

Copyright Statement Detection and Extraction Policy

Identification

We differentiate between authors, holders and copyright statements (as currently suggested by the ScanCode Toolkit).

- Authors are individuals or entities having written or contributing materials.
- Holders are individuals or entities holding rights on the materials.
- Copyright statements are markings of the authors/holders to identify holders and periods in which the materials have been created.

CIR-01: Copyright Identification Rule 01

As explicit copyright statement we identify any statement that conveys right-holder-ship of the conveyed materials. Clear markings are ©, (C), (c) or the term *copyright* itself paired with names of individuals/entities and/or time data (i.e. such as year or range of years).

A mention in text of a right holder is not regarded an explicit copyright statement. We do not induce that the provider of the text intended to make an explicit statement. Yet the individual/entity is regarded a holder.

Extraction

CEP-01: Copyright Statements are extracted “as is”; they are not modified.

The copyright holder has made the statement. An extraction “as is” respects the choices and intentions of the copyright holder.

A technical limitation is different character sets and graphical representations. The extracted copyright statements must try to represent the original statement as far as possible.

Tabs and newlines and tabs are preserved.

Commenting related formatting (systematically trailing // or *) must not be preserved. These are regarded external boundary conditions.

CEP-02: “All rights reserved.” remarks are considered part of the copyright statement.

The “All rights reserved.” Statements are not considered part of the license. As such it is regarded part of the statement by the copyright owner. The statement is

not required in general (since copyright-law is in place in participating countries). Yet, in case the copyright holder provides the remark, the remark is preserved as part of the copyright statement.

In general, all remarks by the copyright holder are regarded part of the copyright statement and are to be preserved.

CEP-03: A block of copyright statements is not decomposed.

A block of copyright statements may list several copyright holders of separate parts or changes. A copyright statement block may indicate a joint work of several copyright holders. A copyright statement may include “All rights reserved” remarks, not well aligned with the individual copyright holders.

The block is preserved since a decomposition requires additional knowledge on the copyrighted material. Preserving the block does not impose additional knowledge or interpret the intentions of the copyright holders. It does not disconnect individual marks from other parts.

CEP-04: Copyright statement on license texts must not be identified for the copyrighted materials or must be identified separately.

Licenses may have a copyright different from the materials supplied under the license. The copyright must be clearly assignable to the subject.

Copyrights of a license must not be ignored but identified as copyrights for the license.

CEP-05: Copyright statements are not consolidated.

I.e. a harmonization of copyright holders or consolidation/merges of time ranges are not allowed.

CEP-06: Copyrights are individually identified.

Copyrights are mapped to the associated license in a later processing step. A copyright may be mentioned several times with different license associations.

In the subsequent documentation it is sufficient to list a copyright statement once in combination with its license association. Repeating the same copyright with the same license is not regarded required.