publications

MDPI

Article

Establishing Genealogies of Born Digital Content: The Suitability of Revision Identifier (RSID) Numbers in MS Word for Forensic Enquiry

Dirk H. R. Spennemann [1,*] and Rudolf J. Spennemann [2]

[1] School of Agricultural, Environmental and Veterinary Sciences, Charles Sturt University, P.O. Box 789, Albury, NSW 2640, Australia
[2] Independent Researcher, Canberra, ACT 2602, Australia
* Correspondence: dspennemann@csu.edu.au

Abstract: Born-digital content is rapidly becoming the norm for literary works, professional reports, academic journal articles, and formal corporate correspondence. From the perspective of digital forensics, there is a need to understand the origin of a document and its entire creation process, from outlining and drafting to editing the final version of the text. Revision save identifier (RSID) numbers embedded in MS Word documents have been used to examine the nature and extent of individual edits within a document. These RSIDs remain logged in the metadata even if the text with which they were associated has been removed. As copies of such files retain the original's RSIDs, this metadata can be used to determine the order in which documents were cloned from each other. As a proof-of-concept, this paper examined over 400 template files generated by a single publisher for manuscript submissions to its journals. The study can show that it is possible to establish genealogies and thus relative chronologies of born digital content by first identifying those documents that share a document (root) RSID and then seriating those RSIDs that are shared between two or more documents.

Keywords: born-digital content; digital forensics; publishing process; MS Word; Open XML standard

Citation: Spennemann, D.H.R.; Spennemann, R.J. Establishing Genealogies of Born Digital Content: The Suitability of Revision Identifier (RSID) Numbers in MS Word for Forensic Enquiry. *Publications* **2023**, *11*, 35. https://doi.org/10.3390/publications11030035

Academic Editor: Costantino Thanos

Received: 2 March 2023
Revised: 12 May 2023
Accepted: 19 June 2023
Published: 25 June 2023

1. Introduction

The personal computing and desktop publishing revolutions of the 1980s and 1990s, coupled with the rise to prominence of the WWW and associated e-mail communications in the late 1990s and early 2000s, have seen a fundamental shift in the generation of documents. The generation of textual material originates overwhelmingly in digital form, from outlining to writing and editing the final version of the text. This "born digital content" is poised to become the sole evidence for an author's writing process for all written materials, be they literary works, professional reports, academic journal articles, or formal corporate correspondence. Akin to heavily edited typewritten manuscripts replete with strike-outs and manual marginal annotations, digital manuscripts can retain evidence of the editing processes of deletion, rephrasing, and addition. Unlike typewritten manuscripts, however, the evidence for these editing processes is not visible to the casual, everyday user.

Two aspects of digital manuscript files created by word processing programs are of interest from a forensic science and digital archeology perspective: their intentional manipulation for the purposes of hiding information (steganography) [1–7] and their use to identify content generation and editing processes that may shed light on the origin of the information contained in the files. The latter is of importance for understanding the nature, sequence, and extent of textual modifications in, say, policy documents. Importantly, in cases where multiple, iteratively generated versions exist, it might be useful to be able to reconstruct the overall editing sequence and thereby arrive at a 'genealogy' of the various documents.

Microsoft Word (henceforth MS Word), supplied as part of the Microsoft Office package of desktop applications and available for the main operating systems of Microsoft Windows and Apple Mac OS, has become one of the most ubiquitously used word processing programs and has attained a monopoly in the corporate, educational, and consumer space [8]. It is therefore useful to examine the suitability of MS Word files for digital forensics inquiries.

Using a large data set of MS Word document template files created by a single publisher, this paper provides a proof-of-concept demonstration to show that document genealogies can indeed be established using "revision save identifiers" (RISD) embedded in MS Word files—and that such genealogies are independent of metadata, such as time stamps, that can be easily manipulated.

1.1. Office Open XML Standard and the Implementation of Revision Save Identifier Numbers

From 2007 onwards, Microsoft implemented the Office Open XML standard [9], where all document components (text, styles, settings, metadata, etc.) are stored in a compressed archive, which is visible to the general user as a .docx file. In compliance with the standard, the application MS Word assigns a revision save identifier (rsid) number to each document (stored in the <rsidRoot> tag in the settings.xml file) and also to every document editing session that begins and ends with a save (or e-mail send) action. These RSID numbers are (i) embedded in a tag that encloses the respective text inserted during that editing session and (ii) stored in the settings.xml file as a sequence of eight-digit hexadecimal numbers. In the document XML file, each RSID is associated with a specific action such as a text insertion or deletion (if the track changes function was activated) as well as specific formatting such as bolding, the insertion of page and section breaks, and subsections of tables.

The Office Open XML standard requires that RSID numbers be "randomly generated based on the current time" and that "every editing session shall be assigned a revision save ID that is larger than all earlier ones in the same file" [9]. As noted elsewhere, MS Word does not faithfully implement the Office Open XML standard, and successive edits are not assigned RSID numbers of increasing value [10]. The Office Open XML standard also stipulates that "[a]n identical <rsid> value between two documents with the same <rsidRoot> shall indicate the same editing sessions" [9]. Thus, if a digital copy of a document is made, either by duplicating via the operating system (file copy) or via a "save as" action in MS Word, then both files will carry the same rootRSID and all other RSID listed in the respective settings XML files. The listed RSID numbers will start to diverge with the addition of new RSIDs, which reflect subsequent edit and save events that are unique to the duplicated files (Figure 1) [9]. Testing has shown that MS Word, both in its standalone desktop and distributed versions (Microsoft 365), does implement this part of the standard [10]. RootRSID numbers, as well as RSIDs assigned to an editing session, stay with a given document, even if this document is being shuttled around between multiple authors (with additional RSIDs allocated as editing progresses, irrespective of the computer used).

```
<w:rsids>
 <w:rsidRoot w:val="00E13154"/>
 <w:rsid w:val="004240E9"/>
 <w:rsid w:val="00E13154"/>
</w:rsids>
```

A

```
<w:rsids>
 <w:rsidRoot w:val="00E13154"/>
 <w:rsid w:val="004240E9"/>
 <w:rsid w:val="00E13154"/>
 <w:rsid w:val="00E51B9F"/>
</w:rsids>
```

B

```
<w:rsids>
 <w:rsidRoot w:val="00E13154"/>
 <w:rsid w:val="004240E9"/>
 <w:rsid w:val="00E13154"/>
 <w:rsid w:val="00E82E7F"/>
 <w:rsid w:val="00FE1D13"/>
</w:rsids>
```

C

Figure 1. Implementation of RSIDRoot and RSID values. (**A**) original file; (**B**,**C**) file copy with subsequent unique edits.

1.2. Using Revision Save Identifier Numbers to Develop Genealogies of Born-Digital Content

The adherence of MS Word to that part of the Office Open XML standard allows us to develop genealogies of document generation. The process of developing these genealogies is best explained through a hypothetical example, which shall be comprised of seven files (F1–F7), each of which contains eleven RSIDs. For the sake of simplicity, these were given alphanumeric identifiers in Figure 2. Each of these seven files contains RSIDs that are unique to that file and some that are shared between one or more other files. The removal of all RSIDs that are not shared between two or more of these files, followed by a seriation, results in the RSID pattern as shown in Figure 3.

F 1	F2	F3	F4	F5	F6	F7
A	D	B	B	B	B	B
C	F	D	D	D	D	D
D	I	F	F	F	F	F
F	K	I	H	H	H	H
G	L	K	I	I	I	I
I	M	M	K	K	J	J
Q	N	P	M	M	K	K
S	O	R	P	P	M	M
S	X	AB	R	R	P	P
W	Y	AC	T	T	R	R
Z	AA	AD	V	AE	V	U

Figure 2. Hypothetical example of the RSID allocation in seven files (F1–F7).

F 1	F2	F3	F4	F5	F6	F7
D	D	D	D	D	D	D
F	F	F	F	F	F	F
I	I	I	I	I	I	I
	K	K	K	K	K	K
	M	M	M	M	M	M
		B	B	B	B	B
		P	P	P	P	P
		R	R	R	R	R
			H	H	H	H
					J	J
			T	T		

Figure 3. RSID pattern after removal of unique RSID and seriation.

In this example, three RSIDs (D, F, and I) are shared by all seven files, with two RSIDs shared by six files, three RSIDs shared by five files, one RSID shared by four files, and two RSIDs shared by two files each (Figure 3). Given that file F1 shares only three RSIDs with the other six files, and those files show other commonalities, F1 is the parent of the other six. Likewise, as F2 shares only five RSIDs with the other children of F1, files that also show other commonalities, F2 must be the parent of the remaining files. This logic allows us to develop a genealogy of files that can be represented in a phylogram (Figure 4). In the case of files F4 to F7, all have one additional RSID in common (RSID H), but differ, with files F4 and F5 sharing one RSID (T) that files F6 and F7 do not have, whereas files F6 and F7 possess a RSID (J) that is not shared with files F4 and F5. Thus, these four files all have a common ancestor, then split and split again. Genealogical priority cannot be resolved at the lowest level, however. While both F4 and F5 have a common direct ancestor, it is

not clear whether F5 was split off from F4 or vice versa. A relative level of time depth can be established based on the number of shared RSIDs (representing the number of edits) between splits in the file sequence.

