Draft proposal to develop a coordinated stock of classification decisions for research institutes and related statistical units

Draft proposal for discussion prepared by Luis Sanz-Menéndez (CSIC, Ministry of Science, Innovation and Universities, Spain, and delegate to CSTP and GSF; luis.sanz@csic.es) and Cecilia Cabello (FECYT, Delegate to NESTI; cecilia.cabello@fecyt.es) with support from Fernando Galindo-Rueda (OECD; fernando.galindo-rueda@oecd.org).

Background and rationale for a study

After considering a range of potential changes, the 2015 Frascati Manual (FM) maintained the institutional classification used in previous editions. Responding to policy interest, the FM classification combines the SNA institutional sector classification (which is based on attributes dealing with the nature of economic production –i.e. whether market based or non-for profit- and the existence of government control – i.e. whether public or private-) with the ad hoc identification of the higher education (HE) sector (units with HE as main economic activity). While this idiosyncratic approach was maintained, it was also decided to refrain from specifying other *ad hoc* sectors -on top of the existing ones- based on other economic activities, such as health or research. Several countries use such ad hoc classifications for their own policy and statistical purposes, for example, implementing specific regimes for specialized research organisations or using different data collection procedures for non-business units providing health services as main economic activity, etc.

In recognition of these multiple needs, the 2015 Frascati Manual edition (chapter 3) introduced a number of improvements and conceptual clarifications in its decision tree – the sequence of criteria to be used in the allocation of units to sectors - relating those criteria more explicitly to classification procedures used in the System of National Accounts. In addition, the manual included some analysis of borderline cases in methodological chapters dedicated to each sector.

The Frascati Manual also encourages the active management of registers of statistical units that capture key attributes relevant for classification, as shown in the Table below. This helps serve multiple purposes, such as organising data collection, undertaking data linking efforts through common identifiers, as well as providing custom tabulations that address different questions.

Table 1. Proposals in FM2017 for tagging units in R&D registers

Table 3.2. **A simplified example of a potential frame structure tagging statistical units on various dimensions**

	Frascati institutional sector	SNA institutional sector[1]	Primary economic activity[1]	Secondary economic activity (if any)[1]	Private / public status[1]	Non-profit institution (NPI)?[1]	Links to other units[1]
Unit A							
Unit B							
...							

1. Can be adopted from other statistical frameworks or sources as enabled by data-sharing agreements, or imputed by the agency compiling the R&D statistics.

The forthcoming Oslo Manual highlighted the potential importance of capturing innovation-related links with research institutes, especially those under some form of governmental control. The manual also highlighted the potential to offer respondents the possibility to identify key partners from a pre-defined list, in order to facilitate the reporting of data and reduce burdens.

Rationale for a project in this area

As noted during the revision there is very limited information about the national differences in the strategies of implementation of those criteria set up by the national Authorities. Given recurrent requests to OECD (mostly coming from OECD policy makers and NSOs) for country-specific information on what type of organisation is covered where, it appears necessary to gather insights about the reliability (in light of potential changes of criteria) and interpretability of the international comparisons on the share of the different institutional sectors in GERD. It is necessary to advance our knowledge of NSO practices in classifying similar types of entities. The information currently provided through the sources and methods database is rather insufficient and it is not always possible to interpret cross and within country variations.

Why focus on research institutes?

Research institutes are not only of high policy importance but also present a number of classification challenges finding themselves in the borderline of many dimensions considered.

Table 2. Research institutes in the FM decision tree

Is HE main activity?#	Is it controlled by government?	Charges econ. significant prices?##	FM classification	Is R&D the main economic activity of the unit?
Yes	Yes	Yes	HE (public)	No*
Yes	Yes	No	HE (public)	No*
Yes	No	Yes	HE (private)	No*
Yes	No	No	HE (private)	No*
No	Yes	Yes	BE (public)	Maybe commercial private RI
No	Yes	No	GOV (public)	Maybe government RI
No	No	Yes	BE (private)	Maybe commercial public RI
No	No	No	PNP (private)	Maybe private non commercial RI

Notes: # Or is it controlled by a unit for which HE is the main activity?. ## Or does it depend on institutions that do so? * May contain subunits principally dedicated to R&D.

First of all, it can be challenging to ascertain if these are separate institutional units or data about them should be captured as part of the broader organization they belong to, as is the case of many institutes in higher education. The public dimension can be fuzzy due to multiple mechanisms by which governments may exert control over the general activity and the research activity more specifically. The engagement of research institutes in market activities is also a source of potential confusion, as economically significant prices may be charged for some but not all their products in proportions that may be highly volatile. A lack of profit motive may not be sufficient to fulfill the requirement for exclusion from the corporate sector, for example if an institute works entirely for industry on the basis of subscriptions. When trying to identify

research institutes, the conduct of R&D as main economic activity may not be obvious either. In sum, there are widespread concerns about how research institutes and especially those with some relationship to Government are being classified.

