

1

00:00:42,560 --> 00:01:31,000

Maximilien Petit

Alors, Joachim Schöpfel, tu es maître de conférence en sciences de l'information et de la communication à l'Université de Lille, et plus précisément au sein du laboratoire Gérico pour groupe d'études et de recherche interdisciplinaire en information et communication. Tu as œuvré, entre autres, au sein de l'Atelier national de reproduction des thèses et également au sein de l'INIST, l'Institut de l'information scientifique et technique du CNRS. Tu travailles précisément sur des questions qui nous motivent dans le podcast C'est pas donné, c'est-à-dire la communication scientifique, la science ouverte, les données de la recherche et tous les dispositifs de l'IST. J'ai eu par exemple le plaisir de te lire dans de nombreux articles et enquêtes et en particulier celle qui concerne l'Université de Lille 3 en 2015, qui est éclairante à plus d'un titre. Joachim, bonjour et merci beaucoup.

2

00:01:32,180 --> 00:01:34,020

Joachim Schöpfel

Bonjour, merci pour l'invitation.

3

00:01:35,300 --> 00:02:16,860

Maximilien Petit

Mireille Brenel, tu es chargée de mission données de la recherche et tu exerces au sein de la cellule données et appel à projets du pôle numérique, recherche et science ouverte de la Dibiso, la direction des bibliothèques de l'information et de la science ouverte de l'Université Paris-Saclay. Tu es donc, pour résumer, en charge de l'offre des services en place pour les chercheurs sur toutes les questions de gestion des données de recherche. Et à ce titre, tu pilotes brillamment un atelier de la donnée, Data Saclay, et avant cela, tu as coordonné au sein de l'université le circuit des thèses pour 14 établissements. Nous avons notamment travaillé ensemble, et c'était précieux, là aussi, sur une enquête consacrée aux données de la recherche. Mireille, bonjour et merci beaucoup.

4

00:02:17,260 --> 00:02:18,980

Mireille Brenel

Bonjour et merci à toi, Maximilien.

5

00:02:19,780 --> 00:02:43,000

Augusto Britto

Alors bonjour, bonjour à vous deux. J'aimerais bien commencer la discussion en te posant une question. Joachim : en termes de culture de la donnée, est-ce qu'il y a, selon toi, des avancées ? Est-ce que les chercheurs en sciences humaines et sociales osent enfin prononcer l'expression « données de la recherche » ou est-ce qu'il subsiste encore une problématique épistémologique, sémantique ?

6

00:02:43,840 --> 00:04:34,040

Joachim Schöpfel

Ça fait trois questions, en fait. C'est pas facile, c'est difficile à donner une réponse de manière générale. Culture, culture de la donnée. Moi, je l'entends comme des pratiques communes et des valeurs communes. Il est certain qu'en SHS, il y a des pratiques très différentes, valeurs peut-être beaucoup moins, valeurs éthiques surtout, parce que ça veut dire faire la bonne recherche. Mais en ce qui concerne les pratiques, évidemment, elles ne sont pas les mêmes entre les archéologues, les psychologues, etc. Donc, quand on a commencé à s'intéresser aux données sur notre campus SHS à Lille, on a bien constaté que tout le monde ne parle pas de données. Pour beaucoup de chercheurs, « données », c'est un terme abstrait, qui ne veut rien dire. En fait, il s'agit de résultats d'enquête. Il s'agit de résultats de fouilles en archéologie, qu'il s'agit de corpus et de documents en histoire ou littérature, des références, etc. Donc, je dirais qu'effectivement, oui, il y a une question sémantique par rapport à la donnée. Ceci étant, je n'ai jamais trouvé personne en sciences humaines et sociales, aucune chercheuse, qui n'ait pas de données, de ce qu'on appelle « données », sauf que vous n'appelez pas nécessairement ça de cette manière-là.

7

00:04:34,240 --> 00:05:12,000

Maximilien Petit

C'est vrai que, et notamment autour du travail de l'enquête avec Mireille, on a rencontré pas mal de chercheurs en SHS. C'est vrai qu'on n'a pas rencontré non plus de chercheurs qui nous disent « je n'ai pas de données ». Par contre, il y en a beaucoup qui nous disent « mais ce que j'ai, ce n'est pas intéressant ou ça n'a pas de valeur, donc pourquoi je travaillerais dessus ? » Ça, c'est quelque chose qui, moi, m'avait frappé. Il y avait parfois des réticences à travailler, à évoquer, ne serait-ce que des questions de méthodologie, « voilà comment je travaille, voilà mon corpus », et de tout de suite abrégé la question en disant « mais de toute façon, ça n'intéresse personne, pourquoi j'en parlerais plus ? » Ça, c'est quelque chose qui m'avait, moi, en tout cas, étonné.

