HAL open science

Eléments de description du paysage médiatique français

Agnès Saulnier

To cite this version:

Agnès Saulnier. Eléments de description du paysage médiatique français. 2021. hal-03343071v1

HAL Id: hal-03343071

https://hal.science/hal-03343071v1

Preprint submitted on 13 Sep 2021 (v1), last revised 29 Oct 2021 (v3)

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.

Eléments de description du paysage médiatique français

Agnès Saulnier
Institut National de l'Audiovisuel
Bry Sur Marne, France
asaulnier@ina.fr

RESUME

Pouvoir s'informer est indispensable au bon fonctionnement de la démocratie mais de nos jours les médias sont frappés par la crise de confiance et la perte de repères. Internet a bousculé l'univers médiatique et pour s'y retrouver il faut se repencher sur la notion de média afin de fournir un nouvel ensemble de critères de description des entités médiatiques. D'où vient l'information ? Comment est-elle financée ? Comment a-t-elle été produite ? Dans quel but ? Par qui est-elle diffusée ? Pour quel public ? Autant de questions essentielles à se poser pour mieux se diriger dans l'écosystème des médias. Notre objet d'étude porte ainsi sur la définition, l'observation et la modélisation des médias d'informations. Modéliser le domaine médiatique est une tâche complexe et délicate que cet article présente dans son ensemble en prenant en compte un niveau très fin de description des médias et l'aspect diachronique lié à l'évolution de la structure juridique des entités médias, de la structure de production et de la ligne éditoriale. Le modèle est testé et affiné à l'aide d'une liste de 3000 sites web d'information.

MOTS-CLES

Média, média d'information, modèle conceptuel, base de connaissance, ontologie, système documentaire

1 Introduction

Jusqu'à la fin des années 1990, les médias d'information contrôlaient toutes les étapes en amont et en aval de la production médiatique. Avec l'arrivée d'Internet, on assiste à un brouillage des pistes. De nouvelles pratiques sociales sont apparues avec le développement de nouveaux accès et modes de production de l'information. Un ensemble de nouveaux acteurs gravite désormais autour des médias traditionnels. Cependant, si Internet a engendré un plus grand pluralisme des sources d'information, ce phénomène n'est pas sans risque. D'une part, avec ces nouveaux canaux, des informations peuvent être diffusées sans contrôle ni responsabilité, créant de fait une concurrence sauvage à l'information journalistique. D'autre part, nous assistons aussi à une concentration de certains médias traditionnels dominés par un petit nombre d'acteurs. De nouveaux systèmes de repères doivent ainsi être développés pour aider les utilisateurs à s'orienter dans ce nouveau paysage médiatique assez protéiforme.

L'enjeu de ce travail est de définir les contours du paysage médiatique actuel, de proposer une grille d'observation des médias d'information sur laquelle peut reposer une modélisation conceptuelle du domaine médiatique. Une liste de 3000 médias français a été recensée afin de pouvoir aider à l'élaboration et aux tests du modèle. Cette modélisation débouchera plus tard sur la définition formelle d'une ontologie et la réalisation d'une base de connaissance médiatique. Il s'agit ainsi de fournir des critères de description des entités médiatiques afin d'aider les différents utilisateurs à répondre aux questions qu'ils se posent : Qui produit l'information ? Qui la finance ? Qui la diffuse ? Comment circule-t-elle ? Comment et pourquoi évolue la ligne éditoriale ? Pour mieux répondre à ces questions, nous avons besoin d'aller vers une plus grande finesse de description que celle des systèmes déjà existants et aussi de tenir compte de l'aspect diachronique lié à l'évolution des médias. La base de connaissances ainsi constituée sera utilisée pour aider le grand public à détecter les fake news ou les biais idéologiques, pour donner accès à une grille de lecture en vue de l'éducation aux médias, et pour analyser l'évolution de la production médiatique par les chercheurs.

2 Contour du paysage médiatique

Qu'appelle-t-on média de nos jours ? C'est à cette première question qu'il convient de répondre tant cette dénomination recouvre maintenant de réalités différentes. La presse, les ondes hertziennes, TF1, Nice-Matin, la radio, la télévision, le cinéma, les journaux, Internet, Facebook ou Twitter, autant d'exemples qui montrent le caractère protéiforme des médias et la difficulté d'établir une catégorie homogène. Avec l'arrivée d'Internet, il convient de s'interroger sur la notion de média car il est difficile de bien cerner les entités médiatiques actuelles. Redéfinir la notion de média est une tâche bien trop ambitieuse mais au regard de définitions déjà existantes, nous allons recenser les diverses définitions et évolutions pour proposer un contour élargi du domaine médiatique « traditionnel » constitué de la presse, la télévision et la radio. Notons que nous n'aborderons que la notion de média

d'information et non les médias faits pour se divertir, se cultiver ou consommer, ce qui exclue de notre étude la publicité ou les livres.

2.1 Approche théorique des médias

La provenance étymologique se situe dans le mot latin « media » (pluriel du mot neutre « medium ») qui signifie « milieux » et par extension « intermédiaires ». Au regard de cette étymologie, un média peut donc tout d'abord être défini comme l'interface entre l'émetteur et le récepteur d'un message. Une autre origine provient vers les années 1950 de l'abréviation du terme mass media, expression américaine utilisée pour désigner toutes les techniques qui permettent d'atteindre une audience étendue, diverse et dispersée. Sous cet angle, le mot média représente les moyens de communications de masse (presse, radio, TV, cinéma et affichage). Historiquement les médias ont souvent été associés à des systèmes techniques mais plusieurs auteurs, comme Olivier Bomsel [1] soutiennent que cela doit être remis en cause avec le numérique.

Si l'on définit les médias comme support de circulation d'un message, ils existent depuis l'Antiquité. Platon, Lessing, Herder, Hegel ou Nietzsche se posent comme autant de précurseurs d'une théorie du médium, aussi différentes soit leurs pensées. Dans les années 1830, Alexis de Tocqueville [2] a présenté les médias en tant que remèdes aux effets pervers de la démocratie et défini leurs rôles selon trois ordres : critiquer le pouvoir, éduquer les foules et créer du lien social. Mais l'étude des médias s'est plus profondément développée avec le début d'une réflexion sur les dispositifs techniques qui reposent essentiellement sur la Théorie des médias, la médiologie et l'archéologie des médias. Marshall McLuhan, l'un des pères fondateur de la Théorie des médias, a popularisé le mot « media » à la fin des années 1960, en donnant un sens très large aux médias, les définissants sous l'angle des supports et non sous l'angle des contenus des messages : « The medium is the message » [3]. Dans la même lignée des théories de McLuhan s'est formé la médiologie française avec Régis Debray et Daniel Bougnoux. Pour eux, la pensée est tributaire de la technique. Le médium n'est alors ni un mass-média, ni un moyen, mais un « dispositif véhiculaire », c'est-à-dire un moyen permettant la transmission d'un message [4].

Ces définitions techno-centrées sont cependant très larges dans le sens où elles risquent ainsi d'identifier les médias à tous les processus de communication, or le domaine des médias exclue la sphère des relations interpersonnelles. Pour cette raison, Olivier Bomsel, dans un article de The Conversation en partenariat avec le séminaire PSL datant de Novembre 2017, appelle média « toute signification, tout message, tout récit destiné à l'usage public ». De nombreux chercheurs en sciences de l'information et de la communication ont proposé des définitions. Selon Patrice Flichy, « un média trouve sa définition dans l'articulation d'une technique, d'un mode de financement et d'un contenu » [5]. Rémy Rieffel [6] ajoute la notion de public et Bernard Miège [7] celle de régularité de publication. Benoit Lafon a repris ces notions et a proposé un acronyme ME.D.I.A.TS [8] pour qualifier cinq dimensions des médias : le Modèle Economique, représente le modèle d'affaire visant à pérenniser l'organisation; le Discours (ou production médiatique) prend en considération l'analyse des contenus en lien avec les pratiques sociales de consommation ; l'Institution qui fait référence à l'organisation du média et son cadre légal ; le terme Acteurs qui désigne les différentes catégories d'acteurs utilisant les médias, à savoir les publics ainsi que les promoteurs d'information visant à être médiatisés (avec tout particulièrement le rôle de collectif visant à médiatiser les causes, le rôle des relations publics ainsi que celui des annonceurs) ; les Techniques en Société qui représentent les dispositifs techniques médiatiques et communicationnels.

Avec le développement d'Internet comme nouveau support de communication, se pose la question de savoir si Internet est un média. Dans sa définition des médias de masse, Wikipédia place Internet au même niveau que la presse, l'affichage, le cinéma, la radiodiffusion et la télévision. Dans le langage courant, Internet est aussi qualifié de média. Des spécialistes des médias le qualifient aussi ainsi, tel Eric Sherer qui affirme qu' « Internet est bien devenu le média global du XXIe siècle » [9]. Cependant, beaucoup d'autres défendent les limites d'Internet en tant que média. Pour Francis Balle, c'est le Web qui est un média (cité dans [10]), ce à quoi Jean-Marie Charon répond qu'il n'est pas possible de conclure pour savoir si Internet est un média ou le support de nouveaux médias, mais on ne peut pas nier pour autant que sur Internet soient présentes diverses informations de type journalistiques [11]. Nous inclurons ainsi Internet en tant que support média, avec toutes les applications associées qui aident à mettre des contenus en ligne, dans notre étude.

