# Submit 20190731 Philipson #558-1538

Dear editors,

I consider I have already responded to the first two issues (numbered by me) in your list, so below following I added responses to the other points as well, indicating also the subsequent changes that have been made in the manuscript to make them more clear. Unfortunately, all the issues could not be addressed within the paper for lack of space. In particular, this affected the issues 10 and 11 in the list below (from Patricia Feeney and Phil Archer).

I am particularly indebted to John Kunze for thorough proof-reading and suggestions for improvement, many of which (– but not all! –) I have accepted and used in this version of the submitted paper, having considered in detail the “tracked-changes” in the google doc from John Kunze, which I send back with comments as a Wordfile (.docx) in this package.

Best regards,

Joakim Philipson

**Från:** Joakim Philipson   
**Skickat:** den 30 juni 2019 12:21  
**Till:** 'kuhntobias@gmail.com' <kuhntobias@gmail.com>  
**Kopia:** 'Alejandra Gonzalez-Beltran' <alejandra.gonzalezbeltran@oerc.ox.ac.uk>  
**Ämne:** SV: [Data Science] Decision letter, #558-1538

Dear Tobias,

I resend this again with a further note that also the first point on your list (bluemarked below) might indicate that you have not read the submitted manuscript in its RASH-rendering, where the indicated passage looks like this:

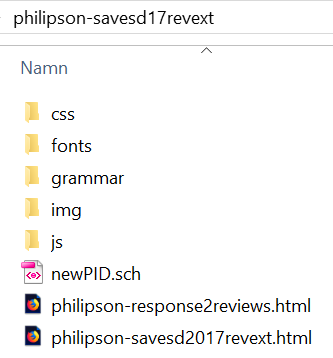
“There are several cases where general data repositories, professing to be FAIR and adhere to accepted metadata standards both for their default output and export formats, nevertheless fail to validate against schemas of these same standards.[[38]](file:///C:\Users\joph9849\Desktop\Konferenser\philipson-savesd2017pubrev\philipson-savesd17revext\philipson-savesd2017revext.html#philipson-2019)”

with the citation [[38]](file:///C:\Users\joph9849\Desktop\Konferenser\philipson-savesd2017pubrev\philipson-savesd17revext\philipson-savesd2017revext.html#philipson-2019) then referring to

Philipson, J. (2019). The Red Queen in the Repository: metadata quality in an ever-changing environment.IDCC 2019. (In press). <https://doi.org/10.5281/zenodo.2276777>

in the reference list, holding a paper and dataset with several examples justifying the statement you say needs justification.

So, in order to have at least a theoretical chance of meeting your new deadline of 20 days, I urgently need to know if this is not enough and if RASH is no longer accepted as submission format! Please remember also that in order to get the RASH rendered properly, the manuscript html-file to be opened must be placed in the same folder/ directory as the css, fonts, grammar, img, js etc. of RASH, thus:



That information is crucial for me to know how to respond to all the issues on your list.

Kind regards,

*Joakim Philipson*

Joakim Philipson

*Ph.D. , MLIS*

*Research Data Analyst*

[*https://orcid.org/0000-0001-5699-994X*](https://orcid.org/0000-0001-5699-994X)

[*https://github.com/StockholmUniversityRDMteam/RDMtoolkit4suRe-use*](https://github.com/StockholmUniversityRDMteam/RDMtoolkit4suRe-use)

*Stockholm University Library*

**Stockholm University** [ <https://ror.org/05f0yaq80> ]  
SE-106 91 Stockholm   
SWEDEN   
Phone: +46-8-16 29 50  
Mobile: +46-72-1464702,

or +46-70-4053127

[joakim.philipson@su.se](mailto:joakim.philipson@su.se)

@JoakimPhilipson

**Från:** Joakim Philipson   
**Skickat:** den 30 juni 2019 00:20  
**Till:** 'Alejandra Gonzalez-Beltran' <[alejandra.gonzalezbeltran@oerc.ox.ac.uk](mailto:alejandra.gonzalezbeltran@oerc.ox.ac.uk)>  
**Kopia:** [iswariya@coredigital.co](mailto:iswariya@coredigital.co); [s.delbecque@iospress.nl](mailto:s.delbecque@iospress.nl); [kuhntobias@gmail.com](mailto:kuhntobias@gmail.com); [michel.dumontier@gmail.com](mailto:michel.dumontier@gmail.com); [m.frohlich@iospress.nl](mailto:m.frohlich@iospress.nl)  
**Ämne:** SV: [Data Science] Decision letter, #558-1538

Dear editors,

I received this notification only on return from vacation without access to the internet and e-mail, but will do what I can to address the issues with my submitted paper (although I fear it will be difficult to meet all the demands within the allotted space for a position paper). However, as I have another deadline to meet for July 8, I will start working seriously on it only after that.

