

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs

Analyse de l'enquête

Ecole Normale Supérieure de Lyon

préparé par	Emmanuel Quemener
contrôlé par	
approuvé par	<p>_____ Date: _____</p>
reference	ENSL-Storage4labs2PDS-100528
version	0 draft
date de version	28 mai 2010
nom de document	ENSL-Storage4labs2PDS-100528.odt

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

Mise à jour

Date	Version	Etat	Pages	Raisons du changement
Initiale	0	Brouillon		
28/05/10	0.1	Brouillon		

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

Table des Matières

1 Introduction.....	6
2 Documents applicables.....	6
3 Glossaire.....	6
4 Démarche.....	6
5 Modélisation du circuit des données numériques.....	7
5.1 Les processus.....	7
5.2 Les données.....	7
5.3 Les « formes » d'entreposage de ces données.....	7
6 Questionnaires.....	8
6.1 De la démarche analytique aux questionnaires.....	8
6.2 Une approche en deux temps.....	8
7 Résultats préliminaires.....	9
7.1 Réponses aux questionnaires.....	9
7.1.1 Premier bilan.....	9
7.1.2 Éléments d'analyse.....	9
8 Analyse.....	9
8.1 Analyse des besoins ENSMédia.....	9
8.1.1 Un contexte.....	9
8.1.2 Des contraintes techniques fortes.....	10
8.1.3 Des volumes considérables.....	10
8.1.4 Une croissance dopée par la HD et la fusion.....	10
8.1.5 Les besoins de la création graphique.....	10
8.2 Analyse de besoins d'UNIS.....	11
8.3 Analyse des besoins de la bibliothèque.....	11
8.3.1 Besoins de la Photothèque.....	11
8.3.2 Besoins du SID.....	11
8.3.3 Besoins de la BIU.....	12
8.3.4 Besoins de la Bibliothèque ENS.....	12
9 Cumul de tous les besoins en stockage.....	12
10 Des spécifications fonctionnelles aux spécifications techniques. .	14
10.1 Vers la clôture du triptyque de l'étude.....	14

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

10.2	Spécifications fonctionnelles.....	14
10.2.1	Pour le stockage.....	14
10.2.2	Pour la sauvegarde.....	15
10.2.3	Pour l'archivage.....	15
11	Retour à la lettre de mission.....	15
12	Conclusion.....	16
13	Annexes.....	17
13.1	Questionnaire « stockage pour les plates-formes d'acquisition».....	17
13.2	Questionnaire « stockage pour les plates-formes de traitement ».....	17
13.3	Questionnaire « stockage pour l'exploitation des résultats ».....	18

Index des illustrations

Illustration 1:	Modélisation du circuit de données numériques.....	8
Illustration 2:	Besoin par entité (en Go) pour chacune des 4 années.....	13
Illustration 3:	Besoin en stockage cumulé par entité pour les 4 années.....	13
Illustration 4:	Croissance du besoin en stockage (en Go).....	14

Index des Tables

Table 1:	Besoin en stockage pour le pôle de diffusion des savoirs, par unité (en Go).....	12
Table 2:	Besoin en stockage pour le pôle de diffusion des savoirs (en Go).....	12
Table 3:	Synthèse des besoins en stockage/sauvegarde/archivage en To.....	16

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

1 Introduction

Fin 2009, la Direction de la Recherche a reçu de la part des laboratoires de biologie de l'établissement une demande de financement pour une infrastructure de stockage.

La Direction de la Recherche a alors mandaté l'auteur pour évaluer les besoins de stockage, de sauvegarde et d'archivage de tous les laboratoires de l'établissement, site Jacques Monod. Le rapport de cette étude a été proposé à la Direction le 31 mars 2010.

A la suite de cela la Directrice des Systèmes d'Information a souhaité que le même travail soit réalisé pour les laboratoires du site Descartes, d'une part, et pour le pôle de diffusion des savoirs, d'autre part.

L'auteur de ce document a donc été, une nouvelle fois, sollicité pour établir une étude en collaboration avec deux personnes du service informatique. La lettre de mission rédigée pour l'occasion présentait les objectifs de ce nouveau travail.

2 Documents applicables

Lettre de mission E. Quemener	http://perso.ens-lyon.fr/emmanuel.quemener/documents/lettre_mission_e.quemener.pdf
Étude des besoins de stockage, site Monod	http://perso.ens-lyon.fr/emmanuel.quemener/documents/ENSL-Storage4labs-100415.pdf

