

Refonte du module statistiques de HAL

Résultats de l'enquête auprès des administrateurs de portail

Une priorité des développements HAL pour 2018 concerne la refonte du module de statistiques. Cela englobe l'amélioration des performances, celle de la fiabilité des indicateurs, l'ergonomie de l'interface d'interrogation et l'exploitation des résultats.

Afin de correspondre au mieux aux besoins de ses utilisateurs, le CCSD a proposé une enquête destinée aux administrateurs de portail. Elle est centrée sur les besoins en indicateurs statistiques des établissements et des institutions, concernant les données gérées dans une archive ouverte.

L'enquête s'est déroulée du 1er au 30 mars 2018 : un lien vers un questionnaire en ligne a été adressé à tous les utilisateurs ayant le privilège « administrateur » dans la base de données, soient 411 adresses au total.

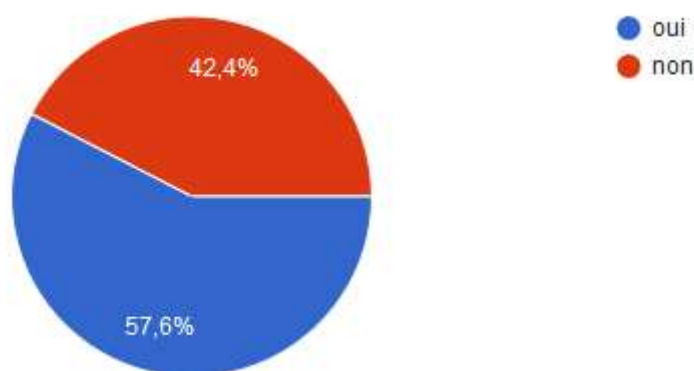
L'objectif en procédant ainsi était de procéder à un nettoyage de la base de données qui s'est révélé moins important que prévu puisque seulement 10 profils ont été mis à jour : adresse mail erronée, retour de l'utilisateur nous informant qu'il n'est plus administrateur.

66 personnes ont répondu et nous les remercions vivement de leur participation. La dernière question permettait de préciser le nom de son portail : certaines personnes administrant plus d'un portail, cela porte à 69 le nombre total de portails représentés par les réponses à ce questionnaire (57 portails dont la liste est en annexe, 12 réponses sans le nom du portail, une personne déclare n'être que gestionnaire de collections). **Soit une proportion de 55 % du nombre actuel de portails.**

Pour recueillir au mieux les besoins, la plupart des questions étaient volontairement ouvertes.

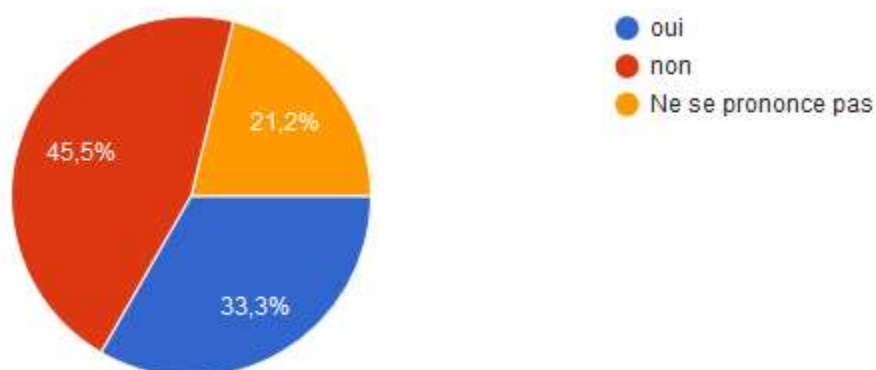
Les premières questions visaient à recueillir plus spécifiquement les besoins en données statistiques, institutionnelles comme spécifiques à l'activité d'administrateur d'un portail.

Question 1 : Fournissez-vous les données concernant l'archive ouverte pour le formulaire ESGBU (Enquête Statistique Générale auprès des Bibliothèques Universitaires) ?



La majorité des 66 répondants sont concernés par l'enquête annuelle à destination des bibliothèques universitaires.