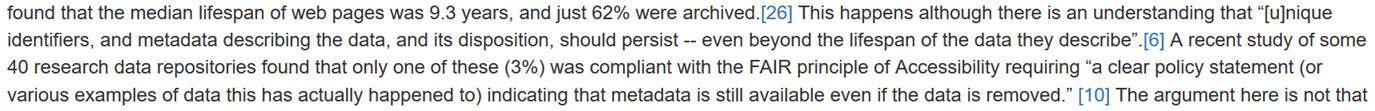
For now, I am just curious to know whether you and reviewers were not able to read the manuscript in the RASH format that it was submitted in (and for which I actually received a price at the SAVE-SD workshops in 2016 and 2017)?

That would explain to me the otherwise incomprehensible point below:

Formatting issues:

* Please, fix the citations, the HTML is not well-formatted. In some cases your citations appear as a link from an underscore symbol and it is even not clear to what statement the citation corresponds to. (e.g. “This happens although there is an understanding that [u]nique identifiers, and metadata describing the data, and its disposition, should persist -- even beyond the lifespan of the data they describe.  A recent study of some 40 research data repositories... ” - in this case, I imagine you are citing the data citation principles to justify the first sentence, but it is not clear what citation is included to justify the statement about the 40 data repositories. Please, revise all citations and that they can be seen properly.

In my RASH rendering of my submitted paper the quotation within parentheses looks like this (here as image to be sure it shows as on my screen):



where the citation numbers are clickable leading to the reference list, but also show full citations on mouseover. The RASH-HTML behind it is this:

This happens although there is an understanding that <q>[u]nique identifiers, and  
 metadata describing the data, and its disposition, should persist -- even beyond  
 the lifespan of the data they describe</q>.<a href="#force11-2014"> </a> A  
 recent study of some 40 research data repositories found that only one of these (3%)  
 was compliant with the FAIR principle of Accessibility requiring <q>a clear policy  
 statement (or various examples of data this has actually happened to) indicating  
 that metadata is still available even if the data is removed.</q>  
 <a href="#dunning-2017"> </a>

with the two references:

<li id="force11-2014" role="doc-biblioentry">  
 <p about="force11-2014" typeof="fabio:WebPage"  
 property="dcterms:bibliographicCitation">Data Citation Synthesis Group,  
 Martone M. (ed.)(2014). <em> Joint Declaration of Data Citation  
 Principles</em> San Diego, CA: FORCE11 <a about="force11-2014"  
 typeof="fabio:WebPage"  
 href="<https://www.force11.org/group/joint-declaration-data-citation-principles-final>">

[https://www.force11.org/group/joint-declaration-data-citation-principles-final</a></p](https://www.force11.org/group/joint-declaration-data-citation-principles-final%3c/a%3e%3c/p)>  
 </li>

<li id="dunning-2017" role="doc-biblioentry">  
 <p about="dunning-2017" typeof="fabio:ConferencePaper"  
 property="dcterms:bibliographicCitation">Dunning, A., de Smaele, M., Böhmer,  
 J. (2017).<em> Are the FAIR Data Principles fair?</em> Practice Paper. 12th  
 International Digital Curation Conference (IDCC 2017), Edinburgh, Scotland,  
 20 - 23 February 2017. <a about="dunning-2017" typeof="fabio:Preprint"  
 href="<https://doi.org/10.5281/zenodo.321423>">[https://doi.org/<span](https://doi.org/%3cspan)  
 about="dunning-2017" property="prism:doi"  
 >10.5281/zenodo.321423</span></a>

* all well-formed RASH-HTML in my webbrowser Firefox Quantum 67.0.4 and in my Oxygen XML editor 20.1 .

If RASH is no longer accepted as a submission format in Data Science, as I was told it was earlier, please tell me so! Then I will have to revise the whole paper.

Other issues on your list might be due to my clearly deficient mastery of the English language, since I experience that the intended meaning of sentences has been lost and utterly misinterpreted. I will try to revise the language here, so as to make my intention more clear, if possible with the aid of John Kunze’s extensive and helpful review.