3 Glossaire

CQOCOQP	Allographe de QOQOCCP (Qui fait Quoi ?, Où ? Quand ? Comment ? Combien ? Et Pourquoi ?)
FLV	Extension de documents au format Flash Vidéo ou vidéo Flash
MP4	Extension de documents au format MPEG-4, une norme de codage d'objets audiovisuels
DVD	Digital Versatile Disc : support de stockage. Les DVD Vidéo permettent, dans leur version standard, un stockage de 2 heures
TIFF	Tag(ged) Image File Format : format d'image au codage sans perte
JPEG	Joint Photographic Experts Group : format d'image au codage avec perte
DGesco	Direction Générale de l'Enseignement Scolaire
EPI	Espace Pédagogique Intégré (Espace Numérique de Travail à destination des élèves et enseignants de l'établissement)
ITA	Personnel Ingénieur, Technicien ou Administratif
UNIS	Universités des Sciences : unité numérique de diffusion scientifique de l'ENS-Lyon
BIU	Bibliothèque Inter Universitaire
SID	Système d'Information Documentaire
BDD	Bases des Données
NAS	Network Attached Storage (stockage accessible par le réseau)
Mb/s	Mega bits par seconde
Mo	Mega Octet (normalement 2 ²⁰ soit 1048576 octets, mais standardisé en 2007 à 1 million d'octets)
Go	Giga Octet (normalement 2 ³⁰ soit 1073741824 octets, mais standardisé en 2007 à 1 milliard d'octets)
To	Tera Octet (normalement 2 ⁴⁰ soit 1099511627776 octets, mais standardisé en 2007 à mille milliards d'octets)
Po	Peta Octet (normalement 2 ⁵⁰ soit 1125899906842624 octets, mais standardisé en 2007 à un million de milliards d'octets)
GE	Gigabit Ethernet (standard réseau permettant un transfert de données symétrique à une vitesse de 125 Mo/s)

4 Démarche

Cette étude reprend autant que possible les éléments produits lors de la précédente étude (Étude des besoins de stockage, site Monod, page 6): le schéma synoptique a été sommairement adapté, les questionnaires légèrement modifiés et réordonnés.

Étant donné la contrainte de temps posée par la fourniture du livrable, l'approche initiale d'un questionnaire en ligne dans lequel les individus saisissent leurs doléances n'a pas été privilégié.

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

Il a donc été préféré le contact direct, sous forme d'un rendez-vous pris avec les personnes les plus à même d'exprimer les besoins en ressources matérielles et logicielles.

5 Modélisation du circuit des données numériques

« *Un bon croquis vaut mieux qu'un long discours.* » disait Napoléon Bonaparte.

Un schéma synoptique (Illustration 1, page 8) permet de visualiser rapidement le circuit suivi de leur genèse à leur diffusion.

Plutôt que de s'axer dans un premier temps sur la nature des données, considérons tout d'abord les processus, abordons ensuite les données.

5.1 Les processus

Quatre processus manipulent ces données à l'aide de plates-formes :

- l'acquisition : la « génération » des données brutes, leur indexation, ... ;
- le traitement : leurs réduction, synthèse, analyse, indexation, ...
- la valorisation : leur transformation en contenu scientifique. Cela constitue le « cœur de métier » du chercheur ;
- l'exploitation : leur diffusion sur tous les médias scientifiques.

5.2 Les données

Ainsi, les données, à la source ou à destination de ces processus, sont finalement de 3 natures distinctes :

- les données « brutes » : directement issues des acquisitions ;
- les données « traitées » : premiers résultats ;
- les données « valorisées » : résultats à destination des travaux de publication.

5.3 Les « formes » d'entreposage de ces données

De plus, le schéma synoptique présente un entrepôt de ces données avec une triple nature :

- le stockage : accès direct aux données. L'utilisateur peut, à tout moment, disposer d'un accès à la totalité de cet espace ;
- la sauvegarde : duplication complète/partielle, synchrone/asynchrone de l'espace de stockage dans un endroit apportant une sécurité aux données en cas de perte du stockage. L'utilisateur final ne peut pas disposer d'un accès au domaine de sauvegarde. Il doit en référer au gestionnaire de la sauvegarde pour disposer à nouveau, sur un espace de stockage, de ses données ;
- l'archivage : état du stockage complet ou partiel, à un instant dans le passé. Ces états sont d'une fréquence et d'une pérennité à définir. L'utilisateur peut demander à restaurer, dans un espace de stockage dédié, un état de son stockage du passé.

