**Traceable Original**

**Journalistic Content (TOJC)**

**Projet IMI**

Etat de l’art

Octobre 2019 - Version 0.1

**Auteurs:**

HES-SO Valais/Wallis - IIG

# Etat de l’art pour premières discussions

Le besoin de regagner la confiance des lecteurs est on ne peut plus d’actualité pour les médias qui doivent de plus faire face au défi de la numérisation. Ces difficultés vont de pair avec les préoccupations des gouvernements pour qui la prolifération de la désinformation est tout aussi nuisible. Dans son rapport de 2018 "A multi-dimensional approach to disinformation"[[1]](#footnote-1), la Commission européenne souligne l'importance d'une approche multi-dimensionnelle, basée sur la transparence, le développement d'outils qui supportent les journalistes pour un travail de qualité, la sauvegarde de l'écosystème des médias au niveau européen, et l'importance de la recherche dans ce domaine afin de suivre l'évolution du phénomène et trouver les moyens efficaces de le contrer. Lors de la conférence nationale "Suisse numérique 2019"[[2]](#footnote-2), début septembre 2019, Mme Sommaruga, cheffe du département de la communication, a souligné les défis relatifs aux médias, l'importance pour notre démocratie des journaux régionaux qui doivent faire face aux grands groupes, ainsi que le soutient du Conseil fédéral pour l'offre payante des médias en ligne.

En 2017, la Commission fédérale des médias (COFEM) a partagé dans un rapport[[3]](#footnote-3) sur l’avenir des médias, son intérêt à mettre sur pied un label de qualité qui permettrait de reconnaître les contenus journalistiques satisfaisant aux normes de qualité minimales et de les distinguer des offres alternatives telles que les blogs ou les stratégies numériques de communication commerciale (marketing de contenu, native advertising, branded content, etc.). Ce rapport souligne que “les utilisateurs ont toujours plus de difficultés à distinguer les contenus véritablement journalistiques des contenus qui ne satisfont pas aux exigences minimales d'un journalisme professionnel (par exemple la vérification des sources, la pertinence, la distance, l'exactitude, la classification et la transparence)”.

Du côté des agences de presse, s'agissant d’une situation qui a décliné depuis le début des années 2000[[4]](#footnote-4), des initiatives concrètes ont déjà vu le jour et/ou sont en cours d’expérimentation. Il faut d'autre part noter que cette problématique est généralisée au niveau du Web, car personne ne veut véhiculer de fausses informations, surtout pas les acteurs majeurs comme les grands réseaux sociaux ou les moteurs de recherche qui y jouent aussi leur crédibilité.

Les solutions proposées s’attaquent au problème sous différents angles :

* Publier les métadonnées des articles, notamment des informations structurées, souvent prisonnière des systèmes internes des médias, pour les rendre ainsi visibles et exploitables par les moteurs de recherche notamment
* Définir un label qualité qui certifie et regroupe les auteurs de bon contenu journalistique (logo, note de qualité sur l'auteur ou l'article, etc.)
* Améliorer la transparence lors de la publication d'articles, notamment en citant les sources et en donnant du contexte
* Juger de la qualité et de l'originalité de l'article par analyse humaine, automatique, voire semi-automatique (Fact Checking, etc.)

Les moteurs de recherche eux-mêmes améliorent leurs algorithmes afin de mieux évaluer le contenu original et pertinent. Au moment d’écrire cet état de l'art, Google vient d’annoncer en Septembre 2019 la modification de leurs algorithmes avec pour but d'éviter qu'un partage ou une recopie d'une information originale ne sorte en premier dans les résultats de recherche[[5]](#footnote-5), d'autant plus avec ce risque que l'information originale soit modifiée ou falsifiée.

Ces géants du Web ont entrepris depuis quelques années de lutter contre la désinformation. La Google News Initiative (GNI[[6]](#footnote-6)) a permis, par son fond de recherche DNI, de soutenir un grand nombre de projets en la matière : 559 projets dans 30 pays européens, pour un total de 115 millions d'Euros[[7]](#footnote-7). La vision du GNI est large :

* Élever et renforcer un journalisme de qualité
* Faire évoluer les modèles commerciaux pour générer une croissance durable
* Donner plus de pouvoir aux organes de presse grâce à l'innovation technologique

Les moteurs de recherche poursuivent aussi la promotion des métadonnées structurées par les fournisseurs de contenu eux-mêmes, données plus facilement exploitables par les machines, ceci par leur initiative schema.org[[8]](#footnote-8) qui rencontre depuis sa mise en œuvre un grand succès. Ils proposent des recommandations particulières pour les éditeurs d'articles[[9]](#footnote-9) ainsi que les fournisseurs de revues dans le cadre du fact-checking[[10]](#footnote-10). Cette démarche est en phase avec le W3C qui travaille sur le Web des données[[11]](#footnote-11), notamment par la mise en œuvre du Linked Data[[12]](#footnote-12) dans le cadre de la vision du Web Sémantique par Tim Berner's Lee lui-même.

