

# Anonymat - Anonymisation

## Histoire anonymat

Commençons par l'étymologie : attesté en français depuis le XVI<sup>e</sup> siècle, *anonyme* dérive du bas latin *anonymus* qui, lui-même, provient du grec *anônumos*, mot forgé à partir du préfixe privatif *an* et de *onoma*, « nom », ce qui signifie littéralement « sans nom ». Rien donc de vraiment neuf derrière ce mot : est anonyme ce qui ne peut-être assigné à un individu reconnaissable.

Ainsi, parle-t-on depuis longtemps de **lettres anonymes** ou d'**œuvres anonymes**.

- Dans un cas, quelqu'un se cache volontairement pour avancer masqué,
- dans l'autre l'identité de l'auteur se perd pour d'obscures raisons.

Quoi qu'il advienne, l'idée demeure et, partout, la connotation se révèle plutôt négative, puisque cela revient au pire à une *dissimulation*, au mieux, à une *perte*.

Rappelons que, pendant longtemps, celui qui se dérobaient délibérément aux regards apparaissait comme suspect : seul un coupable était censé procéder de la sorte. A tout moment, l'individu innocent devait agir au grand jour, en toute transparence, et rendre des comptes. A titre d'illustration, à Athènes, dans l'antiquité, la plupart des votes se faisaient publiquement, à main levée, car les hommes libres devait répondre de leurs choix.

Et, même dans le passé récent, l'anonymat n'avait qu'un caractère dérogatoire. Ainsi en allait-il de l'**accouchement sous X** qui permettait à une mère de mettre au monde un enfant sans donner son nom, pour ne pas établir de lien officiel de filiation et procéder légalement à un abandon. Dans registre similaire, en France, à partir de 1831, la **légion étrangère** engageait ses hommes sous une simple « identité déclarée », sans exiger de preuves, ce qui garantissait l'anonymat et offrait à certains une seconde chance.

Ajoutons à cela des **franchises temporaires**, à caractère plus ou moins ludique, comme les bals masqués, pendant lesquelles l'anonymat s'imposait. Jeu, écart passager, cela restait, bien évidemment exceptionnel. On s'éloignait pour un temps de la morale ordinaire pour mieux y retourner après.

Il arrivait aussi que, dans de rares cas, l'anonymat ait des **connotations positives**, comme pour les dons qui, lorsqu'ils étaient anonymes, attestaient d'une sincère compassion, sans aspiration à une gratification ou à une quelconque reconnaissance. Mais cela demeurait l'exception.

## Actualité de l'anonymat

Aujourd'hui, dans l'univers numérique, le **statut de l'anonymat change**.

Certes, sur internet le jeu demeure. En témoigne l'usage généralisé des pseudonymes dans les jeux en lignes ou les réseaux sociaux. Il arrive même que certains changent d'âge, d'occupation ou de sexe, par simple plaisir. Sans doute retrouvons nous là le caractère ludique de l'anonymat. Si cela perd son caractère exceptionnel, ce demeure un jeu.

Mais, là où l'anonymat avait essentiellement une connotation négative, il prend, de plus en plus souvent, une connotation positive.

Notons d'abord que désormais de **nombreux dons exigent l'anonymat**. Ainsi,

- le don de sperme,
- le don d'organes et, dans une moindre mesure,
- le don de sang,

se font uniquement en France sous couvert d'anonymat.

De même, souvent, l'**expertise** n'apparaît neutre que lorsqu'elle s'accompagne de garanties d'anonymats. Songeons à l'**évaluation scientifique** des articles ou des projets. Cela part du principe que seul celui qui se prononce à l'abri des regards est libre, car sinon, il s'exposerait à des représailles. Cette tendance tient peut-être à la disparition de l'autorité, au sens ancien, car plus personne ne possède l'ascendant suffisant pour décider seul, sans être remis en cause. On recourt donc à des procédures supposées équitables où l'évaluateur agit à l'abri. Ce n'est toutefois pas le lieu, ici, de prolonger cette analyse.

