Response to Pieter Vermeesch (Associate Editor)

> Both the reviewers and myself agree with the need to record raw data in a human-readable ASCII format. You have chosen an XML format, but you could have also chosen JSON. In fact, line 363 of the manuscript points out that "JSON has gained popularity over XML in recent years". So I am curious why you used XML. The example of Section 5 would look cleaner in JSON than XML.

Thank you for this particular question, which we thought would come earlier. In essence, we had three reasons:

1. The predecessor of XLUM is, to some extent, the XSYG format developed years ago in the framework of another software project at Freiberg Instruments. While this format did not gain much attention over the last ten years, there was still some room to test the format and spot flaws. This experience ultimately led to the proposal of the XLUM format.

2. XML is highly standardised and used countless times, particularly in data archives. It is sometimes considered a little bit verbose. Still, the available parsers for XML and JSON are not much different in speed, making it more of a personal preference. But we agree that it will be desirable to have a JSON conversion in the future.

3. We aim to get **the** first format definition out and accepted by the community. If we reach that aim, it will be straightforward to convert data into other formats if needed.

> For example, OSL timestamps and signals are stored in vectors of equal length. The XML code places these two vectors at different hierarchical levels of the 'curve' tags. Using an abridged version of the example: 293 303 313 323 333 343 353 363 373 383 JSON would allow you to store the same values in a two-dimensional array, which better represents the information content of the data. For example:

```
curve: {
duration: 10,
signal: [[1,2,3,4,5,6,7,8,9,10],[293,303,313,323,333,343,353,363,373,383]]
}
```

This is correct, and perhaps more is owed to a personal preference. Here we preferred to separate data from metadata. The data measured by the component are always and exclusively stored in the `<curve/>` node. Everything else we tagged as metadata (including the timestamps).

> Since it is quite easy to convert JSON to XML and vice versa, I would suggest that you consider providing both options. Or instead of specifying a specific format, perhaps you could simply propose a schema that can be applied to any database format?

As you wrote, it is easy to convert between both formats. The key to our proposal is the data structure emphasises component-based data storage and the proposed design. If we were to propose this as XML and JSON simultaneously, this may lead to confusion and require some extra effort we think is not justified at the current stage. If JSON makes it easier (in particular regarding databases): certainly we will write a converter and then propose XLUM as XML and JSON format. But we are not yet there, so first things first.

I hope our reasoning makes somewhat sense. It is not about avoiding the effort, but about communicating a standardised format with the solid basis but without further complicating the approach in the first place.

> Line 7: too many commas

Done.

> Line 21: typo -> "Rechereche"

Done.

> On line 67, there is a sudden change of pace. Whereas the previous paragraphs required no specialised knowledge of OSL, point 2 of this enumerated list suddenly moves on to advanced R packages without properly introducing them. I also wonder if the list really needs 5 items, since items 2 and 3 seem to make broadly the same point.

Thank you for pointing this out. We merged items 2 and 3 and added one more line for the cited R packages. This does not really detail the R packages but presents their idea in a nutshell.

> Line 104: Please avoid double brackets between references.

Now removed. The double brackets were in place to avoid the impression that *Zenodo* was developed by us.

> Line 105: Replace "At last some format conventions, hereafter we will" with "In the remainder of this paper, we will" or something along those lines.

Done.

> Figure 2: What is the purpose of the equation? Either explain it or remove it. Caption of Figure 2: "the minimum and maximum values"... of what?

The equation details the calculation andv "balances" the figure (three panels in each column). It is now mentioned in the figure caption, where we also added "temperature" for the "the minimum and maximum values".

> Section 3 contains a long list of requirements, several of which are redundant. I suggest replacing all instances of "should" and "shall". For example, replace "The format should preserve..." with "The format preserves...".

We removed "should" and "shall" (both expressions used technical requirement documents) and merged the last two points.

> Line 162: remove "nearly"

Done.

Line 165: rephrase "2003, the year of the article by Bortolot and Bluszcz (2003);"

Done.

Line 219 and Equation 1: what is the relationship between 'v' and 'A'?

v spans the array A (now written in the text)

Line 222: instead of using two levels of subscript (e.g., `$A_{(2,1)}_1$`), I suggest using 3-dimensional arrays (e.g., $A_{(2,1,1)}$).

Agreed; I cannot remember anymore why we had chosen this, but your suggestions reads less complicated.

Figure 6: What is the difference between the top and bottom panel? Both have the same horizontal axis and are labelled "PMT-luminescence in [cts] 'measured' ".

We added a y-axis and added more information to the figure caption.

Line 380: "The notion of open data is access and insight" is not a proper sentence.

Corrected.

In general, Section 7 is too wordy and vague. Please sharpen and shorten.

We agree, we shortened the text and sharpen the wording.

Final note: we had responded to the reviewers' comments in the public discussion, all comments have been addressed.

On behalf of all co-authors,

Sebastian Kreutzer, Heidelberg, 2023-03-07