

# Certification Criteria for XML Article text Tagging Services according to SciELO Publishing Schema (SciELO PS)

## Introduction

In the past few years, the SciELO Program has been prioritizing lines of action to strengthen professionalization and internacionalization of publishing and publication services for indexed journals. To achieve this, it is updating and extending the common services of publishing, publication and interoperability that are operated by the SciELO Platform in partnership with international and national companies.

Among SciELO platform services are the text structuring in XML language for storage in databases, meta-data extraction, bibliometric calculations, and interoperability with other Web sources. In 2013, the methodology of the platform's XML text tagging and structuring was updated with the [SciELO Publishing Schema \(SciELO PS\)](#) based on standard [ANSI / NISO Z39.96-2012, JATS: Journal Article Tag Suite \(version 1.0\)](#), derived from the PubMed Central DTD.

[SciELO PS](#) is fully compatible with the standard [ANSI / NISO Z39.96-2012, JATS: Journal Article Tag Suite \(version 1.0\)](#) (and with the PMC DTD), but has additional data elements that are essential for SciELO indexing functions:

- Tagging needs institutional affiliation from all authors, necessary to generate the scientific production indicators per institution and its main units (e.g: University, College, etc.);
- Tagging of data elements that compose the references of the documents cited in articles, indispensable for SciELO's bibliometric indexes, and
- Identification of funding sources for the research.

With the objective of seeking correct application of the article texts structuring schema, SciELO establishes minimum criteria for the companies that render text-tagging services for SciELO journals.

**IMPORTANT:** Each institution is responsible for monitor the documentations launch versions, as well as introduce the modifications on XML production as indicated.

## 1. How to apply for certification

For certification services for tagging text, it is necessary to contact the SciELO team by mail [scielo@scielo.org](mailto:scielo@scielo.org) and request of sample articles for XML production as [SciELO PS](#).

It will be forwarded to the applicant a set of files for XML production containing different types of documents of SciELO Network journals.

The set of files, when finished, will be sent to the team SciELO by sendspace or wetransfer, communicating its sent by email: [scielo@scielo.org](mailto:scielo@scielo.org).

## 2. Sample for submission

The sample files must be marked according to the latest version of [SciELO PS](#) observed [SciELO Brasil Specific Rules](#).

Each sent must contain the material produced in .zip format, here referred to as "data package", bringing together the following files:

- a) XML Files
- b) PDF
- c) Digital Assets<sup>1</sup>

## 3. Evaluation Process:

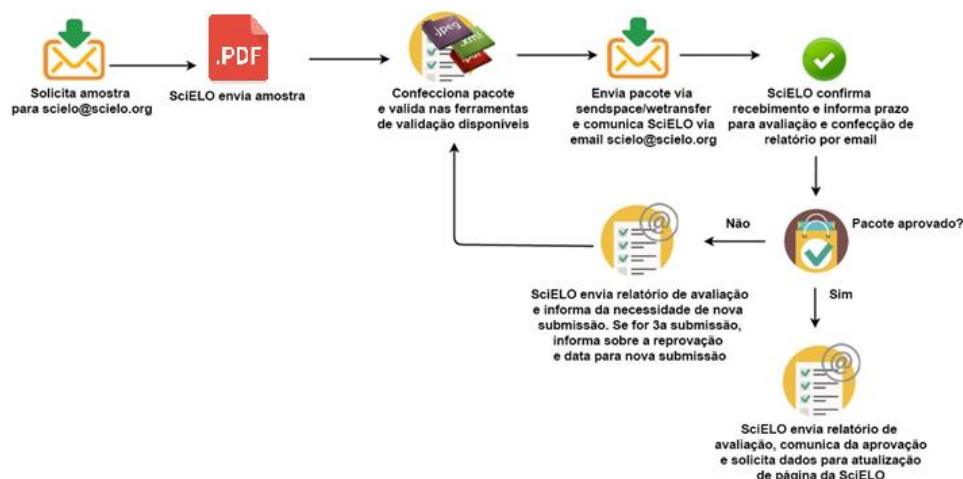
After receiving the data package (XML, PDF, digital assets), the SciELO team will evaluate the compliance of files [SciELO PS](#), the result presented by the validators ([see item 6](#)) and verify the marking of articles elements in the XML document against the PDF.