8

00:05:14,020 --> 00:05:42,540

Augusto Britto

Alors, je vous pose une question à vous deux : lorsqu'il s'agit de SHS, le plus souvent, les enquêtes indiquent que le manque de moyens, moyens informatiques et personnels supports de proximité, est le principal obstacle à une bonne gestion des données de recherche. Il est peu probable que cette réalité s'inverse. De fait, comment faire ? Est-ce qu'il y a des stratégies qui permettent de pallier à ce manque de moyens ?

9

00:05:42,540 --> 00:07:27,000

Mireille Brenel

Alors, en effet, c'est des obstacles qui sont souvent pointés. On a

pu le remarquer dans l'enquête qu'on a fait au sein de l'université Paris-Saclay il y a quelques années, où souvent les chercheurs disaient qu'ils se sentaient seuls face à la gestion de leurs données, qu'il n'y avait pas forcément de personnel pour les accompagner. Et du coup, ça permet de mettre un éclairage sur le fait qu'il y a cette perception des chercheurs qui ont l'impression de ne pas avoir de services adaptés en face, de personnes contacts. Et en fait, ce que ça révèle, c'est qu'il y a peut-être plutôt un souci d'identification et de connaissance de ces services mis à leur disposition. Et donc, c'est pour ça qu'il y a un axe fort à développer, justement, pour les personnels IST, les personnels d'accompagnement, les ateliers de la donnée, sur la communication de nos services et de nos actions, justement, pour que les chercheurs s'en saisissent. Et il faut aussi que nous, on porte à la connaissance des communautés les outils qui sont mis à leur disposition. Et ce, même au niveau national, il y a quand même pas mal d'outils qui existent pour accompagner les chercheurs, notamment dans la rédaction de leur plan de gestion de données, dont ils n'ont pas forcément connaissance. Non, ils ne sont pas seuls pour gérer leurs données. Mais en effet, il y a aussi un besoin de moyens. Mais surtout, une demande d'outils simples des chercheurs pour leur faciliter la vie. Et une fois de plus, ces outils existent. Mais je pense que c'est aussi aux institutions de déployer ces services, de mettre à disposition des outils adaptés. Ce qui peut justement être impulsé par une politique des données claires, qui peut être prise à l'échelle d'une université ou d'un organisme.

10

00:07:28,580 --> 00:08:57,500

Maximilien Petit

Et moi, ce que je vois au quotidien, en tant que documentaliste dans un laboratoire, alors effectivement, il y a ce besoin d'avoir des outils, d'avoir des logiciels, d'avoir des dispositifs d'appui. Et le plus souvent, en tout cas, moi je suis confronté à des chercheurs qui connaissent, ils savent très bien que ça existe. Par contre, alors ils n'iront que si déjà, ils en ont strictement besoin, à l'instant T, pour un projet, pour une démarche bien particulière. Et bon, il faut que je trouve une autre formule, parce que c'est un peu péjoratif. Mais ils ont envie d'abord qu'on leur montre vraiment comment ça marche. Pas prendre par la main, parce que ce n'est pas du tout le cas, mais ils ont vraiment besoin d'avoir un accompagnement humain, déjà vers ces logiciels, vers ces outils, etc. Et ensuite, ils s'en saisissent. En tout cas, moi, ce que je vois, ce qui n'est pas forcément le cas partout et dans les autres disciplines, je travaille particulièrement avec des historiens, des littéraires, des civilisationnistes, en tout cas. Ça fonctionne comme ça, avec eux. Dans le cadre de mon labo. Et aussi, on a parlé tout à l'heure d'équipement, d'instrumentation, parce que c'est vrai qu'on n'a aucun mal à financer, entre guillemets, à financer de l'équipement. Besoin d'un serveur, besoin d'une machine, besoin d'un scanner, on y arrivera. Par contre, quand il y a besoin d'avoir un personnel pour gérer l'équipement, le mettre en relation avec les chercheurs, c'est un peu plus compliqué. J'ai le sentiment que là, dans toutes les stratégies ministérielles, les questions de données,

c'est moins pris en compte. C'est peut-être un petit peu dommage.