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

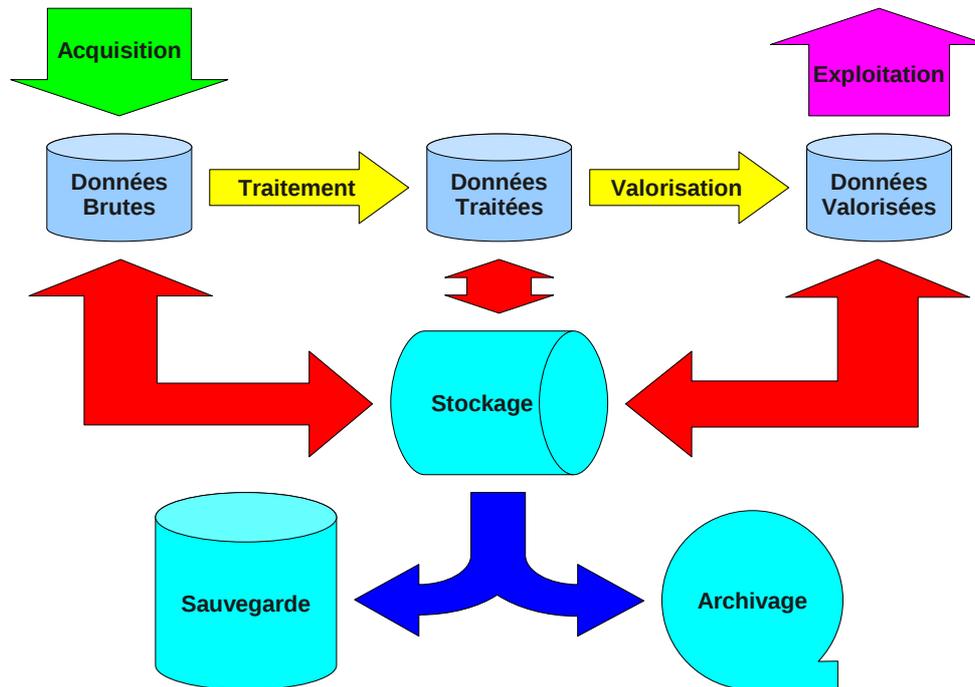


Illustration 1: Modélisation du circuit de données numériques

Pour qu'aucune information importante n'échappe à l'étude préliminaire, a été choisie la collecte basée sur une démarche analytique classique. Elle propose de répondre aux sept questions élémentaires : Pourquoi ? Quoi ? Qui ? Quand ? Où ? Combien ? Comment ?

6 Questionnaires

6.1 De la démarche analytique aux questionnaires

La démarche analytique utilisée, plus connue sous l'allographe CQOCOQP, a été « transformée » en questions simples les plus générales possibles pour que les utilisateurs de plate-forme puissent, quelle que soit leur discipline, répondre simplement.

Les trois questionnaires sont disponibles en annexe de ce document :

- l'acquisition (page 17, chapitre 13.1) ;
- le traitement (page 17, chapitre 13.2) ;
- l'exploitation (page 18, chapitre 13.3).

6.2 Une approche en deux temps

La contrainte de temps ne permettait pas l'utilisation des questionnaires électroniques, comparable à celle entreprise lors de la première étude.

La première étape a été d'identifier les individus susceptibles de répondre aux questionnaires.

La seconde étape fut une rencontre destinée à saisir directement les réponses dans le feuillet d'analyse.

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

7 Résultats préliminaires

7.1 Réponses aux questionnaires

7.1.1 Premier bilan

Au 26 mai 2010, date de rédaction de ce document, sur les quatre composantes du pôle de diffusion des savoirs, trois avaient répondu à l'enquête : ENSMédia, UNIS et la bibliothèque.

Cependant, par rapport aux réponses collectées électroniquement pour la première enquête concernant le site Monod ou directement concernant les laboratoires du site Descartes, seule la rencontre organisée avec ENSMédia a permis d'obtenir des résultats assez fins dépassant la simple expression de besoin en volume de stockage : ainsi, les éléments complémentaires, associés au cycle des données, à leur pérennité, aux différents acteurs engagés dans leur création ou leur exploitation, n'ont pas été collectés, faute d'une réelle discipline dans l'évaluation de l'expression de besoins, essentiellement liée au manque de temps.

Pour UNIS, le principal frein à la moisson d'information a été la perte sèche de toute la connaissance, depuis maintenant plusieurs années, des principaux acteurs informatiques de l'ancien service Pr@tic. Il était très difficile d'apprécier de la qualité des données disponibles : quelles étaient leur âge, leur croissance et leur position dans l'ensemble des processus. Aussi, le seul élément exploitable, pour UNIS côté Monod, demeure l'analyse volumétrique des données existantes.

7.1.2 Éléments d'analyse

Étant donné la forte différence de retour entre les différentes activités du pôle de diffusion des savoirs, il a été préféré une analyse en deux temps.

D'abord, le premier temps reprend les éléments de l'enquête, notamment les questions connexes liées aux contextes.

Ensuite, un second temps est destiné à analyser, à partir des estimations numériques, les volumes nécessaires en stockage, leur projection par extrapolation sur les trois prochaines années.

8 Analyse

8.1 Analyse des besoins ENSMédia

Les éléments suivants ont été constitués lors d'une entrevue le 12 mai 2010 dans les locaux de ENSMédia.

8.1.1 Un contexte

Le pôle ENSMédia a subi la « révolution » du tout numérique ces tous derniers mois. Si l'activité de montage vidéo via des outils informatiques est un domaine que le pôle maîtrise depuis plusieurs années, l'acquisition directement en numérique sur des supports autres que des bandes date seulement des derniers mois.

