Algorithmic Science News: support platform for science journalism

# Algorithmic Science News: support platform for science journalism

**Joana Silva, Maria Fernandes, Arian Pasquali, António Loureiro, Filipa Silva, Pedro Candeias, Helena Lima, Ana Isabel Reis, Bruno Giesteira, Alípio Jorge, Sérgio Nunes**
INESC TEC e Universidade do Porto
joana.rodrigues.silva@gmail.com, mariafernandes118@gmail.com, arianpasquali@gmail.com, ajb.loureiro@gmail.com, filipa.silva@mil.up.pt, pedro.candeias@mil.up.pt, hllima@letras.up.pt, areis@letras.up.pt, bgiesteira@fba.up.pt, amjorge@fc.up.pt, sergio.nunes@fe.up.pt

**Resumo:**
**A plataforma Algorithmic Science News (ASN) é uma nova ferramenta criada por uma equipa multidisciplinar que surgiu da necessidade de reinscrever o papel das notícias sobre ciência no espaço mediático presente. A plataforma tem como objetivo aumentar o número de notícias científicas disponíveis para os editores e reduzir o esforço associado a tarefas mais demoradas como a recolha de dados e análise de artigos científicos, facilitando todo o processo de produção de notícias. A plataforma ASN agrega um conjunto de funcionalidades destinadas a apoiar o trabalho habitual de um jornalista em contextos redatoriais, permitindo a utilização de documentos em repositórios científicos de acesso aberto. O desenvolvimento de algoritmos que trabalham sobre estes repositórios seguiu o propósito de facilitar o acesso e a exploração destas coleções e permitir que os órgãos de comunicação as utilizem como fonte informativa.**
**Neste projeto desenvolveram-se ferramentas de leitura e interpretação, escrita e sugestão semântica. No que concerne**

**à leitura e interpretação, o ASN permite encontrar especialistas relacionados com o artigo científico, resumir as partes mais determinantes, nomeadamente a introdução, objetivos, metodologias e resultados, apresentar definições de termos técnicos e sugerir projetos relacionados com o tema. No que diz respeito à parte de escrita, a plataforma permite escrever notas relacionadas com partes do artigo e ter acesso a sugestões de frases. A utilização de uma versão beta desta plataforma em contextos redatoriais permitirá perceber até que ponto a automação de tarefas associadas à produção jornalística poderá ajudar os meios de comunicação em transição para o digital, assim como contribuir para uma maior eficiência nas tarefas associadas ao jornalismo de ciência, permitindo uma maior massificação e qualidade na produção de notícias científicas no panorama mediático atual.**

**Palavras-chave:**
**Jornalismo, Ciência, Automação, Dashboard, Inovação**

**Abstract:**
**The Algorithmic Science News (ASN) platform is a new tool created by a multidisciplinary team that emerged from the need to reinscribe the role of science news in today's media space. The platform aims not only to increase the number of scientific news available to publishers, but also to reduce the effort associated with more time-consuming tasks such as collecting data and analyzing scientific articles and thereby facilitating the entire news production process. The ASN platform includes a set of features designed to support the journalist's standard procedure in writing contexts, allowing the use of documents in open access scientific repositories. The development of algorithms working on these repositories followed the purpose of facilitating the access and exploration of these collections and allowing the media to use them as an information source.**





**In this project, we developed tools of reading and interpretation, writing and semantic suggestion. With regard to reading and interpretation, ASN helps on finding specialists related to the scientific article, summarizes the most determinant parts, namely the introduction, objectives, methodologies and results, present definitions of technical terms, and suggest projects related to the topic. With regard to the writing part, the platform allows to write notes related to parts of the article and have access to phrases suggestions. The use of a beta version of this platform in writing contexts will allow the realization of the extent to which the automation of tasks associated with journalistic production can help the media in transition to digital, as well as to contribute to a greater efficiency in the tasks associated with science journalism, allowing a greater massification and quality in the production of scientific news in the current media landscape.**

**Keywords:**
**Journalism, Science, Automation, Dashboard, Innovation**

The Algorithmic Science News Agency (ASN) project comes from the fusion of three particular research areas: Journalism, Computer Science and Science Communication, seeking to develop an innovative set of tools that add value to the current panorama. With the advent of the World Wide Web in the early 1990s, the digital distribution of news and the changing economic landscape associated with the sale of news led to a profound transformation of computing within the profession of journalism: instead of being perceived as an exotic interpretation of journalism, or a set of useful tools, computing was instrumental in maintaining the relevance of journalism and its economic sustainability within a different and innovative network information ecosystem (Andersen, 2017: 2). It is possible to define science journalism as a bridge between the lay public and the developments within science.

