Stretching the boundaries of publishing: The Open Web Platform and the alternatives

Nicolaie Constantinescu

Information architect, kosson.ro, Romania,
E-mail: kosson@gmail.com

Abstract: Today's scientific practice makes data the first territory for analysis and publishing becomes the rich context for cross referencing and interpretation. Scholarly publishing practices, editing and publishing in general have been shaped by the printing press and later on fashioned by the launch of the scientific journals starting in 1665 with the Philosophical Transactions by the Royal Society. Trends are mounting towards a self-publishing model, let that be under a known editorial label or in a self-sustained environment under university's press umbrella or by the help of online free existing tools and services. All of the practices lend themselves to one purpose: being indexed, being parsed, mined for data, and emerging as valuable contributions to the field and become relevant. This paper explores some the new tool chains the Open Web Platform offers and some alternatives to be considered in the daily editing workflows. The envelope of the existing technologies is pushed further to step into the future world of deep learning and artificial intelligence. An exploratory tab will be set on the possibilities self-publishing is presenting the researchers in terms of tools and publishing platforms other than the *regular* channels.

Keywords: publishing; editorial trends; editorial tools; editorial services; editorial business models; open web platform; markdown; publishing platforms; html5; css3; epub.

Limits and conditions of today editing and publishing framework

There are confinements with regards to what you may achieve with the traditional file formats and software used to edit and disseminate research findings. There is a continuous search for new, more flexible tools and formats able to achieve more

audience, let that be persons or machines. Having flexible tools and better articulated technologies that allow cross-sectorial information to be expressed is a scope of activities extending way beyond computer and information science. The main goal is to achieve a better expressibility through non-sophisticated toolchains which are flexible enough to accommodate all fields of research geared up towards the final goal: reproducibility.

"[…] the global scientific publication output is growing at a rate of approximately 3% annually. The volume of publications doubles approximately every 24 years"[1] reported a study done in 2014 that has analysed figures coming from natural, medical and health sciences. But there is also one aspect important for understanding of the new publishing trends: "nine out of 10 academic papers—which both often take years to research, compile, submit, and get published, and are a major component by which a scholar's output is measured—contribute little to the academic conversation".[2]

Concurrently, members of scholarly community are "being hoodwinked into writing books nobody can buy"[3], and sometimes acceptance comes as a part of satisfying "publish or perish" cycle.

These findings mark a continuous race to get one's contribution to the most visible spot concerning a certain field of research or in a particular learned society. The best visibility is gained through content exposure (intrinsic to document exposure), and this sends the researchers and the managers to devising and adopting the best strategies and tooling suites concerning exposure, visualisation and mining. It was wildly acknowledged that only Open Access, an accomplished policy throughout European Union[4], enabled through Open Licenses is able to give the best chance for the content to be exposed through distribution channel exponential multiplication offered by the automated and semi-automated processes.

Part of the strategies employed with content exposure are closely linked with copyright involving open licensing enabled through wavering certain rights retaining only those needed for acknowledgement. This leads to a readership empowerment which in turn resolves to extending precious findings and most of all enables reproducibility.

---

[1] Bornmann, Lutz, and Ruediger Mutz. "Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References." arXiv.org, Digital libraries, May 8, 2014. doi:arxiv.org/abs/1402.4578v3.

[2] Gordon, Aaron. "Killing Pigs and Weed Maps: The Mostly Unread World of Academic Papers." Pacific Standard, Books & Culture, March 18, 2014.

[3] "Academics Are Being Hoodwinked into Writing Books Nobody Can Buy." The Guardian, September 4, 2015, sec. Academics Anonymous.
http://www.theguardian.com/higher-education-network/2015/aug/21/academics-you-need-to-be-managed-its-time-to-accept-that.

[4] "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020." Directorate-General for Research & Innovation, March 21, 2017. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

Reproducibility might be an issue if we are to look at the humanities, but even in these fields data would establish, at least a context for further understanding the reasons for a certain result.