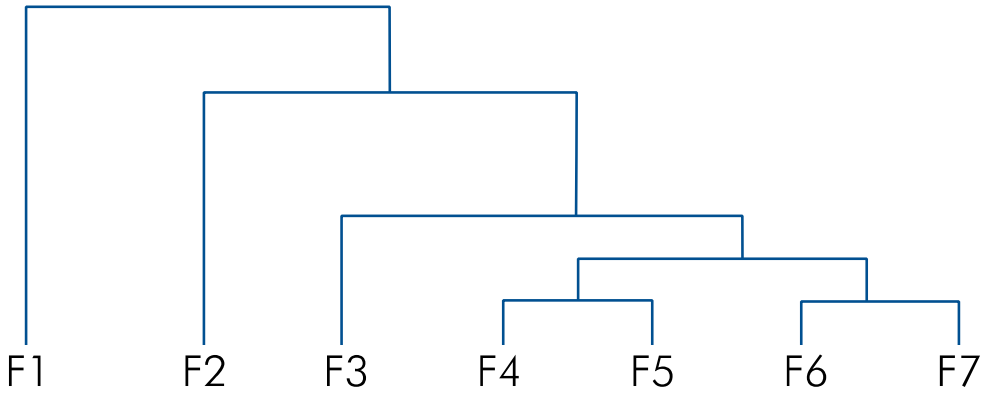

Figure 4. Phylogram of a file genealogy based on the RSIDs shown in Figure 3.

1.3. Aim of the Paper

Drawing on a large data set of MS Word document template files created for more than 400 journals published by a single publisher, this paper provides a proof-of-concept demonstration to show that document genealogies can in fact be established. At the conceptualization of the research, it had been posited that the overall ease of workflow and the time savings derived therefrom predicated that the overwhelming majority of the publisher's journal template files would have been created by cloning and modifying one or more existing templates. The technique discussed in this paper will gain in importance as the need for the conservation, curation, and interpretation of 'born digital' content increases [11–14].

2. Methodology

2.1. Data Acquistion

By October 2022, the publisher Multidisciplinary Digital Publishing Institute (MDPI) (www.mdpi.com, accessed on 11 October 2022) had published 402 journal titles (see Appendix A for details). The publishing model requires authors to submit the manuscript in a close-to-publication-ready format, which entails the use of a journal-specific document template (Figure 5), which can be downloaded from the respective "Instructions for Authors" page. The document template file for each journal was individually downloaded. In addition, the Internet Wayback Machine (www.archive.org/web/, accessed on 10 February 2023) was used to acquire numerous archived copies of template files of several journals dating back to 2010, which covers the majority of journal titles (Figure 6).

2.2. Data Processing

All MS Word files, including MS Word templates (.dot), are compressed XML archives, which can be readily extracted by changing the file type from .dot to .zip.

Using a software tool written in Java, the RSID data (rootRSID and general RSIDs) were extracted from the settings.xml file (in the "word" subfolder) of each journal template file and written to a CSV data file. The same process was executed to extract the document creation and modification dates (file core.xml in the "docProps" subfolder). The RSID CSV file was sorted in MS Excel, with the RSID data for each template with the same root RSID subsequently extracted into separate work sheets for seriation. The seriated data were graphed to create a phylogram for several of the rootRSID.

publications

MDPI

Type of the Paper (Article, Review, Communication, etc.)

Title

Firstname Lastname [1], Firstname Lastname [2] and Firstname Lastname [2,*]

[1] Affiliation 1; e-mail@e-mail.com
[2] Affiliation 2; e-mail@e-mail.com
* Correspondence: e-mail@e-mail.com; Tel.: (optional; include country code; if there are multiple corresponding authors, add author initials)

Abstract: A single paragraph of about 200 words maximum. For research articles, abstracts should give a pertinent overview of the work. We strongly encourage authors to use the following style of structured abstracts, but without headings: (1) Background: Place the question addressed in a broad context and highlight the purpose of the study; (2) Methods: briefly describe the main methods or treatments applied; (3) Results: summarize the article's main findings; (4) Conclusions: indicate the main conclusions or interpretations. The abstract should be an objective representation of the article and it must not contain results that are not presented and substantiated in the main text and should not exaggerate the main conclusions.

Keywords: keyword 1; keyword 2; keyword 3 (List three to ten pertinent keywords specific to the article yet reasonably common within the subject discipline.)

0. How to Use This Template

The template details the sections that can be used in a manuscript. Note that each section has a corresponding style, which can be found in the "Styles" menu of Word. Sections that are not mandatory are listed as such. The section titles given are for articles. Review papers and other article types have a more flexible structure.

Remove this paragraph and start section numbering with 1. For any questions, please contact the editorial office of the journal or support@mdpi.com.

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Publications* **2022**, *10*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The introduction should briefly place the study in a broad context and highlight why it is important. It should define the purpose of the work and its significance. The current state of the research field should be carefully reviewed and key publications cited. Please highlight controversial and diverging hypotheses when necessary. Finally, briefly mention the main aim of the work and highlight the principal conclusions. As far as possible, please keep the introduction comprehensible to scientists outside your particular field of research. References should be numbered in order of appearance and indicated by a numeral or numerals in square brackets—e.g., [1] or [2,3], or [4–6]. See the end of the document for further details on references.

2. Materials and Methods

The Materials and Methods should be described with sufficient details to allow others to replicate and build on the published results. Please note that the publication of your manuscript implicates that you must make all materials, data, computer code, and protocols associated with the publication available to readers. Please disclose at the submission stage any restrictions on the availability of materials or information. New methods and protocols should be described in detail while well-established methods can be briefly described and appropriately cited.

Figure 5. First page of the MDPI journal template as it was current in 2022, containing sample text, images, and embedded stylesheets.

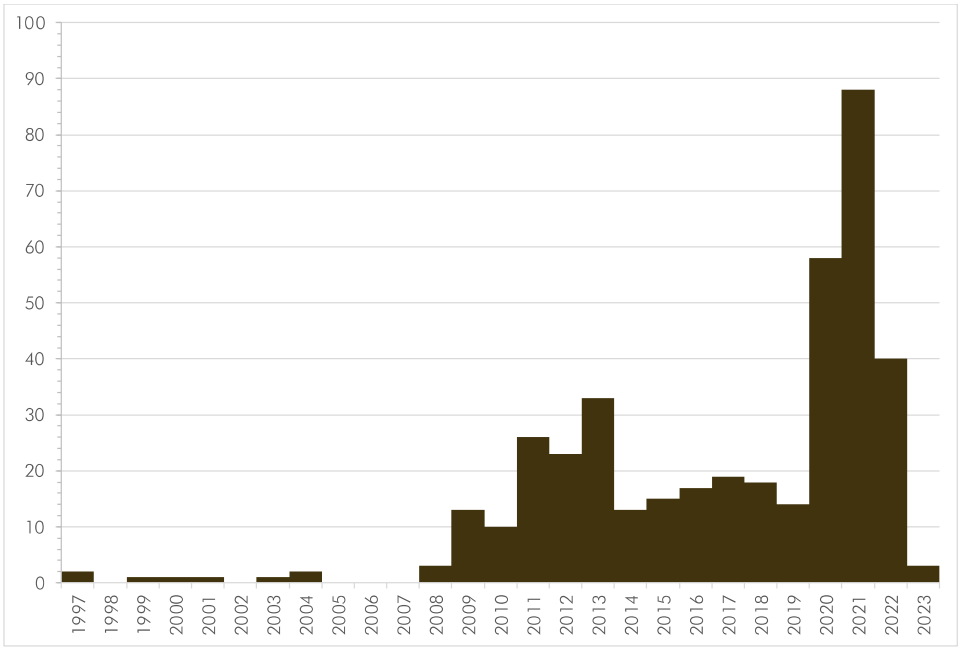

Figure 6. Frequency of journal releases by MDPI.

3. Results

Given the effort involved in setting up a template file for a formal, print-ready manuscript, it was expected that the overwhelming majority of the journal template files would have been created by cloning and modifying a small number of existing templates. The assessment of the rootRSID of the template files for all 402 journal titles listed in the Appendix A, however, found that this was not the case: 212 (51.7%) of all titles carry unique rootRSIDs. Two rootRSIDs account for 29.1% of all template files, representing 62 titles (00D77B5F) and 52 titles (00996419) each. Both were mainly present in journals that began publication in 2020 and 2021. A small number of journals with older commencement dates (2012–2015 period) also have one of these two root RSIDs, which suggests that the templates for these journals were reworked at that time. A further three RSIDs are shared by more than 10 titles each, with another nine RSIDs being shared by two or more titles. Example proof-of-concept phylograms were established for six rootRSIDs (phylograms A–F, Figures 7–12), which account for 42.3% of all journals.

Phylograms A and B reflect examples of successive creations and subsequent modifications of template files. The parent file of Phylogram A, based on root RSID 00A43C68, was created for the *International Journal of Plant Biology*, which in its creation and initial setup underwent 228 discrete edit and save actions. At that point, the file was duplicated for the journal *Alloys* (Figure 7). While the original file underwent another 74 edits until finalized, the template file for *Alloys* was edited six times before another duplicate for the file *Dietetics* was created. While *Alloys* was edited another 14 times until finalization, the template file for *Dietetics* was edited eight times before it was duplicated for the journals *Waste*, *Meteorology*, and *Organoids* (Figure 7). The template for *Waste* was edited three times before it spawned the template for the journal *Entomology*, while the template for *Organoids* became the parent for successive sequences of editing and spawning templates for other journals (Figure 7). Phylogram B, based on root RSID 007B3BA3, shows a similar linear sequence of generational editing and spawning from a single parent after an initial 175 discrete edit and save actions (Figure 8).

Phylogram C, based on root RSID 00A556C8, shows that the template for the journal *Batteries* spawned the template for the *Journal of Cybersecurity and Privacy* after 114 discrete edit and save actions. Both journal templates then went on to spawn subsequent linear sequences of generational editing and spawning (Figure 9).

