# Illustration des apports et limites de l'usage des sources ouvertes à travers le cas de la Russie

*Marie-Gabrielle Bertran*[1]

Les recherches qui s'appuient sur des méthodes d'Osint (*open source intelligence*), « renseignement en sources ouvertes[2] » [Verdi, 2021], peuvent représenter un apport considérable dans le domaine de la recherche en sciences sociales, notamment en géopolitique. Elles s'appuient aujourd'hui majoritairement sur des sources ouvertes disponibles en ligne[3] [Ronzaud et Ruan, 2022], c'est-à-dire sur les documents textuels, iconographiques et audiovisuels qui peuvent être consultés ou récupérés sur le Web [OTAN, 2002b]. Les recherches en Osint peuvent donc être réalisées à distance du terrain de recherche. Elles permettent de collecter des informations qui, pour la plupart, n'auraient pas pu être obtenues par d'autres choix méthodologiques, comme la conduite d'entretiens ethnographiques ou l'étude des archives.

L'usage des sources d'information découvertes lors de la réalisation d'un terrain numérique dans l'espace cyber ou la datasphère [Douzet et Desforges, 2018] tombe cependant sous le coup de plusieurs limites, qui peuvent être d'ordre

1. Doctorante en géopolitique, université Paris 8, ED 401-Sciences sociales, EA 353-IFG-Lab, chercheure au centre de recherche Géopolitique de la datasphère (Geode).

2. En français on parle de Roso (Renseignement d'origine sources ouvertes) [Moutouh et Poirot, 2018, p. 677-681]. L'expression et l'acronyme français semblent rester cantonnés aux milieux militaire et du renseignement, tandis que le domaine civil et les enquêteurs indépendants emploient plus couramment l'acronyme anglais.

3. Nous considérerons ici les méthodes d'Osint qui reposent sur le numérique, bien que cette pratique ait largement précédé l'émergence de l'Internet, notamment dans le domaine du renseignement [OTAN, 2002a, p. 9].

éthique, légal, méthodologique ou scientifique. D'abord, la récupération et l'étude de sources qui ont été obtenues à l'insu d'un groupe d'acteurs ou d'un individu provoquent un déséquilibre des connaissances entre l'objet (ou le sujet) d'une étude et celui qui l'étudie. Elles supposent que le chercheur en sait plus sur ces acteurs que ce qu'ils savent qu'il sait d'eux. Cet avantage cognitif du chercheur par rapport aux acteurs qu'il étudie doit faire l'objet d'une réflexion éthique [Janner-Raimondi, 2015]. Ensuite, les sources d'information obtenues sur Internet peuvent avoir été mises en ligne suite à la réalisation d'un acte délictueux, comme le piratage. Dans ce cas, la récupération et la conservation de ces données peuvent tomber sous le coup de la loi, et être considérées comme une forme de recel[4]. Par ailleurs, lorsque la diffusion de ces sources dites grises[5] est réalisée par d'autres acteurs que ceux qu'elles concernent directement ou à qui elles appartiennent, elle nécessite de prendre en compte leurs motivations.

C'est en particulier le cas dans le cadre des conflits géopolitiques et militaires, où les acteurs ont tendance à diffuser de la donnée pour propager et soutenir leurs narratifs, influencer l'opinion et favoriser leurs actions. Ces pratiques historiques du conflit et de la guerre se voient renouvelées aujourd'hui par la diffusion massive de ces données en ligne, sous format numérique. Depuis février 2022 et le début de la guerre qui oppose l'Ukraine et la Russie, le réseau Internet semble ainsi marqué par une augmentation sans précédent des fuites de données[6]. Elles proviennent en grande partie des cyberattaques qui ont été menées contre des institutions publiques et privées russes par différents groupes de pirates informatiques, dont Anonymous[7]. Surtout, l'usage de ces fuites comme outils géopolitiques et stratégiques (voire militaires) de déstabilisation d'un pays (la Russie) semble n'avoir jamais eu d'équivalent auparavant[8]. On peut d'ailleurs parler ici de véritables opérations de *hack and leak*, c'est-à-dire de piratages suivis de la publication intentionnelle des données dérobées, dans le but de déstabiliser un concurrent, un

4. « Piratage de compte, que faire ? », cybermalveillance.gouv, 15 janvier 2020, <https://www.cybermalveillance.gouv.fr/tous-nos-contenus/fiches-reflexes/piratage-de-compte?msclkid=c2373cc4ae7a11ec8cc034ae4f135a3f#loi-piratage-compte>.

5. Données sensibles et difficiles d'accès, voir *infra*.

6. Le site DDoS Secret (<https://www.ddosecrets.com/wiki/Distributed_Denial_of_Secrets>) en répertorie une partie disponible au lien suivant :
<https://www.ddosecrets.com/wiki/Special:Categories>.

7. Anonymous a notamment publié des indicateurs de ses attaques en temps réel, sur un site qui répertoriait les noms de domaines qu'il ciblait en Russie, et a diffusé les données dérobées lors de ces attaques sur différentes plateformes, dont Twitter et DDoS Secret.