11

00:08:58,040 --> 00:11:45,840

Joachim Schöpfel

Oui, comment dire ? En général, je pense que la mutualisation, de toute manière, c'est la stratégie. C'est de partager des ressources, surtout pour les unités qui n'ont pas de personnel, comme la mienne, et qui ont quand même plus de mal d'acheter des équipements. On n'en a pas. C'est partagé au niveau du campus. Mais je dirais de manière plus pragmatique. De notre côté, effectivement, comme tu dis, on cherche tout ce qu'il y a et on essaye de travailler avec d'autres qui ont cet équipement, d'autres laboratoires, d'autres universités. Et puis, pour le reste, je dirais comme souvent en SHS, quand il n'y a pas de moyens, on fait de "l'artisanat". Je ne parle pas de bricolage. Je parle vraiment du sur-mesure avec les moyens de bord et pas quelque chose qu'on peut implémenter ailleurs. Donc, sur-mesure, j'en ai suffisamment vu dans ce domaine-là. Dans ce domaine de la donnée, les laboratoires se débrouillent. Ce n'est peut-être pas normalisé, ce n'est pas standardisé, mais ça suffisait jusqu'à présent avec des bases de données développées en interne qui fonctionnent comme une sorte de silo. Ce n'est pas bon, mais au moins, ça fonctionne. Que dire d'autre ? Pour le reste, ce que tu disais. Je rappelle que c'est connu, quand même, depuis des décennies, en ce qui concerne les nouvelles technologies, on les accepte si elles sont faciles à prendre en main et si elles sont utiles immédiatement. Donc, le reste, il y a quand même aussi au niveau national des outils pour les SHS qui sont compliqués, qui n'ont rien à voir avec ce que ça pourrait être. Il faut suivre les procédures, il faut avoir un accompagnement, éventuellement une formation. Ce n'est peut-être pas nécessaire. Je vois comment on prend en main aujourd'hui Recherche Data Gouv, ce qui n'a rien à voir, c'est facile. On n'a pas besoin de comprendre grand-chose pour déposer et décrire des données sur cette plateforme.

12

00:11:47,340 --> 00:12:44,980

Mireille Brenel

Alors, c'est vrai que c'est facile, mais je pense qu'il y a quand même un besoin d'accompagnement qui va permettre de faciliter la tâche, justement, des chercheurs. Et on voit que c'est quelque chose qu'ils apprécient beaucoup. Notamment, ça nous arrive, par exemple, pour un chercheur qui veut déposer la première fois sur Recherche Data Gouv, de faire le premier dépôt avec lui pour bien lui expliquer chaque champ de métadonnées, ce qu'il faut renseigner, etc. Bien sûr, il y a des rapports de curation, il y a des documents, etc. Mais de le voir et d'avoir un personnel à côté qui explique, ça va ouvrir la voie pour le reste. Et c'est quelque chose, je trouve, qui vaut pour pas mal d'outils. D'avoir, tout simplement, des démos, même pour les PGD, pour DMP Opidor, pouvoir montrer vraiment rapidement, en 15-30 minutes, comment marche l'outil. Ça permet justement aux chercheurs de mieux les assimiler

et les utiliser peut-être plus facilement.

13

00:12:46,180 --> 00:13:49,700

Maximilien Petit

Parce que justement, Mireille, tu parles de curation, etc. On a parlé aussi un petit peu de métadonnées. Mais c'est quoi le cœur de métier quand on parle des personnels IST, que ce soit des bibliothécaires, des ingénieurs, des ITA, qui sont sur toutes ces questions-là ? On parle de curation, de métadonnées, le travail qui est fait sur Recherche Data Gouv. Donc est-ce que c'est vraiment cet aspect curation et métadonnées qui constitue le cœur de métier et qui peut constituer peut-être le cœur d'un accompagnement ? Est-ce qu'on dépasse parfois ce cœur de métier-là ? Et finalement, quand on le dépasse, comment on fait ? Est-ce que parfois on n'empiète pas sur le terrain de l'informatique, le terrain du juridique, le terrain de la valorisation ? Je parle à l'échelle des universités et des établissements. Et toutes ces compétences-là, finalement, est-ce qu'il faut qu'on les agrège ou est-ce qu'il faut qu'on délègue ? Est-ce qu'il faut qu'on aille chercher ces gens-là en réseau, en collaboration ?

14

00:13:50,600 --> 00:15:25,340

Mireille Brenel

C'est vrai que souvent, quand on pense à la plus-value des personnels IST dans l'accompagnement à la gestion des données, on va souvent penser à ce que tu disais, Maximilien, au travail de curation, de choix des métadonnées, parce que ça s'apparente vraiment au cœur de métier des personnels IST. En fait, on va souvent se dire que référencer, indexer des données, c'est un parallèle avec le catalogage des documents qu'on peut faire, le catalogage des monographies, etc. Mais les métiers évoluent. Le métier de bibliothécaire évolue depuis un certain nombre d'années, on l'aura bien constaté. Aujourd'hui, les personnels IST sont tout à fait à même d'apporter des éléments, des conseils, au moins de premier niveau, sur l'ensemble du cycle de vie de la donnée, voire même d'être, pour certains, des experts de la donnée. On voit qu'aujourd'hui, des bibliothécaires deviennent des data librarian. Après, quand il s'agit de questions beaucoup plus spécifiques, comme tu le disais, de type informatique ou juridique, par exemple, je pense que les personnels IST ont tout à fait la compétence d'avoir un premier vernis, mais c'est là que, justement, faire réseau prend tout son sens. Justement, de pouvoir compter dans un réseau, dans des ateliers de la donnée, des services informatiques, des mésocentres, des services juridiques, pour justement pouvoir bénéficier de leur expertise, soit pour s'en nourrir, soit pour renvoyer nos chercheurs, vers ces experts.

15

00:15:27,180 --> 00:15:44,000

Maximilien Petit

Joachim, dans certains de tes articles, tu évoques, entre guillemets, un discours d'injonction idéologique, qui est parfois

imposé aux chercheurs, sur toutes ces questions de gestion de la donnée. Est-ce que tu peux expliquer un petit peu ce que tu entends par discours d'injonction idéologique ?

