



HAL
open science

Semantic publishing, la sémantique dans la sémiotique des codes sources d'écrits d'écran scientifiques

Gérald Kembellec

► **To cite this version:**

Gérald Kembellec. Semantic publishing, la sémantique dans la sémiotique des codes sources d'écrits d'écran scientifiques. Les Enjeux de l'information et de la communication, 2019, 20 (2). hal-02423752

HAL Id: hal-02423752

<https://hal.science/hal-02423752>

Submitted on 25 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Semantic publishing, la sémantique dans la sémiotique des codes sources d'écrits d'écran scientifiques

Critical Analysis and Foresight of semantic publishing source codes: Semantics in Semiotics

Análisis crítico y previsión de códigos de fuente de publicación semántica: Semántica en semiótica

Gérald Kembellec

Gérald Kembellec est enseignant-chercheur en Sciences de l'Information et de la Communication au sein du Laboratoire Dicen-IdF, Cnam. Ses problématiques de prédilection sont l'IST avec une focale sur le Linked Data scientifique, les graphes de citations en lien avec les méthodes d'*écriture*. Il essaie de lier autant que possible ses réflexions, de manière interdisciplinaire, avec les humanités numériques.

Plan de l'article

Introduction

Sémiotique et *semantic publishing*,

Enjeux et opportunités du *semantic publishing*

Pourquoi parler de sémiotique du code ?

Sémiotique du code : typologie analytique des signes passeurs en *semantic publishing*

Les signes indicateurs de notices

Signalisation de termes, entités nommées et fragments

Les signes communicationnels à destination des médias sociaux

Discussion sur l'usage du *semantic publishing*

Formats, référentiels et gouvernance, une impossible neutralité

Apports pour les chercheurs : alignement sur les primitives savantes de la recherche

Prospective sur le *semantic publishing*

Les points de vigilance

Conclusion

Références bibliographiques

Résumé

Cet article analyse les enjeux du *semantic publishing* en contexte scientifique et examine sous un axe sémiotique les codes sources qui en sont le vecteur de propagation. Sont présentés et discutés les différents signes passeurs qui rendent possible le maillage de l'écriture fragmentaire en réseau : le RDFa, les microdonnées et le JSON-LD par exemple. Leurs usages sont ici analysés et mis en relation avec les besoins et objectifs des chercheurs, qu'ils soient auteurs ou lecteurs. Enfin, le futur du *semantic publishing* scientifique est anticipé de manière critique et des points de vigilance sont évoqués tant sur la gouvernance des autorités et des schémas qui étayent le *linked data* que sur les tentations d'user et d'abuser des bénéfices communicationnels annexes entre médiation et médiatisation.

Mots clés

Analyse sémiotique ; Écriture numérique ; IST ; Semantic publishing ; Code source ; Données liées (Linked Data) ; médiation ; Médiatisation

Abstract

This article analyzes the issues of semantic publishing in a scientific context and examines, under a semiotic axis, the source codes which are the vector of propagation. The various signs that make possible the meshing of fragmented semantic writing: the RDFa, the microdata and the JSON-LD, for example, are presented and discussed. Their uses are analyzed here and related to the needs and objectives of researchers whether authors or readers. Finally, the future of scientific semantic publishing is critically anticipated, and points of vigilance are mentioned both on the governance of the authorities and the schemes that underpin the linked data. We also discuss temptations to use and abuse the ancillary communication benefits between mediation and media coverage.

Keywords

Semiotic analysis Digital writing; IST; Semantic publishing; Source code ; Linked data

Resumen

Este artículo analiza los problemas de la publicación semántica en un contexto científico y examina, bajo un eje semiótico, los códigos fuente que son el vector de propagación. Se presentan y discuten los diversos signos de contrabando que hacen posible la combinación de la red de escritura fragmentada: RDFa, los microdatos y JSON-LD. Sus usos se analizan aquí y se relacionan con las necesidades y objetivos de los investigadores, ya sean autores o lectores. Finalmente, el futuro de la publicación semántica científica se anticipa críticamente y se mencionan puntos de vigilancia tanto en la gobernanza de las autoridades como en los esquemas que sustentan los datos vinculados y en las tentaciones de usar y abusar de los beneficios de la comunicación auxiliar entre Mediación y cobertura mediática.

Palabras clave

Análisis semiótico Escritura digital; IST; Publicación semántica; Código fuente ; Datos vinculados

Introduction

Les enjeux contemporains de la « mise en média de l'information scientifique » au travers des plateformes du Web, y compris celles du Web social, mettent en tension les questions informationnelles et médiatiques. Les états, l'Europe et la communauté scientifique internationale réclament, dans une politique d'ouverture de la science, la mise à disposition proactive des contenus à destination du plus grand nombre. Penser, anticiper la circulation médiatique des productions scientifiques est une gageure info-communicationnelle qui interroge les modalités de mise en œuvre d'une telle politique. En effet, il existe de nombreuses variables à appréhender pour proposer un formalisme qui satisfasse tous les acteurs de l'Information Scientifique et Technique. Comment envisager le *design* de l'information scientifique et concilier les besoins des différents acteurs de la

chaîne de l'IST dans un contexte où il y a autant de facettes à prendre en compte que d'acteurs des documentarisations auctoriale, éditoriale, diffusionnelle et d'appropriation (Zacklad, 2019) ?

Dans cet article, nous présentons une focale sur la sémiotique du code des écrits d'écran (Souchier, 1996 ; Jeanneret et Souchier, 1999) au sein des plateformes d'édition en ligne d'information scientifique et technique (IST). Nous proposons de mettre l'accent sur les modalités d'exposition des métadonnées liées aux écrits scientifiques, tout autant que sur leurs potentiels usages communicationnels. Pour ce faire, nous reprenons un positionnement qui trouve ses racines dans les travaux sur les raisons « graphique » (Goody 1979) et « computationnelle » (Bachimont, 1999), mais aussi dans la « théorie opérationnelle de l'écriture numérique » (Crozat et al., 2011). Sous ce positionnement nous examinerons l'écriture des articles scientifiques avec une « machine à calculer » (Crozat, 2016), pour des humains et d'autres « machines à calculer » (Crozat, 2016). Les méthodes propres au *semantic publishing* peuvent-elles proposer, sinon une réelle « rhétorique hypertextuelle » (Saemmer, 2015), tout au moins un réel bénéfice info-communicationnel dans la sphère scientifique ? La question sous-tendue par cette analyse est le bénéficiaire de ces techniques : l'auteur, l'éditeur ou le lecteur ? Ces enjeux médiatiques en information scientifique seront donc problématisés tout autant dans une optique d'inscription de fragments documentaires que de celle de la documentarisation auctoriale ou encore de celle des industries médiatiques qui éditent et distribuent l'IST, sans oublier enfin celle du lecteur final.

Sémiotique et *semantic publishing*

Enjeux et opportunités semantic publishing

Le *semantic publishing* (Verlaet et Dillaerts, 2016 ; d'après Shotton, 2009) est un modèle de publication qui offre de larges possibilités de mise en œuvre de documents numériques auto-décrits et inscrits dans un vaste maillage conceptuel. L'ajout de métadonnées lisibles par les machines permet aux robots et agents logiciels d'appréhender – sinon le sens – tout du moins la structure et le contexte d'un fragment d'information. En effet, ce dernier est inscrit au sein d'un document et plus largement dans un réseau d'autres fragments, concepts ou auteurs. Cette structure est rendue possible grâce à des « référentiels autorités » de désambiguïsation réputés communs ne laissant aucune place à l'interprétation, à la confusion typologique ou encore à la polysémie.

