Office for National Statistics

Identifying Science and Technology Businesses in Official Statistics

Author Name(s): **James P Harris, Office for National Statistics**

Abstract

Policy makers are interested in identifying Science and Technology (S&T) businesses in official data sources to better understand the contribution that they make to the economy. This article proposes a classification for S&T and for sub-categories such as Life Sciences and Digital Technology, based on the UK Standard Industrial Classification of Economic Activities (SIC) 2007. It includes practical examples and explanations of how to use the approach with different data sources.

1. Introduction

1.1 Science and Technology (S&T) is a key area of interest for policymakers. The UK Parliament has two select committees dealing with it, in the House of Commons and in the House of Lords. Within government, the Department for Business, Innovation & Skills (BIS) leads on S&T, the Council for Science and Technology advises on cross-cutting issues, and the subject is also relevant to other government departments. Local government and the Local Enterprise Partnerships (LEPs) also have an interest in S&T, supporting public-private initiatives such as London Technology Week (held for the first time in June 2014). All of these stakeholders are interested in obtaining evidence, including data, to inform their policy and planning decisions.

1.2 Official statistics use a number of standard classifications which are incorporated into official data sources, allowing evidence from different sources to be compared on the same basis. One of these is the UK Standard Industrial Classification of Economic Activities (UK SIC), which classifies economic activities using a five-level hierarchical framework. Another is the Standard Occupational Classification (SOC), which classifies occupations using a hierarchical structure with four nested tiers. The SIC and the SOC are used in business statistics, household surveys, the Census and the National Accounts.

1.3 However, there is currently no standard classification for S&T or for the sub-categories of interest within S&T such as Life Sciences and Digital Technology. Therefore varying definitions are used by analysts, which makes it difficult to compare sources and assess the evidence.

1.4 This article presents a method for identifying S&T businesses and some S&T sub-categories, as well as those businesses whose activities are not scientific or technical (‘not S&T’). The method has been developed by the Office for National Statistics (ONS) at the request of the Greater London

Authority (GLA). It uses the lowest levels of the UK SIC hierarchy (the 5-digit Sub-classes). The article makes available this S&T classification for other analysts who work with official statistics. We recommend its adoption as a common tool for S&T analyses because this would help comparability when assessing the evidence. However, it is not one of the official classifications of the ONS.

1.5 The method presented here has advantages and disadvantages, as explained in this article. The two main advantages are: first that it uses the UK SIC, which is present in most official data sources; and second that it is relatively simple to construct. The main disadvantage is that it is one-dimensional: it uses disaggregated economic activity (the 5-digit Sub-classes of the SIC) but it does not take into account people's occupations or other characteristics such as educational qualifications. The reason for simplifying in this manner is explained below.

1.6 The remainder of this article defines the proposed S&T classification and its five sub-categories. It discusses how the methodology was developed, building on existing definitions and classifications, and it contains a 'how to' guide for analysts who wish to use the approach. It also considers the main limitations of the approach.

2. Defining S&T and its sub-categories

2.1 The proposed classification has been designed to create a broad definition of S&T which contains five sub-categories, namely:

2.1.1 **Digital Technologies**: the manufacture and repair of computers and electronic components; computer services including software development; internet services; and computer consultancy

2.1.2 **Life Sciences and Healthcare**[1]: medical healthcare services (both human and veterinary); medical research and development (including biotechnologies); and manufacture of pharmaceuticals and medical treatment machinery

2.1.3 **Publishing and Broadcasting**: publishing and telecommunications; specialist graphic design and marketing services; the manufacture and repair of communication equipment; and the use of this equipment by means of broadcasting

2.1.4 **Other scientific/technological manufacture**: precision engineering and the manufacture and repair of equipment for aerospace, defence, automotive, chemical products, engines and machinery (both electrical and non-electrical)

2.1.5 **Other scientific/technological services**: knowledge-intensive services including higher education, engineering, architecture, quantity surveying, aerospace transport services, and non-medical research and development

2.2 Where enough data is available, the five sub-categories can be broken down into more detailed topics. An illustration of the full hierarchy of sub-categories and topics can be found in **Appendix 1: Constituent parts of the proposed S&T classification**.