It will be sent detailed report for the data package correction, if it has.

---

<sup>1</sup>Digital Assets are classified as image files, multimedia and textual content. See in: [https://en.wikipedia.org/wiki/Digital\\_asset](https://en.wikipedia.org/wiki/Digital_asset)

The data package may be re-submitted after submission of the evaluation report, obeying the limit of 2 attempts (Figure 1).



**Figure 1:** Sample application and evaluation for certification of tagging text services according SciELO PS

In case of doubt about marking the article, generation or validation of XML documents, you may contact the SciELO team through discussion list [scielo-xml](mailto:scielo-xml)<sup>2</sup>.

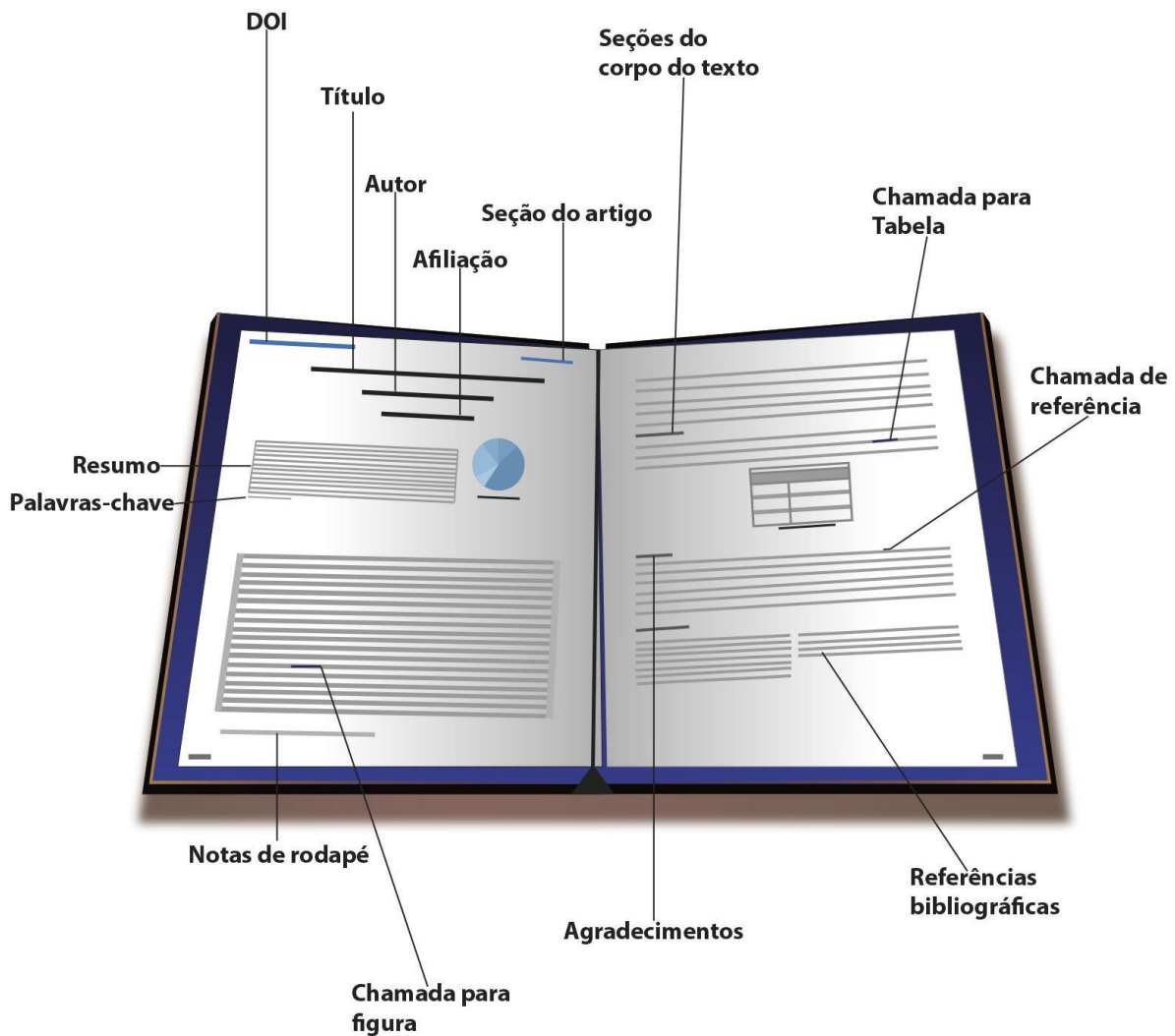
#### 4. Criteria for approval and certification services as tagging text according [SciELO PS](#)