Parallèlement à cette croissance exponentielle des volumes de données à générer puis traiter, le pôle ENSMédia a été complètement oublié de l'infrastructure de stockage déployée sur le site Descartes début 2009. A aucun moment, le responsable de ce service n'a été consulté pour exprimer ses besoins. Il a donc également, par voie de conséquence, été très négligemment occulté lors de la mise en place de l'infrastructure.

L'infrastructure de prises de vue de ENSMédia repose, en plus de son studio d'enregistrement, sur trois amphithéâtres disposant de plusieurs caméras numériques, d'une régie numérique et d'un stockage de toutes ses données numériques directement sur disque dur ou sur support DV (à bande).

De plus, les anciens supports analogiques, à bande notamment, sont numérisés pour ne pas perdre leur contenu.

Côté utilisateur, les personnes réalisant les prises de vue sont des étudiants, des enseignants chercheurs et des personnels ITA. Il en est de même pour leur exploitation dans les traitements et leur exploitation pour la diffusion sur le Web. Cependant, les données mises en ligne sont à destination de tout public.

La durée de conservation des données doit être la plus longue possible, y compris pour les prises de vue. Ainsi, en cas de données non accessibles directement, le délai jugé raisonnable pour disposer à nouveau de ces données est de 24

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

heures (un délai comparable à celui consistant à aller chercher une bande dans une armoire et en extraire le contenu). De ce fait, il est raisonnable de penser que toutes les données de ENSMédia doivent toujours rester disponibles sur des supports à accès aléatoire (en substance des supports comparables à des disques durs) et font partie intégrante du stockage tel qu'il a été défini.

8.1.2 Des contraintes techniques fortes

Les processus de prise de vue (acquisition) et de montage (traitement et encodage) sont des opérations sollicitant beaucoup d'accès au support de stockage, sur de très gros volumes de données. Les débits indispensables pour ce type d'opérations dépassent plusieurs centaines de Mb/s.

Dans l'éventualité d'une utilisation de stockage distant accessible par le réseau informatique traditionnel, il sera indispensable que l'interconnexion entre client de volume disque et serveur soit, au minimum, en Gigabit Ethernet.

Une métrologie fine devra être menée pour ajuster au mieux l'usage de ces volumes à ces utilisations en vidéo numérique.

8.1.3 Des volumes considérables

Concernant la volumétrie, les prises de vue annuelles sont les suivantes :

- 100 heures de vidéo haute définition avec un débit de 24 Mo/s, soit 86400 Mo/heure : environ 9 To ;
- 10 conférences de 25 heures à un débit de 25 Mb/s, soit 11250 Mo/heure : environ 3 To ;
- 1000 heures de cours avec un débit de 25 Mb/s, soit 11250 Mo/heure : environ 11 To.

Ainsi, sur une année, la production exige **un volume nécessaire pour l'acquisition avoisinant les 23 To**.

Les traitements se réalisent sur une régie numérique à l'aide de logiciels, essentiellement de l'éditeur Adobe.

Les prises de vue haute définition servent généralement à la réalisation de DVD, donc disposant d'un débit de 4,5 à 9 Mb/s. Ces prises de vue, tout comme les autres prises de vue bas débit, sont utilisées pour réaliser des montages destinés à la diffusion en ligne, avec des débits beaucoup plus raisonnables : du FLV à 64 Ko/s ou du MP4 à 512 Ko/s.

- 100 heures de DVD à 9 Mb/s : environ 405 Go
- 1350 heures (1000 + 250 + 100) heures de vidéo en ligne au format FLV à 512 Kb/s : environ 311 Go

Ainsi, sur une année, les **opérations de montage et de diffusion nécessitent un volume de 716 Go**.

En conclusion, **le volume nécessaire, au jour d'aujourd'hui, se monte à 24 To**.

Ces volumes sont actuellement stockés sur des supports internes ou externes, à accès direct (disques durs) ou séquentiels (bandes LTO).

8.1.4 Une croissance dopée par la HD et la fusion

La croissance des acquisitions devrait être conditionnée par deux facteurs bien distincts :

- le souhait des professeurs du site Monod de numériser leurs cours et de les placer en ligne d'une part ;
- l'utilisation systématique de vidéo HD (et un volume 4 fois plus important) et le bannissement des bandes DV, d'autre part.

La croissance à 3 ans doit donc, vraisemblablement, décupler. Cependant, devant le colossal volume exprimé (240 To par an), il paraît plus raisonnable de considérer le besoin comme doublant la première année (par la fusion), triplant la suivante et quadruplant la troisième année. Le cumul sur les 4 années se monte alors à **240 To**.

8.1.5 Les besoins de la création graphique

L'unité de création graphique de ENSMédia a également exprimé ses besoins à partir du bilan des réalisations de l'année passée.

Sachant qu'autour de 170 articles de communication événementielle ont été réalisés avec près d'une centaine d'affiches, il peut être estimé très grossièrement le besoin en stockage des compositions (calques) : à près de 500 Mo l'unité, ces réalisations nécessitent autour de **100 Go** d'espace par an.

