

## **ATELIER 2020-2021 de la PLATEFORME ETHIQUE ET BIOSCIENCES**

### **USAGES ETHIQUES DES BIG DATA EN BIOSCIENCES**

La Plateforme « Ethique et Biosciences » de la Génopole® Toulouse, Occitanie, organise chaque année, depuis 2006, un atelier thématique de réflexion éthique, destiné en priorité à la communauté scientifique et ouvert à toute personne intéressée. Il s'agit de donner la parole aux chercheurs, à l'ensemble des professionnels, et au grand public, afin de leur offrir l'opportunité d'interagir avec des spécialistes ayant une expérience dans le domaine. Ces ateliers se déroulent en 3 volets d'une après-midi chacun et analysent les différentes facettes d'un thème choisi. Ils font une large place à la discussion et sont animés par un membre de la plateforme. Chaque atelier forme un tout cohérent ; cependant la participation à l'un de ces volets n'est pas conditionnée par une présence à chacun d'entre eux.

## **PRESENTATION DE L'ATELIER**

Les Big Data - ou données massives - font parler d'elles depuis plusieurs années, compte tenu des nouvelles capacités d'explorations qu'elles permettent et des perspectives d'usages qu'elles ouvrent pour l'ensemble des secteurs d'activités : la santé (ex. CCNE Avis 130: <https://www.ccne-ethique.fr/fr/publications/donnees-massives-et-sante-etat-des-lieux-prospective-et-nouvelles-questions-ethiques>), la sécurité, le marketing, etc.

La science, qui produit et utilise de plus en plus de données, se retrouve au premier rang des acteurs du Big Data. Cette production exponentielle est due aux capacités technologiques contemporaines (informatiques, techniques) permettant de générer des données massives et d'en assurer le stockage. Quant aux usages, ils sont poussés par la « Science de la donnée » postulant que les questions de recherche peuvent être directement formulées à partir de données existantes ; renversant le paradigme prévalant jusqu'à lors de l'acquisition des données *a posteriori* de la formulation des hypothèses de recherche. Un défi majeur persiste pour assurer le traitement de ces données massives en toute fiabilité et sécurité, et de nouveaux outils devront être élaborés pour permettre leur usage et leur traitement de manière responsable et transparente (notamment grâce à l'intelligence artificielle). Ces enjeux techniques se doublent d'une politique de plus en plus fréquente de partage de ces

#### **PLATEFORME ETHIQUE ET BIOSCIENCES**

Génopole® Toulouse Occitanie  
Unité mixte Inserm - Université Toulouse III Paul Sabatier  
UMR 1295 - CERPOP  
Faculté de Médecine, 37 allées Jules Guesde  
31000 Toulouse - Tél. + 33 (0) 5 61 14 59 38  
<http://societal.genotoul.fr/>

#### **Contacts :**

Emmanuelle RIAL-SEBBAG (Responsable scientifique)  
[emmanuelle.rial@univ-tlse3.fr](mailto:emmanuelle.rial@univ-tlse3.fr)  
Gauthier CHASSANG (Responsable opérationnel)  
[gauthier.chassang@gmail.com](mailto:gauthier.chassang@gmail.com)  
Lucie SERRES (Coordinatrice)  
[serres.lucie@gmail.com](mailto:serres.lucie@gmail.com)  
Anne CAMBON-THOMSEN (Fondatrice de la Plateforme)  
[anne.cambon-thomsen@univ-tlse3.fr](mailto:anne.cambon-thomsen@univ-tlse3.fr)

données, de leur mise à disposition et donc de leur ouverture à l'ensemble de la communauté scientifique et de leur réutilisation. Ainsi, du fait notamment du volume et de la variété des données en présence, plusieurs questions éthiques et pratiques se posent pour collecter, stocker, utiliser ces données dans le respect des droits de l'homme, dans l'intérêt général et pour le bien commun.

Les ateliers de la Plateforme visent à aborder les questions éthiques, juridiques et pratiques des Big Data en tant que ressources indispensables au développement des biosciences<sup>1</sup> modernes, depuis la définition du concept des Big Data à la réalité de la recherche de terrain, au travers d'exemples de projets utilisant ces ressources et des débats qui suivront. Qu'appelle-t-on Big Data ? Quelles sont les opportunités d'usages et les enjeux éthiques liés aux Big Data ? Comment assurer un accès et une utilisation éthique des Big Data dans le domaine des biosciences ? Quels défis restent-ils à relever pour que les Big Data soient une ressource accessible et fiable ?

Ces réflexions s'articulent autour de **3 volets** :

**Volet 1 : « Définition des Big Data : mise en commun et partage »**

Jeudi **12 mars 2020** de 13h30 à 17h à l'Université Fédérale Toulouse Midi-Pyrénées, 41 allées Jules Guesde, Toulouse - Salle du Conseil et Salle de réception.