2.2 Fin de l'hégémonie des médias de masse

Les médias de masse ont perdu leur statut hégémonique. Le numérique a remplacé le « journalisme vertical » par un « journalisme horizontal ». Selon l'ancienne logique des médias de masse « one to many », l'information descendait des grands médias vers le public. Cette intégration verticale permettait aux médias de contrôler toute la chaîne de valeur médiatique : création, édition-publication, diffusion, consommation. Avec le flux actuel d'information en réseaux « many to many », l'information provient d'acteurs plus variés, peut être publiée sur différents canaux et circule parmi les internautes sous des formes différentes. Benoît Lafon [8] précise que l'extension du domaine médiatique suit un double mouvement: tout d'abord du côté des médias de diffusion, on assiste à un renouvellement de l'offre grâce aux techniques

numériques (pures players, chaine youtube..) ainsi qu'à multiplication des moyens d'accès (podcast, TV de rattrapage, archives en ligne…); dans le même temps, du côté des réseaux de mise en relation, la chaîne de valeur s'étoffe de nouvelles activités comme la ré-éditorialisation de contenus par des usagers (Facebook, Twitter), l'éditorialisation des échanges, l'éclosion de commentaires et de recommandations sous forme de forum ou de micro-médiatisation. A ceci, il faut ajouter le rôle considérable des nouveaux intermédiaires de l'information. Tous les maillons de la chaine médiatique sont donc bouleversés.

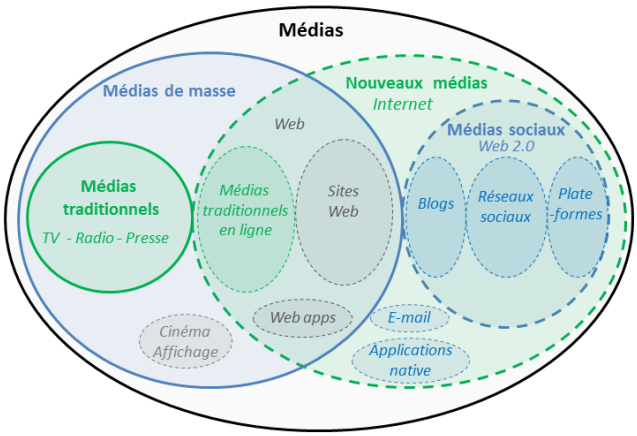

Figure 1 : Cartographie des supports médias

Nous pouvons dresser une cartographie des supports médias (Figure 1). Les médias traditionnels (TV, radio, presse) se voient opposés aux nouveaux médias sur Internet par rapport à la nature des supports. D'autre part les médias de masse se voient opposés aux médias sociaux par rapport au mode de diffusion verticale vs horizontal. A l'intersection se trouve le Web sur Internet qui peut être considéré comme un média de masse qui diffuse un contenu original de une vers plusieurs personnes (la masse). Au contraire, les médias sociaux sur le Web 2.0 ont une technique de diffusion de structure « rhizomatique » où tout le monde est à la fois diffuseur et récepteur. Fred Cavazza, blogueur reconnu dans le domaine des médias sociaux, propose tous les ans sur son blog un panorama détaillé des médias sociaux. Nous avons sélectionné dans les structures les plus couramment utilisées pour partager ou diffuser des informations d'actualité : plateformes de blogs, réseaux sociaux (ex Facebook, Twitter), plateformes de partage (vidéo ex Youtube, photo ex Flickr, podcasts ex iTunes). Nous n'avons pas mentionné le Web 3.0 (web sémantique) car il n'en est qu'à ses débuts. Remarquons que l'email peut aussi être considéré comme un support média comme l'atteste l'exemple de Brief.me.

2.3 Confusion entre information et communication

De nos jours, Internet véhicule un grand nombre d'informations qu'il est important de pouvoir distinguer car le terme information tend à se confondre avec celui de contenu, faisant disparaitre la valeur ajoutée journalistique. François Demers parle de communication publique pour repositionner le journalisme contemporain [12]. Sur le web nous pouvons voir le mélange de sources médiatiques, formelles et informelles que nous avons résumé dans la Figure 2.

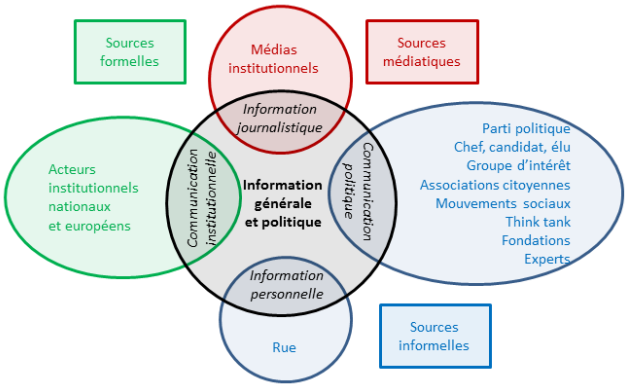

Figure 2 : Information générale et politique sur Internet

Différents types de contenus se partagent les arènes du débats publics : l'information journalistique produit par les médias traditionnelles et les agences de presses, l'information personnelle qui peut prendre la forme de journalisme citoyen, la communication institutionnelle émise par des sources reconnues ou encore la communication politique pratiquée par des organisations ou des personnes politiques. Benoit Lafon a utilisé le concept de médiatisation pour intégrer les nouveaux phénomènes sociaux et communicationnels. « La médiatisation consiste en la mise en média d'individus, de groupes ou d'institutions par la construction de produits médiatiques formalisés, dans une visée stratégique, impliquant des pratiques collectives de consommation » [13, p163]. Nous inclurons ainsi tous ces nouveaux acteurs de médiatisation dans notre domaine en veillant à bien les distinguer.

2.4 Un domaine médiatique étendu

Les médias historiques se sont transformés et le domaine médiatique s'est étendu dans plusieurs directions depuis l'avènement d'Internet (Figure 3). En premier lieu nous avons déjà constaté que les médias actuels ne sont plus assimilables à leurs supports. Ils ont aussi vu se développer une large concurrence sur Internet, non seulement avec la création de médias pure players mais aussi avec la présence de sites d'informations détenus par des acteurs institutionnels, politiques ou même par de simples citoyens. Les médias traditionnels se sont transformés grâce à l'apport des

nouvelles technologies : un média provenant de la presse écrite peut désormais avoir son site web, éditer des blogs de citoyen et même produire des vidéos diffusés notamment sur YouTube. Les formes de journalismes se sont multipliées avec, par exemple, le « live journalisme » ou le journalisme de liens. Enfin les infomédiaires ont bouleversé l'accès à l'information avec l'essor des moteurs de recherche et des réseaux sociaux. C'est donc ce domaine médiatique ainsi étendu qu'il convient de modéliser.

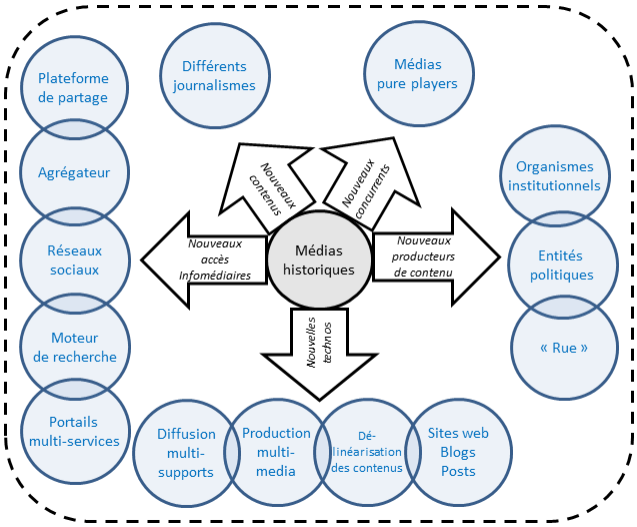

Figure 3 : Extension du domaine médiatique

3 Grille d'observation

Avant de pouvoir proposer une modélisation conceptuelle du domaine médiatique, il est utile de relever tous les termes descriptifs utiles, ainsi que les typologies déjà existantes, pour construire une grille d'observation conjointe des médias traditionnels et issus d'Internet.

3.2 Analyse des besoins

Dans un objectif d'éducation aux médias (ex site du CLEMI) ou pour aider le grand public à s'orienter dans ce nouveau domaine médiatique, il faut pouvoir répondre aux questions posées par Harold Lasswell en 1948, et toujours très actuelles, qui nous indiquent à quoi doit pouvoir répondre la base de données «Qui, dit quoi, à qui, par quel canal, avec quel effet ? ». Ce à quoi nous pouvons ajouter les notions de contexte et de code qui ont été mentionnés dans le modèle de Roman Jakobson en 1963. La détection de fakenews est un problème très complexe à la fois pour les définir et les identifier. Notre base ne prétend pas fournir la solution pour faire le tri entre « bonne » et « fausse » information mais fournir des critères pour aider à cette tâche. Pour repérer les fake news, Claire Wardle [14] membre de la coalition international des journalistes First Draft News, propose d'analyser l'écosystème de l'information suivant trois critères à savoir les types de contenus, les intentions des auteurs et les mécanismes de diffusion. Nous pouvons aussi nous référer à l'état de l'art de Xinyi Zhou et Reza Zafarani [15] qui présentent différentes méthodes de détections manuelles et automatiques sous quatre angles : les fausses connaissances véhiculées, le style d'écriture, les modèles de propagation et la crédibilité de la source. Enfin les chercheurs, mais aussi parfois le grand public, ont besoin de données extérieures par exemple sur les propriétaires, fondateurs ou audiences, ainsi que sur les liens possibles avec les autres publications. Le Tableau 1 synthétise l'ensemble de ces besoins que nous allons détailler par la suite.

