LIBER QUARTERLY
THE JOURNAL OF THE ASSOCIATION OF EUROPEAN RESEARCH LIBRARIES

Vol. 23, no. 4 (2014) 274–299 | ISSN: 1435-5205 | e-ISSN: 2213-056X

Bringing Digital Science Deep Inside the Scientific Article: the Elsevier Article of the Future Project

IJsbrand Jan Aalbersberg

Elsevier, Amsterdam, The Netherlands
IJ.J.Aalbersberg@elsevier.com

Sophia Atzeni

Elsevier, Amsterdam, The Netherlands
S.Atzeni@elsevier.com

Hylke Koers

Elsevier, Amsterdam, The Netherlands
H.Koers@elsevier.com

Beate Specker

Elsevier, Amsterdam, The Netherlands
B.Specker@elsevier.com

Elena Zudilova-Seinstra

Elsevier, Amsterdam, The Netherlands
E.Zudilova-Seinstra@elsevier.com

Abstract

The ICT revolution of the last decades impacted scientific communication as it has impacted many other forms of communications, changing the way in

Igitur publishing | http://liber.library.uu.nl/ | URN:NBN:NL:UI:10-1-116066

which articles are delivered and how they can be discovered. However, the impact of ICT on the research itself has been much more profound, introducing digital tools to the way in which researchers gather data, perform analyses, and exchange results. This brought new, digital forms of research output, and disseminating those calls for changes deeply impact the core format of the scientific article.

In 2009, Elsevier introduced the "Article of the Future" project to define an optimal way for the dissemination of science in the digital age, and in this paper we discuss three of its key dimensions. First we discuss interlinking scientific articles and research data stored with domain-specific data repositories — such interlinking is essential to interpret both article and data efficiently and correctly. We then present easy-to-use 3D visualization tools embedded in online articles: a key example of how the digital article format adds value to scientific communication and helps readers to better understand research results. The last topic covered in this paper is automatic enrichment of journal articles through text-mining or other methods. Here we share insights from a recent survey on the question: how can we find a balance between creating valuable contextual links, without sacrificing the high-quality, peer-reviewed status of published articles?

Key Words: content innovation; research data; data linking; 3D visualization; content enrichment; text mining

1. Introduction

Until the end of the last century, the role of technology in formal scientific communication was not any different from the role of technology in any other type of (print) communication. For example, when the invention of the printing press enabled easier publication and wider distribution, also scientific findings and views got published more easily and distributed more widely (Galilei, 1638; Le Journal des Sçavans, 1655; Philosophical Transactions, 1655). And when photos and colour got introduced in print, these also found their way into the scientific article.

The introduction and proliferation of digital technology, the internet, and the web in the 1980's and 1990's affected science in a different way than it influenced scientific publishing. Whereas science started to use digital tools for recording, processing, and storing the actual scientific *findings* and *results*, scientific publishing applied digital and web tools for the easier and faster *discovery* and *dissemination* of scientific information (achieved through the introduction of online submission systems, the internal use of SGML and XML, the adoption of PDF, the creation of journal web sites, and the implementation of text-search engines).

Despite all these technological developments in scientific publishing, as far as the scientific content was concerned, the scientific article stayed very close to the (same old) print version (though now distributed in a new electronic format, the PDF).

More recently, however, the role of technology in the scientific article has changed. Value has been discovered in the addition of supplementary material in non-traditional formats, with over 90% of STM journals now offering the option to submit such material (PARSE.Insight, 2009–2010). In some cases, interactive functionality is being offered on top of those supplements, most notably for video (Journal of Visualized Experiments, 2006) and audio file types. In addition, high-quality text mining (Müller, Kenny, & Sternberg, 2004) offers the opportunity to enrich article content with definitions, annotations, clarifications, and links to data.

Nevertheless, it is still fair to say that the use of technology in science moved faster than the technological advances in the scientific article: science saw a further explosion in the creation and exchange of digital data, in almost all disciplines and in many discipline-specific formats (PARSE.Insight, 2009–2010; Smit, 2011), while the scientific article has largely remained the same.

Bridging the gap between the traditional print-based one-size-fits-all scientific article and today's reality of discipline-specific digital science has been the objective of quite a number of studies, prototypes, and publication efforts (Journal of Archaeology, 2009; Shotton, Portwin, Klyne, & Miles, 2009). In this same context, Elsevier initiated the Article of the Future project[1] (Aalbersberg, Heeman, Koers, & Zudilova-Seinstra, 2012), with the following objectives:

- to bring discipline-specific science deep into the formal scientific record, i.e. the scientific article;

- to improve scientific communication by publishing the full richness of scientific research;
- to offer authors the right tools for communicating diverse and discipline-specific results; and
- to provide users an optimal reading experience to obtain effectively maximum insight.

This Article of the Future project was launched in 2009 as an initiative of the life sciences' journals of Cell Press, an Elsevier imprint (Marcus, 2010), and in June 2011 Elsevier publicly released thirteen prototypes for seven more discipline-specific scientific domains. Since then, implementation of quite a few of the Article of the Future concepts has taken place in many of Elsevier's journals and on its full-text platform ScienceDirect[2], with a major change in the online presentation of the scientific article in January 2012.

In this paper, we will discuss three different aspects as addressed by the Article of the Future project on bringing digital and discipline-specific science deep into the scientific article. First, we will cover the topic of how technology enables us to connect research articles with external data repositories, the growing store of the actual research data and results. Secondly, we describe some of our technological advances in bringing 3D visualizations into our research articles: molecular structures, mathematical plots, and archaeological models can now all be visualized inside the article to create a faster understanding of and a deeper insight into the research described.

In our last section we take a more reflective approach, and investigate some of the issues that we encounter when applying modern technology to our content in an immoderate fashion. More specifically: in the context of scientific publishers being able to automatically link entities in our content with definitions, annotations, clarifications, and links to data, we address the question whether our authors and readers appreciate and accept that we will automatically enrich articles without author or referee approval. This is not an easy question to answer, but the answers need to be taken into account in the further deployment of these technologies.

We end the paper with some conclusions and outlooks on future developments, in which scientific data and results get even more deeply integrated into the scientific article … of the future.

2. Connecting research articles and primary research data

2.1. Research data, building the shoulders of giants

Data has always been an integral part of scientific research, providing an objective foundation for evolving insights and, occasionally, the spark for a scientific revolution. But while data and its role in science aren't new, the scale on which it is available is changing fundamentally. Powered by the World Wide Web and by the common availability of cheap data storage solutions, the amount of data that is being produced, analysed and shared, has never been larger. Research data is ubiquitous, as exemplified by some well-known projects such as the Human Genome Project and the Large Hadron Collider. Each of these systems is producing, or already has produced, more data than many people believed to be possible just a few decades ago.

And the end is not in sight. Driven by Moore's law and by the continued growth and development of web infrastructure, we may expect that research data will continue to see an accelerated growth in the foreseeable future. In particular, initiatives to hook up sensory data to the web will increasingly provide the research community with easy access to real-time data from a plethora of sources, all over the globe.

Organizing all of this data is, and has been, a challenge — in particular archiving data in such a way that it will remain available for future generations, that it can be easily found by users (both human and machines), and that the context and the metadata which are essential for understanding and re-use are available. Unfortunately, in the absence of universally followed standards and best-practices, researchers deal with data in many ways. The PARSE.Insight study (PARSE.Insight, 2009–2010), which ran from 2008 to 2010 and was co-funded by the EU, revealed that the most popular choices among researchers to store data include the PC at home, the PC at work, a portable hard-drive, or an organizational server — all of which are generally not easy to access from the outside world. Other options, though less popular, are submitting them as supplementary material with a journal article, or uploading data into an institutional or domain-specific data repository.