There are also issues I feel I have already addressed in the last revised submitted paper and in my response to reviewers, e.g. the yellow-marked point about regular expressions below, where I write in the revised paper about DOIs:

“To be sure, there are other regular expression restrictions suggested for DOIs, those that are even more permissive (as DataCite 4.1, with the pattern value for doiType set to "10\..+/.+" [[7]](file:///C:\Users\joph9849\Desktop\Konferenser\philipson-savesd2017pubrev\philipson-savesd17revext\philipson-savesd2017revext.html#datacite4-2017), apart from not being PHP or JavaScript compliant, allowing also inline spaces, or the pattern registered for DOIs at identifiers.org as "^(doi\:)?\d{2}\.\d{4}.\*$" [MIR:00000019](https://www.ebi.ac.uk/miriam/main/collections/MIR:00000019), both of which also allow for the fake DOI above as valid, when tested in [regex101.com](https://regex101.com/)). There are other patterns that are more restrictive, but then obviously not catching all the now prevalent and permitted DOIs by one singular regular expression. [[13]](file:///C:\Users\joph9849\Desktop\Konferenser\philipson-savesd2017pubrev\philipson-savesd17revext\philipson-savesd2017revext.html#fenner-2016), [[18]](file:///C:\Users\joph9849\Desktop\Konferenser\philipson-savesd2017pubrev\philipson-savesd17revext\philipson-savesd2017revext.html#gilmartin-2015)”

See also my response to former Review 4, point 5.

This makes me sometimes suspect that all reviewers and editors have not read thoroughly \***all**\* my responses to former reviews, but perhaps only the one directly addressed to their own review, and may also not have taken note of all the changes in the revised submitted paper. If so, I ask you to take steps to avoid unnecessary redundancies in the issue list, so that I could focus on the most important issues to be revised.

With kind regards,

*Joakim Philipson*

P.S.: Regarding the Twitter references, you will notice that images of the original tweets were in the original paper presented at SAVE-SD 2017, <http://cs.unibo.it/save-sd/2017/papers/html/philipson-savesd2017.html>, but they were removed in this version to save space for the necessary changes in the extended version and addressing some of the former issues. I suspect these “anonymous” references now will have to be removed completely, in order to address this issue and leave further space for others.

Joakim Philipson

*…*

**Från:** [noreply@submissions.iospress.com](mailto:noreply@submissions.iospress.com) <[noreply@submissions.iospress.com](mailto:noreply@submissions.iospress.com)>   
**Skickat:** den 24 juni 2019 07:12  
**Till:** Joakim Philipson <[joakim.philipson@su.se](mailto:joakim.philipson@su.se)>  
**Kopia:** [iswariya@coredigital.co](mailto:iswariya@coredigital.co); [s.delbecque@iospress.nl](mailto:s.delbecque@iospress.nl); [kuhntobias@gmail.com](mailto:kuhntobias@gmail.com); [michel.dumontier@gmail.com](mailto:michel.dumontier@gmail.com); [m.frohlich@iospress.nl](mailto:m.frohlich@iospress.nl)  
**Ämne:** [Data Science] Decision letter, #558-1538

Dear Author,  
  
We have reached a decision for your paper. Please find below the details.  
  
558-1538: Joakim Philipson. PIDs, please play FAIR and identify yourselves! ('Position Paper'). [Click here to follow the link](https://u2275423.ct.sendgrid.net/wf/click?upn=NFwU0TSA-2F23OGsES39oEMn8NQm2GiKo2jBnKurpL5xLxp3t11qM8Dg7EbE0JnvNa0lsWEJ6YEQ1RhGS9dZZVHqvRF6Id-2FzFyLzdloJD4xCSGv3681CCEFDfpMWjkLY-2FR_hNgeXyB53Rxg-2F8hvMV6VWzFo8-2FMUzbYpj9iafp8osFqgbemQabkDw3dyoh1Ga-2FeFPxgQ0N4DXJy7dxAdBLA5T-2F5yN6OTBSJl-2BfZ4bzEJVFMDg11uURjC8tQWKy-2F88Qq-2Bh5rRC9sWnmkbYOe-2FrmBZDyy85xrW3IOyCDW97FlW7j7u6I8iJYfK5-2FQwwioHtQbA3-2B2POz7FEWSiPYEkF7gcaosQDHVBSeUYiNHlvX3t-2FuQ-3D).

We inform you that your paper has been conditionally accepted.

Dear Joakim,

Thanks for your article re-submission and for the responses to reviewers. You will see that from this new round, we received 5 reviews from experts on identifiers from ORCID, ARK, CrossRef, identifiers.org and web

architecture/persistent identifiers. There are varying opinions on the significance and novelty of your contribution, and several suggestions for improvements. In particular, some of the arguments made are not well-justified. I agree with those comments and suggestions for improvement, and I am listing below more issues that must be addressed. As a consequence, my recommendation is to accept the contribution, conditionally to all the changes being incorporated and the paper improved. I also expect to see an enumeration of the changes and justification on how the suggestions were addressed.