16

00:15:45,040 --> 00:19:22,460

Joachim Schöpfel

Pour commencer, je vais utiliser ce terme-là. Pas d'une manière...péjoratif, donc c'est descriptif. Et c'est par rapport à une manière de parler des données qui se distinguent d'autres manières. Idéologique au sens que ça relève des idées. Sachant que, quand on a mené nos interviews et les enquêtes autour des données, on a constaté régulièrement trois types de discours. Et donc, les discours qui relèvent des idées, c'est un discours. Mais il y a deux autres. Qui relèvent des idées, ça veut dire que quelqu'un qui dit...C'était par rapport aux publications, pas aux données, mais c'est du même genre. C'est bien de partager les résultats de la recherche, parce que c'est bien pour la société, ça correspond à mes valeurs, l'ouverture, c'est ce qu'il faut, etc. Donc, ça, c'est des idées que j'ai décrites comme idéologiques. En l'écrivant, je me suis dit, est-ce que ça va être bien perçu ou pas ? Mais je l'ai fait, sachant que, juste à côté, je crois que c'est marqué philosophie et/ou politique. Donc, ce qui va dans le même sens. D'autres manières, c'était au sein des laboratoires qu'on a constaté ça. D'autres manières de parler du partage des résultats. Le premier, c'est parfaitement utilitaire. Il faut partager, surtout là, par rapport aux publications, au niveau des données. On n'a pas encore cette expérience, de manière plus généralisée. Il faut partager, c'est bien de partager les publications, parce que ça a un impact. On sera plus visible, meilleure réputation, plus souvent cité, etc. Donc, ça, c'est parfaitement utilitaire. Ça n'a rien à voir avec des grandes idées philosophiques ou politiques ou idéologiques. Ce ne sont pas des idées. Je veux obtenir quelque chose, un pacte. C'est comme si je signalais...toutes mes publications sur Facebook ou sur Twitter, ou je ne sais pas où. Donc, c'est utilitaire. Et donc, il y a un autre, ce que j'ai appelé, enfin, ce que nous, on a appelé, dans notre équipe, "pragmatique". C'est tout simplement dire, ok, dans notre environnement, il faut y aller. Donc, c'est à cause de l'ANR, c'est à cause des programmes cadres de l'Union européenne. Le ministère souhaite...l'université le souhaite. Dans certaines universités, comme Aix-Marseille, maintenant, il y a un système bonus-malus. Donc, pas pour les données, pour l'instant, mais si je vois bien, mais pour les publications sur HAL. Donc, il faut y aller. Donc, c'est mieux pour nous. Donc, ça non plus, ça ne relève pas des idées. Injonction, ce n'est pas une obligation. Donc, c'est une incitation forte. Voilà.

17

00:19:23,440 --> 00:19:37,820

Augusto Britto

Alors, pour penser un peu plus au travail des IST, est-ce qu'il y a une stratégie efficiente pour assurer les services sur les questions de gestion des données de recherche pour les sciences humaines et

sociales ? Et si la bonne stratégie est de faire réseau, précisément, il faut faire réseau comment et avec qui ?

18

00:19:45,460 --> 00:19:51,520

Mireille Brenel

Alors, pour moi, la meilleure stratégie, elle va se développer en trois points. Donc, tout d'abord... le premier point, c'est de bien connaître nos communautés, leurs spécificités, leurs besoins, pour leur offrir, une fois de plus, une offre de services adaptée et cohérente. Et ensuite, le deuxième axe, justement, c'est de faire réseau, en effet. D'accompagner sur l'ensemble du cycle de vie de la donnée, ça implique de mutualiser différentes compétences. Il faut donc faire réseau autour de différents services et personnels des universités. Justement, il y a un gros enjeu autour de tout ça, c'est qu'on apprenne à tous travailler ensemble, à ce qu'on mutualise nos compétences, à ce qu'on arrive à articuler des services autour de ces multi-compétences et de vraiment créer du lien et une dynamique propice à tout ça. Donc, ça, c'est tout l'objectif des ateliers de la donnée, justement. Et c'est comme ça qu'on va réussir à travailler en réseau autour...des personnels IST et de bibliothèques, des services d'appui à la recherche et de montage de projets de recherche aussi, autour des services juridiques, mais aussi avec les archivistes, les services informatiques et notamment les services du type Mésocentre, qui sont aussi des outils et des relais assez précieux dans le cadre des ateliers de la donnée.

