



# **Le pistage des données dans le monde académique - Retours d'expérience des négociations DEAL**

Dr. Michaela Bilic-Merdes, Fondation Allemande pour la Recherche (DFG)

Printemps Couperin | Paris | 19.03.2025

# Le pistage des données dans le monde académique

User Tracking on Academic Publisher Platforms

*Cody Hanson*  
@codyh  
codyhanson@umn.edu

[User Tracking on Academic Publisher Platforms](#)



## Stop Tracking Science

[Stop Tracking Science](#)

Data tracking in research: aggregation and use or sale of usage data by academic publishers

A briefing paper of the Committee on Scientific Library Services and Information Systems of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)

20 May 2021

Deutsche Forschungsgemeinschaft  
Königstraße 57, 53170 Bonn  
Telefon: +49 228 885-1; Telefax: +49 228 885-2777; postmaster@dfg.de; www.dfg.de



[datentracking\\_papier\\_en.pdf \(dfg.de\)](#)

TRAÇAGE DES DONNÉES  
DANS LA RECHERCHE



07/04/2022

Agrégation et utilisation ou vente des  
données d'usage par les éditeurs  
scientifiques

Une synthèse du Comité des services de bibliothèques  
scientifiques et des systèmes d'information de la  
Deutsche Forschungsgemeinschaft (DFG, Fondation  
allemande pour la recherche)

[Note\\_d'information\\_tracage\\_fr.pdf \(hal.science\)](#)

- **Note de synthèse** du comité des services de bibliothèques scientifiques et des systèmes d'information (AWBI) de la Fondation allemande pour la recherche (DFG) – *et traduction française*
- Présentation des **pratiques actuelles dans le pistage des données (data tracking)**
  - Provoquer la **discussion sur les conséquences** auprès de tous les acteurs de la communauté académique

# Le pistage des données dans le monde académique

## Digital Privacy

Dr. Robert Altschafel, Prof. Dr. Michael Beurskens, Prof. Dr. Jana Dittmann, Prof. Dr. Wolfram Horstmann, Dr. Stefan Kiltz, Prof. Dr. Gerhard Lauer, Judith Ludwig, Dr. Bernhard Mittermaier and Katrin Stump

### Data tracking and DEAL

– On the 2022/2023 negotiations and the consequences for academic libraries –  
In memory of Prof Dr. Gerald Spindler (1960-2023)

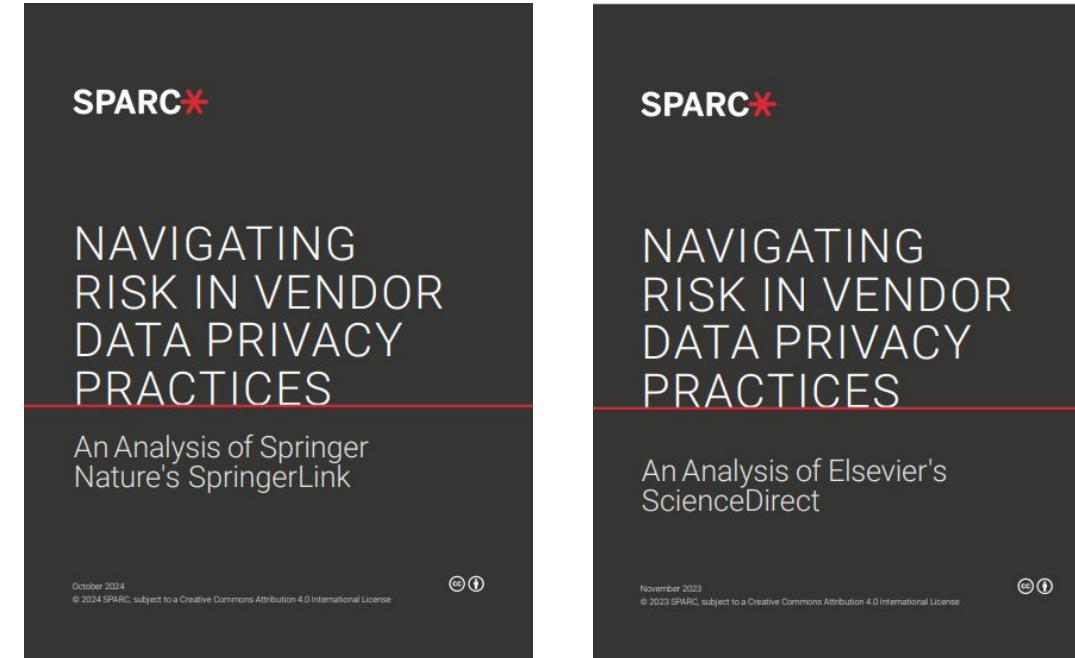
### I. The Maxwell-Garfield system of academic publishing

