

Trialling a new JATS-XML workflow for scientific publishing

Tamir Hassan
Round-Trip PDF Solutions
Vienna, Austria

The *de facto* standard for academic publishing

Wochte kurz darauf

Blogs und im BibCamp-Video

Beyond the PDF

PDF on trial

Wenn es um die Veröffentlichung

Abhandlungen geht

Over the years, it has attracted considerable criticism, being dismissed as 'antithetical to the spirit of the Web'²¹ and 'an insult to science'. Perhaps more colourfully, and with particular reference to the difficulties of using machines to

the PDF has also been described as a 'hamburger that we're trying to turn back into a cow'

Pettifer et al.: Ceci n'est pas un hamburger: modelling and representing the scholarly article, *Learned Publishing* 24:207–220, 2001

FORCE 11 (formerly "Beyond the PDF"), <https://www.force11.org>

Be careful what you wish for

- **Authoritative visual presentation**
- Robust layout
- **Self-contained file**
- Universal format
- Open and fully documented (ISO standard)

- However ...
(usually) no structure

The best of both worlds?

- Embed additional information in PDF
- PDF/A-3 explicitly supports source files
- “Hybrid PDF” functionality in OpenOffice since 2008

- N.B.: *Tagged PDF* is an alternative mechanism for document-generic structure

Workflows are inefficient

- Humanities
 - Submission in Word
 - DTP operator creates final layout by hand
 - Published as PDF
- Natural sciences
 - Submission as PDF
 - Author typesets final version using LaTeX
 - Published as PDF

The experiment

- ACM Symposium on Document Engineering
 - September 2018: Halifax, Nova Scotia, Canada
 - September 2019: Berlin
- Sister conference of XML Prague
 - discount for attendees of both conferences
- First day consists of workshops and tutorials
 - similar to the “Unconference Day”
- *We should practise what we preach*

The experiment (2)

- Format: PDF with content embedded in JATS-XML
- Authoring: Texture
 - <http://substance.io/texture>
 - Written using Web standards
 - Electron or integration into review system
- PDF creation: Pint (Pint Is Not TeX) formatter
 - <https://github.com/tamirhassan/pint-publisher>
 - Complete control over layout, character placement, etc.
 - Future: Tagging and PDF/UA (accessible PDF)
 - Creation of editable PDFs

Aside: Editable PDF

- Tagged PDF:
 - some logical structure
- PDF/UA (accessible PDF):
 - complete logical structure according to definition
 - other requirements such as alt text for images
- Editable PDF:
 - adds layout description according to defined model
 - determines how layout changes with content
- Libraries and archives are set up to work with PDF

Aside: Editable PDF (2)

- Generic format that can represent any print-oriented content
- Application agnostic
- Editing in XML (round-tripping) or via GUI
- Robust typography for minor edits
- Unlike OpenOffice Hybrid PDF, layout is preserved when opening document for editing
- More information on **editablepdf.org**
- *looking for collaborators*

What about reviewing?

- Beyond the scope of this initial pilot
- EasyChair was used
- Necessary to rename XML to PDF
- PDFs were attached for reviewing purposes

Outcomes

- Texture
 - authors needed to learn new software
 - some bugs (e.g. figures) communicated back to developers and fixed
 - no PDF preview (on roadmap)
- Pint
 - layout and e.g. figure placement needed manual adjustment
 - no support for advanced typographic features
- Currently mature alternatives (LyX, LaTeX) perform better

Alternatives

- Authoring
 - LyX → LaTeX → XML
 - Various XML tools for those who prefer to code directly (mostly non-free)
 - Potentially even a Word conversion
 - *However, no possibility to integrate into reviewing system*
- Formatting
 - LaTeX is mature
 - XML formatting engines based on XSL:FO or CSS
 - Headless browser – quality not sufficient
 - *Resulting PDF often not tagged*

Next steps

- Give authors a choice of editor
 - LyX, LaTeX, Texture or JATS-XML by hand
- Larger pilot with structural reviewing
 - Suggestions from the XML community for an EasyChair alternative? :-)
- Further work on formatting
 - Authors require a PDF preview
- Development and formal definition of Editable PDF format