Improving the availability of data has a tremendous potential to accelerate the pace of research and prevent errors. Data is often essential to reproduce results, and helps to efficiently and accurately build further on existing work. On the flip side, the absence of data with research publications makes it more difficult to detect research flaws or even scientific misconduct, a point that has been painfully demonstrated by recent scandals (Stroebe, Postmes, & Spears, 2012). In light of the potential of data to accelerate research, it will come as no surprise that funding bodies are increasingly sensitive to research data and have started to place additional data management requirements for grant applications (NSF, n.d.).

A point to highlight is that making research data available is not only good scientific practice, it also brings benefits to researchers who are looking to create impact. Recent studies (Piwowar, Day, & Fridsma, 2007; Henneken & Accomazzi, 2011) indicate that there is a correlation between publications that include the underlying data and a higher citation count. Notwithstanding these results, many researchers are still reluctant to share their data out of fear this may negatively impact their scientific edge (PARSE.Insight, 2009–2010).

2.2. Interlinking data and scholarly articles: a win-win situation

So, in this landscape of rapidly changing technology, massive opportunities and equally impressive organizational and behavioural challenges — how can a publisher of scholarly journals help?

Traditionally, most publishers are supporting authors who want to share their data by giving them the possibility to upload it as supplementary material with the article. This mechanism ensures the long-term availability of data and, by connecting the data directly to the article, it also provides a way to systematically capture metadata and the "why-what-how" narrative that is often necessary to understand and re-use data. However, such supplementary material is typically stored with the article which limits its utility because it can be hard to find and is often not interoperable with similar data sets.

As an alternative to the supplementary material option, Elsevier is keen to work with domain-specific data repositories to encourage authors of journal articles to upload their data to the repository and then interlink data and their published article. This allows the data to enjoy a number of very specific value elements that a data repository can bring, while preserving the valuable connection to the article. Domain-specific data repositories are usually run by experts in the field who have the best expertise about organizing a particular type of data. This enables additional services that help researchers locate data more efficiently and also helps to make data sets better interoperable, so that researchers can combine data sets to meet their specific needs. With the data set lodged at a repository, interlinking it with an article adds value by increasing the visibility and discoverability of both data and the article. In addition, journal articles often provide essential context to data sets, and so linking from the data to the article can alleviate the concern of many researchers that their data might be used incorrectly (PARSE.Insight, 2009–2010).

There are a number of ways in which Elsevier is interlinking data and articles on ScienceDirect, which can be categorized into three classes (Aalbersberg, Dunham, & Koers, 2011): linking through in-text data set identifiers or accession numbers, linking through data repository banners next to the article, and linking to repository-specific applications that are integrated into the article view.

2.3. Data linking through identifiers or repository banners

The first way to link articles to data sets is through data identifiers that may be inserted at any point in the paper, essentially providing authors with a mechanism to refer to, or "cite" data sets. That may be the data that underpins the research presented in the article, but it may also be existing data that is relevant to the work. Authors are required to use a specific syntax when they use data identifiers for accuracy and precision. The generic pattern is *[Database]: [Identifier]*, for example "PDB: 1FS3" to refer to the Protein Data Bank (PDB) with accession number 1FS3 (see also Figure 1). Data set identifiers can be inserted at any place in the article to give authors full flexibility in determining the most appropriate context.

A special case of linking through data set identifiers is the data DOI (Digital Object Identifier). An initiative from DataCite[3], the data DOI was created as a universal identifier for data sets in order to facilitate interoperability between data repositories and to establish a common way to cite data from journal articles. Elsevier supports data DOI's, so that any data DOI occurring in the article will automatically link out to the relevant data website.

Fig. 1: Example of a data accession number included in the article by the author. In this case, this is an accession number from the Protein Data Bank, which the author has included in the figure caption. Taken from http://www.sciencedirect.com/science/article/pii/S0032386113002462.

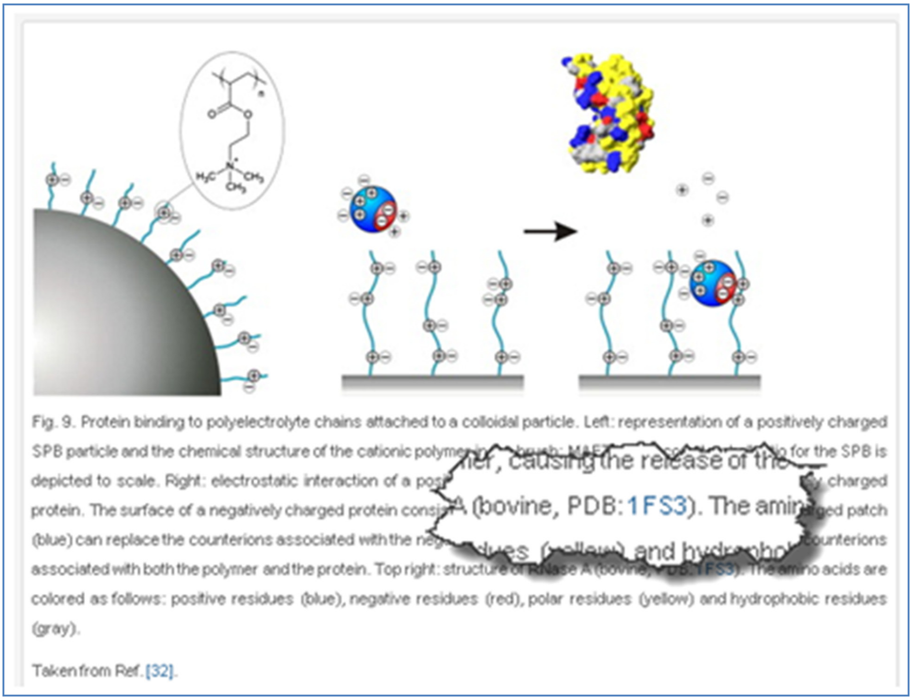


Fig. 9. Protein binding to polyelectrolyte chains attached to a colloidal particle. Left: representation of a positively charged
SPB particle and the chemical structure of the cationic polyme for the SPB is
depicted to scale. Right: electrostatic interaction of a pos charged
protein. The surface of a negatively charged protein cons ged patch
(blue) can replace the counterions associated with the ne counterions
associated with both the polymer and the protein. Top right: structure of RNase A (bovine, PDB: 1FS3). The amino acids are
colored as follows: positive residues (blue), negative residues (red), polar residues (yellow) and hydrophobic residues
(gray).

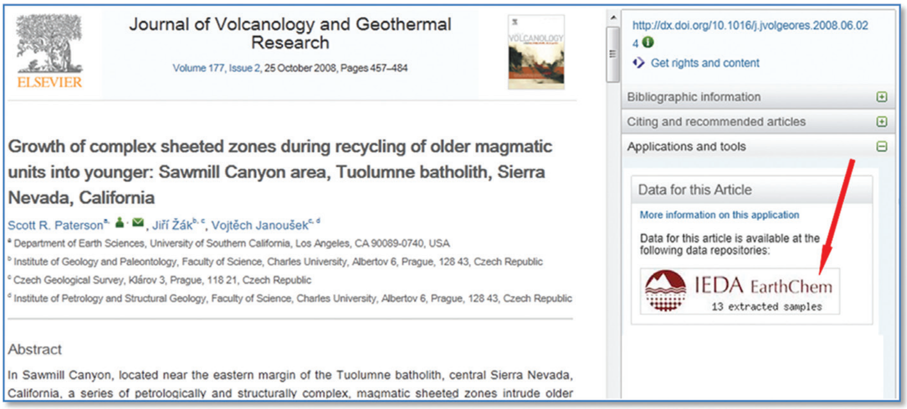
ner, causing the release of the
A (bovine, PDB: 1FS3). The amin
idues (yellow) and hydropho

Taken from Ref. [32].