**Volet 2 : « Big Data et innovations thérapeutiques »**

Jeudi **4 mars 2021** de 13h30 à 17h – En visioconférence.

**Volet 3 : « Enjeux socio-économiques des Big Data »**

Jeudi **20 mai 2021** de 13h30 à 16h – En visioconférence.

*[Le format sera adapté selon le contexte sanitaire (en présentiel ou en visioconférence)]*

**Inscription gratuite mais obligatoire** par envoi d'une [fiche d'inscription](#) à Lucie Serres ([serres.lucie@gmail.com](mailto:serres.lucie@gmail.com)).

---

<sup>1</sup> Biosciences, ou sciences de la vie, c'est à dire l'ensemble des études et disciplines scientifiques qui s'intéressent à la vie, qu'elle soit animale, végétale, humaine (ex: la zoologie, la botanique, la génétique, l'agronomie, la paléontologie, la recherche médicale, l'écologie...) ou qui utilisent des êtres vivants dans différentes applications (ex : biotechnologies).

**PLATEFORME ETHIQUE ET BIOSCIENCES**

Génopole® Toulouse Occitanie  
Unité mixte Inserm - Université Toulouse III Paul Sabatier  
UMR 1295 - CERPOP  
Faculté de Médecine, 37 allées Jules Guesde  
31000 Toulouse - Tél. + 33 (0) 5 61 14 59 38  
<http://societal.genotoul.fr/>

**Contacts :**

Emmanuelle RIAL-SEBBAG (Responsable scientifique)  
[emmanuelle.rial@univ-tlse3.fr](mailto:emmanuelle.rial@univ-tlse3.fr)  
Gauthier CHASSANG (Responsable opérationnel)  
[gauthier.chassang@gmail.com](mailto:gauthier.chassang@gmail.com)  
Lucie SERRES (Coordinatrice)  
[serres.lucie@gmail.com](mailto:serres.lucie@gmail.com)  
Anne CAMBON-THOMSEN (Fondatrice de la Plateforme)  
[anne.cambon-thomsen@univ-tlse3.fr](mailto:anne.cambon-thomsen@univ-tlse3.fr)

## Volet 3 : « Enjeux socio-économiques des Big Data »

Animatrice : **Laurence Mabile**, Ingénieur de recherche, Docteur en biochimie, Université Toulouse III Paul Sabatier, CERPOP - UMR 1295, Université de Toulouse, INSERM, UPS, Equipe BIOETHICS et Equipe EQUITY, Membre de la Plateforme Ethique et Biosciences (Genotoul Societal).

Intervenants :

- **Michelle Kelly-Irving**, Épidémiologiste, Chargée de recherche INSERM, CERPOP - UMR 1295, Université de Toulouse, INSERM, UPS, Responsable de l'équipe EQUITY, Directrice de l'IFERISS.
- **Bertrand Monthubert**, Conseiller régional délégué à l'enseignement supérieur et à la recherche, Président d'Occitanie Data, Membre du groupe « Gouvernance des données » du Partenariat Mondial de l'Intelligence Artificielle.

Date : le jeudi 20 mai 2021 de 13h30 à 16h.

Format : En visioconférence (suite à votre inscription un lien de connexion vous sera personnellement adressé par mail avant l'atelier).

## PRESENTATION DU VOLET

Les Big Data, ou données massives, se développent aujourd'hui à tous les niveaux que ce soit en termes de capacités de collecte, de stockage et d'analyse ou d'applications à des domaines variés. Au-delà des aspects réglementaires, cette évolution n'est pas sans impact sur son environnement et notamment au niveau socio-économique. L'accès à cette multitude de données peut être source de développements socio-économiques vis-à-vis de secteurs scientifiques, technologiques ou financiers (ex : augmentation des connaissances, élaboration de modèles, mises en applications...) ; et être à la fois source d'interrogations sur la fiabilité de ces données et les risques que celles-ci sont susceptibles d'engendrer.

Ces Big Data et les analyses qui en découlent sont-elles représentatives de la diversité socio-économique existante ? Cela peut-il avoir un impact sur l'évolution de la science ou la confiance portée aux concepteurs d'applications ou de technologies ? Les données manquantes ou inaccessibles ne peuvent-elles pas engendrer des biais d'analyses dont nous ne mesurons pas encore les impacts ? Nous proposons d'aborder cette thématique au travers de deux interventions d'experts du domaine, suivies d'un débat avec les participants.