The late comer in this concerto is data as a base component, and how to design *Data Management Plans*[5] in such a way that the research efforts would establish extended context needed for higher reproducibility. Some data policies require even to have the data sets accompanying the articles right at the moment of publication[6].

Reign of LaTeX, XML schema based formats, HTML and PDF seems to go well and undisturbed being the majority of digital resources aggregated in the worldwide repositories where the scholarly scientific output usually lands. This is set to be changed profoundly with the continuous rise of the Open Web Platform being "the collection of open (royalty-free) technologies[7] which enables the Web". Using the Open Web Platform, everyone has the right to implement a software component of the Web without requiring any approvals or waiving license fees"[8]. OWP offers a better integrated platform for information and data, let that be metadata, data microformats or canonical datasets based on Comma Separated Values (CSV) or JavaScript Open Notation (JSON).

"One of the biggest bottlenecks in Open Access publishing is typesetting. It shouldn't be" says Rob Welsh, Product Designer and Co-Founder of Scholastica[9]. Besides bottlenecks there is a lot of documented wasted time in peer-review processes working against "disseminating information useful for researchers and society."[10] There is a whole lot of time spent by the researchers in putting together the research results and formatting and often times adapting their own writing behaviours to the editor's publishing conditions. Picking the right tools to write and make sense of the data collected could be seen as an overhead, but in the end these personalised work-flows need another shape shifting in order to respond to publishing rigours.

Let us not forget the fact that most of citations need to be resolved because the vast majority are hyperlinks. No matter the mechanism employed, the citations should be regarded as context understanding enablers for their essential role of setting the piece of information among the sources which led to it.

---

[5] "Guidelines on FAIR Data Management in Horizon 2020." EUROPEAN COMMISSION, Directorate-General for Research & Innovation, n.d. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[6] Bloom, Theo. "Data Access for the Open Access Literature: PLOS's Data Policy." Public Library of Science. Accessed September 17, 2015. https://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/.

[7] Browser Technologies. http://www.w3.org/wiki/BrowserTechnologies.

[8] Open Web Platform. http://www.w3.org/wiki/Open_Web_Platform.

[9] Scholastica is a journal and publishing platform. http://blog.scholasticahq.com/post/85756650973/one-of-the-biggest-bottlenecks-in-open-access.

[10] Shashok, Karen. "Authors' Editors in the 21st Century: Promoters of Publication Quality and Efficiency." European Science Editing 40, no. 3 (August 2014): 60–62.

Although there are many possibilities to realise this wider context, even to this day simple enabling technologies like Cool URIs needed to establish a true semantic web, are efforts with little impact. "I fully agree with him and others that publishers need to think and act a lot more like web developers"[11].

The technologies used and the final medium should acknowledge that "most of the time, a person sits down at her personal computer not to create, but to *read, observe, study, explore, make cognitive connections, and ultimately come to an understanding*. This person is not seeking to make her mark upon the world, but to rearrange her own neurons. The computer becomes a medium for asking questions, making comparisons, and drawing conclusions—that is, for *learning.*"[12]

Editing, presentation, interactivity, disseminating, reuse

The fast pace of research dissemination is guarded by the continuous debate of a needed capturing context for the whole research output starting from data up to publication and presentational phase.

In the same time, there is a need to use simpler file formats both understandable by the machine and human. The classical steps of putting together a finding has translated the final form of having a paper, and slowly migrates to being able of capturing a workflow with all the pieces this open science paradigm entails: datasets, lab notes, blog posts, debates on social media, presentations, software and all the versions of it, the preprint and if accepted a version for print. But all these apparently chaotic steps are the expression of the modern ways in which science is conducted which leads to some interesting controversies linked to the need for fixity[13] in a dynamic web pages world let alone publishing under a widely known label or enter indie row. It is easy to acknowledge the fact all the scientific content ends up on the web, in a repository or on a webpage, be that a dynamically generated by a content management system or a simple version of HTML.