The second mechanism through which Elsevier is linking with data repositories is by displaying banners next to the article on ScienceDirect, as illustrated in Figure 2. The display of the banner is triggered by a real-time query from ScienceDirect to the data repository server. This process requires that the data repository keeps track of which data set belongs to which article — which quite often is realised, in particular for data repositories that extract data from the literature through data curators.

An additional benefit of real-time banner linking is that it allows for data to be made available and linked after the article has already been published.

Fig. 2: Example of a banner linking out to a data repository (in this case EarthChem). A similar banner will be displayed with all articles on ScienceDirect for which the data repository has relevant data sets. Taken from http://www.sciencedirect.com/science/article/pii/S037702730800348X (the red arrow is added for illustration purposes).



Journal of Volcanology and Geothermal Research

Volume 177, Issue 2, 25 October 2008, Pages 457–484

Growth of complex sheeted zones during recycling of older magmatic units into younger: Sawmill Canyon area, Tuolumne batholith, Sierra Nevada, California

Scott R. Paterson[a], Jiří Žák[b, c], Vojtěch Janoušek[c, d]

[a] Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089-0740, USA
[b] Institute of Geology and Paleontology, Faculty of Science, Charles University, Albertov 6, Prague, 128 43, Czech Republic
[c] Czech Geological Survey, Klárov 3, Prague, 118 21, Czech Republic
[d] Institute of Petrology and Structural Geology, Faculty of Science, Charles University, Albertov 6, Prague, 128 43, Czech Republic

Abstract

In Sawmill Canyon, located near the eastern margin of the Tuolumne batholith, central Sierra Nevada, California, a series of petrologically and structurally complex, magmatic sheeted zones intrude older

http://dx.doi.org/10.1016/j.jvolgeores.2008.06.024

Data for this Article

More information on this application

Data for this article is available at the following data repositories:

IEDA EarthChem

13 extracted samples

This supports authors who desire to make their data available only after an embargo period, and also makes it possible to retro-digitize data for articles that have been published years ago [see e.g. Elsevier Research Data Services (2013) for such an initiative in the geosciences].

Irrespective of the kind of linking, the real value lies in connecting articles and data sets that are specifically relevant to each other — as opposed to pointing readers to a general search portal where such a sense of connection would be lost. For this reason, Elsevier works closely together with supported data repositories to make sure that links to data actually point to landing pages that are specifically relevant to the article.

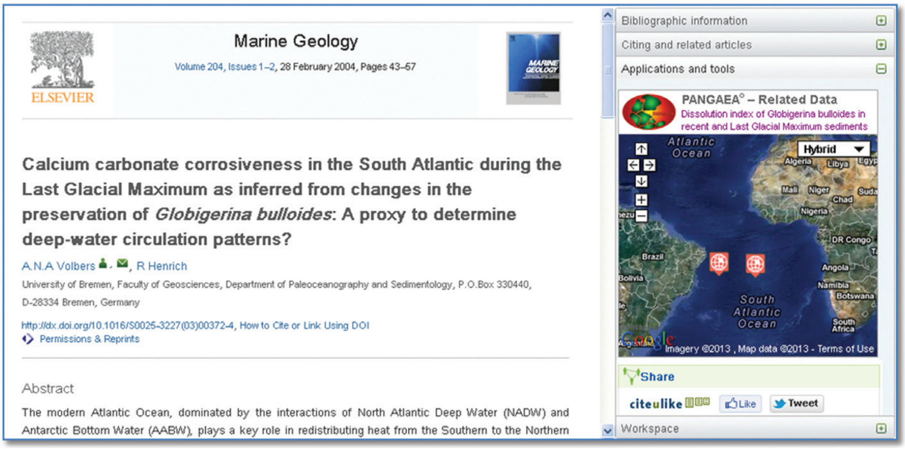
2.4. Integrated data visualization tools

The third way to interlink data and articles utilizes an extensible application framework to develop fully customized data viewers on ScienceDirect. In close collaboration with the data repository, Elsevier has developed a number of such viewers that link out to the data repository's web server, pull in a selection of the available (meta-)data, and display that next to the article on ScienceDirect. This enriches the user experience when reading the article, and also helps readers determine if the full data record that is available under a link will be relevant for them. Examples of these visualization tools include the Protein Viewer and the Genome Viewer on ScienceDirect.

Another example of an integrated data visualization tool is the interactive map viewer that was developed by Elsevier and PANGAEA, a data repository for earth & environmental science. As illustrated in Figure 3, this application places the location of relevant data sets at PANGAEA on an interactive map so that the reader can get a better sense of the data and how that might be relevant for them, directly from within the article page view. The full data record and data sets are available at the PANGAEA portal under a link from the application.

At the moment of writing, Elsevier actively supports linking schemes with over 30 data repositories covering all fields of science, technology, and medical research[4].

Fig. 3: The PANGAEA data viewer shown next to an article on ScienceDirect. The place markers indicate the region of interest for data at PANGAEA that was used for the research presented in this paper. Taken from http://www.sciencedirect.com/science/article/pii/S0025322703003724.



ELSEVIER

Marine Geology

Volume 204, Issues 1–2, 28 February 2004, Pages 43–57

Calcium carbonate corrosiveness in the South Atlantic during the Last Glacial Maximum as inferred from changes in the preservation of *Globigerina bulloides*: A proxy to determine deep-water circulation patterns?

A.N.A Volbers, R Henrich

University of Bremen, Faculty of Geosciences, Department of Paleoceanography and Sedimentology, P.O.Box 330440, D-28334 Bremen, Germany

http://dx.doi.org/10.1016/S0025-3227(03)00372-4, How to Cite or Link Using DOI

Permissions & Reprints

Abstract

The modern Atlantic Ocean, dominated by the interactions of North Atlantic Deep Water (NADW) and Antarctic Bottom Water (AABW), plays a key role in redistributing heat from the Southern to the Northern

PANGAEA° – Related Data

Dissolution index of Globigerina bulloides in recent and Last Glacial Maximum sediments

3. 3D visualization in research articles on ScienceDirect

Authors in many different scientific disciplines deal with 3D data. 3D visualization is an important tool for understanding complex structures, dynamic simulations and research discoveries. In the traditional scientific publication in print, 3D models are "flattened" into static 2D images, which significantly reduce the value of the author's analysis and the level of interactivity and insight of the reader, as they can only capture one specific projection of a 3D object. Hence, embedding 3D visualization tools in online research articles would be extremely useful.

To support this need, Elsevier offers several online 3D viewers appearing inside its online scientific articles in various journals on ScienceDirect. When embedded in the article, 3D visualization tools allow readers to interactively explore 3D objects without interrupting the reading process. The functionality of each 3D viewer is carefully thought out in order to support the domain-specific needs that are required for getting an optimal understanding of the 3D data. When an even more thorough investigation of the data is needed, the reader can download the original dataset as provided by the author, and explore it using visualization tools that he/she normally utilizes in everyday activities.