Question 2 : Votre établissement ou institution vous demandent-ils des indicateurs sur l'archive ouverte autres que ceux demandés dans le formulaire ESGBU (nombre de notices et documents fulltext, leur accroissement annuel, nombre de visites sur le portail, nombre de téléchargements)



En-dehors des informations demandées dans l'ESGBU, seulement un tiers des 66 répondants sont sollicités par leur établissement ou institution pour d'autres indicateurs. La question suivante permettait d'exprimer librement les demandes institutionnelles.

Question 3 : Si vous avez répondu oui cette question, merci de préciser lesquels et à quelle fréquence

21 réponses.

Afin de mieux analyser les réponses à cette question, nous différencions les indicateurs selon plusieurs types :

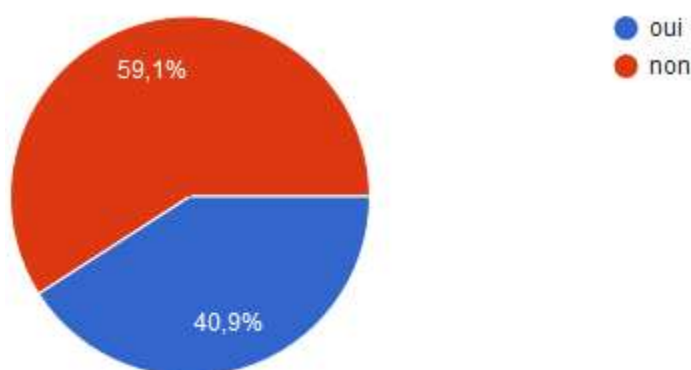
- Les statistiques de consultation des documents (notices et fulltext) : 5 réponses
 - Répartition des consultations par pays
 - Nb de visites des notices
 - Nb de téléchargements des fulltexts
 - Consultations par referrer (page par laquelle on arrive sur le document : moteur de recherche, etc)
- Les statistiques de consultation d'un site
 - Nb de visites d'un portail
- Les statistiques d'usage/stock (liées aux données de HAL, pas de consultation)
 - Nb de dépôts réalisés dans un portail
 - Nb de notices
 - Nb de fulltext
 - Nb de documents par type de dépôt (notice/fulltext)
 - Accroissement du nombre de documents / notices
 - Nb de documents par laboratoire d'une institution
 - Répartition du nb de notices/fulltext
 - par type de document
 - par domaine
 - par collection liée à un portail
 - par sous-structure d'une structure
 - par programme de recherche
 - par équipe de recherche
 - par langue du document
 - Taux de fulltext
 - Nb de copublications avec X
 - Nb de notices modérées dans un portail
 - Indicateurs par projets IDEX
 - Indicateurs par institutions d'une COMUE
 - Répartitions par auteurs interne / externe

Un certain nombre de demandes rejoignent les indicateurs de l'enquête ESGBU. La fréquence à laquelle sont demandés des indicateurs est majoritairement annuelle.

D'autres réponses externes au périmètre de HAL ont été formulées :

- Indicateurs liés aux critères HCERES
- Indicateurs liés au classement de Shanghai (dépendants du WOS)
- Indicateurs liés aux critères WOS
- Indicateurs sur les productions d'ExtractHAL

Question 4 : Pour votre gestion régulière de l'archive, avez-vous besoin d'indicateurs autres que ceux demandés par votre établissement ou institution ?



Une majorité des 66 répondants déclarent ne pas avoir besoin de statistiques supplémentaires. La question suivante permettait aux autres répondants de formuler librement leurs besoins pour gérer au mieux leur portail.

Question 5 : Si vous avez répondu oui cette question, merci de préciser lesquels et à quelle fréquence

24 réponses, seulement 3 personnes n'ont pas détaillé leurs besoins.

De la même façon que la question 3, nous différencions les réponses par type de statistiques :

- Consultation des documents
 - Temps de consultation des vidéos
- Statistiques d'usage
 - Nombre de notices / fulltext
 - dans un portail
 - par collection liées à un portail
 - par tampon (département, auteur, etc)
 - Accroissement du nombre de dépôt par mois
 - Nombre d'articles modérés par modérateur
 - Nombre de dépôts par déposant
 - Répartition par portail de dépôt pour un corpus
 - Répartition des dépôts
 - par type de document
 - par langue du document
 - par domaine (racine ou sous-domaine)
 - avec ou sans DOI
 - par département
 - par programme de recherche
 - par année de publication
 - par année de dépôt
 - par type de licence
 - par statuts dans HAL

