



La numérisation des thèses de doctorat françaises - État des lieux

Enquête 2020

Ensemble, avec nos réseaux,
réinventons le service public des données

abes.fr

Sommaire

Introduction	4
Objectifs de l'enquête	4
Contenu de l'enquête	5
Pilotage du projet de numérisation des thèses	5
Typologie des fonds de thèses numérisés.....	5
Numérisation.....	5
Au préalable.....	5
Prestation.....	5
Ressources humaines et financières	5
Aspects techniques	6
Conservation des fichiers numérisés.....	6
Signalement des fichiers numérisés.....	6
Diffusion des fichiers numérisés	7
Utilisation du corpus numérisé	7
Bilan	7
Résultats de l'enquête	8
Liste des répondants.....	8
Pilotage du projet de numérisation des thèses	9
Question 1 : objectifs des chantiers de numérisation des thèses.....	9
Question 2 : portage et suivi du projet.....	9
Typologie des fonds de thèses numérisés.....	10
Question 3 : période(s), discipline(s) et volumétrie des fonds numérisés	10
Question 4 : type(s) de thèses numérisées	12
Numérisation.....	13
Au préalable.....	13
Question 5 : collecte des fichiers numériques de thèses auprès des auteurs avant numérisation.....	13
Question 6 : récupération des fichiers de thèses	13
Question 7 : nombre de fichiers récupérés	13
Prestation.....	13
Question 8 : appel à un prestataire extérieur pour la numérisation.....	13
Question 9 : liste des prestataires sollicités.....	14
Ressources humaines et financières	14
Question 10 : financements	14
Question 11 : Avez-vous eu recours à un ou des recrutements pour mener à bien le projet ?	15
Aspects techniques	16

Question 12 : supports numérisés.....	16
Question 13 : raisons du choix du support.....	17
Question 14 : bilan du choix du support	17
Question 15 : recours à la numérisation destructive.....	18
Question 16 : coloris des fichiers numérisés.....	18
Question 17 : résolution des fichiers numérisés	19
Question 18 : format(s) de numérisation.....	19
Question 19 : océrisation des fichiers numérisés.....	20
Conservation des fichiers numérisés.....	20
Question 20 : conservation des fichiers numérisés.....	20
Signalement des fichiers numérisés.....	21
Question 21 : signalement des thèses numérisées	21
Diffusion des fichiers numérisés	21
Question 22 : diffusion des fichiers.....	21
Questions 23 et 23 bis : recueil de l'accord des auteurs ou des ayants droits	22
Question 24 : organisation du recueil des autorisations de diffusion.....	22
Question 25 : bilan de l'organisation du recueil des autorisations de diffusion	23
Question 26 : format(s) de diffusion des fichiers numérisés.....	23
Question 27 : plateforme(s) de diffusion	24
Utilisation du corpus numérisé	25
Question 28 : services attachés aux corpus de thèses numérisées.....	25
Question 29 : usages du corpus de thèses numérisées.....	25
Bilan.....	26
Question 30 : bilan du projet de numérisation des thèses	26
Synthèse des résultats de l'enquête.....	27
Conclusion.....	28

Introduction

Menée du 8 au 31 janvier 2020, cette enquête s'adressait aux établissements de l'Enseignement Supérieur et de la Recherche (ESR) membres des réseaux de l'Abes ayant pour caractéristique de conserver des fonds de thèses (anciennes ou modernes) et ayant entrepris – ou achevé – de numériser tout ou partie de ces fonds.

L'enquête a recueilli **24 réponses valides** (dont trois réponses incomplètes).

Objectifs de l'enquête

- **recenser** les établissements ayant entrepris ou achevé de numériser leurs fonds de thèses
- **dresser une typologie des fonds de thèses numérisés** : type de thèses, périodes de soutenance, disciplines, volumétrie
- **dresser un état des lieux des modalités techniques et organisationnelles** mises en œuvre dans le cadre de la numérisation des documents proprement dite mais également de l'exploitation, de la conservation et de la diffusion des fichiers numérisés

Contenu de l'enquête

Pilotage du projet de numérisation des thèses

- **Question 1** : Quels étaient les objectifs de votre projet de numérisation des thèses ?
 - Maintenir et faciliter l'accès à la collection
 - Améliorer la diffusion, la visibilité et la valorisation des thèses
 - Participer à un projet de recherche
 - Répondre à une commande de l'établissement
 - Améliorer la conservation des documents
 - Autre (précisez)
- **Question 2** : Comment a été organisé le portage et le suivi du projet ?

Typologie des fonds de thèses numérisés

- **Question 3** : Quel est le périmètre du fonds de thèses que vous avez numérisé ?
 - Période(s)
 - Discipline(s)
 - Volumétrie
- **Question 4** : Quel(s) type(s) de thèses avez-vous numérisés ?
 - Thèses de doctorat
 - Thèses d'exercice
 - Autre (précisez)

Numérisation

Au préalable

- **Question 5** : Avant la numérisation, avez-vous procédé à une collecte des éventuels fichiers numériques de thèses auprès des auteurs ?
 - Collecte préalable
 - Pas de collecte préalable
- Si le répondant a indiqué qu'il a collecté les fichiers numériques des thèses :
 - **Question 6** : Comment avez-vous procédé pour récupérer les fichiers ?
 - **Question 7** : Combien de fichiers avez-vous pu récupérer ?

Prestation

- **Question 8** : Avez-vous fait appel à un prestataire extérieur pour réaliser la numérisation ?
 - Numérisation réalisée en interne
 - Prestation de service pour la numérisation de l'ensemble du fonds
 - Prestation de service pour la numérisation d'une partie du fonds (précisez)
- Si le répondant a indiqué qu'il a fait appel à une prestation :
 - **Question 9** : A quel(s) prestataire(s) avez-vous fait appel ?

Ressources humaines et financières

- **Question 10** : Avez-vous bénéficié d'un ou de plusieurs financements ?
 - Projets non financés, menés sur fonds propre
 - Projets financés
- **Question 11** : Avez-vous eu recours à un/des recrutement(s) pour mener à bien le projet ?
 - Projets menés sans recrutement de personnels supplémentaires
 - Projets menés avec recrutement de personnels (précisez)

Aspects techniques

- **Question 12** : A partir de quel(s) support(s) avez-vous procédé à la numérisation des thèses ?
 - Exemplaire imprimé, pour toutes les thèses
 - Exemplaire imprimé, pour une partie des thèses
 - Microfiche établissement, pour toutes les thèses
 - Microfiche établissement, pour une partie des thèses
 - Microfiche mère (détenue par l'ANRT), pour toutes les thèses
 - Microfiche mère (détenue par l'ANRT), pour une partie des thèses
 - Précisions
- **Question 13** : Pour quelles raisons avez-vous opéré ce choix de support ?
- **Question 14** : Quel bilan tirez-vous de ce choix de support (facilité de numérisation, coût, rendu final, etc.) ?
- **Question 15** : Avez-vous procédé à une numérisation destructive ?
 - Numérisation non destructive
 - Numérisation destructive pour toutes les thèses
 - Numérisation destructive pour une partie des thèses (précisez)
- **Question 16** : Comment avez-vous numérisé les documents en termes de coloris ?
 - En noir et blanc
 - En niveau de gris
 - En couleur
 - Cela dépend
- **Question 17** : Quelle résolution avez-vous choisie ?
 - 200 dpi
 - 300 dpi
 - 400 dpi
 - 500 dpi
 - 600 dpi
 - Cela dépend
- **Question 18** : Sous quel(s) format(s) image avez-vous numérisé les documents ?
 - .tiff
 - .png
 - .jpeg
 - Autre (précisez)
- **Question 19** : Avez-vous procédé à l'océrisation des fichiers numérisés ?
 - Fichiers non océrisés
 - Fichiers océrisés

Conservation des fichiers numérisés

- **Question 20** : Comment assurez-vous la conservation des fichiers numérisés ?
 - Stockage sur des serveurs internes / appartenant à l'établissement
 - Stockage sur des serveurs externes / n'appartenant pas à l'établissement
 - Archivage pérenne chez un opérateur public ou privé (précisez)

Signalement des fichiers numérisés

- **Question 21** : Dans quel(s) outil(s) signalez-vous les documents numérisés ?
 - Dans le Sudoc
 - Dans le catalogue de l'établissement
 - Dans la bibliothèque numérique de l'établissement
 - Sur une plateforme d'archives ouvertes
 - Autre (précisez)

Diffusion des fichiers numérisés

- **Question 22** : Les fichiers sont-ils diffusés ?
 - En accès libre sur internet
 - En accès libre sur internet OU en accès restreint sur intranet
 - En accès restreint sur intranet
 - Pas de diffusion
- Si le répondant a choisi « En accès libre sur internet » ou « En accès restreint sur intranet » à la question 22 :
 - **Question 23** : Avez-vous recueilli l'accord des auteurs ou des ayants droits ?
 - Demande d'autorisation de diffusion adressée à l'auteur
 - Pas de demande d'autorisation de diffusion adressée à l'auteur
- Si le répondant a choisi « En accès libre sur internet OU en accès restreint sur intranet » à la question 22 :
 - **Question 23 bis** : Avez-vous recueilli l'accord des auteurs ou des ayants droits ?
 - Demande d'autorisation de diffusion adressée à l'auteur uniquement pour la diffusion sur internet.
 - Demande d'autorisation de diffusion adressée à l'auteur pour la diffusion sur internet ET la diffusion sur intranet.
 - Pas de demande d'autorisation de diffusion adressée à l'auteur
- Si le répondant a indiqué qu'il a recueilli l'autorisation de diffusion des ayants droits aux questions 23 et 23 bis :
 - **Question 24** : Comment avez-vous procédé ?
 - **Question 25** : Quel bilan tirez-vous de la méthode que vous avez utilisée pour recueillir l'accord des ayants droits ?
- Si le répondant a indiqué « En accès libre sur internet » à la question 22 :
 - **Question 26** : Sous quel(s) format(s) diffusez-vous les thèses numérisées ?
 - .pdf
 - .tiff
 - .png
 - .jpeg
 - Autre (précisez)
 - **Question 27** : Quelle(s) plateforme(s) de diffusion utilisez-vous ?

Utilisation du corpus numérisé

- **Question 28** : Avez-vous rattaché de nouveaux services à ce corpus de thèses numérisées ?
- **Question 29** : Quels usages sont faits aujourd'hui de ce corpus de thèses numérisés (retour des usagers, usages constatés, données statistiques, etc.) ?

Bilan

- **Question 30** : Quel bilan tirez-vous de votre projet de numérisation des thèses (rendu final, difficultés rencontrées, satisfaction, etc.) ?

Résultats de l'enquête

Si de grandes tendances se dégagent des résultats de l'enquête concernant les choix techniques et stratégiques opérés par les établissements en termes de numérisation, de diffusion, de signalement et d'archivage, les chantiers s'avèrent suffisamment différents les uns des autres pour être riches d'enseignements.

Les bilans, positifs comme négatifs, que les établissements tirent de leur expérience, permettent d'identifier les écueils à éviter et les bonnes pratiques à adopter.

Liste des répondants

23 établissements ont indiqué avoir mené, ou être en train de mener, un ou plusieurs projets de numérisation des thèses. Dans 18 cas sur 23, il s'agit d'universités, ce qui représente **26% des 69 universités habilitées à délivrer le doctorat** en propre ou au sein d'un regroupement d'établissements (ComUE ou regroupement expérimental).

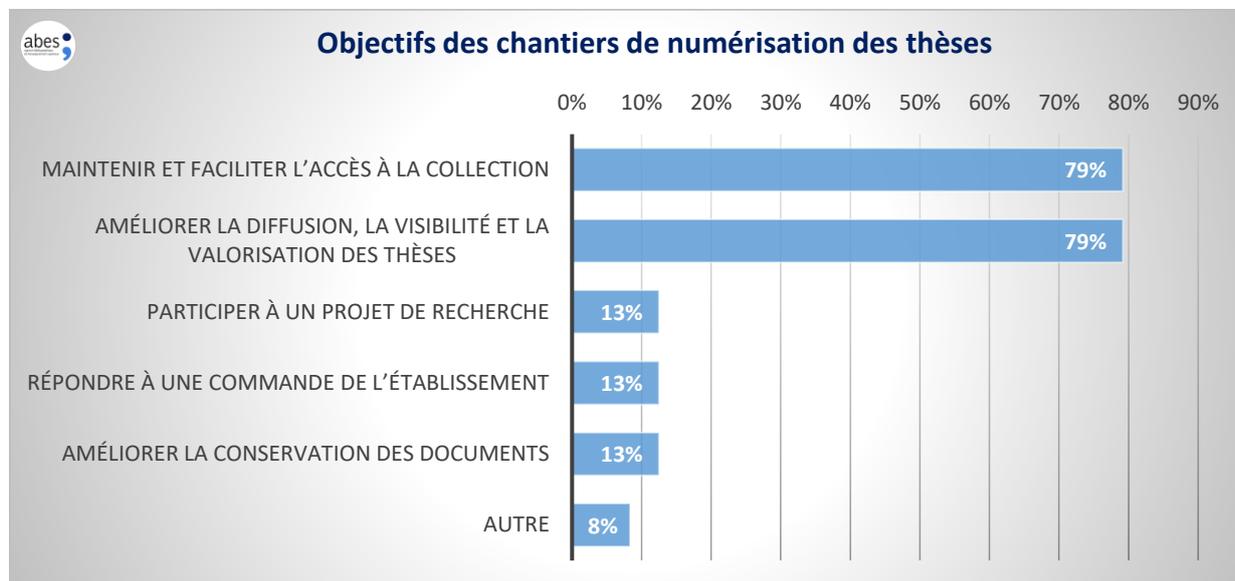
- Aix-Marseille Université
- CentraleSupélec
- Ecole Centrale de Lyon
- Mines ParisTech
- Observatoire de Paris
- Sciences Po Paris
- Université de Bordeaux
- Université de Bordeaux Montaigne
- Université de Cergy-Pontoise
- Université Grenoble Alpes
- Université de La Réunion
- Université de Limoges
- Université de Lorraine
- Université Lyon 1 et Observatoire de Lyon
- Université Lyon 2
- Université du Mans
- Université de Paris
- Université Paris-Dauphine
- Université Paris 8 Vincennes Sainte-Denis
- Université de Perpignan
- Université de Poitiers
- Université Polytechnique Hauts-de-France
- Université de Strasbourg

Pilotage du projet de numérisation des thèses

Question 1 : objectifs des chantiers de numérisation des thèses

Les projets de numérisation ont pour principaux objectifs :

- **faciliter l'accès aux thèses** ;
- **valoriser les fonds**, en améliorant leur diffusion et leur visibilité en ligne.



