

Document Labeling Policy

For Document Publication Distribution
Statements

January 14, 2020- v1

Distribution: Public

Author: Andrew Adams



1 Executive Summary

In response to incident '2019-10-02_01', a scheme for labeling documents was developed. The goal of this scheme (and policy to implement it) is to ensure that future documents produced by Trusted CI are published in the manner identified by the document's maintainer. The scheme and policy should be simple to employ by the maintainer, it should not be too obtrusive on a document's title page and it should be sufficiently tractable by the publications' liaison in both identifying and parsing the distribution label specified. Moreover, it should not conflict with existing distribution statements, thus, providing a query-able string component to allow batch processing of Trusted CI's numerous documents.

Several systems ranging from the canonical classifiers used to mark confidential documents as outlined in various executive orders, e.g., 'Confidential', 'Restricted', to more robust province-based systems were surveyed. Our solution to a document distribution statement, defined in section 3, is comprised of elements from the schemes we surveyed. A brief description of the methods surveys, as well as an explanation as to why said systems were not suitable for our use are outlined in section 2. Finally, guidelines for implementing our scheme are found in section 4.

2 Survey of existing document classifying schemes

Below is a survey of document classification/distribution systems that were surveyed. For each, we briefly describe the system, then discuss why said scheme is inadequate for Trusted CI. If you wish to skip directly to the scheme chosen, please see section 3.

Executive order classifiers - simple, one- or two-word labels, e.g., 'Confidential' and 'Restricted' used to identify a document's security classification. Although the scheme satisfies ease of use, they lack additional necessary attributes, e.g., a date signifying when an embargo on a document can/should be lifted.

Superintendent of Documents Classification System (SuDoc)¹ - a mature scheme that allows for multiple attributes, the author symbol, series designation, and book number, to be encoded

¹ <https://www.angelo.edu/services/library/handouts/sudoc.php>

within the label. The system uses a single, two or three letter prefix to mark the author or province of the document, e.g., 'A' stands for Agriculture and 'LC' for Library of Congress. Further refinements to the author are included as the series designation. A colon then separates the province from the book number, encoded as a type and usually a year (minus the thousandths place). For example, the SuDoc number 'A 1.35:R 42/995' identifies a committee in the Dept. of Agriculture ('A 1.35:') and the book 'R 42' published in 1995. The primary issue with these types of systems, though, is their complexity. That is, its grammar is comprised of multiple delimiters (':', ':" and '/') which all have specific meaning. Moreover, the numeric codes that represent the series and sub jurisdictions of the author and the book type require an index.

United Nations scheme - a variation of the SuDoc province-based system in use at the United Nations,² was additionally examined. It possessed two or more sections, or segments, delimited solely by a '/'. The first segment, similar to SuDoc is a one, two or three letter code to identify the author. The second segment is for further refinement to the author (e.g., a subassembly) and is optional. The determinant for its presence is a simple check to see if the first character in the second segment is non-numeric. The next segment (the second if no author refinement) is the year of publication. The final segment, if necessary, is the document number. Of interest here, is that an additional segment may occur between the year and document for including an attribute, e.g., a distribution classification! That is, the label 'A/S35/2019/R' would identify a document produced by a subassembly of the General Assembly in 1995 that is *restricted*. A variant of the United Nations Classification System, where the province would be 'TCI' for Trusted CI, the year segment would be the date of desired publication in YYYY-MM-DD format, e.g., 2020-01-15, and the last segment would be any modifying attribute (i.e., classifiers) that should be applied to the document, including: 'INTERNAL' for Internal Use Only, 'EMBARGOED' for embargoed, and 'PUBLIC' for public, was considered. However, it too was deemed too complex.

Traffic Light Protocol - a solution discussed was to use rely on simple words, as in the executive orders, but to color-code them. For example, '**Confidential**' (or based on our current Classification Policy, '**Internal Use Only**') would be printed using red. '**Embargoed**' would be printed in amber, and '**Public**' would be in green; weakly akin to the Traffic Light Protocol³.

DTIC Distribution Statements - a well-defined human-readable statement describing the intended audience, the reason a control is in place, a date, and the controlling authority, e.g., Distribution authorized to IU only; Restricted information ; 01 Jan 20. Although this scheme is clear, it appears more cumbersome than is necessary for Trusted CI. Moreover, the date could

² <https://guides.library.illinois.edu/c.php?g=439687&p=3006514>

³ <https://www.us-cert.gov/tlp>

be interpreted ambiguously; that is, is it the date of release or the date the controls were set in place.

3 Document Labeling Scheme

After surveying and reviewing the aforementioned schemes, the ISO and leadership decided that simple and concise was best. Moreover, it was agreed that color-coding aided readability significantly. Hence, we settled on a single word *'Distribution'* followed by a colon to announce a distribution statement. This prefix not only announces a distribution statement, but it gives us uniformity. The document's state is then specified with either a single word or short clause, color-coded. Specifically:

- for documents with no restrictions, the color-coded word *'public'* is used,
- for internal private/confidential documents, the color-code phrase *'Internal Use Only'* is used,
- for documents that have restrictions, the color-coded phrase *'Release to'* followed by the target is used,
- And if the restriction is guided by a date, i.e., the document is embargoed, then following the *'Release to'* & target clause an *'after'* word and date should be appended.

Examples;

- a document for public disclosure would have a label affixed to the title page of *'Distribution: Public'*
- a document for distribution to just a third party (e.g., an engagee) would have the label (in the case of being restricted to just the NSF) *'Distribution: Release to NSF'*
- a document embargoed until Jan. 15, 2020 would have the label *'Distribution: Release to public after Jan 15 2020'*
- an embargoed document for distribution to just a third party would have the label *'Distribution: Release to NSF after Jan 15 2020'*
- A document intended to remain internal should be marked with *'Distribution: Internal Use Only'*

Note, since bright green (*sample*) is hard to read/see, please don't use it. If unsure, it is acceptable to color-code 'Public' with **black**.

Caveat: 'Distribution:' was chosen over 'Trusted CI Distribution to' for three reasons: first, the former is shorter and simpler, yet still unique (for searching); second, this policy is tailored for Trusted CI (i.e., the releasing authority will never change); finally, we wanted a different clause to allow us to retroactively search for older distribution statements if ever there was a need to update them.

4 Procedures for publication

Documents using Trusted CI's short and long report format already include a place-holder for the distribution statement. This labeling policy should be applied in that place-holder. Other documents and slides should include this distribution scheme in the title page, preferably on the lower right.

For data that is immutable (e.g., it's not text or it's owned by someone else), then the distribution statement needs to be appended to the filename (or to the parent folder if editing the filename is impractical).

It is up to the document's author to ensure that the distribution statement is correct. The publication's liaison used is responsible for checking the distribution statement prior to uploading the document and verify that it does indeed satisfy the expected release.

Note, documents that do not include a distribution are, by default, public.