**En politique**, on pense aussi que l'anonymat est nécessaire pour pouvoir agir de façon désintéressée, au regard de ses seules convictions, sans revendiquer quoi que ce soit d'autre, sans subir de pression non plus, ni a fortiori de rétorsions. C'est ainsi que de nos jours, un vote n'apparaît démocratique que s'il est anonyme, à savoir si l'on n'est pas en mesure de remonter à l'identité des personnes qui ont voté de telle ou telle façon. À ce titre, il est révélateur qu'un groupe de hackers subversifs procédant à des attaques informatiques sur les réseaux pour défendre leurs opinions se soit appelé les « **Anonymous** » et adopte comme emblème un masque de carnaval.

Plus généralement, avec l'essor sans précédent de l'individualisme, on revendique l'anonymat pour **protéger sa vie privée** et agir à son bon vouloir, sans contrainte. Cela apparaît comme le *garant de la liberté* entendue comme possibilité de se comporter à sa guise, tant que cela n'enfreint pas les lois. Cela porte entre autre sur les traces que nous laissons, à tout moment, lorsque nous payons avec nos cartes de crédit, lorsque nous voyageons avec nos passes magiques, lorsque nous ouvrons nos téléphones portables, lorsque nous appelons nos correspondants, lorsque nous allons chez le médecin ou que nous achetons des médicaments, etc. Nous ne souhaiterions pas que l'on utilise ces informations à notre insu, pour nuire à notre personne d'une quelconque façon.

## Limites de l'anonymat

### Limites intrinsèques

La première limite vient de données qui sont si intimement liées à la personne qu'il est impossible, avec elles, de conserver l'anonymat. C'est en particulier le cas de données biométriques en général, par exemple des empreintes digitales, de la posture ou, pire, des données génomiques. Dans cette dernière éventualité la séquence ADN identifie non seulement l'individu, sans équivoque, mais aussi l'ensemble de son lignage, à savoir son ascendance et sa descendance. Nous verrons, dans la suite de la journée, les précautions qu'il convient de prendre avec des données de cet ordre. Nous discuterons aussi des conséquences que cela induit sur les données elles-mêmes.

### Limites éthiques

Notons toutefois que cette aspiration à l'anonymat se heurte non seulement à des difficultés intrinsèques qui tiennent à des données qui trahissent directement l'individu, sans qu'il y ait de dissimulation possible, mais à des difficultés éthiques qui font que dans certains cas de figure, il est souhaitable, pour des raisons éthiques, de **dévoiler des informations sur les personnes**, sans rien masquer.

Cela peut tenir à une **volonté assez partagée aujourd'hui de transparence**. En effet, même si nous souhaitons tous demeurer anonymes dans notre vie de tous les jours pour protéger notre intimité, nous exigeons aussi des personnages politiques et, plus généralement, de toutes les femmes et de tous les hommes investis d'une quelconque responsabilité, par exemple du médecin, de l'avocat, du professeur, de l'expert, etc. qu'ils se comportent au vu et au su de tout le monde, sans rien dissimuler. Nous voulons tout savoir d'eux, de leur passé, de leurs occupations, de leur vie affective et intime afin de nous assurer que n'entre aucun conflit d'intérêt dans leurs décisions. Tel est certainement la tension qui se fait jour derrière notre besoin contemporain d'anonymat.

Il y aurait certainement beaucoup à dire sur cette tension. Ce n'est pas le lieu ici. Mais, je souhaiterais tout de même mentionner la position de la *quadrature du net* selon laquelle il y aurait des puissants, dont on exige la transparence, et des « petits » qui ont droit à la vie privée. Or, cette vision est discutable, car nous sommes tous à la fois des personnes privées qui ont droit à leur intimité, et des personnes investies, dans leur activité professionnelle, d'une fonction publique. Cela signifie qu'il faut trouver des compromis entre le besoin de protection de la vie privée et l'exigence de transparence.

À cette aspiration à la transparence, on doit ajouter une demande très forte de sécurité, qu'il s'agisse de sécurité intérieure, pour lutter contre la criminalité ou le terrorisme, ou de sécurité sanitaire, pour lutter contre les épidémies, mais aussi pour protéger la santé individuelle. Là encore, il existe des compromis entre une vision sécuritaire très forte, qui prétexterait de la défense collective pour faire irruption à tout moment dans notre vie privée, et une absence totale de contrôle, qui conduirait à mettre en danger les individus du fait de l'impossibilité de dévoiler des informations les réseaux maffieux ou terroristes, au nom de la protection de la liberté individuelle. Il en va de même pour la protection de la santé individuelle : il y a un compromis à faire entre la protection de l'individu et le besoin de connaître son état de santé pour le protéger.