Question 2 : portage et suivi du projet

Dans la grande majorité des cas, les projets de numérisation des thèses sont confiés aux **services documentaires des établissements**. Ainsi, douze établissements indiquent que le projet a été piloté par le Service Commun de la Documentation (SCD) seul.

Dans trois cas, le projet a été piloté par plusieurs services documentaires, dont le SCD :

- Bordeaux : Service de coopération documentaire (service interuniversitaire) en collaboration avec le SCD de l'Université de Bordeaux
- Lyon : Bibliothèque Math-Info-Recherche en collaboration avec la Bibliothèque Diderot de Lyon (service interuniversitaire)
- Strasbourg : SCD de l'Université de Strasbourg en collaboration avec la Bibliothèque Nationale Universitaire de Strasbourg (BNUS) et la médiathèque de l'EuroMétropole

En revanche, à l'Université Lyon 2, le projet de numérisation a été mené par l'ERAD (Edition, Reproduction & Archivage du Document), service qui n'est pas lié au SCD.

Au sein des SCD, le pilotage est généralement confié à un **personnel de catégorie A**, conservateur, bibliothécaire ou ingénieur, le plus souvent responsable du service des thèses et/ou de la documentation numérique, ou du service numérisation.

A Strasbourg, le portage scientifique du projet a été assuré par un groupe de chercheurs.

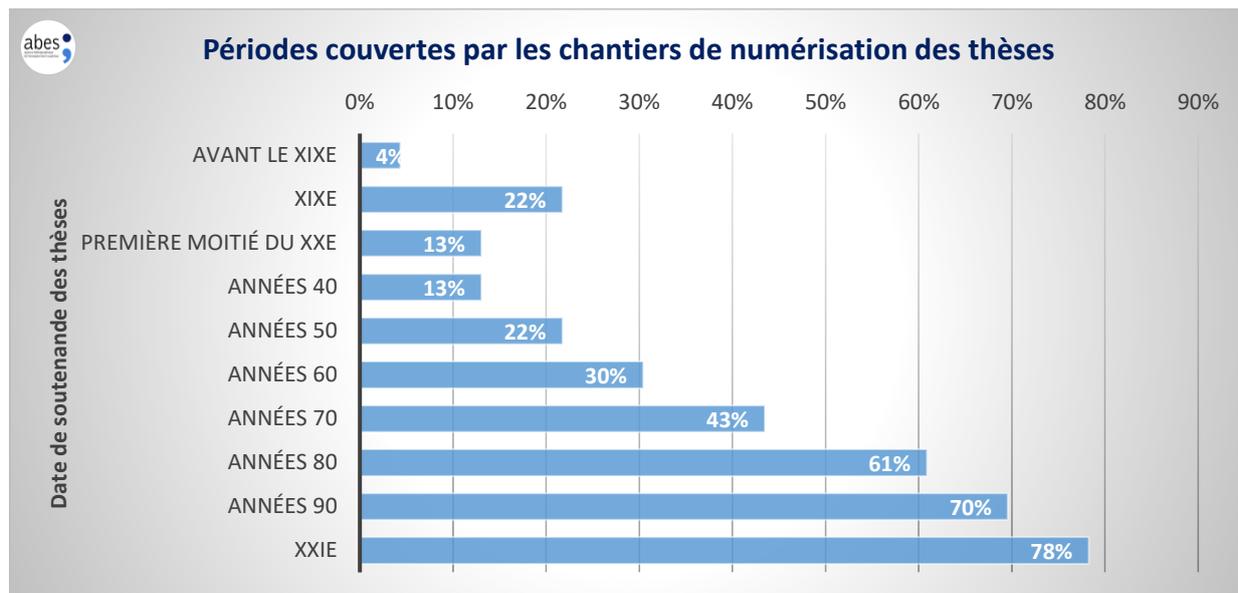
Typologie des fonds de thèses numérisés

Question 3 : période(s), discipline(s) et volumétrie des fonds numérisés

Période(s) couverte(s) par les projets de numérisation

La majorité des chantiers de numérisation porte sur les **thèses soutenues entre les années 1980 et le passage au dépôt national électronique des thèses**, entre 2006 et 2016. Les thèses anciennes représentent :

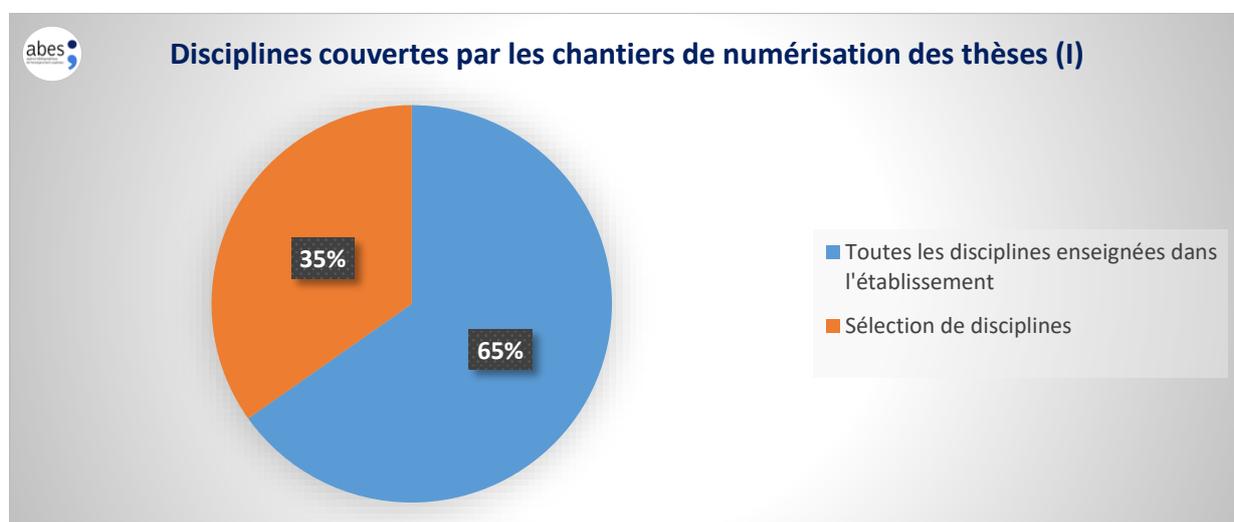
- un peu plus de 20% des chantiers de numérisation pour les thèses du XIX^e siècle
- un peu plus de 10% des chantiers pour les thèses du début du XX^e siècle
- et 20% pour les thèses des années 1950s



Discipline(s) couverte(s) par les projets de numérisation

2/3 des établissements ayant procédé à la numérisation totale ou partielle de leur fonds de thèses ont opéré **sans distinction de disciplines**. Parmi eux, on dénombre 8 établissements non pluridisciplinaires et 6 établissements pluridisciplinaires.

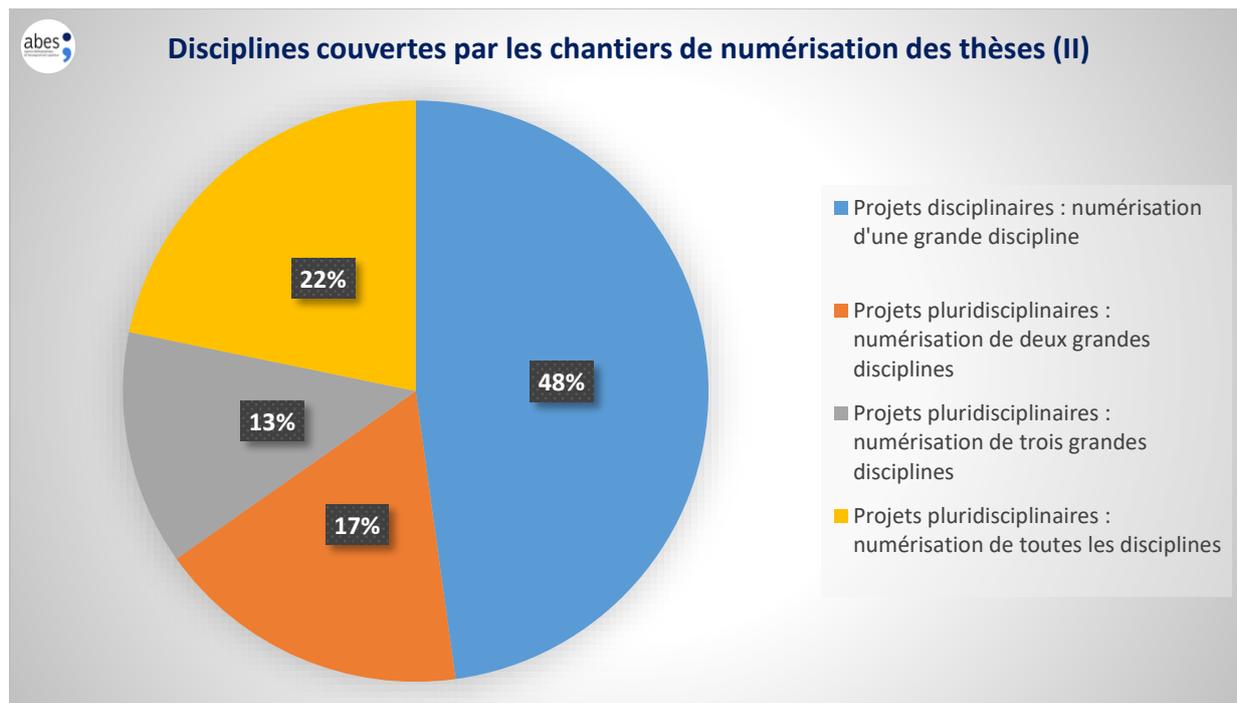
En revanche, un tiers des établissements (9 établissements pluridisciplinaires) ont choisi de faire porter leur projet de numérisation sur certains segments disciplinaires précis plutôt que sur la totalité de leur fonds (ex : Mathématiques pour l'Université de Paris ; Médecine et Pharmacie pour l'Université de Limoges ; Droit, Economie, Politique pour Aix Marseille Université).



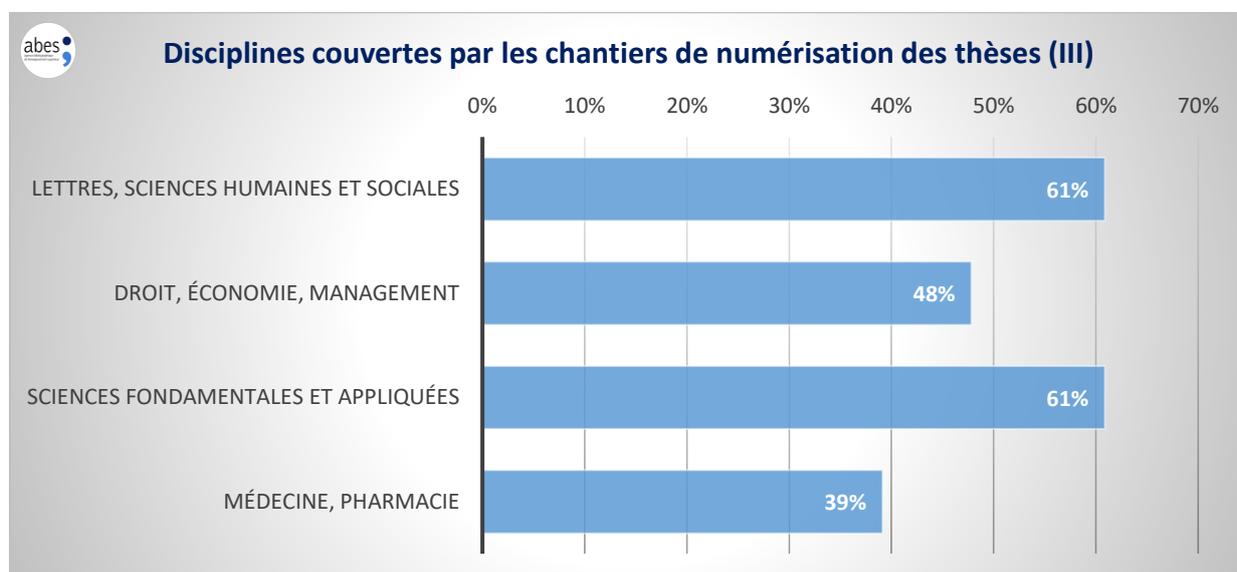
La majorité des projets de numérisation (8 sur 11 projets) sont portés par des établissements spécialisés (universités non pluridisciplinaires, écoles d'ingénieur).

Près de la moitié des projets portent sur une seule grande famille disciplinaire : Lettres et Sciences humaines ; Droit, Economie, Management ; Sciences fondamentales et appliquées ; Médecine, Pharmacie.

Seuls deux établissements pluridisciplinaires ont choisi de faire porter la numérisation de leurs thèses sur un **seul segment disciplinaire**. 5 établissements pluridisciplinaires ont numérisé l'ensemble de leurs fonds sans distinction de discipline. 3 autres ont uniquement exclu les thèses de médecine et pharmacie de leurs chantiers de numérisation.

