Publication history: a double DOI-based method to store and/or monitor information about published and corrected academic literature

Jaime A. Teixeira da Silva[1], Serhii Nazarovets[2]

[1] Independent researcher, P. O. Box 7, Ikenobe 3011-2, Kagawa-ken, 761-0799, Japan; jaimetex@yahoo.com
[2] Borys Grinchenko Kyiv University, 18/2 Bulvarno-Kudriavska Str., 04053 Kyiv, Ukraine; serhii.nazarovets@gmail.com

Abstract

The status of published literature can change at any time in its history following publication, although current existing structures in academic publishing, despite the existence of some robust tools – such as the digital object identifier (DOI) – to record those changes, appear to be insufficiently robust, or their use is inconsistent or inefficient, to deal with multiple corrections of the literature. In this paper, an information storage method and corrective measure or tool is proposed – the "publication history" – that considers the full history and background of a paper's publication record. The "publication history" is adjusted to record changes to a paper over time, and is thus a "live" document, i.e., always open to modification and updating. In theory, the "publication history" would accommodate, in a single document (PDF and HTML formats), information about pre-publication (e.g., preprints) and post-publication events, including submission, resubmission, acceptance dates, handling editors, peer review format, corrections, expressions of concern and retractions. The "publication history" employs two DOIs, one for the paper and one for any and all edits, to document these changes. Our proposal offers one possible solution to fortify the integrity of peer review and the publication process pre- and post-peer review. The double DOI-based "publication history" can be applied to any document.

Keywords: DDPH (double-DOI "publication history"); open access; open peer review; post-publication peer review; predatory publishing; preprint; retraction

This is a preprint of an article published by University of Toronto Press (UTP) in *Journal of Scholarly Publishing* available online at https://doi.org/10.3138/jsp-2017-0017

Sources of fluidity of information in an academic paper that may limit or restrict its usefulness

The academic literature, like research, is mostly dynamic, but its existence should be indefinite, to preserve the published record. Consequently, ideas evolve, some may become outdated and fall to the wayside in terms of use, and others yet might be disproved as new theories emerge. An academic idea, like a publication, is thus not terminal (Popper, 2002). It does not – and should not – cease to exist, but its use may be limited through disuse, correction (Blachowicz, 1995), elimination or retraction (Hotez, 2020). Uncited literature (Liang et al., 2015) can still be read, and is thus not necessarily useless merely because it is uncited. From the perspective of neoliberal academia (Whelan, 2015), an end-life – including through retraction – might be assigned to works, publications, ideas, and/or theories. Even if a concept ceases to be popular and be cited or if a paper loses moral validity and credibility, through retraction, the paper itself (i.e., the document) should not cease to exist, much like a historical document, so that its merits and demerits can always be appreciated even beyond death of the current generation that was involved in the curation of that document (authors, peer reviewers, editors, publishers). Simply because a paper has been retracted does not necessarily mean that it cannot, or should not, be cited or used, and there are multiple reasons, some academically valid, others invalid or unscholarly, as to why retracted papers continue to be cited (Teixeira da Silva and Bornemann-Cimenti, 2017). An appreciation of these concepts is central to this paper.

The birth of an academic paper might be imperfect because an academic paper is an artificial construct, so there is a risk that its fundaments may be flawed, or could be subjected to changes and challenges post-publication. An academic paper is "constructed" by humans, except where artificial intelligence may be used to create a fake author or paper (de Vries, 2020), or a paper generated by software to create nonesense texts (Cabanac and Labbé, 2021), so because it carries inherent biases, flaws, and subjective interpretations, an academic paper can be subjective[1]. Using current classification systems, the most subjective of academic papers are editorials (editors' subjectivity), letters to the editor (authors' subjectivity, usually in response to a published paper), or opinion papers or commentaries (authors' subjectivity with regard to general themes). To reduce – but never eliminate – subjectivity, a core aspect of the reproducibility movement (Forstmeier et al., 2017), artificial parameters and constructs are added, for example in original research, to try and make it more objective (i.e., less influenced by human biases). However, since objectivity's robustness might also be subjective (Williams, 2015), a paper that attempts to eliminate subjectivity through the use of objectivity may introduce a double layer of subjectivity. The existence of subjectivity and objectivity in an academic paper suggests that its construct can be subjected to challenge and change at any time of its existence, even post-publication.

A paper's subjectivity may be fortified by introducing robust evidence rather than relying exclusively on opinion. This could involve greater reliance on the published peer-reviewed literature to assert facts or support statements, or might include more robust analyses to support claims (Abramo and D'Angelo, 2011). Lack of reproducibility grounded in potentially false findings (Ioannidis, 2005) that underlie the "replication crisis" (Hillary and Medaglia, 2020) may be caused by insufficiently strong measures to assess significance (Hardwicke and Ioannidis, 2019). Thus an academic paper might be flawed at various levels (thematically, methodologically, etc.), might contain other flaws, inaccuracies, spin or hype (Millar et al., 2020). Although peer review is typically the stage at which such flaws are detected, and corrected, in the publication of an academic paper, preprints have allowed critique and correction to take place prior to formal peer review, while post-publication peer review (PPPR) is a corrective measure in academic publishing following the publication of a peer-reviewed paper (Teixeira da Silva et al., 2017) that detects and tackles the flaws described above (Teixeira da Silva and Dobránszki, 2015), including misconduct and fraud (Campos-Varela et al., 2020).

Are current corrective measures for the academic literature sufficiently robust and flexible?