2.3 The proposed classification was originally developed for London at the request of the GLA. In London, when using business statistics sources such as the Inter-Departmental Business Register (see below: **How to use the proposed S&T classification**) it is possible to look at S&T by small area geographies such as Middle-layer Super Output Areas (MSOAs). It may not be possible to analyse the sub-categories at MSOA level, but it should possible at higher geographies such as local and unitary authorities. Information on more detailed topics within the sub-categories can be provided across London as a whole.

Notes

1. Healthcare has been included in the same category as Life Sciences for two reasons: 1) healthcare businesses are most likely to be the primary users of Life Science outputs, and 2) 'Life Sciences' is not large enough to be analysed on its own. The two individual sub-categories (Life Sciences and Healthcare) can still be created, but this will only be possible where there are sufficient numbers of workplaces.

3. Overview of the chosen method

3.1 The proposed S&T classification is based on the UK Standard Industrial Classification of Economic Activities 2007 (SIC07), which is the most recent version of the SIC. The SIC07 is designed as a five-level hierarchy formed of Sections (at the highest level), Divisions, Groups, Classes and Sub-classes (at the lowest level). Sub-classes fit within Classes which fit within Groups, which in turn fit into Divisions, which fit into Sections.

3.2 The main reason that the SIC07 was chosen as the basis for our proposed S&T classification is that the SIC07 variable already exists in most official data sources and, at the lowest level of the hierarchy, it can be used to identify S&T businesses. Another advantage of using the SIC07 is that it is the national version of the statistical classification of economic activities in the European Community (known as NACE, from the French acronym), with some additional Sub-classes specific to the UK economy. The latest NACE classification is revision 2.

4. Comparison with other definitions

4.1 There are several other definitions of S&T in use. In order to develop a consistent approach based on a common understanding of the meaning of S&T, this project began by comparing the existing definitions. Specifically, the following approaches based on the SIC were considered:

4.1.1 the Organisation for Economic Co-operation and Development (OECD, via Eurostat)

4.1.2 various definitions provided by members of the Greater London Authority (GLA)

4.1.3 the UK Innovation Survey

4.1.4 the Office for Life Sciences

4.1.5 the London Science and Technology Evidence Base

4.1.6 the Southern Cluster Policy Paper

4.2 Although the proposed classification uses the SIC07, some of the approaches listed above were based on earlier versions of the SIC such as the SIC03, or on the NACE. These definitions have been converted onto a SIC07 basis as closely as possible for analytical purposes.

4.3 Some definitions are based on the SOC (occupational classification) rather than the SIC. For instance, Business, Innovation & Skills (BIS) STEM[1] analyses use a SOC-based approach. Recent reports such as Tech Nation 2015 and The Geography of the UK's Creative and High-Tech Economies analyse data mainly on a SOC or STEM basis. The issues involved in choosing between the SIC and the SOC are discussed below (**Main limitations of the S&T classification**).

4.4 The workbook associated with this document contains the full SIC07 structure, the proposed S&T classification (by sub-category and detailed topic) and the economic activities that are considered 'not S&T'. In addition, it compares the proposed S&T classification with some of the existing definitions.

Notes

1. STEM stands for Science, Technology, Engineering and Maths. The STEM approach was originally developed for analyses relating to educational qualifications, curriculum choices and skills.

5. Detailed methodology

5.1 The proposed S&T classification is based on assigning each SIC07 5-digit code to a S&T category, sub-category and topic using the description of that code, the skills and equipment relating to that code and how other approaches have classified it.