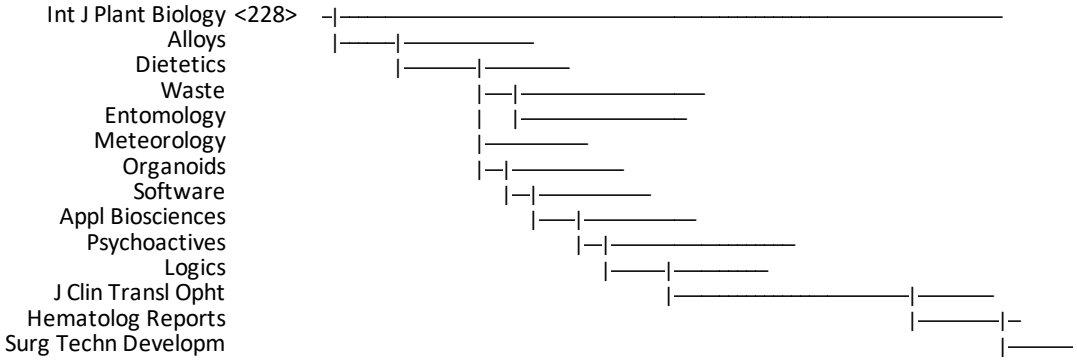

Figure 7. Phylogram (A) for root RSID 00A43C68. All journals commenced in 2022.

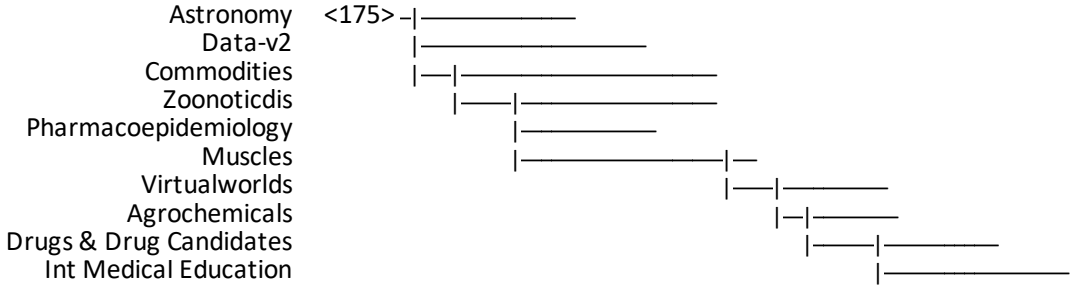

Figure 8. Phylogram (B) for root RSID 007B3BA3. All journals commenced in 2022.

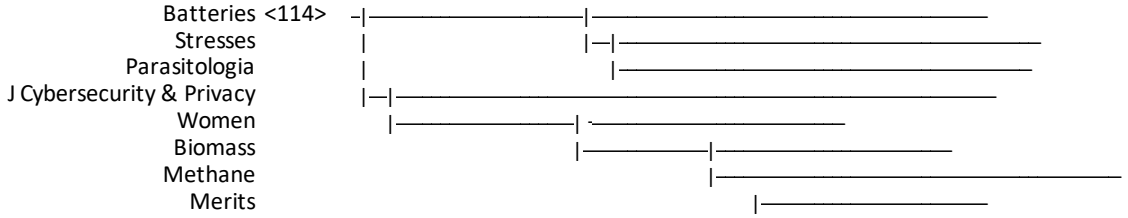

Figure 9. Phylogram (C) for root RSID 00A556C8. Most journals commenced in 2021.

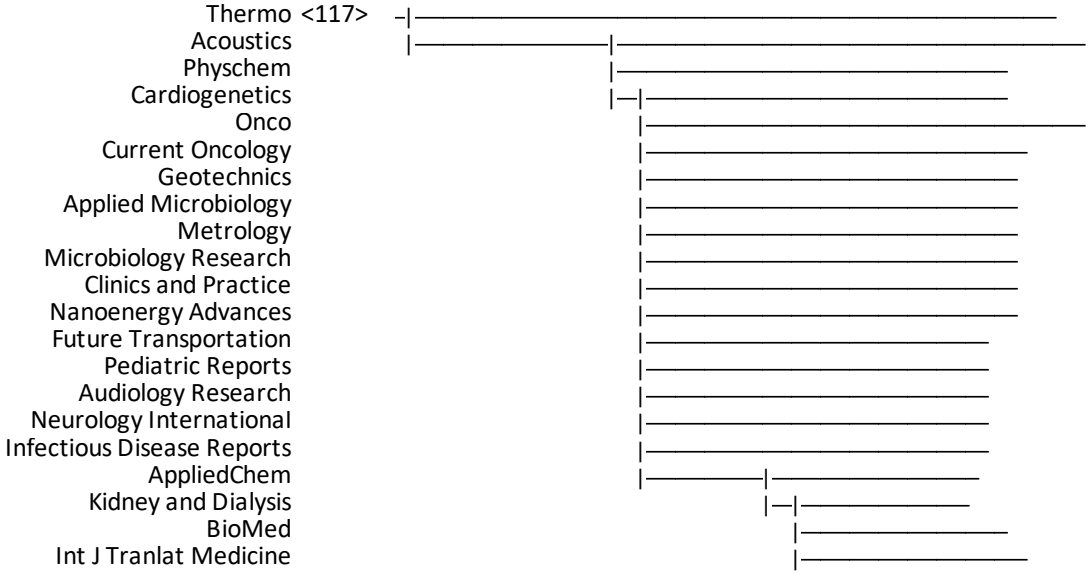

Figure 10. Phylogram (D) for root RSID 00E13154. Most journals commenced in 2021.

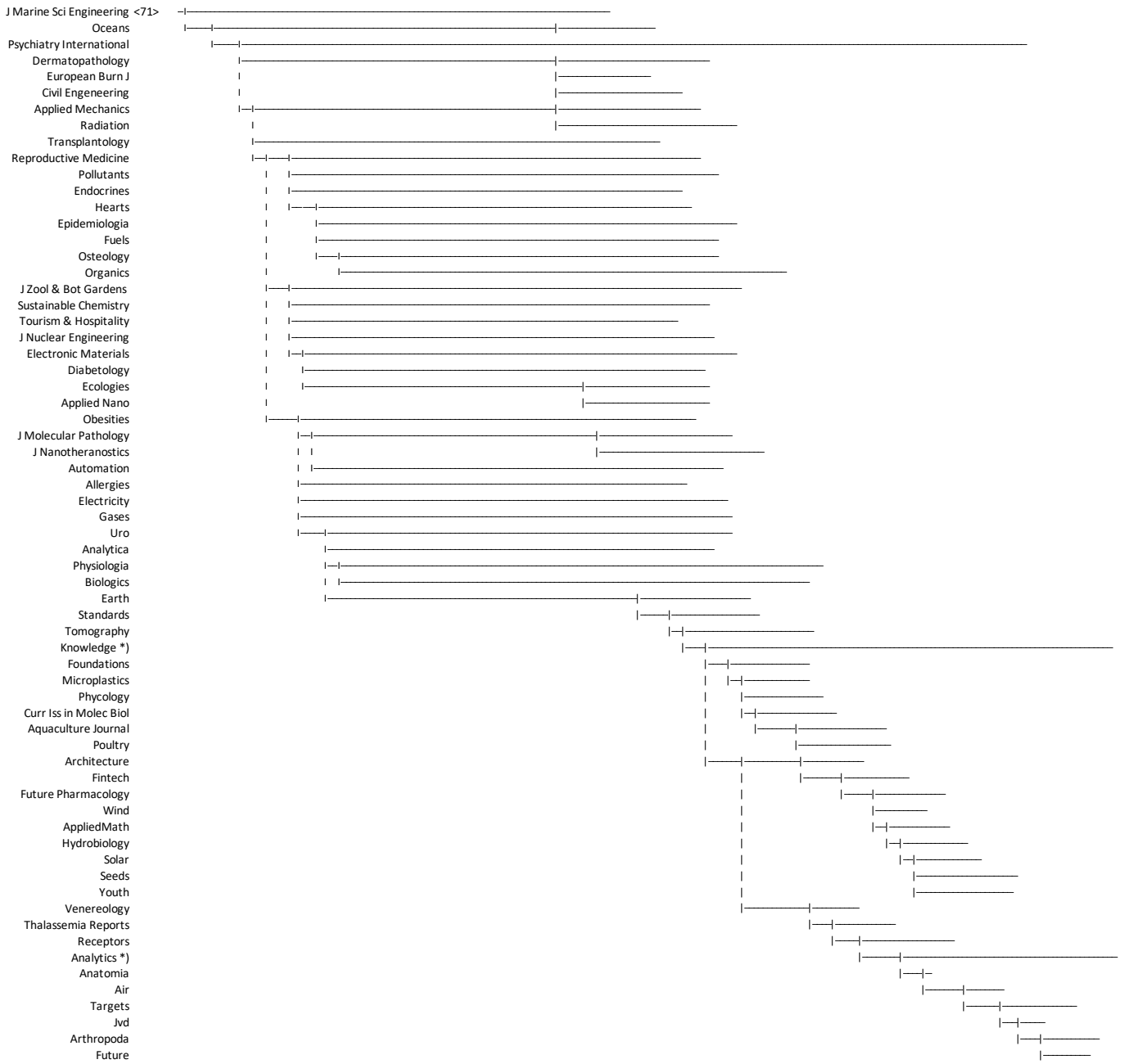

Figure 11. Phylogram (E) for root RSID 00D77B5F. Most journals commenced in 2021. Titles marked with an asterisk (*) indicate journals where the 2023 template file was used.

Phylogram D, based on root RSID 00E13154, shows a very different workflow. The parent was the journal *Thermo*, which, after 117 discrete edits and save actions, spawned the template for the journal *Acoustics* (Figure 10). This template underwent 20 edit and save actions when it spawned the templates for the journals *Psychem* and *Cardiogenetics*. The template for the latter journal was edited twice and then formed the parent for the templates for 14 different journal titles. One of these spawned another, which in turn spawned two more. The 14 journals spawned from *Cardiogenetics* all underwent between 35 and 38 subsequent edit and save events.