- Proportion d'auto-archivage
- Nombre de contributeurs actifs d'une structure
- Accroissement du nombre d'idHAL créés dans la structure par an
- Accroissement du nombre de contributeurs sur 1 an
- Nombre de collaborations nationales / européennes / internationales

Une réponse synthétise sous forme de bilan les indicateurs nécessaires :

- Bilan par laboratoire [Evolution des dépôts, répartition par type de dépôt, par type de document, par auteur, dépôts réalisés par les auteurs eux-mêmes, consultations de fiches, téléchargements]
- Bilan par axe de recherche du labo [Evolution des dépôts, répartition par type de dépôt, par type de document, par auteur, dépôts réalisés par les auteurs eux-mêmes, consultations de fiches, téléchargements]

Statistiques sur les utilisateurs

- documentaliste de laboratoire, forts déposants, limitation des indicateurs à la liste des auteurs du labo
- ratio nombre (enseignants)-chercheurs d'une structure / publications ?

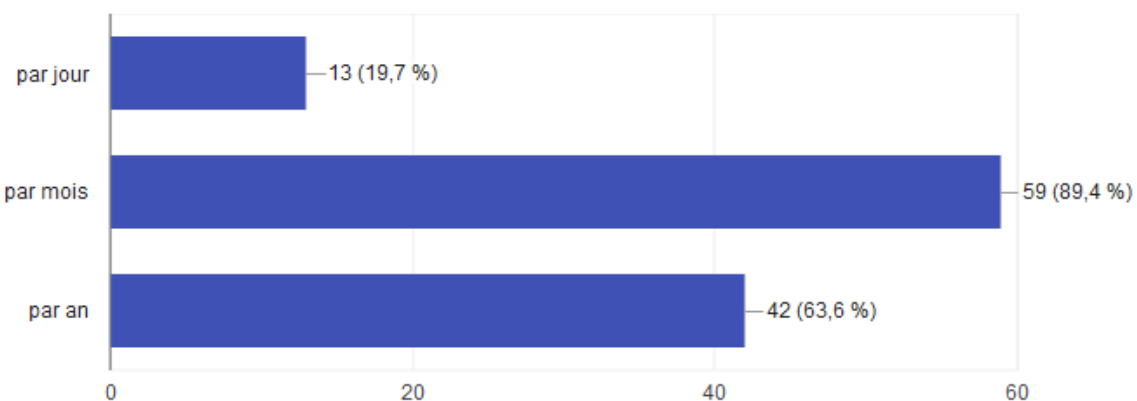
Une réponse formule une donnée externe au périmètre de HAL :

- Nombre de publications référencées dans le WOS

En dehors des contenus, trois fonctionnalités ont été évoquées :

- Export du CV (*fonctionnalité déjà présente : widget à choisir par l'auteur*)
- Export CSV des indicateurs
- Nombre/Liste de dépôts avec des métadonnées non valides

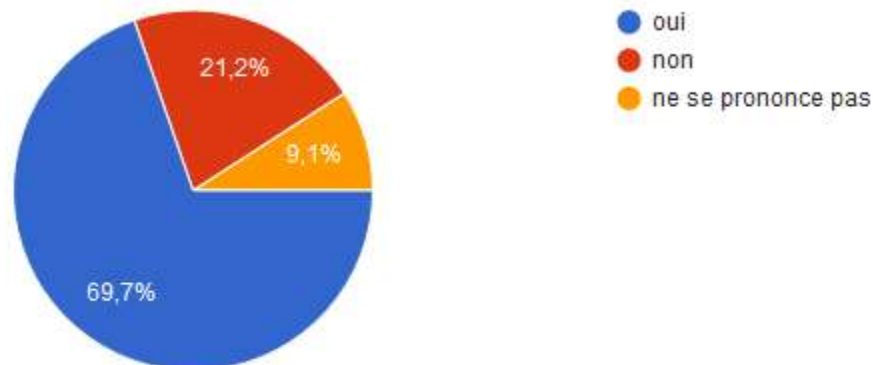
Question 6 : Trouvez-vous utile d'avoir des statistiques de consultation du portail et des documents



La granularité mensuelle des statistiques de consultation convient à près de 90% des 66 répondants et la majorité trouve utile les statistiques de consultation par an. Le besoin de statistiques par jour est à relativiser car 6 des personnes qui trouvent cette granularité utile ont pourtant répondu non aux questions 3 et 5.