Texture

⌂ ⌕ B I 🔍 x_2 x^2 </> 🗨️ 📄 💰 Cite View

Abstract

Introduction

Results

Discussion

- A- Minimization of laser power
- B- Volumetric Imaging
- C- Minimizing Cross-talk between Imaging and Photostimulation
- D- Nonspecific Activation

Future outlook

References

Simultaneous two-photon imaging and two-photon optogenetics of cortical circuits in three dimensions

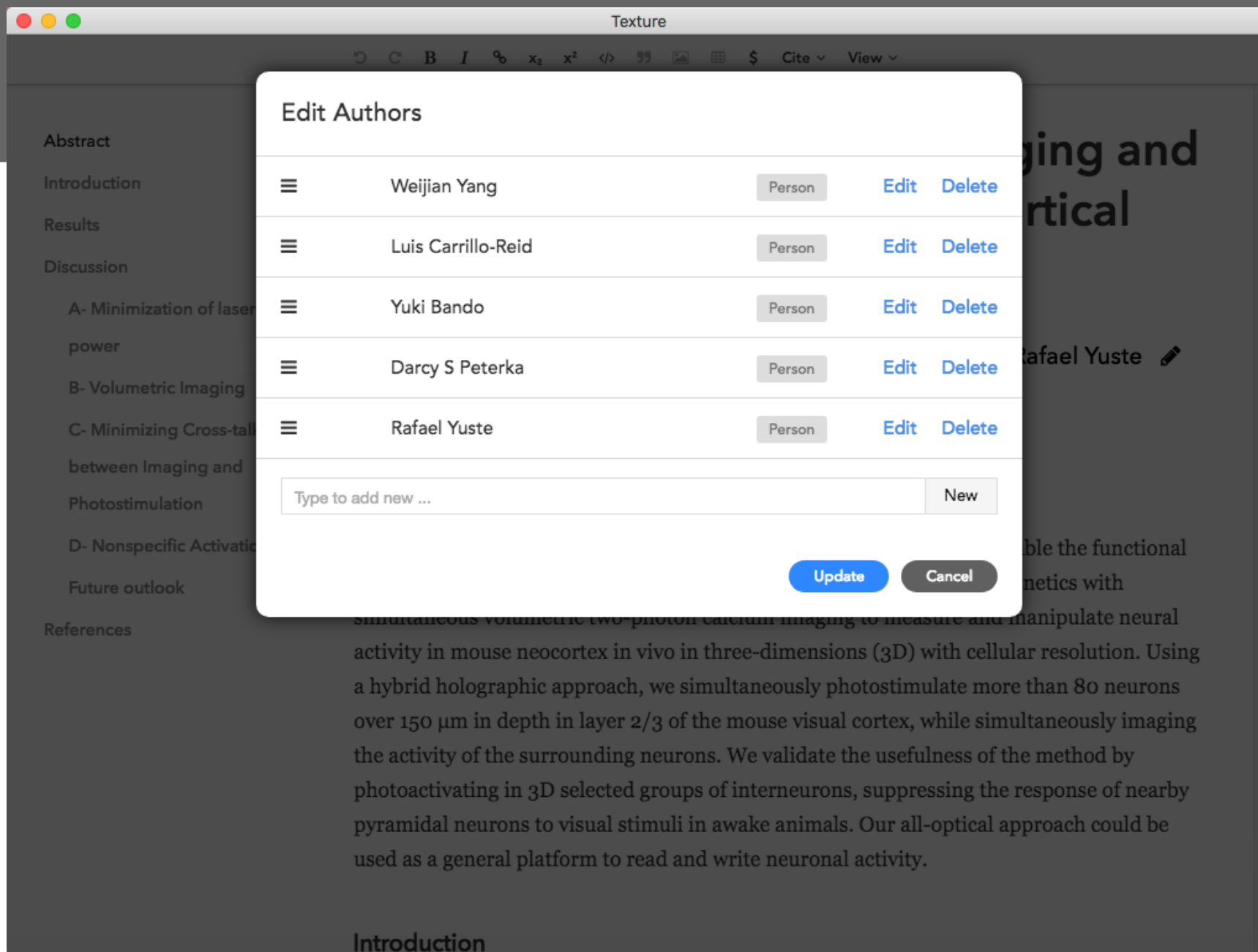
Weijian Yang, Luis Carrillo-Reid, Yuki Bando, Darcy S Peterka, Rafael Yuste ✎