Cette rédaction sémantique, hautement technique pour le profane peut se faire par délégation scripturaire au moyen de nouveaux architectes. Ces derniers sont de véritables chaînes éditoriales sémantiques offrant une assistance structurelle à la rédaction, convoquant tout à la fois la liaison, l'inscription et la discrétisation des données (Bachimont, 2010 ; Crozat, 2012). Cette nouvelle méthode d'écriture commence à trouver un écho dans le monde scientifique avec des revues nativement numériques comme *PeerJ* dont les contenus sont individuellement accessibles et parfois sémantisés (Broudoux et Kembellec, 2017). Dans certaines revues, chaque paragraphe, illustration, tableau ou encore analyse de lecteur est identifiable et accessible par une adresse de ressource unique (URI). Chacun de ces fragments est ouvert aux commentaires, eux-mêmes identifiables et donc possibles à lier sémantiquement. Nous avons proposé un modèle (Kembellec et Bottini, 2017) qui descendait plus loin en granularité pour permettre à la fois la typologie et l'identification de fragments d'articles scientifiques en réseau. Les possibles usages du *semantic publishing* pour un chercheur-usager équipé, au sens de Bigot (2018), d'un dispositif de lecture augmenté sont multiples¹. Le lecteur peut ainsi (1) *a minima* acquérir et indexer automatiquement des notices

.....

¹ Par dispositif de lecture augmentée, nous entendons des équipements techniques qui s'agrègent au navigateur hypertexte pour éclairer sur les différentes sources de description. Cela peut aller du simple *plug-in* de gestion de références

bibliographiques. Il pourrait (2) encore désambiguïser les propos de l'auteur (3), mieux saisir ses intentions par la typologie des concepts mobilisés ou par la structuration du document en analysant l'agencement des fragments. En effet, le paradigme du *semantic publishing* qui lie sémantique et sémiotique au sens de Choi (2001), renégocie le document scientifique dans un réseau méta-informationnel qui n'est pas sans rappeler un certain « tissage » pour faire référence à la théorie du texte de Roland Barthes (Simon, 2015 ; Barthes, 1973). Enfin (4), la sémantique des contenus scientifiques permet de partager de manière simplifiée l'URL de documents scientifiques sur ses réseaux socionumériques de prédilection avec l'assurance d'une auto-description (Goyet, 2017).

Les deuxième et troisième propositions sont placées au conditionnel, car s'il existe des possibilités de réaliser ce fantasme de chercheur (i.e. Auer et al, 2018), la réalité des pratiques éditoriales scientifiques est encore bien loin d'exploiter pleinement le potentiel réel du Web des données liées. La quatrième possibilité relève plus de la communication que d'une réelle sémantisation de contenus scientifiques, nous y reviendrons dans une analyse sociale de ces pratiques et des potentiels effets à moyen terme.

Pourquoi parler de sémiotique du code ?

Avec le nouveau mode de publication qu'est le *semantic publishing*, la structuration des contenus d'un article scientifique va au-delà de son simple plan. Les fragments, comme unités de pensées structurées, sont agencés pour construire une réflexion : un plan de type IMRAD va, par exemple, présenter une typologie de parties clairement identifiables quant à leur fonction. Nous avons montré qu'il était possible d'anticiper également d'autres usages à des fragments qui ne se superposeraient pas aux parties ou même aux paragraphes en présentant la péricope, qui est un fragment autonome qui peut avoir une autonomie de sens et peut être transcluse (au sens de l'intertextualité) dans un nouveau document sans perte de contexte (Kembellec et Bottini, 2017). Pour descendre encore plus loin en granularité, certains termes ou entités nommées du texte d'un article trouvent possiblement des références définitoires multiples. Par exemple, si sous un angle structuraliste l'on analyse le signifiant « documentarisation », on trouve au sein même des sciences de l'information, en documentation des signifiés différents selon les écoles de pensées. De même, pour les entités nommées, lorsque l'on se réfère à Christian Jacob, il est aisé de comprendre selon le contexte de l'article qu'il s'agit d'un chercheur en histoire des écritures ou d'un homme politique. Cependant, un outil d'analyse en lecture augmentée fera difficilement cette désambiguïisation sans l'implémentation sémantique d'un référentiel d'autorités comme l'ISNI, VIAF, ORCID ou IdRef².

Pour permettre à un outil de comprendre — tout du moins d'inférer si l'on veut éviter tout anthropomorphisme — qu'une portion de texte, une figure ou un tableau constitue une unité de sens traitant d'un sujet en particulier, il faut que l'auteur puisse expliciter son inférence dans un format compréhensible par la machine : un code source. Goyet rappelait à ce sujet que le code informatique est une écriture (Goyet, 2017 ; Herrenschmidt, 2007), ses composantes en sont donc des signes, d'où la qualification de leur étude de « sémiotique ». En effet, les modalités d'inscription et de signalisation des fragments au sein d'un dispositif éditorial scientifique sont d'importance. Lors d'une

[suite de la note]

bibliographiques (Zotero) jusqu'à des greffons plus complexes qui détectent tous types de liaisons de données comme *OpenLink Data Sniffer*.

² *L'International Standard Name Identifier (ISNI)* est l'autorité internationale qui permet d'identifier les identités publiques des personnes ou des organismes impliqués dans la création, la production ou la gestion et la distribution de contenus intellectuels et artistiques. Identifiants et Référentiels pour l'Enseignement supérieur et la Recherche (IdRef) est un service d'autorité scientifique français, ce que *l'Open Research and Contributor ID (ORCID)* est à l'international. Le *Virtual International Authority File (VIAF)* est un service destiné aux bibliothèques qui mutualise et simplifie l'accès aux principaux fichiers d'autorité de noms du monde entier.

lecture attentive l'utilisateur va comprendre, tout du moins interpréter, la pensée de l'auteur, saisir l'agencement de ses réflexions en lien avec ses références. Un outil d'analyse en lecture équipée (ibid) n'aura pas la capacité d'inférence et de rebond rêvée par Bush (1945) dans son *Mémex* : c'est ici qu'interviennent les signes passeurs de connaissance qui sont mis en œuvre par la conversion techno-sémiotique (Doueïhi, 2008) des écrits au moyen de la délégation d'écriture des divers dispositifs scripturaux en ligne (Jeanne-Perrier, 2006). Selon la méthodologie structurale d'inclusion de métadonnées sémantiques, la typologie des acteurs d'autorités de référencement et le choix des vocabulaires, l'inscription des fragments informationnels et de leur documentation ne se fera pas de la même façon. Pour clarifier le propos, plusieurs codes sources d'écrits d'écran hypertextes pourront amener au même résultat visuel à l'affichage-écran. Cependant, une analyse sémiotique de leurs codes sources pourra informer le lecteur attentif sur (1) le type d'inférences que souhaite réaliser l'auteur (2), un potentiel parti pris dans la gouvernance des systèmes d'information, mais aussi (3) ses objectifs communicationnels de diffusion de l'information (Sire, 2018).

Sémiotique du code : typologie analytique des signes passeurs en *semantic publishing*

Nous proposons ici une typologie analytique illustrée de ces divers signes qui exposent la sémantique de leurs contenus en contexte scientifique.

Les signes indicateurs de notices

La première catégorie que nous analyserons est celle centrée sur la signalisation de notices bibliographiques. Les cas d'usages sont liés au repérage et à l'extraction des références.