As described above, "artificial" (subjective) corrals have been placed around how academic papers should be published, and additional "artificial" (subjective) corrals have also been put in place to describe how they could or should be corrected when faced with a flaw or a challenge. These are broadly categorized as errata,

[1] Artificial intelligence can also be used for positive purposes. For example, in 2019, Springer Nature published what may very well be the first machine-generated book: https://group.springernature.com/gp/group/media/press-releases/springer-nature-machine-generated-book/16590134 (April 2, 2019; last accessed: December 12, 2021)

expressions of concern (EoCs) and retractions (Teixeira da Silva, 2020a). However, even this scope of corrective measures is still insufficient to effectively adapt to an evolving literature and, as explained next in ideas and using select examples, requires reform to deal with the "artificial" (subjective) corrals that have been put in place. An analysis of 72,069 papers in three journals and about three authors suggested that PPPR resulted in greater correction of the literature than negative citations to flawed papers (Bordignon, 2020). Compounding these limitations, there is also still a negative stigma and punitive ideology associated with the correction of the literature (Shashok and Matarese, 2018), especially retractions (Vuong, 2020; Teixeira da Silva and Al-Khatib, 2021).

An emergent problem is the lack of assignment of a digital object identifier (DOI) to an EoC or its euphemistic disguise as an "editor's note" that may be restricted exclusively to the HTML version of a paper, perhaps as a way to reduce the negative stigmatization of the journal or authors, but ultimately eroding the integrity of the paper's publication history because it cannot be uniquely, or independently, verified (Teixeira da Silva, 2022). In theory, such HTML versions and notices or addenda can be modified, manipulated or erased at any time, without academia or the public even being aware, while papers may occasionally be removed or withdrawn, without any transparent reason provided by the publisher, i.e., silent or stealth retractions, despite the existence of a DOI (Teixeira da Silva, 2016). The current mind-set is "blind trust in the publisher". If some of these events were in fact to happen post-publication, then they might be interpreted as a "predatory" publishing characteristic because they involve the opaque manipulation of the published record, or the version of record (VoR), thereby misleading academics and the public (Teixeira da Silva et al., 2019). Such changes to the published VoR are external to the National Information Standards Organization (NISO)[2] framework of VoRs (Morgan, 2008), and may be difficult to distinguish from publishing practices that also involve multiple VoRs, and tweaks and edits to even the final published version, as well as between a preprint and a published version, paginated in print, or online (Krell, 2015). NISO is still struggling to effectively implement policies that can address "disorder in scholarly communication", at least related to COVID-19 (Schneider et al., 2021), despite improvements to document flow and exchange (Alves, 2020).

Bibliometric integrity is compromised when changes to HTML or PDF versions are made opaquely, i.e., without a formal record and absent transparency about those changes. When the integrity of corrective bibliometric components is compromised, so too is the wider integrity of the academic literature.[3] A paper might be retracted then replaced, or might pass through a series of corrective measures, such as the Mehra et al. (2020) paper, which suffered a correction, an EoC, and then finally a retraction.

A practical solution is thus needed that can accommodate a wide range of corrective measures, both in terms of number and scope. In this paper, a solution is proposed, referred to as the double-DOI "publication history" (DDPH). Prior to discussing DDPH, we briefly review the strengths and weaknesses of the tool that is fundamental to this bibliometric solution, namely the DOI.

Strengths and weaknesses of the DOI, and competing data storage and integrity systems

Modern scientific communication is closely linked to the dynamic digital environment, which requires permanent identifiers to ensure reliable access to electronic resources. The DOI[4] is a permanent, unique, alphanumeric string that identifies objects and is widely used for the identification of academic publications and related information, and has become an important bibliometric tool in librarianship and for academic and scientific communities. The DOI system is administered by a non-profit organization, the International DOI Foundation (IDF). The DOI rapidly gained global prominence and recognition, and its first application was in 2000, and approved as an ISO standard in 2010.[5] DOI registration agencies, such as Crossref or the China National Knowledge Infrastructure (CNKI), maintain a central DOI directory that acts as an intermediary between users and depositors of content. The DOI syntax consists of a prefix that identifies the clients of the

[2] Our proposal could be helpful for organizations like NISO involved in publication record integrity: https://www.niso.org/publications/standards (last accessed: December 12, 2021).

[3] The issues we raise, including the outdated and inflexible choices proposed by NISO, have practical applications, for example the advice offered to authors and editors regarding literature correction, by Taylor & Francis: https://authorservices.taylorandfrancis.com/publishing-your-research/after-publication/corrections-to-published-articles/# (last accessed: December 12, 2021).

[4] Even though the abbreviation DOI is a registered trademark (i.e., DOI®), for simplification, it is simply referred to as DOI in this paper.

[5] https://www.doi.org/doi_handbook/7_IDF.html (last updated: March 8, 2018; last accessed: December 12, 2021)

DOI registration agencies and a suffix that helps users to identify and find relevant objects. After assigning a persistent identifier to an object (e.g., a preprint or scientific paper), the depositor registers its current location on the Internet in the DOI system. Once registration is completed, the DOI remains unchanged forever. If an object changes its location, the depositor must update the DOI record by specifying a new URL (Paskin, 2005). Dysfunctional URLs (e.g., 404 error) that can be found in academic papers, such as their reference lists, a phenomenon referred to as reference rot (Teixeira da Silva, 2021), suggests that a portion of depositors are not conforming to this requirement. Thus, a user request for a specific digital object is sent to the DOI server, the server finds the DOI record and the current URL of the associated object, and returns to the user's browser information on how to access the full text and other related information.

Each DOI is associated with metadata related to the content of the object, and DOI registration agencies ensure that depositors provide appropriate metadata for each registered DOI. The DOI identifier must be associated with metadata that describes what is being identified. The DOI standard sets a certain minimum set of metadata, supported by an XML Schema, and also assists registration agencies in creating their own schemes[6]. Using machine or human interfaces when registering a DOI, registration agency clients may be depositing only a basic set of metadata according to the type of deposit, which includes fields such as dates, names, authors and online locations. However, to unleash the full potential of persistent identifiers, the scientific community should use rich metadata and provide as much information about the deposit as possible, to make the content easier to find and thus increase the chances of its further detection, use and citation. For example, the Participation Reports[7] service from Crossref allows anybody to independently check the richness of the metadata of publishers that use this service, in particular to find the percentage of content items that contain lists of links in their metadata, annotations, ORCID IDs, funder registry IDs, license URLs, funding award numbers, and more. Today, leading scientific publishers prefer to work with the Crossref registration agency (Lammey, 2020), which is a major source of scientific data for publishers, scholars, librarians and funders. Crossref had over 13,000 members from 120 countries, its database contained more than 106 million records, and it is expanding at an average rate of 11% per year (Hendricks et al., 2020).