I am considering that this is a position paper, and the journals’ submission guidelines indicate: “We accept position papers presenting discussions and viewpoints around Data Science topics. These papers do not need an evaluation, but need to present relevant and novel discussion points in a thorough manner.” (see [https:/!/datasciencehub.net/content/guidelines-authors](https://u2275423.ct.sendgrid.net/wf/click?upn=NFwU0TSA-2F23OGsES39oEMn8NQm2GiKo2jBnKurpL5xJfbuOH8qIqI3zlkqqYbkStyHVIbhUIGz9K-2Ff71sWv7Yw-3D-3D_hNgeXyB53Rxg-2F8hvMV6VWzFo8-2FMUzbYpj9iafp8osFqgbemQabkDw3dyoh1Ga-2FeF-2BLZQXQCbqwucjcQoAKbTRBplaMsLrxvjC5NPqUWFzWl-2B8RJxmwY3WikcrSIIGv0XfYxJLphF76wmAzvGXPyARbzNiVPM9hwBiEyU-2F8QvmPLltsthyJm68yz-2F8bBUOcWOgTWCNDg0xc02ttO7N4U9xXJRmL13ScgxaZtpsgemhUg-3D))

Thus, I strongly encourage you to present the discussion **justifying all your arguments in a thorough manner** as part of the condition for acceptance.

1. I find that many of your statements in the paper (as well as in the response to reviewers) are not well-justified. For example, in the section about FAIR principles you indicate “There are several cases where general data repositories, professing to be FAIR and adhere to accepted metadata standards both for their default output and export formats, nevertheless fail to validate against schemas of these same standards.” This statement is not justified neither with examples nor with a citation (also see issue with citations that I raise in the formatting issues below) and this should be addressed. If you say “there are several cases”, you should provide examples of those cases, including what export formats and why they fail validation, and/or a reference that provides those examples.
2. You refer to “validability” of identifiers, and refer to regular expressions. There are already systems that maintain such expression for identifier validation, such as identifiers.org (see for example the entry for DOI and its regular expression: [https://www.ebi.ac.uk/miriam/main/collections/MIR:00000019](https://u2275423.ct.sendgrid.net/wf/click?upn=NFwU0TSA-2F23OGsES39oEMpT4qz9SVOesrABU984hP6rziXtYSp4dvhEYnge5L91ZzgiGYyle2GX6gTOu0bT2N1j-2FY9OC73QsGDuGrvXBl-2BI-3D_hNgeXyB53Rxg-2F8hvMV6VWzFo8-2FMUzbYpj9iafp8osFqgbemQabkDw3dyoh1Ga-2FeFs8ZlW2SFtQx18fXcsmsaAoy2f2GLDxNg8vOz3xt1qHaOzcHG1Vbfeds7i-2BAur8fURTFz4S5crDE-2FcKQfiPs7qEOUYMtbKTW0s8cy4al6tRskperES7CTjS92P0u2MKySsJkoMZpoQVwDPnABjjpU1MENBA-2BIc8fcCe22z-2FxZYt0-3D)). How does this affect your arguments? Justify.
3. In the text, you are referring to interoperability and then say “This is also used by *fairmetrics.org* as a measure of Findability.” - how is interoperability used as a measure of findability? I don’t think that is correct and you provide no explanation.

ANSWER: The preceding context of this statement in the manuscript is: “Figuring prominently in the explications of all these principles, particularly for interoperability, is the requirement that metadata should be machine actionable, a conditio sine qua non for FAIRness.] [[14]](file:///C:\Users\joph9849\Desktop\philipson-savesd17revext\philipson-savesd2017revext.html#force11-2016) This is also used by fairmetrics.org as a measure of Findability.[[11]](file:///C:\Users\joph9849\Desktop\philipson-savesd17revext\philipson-savesd2017revext.html#FM_F2)” The intended meaning is of course that being *machine actionable* is “a conditio sine qua non for FAIRness” that is also used as “a measure of Findability”, as indicated by those two references (again requiring RASH to be seen). [But, as a further illustration of the prevalence of PIDs, the second reference [11], <https://purl.org/fair-metrics/FM_F2> despite being a “PURL” , now gives a 404, so this reference has now been changed.] To make this clear, the misinterpreted sentenced has been changed accordingly: “Providing machine -readable metadata is also used by fairmetrics.org as one measure of Findability, FM2\_2.[[11]](file:///C:\Users\joph9849\Desktop\philipson-savesd17revext\philipson-savesd2017revext.html#FM_F2)” [The initial reference was to a statement in the now archived Zenodo Gen 1. version of FAIR metrics [**https://doi.org/10.5281/zenodo.1305060**](https://doi.org/10.5281/zenodo.1305060)**,** where the FM\_F2 reads (in html):

“What is being measured? The availability of machine-readable metadata that describes a digital resource.”