Métadonnées administratives (identification)
Aspect technique (format et transmission du message)
Aspect contenu (modèle éditorial)
Aspect organisation (propriétaire, éditeur, statuts juridiques)
Aspect économique (modèle financier)
Aspect usage (type de producteurs, auteurs-contributeurs)

Tableau 1 : Dimensions de la grille d'observation

La base de données quant à elle devra pouvoir répondre à des requêtes spécifiques. Nous pouvons nous référer aux opérations classiques concernant les systèmes documentaires, même si nous ne nous positionnons pas au niveau des documents mais au niveau des publications médiatiques (titres produits par un média). Le FRBR a émis des spécifications (trouver, identifier, choisir, obtenir) qui sont relatées par la BNF. Nous pouvons ajouter celle de l'IFLA Library Reference Model (LRM) de 2017 (naviguer) et celle du modèle FRAD (justifier). Enfin nous pouvons ajouter des besoins utilisateurs plus spécifiques : comparer des publications entre elles, évaluer la qualité pressentie d'un contenu (à l'aide de comparaisons ou de données extérieures).

3.2 Modèles de référence

S'il existe beaucoup de travaux théoriques sur l'analyse des médias, peu d'entre eux tentent de modéliser en même temps les anciens et les nouveaux médias. Au sein de l'ouvrage collectif « Médias et médiatisation » [13], Benoit Lafon a présenté les principaux concepts et a réalisé une synthèse des travaux de recherche sur ce sujet. Il a lui-même proposé le modèle ME.D.I.A.TS que nous avons déjà cité pour définir les médias. De son coté, Jean-Marie Charon [16] a aussi étudié l'écosystème médiatique et a modélisé les différents acteurs de l'information, en précisant les rôles, leur organisation, le contenu de leur activité, leur production ainsi que le rapport au public et les interrelations qu'ils ont les uns avec les autres (compétitions, mutualisation…). Les acteurs concernent aussi bien les éditeurs, les agences que les centres de formation, les infomédiaires, les annonceurs ou bien même le public. Enfin, Axel Bruns [17] a étudié les processus de communication d'informations avec l'intégration des plateformes de médias sociaux et leur impact sur le journalisme. Tous ces travaux

sont très enrichissants mais ne fournissent pas de grille directement exploitable.

Sadok Hammami [18] a quant à lui réalisé une grille d'observation pour étudier la presse électronique tunisienne. Franck Rebillard et Ghislaine Chartron [19] ont aussi proposé une grille d'analyse des médias numériques formalisée par des « modèles de publication » qui s'appuie sur les acteurs, les modes de financement et la valeur des contenus. Plus tard, en 2010, Emmanuel Marty, Franck Rebillard, Nikos Smyrnaios et Annelise Touboul [20] ont proposé une typologie des sites web d'information générale et politique qui reposait sur les catégories suivantes :

- presse en ligne (versions internet de médias existants),
- webzines (publications collectives internet),
- blogs (publications individuelles internet),
- sites participatifs (publications collaboratives internet),
- portails (composantes informationnelles de plates-formes multiservices),
- agrégateurs (regroupements automatisés d'informations d'actualité).

Cette typologie a été mise à jour en 2012 par Emmanuel Marty, Franck Rebillard, Stéphanie Pouchot et Thierry Lafouge [21] qui ont cartographié l'espace web à partir de l'analyse de 200 sites web d'information générale et politique à caractère essentiellement national. Ils ont proposé une typologie plus restreinte, à laquelle il faut ajouter le caractère professionnel ou amateur.

- médias en ligne (provenant de la presse, TV ou radio),
- sites natifs de l'Internet (webzine et site participatif),
- blogs,
- infomédiaires (portails et agrégateurs).

Pour terminer nous pouvons citer une infographie du paysage médiatique, délivrant une modélisation beaucoup plus subjective et poétique, qui a été donnée par Anne-Sophie Novel dans son blog lesmediaslemondeetmoi.com. Cette carte représente le paysage protéiforme médiatique en fonction de ses propriétés. Elle n'est bien sûr pas exhaustive mais montre bien la diversité des médias. Tous ces travaux sont très intéressants à prendre en compte mais nous souhaitons aller plus loin dans le niveau de description.

3.3 Aspect technique

L'aspect technique décrit le type de support ainsi que le format physique de la publication. Du point de vue des supports, comme nous l'avons déjà vu, Internet fait partie du domaine médiatique étudié, avec tout particulièrement le Web et les médias sociaux. Les formats physiques de publication dépendent directement des supports de diffusion.

Les deux formats de publication de la presse écrite (format papier) les plus connus sont le journal et le magazine. On peut ajouter à cette liste les revues, les mooks, les newsletters ou les bulletins de presse. Si on veut obtenir une plus grande précision de description, les cahiers correspondent à des pages spéciales dans les journaux. Pour la télévision, comme pour la radio, nous n'avons pas besoin de décrire les normes de transmission (numérique, analogique) ni les modes de diffusion (ondes hertziennes, ADSL, …). Il pourra aussi être utile de tenir compte des formats de diffusion de contenu audiovisuel : podcast, streaming, live, replay, vidéo à la demande.

3.4 Aspect contenu

L'aspect contenu repose sur la ligne éditoriale du média qui qui détermine la manière d'adresser un message au public. La ligne éditoriale définit à la fois la nature des contenus proposés ainsi que leur mise en forme.

La presse dispose déjà de nombreux systèmes de classification. La nature peut tout d'abord correspondre à la cible du public. On distingue ainsi la presse généraliste (d'information générale et politique), de la presse spécialisée (correspondante à l'âge des lecteurs ou à leur centre d'intérêt), et la presse professionnelle (orientée métier). La presse professionnelle tendant à disparaitre sur le Web pour se fondre avec la presse spécialisée. La nature du contenu peut aussi être définie par sa thématique, mais il existe autant de listes de thématiques que de sources qui les utilisent (BNF Gallica, Press directory, Catalogue des Messageries Lyonnaises de Presse, Presstalis, Giga Press…). Ensuite, la presse est couramment classifiée par périodicité (quotidien, hebdomadaire..) et par périmètre (national, régional…). Il existe un grand nombre de familles de presse qui proposent des classifications exclusives considérant ces différents critères. Nous pouvons nous référer aux typologies de Patrick Eveno [22], Jeanne Bordeau [23] ainsi que celle utilisée par l'ACPM, Press directoty, Giga Presse, Presstalis, Presse News ainsi que le ministère de la culture (La-presse-en-chiffres) et le Sénat. Ces familles s'accordent généralement sur le découpage en PQN (Presse Quotidienne Nationale), PQR et PHR (Presse Quotidienne et Hebdomadaire Régionale), PGI (Presse Gratuite d'Information). En revanche d'autres familles incluent dans cette typologie des notions de thématique ou de public cible, certaines utilisent la notion de magazine tandis que d'autres préfèrent la classe PHN (Presse Hebdomadaire Nationale). Parfois sont ajoutées des classes supplémentaires (presse associative, syndicale..). La presse en ligne ne contient pas que du texte et des photos mais aussi de l'audio et de la vidéo. Les sites web de presse traditionnelle peuvent faire référence au classement du titre papier associé. Pour les autres sites d'information, il n'existe pas de famille qui soit déjà en mode d'exploitation.

La radio admet aussi des familles reconnues qui reposent sur plusieurs critères. Tout d'abord l'Inathèque propose une classification en radio généraliste publique, généraliste privée

et thématique (économie, musique, divertissement). Le classement en radios de service public et en radios privées sont quant à elles divisée par le Conseil supérieur de l'audiovisuel en 5 catégories (A : radios associatives ; B : radios commerciales indépendantes ; C : radios commerciales locales ou régionales ; D : radios commerciales thématiques à vocation nationale ; E : radios commerciales généralistes à vocation nationale). Mais les radios peuvent aussi être classées en fonction du type de programmes qu'elles diffusent. Pour la télévision, le CSA propose une catégorisation en chaînes nationales gratuites, nationales payantes et locales ou régionales. L'Inathèque a développé une classification pour la consultation de ses archives qui repose essentiellement sur le critère de cible du public (généraliste, information continue, jeunesse, sports et loisirs,… ainsi que territoires pour les télévisions régionales ou étrangères). Les émissions audiovisuelles peuvent aussi être classifiées par thématique (magazine TV, Talkshow, JT, documentaire…). Nous pouvons nous référer à l'Ina ou directement aux chaînes.

En revanche, pour Internet, il n'existe pas de famille officiellement reconnue. Nous pouvons nous inspirer de la littérature [21] pour proposer une taxonomie. Nous pouvons ainsi distinguer les sites Web, les blogs et les portails, auquel nous pouvons ajouter les Web TV et radio ainsi que les plateformes de contenu qui gèrent des contenus qu'elles ne créent pas. Nous distinguons aussi le portail d'actualité qui rassemble un ensemble d'actualités provenant d'autres sites web et qui propose un lien vers ces sites (Google actualités), le portail Web qui contient des ressources et des services et sert de passerelle vers le Web (Google, Yahoo), et le portail média qui regroupe plusieurs médias appartenant à un même réseau (lefigaro.fr, francetvinfo.fr, reussir.fr, mutu.fr).

La notion de genre journalistique permet aussi de classer les différents contenus des productions médiatiques. Toutes les productions ne se ressemblent pas car elles peuvent avoir été réalisées avec des techniques différentes pour des finalités différentes. Le reportage, la brève, l'interview ou l'éditorial en sont des exemples. Comme le relate Roselyne Ringoot et Jean-Michel Utard. « Le genre retenu va déterminer en amont le processus de production (modalités de recueil de l'information, temps passé) et en aval le type d'écriture » [24]. Ainsi les genres peuvent aider à trier les différents organes d'information car chaque journal va exploiter différemment les potentiels des genres : « la typologie des genres est une question capitale pour le discours de la presse, car elle touche à l'identité même du journal ». La définition des genres journalistiques est ainsi à mettre en relation avec les catégories du journalisme. Nous pouvons par exemple citer celle d'Erik Neveu [25]. Journalistes et linguistes ont proposé de nombreuses typologies des genres journalistiques, la plupart étant plus particulièrement orientées vers la presse écrite [26, 27, 28, 29].