Une autre approche en cours consiste à mettre sur pied un label reconnaissant le journalisme de qualité. Le Trust Project[[13]](#footnote-13) évalue la fiabilité d'une agence de presse qui, pour obtenir leur label, doit répondre à 8 Indicateurs (todo : à lister) de confiance. Le projet est mené par la journaliste Sally Lehrman[[14]](#footnote-14), co-fondé par Richard Gingras qui dirige Google News et donc soutenu par GNI, et regroupe 120 agences de presse. La mise en œuvre et la définition de ces indicateurs est basée sur une étude de 2 ans menée aux Etats-Unis et en Europe, et orientée satisfaction des lecteurs[[15]](#footnote-15) dont il ressort : "En fait, ils ne veulent pas simplement éliminer les imposteurs. Ils veulent savoir qui a écrit ou produit une histoire, quelle expertise ils ont et si l'éditeur a un agenda". Ce projet réuni des journaux tels le Washington Post, The Economist, Mic, the Globe and Mail (Canada) et Zeit Online (Allemagne), mais aussi les géants du Web comme Google, Facebook et Bing qui utilisent les indicateurs de différentes manières. Le projet fourni un logo sécurisé à l'attention des agences agréées, ainsi que des plugins WordPress et Drupal pour générer du marquage basé sur schema.org.

Dans la même idée de promouvoir du journalisme de qualité, un projet du même type, soutenu lui aussi par une centaine d'agences ainsi que le GNI, est le Journalism Trust Initiative (JTI). Il est mené par Reporters sans frontières et ses partenaires Agence France Presse (AFP), European Broadcasting Union (EBU), et Global Editors Network (GEN). Leur but est "de soutenir la liberté d’opinion universelle et individuelle grâce à l’accès à l’information et à des médias indépendants et pluralistes". Le JTI évoque une collaboration souhaitée avec le Trust Project, mais aucun détail n'est vraiment disponible à ce sujet. Ce projet devrait se terminer fin 2019 avec comme résultat un Workshop Agreement du Comité Européen de Standardisation (CEN[[16]](#footnote-16)) qui est ouvert pour une révision publique[[17]](#footnote-17) dès juillet 2019.

Le Credibility Coalition[[18]](#footnote-18), un autre groupe interdisciplinaire de libraires, journalistes et chercheurs, pousse la collaboration avec le W3C en créant le Credible Web Community Group[[19]](#footnote-19). (Zhang et al., 2018), leur papier de recherche qui a aussi contribué au développement du GNI, souligne le besoin de fédérer tous les types d'initiatives en cours : les labels de qualité (comme le Trust Project), les techniques de scoring et labellisation automatique, la vérification des faits (fact-checking) par des experts, les compagnes pour améliorer l'alphabétisation et le savoir des lecteurs ainsi que les annotations par crowdsourcing. Ils proposent des indicateurs de crédibilité qui peuvent être évalués par tout un chacun, en se fiant au jugement humain et avec un peu d'entrainement, sans devoir être un expert du domaine. En septembre 2019 les activités de ce groupe sont en cours d'organisation.

L'approche de NewsGuard[[20]](#footnote-20) va dans le même sens, mais il s'agit d'une approche commerciale où des journalistes donnent des notes aux différentes sources en fonction de 9 critères (todo : à lister), pour indiquer s’ils font du bon journalisme ou non. Leurs analystes ont pour l'instant noté des agences des U.S, Royaume-Unis, Allemagne, France et Italie, et d'autres pays sont à venir. Le projet est développé par une startup co-fondée par Knight Foundation, l'agence de presse Publicis Groupe et différents individus[[21]](#footnote-21). Concrètement ils fournissent un plugin pour navigateur qui affiche une icône propre au site consulté (Figure 1) et qui permet d'obtenir plus d'information (full nutrition label) sur leur analyse de ce fournisseur d'information.