8. Au vu de l'ampleur du phénomène dans le cadre d'un conflit militaire qui a lieu à une époque où l'usage de l'Internet et des réseaux sociaux se généralise.

adversaire ou un ennemi[9]. Si la quantité massive des données qui sont publiées dans ce cadre est difficile à analyser, elles peuvent néanmoins permettre aux chercheurs d'accéder aux processus de décision, au fonctionnement et aux perceptions internes de certaines institutions russes publiques et privées, telles que la branche régionale de l'agence de régulation de l'Internet russe RosKomNadzor au Bachkortostan ou la banque Sber[10].

Nous verrons donc d'abord que les informations utiles à la réalisation d'un travail de recherche en géopolitique sont généralement plus accessibles qu'il n'y paraît dans le cas de la Russie, grâce aux données disponibles en ligne, notamment issues de fuites de données. Nous verrons cependant que ces informations doivent toujours être considérées avec un regard critique et reproduites avec précaution dans un travail de recherche, en particulier lorsqu'elles proviennent de sources dites grises (dont les données divulguées font partie). La disponibilité en ligne de ces sources dont l'origine est peu ou mal connue, voire inconnue, n'est jamais neutre en effet, c'est-à-dire sans cause ou sans objectif, déclaré ou dissimulé.

## Le RuNet : une source d'information importante sur la Russie

La Russie est encore souvent perçue comme un pays imprégné de l'esprit de la guerre froide, auquel on a tendance à associer le régime du secret. La nécessité de garder des informations secrètes durant cette période était en effet sans cesse rappelée aux citoyens soviétiques, notamment au moyen d'affiches placardées dans les rues et de consignes à appliquer au sein des administrations [Rosenfeldt, 2009].

9. L'usage de la diffusion publique du renseignement depuis janvier 2022 par le gouvernement des États-Unis pour anticiper et contenir les actions de l'armée russe fait d'ailleurs partie de ce nouveau cadre, où la publication de données sensibles n'est plus seulement un instrument de communication et d'influence, mais également un outil *militaire* à proprement parler. « Guerre en Ukraine : quand les États-Unis dégainent l'arme du renseignement militaire », France 24.

10. <https://ddosecrets.com/wiki/Roskomnadzor> et <https://ddosecrets.com/wiki/Sberbank_of_Russia>. L'étude de ces données diffusées publiquement suite à des piratages présente néanmoins plusieurs risques et limites, dont ceux qui ont été évoqués *supra*, mais aussi le risque informatique et technique que les documents publiés aient été vérolés de manière à diffuser des programmes informatiques malveillants.

*Des données accessibles sur le RuNet malgré un resserrement du contrôle de la donnée en Russie*

Cette tendance semble réapparaître depuis les années 2010, à travers la mise en place de lois, mesures gouvernementales et décrets présidentiels destinés à accroître le contrôle des autorités sur la diffusion des données dans le pays et en dehors, en particulier *via* le réseau Internet[11]. Les dernières directives du Service fédéral de sécurité russe (FSB) sur la divulgation d'informations considérées comme sensibles ou stratégiques en témoignent [FSB, 2021]. Elles correspondent à une liste de soixante sujets, qui ont été catégorisés comme étant trop sensibles pour être abordés avec des étrangers ou des médias. Les citoyens russes qui s'en rendraient coupables peuvent non seulement être considérés comme des « agents de l'étranger[12] » par les autorités judiciaires, mais également être accusés de haute trahison et de divulgation de secrets d'État. L'arrestation en septembre 2021 du directeur de l'entreprise de sécurité informatique Group-IB, Ilja Satchkov, pour ces deux motifs est un signe du resserrement du contrôle de la donnée sur la cybersécurité en Russie par le biais de cette directive[13]. Cette tendance à l'accroissement systématique du contrôle de l'information par les autorités se fait d'ailleurs parfois au détriment de la sécurité du pays, qu'elle vise pourtant à assurer. Le décret présidentiel n° 98 du 2 mars 2018, « Sur les amendements à la liste des données [ou informations] classées comme secret d'État, approuvée par le décret présidentiel n° 1203 du 30/11/1995 » [Présidence de la Fédération de Russie, 2018], en est un bon exemple. Il a étendu la définition du secret d'État (*gosudarstvennaja tajna*) à l'ensemble des données qui portent sur les infrastructures informatiques russes considérées comme « critiques[14] », dans le but de les protéger en réduisant la connaissance que pourraient en avoir de potentiels attaquants. Sa mise en œuvre implique néanmoins une restriction du partage de l'information entre experts en cybersécurité, pourtant nécessaire à la lutte contre les nouveaux modes d'attaques.

À l'inverse, certaines données, souvent à caractère personnel, sont facilement accessibles sur l'Internet russophone (ou RuNet), alors qu'elles seraient

11. Il s'est d'ailleurs fortement accru en 2022 depuis l'annonce officielle de l'offensive militaire en Ukraine le 24 février, notamment dans le cas des médias (voir *infra*).

12. « Le FSB a approuvé une liste d'informations dont le transfert en dehors du territoire peut conduire à être considéré comme agent de l'étranger », *Kommersant*, 30 septembre 2021.

