

University hospitals:

- Teknisk/klinisk personale

Other government sector:

- Teknisk personale

Other Support Staff include the subcategories (which are only available in Danish terms):

Universities:

- Administrativt personale
- Scholarships

University hospitals:

- Administrativt personale
- Scholarships

Other government sector:

- Administrativt personale
- Scholarships

2.3 Sector coverage

Statistical units are according to the Frascati Manual classified in 4 sectors:

- Business Enterprise Sector
- Higher Education Sector
- Government Sector
- Private non-profit Sector

The statistics on R&D in the public sector covers the last 3 sector mentioned.

Because of their relation to the universities the university hospitals are classified under the Higher Education sector, which hence cover most of the R&D activities in Danish hospitals. The government sector includes smaller hospitals, museums and archives and centers and sectoral research institutions not covered by universities. Private non-profit organisations are covered since they do not operate on the market. The main distinction regarding market and non-market activities is based on National Accounts definitions.

2.4 Statistical concepts and definitions

Smallest Homogeneous Unit: The statistical unit is defined in the Frascati manual as being the smallest homogeneous unit predominantly involved in only one of the six fields of science and for which a complete (or almost complete) set of factor input data can be obtained. In the Danish terminology this means the individual institute in universities, a centre and the individual department in university hospitals.

Research and Development (R&D): Research and development (R&D) comprise creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications.

Type of Research and Development: Three types of R&D may be distinguished:

- **Basis research:** experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view.
- **Applied research:** original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific practical aim or objective.
- **Experimental development:** systematic work, drawing on existing knowledge gained from research and/or practical experience, which is directed to producing new materials, products or devices, to installing new processes, systems and services, or to improving substantially those already produced or installed.

Field of Science: The grouping to be used to divide the activities in the HES sector (Higher Education) and in the PNP sector (Private Non Profit) is layed down in the Frascati Manual. The

described Fields of Science are used in Danish statistics for the remaining part of the government sector as well.

The 6 Fields of Science are:

- Natural Sciences
- Engineering and Technology
- Medical Sciences
- Agricultural Sciences
- Social Sciences
- Humanities

R&D expenditures: According to the Frascati Manual expenditures are all expenditures for R&D performed within a statistical unit or sector of the economy during a specific period, whatever the source of funds.

The expenditures cover not only the direct costs for e.g. wages to researchers but also parts of the indirect costs such as IT-support, electricity and heating and various administrative services. In cases of problems with measuring precisely the indirect costs an estimate based on the relative R&D activity of the statistical unit may be applied. In that case a measure based on R&D personnel (full time equivalents) is recommended by Statistics Denmark.

Sectors: Statistical units are according to the Frascati Manual classified in 4 sectors:

- Business Enterprise Sector
- Higher Education Sector
- Government Sector
- Private non-profit Sector

The statistics on R&D in the public sector covers the last 3 sectors mentioned.

Because of their relation to the universities the university hospitals are classified under the Higher Education sector, which hence covers most of the R&D activities in Danish hospitals. The government sector includes smaller hospitals, museums and archives and centers and sectoral research institutions not covered by universities. Private non-profit organisations are covered since they do not operate on the market. The main distinction regarding market and non-market activities is based on National Accounts definitions.

2.5 Statistical unit

The statistical unit is defined in the Frascati manual as being the smallest homogeneous unit predominantly involved in only one of the six fields of science and for which a complete (or almost complete) set of factor input data can be obtained. In the Danish terminology this means the individual institute in universities, a centre and the individual department in university hospitals.

2.6 Statistical population

The population to be measured is all non-market oriented government units and private non-profit units which are having R&D activities at a level of at least one person measured in full time equivalents. The size of the population is around 700.

2.7 Reference area

Denmark.

2.8 Time coverage

2007-2016

2.9 Base period

Not relevant for these statistics.

2.10 Unit of measure

All expenditures are measured in 1.000 DKK and personel in numbers and full time equivalents.

2.11 Reference period

The statistics are compiled annually.

2.12 Frequency of dissemination

The statistics are published annually.

