

Cahiers français

#Politiques publiques

La politique maritime
de la France

#En débat

Faut-il avoir peur de la
dette publique ?

#Le point sur

La reconnaissance
de la langue des signes

Le règne des données



#Sommaire

05

#Politiques publiques

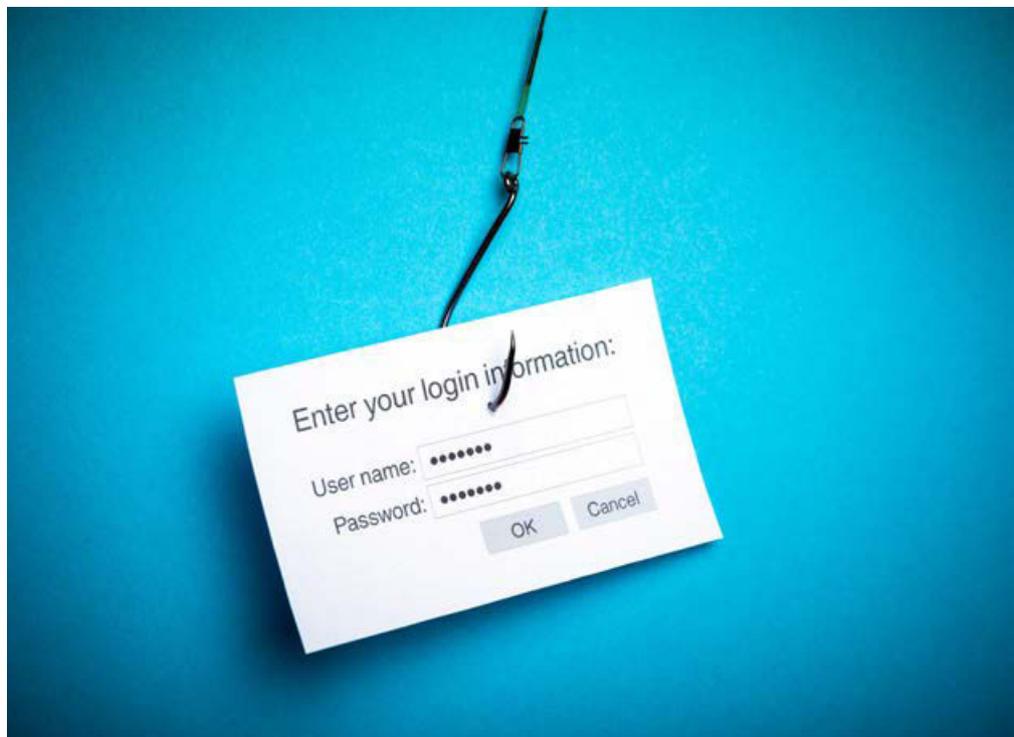
La nouvelle politique maritime de la France

Nicolas Boillet



© DAMEDIAS-STOCK.ADOBE.COM

© BLOOMICON-STOCK.ADOBE.COM



15 #Dossier

Le règne des données

16/ La donnée numérique : un bien économique comme les autres ?

Edmond Baranes

24/ Le RGPD : qu'a-t-il vraiment changé pour l'internaute ?

Jean-Luc Sauron

36/ Les données publiques : qu'en est-il de leur ouverture et de leur réutilisation ?

Mathilde Hoang

46/ Le capitalisme de plateforme

Henri Isaac

56/ La politique de la concurrence face au défi des entreprises du numérique

Marc Bourreau, Guillaume Thébaudin

68/ Les mégadonnées et l'essor de l'intelligence artificielle

Stephan Cléménçon

78/ Données et intelligence artificielle : quels enjeux pour la cybersécurité des États ?

Amandine Lévêque

90

#Les plus de la rédaction

90 / Ce qu'il faut retenir

91 / Les chiffres clés

92 / Les mots du dossier

93 / Le dossier en dessins

94 / Les dates clés

95 / Pour en savoir plus

97 #En débat

Faut-il avoir peur de la dette publique ?

Nicolas Dufrêne, Guido Hülsmann



© ONODJI-STOCK.ADOBE.COM

107 #Le Point sur

De la réforme de l'État à la transformation de l'action publique

Vie-publique.fr



© HAMILTON/REA



JENNIFER LESCOUET/CC BY-SA 4.0

113 #C'était en... 2005

La reconnaissance de la langue des signes comme langue à part entière

Olivier Schetrit



→ Retrouvez l'univers Cahiers français sur
www.vie-publique.fr/cahiers-francais
→ Les fiches au format mobile





Le règne des données

16

La donnée numérique :
un bien économique
comme les autres ?