13. Thomas Brewster, « Russia arrests one of its biggest cyber stars on treason charges », *Forbes*, 29 septembre 2021.

14. Les données considérées comme les plus sensibles concernent le nombre d'attaques subies par les infrastructures informatiques des institutions publiques, et les mécanismes de protection employés par ces institutions contre ces attaques.

considérées comme particulièrement sensibles au sein de l'Union européenne suivant le Règlement général sur la protection des données (RGPD) du 14 avril 2016. Cette différence entre les systèmes de protection des données russe et européen semble liée à une différence fondamentale de perception de la sécurité des personnes (individus ou entreprises) qui semble avoir émergé dans les années 1990, et qui induit des différences de législation et de pratiques. En Russie, les années 1990 ont été caractérisées en effet par un démantèlement des infrastructures étatiques et administratives soviétiques, qui a rendu difficile, pour les nouveaux citoyens russes, d'accéder aux services publics et de protection juridique [Ledeneva, 2006].

Le début des années 2000 a donc été marqué par une tentative de reprise en main des services et des administrations publics par les autorités étatiques[15], et par un processus de normalisation de leur fonctionnement. Ce processus est notamment passé par la création de bases de données publiques centralisées sur Internet, pour permettre aux citoyens russes d'accéder à ces services et de faire valoir leurs droits. Ces bases de données ont donc été mises en ligne à une époque où la prise en compte des risques liés à l'accessibilité des données personnelles était moindre, car les citoyens russes étaient peu nombreux à posséder un ordinateur, encore largement réservé au secteur industriel et à l'administration. Parallèlement, le besoin d'obtenir des données personnelles et administratives pour pouvoir accéder aux aides publiques et juridiques était important, en raison des difficultés économiques et sociales induites par le changement brutal de structure économique du pays au début des années 1990 et l'adoption d'un système économique de marché fortement libéralisé. Aujourd'hui encore, les citoyens russes sont habitués à rechercher des informations personnelles ou légales dans ces bases de données pour vérifier la fiabilité d'une entreprise ou d'un produit, porter plainte, attaquer un individu ou une institution en justice, ou encore faire une demande de prêt ou de crédit. De nombreuses données publiques et privées sont donc accessibles en source ouverte sur le RuNet, notamment à partir des sites internet suivants : <http ://pravo.gov.ru> pour les documents officiels, mesures, décrets et lois. Les sites internet des institutions publiques, universités et centres de recherche, entreprises, etc., <http ://kremlin.ru>, <https ://culture.gov.ru>, <https ://www.msu.ru>, etc.

GosZakupki, le registre fédéral des contrats et marchés publics (<https ://zakupki.gov.ru>), les registres fédéraux et régionaux d'entreprises, qui sont généralement accessibles en ligne grâce à des sites qui agrègent ces registres dans leurs bases de

15. Qui a été au cœur des deux premiers mandats présidentiels de Vladimir Poutine entre 1999 et 2008.

données (sbis.ru, list-org.ru, TAdviser, etc.)[16]. Les sites des agences de presse et médias en ligne. Parmi les agences de presse publiques et privées les plus importantes en Russie, on peut citer Tass (<https://tass.ru>), RIA Novosti (<https://ria.ru>) et Interfax (<https://interfax.com>). Certains médias proposent un contenu généraliste, tels que le journal *Kommersant* (<https://www.kommersant.ru>) pour l'économie et les politiques intérieure et internationale. D'autres proposent un contenu orienté ou politiquement situé, tels que le journal *Komsomolskaja Pravda* (<https://www.kp.ru>), qui propose des lignes politique et éditoriale favorables aux politiques gouvernementales. À l'inverse, l'organe de presse indépendant Dozhd' (<https://tvrain.ru>) est connu pour avoir adopté une ligne politique et éditoriale critique, voire en opposition au gouvernement, tandis que le journal *Novaja Gazeta* (<https://novayagazeta.ru>) est réputé pour proposer des enquêtes journalistiques sans concession sur les milieux économiques et politiques. Ces deux médias ont d'ailleurs cessé d'émettre en Russie depuis 2022 et le début du conflit en Ukraine. Le site et l'application Internet de la chaîne Dozhd'ont été officiellement bloqués à partir du 1er mars 2022 à la demande du Parquet général russe[17], tandis que le journal *Novaja Gazeta* a déclaré le 28 mars suspendre ses activités en Russie jusqu'à la fin du conflit après avoir reçu plusieurs avertissements de la part des autorités[18]. Ces changements dans l'espace médiatique russe sont le signe le plus patent du resserrement du contrôle de la donnée en Russie, qui est plus que jamais un enjeu central pour les autorités dans le cadre du conflit en Ukraine. Ils vont de pair avec l'interdiction d'employer les mots « guerre » (*vojna*), « invasion » (*vtorzenie*), « offensive » ou « attaque » (*napadenie*) et « déclaration de guerre » (*objavlenie vojny*) dans toute publication ou discours au sujet du conflit. L'usage de l'expression « opération militaire spéciale » (*special'naja voennaja operacia*) est également préconisé par les autorités pour éviter que l'armée russe ne pâtisse de l'image de l'assaillant[19].

Les réseaux sociaux (VK. com, OK. ru, Twitter, Facebook...) et forums russes et russophones, notamment accessibles à partir du réseau Tor. L'interdiction de l'usage

16. Ces données sont généralement déclaratives. Elles doivent donc être considérées avec précaution, notamment dans le cas des bilans commerciaux et fiscaux des entreprises.