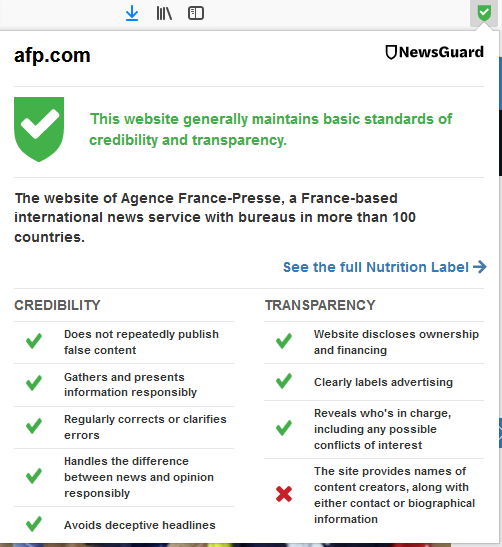


Figure 1- Note de NewsGuard pour le site AFP

Du côté de la traçabilité de l'information, la plupart des initiatives soulignent l'importance de la transparence, qui est aussi la première action proposée par ce rapport 2018 de la Commission européenne. Actuellement cette transparence repose entre autres sur la mise en ligne de données structurées afin d'expliciter d'où vient l'information, qui en sont les auteurs, les sources, etc. La bonne nouvelle est que l'on trouve ici une mise en œuvre commune aux différentes initiatives, basée sur le vocabulaire de schema.org cité ci-dessus. Une approche pionnière en matière de transparence par la publication de métadonnées structurées fut celle du New York Time et de l'International Press Telecommunications Council, nommée rNews[[22]](#footnote-22), qui a cherché dès le départ la collaboration avec schema.org et est cité sur leur page de recommandation pour les articles de journaux[[23]](#footnote-23).

De cette nécessité de transparence, de citation des sources et d'enrichissement du contexte est née la plateforme DocumentCloud[[24]](#footnote-24) analysée par (Mor & Reich, 2018). Cette plateforme destinée aux journalistes leur permet d'y stocker et d'y annoter des documents qu'ils joindront aux articles afin de fournir du contexte (des preuves, du complément d'information) pour regagner la confiance des lecteurs. Les 2-3 millions de documents transférés par des centaines d'agences, pour la plupart américaines mais aussi Agence France Presse et SwissInfo.ch, témoignent du succès de l'initiative qui palie au problème de citation de source par lien hypertexte : la mise à disposition d'un fichier au format .pdf n'entrainera pas le lecteur vers un autre site, et garanti aussi que cette source n'est pas modifiée dans le temps.

Un autre projet concret bien avancé en la matière mais de bien plus petite envergure est le Transparency Journalism[[25]](#footnote-25). Il a reçu un financement du fond DNI[[26]](#footnote-26), ce qui a permis la réalisation d'un outil open source nommé TJ Tool. Il s'agit d'un plugin WordPress qui aide les journalistes à publier l'article et les données structurées, tout en y apportant de la transparence et du contexte. Cet outil démontré dans des vidéos de présentation de leur site est mis en œuvre par le média espagnol Publico[[27]](#footnote-27). Il n'est actuellement disponible qu'en espagnol et catalan, et ne semble pas utilisé par d'autres agences que Publico. Chaque article est publié avec une icône permettant d'accéder à la "Carte de transparence" de l'article (voir figure X) qui présente une note sur 100%.