Science journalism is a form of communication that aims to equalize informative access to complex contents, only available, in a first stage, to the scientific community. Scientific journalism or science journalism arises from the need to '*democratize access to scientific knowledge and establish conditions for [the] so-called scientific literacy. It contributes, therefore, to include the citizens in the debate on specialized subjects that can impact their life and their work*' (Bueno, 2010: 5). As far as science news is concerned, the journalist is confronted with a large number of scientific studies that are not always in the area of his knowledge. This leads to taking more time and attention to understand technical concepts and to seek validation of results with authors and researchers who are not always geographically close.

If, within this process, we take into account the use of repositories, instead of the press releases issued by communication agencies, the time spent performing the tasks inherent to the journalistic process, rises considerably. There are certain tasks in the journalistic procedure that are standardized and that can benefit from a computational support insofar as this support allows a reduction of the time in some slow and less appealing tasks. It is through computer based tools, namely in the development





of a dashboard that replicates some tasks of science journalism, that the ASN project has developed its features. The use of dashboards has been growing with the transition from paper to digital and today plays a crucial role in making journalism processes more profitable in an era where speed is valued and necessary for efficient management of the dissemination of news content.

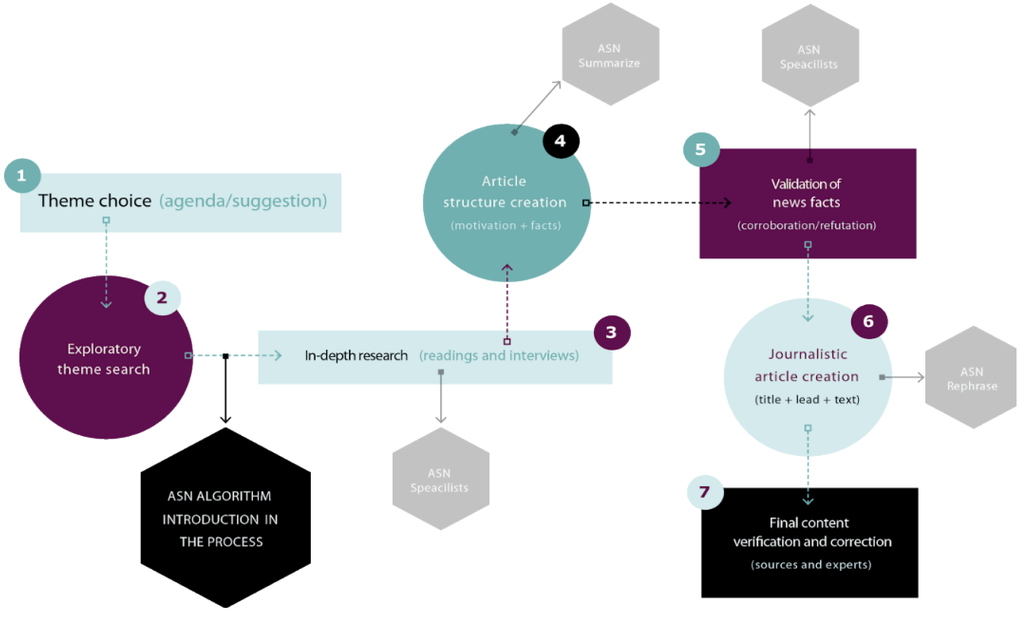
> 'Internal product platforms are those reusable tools that enable an organization to work more efficiently in their production processes. For news organizations, this might include everything from publishing tools, to information gathering and management systems, to analytics dashboards' (Diakopoulos, 2016: 3).