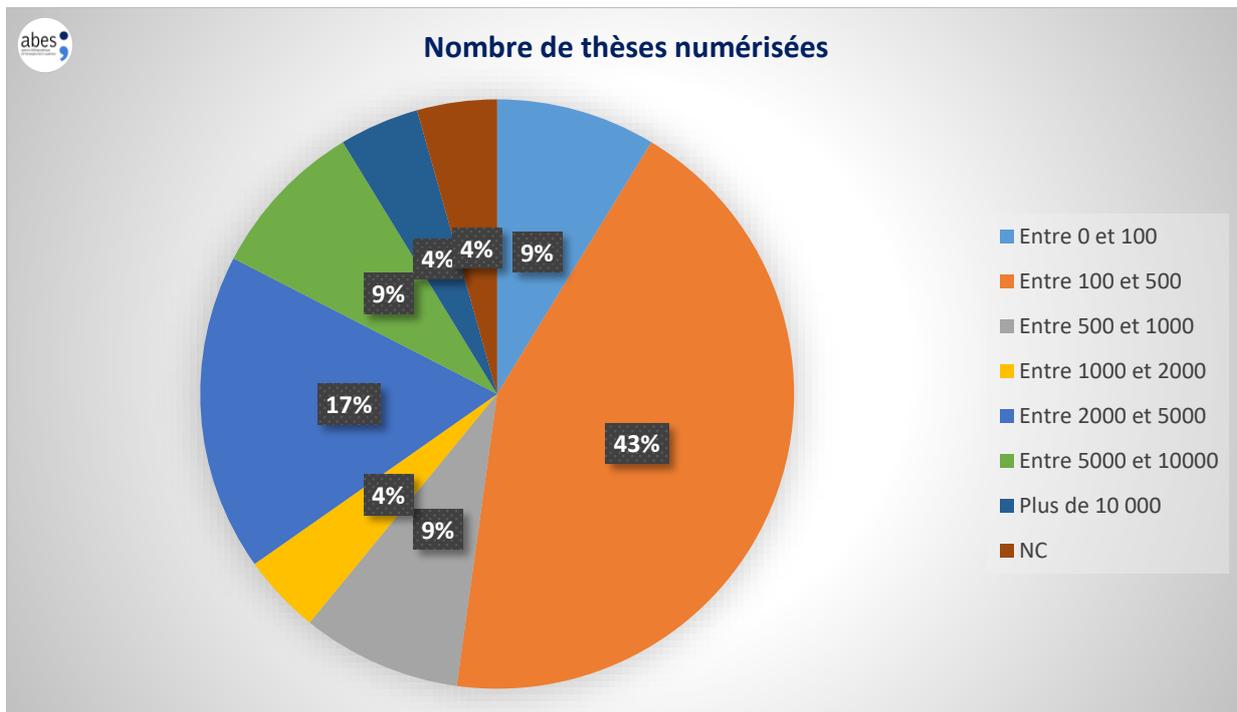


2/3 des chantiers incluent des thèses en Sciences fondamentales et appliquées. Il en est de même pour les thèses en Lettres et Sciences Humaines. Les thèses en Droit, Economie, Management, quant à elles, sont incluses dans près de la moitié des projets, et les thèses en Médecine, Pharmacie dans 39% d'entre eux.



Volumétrie des fonds numérisés

51% des projets de numérisation portent sur des fonds dont la **volumétrie oscille entre 10 et 500 thèses**. 35% dépassent le millier de volumes numérisés.

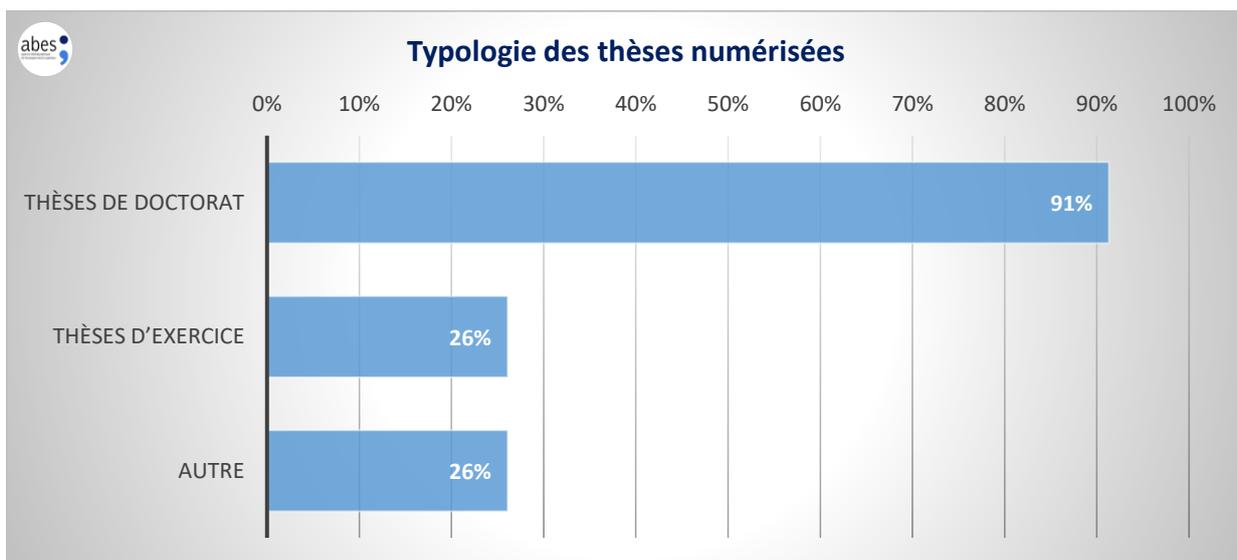


Question 4 : type(s) de thèses numérisées

Si 91% des projets portent sur la numérisation des **thèses de doctorat**, les thèses d'exercice occupent quant à elle un quart des projets, ce qui n'est pas négligeable.

Les autres documents numérisés sont :

- des thèses antérieures à la réforme du doctorat de 1985 : les thèses de doctorat d'Etat, les thèses de 3^e cycle et les thèses de licence du XIX^e siècle
- des mémoires
- des HDR
- des thèses « d'établissement » (hors cursus LMD).

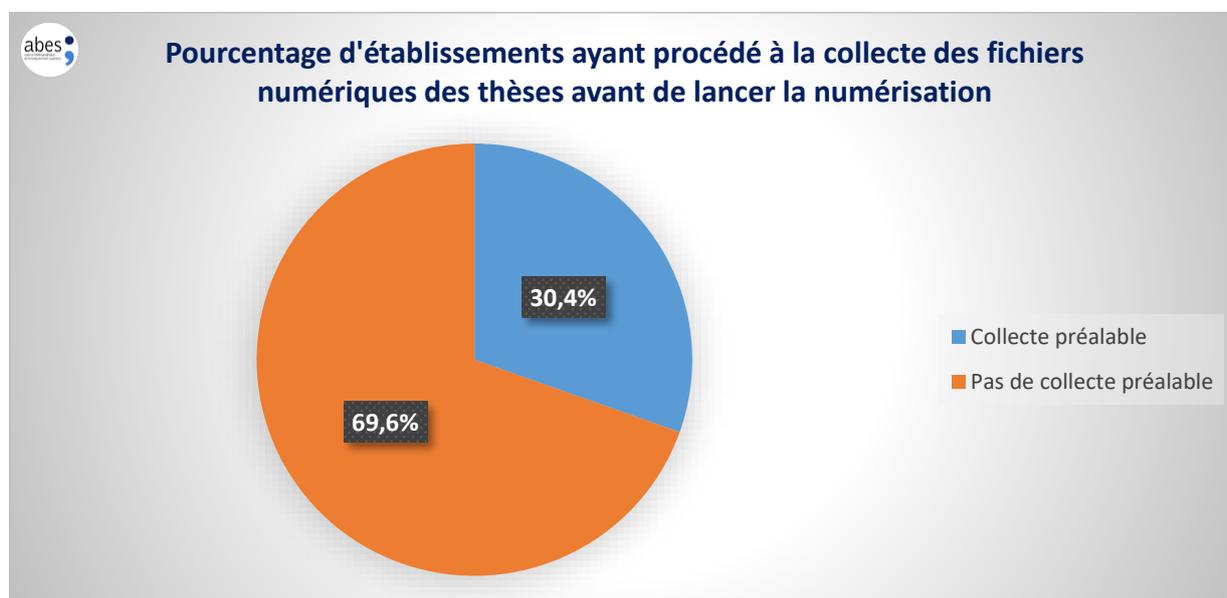


Numérisation

Au préalable

Question 5 : collecte des fichiers numériques de thèses auprès des auteurs avant numérisation

Seul **un tiers des établissements** a entrepris de collecter les fichiers nativement numériques des thèses avant de procéder à la numérisation des documents.



Question 6 : récupération des fichiers de thèses

Les établissements qui ont collecté les fichiers avant numérisation ont procédé de deux manières ;

- **concomitamment au dépôt imprimé de la thèse** (deux tiers des cas)
- **a posteriori en contactant les docteurs par mail** (un peu moins d'un tiers des cas)

Question 7 : nombre de fichiers récupérés

Dans le cas où les établissements ont procédé à un double dépôt électronique/imprimé, toutes les thèses ont pu être récupérées au format électronique. L'un de ces établissements signale néanmoins que les fichiers ainsi récupérés (quelques dizaines) se sont avérés inexploitable car différents de la version imprimée (éléments manquants, mise en page différente).

Sur les deux établissements ayant contacté les docteurs a posteriori par mail, l'un n'a pas pu indiquer le nombre de thèses récupérées ; le second a signalé un taux de recouvrement relativement faible (14%).

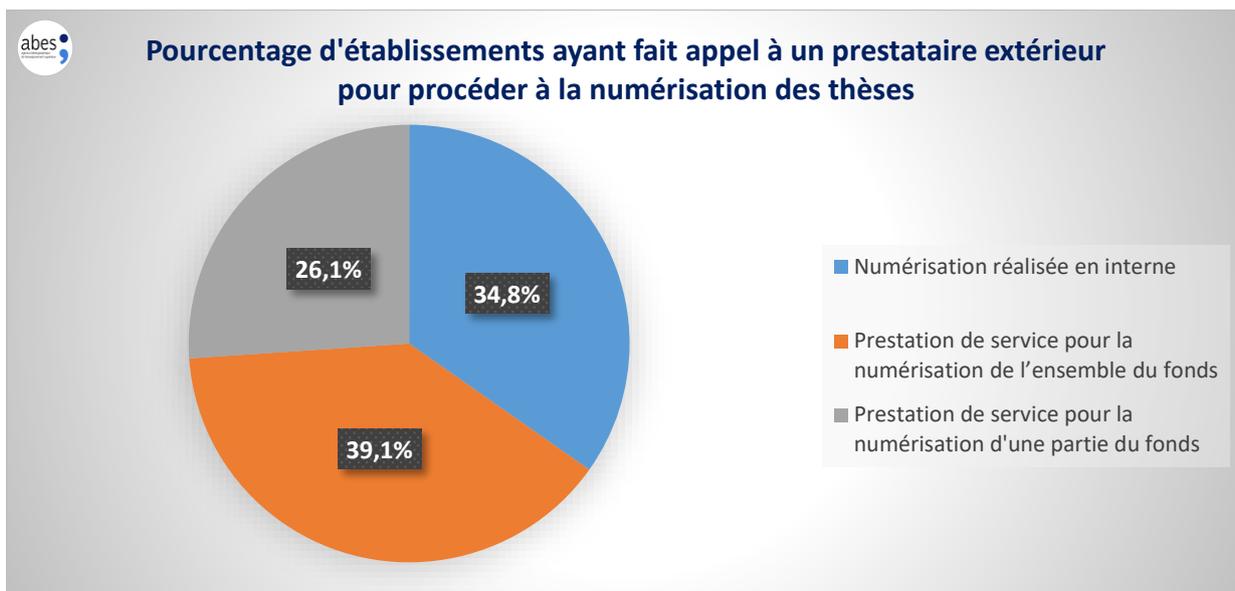
Prestation

Question 8 : appel à un prestataire extérieur pour la numérisation

Dans **35%** des cas, les établissements ont réalisé **la numérisation de leur fonds en interne**, sur leur propre banc de numérisation. **39% ont fait appel à un prestataire pour numériser** l'ensemble de leur fonds et **26% pour en numériser une partie** seulement.