Les métadonnées d'entête de fichier HTML

Les balises <meta> autorisent l'usage d'un ensemble normalisé de métadonnées, souvent du Dublin Core simple ou qualifié décrivant le document dans lequel elles sont intégrées (Kunze, 1999 ; Greenberg, 2003 ; Turner et Brackbill, 2011). Ce premier niveau de sémantique des contenus ne permet de décrire que le document lui-même de manière monolithique et donc n'expose qu'une seule notice, mais de manière très fine. Cette technologie est utilisée par exemple pour exposer un billet de blog de recherche sur Hypothèses.org ou sur Hal et en proposer une indexation éditorialisée. La mise en œuvre des métadonnées embarquées à l'entête du document HTML est très simple : elles sont placées entre les balises HTML <head> et </head>. Chaque balise <meta> comprenant un premier attribut *name* pour typer la métadonnée et un deuxième attribut *value* pour donner sa valeur.

```

1 <!DOCTYPE html>
2 <html lang="fr-FR">
3   <head>
4     <meta charset="utf-8" />
5     <title>Étude comparée de quatre logiciels de gestion de références
6     bibliographiques libres ou gratuits.</title>
7     <!-- Métadonnées de l'industrie de la Presse : Highwire Press Tags -->
8     <meta name="citation_title"
9           content="Étude comparée de quatre logiciels de gestion de références
10            bibliographiques libres ou gratuits">
11     <meta name="citation_authors" content="Kembellec, Gérard; Scopsi, Claire">
12     <meta name="citation_journal_title" content="Documentation et Bibliothèques">
13     <meta name="citation_publisher" content="ASTED">
14     <meta name="citation_issue" content="58">
15     <meta name="citation_volume" content="4">
16     <meta name="citation_firstpage" content="187">
17     <meta name="citation_lastpage" content="197">
18     <meta name="citation_date" content="2012-12-12">
19     <meta name="citation_language" content="fr">
20     <meta name="citation_issn" content="0315-2340">
21   </head>
22   <body>
23     ...
24   </body>
25 </html>

```

Figure 1. Signaler les métadonnées d'en tête.

Google Scholar encourage l'usage d'autres vocabulaires professionnels de presse comme les *Prism Tags* ou *Highwire Press Tags*³ plutôt que celui du *Dublin Core* plébiscité par le monde la documentation. Il est possible donc, à la lecture du code source du document scientifique, d'émettre une hypothèse sur l'objectif éditorial de la documentation de la page. Si les signes passeurs d'information utilisent un ou des vocabulaires préconisés par la firme *Google* plutôt que ceux du *Dublin Core Metadata Initiative*, les problématiques de visibilité sont d'une importance au moins égale à la simple mise à disposition de notices pour les chercheurs. L'exemple présenté en figure 1 utilise le vocabulaire *Highwire Press Tags*.

Le Context Object in Span (COinS)

Le COinS est l'implémentation en HTML de l'*OpenURL*, norme NISO Z39-88. Il va plutôt avoir pour objectif de mettre en avant la bibliographie du document et non le document en lui-même. Ce type de format de métadonnées, comme leur nom l'indique, s'inscrit uniquement dans des balises « *inline* » (c.-à-d. sans saut de ligne) de type ``. Cette balise n'est pas porteuse de sens, son usage est lié au contexte d'utilisation. Ici, l'encadrement par les balises `` matérialise l'ancrage de métadonnées pour un item de bibliographie ou un appel de citation. L'exemple en figure 2 présente un élément de bibliographie, encadré par des balises `<cite>` (qui ont intrinsèquement du sens) dans une balise ``. Si l'on décompose *les signes*, on remarque que les variables (en gras noir) et sont séparées par le signe *&*; et que tous les caractères spéciaux sont encodés pour les besoins de l'usage du protocole HTTP. Cette deuxième catégorie de signes passeurs d'information est à considérer comme étant potentiellement : (a) une documentation altruiste afin de faciliter le repérage technique des notices par *Zotero* ou (b) plus probablement dans le cas des moteurs de recherche ou des éditeurs scientifiques commerciaux comme un service de création de valeur fidélisant le chercheur-usager-client.

.....

³ *Prism* et *Highwire Press Tags* sont des collections de métadonnées issues de l'industrie de la presse en ligne qui peuvent être détournées pour la documentarisation éditoriale scientifique. Voir la documentation de *Google Scholar* qui explicite comment maximiser les chances d'indexation avec l'inclusion de ces métadonnées et l'analyse de la cause de ces prescriptions par Arlitsch et O'Brien (2012) : <https://scholar.google.com/intl/en/scholar/inclusion.html#indexing>.

```

<span
class='Z3988'
title='url_ver=Z39.88-2004&amp;
ctx_ver=Z39.88-2004&amp;
rft_id=info%3Aasid%2Fzotero.org%3A2&amp;
rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Adissertation&amp;
rft.title=Etude%20sur%20la%20valorisation%20des%20logiciels%20de%20gestion%20de%
20r%C3%A9f%C3%A9rences%20bibliographiques%20(LGRB)%20dans%20le%20milieu%
20des%20biblioth%C3%A8ques%20universitaires%20romandes%2C%20et%20sur%20les%
20avantages%20et%20limites%20du%20plug-in%20Zotero&amp;
rft.inst=Haute%20%C3%A9cole%20de%20gestion%20de%20Gen%C3%A8ve&amp;
rft.aufirst=Robin&amp;
rft.aulast=Masur&amp;
rft.au=Robin%20Masur&amp;
rft.date=2009'>
<cite>Étude sur la valorisation des logiciels de gestion de références bibliographiques
(LGRB) dans le milieu des bibliothèques universitaires romandes, et sur les avantages et
limites du plug-in Zotero</cite>. Genève : Haute école de gestion de Genève (HEG-GE).
</span>

```

Figure 2. COinS implémenté

Signalisation de termes, entités nommées et fragments

Sire a récemment proposé de présenter les différentes techniques de sémantisation des contenus en ligne (2018), rendant possible leur discrétisation et par là une réelle sémantique des fragments infodocumentaires scientifiques. Nous présentons dans cette sous-partie une synthèse sélective de ces modèles applicables en situation de *semantic publishing*. L'intégration de liaisons descriptives entre autorités, vocabulaires et fragments informationnels est une condition nécessaire à la systématisation de la sémantique des contenus. Cette condition nécessite qu'un certain nombre de prérequis soient remplis et inscrits sous forme sémio-rhétorique faisant ainsi coïncider fond et forme (Bouchardon et al., 2011, p. 14).

- 1) Tout d'abord, les contenus d'un document (ou fragments) doivent, pour satisfaire aux règles du FAIR¹, former un ensemble discret : ils doivent être accessibles de manière unique à travers d'URI (Bachimont, 1999, p 200 ; Goyet, 2017, p. 68 ; Sire, 2018, p. 147).
- 2) Ensuite, il convient que chaque élément décrit par l'auteur ou l'éditeur use d'un référentiel composé de vocabulaires, de grammaires, et d'autorités qui soient communes aux différents acteurs - qu'ils soient humains ou machines.

À partir de ces règles fondamentales, il faut analyser les formats qui se basent sur ce modèle dit de « triplets », pour implémenter la sémantique au sein d'un document scientifique, ce que nous proposons ici. Le moyen le plus courant est d'implémenter le modèle abstrait *Resource Description Framework* (RDF) comme méthode de description au sein d'un hypertexte scientifique. Ce modèle est tout à la fois un système de pensée par graphe conceptuel et un langage de description dérivé du XML déclinable en diverses implémentations. Ce système se base sur le concept de triplet évoqué plus haut que l'on va définir comme l'association d'un sujet, avec un objet au moyen d'un prédicat. Le *sujet* représente la ressource à décrire, dans le cas qui nous intéresse le fragment de texte (ou la ressource média dans un sens large : une illustration par exemple). Le *prédicat* représente un type de propriété applicable à cette ressource, cette propriété étant définie sous la forme d'un élément de descripteur accessible en ligne au travers d'une URL. Enfin, l'*objet* représente une donnée ou une autre ressource : c'est la valeur de la propriété. L'objet va avoir pour valeur soit identifiant fourni par une autorité (ISNI, ISBN...) dans ce cas l'on parlera d'URN soit une valeur textuelle ou numérique

.....

¹ Nous expliciterons dans la partie « Les points de vigilance », les bonnes pratiques d'accessibilité du FAIR.

dite « littérale », ou encore une URL d'identification sensée définir le sujet documenté. Pour donner un exemple, le *sujet* du présent fragment textuel a pour *prédicat* de définir l'*objet* de triplet RDF. Toutes les méthodes présentées ci-après se basent sur ce principe de triplet RDF, les éléments de divergence, outre le substrat technique, vont être liés au mode de gouvernance des organismes liés tout autant que les objectifs affichés par lesdits organismes (Sire, 2018).