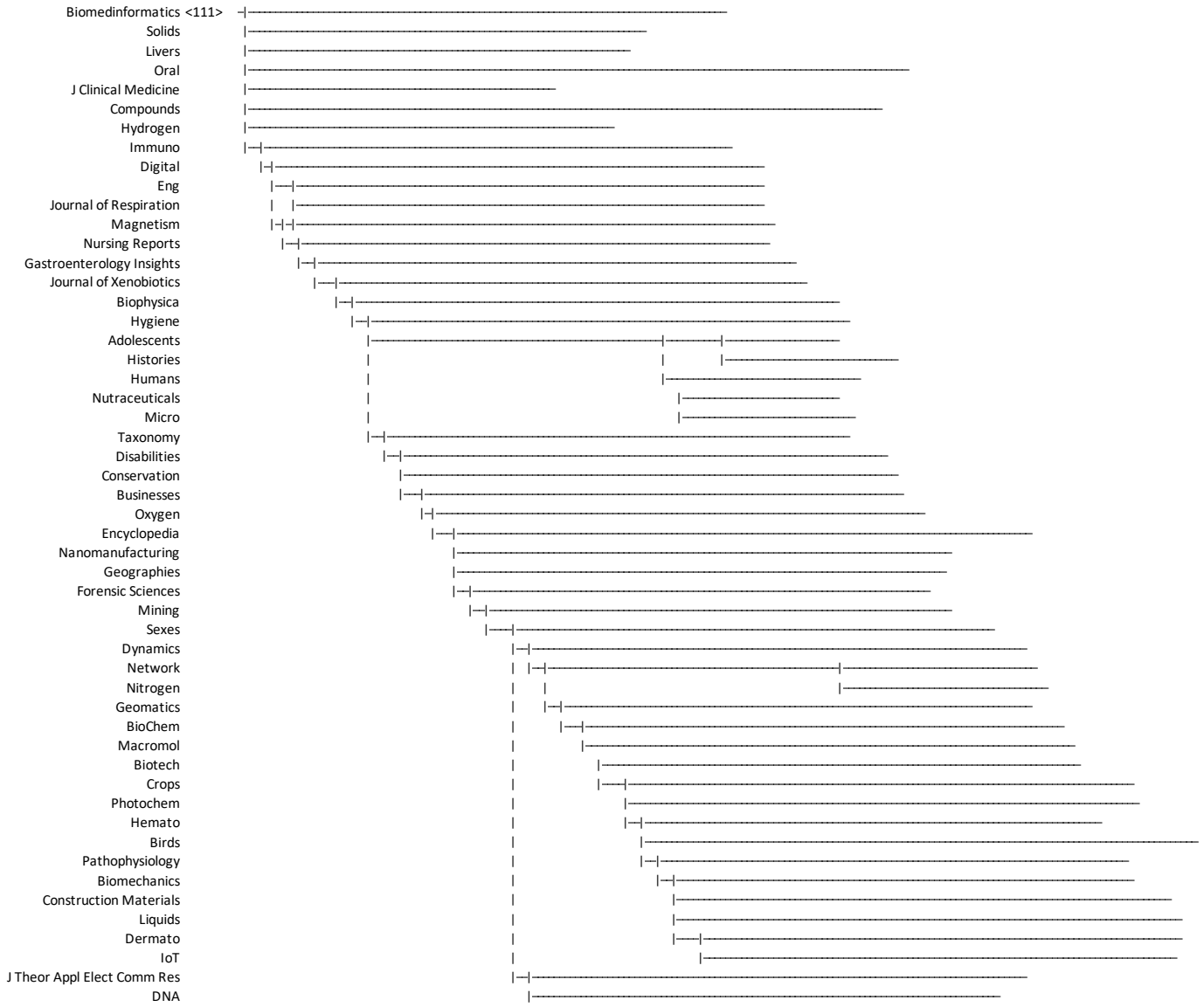

Figure 12. Phylogram (F) for root RSID 00996419. Most journals commenced in 2021.

Phylograms E and F are, in essence, merely more complex variations of the previous phylograms. Phylogram E, based on root RSID 00D77B5F and representing 62 titles, commenced as the template file for the *Journal of Marine Science and Engineering,* which, after 71 discrete edit and save actions, spawned *Oceans*, which, after five edit and save actions, spawned *Psychiatry International* (Figure 11). That journal template spawned two journals, one of which, Applied Mechanics, spawned another two journals, one of which was *Reproductive Medicine*. So far, the pattern follows that shown in Phylograms A–C. The template for *Reproductive Medicine* spawned the *Journal of Zoological and Botanical Gardens* and the journal *Obesities.* Each of these three journals was modified for four to six edit and save actions and then acted as the parent for numerous other journal templates, often created via file copies. *Obesities* acted as the parent for successive templates. All of these underwent considerable individual edits. Of these templates, the one for *Earth* is of interest because, after 68 discrete edit and save actions, this template spawned numerous other templates. Of significance is that, in that section of the genealogy of Phylogram E, the majority of templates are not spawned by large-scale copy file actions but by spawning following short runs of four to six edit and save actions.

Phylogram F, based on root RSID 00996419 and representing 52 titles, starts off as the template for *Biomedinformatics*, which after 111 discrete edit and save actions spawned templates for seven journal titles (Figure 12). One of these, *Immuno*, served as the parent for all other titles. What follows is a long sequence of a journal spawning a new template, which, following a short run of two to five edit and save actions, spawns a further new template. Two journals (*Hygiene* and *Sexes*) spawn two new titles each, which then spawn further templates and create their own descendant branches.

Theoretically, the time stamp data for document creation (dcterms:created) and modification dates (dcterms:modified) contained in the file core.xml (in the "docProps" subfolder) provide the time bracket between the initial creation of the file and the last save event. The document creation dates would therefore allow for an absolute measure of time for the relative file generation sequence established based on the seriation of the RSID numbers. The extracted document creation dates, however, show that the vast majority of template files (90.4%) were created within less than ten minutes of each other (Table 1, 2021-12-27, 04:45 to 04:54 GMT) at an average of 41 files/minute. This can only have been achieved when a series of completed template files, created as .docx files, were opened and saved as.dot templates. This save action will preserve all RSID codes and allocations embedded in the document but will generate a file with a new creation date. This is reinforced by the observation that 94.3% of these files were never modified. Of the remaining 9.6% with creation dates other than 2021-12-27, 82.1% were never modified, suggesting that the low percentage of files that were modified were those where some elements were overlooked in the final QA process.

Table 1. File creation dates and times (GMT) as extracted from the core.xml file.

Time	*n*
2021-04-17 13:20:00	1
2021-12-27 04:45:00	15
2021-12-27 04:46:00	45
2021-12-27 04:47:00	45
2021-12-27 04:48:00	42
2021-12-27 04:49:00	44
2021-12-27 04:50:00	36
2021-12-27 04:51:00	39
2021-12-27 04:52:00	40
2021-12-27 04:53:00	41
2021-12-27 04:54:00	19
2021-12-28 00:43:00	15
2022-06-07 10:01:00	1
2022-06-07 10:03:00	5
2022-10-09 04:38:00	3
2022-10-09 04:39:00	4
2022-10-09 04:40:00	4
2022-10-09 04:41:00	4
2022-12-27 04:35:00	1
2022-12-27 04:38:00	1
Total	405

A small number of templates were finalized on 2022-10-09 between 04:38 and 04:41 (Table 1). Several of these belong to the end section of Phylogram E (Figure 13). While the relative chronology shows that the templates for the last five journals in the genealogy were successively cloned for the template for the journal *Anatomia*, adding the creation dates shows that *Anatomia* and the preceding journals were completed on 27 December 2021, while the templates for the new journals were added on 10 September 2022. At the same time, the template for the journal *Receptors* was also revisited and a new template

saved. At the end of 2022, the template for the journal *Analytics* was revisited, and a new template was also saved.

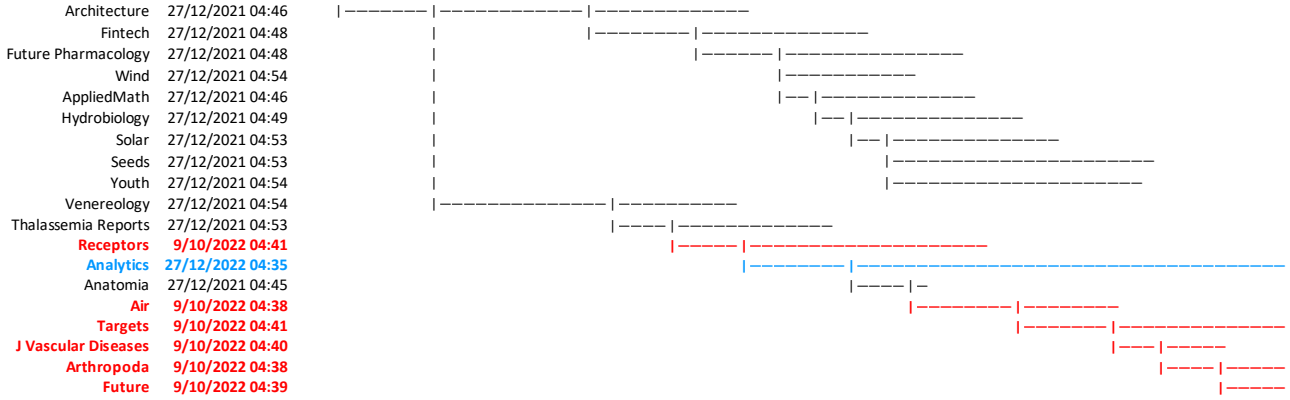

Figure 13. End section of Phylogram E with document creation dates added.

The publication of the journals *Future* and *Journal of Vascular Diseases* started in 2022, while the journals *Air*, *Arthropoda*, and *Targets* formally commenced in 2023. Clearly, however, preparations for this had commenced long before.

Historic Trajectories of Selected Journals

In view of the findings that the majority of journals launched before 2021 tend to exhibit unique rootRSIDs but that a large number of journals launched in 2021 and 2022 used templates that had been created from a small number of rootRSIDs, it was decided to trace back the template files for a select number of journal titles. In order to be able to consider a longer historic trajectory, titles were chosen that had been founded in 2009 and 2010. The choice of journals was restricted by the availability of older versions of template files that had archived by the Wayback Machine. Overall, while most of the journals' main landing pages as well as the "Instructions for Authors" pages were regularly archived, this did not extend to the linked download files. As a result, the availability of past templates was patchy (Table 2).