Conséquences du numérique, de nouvelles pratiques journalistiques sont nées récemment. Le live journalisme répond au besoin d'instantanéité du journalisme « conventionnel » tandis que le « journalisme au long cours » se renouvelle grâce à de nouveaux formats et de nouvelles formes de narration. Le journaliste Louis Haushalter a proposé dans un article d'Owni du 23 septembre 210 une typologie assez complète qui distingue six types de Webjournalisme : le « flash journalism » (traitement de l'information brute), le « live journalism » (forme de journalisme textuel ou vidéo qui consiste à faire vivre les évènements au public en direct), le « narrative journalism » (genre hybride entre le journalisme et la littérature), le « data journalism » (données brutes livrées aux internautes avec différentes méthodes de visualisation, tableaux, schémas, infographies interactives), le « blog journalism » (technique du blog appliquée au journalisme professionnel ou amateur), le « talk journalism » (journalisme d'interaction qui rassemble le traitement des réactions et témoignages des internautes). Nous pouvons ajouter à cette liste le « link journalisme » (travail manuel de veille pour proposer des liens sur un sujet précis) et le « fact-checking » (travail de vérification des médias et de contrôle de la véracité des propos tenus par des personnalités politiques dans l'espace public). Ces catégories ne sont bien sûr pas exclusives. Par exemple, les publications de pure players type *Rue89* sont souvent au croisement du narrative journalism, du talk journalism et – parfois – du data journalism. A côté de ces différents genres journalistiques se trouvent aussi sur les web des agrégateurs automatiques d'actualité ou des portails multi-services. Techniquement les portails multi-services proposent des informations achetées à des tiers (actu.orange.fr propose des dépêches de l'AFP) tandis que les agrégateurs de contenu proposent une sélection d'hyperliens qui renvoient vers des informations éditées par des sites d'information extérieurs.

3.5 Aspect organisation

L'aspect organisation (ou institutionnel) recouvre le cadre légal dans lequel le média doit exercer son activité. Plusieurs acteurs sont impliqués dans cette tâche. Le rôle de l'éditeur est central dans un média. En effet, le producteur de contenu peut tout à fait être indépendant du média contrairement à l'éditeur. C'est l'éditeur qui définit les règles éditoriales. La diffusion de contenu est assuré par les distributeurs (Presstalis) pour la presse et des diffuseurs (TDF) pour les publications radio et télé. L'éditeur agit donc comme un médiateur entre les producteurs de contenu et les consommateurs. Michaël E. Sinatra et Marcello Vitali-Rosati [30] leur ajoutent comme fonctions le fait de légitimer les contenus et de les mettre en circulation.

Du point de vue légal, l'entreprise de presse n'est pas obligatoirement définie ni régie par un statut particulier. « Aucune obligation particulière n'est prévue quant à la forme

juridique que doit revêtir une société éditant une publication de presse, qu'elle soit éditée sous forme papier ou sous forme électronique » telles sont les recommandations apparaissant sur le site culture.gouv.fr[1] . Il n'est donc pas aisé d'identifier ces sociétés. Néanmoins, pour la presse papier, il existe un régime de déclaration prévu par la loi du 29 juillet 1881, ainsi que des formalités de dépôts (dépôt administratif et judiciaire, dépôt légal auprès de la Bibliothèque nationale qui fournit l'attribution d'un code d'identification de la publication ISSN, dépôt auprès de la DGMIC pour une diffusion nationale). L'enregistrement auprès de la CPPAP ouvre des droits, à des services d'aide à la presse proposés par le ministère de la culture. En ce qui concerne la presse en ligne, le décret n° 2009-1340 du 29 octobre 2009 est venu préciser la définition du service de presse en ligne figurant à l'article 1er de la loi n°86-897 du 1er août 1986 portant réforme du régime juridique de la presse. Il s'applique à tout service de communication au public en ligne, édité à titre professionnel par une personne physique ou morale, qui a la maîtrise éditoriale de son contenu. Ce statut accorde aussi des avantages notamment fiscaux ou d'accès à des fonds d'aide. La CPPAP gère aussi la reconnaissance de service de presse en ligne. Les sites Internet personnels et les blogs, édités à titre non professionnel, ne sont pas concernés par ce statut de presse en ligne relevant de la CPPAP.

Jean-Marie Charon [31] fait remarquer qu'il est important de distinguer les éditeurs de contenu, qui ont fait le choix de s'inscrire dans le cadre éditorial et juridique des éditeurs de presse en ligne (Mediapart), des éditeurs de services (Linternaute, CCM Benchmark, Aufeminin.com) qui n'ont pas signé la convention collective des journalistes mais celle beaucoup moins contraignante des secteurs de services (Syntec). Le journaliste professionnel se voit remplacé par un rédacteur de contenu qui n'a pas des obligations de rédaction (nouveauté, vérification des informations, objectivité). Notons pour terminer que si un blog n'a pas un statut de presse en ligne, il est néanmoins un « éditeur de services de communication publique en ligne » et à ce titre, il est soumis à des lois comme tout site Internet. Il faut se référer à la Loi pour la confiance dans l'économie numérique (LCEN)[2], qui encadre la responsabilité du bloggeur en cas de litige relatif à des contenus illicites ou préjudiciables. Ainsi par exemple, il est obligatoire de publier des « mentions légales » sur un blog comme sur tout autre site web. C'est toujours l'éditeur qui reste responsable du traitement de l'information. La nature du Producteur quant à elle permet de mieux distinguer les contenus journalistiques des autres contenus.

[1] https://www.culture.gouv.fr/Sites-thematiques/Presse/Aides-a-la-Presse/Presse-papier-presse-en-ligne-quel-statut, consulté le 14 juin 2021
[2] https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000801164/, consulté le 14 juin 2021

Il existe d'autres acteurs externes qui participent à l'organisation des médias : les propriétaires et groupes médias qui possèdent les médias, ainsi que les annonceurs, les régies publicitaires, et encore les marques. Regardons de plus près le rôle des marques. Les marques jouent en effet un rôle important car les médias traditionnels confrontés au chamboulement de leur activité et aux nouveaux usages (réseaux sociaux, multi supports, délinéarisation des contenus, programmes à la carte) doivent repenser leur identité. La marque permet aux médias d'émerger dans la bataille de l'attention et de fidéliser les consommateurs sur l'ensemble des supports. Ainsi que le soutient Olivier Bomsel [32], la marque permet de signaliser une offre quand celle-ci est éclatée sur le web.

3.6 Aspect économique

Comme l'affirme Pierre Rimbert dans un article du Monde diplomatique datant de décembre 2014, les médias sont soumis à une ambivalence : « l'information est pensée comme un bien public, mais produite comme une marchandise ». Les médias sont soumis à la loi du marché et doivent trouver un moyen de se financer faute de quoi ils sont voués à disparaître.

Avant Internet, il existait deux modèles d'affaire, le modèle éditorial fondé sur la vente à l'unité (livre, disque) et le modèle de flot fondé sur la programmation et le flux (TV-radio). La presse quant à elle, pouvait s'appuyer sur un double financement, à savoir celui des lecteurs et celui des annonceurs. Ainsi pour s'informer sur le sport par exemple, le lecteur achetait le journal et ne lisait que les articles relatifs au sport, tout en supportant financièrement l'intégralité du journal, en particulier les reportages plus chers à produire. Et la vente des journaux en kiosque ou par abonnement était complétée par le revenu de la publicité et des petites annonces. Ce modèle à double financement a été décrit par Nathalie Sonnac [33] en tant que « marché à deux versants » car la vente d'un double produit génère des effets de réseau croisés entre le marché des médias et le marché de la publicité. Exception typiquement française, l'Etat apporte aussi une troisième source de subvention par l'intermédiaire d'aide à la presse pour tenter de garantir un certain pluralisme des médias. En ce qui concerne l'audiovisuel, le secteur public tient une place importante dans l'offre d'information avec France Télévisions et Radio France qui détiennent la plus grande rédaction de France. L'AFP, est également financée par l'Etat.