After the end of the Second World War, a structure of academic publishing developed that can be specifically referred to as the *Maxwell-Garfield system*. This refers to two post-1945 developments that were initially only indirectly linked to one another. The name *Robert Maxwell* is associated with a particularly ruthless commercialisation of academic publishing (Miranda 2001) where costs are allocated solely to the public and profits solely to the publishers. With the rapid growth of universities, this business model – described by Deutsche Bank as “bizarre” (Deutsche Bank 2005; Klein 2019) – gave rise to an oligopoly of a small number of publishers, especially in the academic journal sector, while also resulting in drastic price increases for libraries and low costs for publishers. Meanwhile, the name *Eugen Garfield* is associated with an index that was initially developed to help libraries sort their acquisitions (Garfield 1955; Cronin & Atkins 2000), but which rapidly became an instrument of control, favouring the large publishers who adopted this index early on in guiding their pricing policy; at the same time, it enabled these publishers to make data available to university administrations as a basis on which to make decisions on research policy investments and structural developments.

<sup>1</sup> Members of the Ad Hoc Working Group of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) on Data Tracking in Research under the direction of Dr. Michaela Bllic-Merdes: Dr. Robert Altschafel, Otto von Guericke University Magdeburg; Prof. Dr. Michael Beurskens, University of Passau; Prof. Dr. Jana Dittmann, Otto von Guericke University of Magdeburg; Prof. Dr. Wolfram Horstmann, FIZ Karlsruhe – Leibniz Institute for Information Infrastructure; Dr. Stefan Kiltz, Otto von Guericke University of Magdeburg; Prof. Dr. Gerhard Lauer, University of Mainz; Judith Ludwig, German National Library of Science and Technology (TIB), Hanover; Dr. Bernhard Mittermaier, Forschungszentrum Jülich; Dr. Hildegarde Schäffler, Bavarian State Library, Munich; Prof. Dr. Gerald Spindler, University of Göttingen and Katrin Stump, Saxon State and University Library, Dresden.

Recht und Zugang, issue 1-2024 DOI: 10.5771/2699-1284-2024-1-23  
<https://doi.org/10.5771/2699-1284-2024-1-23>, am 04.07.2024, 11:15:57  
Open Access – CC BY-NC-ND <https://www.norusis-eLibrary.de/ojs/>

[Data tracking and DEAL: On the 2022/2023 negotiations and the consequences for academic libraries](#)



## ► Enjeux du pistage des données dans le contexte des négociations DEAL avec les éditeurs Elsevier, Springer Nature and Wiley

- Inclusion de clauses de protection des données dans les licences des abonnements
- Consultation et évaluation des résultats par le groupe de travail “Data Tracking” de la DFG et DEAL

# Évolution de la responsabilité en matière de protection de la vie privée

## Le monde des bibliothèques analogiques



[Pixabay](#)

- ▶ Environnement protecteur de la vie privée
- ▶ Normes élevées de protection
- ▶ Responsabilité et contrôle des bibliothécaires

## Plateformes numériques de l'information



DALL-E

- ▶ Collecte intrusive de données d'usage à personnel
- ▶ « Normes » commerciales / « surveillance commerciale »
- ▶ Contrôle par les éditeurs / hors de contrôle pour les tierces parties

*Risque d'utilisation abusive des données collectées*

**Assumez-en la responsabilité !**

# Transformation en entreprises d'analyse de données

## Évolution du modèle des éditeurs académiques : de fournisseurs de contenus à entreprises d'analyse de données

- ▶ Plateformes d'information, bases de données et outils de recherche, fournissant **services & analyses** sur l'ensemble **du cycle de la recherche**
- ▶ Des ensembles de données complètes sont compilés et aux mains de quelques grandes sociétés commerciales
- ▶ Systèmes d'**Information sur la Recherche (CRIS)**
- ▶ Solutions d'**analyse** et de **recherche** basées sur l'**IA**



[Research - Business School Library - LibGuides at Edinburgh Napier University](#)

 <https://www.elsevier.com> :

Elsevier | An Information Analytics Business

Elsevier is a global information analytics company that helps institutions and professionals progress science, advance healthcare and improve performance.