17. La chaîne semble néanmoins reprendre ses activités depuis la Géorgie où se sont expatriés plusieurs de ses journalistes. Voir l'article de Pjotr Sauer et Ruth Michaelson, « "It was game over" : Russian journalists flee to Istanbul after Putin's shutdown », *The Guardian*, 18 mars 2022.

18. Une nouvelle édition, *Novaja Gazeta Europe*, a été lancée le 20 avril 2022. Officiellement indépendante de l'édition originale russe, elle est éditée par une équipe de journalistes expatriés à Riga en Lettonie, <https://www.liberation.fr/economie/medias/suspendue-en-russie-la-novaia-gazeta-ecrit-une-nouvelle-page-en-lettonie20220418_UKLKSCCGBVDEZBEEMGO2PUWDDY/>.

19. Auprès des citoyens de la Fédération de Russie en particulier.

du réseau Tor est évoquée néanmoins depuis plusieurs années par les autorités en Russie dans le cadre du resserrement du contrôle de la donnée et des internautes. Le site du projet Tor, à partir duquel il est possible de télécharger le moteur de recherche qui permet d'accéder au réseau (Tor Browser) a été bloqué début décembre 2021[20]. Les sites qui proposent aux internautes de récupérer les informations personnelles d'individus à partir de sources ouvertes ou de bases de données fermées, pour un tarif donné ou en fonction d'un abonnement, tels que <https://rfpoisk.ru>, <https://checkperson.ru>, <https://namebook.club>, <http://spra.vkaru.net> pour les adresses et numéros de téléphones, ou <https://checklic.ru>.

Lorsqu'elles ne sont pas directement accessibles sur les portails des institutions publiques, ces données peuvent faire l'objet d'un commerce illégal. Certains fonctionnaires au salaire peu élevé mettent ainsi en vente des données ou des prestations de récupération de données personnelles, comme le montre le « dossier » publié par la fondation Open Russia[21] en juin 2020, qui présente la liste des prix proposés par des agents du FSB à des clients privés pour des prestations de ce type[22].

### *La problématique des « sources grises » à travers le cas du piratage de l'entreprise SyTech*

Les données obtenues par des méthodes de ce type sont dites grises. L'expression « données grises » ou « sources grises » est tirée du vocabulaire du renseignement économique, qui distingue l'information blanche, ouverte et accessible à tous, l'information noire, protégée et dont la diffusion est interdite ou très restreinte, et l'information grise, relativement sensible et difficile d'accès [Pelletier et Cuenot, 2013, p. 98]. L'information grise correspond donc à des données qui peuvent avoir été dérobées et revendues, mais aussi publiées par certains acteurs à l'origine de fuites de données.

Or les problèmes que posent le traitement et l'analyse des sources ou données grises sont particulièrement saillants dans le cas des fuites de données. La fuite des données de l'entreprise de cybersécurité russe SyTech en est un bon exemple. Ces données[23] ont été rendues publiques sur Twitter, après le piratage du système Jira de stockage et de gestion des projets de l'entreprise à partir de son serveur

20. Alexandre Horn, « Le site de Tor est bloqué en Russie, la lutte contre le navigateur anonyme continue », Numerama, 8 décembre 2021.

21. Créée par l'oligarque Mikhaïl Khodorkovski, une figure de l'opposition au gouvernement russe.

22. <https://fsb.dossier.center/>.

23. Un ensemble de vingt dossiers contenant cent dix-neuf documents.

principal Active Directory le 13 juillet 2019, par 0v1ru$, un groupe de pirates informatiques inconnu jusqu'alors[24]. D'après 0v1ru$, cette attaque et la diffusion de ces données visaient à dénoncer la réalisation de projets de surveillance des internautes par SyTech pour le compte du FSB. Pour augmenter la portée de son message, 0V1ru$ a confié une partie des 7,5 téraoctets de données dérobées[25] à Digital Revolution, un groupe de pirates informatiques connu pour avoir attaqué et publié les données d'un autre sous-traitant du FSB en décembre 2018, l'entreprise Centre de recherche Kvant, avec les mêmes objectifs. Digital Revolution a donc utilisé son compte Twitter le 22 juillet 2019 pour propager la nouvelle du piratage et partager les données de SyTech grâce à un lien vers le site de stockage Mega où il les avait placées. 0V1ru$ a également contacté *via* Twitter un ensemble de médias et de journalistes réputés pour leur indépendance ou leur ton critique vis-à-vis du gouvernement russe, tels que BBC News Russia, TJournal, Roman Dobrokhotov (journaliste du média The Insider) et des figures de la défense des libertés et droits des internautes, telles que l'association Roskomsvoboda, le Parti pirate russe et le directeur de l'entreprise Red Shield VPN. Avec ces prises de contact, le groupe avait visiblement pour objectif de médiatiser au maximum son piratage. Il a notamment utilisé les mentions de comptes Twitter précédées d'un @ (qui permettent de signaler à un autre compte et utilisateur que l'on a posté un message qui s'adresse à lui ou le mentionne) et des hashtags ou mots-dièses (précédés d'un #, qui permettent de forger une expression que les autres utilisateurs vont pouvoir reprendre et diffuser) pour augmenter l'impact de ses publications.