De nos jours, le modèle est passé de l'économie matérielle à celui de l'économie immatérielle. Si nous dressons un constat de nos jours, nous pouvons tout d'abord constater que les géants du web, les fameux GAFA (Google, Amazon, Facebook, Apple), ont récupéré une bonne partie du revenu provenant des petites annonces et de la publicité. Dans leur livre [34],

Gérôme Guibert, Franck Rebillard, Fabrice Rochelandet expliquent qu'une concentration économique s'est formée autour de ces plateformes. Ensuite, l'offre s'est démultipliée. Pour s'informer sur le sport, le lecteur consulte maintenant un média spécialisé et n'achète plus un journal pour lire juste une rubrique. Une partie du public est donc perdue. Enfin, le modèle de la gratuité a capturé une bonne part du lectorat. Le modèle de double financement avait commencé à vaciller au début des années 2000 avec la naissance des journaux gratuits comme Métro, mais celui touchait un public différent principalement constitué de non-lecteurs ou de lecteurs occasionnels de la presse payante. La gratuité de l'information sur le Web a en revanche révolutionné le mode de consommation de l'information. Une partie du public est maintenant convaincu qu'il ne faut pas payer pour avoir de l'information. Et cela n'est pas sans conséquence sur le contenu produit par les médias, car pour un média exclusivement financé par la publicité, il faut offrir aux annonceurs un produit attractif, où l'audience est l'indicateur principal de sa mesure. Or maximiser l'audience revient à choisir des contenus qui plaisent au plus grand nombre, ce qui ne favorise pas une diversification des contenus mais au contraire une uniformisation des contenus. Le modèle économique agit donc sur le contenu, alors que l'objectif des médias traditionnels est d'informer, l'objectif des médias gratuits, ainsi que des réseaux et plateforme, est de faire le plus d'audience possible pour maximiser le revenu apporté par les annonceurs. Cet effet réseau a été détaillé par Nathalie Sonnac [35, 36]. La perte de valeur a engendré un changement dans le modèle économique qui est passé d'une économie de la rareté à une économie de l'abondance. Ainsi, l'attention devient la plus précieuse ressource de valeur. Yves Citton [37] explique ainsi que de nos jours nous avons accès à une quantité d'informations bien supérieure aux capacités attentionnelles dont nous disposons pour en prendre connaissance. Dès lors l'attention devient la nouvelle rareté. Et cette rareté se situe non plus du côté de la production des informations mais du côté de la réception. L'attention peut se calculer par exemple avec le Google Page Rank qui a remplacé l'audimat. Dominique Cardon a proposé plusieurs métriques de l'attention qui sont la popularité, l'autorité, la réputation et la prédiction [38]. Diverses traces laissées par l'attention des internautes peuvent servir de données à collecter (vues, liens, likes, recommandations). Dans ce modèle, les médias ou sites d'info fournissent ainsi de l'information que les GAFA transforment en attention qui peut être ensuite revendue à des annonceurs. Comme l'a rappelé Tim Wu, dans son livre Les marchands d'attention [39], l'exploitation de l'attention ne relève pas d'une démarche d'information mais d'un processus de création de valeur. Les GAFA font ainsi un usage commercial des données privées qu'ils collectent gratuitement. En plus de s'enrichir par les publicités, ils s'enrichissent aussi par les données qu'ils pourront revendre. Ainsi, les logiques économiques réduisent l'information à sa seule valeur commerciale. C'est ce que Bourdieu avait dénoncé en développant le concept de « marchandisation de l'information ». Une information ne devient pertinente que si elle est rentable, ce qui sert les intérêts des propriétaires au détriment du collectif.

Face au problème économique, les médias traditionnelles ont développé différentes solutions qui reposent sur une optimisation et réduction des coûts de production (optimisation organisationnelle, diminution de la taille des rédactions), des tentatives de consolidation de marché et d'économie d'échelle (concentration des médias) ainsi que l'investissement dans des solutions techniques apportées par le numérique (diversification des supports et des services, développement de portails, personnalisation des contenus). De nombreuses données sont ainsi intéressantes à collecter, qui concernent aussi bien les modes de paiements (gratuité, abonnement, mode freemium..), les subventions publiques ou les plans d'aide ponctuels des GAFAM, les publicités, l'appel au don ou le crowdfunding, mais aussi la proposition de services et de formations, ou encore les mouvements de fusion. Il sera aussi utile d'intégrer des données extérieures sur les audiences, diffusions, fréquentation, nombre d'abonnés, provenant par exemple de l'ACPM, Alexa ou Médiamétrie.

3.7 Aspect usage

Les caractéristiques des producteurs de contenu ainsi que des auteurs/contributeurs sont des éléments très utiles pour prédire les usages escomptés. Journalistes professionnels et communicants se retrouvent en effet sur le Web et la frontière entre ces deux domaines devient de moins en moins visible. Leurs productions se voient forcées de cohabiter. Les communicants ont essentiellement pour rôle de promouvoir à l'image des institutions politiques et organisations officielles. Leur intérêt peut-être à la fois social et politique. Des experts et des citoyens peuvent aussi développer le même intérêt. En politique peuvent aussi cohabiter des acteurs officiels à coté de militants ou de groupes d'intérêt politique. Enfin il ne faut pas oublier un troisième intérêt, financier cette fois-ci, qui comme nous l'avons déjà vu motivent tout spécialement les services liés aux moteurs de recherche ou aux réseaux sociaux. La Figure 4 qui recense ces usages est cependant très théorique car derrière le journalisme, et tout particulièrement le journalisme de communication, on peut voir les intérêts financiers des annonceurs et publicitaires. De même les médias citoyens peuvent avoir un intérêt financier par exemple dans les cas de site de « clickbait ».

Notons que la notion de journalisme professionnel est difficile à définir. Nous nous reposerons donc sur la loi Brachard 1935 à laquelle nous pouvons ajouter les notions de déontologie et les méthodologies journalistiques. Rappelons-le, le journalisme doit être mené en respectant une certaine éthique. Armande Saint-Jean [40] affirme que « Les

fondements philosophiques sur lesquels s'appuie l'éthique journalistique peuvent être ramenés à trois concepts clés : la liberté, la vérité et la responsabilité. » A côté des professionnels, les amateurs ont pris une place indéniable dans les médias mais n'ont pas les mêmes contraintes de publication. Il est donc important de pouvoir les identifier. Ils peuvent jouer à la fois des rôles au sein de la création, de l'édition et de la diffusion de la production journalistique (Figure 5). La frontière avec le journalisme professionnel est parfois difficile à trouver, comme dans le cas du youtubeur et journaliste Hugo Travers. En effet il exerce du journalisme citoyen lorsqu'il s'exprime aux jeunes sur sa chaine YouTube sous le pseudo Hugo Décrypte, mais il réalise cependant un travail journalistique très sérieux d'enquête et d'interview pour sa chaîne et il a d'autre part rejoint la chaîne LCI en tant que chroniquer.

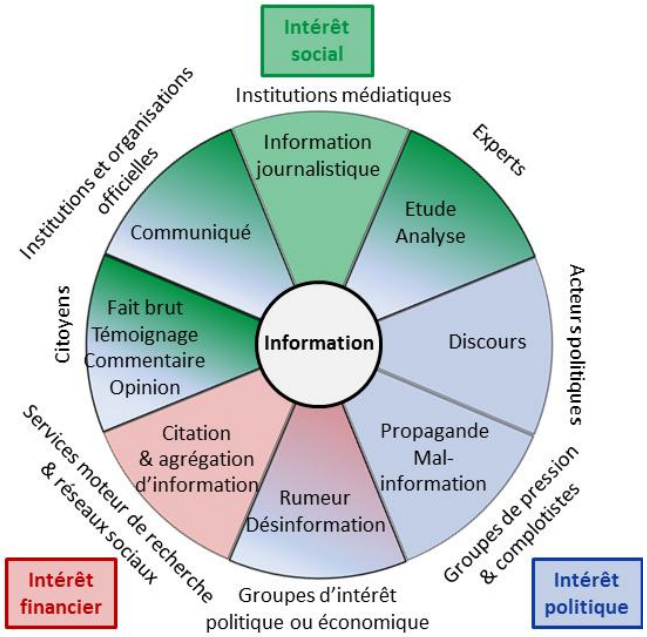

Figure 4 : Intérêts divergents sur le Web

Regardons de plus près les types de journalisme sur le Web. Le journalisme citoyen est une production ou diffusion d'informations d'actualité gérées par des non-professionnels. Nous avons choisi d'inclure l'enjeu social dans la définition du journalisme citoyen afin de le séparer des publications de communication politique. Il existe par exemple de nombreux blogs, comme celui de Maître Eolas. Le Crieur de la Villeneuve qui est un journal du quartier de la Villeneuve à Grenoble, citoyen et participatif, accréditée par la SPILL et la CPPAP sera classé Journalisme citoyen et participatif. Le journalisme social est une extension du journalisme citoyen car il ajoute une validation des productions amateurs des professionnels. Café Babel et Blasting News en sont des exemples significatifs. Le journalisme participatif implique la participation d'amateurs extérieurs au média, média qui peut être lui-même citoyen ou professionnel. Nous distinguons plusieurs niveaux de participation intégrée (ex Rue 89) ou juxtaposée (ex Le blog du Monde). Remarquons qu'il existe aussi des plateformes communautaires de journalisme citoyen participatif (ex Agoravox) qui se rapprochent très fortement du journalisme social. La différence repose sur la validation des contenus qui n'est pas faite par des professionnels mais par des rédacteurs amateurs qui ont déjà publiés au moins 4 articles. Le journalisme professionnel est quant à lui illustré aussi bien par les sites en ligne de médias traditionnels (ex LeMonde.fr) que par des pure players (ex Médiapart). Enfin le journalisme collaboratif correspond à la collaboration de plusieurs médias professionnels dans le cadre d'enquêtes. Il est pratiqué par des journalistes professionnels et parfois amateurs. CrossCheck a par exemple été un projet de journalisme collaboratif pour lutter contre la désinformation en ligne pendant les dix semaines qui ont précédées les élections présidentielles de 2017.