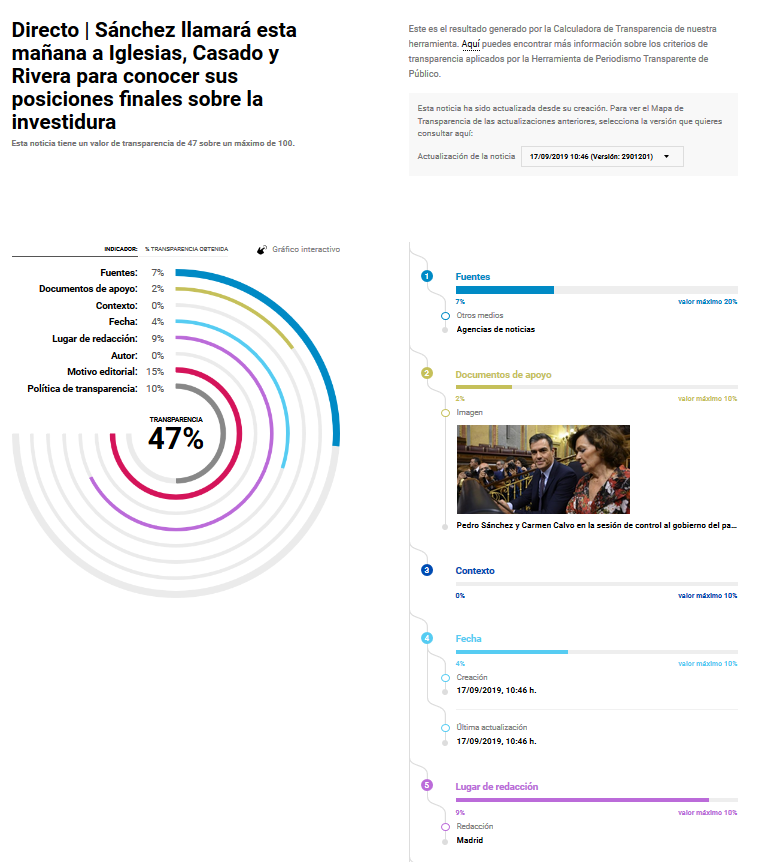


Figure 2- Carte de transparence de TJ Tool

Les 8 indicateurs composants cette note sont :

* Rendre publique notre politique éditoriale
* Expliquer pourquoi nous choisissons une certaine manière de dire les choses
* Indiquer le lieu où nous écrivons l'information
* Byline (signature)[[28]](#footnote-28): donne-t-il un nom complet, des initiales ou une dénomination générale pour la salle de presse
* Sources mentionnées - Personnes, institutions, autres médias et même une source non identifiée
* Joindre des documents de support (photos, doc, vidéos, audios, graphiques…)
* Des liens vers d'autres documents contextuels
* Donner les dates de publication et de mise à jour, ainsi que les dates de tous les documents

Concernant l'identification des articles, les projets qui le font reposent sur leurs adresses Web de publication (URLs), ce qui colle parfaitement avec la vision W3C du Web des données/Linked Data[[29]](#footnote-29). Malgré tout certaines plateformes proposent la gestion de contenu et des documents ainsi que d’identifiants. C’est le cas de figShare[[30]](#footnote-30) qui permet d‘héberger le contenu des chercheurs scientifiques mais aussi des éditeurs, et qui gère automatiquement les identifiants DOI pour ce contenu. Il s’agit ici d’une plateforme commerciale qui fournit un API et qui pourrait jouer un rôle similaire au DocumentCloud mentionné précédemment.

Il est à noter que de la recherche et des applications sont en cours pour voir l'implémentation des technologies Block-Chain afin de fournir des identifiants immuables pour les articles. Comme il ne s'agit pas du cœur de notre projet, nous citerons simplement le projet Pix[[31]](#footnote-31), soutenu par le fond DNI, le News Provenance Project[[32]](#footnote-32) basé sur la solution d'IBM[[33]](#footnote-33), et enfin la solution safe.press[[34]](#footnote-34) adoptée par Orange[[35]](#footnote-35) qui le présente comme "le premier label de confiance numérique au monde pour les informations en ligne, sécurisé par la blockchain". Les arguments avancés pour la transparence basée sur cette technologie sont la traçabilité, l'approche décentralisée et l'aspect immuable de cette information.

Pouvons-nous à trouver des infos sur ceci :

Les documents JTI parle d’une initiative Suisse ISO 9001:  
“A second project launched in Switzerland about ten years ago adopted an existing, general standard on Quality Management Systems (ISO9001) and attempted to apply it to the media industry. A respective entity was created to sell certification, but ceased operations in 2016.”  
  
Il serait peut-être intéressant de voir de quoi il s'agissait, pourquoi cela s'est terminé, mais je n’ai rien trouvé si ce n’est cette mention du projet JTI.

# Autres informations pour discussions

Voici encore deux points sensibles dont il faut, à notre avis, tenir compte dans ce projet. D'une part l'importance de fédérer les approches afin d'éviter un effet contre-productif si le lecteur se retrouve face à une multitude de solutions/logos/schémas à devoir comprendre et analyser pour les différents médias qu'il consulte, voire mettre en œuvre une solution swiss-made si cela fait du sens. Et d'autre part veiller à ce que ces initiatives sauvegardent la liberté d'opinion, comme le rappel Reporters sans frontières dans leur conclusion du rapport 2018 de la Commission européenne :

**"Reporters sans Frontières** – RSF voted in favour of the report with the following motivation:  
RSF supports most parts of the report, for example the description of the problem and the underlying principles and freedoms at stake. As regards the recommendations, however, the composition of the HLEG and also procedural limitations, namely the consensual approach, left little room for dissenting opinions. RSF remains firmly opposed to the approach of a ‘Coalition’ as long as its mandate, the appointment procedures and governance mechanisms remain unclear. Given the sensitivity of the matters at stake, such a supra-national entity should be launched only with robust safeguards against undue influence in place."