The new publishing open technologies, are currently amassing to realising the long-envisioned goal of having a modular expressive set of instruments able to combine and complement to realize the full potential of research and innovation expressions. All the scientific output should be accounted through the eyes of the machines as pure data if we are to ripe all the benefits. Text and Data Mining considers machines to be first class citizens of the digital world capable of offering a rich context for the humans

---

[11] McCoy, Bill. "Portable Documents for the Open Web (Part 1)." TOC Tools of Change for Publishing, August 16, 2012. http://toc.oreilly.com/2012/08/portable-documents-for-the-open-web-part-1.html.

[12] Victor, Bret. "Magic Ink: Information Software and the Graphical Interface." Blog. Bret Victor, September 16, 2015. http://worrydream.com/#!/MagicInk.

[13] Anderson, Kent. "The Mirage of Fixity — Selling an Idea before Understanding the Concept." *The Scholarly Kitchen*, n.d. https://scholarlykitchen.sspnet.org/2012/01/09/the-mirage-of-fixity/.

to evolve a broader insight many times soon with the help of a neural network upon which deep learning algorithms are processing research inputs, and possibly having it all fed into an AI.[14]

Overview of the existing building blocks

There are several technologies that lead to a new upheaval. All of them are linked to the web regarded as a platform rather than a distribution network. Some of them are combined for the sum of parts offers better flexibility like in the case of Web Publications: *A Web Publication (WP) is a collection of one or more constituent resources, organized together in a uniquely identifiable grouping that may be presented using standard Open Web Platform technologies*[15]. This idea is not a new one as Object Reuse and Exchange standards issued by the Open Archives Initiative proposed a similar approach in 2008: "Open Archives Initiative Object Reuse and Exchange (OAI-ORE) defines standards for the description and exchange of aggregations of Web resources"[16]. Although both are similar they share the same traits of the *Architecture of the World Wide Web.*

In order to explore future possibilities, there is a need to understand better the parts because all of them in their own right have a transformative effect upon the others.

Open Web Platform: HTML5, CSS3, JavaScript and SVG

HTML is a recommendation[17] of W3C meant to evolve the entire web technologies ecosystem starting with the markup and ending with the different APIs[18]. Together with Cascading Style Sheets 3, JavaScript (an implementation of ECMAscript standard), and Scalable Vector Graphics (SVG) form the Open Web Platform.

One might argue it is only the beginning of the implementation of the new technologies, but all the browser producers are actively supporting it. Also, there are a lot of good editors that would dramatically reduce authoring time in order to arrive to a fine articulated solution. HTML5 technologies act as the enabling context and JavaScript as a glue between the other components of the Open Web Platform. World Wide Web Consortium – www.w3.org – began to show interest to the publishing issues since

---

[14] McGuire, Hugh. "What Books Can Learn from the Web / What the Web Can Learn from Books." Medium, April 21, 2016. https://medium.com/@hughmcguire/what-books-can-learn-from-the-web-what-the-web-can-learn-from-books-64670171908f.

[15] "Web Publications for the Open Web Platform: Vision and Technical Challenges." W3C - Interest Group, May 2, 2017. https://w3c.github.io/dpub-pwp/.

[16] "ORE User Guide - Primer." Open Archives, October 17, 2008. https://www.openarchives.org/ore/1.0/primer.

[17] HTML5: a vocabulary and associated APIs for HTML and XHTML. http://www.w3.org/TR/2014/REC-html5-20141028/.

[18] "Application Programming Interface." Wikipedia, n.d. https://en.wikipedia.org/wiki/Application_programming_interface.

June 2013, when Digital Publishing Interest Group was initiated (https://www.w3.org/dpub/IG/wiki/Main_Page). The activity of the group is essential for understanding the shape of the digital publishing using OWP in the future. The use of these technologies let to many attempts in finding a formula able to fit the bill of the rich expression palette of the traditional publication world and the innovative and experimental one. Some of them will be reviewed for a better understanding of the different avatars, but first the building blocks.

HTML5

The latest iteration of the standard says that *HTML is the World Wide Web's core markup language. Originally, HTML was primarily designed as a language for semantically describing scientific documents*[19]. HTML5 is the context in which many other web technologies are bind like CSS and JavaScript. To the markup structures used to represent the information, there is a programmatic node equivalent for every piece of it called Document Object Model. DOM is the key to realize dynamic and interactive documents because every node is manipulable and mutable.