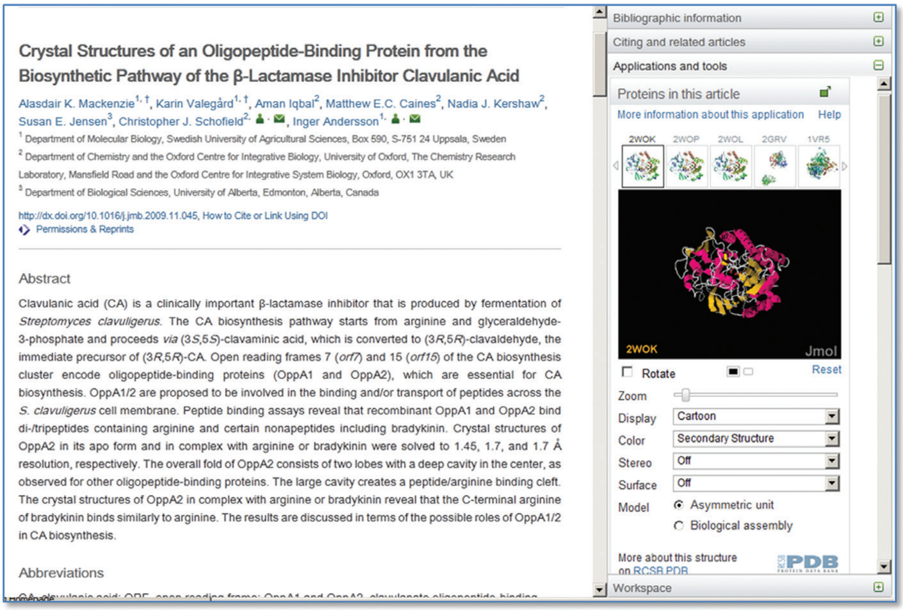
To be able to visualize the 3D data next to the article, it has to be either submitted with the article as supplementary files or uploaded to one of the external data repositories that Elsevier partners with.

3.1. The Protein Viewer

The Protein Viewer is the first embedded 3D visualization tool developed by Elsevier. The Protein Viewer makes use of the Jmol applet[5], which is an open-source Java viewer for 3D chemical structures, with features for chemicals, crystals, materials and bio-molecules. Protein structure files (in **PDB** format) used for the interactive visualization are obtained from the RCSB Protein Data Bank. Each file is processed on-the-fly.

Articles in more than 60 Elsevier journals, which contain protein identifiers, display their structure using the Protein Viewer application displayed alongside the article, in the right hand side panel (see Figure 4).

Fig. 4: The Protein Viewer embedded next to the article on ScienceDirect. Taken from http://www.sciencedirect.com/science/article/pii/S0022283609014399.



Crystal Structures of an Oligopeptide-Binding Protein from the Biosynthetic Pathway of the β-Lactamase Inhibitor Clavulanic Acid

Alasdair K. Mackenzie[1, †], Karin Valegård[1, †], Aman Iqbal[2], Matthew E.C. Caines[2], Nadia J. Kershaw[2], Susan E. Jensen[3], Christopher J. Schofield[2], Inger Andersson[1]

[1] Department of Molecular Biology, Swedish University of Agricultural Sciences, Box 590, S-751 24 Uppsala, Sweden

[2] Department of Chemistry and the Oxford Centre for Integrative Biology, University of Oxford, The Chemistry Research Laboratory, Mansfield Road and the Oxford Centre for Integrative System Biology, Oxford, OX1 3TA, UK

[3] Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

http://dx.doi.org/10.1016/j.jmb.2009.11.045, How to Cite or Link Using DOI

Permissions & Reprints

Abstract

Clavulanic acid (CA) is a clinically important β-lactamase inhibitor that is produced by fermentation of *Streptomyces clavuligerus*. The CA biosynthesis pathway starts from arginine and glyceraldehyde-3-phosphate and proceeds *via* (3*S*,5*S*)-clavaminic acid, which is converted to (3*R*,5*R*)-clavaldehyde, the immediate precursor of (3*R*,5*R*)-CA. Open reading frames 7 (*orf7*) and 15 (*orf15*) of the CA biosynthesis cluster encode oligopeptide-binding proteins (OppA1 and OppA2), which are essential for CA biosynthesis. OppA1/2 are proposed to be involved in the binding and/or transport of peptides across the *S. clavuligerus* cell membrane. Peptide binding assays reveal that recombinant OppA1 and OppA2 bind di-/tripeptides containing arginine and certain nonapeptides including bradykinin. Crystal structures of OppA2 in its apo form and in complex with arginine or bradykinin were solved to 1.45, 1.7, and 1.7 Å resolution, respectively. The overall fold of OppA2 consists of two lobes with a deep cavity in the center, as observed for other oligopeptide-binding proteins. The large cavity creates a peptide/arginine binding cleft. The crystal structures of OppA2 in complex with arginine or bradykinin reveal that the C-terminal arginine of bradykinin binds similarly to arginine. The results are discussed in terms of the possible roles of OppA1/2 in CA biosynthesis.

Abbreviations

CA, clavulanic acid; ORF, open reading frame; OppA1 and OppA2, clavulanate oligopeptide-binding

Proteins in this article

More information about this application Help

☐ Rotate

Zoom

Display: Cartoon

Color: Secondary Structure

Stereo: Off

Surface: Off

Model: ● Asymmetric unit ○ Biological assembly

More about this structure on RCSB PDB

Using the thumbnail menu on top of the viewer, the user can browse through all protein models mentioned in the article and interactively explore each of them: scale the model, change background colour, experiment with different display and colour schemes, apply surface rendering, switch between "biological assembly" and "asymmetric unit", and even view structures in 3D stereo mode. It is also possible to open a larger Protein Viewer in a new window.

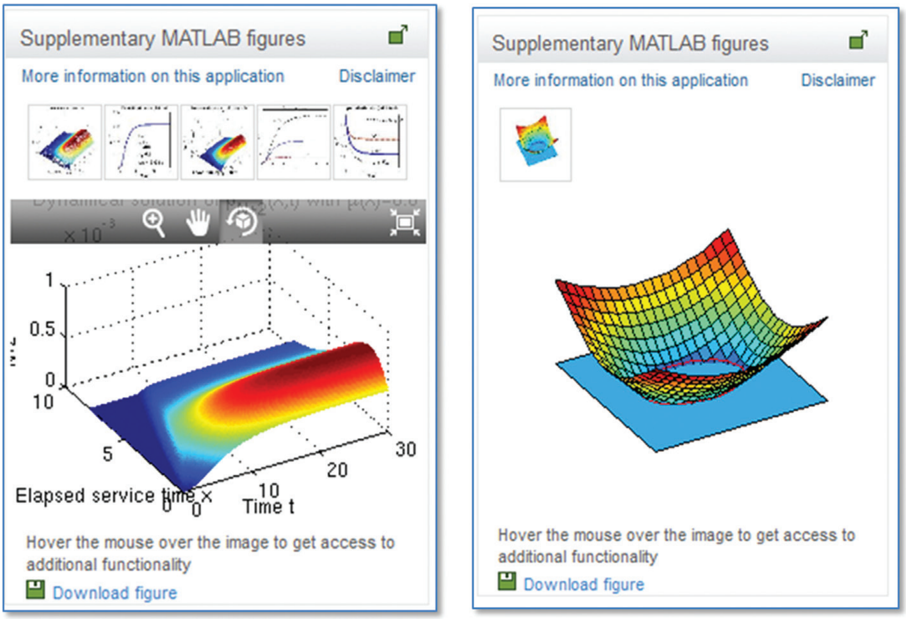
3.2. The MATLAB figure viewer

The Java-based MATLAB figure viewer, launched on ScienceDirect in 2012, enables authors to enrich and extend their articles by making their figures interactive and the underlying data better accessible.