DOIs are of critical importance in building reliable links between scientific documents and help make scientific data accessible, interoperable, findable, and reusable in accordance with FAIR principles (Wilkinson et al., 2016; Hauschke et al., 2021). Additionally, for users' convenience, DOI content can be used at any level of publication detail, DOI-related metadata can specify different document formats (DeRisi et al., 2003), or link a paper's translation to its original version (Birukou and Skalaban, 2020). Curiously, even though preprints and their final peer-reviewed versions have independent DOIs, there is still a wide gap in a connection between both versions of papers (Cabanac et al., 2021), suggesting that the current DOI-based bibliometric tool for librarianship is deficient. Apart from that, DOIs are successfully used in a library and integrity systems, they have considerably expanded users' access to open content not subscribed by libraries, have increased the usage of electronic resources, made it easier for users to find relevant links to full-text publications (Wang, 2007), are an integral part of the access to black or pirate open access content, such as through Sci-Hub (Greshake, 2017), and are also crucial for altmetrics tools (Roemer and Borchardt, 2015).

Since DOIs are not being assigned to all scientific documents (typically old publications might not have them), publishers do not always promptly update metadata, and only limited DOI-related metadata of scientific objects is open for reuse. Some journals have only a portion of their content linked to DOI, leaving DOI-unlined literature bibliometrically isolated and unlinked in the wider information stream (Teixeira da Silva, 2020b). These issues create a number of barriers to effective scientific communication. Initiatives such as the Initiative for Open Abstracts and the Initiative for Open Citations (Peroni and Shotton, 2020) are attempting to improve this situation, but the openness of scientific metadata is still sub-optimal. Crossref currently suggests the assignment of a separate DOI for different versions of a paper.[8]

However, despite its good intentions, in theory, this might become a laborious and messy process as a

[6] https://www.doi.org/doi_handbook/4_Data_Model.html#4.3 (last updated: March 21, 2017; last accessed: December 12, 2021)
[7] https://www.crossref.org/members/prep/ (last accessed: December 12, 2021)
[8] https://www.crossref.org/education/crossmark/version-control-corrections-and-retractions/#00327 (last updated: April 8, 2020; last accessed: December 12, 2021)

paper evolves, potentially involving multiple (or too many) DOIs. As an extreme example, an EoC might be corrected, so a DOI is needed for the erratum of an EoC, and if the correction (or original paper) are re-corrected, then a separate DOI is needed, and other DOIs for any possible future corrections, a separate DOI is needed if the paper is retracted, and yet another DOI if the retraction is corrected. In a recent example of an IOP Press journal, a single retraction notice was issued for approximately four dozen papers, and a single DOI was assigned to all retractions (*Journal of Intelligent & Fuzzy Systems*, 2021). There is one serious *faux pas* in that retraction notice, namely that each retracted paper should carry a separate and independent DOI for its own retraction notice. Since DOIs are not free, it is possible that this "clustered" DOI approach may be a cost-saving strategy. Furthermore, careful scrutiny of the retraction notice indicates that many of the Chinese authors' names are incorrectly indexed, i.e., first names are listed and not family names, suggesting that this retraction notice needs major corrections to accurately represent academics' names. For these reasons, and based on these very practical limitations and problems, we propose in this paper the DDPH as one possible simplified mechanism to deal with such fluctuations in bibliometric content.

Introducing "publication history" as a literary corrective measure: applications, advantages, limitations, and accountability

The DDPH involves the assignment of a single DOI (DOI1) to a paper and a separate DOI (DOI2) to its history (Table 1). The DDPH is DOI-based because, more than any other currently available tool to digitally track a document, the DOI is widely used, particularly in academic publishing (Boudry and Chartron, 2017). In the DDPH, a paper can be a preprint or a peer-reviewed document. Each new version of a preprint can be assigned a suffix to indicate that version.[9]

For example, at Zenodo, not only is each document assigned an individual DOI, so too is each version and/or upload. However, we question the efficiency and sustainability of this model. Not only does each DOI carry a cost[10] (which is currently born by CERN, which hosts and manages Zenodo), in cases where there are multiple versions, an excessive number of documents can make it difficult for readers and users to navigate which version they should cite. As one example, Vergoulis et al. (2021a), which is a dataset, carries one DOI (10.5281/zenodo.3723281), the website[11] claims that there are 78 versions, but the latest version is indicated as being version #70 (December 5, 2021), and all indicated versions[12] have different DOIs. Moreover, on the same page, the preprint is indicated with two separate citations, the first to Zenodo (Vergoulis et al., 2021a), and the second to *bioRxiv*, which is a preprint server (Vergoulis et al., 2021b). On that Zenodo page, it is indicated that Vergoulis et al. (2021a) is a supplement of version 2 of the Vergoulis et al. (2021b) *bioRxiv* preprint. However, the indicated DOI leads to version 4 of the preprint (see other examples of preprint versioning next). When one traces to the *bioRxiv* preprint, the reader learns that the preprint has been published as a peer-reviewed paper (Vergoulis et al., 2021c), information that cannot be gleaned anywhere on the Zenodo website. The DOI versioning used at Zenodo[13] for documents, including preprints, would be incompatible with the system used by MDPI's preprint.org, as indicated in an example below. This indicates that even though DOIs have been used to trace different versions of a paper from the preprint stage to the peer-reviewed paper stage, in the case of Vergoulis et al., this involves three different websites, including two independent preprint servers and a single publisher, i.e., information is ineffectively linked. Moreover, there is no document that centrally records all events that occurred in the transition of the preprint(s) to the peer-reviewed version, which is where the DDPH system we propose carries value.