5.2 For each 5-digit code, the following decision process has been followed:

5.2.1 How is this code classified according to the OECD?

If the code is classified in one of the relevant S&T categories of the OECD definition (high technology manufacturing, medium-high technology manufacturing, and scientific parts of knowledge intensive services) then it should be included in the proposed S&T classification. However, the OECD definition was built on NACE revision 1.1 and S&T categories were selected at a higher level (in terms of the SIC hierarchy) than that used by the proposed S&T classification. This broad-brush approach meant that the OECD definition included some economic activities at lower levels of the SIC hierarchy that were not scientific or technological. In such cases, it was decided not to follow the OECD approach.

5.2.2 How is this code classified according to others?

The classifications of a variety of other approaches were collated and considered (see above: **Comparison with other definitions**). If sufficient numbers of other approaches agreed that a 5-

digit code was part of S&T (even if the OECD approach did not include it), then it was classified accordingly.

5.2.3 What is the definition of this code? What activities does it contain, what is its output, what skills are required during output, and what technology/research is required to do it?

Every SIC07 code has explanatory notes describing the type of activity performed. This is available on the Neighbourhood Statistics website. This information was used to make a judgement about the position of the code in the proposed S&T classification. An attempt was also made to understand the employee skills and equipment required for the activity. If an industry involved sufficient levels of skill, technological equipment or processes, or scientific understanding and development, then it was included. In cases where the nature of the activities, output or required skills was in doubt, additional research was carried out in order to make a decision.

5.2.4 Are the activities of this code themselves scientific or technical, or do they simply employ scientific or technical equipment during non-scientific/non-technical processes?

Some industries, such as manufacture of electronic computer components, produce equipment that is clearly technological and require specialist skills and materials during production. Other industries simply make use of technology without requiring specialist knowledge or equipment, such as use of computerised cash register systems by a food retailer. During the decision-making process, the former were classified as S&T whereas the latter were not.

5.2.5 Do the other codes in this section fit into a specific sub-category, and if so should this code fit into the same sub-category?

Where possible, Sub-classes belonging to the same Class, Classes belonging to the same Group and Groups belonging to the same Division within the SIC07 hierarchy were kept together. Classifying at higher levels makes it easier to extract data, helps comparability with other definitions that only use higher levels. For example, if one Division contained two Groups which each contained five Classes, it would be better if the entire Division were assigned to a category, rather than assigning each of the ten Classes separately. However, in some cases it was clear that while one Class was scientific or technological, its companion Class was not; or that one Class fitted into a particular sub-category of the S&T classification while its companion fitted into a different sub-category. Thus, components of SIC Groups and Divisions were kept together where possible, but this rule was used flexibly.

Classifying at higher levels within the SIC hierarchy also makes the proposed S&T classification less vulnerable to breaches of confidentiality. It is easier to identify individual people or businesses from low-level breakdowns than when data is aggregated to a higher level. Disclosure control procedures need to be applied in all cases (see below: **How to use the proposed S&T classification**), but these are more challenging when the data is disaggregated to low levels.

5.3 Once every SIC07 code had been assigned to a category, the proposed categories were peer reviewed and compared again with the existing classifications. Decisions were made regarding suggested re-categorisation and the proposed S&T classification was finalised.

6. How to use the proposed S&T classification

6.1 The workbook associated with this article contains the list of all SIC07 5-digit codes and the S&T category, sub-category and topic that they fit into. When analysing microdata, analysts can use this to categorise the SIC07 codes in their data file and output results on the S&T basis. Further guidance on how to use this list is contained in the 'Notes' sheet of the workbook associated with this article.

6.2 This approach only works on data collected on a SIC07 basis. If the analyst is dealing with earlier versions of the SIC, they may be able to convert the data onto a SIC07 basis. Guidance on how to achieve this is available at: www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/standard-industrial-classification/index.html

6.3 It is not possible to apply the classification to the higher levels of the SIC07 hierarchy, eg Divisions, as these comprise numerous individual Classes and Sub-classes which may be split between S&T categories. It is, therefore, necessary for analysts to use the lowest-level (5-digit) SIC07 codes.