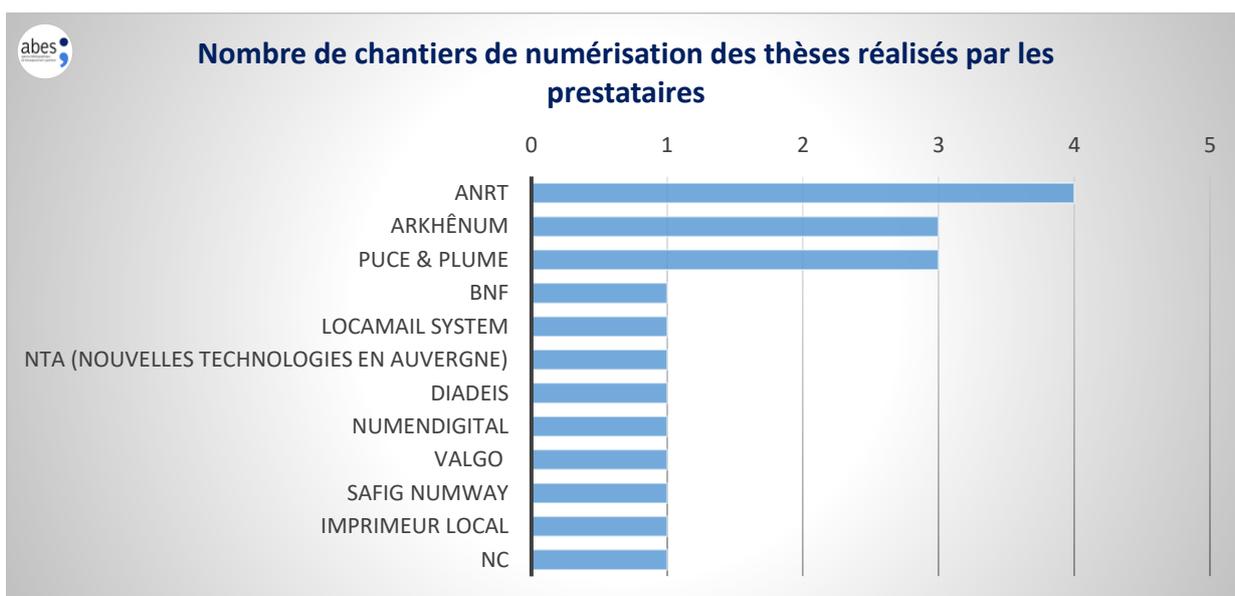
Entrent dans ce dernier cas de figure les établissements :

- ayant mené **plusieurs projets de numérisation réussis** : certains projets menés en interne, d'autres non
- ayant confié à un prestataire la numérisation en masse de la majorité des documents et **numérisé en interne les documents les plus fragiles** sur le plan matériel
- ayant **fait appel à l'ANRT pour récupérer les fichiers numériques** utilisés dans la fabrication des microfiches de thèse les plus récentes



Question 9 : liste des prestataires sollicités

L'enquête recense une **dizaine de prestataires** auxquels les établissements ont fait appel. Certains prestataires ont été sollicités à plusieurs reprises (ANRT, Arkhênum et Puce & Plume).



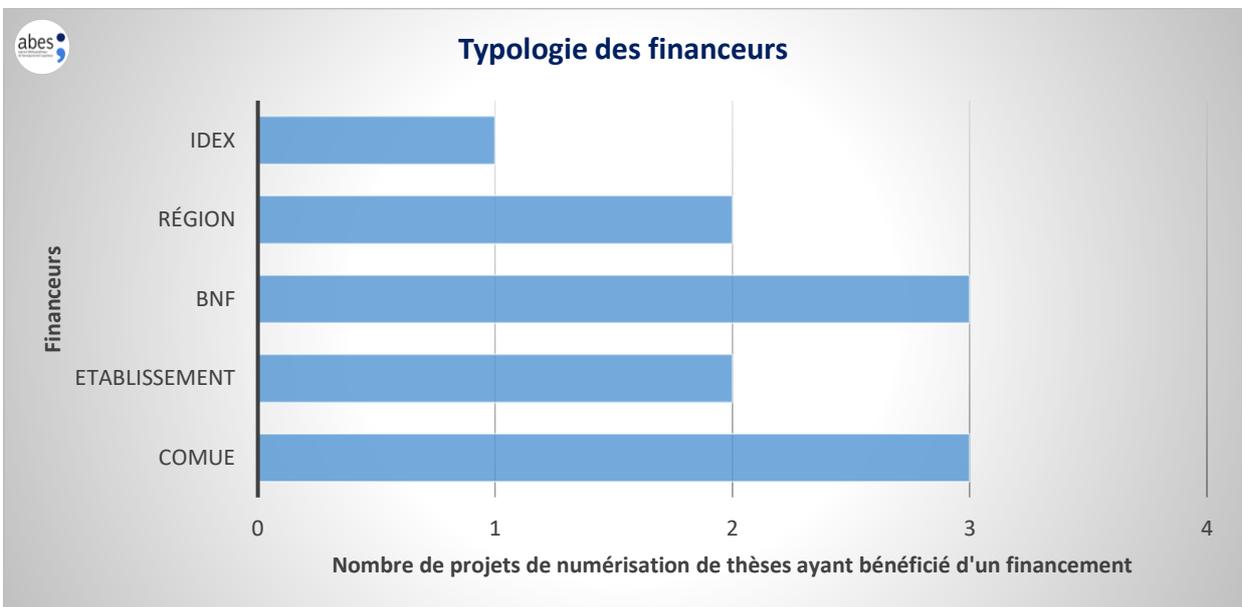
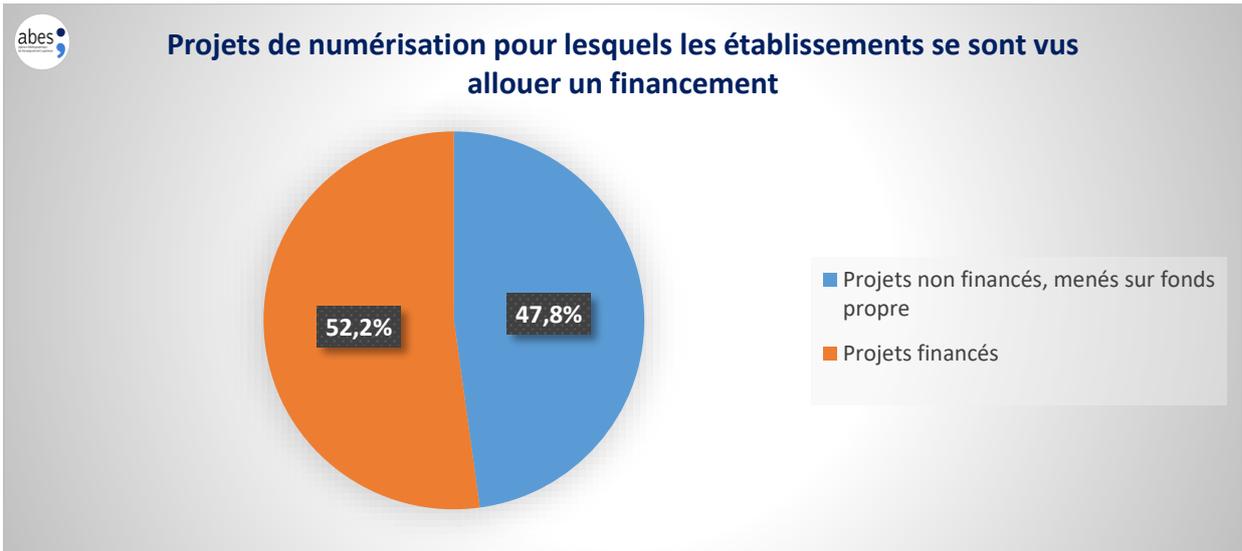
Ressources humaines et financières

Question 10 : financements

Un peu plus de la moitié des projets de numérisation (52%) ont bénéficié d'un financement.

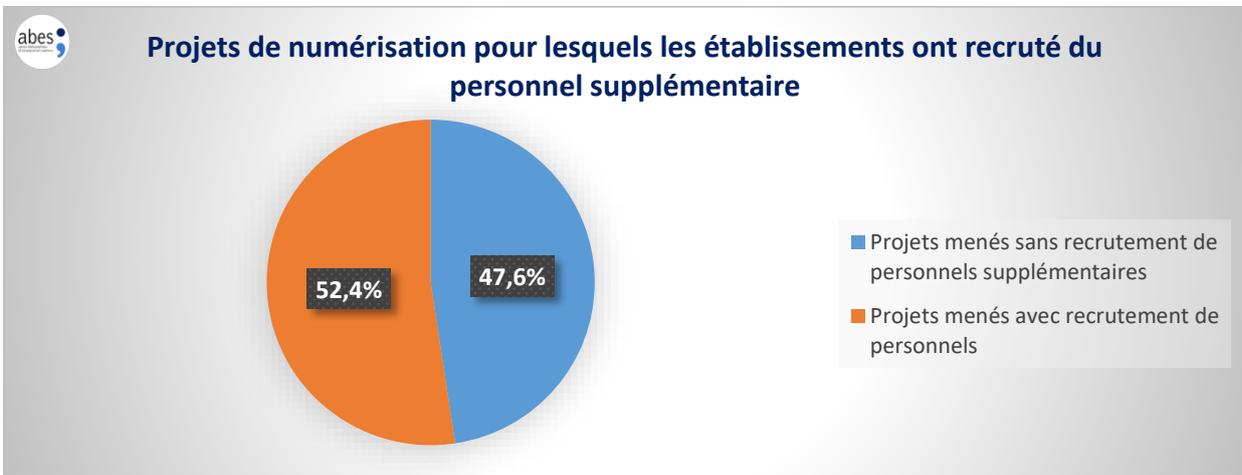
Les financements proviennent :

- de **l'enseignement supérieur et de la recherche** : ComUE (3), universités de rattachement des SCD (2), financement IdEx (1)
- de la **BnF** (3) : pour les projets qui portent sur la numérisation des thèses anciennes (XIX^e et début XX^e siècle)
- des **régions** (2)



Question 11 : Avez-vous eu recours à un ou des recrutements pour mener à bien le projet ?

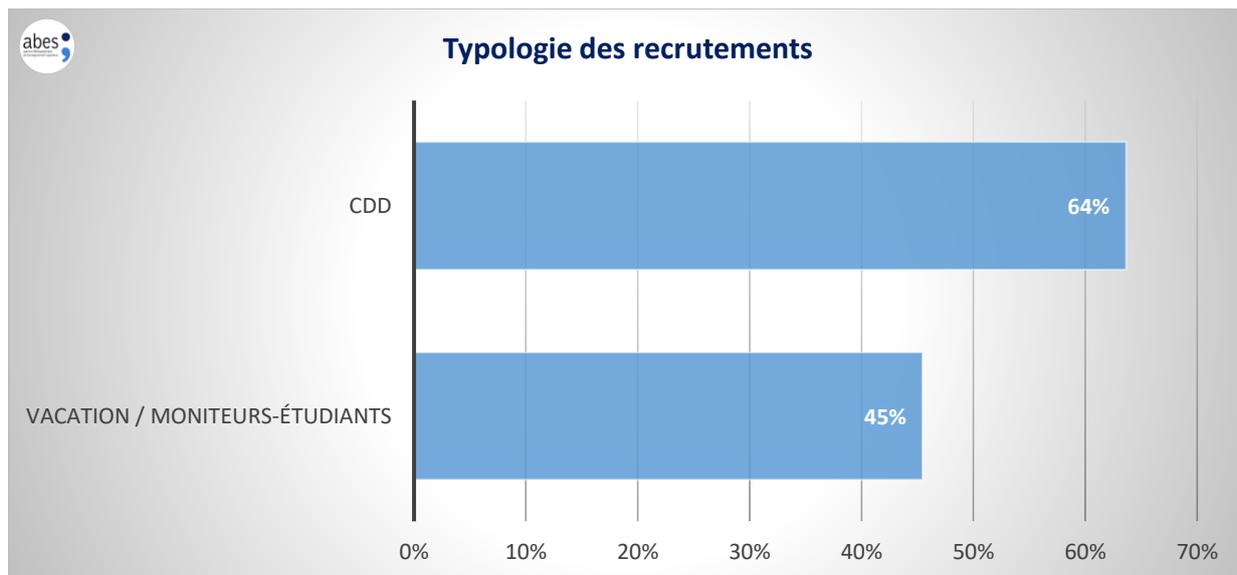
Un peu plus de la moitié des projets de numérisation (52%) ont bénéficié d'un recrutement de personnels.



Les recrutements prennent la forme, majoritairement, de **CDD** (64%), ainsi que de **vacations, avec le recours aux moniteurs étudiants** (45%). Les établissements combinent parfois les deux types de recrutements.

Dans le cas des CDD, 43% des établissements ont eu recours à un seul recrutement, pour toute la durée du projet, alors que 57% ont pu recruter plusieurs CDD, simultanément ou successivement, pour quelques mois, à certaines phases du projet (pour le signalement du corpus numérisé par exemple).

Dans le cas des vacations, les établissements font appel à plusieurs vacataires en même temps, pour quelques heures seulement, ou pour toute une année.



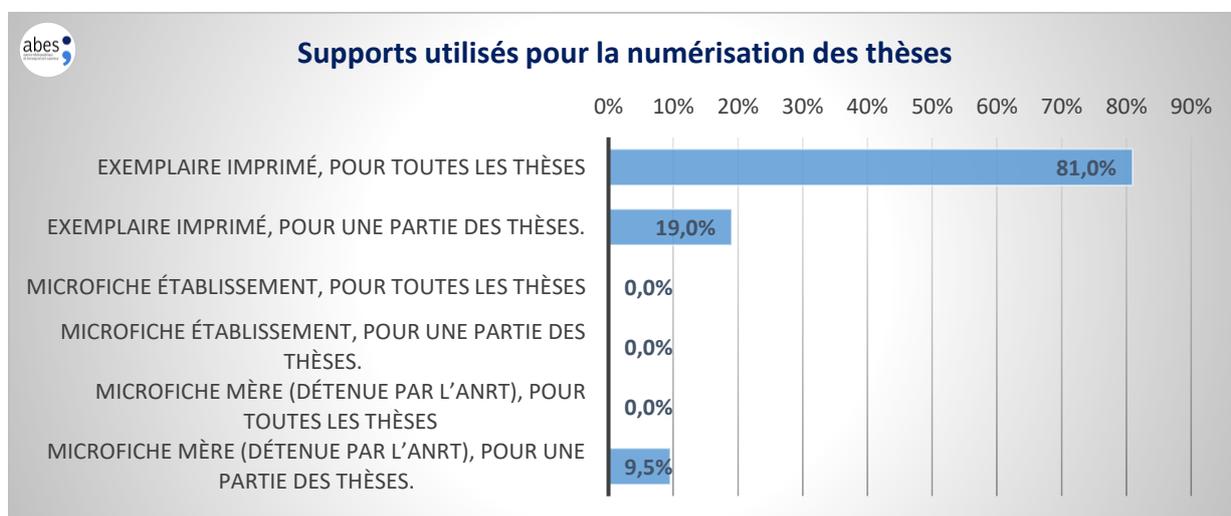
Aspects techniques

Question 12 : supports numérisés

Tous les projets de numérisation ont été menés, pour tout ou partie, à partir **des exemplaires imprimés des thèses**.