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

8.2 Analyse de besoins d'UNIS

Les éléments suivants ont été constitués :

- lors d'une rencontre avec la responsable d'UNIS, Nathalie Alazard, le 15 avril 2010 ;
- à partir de la redirection d'une demande de Raphaëlle Caillet auprès de la responsable d'UNIS, le 12 mai 2010.

Les équipements actuels d'UNIS distribuent essentiellement des sites de diffusion des connaissances, notamment les sites DGesco, et l'EPI.

Les volumes utilisés, comptant les cours mis à disposition et les bibliothèques de vidéo, sont de l'ordre de **2 To** sur les équipements du site Monod. En considérant une croissance raisonnable de 10% par an, nous parvenons à un montant à 3 ans de près de **2,7 To**.

Sur le site Descartes, des volumes de l'ordre de **1,5 To** sont également utilisés. Une partie est stockée sur le PC de l'utilisateur et donc sauvegardé, jusqu'à présent, par le service informatique. Semaine 19, il a été demandé par la Direction des Systèmes d'Information de libérer cet espace. Un espace de substitution est donc maintenant nécessaire pour garantir la pérennité de ces données.

De plus, il a été analysé par ces sites DGesco que, en 5 mois, environ 350 Go d'espace disque ont été utilisés. En extrapolant, nous parvenons à un montant de **2,5 To** à 3 ans.

8.3 Analyse des besoins de la bibliothèque

La bibliothèque a exprimé ses besoins pour quatre segments :

- la photothèque ;
- le Système d'Information Documentaire ou SID ;
- la Bibliothèque Inter Universitaire ou BIU ;
- la Bibliothèque ENS.

Les besoins se sont exprimés lors d'une rencontre entre Thomas Bellebois, de la DSI et les personnels de la bibliothèque, le 20 mai 2010, concernant la photothèque.

Les besoins des autres segments ont été détaillés dans un message de Danielle Roger, le 25 mai 2010.

8.3.1 Besoins de la Photothèque

Les besoins de la photothèque s'articulent autour de trois bases d'images :

- Première base « territoire et formes urbaines » et : 10 To de besoins à 3 ans
 - 300 Mo par image (130 Mo + 80 Mo en TIFF et plusieurs déclinaisons JPEG) et 10000 images : 3 To ;
 - Une croissance presque comparable au besoin initial chaque année ;
- Seconde base « mémoires et actualités » : 8 To de besoins à 3 ans
 - Besoin initial de 5 To
 - Croissance de 1 To par an
- Troisième base « art et culture » : 4 To
 - Besoin initial : 3 To
 - Croissance de 1 To à 3 ans

Les besoins de la photothèque sont à hauteur de 11 To initialement et 22 To à 3 ans.

8.3.2 Besoins du SID

Le besoin pour le SID n'a été exprimé que via le produit Digitool : « En ce qui concerne le stockage sur le serveur Digitool, on peut évaluer le besoin à environ **600 Go** minimum ».

Compte tenu de l'extrapolation proposée (de 15% d'augmentation à la troisième année, comptant une croissance

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

linéaire), nous parvenons à un montant de **780 Go à trois ans**.

8.3.3 Besoins de la BIU

Les besoins du SID ont été très sommairement exprimés : « **700 Go** pour 17000 fichiers TIFF».

Compte tenu de l'extrapolation proposée (de 15% d'augmentation à la troisième année, comptant une croissance linéaire), nous parvenons à un montant de **910 Go à trois ans**.

8.3.4 Besoins de la Bibliothèque ENS

Les besoins de la bibliothèque ENS ont également été très sommairement exprimés : « 12 000 pages avec trois vues chacune en format TIFF». Soit, « Par extrapolation à partir des 375 premières pages numérisées (nous sommes en tout début de l'opération), on trouve 921 Go», soit autour de **1 To**.

Compte tenu de l'extrapolation proposée (de 15%), nous parvenons à un montant de **1200 Go à trois ans**.

9 Cumul de tous les besoins en stockage

Le tableau 1, page 12 présente les besoins exprimés des différentes composantes du pôle de diffusion des savoirs, par domaine d'activité.