Despite the existence of some models that employ a "system" similar to what we are proposing, there are caveats, limitations, and distinct problems, as explained above for Zenodo, and as explained next, with other examples. For example, Teixeira da Silva (2020), a preprint at preprint.org, has a separate DOI for each version of the preprint, as indicated by the suffix "v1" and "v2" that follow the "base" DOI

[9] Common practice, at least at Zenodo and F1000

[10] See limitations section for a greater discussion.

[11] https://zenodo.org/record/5759169

[12] Even though the DOI indicated as the "base" for citation to all versions is 10.5281/zenodo.3723281, the DOI for the latest version 70 is 10.5281/zenodo.5759169, the DOI for version 69 is 10.5281/zenodo.5725900, version 68 is 10.5281/zenodo.5718035, etc. all of which are disjunct numbers with no logical or related sequence.

[13] https://help.zenodo.org/#versioning

10.20944/preprints201708.0029.[14] In contrast, for example, Siler (2020), which was published in Wiley's *Journal of the Association for Information Science and Technology* (*JASIST*), has 12 versions of a preprint on OSF's *SocArXiv* preprint server, but only a single DOI is assigned as a "base" (i.e., 10.31235/osf.io/6r274), with each version being downloadable as a separate document (PDF or Word).[15] This is problematic for several reasons: 1) each version of the preprint does not carry its own independent DOI, but should, for clarity and bibliometric distinction; 2) the publication history for the preprint's DOI at Crossref does not link to the final published paper in the *JASIST*, and *vice versa*, while the preprint entry shows zero information about the preprint's history, for example that it has 12 versions[16]; 3) the *JASIST* website[17] does not indicate to readers and the public that the peer-reviewed paper was preceded by 12 versions of a preprint, nor does it indicate where this important publication history can be found; 4) the final published version does not indicate that it was preprinted. This case indicates that the current bibliometric system, at least in terms of the link between preprints and published versions in peer-reviewed journals, has very serious deficiencies.

In our DDPH proposal, any publishing-related action or version is recorded in the DOI2, which is in a dynamic or "live" state of updates. For example, if a submission is made to a preprint server, this is recorded, and if edits are made to the preprint, new versions are published, and any change, including dates, are recorded in the DOI2 (Table 1). The numbering of the DOI1 and DOI2 are independent of each other. We propose that Crossref manage DDPH, but that the publisher (or preprint server) that is handling the paper at that moment in its history is responsible for updating the DOI2, as a centralized measure to ensure accountability and responsibility, with financial fines handed out to non-compliant Crossref members or participants.

The DDPH can be quickly applied to many specialized applications:

1. citation and reference managers that allow to import/export, store and manage bibliographic information (e.g., Zotero, JabRef, Mendeley, Endnote);
2. library discovery systems that provide a single search window to segregated open content and resources subscribed by a library (e.g., ProQuest-ExLibris Primo, ProQuest-ExLibris Summon, EBSCO Discovery Service, WorldCat Discovery);
3. software for constructing co-authorship and citation-based networks (e.g., VOSviewer, CiteSpace);
4. as a data source for altmetrics services (e.g., Altmetric, Impactstory, Plum Analytics).

The most evident envisioned advantages of DDPH are:

1. practical use by all participants and stakeholders of the academic publishing process either for confirmation, verification, or evaluation: authors, peer reviewers, publishers, librarians, funders, service and infrastructure providers;
2. fortification of a document's history, centralization of information, and thus transparency through the use of a popular, persistent, unique and actionable tool, the DOI;
3. potential stimulation of the expanded use of DOI to the Chinese literature where the DOI has been weakly adopted (Wang et al., 2018);
4. the downstream fortification of the validity and reliability of a citation or reference (Mayernik and Maull, 2017);
5. as a result of the above advantages, the ability to independently trace if the VoR was manipulated, or not, which may be useful for ethics investigations where a fine-scale appreciation of the historical evolution of a paper might be needed;
6. an open/public tool that allows authors to hold editors and publishers accountable for any manipulative or predatory practices, and *vice versa*;
7. despite the existence of a wealth of data identification systems, some more elaborate than others (Lee and Stvilia, 2014), to the authors' knowledge, none has yet been able to effectively detect and record changes in VoR, if at all, nor is there a single approach today to support all uses of dynamic data citation

[14] v1: http://doi.org/10.20944/preprints201708.0029.v1; v2: http://doi.org/10.20944/preprints201708.0029.v2
[15] https://osf.io/preprints/socarxiv/6r274/
[16] https://search.crossref.org/?q=10.31235%2Fosf.io%2F6r274&from_ui=yes; https://search.crossref.org/?from_ui=yes&q=10.1002/asi.24339
[17] https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24339

(Bellini, 2019).

8. the use of this tool for HTML content, such as publishers' guidelines, ethics guidelines such as by COPE [18] or the ICMJE[19], journal's instructions for authors, lists of editors on editor boards, etc. (in essence, any formal information associated with a publishing entity, e.g., journal or publisher).
9. an analysis of the information about the most common reasons for making changes to documents is likely to help improve the editorial process and allow editors to avoid common mistakes.

The most obvious envisioned limitations are:

1. the inability to apply this tool to previously published papers for which no digital record exists, or to retroactively apply DOI2 to an already published paper without a digital fingerprint but based on what exists in the public domain and in databases; this may reduce the reliability and veracity of that paper's DOI2, thereby reducing its use;
2. the IDF protects all intellectual property rights related to the DOI system, and supports the development and promotion of the DOI system, so any improvements to the DOI system are the exclusive prerogative of the IDF;
3. the reliance on a third party or private firm, or centralized body (e.g., IDF, CNKI, Crossref, etc.) to manage DDPH; the majority of the scientific community and stakeholders would need to support this initiative, and any support of the DDPH must ensure constant uninterrupted access to data through an API, so that stated benefits can truly work;
4. related to 2, structural and financial sustainability, and funding for long-term use.
5. this system on its own might not be sufficiently robust for recording the historical development of source codes, computer or software, or even genetic codes, which might require a more open source network that also relies on the integration of identifiers for digital objects and digital identifiers of objects, such as the DOI (Di Cosmo et al., 2020; Katz et al., 2020);
6. the technical and financial support of the "publication history" (DDPH) will surely require additional publishing costs that might increase APCs of open access journals.
7. DOIs are not free. For example, at Crossref, metadata depositors must first become members, so there are membership fees, and then there are also fees for each DOI[20].
8. accountability and transparency may be challenging. Is a "carrot and stick" approach employed to ensure compliance, and how are infractions dealt with? For example, should members be incurred a financial fine if they are not DDPH-compliant?

