

Unclassified

English - Or. English

9 November 2021

**Development Co-operation Directorate  
Development Assistance Committee**

## **DAC Working Party on Development Finance Statistics**

### **Further development of keywords in the CRS data collection**

Informal WP-STAT meeting  
Hybrid meeting, 22-24 November 2021

In 2020, the WP-STAT approved the introduction of a keyword field in the CRS to track COVID-19 activities and subsequently other possible items of interest [DCD/DAC/STAT(2020)37].

This note, taking into account preliminary results of the ongoing data collection, discusses the possible expansion of the keywords.

The paper is presented for DISCUSSION at the meeting. If the necessary consensus emerges, it will be subsequently submitted for APPROVAL via the written procedure on a non-objection basis.

Giorgio Gualberti - Email: [Giorgio.Gualberti@oecd.org](mailto:Giorgio.Gualberti@oecd.org)

**JT03485004**

## Further development of keywords in the CRS data collection

### 1. Introduction

1. The keyword field was introduced in the CRS in 2020 to track activities related to COVID-19 control and relief [DCD/DAC/STAT(2020)37]. As of November 2021, preliminary results from the first data collection point to a widespread adoption of the keyword field to track COVID-19 related activities.
2. The keyword field could be used to track other topics. Different approaches are possible in this regard: a top-down approach, a bottom-up approach, or a combination of the two. The top-down approach (the one currently implemented to track COVID-19) entails a list of agreed keywords, the definitions of which are inserted in the Annex 13 of the Statistical Reporting Directives [DCD/DAC/STAT(2020)44/ADD1/FINAL]. A bottom-up approach would allow members greater flexibility in the use of keywords, allowing them to autonomously select and report keywords without prior agreement by the WP-STAT. Finally, a mixed approach would entail the coexistence between agreed keywords (included in the Annex 13) and bottom-up keywords (submitted autonomously by members).
3. This note discusses the adoption of a mixed approach to the use of the keywords field, and its possible practical implementation.
4. The paper is presented for **discussion** at the WP-STAT of 22-24 November 2020. If the necessary consensus emerges, it will be subsequently submitted for **approval** via the written procedure on a non-objection basis.

### 2. The keyword field in the Statistical Directives

5. The keyword field was introduced in the CRS in 2020 with the objective of tracking COVID-19 related activities and other possible topics of interest [DCD/DAC/STAT(2020)37]. The keyword field is described in the Statistical Directives in the following way:

#### Box 1. Keywords field description in the Statistical Directives [DCD/DAC/STAT(2020)44/FINAL, Chapter IV, Section VI, paragraphs 213-217]

##### Concept

Keywords are short strings of text that signal that the activity corresponds to an agreed definition, such as pursuing a specific policy objective or having a certain intrinsic characteristic.

The keyword field accepts both keywords and hashtags (i.e. keywords prefixed by the symbol #). The hashtags are defined in the 'keyword definitions table' (see Annex 13).

##### Structure and logic

An activity can be associated with one or more of the agreed hashtags. In addition to the agreed hashtags, keywords with the same definition but that might be written with

alternate spelling or in another language can be reported. The Secretariat maintains a lookup table (published in the community space) that links the agreed hashtags with keywords reported by members, and consolidates the data prior to their publication.

### **Field of application**

By default, keywords are associated with any activity reported in the CRS database. The 'keyword definitions table' can restrict the application of a specific hashtag to a subset of activities.

Reporting keywords is voluntary.

## **3. Possible approaches for the use of the keyword field**

6. The keyword field is a flexible tool designed to capture the cross-dimensionality of an activity. We distinguish three possible approaches for its use.

- **Top-down.** The keyword field is currently implemented following a top-down approach. While reporting keywords remains voluntary, members agree on the list of keywords collected in the CRS and on their definitions. With this approach, the keyword field functions the same way as some other cross-cutting fields in the CRS, such as the flags (e.g. the programme-based approach flag). The main advantage of the top-down approach is that it provides for consistent utilisation of the agreed keywords and remains space-efficient in terms of reporting format. One single field containing the keywords could be used to provide information on different agreed topics. On the other hand, a top-down approach does not allow innovative uses of keywords by members.
- **Bottom-up.** This approach does not put restrictions on which keywords are reported. Bottom-up approaches, often referred to as folksonomies, are a method of categorising information based on the spontaneous submission of keywords by the institution submitting the data. Folksonomies are open systems to describe data, as opposed to the more traditional taxonomies that are coherent, comprehensive and controlled systems to categorise data. One advantage of folksonomies is the freedom they provide to users to compile additional information, thus spurring innovation. Bottom-up approaches could provide additional transparency to the data, but cannot offer the consistency that is proper of statistical measures with agreed definitions. Also, keywords reported following this approach would not be comparable across members.
- **Hybrid.** Hybrid approaches to keywords reunite the characteristics of bottom-up and top-down systems described above. This entails the coexistence of agreed keywords, with common definitions, and of user-generated keywords, which are spontaneously added by the data providers. These two sets of keywords could be differentiated in the reporting by a symbol (such as a hashtag). The advantage of a hybrid system is that it can provide both the consistency of a top-down approach and the innovation of a bottom-up system. A hybrid system would need appropriate documentation for the users to understand the differences between hashtags and keywords, especially the fact that only the former is based on an agreed statistical methodology.