### PLATEFORME ETHIQUE ET BIOSCIENCES

Génopole® Toulouse Occitanie  
Unité mixte Inserm - Université Toulouse III Paul Sabatier  
UMR 1295 - CERPOP  
Faculté de Médecine, 37 allées Jules Guesde  
31000 Toulouse - Tél. + 33 (0) 5 61 14 59 38  
<http://societal.genotoul.fr/>

### Contacts :

Emmanuelle RIAL-SEBBAG (Responsable scientifique)  
[emmanuelle.rial@univ-tlse3.fr](mailto:emmanuelle.rial@univ-tlse3.fr)  
Gauthier CHASSANG (Responsable opérationnel)  
[gauthier.chassang@gmail.com](mailto:gauthier.chassang@gmail.com)  
Lucie SERRES (Coordinatrice)  
[serres.lucie@gmail.com](mailto:serres.lucie@gmail.com)  
Anne CAMBON-THOMSEN (Fondatrice de la Plateforme)  
[anne.cambon-thomsen@univ-tlse3.fr](mailto:anne.cambon-thomsen@univ-tlse3.fr)

## Résumé de Bertrand Monthubert – « Occitanie Data : comment l'éthique est un levier du développement des Big Data »

Alors que le Big Data est un champ d'application qui a été fortement poussé par des acteurs scientifiques et technologiques, grâce à des succès applicatifs importants, la question de l'éthique a commencé à émerger depuis plusieurs années, suite à des utilisations désastreuses de ces techniques. Si un cadre réglementaire de protection des données personnelles s'est progressivement mis en place à travers le monde (RGPD en Europe, Act on the Protection of Personal Information au Japon, California's Consumer Privacy Act...), une perception grandissante est qu'il y a une tension entre la dynamique d'innovation et celle de la protection des données personnelles, conduisant même le Partenariat Mondial de l'Intelligence Artificielle à travailler sur le thème « Squaring the circle of achieving an innovation-friendly climate and a high level of protection in data privacy law ».

Nous verrons comment l'éthique doit intervenir dans le débat, qui ne peut rester au seul niveau réglementaire. Le respect d'un cadre éthique n'est pas seulement une contrainte, c'est aussi un levier qui fournit un cadre d'analyse permettant de légitimer le recours au Big Data, en utilisant des données personnelles, dans certaines situations que la seule analyse réglementaire ne permet pas de résoudre. C'est aussi un élément fondamental dans la création d'un cadre de confiance, dont la nécessité s'est accrue au fur et à mesure que l'atteinte aux intérêts des individus s'est dévoilée. Aujourd'hui, la confiance est apparue au sommet des préoccupations des acteurs du Big Data et de l'Intelligence Artificielle.

A travers l'analyse des travaux d'Occitanie Data dont la Commission Ethique animée par Emmanuelle Rial-Sebbag a élaboré une *Charte éthique pour un développement responsable de l'économie de la donnée*, et du Partenariat Mondial de l'Intelligence Artificielle, nous observerons la manière dont les questions éthiques sont devenues une condition du développement des Big Data.

### PLATEFORME ETHIQUE ET BIOSCIENCES

Géopole® Toulouse Occitanie  
Unité mixte Inserm - Université Toulouse III Paul Sabatier  
UMR 1295 - CERPOP  
Faculté de Médecine, 37 allées Jules Guesde  
31000 Toulouse - Tél. + 33 (0) 5 61 14 59 38  
<http://societal.genotoul.fr/>

### Contacts :

Emmanuelle RIAL-SEBBAG (Responsable scientifique)  
[emmanuelle.rial@univ-tlse3.fr](mailto:emmanuelle.rial@univ-tlse3.fr)  
Gauthier CHASSANG (Responsable opérationnel)  
[gauthier.chassang@gmail.com](mailto:gauthier.chassang@gmail.com)  
Lucie SERRES (Coordinatrice)  
[serres.lucie@gmail.com](mailto:serres.lucie@gmail.com)  
Anne CAMBON-THOMSEN (Fondatrice de la Plateforme)  
[anne.cambon-thomsen@univ-tlse3.fr](mailto:anne.cambon-thomsen@univ-tlse3.fr)

## Résumé de Michelle Kelly-Irving – « Qui, quoi, où ? Prise en compte des facteurs sociaux et socio-économiques dans l'évaluation des données, notamment dans le contexte de la crise sanitaire liée à la Covid 19 »

Malgré l'apparente disponibilité et l'accès à de vastes quantités de données, la crise de la Covid 19 a mis en évidence, entre autres, que nous n'avons toujours pas accès aux données dont nous avons besoin, notamment en santé publique. J'évoquerai dans cette présentation que l'on ne peut pas supposer que les données auxquelles nous avons accès sont celles dont nous avons besoin. En termes social et socio-économique, *qui* est représenté dans les bases de données auxquelles nous avons accès ? Quels types de données manquent ? Quelles personnes, populations et territoires sont absents ? Les données auxquelles nous avons accès peuvent au final amplifier les biais et discriminations sociales, ainsi, les analyses et algorithmes produits à partir des grandes bases de données peuvent être source de reproduction des inégalités sociales et socio-économiques, déjà bien ancrées dans nos sociétés. Je prendrai des exemples pour examiner ces biais afin de réfléchir aux questions éthiques qui en découlent.