6.4 Following is a non-exhaustive list of official microdata sources with which the proposed S&T classification can be used, because they contain a SIC07 variable from which the classification can be constructed:

6.4.1 Inter-Departmental Business Register (IDBR)

6.4.2 Business Register and Employment Survey (BRES)

6.4.3 Annual Population Survey (APS)

6.4.4 Labour Force Survey (LFS)

6.4.5 The 2011 Census

6.4.6 Annual Survey of Hours and Earnings (ASHE)

6.4.7 UK Innovation Survey (UKIS)

6.4.8 Companies House business data

6.5 When using these microdata sources, the confidentiality of individual respondents must be protected. 'Disclosure control' rules must be applied to output tables according to the rules set out for the data source. These include the threshold rule (where values must be built upon sufficient numbers of records) and the p% rule (where the largest values in a cell must not dominate the value of the cell). The analyst must also check that it is not possible to deduce confidential information from any part of the output tables or from comparison with previously published tables. This is particularly important when breaking down statistics to low geographies or small analytical

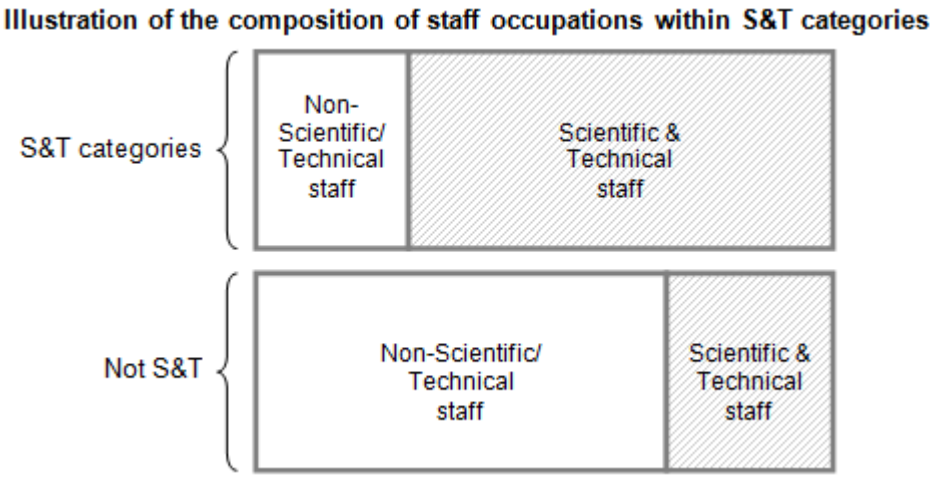
groupings. For further information about these rules and how to apply them, please contact the ONS Statistical Disclosure Control team.

7. Main limitations of the S&T classification

7.1 Businesses generally employ staff with a variety of skills including specialists, such as engineers and doctors, and non-specialists, such as administrative staff and facilities management. Ideally a classification of S&T would take this into account, such that it included only the specialist staff in S&T industries.

7.2 However, a classification based on both of these dimensions would be technically challenging and restricted in terms of the microdata sources with which it could be used. In practice it makes sense to choose between the SIC and the SOC. The proposed S&T classification was developed mainly to analyse official business statistics. Most business statistics use the SIC07 but contain little or no information about occupations or job descriptions. Therefore, a decision was made to classify businesses using their industrial classification. This means some staff in businesses classed as S&T in the proposed classification will neither have S&T skills nor work on S&T subjects. The reverse is also true: some staff working in non-S&T businesses will have S&T skills or will be employed in S&T activities (see Illustration).

Illustration of the composition of staff occupations within S&T categories



7.3 Nevertheless, for data sources which do include the SOC as well as the SIC, such as the 2011 Census, the LFS, the APS and the ASHE, it should be possible (sample size permitting) for analysts to cross-tabulate the proposed S&T classification with information about occupations. In the 2011 Census, the LFS and the APS, it may also be possible to cross-tabulate the proposed S&T classification with information on educational qualifications. Similarly, the BIS's STEM approach, which is based on occupation, can be cross-tabulated with information about industries if the data source contains the SIC as well as the SOC.