With regard to journalism specialized in science, ASN is pioneer in transforming a large volume of information, present in the main news sources, into smaller blocks that are easier to interpret by journalists.

## The Standard Journalistic Procedure

The journalist who writes or 'constructs' a general narrative follows the same steps as a journalist who has specialized in portraying science subjects. *'Many obstacles confront the (...) journalist, including limitations of time and space, editorial priorities, and the need to create stories that are compelling enough to warrant space in a publication'* (Gordon et al, 1999, 1). The usual journalist procedure (described in the diagram below) is a way to understand what are the usual steps of a journalist from the moment he knows the theme to explore to the design and publication of news.

**Figure 1. Journalistic workflow inspired by *A Journalist's Guide to Writing Health Stories from* Gordon et al.**




**Source: Own elaboration**

It all begins with the selection of the topic the journalist must investigate to create his story. Often this choice is up to the editor-in-chief, who might have very different purposes in mind depending on the target audience. In a magazine or media organization dedicated solely to the dissemination of science content, the most popular topics among readers are in the area of medicine and biology. Bauer in his 2000 work, Rooyen in his 2002 work and Bucchi and Mazzolini in his 2003 work (quoted in Badenschier and Wormer, 2012: 61) state that *'regarding the top fields of science covered by media, there are typical patterns which seem to be internationally consistent. Altogether medicine/health and biology dominate science coverage worldwide*'.

After determining the central theme of the investigation, the journalist resorts to the sources that are most reliable to do an exploratory research on the subject. When the information the





journalist wants to analyze is found in a press release, the remaining procedures may or may not happen, since the essential information to keep is in these documents. Generally, press releases are written by communication professionals, who make information treatment more accessible, specifically '*(...) the title of a scientific article, [and] the text, [which] are usually far too complex for the common reader. A press release simplifies the information and interprets it in a context that transforms it into news*' (Semir, 2000: 3).

In case there are no press releases, it is common for the exploratory research to try to find news that have already covered the same topic. From the moment that the exploratory research provides the data of interest to the news publication, the possibility of deepening the theme through the full reading of the scientific article is evaluated. It is at this stage that the ASN platform intends to support the journalist, through the analysis and summarization of the information contained in the scientific article. In order to shorten the time for reading, analysing and interpreting the paper, the ASN platform will present more objective and concrete data about the essence of the article to be explored. Additionally, the display of contact data regarding the author and institution where the study was developed provided by the ASN platform will allow a shortcut to develop news that would otherwise have consumed more time and dedication. Regarding the creation of the news skeleton, that is, the essential parts to be addressed in the body of the news article, the ASN platform will allow access to a brief summary of what appears in the article, identifying more clearly which methods are used, what are the research objectives and the results obtained.

Finally, the ASN platform aims to facilitate the process of creating narratives that make up the news piece through the possibility of rephrasing and construction of textual suggestions. Following the publication of the news article in media organizations or in the ASN news library, it will be possible to cross-reference information and data likely to be included in other science journalism pieces or projects.

## Existing Dashboards in Journalism

The ASN platform presents itself as a complex dashboard that puts in the same place resources and suggestions that allow to shorten the time spent in the analysis, interpretation and transformation of data into news stories. Hermida says that

> 'more than other media, such as television or newspapers, digital platforms can offer science journalism a greater diversity of coverage and voices. The multimedia, nonlinear, and networked nature of online journalism is forcing journalists to rethink storytelling for a digital age' (2010: 86).

The use of dashboards in newsrooms is not new, so the analysis of pre-existing tools in this scope will allow to frame the ASN project in scientific research and media innovation. With regard to platforms, more specifically dashboards used in newsrooms, it is crucial to identify what has already been created to obtain a more realistic perception of the knowledge developed and the application of this knowledge in real contexts. Within the types of dashboard that help the journalist in the usual procedure of creating news we identify Tweetdeck and Netvibes.