Alternatives or supplements to the DDHP: blockchain and non-fungible tokens

A good alternative to effectively tracking the background of a published article may be for an academic publisher to use blockchain. A blockchain is a list of records (blocks) stored on many independent computers that contain timestamps and transaction data that are linked together by cryptographic tools. Since each block must contain information about the previous block, these records cannot be changed or deleted without interfering with the work of all other blocks (van Rossum, 2018; Wang and Zhao, 2021). Consequently, only new blocks can be added. This would provide a layer of records "security" to our DDPH proposal.

Their stability, transparency and decentralization make blockchain technology attractive and promising for use in scientific communication (Bellini, 2019). An author's manuscript can be considered as a digital asset that is transferred (transaction) between different participants in the scientific publication process, and all details of the asset transfers and actions agreed upon by participants will be securely recorded in 'smart contracts'. A promising way to use blockchain features in scientific communication are developments based on the idea of using digital shortages and a unique subclass of digital assets, fungible token (FTs) (Kosmarski and Gordiychuk, 2020), and developments that offer to use another subclass of digital assets, non-fungible tokens (NFTs), that can abstract the complex properties of scientific objects and be used as persistent identifiers (Wang and Zhao, 2021).

However, today the legal status of smart contracts in many countries remains unclear and they are not

[18] https://publicationethics.org/guidance/Guidelines
[19] Recommendations: http://www.icmje.org/recommendations/; disclosure of interest: http://www.icmje.org/disclosure-of-interest/
[20] https://www.crossref.org/fees/#member-fees (last updated: March 15, 2019; last accessed: December 12, 2021)

yet legally binding. The scientific community tends to be characterized by certain conservatism, so it might still be at the beginning of the road to embrace revolutionary changes (van Rossum, 2018) that encompass blockchain, FTs, and NFTs. Instead, our proposed DDPH based on the DOI system that is already accepted by the scientific community, is a simple but effective way to bridge current gaps. In addition, the DDPH is equally suitable for gold open access journals (Kosmarski and Gordiychuk, 2020), diamond open access journals, as well as traditional subscription journals.

Integration of the "publication history" into existing flow models, indexing agencies and platforms

Crossref allows its clients to specify the relationship between different objects, as well as manage changes in a document, file or dataset. Relationships are specified in the object metadata for a particular DOI and can be specified and updated.[21] It may be possible to use existing Crossref capabilities to implement "publication history" (DDPH). Ideally, the list of Crossref Relationship types should be expanded and add a new type that accomodates "publication history" should be included. However, one of the available types can also be used to specify a link to the publication history, such as the Comment type, as indicated by the code denoted next:

```
<program xmlns="https://www.crossref.org/relations.xsd">
<related_item>
<description>Publication history</description>
<intra_work_relation relationship-type="isCommentOn" identifier-type="doi">
10.1234/12345678</intra_work_relation>
</related_item>
</program>
```

Scientific publishers, in addition to receiving a DOI1 for articles, can also receive a DOI2 for DDPH and indicate in DOI1 the connection with DOI2, and all changes in the publication must be promptly (within days or at most a week or two) recorded by the publisher in its DDPH file. With the Crossref REST API, an article's DDPH can be easily integrated on journal websites, references and citation databases, scientific literature aggregator platforms, and display DOI2 links for DDPH on the publication page. In this way, it will be possible for both humans and computer algorithms to quickly access the full "publication history" of a preprint, pee-reviewed paper or any document that is relevant to academic publishing.

Conclusions and recommendations

The academic literature is currently in a very fluid and dynamic state of flux, whether this be a constructive and developmental process (for example preprints to peer-reviewed papers), or a deconstructive process such as retractions. As this fluidity increases, so too does the volume of information and documents. Even though the DOI is the most reliable tool to document and map information and documents, it is not used by all stakeholders, nor by "fringe" publishing parties such as predatory journals/publishers or low-cost publishing operations. There are also ample problems and limitations associated with inter-DOI links or bridges between documents, or the lack of digital records for essential aspects of the publication parameters such as peer review reports in open peer review, journals' instructions for authors, or editor board constituency, aspects that our proposal hypothetically has the ability to address. There are several dominant players in the publishing ecosystem that have ample infrastructure and resources to implement DDPH, but this would require an appreciation of its strengths and weaknesses, participation by all stakeholders, testing, and implementation with an appropriate verification protocol to ensure accountability and transparency. We agree with Couldry and Hepp (2016) that nowadays, the autonomy of social life is increasingly subject to the imperatives of technological systems, which are determined by the commercial interests of their developers, and this potentially dangerous trend can be seen in the current system of scientific communication. However, in our opinion, if anyone can see this threat, counter it, and take responsibility for doing the critical work of constructing reality, it is the scientific community. Thus, it is scholars who must initiate the process of integrating the "history of publications" into current publishing models, without waiting for a commercial player to offer their vision for solving this problem.

[21] Relationships between different research objects: https://www.crossref.org/education/content-registration/structural-metadata/relationships/ (last updated: April 8, 2020; last accessed: December 12, 2021)

Conflicts of interest

The authors declare no conflicts of interest relevant to this topic.

Authors' contributions

The authors contributed equally to the intellectual discussion underlying this paper, literature exploration, writing, reviews and editing, and accept responsibility for the content and interpretation.