19

00:21:17,200 --> 00:22:43,640

Maximilien Petit

Je tiens à dire, pour le voir au quotidien, le fait de ne plus fonctionner en silo, mais bien de fonctionner en réseau et de connaître, en tant que professionnel IST et en tant que IGE chargé de ressources documentaires, documentaliste, le fait de connaître les personnes, les personnes et leurs compétences et puis leurs services. De dire, bon, là, je dois répondre à un besoin d'un chercheur à un instant T, avec des questions qui, moi, me dépassent, parce que ça relève d'une technique de recherche que je ne maîtrise pas, d'une discipline que je ne connais pas, d'un outil que je ne connais pas. Mais maintenant, connaître les noms, connaître ces gens qui ont ce...leur spécialité, et de pouvoir les contacter et de travailler ensemble, ce n'est pas si facile que ça, parce que c'est super, c'est super au quotidien, et je trouve qu'on le fait de mieux en mieux. Et je pense que le défi, il est là, parce que hors réseau, plusieurs années, quand on travaillait sur les questions de données et qu'il n'y avait pas ces ateliers, qu'il n'y avait pas ces réseaux, c'était un peu, chacun se débrouille dans son coin. Chacun fait, on a évoqué tout à l'heure, des solutions qui fonctionnent à l'instant T, et c'est vrai, parce qu'il faut répondre, il faut répondre à un besoin, et on fait avec les moyens du bord. Et effectivement, je suis d'accord aussi, c'est un travail qui a de la valeur, qui est difficile de mutualiser, mais qui a quand même énormément de valeur. C'est un travail qui, parfois, mérite qu'on en

parle aussi, quand même. Et maintenant, le fait de pouvoir travailler avec des gens avec des compétences beaucoup plus larges, et des gens qu'on connaît, et plus seulement des services un peu désincarnés, c'est quand même très précieux.

20

00:22:44,540 --> 00:24:01,000

Mireille Brenel

Oui, je suis tout à fait d'accord avec toi, Maximilien, ça a vraiment été une plus-value, notamment au sein de notre atelier, d'identifier les services, comme tu le disais, on le connaissait déjà, mais de travailler...aussi, entre professionnels, en tant qu'individus identifiés, et de vraiment travailler en équipe, ça permet également de construire des services aussi ensemble, d'avoir des idées de services, ce qui est assez précieux, et qui nous nourrit, nous, énormément, mais bénéficie d'autant plus pour les communautés, ce qui est aussi très important, parce que tout ce travail qu'on fait, c'est avant tout pour nos chercheurs. C'est vrai qu'il s'apporte une plus-value. Et juste pour terminer sur la question, le troisième point pour définir une stratégie efficiente, il me semble que c'est aussi tous les enjeux de communication. Justement, une fois qu'on a monté l'offre de services et qu'on est prêt à dispenser tout ça aux chercheurs, il faut faire connaître ces services, permettre aux communautés de nous identifier, de justement savoir qu'elles ne sont pas seules face à la gestion de leurs données, et qu'il y a justement des personnels compétents qui sont disponibles pour les orienter et pour leur faciliter la tâche.

21

00:24:03,000 --> 00:24:56,420

Maximilien Petit

Parce qu'on parle du cadre institutionnel, donc organismes, financeurs, ministères, on a parlé tout à l'heure d'incitation et même de forte incitation. Il y a tous ces mécanismes-là que je vois aussi au quotidien, de sensibilisation, d'incitation, et puis ensuite d'obligation, sur toutes les questions qui concernent les données, la science ouverte. Il y a un exemple très récent, en tout cas en Ile-de-France, il est désormais presque obligatoire d'utiliser HAL pour la collecte des notices bibliographiques dans le cadre de la dernière évaluation HCERES. Est-ce qu'on va observer selon vous le même mécanisme pour les données ? Est-ce que les unités de recherche devront, d'ici cinq ans, déposer sur Recherche Data Gouv des jeux de données pour remplir des dossiers d'évaluation ? Et si c'est le cas, est-ce que, selon vous, c'est une bonne stratégie ?

22

00:24:56,420 --> 00:29:45,600

Joachim Schöpfel

Je ne pense pas. Je ne pense pas que ça va aller dans ce sens-là, pour plusieurs raisons. La première raison, c'est que je crois qu'il y a... Enfin, peut-être que c'est juste dans ma tête, mais il y a un positionnement de Recherche Data Gouv qui n'est pas clair. Moi, j'ai entendu au ministère...On a parlé des trous dans la raquette. C'est-