24

Le RGPD : qu'a-t-il
vraiment changé
pour l'internaute ?

36

Les données publiques :
qu'en est-il de leur ouverture
et de leur réutilisation ?

46

Le capitalisme
de plateforme

56

La politique de la concurrence
face au défi des entreprises
du numérique

68

Les mégadonnées
et l'essor de l'intelligence
artificielle

78

Données et intelligence
artificielle : quels enjeux
pour la cybersécurité
des États ?

90

Les plus de la rédaction

- Ce qu'il faut retenir
- Les chiffres clés
- Les mots du dossier
- Le dossier en dessins
- Les dates clés
- Pour en savoir plus

La donnée numérique : un bien économique comme les autres ?

Edmond Baranes

Professeur d'économie à l'université de Montpellier

La donnée occupe une place centrale dans nos sociétés. Elle est un outil stratégique pour les entreprises et une opportunité pour la croissance économique, mais elle pose également des questions en matière de respect de la vie privée. Est-elle un bien économique comme les autres ?

La prolifération des données apparaît comme un révélateur de la transformation numérique de nos sociétés. Celle-ci a véritablement commencé, il y a près de quinze ans, avec l'apparition des équipements connectés (téléphones intelligents, tablettes) et des premières applications. Aujourd'hui, l'avènement des réseaux sociaux, le développement de l'internet des objets (objets reliés par des capteurs interconnectés) et les progrès de l'intelligence artificielle (IA) renforcent ce phénomène. La collecte massive et le traitement des données entraînent de profonds changements au niveau économique, mais aussi social, culturel et politique. Cette évolution est souvent qualifiée de « révolution » de la donnée ou de nouvelle révolution industrielle.

La donnée est partout et ses sources sont multiples. Elle est incontestablement

devenue une ressource importante pour l'économie. Son accumulation permet d'accroître la qualité de l'information et le niveau des connaissances, ce qui joue ainsi un rôle fondamental pour l'innovation et la croissance. L'Organisation de coopération et de développements économiques (OCDE), dans un rapport publié en 2015 (*Data-Driven Innovation : Big Data for Growth and Well-Being*, OCDE, 2015), recense de manière assez précise les effets potentiels de l'usage en masse des données sur les performances des entreprises et leur capacité à innover, sur la croissance économique et sur le bien-être des consommateurs. S'il est clair que les données apparaissent comme un enjeu majeur pour l'économie, leur utilisation pose de nombreuses questions, notamment en matière de sécurité et de protection de la vie privée. Récemment, l'Union européenne s'est dotée de plusieurs règlements pour renforcer

la confiance dans le numérique, en particulier le règlement général sur la protection des données (RGPD) du 27 avril 2016, entré en vigueur en mai 2018 en France, et le règlement sur la cybersécurité du 17 avril 2019.

L'abondance et la diversité des données

L'accumulation et l'usage des données ne sont pas des faits nouveaux. Les entreprises et les États ont toujours cherché à collecter et à stocker des informations sous forme de données, pour améliorer leurs modèles économiques, pour évaluer des politiques en cours ou en mettre en œuvre de nouvelles. Grâce aux avancées technologiques, les opérations de collecte et de traitement sont nombreuses et mieux organisées, et la diffusion des informations peut se faire à une plus grande échelle. Ces dernières décennies, les progrès dans ce domaine ont permis de réduire considérablement les coûts de collecte et de traitement. Par exemple, les algorithmes développés améliorent sans cesse les capacités de production et d'analyse des données. La mise en réseau de l'ensemble des équipements numériques (ordinateurs, téléphones intelligents, capteurs) produit un effet démultiplicateur, en permettant le croisement de différents types de données. En effet, ce sont les masses des données disponibles et leur très grande variété qui offrent d'innombrables possibilités d'innovation et de développement économiques.