References

Abramo, G., D'Angelo, C.A. (2011). Evaluating research: from informed peer review to bibliometrics. *Scientometrics* 87(3): 499-514. https://doi.org/10.1007/s11192-011-0352-7

Alves, T. (2020). Manuscript exchange common approach: Why we need it, what is it, and what's next? *Science Editor* 43(4), 109-112. https://doi.org/10.36591/SE-D-4304-109

Bellini, E. (2019). A blockchain based trusted persistent identifier system for big data in science. *Foundations of Computing and Decision Sciences* 44(4): 351–377. https://doi.org/10.2478/fcds-2019-0018

Birukou, A.A., Skalaban, A.B. (2020). Technical aspects of publication in several languages – how to link digital object identifiers (DOIs). *Science Editor and Publisher* 5(1): 29–39 (in Russian with English abstract). https://doi.org/10.24069/2542-0267-2020-1-29-39

Blachowicz, J. (1995). Elimination, correction and Popper's evolutionary epistemology. *International Studies in the Philosophy of Science* 9(1): 5-17. https://doi.org/10.1080/02698599508573503

Bordignon, F. (2020). Self-correction of science: a comparative study of negative citations and post-publication peer review. *Scientometrics* 124(2), 1225–1239. https://doi.org/10.1007/s11192-020-03536-z

Boudry, C., Chartron, G. (2017). Availability of digital object identifiers in publications archived by PubMed. *Scientometrics* 110(3): 1453–1469. https://doi.org/10.1007/s11192-016-2225-6

Cabanac, G., Labbé, C. (2021). Prevalence of nonsensical algorithmically generated papers in the scientific literature. *Learned Publishing* 72(12): 1461–1476. http://doi.org/10.1002/asi.24495

Cabanac, G., Oikonomidi, T., Boutron, I. (2021). Day-to-day discovery of preprint-publication links. *Scientometrics*, *126*(6), 5285–5304. https://doi.org/10.1007/s11192-021-03900-7

Campos-Varela, I., Villaverde-Castañeda, R., Ruano-Raviña, A. (2020). Retraction of publications: a study of biomedical journals retracting publications based on impact factor and journal category. *Gaceta Sanitaria*, *34*(5), 430-434. https://doi.org/10.1016/j.gaceta.2019.05.008

Couldry, N., Hepp, A. (2016). The Mediated Construction of Reality. Cambridge: Polity Press.

de Vries, K. (2020). You never fake alone. Creative AI in action. *Information, Communication & Society* 23(14): 2110-2127. https://doi.org/10.1080/1369118X.2020.1754877

DeRisi, S., Kennison, R., Twyman, N. (2003). The what and why's of DOIs. *PLoS Biology* 1(2): E57. https://doi.org/10.1371/journal.pbio.0000057

Di Cosmo, R., Gruenpeter, M., Zacchiroli, S. (2020). Referencing source code artifacts: a separate concern in software citation. *Computing in Science & Engineering* 22(2): 33-43. https://doi.org/10.1109/MCSE.2019.2963148

Forstmeier, W., Wagenmakers, E.J., Parker, T.H. (2017). Detecting and avoiding likely false-positive findings - a practical guide. *Biological Reviews of the Cambridge Philosophical Society* 92(4): 1941–1968. https://doi.org/10.1111/brv.12315

Greshake B. (2017). Looking into Pandora's Box: The content of *Sci-Hub* and its usage. *F1000Research* 6: 541. https://doi.org/10.12688/f1000research.11366.1

Hardwicke, T.E., Ioannidis, J.P.A. (2019). Petitions in scientific argumentation: Dissecting the request to retire statistical significance. *European Journal of Clinical Investigation* 49(10): e13162. https://doi.org/10.1111/eci.13162

Hauschke, C., Nazarovets, S., Altemeier, F., Kaliuzhna, N. (2021). Roadmap to FAIR research information in open infrastructures. *Journal of Library Metadata* (in press). https://doi.org/10.1080/19386389.2021.1999156

Hendricks, G., Tkaczyk, D., Lin, J., Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies* 1(1): 414–427. https://doi.org/10.1162/qss_a_00022

Hillary, F.G., Medaglia, J.D. (2020). What the replication crisis means for intervention science. *International Journal of Psychophysiology* 154: 3–5. https://doi.org/10.1016/j.ijpsycho.2019.05.006

Hotez, P.J. (2020). Anti-science extremism in America: escalating and globalizing. *Microbes and Infection* 22(10): 505-507. https://doi.org/10.1016/j.micinf.2020.09.005

Ioannidis, J.P.A. (2005). Why most published research findings are false. *PLoS Medicine* 2(8): e124. https://doi.org/10.1371/journal.pmed.0020124

Journal of Intelligent & Fuzzy Systems (2021). Retraction notice regarding several articles published in the *Journal of Intelligent & Fuzzy Systems*. *Journal of Intelligent & Fuzzy Systems* 41(5), 5769–5772.