Le RDFa : le respect des règles institutionnelles

Ce modèle de structure sémantique implémentable au sein du HTML 5 dérive du xHTML (plus précisément du RDF intégré au xHTML). Il est proposé au sein du *World Wide Web Consortium* (W3C), l'organisme en charge de standardisation du Web, fondé en 1994 par son inventeur et qui se charge de réguler les divers formats informatiques qui y cohabitent. Certaines descriptions peu flatteuses de la gouvernance du W3C font paraître cet organisme comme autocratique : la régulation des discussions appartenant *in fine* à Tim Berners Lee (Malcom, 2008). Toutefois cette organisation, comme le rappelle Sire, citant Russel, a le mérite incontestable d'éviter que les formats du Web ne tombent sous la coupe des industriels avec des dépôts de brevets (Russell, 2003).

L'intégration du RDFa au sein du HTML5 n'influe pas sur le rendu visuel du document dans le navigateur, sauf à user de *plug-ins* dédiés à cet usage. Les signes qui permettent de l'implémenter sont facilement identifiables lors d'une analyse sémiotique du code source⁵ : des attributs sont ajoutés aux balises HTML et vont indiquer les schémas, vocabulaires ou espaces de nommages convoqués et ainsi décrire et discrétiser les fragments sous forme de triplets. Dans l'exemple fictif en figure 3, la ligne 9 propose au sein de la balise <section> l'usage du schéma « *ScholarlyArticle* » pour décrire la citation. Il est précisé entre les lignes 11 et 15 que l'auteur est une « *Person* », pas une institution. De plus, cet auteur est spécifié comme étant celui présenté dans le référentiel d'autorité VIAF sous l'identifiant 40280446, ce qui évite toute ambiguïté d'homonymie. L'humain verra donc l'entité nommée « Lawrence Lessig » inscrite sur son écrit d'écran et aura sa représentation mentale de la personne référée. La machine reproduira une démarche analytique « similaire » grâce au triplet sémantique pointant vers le serveur VIAF : le triangle sémiotique est donc respecté⁶ dans la dimension sémantique de Morris pour un usage informatif (Morris, 1946). On remarque aussi un bornage temporel différent de la simple année dans les lignes 17-19 au sein de la balise <date> : son attribut *property* est renseigné avec *temporalCoverage*, une durée qui va être évaluée à 1999-2001. Cela correspond bien à un équivalent du texte à destination des humains : « le tournant du siècle ».

.....

⁵ Documentation officielle, accédée le 27 février 2019 : <http://www.w3.org/TR/rdfa-in-html/>

⁶ Nous ne discuterons ici le fait que l'ambiguïté puisse persister pour l'humain qui ne partagerait pas le même référentiel. Dans le cadre d'une lecture scientifique, nous portons comme présupposé que la désambiguïsation par les humains va de soi, comme dans l'exemple « Christian Jacob » présenté précédemment.

```

1 <!doctype html>
2 <html version="HTML+RDFa 1.1" lang="fr">
3 <head>
4   <meta charset="utf-8">
5   <title>Article au format RDFa</title>
6 </head>
7 <body>
8   Rappelons que l'article
9   <section vocab="http://schema.org/" typeof="ScholarlyArticle" resource="#citation">
10  <cite property="headline">Code Is Law : On Liberty in Cyberspace</cite>
11  par <span property="author" typeof="Person">
12    <span property="name">
13      <a property="sameAs" href="http://viaf.org/viaf/40280446">Lawrence Lessig</a>
14    </span>
15  </span>
16  a été écrit dès
17  <date property="temporalCoverage" content="1999/2001">
18    le tournant du siècle
19  </date>
20  
22 </section>
23 </body>
24 </html>

```

Figure 3. RDFa implémenté

Les micro-données et la sémantique orientée interaction et référencement : Microdata et Schema.org

Ces métadonnées sont nées d'un désaccord de gouvernance au sein du W3C (Sire, 2018) : d'un côté Tim Berners Lee souhaitait privilégier l'usage du xHTML comme vecteur de sémantisation du Web, de l'autre l'initiative de créer un langage qui l'intégrerait nativement de manière simplifiée : le HTML5. Cela a mené à la création du *Web Hypertext Application Technology Working Group* (WHATWG) soutenu par *Google*. Les règles des microdonnées ou *microdata* sont nées des mêmes acteurs, qui sont présents dans les deux groupes, mais avec des tensions certaines. Ceci explique une multiplicité des formats d'inclusion de métadonnées et données sémantiques dans le HTML5. Dans un même temps, les principaux moteurs de recherche mondiaux se sont rapprochés pour créer des schémas de description du réel, des référentiels dont l'usage est encouragé dès 2011 par *Google* comme ayant un impact fort sur l'indexation. Il est cependant possible d'utiliser conjointement aux *microdatas* d'autres référentiels de description comme le *Dublin Core* qualifié.

La sémantique des micro données est portée, outre par les balises HTML5 elles-mêmes, par les attributs suivants :

- **itemscope** : qui signifie le champ sémantique de l'élément ;
- **itemtype** : qui précise le type d'élément concerné ;
- **itemprop** : définit la propriété, de l'objet défini ;
- **itemid** : référence un identifiant unique, comme un URN fourni par un organisme d'autorité de type ORCID, ISNI, ISSN ou encore ISBN.

Dans la figure 4, nous avons composé le contenu d'un paragraphe qui donne une opinion positive sur un article et ce qu'en voit le robot de recherche sémantique de *Google*. Ici donc, la représentation visuelle du terme « Nous » augmentée de son code source inscrit dans la balise `` et sera traduite par les machines comme : la personne dont l'identifiant ORCID est 0000-0003-3036-6989. Ce dernier cite l'article « *Code is Law* » et abonde dans le sens de la référence. Nous notons ici aussi un usage informatif des triplets, ce qui semble aller dans le même sens que le RDFa.

```

1 <p itemscope itemtype="http://schema.org/AgreeAction">
2   <span itemprop="agent"
3     itemscope itemtype="http://schema.org/Person"
4     itemid="https://orcid.org/0000-0003-3036-6989">
5     Nous
6   </span> abondans dans le sens de l'article
7   <cite itemscope
8     itemtype="http://schema.org/ScholarlyArticle"
9     itemprop="object">
10    
12    <span itemscope itemprop="author"
13      itemtype="http://schema.org/Person">Lawrence Lessig
14    </span>
15    <span itemprop="headline">
16    Code Is Law : On Liberty in Cyberspace</span>
17    <em itemprop="isPartOf"
18      itemscope itemtype="http://schema.org/Periodical">
19      <span itemprop="name">The Industry Standard</span>
20    </em>
21    <span itemprop="isPartOf"
22      itemscope
23      itemtype="http://schema.org/PublicationIssue">
24      no. <span itemprop="issueNumber">18</span>
25    </span>
26  </cite>
27 </p>

```

AgreeAction		0 ERREUR	0 AVERTISSEMENT
@type	AgreeAction		
agent			
@type	Person		
@id	https://orcid.org/0000-0003-3036-6989		
object			
@type	ScholarlyArticle		
image	https://theindustrystandard.com/logo.png		
headline	Code Is Law : On Liberty in Cyberspace		
author			
@type	Person		
isPartOf			
@type	Periodical		
name	The Industry Standard		
isPartOf			
@type	PublicationIssue		
issueNumber	18		