Il existe plusieurs types de données, de nature et de format différents. Tout d'abord, les données peuvent être regroupées en fonction du type d'informations qu'elles apportent (sur les individus, les administrations, les entreprises, les objets). Jusqu'à présent, il est surtout question dans les débats de données personnelles, dans la mesure où leur collecte et leur diffusion requièrent



de trouver un équilibre entre le respect de la vie privée des personnes et les bénéfices commerciaux et sociaux qui peuvent être obtenus. Mais, les données peuvent aussi être non personnelles : il s'agit de toutes celles qui ne peuvent pas être rattachées à une personne physique identifiable. C'est le cas des données personnelles rendues anonymes ou de celles obtenues à partir de certains objets connectés (par exemple, machines industrielles) sans possibilité d'identifier l'utilisateur.

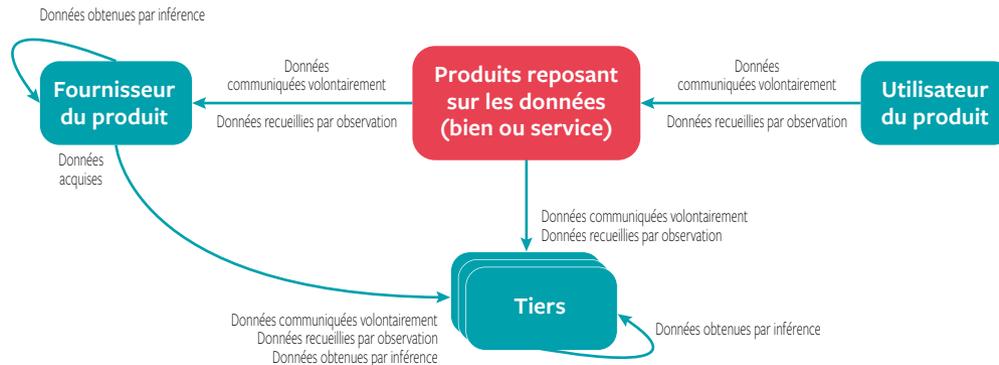
Nés au milieu des années 2000, les UMPCs (pour Ultra Mobile PC) peuvent être considérés comme une nouvelle catégorie d'ordinateurs portables
FREAKEDENOUGH/CC BY-SA 4.0

“

Les masses des données disponibles et leur très grande variété offrent d'innombrables possibilités d'innovation et de développement économiques

On distingue également les données structurées et non structurées, suivant qu'elles sont organisées ou non en base de données. Les données sont disponibles sous forme structurée lorsque les informations sont

Classification des données en fonction des méthodes de collecte



Sources : OCDE, *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*, Éditions OCDE, 2019.

classées dans des champs explicites (nom, adresse, âge), qui peuvent être liés les uns aux autres, ce qui facilite le traitement et l'analyse de ces données. Les données non structurées ne suivent ni typologie ni format précis. En d'autres termes, ce sont toutes celles qui ne sont pas facilement formatées, par exemple les fichiers audio et vidéo ou encore celles contenues dans les messageries.

Les données peuvent enfin être regroupées en fonction des méthodes de collecte utilisées. Il s'agit alors de classer séparément celles fournies volontairement par les individus (au moment de la création d'un compte sur un réseau social, lors du paiement des achats sur un site de commerce en ligne), celles collectées à partir de l'observation du comportement des individus (comportement d'un internaute lors de la navigation en ligne, données de localisation...), celles inférées à partir de données existantes (sur l'état de santé, l'orientation politique...) et enfin celles acquises auprès de tiers (entreprises spécialisées dans la vente de données, mises à disposition par les administrations...). Les

données forment donc un ensemble très hétérogène, renvoyant à des caractéristiques économiques bien spécifiques, qui font de la donnée numérique un bien singulier.

La donnée, un bien intermédiaire particulier

Le rôle joué par la donnée dans l'économie est complexe. Elle peut être un bien intermédiaire que l'entreprise combine avec les autres facteurs de production (travail, capital, énergie, terre) pour innover et élargir sa gamme de biens et services. La donnée peut aussi être une source d'information, car elle augmente la qualité et la quantité des connaissances. La collecte et l'exploitation des données permettent de réduire les asymétries d'information entre les agents économiques, améliorant ainsi l'efficacité économique (sur les marchés ou des politiques économiques). Cependant, elles peuvent être également sources de pratiques anticoncurrentielles et induire des effets néfastes sur le bien-être social, qui permet