à-dire que nous, les chercheurs, on a nos propres infrastructures, nos plateformes, nos entrepôts. Et si jamais on n'en a pas, comme pas malheur, dans notre domaine des sciences de l'info, il n'y a pratiquement rien. Et donc, à ce moment-là, on peut aller là. On devrait aller là. Pour ne pas aller, je ne sais pas, sur Figshare, par exemple. De l'autre côté, j'ai l'impression que ça devient... C'est peut-être pas une vocation, mais je pense que ça a le potentiel, disons ça comme ça, de devenir l'entrepôt par défaut. On met là, sauf si on a quelque chose ailleurs, un entrepôt plus intéressant ou plus visible. Ce qui est toujours reproché par une partie des chercheurs, par exemple, par rapport à HAL, qui disent que HAL, c'est franco-français. Donc, on préfère aller ailleurs, par exemple, sur arXiv. Donc, je ne sais pas si ça va être un problème aussi pour Recherche Data Gouv mais ce positionnement entre le trou dans la raquette et par défaut, il faut aller là, ou on devrait aller là. Donc, je crois qu'il faudrait être clair là-dessus. Et notamment dans notre discipline, la SHS, je crois que ce n'est pas clair. Par rapport à l'existence aussi de Huma-Num et de Nakala. Mais bon, ça, c'est une partie. Une autre partie. Contrairement aux publications, je crois que déposer des données, ce n'est pas aussi simple que ça. Ça a été abordé tout à l'heure, la question des données à caractère personnel. Il y a d'autres problèmes. Quand je regarde le livre, super, d'Agnès Robin, sur les aspects juridiques, c'est le meilleur livre qui a été écrit en français jusqu'à présent sur les données. À mon avis, ça restera aussi. Quand je regarde, je ne sais pas combien de centaines de pages sur des questions juridiques et légales. Je les ai tous lus, ces pages-là. Je crois que rien que pour cette raison-là, ce n'est pas aussi simple. Ce n'est pas aussi simple non plus de comparer, je ne sais pas, un laboratoire avec un gros équipement qui produit des tonnes de données et à côté, un petit laboratoire en sciences de l'éduc ou, je ne sais pas, en communication avec des enquêtes, des interviews dont une partie ne peut pas être déposée, une partie peut l'être, mais c'est... Comment on veut comparer ? Comment on veut comparer ? Donc, pour cette raison-là... Je pense que ça ne va pas aller dans le même sens que pour HAL et puis cette injonction de tout mettre là, de tout signaler là. Mais je pense que ce qui va se mettre en place de plus en plus, c'est la nécessité de justifier, d'expliquer ce qu'on fait avec les données. Où sont les données ? Pourquoi elles sont là ? Pourquoi elles ne sont pas là ? Pourquoi elles ne sont pas en ligne ? Pourquoi il y a un accès restreint, etc. Ça, oui. Et je pense lié à ça, et j'attends que ça se mette en place. Je crois que c'est la partie la plus compliquée de tout ce projet de l'écosystème Recherche Data Gouv. Je ne sais pas comment ça s'appelle maintenant, mais c'est cette espèce de méta-catalogue qui doit se greffer dessus. C'est-à-dire que... Où ? Mais je crois que ça va être la partie la plus compliquée de tout. Où ? Je crois que c'est l'idée du ministère et du CNRS, où on a une sorte de vue d'ensemble sur toutes les données produites par la recherche française. Que ce soit dans Zenodo, que ce soit sur figshare, ou Mendeley, ou Recherche Data Gouv, ou ailleurs. C'est ça, je crois, qu'ils attendent. Mais je pense que...

00:29:45,600 --> 00:29:52,600

Maximilien Petit

Techniquement, un énorme défi. Parce que quand je vois les historiens, souvent on dit, ils ne déposent pas sur Nakala, ils ne déposent pas sur Recherche Data Gouv. Mais pourtant, ils ont des bases de données, avec des données extrêmement enrichissantes et travaillées. Mais effectivement, des bases de données confidentielles...

24

00:30:05,280 --> 00:30:12,100

Joachim Schöpfel

Cette partie-là, pour moi, scientifiquement, n'a pas de sens. Mais ça fait sens du point de vue monitoring. C'est-à-dire de savoir ce que les chercheurs font avec les données, d'avoir une vue d'ensemble. Mais on va voir. Vous allez voir. Quand ça se mettra en place. C'est déjà assez compliqué pour les publications d'avoir une vue d'ensemble.

25

00:30:29,680 --> 00:30:34,100

Maximilien Petit

Oui. Et toi, Mireille, qu'est-ce que tu en penses de tout ça ?

26

00:30:36,620 --> 00:33:33,660

Mireille Brenel

Ce que j'en pense, c'est que le contexte actuel, notamment des politiques impulsées par les financeurs, par les différents plans nationaux pour la science ouverte, incitent justement fortement à partager les données de recherche. Il en va de même pour les éditeurs, qui demandent de plus en plus la publication des données en lien avec les articles. Donc on voit qu'il y a vraiment une dynamique en place, dans le sens d'une ouverture de plus en plus importante des données. Alors, là où je m'interroge, c'est que peut-être on va, à plus ou moins long terme, vers une incitation forte au dépôt des données pour remplir les dossiers d'évaluation. Peut-être. Je ne sais pas exactement. En tout cas de mon point de vue, ça ne me paraît pas complètement incongru. Mais si c'est le cas, une fois de plus, il faut accompagner et soutenir les communautés dans ce sens. Ne pas les laisser seules face à ce processus d'ouverture. Et on voit que les moyens aussi mis en place ces dernières années vont dans ce sens, notamment avec la structuration de services d'accompagnement à la gestion des données de plus en plus importantes dans les universités. Et justement, avec la mise en place de l'écosystème Recherche Data Gouv, qui œuvre également pleinement dans ce sens. Et donc, si ce virage est amené à être opéré, je crois qu'il y aura des infrastructures, des services d'accompagnement prêts à soutenir les équipes de recherche. Mais là où je rejoins Joachim aussi, ce qu'il faut faire attention, c'est que Recherche DataGouv n'a pas vocation à collecter tous les jeux de données produits, contrairement à HAL. Et je pense que, oui, il y a plus un intérêt à peut-être...plus prendre en compte tous les aspects de données dans les dossiers d'évaluation, expliquer, comme