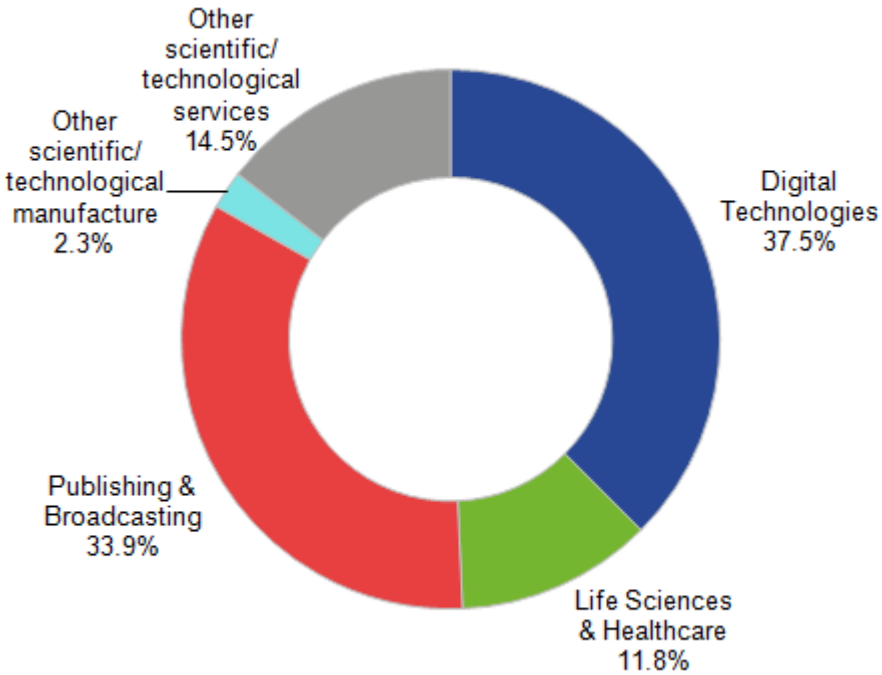
8. S&T in London and the rest of the UK

8.1 The proposed S&T classification was initially developed for London's economy at the request of the GLA. This section shows some results for London and compares them with the picture for the UK as a whole. Similar analyses can be produced for other parts of the UK.

8.2 Applying the proposed S&T classification to the IDBR shows that there were 102,105 workplaces in London classified as S&T in 2014. This was 22% of all workplaces in London and accounted for 21% of employee jobs in the capital.

8.3 Figures 1 and 2 show the proportion of workplaces and of employee jobs in each of the five categories within S&T in London in 2014. The largest proportion of S&T workplaces was in Digital Technologies (37%), closely followed by Publishing & Broadcasting at 34%. However, Digital Technologies only accounted for 18% of S&T employee jobs, while the proportions of employee jobs in Life Sciences & Healthcare and in Other scientific/technological services were higher than their respective shares of workplaces. This indicates that there are differences in the average number of employees per workplace in these sub-categories. In Digital Technologies there were some 4 employees per workplace in 2014 and in Publishing & Broadcasting and Other scientific/ technological manufacture the average was 8 employees per workplace. By contrast, in Other scientific/technological services and in Life Sciences & Healthcare there were on average 14 and 22 employees per workplace respectively.

Figure 1: Breakdown of Science and Technology workplaces in London, 2014



Source: Inter Departmental Business Register (IDBR) - Office for National Statistics