Cette asymétrie d'information relève d'une situation dans laquelle l'entreprise dispose de peu d'informations pertinentes sur les consommateurs. Les données collectées permettent alors à l'entreprise d'améliorer ses connaissances sur le consommateur, pour ensuite mieux adapter ses produits à ses besoins ou cibler efficacement le prix auquel le produit sera vendu. En matière de politique économique, cette meilleure connaissance des entreprises et des consommateurs permet également de mieux définir les actions mises en œuvre et ainsi d'améliorer son efficacité.

de mesurer les gains économiques des entreprises et des consommateurs. Sur ce point, un rapport de l'Autorité de la concurrence et du Bundeskartellamt, publié en 2016, fournit une analyse détaillée de la manière dont l'utilisation des données par les entreprises peut conduire à un renforcement des pouvoirs de marché (la position d'un acteur sur un marché) et, ainsi, créer des distorsions de concurrence qui peuvent nuire aux consommateurs finaux (*Droit de la concurrence et données*, rapport de l'Autorité de la concurrence et du Bundeskartellamt, mai 2016).

“

La comparaison de la donnée avec une matière première est peu pertinente

La donnée est souvent désignée comme le nouvel « or noir » de l'économie ou le pétrole du XXI^e siècle. Ces métaphores donnent faussement l'idée qu'elle serait une matière première rare, disponible en quantité limitée, et source de richesse pour celui qui la détient. La comparaison avec une matière première est toutefois peu pertinente au regard de ses caractéristiques économiques qui en font un bien économique particulier.

À première vue, la donnée peut être considérée comme une marchandise, c'est-à-dire un bien échangeable sur un marché contre un prix. Dans ce cas, la transaction entre le vendeur et l'acheteur permet de révéler sa valeur d'échange. Certaines données collectées par des entreprises peuvent en effet être acquises par d'autres sociétés pour améliorer leurs produits existants ou conquérir de nouveaux



marchés. Les nouvelles technologies d'exploitation des données ont permis le développement d'entreprises spécialisées (les courtiers en données) dans la vente de données (par exemple : Criteo, Axciom ou Experian). Il arrive également que des sociétés spécialisées dans la collecte de données proposent aux utilisateurs de services numériques une rémunération en échange de leurs données. Cette situation reste encore relativement limitée au regard de la masse de données disponibles. En réalité, ce ne sont pas tant les données en elles-mêmes qui font l'objet d'une marchandisation que le traitement en masse de données de nature et de format différents répondant à des besoins spécifiques.

Les données constituent tout d'abord un bien immatériel dont la production nécessite de supporter un coût fixe, notamment pour l'installation de capteurs et d'une infrastructure associée, et un coût variable, pour la collecte, le stockage et la sécurité. Le coût marginal de la collecte (c'est-à-dire de l'obtention d'une donnée supplémentaire) est très faible dans la mesure où il s'agit souvent d'un sous-produit d'une activité de production ou de consommation. La collecte de données personnelles peut aussi donner lieu à un coût additionnel, celui de la compensation des individus et, en général, celle-ci est indirecte. Elle traduit le principe même des

Parmi les thèmes abordés en 2019 par le colloque « Données de santé et compétitivité, quels défis pour la technique et le droit ? » figurent la valorisation et la protection des données, mais aussi la protection de la vie privée

MINES PARISTECH/S. BODA/
FLICKR/CC BY-NC 2.0

Coût fixe : coût qui ne dépend pas de la quantité produite mais nécessaire au fonctionnement de l'entreprise (par exemple, loyers ou coûts administratifs).
Coût variable : coût qui dépend de la quantité produite (par exemple, coût des matières premières).
Coût marginal : coût supplémentaire induit par la dernière unité produite.



Siège social de LiveRamp-Axiom (Little Rock, États-Unis), l'une des plus importantes entreprises de stockage de données sur les consommateurs

© JUSTIN BOLLE/THE NEW YORK TIMES-REDUX-REA

marchés bifaces (qui mettent en relation deux types d'utilisateurs différents) dans lesquels les données sont collectées par une plateforme en échange d'un service numérique dont l'accès est offert sans coût à l'utilisateur. L'offre de ce service induit des coûts pour le collecteur, ces données font ensuite l'objet d'une transaction sur un marché séparé, par exemple celui des annonceurs. Le coût pour collecter ces données est alors lié à l'importance qu'accorde l'utilisateur de la plateforme au caractère privé de ces données. Par ailleurs, les coûts de stockage sont plus importants, bien qu'ils tendent à diminuer avec les possibilités du stockage distant (les services en nuage). En revanche, la sécurisation des données collectées devient un enjeu important et nécessite de supporter des coûts relativement élevés pour en assurer la protection dans un contexte où les cyberattaques se multiplient.