**Figure 2 and 3. Tweetdeck and Netvibes dashboard**





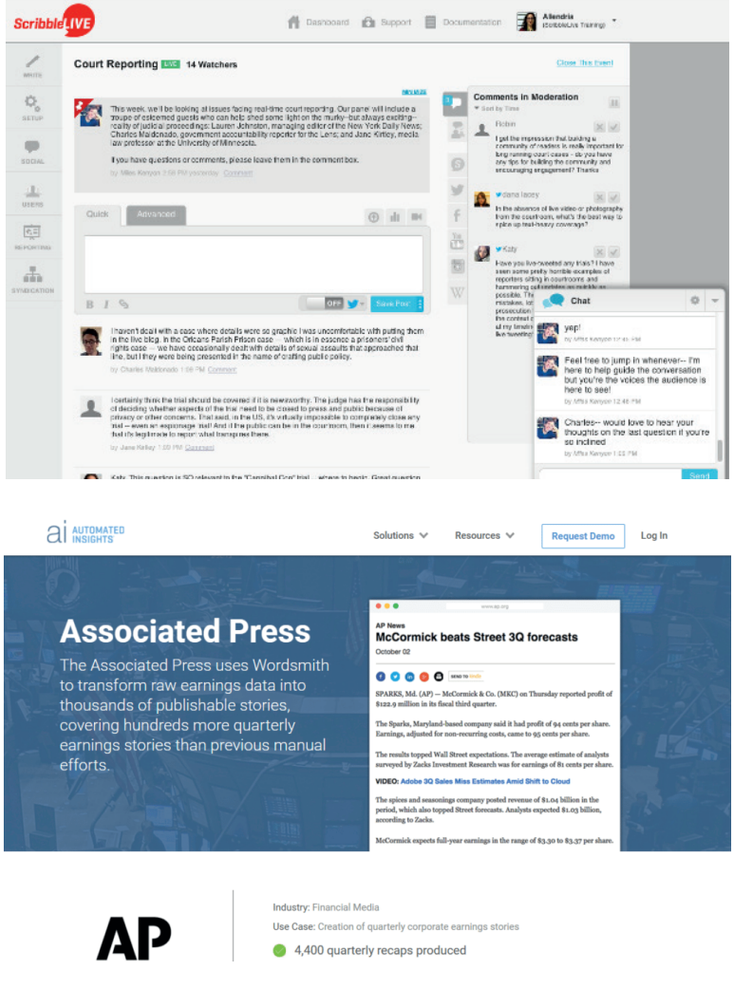

**Source: https://tweetdeck.twitter.com/, https://www.netvibes.com/en**

These two dashboards promise to reduce search time, as *'RSS readers like Netvibes mean that we do not need to check multiple websites or perform multiple searches to see if new information has been published or shared'* (Bradshaw, 2015). Tweetdeck and also Hootsuite focus more on their activity in social networks, with the possibility to publish through the dashboard *'and some cases access analytics'* (Bradshaw 2015). The *ScribbleLive* platform, launched in 2008 by two employees of Canadian television network, Michael de Monte and Jonathan Keebler, focus its purpose in real time storytelling. Among its clients we can list Reuters, MSN, CBS and CNN. This platform allows journalists to cover events in real time using their own content or user generated content. Publications can be updated not only by the online interface, but also through iPhone, Android, Blackberry, SMS and voice-mail applications. It allows, in a way, to write a story based on isolated facts that appear online. There is also the possibility for journalists to publish their content directly through the computer and mobile phone, with the possibility of editing this content in real-time (Guerrini, 2013: 16).

In 2014, the Associated Press publishers started testing simple

algorithms from software provided by Automated Insights in Durham, North Carolina, which generated earnings reports on listed companies. This software is used to generate news assuring the continuity of news operations by increasing efficiency with automation of certain editorial tasks, which are more time-consuming and laborious (Marconi and Siegman, 2017: 16).