Authors have to upload MATLAB **FIG** files created with MATLAB (a software product from The MathWorks, Inc.[6]) to the Elsevier Editorial System as supplementary files along with their original article manuscript. The application will then automatically generate interactive figures from the **FIG** files provided by the author and include these in the online article on ScienceDirect.

MATLAB **FIG** files provided by the authors may contain experimental data, numerical results, a visualization of a model, etc. The **FIG** file format captures not only the visual information but also the underlying data. Thanks to the MATLAB figure viewer (see Figure 5) displayed next to the article, it becomes possible to view a figure at maximum accuracy at all levels of zoom and from all viewpoints, and also to download the data for validation or re-use. This helps readers to quickly understand the relevance of a research paper and to interactively explore research data for deeper insights. The **FIG** files can also be downloaded for further analysis.

Fig. 5: Two examples of the MATLAB figure viewer. Taken from http://www.sciencedirect.com/science/article/pii/S0307904X12004854 (left) and http://www.sciencedirect.com/science/article/pii/S0045782512002198 (right).



Supplementary MATLAB figures

More information on this application Disclaimer

Hover the mouse over the image to get access to additional functionality

Download figure

Supplementary MATLAB figures

More information on this application Disclaimer

Hover the mouse over the image to get access to additional functionality

Download figure

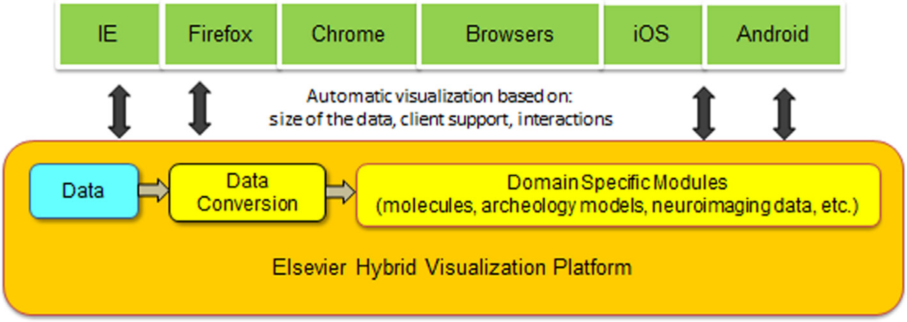
3.3. A visualization platform for online 3D visualization

To address the growing demand in online 3D visualization across various journals and scientific domains, Elsevier launched an Article of the Future subproject aiming at building a 3D visualization infrastructure supporting different data formats, as well as domain-specific visualization techniques and interaction styles. The goal of this subproject is to provide a generic online visualization environment that allows ScienceDirect users to view and interact with small or large 3D datasets submitted with the article. Furthermore, in the current multi-device reader environment, an additional requirement was that the resulting visualizations should be usable from all major devices, including mobile phones, tablets, laptops, and desktops. This project is built in close collaboration with Kitware SAS[7], which serves as the 3D visualization service provider.

The initial visualization infrastructure (Figure 6) is already available and provides a hybrid visualization solution that combines local (WebGL) and remote (ParaViewWeb) rendering techniques, allowing the user to view and interact with small to massive 3D datasets on a large number of devices without any additional plug-in. Depending on the web browser and the size of the 3D data, an optimal user support is offered such that a real-time user interaction with even very big 3D models is guaranteed.

The two visualization modules, which have been completed so far, offer two embedded viewers on ScienceDirect: the 3D molecular viewer and the 3D

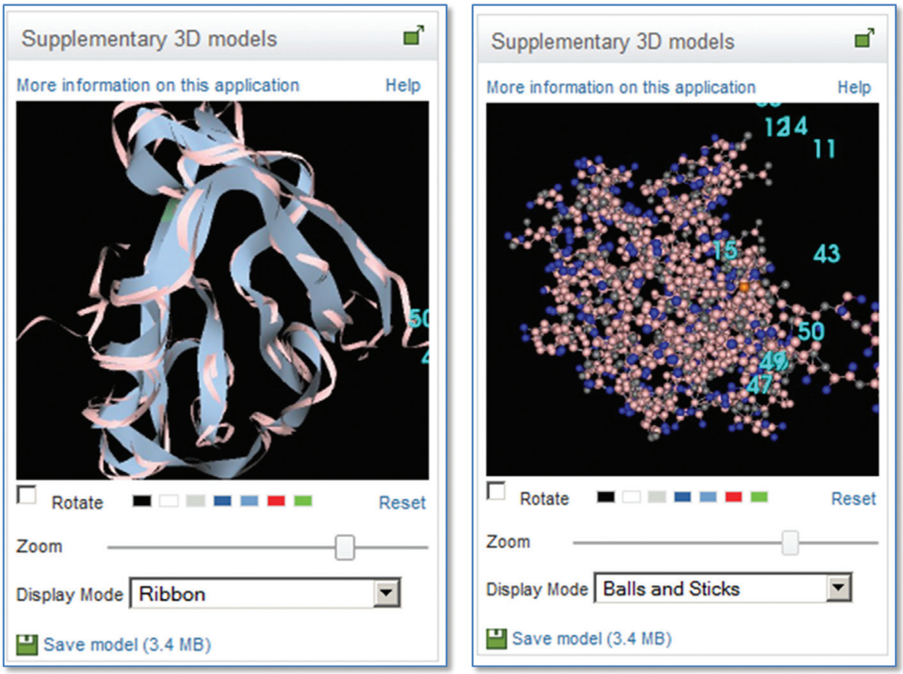
Fig. 6: The architecture of the expandable Elsevier Hybrid Visualization Platform developed in collaboration with Kitware SAS.


IE
Firefox
Chrome
Browsers
iOS
Android
Automatic visualization based on:
size of the data, client support, interactions
Data
Data Conversion
Domain Specific Modules
(molecules, archeology models, neuroimaging data, etc.)
Elsevier Hybrid Visualization Platform

archaeological viewer. Both viewers are displayed next to the relevant online article, in the right hand side panel; it is also possible to open both viewers in a full-screen mode. The 3D molecular/archaeological models are uploaded as supplementary files to the Elsevier Editorial System by the authors, after which these 3D models are pre-processed if the article is accepted.

The 3D molecular viewer visualizes molecular structures and supports **PDB**, **PSE**, and **MOL/MOL2** data formats. It allows the 3D molecular models to be explored using the two most common visualization techniques "ribbons" (Figure 7-left) and "balls-and-sticks" (Figure 7-right). With the viewer, the reader can browse through 3D molecular models using the thumbnail menu on top, and zoom into each model, rotate and pan the model, change display settings, and download original data files.

Fig. 7: "Ribbons" (left) and "balls-and-sticks" (right) visualizations of a 3D molecular model. Taken from http://www.sciencedirect.com/science/article/pii/S0969212612003814.