La donnée numérique : un bien public ?

Une caractéristique économique spécifique de la donnée est que son utilisation par un agent économique ne la détruit pas (contrairement aux matières premières) et ne prive donc pas d'autres agents de son usage. De ce point de vue, elle est un bien économique « non rival » : les données ne sont que très rarement vendues ou achetées en tant que telles ; c'est l'accès aux données, pour des usages spécifiques, qui fait l'objet d'une transaction. La question économique pertinente n'est donc pas la propriété des données, mais l'accès à ces données.

Les données peuvent être utilisées par plusieurs agents, simultanément, sans risque d'épuisement, ce qui permet d'accroître les

rendements d'échelle dès lors qu'elles sont partagées. Cette caractéristique encourage la politique ambitieuse d'ouverture des données (*open data*), aussi bien publiques que privées, que la numérisation des données a rendu possible en réduisant fortement les coûts de duplication et de transfert.

Les décisions privées peuvent toutefois limiter cette non-rivalité. Les entreprises peuvent être incitées à limiter ou contrôler l'accès aux données qu'elles détiennent (qu'il s'agisse de données de production ou de consommation). Ce comportement d'accumulation stratégique limite la contestabilité (c'est-à-dire la possibilité pour de nouvelles entreprises d'entrer sur un marché) et les bénéfices sociaux qui pourraient résulter d'une concurrence plus intense ou d'une meilleure organisation de ces marchés.

En matière d'ouverture des données, la France fait figure de bon élève au sein des pays de l'OCDE ; l'indice OURdata la classe en effet en deuxième position, après la Corée du Sud. Ce résultat récompense notamment les efforts déployés par Etalab, département de la Direction interministérielle du numérique (DINUM), en charge de la conception et de la mise en œuvre de la stratégie de l'État dans le domaine de la donnée.

La possibilité de partager largement les données génère des externalités (des effets externes qui affectent le bien-être d'un agent économique sans qu'il y ait de compensation financière) positives ou négatives. Les bénéfices induits par le partage d'information entre certains agents économiques peuvent s'étendre à toute la société. Par exemple, l'analyse des requêtes des internautes et des discussions sur les forums spécialisés aident à améliorer la surveillance des médicaments et la prévention du risque des effets indésirables. Dans le domaine du commerce en ligne, les systèmes de notation et de recommandation permettent aux internautes d'améliorer leur

connaissance des produits ou d'affiner leurs besoins. Mais, les données ou leur exploitation peuvent aussi être à la source d'externalités négatives (divulgaration d'informations intimes, vol d'identité numérique, publicités indésirables). Par conséquent, des règlements clarifient les droits et obligations des différentes parties et fixent les modalités d'une meilleure protection des données et de la vie privée.

Ce raisonnement pourrait conduire à considérer la donnée comme un bien public. Nous avons effectivement expliqué en quoi elle est un bien non rival. Pour être considérée comme un bien public, elle doit être aussi un bien non excluible : un individu ne peut, par aucun moyen, être privé de son utilisation. Or, il existe de nombreuses possibilités de limiter l'accès aux données. Tout d'abord, c'est le principe même du modèle économique des plateformes numériques : elles offrent un service en apparence gratuit aux internautes, mais vendent l'accès aux données collectées à d'autres usagers (par exemple à des annonceurs). Dans ce cas, la donnée peut être considérée comme un « bien de club », c'est-à-dire un bien excluible et non rival selon la classification proposée par la titulaire du prix Nobel d'économie Elinor Ostrom, puisqu'il faut payer pour accéder à cette donnée. Néanmoins, le fait de l'utiliser ne prive pas les autres « membres du club » de son usage. Il est également possible de limiter l'accès aux données en les stockant sur des systèmes interconnectés (notamment les services informatiques en nuage) nécessitant des investissements conséquents pour les agents économiques qui souhaitent y accéder. Cela pose aussi la question des incitations des collecteurs de données à investir pour sécuriser l'accès à ces données et réduire ainsi des coûts liés à des problèmes de cybersécurité.