Les microfiches mères détenues par l'ANRT ont été utilisées comme support de numérisation dans deux projets, en compléments des exemplaires imprimés.

D'autres projets ont couplé numérisation des thèses imprimées et utilisation des fichiers de thèses nativement numériques déposés par les docteurs en parallèle du dépôt imprimé.



Question 13 : raisons du choix du support

Les établissements ont choisi les exemplaires imprimés comme support de numérisation des thèses principalement en raison :

- de la **disponibilité du support**
- de la **facilité de manipulation et de numérisation** des exemplaires imprimés

Les autres arguments évoqués sont : le bon état des supports, la qualité du rendu après numérisation, l'exhaustivité du corpus imprimé, les contraintes imposées par le prestataire.

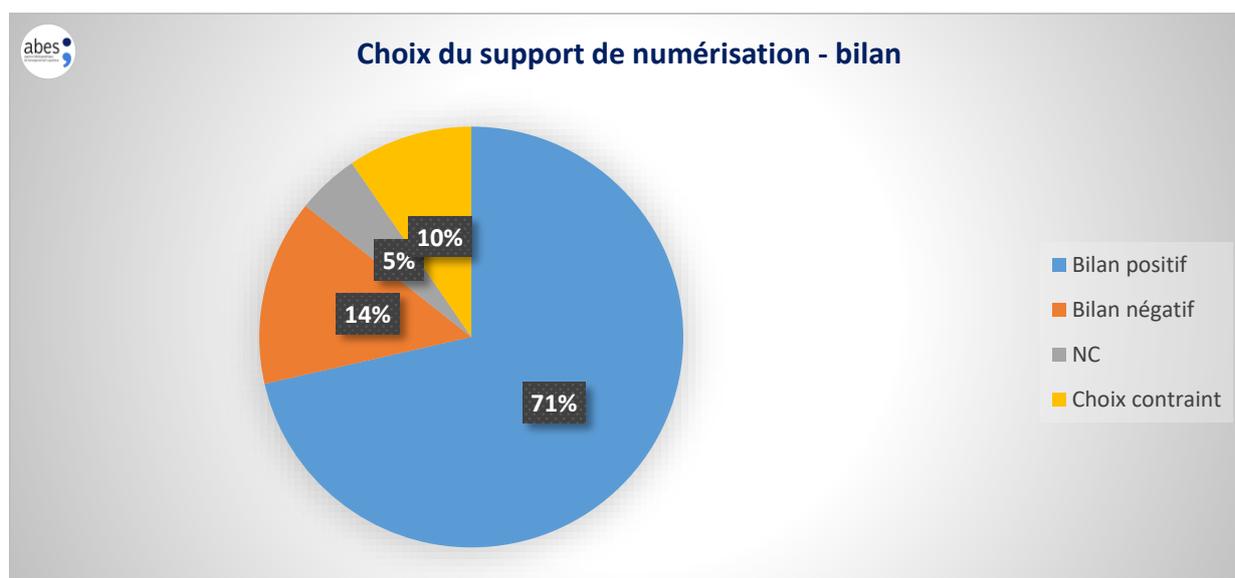
Le choix du recours aux microfiches, quant à lui, a été induit : par la **facilité de manipulation et de numérisation** du support, par le recours à **l'ANRT comme prestataire** de numérisation, par les contraintes exceptionnelles de numérisation en environnement amianté (cas des thèses polluées à l'amiante de Bordeaux Montaigne). A noter que l'un des établissements ayant eu recours aux microfiches comme support de numérisation a signalé le mauvais rendu des illustrations.

Question 14 : bilan du choix du support

71% des établissements tirent un bilan positif du choix de support de numérisation qu'ils ont effectué, dont l'un des deux établissements qui a eu recours aux microfiches. Les points positifs relevés par les établissements ayant eu recours à l'imprimé sont d'abord la **facilité d'utilisation** et le **bon rendu final**, mais aussi la **disponibilité** du support et le **caractère économique** de son utilisation. L'établissement qui a eu recours aux microfiches indique quant à lui que ce choix de support permet de ne pas détruire ou abîmer les exemplaires imprimés, en plus d'être économique et facile d'utilisation.

Les établissements qui tirent un bilan plus négatif de leur choix se plaignent quant à eux :

- d'un **rendu final moyen** : soit parce qu'ils ont numérisé des microfiches ; soit parce que la numérisation a été effectuée au début des années 2000 ; soit parce que la numérisation a été mal effectuée
- de la **difficulté à manipuler les thèses anciennes** imprimées, qui sont fragiles ou en mauvais état

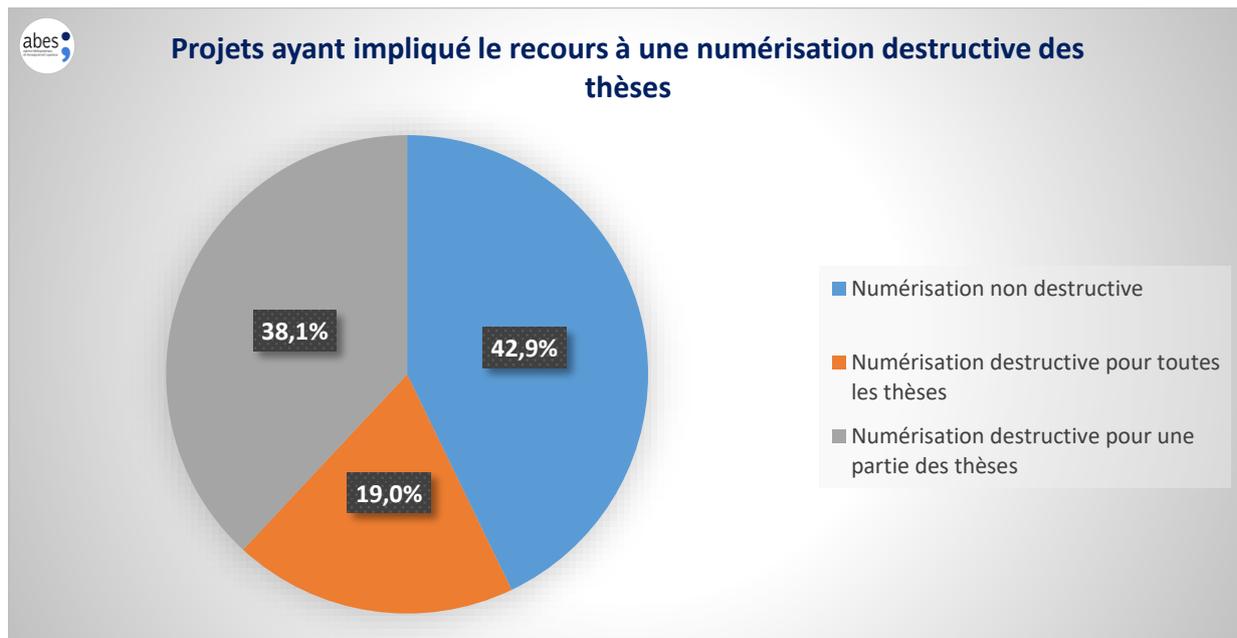


Question 15 : recours à la numérisation destructive

43% des établissements ont procédé à une numérisation **non destructive des thèses**.

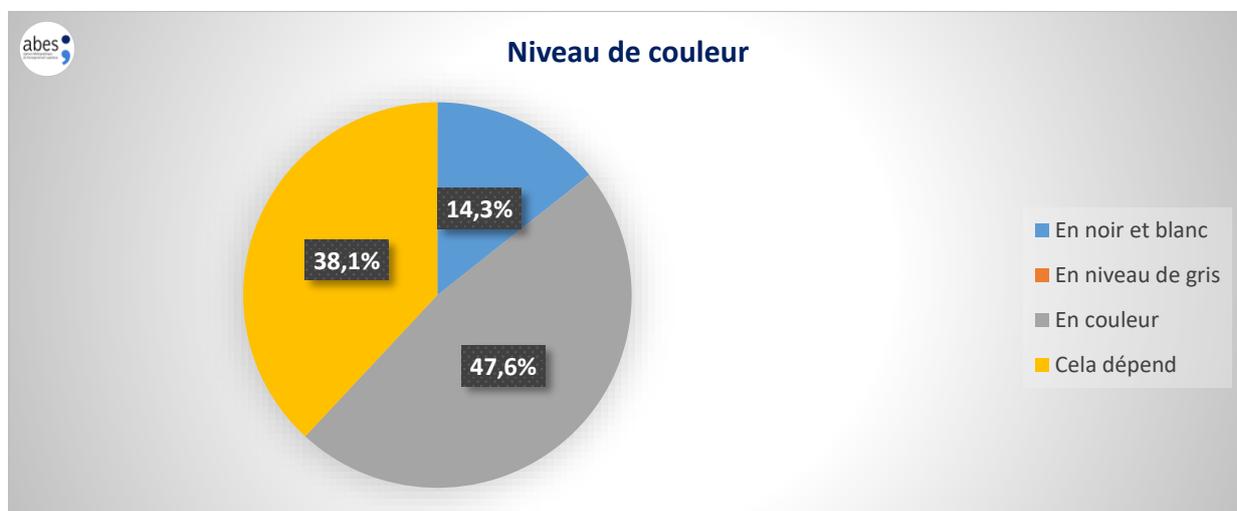
19% ont procédé à une numérisation destructive pour l'ensemble de leur fonds et **38% pour une partie seulement**.

Les établissements procèdent à une numérisation destructive lorsqu'ils disposent de plusieurs exemplaires imprimés de la thèse : l'un des exemplaires est conservé et le second exemplaire est détruit. Lorsqu'il n'existe plus qu'un exemplaire de la thèse, les établissements ne procèdent pas à sa destruction : ils le massicotent, pour permettre sa numérisation, puis le font relier, ou le conservent en feuilles, dans une pochette.



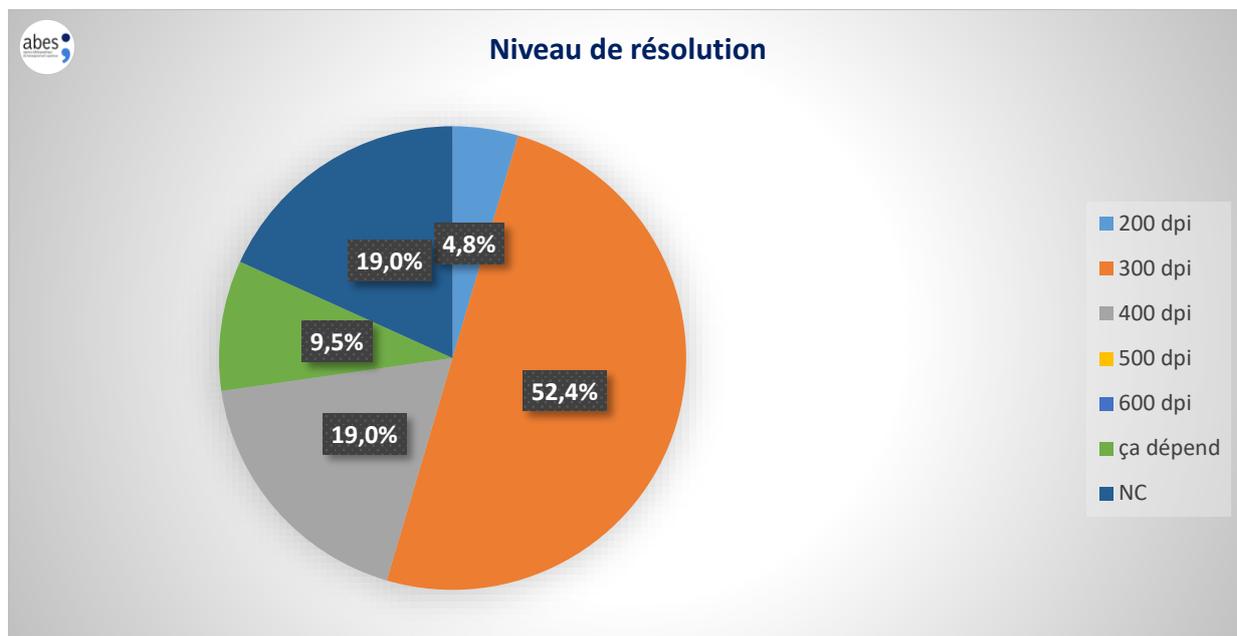
Question 16 : coloris des fichiers numérisés

Près de la **moitié des établissements** a procédé à une **numérisation en couleurs** des thèses. 14% ont procédé à une numérisation uniquement en noir et blanc. Les 38% restants ont choisi de numériser les thèses **majoritairement en noir et blanc**, mais de numériser **en couleurs la couverture, les pages illustrées, et les pages qui comportent des figures**. Certains bancs de numérisation détectent la présence de couleurs et gèrent automatiquement la numérisation en noir et blanc ou en couleurs.