CSS3 modules

When it comes to the first possible useful aggregation using web technologies, the best combination yet is pure HTML5 and CSS3[20]. For the publishing sector, the magic comes from the CSS Paged Media Module Level 3 which allows arrangement of a webpage for print. One possible experimentation for these two main components is PubCSS[21].

JavaScript – ECMAScript standard

JavaScript is the programming language used in the browser to manipulate the elements on a HTML page through the Document Object Model. In publishing sector, it has its uses when it comes to interactivity, animation of the SVG nodes and as a general enabler for other technologies coupled with the HTML APIs.

SVG

Scalable Vector Graphics (http://www.w3.org/Graphics/SVG/) is another markup language for rendering 2D vector images. The advantage of using SVG is the capability of generating animations. The technology is used as the basic layer for a whole range of libraries aimed at producing illustrations and diagrams.

---

[19] https://www.w3.org/TR/html51/introduction.html#introduction.
[20] https://www.w3.org/Style/CSS/specs.en.html.
[21] Park, Thomas. "PubCSS: Formatting Academic Publications in HTML & CSS." Personal blog. Thomas Park, n.d. http://thomaspark.co/2015/01/pubcss-formatting-academic-publications-in-html-css.

The technologies presented are only the components for realizing an articulated platform for publishing scientific information using the Open Web Platform.

Useful additions for the STEM – Science, technology, engineering and mathematics tool chain

MathML

MathML is a markup XML language released in 1999. MathML is part of the HTML5 namespace meaning that you may drop MathML code into HTML code and the user agent will render correctly. If there are some older technologies involved there is a great JavaScript polyfill library called MathJax that will do the proper rendering. In June 2015, MathML 3.0 became ISO/IEC International Standard (ISO/IEC 40314:2015). "The standardization of MathML is key to "friction-free" interoperability among software components for e-learning and technical collaboration in the academic world". The best place to start learning about it is http://www.w3.org/Math/. For LaTeX users, there are converters hence diminishing the burden of migration.

MathJax

MathJax[22] is the product of MathJax Consortium and it is an open source JavaScript platform for high-quality display of mathematical notation across all browsers. Its structure is highly modular to accommodate the use of LaTeX and MathML directly into the HTML source. MathJax is the primary option for the mathematical representation for all major science publishers and some more.

Chemical Markup Language – CML and ChemDoodle

CML is an XML schema designed to support the chemical representations and is accepted by most of the publishers. ChemDoodle Web Components is a JavaScript library released by iChemLabs under a GPL license. It is making use of the WebGL and canvas element.

LaTeX

"LaTeX is a high-quality typesetting system; it includes features designed for the production of technical and scientific documentation. LaTeX is the *de facto* standard for the communication and publication of scientific documents"[23]. You may take into account LaTeX when you prepare documents. LaTeX is here to stay, being more and more included in the online documents through different plug-in mechanisms. It is not unusual to see fragments of LaTeX in web documents rendered through special parsing scripts.

---

[22] https://www.mathjax.org/.
[23] LaTeX – A document preparation system. http://www.latex-project.org/.

The boundary stretched – online and offline boundary blurred

From the researcher point of view much of the format and afterwards the publication paths are dictated in large amounts by the field or by the sphere of the multidisciplinary road taken.

What binds the practices together is the need to arrive at a version of the document that is acceptable by the publisher. Usually the editorial end of line is a Portable Document Format despite the start and intermediate forms and file formats distilled in the end in the PDF version. Through the use of conversion tools like Pandoc, markdown, R and LaTex files melt into a final version, let that be Microsoft Word (.docx), Open Document Format (.odf), PDF or HTML for the matter.

The Open Web Platform is offering everything in order to present enticing opportunities for text structuring as well as data embedding and on-the-fly visualizations. There is also the capacity of further aggregation and a large cover in terms of standards and fields for descriptive, administrative and technical metadata. Here we will explore some of the technical solutions developed using OWP. First, some of the developments are there to form the main traits of the current trend and the approach of every individual initiative, editorial or experimental leaves a picture rendering the toils of the beginnings.