AM	Journalisme citoyen	• Production ou diffusion d'info d'actualité par des non-professionnels avec un enjeu social • Blog, site web, web TV, tweet,….
PRO-AM	Journalisme social	• Journalisme citoyen encadré par des professionnels • Création amateur d'article, reportage TV, interviews, … validée par des professionnels
AM ou PRO	Journalisme participatif	• Journalisme citoyen ou pro avec participation d'amateurs • Participation juxtaposée : blog, commentaire, chat, forum • Participation intégrée : co-production de contenu
PRO	Journalisme collaboratif	• Collaboration entre journalistes professionnels appartenant à des médias différents • Enquêtes, investigation, …
	Journalisme professionnel	• Journalisme exercé à « titre professionnel » • Productions journalistiques gérées par des professionnels (selon loi Brachard 1935) • Médias traditionnels en ligne, medias pure player

Figure 5 : Type de journalisme sur le Web

Afin de décrire les différents types de journalisme présent sur un site Web, il est utile de pouvoir décrire très finement la structure du site pour le cas échéant distinguer un blog d'un site de média en ligne. L'ajout de données extérieurs peut aussi apporter des indications précieuses pour mieux cerner un média. Ainsi nous pourrons lier des informations relatives au fondateur du média, à des récompenses ou condamnations. Il est aussi important d'indiquer les médias parodiques ou satiriques ainsi que les médias d'opinion. L'orientation politique n'est pas facile à trouver mais quand cela est

possible elle pourra aussi être indiquée avec l'origine de la source. Nous pouvons aussi remarquer que la fachosphère qui est constitué par l'extrême droite sur Internet, constitue un sous-groupe intéressant à analyser car elle est très présente sur le Web. La fachosphère se nomme elle-même site de « réinformation » et se présente comme une alternative aux médias traditionnels. Selon Gaël Stephan et Ysé Vauchez [41], les sites de réinformation ont en effet un double positionnement à la fois informationnel et militant. Le terme de fachosphère regroupe des familles très différentes mais les sites se référencent beaucoup entre eux. De manière plus générale, les liens du média étudié vers les sites amis, ou au contraire les liens de listes extérieures (CPPAP, SPILL, syndicats, associations…) sont des indicateurs à répertorier. Il est même possible de décrire encore plus finement les liens en analysant les pourcentages de copié-collé comme l'ont fait Julia Cagé, Marie-Luce Viaud et Nicolas Hervé [42].

4 Modélisation du domaine médiatique

Le modèle conceptuel de la base de données est une représentation simplifiée de la réalité qui est destinée à réaliser des tâches particulières. Notre modèle doit décrire de manière homogène des médias très divers et tenir compte de contraintes spécifiques :

- Niveau très fin de description (blog d'un site de presse).
- Aspect diachronique (changement de nom, de règle éditoriale, fusion avec un autre média).
- Notion de provenance (valeurs multiples provenant de plusieurs familles de presse).

Commençons par répertorier les modèles existants répondant à nos usages afin de les appliquer, de les adapter, de les assembler ou de s'en inspirer pour proposer notre propre modèle.

4.1 Ontologies et bases de connaissance de référence

Du point de vue des bases de connaissances généralistes déjà existantes, Wikidata et DBpedia sont à ce jour probablement les plus complètes. Elles fournissent un grand ensemble de connaissances semi-structurées sur le principe du Linked Open Data. Elles référencent ainsi un grand nombre de médias via l'utilisation d'identifiants uniques mis en relation entre bases et fournissent leur lot d'informations sous forme d'ensemble de couples propriétés-valeurs. Nous proposons ici une structure plus fine de description. Par exemple lemonde.fr n'est pas décrit comme une entité de Wikidata. Il manque donc une finesse de description des sous-entités qui ont des propriétés spécifiques. En revanche, nous pouvons remarquer que la notion de provenance est centrale dans Wikidata, ce qui permet de sourcer et de dater une instance. Notons que la BNF constitue aussi une ressource intéressante mais elle est axée sur les œuvres littéraires et non sur les périodiques.

En ce qui concerne les ontologies, Emilio Sanfilippo [43] a effectué un état de l'art assez général sur celles dédiées aux entités d'information. Les ontologies sont constituées de classes et de propriétés. Les classes représentent des concepts (objets, concrets ou abstraits, du monde réel) qui peuvent être organisées en taxonomie (hiérarchie). Les propriétés incluent des attributs (ayant pour valeur un texte libre ou un type de données particulier) et des relations (vers d'autres classes ou ressources identifiables). Regardons les ontologies spécialisées dans les systèmes documentaires. Le Dublin Core, la plus ancienne des ontologies du domaine documentaire, est toujours utile pour décrire un document bibliographique sur le web. Elle est toujours très utilisée et propose des extensions. BIBO gère l'aspect physique des items. CIDOC CRM est un modèle assez générique fondé sur la modélisation événementielle mais est spécialisé dans la description des objets patrimoniaux. FRBR est quant à lui particulièrement bien adapté aux bibliothèques. Il fournit une représentation simplifiée et schématique de l'activité de catalogage en s'appuyant sur trois groupes d'entités : les œuvres, les agents et les sujets. Les œuvres sont décomposées en différents niveaux : œuvre (création intellectuelle), expression (édition), manifestation (publication), et item (exemplaire matériel). Cependant on lui a souvent reproché de ne pas être bien adapté aux publications en série comme l'ont démontré Patrick Le Bœuf et François-Xavier Pelegrin [44] avec l'exemple du Petit Phare. Une harmonisation des modèles CIDOC et FRBR a été faite avec $FRBR_{OO}$ qui a lui-même débouché sur $PRESS_{OO}$, modèle mieux adapté aux périodiques avec la classe règle éditoriale Z12 où chaque paramètre (fréquence, formatage, titre…) conduit à une instanciation de la règle, ce qui permet de ne pas répéter toutes les caractéristiques d'une publication en série à chaque fois que l'une seule change. Pour terminer citons BibFrame qui se rapproche de FRBR mais ne modélise que trois classes d'objets : œuvre, instance et item. Il existe d'autres ontologies utiles. FOAF est une ontologie modélisant les personnes et les relations qu'elles entretiennent entre elles. Une extension a été proposée pour les réseaux sociaux [45]. SIOC permet aussi de décrire les objets utilisés dans les réseaux sociaux ainsi que leur interaction.

4.2 Proposition de modélisation

Nous souhaitons proposer une modélisation du domaine médiatique qui repose sur un aspect diachronique et qui permette une grande exhaustivité de description des entités. Les concepts principaux, que l'on retrouve aussi sous le nom de classes, correspondent aux éléments de la chaine de production médiatique : Producteur de contenu, Editeur, Participant, Maison mère et Propriétaire (Figure 6). Nous nous sommes inspirés de BibFrame et FRBR, en distinguant le Produit médiatique (œuvre) de ses Publications médiatiques (instance), ce qui permet d'identifier et de relier les publications sur différents supports (presse papier, web,

réseaux sociaux…) à un même produit. Afin de rendre la base un peu plus accessible et interopérable, chacune des classes est typée à l'aide du schéma (RDF>type) et de classes (FOAF :organization…) déjà existantes sur le Web. Les instances font aussi référence à des URIs (identifiant Wikidata, FOAF…) pour identifier les ressources sur le Web.

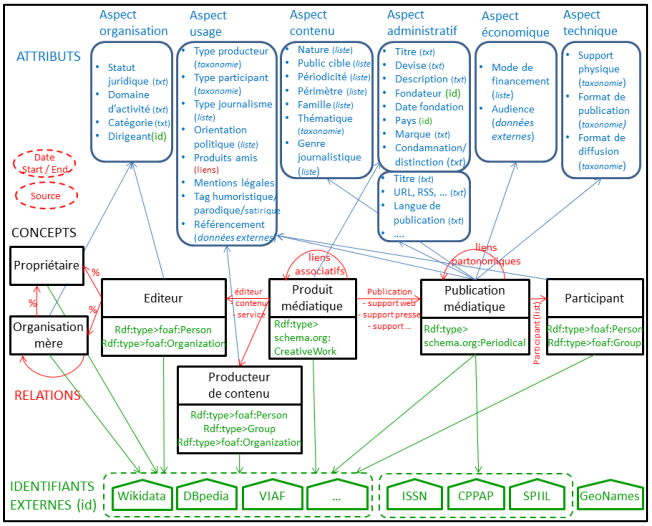

Figure 6 : Schéma général du modèle conceptuel

Pour être le plus exhaustif possible, nous avons introduit des relations partonomiques «tout-parties » qui permettent notamment de décrire le blog d'un site de presse en ligne, ce qui permet dans cet exemple de distinguer les participations des journalistes de celles des internautes. Les relations associatives servent quant à elles à associer des produits (supplément de…) ou à décrire des restructurations de médias (fusion…) tandis que les relations hiérarchiques « est un » peuvent décrire des liens d'appartenance (hiérarchie). Remarquons que les relations peuvent aussi être spécialisées en sous-relations, ce qui permet de distinguer un éditeur de service d'un éditeur de contenu, les publications sur différents supports ou encore les différents rôles de participant. Plus précisément, détaillons la relation Participant. Cette relation se développe en sous-relations pour définir les différents rôles possibles que la personne opère dans la publication (rôle chef de rédaction, rôle journaliste, rôle rédacteur, …). Elle pointe vers la classe Participant qui peut être de type personne ou groupe. Une instance de Personne pourra posséder des attributs relatifs à cette entité nommée (date de naissance, activité…) et des URIs (identifiants Wikidata, VIAF…) pour l'identifier sur le Web. Remarquons que celle-ci sera autonome, c'est-à-dire que ses valeurs seront indépendantes du contexte lié à la publication. Il sera alors possible de récupérer d'autres propriétés la concernant sur le Web pour enrichir la base. Deux publications pourront ainsi pointer vers la même personne si elle intervient dans les deux publications même avec un rôle différent. Notons enfin que les propriétés peuvent aussi porter des attributs, comme un pourcentage d'appartenance en ce qui concerne la relation Propriété.