Question 17 : résolution des fichiers numérisés

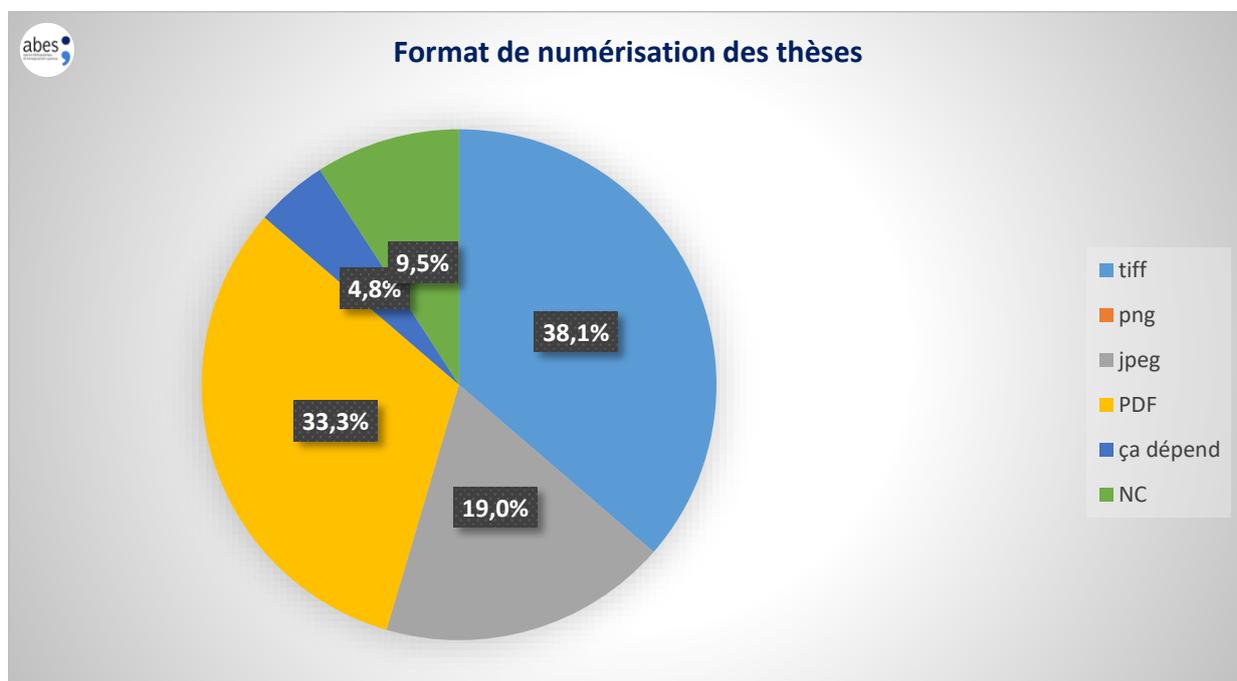
La majorité des établissements a procédé à une numérisation des thèses en **300 dpi** (52,4%) ou **400 dpi** (19%).



Question 18 : format(s) de numérisation

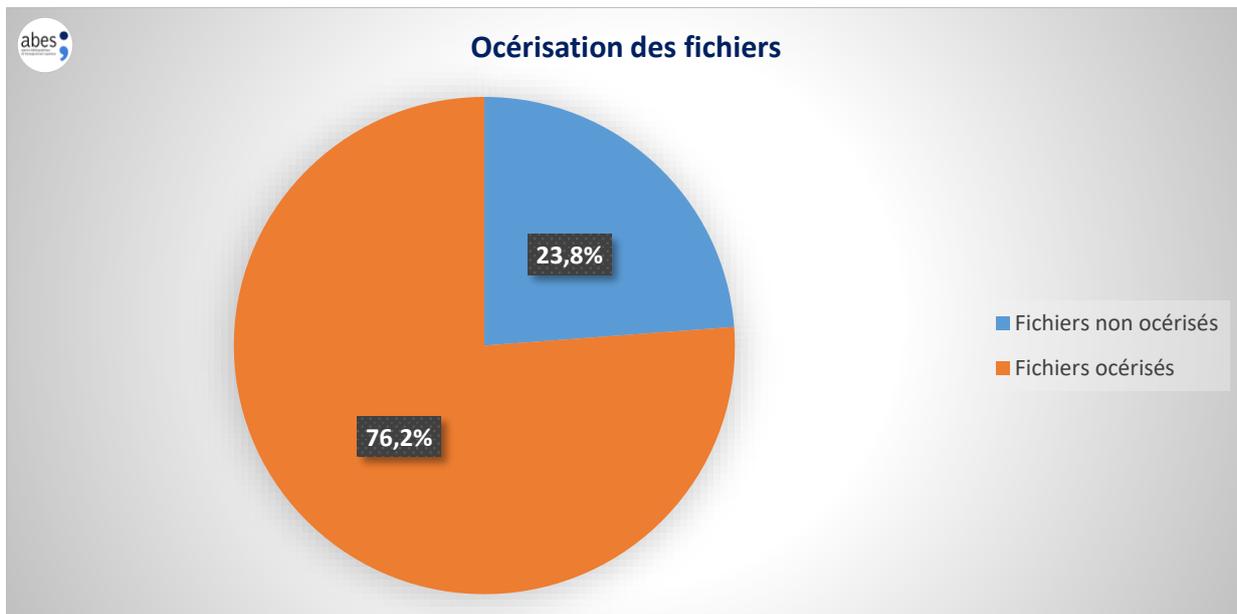
Les trois formats de numérisation utilisés sont : le **.tiff** (38%), le **.pdf** (33%) et le **.jpeg** (19%).

Dans le cas du PDF, le PDF/A a été utilisé dans 2 projets de numérisation sur 7.



Question 19 : océrisation des fichiers numérisés

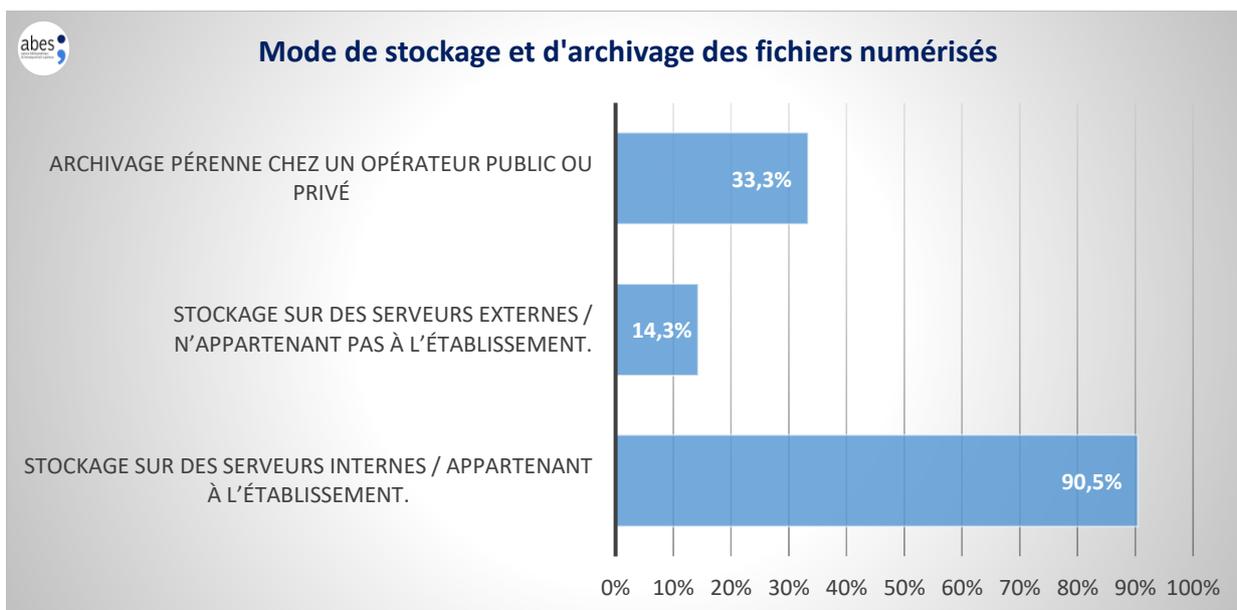
Les **trois quarts des projets** de numérisation s'accompagnent de l'océrisation des fichiers numérisés.



Conservation des fichiers numérisés

Question 20 : conservation des fichiers numérisés

Dans la plupart des cas (91%), les **fichiers numérisés sont simplement stockés sur des serveurs internes**. Néanmoins, un peu plus de **33%** des établissements ont procédé, parfois en plus d'un stockage des fichiers en interne, à un **archivage pérenne des documents : principalement au CINES**, mais aussi dans l'archive SPAR de la BnF lorsque cette dernière a réalisé la numérisation (lorsque la BnF subventionne sans réaliser elle-même la numérisation, il revient à l'établissement d'assurer l'archivage pérenne des fichiers).

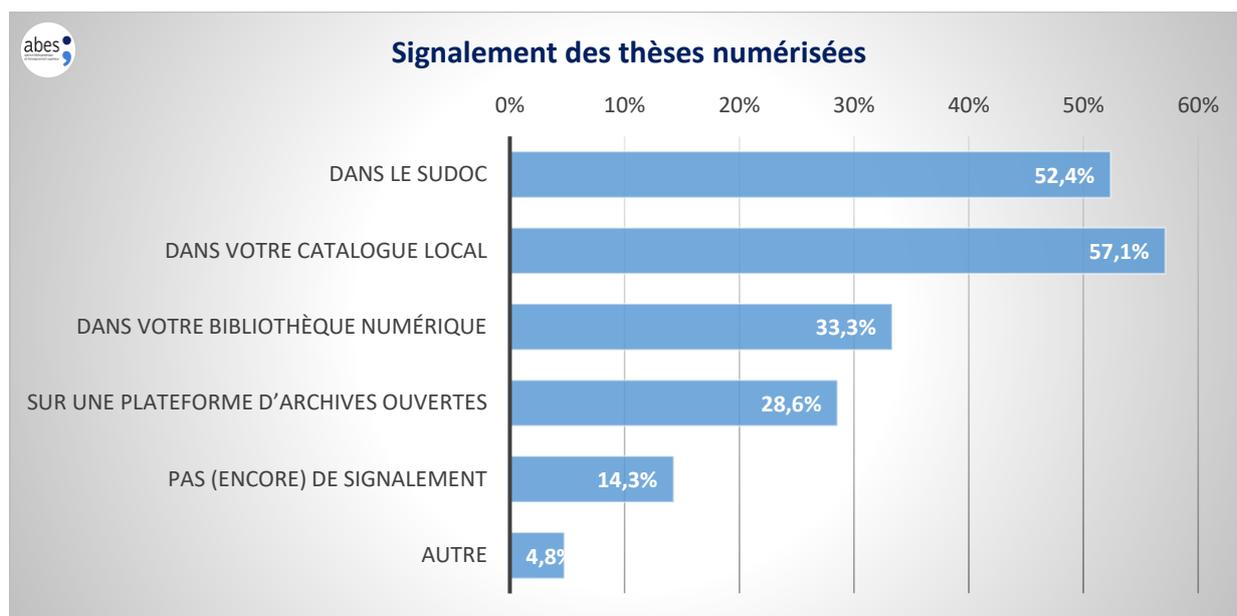


Signalement des fichiers numérisés

Question 21 : signalement des thèses numérisées

Un peu plus de la moitié des projets de numérisation des thèses aboutissent au **signalement des fichiers numérisés dans le Sudoc et dans le catalogue de l'établissement**. Le signalement se fait également dans les bibliothèques numériques des établissements, ou sur une plateforme d'archives ouvertes, HAL en l'occurrence.

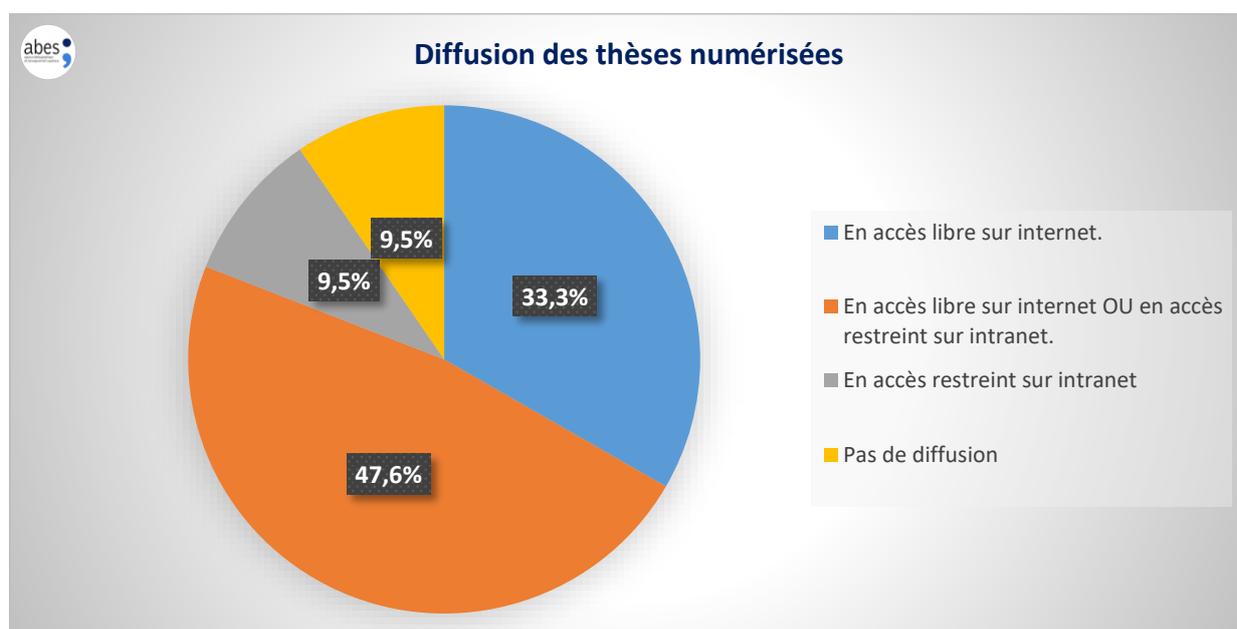
À noter : 14% des projets de numérisation ne s'accompagnent pas du signalement des fonds numérisés.