Cet objectif de diffusion la plus large possible est fréquent, si ce n'est systématique dans le cas des fuites de données[26]. Comme dans le cas de SyTech, ces données sont souvent publiées sur le mode de la révélation ou de la dénonciation. Elles peuvent ainsi servir à dénoncer ou à discréditer des individus ou des entreprises pour répondre à des objectifs politiques, militants (notamment dans le cas des lanceurs d'alerte), ou économiques et commerciaux [Gastineau et Vasset, 2017]. Dans le cas de SyTech, outre les projets de surveillance des internautes, Digital Revolution révélait, par exemple, que l'entreprise avait surfacturé des projets au FSB, qui les avait payés plus cher que ce qu'ils valaient d'un point de vue technique et du point de vue de leur apport en termes d'innovation. L'identité réelle de 0v1ru$ étant inconnue, rien ne pouvait donc permettre d'écarter la

24. Ou un individu déclarant être un groupe.

25. L'équivalent d'environ 850 à 1 750 vidéos de films.

26. Qui constituent une partie seulement des données grises que l'on peut trouver en ligne. Dans la majorité des cas, ces données sont disponibles suite à des négligences ou des erreurs de la part des administrateurs des sites et serveurs sur lesquels elles sont hébergées.

possibilité qu'il ait été employé par une entreprise concurrente pour discréditer SyTech et récupérer ses contrats avec le FSB.

L'étude des données qui sont mises en ligne dans le cadre de fuites de données amène ainsi à se poser la question des objectifs de ceux qui les publient, et à envisager la possibilité que leur publication même soit un instrument dans un projet d'influence. Si cette problématique se pose pour l'ensemble des données qui sont disponibles en source ouverte, elle concerne donc en particulier les données ou sources grises. Les données grises diffusées en sources ouvertes le sont en effet toujours de manière intentionnelle, puisque leur diffusion suppose qu'un acteur a décidé de les rendre publiques, alors que l'accès à ces données était restreint. La possibilité que l'acteur qui diffuse des données grises le fasse dans le but de manipuler ou d'influencer un public cible doit donc nécessairement être prise en compte par le chercheur qui s'en sert, quelles que soient la valeur et l'importance de leurs révélations, leur véracité, et l'authenticité des documents qui servent à les appuyer.

## Limites de l'usage des sources grises dans le travail de recherche : l'instrumentalisation (géo)politique des fuites de données, un risque à prendre en compte

Cette problématique se pose *a fortiori* pour la recherche en géopolitique, qui repose sur l'étude des conflits et rivalités de pouvoir entre des acteurs dont les perceptions se concurrencent ou s'opposent [Lacoste, 2016]. Ces acteurs peuvent avoir en effet pour objectif d'influencer ou de modifier les perceptions qui seraient contraires à la réalisation de leurs intérêts en diffusant des éléments qui confortent leur discours (on parle couramment de récits ou de « narratifs ») auprès des médias, mais aussi auprès des chercheurs qui les étudient. Dans leur ouvrage *Armes de déstabilisation massive*, les journalistes Pierre Gastineau et Philippe Vasset abordent précisément la question de l'usage des fuites de données comme instruments d'influence géopolitique [Gastineau et Vasset, 2017]. Ils décrivent notamment leur rencontre avec un ancien agent du renseignement technique israélien, à qui ils ont soumis l'hypothèse de l'existence d'opérations de publication sur Internet de données dérobées pour déstabiliser un pays en créant un scandale interne[27] :

> « [Q]uestionné sur l'utilité des fuites massives de mèls pour affaiblir un pays ennemi, [un ancien du renseignement technique israélien rencontré à Tel Aviv] refuse obstinément de considérer qu'il s'agit d'une tactique valide : "Le premier but d'un service de renseignement, c'est d'infiltrer ses ennemis et de préserver au maximum ses

27. C'est-à-dire d'opérations de *hack and leak* à buts géopolitiques.

> accès secrets. Organiser une fuite massive de documents serait, de ce point de vue, totalement contre-productif." Mais ces préventions n'ont plus cours dans un monde où l'information, comme les opinions publiques, est de plus en plus [mondialisée]. Et où l'impératif de discrétion a été remplacé par le déni plausible : les piratages massifs étant systématiquement menés par des hackers, les États peuvent toujours nier toute implication. »

Leurs remarques soulignent ainsi les difficultés qui se posent lors de l'étude de sources grises récupérées sur le Web après une fuite de données. Elles indiquent que le risque d'instrumentalisation géopolitique des données irrégulièrement divulguées est d'autant plus important que leur dissémination peut se faire anonymement sur Internet. Pour la recherche, cette problématique se traduit par la question de la réutilisation, du traitement et de la présentation des sources lorsque leur provenance, leur authenticité ou la cause de leur disponibilité en ligne en tant que sources ouvertes, c'est-à-dire les motifs derrière leur diffusion, sont incertaines.