**Figure 4 and 5. ScribbleLive and Associated Press dashboard**



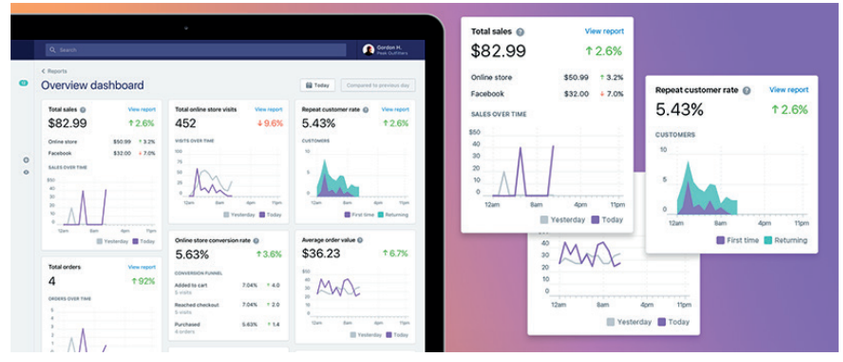

**Source: https://www.scribblelive.com/, https://www.ap.org/en-gb/**





The Overview tool offers the ability to cluster documents and view those clusters. This platform allows the custom creation of visual data, providing a lot more flexibility and adaptability to the needs of different kinds of investigations. DocumentCloud project serves as a repository for collecting and annotating documents. The documents become a substrate for investigative journalism work that can be built upon using the API.

**Figure 6. Overview dashboard**



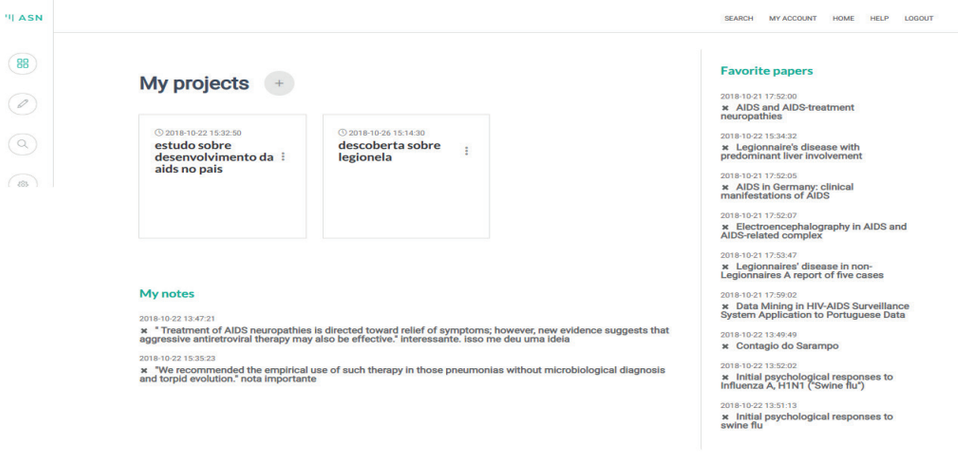

**Source: Screen caption of The Ovewrview platform**

## The Algorithmic Science News Platform

The ASN platform presents itself as a set of tools that articulate each other in order to reduce the time spent on more demanding tasks of the journalistic procedure, as well as to support the completion of more complex tasks, through the use of digital automation tools. In the early stages of development of the dashboard, algorithmic mechanisms were used in order to automatically produce scientific content news. As this process raised certain obstacles to the full development of the algorithm, it was outlined that a semi-automated device would be developed to support the story-making procedure in scientific journalism. The first output of this project, developed by our research team, was an instrument for research and analysis of the scientific articles found in

the open repositories of the Universities of Oporto, Minho and Trás-os-Montes and Alto-Douro (Figure 7). This tool allows the search of articles, by keywords and by the name of the author of the article. It also allows to filter searches by applying pre-determined filters. All the features that this primary dashboard presents mainly help the phases of preliminary exploration and further research from the scientific article.