2.13 Legal acts and other agreements

The data are collected in accordance with section 8 of the Act on Statistics Denmark (Consolidated act No 599 , June 22, 2000). Participation in the survey is voluntary up to and including the reference year 2006.

Data are collected in accordance with Decision No 1608/2003/EC of the European Parliament and of the Council of 22 July 2003 concerning the production and development of Community statistics on science and technology And Commission regulation (EC) No 995/2012 implementing Decision No 1608/2003 of the European Parliament and of the Council as regards statistics on science and technology.

2.14 Cost and burden

The response burden on public institutions is not calculated.

2.15 Comment

www.dst.dk/fui.

3 Statistical processing

The Statistics are compiled annually on the basis of questionnaires. The data collected are validated very carefully focusing on a number of prioritized variables, notably the R&D expenses. This is performed at macro- as well as micro level.

3.1 Source data

The Statistics are compiled on the basis of questionnaires. There are separate questionnaires for university hospitals and for other institutions.

3.2 Frequency of data collection

Annual.

3.3 Data collection

Web-questionnaire. A paper version is available on request.

3.4 Data validation

A number of variables has been chosen for a intense data validation based on the experience regarding the impact in the total results.

The variables prioritized for the detailed data validation are:

- R&D labour costs
- Other R&D current costs
- R&D expenditures in total
- R&D personel in full time equivalents
- Average labour costs per person (full time equivalents). This variable is calculated using R&D labour costs and R&D personel in full time equivalents.

The variables are numerical and errors in the reported data may have considerable effects in the totals. Hence, they are given special attention in the data validation process.

The remaining variables are also validated and in those cases good estimates for the corrections are pro-rata calculations in case of relative distributions or reportings from the preceding year.

In *the micro validation* the individual reporting is validated in several levels:

- at the level of variables
- at the level of crossing variables
- at the level of time (years)
- comparison with other data sources

Each variable is validated against its definitions. This is followed by a cross checking with combination of relevant variables. If available, information from historic reporting is introduced in the validation of consistency.

Basically there is a distinction in two types of errors. Logical errors (where there is a contradiction in the data for comparable questions) and potential errors (where there is a probable but not

necessarily real error).

Macro validation is based on a comparison of each reporting with the aggregates. The validation is performed using the software Banff which is developed by Statistics Canada.

Various forms of *imputation* is used:

- Imputation of missing values
- Pro rata imputation
- Imputation for missing reporting

Imputation for missing values is applied for more questions and primarily those representing relative distributions of R&D activities by Fields of Science, purpose and Strategic area. In general, values are imputed by use of information from the preceding year. In cases where reliable data on labour costs is not available an estimate is imputed based on the number of R&D personel.

Pro-rata imputation is applied primarily for the mentioned questions representing relative distributions, in cases where the sums are not equal to 100. Furthermore, pro-rata imputation is used if R&D expenditures covered by external sources, exceeds the total expenditures.

Imputation for missing reporting, which are cases where the full questionnaire is filled in automatically due to a non-response even after several re-contacts, is very seldomly occurring due to the very high response rate.

Statistics Denmark have established at University and University hospitals agreements with local contacts who are able to provide further information at the very detailed level if necessary. The contacts also collect the data for their institution which ensure an efficient contact to the relevant statistical units which may otherwise cause problems. The procedure enables a streamlining of the data collection process for institutions with multiple statistical units and lay ground for a high degree of continuity. The population is constructed based on the information from the preceding year. New units are basically identified by information on structural changes reported by the contacts in the universities and the university hospitals. For the remaining part of the population similar kind of information is used and also a surveying of potential new relevant units is performed.

3.5 Data compilation

The main aim of the data collection is to produce statistics on Research and development activities in the Danish public sector. Data has to be:

- True and covering the R&D activities in the public sector and private non-profit (PNP) sector as a whole.
- Suitable for statistics at more detailed level as well.
- Burden the respondents as less as possible

The statistics is based on a census of units in the public sector and PNP's performing R&D activities. The reporting of data is mandatory according the the Act on Statistics Denmark paragraphs 6 and 8.