https://doi.org/10.3233/JIFS-219218
Katz, D.S., Hong, N.P.C., Clark, T., Fenner, M., Martone, M.E. (2020). Software and data citation. *Computing in Science & Engineering* 22(2): 4-7. https://doi.org/10.1109/MCSE.2020.2969730
Krell, F-T. (2015). A mixed bag: when are early online publications available for nomenclatural purposes? *The Bulletin of Zoological Nomenclature* 72(1): 19-32. https://doi.org/10.21805/bzn.v72i1.a14
Kosmarski, A., Gordiychuk, N. (2020). Token‐curated registry in a scholarly journal: Can blockchain support journal communities? *Learned Publishing*, 33(3), 333–339. https://doi.org/10.1002/leap.1302
Lammey, R. (2020). Crossref at 20 years: what do the community need? *Science Editing* 7(2): 125–129. https://doi.org/10.6087/kcse.206
Lee, D.J., Stvilia, B. (2014). Developing a data identifier taxonomy. *Cataloging & Classification Quarterly* 52(3), 303–336. https://doi.org/10.1080/01639374.2014.880166
Liang, L., Zhong, Z., Rousseau, R. (2015). Uncited papers, uncited authors and uncited topics: A case study in library and information science. *Journal of Informetrics* 9(1): 50–58. https://doi.org/10.1016/j.joi.2014.11.001
Mayernik, M.S., Maull, K.E. (2017). Assessing the uptake of persistent identifiers by research infrastructure users. *PLoS ONE* 12(4): e0175418. https://doi.org/10.1371/journal.pone.0175418
Mehra, M.R., Desai, S.S., Ruschitzka, F., Patel, A.N. (2020b). Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *The Lancet* https://doi.org/10.1016/S0140-6736(20)31180-6; erratum https://doi.org/10.1016/S0140-6736(20)31249-6; expression of concern https://doi.org/10.1016/S0140-6736(20)31290-3; retraction https://doi.org/10.1016/S0140-6736(20)31324-6
Millar, N., Budgell, B., Salager-Meyer, F. (2020). Hype in reports of clinical research: The authors' perspectives. *English for Specific Purposes*, 60, 53-64. https://doi.org/10.1016/j.esp.2020.07.001
Morgan, C. (2008). Journal article version nomenclature: the NISO/ALPSP recommendations. *Learned Publishing* 21(4): 273–277. http://doi.org/10.1087/095315108x356699
Paskin, N. (2005). Digital Object Identifiers for scientific data. *Data Science Journal* 4: 12–20. https://doi.org/10.2481/dsj.4.12
Peroni, S., Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies* 1(1): 428–444. https://doi.org/10.1162/qss_a_00023
Popper, K. (2002). The Logic of Scientific Discovery. Routledge Classics
Roemer R.C., Borchardt, R. (2015). Altmetrics and the role of librarians. *Library Technology Reports* 51(5): 31-37. https://doi.org/10.5860/ltr.51n5
Schneider, J. Avissar-Whiting, M., Bakker, C., Heckner, H., Massip, S., Townsend, R., Woods, N. D. (2021). Addressing disorder in scholarly communication: strategies from NISO 2021. *Information Services & Use*, (in press). https://doi.org/10.3233/ISU-210113
Shashok, K., Matarese, V. (2018). Post-publication peer review in biomedical journals: overcoming obstacles and disincentives to knowledge sharing. *Roars Transactions* 6(1): 10125. http://doi.org/10.13130/2282-5398/10125
Siler, K. (2020). Demarcating spectrums of predatory publishing: Economic and institutional sources of academic legitimacy. *Journal of the Association for Information Science and Technology*, 71(11), 1386-1401. https://doi.org/10.1002/asi.24339
Teixeira da Silva, J.A. (2016) Silent or stealth retractions, the dangerous voices of the unknown, deleted literature. *Publishing Research Quarterly* 32(1): 44-53. http://doi.org/10.1007/s12109-015-9439-y
Teixeira da Silva, J.A. (2020a). Evolution in the correction of the literature: preprints, manuscript versioning, error amendment, and retract and replace. *Preprints.org* (preprint, not peer reviewed) http://doi.org/10.20944/preprints201708.0029.v2
Teixeira da Silva, J.A. (2020b). Suggestions for fortifying the discoverability of papers published in *European Science Editing*. *European Science Editing* 46: e57377. http://doi.org/10.3897/ese.2020.e57377
Teixeira da Silva, J.A. (2021). Reference rot degrades information preservation and induces the loss of intellectual integrity. *Epistēmēs Metron Logos* 6: 1-6. http://doi.org/10.12681/eml.25792
Teixeira da Silva, J.A. (2022). Two disturbing trends about expressions of concern. *Medical Journal Armed Forces India* (in press) http://doi.org/10.1016/j.mjafi.2020.12.004
Teixeira da Silva, J.A., Al-Khatib, A. (2021). Ending the retraction stigma: encouraging the reporting of errors in the biomedical record. *Research Ethics* 17(2): 251-259. http://doi.org/10.1177/1747016118802970
Teixeira da Silva, J.A., Al-Khatib, A., Dobránszki, J. (2017). Fortifying the corrective nature of post-publication peer review: identifying weakness, use of journal clubs, and rewarding conscientious behavior. *Science and Engineering Ethics* 23(4): 1213-1226. http://doi.org/10.1007/s11948-016-9854-2
Teixeira da Silva, J.A., Bornemann-Cimenti, H. (2017). Why do some retracted papers continue to be cited? *Scientometrics* 110(1): 365-370. http://doi.org/10.1007/s11192-016-2178-9
Teixeira da Silva, J.A., Dobránszki, J. (2015). Problems with traditional science publishing and finding a wider niche for post-publication peer review. *Accountability in Research* 22(1): 22-40.