**Figure 7. ASN Open Screen**




**Source: Screen caption of the ASN platform**

In terms of the philosophy of the tool, it is important to emphasize the main objective of the whole project, which consists in the development of a tool that allows media innovation and reinsertion of the role of the repositories, not only by the academic community, but also by the media communication and media diffusion communities. To describe all the features of what has been developed, we will separate the functionalities of the dashboard, in specific purposes within the usual tasks of journalists. In the first stage it is possible to use the dashboard to find scientific articles, experts and work related to a specific topic. For this purpose there is a search box (similar to a search engine) that allows





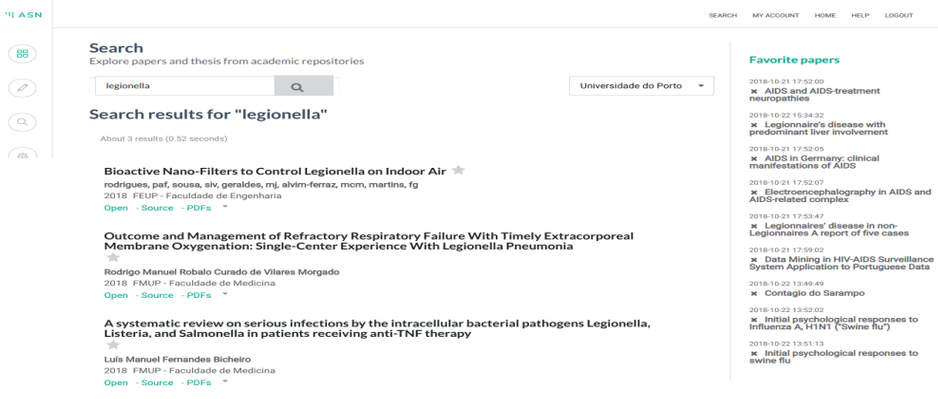
writing the topic to look for and get a list of related articles and authors (Figure 8). Eventually this functionality will also be useful in more advanced phases of research, particularly when further research is required on a particular topic or the cross-linking of research to determine the accuracy of some assumptions.

When selecting a specific article, the ASN dashboard allows to select more relevant text excerpts and write an explanatory note next to the selection. From the moment an article is selected, more functionalities unfold from those initially described, allowing a better understanding of the data included in the scientific article. Regarding data visualization, the ASN dashboard allows you to open the document on the dashboard itself, download the article file and save the link in bookmarks. Saved articles are displayed chronologically. The presentation of the articles can be changed between the articles with more relevance or the more recent articles. When a journalist has to dig up a subject to discover certain information, or to deepen the knowledge through investigation, the methods used depend on the type of news to be developed and the type of medium that will be disseminated on. On the ASN dashboard, the journalist (potential user of the tool) is able to guarantee a better analysis and deepening of information, through utilities developed to facilitate the understanding of a certain subject (Figure 8).

Regarding the analysis of the article, the ASN dashboard is divided into two distinct screens. In one panel it is possible to analyze and annotate the text of the article in question. On a second panel it is possible to find specialists, obtain a summary of the article, taking into account the introduction, methods, results and conclusions, and have access to definitions of more complex terms and articles related to the article under review. This panel of support to the reading and in-depth research of the subject, allows an articulation of utensils that can facilitate the work of the journalist, being therefore necessary to be executed in a single window. In the tab that allows to find the specialists, is shown information about the institution that hosted the investigation, the year of publication of the article, other investigations of the same

author (Figure 9), and his contacts (e.g. email). In what concerns the functionality of obtaining a summary of the text under review, a number of sentences is collected to reflect the essence of the introduction chapter, a basic reference to the methods used, the summary presentation of the results obtained in this study and the main conclusions of the research.

**Figure 8. ASN Search Screen**




**Source: Screen caption of the ASN platform**

In the definitions tab, a more general, more accessible language definition of more complex scientific terms is presented and has to be simplified for audience understanding. Finally, in the papers tab, a list of papers is presented that somehow has points of intersection with the article under analysis. All the tools described throughout the text are pertinent and fundamental in the processes of in-depth reading, content analysis and verification of the facts. Usually, the journalist checks the information to be published, through field research, interviews, or physical interactions with the sources of his news. Here, the contact of the investigators can be decisive in the discovery and verification of the facts. Writing a news report presupposes a basic knowledge of the topic to be dealt with, a solid fieldwork, with validation of the facts and collection of declarations that validate the content. Because it is a phase that highlights the capabilities of analysis, synthesis and interpretation of the journalist in a given subject, text-formatting tools are usually used as in the case of Microsoft Office Word. The ASN dashboard does not propose to take the place of the most used tools for writing of text, but to complement them with utensils to help the work of the journalist.