HTMLBook[24]

Is a XHTM5-based standard for authoring print and digital books. HTMLBook is a subset of XHTML5. Many of today O'Reilly books are published this way. The tool employed is called Atlas[25] and it is also available to the general public. The best case to date for this new editing platform is Lea Verou's[26] book *CSS Secrets: Better Solutions to Everyday Web Design Problems.* Even O'Reilly's new editing tools platform makes use of the Git for version control purposes.

EPUB3 and Web Publications

EPUB is a standard proposed by the International Digital Publishing Forum, and is intended to be a general purpose document format. It is a combination of XHTML5, SVG 1.1, CSS3, JavaScript and XML to produce a special zipped file.

Out of the many formats available, EPUB is the one striding to become the best for distribution and interchange. The key feature is the use of Web Standards. In January, 2017, the efforts of the IDPF were transferred to the World Wide Web Consortium (W3C). Immediately after, in February, 2017, the efforts of both organisations were

---

[24] http://oreillymedia.github.io/HTMLBook/.
[25] https://atlas.oreilly.com/.
[26] Invited Expert in the W3C CSS Working Group.

combined into a *New Roadmap for Future of Publishing is Underway as W3C and IDPF Officially Combine*[27]. Publishing Working Group established soon after published a Charter with the goal of having "all publications—with all their specificities and traditions—to become first-class entities on the Web"[28].

EPUB+WEB

Is a "vision"[29] fostered by the W3C to advance the establishment of a document representation to be used in the context of Open Web Platform. This public working draft has a section dedicated to scholarly publishers and STM publishers. This vision achieved a new embodiment in the form of the Web Publication.

Web Publications

Given the web technologies, scholars engaged to find a unifying answer and one notable example is the concept of "book in browser"[30] which was followed independently by an intensified activity giving birth to a new community around the new concept of "web publication" enabled by the Open Web Platform.
A Web Publication (WP) is a collection of one or more constituent resources, organized together in a uniquely identifiable grouping, and presented using standard Open Web Platform technologies[31].
The example of Organisation for Economic Co-operation and Development is showing already an integration platform for the Web Publications in the electronic library of OECD accessible at the following link http://www.oecd-ilibrary.org/. Besides the classic PDF, the document is represented as a Web Publication in the form of an HTML version and also as an aggregated form as EPUB.

The case of lab notes

In case the researchers need to document their laboratory work, there are many tools, but only a handful are also capable to offer dynamic interactions. I have chosen the case of Jupyter Notebooks as it is a popular choice thee days. Jupyter Notebooks (https://jupyter.org/) is introducing yet another open document format based on JSON.

---

[27] "New Roadmap for Future of Publishing Is Underway as W3C and IDPF Officially Combine." W3C, February 1, 2017. https://www.w3.org/2017/01/pressrelease-idpf-w3c-combination.html.en.
[28] "Publishing Working Group Charter." W3C - Publishing Working Group, n.d. https://www.w3.org/2017/04/publ-wg-charter/.
[29] Advancing Portable Documents for the Open Web Platform: EPUB+WEB. White Paper. Unofficial draft. 30 June 2015.
[30] Montana State University Library developed a project aiming to explore the combination of web technologies in search for a unified answer: http://arc.lib.montana.edu/book/.
[31] "Web Publications Use Cases and Requirements." W3C, May 2, 2017. https://www.w3.org/TR/pwp-ucr/.

It includes a record of the user's sessions code sequences needed, text, mathematical equations and rich output. To render the content, a Jupyter notebook needs a specialized piece of software called kernels. The kernels are in charge with running the embedded fragments of code and the text, all being transmitted to the kernel using a specialized protocol based on JSON. More example could be consulted at the following link https://github.com/jupyter/jupyter/wiki/A-gallery-of-interesting-Jupyter-Notebooks.