Unité	Domaine	Année 0	Année 1	Année 2	Année 3	4 années
ENSMédia	Vidéo	24000	48000	72000	96000	240000
	Infographie	100	100	100	100	400
UNIS	Site Monod	2000	200	220	242	2662
	Site Descartes	1500	840	840	840	4020
Bibliothèque	Photothèque	11000	3000	4000	4000	22000
	SID	600	30	60	90	780
	BIU	700	35	70	105	910
	Bibliothèque ENS	1000	50	100	150	1300
Total		40900	52255	77390	101527	272072

Table 1: Besoin en stockage pour le pôle de diffusion des savoirs, par unité (en Go)

Unité	Année 0	Année 1	Année 2	Année 3	4 Années
ENSMédia	24100	48100	72100	96100	240400
UNIS	3500	1040	1060	1082	6682
Bibliothèque	13300	3115	4230	4345	24990
Total	40900	52255	77390	101527	272072

Table 2: Besoin en stockage pour le pôle de diffusion des savoirs (en Go)

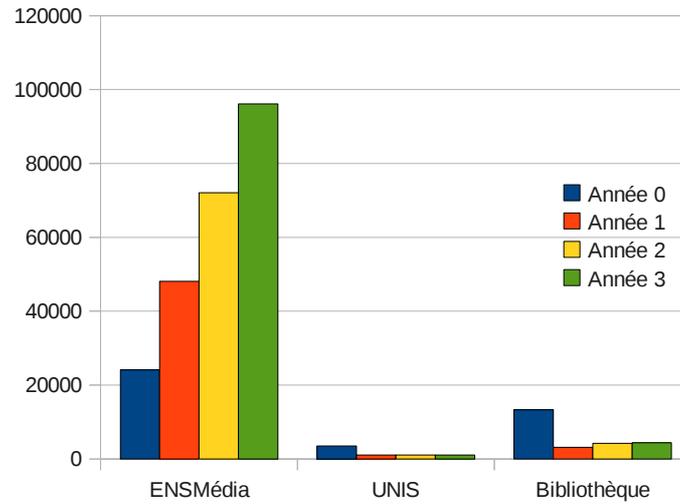


Illustration 2: Besoin par entité (en Go) pour chacune des 4 années

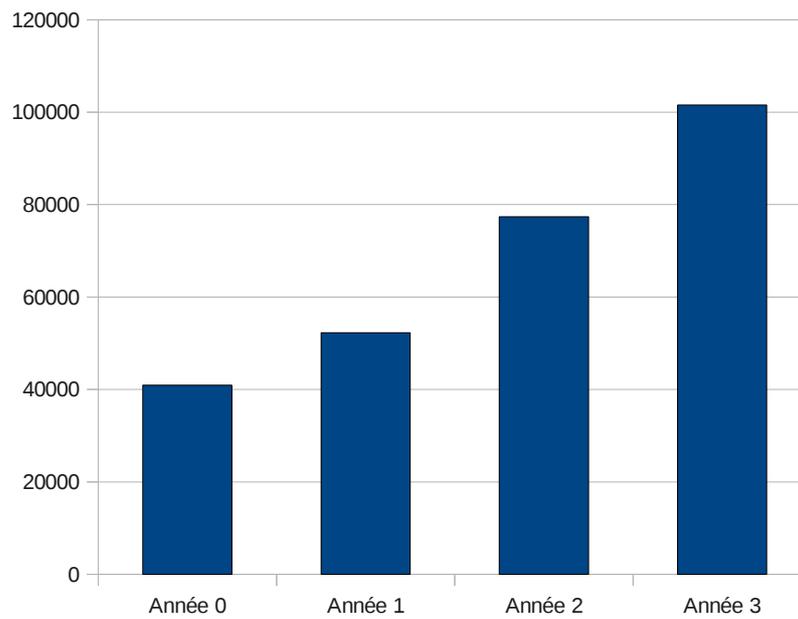


Illustration 3: Besoin en stockage cumulé par entité pour les 4 années

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

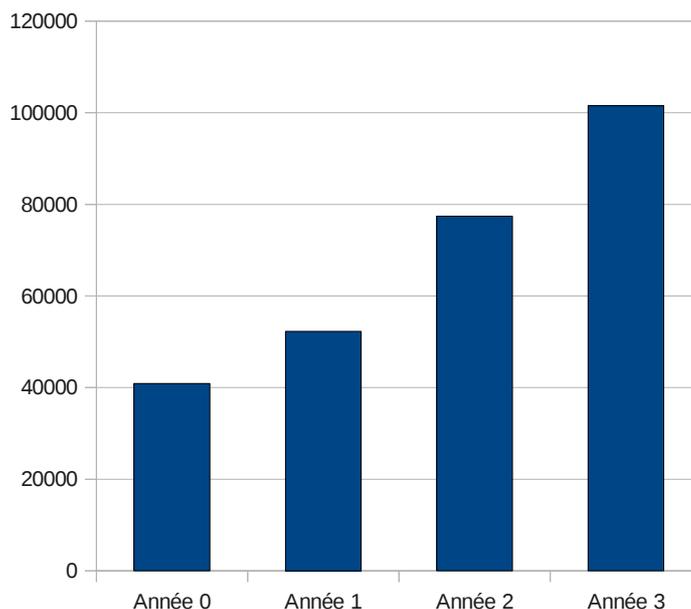


Illustration 4: Croissance du besoin en stockage (en Go)

10 Des spécifications fonctionnelles aux spécifications techniques

10.1 Vers la clôture du triptyque de l'étude

Cette étude a été menée suivant le triptyque traditionnel de gestion de projet : « où en est-on ? », « où va-t-on ? », « comment y-va-t-on ? ».