### *Les cas SyTech et Sands : hacktivisme*[28] *ou instrumentalisation géopolitique ?*

Le cas du piratage et de la fuite des données de l'entreprise SyTech est intéressant de ce point de vue. Dans le cas de ce piratage, il était possible d'envisager en effet que 0v1ru$ ne soit pas un cybermilitant[29] opposé à la surveillance des internautes en Russie, contrairement à la représentation qu'il donnait de lui-même, mais un individu ou un groupe employé par une entreprise concurrente. Certains spécialistes de l'entreprise de cybersécurité russe Kaspersky Lab ont également envisagé la possibilité que 0v1ru$ ait été une *persona*[30] créée par un groupe de pirates informatiques iraniens sponsorisés par leur État (on parle d'APT, Advanced Persistent Threat), surnommé APT34, OilRig ou Helix Kitten. Les techniques et outils de piratage (*tactics, techniques, and procedures* – TTPs) employés par 0v1ru$, et sa méthode de diffusion des données sur les réseaux sociaux à l'aide d'un compte Twitter spécifiquement créé pour ce faire, pouvaient correspondre à leurs méthodes. Des cas similaires avaient en outre été observés les années précédentes. Parmi ces cas, on comptait notamment celui d'une cyberattaque contre un hôtel-casino du groupe Sands à Las Vegas en février 2014. Pour réaliser cette attaque, les exécutants avaient employé des outils préconçus et disponibles en

28. On parle aussi de cyberactivisme ou de cybermilitantisme.

29. On parle aussi de hacktiviste ou de cyberactiviste.

30. C'est-à-dire une personne fictive (groupe ou individu) créée de toutes pièces sur les réseaux sociaux.



*Hérodote*, n° 186, La Découverte, 3e trimestre 2022.

source ouverte (*open source*) [Le Deuff, 2021], comme dans le cas du piratage de SyTech. Ces outils comprenaient notamment des programmes utilisés pour tester la sécurité d'infrastructures informatiques à la demande de clients[31], tels que les logiciels Impacket et ProxyChains utilisés par 0v1ru$. Ces programmes leur avaient permis de réaliser une escalade de privilèges[32] sur une centaine de serveurs du groupe Sands, d'en exfiltrer les données, puis d'en nettoyer les disques : une série d'étapes également suivie par 0v1ru$ en 2019. Une fois l'attaque terminée, les attaquants ont également entrepris de défacer (*deface*)[33] le site internet de Sands, avant de publier les données dérobées sur les réseaux sociaux en affirmant être des hacktivistes groupés sous le nom de Anti WMD Team (Anti Weapons of Mass Destruction Team – « Groupe de lutte contre les armes de destruction massive »). L'image publiée sur le site web piraté et les messages du groupe sur Twitter dénonçaient les propos tenus par Sheldon Adelson, P.-D.G. des hôtels-casinos Sands qui avait critiqué quelques mois auparavant (en octobre 2013) l'hostilité de l'État iranien vis-à-vis de l'État d'Israël, et appelé les États-Unis à prendre des positions « plus fortes » en menaçant l'Iran de frappe nucléaire[34].

S'appuyant sur les revendications apparemment pacifistes du groupe de cybermilitants autoproclamés, l'entreprise Dell SecureWorks avait conclu à une attaque menée par des hacktivistes iraniens indépendants, c'est-à-dire non sponsorisés par

31. On parle de tests de pénétration (*penetration testing*, *PenTesting* ou *PenTest* en anglais). Ces missions sont souvent commandées par des entreprises ou des institutions publiques qui cherchent à éprouver la sécurité de leurs installations informatiques face à de potentielles tentatives de piratage. Elles peuvent être déléguées à des entreprises spécialisées, qui confient à leurs équipes la réalisation de ces missions sur contrat, ou à des équipes internes lors d'exercices qui sont parfois appelés *red teaming* (l'équipe qui joue le rôle des attaquants prend le statut de *red team*, équipe des rouges, face à la *blue team*, équipe des bleus, qui va devoir déjouer l'attaque et assurer la sécurité de ses installations).

32. Les systèmes informatiques (systèmes d'exploitation, logiciels et applications) qui contrôlent le fonctionnement des ressources physiques et virtuelles d'une machine (composants physiques, processus etc.) reposent généralement sur une structure hiérarchisée entre différentes instances (utilisateurs, groupes et administrateurs), qui disposent de droits spécifiques sur ces ressources. L'*escalade de privilèges* consiste à profiter des failles d'un système informatique pour obtenir les droits les plus élevés possibles sur les ressources d'une machine, de manière à pouvoir la contrôler.

33. Pratique qui consiste à modifier la première page d'un site en publiant un contenu qui indique qu'il a été piraté.

34. Il avait été jusqu'à « suggérer de faire exploser une bombe nucléaire dans le désert du Nevada pour avertir de ce qui pourrait arriver si Téhéran continuait [à développer] son programme nucléaire », *in* Russell Brandom, « Iran hacked the Sands Hotel earlier this year, causing over $40 millions in damage », The Verge, 11 décembre 2014.

leur gouvernement[35], un statut similaire à celui de l'acteur russophone 0v1ru$. Pourtant, le motif déclaré par les attaquants pour justifier leur attaque n'était en rien pacifiste. Il s'agissait en effet de punir un discours hostile à l'Iran sur un sujet de défense, le nucléaire. Le véritable motif de l'attaque était donc géopolitique, puisque les attaquants entraient ostensiblement en confrontation sur les terrains cyber, numérique et informationnel, avec un discours perçu comme hostile envers l'État iranien, voire dangereux pour l'Iran. Malgré les conclusions de Dell SecureWorks, les équipes de recherche de l'entreprise Kaspersky Lab ont présenté une autre analyse. Selon elles, une vue d'ensemble de la procédure suivie par les attaquants et des éléments de leur infrastructure (notamment des serveurs de Commande et de Contrôle, C2) concordaient avec ceux utilisés dans des attaques menées par la suite au Moyen-Orient, et indiquaient que les exécutants appartenaient bien à une APT sponsorisée par l'État iranien, probablement APT34. Le motif, ou la représentation, du hacktivisme a donc été utilisé par APT34 comme une couverture pour une action motivée par des intérêts étatiques, ici géopolitiques.