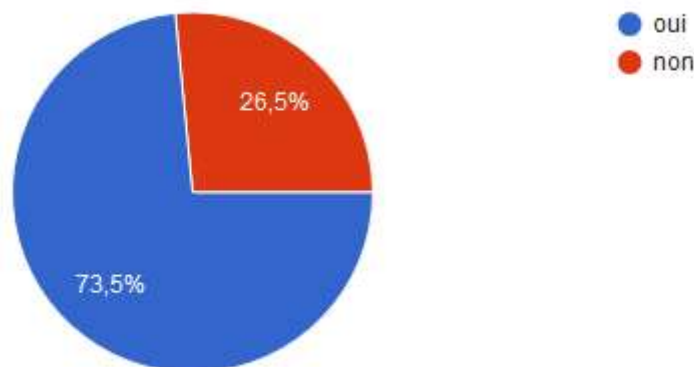
Les questions suivantes visent à recueillir l'avis des administrateurs sur les outils actuellement disponibles.

Question 7 : Utilisez-vous l'interface de statistiques actuelle de HAL ?



Question 8 : Si vous avez répondu oui, utilisez-vous les filtres pour constituer le corpus de données (sélection des ressources) ?

49 réponses



Une grande majorité utilise la fonction de filtres pour construire le corpus sur lequel obtenir des statistiques. La question suivante permettait de préciser quelles sont les métadonnées utilisées fréquemment.

Question 9 : Si vous utilisez l'interface actuelle, quelles sont les métadonnées que vous utilisez le plus souvent pour obtenir les statistiques ?

28 réponses

- Type de document (13 réponses)
- Date de publication (6 réponses)
- Structure (7 réponses)
- Revue (5 réponses)
- Auteur (5 réponses)
- Type de dépôt (6 réponses)
- Date de dépôt

- Projet ANR
- Projet Européen
- Contributeur (Déposant)
- Couleur Sherpa/Romeo
- Consultation des ressources
- Evolution du nombre de publication au cours de l'année
- Auto-archivage
- Portail de dépôt
- Pays de provenance
- Titre de la conférence
- Etat dans le référentiel
- Code de collection(s) pour obtenir toutes les unités d'un établissement
- Collection
- Domaine