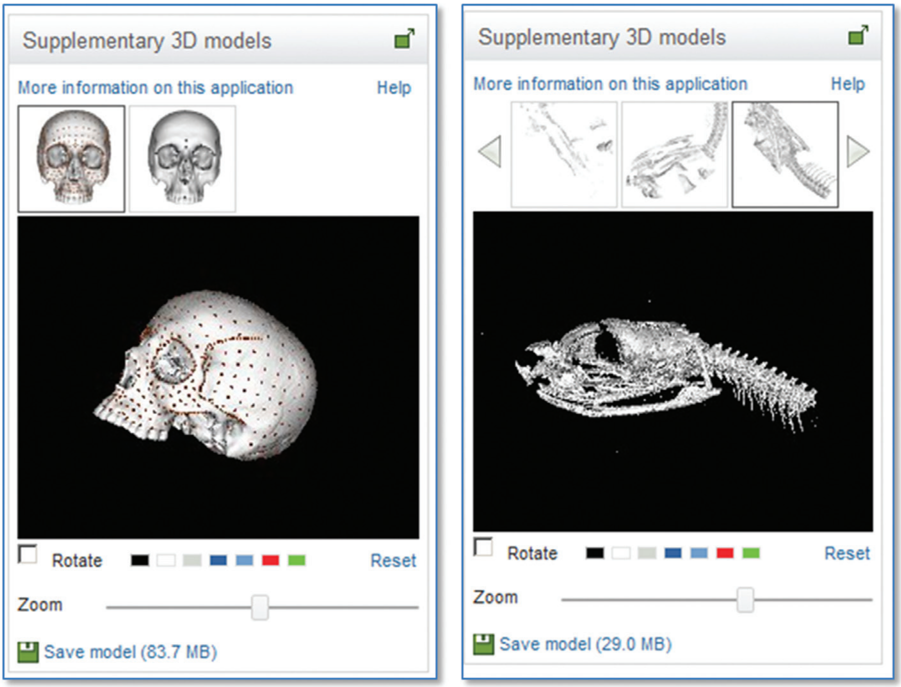


Display Mode Ribbon

Display Mode Balls and Sticks

The archaeological viewer visualizes 3D models submitted in **PLY** and **OBJ** formats. The surface rendering technique is applied to display 3D data, and includes texture and material properties support. Using the viewer, the reader can scale the 3D models described in the article, rotate and pan the model using a mouse, use the auto-rotate option, change display settings, and download original data files. Two examples of 3D archaeological models visualized with the viewer can be found in Figure 8. This viewer has been developed to support the new Elsevier Journal *"Digital Applications in Archaeology and Cultural Heritage" (DAACH)*. DAACH offers scientists the opportunity to publish their models online with full interactivity, such that readers can explore them at will. It is unique in its focus on the application of 3D modelling to cultural heritage (DAACH, 2013).

Fig. 8: Surface rendering of 3D archaeological models. Taken from http://www.sciencedirect.com/science/article/pii/S2212054813000027 (left) and http://www.sciencedirect.com/science/article/pii/S0944200612000931 (right).



The Elsevier hybrid visualization platform and domain-specific modules provide a solid basis for the ongoing enrichment of content within Elsevier's journals. The next 3D module for neuroimaging data in NIfTI format is currently being developed and will become available in online articles of selected neuroscience journals shortly. The value of this new module is supported by a survey amongst 223 neuroscientists (all authors of the piloted journals), of which 80% confirmed that they will use the 3D neuroimaging data viewer if it is available next to the article and provides access to the data that the article deals with. Seventy-five percentage of the survey respondents agreed that getting access to neuroimaging data is essential for a better understanding of experimental research presented in an article.

4. Author and editor views on automatically created links

Traditionally, an author who submitted an article was solely responsible for creating the article content: text, images, and data output such as graphs and tables. However, as technologies developed, today this responsibility also extends to the submission of supplementary data (like 3D models, see above) and the addition of links to related article content (like data sets, see above). However, sometimes modern technology can actually take over the latter task: with modern text-mining tools [like the one described in Müller et al. (2004)], such relevant links can be automatically generated, and in this way, the scientific article gets enriched without any effort from an author, editor, or reviewer.

Even though automatically generated links give great opportunities for advancing the understanding and deepening the insight of a reader, a fundamental question is in how far text-mining technology should create links in a peer-reviewed scientific article. And thus it is important for the Article of the Future project to find a careful balance between adding real value to a scientific article through the automatic addition of links and at the same time keeping the integrity of author creation and peer review intact. As authors and editors are the key stakeholders in the process of content creation and peer review, finding this proper balance led us to posing the following questions to them:

- What is their perception of automatically created links?
- Do they perceive such links as intruding the original article content?
- What is their view on automatically generated links that are incorrect?

To find the answers to these questions we have run an online survey.

4.1. Survey method and design

The online survey was created using web-based survey software. One version of the survey was sent to authors and it emphasized that the survey was about an article that they had published. A second version was directed at editors and had the same questions, but emphasized that the survey was about articles published in one of their journals. Both survey versions consisted of 17, mainly closed, questions and were divided into three parts.

First, we explored the perception, awareness, and expectations of (automatically created) links in an online article. The second part consisted of different statements about automatically generated links, each requesting for a rating between *strongly agree* and *strongly disagree*. The topics for these statements included: (i) concerns about erroneous links, (ii) usefulness of automatically generated links, and (iii) the appearance of author-provided and automatically generated links. The third and last part of the survey contained questions about the authors' or editors' background (experience, research discipline, and more).

Next to these questions, an example to an online article on ScienceDirect with automatically generated links was provided (http://www.articleofthefuture.com/S0031018208004690/). Also the following two definitions were included in the survey:

- *Author provided links* are provided by the article's author to identify unique scientific concepts, usually with 100% precision.
- *Automatically generated links* are created by a text-mining process to identify certain scientific concepts in the article, sometimes resulting in less precision when compared to author-provided links.

These definitions were included to make sure that the respondents interpreted the meaning of these two types of links in exactly the same way. And though we did ask about author-provided links, the main goal of the survey was to understand the perception and awareness of authors and editors about the second type of link: the automatically generated links.

4.2. Results — background information

In total 63 respondents completed one of the versions of the online survey: 47 authors and 16 editors. Although the sample did vary in research domain, current position, and number of years working in the field, the majority of the editors (71%) are professors, who work in their field for more than 8 years. Thirteen out of 63 of the respondents work in the earth sciences domain.

4.3. Results — perception and awareness of links in online article

94% of the responding authors and 75% of the responding editors looked at an online version of one of their articles. 80% of the authors and 75% of the editors knew that links can be automatically added to articles in ScienceDirect.

63% of the authors indicated that they noticed the links in their online article and used them. Out of these, 97% found the links to the related information useful; the main reasons they gave for this were: (i) it is a quick way to get more information, (ii) it saves time, and (iii) it gives additional information about the subject one is reading. 18% of the authors noticed the links but never used them and 16.3% never noticed the links. Some authors commented that the links were a waste of time, distracting, or had no need to use them.

Fewer editors had noticed the links within the online article (44%). 71% of them indicated that links are useful, because they provide complementary information or allow one to replot data. However, it was also noted that these links can lead to distractions: "they sometimes get clicked by accident, and they take me away from what I am reading, fragmenting focus". The majority of editors noticed the links but never used them (25%) or never noticed them at all (25%).

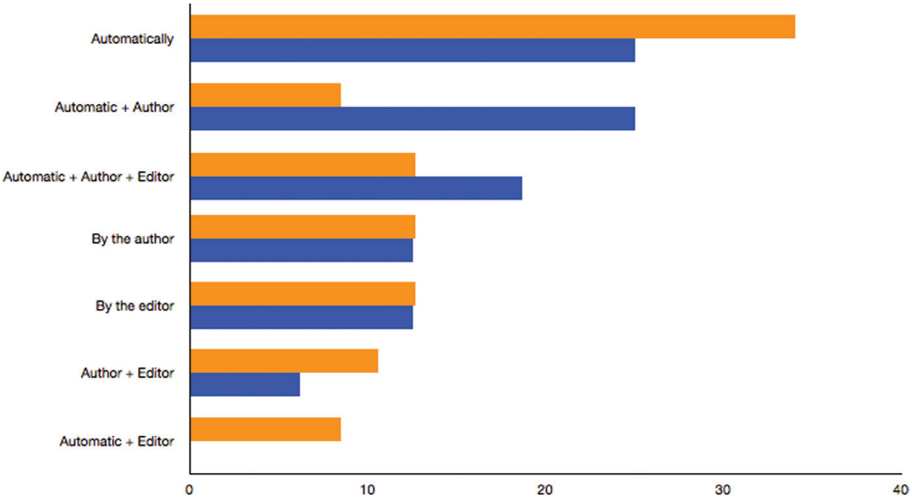
Figure 9 shows that only 34% of the authors and 25% of the editors expect that online links are completely automatically generated. Most authors and editors do expect that there is some form of human involvement (from an editor, author, or both) when links are created.