A relevance challenge

The Frascati Manual classification has the challenge of being fundamentally applied to units for which data are not systematically made available in the public domain, although some NSO do publish data at the level if institutions for public sector organisations under transparency requirements. However, publicly available organisational registers can be increasingly found for all sectors and include in some cases data that is purported to represent R&D or other STI related activities. Users of such resources have a compelling need to classify such units and the Frascati guidance, while helpful, may turn to be implemented in very different ways depending on the information available. In recent years, resources of this type covering several countries have become available.

For instance, institutions featuring in bibliometric data (as affiliations and funders) are commonly classified their providers. A highly population classification is that found in a number of data infrastructures in this area is that used in the Crossref funder database,[1] which has become a de facto competing standard which renders data interoperability with R&D statistics a challenge.

Table 3. Crossref funder data

Crossref funder sectors	Number of records
Academic	2871
Corporate	1359
Federal/national government	2239
Foundation	4858
Government non-federal	3861
International	182
Other non-profit	1971
Professional associations and societies	1672
Total	**19013**

---

[1] Crossref is a non for profit consortium of academic publishers. See https://www.crossref.org/services/funder-registry/ . The Funder Registry provides a common taxonomy of international funding body names for funding data participants to use to normalise funder names and IDs for deposit with Crossref. The list is intended to be used to present authors with a pre-populated "Funder Name" option at the time of submission, and can also be used to match funder names extracted from papers. The list is available to download as an RDF file, and is freely available under a CC0 license waiver. NESTI delegates can examine an extraction of the Crossref funders database containing institution name, country of affiliation, and assigned sector, in the NESTI NET community space at https://community.oecd.org/docs/DOC-142046 The Crossref funding database is originally based on data provided by Elsevier, with strong emphasis on institutions named in acknowledgements in scientific publications. The interest of institutions is to trace scientific outputs to their own funding sources. While funders are not necessarily performers, the database may ultimately may also be a source of data on R&D performing organisations because authors often acknowledge the support provided by the institutions they work for or assign dedicated resources to a project, so these eventually become enumerated.

The GRID institutional database[2] by Digital Science provides records classified in the following way:

Table 4. Classification used in GRID data

GRID sector	Description	# records
Education	An educational institution where research takes place. Can grant degrees and includes faculties, departments and schools.	17965
Healthcare	A health related facility where patients are treated. Includes hospitals, medical centres, health centres, treatment center. Includes trusts and healthcare systems.	12103
Company	Business entity with the aim of gaining profit.	24870
Archive	Repository of documents, artifacts, or specimens. Includes libraries and museums that are not part of a university.	2256
Nonprofit	Organisation that uses its surplus revenue to achieve its goals. Includes charities and other non-government research funding bodies.	10024
Government	An organisation operated mainly by the government of one or multiple countries.	5202
Facility	A building or facility dedicated to research of a specific area, usually contains specialised equipment. Includes telescopes, observatories and particle accelerators.	7262
Other	Used in cases where none of the previously mentioned types are suitable.	6784
Total		**86 466**

Both GRID and Crossref infrastructures are being taken note of as part of current work for the component of the DSIP project looking at use of global, persistent and unique identifiers in the area of science and innovation.[3] The emergence of these global resources should be, as noted as the 2016 Blue Sky forum highlighted, a major point for reflection for NSOs in terms how they approach cross boundary phenomena and collaborate with each other in contexts such as the OECD and NESTI in particular.

The project

The general objective for the proposal is to assess the degree of comparability of R&D international statistics in terms of the institutional sector classifications, but the assessment will also allow us to better estimate the evolution of the real role of Governments in R&D on similar grounds. The project can also help promote a more in depth discussion on issues relating to the potential use and or development of global data infrastructures and registers, and how NSOs working on STI data and indicators relate to them.

---

[2] GRID https://www.grid.ac/ is comprised of a worldwide collection of institutes associated with academic research. IDs and metadata are freely available for use under the Creative Commons Public Domain 1.0 International licence. The institutes contained are distinguished by a unique identifier, GRID ID. Each unique GRID record also contains relevant metadata, as well as relationships between associated institutes. GRID records can display relationships between each other (subordinate associations, called a parent-child relationship and relationships that describes other associations, called a related relationship).

[3] The ORCID consortium has also been looking into this issue.

The task

The specific objective would be to construct a database drawing on names of institutional units based on open data sources, incorporating attributes assigned by NESTI contacts within countries. The proposal is to select a set of relevant R&D units in each of the countries selected for the test, and to ask the NESTI delegates in which institutional sector have those units been classified. This means the effective application of the FM recommendation presented in Table 1 above to a concrete list of institutions.

The initial idea would be to pilot this approach with a smaller group of volunteer countries.