Table 2. File format of template files 2010–2023 available for selected journals founded in 2009 and 2010.

	Journal							
Year	***Cancers***	***Diversity***	***Information***	***Micromachines***	***Nutrients***	***Remote Sensing***	***Sustainability***	***Water***
2010						doc		
2011						doc		
2012	doc	doc		doc	doc	doc	doc	
2013								
2014	doc	doc				doc		doc
2015	doc	doc	doc	doc	doc docx [29 Sep]	doc	doc	doc
2016		docx	docx		docx	docx	docx	docx
2017		docx	docx		docx	docx	docx	docx
2018	doc	doc	doc	docx [15 Jan] doc [1 May]		doc	doc	
2019								
2020								
2021	doc			docx		doc	doc [4 Feb] docx [27 Sep]	doc [18 Apr] docx [22 Jun]
2022	docx	docx	docx	docx	docx	docx	docx	docx
2023	docx	docx	docx	docx	docx	docx	docx	docx

The early template files of the 2010 to 2015 period were based on the .doc format. This makes sense from a publisher's perspective, as the public adoption of the .docx format of files based on the Office Open XML standard, as introduced by Microsoft in 2007, was slow, and the publisher had to ensure compatibility. From late 2015 on, templates used the XML-based file format. Intriguingly, this format seems to have been abandoned during the second quarter of 2018 with the reintroduction of the .doc format. While there are no archived template files for any of the eight chosen journals for the years 2019 and 2020, template files created for other journals (as retrieved via the Wayback Machine) show that the .doc format was used both in 2019 (*Heritage*) and 2020 (*Nanomaterials*). The rationale for the temporary switchback to .doc format templates remains unclear. Template files changed back to the XML based format during the second quarter of 2021 (Table 2, date supported by other journal titles).

An examination of the documents' XML files for the eight journals (Table 3) showed that the template files before and after the 2018–2021 return to the .doc file format differ significantly. The files from 2021 onwards all have in common that the rootRSID for each title remains the same for the 2021, 2022, and 2023 templates and that the total number of RSIDs listed in the templates increases each year. Clearly, the 2022 XML-based template for a journal was developed by cloning and editing the 2021 journal template, while the 2023 XML based template was developed by cloning and editing the 2022 template. This conforms with the expectations for document genealogies as discussed above.

Table 3. Root RSID, total number of RSIDs listed in the RSID table, and file creation and modification dates (in italics) of the XML files of selected journals founded in 2009 and 2010.

	Journal							
Year	***Cancers***	***Diversity***	***Information***	***Micromachines***	***Nutrients***	***Remote Sensing***	***Sustainability***	***Water***
2015					00486F32 [1345] 2015-09-29			
2016		005835BC [1413] 2016-01-21	006B0BD2 [1414] 2016-01-21		00ED1887 [1372] 2016-05-03	001D1B57 [1396] 2016-01-21	001F4825 [1396] 2016-01-21	00F2211C [1420] 2016-05-03 *2016-11-04*
2017		00740F17 [1432] 2017-01-11	0048362C [1446] 2016-05-03 *2017-06-06*		00ED1887 [1406] 2016-05-03 *2017-09-13*	005D4A69 [1431] 2016-05-03 *2017-06-06*	00E477C2 [1429] 2016-05-03 *2017-06-06*	00F2211C [1442] 2016-05-03 *2017-06-06*
2018				00A657BD [1418] 2018-01-08				
Hiatus								
2021				00AE0E56 [155] 2021-07-23			00CD3203 [153] 2021-07-23	00746129 [153] 2021-04-17
2022	003A55A8 [195] 2022-06-07	007B1618 [171] 2022-06-07	004E6907 [172] 2021-12-27	00AE0E56 [170] 2022-06-07	002549C8 [166] 2021-12-27	0014356F [172] 2022-06-07	00CD3203 [167] 2021-12-27 *2022-01-07*	00746129 [170] 2022-06-07
2023	003A55A8 [207] 2022-06-07	007B1618 [184] 2022-12-27	004E6907 [187] 2022-12-27	00AE0E56 [182] 2022-12-27	002549C8 [182] 2022-12-27	0014356F [184] 2022-12-27	00CD3203 [182] 2022-12-27	00746129 [182] 2022-12-27

All XML-based template files offered in 2016, 2017, and 2018, however, carry different rootRSIDs for each year (except for the journal *Nutrients*) and exhibit significantly larger

numbers of RSIDs (average 1416 RSIDs) than the templates from the 2021–2023 period (average 172 RSIDs). Looking at the file creation dates, numerous files were created on the same date (2016-05-03), suggesting the same QA processes were at work in 2016 as they were in 2022.

4. Discussion

The foregoing section has shown that the combination of rootRSIDs and general RSIDs contained in MS Word XML files can be used for seriation and that phylograms can be derived that represent document genealogies and relative chronological sequences. The phylograms clearly demonstrated whether templates were generated by copying an existing template file several times (e.g., Phylogram D, Figure 10), or whether creation sequences were largely iterative, with new templates spawned following minor edits (e.g., Phylogram F, Figure 12).

Theoretically, the extraction of file creation dates would allow us to place these relative chronological sequences on an absolute time scale, with the caveat (from a digital forensics perspective) that file creation dates can be manipulated. A relative chronological sequence could not be demonstrated for the present case study as 90.4% of all template files were created within less than ten minutes of each other. While the bulk of the data was not suitable for establishing absolute chronologies, it provided an insight into the production process. The data suggests that the individual template files had been cloned, modified, and completed after passing quality assurance. These finalized files were then saved from .docx to the .dot format for use as templates, with the majority of these not undergoing a single further edit. Under "normal" document generation circumstances, such actions would be extremely rare, and absolute chronologies could be developed. The end section of Phylogram E (Figure 13) can serve to support the proof of concept.

The file creation date stamps have limitations, however, as the dcterms:created and dcterms:modified tags of MS Word record the computer time only in full minutes. Thus, they do not allow us to distinguish between a file copy using the operating system (and subsequent renaming), which would have a time stamp that is identical to the original to the second, and a copy made via a "save as" action, which would be likely to deviate by one or more seconds, assuming that the "save as" action occurs immediately after the final save of the parent file. If, however, sufficient time (for the computer clock to move to the next full minute) is allowed to elapse while the parent file is still open, or if that amount of time elapses between the last save (and close) of the parent file and the 'save as' action of the reopened parent file, then the time stamps of the parent and the child files will diverge (verified by testing). The modification date stamp would indicate the time spent modifying the relevant child document.

Since the file data are coded to Z(ulu) time (i.e., GMT), global time zone differences can be ignored, a matter that would come into play when considering documents derived from international collaboration or contract cheating drawing on international tutors.

While the MS Word XML also allows for the document's creator to be identified via the <dc:creator> tag in the core.xml file, all creator IDs were, as expected, set to the corporate MDPI term. Consequently, individual editors cannot be identified from the metadata. Even though it is not possible to assess how many people worked on the creation of the templates represented by Phylograms A–C and attribute the changes to specific authors, the unique pattern of Phylogram D suggests, albeit not conclusively, that the same author duplicated the template for the journal *Cardiogenetics* before editing them one-by-one.

5. Conclusions

Since 2007, Microsoft has implemented the Office Open XML standard in its suite of desktop applications. Given the near monopoly held by MS Word in the corporate, educational, and consumer spaces, this provides opportunities for digital forensics. From a document forensics perspective, the allocation of revision save identifiers (RSIDs) to reflect

a range of textual and formatting edits allows us to understand much of the generation, writing, and editing history of a MS Word document.

As this paper has demonstrated, the allocation of a unique document RSID (rootRSID) to each MS Word file, which persists even when that file is duplicated via "save as" events or via direct file copy in the operating system, forms the basis for document genealogies. Starting with common "ancestor" document content, and thereby the RSID associated with the new content, they will begin to diverge. Subsequent copying will spawn new generations with shared but diverging combinations of RSIDs. A seriation of the shared RSIDs contained in the XML metadata of MS Word files can be used to develop phylograms, which represent the relative genealogical relationship between documents created from the same ancestor file. Differences in document creation dates then allow this relative relationship to be translated into absolute chronological dimensions.

An examination of over 400 template files produced by the publisher MDPI, carried out as a proof-of-concept exercise, validated the theoretical model. Moreover, several of the phylograms, combined with document creation timestamp information, also provided insights into the publishers' template production processes, underscoring the potential of RSIDs for digital forensic inquiry.

The findings of this proof-of-concept exercise have wide-ranging implications. The unique allocation of document RSIDs, coupled with the sequential addition of RSIDs associated with new content, allows for the generation of parent-child relationships between documents. Importantly, these genealogies are methodologically independent of other metadata, in particular time stamps, that are easily manipulated. Where multiple versions exist that exhibit conflicting text, the establishment of document genealogies allows an investigator to understand from which document the deviating file was forked.

Positioning genealogically one file between two others, where the creation dates of all files are known, may confirm or contradict the metadata embedded in the file, thereby potentially highlighting willful manipulation of dates.

Finally, from a future historian's point of view, the ability to establish genealogies of iteratively generated "draft" documents will give insight into policy development and formulation processes, akin to interpreting the strike-outs and file annotations encountered in traditional paper-based archival documents.

Author Contributions: Conceptualization, D.H.R.S.; methodology, D.H.R.S.; software, R.J.S.; formal analysis, D.H.R.S.; data acquisition: D.H.R.S. and R.J.S.; data curation, D.H.R.S.; writing—original draft preparation, D.H.R.S. and R.J.S.; writing—review and editing, D.H.R.S. and R.J.S.; visualization, D.H.R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Journal titles published by MDPI. Titles marked with an asterisk (*) indicate journals where the 2023 template file was used.