disait Joachim, pourquoi on a ouvert ou pas, pourquoi on a partagé ou pas, justifier ses choix. Ce qui reviendrait peut-être à associer le PGD, enfin, je ne sais pas, au dossier d'évaluation. Mais une fois de plus, il faudrait réassurer les communautés de garder en tête l'adage "aussi ouvert que possible et aussi fermé que nécessaire". D'autant plus en sciences humaines et sociales, où je pense qu'il faut assez les rassurer là-dessus, sur le fait que tous les jeux de données n'ont pas vocation à être partagés. Par contre, je crois qu'il y a aussi un enjeu pour les institutions à recenser leurs jeux de données. Et il est probable que le module catalogue de Recherche Data Gouv, dont parlait Joachim, ait pour objectif de tendre à l'exhaustivité dans les prochaines années, justement. Et ça pourra justement être un outil très pratique pour les établissements qui voudront avoir une vitrine pour leurs données, tout comme HAL est une vitrine pour les publications d'une institution.

27

00:33:34,660 --> 00:33:48,860

Maximilien Petit

Justement, pour rester sur ces questions de plan de gestion de données, moi, ce qui m'intéresse de savoir, c'est si les chercheurs en sciences humaines et sociales sollicitent avant tout des recommandations, c'est-à-dire on a besoin qu'on relise leur plan de gestion de données, ou est-ce qu'ils espèreraient plutôt une sorte de service clé en main, qui n'est pas seulement de la relecture, mais qui est vraiment la prise en charge d'écriture sur certaines parties des plans de gestion de données. Parce que plus souvent, peut-être, dans les établissements, on en est au stade relecture, pour tout un tas de questions, questions de légitimité aussi, peut-être. Est-ce qu'il s'agit de passer à une autre étape à ce niveau ?

28

00:34:13,340 --> 00:35:15,040

Mireille Brenel

Alors, en effet, tu as raison pour l'instant. Enfin, pour l'instant. Pas que pour l'instant, mais... On propose surtout des recommandations, des conseils, des relectures. Mais il arrive, et on sent bien que des fois, les chercheurs aimeraient bien que le service aille plus loin, et que, pourquoi pas, on rédige pour eux leur plan de gestion de données. Après, moi, je ne suis pas tout à fait d'accord avec ce principe, et je ne pense pas que ce soit notre rôle. Pour moi, le plan de gestion de données, c'est souvent vu comme un document administratif, mais c'est avant tout un travail scientifique. Et pour moi, seuls les porteurs de projets sont en mesure de réaliser ce travail-là, et l'intérêt serait très limité si quelqu'un, en dehors du projet, se mettait à rédiger le plan de gestion de données. Donc, en tant que personnel IST, pour moi, notre rôle, c'est de les accompagner, les conseiller, leur donner des recommandations, et les orienter vers des outils adaptés ou des experts compétents, mais pas rédiger le plan de gestion de données.

29

00:35:15,040 --> 00:37:04,940

Joachim Schöpfel

Je serais plus nuancé sur un point. Tu parles comme si tous les personnels IST seraient au même endroit. Mais je pense qu'un documentaliste ou un ingénieur de données qui est dans un laboratoire et qui est intégré dans les projets, en SHS c'est quand même courant, à ce moment-là, il n'y a aucun problème pour que ce soit cette personne-là à rédiger un plan de gestion. C'est complètement différent si c'est une équipe du SCD, et bien ça, je suis d'accord. De toutes manières, pour avoir toutes les informations pour rédiger un plan de gestion d'un projet qu'on ne connaît pas, ça prend plus de temps de discuter avec les gens ou de faire des allers-retours que de l'écrire. Mais là où les personnels font partie des projets, à mon avis, c'est tout à fait un travail de professionnel. Il n'y a pas besoin d'avoir un chercheur. Donc, on a monté un projet dans le cadre de l'appel à projets sur la science ouverte. Là, on a clairement indiqué que dans ce projet-là, il y a une personne, c'est une bibliothécaire, qui est la responsable des données dans le cadre de ce projet. Et le plan de gestion sera rédigé par elle. Et pas la personne responsable scientifique du projet. Donc, on a réparti le travail comme ça.

30

00:37:06,680 --> 00:37:10,120

Augusto Britto

Donc, Joachim, tu as évoqué la formation aux doctorants. Par rapport à l'acculturation aux données et aux bonnes pratiques de gestion, est-ce qu'il faudrait aussi penser déjà au niveau du master ?