The two databases listed above are potential candidates for use as part of this work, but require further refinement to narrow down the country lists for NSO assessment. In addition to these, there is a curated set of data from Orgreg RISIS project that would be ready for a feasibility assessment but it is focused on European countries.[4]

The resulting database would provide an initial basis for methodological discussion within NESTI, with a view to assessing the mutual consistency of decisions. This combined cross-country resource could be useful for individual countries as they try to map international collaborations and partnerships for units within their countries. Subject to agreement at NESTI, the database could be eventually made available as public resource on an experimental basis. These are points for discussion.

The suggested set of questions for each of the institutions under consideration could be the following sequence:

Is the Unit included in your Directory as "independent" institutional Unit?

> YES / NO (If not explain)(END)

1. **If YES, in which of the institutional sectors has the unit been included for the 2016 R&D data collection?**
   a) **What is the main activity of the institutional unit?**

   EXPLAIN

   b) **Is the main activity of the institutional unit ISIC 72 (or 74)?**

---

[4] ORGREG (Organisation Register) http://risis.eu/orgreg/ OrgReg is a public facility, which provides a comprehensive register of public-sector research and higher education organisations in European countries. It is a public resource whose main function is to allow integrating different RISIS datasets at the level of actors through the definition of a common list of organisations and the use of organizational IDs that are used consistently in the RISIS datasets providing data at the level of organizational actors. Distant access to OrgReg data is possible within RISIS through the CorTexT authentication system. Users can create a CorTexT account here and accept the RISIS code and conduct for usage of data. Afterwards, users can access the OrgReg data through orgreg.joanneum.at by selecting the Register tab on the top of the page. Data downloaded from Orgreg are subject to the RISIS Code of Conduct. In particular, following conditions of usage are crucial: They can be used solely for purposes of scientific research. Any use for commercial purposes, like the provision of for-paying services, is excluded. Data have to be published in a way, which makes impossible the identification of the specific entity considered and the attribution of a cell value to that entity.

YES/ NO (If not which?)

c) Has the institutional unit public status (is it controlled by the Government)?

YES / NO (comments)

d) Does it operate on a non-commercial, non-profit basis?

YES / NO (comments)

2. **For the <u>total</u> of the "institutional units" included in your country list:**

a) What is the aggregate total expenditure in R&D? (in national currency)

- What is its share (%) in the national GERD?

3. **For the total of the "institutional units" included in your country that you classified in the institutional sector of Government:**

a) What is the aggregate total expenditure in R&D? (in national currency)

- What is its share (%) in the Government Expenditure in R&D sector?

Questions to NESTI delegates

NESTI delegates are invited to consider the issues raised in this proposal:

- Is register classification information publicly available in your country for public sector organisations doing R&D?
- Should NESTI engage more actively in understanding and potentially using available registers in the area of STI, subjecting them to further investigation for comprehensiveness and consistency?
- Would you like to participate in the proposed pilot analysis of research institutions in your country and making the information available to other NESTI delegates?

Annex Figure 1. Implied distribution of scientific authors publishing in 2017 by Frascati sector in the provisional ISSA2018 sample

(As percentage of respondents, unweighted, provisional results)
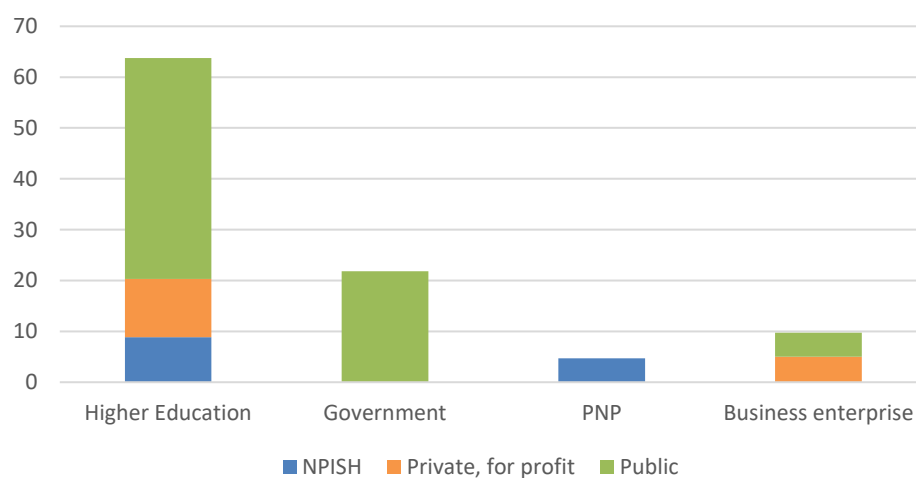
Source: OECD, preliminary analysis of ISSA 2018 data, based on questions about main activity of employer organization, potential government affiliation and potential for profit status.

Note: This figure shows how it is possible to use separate questions to characterize the sectoral affiliation of units in a specific database using the Frascati manual classification.