The aim for Statistics Denmark is that reporting of data is digital. This is provided via the web portal <http://www.Virk.dk>. Beside this the reporting of most data for universities and university hospitals is also in electronic format, notably spread sheets. Finally the questionnaire is available in a paper version on request..

The aim of the further data treatment is to treat the data collected and validate for errors and missing data to bring it to a quality adequate to give a true picture of the R&D activities, also over time.

Due to the large questionnaire (600 variables) the process of data management is comprehensive. As the number of respondents exceeds 700 units it then gives a first set of results of more than 420.000 cells and hence a considerable number of potential errors or misunderstandings in the reported information.

This implies that:

- Data validation as well as data correction as much as possible is carried out automatically. Since many of the questions are interrelated the correction need to be performed in a planned and systematic manner.
- It is necessary to take in to consideration that some questions have greater impact on the overall picture than others, and then perform a prioritized validation. Tools for the validation is based on the statistical software SAS.

Many of the questions are dealing with the same issues. It gives advantages in the sense that it provides the basis for internal verification in the questionnaire itself. At the same time, it is a source for identifying errors across questions with contradictory information.

3.6 Adjustment

No further adjustments are done.

4 Relevance

There is a general political objective of increasing the share of GDP, which is applied for research and development, and in the light of this objective, there is great interest in compiling statistics on research and development. The main users are Ministries, public authorities, business organisations, researchers, private business enterprises and the media. In addition, the European Commission (The statistical office Eurostat) and the OECD are important users. Data are available for purposes of research.

4.1 User Needs

There is a general political objective of increasing the share of GDP, which is applied for research and development, and in the light of this objective, there is great interest in compiling statistics on research and development. The main users are Ministries, public authorities, business organisations, researchers, private business enterprises and the media. In addition, the European Commission (The statistical office Eurostat) and the OECD are important users.

Data are available for purposes of research.

4.2 User Satisfaction

The main users of the statistics are represented in a Users Committee that meets regularly. No specific survey regarding user satisfaction has been carried out.

4.3 Data completeness rate

There are no data cells missing in transmissions to Eurostat.

5 Accuracy and reliability

The reliability is high. The statistics is based on a census of 700 public institutions and the response rate is 99 percent. The collected data is going through an intense validation process and except for a single question the imputation rate for missing values is relatively moderate. Imputation could e.g. be the use of information reported by other institutions to replace missing information.

5.1 Overall accuracy

The statistics is based on a census of 700 public institutions performing research and development and the response rate is 99 percent. Despite the large number of variables the measurement error is considered to be small which is based on the very comprehensive data validation process.

5.2 Sampling error

The statistics is not based on a sample.

5.3 Non-sampling error

The non-sampling error is low which is primarily due to the very high coverage. Furthermore, there is no over-coverage. Despite the large number of variables the measurement error is considered to be small which is based on the very comprehensive data validation process.

5.4 Quality management

Statistics Denmark follows the recommendations on organisation and management of quality given in the Code of Practice for European Statistics (CoP) and the implementation guidelines given in the Quality Assurance Framework of the European Statistical System (QAF). A Working Group on Quality and a central quality assurance function have been established to continuously carry through control of products and processes.

5.5 Quality assurance

Statistics Denmark follows the principles in the Code of Practice for European Statistics (CoP) and uses the Quality Assurance Framework of the European Statistical System (QAF) for the implementation of the principles. This involves continuous decentralized and central control of products and processes based on documentation following international standards. The central quality assurance function reports to the Working Group on Quality. Reports include suggestions for improvement that are assessed, decided and subsequently implemented.

5.6 Quality assessment

The overall quality of the statistics is considered to be high. This is based on the fact that it is a census with a very high response rate. Non-response is only 1 percent. The collection of the data is for the dominant respondents Universities and University hospitals organized through close collaboration with central contact persons. The collected data is going through an intense validation process and only for a single question (R&D distributed by strategic topics) the imputation rate is considerable, around 15 percent.

5.7 Data revision - policy

Statistics Denmark revises published figures in accordance with the [Revision Policy for Statistics Denmark](#). The common procedures and principles of the Revision Policy are for some statistics supplemented by a specific revision practice.