31

00:37:20,480 --> 00:40:54,000

Joachim Schöpfel

Pour moi, la question est, comment dire, entre guillemets, c'est une question de développement de services, de marketing. C'est-à-dire que si on n'a pas assez de ressources, et je crois que c'est le cas partout, c'est qu'il faut cibler. On ne peut pas tout faire en même temps. Il faut cibler. C'est exactement comme ça qu'on a procédé. À Lille 3, je parle bien avant la fusion, on a clairement dit qu'on cible les doctorants. Évidemment, il y a une raison pour ça. Investissement de l'avenir et tout ça, ça fait partie des compétences basiques aujourd'hui qu'on doit maîtriser si on veut faire de la recherche. Mais dès le départ, on n'avait pas envisagé de former... Tous les chefs de projet, tous les anciens chercheurs, tous les directeurs de laboratoire, on aurait pu faire autrement. On a dit non, cela. Et en pensant aussi un petit peu comme ce qu'on appelle le marketing viral, si les autres, les vieux, viennent, et effectivement, il y en a des qui sont venus, ils ont dit "pourquoi pas nous ?" On va voir. Donc, soit on les intègre dans nos formations, soit on fait ça d'une autre manière. Mais c'est pour cibler. Et pour l'instant, ça a porté ses fruits. Je vois ça au niveau national. Donc, j'ai vu la démarche du collège données du comité pour la science ouverte, qui a mis ça en avant avec des guides et tout ça. Voilà. Maintenant, tu as posé la question des masters. Je suis partagé. Je suis partagé. Donc, je suis même

responsable d'une première année de master. Partagé entre... Oui, il faut une culture de données. Notamment dans nos domaines, là-bas, on forme des documentalistes, des bibliothécaires, des archivistes, des gens qui travaillent quelque part avec l'information et les documents. Il faut... Eh bien, on est en train de développer ça. Il leur faut une culture de la donnée. Que ce soit une donnée de recherche, une donnée publique, une donnée ouverte, qu'importe. Il leur faut ça. Ça, c'est important. Maintenant, former tous les étudiants au master à des questions de données de la recherche, je suppose que dans la plupart des cas, il y a d'autres priorités. Et que ça existe comme une formation optionnelle. Notamment là où il y a des parcours de recherche en deuxième année. Pourquoi pas ? Pourquoi pas ? Mais je crois que ce qui est plus important pour eux, c'est cette culture de la donnée de manière générale, qu'ils sachent de quoi il s'agit, qu'ils aient une sorte de sensibilisation à l'enjeu des métadonnées, des identifiants, des méthodes d'analyse, de visualisations de données. Je crois que ça, oui.

32

00:40:54,400 --> 00:43:01,300

Mireille Brenel

Alors, je vais faire beaucoup de répétitions, je pense, mais je rejoins Joachim sur absolument tout ce qu'il vient de dire. Bien sûr que dans les actions de sensibilisation, de formation, il faut qu'on s'adresse aux chercheurs, aux ingénieurs d'études, aux ingénieurs de recherche, aux personnels de laboratoire, etc. Mais bien sûr également aux doctorants. C'est aussi une priorité qu'on a mis de notre côté de former l'ensemble des doctorants de l'université à la science ouverte et à la gestion des données. Comme tu le disais, Joachim, les doctorants, c'est les chercheurs de demain. Et bien souvent, c'est aussi eux qui peuvent être amenés à gérer, dans le cadre de leur thèse, les données en lien avec leur projet, voire même la rédaction du plan de gestion de données. Donc, c'est très important qu'ils aient ce bagage-là. Et on se dit également que d'acculturer cette génération, c'est aussi un bon investissement sur l'avenir. Et on compte également sur eux pour impulser les bonnes pratiques dans leur laboratoire. Se faire un peu... Les porte-paroles de ce qu'ils ont appris avec nous en formation, réussir à le calquer dans leur travail de recherche. Après, en ce qui concerne les étudiants en master, alors en effet, ça pourrait être intéressant de réfléchir à les sensibiliser. Enfin, à l'université, entre collègues, on en a déjà parlé plusieurs fois. Mais je suis aussi d'accord pour dire que c'est peut-être une question de priorité aussi. Et la priorité, elle est donnée sur les équipes de recherche, les doctorants. Et là où je rejoins encore une fois Joachim, c'est que pour moi, l'idée, ce serait, sachant que tous les étudiants en master ne sont pas non plus amenés à faire de la recherche, peut-être proposer des formations sur la gestion des données de recherche, mais de façon optionnelle, qu'ils aient juste un premier vernis, au moins connaissance, que ça commence un peu à germer. Mais en effet, je suis d'accord aussi que ce serait embêtant qu'ils passent à côté. Mais peut-être proposer dans un premier temps quelque chose d'optionnel pour ceux que ça intéresse, qui puissent justement bénéficier de nos formations. Oui, je pense que c'est une

piste à garder en tête.