Question 10 : Si vous exportez les données, merci de préciser pour quels usages

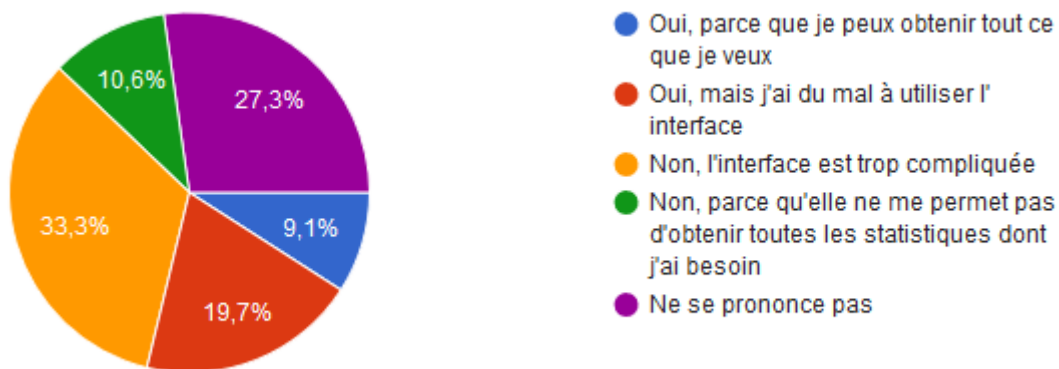
29 réponses

- Rapport statistique annuel
- Bilans statistiques.
- Enquêtes ESGBU
- Bilans annuels du laboratoire
- Pour les enquêtes, pour l'HCERES
- Perspective d'évaluation HCERES
- Pour le pilotage des services à la recherche
- Rapports d'usage pour les unités qui ont une collection, stats sur le portail pour la direction de l'Université
- Demande de rapport de la direction rapport annuel de la bibliothèque, rapport Couperin
- A terme pour des évaluations, rapports d'activité
- Production de liste via ExtracHAL
- Indicateurs de suivi pour le comité de direction
- Analyse
- Suivi des dépôts par structure
- Analyse bibliométrique
- consultation des dépôts du portail (notices / téléchargements)
- Pour faire des comparaisons avec les données du WoS ou d'autres bases biblio
- Evolution des dépôts du portail ou d'une collection HAL de labo (14 collections au total) et consultation des dépôts sur une période ; etc.
- Constitution de la BDD des publications du laboratoire
- Recensement des dépôts
- Capitalisation de données dans une base externe (analyse bibliométrique)
- Pour les retravailler dans Excel
- Construction de tableaux de bord, réalisation de tableaux de suivi
- Présentation du portail et projection des futures ressources à déposer
- Data visualisation pour des présentations (graphique, illustrations)
- Mise en forme graphique des données
- J'exporte systématiquement pour traiter les données dans un fichier excel

(production d'infographies pour alimenter la page d'accueil du portail)

- Demandes ponctuelles de la VP recherche sur les dépôts dans HAL par les laboratoires par exemple
- CV d'enseignant-chercheur
- Filtre et comptage
- Générer des graphiques

Question 11 : Etes-vous satisfait de l'interface actuelle ? (A noter : La refonte du module statistiques ne limitera plus à 25 000 résultats maximum et améliorera la détection des robots)



La formulation de la question retirait les principales critiques généralement exprimées : la limite du nombre de documents qui compose le corpus sur lequel on veut obtenir des statistiques, et l'élimination des hits de consultation en provenance des robots.

Le répondant devait faire un choix parmi 5 items. Un tiers trouve l'interface trop compliquée : elle a en effet été conçue pour être très paramétrable (constitution du corpus avec les filtres, liste complète des métadonnées, etc.), le risque étant qu'elle peut être difficile à prendre en main : 19,7% ont du mal à utiliser l'interface. 10,6% ont besoin d'autres données qui ne sont pas proposées.

Question 12 : Utilisez-vous la plateforme Piwik pour le suivi d'audience de votre portail ?



Cette plateforme n'est proposée systématiquement aux nouveaux portails que depuis 2017 : ce n'est donc pas étonnant que la majorité ne la connaisse pas.

Pour ceux qui se servent de cette plateforme, la question suivante permet de formuler librement les données qu'ils utilisent.

Question 13 : Si vous avez répondu oui, quelles statistiques vous sont le plus utiles ?

7 réponses

le nombre de visites sur le portail

nombre de visites

celle demandée par l'ESGBU / ERE

Nb visiteurs uniques; Nb d'actions

Visites, pages vues, provenances, téléchargements, titres des pages

Nombre de visiteurs, Nombre de visiteurs uniques

Nombre de visiteurs, provenance, types de référents

Annexe : Liste des portails

Campus AAR	IMT Atlantique
AAI(Archives Audiovisuelles de l'INALCO)	Lab-STICC
CENTRALE SUPELEC	Lirmm Mines Albi
E.N.S	Montpellier SupAgro
EDF	<i>Nous ne disposons que d'une collection</i>
EHESP	HAL Rennes 2
HAL Agrocampus Ouest	HAL-UDL
HAL AMU	Publi-EIAH
HAL CSTB	EIAH
HAL ENAC	Telearn
HAL INSA Toulouse	Sciences Po
HAL Lyon 3	Université Paris Nanterre
HAL Normandie Université	Univ-Montpellier
HAL Paris 1	URCA (Université de Reims Champagne
HAL PARIS 3	Ardenne)
HAL Paris 13	
HAL Paris 8	
HAL Polytechnique	
HAL Pastel	
HAL UCA	
HAL UG	
HAL UNIV-PARIS 5 (HAL DESCARTES)	
HAL UPEC UPEM	
HAL UT2 hal-univ-tlse2	
HAL Univ Brest	
HAL-Anses	
HAL-Artois	
HAL-CEA	
HAL-ENSTA	
HAL-IFPEN	
HAL-INRIA	
HAL-INSU	
HAL-IRD	
HAL-LAAS	
HAL-Obspm	
HAL-pasteur	
HAL-RIIP	
HAL-SHS	
HAL-Univ-Nantes	
HAL-UPS	
HAL-UVSQ	
HAL AgroSup Dijon	
HAL BNF	
IFP Energies nouvelles	