La complétude des attributs contribue aussi à une plus grande exhaustivité de description. Les attributs ont pour valeur un texte libre ou une catégorie appartenant à une typologie (liste de vocabulaire contrôlé ou taxonomie). Les typologies reposent sur le principe d'exclusivité des différentes catégories proposées. En revanche, suivant leur nature, un attribut pourra être attribué à une catégorie unique ou plusieurs catégories de la typologie. Par exemple pour la nature des producteurs de contenu, il s'agit de distinguer les constituants des arènes du débat public (Figure 2) et de les organiser dans une taxonomie. Enfin les autres attributs servent à décrire les diverses caractéristiques que nous avons identifiées (technique, contenu, organisation, modèle économique, public, usage). Nous pouvons nous inspirer du modèle PRESS00 pour les règles éditoriales. Nous y ajoutons, entre autre, une typologie de types de journalisme et de nature de contenu. A titre d'exemple, les attributs concernant le modèle économique décrivent aussi bien les subventions extérieures reçues que les modes de paiement (abonnement, freemium…), le financement par crowdfunding, l'utilisation de publicité, la proposition de services et d'évènementiels, ou encore l'appel au don. Des métadonnées extérieures peuvent aussi enrichir le modèle en procurant des données d'audience (ACPM, Alexa). D'autre part, des indications d'orientations politiques, des référencements (CPPAP, SPIIL, syndicats…), tout comme le signalement de récompenses ou de condamnations, sont autant d'éléments précieux pour obtenir des critères de reconnaissance des médias Pour terminer, notons que pour rendre la base plus interopérable, il sera possible de faire appel, dès que possible, à des relations d'ontologies déjà existantes, comme par exemple celle du Dublin Core pour l'attribut Titre.

En ce qui concerne la modélisation du temps, il s'agit de poser un intervalle temporel sur chaque propriété. Pour accrocher une date, tout comme une source de provenance, il est par exemple techniquement possible de réifier les propriétés afin qu'elles portent certains de ces attributs.

4.3 Tests à l'aide d'une liste d'autorité

Une liste d'autorité constitué de 3000 sites web d'information, et 1000 titres de presse papier en relation, a été constituée en croisant des sources de données de différents types : CPPAP, BNF, SPIIL, Upreg, ACPM, encyclopédies et annuaires (Giga presse, Press directory, Giga Presse, Decodex,…), projets EMM ou Netguide.com, Alexa (agence de référencement des sites internet), articles de presse (liste de sites d'extrême droite, de presse alternative, de blogs…). Cette première liste a été complétée par des sites amis apparaissant sur la première page des sites web. Cette liste contient les titres, URL, et les descripteurs produits par les différentes sources de données.

Elle reflète des profils très divers à savoir des sites web d'organisme de presse ou audiovisuel, des médias pure player, des sites d'agrégation, des blogs personnels, ainsi que des sites d'information produits et édités par diverses organisations ou institutions (parti politique, ONG,…).

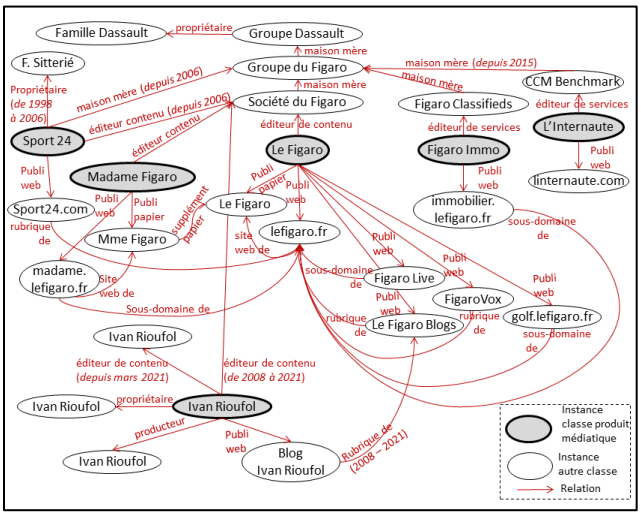

Figure 7 : Extraits Relations de la modélisation du *Figaro*

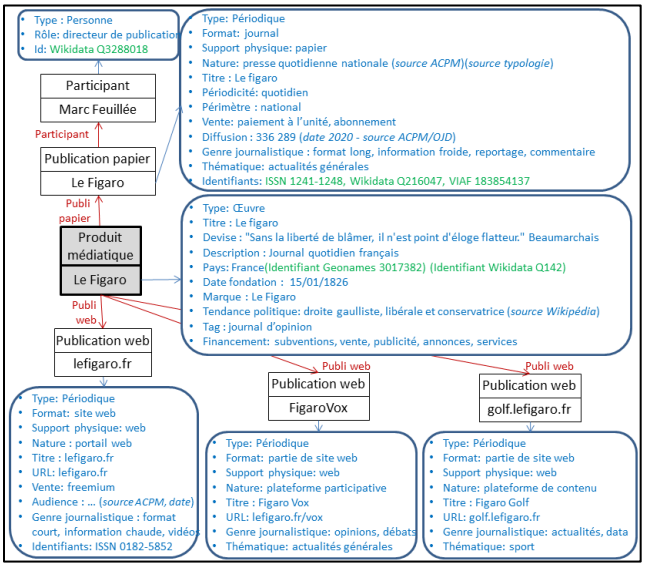

Figure 8 : Extraits Attributs de la modélisation du *Figaro*

Cette liste d'autorité, destinée à fournir des instances dans la future base de connaissance, a servi à deux types de test. Tout d'abord, une sélection d'une vingtaine de sites représentatifs de la liste a permis d'évaluer la structure générale et la souplesse du modèle conceptuel. A titre d'illustration, les figures 7 et 8 retracent un extrait de la modélisation du *Figaro* qui est un exemple emblématique de stratégie éditoriale [46]. Par manque de place, cette illustration a omis des classes, des relations et des attributs, mais atteste néanmoins de la nécessité d'avoir accès à une modélisation très fine. Nous pouvons voir la cohabitation de CCM Benchmark, spécialisé dans les services, au côté de la Société du Figaro spécialisé dans les médias. CCM Benchmark édite en effet *linternaute.com*, produit par des rédacteurs ne relevant pas de la convention collective des journalistes. Différents types de contributeurs interviennent dans le site, c'est pourquoi il est utile de pouvoir distinguer *Figarovox* qui est un forum (rédacteurs non professionnels encadrés par des professionnels), *Figaro blogs* (rédigés par des bloggeurs), *Figaro golf* qui est une plateforme mettant à la fois à disposition des articles de journalistes et des données extérieurs (résultat), *Figaro Live* (type de journalisme particulier). Enfin, nous pouvons modéliser le *blog d'Ivan Rioufol* qui a été édité sur le site du *Figaro* avant d'être édité par l'auteur en personne depuis mars 2021. L'historique du blog permet ainsi d'enrichir le contexte autour du blog.

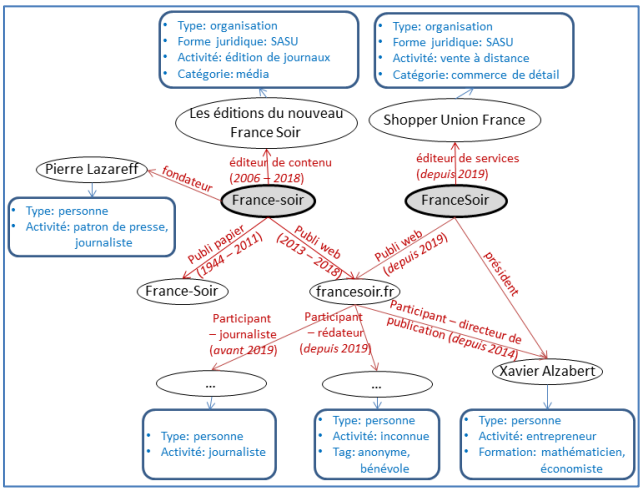

Figure 9 : Extraits de la modélisation de *francesoir.fr*

A côté de l'exemple du *Figaro* qui illustre le besoin de disposer d'une modélisation la plus précise possible, nous pouvons citer *francesoir.fr* (figure 9) qui démontre l'importance de la modélisation diachronique. La publication *francesoir.fr* est en effet reliée à deux produits médiatiques différents, *France-Soir* (2013-2019) et *FranceSoir* (depuis 2019). Plus précisément, ce site est dirigé depuis 2019 par l'entrepreneur Xavier Azalbert qui a repris la marque et le site web du journal quotidien *France-Soir* dont le format papier a disparu. Il est édité par l'éditeur de service Shopper Union France. De nos jours, les contributeurs ne sont plus des journalistes mais des bénévoles, le plus souvent anonymes. Remarquons que les deux modélisations du *figaro.fr* et de *francesoir.fr* permettent ainsi de bien différencier les contributions des journalistes professionnels des autres contributions (rédacteur, blogueur, journalisme participatif). Citons aussi le site *Presse-citron* qui est passé de la nature de blog personnel rédigé et édité par son fondateur Eric Dupin à la nature de site web avec une équipe de rédaction en 2011, et qui a été racheté en 2018 par

Keleops AG. Les exemples sont nombreux pour démontrer que les médias sont des entités évolutives qui présentent de nombreux changements dans le temps affectant aussi bien les attributs que les relations.

France Info est un autre exemple emblématique qui démontre cette fois l'importance de la classe Producteur. En effet, plusieurs producteurs, à savoir France Télévisions, Radio France, France Médias Monde et l'INA, participent à la création de contenus. Dans d'autres cas, la classe Producteur permet d'identifier un mouvement politique ou un syndicat derrière un produit médiatique. La classe Participant, quant à elle, peut jouer un rôle dans l'analyse des relations entre médias. Par exemple, Youssef Hindi (essayiste et conférencier) fédère autour de lui plusieurs publications. Il est en effet contributeur du site *Egalité et réconciliation*, et intervient aussi sur *Algérie patriotique*, le *Cercle des volontaire*, *Arrêt sur Info* et *Culture populaire*, faisant ainsi apparaitre des liens supplémentaires entre des publications.