In the new Gen 2 it reads in part [ (**<https://github.com/FAIRMetrics/Metrics/blob/master/FM_F2>),** where you yourself is listed as a contributor, so you should know]:

fm:measuring "The availability of machine-readable metadata that describes a digital resource.

…

fm:requirements "A URL to a document that contains machine-readable metadata for the digital resource. Furthermore, the file format must be specified."

1. You insist that the FAIR principles don’t refer to findability, but you say: “However, the FAIR principles do not say anything *explicitly* about *validation*. Particularly for the principles of *Interoperability* and *Re-usability*, it is crucial that metadata can be properly validated against a schema, as adhering to an accepted metadata standard.” So, you are providing a counter-argument that in fact the FAIR principles refer to validation explicitly. Please, clarify.

ANSWER: I do nothing of the sort! Where do you find justification for that claim? I explicitly cite the FAIR principles as they are, including those about being Findable. But, I do suggest that the principle of Findability could be amended and widened, as could those of Interoperability and Re-usability:

“ The "novelty" of the paper, if any, is then the "widening" of the FAIR principles to include also Findability as rate of distribution or dissemination (e.g. as measured by means of 'googling') and Interoperability or Re-usability to include also 'validatability'.”

Is this what you are referring to? If so, this has now beens slightly revised to make it clearer that it is the widening of the already existing principle of Findability that I intended:

“The "novelty" of the paper, if any, is then the "widening" of the FAIR principles to make Findability include also rate of distribution or dissemination (e.g. as measured by means of 'googling') and Interoperability or Re-usability to include also 'validatability'.”

The second color-marked sentence cited by you above, which has little to do with Findability, on the other hand is an argument that I myself am offering, it is NOT something that is explicit in the FAIR-principles.

So in no way it constitutes a “counter-argument” as you claim. I have now made it clear in the text that this is my own argument. At the same time you seem to overlook my contention, as a response to your earlier metareview, that

“It has been remarked that this is already implied by the FAIR principle R1.3 above, but if so, only indirectly and still open to interpretation.”

and the ensuing discussion of what that means.

1. You introduce examples using Life Science Identifiers (LSID) but do not discuss the issues around them. You can check the Wikipedia entry about LSID ([https://en.wikipedia.org/wiki/LSID](https://u2275423.ct.sendgrid.net/wf/click?upn=NFwU0TSA-2F23OGsES39oEMrM64uuD-2FXsJb6uobOOzMG9jzE-2FAGvolnCnmp2bFK3AR_hNgeXyB53Rxg-2F8hvMV6VWzFo8-2FMUzbYpj9iafp8osFqgbemQabkDw3dyoh1Ga-2FeFAoXqNuM9inz4gUa7F9mf2N-2B0CknKZiqHlqq5-2FxWDvKjv-2FV2dDPg6xfdnv33iQYwKsW8vKUxyEp3TkViLsf8YNIDmvFfkz1UvoinLq6OcVq7VqKkkUdwFR6IIJgaCqp8Cw95yWcUQbiL7SlVgQ9jcw9NB9r70tzrMJBc0ww2AEHY-3D)) and in particular, see “Controversy over the use of LSIDs” - [https://lists.w3.org/Archives/Public/www-tag/2006Jul/0041](https://u2275423.ct.sendgrid.net/wf/click?upn=NFwU0TSA-2F23OGsES39oEMuSWuur8sk6cFzLQNLFcFd7xV-2Bjkq3RzEUOK50b-2FX7m6QW3C-2Bo1ZJapHS9qDEiQmqkhSZoD5RbKxjVVH3XWpvdU-3D_hNgeXyB53Rxg-2F8hvMV6VWzFo8-2FMUzbYpj9iafp8osFqgbemQabkDw3dyoh1Ga-2FeF2Lcplxrz0rI6mGRKdE6jpixKoWdoKXJ-2FyDI3ahQaAxIMDzxFVQ0P0oJJC-2Blufhs-2F2rDG6GDR8hJvpqlltKKAP3wCu1qk-2BQFluVFVUym8sw8pWJvxeVJR-2BA6esH0YERsQxqyROb-2Fo0SF7MqT7WT6s2CbWtvIRsF4zeR9h3fcyyic-3D); How does this discussion affect your arguments? Include a justification.  
   ANSWER: I happen to mention almost in passing, ONE example of a composite URN-LSID-UUID, I am not sure if that justifies a whole discussion about LSIDs in particular – that would mean having to leave out something else, as I am already fast approaching the space limit for a position paper.