Publish with Elsevier • Elsevier Connect • Elsevier for Reviewers • Books & Journals

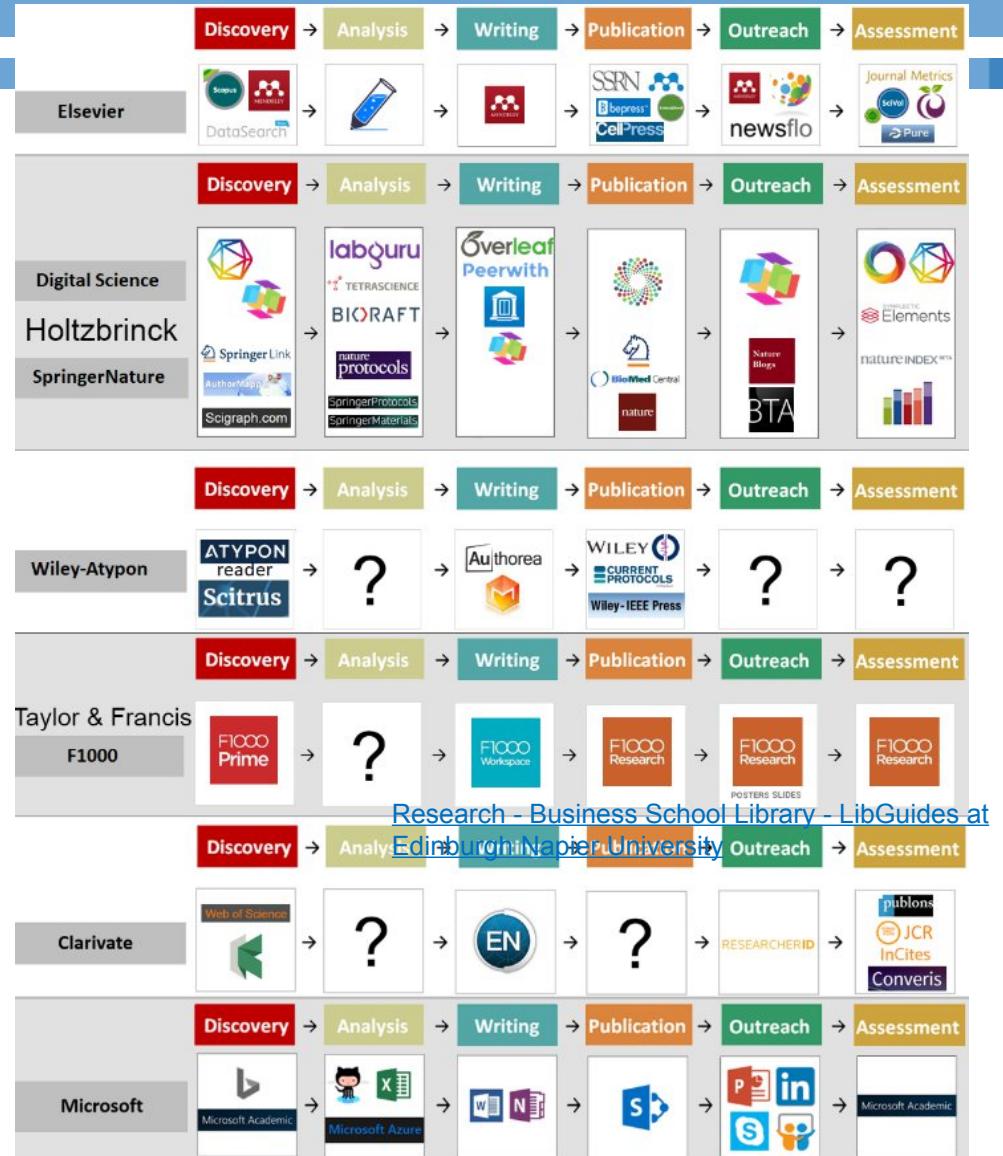
# Transformation en entreprises d'analyse de données

## La souveraineté des données

- ▶ Les « entreprises d'analyse de données » produisent des données sur les **connaissances**, les **développements de la recherche** et les **parties prenantes**.
- ▶ **Verrouillage** dans des **systèmes propriétaires** de toutes les activités des chaînes de production de la recherche
- ▶ Outils de prise de décision et de contrôle aux mains d'entités commerciales

Des services groupés pratiques

Qui exerce le contrôle sur les données ?