Journal	ISSN	Started	Years	Rsid Root	Qty RSID
Acoustics	2624-599X	2019	4	00E13154	184
Actuators	2076-0825	2012	11	00FE3A26	190
Administrative Sciences	2076-3387	2011	12	00A74D3F	195
Adolescents	2673-7051	2021	2	996419	208
Aerospace	2226-4310	2014	9	00577FC8	177

Table A1. *Cont.*

Journal	ISSN	Started	Years	Rsid Root	Qty RSID
Agriculture	2077-0472	2011	12	0038174E	167
AgriEngineering	2624-7402	2019	4	000A28BC	171
Agrochemicals	2813-3145	2022	1	007B3BA3	218
Agronomy	2073-4395	2011	12	00CD22B7	173
AI	2673-2688	2020	3	009039D6	168
Air	2813-4168	2023	0	00D77B5F	234
Algorithms	1999-4893	2008	15	00F87C77	168
Allergies	2313-5786	2021	2	00D77B5F	176
Alloys	2674-063X	2022	1	00A43C68	248
Analytica	2673-4532	2020	3	00D77B5F	182
Analytics *	2813-2203	2022	1	00D77B5F	189
Anatomia	2813-0545	2022	1	00D77B5F	219
Animals	2076-2615	2011	12	00E92783	187
Antibiotics	2079-6382	2012	11	00F23B21	186
Antibodies	2073-4468	2012	11	002D4D3E	163
Antioxidants	2076-3921	2012	11	00471DC3	167
Applied Biosciences	2813-0464	2022	1	00A43C68	260
Applied Mechanics	2673-3161	2020	3	00D77B5F	181
Applied Microbiology	2673-8007	2021	2	00E13154	176
Applied Nano	2673-3501	2020	3	00D77B5F	180
Applied Sciences	2076-3417	2011	12	009039D6	161
Applied System Innovation	2571-5577	2018	5	005F047F	162
AppliedChem	2673-9623	2021	2	00E13154	171
AppliedMath	2673-9909	2021	2	00D77B5F	223
Aquaculture Journal	2673-9496	2021	2	00D77B5F	210
Architecture	2673-8945	2021	2	00D77B5F	207
Arthropoda	2813-3323	2023	0	00D77B5F	252
Arts	2076-0752	2012	11	001427EB	178
Astronomy	2674-0346	2022	1	007B3BA3	191
Atmosphere	2073-4433	2010	13	002E6028	165
Atoms	2218-2004	2013	10	00436D99	173
Audiology Research	2039-4349	2020	3	00E13154	173
Automation	2673-4052	2020	3	00D77B5F	184
Axioms	2075-1680	2012	11	002C4BCD	164
Bacteria	2674-1334	2022	1	00216F54	181
Batteries	2313-0105	2015	8	00A556C8	182
Behavioral Sciences	2076-328X	2011	12	00174C1F	167
Beverages	2306-5710	2015	8	00AA7B35	166
Big Data and Cognitive Computing	2504-2289	2017	6	00B53EA5	165
BioChem	2673-6411	2021	2	996419	242
Bioengineering	2306-5354	2014	9	003709B0	165

Table A1. *Cont.*

Journal	ISSN	Started	Years	Rsid Root	Qty RSID
Biologics	2673-8449	2021	2	00D77B5F	201
Biology	2079-7737	2012	11	674472	191
Biomass	2673-8783	2021	2	00A556C8	177
Biomechanics	2673-7078	2021	2	996419	250
BioMed	2673-8430	2021	2	00E13154	173
Biomedicines	2227-9059	2013	10	0094171B	167
BioMedInformatics	2673-7426	2021	2	996419	199
Biomimetics	2313-7673	2016	7	00E72C35	164
Biomolecules	2218-273X	2011	12	000343DC	163
Biophysica	2673-4125	2021	2	996419	213
Biosensors	2079-6374	2011	12	00F4755F	164
BioTech	2673-6284	2012	11	996419	244
Birds	2673-6004	2020	3	996419	264
Brain Sciences	2076-3425	2011	12	009A6647	167
Buildings	2075-5309	2011	12	00CD5F86	172
Businesses	2673-7116	2021	2	996419	221
C	2311-5629	2015	8	00BE2C00	168
Cancers	2072-6694	2009	14	003A55A8	191
Cardiogenetics	2035-8148	2011	12	00E13154	175
Catalysts	2073-4344	2011	12	8.07E+39	185
Cells	2073-4409	2012	11	00657CE6	169
Ceramics	2571-6131	2018	5	007D2828	160
Challenges	2078-1547	2010	13	00F27B93	186
ChemEngineering	2305-7084	2017	6	0005626D	166
Chemistry	2624-8549	2019	4	004B1B47	176
Chemosensors	2227-9040	2013	10	00677D23	170
Children	2227-9067	2014	9	240000000	165
Chips	2674-0729	2022	1	00E61C2B	243
CivilEng	2673-4109	2020	3	00D77B5F	178
Clean Technologies	2571-8797	2019	4	00AC6D50	167
Climate	2225-1154	2013	10	000A28BC	167
Clinical and Translational Neuroscience	2514-183X	2017	6	00E41938	216
Clinics and Practice	2039-7283	2021	2	00E13154	175
Clocks & Sleep	2624-5175	2019	4	00646BFC	165
Coasts	2673-964X	2021	2	00E33B02	209
Coatings	2079-6412	2011	12	325048	165
Colloids and Interfaces	2504-5377	2017	6	00CB52FE	164
Colorants	2079-6447	2022	1	007D153F	212
Commodities	2813-2432	2022	1	007B3BA3	204
Compounds	2673-6918	2021	2	996419	229
Computation	2079-3197	2013	10	000E2F14	170

Table A1. *Cont.*

Journal	ISSN	Started	Years	Rsid Root	Qty RSID
Computer Sciences & Mathematics Forum	2813-0324	2022	1	00FC4602	195
Computers	2073-431X	2012	11	000662F8	181
Condensed Matter	2410-3896	2016	7	00825EA8	173
Conservation	2673-7159	2021	2	996419	221
Construction Materials	2673-7108	2021	2	996419	257
Corrosion and Materials Degradation	2624-5558	2020	3	00BC2AFA	170
Cosmetics	2079-9284	2014	9	00B0701F	165
COVID	2673-8112	2021	2	454992	224
Crops	2673-7655	2021	2	996419	253
Cryptography	2410-387X	2017	6	00EA797C	164
Crystals	2073-4352	2011	12	945307	168
Current Issues in Molecular Biology *	1467-3045	2021	2	00D77B5F	200
Current Oncology	1718-7729	2021	2	00E13154	177
Dairy	2624-862X	2020	3	225164	129
Data	2306-5729	2016	7	006934C1	158
Dentistry Journal	2304-6767	2013	10	00014B7A	165
Dermato	2673-6179	2021	2	996419	258
Dermatopathology	2296-3529	2014	9	00D77B5F	185
Designs	2411-9660	2017	6	00E17A65	168
Diabetology	2673-4540	2020	3	00D77B5F	180
Diagnostics	2075-4418	2011	12	00D85F80	165
Dietetics	2674-0311	2022	1	00A43C68	251
Digital	2673-6470	2021	2	996419	204
Disabilities	2673-7272	2021	2	996419	219
Diseases	2079-9721	2013	10	009A11EC	161
Diversity	1424-2818	2009	14	007B1618	167
DNA	2673-8856	2021	2	996419	233
Drones	2504-446X	2017	6	001F6ABF	166
Drugs and Drug Candidates	2813-2998	2022	1	007B3BA3	227
Dynamics	2673-8716	2021	2	996419	238
Earth	2673-4834	2020	3	00D77B5F	188
Ecologies	2673-4133	2020	3	00D77B5F	179
Econometrics	2225-1146	2013	10	000746AD	169
Economies	2227-7099	2013	10	00101D43	171
Education Sciences	2227-7102	2011	12	006F6234	167
Electricity	2673-4826	2020	3	00D77B5F	185
Electrochem	2673-3293	2020	3	0047331F	163
Electronic Materials	2673-3978	2020	3	00D77B5F	187
Electronics	2079-9292	2012	11	003A4C09	164
Encyclopedia	2673-8392	2021	2	996419	243
Endocrines	2673-396X	2020	3	00D77B5F	177

Table A1. *Cont.*

Journal	ISSN	Started	Years	Rsid Root	Qty RSID
Energies	1996-1073	2008	15	00B402AB	167
Eng	2673-4117	2020	3	996419	203
Entomology	2813-0502	2022	1	00A43C68	262
Entropy	1099-4300	1999	24	004B1B47	168
Environments	2076-3298	2014	9	003F7165	168
Epidemiologia	2673-3986	2020	3	00D77B5F	186
Epigenomes	2075-4655	2017	6	00C95365	172
European Burn Journal	2673-1991	2020	3	00D77B5F	172
European Journal of Investigation in Health, Psychology and Education	2254-9625	2020	3	0046726B	164
Fermentation	2311-5637	2015	8	005523F9	159
Fibers	2079-6439	2013	10	00247A62	164
FinTech	2674-1032	2022	1	00D77B5F	216
Fire	2571-6255	2018	5	584230	161
Fishes	2410-3888	2016	7	673164	167
Fluids	2311-5521	2016	7	931549	170
Foods	2304-8158	2012	11	243653	166
Forecasting	2571-9394	2019	4	004665D0	165
Forensic Sciences	2673-6756	2021	2	996419	223
Forests	1999-4907	2010	13	664321	174
Foundations	2673-9321	2021	2	00D77B5F	196
Fractal and Fractional	2504-3110	2017	6	00B47BA5	160
Fuels	2673-3994	2020	3	00D77B5F	182
Future	2813-2882	2022	1	00D77B5F	249
Future Internet	1999-5903	2009	14	1.56E+84	171
Future Pharmacology	2673-9879	2021	2	00D77B5F	223
Future Transportation	2673-7590	2021	2	00E13154	173
Galaxies	2075-4434	2013	10	001473F0	172
Games	2073-4336	2010	13	00CE4D7C	171
Gases	2673-5628	2021	2	00D77B5F	186
Gastroenterology Insights	2036-7422	2020	3	996419	207
Gastrointestinal Disorders	2624-5647	2019	4	342350	170
Gels	2310-2861	2015	8	0055583E	174
Genealogy	2313-5778	2017	6	636716	174
Genes	2073-4425	2010	13	006E06CC	163
Geographies	2673-7086	2021	2	996419	227
GeoHazards	2624-795X	2020	3	150594	112
Geomatics	2673-7418	2021	2	996419	237
Geosciences	2076-3263	2011	12	009356FF	166
Geotechnics	2673-7094	2021	2	00E13154	176
Geriatrics	2308-3417	2016	7	007D4A75	157