Diffusion des fichiers numérisés

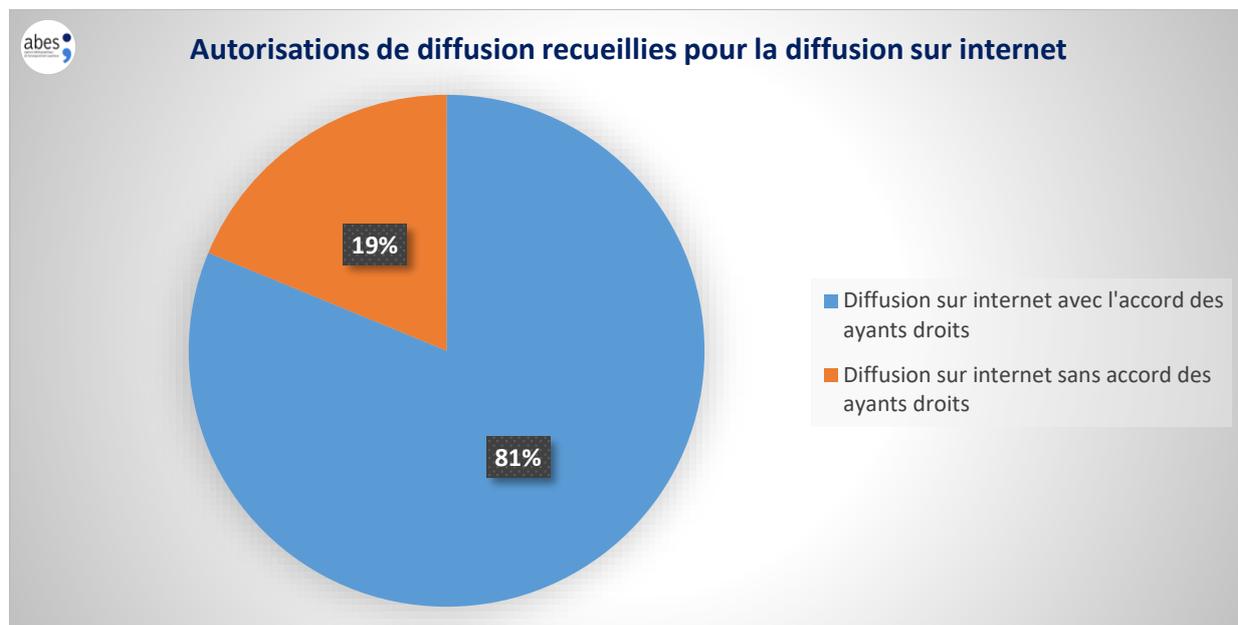
Question 22 : diffusion des fichiers

La grande majorité des établissements a choisi de diffuser les fichiers de thèses numérisées : 1/3 sur internet, 10% uniquement sur intranet et près de la moitié (47,6%) à la fois sur internet et sur intranet, en fonction des autorisations de diffusion données par les auteurs.

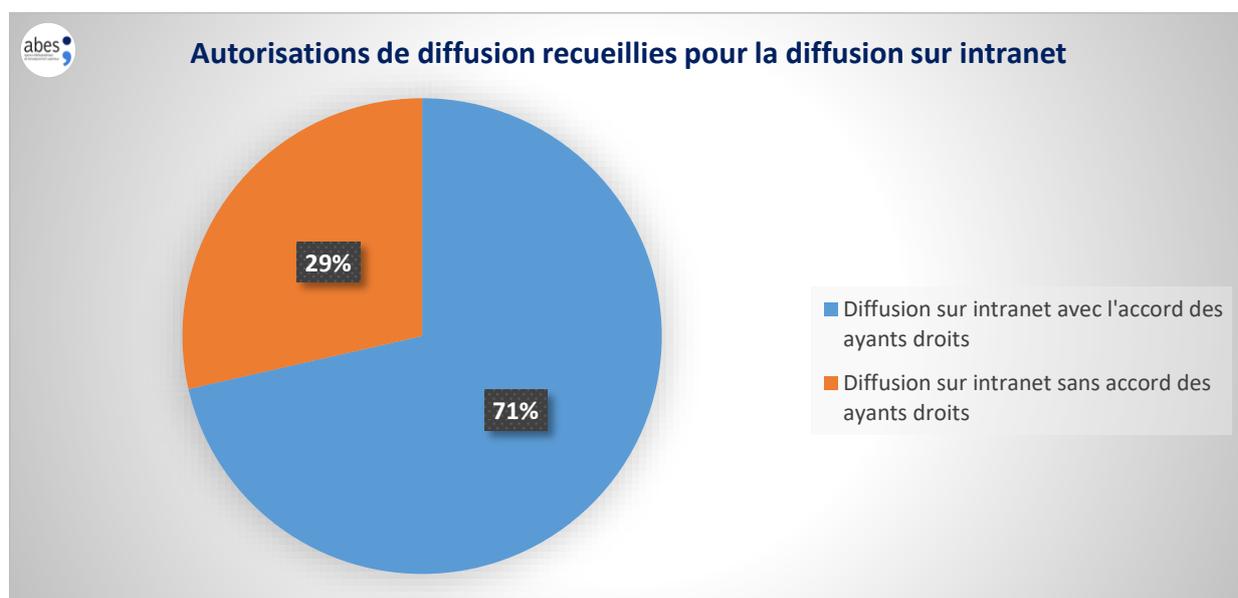


Questions 23 et 23 bis : recueil de l'accord des auteurs ou des ayants droits

Lorsque les établissements font le choix de diffuser les thèses sur internet, ils demandent, dans 80% des cas, l'autorisation aux ayants droits.



Lorsque les établissements font le choix de diffuser les thèses sur intranet, ils demandent, dans 70% des cas, l'autorisation aux ayants droits.



Question 24 : organisation du recueil des autorisations de diffusion

Dans le cas des établissements qui ont procédé à un double dépôt imprimé/électronique, les autorisations de diffusion ont pu être recueillies au moment du dépôt imprimé de la thèse. Mais dans la majorité des cas, les établissements ont recueilli les autorisations de diffusion a posteriori, dans le cadre du projet de numérisation.

La recherche des auteurs s'est effectuée avec l'aide de l'administration, des services de scolarité, des annuaires d'anciens étudiants, des informations disponibles sur les sites web institutionnels et des listes de diffusion internes aux établissements, mais aussi à partir de sources d'information externes : annuaire de l'ordre des médecins et de l'ordre des

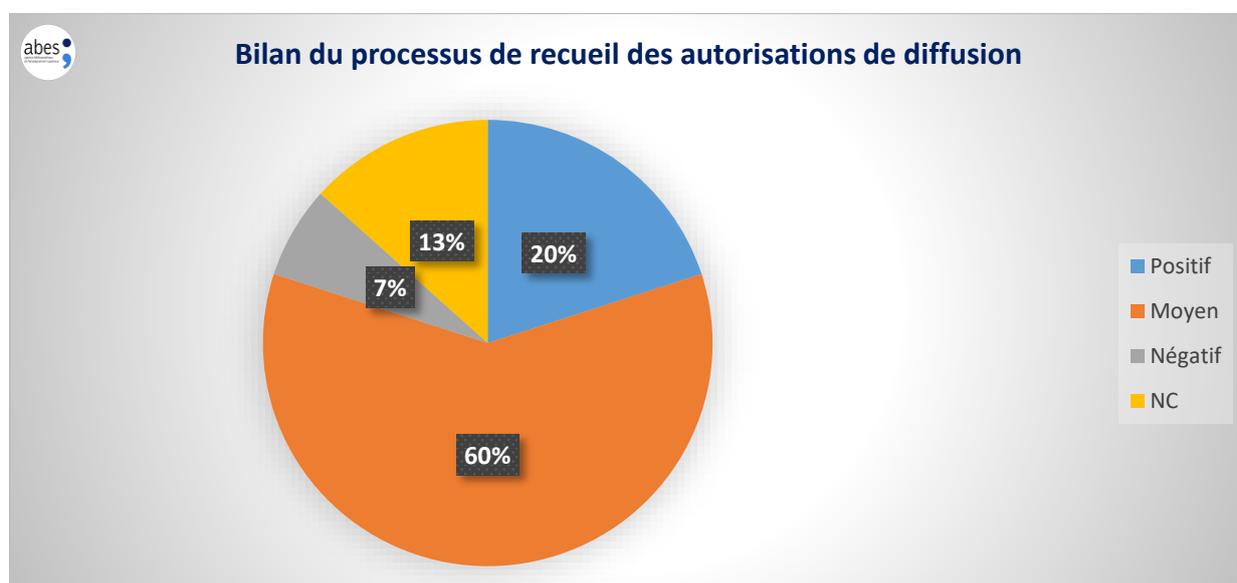
pharmaciens, réseaux sociaux, presse. Les docteurs sont ensuite **contactés par mail**, dans la plupart des cas, ou par téléphone.

Deux établissements ont également lancé un **appel général** aux docteurs sur le site du SCD ou de l'université.

Sans réponse de la part des docteurs, deux établissements ont fait le choix de **diffuser les thèses sur intranet**, un autre établissement a fait le choix de diffuser les thèses **sur internet**, la **diffusion étant assortie d'un système d'opt-out** pour les docteurs qui souhaiteraient s'y opposer a posteriori.

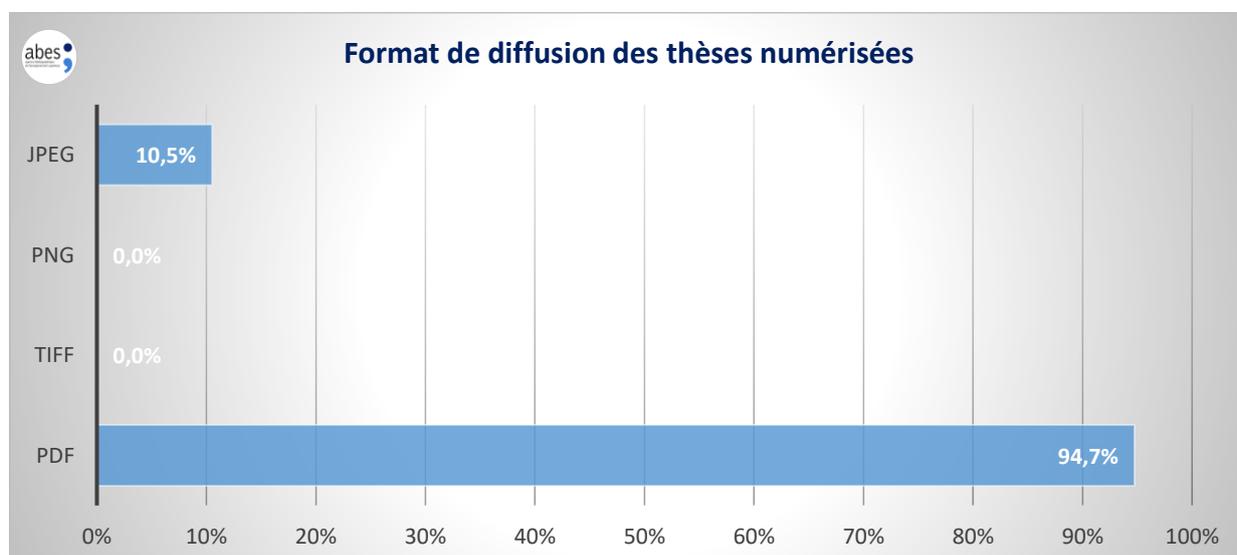
Question 25 : bilan de l'organisation du recueil des autorisations de diffusion

Parmi ceux qui ont dû retrouver et contacter les docteurs a posteriori, les **deux tiers** des établissements tirent un **bilan mitigé** du processus mis en place pour recueillir les autorisations de diffusion : ils soulignent le caractère particulièrement **chronophage** et **laborieux** de ce chantier, ainsi que le **faible taux de retours**.



Question 26 : format(s) de diffusion des fichiers numérisés

Le format de diffusion des fichiers utilisé est, pour la quasi-totalité des fonds de thèses numérisés, le format **.pdf**.