# Qu'est-ce que le pistage des données ?

## Données d'utilisation et traces à caractère personnel

- ▶ Capture et **stockage des données des chercheurs** à l'utilisation des services d'information, par ex. Pendant une recherche documentaire
- ▶ **Création de profils personnalisés**, données d'accès et d'usage, durée d'utilisation, téléchargements, mouvements de la souris, etc.
- ▶ **Information partiellement insuffisante** fournie aux utilisateurs, et **sans consentement actif**
- ▶ Les données peuvent être **agrégées et enrichies** avec des données additionnelles issues d'environnements non académiques, pour créer des profils



**Les plateformes académiques collectent, partagent et utilisent les données enregistrées et les informations à caractère personnel de manière souvent opaque et non réglementée**

# Comment les données sont-elles collectées?

## Le pistage par les éditeurs et les tierces-parties



### ► Par les éditeurs : propriétaire (ou *First-Party*)

- Les **cookies propriétaires** : traces directement laissées par les usagers
- Les **pixels espions** permettent une collecte exhaustive et une **analyse de l'usage et de l'intérêt en temps réel**

### ► Par les fournisseurs de services commerciaux, les courtiers en données, les plateformes de réseaux sociaux, etc. : tiers (ou *Third-Party*)

- Les **cookies tiers** : le code Javascript a accès au modèle objet du document (DOM) de la page consultée, peut accéder au texte avec lequel les utilisateurs interagissent, ce qu'ils consultent ensuite, les mots de recherche qu'ils saisissent, etc.
- Le **relevé d'empreintes numériques du navigateur** collecte des **données de localisation et de comportement**

### ► Flux d'enchères, balises, balayage de ports, pistage des pdf et autres techniques

- Collecte en arrière-plan de données sur la **localisation**, les **appareils**, les **données utilisées** ; **géo-repérage** et **ciblage IP**
- Relier **des données à un identifiant en temps réel**; **agrégation de données**
- Collecte de **données biométriques** (par ex. vitesse de frappe, types de mouvements de souris) pour personnaliser des profils usagers malgré l'utilisation de serveurs proxy et de tunnels VPN

# Pourquoi est-ce problématique ?

- ▶ Usage d'une vaste gamme de **méthodes invasives de pistage web** pour pister les comportements des usagers à l'intérieur, à l'extérieur et **au delà de la plateforme** de ressources académiques
- ▶ **Collecte extensive** d'un large éventail de données (d'usages) personnelles
- ▶ Le traitement, le partage, l'échange, la vente de données par des tierces parties et les fournisseurs de services de pistage sont souvent **opaques et non réglementés**, et se trouvent **en dehors de toute influence ou contrôle de la part de l'éditeur**
  - Moteurs de recherche
  - Fournisseurs de solutions de « sécurité », autres fournisseurs
  - Plateformes de réseaux sociaux
  - Agrégateurs de données personnelles et de profils comme Google, Adobe, Cloudflare, New Relic, etc.
  - Courtiers en données (*Data brokers*), acteurs du secteur AdTech, plateformes d'enchères en temps réel (RTB), etc.
- ▶ Les données peuvent être **agrégées et enrichies** avec des données complémentaires d'autres environnements pour **générer des profils**

**Beaucoup sont évalués comme sensibles du point de vue de la vie privée**

**Possibilité de constituer profils individuels et identifiables de chercheurs**

# Les possibles conséquences du pistage et de l'agrégation des données

- ▶ Dangers pour la liberté de la science et celle de la recherche et de l'enseignement
- ▶ Droit à l'autodétermination en matière d'information / interférence avec l'anonymat des chercheurs garanti par la loi sur la protection des données
- ▶ Possible utilisation abusive des données et espionnage scientifique
- ▶ Risques de discriminations, d'exclusions, autres implications sociales



**Réflexion éthique nécessaire, exploration des cadres techniques et légaux & modification des pratiques actuelles**



ALLIANZ DER  
WISSENSCHAFTSORGANISATIONEN



- ▶ Ce n'est pas une question uniquement technique ou juridique, cela touche à l'éthique et aux politiques de recherche
- ▶ La collecte intrusive de données personnelles et le pistage sont incompatibles avec les normes de protection des bibliothèques
- ▶ Importance de la transparence et de la confiance en des infrastructures scientifiques éthiquement intègres

# Comment changer les mauvaises pratiques actuelles ?

Recourir à des **négociations nationales consortiales** pour prendre en charge le problème du pistage de données avec les éditeurs académiques principaux