http://doi.org/10.1080/08989621.2014.899909
Teixeira da Silva, J.A., Dobránszki, J., Tsigaris, P., Al-Khatib, A. (2019) Predatory and exploitative behaviour in academic publishing: An assessment. *The Journal of Academic Librarianship* 45(6): 102071. http://doi.org/10.1016/j.acalib.2019.102071
van Rossum, J. (2018). The blockchain and its potential for science and academic publishing. *Information Services and Use* 38(1–2): 95–98. https://doi.org/10.3233/ISU-180003
Vergoulis, T., Kanellos, I., Chatzopoulos, S., Karidi, D. P., Dalamagas, T. (2021a). BIP4COVID19: Impact metrics and indicators for coronavirus related publications (Version 70) [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.5759169
Vergoulis, T., Kanellos, I., Chatzopoulos, S., Karidi, D. P., Dalamagas, T. (2021b). BIP4COVID19: Releasing impact measures for articles relevant to COVID-19. *bioRxiv* (preprint, not peer reviewed, version 4) https://doi.org/10.1101/2020.04.11.037093
Vergoulis, T., Kanellos, I., Chatzopoulos, S., Karidi, D. P., Dalamagas, T. (2021c). BIP4COVID19: Releasing impact measures for articles relevant to COVID-19. *Quantitative Science Studies* (in press) https://doi.org/10.1162/qss_a_00169
Vuong, Q-H. (2020). Reform retractions to make them more transparent. *Nature* 582(7811): 149. https://doi.org/10.1038/d41586-020-01694-x
Wang, J. (2007). Digital Object Identifiers and their use in libraries. *Serials Review* 33(3): 161–164. https://doi.org/10.1080/00987913.2007.10765116
Wang, W., Deng, L., You, B., Zhang, P., Chen, Y. (2017). Digital object identifier and its use in core Chinese academic journals: A Chinese perspective. *Learned Publishing* 31(2): 149–154. https://doi.org/10.1002/leap.1137
Wang, Y., Zhao, L. (2021). Blockchain for scholarly journal evaluation: Potential and prospects. *Learned Publishing*, 34(4), 682–687. https://doi.org/10.1002/leap.1408
Whelan, A. (2015). Academic critique of neoliberal academia. *Sites* 12(1): 130-152. http://dx.doi.org/nd.nnnu2/sites-volnoissni
Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 160018. https://doi.org/10.1038/sdata.2016.18
Williams, M. (2015). Situated objectivity, values and realism. *European Journal of Social Theory* 18(1): 76-92. https://doi.org/10.1177/1368431014549681

Table 1 Hypothetical chronological trajectory of the "publication history" of preprints, peer-reviewed papers, guidelines, editor boards, or documents using a double DOI-based system for improved transparency and accountability in the scholarly and academic publishing information stream. All indicated numbers are purely hypothetical.

Publication event / document	DOI1 + version #	DOI2: information recorded chronologically
Preprint route		
Preprint	DOI1v1	Submission date of original; preprint server name; names of any handlers (quasi-editors).
	DOI1v2	Submission date of revised version; names of any handlers (quasi-editors).
	DOI1v*n*[1]	Same information for each preprint version as for DOI1v2.
Preprint + co-submission to peer-reviewed journal	DOI1v*n*+1[2]	Submission date to journal (peer-reviewed or not); journal and publisher name; names of any handlers (editors and peer reviewers if OPR); peer review model (e.g., single-blind, double-blind, OPR, etc.).
	DOI1v*n*+*x*[3]	Return date to author, revision date back to journal; acceptance and/or rejection dates; additional information about peer review process, editors, or peer reviewers that may be different to DOI1v*n*+1 (or DOI1v*n*+*x*–1 for each previous version).
Journal route (peer reviewed, or not)[4]		
Original submission	None	None
Revisions (*n*)[5]	None	None
Accepted version	DOI1v1	Acceptance date. Also, retrospective information: Submission and resubmission date to journal (peer-reviewed or not); journal and publisher name; names of any handlers (editors and peer reviewers if OPR); peer review model (e.g., single-blind, double-blind, OPR, etc.); return date to author, revision date back to journal; acceptance and/or rejection dates; additional information about peer review process, editors, or peer reviewers that may be different to the original submission.
Revised versions	DOI1v*n*	Dates in which errors, author notes (e.g., changes in authorship), editorial notes, addenda, corrigenda, errata, EoCs, retractions, R&R, R&RE, etc. In each case, the submission, resubmission, and other dates when an edit is made, as well as names of any handlers (e.g., editors) are noted.
Ethical guidelines (e.g., COPE, ICMJE)		
Original draft	DOI1	Authors of guideline, notes, date of publication.
Revised versions	DOI1v*n*	Changes to authorship, edits to content, date of publication.
Journal editor board		
Original constituency	DOI1	Editor board constituency; process used to vote editors; ORCID; CV (PDF or HTML); institutional affiliation; COIs; dates of any changes.
Changes to editor board	DOI1v*n*	Removal of editors; reasons for removal; changes to any metadata of DOI1.
Instructions for authors		
Original instructions	DOI1	All clauses of the instructions and guidelines.
Revised versions	DOI1v*n*	Specification of changes to any clause (PDF or HTML); rationale for changes; authorship of changes; dates of any changes.
General		
Hypothetical examples	DOI1[6]: https//doi.org/10.99999/exp1	DOI2[6]: https//doi.org/10.99999/exp2

Notes:
[1] In principle, there is no limit to the number of versions of a preprint, unless a limit is established by the preprint server, allowing it to be updated continuously, and theoretically, indefinitely, without ever needing to be published in a peer-reviewed journal.
[2] When a preprint is submitted to a journal, a new version of the same DOI (DOI1) is assigned, to indicate a new and independent event; even if a preprint is updated (provided that this does not violate the rules of a preprint server) while a preprint is being considered for publication in a peer-reviewed journal, the "base" DOI, i.e., DOI1, does not change, and the version simply reflects the chronological or historical sequence, i.e., the numbering of the version is directly related to, and dependent on, the chronological event, precisely to ensure verification and accuracy.
[3] With any new edits or steps in the processing of the paper or document, a new version is created for DOI1, and the changes that occur in that version are recorded in DOI2 with as much detail as possible.
[4] The publishing format of the paper or journal is irrelevant: it can be open access, subscription, or a hybrid model, without impacting how the DOI1 and DOI2 of the "publication history" are recorded.
[5] These could include some of the categories suggested by Crossref, such as pending publication, advanced online publication or ahead of print, or author accepted manuscript: https://www.crossref.org/education/crossmark/version-control-corrections-and-retractions/#00327 (last updated: April 8, 2020; last accessed: December 16, 2021)
[6] These publisher DOIs do not exist; they and the suffixes used in these examples are merely for illustrative purposes
Abbreviations: CV, *curriculum vitae*; EoC, expression of concern; *n*, additional versions; OPR, open peer review; ORCID, Open Researcher and Contributor ID; R&R, retract and replace; R&RE, retract and republish; v, version; x, additional revised versions that are not the original submission to the journal.