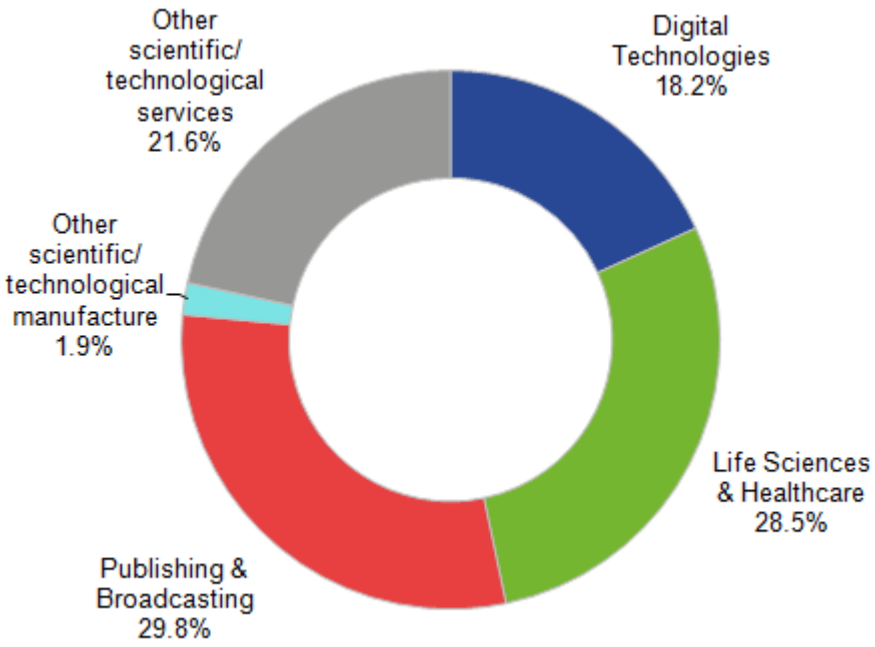
Notes:

1. Workplaces with zero employment and enterprises with zero employment have been excluded from this analysis.
2. London identified using administrative geographic boundaries in the National Statistics Postcode Lookup (NSPL) February 2014.

Download chart

XLS XLS format
(198 Kb)

Figure 2: Breakdown of Science and Technology employees in London, 2014



Source: Inter Departmental Business Register (IDBR) - Office for National Statistics

Notes:

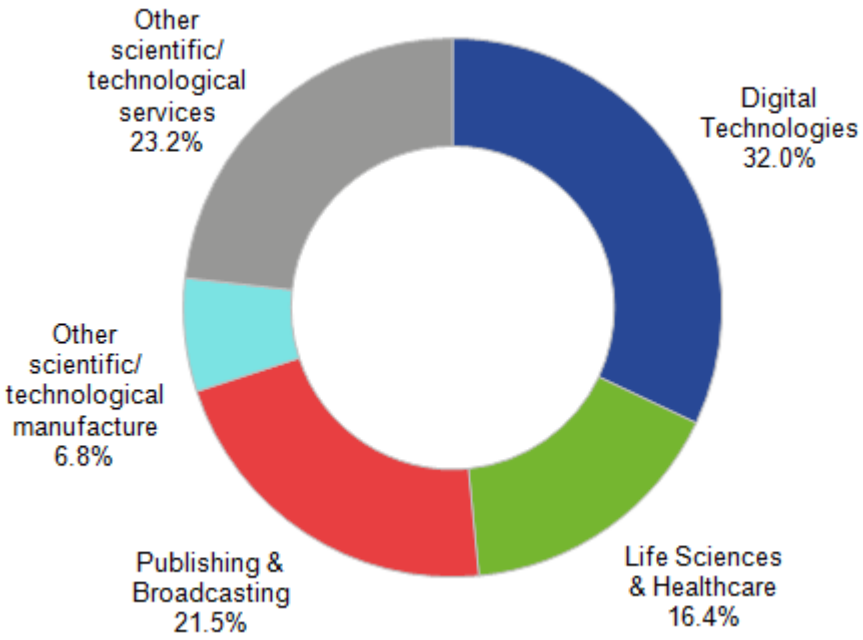
1. Workplaces with zero employment and enterprises with zero employment have been excluded from this analysis.
2. London identified using administrative geographic boundaries in the National Statistics Postcode Lookup (NSPL) February 2014.