Have you got a suggestion as to what could be readily left out then? Having said that, I do recognize the relevance of the discussion about

“the explicit separation of data from metadata; specification of a method for discovering multiple locations for data-retrieval; and the ability to discover multiple independent sources of metadata for any identified thing were crucial parts of the LSID and its resolution specification that have not successfully been mimicked by an HTTP-only approach”

from the Wikipedia doc that you refer to, (reminding me a little of the properties pertaining to ISBNs), so maybe that might be an issue for a follow up article to this, if space does not allow inclusion here. Especially since the LSID seems to have failed in reaching those objectives, it might be interesting to try to find out why.

Anyway, I just include a short reference to the Wikipedia page here.

1. In the section ‘Resolvability or findability?’, you mention that “FAIR principles, the focus is very much on *resolvability* of identifiers despite the general awareness of phenomena like 'link rot' and 'reference rot'. What is the basis for this claim? The FAIR guiding principles don’t refer to link rot issues.

ANSWER: This again is a misinterpretation of what I intended to say, given the references (again requiring RASH to be seen), that there is a general awareness of the phenomenon of ‘link rot’, which FAIR principles does NOT address. So I have rephrased this, hoping the meaning will be clearer.

1. You mention ‘When someone in an ensuing Twitter conversation complained about this, an answering tweet seemed to mean, that was the price we have to pay for something as useful as DOIs. ‘ Twitter could provide some anecdotal material, but a tweet is not a  a good reference for justifying a claim for a scientific article. In addition, you say that they “seem to mean” - this interpretation again doesn’t help in making a case. Moreover, the tweets are not referenced. But please, use more reliable references to justify your arguments instead of tweets.

ANSWER: I have removed most of this section, also to save space. The references to the tweets and images were found in the original paper [[37]](file:///C:\Users\joph9849\Desktop\philipson-savesd17revext\philipson-savesd2017revext.html#philipson-2017), they were removed earlier already to save space.

1. About your proposal of a new identifier schema that maintains context in the identifier, and thus it is not opaque, I would like to see an explanation on how your scheme would handle the identification of objects that might change or evolve in the future. For instance, consider the identification of genes, whose information may evolve in time according to new scientific discoveries being made about it. Also, I would like to see a presentation on how your proposal improves the other identifier schemes, and how improves FAIRness.

ANSWER: I am not sure I understand this point, is it a question of *versioning* of identifiers? To be honest, I had such ideas in the first drafts of the manuscript, but then, in view of the discussion of the different strategies employed e.g. by Figahare and Zenodo in this regard, makin g version part of the DOI-URI (Figshare) or having a newly minted PID for every version (Zenodo; cf. ISBNs for new editions) I decided to cut it out, and go for the latter strategy. I do not see that present PIDs are any better at addressing this problem. It is a fact that science and thus our knowledge about objects described by science advance, so this is not something unique for things like genes. But, without knowing the domain, I believe present PIDs used for genes, such as e.g. [CBS](http://genome.ucsc.edu/cgi-bin/hgTracks?clade=vertebrate&org=Human&db=hg17&position=CBS) (ENSG00000160200) in ENSEMBL have their records updated with the progress of science? And I guess the same should be true of e.g. of e.g. proteins PIDs in UniProt, like e.g. <https://www.uniprot.org/uniprot/P35520>. Both of these types of identifiers should be possible to integrate with the model presented in the paper (given they seem to have their own fixed string-length?) in the same way as the IGSN identifiers, so I really do not see the need to discuss this in particular.

1. Please address all the suggestions made by John Kunze and the modifications proposed in the Google document,

ANSWER: I have at least tried to address all the suggestions made by John Kunze in the attached document with my commentaries, and in the submitted manuscript when judged appropriate.

including addressing his points around weak arguments, such as

* + It is not clear why, “for example, the argument that usability and persistence depend on validatability.”