Figure 4. Microdonnées implémentées et ce qu'en comprend Google

La sémantique dans les pages HTML5 et les Notebooks grâce format JSON-LD

Une dernière méthode que nous analyserons est le format JSON-LD qui s'intègre à la page HTML, soit dans l'entête, soit dans un fichier à part qui sera chargé par le navigateur (avec un *plug-in*) ou le moteur d'indexation. Ce format de code source a pour avantage certain une forte compacité qui n'alourdit pas le code source HTML tout en autorisant une description fine de la page tout en étant décorrélée du code de celle-ci. Dans ce cas précis, le sens du code est le même que pour les microdonnées et l'analyse par le moteur est la même. Notons que le format JSON-LD est utilisé dans le cadre de la documentation les différents fragments d'un *notebook* en science reproductible de type « *JupyterHub* »⁷.

```

1 <script type="application/ld+json">
2 {
3   "@context": "http://schema.org",
4   "@type": "AgreeAction",
5   "agent": {
6     "@type": "Person",
7     "name": "Gérald",
8     "familyName": "Kembellec",
9     "sameAs": "https://orcid.org/0000-0003-3036-6989"
10  },
11  "object": {
12    "@type": "ScholarlyArticle",
13    "headline": "Code Is Law : On Liberty in Cyberspace",
14    "author": "Lawrence Lessig.",
15    "image": "https://theindustrystandard.com/logo.png"
16  }
17 }
18 </script>

```

AgreeAction		0 ERREUR	0 AVERTISSEMENT
@type	AgreeAction		
agent			
@type	Person		
name	Gérald		
familyName	Kembellec		
sameAs	https://orcid.org/0000-0003-3036-6989		
object			
@type	ScholarlyArticle		
headline	Code Is Law : On Liberty in Cyberspace		
image	https://theindustrystandard.com/logo.png		
author			
@type	Thing		
name	Lawrence Lessig.		

Figure 5. JSON-LD et son analyse par Google

Les signes communicationnels à destination des médias sociaux

Les documents en ligne peuvent également intégrer un type de métadonnées reconnaissables par les principaux médias sociaux qui fonctionnent tous sur le même modèle : une description du document comprenant les informations générales telles que le titre, l'URL, les auteurs, une illustration et un résumé (Goyet, 2017). Ces métadonnées s'implémentent sous deux formes : L'*Open Graph* et les *Summary Cards* de la famille des *Twitter Cards*. Le protocole *Open Graph* (<http://ogp.me/>) est à

.....

⁷ Un *notebook*, littéralement un carnet (de recherche), est un document scientifique qui permet de réaliser un article incluant texte (en *markdown* ou en HTML), données et exploitation par des outils de programmation et de présentation graphique. L'avantage de cette *Linked Science* est de proposer ses données et sa démonstration de manière reproductible, fragment par fragment, voir <https://jupyter.org/hub>

destination de *Facebook* et de *LinkedIn*. Basé sur le RDFa et créé par Facebook, *Open Graph* permet à toute page Web de devenir un « objet riche dans un graphe social » (Haugen 2010). Haugen décrit ce protocole comme conçu pour être « utile à tout le monde », ce qui explique son usage par *LinkedIn* par exemple. Il est à noter qu'il existe une fonctionnalité du gestionnaire de contenus *WordPress* pour en automatiser l'intégration dans les billets, ce qui peut être *in fine* applicable à des documents scientifiques ou des carnets de recherche par exemple. Les *Twitter Cards* de type *summary card* sont, dans les mêmes conditions, porteuses d'un discours d'accompagnement du document hypertexte en ligne, mais à destination exclusive de Twitter.

```
<!-- Les métas pour Twitter -->
<meta name="twitter:image" content="https://dfzljdn9uc3pi.cloudfront.net/2017/cs-132/1/fig-1-full.png">
<meta name="twitter:card" content="summary_large_image">
<meta name="twitter:url" content="https://peerj.com/articles/cs-132">
<meta name="twitter:site" content="@thePeerJ">
<meta name="twitter:title" content="Research Articles in Simplified HTML: a Web-first format for HTML-based scholarly articles">
<meta name="twitter:description" content="Purpose This paper introduces the Research Articles in Simplified HTML (or RASH), which is a Web-first format for writing HTML-based scholarly papers">

<!-- Les métas pour Facebook -->
<meta property="og:type" content="article">
<meta property="og:image" content="https://dfzljdn9uc3pi.cloudfront.net/2017/cs-132/1/fig-1-full.png">
<meta property="og:type" content="article">
<meta property="og:title" content="Research Articles in Simplified HTML: a Web-first format for HTML-based scholarly articles">
<meta property="og:url" content="https://peerj.com/articles/cs-132">
<meta property="og:site_name" content="PeerJ Computer Science">
<meta property="og:description" content="Purpose This paper introduces the Research Articles in Simplified HTML (or RASH), which is a Web-first format for writing HTML-based scholarly papers">
```

Figure 6. Twitter Card et Open Graph en situation d'usage, adapté du code d'un article de la revue *PeerJ*

Dans les deux cas cités, grâce aux signes inclus dans l'entête de la page support, si l'URL du carnet de recherche ou de l'article en ligne est publiée sur un média social, une illustration et un texte seront les supports visuels de l'hyperlien. L'exemple ci-dessus, adapté de l'entête du code source HTML d'un article de la revue en ligne *PeerJ*, présente un aperçu des signes qui permettent de faire circuler les principales informations de l'article sur les réseaux sociaux : un visuel, le titre et un résumé court en plus de l'hyperlien. Il ne s'agit pas là d'une exposition de notice, mais bien d'un « *teaser* » invitant à un engagement pour reprendre un terme issu du *marketing* numérique.

Discussion sur l'usage du *semantic publishing*

Des questions émergent naturellement sur l'impact que l'intégration de ces nouvelles méthodes d'inscriptions numériques en ligne a sur la chaîne des acteurs de l'IST : auteurs, éditeurs, lecteurs. Quels en sont les réels enjeux implicites pour le monde, au sens large, de la recherche ?

Formats, référentiel et gouvernance, une impossible neutralité

Goyet (2017), interprétant Lessig dans son célèbre article « *Code is Law* » (1999) avait avancé que les choix concernant l'écriture du code ne sont pas neutres : ce sont des choix techniques bien sûr, mais aussi de convictions. Nous avons présenté que des scissions ont pu intervenir dans les organismes de standardisation du Web pour des questions de gouvernance, de conflits personnels et d'éthique. Outre les potentialités techniques, qui ne sont pas neutres, ce choix est porteur de valeurs morales et culturelles. Donc, sans prendre part à ces débats, l'usage d'un type de modalité d'ancrage des données (xHTML vs HTML5) et métadonnées (RDFa vs Microdonnées ou JSON-LD) positionne l'acteur, possiblement à son insu. De plus, le choix des modèles de données se rajoute comme facteur positionnant l'acteur : une utilisation de *schema.org* pourrait être perçue comme féale au conglomérat des moteurs de recherche, alors que celle du Dublin Core pourrait être plébiscitée moralement, mais discutée techniquement. Enfin, le choix des autorités d'identification n'est absolument pas neutre non plus : si les autorités décrites en note 2 communiquent, mais l'adhésion à l'une ou l'autre peut parfois se faire sur des facteurs aussi triviaux que le chauvinisme : en effet certaines autorités sont nationales, d'autres étrangères ou encore internationales (Boulet, 2016). Ces analyses de solutions de formats, de référentiels et d'autorités sortent de la sphère technique et toute proportion gardée, sont assez comparables aux discussions passionnées d'adhésion entre usagers de systèmes libres et propriétaires. Il existe une multitude de paramètres à prendre en compte dont

l'usabilité, les politiques institutionnelles et nationales, tout autant que les réalités du marché de l'édition et des usages de chercheurs. Lors de telles discussions, il faut s'attendre à une analyse critique des dispositifs d'écriture et de mise en ligne de documents scientifiques sur des aspects qui sont bien éloignés de la recherche.

Apports pour les chercheurs : alignement sur les primitives savantes de la recherche

Pour la première fois de l'histoire des savoirs scientifiques, la sémantique des contenus est privilégiée à la simple évolution d'une grammaire visuelle issue de la mémoire « de formes » des supports d'inscription précédents (Jeanneret, 2006).