## Objectifs



- ▶ Transparency sur la collecte, le traitement, le stockage, etc. des données à caractère personnel
- ▶ Traitement juridique approprié de la **gestion des cookies** et **gestion du consentement** (utilisateurs **pleinement informés** et **consistant de façon active**)
- ▶ Réduction du pistage / Par défaut: **PAS DE PISTAGE**
  - Pas d'utilisation de technologies de pistage invasives
  - Pas de pistage par des tierces parties
  - Le pistage **seulement si nécessaire** pour **l'exécution de l'accord, au respect** d'une obligation légale ou d'un **intérêt légitime** (avec la nécessité de l'expliquer)
- ▶ Minimisation des données – réduire au minimum la quantité de données collectées et les acteurs en charge de la collecte

## Le projet DEAL – Conférence des recteurs (HRK)

- ▶ Initié pour négocier des **accords transformants nationaux de type « publish & read »** avec les éditeurs principaux de revues académiques qui permettent la dissémination ouverte des résultats de la recherche allemande
- ▶ Sortir du modèle d'abonnement de type « **paywall** » (barrière payante) en faveur de **services de publication en accès ouvert**
- ▶ Négociations pour le compte de **tous les établissements allemands**, dont les universités, les institutions de recherche et les bibliothèques d'état et régionales



Springer Nature 2020-2023

## Négociations DEAL 2024-2028 en 2022/2023 avec

Elsevier  
Springer Nature  
Wiley

A screenshot of the Wiley website page for the 'Open access agreement for authors at eligible German institutions'. The page features a dark blue header with navigation links: 'Find Journal', 'Prepare', 'Licensing', 'Open Access', 'Publication', and 'Promotion'. Below the header, there is a section titled 'Open access agreement for authors at eligible German institutions' with text about the agreement's terms. A small image shows three people in a library setting with the text 'Freely available to read, download and share'. At the bottom, there is a video player with the caption 'Wenn Sie möchten, dass Ihre Arbeit mehr gelesen, zitiert und geteilt wird, ist Open Access für Sie die richtige Wahl.'

Wiley 2019-2023

# Négociations DEAL pour la protection des données

## ► Groupe de travail « Data Tracking and DEAL » de la DFG

- AWBI, négociateurs DEAL, chercheurs, bibliothécaires, experts techniques et juridiques
- Ancrer les aspects « vie privée » dans les [accords de licence \(DEAL\)](#)
- Garantir la [co-détermination scientifique](#) et la [transparence](#) dans le traitement des données à caractère personnel
- [Sensibilisation au pistage de la science](#) – bibliothécaires, institutions académiques, chercheurs
- Établir des [bonnes pratiques](#) et des stratégies sur la manière de traiter le pistage de données
- [Avis juridique](#) = responsabilité conjointe des bibliothèques et des maisons d'édition en cas de litige

## ► Groupes de travail : DEAL & Éditeurs sur la “Protection des données et le pistage”

- Un groupe distinct pour chaque éditeur
- Discussion, consultation et préparation de la formulation d'un contrat type



**Groupe de travail DFG**

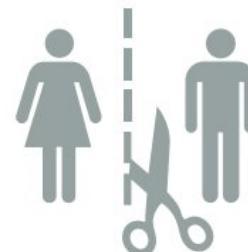
# (Pas de) responsabilité conjointe entre les bibliothèques et les éditeurs

Existe-t-il une **responsabilité conjointe des bibliothèques et des sociétés d'édition** en cas de litige ?

Relation de Contrôleur-à-Contrôleur

vs.

Responsabilité conjointe



Éditeurs

Les avis juridiques diffèrent



DEAL / Avis juridique DFG

- ▶ Sécurité juridique manquante au niveau national & européen
  - Décision de justice ou intervention des agences nationales de protection des données nécessaire
  - En Allemagne : plainte déposée par le Prof. Björn Brembs contre Nomos, Springer et Wiley par le biais de la structure [Gesellschaft für Freiheitsrechte e.V.](#)
- ▶ Compromis & solution alternative : responsabilités réglementées dans l'accord – [clause d'indemnisation](#)

# Aspects importants de la protection des données dans les contrats DEAL

	Elsevier	Wiley	Springer Nature
<b>Responsabilité conjointe</b>	Refusée par tous les éditeurs, solution de remplacement		
<b>Transfert de données (international)</b>	Autorisé conformément à l'art. 44 et suivants du RGPD		Autorisé uniquement après accord de DEAL
<b>Utilisation d'outils de pistage</b>	Autorisé pour des intérêts légitimes		Uniquement si cela est nécessaire à l'exécution du contrat
<b>Récupération de Cookies</b>	Doivent s'abstenir de	Font des efforts raisonnables	Doivent s'abstenir de
<b>Durée de stockage des Cookies</b>	Aucune limitation		1 mois max.
<b>Minimisation / Anonymisation des données collectées</b>	Adresse IP complète : pas plus longtemps que nécessaire (?) ; Atelier commun éditeurs/DEAL		Adresse IP complète : 48 h max.