## Anonymisation et Nécessité de l'anonymat

Certains affirment, qu'en pratique, l'anonymat n'est pas nécessaire aux données de recherche, car lorsque les chercheurs qui travaillent dessus en ont connaissance, ils s'engagent à n'en faire usage que pour les besoins de leurs recherches, et que celles-ci ne visent que l'avancée de connaissances générales. Il n'y aurait donc rien à craindre pour l'intimité de l'individu. Autrement dit, les données de recherche auraient au sens juridique un statut « exorbitant », en cela qu'il sortirait du droit commun. Cela signifie que les chercheurs entreraient dans une forme de confidentialité, ce qui signifie qu'ils ne mésuseraient pas des données personnelles, au risque sinon d'être fortement punis, mais aussi qu'ils ne les communiqueraient pas.

À cela, on peut opposer deux arguments.

D'une part, pour des raisons techniques, il se peut que des personnes mal intentionnées volent les données. Il faut donc s'assurer de la **sécurité des systèmes informatiques** où les données seront stockées. Or, cette sécurité n'est ni assurée, ni certaine, car des pirates informatiques peuvent s'introduire dans les systèmes les plus fermés pour voler les données. Cette sécurité paraît d'autant plus difficile à assurer que les données sont manipulées par beaucoup de personnes. De plus, les logiciels de traitement des données peuvent eux-mêmes transmettre les données.

À cet égard, on doit noter que beaucoup d'algorithmes d'apprentissage machine sont libres de droit, comme la **bibliothèque TensorFlow** de Google Brain ou **OpenAI** d'Elon Musk, ce qui paraît positif. Ils sont accessibles sur les sites des grands acteurs de l'internet, comme Google, qui les mettent à disposition de tous avec des ressources de calcul, comme le **tensor processing unit (TPU)** qui exécute très efficacement l'apprentissage sur les réseaux de neurones. Rien ne nous garantit toutefois que les données ne seront pas volées.

D'autre part, la science ouverte est de plus en plus en vogue. Il est aujourd'hui demandé que les travaux issus de recherches financées sur des fonds publics européens soient en accès libres. Cela va de pair avec la volonté de **partage des données de recherche**. Outre son caractère généreux, ce partage se justifie au plan technique par la capacité de traiter d'immenses masses de données avec les techniques d'apprentissage machine, et surtout par le fait que la qualité des résultats dépend de la quantité d'exemples. Il s'ensuit qu'il devient nécessaire de partager les données de recherche, en particulier les données personnelles. Or, avec ce partage des données personnels, et cette ouverture des données, il n'y a plus de confidentialité possible. Il faut donc essayer de **rendre les données anonymes**. Dispose-t-on de techniques qui permettent de le faire de façon fiable, sans possibilité de remonter aux personnes qui sont à l'origine des données et sans dégrader ces mêmes données ? C'est ce qui va faire l'objet de nos réflexions aujourd'hui.

Ceci étant, avant de commencer, il convient de rappeler les difficultés de l'opération d'**anonymisation** qui empêche de retrouver la personne, c'est-à-dire de l'**identifier**, au moyen d'une opération symétrique de l'anonymisation, la **ré-identification**, qui vise à retrouver l'identité de la personne, même quand elle paraît masquée.

En effet, l'élimination des noms ou leur remplacement par un autre ne suffisent pas. Cette dernière opération qui s'appelle la pseudonymisation — remplacement d'un nom par un pseudo — risque d'être détectée. De plus, il peut y avoir des incohérences, qui font que l'on identifiera l'existence de transformations. À cela s'ajoute la possibilité de

**retrouver — ou d'identifier, opération symétrique de l'anonymisation — les personnes par croisement.** Enfin, la dégradation de la qualité des données peut sinon empêcher leur exploitation, tout au moins diminuer les performances de l'apprentissage. Mais, je n'en dit pas plus car cela fait l'objet de la journée.

Quatre volets seront envisagés aujourd'hui :

1. Un volet juridique (définition)
2. Un volet technique, qui présentera les techniques informatiques d'anonymisation en montrant les limitations des différents procédés.
3. Domaines d'application : génomique, santé
4. Industries
5. Données éducatives, particulièrement sensibles