ANSWER: This part of the paper, somewhat revised, is meant to address this issue:

“Failed validation, e.g. due to simple typos or wrong namespace, may help explain why an identifier or URI does not resolve as expected. Validation is also important for the possibility to export metadata to another format, thereby promoting the re-use of data, without exporting also potential errors. Resistance to transcription errors, e.g. by means of a restricted character set, using base32 for encoding, and fixed string-length (suffix has 2 times 4 characters, separated by a hyphen), has been promoted as an advantage of so-called "cool DOIs".[[14]](file:///C:\Users\joph9849\Desktop\philipson-savesd17revext\philipson-savesd2017revext.html#fenner-2016) These are precisely the kind of properties that make PIDs eminently "validatable", and thereby machine-actionable, in the sense of making it possible for a machine to decide of what type a given PID is (cool DOI,ISBN,ISSN...), or - as is seldom the case in my experince - if it already comes "typed", whether it is true to its given type. A real use-case at the National Library of Sweden proved this information to be crucial in order to export error-free metadata to a new environment, to make it searchable, findable and accessible through the library catalog, thus promoting a wider distribution and use of said PID. Although transformation or harvesting of metadata might be possible even without validation, trust in the results and quality as well as the eventual findability of the data (and so again the re-usability) might be seriously affected. The use of standardized, widely distributed PIDs are likely to enhance the chances of finding metadata for a resource, even when the PID-URI fails resolution.”

But, as I am also answering in a comment to John Kunze (in *tracked-changes.docx*):

[Please note, though, that I am not saying that inherent meaning and “validatability” directly CAUSE persistence and resilience; I only suggest these may be elements promoting the continued and wide-spread USE of PIDs, which I (with Van de Sompel et al. [[43]](file:///C:\Users\joph9849\Desktop\philipson-savesd17revext\philipson-savesd2017revext.html#vandesompel-2016a)) do claim is what makes them persistent and resilient.]

* + “it is not clear why the object type, and registrant "modules" proposed for the PID could not live next to, but outside the PID, in a citation.”

ANSWER: First, the registrant “module” in the model is suggested to be optional,

but if used adds data provenance information, which in particular is important in digital preservation measures Secondly, this remark somewhat begs the question, since it presupposes that we already have a citation and metadata at hand (assuming again a working resolvable link), while the model is trying to promote the USE of PIDs for citations, independently of the current resolvability status. An objectType and possibly even registrant inherent to the PID, together with validation, would at least give some clues about its trustworthiness and source, as well as the resource it is supposed to identify.

1. Also, address Patricia Feeney’s points, including full justification for:

“case for non-opaque identifiers was not clearly stated,

the author still conflates accessibility with discoverability and doesn't address arguments for opaque identifiers.”

ANSWER: I believe I did answer this in my submitted Response to reviews 2:

I believe the onus is on those advocating opaque PIDs to show somehow that they are also more persistent than PIDs with some sort of semantic content. The argument put forward here is rather based on the observation that it is the continued use of PIDs that makes them persistent,[[43]](file:///C:\Users\joph9849\Desktop\philipson-savesd17revext\ds-supplementary-558-853.html#vandesompel-2016a) with the possible implication, based on the case of ISBNs, that an inherent semantic structure will potentially enhance the chances of continued use. Admittedly, there is no conclusive proof of this assumption in the paper, but even so, the semantic content proposed here (in section 7) as part of the modular, contextual new PIDs comes primarily from the namespace prefixes and the associated object types, making these PIDs at least easier to interpret also in the future, even if or when they are no longer resolvable.(PIDs are part of metadata, and as has been observed, the challenge for long-term preservation of metadata is to keep them consistently and correctly interpretable over time.[[31]](file:///C:\Users\joph9849\Desktop\philipson-savesd17revext\ds-supplementary-558-853.html#li_sugimoto-2014)

I understand this was not accepted, as a response to the claim that the paper

“doesn’t directly tackle commonly accepted reasons to keep identifiers opaque (for example: persistent identifiers need to remain the same over time and the metadata used to describe the item being identified may change; non-opaque identifiers may lead users to assume details about potential identifiers).

But then I ask the reviewer, with the same right, to provide evidence justifying e.g. the last statement above. I would be very interested indeed to see a thorough comparative research study of the respective “persistence” of various opaque PIDs vs. “semantic” PIDs (such as ISBNs), in terms of F-A-I-R parameters, at different intervals, 5 – 10 – 20 years after being minted. Then we could have a serious discussion, and I would certainly be prepared to accept as a fact that opaque PIDs are more persistent, if the evidence proved them to be so. But as yet, I know of no such evidence, neither do I have any scientific evidence to the contrary. I cannot at least provide anything of the sort, this being a *position paper*, not a full-fledged research study. (This also makes it impossible for me to meet all demands from Phil Archer, I simply cannot provide e.g. “diagrams and tables when relevant (e.g. to show the longevity of different PIDs and the accuracy with which they lead to the identified thing” since I do not have that data).

