

## iThenticate Similarity Report README

Each manuscript newly submitted to the Copernicus Office Editor undergoes a File Validation before the Call-for-Editors process begins.

Besides a technical check of the submitted files (abstract, manuscript, supplement material [optional], as well as author's response file [optional]), the manuscript PDF file will be checked against the **iThenticate.com** scientific publication database. iThenticate extracts the actual text of the manuscript and compares it to three different sources:

- **CrossCheck:** Organized by CrossRef, the CrossCheck database contains about 30 million documents from scientific journals, conference proceedings and books. All main publishers in STM (Science, Techniques, Medicine) submit their published articles to CrossRef and register their DOIs through this agency
- **Internet:** iThenticate indexes billions of web pages and stores the content of the eight previous years. Everyday they index several million web documents
- **Publications:** Third party periodicals different from CrossCheck are indexed containing EBSCO and other databases.

iThenticate provides Copernicus Publications with a Similarity Report for each submitted manuscript providing (a) the percentage of similarity resulting from counted words, and (b) the mark-up of the various sections, expressions, or terms which were found in other documents.

Copernicus Publications checks the Similarity Report and, if applicable, the report will be uploaded into the Copernicus Office Editor to be shown to the handling editor of the respective manuscript.

### Case 1: No/insignificant Similarities

The MS Records show the event *iThenticate.com Similarity Report completed (DATE) similarities negligible / not found*:

<b>Initial Submission</b>	
<b>Editor Call First Choice</b>	(20 Jul 2011)
<b>iThenticate.com Similarity Report completed</b>	(20 Jul 2011) similarities negligible / not found
<b>Uploaded Files validated</b>	(20 Jul 2011) by Anna Mirena Feist-Polner
<b>File Upload</b>	(20 Jul 2011) <input type="checkbox"/> Abstract <input type="checkbox"/> Manuscript <input type="checkbox"/> Supplement

## Case 2: Significant Similarities

The MS Records show the event *iThenticate.com Similarity Report completed (DATE)* → Report:

Initial Submission	
Editor Call First Choice (20 Jul 2011)	
Uploaded Files validated (20 Jul 2011) by Anna Mirena Feist-Polner	
iThenticate.com Similarity Report completed (20 Jul 2011) <a href="#">Report</a>	
File Upload (19 Jul 2011)	<a href="#">Abstract</a> <a href="#">Manuscript</a>

When downloading the report, the handling editor will receive a PDF file containing the following information:

**sources:**

---

1 196 words / 3% - CrossCheck  
[M. Rasmussen, "Open Access Journals using Public Peer-Review", Example Journal, 2011](#)

---

2 180 words / 3% - CrossCheck  
[N. Deisel, "Open Access Publishing in Europe", Sample Conference Proceedings, 2010](#)

---

3 167 words / 2% - CrossCheck  
[M. Rasmussen, "To ensure Availability and Quality of Research Data through Open Access and Public Peer-Review", Geophysical Research Abstracts, Vol. 13, EGU2011-13647, 2011](#)

---

**paper text:**

The Open Access Business Model of Copernicus  
 Martin Rasmussen 1 , Nadine Deisel 1

**Copernicus Publications**, Copernicus GmbH, **Bahnhofsallee 1e, 37081 Göttingen,** 3  
**Germany, Correspondence to:** [info@copernicus.org](mailto:info@copernicus.org)

The Similarity Report first lists the major sources of similarity. Each source is marked by a number as well as by a specific colour. The list provides the author

and title of the source, links to the source (original document / journal article), and provides the information about the number of similar words and their percentage to the overall number of words in the article that was checked.

The example above shows relatively small similarities of 2–3% and marks the relevant sections and words of analogousness. Thereby, the difficulty is (a) the distribution of similarities, and (b) the content.

- (a) The percentage of similarity is based on the number of same words. These may be distributed over several pages and sections, which is not critical, or may result from only a few paragraphs with high similarity, which is rather critical as it could imply abuse of intellectual property.
- (b) Similarity may result from common terms and expressions which are not critical but may sum up to a number of words high enough to provide a significant percentage of similarity. In the example above, similarities were marked blue as source 3 in the author's affiliation, which is also not critical.

The decision whether a manuscript should be rejected because of fraud or should proceed to the peer-review process belongs to the handling editor. The Similarity Report is also made available to referees.