Respondents were not unanimous on what kind of information they wanted to be added to the online scientific article: the results were distributed between the different possibilities that we suggested. For authors, the majority were interested in general background information (63%), data from the author of the article (55%), definitions (47%), factual data (37%), and information from external sources (33%). For editors, the highest interest was in getting data from the author of the article and seeing definitions (both 56%).

4.4. Results — statements about automatically generated links

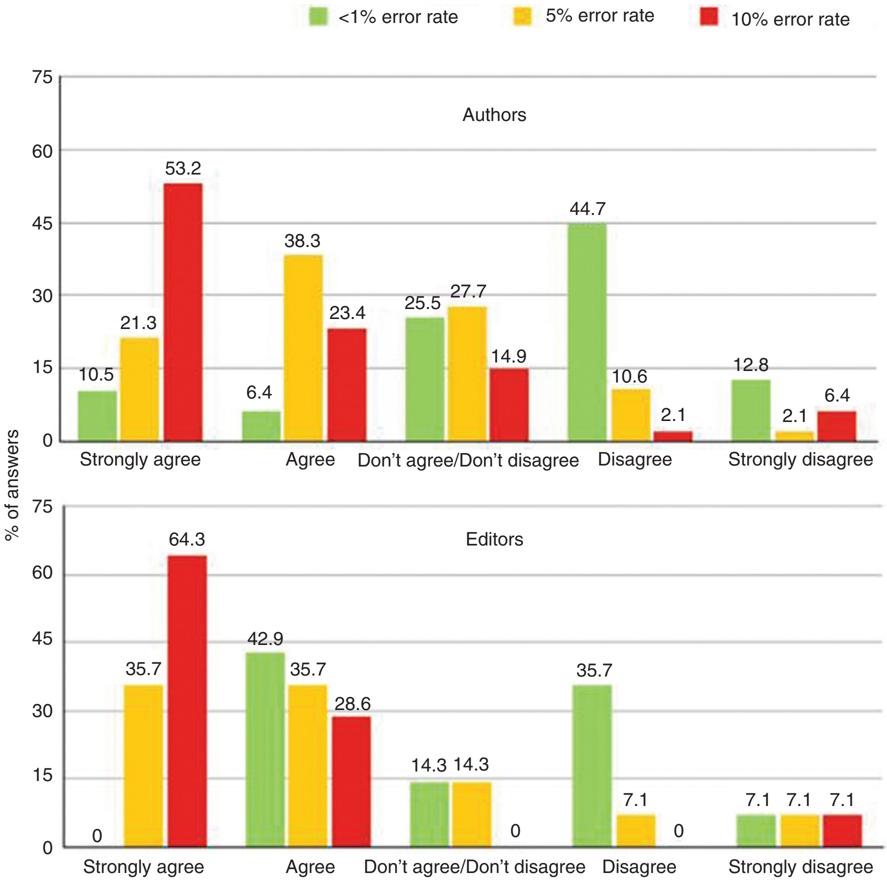
Through measuring the level of (dis-)agreement with explicit statements that we provided, the online survey explored whether authors and editors have concerns about automatically generated links in a scientific article and if so, what these concerns are.

Fig. 9: Question: Where do you expect these links are coming from?



Overall, authors and editors do have concerns when automatically generated links are added to the online version of a scientific article, with their main concern a term that is incorrectly linked. 96% of the authors and 93% of the editors indicated that they "strongly agree" or "agree" with this statement. We also asked to indicate at what error rate percentage they would get concerned: a 5% error rate is already a reason of concern for 60% of the authors and even 70% of the editors (see Figure 10). Other errors, like a missing link, are not a big concern.

Fig. 10: Statement: An error rate of <1%, <5%, or <10% would give me concerns. The green graph bar indicates an error rate <1%, yellow indicates a 5% error rate and red a 10% error rate.



Despite the concerns about erroneous links, both authors and editors do find the links helpful. The majority of them agree with the following statements: "links add more context and understanding for the reader" and "links put more relevant and related information within reach of just 1 click" (see Figure 11). The majority of the authors and about half of the editors indicated that they would want to have links in PDFs as well.

When it comes to the appearance of links in the online scientific article, there is a clear need to distinguish author-provided links from automatically generated ones (see Figure 12). How these links should be distinguished according to the authors and editors did not become clear from the survey: there was a high level of support for both statements "adding a disclaimer at the automatically generated information" and "visually distinguish between the different links".

4.5. Discussion

The concerns as stated at the start of this section are confirmed by this study. The automatic enrichment of the scientific article is helpful and useful. However, the technologies involved should be used with caution and carefulness. For both authors and editors it is not required that all links in a scientific article are author-provided or peer-reviewed, as long as they know how each

Fig. 11: Statement: Links add more context and understanding for the reader.

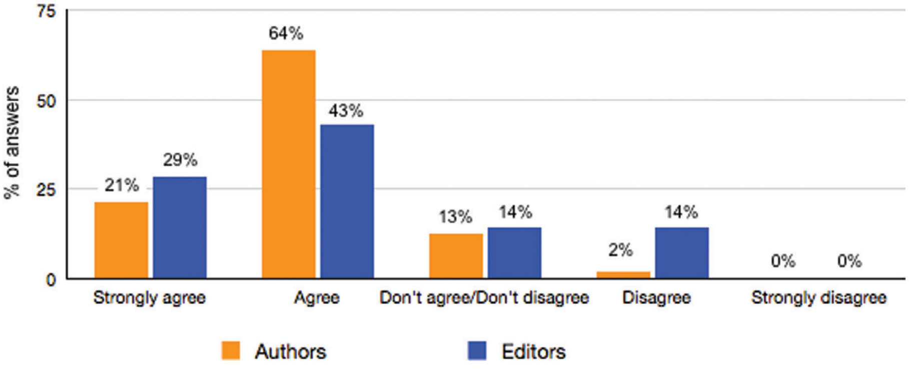
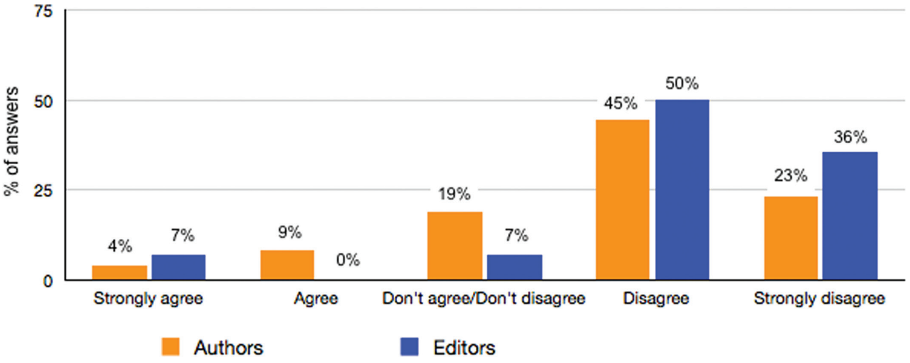


Fig. 12: Statement: There is no need to distinguish author-provided and automatically generated links on the article page.



link came to be: it needs to be clear whether a link is author-provided or automatically generated.