It will be approved the data package is:

- in accordance with [SciELO PS](#);
- valid validation tools offered, without making errors ([see item 6](#)); and
- with its contents identified / marked appropriately (Figure 2).

<sup>2</sup> The SciELO - XML is a SciELO group for discussion and clarification of doubts related XML SciELO processes. To become a member, contact [scielo@scielo.org](mailto:scielo@scielo.org) requesting registration.

PS: Before submitting a question in the discussion list, check if the doubts have already been answered.



**Figure 2:** Content to be marked and evaluated for certification

In case of failure, a new application for submission may be made after 6 months after the last communication by SciELO.

## 5. Continuing Evaluation

Data packages sent to SciELO, after certification, will be continuously evaluated in considering:

- XML support structure according to the [SciELO PS](#) (Most current version)
- Frequent errors;
- Delivery time material, including corrections.

In case of unsatisfactory performance in one or more items, certification will be discredited and will only be possible to request a reevaluation six months after the communication of the loss of certification.

## 6. Tools SciELO

They were developed by the SciELO team the following support tools for producing XML packages:

- [Wayta](#) - Web-service for standardizing names of institutions and country names;
- [PC-Programs](#) - Set of tools SciELO Methodology to manage journals, numbers and data items, to publish on the site;
- [Package Maker](#)- Validation Tool for SciELO data packages. It presents content dialing errors and conformity of the structure of XML;
- [Packtools](#) - Python library that adds features and utilities for a data package handling SciELO PS and XMLs SciELO PS;
  - [StyleChecker](#) - Web-based tool that presents detailed report on the conformity of an XML data in relation to the SciELO PS specification;

## 7. Documentation and Support Material

In the list below are indicated document references and tools that assist in understanding and in generating XML files.

- *SciELO Publishing Schema* (SciELO PS): <[http://docs.scielo.org/projects/scielo-publishing-schema/pt\\_BR/latest/](http://docs.scielo.org/projects/scielo-publishing-schema/pt_BR/latest/)>
- Welcome to SciELO PC Programs' documentation: <[http://docs.scielo.org/projects/scielo-pc-programs/en/latest/pt\\_how\\_to\\_use\\_the\\_programs.html](http://docs.scielo.org/projects/scielo-pc-programs/en/latest/pt_how_to_use_the_programs.html)>
- JATS 1.0
- <<http://jats.nlm.nih.gov/publishing/tag-library/1.0/>>
- JATS 1.1
- <<http://jats.nlm.nih.gov/publishing/tag-library/1.1/>>
- SciELO Style Checker
- <<http://manager.scielo.org/tools/validators/stylechecker/>>

- Release models: To check the specifications for the development of DOI number and specifications documents as erratum, continued publication ahead of print etc, see the link below:

<http://www.scielo.org/php/level.php?lang=pt&component=56&item=57>