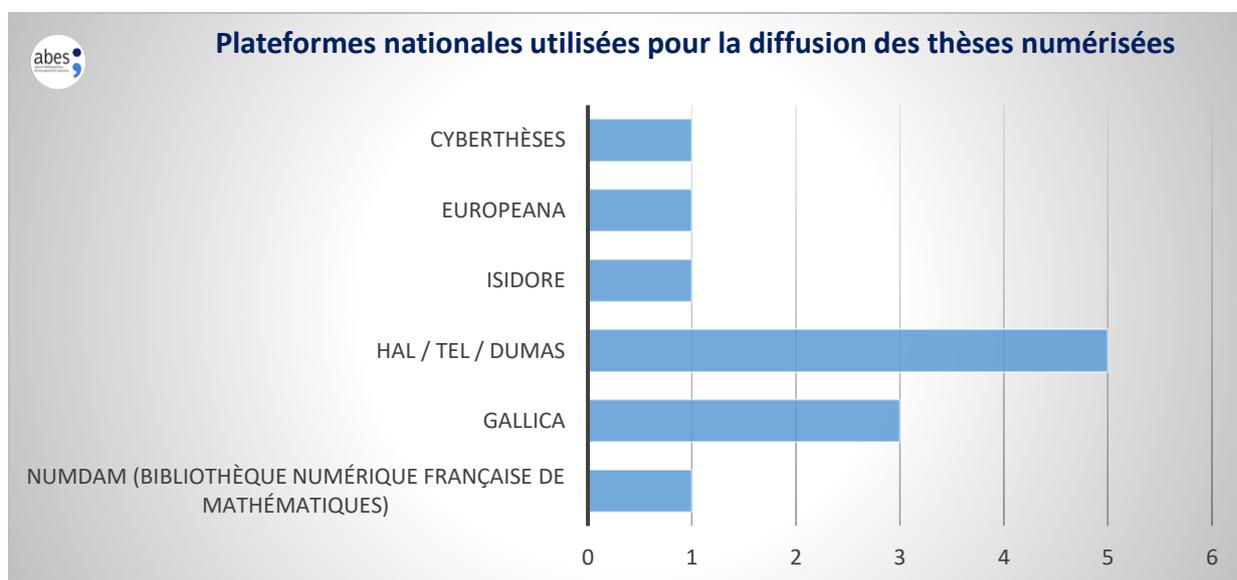
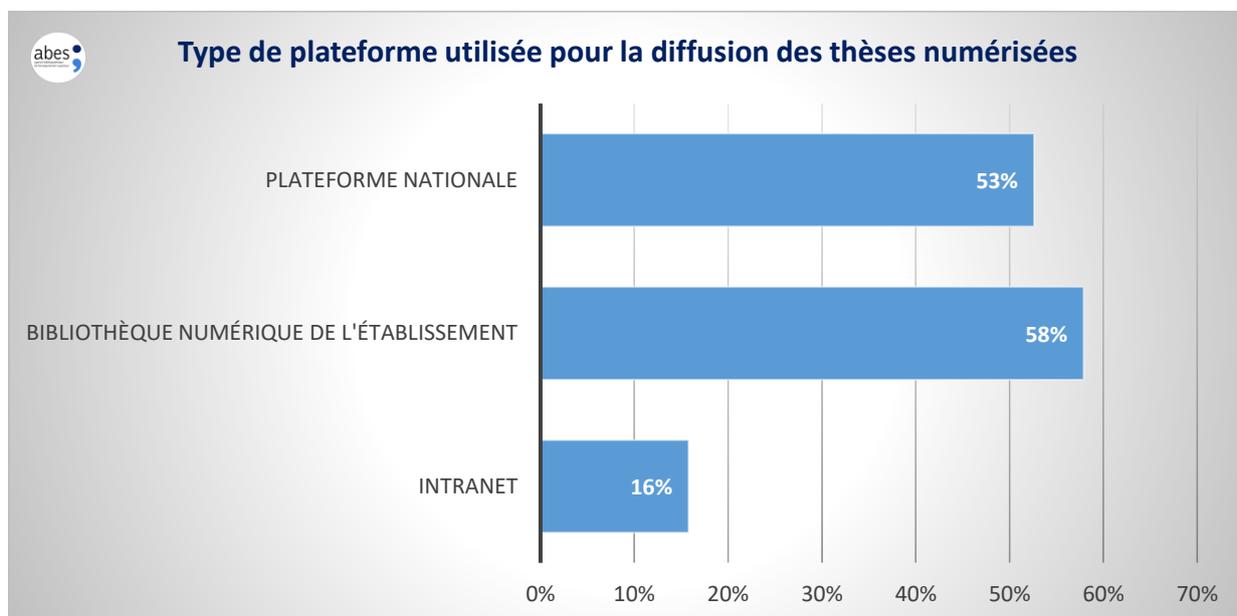


Question 27 : plateforme(s) de diffusion

58% des établissements diffusent les thèses numérisées sur leur propre **bibliothèque numérique**, 53% sur une **plateforme nationale**, et 16% **uniquement sur intranet**. Un quart des établissements combinent plusieurs modes de diffusion.

Les plateformes de diffusion nationales les plus utilisées sont :

- [HAL/TEL](#), maintenue par le [CCSD](#), pour la diffusion des thèses modernes
- [Gallica](#), maintenue par la [BnF](#), pour la diffusion des thèses anciennes.



Utilisation du corpus numérisé

Question 28 : services attachés aux corpus de thèses numérisées

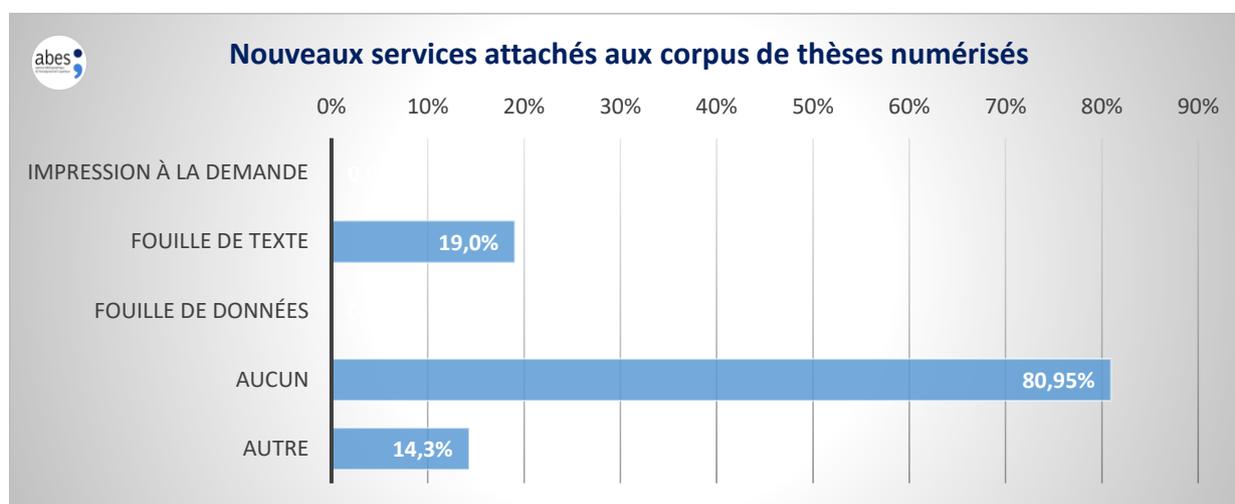
80% des établissements indiquent n'avoir rattaché aucun service particulier aux corpus de thèses numérisées.

4 établissements proposent quant à eux un service de **fouille de texte**, auquel peut s'adjoindre :

- l'utilisation des fichiers numérisés pour répondre aux **demandes de PEB** (dans un cas)
- la constitution de corpus et de bases de données, avec en projet l'enrichissement des métadonnées (dans deux cas)

Ainsi, le SCD de Lyon 1 indique avoir confié son fonds de thèses numérisées à un chercheur en informatique, qui utilise le corpus pour **expérimenter différents algorithmes destinés à enrichir automatiquement des métadonnées** :

- détection de liens vers des référentiels (VIAF, WorldCat, DBpedia, Wikidata, etc.) et des images
- détection automatique de mots-clés RAMEAU
- reconnaissance d'images
- détection de documents liés



Question 29 : usages du corpus de thèses numérisées

Six établissements indiquent disposer de **statistiques de consultation** de leur corpus de thèses numérisées, lesquelles sont **bonnes**.

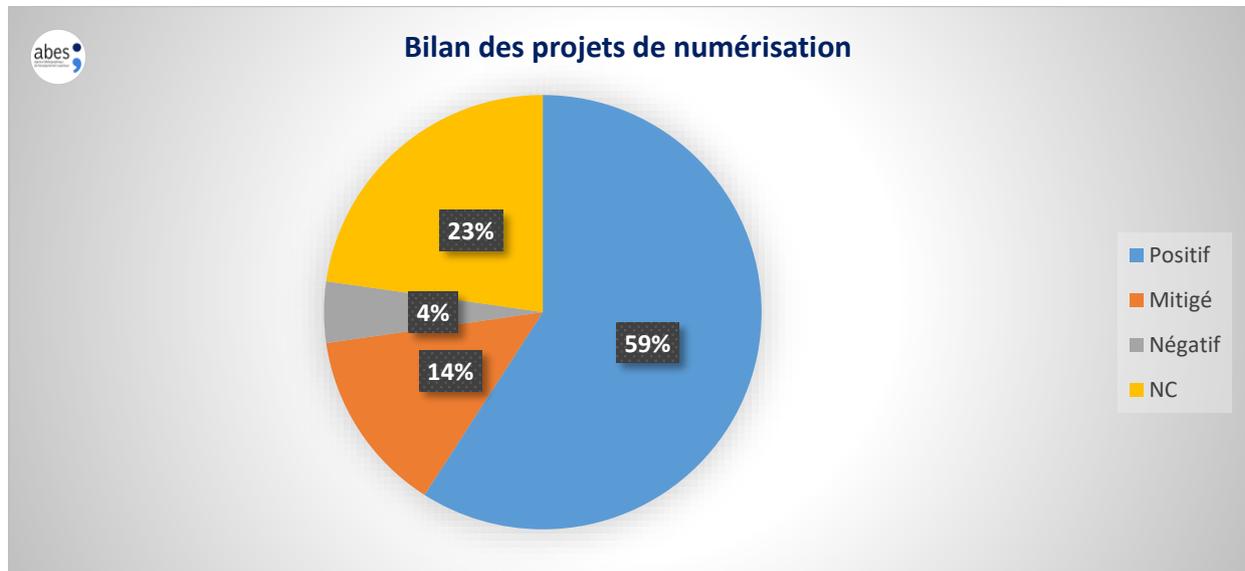
Sept autres établissements mentionnent **l'impact que la numérisation des thèses et leur diffusion ont eu sur le PEB** : les demandes de PEB ont diminué pour les thèses diffusées librement sur internet, alors qu'elles ont augmenté pour les thèses dont la version numérique n'est disponible que sur l'intranet de l'établissement.

Enfin, les corpus numérisés sont utilisés par les **chercheurs qui souhaitent étudier l'histoire des disciplines universitaires ou l'histoire de la thèse comme genre littéraire et écrit scientifique**.

Bilan

Question 30 : bilan du projet de numérisation des thèses

59% des établissements tirent un **bilan positif** de leur projet de numérisation des thèses, **14%** un **bilan mitigé**, et seulement **4%** un **bilan négatif**.



Les principaux **points positifs** évoqués sont : la **disponibilité** des documents, leur **visibilité**, la **valorisation** des travaux de recherche, le fait de répondre aux **besoins** des utilisateurs, l'aide à la **conservation** des documents que représente la numérisation, la **bonne qualité** des documents numérisés.

Certains établissements mentionnent également : la mise en **cohérence des collections**, avec la proposition d'un corpus complet de thèses électroniques ; l'accélération des **chantiers de rétroconversion** ; la complétion de **Gallica** ; le **désherbage** des exemplaires surnuméraires des thèses, destiné à libérer des espaces de stockage.

Les principaux **points négatifs** évoqués sont : le **caractère long, chronophage et laborieux des chantiers**, notamment la **recherche des auteurs** pour la collecte des autorisations de diffusion, mais également la préparation des corpus, la numérisation proprement dite et le signalement.

Les établissements mentionnent également : les **coûts** ; la **manipulation difficile des thèses anciennes**, fragiles et dans certains cas, le **non signalement** des collections et la **non diffusion** de certaines thèses ; l'**absence d'archivage pérenne** ; la qualité moyenne de la numérisation.

Synthèse des résultats de l'enquête

Les grandes tendances identifiées par l'enquête en termes de numérisation, diffusion, signalement et archivage des thèses sont les suivantes :

- *objectifs du projet* : **faciliter l'accès aux documents et valoriser les fonds**
- *pilotage* : assuré **par le SCD** de l'établissement
- *typologie des fonds numérisés* :
 - fonds de **volumétrie moyenne** (quelques centaines de thèses)
 - constitués de **thèses de doctorat soutenues depuis les années 1980s**
 - **tous types de disciplines** confondues
- *ressources humaines et financières* :
 - **projet financé** par l'établissement, la ComUE ou la BnF (thèses anciennes)
 - **recrutement** de plusieurs CDD sur certaines phases du projet, pendant quelques mois, ou d'un CDD sur toute la durée du projet
- *numérisation* :
 - pas de collecte préalable des fichiers de thèses
 - appel à un **prestataire extérieur** pour effectuer la numérisation
 - **numérisation destructive des exemplaires imprimés surnuméraires** des thèses, avec conservation d'un seul exemplaire en rayon
 - soit en **couleurs**, soit en **noir et blanc avec des illustrations en couleurs**
 - en **300** ou **400 dpi**
 - au format **.tiff**, ou **.pdf**
 - avec **océrisation**
- *archivage* :
 - **pas d'archivage pérenne**
 - stockage sur des serveurs internes
- *signalement* :
 - dans le **Sudoc**
 - dans le **catalogue local** de la bibliothèque
- *diffusion* :
 - en **libre accès sur internet** ou en **accès restreint sur intranet**
 - sur **autorisation des ayants droits** recueillies par mail
 - au format **.pdf**
 - sur la bibliothèque numérique de l'établissement, et/ou sur TEL
- *utilisation du corpus* :
 - pas de nouveaux services attachés au corpus numérisé dans la majorité des cas (service de fouille de texte dans 4 cas seulement)
 - consultation satisfaisante
 - impact sur le PEB
 - utilisation par les chercheurs

Conclusion

La présente enquête a permis de recenser les différents projets de numérisation des thèses menés par les établissements de l'enseignement supérieur et de dresser un bilan des choix techniques et stratégiques opérés. Elle constitue également le point de départ de l'étude de faisabilité que l'Abes et l'Université de Lille vont mener en 2021 dans le cadre d'un projet national visant à numériser l'ensemble des thèses de doctorat soutenues en France depuis 1985.

L'expérience acquise par les établissements, les points positifs et négatifs relevés, notamment en ce qui concerne le choix des supports à numériser ainsi que la question de la diffusion des fichiers et de la collecte des autorisations de diffusion, vont permettre d'alimenter la réflexion et d'instruire à la fois l'opportunité de mener à bien un tel projet, ainsi que les conditions, à la fois techniques, logistiques, matérielles et financières, de faisabilité du projet.



abes ;

Enquête 2020 réalisée par le service des Thèses
département Métadonnées et Services aux Réseaux - Contact : theses@abes.fr

En collaboration avec le SCD de l'Université de Lille



Agence Bibliographique de l'Enseignement Supérieur

avenue du Professeur Jean-Louis Viala - CS 84308 - 34193 Montpellier Cedex 5 - Tél. 04 67 54 84 10