Il convient maintenant d'agrèger les besoins, tant qualitatifs que quantitatifs, pour en sortir une liste de spécifications fonctionnelles destinées à être, ensuite, déclinées en spécifications techniques purement informatiques.

10.2 Spécifications fonctionnelles

Il est difficile, pour le pôle de diffusion des savoirs, d'extraire des spécifications fonctionnelles ou techniques comme elles l'avaient été lors de la première étude.

Comme nous l'avons précisé au début de ce document, à l'exception d'ENSMédia, il n'a pas été possible de mener une étude exhaustive de tout le périmètre, notamment touchant aux processus manipulant les données.

Cependant, un certain nombre de spécifications peuvent être reprises ici, simplement parce que les utilisateurs les considèrent comme tacites de leur usage des espaces de stockage mis à leur disposition.

10.2.1 Pour le stockage

Pour le « front office » (le « salon », une affaire d'utilisateurs et leur appropriation des outils) :

- une gestion fine de l'accès aux données :
 - droits des utilisateurs pour les accès en écriture à partir des postes de manipulation,
 - droits des utilisateurs pour les accès en lecture à partir des postes de traitement,
 - droits pour les responsables de ces utilisateurs d'accéder à leurs données après leur départ ;
- une accessibilité des données dépassant le cadre du service :
 - espace accessible de l'extérieur de manière simple et sécurisée ;
- une indexation indispensable des acquisitions et des traitements pour assurer la pérennité des données ;
- une abstraction des volumes de stockage ;

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

- une mise à disposition rapide pour faire face aux demandes urgentes et limiter le risque de perte par casse ;
 - il a été précisé que, si la donnée n'était pas accessible directement, elle devait l'être sous un délai ne dépassant pas 24 heures.

Pour le « back-office » (la « cuisine », une affaire d'informaticiens et la mise à disposition des outils) :

- une amélioration des conditions de transfert des données notamment par la généralisation d'une connectivité haut débit (GE minimum) ;
 - cette condition est indispensable pour la réalisation des montages vidéos. Les débits de disque réseau doivent être comparables, à tout moment, à ceux d'un disque local, voire supérieur ;
- une souscription la plus large possible du contrat de maintenance ;
- des procédures simplifiées pour la mise à disposition ou l'extension d'un volume de stockage ;
- des procédures simplifiées pour la restauration d'un volume de stockage ;
- une « scalabilité » de la solution de stockage pour l'étendre chaque année en fonction des besoins.

10.2.2 Pour la sauvegarde

Cette opération est purement informatique : elle vise à sauvegarder les données (et permettre leur restauration) en cas de destruction d'un élément (ou plusieurs) du stockage.

Dans le cas idéal, elle exige :

- une séparation physique du stockage primaire ;
- une représentation la plus synchrone possible des données originelles.

10.2.3 Pour l'archivage

Cette opération, pour être la plus efficace possible, se base sur les informations mentionnées par les utilisateurs dans l'indexation de leurs données.

Deux approches sont possibles :

- un archivage basé sur le stockage originel ;
- un archivage basé sur la sauvegarde.

Dans les deux cas, les archives peuvent prendre la forme :

- d'une série d'instantanés pris suivant une politique pré-établie : elle est alors intégrée à la solution de stockage ;
- une copie complète sur un support tierce, archivée physiquement : elle utilise généralement une technologie à base de supports à accès séquentiels, comme les bandes.

11 Retour à la lettre de mission

La lettre de mission du 21 avril 2010 demandait de mener une étude sur les besoins de stockage, sauvegarde et archivage des laboratoires du site Descartes et du pôle de diffusion des savoirs dans son ensemble.

Aujourd'hui, les besoins du pôle de diffusion des savoirs intégrant ENSMédia, UNIS et la bibliothèque s'expriment à hauteur de **41 To**.

A trois ans, le cumul de ce montant et des 3 années suivantes, compte-tenu des objectifs de croissance, approche les **272 To**.

En considérant une sauvegarde comparable au stockage, les besoins en sauvegarde sont de **41 To** la première année et atteignent **272 To** pour 4 années d'exploitation.

Si un archivage sur disques durs est privilégié, suivant la même approche que celle proposée pour les laboratoires du site Monod, les besoins en archivage (pour stockage et sauvegarde) sont de **16 To** pour la première année et **109 To** pour 4 années.

Ainsi, une solution de stockage, sauvegarde et archivage parfaitement comparable à celle du site Monod, pour le pôle de diffusion des savoirs, se monterait à **98 To pour la première année** et **653 To sur 4 années d'exploitation**.