Embedding data

There is a leap of faith every researcher has to take: regard your text as data, as structured data capable for much more than conveying information. The research output most than any human information and knowledge productions must become intelligent, context-reach and extendible. Structured and unstructured data are the opportunities to realise a true semantic web[32]. HTML5 already allows data embedding through the use of special baked-in attribute: "data-something". But there is a much more to be done and the text could be further endowed with "intelligent" marking through attributes or tagging as it is the case of Text Encoding Initiative. There is more.

Using the existing technologies already presented, there is a path to add structured data to the existing markup and this is possible due to the use of RDFa - Resource Description Framework in Attributes. "RDFa, like Microformats and Microdata, enables us to talk about *things* on the Web such that a machine can understand what we are saying[33]". If only the subset proposed in RDFa Lite finds a way into the documents this minimal step would represent a leap forward. Use of microdata[34] would add further valuable data into the document. For a preview of what could be accomplished http://rdfa.info/play/ should be accessed, and further down this path http://rdfa.info/ should be acknowledged as a valuable portal.

Schema.org is a consistent effort of the web commercial services to create a vocabulary compatible with RDFa, microdata and JSON-LD. This vocabulary is extended in every walk of life, but it is capable of expressing bibliographic[35] relationships and periodical structure as of 2014[36] having a group[37] dedicated to bringing this type of data expressiveness into HTML documents. Going a bit deeper we find RDF – Resource Description Framework which is the "standard model for data interchange on the

---

[32] http://www.w3.org/standards/semanticweb/.

[33] http://www.w3.org/TR/rdfa-lite/.

[34] http://www.w3.org/TR/microdata/.

[35] http://www.w3.org/community/schemabibex/wiki/Bib.schema.org-1.0.

[36] Wallis, Richard, and Dan Scot. "Schema.org Support for Bibliographic Relationships and Periodicals." Schema Blog, February 9, 2014. http://blog.schema.org/2014/09/schemaorg-support-for-bibliographic_2.html.

[37] https://www.w3.org/community/schemabibex/.

Web[38]". It allows structured or semi-structured data to be mixed. RDF 1.1, the latest version was published in 2014 as a series of W3C (World Wide Web Consortium) Recommendations[39] and Working Group Notes. RDF Schema which is an XML schema provides a "data modelling vocabulary for RDF data.

One important alternative to the OWP: Markdown

If there is a growing movement of establishing the right web framework for expressing scientific literature, there is also anther towards simplifying the whole writing and editing experience. The most popular format growing in importance is Markdown. This is a "a plain text formatting syntax"[40] aiming to realize readability in the best possible way without markup excess. At this point there are many flavours of markdown dictated by the needs of their adopters (GitHub[41], reddit, Diaspora, Stack Exchange, OpenStreetMap and Source Forge[42]). Markdown is one of the fastest growing syntax for document "intended for one purpose: to be used as a format for *writing* for the web"[43].

Markdown is eroding the traditional approaches. It is more flexible and taxes very little. There is also a fantastic opportunity and this translates the text written in Markdown to be treated as data having a minimal markup to indicate structure.

An entire ecosystem grew gradually and also editorial businesses start to appear. For getting across to the known file formats, there was a need to develop software behaving like a digital Rosetta Stone. One having a large footprint in the open community is Pandoc (command line). Pandoc is a universal document converter answering the need of transforming one markup format into another. There are even collections of templates[44] designed for different types of documents.

For the scholarly publishing a special Pandoc adaptation is available called Pandoc Scholar[45] which gives the instruments to easily introduce citations because it has integrated CiTO, the Citation Typing Ontology (http://www.sparontologies.net/ontologies/cito/source.html).

---

[38] http://www.w3.org/RDF/.

[39] http://www.w3.org/standards/techs/rdf#w3c_all.

[40] Markdown. Grubber, John. Daring Fireball. http://daringfireball.net/projects/markdown/.

[41] https://github.com/markdown.

[42] Markdown. https://en.wikipedia.org/wiki/Markdown.

[43] Markdown: Syntax. Grubber, John. Daring Fireball.
   https://daringfireball.net/projects/markdown/syntax#philosophy.