Columbia University, United States

Abstract

The simultaneous imaging and manipulating of neural activity could enable the functional dissection of neural circuits. Here we have combined two-photon optogenetics with simultaneous volumetric two-photon calcium imaging to measure and manipulate neural activity in mouse neocortex in vivo in three-dimensions (3D) with cellular resolution. Using a hybrid holographic approach, we simultaneously photostimulate more than 80 neurons over 150 μm in depth in layer 2/3 of the mouse visual cortex, while simultaneously imaging the activity of the surrounding neurons. We validate the usefulness of the method by photoactivating in 3D selected groups of interneurons, suppressing the response of nearby pyramidal neurons to visual stimuli in awake animals. Our all-optical approach could be used as a general platform to read and write neuronal activity.

Introduction



editable-portable.pdf — Okular

File Edit View Go Bookmarks Tools Settings Help

Previous Next Fit Width Zoom Out Zoom In Browse Zoom

Thumbnails

Contents

Thumbnails

Reviews

Bookmarks

This document has embedded files. [Click here to view them.](#)

Name	Description	Size	Created	Modified
editable-portable.src.xml		40.6 KiB	Tuesday, July 3, 2018 12:54:08 PM UTC	Tuesday, July 3, 2018
editable-portable.phys.xml		21.1 KiB	Tuesday, July 3, 2018 12:54:08 PM UTC	Tuesday, July 3, 2018

Save View Close

Towards a Universally Editable Portable Document Format

Tamir Hassan
Round-Trip PDF Solutions
Vienna, Austria
tamir@roundtrippdf.com

ABSTRACT

PDF is the established format for the exchange of final-form print-oriented documents on the Web, and for a good reason: it is the only format that guarantees the preservation of layout across different platforms, systems and viewing devices. Its main disadvantage, however, is that a document, once converted to PDF, is very difficult to edit. As of today (2018), there is still no universal format for the exchange of editable formatted text documents on the Web; users can only exchange the application's source files, which do not benefit from the robustness and portability of PDF.

This position paper describes how we can engineer such an editable format based on some of the principles of PDF. We begin by analysing the current status quo, and provide a summary of current approaches for editing existing PDFs, other relevant document formats, and ways to embed the document's structure into the PDF itself. We then ask ourselves what it really means for a formatted document to be editable, and discuss the related problem of enabling WYSIWYG direct manipulation even in cases where layout is usually computed or optimized using offline or batch methods (as is common with long-form documents).

After defining our goals, we propose a framework for creating such editable portable documents and present a prototype tool that

ing devices. This is thanks to PDF's roots in PostScript, a printer language on which it is based. The widespread availability of print functions in software at the time (PDF was introduced in 1993) meant that PDF was the "lowest common denominator" among documents from different sources; any document that could also be printed could be effortlessly converted to PDF.

However, this is also PDF's main drawback, and this remains true to this day: the resulting document is essentially a vector graphic, and is difficult to edit, at least in such a meaningful way that would have been possible by the authoring application, such as a word processor. Users wishing to exchange editable documents are therefore still forced to use their native application's format, and do not benefit from the robustness, openness and portability that people have now come to expect from a modern document format.

This paper proposes a framework to tackle this problem, working towards an open, *universal* format that can be edited by a variety of applications conforming to a commonly agreed specification. The following subsection begins by introducing the current state of the art: previous approaches to editing PDFs, relevant developments in the PDF format, and discusses the challenges of a universally editable format with reference to current editable document

1 of 4

Thank you!

Tamir Hassan
Round-Trip PDF Solutions
Vienna, Austria

tamirhassan.com
editablepdf.org