# Bibliographie

Amy Zhang, Aditya Ranganathan, Sarah Emlen Metz, Scott Appling, Connie Moon Sehat, Norman Gilmore, Nick B. Adams, Emmanuel Vincent, Jennifer 8. Lee, Martin Robbins, Ed Bice, Sandro Hawke, and David Karger. A Structured Response to Misinformation: Defining and Annotating Credibility Indicators in News Articles. The Web Conference, April 2018.

Mor, N., & Reich, Z. (2018). From “Trust Me” to “Show Me” Journalism: Can DocumentCloud help to restore the deteriorating credibility of news?. *Journalism Practice*, *12*(9), 1091-1108.

1. <https://blog.mozilla.org/netpolicy/files/2018/03/HLEG-on-Fake-News_Final-Report.pdf> [↑](#footnote-ref-1)
2. <https://strategy.digitaldialog.swiss/fr/conference/> [↑](#footnote-ref-2)
3. <https://www.emek.admin.ch/fr/actuel/apercu/> [↑](#footnote-ref-3)
4. <https://kf-site-production.s3.amazonaws.com/publications/pdfs/000/000/242/original/KnightFoundation_AmericansViews_Client_Report_010917_Final_Updated.pdf> [↑](#footnote-ref-4)
5. <https://searchengineland.com/change-log-google-updated-its-search-quality-rater-guidelines-321549> [↑](#footnote-ref-5)
6. <https://newsinitiative.withgoogle.com/> [↑](#footnote-ref-6)
7. <https://newsinitiative.withgoogle.com/dnifund/dni-projects/> [↑](#footnote-ref-7)
8. <https://schema.org/> [↑](#footnote-ref-8)
9. <https://schema.org/docs/news.html> [↑](#footnote-ref-9)
10. <https://schema.org/ClaimReview> [↑](#footnote-ref-10)
11. <https://www.w3.org/2013/data/> [↑](#footnote-ref-11)
12. <https://www.w3.org/standards/semanticweb/data> [↑](#footnote-ref-12)
13. <https://thetrustproject.org> [↑](#footnote-ref-13)
14. <https://thetrustproject.org/sally-lehrman/> [↑](#footnote-ref-14)
15. <https://thetrustproject.org/user-interviews-research-and-reports/> [↑](#footnote-ref-15)
16. <https://www.cen.eu/Pages/default.aspx> [↑](#footnote-ref-16)
17. <https://www.cen.eu/news/workshops/Pages/WS-2019-013.aspx> [↑](#footnote-ref-17)
18. <https://credibilitycoalition.org/> [↑](#footnote-ref-18)
19. <https://www.w3.org/community/credibility/> [↑](#footnote-ref-19)
20. <https://www.newsguardtech.com/> [↑](#footnote-ref-20)
21. <https://www.newsguardtech.com/about/our-investors/> [↑](#footnote-ref-21)
22. <https://iptc.org/standards/rnews/> [↑](#footnote-ref-22)
23. <https://schema.org/docs/news.html> [↑](#footnote-ref-23)
24. <https://www.documentcloud.org/> [↑](#footnote-ref-24)
25. <https://www.transparentjournalism.org/> [↑](#footnote-ref-25)
26. <https://newsinitiative.withgoogle.com/dnifund/dni-projects/transparent-journalism-tool-tj-tool/> [↑](#footnote-ref-26)
27. <https://www.publico.es/> [↑](#footnote-ref-27)
28. <https://en.wikipedia.org/wiki/Byline> [↑](#footnote-ref-28)
29. <https://www.w3.org/DesignIssues/LinkedData.html> [↑](#footnote-ref-29)
30. https://figshare.com/services/publishers [↑](#footnote-ref-30)
31. <https://newsinitiative.withgoogle.com/dnifund/dni-projects/pix-t-the-blockchain-driven-marketplace/> [↑](#footnote-ref-31)
32. <https://www.newsprovenanceproject.com/> [↑](#footnote-ref-32)
33. <https://www.ibm.com/blockchain> [↑](#footnote-ref-33)
34. <https://safe.press/> [↑](#footnote-ref-34)
35. <https://www.orange.com/en/Press-Room/press-releases/press-releases-2019/Orange-launches-Safe.press-the-world-s-first-digital-trust-label-for-online-news-secured-by-the-blockchain> [↑](#footnote-ref-35)