Download chart

XLS XLS format
(198 Kb)

8.4 In 2014, London contained 22% of the UK's S&T workplaces and employed 17% of the UK's S&T employees. Figures 3 and 4 show that Life Sciences & Healthcare and Other scientific/ technological manufacture are more important in other parts of the UK than in London. Digital Technologies have a slightly higher weight in London than in the UK as a whole, while Publishing & Broadcasting is much more important in London than in other parts of the country. This analysis only shows numbers of workplaces and employee jobs; it takes no account of the relative performance of businesses across the country.

Figure 3: Breakdown of Science and Technology workplaces in the UK, 2014



Source: Inter Departmental Business Register (IDBR) - Office for National Statistics

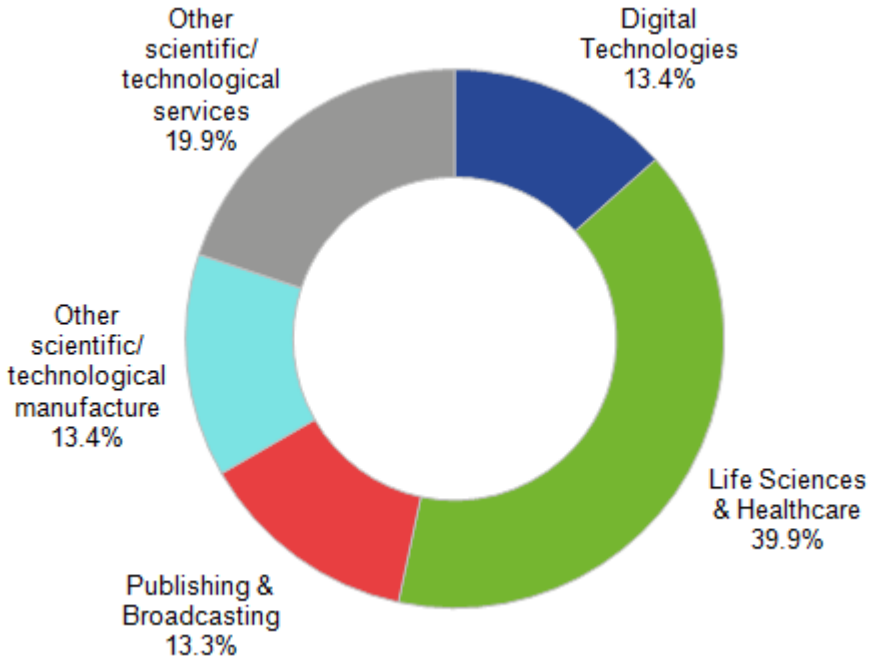
Notes:

1. Workplaces with zero employment and enterprises with zero employment have been excluded from this analysis.

Download chart

XLS XLS format
(198 Kb)

Figure 4: Breakdown of Science and Technology employees in the UK, 2014



Source: Inter Departmental Business Register (IDBR) - Office for National Statistics

Notes:

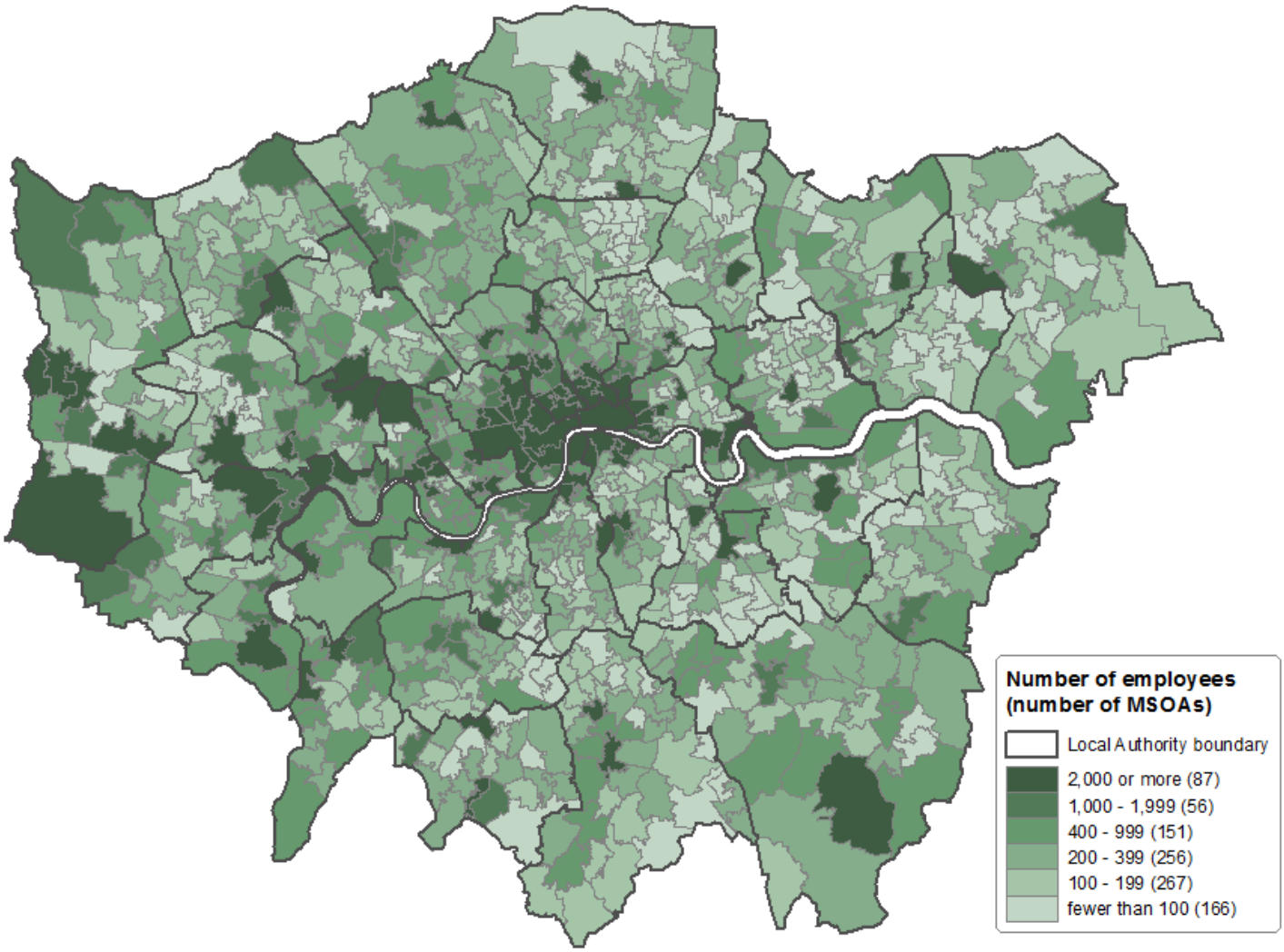
1. Workplaces with zero employment and enterprises with zero employment have been excluded from this analysis.

Download chart

XLS XLS format
(198 Kb)

8.5 Figure 5 shows where S&T employee jobs are located in London. There are high concentrations in central London around Westminster and The City, and also along the Westway, the M4 motorway and around Heathrow. There are also small clusters of S&T employment spread across the capital, including in town centres such as Richmond, Croydon and Romford.

Figure 5: Number of employees working in Science and Technology in London, 2014



Source: Inter Departmental Business Register (IDBR) - Office for National Statistics

Notes:

1. MSOA refers to Middle-layer Super Output Area boundaries

2. Contains National Statistics data © Crown copyright and database right 2015
3. Contains Ordnance Survey data © Crown copyright and database right 2015

Download map

PNG **PNG format**
(378.6 Kb)

Background notes

1. Details of the policy governing the release of new data are available by visiting www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html or from the Media Relations Office email: media.relations@ons.gsi.gov.uk

Copyright

This document is also available on our website at www.ons.gov.uk.

Appendix 1: Constituent parts of the proposed S&T classification

The diagram shows the five sub-categories that make up S&T and the constituent parts of those categories (the topics). The grey boxes provide a further description of the wider term directly above. Note you can click on the diagram to view a larger version of it.

Constituent parts of the proposed S&T classification