And although the online survey was very clear about automatically generated links, authors and editors do still expect that there is some human involvement when links are added. They expect that the links are reviewed and approved by the author. Under the following conditions respondents are fine with having automatically generated links shown on the article page: they have to be clearly distinguishable and there is only a very small error rate.

5. Conclusion

In this paper we have highlighted three different angles of the Elsevier Article of the Future project — all three related to the recent changes to the scientific article through the extensive use of technology in article content.

First we reviewed different possibilities for how the continuously growing body of research data in external data repositories can be connected to the formal scientific record — making this data more discoverable and enabling the reader to inspect data right from the article. As research data becomes more and more important for both validation of the research presented and for re-use in new research initiated, connections like these will be crucial for the advancement of science in the future.

Then we described some great examples of how today's scientific articles outgrew the traditional paper/print format. Fully interactive mathematical plots and archaeological models replace the traditional flat graphs and images in such articles, providing readers deeper insights into the research presented through detailed inspection capabilities of the research outcomes. These developments in visualisation and interactivity are exciting, and can be seen as front-runners of fully executable papers: articles that offer the possibility to re-run and modify complete computational experiments in the context of the article (Nowakowski et al., 2011; Spagnuolo & Veltkamp, 2013).

At the end of this paper, we reflected on how far scientific publishers can apply automatic technologies (e.g. text-mining) to enrich the scientific article, *without* the endorsement of the author, reviewer, and/or editor. Such enrichments can be of high value and are certainly appreciated, but can they simply be added without peer review? Many authors and editors don't think so and thus there is a challenge for publishers on how to balance the value of automatic enrichment against the scientific peer-reviewed status of the article, especially in the context of not overloading authors and reviewers with extra work. A topic that has not yet been concluded, and which is also very relevant in the further development of e.g. structured digital abstracts (Ceol, Chatr-Aryamontri, Licata, & Cesareni, 2008) and curated nanopublications (Groth, Gibson, & Velterop, 2010): what could and should be done manually, what automatically, and what other in-between alternatives are there available?

From these different angles of the modern scientific article, it can only be concluded that modern technology enables science to become deeper integrated into the article, providing a faster understanding and deeper insights to the scientific reader. However, it is also clear that we are only at the beginning of this re-definition of the scientific article and that — given the different communities of authors, editors, reviewers, and readers — a careful balance in how technological capabilities are being implemented needs to be taken.

References

Aalbersberg, IJ. J., Dunham, J., & Koers, H. (2011). Connecting scientific articles with research data: New directions in online scholarly publishing. In Proceedings of the 1st ICSU World Data Systems Conference, Kyoto (2011). Retrieved May 8, 2013, from http://isds.nict.go.jp/wds-kyoto-2011.org/pdf/IS704.pdf.

Aalbersberg, IJ. J., Heeman, F., Koers, H., & Zudilova-Seinstra, E. (2012). Elsevier's article of the future enhancing the user experience and integrating data through applications. *Insights: the UKSG Journal, 25* (1), 33–43. doi: 10.1629/2048-7754.25.1.33.

Ceol, A., Chatr-Aryamontri, A., Licata, L., & Cesareni, G. (2008). Linking entries in protein interaction database to structured text: The FEBS Letters experiment. *FEBS Letters, 582* (8), 1171–1177. doi: 10.1016/j.febslet.2008.02.071.

DAACH. (2013). *Digital Applications in Archaeology and Cultural Heritag.* Elsevier, Oxford.

Elsevier Research Data Services. (2013). The 2013 international data rescue award. Retrieved May 8, 2013, from http://researchdata.elsevier.com/datachallenge.

Galilei, G. (1638). *Discorsi e dimostrazioni matematiche, intorno à due nuove scienze*. Leiden: Elzevir.

Groth, P., Gibson, A., & Velterop, J. (2010). The anatomy of a nanopublication. *Information Services and Use, 30* (1–2), 51–56. doi: 10.3233/ISU-2010-0613. Retrieved May 8, 2013, from http://iospress.metapress.com/content/ftkh21q50t521wm2/fulltext.pdf.

Henneken, E. A., & Accomazzi, A. (2011). *Linking to data — Effect on citation rates in astronomy*, arXiv:1111.3618v1 [cs.DL]. Retrieved May 8, 2013, from http://arxiv.org/pdf/1111.3618v1.pdf.

Journal of Archaeology in the Low Countries. (2009). Amsterdam: Amsterdam University Press.

Journal of Visualized Experiments, MYJoVE Corporation, Cambridge MA. (2006).

Le Journal des Sçavans. (1655). Paris: Jean Cusson.

Marcus, E. (2010). A publishing odyssey. *Cell, 140* (1), 9.

Müller, H.-M., Kenny, E. E., & Sternberg, P. W. (2004). Textpresso: An ontology-based information retrieval and extraction system for biological literature. *PLOS Biology, 2* (11), e309. doi: 10.1371/journal.pbio.0020309. Retrieved May 8, 2013 from http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.0020309.

Nowakowski, P., Ciepiela, E., Harężlak, D., Kocot, J., Kasztelnik, M., Bartyński, T. et al. (2011). The collage authoring environment. *Procedia Computer Science, 4*, 608–617. doi: 10.1016/j.procs.2011.04.064.

NSF. (n.d.). *Data archiving policy*. Retrieved May 8, 2013, from http://www.nsf.gov/sbe/ses/common/archive.jsp.

PARSE.Insight. (2009–2010). *Insight into digital preservation of research output in Europe. Survey report.* Retrieved May 8, 2013, from http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf and http://www.parse-insight.eu/downloads/PARSE-Insight_D3-6_InsightReport.pdf.

Philosophical Transactions of the Royal Society. (1655). London: Royal Society.

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLOS ONE, 2* (3), e308. doi: 10.1371/journal.pone.0000308. Retrieved May 8, 2013, from http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0000308.

Shotton, D., Portwin, K., Klyne, G., & Miles, A. (2009). Adventures in semantic publishing: Exemplar semantic enhancements of a research article. *PLOS Computational Biology, 5* (4), e1000361. doi: 10.1371/journal.pcbi.1000361. Retrieved May 8, 2013, from http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000361.

Smit, E. (2011). Abelard and Héloise: Why data and publications belong together. *D-Lib Magazine, 17* (1/2). doi: 10.1045/january2011-smit. Retrieved May 8, 2013, from http://www.dlib.org/dlib/january11/smit/01smit.html.

Spagnuolo, M., & Veltkamp, R. (2013). Special issue on executable papers for 3D object retrieval. *Computers and Graphics*, *37* (5), A7–A8. Retrieved October 6, 2013, from http://www.sciencedirect.com/science/article/pii/S0097849313000587.

Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science, 7* (6), 670–688. doi: 10.1177/1745691612460687. Retrieved May 8, 2013, from http://pps.sagepub.com/content/7/6/670.full.pdf+html.

Notes

[1] http://www.articleofthefuture.com

[2] http://www.sciencedirect.com

[3] http://www.datacite.org

[4] Elseyier Database Linking: http://www.elsevier.com/databaselinking

[5] http://jmol.sourceforge.net

[6] http://www.mathworks.com/products/matlab

[7] http://www.kitware.com