Nous avons précédemment présenté les fonctions les plus courantes de l'action et de la coopération en termes de recherche, d'analyse, de synthèse et de partage de l'information scientifique (Kembellec et Bottini 2017) au travers des primitives savantes du métier de chercheur par une synthèse de la littérature (Unsworth, 2000 ; Palmer et al. 2009 ; Hennicke et al. ; Nakakoji et al. 2005). Nous proposons ici, une mise en relation des principales primitives avec les signes passeurs du *semantic publishing* évoqués précédemment sous la forme d'un tableau synoptique en figure 7. Nous y avons grisé les parties offrant un réel bénéfice informationnel en contexte scientifique. Ce peut être une désambiguïsation conceptuelle ou des éléments de structuration documentaires signalant la nature ou la fonction d'un fragment sémantique hyperliable.

Prospective sur le semantic publishing

Les bénéfices du *semantic publishing* pour la compréhension et la circulation des savoirs scientifiques sont évidents, le temps héroïque du Web 1.0 est bien révolu et les écrits d'écran ne sont plus le seul vecteur de communication numérique en réseau.

Au-delà même des réflexions d'usages, d'enjeux et de techniques, cette tendance de renouveau en médiation de l'information scientifique offre une proximité conceptuelle partielle avec un Mémex scientifique. Avec les facilités computationnelles et interprétatives offertes par le *semantic publishing* et l'assistance d'outils de lecture adaptés, il semble évident que de nouvelles possibilités de conversion de données en connaissances s'ouvrent (Gartner, 2016, p.10). L'ambiguïté entre les termes et également entre les entités nommées disparaît, les fragments informationnels sont décrits et reliés : on assiste à la naissance d'information enrichie et découvrable avec davantage de visibilité et de citabilité.

De nouvelles métriques en citation ?

Avec la discrétisation des appels de citation (ou des bibliographies d'articles en ligne), un affinage des contextes de citation ouvre la perspective d'une possible (re)théorisation de la citation en prenant en compte des métriques qualitatives autres que celles initiées par Garfield (1955, 1972) puis par Hirsch (2005) : l'emplacement de la citation (Bertin et Atanassova, 2018), l'intention de l'auteur (Xiao et al. 2018) qui peut être perceptible par un modèle de rhétorique du code similaire à celui de l'hypertexte tel que théorisé par Saemmer (2012, 2015).

Types d'inclusions	Objectifs	Primitives (SP) et activités savantes (SA)	Types de signes Passeurs
Exposition de notices	Indexer, collecter et archiver des notices	<u>Palmer et al.</u> <i>Searching: browsing, probing, accessing</i> <i>Collecting: Gathering, organizing</i> <u>Hennicke et al.</u> <i>SP: exploration, discovering</i> <i>SA: Searching, discovering, foraging, monitoring</i> <u>Nakakoji et al.</u> Niveaux I et V : Collecte et archivage pour usage ultérieur	COinS + POSH <i>Metadata in header</i> et/ou inclusion d'un DOI
Désambiguïisation avec des autorités	Désambigüiser des termes et entités nommées. Offrir des possibilités de lien avec d'autres fragments de documents	<u>Palmer et al.</u> <i>Reading: scanning, assessing</i> <i>Writing: assembling, disseminating</i> <u>Hennicke et al.</u> <i>SP: exploration, discovering</i> <i>+ Interpretative modeling</i> <i>SA: discovering, contextualizing, conceptualizing, translating, referring/linking</i> <u>Nakakoji et al.</u> Niveaux I, II et IV : Collecte(discrétisation)des fragments Compréhension, subjectivation, tissage de liens	Microdata ou RDFa et/ou JSON-LD + schemas ou Dublin-Core
Typage et rhétorique de fragments	Offrir une meilleure visibilité sur le type des fragments et leur valeur au sein du document et du réseau scientifique	<u>Palmer et al.</u> <i>Reading: scanning, assessing</i> <i>Writing: assembling, disseminating</i> <u>Hennicke et al.</u> <i>SP: exploration, discovering</i> <i>+ Interpretative modeling</i> <i>SA: discovering, contextualizing, conceptualizing, translating, referring/linking</i> <u>Nakakoji et al.</u> Niveaux I, II, IV et V : Collecte(discrétisation)des fragments Compréhension, subjectivation, tissage de liens et articulation des fragments dans le document ET dans le réseau scientifique	Usage de classes HTML5 + CSS + Microdata ou RDFa et/ou JSON-LD + schema.org ou Dublin-Core
Typage de documents et description pour les médias sociaux (<i>altmetrics</i>)	Communiquer et gagner en visibilité en offrant un résumé du document scientifique dans le même temps que son URI sur les médias sociaux	<u>Palmer et al.</u> <i>Reading: assessing, scanning</i> <i>Writing: disseminating</i> <i>Collaborating: networking, consulting</i> <u>Hennicke et al.</u> <i>SP: exploration, discovering</i> <i>+ externalization</i> <i>SA: discovering, foraging, sharing</i> <i>synthetizing, sampling, publishing, referring/linking, illustrating, disseminating</i> <u>Nakakoji et al.</u> I et V: Collecte et archivage pour usage ultérieur	Twitter Cards et/ou Open Graph et/ou DOI

Figure 7. Corrélation entre signes passeurs et synthèse des besoins et activités des chercheurs

(Kembellec et Bottini, 2017)

De nouveaux architextes scientifiques ?

Ce scénario prospectif de la publication scientifique en ligne ne peut s'envisager qu'avec des outils d'édition, d'inscription, d'interaction et les équipements de circulation (Goyet, 2007). À ce jour, malgré quelques preuves de concepts (Capadisli et al. 2015, Kembellec et Bottini, 2017), il reste encore à mettre en œuvre et à déployer un réel architexte scientifique pensé pour la sémantique des fragments *by design* pour aller dans le sens du *What You See Is What You Mean* cher à Crozat. De plus, il ne peut exister de documents scientifiques en ligne dont les fragments sont reliés entre eux que si lesdits documents sont accessibles : l'*Open Access* en édition scientifique est donc un prérequis au *semantic publishing*, point encouragé, voire prescrit par les politiques nationales et européennes.

Un travail considérable de documentarisation doit être pensé pour réaliser ce type d'enrichissement et il ne peut être pensé que par l'auteur ou en relation avec l'auteur pour ne pas dénaturer sa pensée. Il faut se poser la question de l'engagement des acteurs de l'IST : les auteurs seraient-ils prêts à réaliser eux-mêmes ou participer à la qualification sémantique s'ils en saisissent les enjeux et les avantages ? De même, les éditeurs privés ou institutionnels veulent-ils se saisir et financer ce genre de problématiques ? Nous gageons que des projets de recherche en *semantic publishing* comme NumeRev (<https://numerev.com/>) ouvriront une voie dans cette direction.