# Enseignements à retenir

## Nouveauté : une clause de protection des données dans un accord Accès Ouvert transformant

- ▶ Un potentiel d'amélioration !!!
- ▶ D'autres **pays européens et consortiums** reprennent ce sujet dans leurs négociations et utilisent les contrats DEAL comme exemples (par ex. le nouveau contrat SURF-Elsevier – **nouveau « standard »**)
- ▶ Inclure les revendications relatives à la protection de la vie privée dans la **négociation principale** et qu'elles soient soumises aux **conditions commerciales** de l'accord et constituent un élément essentiel de son **cadre juridique et éthique**
- ▶ Nécessité d'instaurer un **DIALOGUE** avec les éditeurs académiques & les fournisseurs
  - **Ateliers DEAL** avec les éditeurs sur les sujets liés à la confidentialité des données
  - Ateliers prévus initialement (et contractuellement) pour le seul thème de la conservation des adresses IP, mais qui vont en fait bien au delà
- ▶ **Dialogue international et sensibilisation** des éditeurs scientifiques / Association STM et tous les fournisseurs d'informations académiques

# Incitation & intérêt pour les éditeurs académiques

**TRANSPARENCE** à toutes les étapes du cycle de vie des données (d'usage) personnelles

- ▶ **Exigences particulièrement élevées** pour les normes de protection de la vie privée pour les chercheurs et les universitaires
- ▶ « liberté individuelle d'utilisation » / **anonymat d'un chercheur** sur une plateforme d'information académique
- ▶ Les éditeurs scientifiques doivent être **dignes de confiance, fiables et intègres**
- ▶ Un « **Code de conduite** » volontaire et éthique pour aller **au delà de la seule conformité** et protéger la Science et les chercheurs
- ▶ Un **environnement académique de recherche et de lecture sûr et respectueux de la vie privée** sans surveillance
- ▶ **MINIMISATION** des données personnelles dans tous leurs aspects
- ▶ Transition d'un système de pistage intensif à une plateforme d'informations scientifiques **respectueuse de la vie privée et digne de confiance**

## ▶ Idées & Suggestions

- Par défaut : **Pas de pistage**
- **Un vrai choix** pour les usagers : opter pour des « **services sur mesure** » par ex. Historique de recherche, réseaux sociaux spécifiques, etc.
- **Pas de pistage** par des tierces parties
- « **liste positive** » de fournisseurs de services de pistage / tierces parties
- Un « **comité de protection des données** » intégré sur le plan éthique
- **Évaluation critique** de la nécessité et de la conformité de tous les fournisseurs et tierces parties de **technologies de pistage**
- Des **normes éthiques élevées** en matière de protection de la vie privée
- ...

# Recommandations & Perspectives

- ▶ Contrôle permanent du pistage, de son étendue et des technologies employées
  - *Allianz Task Force Data Tracking – document prévu pour la fin 2025*
- ▶ Appliquer les **précautions techniques** au niveau individuel et institutionnel
- ▶ Sensibiliser au thème du pistage des données dans la recherche
- ▶ Développer l'**Accès Ouvert** et créer des **infrastructures non commerciales et protectrices de la vie privée**
- ▶ Approche **internationale** sur le pistage et la souveraineté des données



**Le « pistage » et l'*analyse de la science* peuvent être utiles  
si l'utilisation des données se fait  
dans le **respect de la vie privée**, en **tenant compte des risques**  
et sous la **souveraineté de la science****



# Merci pour votre attention !

## Plus d'informations

- ▶ Sur la DFG : [DFG - Deutsche Forschungsgemeinschaft](https://www.dfg.de)
- ▶ Sur la synthèse de l'AWBI : [Data tracking in research](https://www.awbi.de/de/veroeffentlichungen/whitepaper-data-tracking-in-research)
- ▶ Sur l'article DEAL sur la clause « vie privée » dans les négociations : <https://zenodo.org/records/13837664>
- ▶ sur DEAL: [About DEAL - DEAL Konsortium \(deal-konsortium.de\)](https://deal-konsortium.de)