Par ailleurs, nous réfléchissons aux questions éthiques autour de la vie privée, de l'accessibilité et de la protection des données, des procédures mises en place pour protéger les informations sur les individus. En prenant l'exemple de la crise sanitaire de la Covid, je soulignerai que les données inaccessibles ou inexistantes nous empêchent de pouvoir examiner en temps réel l'évolution de l'épidémie et les inégalités qui la caractérise. Les possibilités qui s'offrent à nous en concevant, collectant et analysant les données de manière responsable sont immenses. Nous pourrions observer des tendances et même simuler des interventions visant à redresser les inégalités. Mais les lois et procédures peuvent nous empêcher d'avoir accès aux données pour ce faire. Quelles sont les questions éthiques sur le fait de ne pas accéder à des données ou les utiliser pour ce faire, alors que c'est possible ?

### PLATEFORME ETHIQUE ET BIOSCIENCES

Génope® Toulouse Occitanie  
Unité mixte Inserm - Université Toulouse III Paul Sabatier  
UMR 1295 - CERPOP  
Faculté de Médecine, 37 allées Jules Guesde  
31000 Toulouse - Tél. + 33 (0) 5 61 14 59 38  
<http://societal.genotoul.fr/>

### Contacts :

Emmanuelle RIAL-SEBBAG (Responsable scientifique)  
[emmanuelle.rial@univ-tlse3.fr](mailto:emmanuelle.rial@univ-tlse3.fr)  
Gauthier CHASSANG (Responsable opérationnel)  
[gauthier.chassang@gmail.com](mailto:gauthier.chassang@gmail.com)  
Lucie SERRES (Coordinatrice)  
[serres.lucie@gmail.com](mailto:serres.lucie@gmail.com)  
Anne CAMBON-THOMSEN (Fondatrice de la Plateforme)  
[anne.cambon-thomsen@univ-tlse3.fr](mailto:anne.cambon-thomsen@univ-tlse3.fr)

## PROGRAMME

- 13h30 *Accueil et présentation des participants*
- 13h40 **Présentation du volet 3 de l'atelier par Laurence Mabile**
- 13h50 **Bertrand Monthubert**, Conseiller régional délégué à l'enseignement supérieur et à la recherche, Président d'Occitanie Data, Membre du groupe « Gouvernance des données » du Partenariat Mondial de l'Intelligence Artificielle  
**« Occitanie Data : comment l'éthique est un levier du développement des Big Data »**
- 14h20 Discussions
- 14h30 **Michelle Kelly-Irving**, Épidémiologiste, Chargée de recherche INSERM, CERPOP - UMR 1295, Université de Toulouse, INSERM, UPS, Responsable de l'équipe EQUITY, Directrice de l'IFERISS  
**« Qui, quoi, où ? Prise en compte des facteurs sociaux et socio-économiques dans l'évaluation des données, notamment dans le contexte de la crise sanitaire liée à la Covid 19 »**
- 15h00 Discussions
- 15h10 *Pause*
- 15h20 **Atelier de réflexion**
- 16h00 *Fin de l'atelier*

### PLATEFORME ETHIQUE ET BIOSCIENCES

Génopole® Toulouse Occitanie  
Unité mixte Inserm - Université Toulouse III Paul Sabatier  
UMR 1295 - CERPOP  
Faculté de Médecine, 37 allées Jules Guesde  
31000 Toulouse - Tél. + 33 (0) 5 61 14 59 38  
<http://societal.genotoul.fr/>

### Contacts :

Emmanuelle RIAL-SEBBAG (Responsable scientifique)  
[emmanuelle.rial@univ-tlse3.fr](mailto:emmanuelle.rial@univ-tlse3.fr)  
Gauthier CHASSANG (Responsable opérationnel)  
[gauthier.chassang@gmail.com](mailto:gauthier.chassang@gmail.com)  
Lucie SERRES (Coordinatrice)  
[serres.lucie@gmail.com](mailto:serres.lucie@gmail.com)  
Anne CAMBON-THOMSEN (Fondatrice de la Plateforme)  
[anne.cambon-thomsen@univ-tlse3.fr](mailto:anne.cambon-thomsen@univ-tlse3.fr)