About PIDs needing to remain the same over time, while metadata describing the item may change, I really do not see the problem with the kind of structural meaning that is inherent in ISBNs, or imparted by the model suggested here, essentially through namespace prefix, objectType, string-length and character-set restrictions. This should not be affected by changes in metadata. BUT, if metadata changes are due to changes of the **object** being identified and described by metadata, there should be a NEW PID to identify that new object (just as with ISBNs for books that changed title or edition.)

[This might call for an extensive discussion of *versioning* strategies for PIDs, that would perhaps lead us astray, but in general, if it is desirable for PIDs to show somehow the *relationship* to *other* in some sense “similar” objects identified by another PID, then one might go for the Figshare strategy of making version part of the PID string, else one would be more inclined to choose the strict policy of Zenodo, where PIDs for different versions of “the same” object are independent of each other.]

Unfortunately, although I have deleted other large parts of the formerly submitted paper to save space, there is simply not enough space left, after trying to meet the other conditions, for this whole argument. But in the RASH manuscript I have added and commented out the following paragraph:

<!-- The persistence of PIDs that acquire meaning in this model   
 through namespacePrefixes and objectTypes, should not be affected by possible  
 changes in metadata, as has been argued in defense of opaque PIDs.   
 But, if metadata changes are due to changes of the <em>object</em> being   
 identified and described by that metadata, there should be a NEW PID to identify   
 that new object. -->

1. Also address all the suggestions made by Phil Archer :
   * Improve the presentation including diagrams and tables when relevant (e.g. to show the longevity of different PIDs and the accuracy with which they lead to the identified thing

ANSWER: As indicated above, I find this is simply asking too much from a position paper. I simply do not have that data yet, it would take a major research effort to collect these data first.

* + Switch the order of the presentation to show your proposal of the PID format first; I would suggest that you also provide a comparison table to show how your PID format would improve the issues that you highlight

ANSWER: I cannot see how this would improve anything. On the contrary, I believe the overview of some prevalent existing PIDs in use gives a background to the suggested model, which aims to combine some inherent structural meaning with “validatability” to promote future use and “interpretability”, and thus, - persistence, properties seen to be lacking (in various degree) in other PIDs.

* + Address the issue raised around ISBNs

ANSWER:If the “issue” here is “However, sadly, it is not true that ISBNs provide a 1-1 mapping.”, I do not think I ever claimed ISBNs are globally unique, if that is the intended meaning of this remark. I am well aware of the fact that ISBNs are sometimes misused or are erroneously used and assigned – in MARC based library catalogs there is even a special subfield, 020 $z (if I remember correctly) for “wrong ISBN” (which, unfortunately, also has been used for other purposes, e.g. corresponding print-ISBN for a digital edition of a work). But I fail to see that this is particularly relevant in this context, when discussing persistence. Nevertheless, I would have been willing to add the following paragraph to address this issue, but there is simply not enough space left for that, so “commented it out” in the RASH manuscript:

<!-- <p>Now, unfortunately, not all ISBNs that are in use have proved to be globally unique; it sometimes happens, for example, that a publisher wrongly assigns to a book an  
 ISBN, that has already been used for another book, even by another publisher in  
 another country. But, due to the generally wide distribution of ISBNs, there are  
 good chances to detect and register these errors. The library community has even  
 dedicated special database "fields" for such errors.</p> -->

Formatting issues:

* Please, fix the citations, the HTML is not well-formatted. In some cases your citations appear as a link from an underscore symbol and it is even not clear to what statement the citation corresponds to. (e.g. “This happens although there is an understanding that [u]nique identifiers, and metadata describing the data, and its disposition, should persist -- even beyond the lifespan of the data they describe.  A recent study of some 40 research data repositories... ” - in this case, I imagine you are citing the data citation principles to justify the first sentence, but it is not clear what citation is included to justify the statement about the 40 data repositories. Please, revise all citations and that they can be seen properly.
* ANSWER: I have already responded to this as being due to not reading the paper in its RASH rendering. Please reread it in RASH.
* The problem with the citations may also be the reason why the paper shows as you introduce many acronyms without indicating what they stand for. For example, the introduction mentions ORCIDs, RORs ARK, DOI, UUID but the references not included. Please, fix these to include citations especially on the first mention of each acronym. In addition, the paper would benefit from including a glossary listing all the acronyms and their definitions.

ANSWER: I have spelled out acronyms on first occurrence in most cases now. But a glossary I think is over the top, anyway there is no extra space left for that.

I look forward to receiving a thoroughly revised version of your article.

Many thanks,