Table A1. *Cont.*

Journal	ISSN	Started	Years	Rsid Root	Qty RSID
Healthcare	2227-9032	2013	10	00CF475D	163
Hearts	2673-3846	2020	3	00D77B5F	177
Hemato	2673-6357	2020	3	996419	246
Hematology Reports	2038-8330	2022	1	00A43C68	292
Heritage	2571-9408	2018	5	006B0F68	167
Histories	2409-9252	2021	2	996419	221
Horticulturae	2311-7524	2015	8	00E75ADD	168
Humanities	2076-0787	2012	11	006D690C	180
Humans	2673-9461	2021	2	996419	215
Hydrobiology	2673-9917	2022	1	00D77B5F	226
Hydrogen	2673-4141	2020	3	996419	178
Hydrology	2306-5338	2014	9	00D82A38	162
Hygiene	2673-947X	2021	2	996419	214
Immuno	2673-5601	2021	2	996419	199
Infectious Disease Reports	2036-7449	2020	3	00E13154	173
Informatics	2227-9709	2014	9	215152	163
Information	2078-2489	2010	13	004E6907	170
Infrastructures	2412-3811	2016	7	006C4820	168
Inorganics	2304-6740	2013	10	008C6560	190
Insects	2075-4450	2010	13	00AC7940	185
Instruments	2410-390X	2017	6	008055B1	164
International Journal of Environmental Research and Public Health	1660-4601	2004	19	0046726B	161
International Journal of Financial Studies	2227-7072	2013	10	005A723C	174
International Journal of Molecular Sciences	1422-0067	2000	23	00646BFC	172
International Journal of Neonatal Screening	2409-515X	2015	8	00E62EDF	163
International Journal of Plant Biology	2037-0164	2022	1	00A43C68	301
International Journal of Translational Medicine	2673-8937	2021	2	00E13154	175
International Journal of Turbomachinery, Propulsion and Power	2504-186X	2016	7	008F66DB	167
International Medical Education	2813-141X	2022	1	007B3BA3	234
Inventions	2411-5134	2016	7	008A6172	160
IoT	2624-831X	2020	3	996419	257
ISPRS International Journal of Geo-Information	2220-9964	2012	11	00F00EA7	176
J	2571-8800	2018	5	00F56BF0	166
Journal of Ageing and Longevity	2673-9259	2021	2	00D6249B	205
Journal of Cardiovascular Development and Disease	2308-3425	2014	9	0060168B	165
Journal of Clinical & Translational Ophthalmology	2813-1053	2023	0	00A43C68	290

Table A1. *Cont.*

Journal	ISSN	Started	Years	Rsid Root	Qty RSID
Journal of Clinical Medicine	2077-0383	2012	11	996419	167
Journal of Composites Science	2504-477X	2017	6	00C93333	170
Journal of Cybersecurity and Privacy	2624-800X	2021	2	00A556C8	184
Journal of Developmental Biology	2221-3759	2013	10	00D25EEF	161
Journal of Functional Biomaterials	2079-4983	2010	13	5750000000	168
Journal of Functional Morphology and Kinesiology	2411-5142	2016	7	00B52CEC	163
Journal of Fungi	2309-608X	2015	8	00EB1C2F	166
Journal of Imaging	2313-433X	2015	8	00D02B8A	169
Journal of Intelligence	2079-3200	2013	10	00F34BFF	185
Journal of Low Power Electronics and Applications	2079-9268	2011	12	008C2503	179
Journal of Manufacturing and Materials Processing	2504-4494	2017	6	6730000000	164
Journal of Marine Science and Engineering	2077-1312	2013	10	00D77B5F	165
Journal of Molecular Pathology	2673-5261	2020	3	00D77B5F	185
Journal of Nanotheranostics	2624-845X	2020	3	00D77B5F	192
Journal of Nuclear Engineering	2673-4362	2020	3	00D77B5F	183
Journal of Open Innovation: Technology, Market, and Complexity	2199-8531	2015	8	00A63295	172
Journal of Otorhinolaryngology, Hearing and Balance Medicine	2504-463X	2018	5	00915AFC	163
Journal of Personalized Medicine	2075-4426	2011	12	00B3376D	171
Journal of Respiration	2673-527X	2021	2	996419	203
Journal of Risk and Financial Management	1911-8074	2013	10	005574DA	175
Journal of Sensor and Actuator Networks	2224-2708	2012	11	00395DA4	162
Journal of Theoretical and Applied Electronic Commerce Research	0718-1876	2021	2	996419	238
Journal of Vascular Diseases	2813-2475	2022	1	00D77B5F	241
Journal of Xenobiotics	2039-4713	2020	3	996419	208
Journal of Zoological and Botanical Gardens	2673-5636	2020	3	00D77B5F	189
Journalism and Media	2673-5172	2020	3	0091161E	186
Kidney and Dialysis	2673-8236	2021	2	00E13154	169
Knowledge *	2673-9585	2021	2	00D77B5F	189
Land	2073-445X	2012	11	00C7668C	177
Languages	2226-471X	2016	7	002356C3	171
Laws	2075-471X	2012	11	00D5113E	183
Life	2075-1729	2011	12	007B260B	169
Liquids	2673-8015	2021	2	996419	259
Literature	2410-9789	2021	2	00E078F8	180
Livers	2673-4389	2021	2	996419	181
Logics	2813-0405	2022	1	00A43C68	266
Logistics	2305-6290	2017	6	606654	174

Table A1. *Cont.*

Journal	ISSN	Started	Years	Rsid Root	Qty RSID
Lubricants	2075-4442	2013	10	0015698D	172
Machine Learning and Knowledge Extraction	2504-4990	2019	4	00374A4D	163
Machines	2075-1702	2013	10	00680A13	166
Macromol	2673-6209	2021	2	996419	243
Magnetism	2673-8724	2021	2	996419	204
Magnetochemistry	2312-7481	2015	8	5.27E+92	179
Marine Drugs	1660-3397	2003	20	004E3F3D	182
Materials	1996-1944	2008	15	00F42E7B	165
Mathematical and Computational Applications	2297-8747	2016	7	006D3275	162
Mathematics	2227-7390	2013	10	463466	170
Medical Sciences	2076-3271	2013	10	009B04B5	163
Medicina	1648-9144	2018	5	00C12256	1615
Medicines	2305-6320	2014	9	00A47112	165
Membranes	2077-0375	2011	12	00F407C6	169
Merits	2673-8104	2021	2	00A556C8	180
Metabolites	2218-1989	2011	12	003D32E5	193
Metals	2075-4701	2011	12	00F0091E	166
Meteorology	2674-0494	2022	1	00A43C68	252
Methane	2674-0389	2022	1	00A556C8	195
Methods and Protocols	2409-9279	2018	5	006E1031	175
Metrology	2673-8244	2021	2	00E13154	176
Micro	2673-8023	2021	2	996419	213
Microbiology Research	2036-7481	2020	3	00E13154	175
Micromachines	2072-666X	2010	13	00AE0E56	166
Microorganisms	2076-2607	2013	10	002B1611	166
Microplastics	2673-8929	2022	1	00D77B5F	195
Minerals	2075-163X	2011	12	002E61F0	166
Mining	2673-6489	2021	2	996419	226
Modelling	2673-3951	2020	3	00B44C52	172
Molbank	1422-8599	1997	26	9.454E+12	177
Molecules	1420-3049	1997	26	00605BAA	182
Multimodal Technologies and Interaction	2414-4088	2017	6	00934DC6	165
Muscles	2813-0413	2022	1	007B3BA3	206
Nanoenergy Advances	2673-706X	2021	2	00E13154	176
Nanomanufacturing	2673-687X	2021	2	996419	228
Nanomaterials	2079-4991	2011	12	196430	168
Network	2673-8732	2021	2	996419	238
Neuroglia	2571-6980	2018	5	001F5D82	166
Neurology International	2035-8377	2020	3	00E13154	173
NeuroSci	2673-4087	2020	3	009039D6	171