Les points de vigilance

La diffusion et la communication des articles scientifiques sont devenues des enjeux en dehors même des cercles scientifiques. Le célèbre *publish or perish* toujours d'actualité côtoie d'autres réalités dues à l'explosion de l'offre en littérature scientifique, à la nécessité de vulgariser et d'exister dans la sphère publique : on parle maintenant aussi de *Get Visible Or Vanish*, de *Deploy Or Die* et même de *Publicize Or Perish*. La diffusion se fait aussi par les moteurs de recherche grand public, les blogs de chercheurs et les réseaux sociaux. Pour ces supports de diffusion, qu'ils soient grand public ou dédiés à la science, l'enjeu communicationnel devient tellement important que de nouvelles métriques comme les *altmetrics* s'invitent dans l'espace du monde social de la recherche. Les méthodes d'indexation naturelles actuelles privilégient la qualité du formalisme et la sémantique de description des contenus au même titre que les contenus eux-mêmes (Kembellec, 2016). Dès lors, la sémiotique du code permet de questionner la limite parfois ténue que l'on peut établir entre formalisation (scénarisation) et publicisation (mercatique) des articles scientifiques sur les plateformes d'édition et sur les blogs de chercheurs. Les règles du FAIR préconisent que la donnée scientifique soit « trouvable, accessible, interopérable et réutilisable » respectant ainsi les enjeux de *l'Open Data Scientifique* (Moore, 2014 ; Wilkinson et al., 2016, Schöpfel, 2018). L'effet corolaire de cette structuration est, outre d'obtenir un *pagerank* élevé pour les moteurs de recherche, d'ouvrir des perspectives communicationnelles extérieures au document au travers du réseau des médias sociaux (Infante-Moro et al., 2015). Sans procès d'intention des chercheurs et des éditeurs de plateformes de diffusion en IST, il convient de mettre en garde la communauté contre une possible confusion entre « industries médiatiques » et « médiatisantes ». Cette tentation peut être inhérente à pression d'évaluation des chercheurs, mais aussi à leurs égos (Anderson, 2014, Branford et Mallucci, 2015). Cela procède de ce que Jeanne-Perrier (2015) identifie comme une redondance informationnelle à vocation de « *lobbying* discursif et sémiotique ». Il convient, *a minima*, de prévenir des tentations de détournement d'usage de ces signes passeurs : entre médiation et médiatisation, que fond et la forme gardent chacun la place qui leur revient.

Enfin, il faut questionner la délégation d'écriture scientifique à de futurs architectes de *semantic publishing* qui pourrait brider l'innovation intellectuelle et les ruptures scientifiques par l'usage de référentiels trop rigides, laissant peu de place à la pensée singulière qui ne saurait être référencée faute de préexister. En termes de lecture « augmentée », il est à envisager également que l'esprit critique soit possiblement bridé pour les mêmes raisons. Ce point est crucial, car il faut également discuter sérieusement le *semantic publishing* sous l'axe de la liberté d'interpréter et de l'esprit critique, mais aussi de l'éthique. En effet, que penser des différents référentiels utilisés par les descriptions conceptuelles de primitives sémantiques et les identifications systématiques ? Qui les édite ? Avec quelle liberté ? Leur sélection a-t-elle (vraiment) une importance ? Autant de questions dont la communauté de l'IST devra se saisir pour revendiquer une autonomie de pensée en contexte de publication sémantique en réseau.

Conclusion

Cet article a établi l'intérêt de l'analyse sémiotique des codes sources d'écrits d'écrans comme porteurs de sens pour les auteurs, éditeurs ou lecteurs. En effet, les sélections des types de signes

passseurs, la méthode de fragmentation et les référentiels d'autorité sont autant de choix loin d'être neutres. Dans la focale du *semantic publishing*, leurs usages répondent à des besoins scientifiques précis en termes informationnels et communicationnels, tant pour l'auteur, l'éditeur que pour le lecteur. Dans un même temps, ils positionnent parfois l'auteur ou l'institution éditrice dans un contexte complexe dont les enjeux en matière de gouvernance peuvent être perçus comme « engageants ».

Nous concluons en reprenant au compte du *semantic publishing* deux idées à mettre en tension. D'un côté Saemmer, reprenant Mesquita, questionnait en situation de Web-journalisme l'évolution d'un usage auctorial de l'hypertexte. Elle convoquait des courants de pensées critiques : l'émergence d'une « vérité dialogique » entre la forme et le fond par des anticipations de pratiques de lecture (Saemmer, 2015 ; Mesquita, 2005). De l'autre côté, nous avons évoqué la crainte de voir se renforcer des usages du *semantic publishing* à visée uniquement diffusionnelle : cela irait dans le sens d'une « circulation circulatoire de l'information », sur le modèle des médias traditionnels fustigés par Bourdieu, induisant une uniformisation et un appauvrissement de la pensée (Bourdieu, 2002 ; Rebillard, 2006). La vérité du *semantic publishing* se trouvera certainement à mi-chemin, entre structuration documentaire forte, avec des données décrites et liées, mais aussi une tentative d'influer sur le calcul des métriques de référencement de moteurs de recherche et des médias sociaux, scientifiques ou non. Cette supposition augure d'un inévitable débat au sein de la communauté des SIC entre les enjeux communicationnels et informationnels relatifs à la médiation et la circulation des productions du *semantic publishing* en contexte scientifique.

Références bibliographiques

Anderson, K., (2014). Exaggerated claims - Has 'publish or perish' become 'publicize or perish'? Available at: <http://scholarlykitchen.sspnet.org/2014/12/15/exaggerated-claims-has-publish-or-perish-become-publicize-or-perish/>.

Arlitsch, K., O'Brien, P. S. (2012). Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar, In *Library Hi Tech*, 30(1), pp. 60 - 81 <http://dx.doi.org/10.1108/07378831211213210>

Auer, S., Kovtun, V., Prinz, M., Kasprzik, A., Stocker, M., & Vidal, M. E. (2018). Towards a Knowledge Graph for Science. In *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics* (p. 1:1 - 1:6). New York, NY, USA: ACM. <https://doi.org/10.1145/3227609.3227689>

Bachimont, B. (1999). Du texte à l'hypotexte: Les parcours de la mémoire documentaire. In *Raisons, Idéologies, Pratiques*, numéro spécial sur la mémoire collective, 13(2), pp. 196-219.

Bachimont, B. (2010). « Le sens de la technique : le numérique et le calcul », Paris, Les Belles Lettres.

Barthes, Roland (1973). (Théorie du) texte, (2002), Œuvres complètes IV. Paris: Seuil

Bertin, M., Atanassova, I. (2018). Recommending Scientific Papers: The Role of Citation Contexts. In *Proceedings of the 1st International Conference on Digital Tools & Uses Congress* (p. 6:1 - 6:4). New York, NY, USA: ACM. <https://doi.org/10.1145/3240117.3240123>

Bigot, J.-É. (2018). *Instruments, pratiques et enjeux d'une recherche numériquement équipée en sciences humaines et sociales*. Université de Technologie de Compiègne, Compiègne.

Bouchardon, S., Cailleau, I., Crozat, S., Bachimont, B., & Thibaud, H. (2011). Explorer les possibles de l'écriture multimédia. In *Les Enjeux de l'Information et de la Communication* (2), 11-24.

Boulet, V. (2016). De la SDN à la Nuit debout : Les métadonnées et les enjeux de gouvernance internationale. *I2D Information, données et documents*, 53(2), pp. 35- 36.

Bourdieu, P. (2002) « Les conditions sociales de la circulation internationale des idées », In. Actes de la recherche en sciences sociales, 145, p. 3-8.

Branford, O. A., Mallucci, P., (2015), Publicize or Perish! A Guide to Social Media Promotion of Scientific Articles: Featuring the Plastic and Reconstructive Surgery 'Author Tool Kit', *In Plastic and Reconstructive Surgery*, 136(4), pp. 579-581

Broudoux, E., & Kembellec, G. (2017). Introduction à l'écritecture scientifique et aux modalités techniques de son augmentation. In *Écritecture augmentée dans les communautés scientifiques. Humanités numériques et construction des savoirs*. (p. 19- 37). London: ISTE.

Bush, V. (1945). As we may think, In *The Atlantic Monthly*, 176(1).

Capadislı, S., Auer, S., & Riedl, R. (2015). This 'Paper' is a Demo. In F. Gandon, C. Guéret, S. Villata, J. Breslin, C. Faron-Zucker, & A. Zimmermann (Éd.), *The Semantic Web: ESWC 2015 Satellite Events* (p. 26- 30). Portorož, Slovenia: Springer International Publishing.

Choi, Y.-H., 2001, Sémiotique et sémantique, in *Linx*, n°44, pp. 75-84.

Crozat, S., Bachimont, B., Cailleau, I., Bouchardon, S., & Gaillard, L. (2011). Éléments pour une théorie opérationnelle de l'écriture numérique. *Document numérique*, 14(3), pp. 9-33.