Suivant ce constat, il apparaît qu'une étude naïve des données obtenues par le biais de méthodes d'Osint, qui ne procède pas par leur évaluation et la vérification préalable de leur source, fait courir au chercheur le risque d'être instrumentalisé. En reproduisant et en publiant ces données dans le cadre de leur étude, il entre en effet dans un système de diffusion de ces données qui peut répondre à des objectifs politiques, géopolitiques ou économiques, au sein d'un projet d'influence[36] [Gastineau et Vasset, 2017]. Dans ce cas, le chercheur reproduit et diffuse vers un nouveau public des données qu'il présente comme des éléments d'information. Informations qui, comme l'étymologie du mot l'indique[37], vont in-former (la pensée de) celui qui les reçoit ou les perçoit comme telles, c'est-à-dire influencer son opinion, voire influer sur (le cours de) ses actions. Le chercheur se retrouve alors dans une position similaire à celle du journaliste d'investigation, en cela qu'il doit nécessairement se distancier des données qu'il étudie et présente comme de potentiels éléments d'information, et avertir ses lecteurs des incertitudes qui entourent leur origine.

35. Sean Gallagher, « Iranian hackers used Visual Basic malware to wipe Vegas casino's network », ArsTechnica, 12 décembre 2014.

36. Qui viserait, par exemple, à discréditer un adversaire politique lors d'une campagne électorale, une entreprise concurrente dans le cadre d'un appel d'offres pour un marché public, un diplomate dans le cadre de ses fonctions, ou encore un gouvernement, un État ou une population dans le cadre d'un conflit.

37. Le terme *information* provient du mot latin *informatio*, qui signifie d'abord dessin, esquisse, puis idée, conception, représentation d'une idée par l'image d'un mot. Il est lié au verbe latin *informare*, qui signifie *donner (une) forme à*, façonner, disposer, organiser, ou se former une idée de, former dans l'esprit, se représenter par la pensée.

Malgré ces limites, les méthodes de recherche qui s'appuient sur l'Osint ne sont pas exemptes d'intérêt scientifique. Leur mise en œuvre est justifiable d'un point de vue méthodologique, grâce notamment à la possibilité de leur reproduction, qui répond au critère scientifique de la reproductibilité. La recherche de données disponibles en ligne présente en effet cet avantage majeur qu'elle peut être retracée, et ainsi reproduite de bout en bout. En ce sens, elle peut être considérée comme fiable (suivant le critère méthodologique de la fiabilité, *reliability*) [Long et Johnson, 2000][38]. Elle rend possible, par ailleurs, la vérification des données par le biais des pratiques de triangulation [Long et Johnson, 2000, p. 34]. La triangulation consiste en la comparaison d'éléments d'information qui proviennent d'au moins trois sources différentes. Malgré ses limites[39] [Golafshani, 2003, p. 603], cette méthode comparative peut permettre de déceler des incohérences entre des données obtenues par le biais de différentes sources (par exemple, entre un entretien de recherche, un article de journal et un document officiel), et ainsi de valider ou d'invalider ces données. Ainsi, s'il est rarement possible d'affirmer ou de confirmer la fiabilité (*reliability*) des données obtenues dans le cadre d'une recherche qualitative, leur validité (*validity*) – ou leur invalidité – peut parfois être vérifiée. L'évaluation de la validité de ces données par des « méthodes alternatives de vérification et d'assurance », telles que la triangulation, peut donc permettre de considérer comme fiables dans une certaine mesure les analyses développées par le chercheur [Long et Johnson, 2000, p. 35].

## Conclusion

Avec la numérisation croissante des échanges, des pratiques et processus professionnels et de la gestion des tâches personnelles au quotidien, les sources ouvertes constituent aujourd'hui une ressource incontournable pour la recherche en géopolitique, et plus généralement en sciences humaines et sociales. Elles permettent au chercheur d'appréhender son terrain sous un angle complémentaire, en lui offrant accès à des éléments qui lui auraient sans doute échappé lors de la réalisation d'un terrain physique traditionnel, notamment à travers ses entretiens,

38. L'usage du critère de *fiabilité* dans les sciences qualitatives a été souvent discuté néanmoins [Golafshani, 2003, p. 601 ; Stenbacka, 2001, p. 552].

39. Tony Long et Martin Johnson expliquent que la vérification de données provenant de différentes sources, par la seule méthode de la triangulation, ne permet pas d'affirmer avec certitude que certaines données sont *vraies* (*true*) tandis que d'autres sont *fausses* (*false*). Pour cela, il faut d'abord vérifier qu'elles peuvent être utilisées pour répondre à la même « question », qu'elles portent sur le même sujet, et ne décrivent pas un « phénomène différent » [Long et Johnson, 2000, p. 35].

ses observations participantes ou ses recherches dans des archives. Elles lui permettent également de pallier l'absence de terrain de recherche non numérique en cas de difficultés d'accès à sa zone de recherche, notamment lorsque cette zone ou le sujet étudié présentent des risques[40].