5.8 Data revision practice

Data for the most recent year are published as preliminary figures. The reference years 2007-2015 are published as final statistics. At the publication of the 2017 statistics the 2016 statistics will be published as final data.

6 Timeliness and punctuality

National release is done 16 months after the end of the reference period. It is the aim to publish preliminary figures 12 months after the end of the reference period. Final figures are published 12 months after the preliminary figures.

Transmission of data to Eurostat follows rules laid down in the relevant legal act.

The release of the statistics is punctual.

6.1 Timeliness and time lag - final results

National release is done within 12 months after the end of the reference period, although for 2016 there has been a slight delay. Final figures are published 12 months after the preliminary figures.

Transmission of data to Eurostat follows rules specified in the relevant legal act.

6.2 Punctuality

There is no difference in the date of foreseen and actual publication.

7 Comparability

The statistics are comparable from the years 1997 to 2006 and 2007 onwards.

7.1 Comparability - geographical

Eurostat and OECD publish a number of manuals providing guidelines for the collection and processing of data and the definition of concepts. The EU legal acts refers to the manuals as guidelines for the statistics.

The manuals are:

- The Frascati Manual, dealing with R&D statistics
- The Oslo Manual, dealing with innovation statistics
- The Canberra Manual, dealing with measuring human resources within Science and Technology

The EU requirements are not defining how the information is provided and therefore methodological differences among countries may exist which affect comparability.

In comparison with statistics from other countries one should be aware that structural differences may have considerable impact. In the case of Denmark the university hospitals are included in the Higher Education Sector (HES), which means that this sector is relatively large compared to the remaining Government Sector (GOV). Other examples exist regarding differences in the content of HES and GOV.

7.2 Comparability over time

The statistics are comparable from the years 1997 to 2006 and from 2007 onwards.

7.3 Coherence - cross domain

Besides the statistics concerning the public sector there are also statistics regarding the business sector and in National Accounts.

7.4 Coherence - internal

Internal consistence exists.

8 Accessibility and clarity

The statistics are published in *Nyt fra Danmarks Statistik* (News from Statistics Denmark), are available from Statistics Denmark's website at: <http://www.dst.dk/fui> and from the database StatBank Denmark.

8.1 Release calendar

The publication date appears in the release calendar. The date is confirmed in the weeks before.

8.2 Release calendar access

The Release Calendar can be accessed on our English website: [Release Calendar](#).

8.3 User access

Statistics are always published at 8:00 a.m. at the day announced in the release calendar. No one outside of Statistics Denmark can access the statistics before they are published.

8.4 News release

News from Statistics Denmark available at <http://www.dst.dk>

8.5 Publications

Results of the statistics are published in an annual publication regarding Innovation and Research.

8.6 On-line database

The results from the survey is published in a newsletter from Statistics Denmark (in Danish only).

8.7 Micro-data access

Data are available for researchers in a specific system in Statistics Denmark.

8.8 Other

The results are available also on Eurostat and OECD websites.

8.9 Confidentiality - policy

In principle information from public authorities is not confidential. In the statistics published only information for the universities are on the individual level.

8.10 Confidentiality - data treatment

In principle information from public authorities is not confidential. In the statistics published only information for the universities are on the individual level.

8.11 Documentation on methodology

<https://www.dst.dk/fui>. Documents only available in Danish.

8.12 Quality documentation

Results from the quality evaluation of products and selected processes are available in detail for each statistics and in summary reports for the Working Group on Quality.

9 Contact

The administrative placement of this statistic is in the division of Business Development. The person responsible is Casper Larsen, tel. +45 3917 3555, e-mail: cla@dst.dk

9.1 Contact organisation

Statistics Denmark

9.2 Contact organisation unit

Business Development, Business Statistics

9.3 Contact name

Casper Larsen

9.4 Contact person function

Responsible for the statistics

9.5 Contact mail address

Sejrøgade 11, 2100 Copenhagen

9.6 Contact email address

cla@dst.dk

9.7 Contact phone number

+45 3917 3555

9.8 Contact fax number

+45 39 17 39 99