In the first instance, the ASN user can read, analyze, annotate and obtain relevant information from a particular essay, investigation or scientific project (Figure 9). At another stage it is possible to enter the writing mode on the dashboard and get a long box of text to enter the intended narrative. But most relevant in this panel are the features that appear next to the text input box. It is possible to consult the notes, previously inserted, in the reader of the articles, it is possible to rewrite certain words or phrases, get definition of terms and articles relevant to the writing of the news. Concerning notes, they are displayed chronologically, giving a brief summary of the note in question and which of the articles refers to the note. Regarding the possibility of rewriting parts of the text, the journalist will have to select the desired word or phrase and wait for the automatic generation of text to reveal a simpler definition of what he wanted to inform. Concerning the definitions and papers tab, these tools are similar to those that appear on the reading and understand panel, i.e. obtaining more basic definitions of complex terms and articles that have parts intersecting with the one under analysis.

**Figure 9. ASN Reader Screen**


**Source: Screen caption of the ASN platform**

The tools that have been developed to help in the process of writing news assume an integration of automated processes that will help the journalist to develop news content using his capacity and knowledge allied to the results suggested by the platform. One of the most important processes in the journalist's work is to apply his knowledge, research, conduct interviews and fieldwork so that this information reaches the target audience. The writing process is one of the most sensitive, but also more transformative in the life of a journalist. The purpose of a journalist is to use its ability to analyze and interpret events and create narratives capable of informing and clarifying a subject, allowing information access to all. In the case of science journalism this work becomes strictly necessary, insofar as the technical and scientific language used to describe the investigations is not perceptible to the majority of the population.





## Conclusion

The ASN tool proposes, with the application of its functionalities, a new way of looking at scientific knowledge and dissemination of this same knowledge. It presupposes a democratization of information access, through the easier diffusion of scientific content and the adaptation of the language to a discourse more popular and accessible to all. By focusing on innovation through digital and technological support of automation processes, the ASN dashboard reinvents the role of science journalism, incorporating in the process of digital dissemination the consumption of more appealing contents. The ASN dashboard thus positions itself as the right arm of innovative and democratic scientific journalism.

## References

ANDERSEN, C.W. 2017. Computational journalism. In the *International encyclopedia of journalism studies*.

BADENSCHIER, F. and WORMER, H. 2012. "Issue selection in science journalism: towards a special theory of news values for science news?". In the sciences media connection – public communication and its repercussions, sociology of the sciences yearbook 28. Acedido em 15 de abril de 2018. DOI: 10.10007/987-94-007-2085-5-4.

BRADSHAW, P. 2015. "Dashboards and journalism: why we need to do better". In Online Journalism. Acedido em 20 de maio de 2018. https://onlinejournalismblog.com/2015/07/01/dashboards-and-journalism-why-we-need-to-do-better/

BUENO, W. 2010. "Comunicação científica e divulgação científica: aproximações e rupturas conceituais". In Londrina. Acedido em 20 de abril de 2018. DOI: 10.5433/1981-8920.2010u15resp.p1.

DIAKOPOULOS, N. 2016. "Computational journalism and the emergence of news platforms". In *The Routledge Companion do Digital Journalism Studies*. Eds. Scott Eldridge II and Bob Franklin. Fondef D08I1082.

GÓMEZ-ZARÁ et al. 2015. *Work Key News: a web platform to support journalistic processes and editorial meetings*.

GORDON et al. 1999. "A Journalist's Guide to Writing Health Stories". In *AMWA Journal*, Vol. 14, No. 1.

GUERRINI, F. 2013. "Newsroom curators and independent storytellers: content curation as a new form of journalism". In *Reuters Institute Fellowship Paper*. University of Oxford.

HERMIDA, A. 2010. *Revitalizing science journalism for a digital age. In Science and the Media*. Pp. 80-87.

MARCONI, F. and SIEGMAN, A. 2017. "The future of augmented journalism: A guide for newsrooms in the age of smart machines". In *Associated Press Insights*. Acedido em 15 de outubro de 2018. https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf



# Automated news production within the uses and professional practices