## **4. The initial discussion on the expansion of the keyword field**

7. At the March 2021 WP-STAT meeting, members held a first discussion on the possible expansion of the keywords field to cover other topics. The Secretariat's

presentation proposed further development of keywords based on an hybrid approach, i.e. including in the CRS a set of 'agreed keywords' listed in Annex 13 of the Statistical Directives (and preceded by an hashtag) and the possibility of accepting other keywords spontaneously submitted by members.

8. Most members reacted positively to the proposal from the Secretariat, but two members expressed concerns pointing out possible data quality issues and the need to evaluate the actual implementation of the COVID-19 keyword before expanding the data collection using other keywords. The WP-STAT agreed to further discuss this topic. [See DCD/DAC/STAT/M(2021)1/FINAL, paragraph 35.]

## 5. Initial feedback from COVID-19 reporting

9. An initial review of the use of the COVID-19 keywords shows that the keyword field is well-reported by members. While at the time of writing this paper the 2020 data processing is still ongoing, over 2 500 activities from nine members marked with the #COVID-19 hashtags have already been processed by the Secretariat.

10. In line with the Statistical Directives, some members have used alternate spelling, such as “(#) Covid-19”, “#Covid-19”, “19c”, or “#COVID19”. These alternate spellings have been unified by the Secretariat to “#COVID-19”, which will be the only one shown in publicly available data. A look-up table, which keeps track of all the different spelling options used by members, will be prepared and uploaded on the community space. For the time being only a few members have reported other keywords, but these have not been uploaded in the CRS database.

11. The Secretariat has not noted any particular problems in the use of the keyword field to track members' COVID-19 activities. A proper assessment of its implementation can however only be provided at the end of the data processing and validation (e.g. how many members reported the keyword, what analysis can be done based on the data collected). The Secretariat will provide members with feedback on this item in its regular report to members on reporting issues.

## 6. The proposal

12. This note invites members to consider the introduction of a hybrid keyword system in the CRS. As noted above, the hybrid system could serve both the needs of homogeneity and consistency of the data, but also open room for innovation in the reporting on development co-operation, allowing members (or groups of members) to spontaneously add keywords in their reporting.

13. The hybrid system proposed to manage the keywords has the following characteristics:

- **Agreed keywords (hashtags)** are approved by WP-STAT and valid for all members. The list of agreed keywords and their respective definitions are maintained in the Annex 13 of the Statistical Directives. Agreed keywords will be preceded by a hashtag (#), and are denominated 'hashtags'. Moreover, the Secretariat will maintain a look-up table to reconcile the hashtags submitted with alternate spelling options.
- **Spontaneous keywords** could be a freely chosen and reported by members but should not be preceded by a hashtag (#). Spontaneous keywords could be used to highlight a publicly relevant aspect of the activity, such as a policy objective, a specific process, the adoption

of a standard, the involvement of stakeholder group, the belonging to a particular set of activities or other intrinsic characteristics of the activity.

- **Spontaneous keywords could be upgraded to hashtags** if members find that they are relevant for all reporters. In this case, they would be added to Annex 13 among the others if members agree to officialise them.

14. To allow the reporting of spontaneous keywords and to differentiate them from agreed hashtags, the Reporting Directives (Chapter IV, Section VI, paragraphs 213-217) would be modified as follows.

**Box 2. Proposed amendments to the Statistical Directives (~~strikethrough~~ indicates deleted text and underlining added text; note also that some text has been moved)**

**Concept**

The keyword field accepts both keywords and hashtags (~~i.e. keywords prefixed by the symbol #~~).

Keywords and hashtags are short strings of text that describe a publicly relevant aspect of the activities, such as pursuing a specific policy objective, or having a certain intrinsic characteristic, that could not be otherwise reported with the existing taxonomies.

Hashtags are keywords prefixed by the symbol #. The hashtags are agreed among members and are defined in the '~~keyword~~ Hashtags definitions table' (see Annex 13).

Keywords are spontaneously submitted by members. Keywords cannot be prefixed by the symbol #.

**Structure and logic**

An activity can be associated with one or more of the agreed hashtags or keywords. In addition to the agreed hashtags, ~~keywords~~ other hashtags with the same definition ~~but that might be written with~~ alternate spelling or in another language can be reported. The Secretariat maintains a lookup table (published in the community space) that links the agreed hashtags with ~~keywords~~ the equivalent hashtags reported by members, and consolidates the data prior to their publication.

**Field of application**

By default, hashtags and keywords ~~are~~ could be associated with any activity reported in the CRS database. The '~~keyword~~ Hashtags definitions table' can restrict the application of a specific hashtag to a subset of activities.

Reporting hashtags or keywords is voluntary.

## 7. Next steps

15. Members are invited to express their views on the proposed extension of keywords to user-submitted keywords. If the necessary consensus emerges at the meeting, the proposal will be submitted for approval via the written procedure on a non-objection basis.