[44] https://github.com/jgm/pandoc/wiki/User-contributed-templates.

[45] Krewinkel A., Winkler R. (2017) Formatting Open Science: agilely creating multiple document formats for academic manuscripts with Pandoc Scholar. PeerJ Computer Science 3:e112 https://doi.org/10.7717/peerj-cs.112.

Self-publishing and dynamic dissemination

Signals of new publishing patterns are seen everywhere. Researchers are using their own blogs hosted by themselves or by the research institutions they are affiliated. Besides blogs, there are universities portals having dedicated pages where many researchers choose at least to keep a record of their published results. Some choose to use online publishing services and some keep online a record of their entire work in the form of complex web sites.

Nanopublications

"A nanopublication is the smallest unit of publishable information: an assertion about anything that can be uniquely identified and attributed to its author[46]". Nanopublications are the works of Concept Web Alliance, a collaborative community *that is actively addressing the challenges associated with the production of unprecedented volumes of academic and professional data[47].*In short as the community behind it states, nanopublications are to be regarded as "core scientific statements with associated context".

Crowdfunding pattern

Kick-starter offers a resounding example in the field of computer science. On the 1st of May 2013, Kyle Simpson, an Open Web Evangelist from Austin, Texas, asked the community to back his project of writing a series of books named "You Don't Know JS". Not only he got successful with the campaign, but he successfully published the books with O'Reilly Media[48], and the interesting part is that he was allowed to publish the books on Github[49] as well under a Creative Commons license. This move and new approach to publishing led to a continuous improvement of the series, being considered now as one of the best authority sources in JavaScript programming.

Just put it out there – self-publishing for the commons

One good example comes is *Eloquent JavaScript* by Marijn Haverbeke. This book is licensed under Attribution-Non Commercial 3.0 Unported. The book is published in print with No Starch Press[50]. Besides packing a version of the book as PDF and Epub there is also the online version as HTML5[51].

---

[46] "What Is a Nanopublication." People, Tools and Knowhow for Datapublishing. Accessed September 20, 2015. http://nanopub.org/wordpress/?page_id=65.

[47] http://www.nbic.nl/about-nbic/affiliated-organisations/cwa/introduction/.

[48] http://search.oreilly.com/?q=Kyle+Sympson&x=0&y=0.

[49] https://github.com/getify/You-Dont-Know-JS.

[50] https://www.nostarch.com/ejs2.

[51] http://eloquentjavascript.net/index.html.

Other example comes from Cody Lindley having his books online as HTML5 documents http://www.domenlightenment.com/. The online version are the "pre-edited / draft form" of the O'Reilly edited version (printed, epub, mobi, pdf). The HTML5 files are kept as online versioned variants.

Other example is Python Data Science Handbook by Jake VanderPlas https://github.com/jakevdp/PythonDataScienceHandbook.

With the growing numbers of alternatives for writing and editing scientific contributions, the number of editorial platforms empowering a new bread of writers are growing to sustain these practices. One interesting example is the Leanpub, a platform allowing "to write, publish and sell in-progress and completed books". The generated output from your contributed plain text is a PDF, an EPUB or a MOBI version of your content. Markdown content is allowed as well as Word files.

O'Reilly Media Inc. has opened its publication workflow to the individual authors and set out a model for self-publishing that is easy using the new formats. The platform is called Atlas and offers all the modern publishing instruments needed for the scholarly publishing. There is some ingredient that is common for any modern publishing system: the tools for versioning the content. O'Reilly is making use of GitHub integration allowing the author to feed into the system the content available in the GitHub account. Atlas is not the only platform that has understood the need for versioning of the content. Atlas content is managed with Git. As many others follow Atlas developers understood that the movement of ideas and changes occurs at a ever growing beet rate to which a publishing, a self-publishing ecosystem must be set.

GitHub is a platform preferred more and more for content creation, not only for writing software. More often is the place where code, data and ideas sum-up is put as a perpetual versioned resource. GitHub is not a platform only to create, but also to curate literature and a vivid example would be Project GITemberg declaring to be "a Free and Open, Collaborative, Trackable and Scriptable digital library"[52].