Table A1. *Cont.*

Journal	ISSN	Started	Years	Rsid Root	Qty RSID
Nitrogen	2504-3129	2020	3	996419	240
Non-Coding RNA	2311-553X	2015	8	00111D09	177
Nursing Reports	2039-4403	2020	3	996419	203
Nutraceuticals	1661-3821	2021	2	996419	210
Nutrients	2072-6643	2009	14	002549C8	165
Obesities	2673-4168	2021	2	00D77B5F	179
Oceans	2673-1924	2020	3	00D77B5F	174
Onco	2673-7523	2021	2	00E13154	183
Optics	2673-3269	2020	3	009039D6	166
Oral	2673-6373	2021	2	996419	233
Organics	2673-401X	2020	3	00D77B5F	196
Organoids	2674-1172	2022	1	00A43C68	255
Osteology	2673-4036	2021	2	00D77B5F	181
Oxygen	2673-9801	2021	2	996419	224
Parasitologia	2673-6772	2021	2	00A556C8	187
Particles	2571-712X	2018	5	00A9552F	165
Pathogens	2076-0817	2012	11	00AC4D97	179
Pathophysiology	1873-149X	2020	3	996419	250
Pediatric Reports	2036-7503	2020	3	00E13154	173
Pharmaceuticals	1424-8247	2004	19	533599	179
Pharmaceutics	1999-4923	2009	14	00B415BD	165
Pharmacoepidemiology	2813-0618	2022	1	007B3BA3	197
Pharmacy	2226-4787	2013	10	00EE36B5	159
Philosophies	2409-9287	2016	7	000C7DEF	175
Photochem	2673-7256	2021	2	996419	254
Photonics	2304-6732	2014	9	00405BD9	167
Phycology	2673-9410	2021	2	00D77B5F	198
Physchem	2673-7167	2021	2	00E13154	176
Physics	2624-8174	2019	4	004B1B47	171
Physiologia	2673-9488	2021	2	00D77B5F	204
Plants	2223-7747	2012	11	005442C6	183
Plasma	2571-6182	2018	5	006060C0	165
Pollutants	2673-4672	2021	2	00D77B5F	184
Polymers	2073-4360	2009	14	00606C84	166
Polysaccharides	2673-4176	2020	3	00090C72	164
Poultry	2674-1164	2022	1	00D77B5F	211
Powders	2674-0516	2022	1	0045614A	213
Processes	2227-9717	2013	10	008712FD	166
Prosthesis	2673-1592	2019	4	00AC4D97	184
Proteomes	2227-7382	2013	10	006B242D	159
Psych	2624-8611	2019	4	00CD3203	166

Table A1. *Cont.*

Journal	ISSN	Started	Years	Rsid Root	Qty RSID
Psychiatry International	2673-5318	2020	3	00D77B5F	188
Psychoactives	2813-1851	2022	1	00A43C68	270
Publications	2304-6775	2013	10	00A900D9	162
Quantum Beam Science	2412-382X	2017	6	00C746A7	162
Quantum Reports	2624-960X	2019	4	0029228C	175
Quaternary	2571-550X	2018	5	150594	168
Radiation	2673-592X	2021	2	00D77B5F	189
Reactions	2624-781X	2020	3	009D2E2A	164
Receptors	2813-2564	2022	1	00D77B5F	224
Recycling	2313-4321	2016	7	00624D3C	173
Religions	2077-1444	2010	13	0085252F	182
Remote Sensing	2072-4292	2009	14	0014356F	168
Reports	2571-841X	2018	5	000913B6	164
Reproductive Medicine	2673-3897	2020	3	00D77B5F	180
Resources	2079-9276	2012	11	00753FC0	168
Rheumato	2674-0621	2021	2	00572C89	211
Risks	2227-9091	2013	10	00DC6882	178
Robotics	2218-6581	2012	11	005F4D95	162
Ruminants	2673-933X	2021	2	002A4806	217
Safety	2313-576X	2015	8	00A124B4	171
Sci	2413-4155	2019	4	00A70054	182
Scientia Pharmaceutica	2218-0532	2016	7	00A415DE	159
Seeds	2674-1024	2022	1	00D77B5F	236
Sensors	1424-8220	2001	22	0028224B	167
Separations	2297-8739	2014	9	4.5E+12	168
Sexes	2411-5118	2020	3	996419	233
Signals	2624-6120	2020	3	0028224B	166
Sinusitis	2673-351X	2016	7	004A282F	163
Smart Cities	2624-6511	2018	5	00CD3203	162
Social Sciences	2076-0760	2012	11	005D784D	159
Societies	2075-4698	2011	12	009450AE	178
Software	2674-113X	2022	1	00A43C68	257
Soil Systems	2571-8789	2017	6	008542B8	160
Solar	2673-9941	2021	2	00D77B5F	227
Solids	2673-6497	2020	3	996419	184
Sports	2075-4663	2013	10	00FF1CF7	167
Standards	2305-6703	2021	2	00D77B5F	189
Stats	2571-905X	2018	5	00273EC6	162
Stresses	2673-7140	2021	2	00A556C8	188
Surfaces	2571-9637	2018	5	00680A22	162
Surgeries	2673-4095	2020	3	006F2B1D	166

Table A1. *Cont.*

Journal	ISSN	Started	Years	Rsid Root	Qty RSID
Surgical Techniques Development	2038-9582	2022	1	00A43C68	298
Sustainability	2071-1050	2009	14	00CD3203	166
Sustainable Chemistry	2673-4079	2020	3	00D77B5F	182
Symmetry	2073-8994	2009	14	00706D40	166
SynBio	2674-0583	2022	1	00DB7B35	187
Systems	2079-8954	2013	10	173241	168
Targets	2813-3137	2023	0	00D77B5F	248
Taxonomy	2673-6500	2021	2	996419	213
Technologies	2227-7080	2013	10	00C94DDE	162
Telecom	2673-4001	2020	3	009039D6	177
Textiles	2673-7248	2021	2	00CD647B	247
Thalassemia Reports	2039-4365	2022	1	00D77B5F	210
Thermo	2673-7264	2021	2	00E13154	183
Tomography	2379-139X	2015	8	00D77B5F	199
Tourism and Hospitality	2673-5768	2020	3	00D77B5F	175
Toxics	2305-6304	2013	10	00FE4A36	163
Toxins	2072-6651	2009	14	005A5E33	189
Transplantology	2673-3943	2020	3	00D77B5F	173
Trauma Care	2673-866X	2021	2	00A47533	209
Tropical Medicine and Infectious Disease	2414-6366	2016	7	00FE3088	159
Universe	2218-1997	2015	8	00E07EBB	168
Urban Science	2413-8851	2017	6	001D13CB	162
Uro	2673-4397	2021	2	00D77B5F	185
Vaccines	2076-393X	2013	10	003D727B	168
Vehicles	2624-8921	2019	4	686168	166
Venereology	2674-0710	2022	1	00D77B5F	206
Veterinary Sciences	2306-7381	2014	9	004F05D7	165
Vibration	2571-631X	2018	5	00A93006	165
Virtual Worlds	2813-2084	2022	1	007B3BA3	218
Viruses	1999-4915	2009	14	001428F9	171
Vision	2411-5150	2017	6	00094BD9	166
Waste	2813-0391	2022	1	00A43C68	264
Water	2073-4441	2009	14	746129	166
Wind	2674-032X	2021	2	00D77B5F	219
Women	2673-4184	2021	2	00A556C8	166
World	2673-4060	2020	3	009039D6	175
World Electric Vehicle Journal	2032-6653	2018	5	00CD3203	172
Youth	2673-995X	2021	2	00D77B5F	234
Zoonotic Diseases	2813-0227	2021	2	007B3BA3	203

References

1. Lee, H.; Lee, H.-W. Hidden message detection in MS-Word file by analyzing abnormal file structure. In Proceedings of the 2020 International Conference on Green and Human Information Technology (ICGHIT), Hanoi, Vietnam, 5–7 February 2020; pp. 54–57.
2. Lee, H.; Lee, H.-W. Forgery Detection Mechanism with Abnormal Structure Analysis on Office Open XML based MS-Word File. *Int. J. Adv. Smart Converg.* **2019**, *8*, 47–57.
3. Dasare, A.; Dhore, M. Secured Approach for Hiding Data in MS Word Document Using MCDFF. In Proceedings of the 2015 International Conference on Computing Communication Control and Automation, Pune, India, 26–27 February 2015; pp. 296–300.
4. Stojanov, I.; Mileva, A.; Stojanovic, I. A new property coding in text steganography of Microsoft Word documents. In Proceedings of the Securware 2014: The Eighth International Conference on Emerging Security Information, Systems and Technologies, Lisbon, Portugal, 16–20 November 2014.
5. Park, B.; Park, J.; Lee, S. Data concealment and detection in Microsoft Office 2007 files. *Digit. Investig.* **2009**, *5*, 104–114. [CrossRef]
6. Hong, K.; Cho, J.; Kim, S.; Kim, J. OOXML 문서에 대한 향상된 데이터 은닉 및 탐지 방법 [Improved Data Concealing and Detecting Methods for OOXML Document]. *J. Korea Inst. Inf. Secur. Cryptol.* **2017**, *27*, 489–499. (In Korean) [CrossRef]
7. Raffay, M.A. Data Hiding and Detection in Office Open XML (OOXML) Documents. Ph.D. Thesis, University of Ontario Institute of Technology, Oshawa, ON, Canada, 2011.
8. Mahlow, C.; Piotrowski, M. Academic writing and publishing beyond documents. In Proceedings of the 22nd ACM Symposium on Document Engineering, San Jose, CA, USA, 20–23 September 2022; pp. 1–4.
9. ECMA. Office Open XML File Formats—Fundamentals and Markup Language Reference [ECMA-376-1:2016]. Available online: http://www.ecma-international.org/publications/standards/Ecma-376.htm (accessed on 15 November 2021).
10. Spennemann, D.H.R.; Singh, C. The generation of revision identifier (rsid) numbers in MS Word—Implications for document analysis. *Int. J. Digit. curation under review*.
11. Ries, T. "die geräte klüger als ihre besitzer": Philologische Durchblicke hinter die Schreibszene des Graphical User Interface. *Editio* **2010**, *24*, 149–199. [CrossRef]
12. Ries, T. The rationale of the born-digital dossier génétique: Digital forensics and the writing process: With examples from the Thomas Kling Archive. *Digit. Scholarsh. Humanit.* **2018**, *33*, 391–424. [CrossRef]
13. Lebrave, J.-L. Computer forensics: La critique génétique et l'écriture numérique. *Genes. Manuscr.–Rech.–Invent.* **2011**, *33*, 137–147. [CrossRef]
14. Crasson, A.; Lebrave, J.-L.; Pedrazzi, J. Le «siliscrit» de Jacques Derrida. Exploration d'une archive nativement numérique. *Genes. Manuscr.–Rech.–Invent.* **2019**, *49*, 7–12. [CrossRef]