#Dossier



Inauguration de la Cité de l'Objet connecté, par François Hollande, le 12 juin 2015 à Saint-Sylvain-d'Anjou (Maine-et-Loire)

LAURENT NEYSSENSAS/
FLICKR/PUBLIC DOMAIN

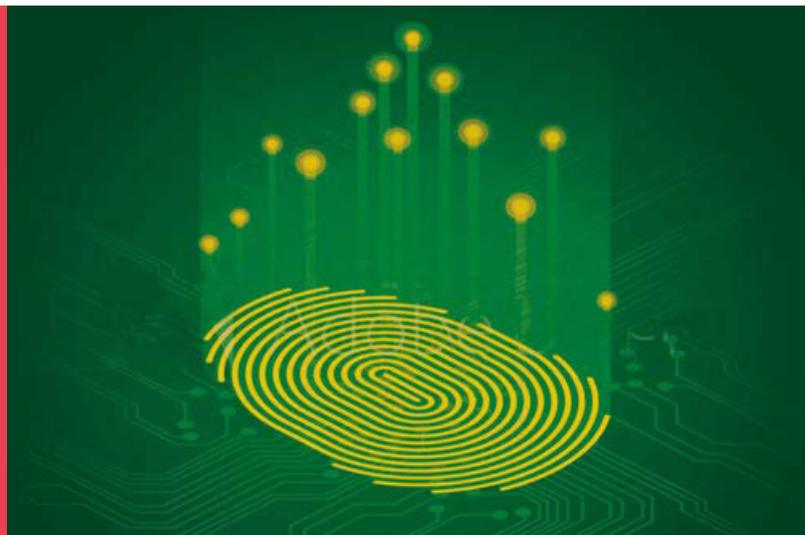
Enfin, la donnée peut avoir les caractéristiques d'un bien commun. Ce terme, mis en lumière par le philosophe Thomas d'Aquin, désigne en économie un bien partagé, dont l'usage et l'exploitation sont gérés par une communauté. Il ne s'agit pas, dans ce cas, de considérer la donnée comme une ressource partagée soumise à un risque de surexploitation et de destruction (comme c'est le cas pour un bien rival). Il s'agit de considérer les données comme des « communs » au sens de « communs informationnels », c'est-à-dire comme des ressources produites et gérées par une communauté d'intérêts, de manière coopérative (Elinor Ostrom, *Governing the Commons : The Evolution of Institutions for Collective Action*, Cambridge University Press, 1991). Ces « communs » peuvent être constitués par une communauté de contributeurs (*OpenStreetMap* en est un exemple), par l'administration (*open data* public) ou par l'ouverture de données privées.

Dans ce dernier cas, sont concernées les données privées ouvertes qui sont mises à disposition soit librement par les entreprises, soit à l'initiative de l'acteur public (les données d'intérêt général). La notion de données d'intérêt général (DIG) a été introduite en France par la loi pour une République numérique du 7 octobre 2016. L'ouverture de ces données est justifiée par leur intérêt pour l'amélioration des politiques publiques (Conseil général de l'économie, *Rapport relatif aux données d'intérêt général*, septembre 2015). Les domaines particulièrement concernés sont les transports, la mobilité, l'énergie, l'environnement et la santé.

La réalité de la donnée numérique est donc complexe et son caractère protéiforme la distingue des autres biens. Les interrogations sur ses caractéristiques économiques renvoient à la question de la valorisation qu'il convient d'articuler avec celle de la protection des données. #

Le règne des données

Une donnée numérique, qu'est-ce que c'est ?
Quels sont les bienfaits de l'*open data* ? Pourquoi
les plateformes numériques peuvent-elles devenir
si puissantes ? Comment les réguler ? Et comment
fonctionne l'intelligence artificielle ? La thématique
des données numériques, il est vrai, interroge –
Cahiers français vous donne les clés pour comprendre.



#Politiques publiques

La politique maritime de la France

#En débat

Faut-il avoir peur de la
dette publique ?

#Le point sur

La reconnaissance
de la langue des signes

NUMÉROS À PARAÎTRE

Démocratie : crise ou renouveau ?, n° 420 (mars-avril 2021)

La gouvernance des déchets, n° 421 (mai-juin 2021)

Diffusion

DILISCO

La documentation Française

Téléphone : 01 40 15 70 10

www.ladocumentationfrancaise.fr

Directrice de la publication

Anne Duclos-Grisier

ISBN (papier) : 978-2-11-157419-9

ISBN (PDF web) : 090-0-00-404190-3

ISBN (ePub) : 978-2-11-157317-8