All those publishing platforms supporting integration with GitHub have understood one essential truth of the current editorial flow: text is data. Data is mutable and information is evolved continuously.

Discussion

A better adoption of new technologies, formats and models are to be impacted by influencers of the scientific fields. The new scientific rock stars will be those able to cross cut several disciplines in a growing data analysis capacity built by software and hardware advanced skills. How to accelerate the process?! By taking into account the power laws. Let's identify the hubs for open science and convert them into shift processes with regards on how science knowledge is recorded.

---

[52] https://www.gitenberg.org/.

Today scholarly communication is no more an effort of aggregation and quantification indices. It is how to expose content, if possible linked content in better ways both for humans and for machines. *As the amount of scholarly communication increases, it is increasingly difficult for specific core scientific statements to be found, connected and curated. Additionally, the redundancy of these statements in multiple fora makes it difficult to determine attribution, quality, and provenance[53].*

A possible trend for publishing is reserved to the mutable document and data if the data is the document. In fact, mutable to a form that will enable others to "see further" through possible re-use, forking and versioning. The is a nuance in the sense that a difference has to be made between the peer-reviewed publication of one author as part of the scientific record which is immutable and the research database or data set that is mutable: "seamless integration between the two is not desirable[54]".

"The future of building interactive user interfaces should lie not in the hands of programmers, but in the hands of the expert of a given field. [...] Using technology to create increased levels of abstraction for the documentation of science can push education and society forward[55]".

There is also the need to mine the scientific facts already existing in the scientific literature. "The right to read is the right to mine!" is the case of the ContentMine[56], a project designing software[57] instruments and building a community dedicated to the goal of extracting 100.000.000 facts from the scientific literature that will be put in Wikidata.

It is wise as a creator, as a researcher to retain all documents produced alone or in collaboration. The reasons are multiple, but the biggest threat is web services evolution sometimes leading to site-deaths[58] taking off and breaking all the permalinks[59] (links retrieving a single post) in the process. Another reason for self-publishing is linked to ownership of the resources created, which in turn leads to a direct chain of citations. And one of the most valuable gain is having the URLs of the resources in your domain. In time, this resolves to better searchability and better metrics through local analytic.

---

[53] http://www.nbic.nl/about-nbic/affiliated-organisations/cwa/introduction/.

[54] Shotton, David, Katie Portwin, Graham Klyne, and Alistair Miles. "Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article." Plos Computational Biology, Open Access Collection, April 17, 2009. doi:10.1371/ journal.pcbi.1000361.

[55] Lynch, Dan. The Art of Digital Publishing A Foundation of Combined Standards to Support the Future of Publishing. Accessed September 15, 2015.
http://www.mathapedia.com/books/31.

[56] http://contentmine.org/.

[57] https://bitbucket.org/petermr/ami/wiki/Home.

[58] http://indiewebcamp.com/site-deaths.

[59] http://indiewebcamp.com/permalinks.

A perfect scenario would be for a research output to be seamlessly integrated with the others via semantic technologies in such a way that the contribution would be instantly recognizable. That would be the very moment when everyone, no matter the field will be interested to contribute.

Future of publishing is interwoven with the capacity of transforming a "publication" – research output – into a hub, a proxy or a "small world" that is highly connected with others on the semantic levels. There is already a common denominator: HTML5 with all its components and browser APIs. Due to the fragmentation in the market of electronic publishing dictated by the commercial interests, a wide adoption of HTML5 as a universal expression format is lagging behind. All the components for rich expression of scientific ideas regardless of the field are there to be used in any means necessary. In turn, there are efforts aiming towards adapting to new "industry standards" like EPUB3.

The trend is set by the use and embedding of the data related to the content – the case of STM, and in the case of humanities the new semantics of HTML5 markup allows a better expression and a ubiquitous presence, not to mention a better preservation. Most of the drawback are directly linked to the print form, most of the hesitations are pointing the culprit leading to supplementary resources being allocated to fit the paper.