En second lieu, la liste a aussi été analysée à la main dans son ensemble pour aider à définir les typologies d'attributs. La définition des typologies est un travail long et itératif pour réussir à tenir compte conjointement de tous les profils de site. Les descripteurs des différentes sources ont été exploités dans ce travail. La liste a été particulièrement utile pour affiner les typologies de nature de producteur, de thématique et de format de publication. Les attributs (figure 6) ont été partiellement testés sur les deux-tiers de la liste : 1148 sites de presse, 243 sites de média audiovisuel, 310 médias pure player, 77 blogs personnels et 122 sites d'information divers. Le dernier tiers de la liste, formé d'un titre et d'un url, reste à décrire.

5 Conclusion

Les journalistes professionnels n'ont plus le monopole de l'information, l'éditeur de contenu se voit concurrencé par des éditeurs de service, et les supports de diffusion se sont multipliés. Afin de pouvoir modéliser ce nouveau paysage médiatique protéiforme, il faut tout d'abord définir son périmètre, puis observer son fonctionnement pour identifier les entités à décrire à l'aide d'une grille, et enfin proposer un modèle où les entités sont représentées dans des classes, leur lien entre elles dans des propriétés et leurs caractéristiques dans des attributs. La modélisation obtenue permet de réaliser une classification commune pour un ensemble de médias très différents. La particularité du modèle obtenu repose sur un très fin niveau de description et sur l'évolution diachronique des entités La démarche mise en place s'appuie sur trois spécificités : l'analyse très fine du domaine médiatique, l'utilisation de ressources disponibles sur le web (ontologie, URI) et le développement à la main de typologie à l'aide d'une liste d'autorité. Dans un futur travail, ce modèle devra être implémenté sous la forme d'une base de données avec des règles externes ou d'une base de connaissances structurée à partir d'une ontologie formelle. La recherche diachronique sera une question importante à évaluer.

REFERENCES

[1] Olivier Bomsel, 2017, Ecosystème de médias. Approche économique et institutionnelle des médias, *Séminaire PSL*, 06 décembre 2017.
[2] Alexis de Tocqueville, 1835, Democracy in America, New York.
[3] Marshall McLuhan, 1964, Understanding Media, The Extensions of Man, *McGraw-Hill Book Company*, New York, 360 p.
[4] Régis Debray, 2000, Introduction à la médiologie, PUF.
[5] Patrice Flichy, 2008, Technique, usage et représentations, *Réseaux*, Volume 148-149, Issue 2, pp 147-174.
[6] Rémy Rieffel, 2005, Que sont les médias?. Paris, Gallimard, *coll. Folio actuel*, 534 p.
[7] Bernard Miège, 2010, L'espace public contemporain. Approche Info - Communicationnelle, Presses universitaires de Grenoble, coll. *Communication medias societe* , 227 p., EAN : 9782706116179.
[8] Lafon Benoit, 2017, Médias sociaux : l'extension du domaine médiatique par l'industrialisation du relationnel, *Les Enjeux de l'Information et de la Communication*, n°18/3A, pp 53-64.
[9] Eric Sherer, 2011, A-t-on encore besoin des journalistes ? Manifeste pour un journalisme augmenté, PUF, p8
[10] Nathalie Sonnac, Jean Gabszewiez, 2013, Le Produit médiatique, L'industrie des médias à l'ère du numérique, p 7-20.
[11] Jean-Marie Charon, 2011, Les médias à l'ère du numérique, Les cahiers du journalisme N°22/23, Automne 2011, p 15-27.
[12] François Demers, 2008, La communication publique, un concept pour repositionner le journalisme contemporain, *Les cahiers du journalisme*. N°18, pp 208-230.
[13] Benoit Lafon, 2019, Médias et médiatisation. Analyser les médias imprimés, audiovisuels, numériques, *Communication en +*, Presses universitaires de Grenoble, 310p.
[14] Claire Warle, Hossein Derakhshan, 2017, Thinking about 'information disorder': formats of misinformation, disinformation, and mal-information, Council of Europe report DGI (2017)09
[15] Xinyi Zhou, Reza Zafarani, 2020, A survey of fake news: Fundamental theories, detection methods, and opportunities, ACM Computing Surveys, Volume 53, Issue 5, October 2020, Article No.: 109, pp 1–40, https://doi.org/10.1145/3395046
[16] Presse et numérique - L'invention d'un nouvel écosystème, *Rapport à Madame la Ministre de la culture et de la communication*.
[17] Axel Bruns, 2018, Gatewatching and News Curation: Journalism, Social Media, and the Public Sphere, *Digital Journalism*, 7(4), pp. 543–544
[18] Sadok Hammami, 2006, Eléments d'une méthodologie d'analyse de la presse électronique, Revue Tunisienne de communication, Vol 1, N°6.
[19] Franck Rebillard, Ghislaine Chartron, 2004, Quels modèles pour la publication sur le web? Le cas des contenus informationnels et culturels.. Jun 2004. sic_00000986
[20] Emmanuel Marty, Franck Rebillard, Nikos Smyrnaios et Annelise Touboul, 2010, Variété et distribution des sujets d'actualité sur Internet. Une analyse quantitative de l'information en ligne, *Mots. Les langages du politique* [En ligne], 93 | 2010, mis en ligne le 01 octobre 2012, consulté le 11 juin 2021. URL : http://journals.openedition.org/mots/19832 ; DOI : https://doi.org/10.4000/mots.19832
[21] Emmanuel Marty, Franck Rebillard, Stéphanie Pouchot, Thierry Lafouge, 2012, Diversité et concentration de l'information sur le web, Une analyse à grande échelle des sites d'actualité français, Réseaux 2012/6, N°176, pp 27-72.
[22] (Patrick Eveno, 2010, La presse, Paris: PUF, p 45
[23] Jeanne Bordeau, 2008, La Veille média et la revue de presse, Paris, Eyrolles.
[24] Roselyne Ringoot, Jean-Michel Utard, 2006, Genres journalistiques et « dispersion » du journalisme, *Le journalisme en invention*, Presse universitaire de Rennes, p21-47.
[25] Erik Neveu, 2001, Sociologie du journalisme, *La Découverte, coll. Repères,* Paris.
[26] José de Broucker 1995, Pratique de l'information et écritures journalistiques, Paris, CFP.
[27] Jean-Luc Martin-Lagarette 1994, Guide de l'écriture journalistique. Ecrire, informer, convaincre, Paris, Syros.
[28] Henri Montant 1994, Commentaires et humeurs, CFPJ, Paris
[29] Yves Agnès,2002, Manuel de journalisme. L'écrit et le numérique ; Paris, La Découverte

[30] Michaël E. Sinatra, Marcello Vitali-Rosati, 2014, Pratiques de l'édition numérique, Montréal, Les Presses de l'Université de Montréal, ISBN 978-2-7606-3203-5.
[31] Jean-Marie Charon, 2016, La presse d'information multisupport, . Imaginer, concevroir, expérimenter, créer, Uppr.
[32] Olivier Bomsel, 2007, Gratuit ! Du déploiement de l'économie numérique, *Folio Actuel*, Gallimard.
[33] Nathalie Sonnac, 2006, Les médias : une industrie à part entière et entièrement à part, *Questions de communication*, 9 | 2006, p455-473, https://doi.org/10.4000/questionsdecommunication.7947
[34] Gérôme Guibert, Franck Rebillard, Fabrice Rochelandet, 2016, Médias, culture et numérique. Approches socioéconomiques, Paris, Armand Colin, coll. « Cursus »,, 237 p., ISBN : 978-2-200-61454-6.
[35] Nathalie Sonnac, 2009, L'économie de la presse: vers un nouveau modèle d'affaires, *Les cahiers du journalisme* N°20, automne 2009.
[36] Nathalie Sonnac,, 2013, L'écosystème des médias. *Communication*. DOI: 10.4000/communication.5030
[37] Yves Citton, 2014, L'économie de l'attention: nouvel horizon du capitalisme ?, *La découverte*, 287p.
[38] Dominique Cardon, 2015, A quoi rêvent les algorithmes : nos vies à l'heure des big data, Paris, Seuil, dl 2015, 105 p. (ISBN 978-2-02-127996-2)
[39] Tim Wu, 2016, The Attention Merchants: The Epic Scramble to Get Inside Our Heads, New York, NY: Alfred A. Knop
[40] Armande Saint-Jean, 2002, Éthique de l'information: fondements et pratiques au Québec depuis 1960, Presses de l'Université de Montréal.
[41] Stephan Gaël et Vauchez Ysé, 2019,, « Réinformation » Publictionnaire. Dictionnaire encyclopédique et critique des publics. Mis en ligne le 14 novembre 2019. Accès : http://publictionnaire.huma-num.fr/notice/reinformation ,
[42] Julia Cagé, Marie-Luce Viaud, Nicolas Hervé, 2017, L'information à tout prix, INA Editions.
[43] Emilio Sanfilippo, 2021, Ontologies for information entities: State of the art and open challenges, Applied Ontology, vol. 16, no. 2, pp. 111-135. DOI: 10.3233/AO-210246
[44] Patrick Le Bœuf, 2015, A basic introduction to FRBRoo and PRESSoo, library.ifla.org
[45] Asmae El Kassiri, Fatima-Zahra Belouadha, 2017, A FOAF ontology extension to meet online social networks presentation and analysis, IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), DOI:10.1109/ICPCSI.2017.8392287
[46] Alexandre Joux, 2017, Stratégies de marques et stratégies éditoriales du Groupe Figaro, Du portail à l'écosystème intégré par le Data Management, *Réseaux* 2017/5 (n°205), pp 117-143.