Leur usage dans des travaux de recherche doit cependant prendre en compte certaines limites et respecter certaines précautions. Comme pour tout travail de recherche destiné à des productions scientifiques, ces sources d'informations doivent être analysées sous un angle critique, avec une prise de distance lors de leur (re)production ou de leur citation. Cette prise de distance est d'autant plus importante lorsque les sources étudiées proviennent de fuites de données, c'est-à-dire lorsqu'elles correspondent à un contenu diffusé par des acteurs intéressés. En ce sens, il est nécessaire de considérer la possibilité d'une instrumentalisation politique, géopolitique, diplomatique ou encore économique des sources qui sont mises en ligne et diffusées dans le cyberespace. Les raisons de leur présence sur le Web en tant que sources ouvertes doivent toujours être interrogées à l'aune d'une problématique éminemment géopolitique : celle de l'intention des acteurs qui les diffusent suivant une certaine perception de leurs intérêts.

## Bibliographie


BOUMAZA M. et CAMPANA A. (2007), « Enquêter en milieu "difficile" : introduction », *Revue française de science politique*, vol. 57, n° 1.

DOUZET F. et DESFORGES A. (2018), « Du cyberespace à la datasphère. Le nouveau front pionnier de la géographie », *Netcom*, vol. 32, n°1/2, p. 87-108.

FSB, SERVICE FÉDÉRAL DE SÉCURITÉ DE LA FÉDÉRATION DE RUSSIE – FEDERAL′NAJA SLUŽBA BEZOPASNOTI (2021), « Ordonnance du Service fédéral de sécurité de la Fédération de Russie n° 379 du 28 septembre 2021 "Sur l'approbation de la liste des informations dans le domaine des activités militaires et militaro-techniques de la Fédération de Russie, qui, dès réception par un État étranger, ses organes étatiques, une organisation internationale ou étrangère, des citoyens étrangers ou des apatrides, peuvent être utilisées contre la sécurité de la Fédération de Russie" (Enregistrée le 30 septembre 2021 sous le n° 65202) », URL : <http://publication.pravo.gov.ru/Document/View/0001202109300048?index=0&rangeSize=1>.

GASTINEAU P. et VASSET P. (2017), *Armes de déstabilisation massive*, Paris, Fayard.

GOLAFSHANI N. (2003), « Understanding reliability and validity in qualitative research », *The Qualitative Report*, vol. 8, n° 4, p. 597-606.


40. Néanmoins, les notions de *risque* et de *terrain difficile* peuvent être débattues [Boumaza et Campana, 2007, p. 8-12].

JANNER-RAIMONDI M. (2015), « Questions d'éthique et entretien : une approche philosophique », *in* BEDOIN D. et SCELLES R. (dir.), *S'exprimer et se faire comprendre*, Toulouse, Érès, p. 27-47.

LACOSTE Y. (2016), *La géographie, ça sert, d'abord, à faire la guerre*, Paris, La Découverte, « Poche ».

LEDENEVA A. (2006), *How Russia Really Works. The Informal Practices That Shaped Post-Soviet Politics and Business*, Ithaque/Londres, Cornell University Press.

LE DEUFF O. (2021), « L'Open Source Intelligence (OSINT) : origine, définitions et portée, entre convergence professionnelle et accessibilité à l'information », *I2D – Information, données & documents*, n°. 1, p. 14-20.

LONG T. et JOHNSON M. (2000), « Rigour, reliability and validity in qualitative research », *Clinical Effectiveness in Nursing*, vol. 4, n° 1, p. 30-37.

MOUTOUH H. et POIROT J. (2018), *Dictionnaire du renseignement*, Paris, Perrin.

ORGANISATION DU TRAITÉ DE L'ATLANTIQUE NORD (2002a), « NATO Open Source Intelligence reader », *NATO OSINT Reader*.

— (2002b), « Intelligence exploitation of the Internet », *NATO OSINT Handbook*.

PELLETIER A. et CUENOT P. (2013), *Intelligence économique, mode d'emploi. Maîtrisez l'information stratégique*, Paris, Pearson France.

PRÉSIDENCE DE LA FÉDÉRATION DE RUSSIE (2018), « Décret présidentiel (*Ukaz*) n° 98 "Sur les amendements à la liste des données [ou informations] classées comme secret d'État, approuvée par le décret présidentiel n° 1203 du 30/11/1995" », disponible sur <http://publication.pravo.gov.ru/Document/View/0001201803020009?index=0&rangeSize=1>.

RONZAUD L. et RUAN L. (2022), « World Wild Web : une typologie non exhaustive des méthodes d'enquête et d'analyse des campagnes d'influence sur les réseaux sociaux », *Hérodote*, n° 186.

ROSENFELDT N. E. (2009), *The « Special » World. Stalin's Power Apparatus and the Soviet System's Secret Structures of Communication*, vol. 1, Copenhague, Museum Tusculanum Press/université de Copenhague.

STENBACKA C. (2001), « Qualitative research requires quality concepts of its own », *Management Decision*, vol. 39, n° 7.

VERDI U. (2021), « Cycle et vocabulaire de l'Open Source Intelligence (OSINT) : le cas des services de renseignement », *I2D – Information, données & documents*, n° 1, p. 21-24.