Crozat S. (2012). « Chaînes éditoriales et rééditorialisation de contenus numériques ». Lisette Calderan, Pascale Laurent, Hélène Lowinger et Jacques Millet : *Le document numérique à l'heure du web*, ADBS, pp.179-220, Sciences et techniques de l'information ; Le document numérique à l'heure du web de données.

Crozat, S. (2016), Écrire avec une machine à calculer, écrire pour une machine à calculer », *I2D-Information, données & documents*, 2016, 53(2), pp. 62-64.

Doueïhi M. (2008), *La grande conversion numérique*, Paris, Éditions du Seuil.

Garfield, E. (1955). Citation indexes for science. *Science*, 122(3159), 108-111.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178(4060), 471-479.

Gartner, R. (2016). *Metadata : Shaping Knowledge from Antiquity to the Semantic Web*. Springer.

Goody, J. (1979). La raison graphique : la domestication de la pensée sauvage. (J. Bazin & A. Bensa, Trad.). les Éditions de Minuit.

Goyet S. (2017), Outils d'écriture du web et industrie du texte. Du code informatique comme pratique lettrée, In *Réseaux* 6 (206), pp. 61-94.

Greenberg, J. (2003). Metadata and the world wide web. *Encyclopedia of library and information science*, 3, 1876-1888.

Haugen, A. (2010). The Open Graph Protocol Design Decisions. In P. F. Patel-Schneider et al. (Éd.), *The Semantic Web - ISWC 2010* (p. 338). Springer Berlin Heidelberg.

Hennicke, S., Gradmann, S., Dill, K., Tschumpel, G., Thoden, K., Morbidoni, C., Pichler, A. (2015). D3.4 - *Research Report on DH Scholarly Primitives*. D2ME Project.

Herrenschmidt, C. (2007). Les trois écritures, langue, nombre, code. Paris : Gallimard.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences*, 102(46), 16569-16572.

Infante-Moro, A., Zavate, A., & Infante-Moro, J.-C. (2015). The influence/impact of Semantic Web technologies on Social Media. *International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC)*, 2(2), pp. 18- 30.

- Jeanne-Perrier V. (2006), « Des outils d'écriture aux pouvoirs exorbitants ? », *Réseaux*, n° 137, pp. 97-131.
- Jeanne-Perrier, V. (2015). Des discours et des signes : pour une évaluation critique de Twitter et de Facebook. *Communication langages*, 183(1), pp. 111- 131.
- Jeanneret Y., Souchier E. (1999), « Pour une poétique de l'écrit d'écran », *Xoana*, n°6, pp. 97-107.
- Jeanneret Y. (2006), 'Ceci n'est pas une page, ceci n'est pas un site', in *Médiamorphoses*, n° 16, pp. 88-92.
- Jeanneret Y. (2014), « Critique de la trivialité : les médiations de la communication, enjeu de pouvoir », Paris, Éditions Non Standard.
- Kembellec, G. (2016). « Que voit réellement Google de la sémantique des pages web ? ». in *I2D - Information, données & documents*, 53(2), p. 65-65.
- Kembellec, G., & Bottini, T. (2017). Réflexions sur le fragment dans les pratiques scientifiques en ligne : entre matérialité documentaire et péricope. In *20° Colloque International sur le Document Numérique : CiDE. 20*.
- Kunze, J. (1999). Encoding dublin core metadata in HTML, In. *Network Working Group* (rfc2731), consulté à l'adresse <http://buildbot.tools.ietf.org/html/rfc2731>, 23 p.
- Lee B., Lipuma E. (2002), « Cultures of Circulation: The Imaginations of Modernity », in *Public Culture*, 14(1), pp. 191-213.
- Lessig, L. (1999). The limits in open code: regulatory standards and the future of the net. In. *Berkeley Tech. Law Journal*, 14(2), 759- 769.
- Malcolm, J. (2008). *Multi-stakeholder governance and the Internet Governance Forum*. Terminus Press.
- Merton, R. K. (1968). 'The Matthew effect in science: The reward and communication systems of science are considered'. In *Science*, 159(3810), pp. 56-63.
- Mesquita, M. (2005). « Théories et pratiques du journalisme du télégraphe à l'hypertexte ». In. *Recherches en communication* 23, p. 203-232.
- Morris, C. (1946). *Signs, language and behavior*. Oxford, England: Prentice-Hall.
- Nakakoji, K., Yamamoto, Y., Akaishi, M., Hori, K. (2015). Interaction design for scholarly writing: hypertext representations as a means for creative knowledge work. In *The New Review of Hypermedia and Multimedia*, Special issue: Scholarly hypermedia, 11(1), Taylor & Francis.
- Palmer, C. L., Tefteau, L. C., & Pirmann, C. M. (2009). Scholarly information practices in the online environment: Themes from the Literature and Implications for Library Service Development (research No. 298733494) (p. 59). Dublin, Ohio: OCLC Research. Consulté à l'adresse www.oclc.org/programs/publications/reports/2009-02.pdf
- Prime-Claverie, C. (2016). Linked science et Web de données. *I2D Information, données documents*, Volume 53(2), pp. 42- 43.
- Rebillard, F. (2006). « Du traitement de l'information à son retraitement ». In *Réseaux* 137, p. 29-68.
- Romm, J. (2009). « Publicize or Perish ». In *Physics World* 22(10), p. 22-23
- Russell, A. L. (2003). *The W3C and Its Patent Policy Controversy: A Case Study of Authority and Legitimacy in Internet Governance* (SSRN Scholarly Paper n°. ID 2056900). Rochester, NY: Social Science Research Network
- Saemmer, A. (2012). Étude sémio-rhétorique du rôle de l'hypertexte dans le discours journalistique. in *Médiation et information*, (34).

- Saemmer, A. (2015). *Rhétorique du texte numérique. Figures de la lecture, anticipations de pratiques*. Villeurbanne: ENSSIB.
- Shotton, D. (2009), 'Semantic publishing: the coming revolution', in *scientific journal publishing. Learned Publishing*, 2009, 22(2), pp. 85-94.
- Simon, J. (2015). Le discours hypertextualisé: Une notion essentielle pour l'analyse du Web. In *Le numérique à l'ère de l'Internet des objets, de l'hypertexte à l'hyper-objet*, pp. 3-20. ISTE Editions.
- Sire, G. (2018). Web sémantique : les politiques du sens et la rhétorique des données. *Les Enjeux de l'information et de la communication*, 19/2(2), pp. 147-160.
- Souchier E. (1996), « L'écrit d'écran, pratiques d'écriture & informatique », *Communication & langages*, n° 107, pp. 105-19.
- Souchier E. (1998). « L'image du texte pour une théorie de l'énonciation éditoriale », *Les Cahiers de médiologie*, n°2, p.137-145.
- Souchier E., Jeanneret Y. (2002), 'Écriture numérique ou médias informatisés ?' *Pour la Science*, n° 33, pp. 100-105.
- Turner, T. P., & Brackbill, L. (2011). Rising to the top: evaluating the use of the HTML meta tag to improve retrieval of World Wide Web documents through Internet search engines. *Library Resources & Technical Services*, 42(4), pp. 258-271.
- Unsworth, J. (2000) 'Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this?' in *symposium on 'Humanities Computing: formal methods, experimental practice'*, King's College, London, May 13, 2000.
- Verlaet L. et Dillaerts H. (2016), « L'enjeu du Web de données pour l'édition scientifique » in. *I2D-Information, données & documents*, 53(2), p. 49.
- Xiao, M., Shi, Z., & Wang, S. (2018). The Impact on Citation Analysis Based on Ontology and Linked Data. *Scientometrics*.
- Zacklad, M. (2019). Le design de l'information : textualisation, documentarisation, auctorialisation. In. *Communication & langages*, 199(1), 37-64. doi:10.3917/comla1.199.0037.