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

Usage	Besoin immédiat	Besoin sur 4 années
Stockage	41	272
Sauvegarde	41	272
Archivage	16	109
Total	98	653

Table 3: Synthèse des besoins en stockage/sauvegarde/archivage en To

12 Conclusion

Les besoins des unités du pôle de diffusion des savoirs se sont, malgré tout, exprimés, certes de façon plus sommaire que les laboratoires du site Monod. L'utilisation du schéma synoptique de traitement, présentant le cycle de la donnée numérique comme étant son acquisition, son traitement, sa valorisation et son exploitation s'est largement retrouvés dans les différentes activités des composantes de ce pôle, notamment dans l'activité d'ENSMédia.

L'analyse des besoins a été menée en cherchant, le mieux possible, à cerner quels étaient les usages et les objectifs de croissance. Cependant, l'auteur ne peut que regretter que le grain de l'étude, comparé à celui de la première, ne soit pas aussi fin.

Il convient maintenant, à partir de ces chiffres, de s'entendre sur une stratégie de sauvegarde et d'archivage pour ces volumes très conséquents. Ces volumes sont donc à ajouter à tous les volumes issus des études liées auprès des laboratoires du site Monod et du site Descartes.

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

13 Annexes

13.1 Questionnaire « stockage pour les plates-formes d'acquisition »

1. Nom de l'unité
2. Principales missions (fonctions, activités)
3. Quel est votre nom (interviewé) ?
4. Quel est le type d'équipement d'acquisition (station de travail, poste de travail, ...) ?
5. Quel type de logiciel utilisez-vous pour les acquisitions ?
6. Quel est le type de données en sortie de l'équipement d'acquisition (image, vidéo, base de données, ...) ?
7. Qui réalise l'acquisition des données ?
 1. Stagiaire
 2. Étudiant
 3. Doctorant
 4. Chercheur
 5. Enseignant-Chercheur
 6. Personnel Technique
 7. Prestataire externe
 8. Autre :
8. Qui exploite les données brutes de la plate-forme ?
 1. Stagiaire
 2. Étudiant
 3. Doctorant
 4. Chercheur
 5. Enseignant-Chercheur
 6. Personnel Technique
 7. Prestataire externe
 8. Autre :
9. Quel est le nombre d'acquisitions réalisées par mois / an ?
10. Quel est le volume de données moyen généré par une acquisition ?
11. Où sont stockés les résultats des acquisitions ?
12. Existe-t-il des contraintes sur les acquisitions ?
13. Combien de temps les données doivent-elles être conservées ?
14. Avez-vous besoin d'accès immédiats à l'ensemble des acquisitions ?
15. Si non, quel est le délai de mise à disposition à envisager, en fonction de la date d'acquisition ?
16. Par combien comptez-vous multiplier le nombre d'acquisitions sur les 3 prochaines années ?
17. Le réseau actuel (100 Mb/s) est-il satisfaisant en cas de transfert de données (dépasse le 1/4 d'heure) ?

13.2 Questionnaire « stockage pour les plates-formes de traitement »

1. Nom de l'unité

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

2. Principales missions (fonctions, activités)
3. Quel est votre nom (interviewé) ?
4. Quel est le type d'équipement de traitement (station de travail, poste de travail, ...) ?
5. Quel type de logiciel utilisez-vous pour les traitements ?
6. Quel est le type de données en entrée de l'équipement de traitement (image, vidéo, base de données, ...) ?
7. Quel est le type de données en sortie de l'équipement de traitement (image, vidéo, base de données, ...) ?
8. Qui réalise le traitement des données ?
 1. Stagiaire
 2. Étudiant
 3. Doctorant
 4. Chercheur
 5. Enseignant-Chercheur
 6. Personnel Technique
 7. Prestataire externe
 8. Autre :
9. Qui exploite des données traitées de la plate-forme ?
 1. Stagiaire
 2. Étudiant
 3. Doctorant
 4. Chercheur
 5. Enseignant-Chercheur
 6. Personnel Technique
 7. Prestataire externe
 8. Autre :
10. Quel est le nombre de traitements réalisées par mois / an ?
11. Quel est le volume de données moyen généré par un traitement ?
12. Où sont stockés les résultats des traitements ?
13. Existe-t-il des contraintes sur les traitements ?
14. Combien de temps les données doivent-elles être conservées ?
15. Avez-vous besoin d'accès immédiats à l'ensemble des traitements ?
16. Si non, quel est le délai de mise à disposition à envisager, en fonction de la date de traitement ?
17. Par combien comptez-vous multiplier le nombre de traitements sur les 3 prochaines années ?
18. Le réseau actuel (100 Mb/s) est-il satisfaisant en cas de transfert de données (dépasse le 1/4 d'heure) ?

13.3 Questionnaire « stockage pour l'exploitation des résultats »

1. Quel est votre nom ?
2. Quelle est l'unité à l'origine de l'exploitation ?
3. Quel nom donnez-vous à l'exploitation de ces résultats ?
4. Quel type de logiciel utilisez-vous pour l'exploitation ?
5. Quel est le type de données utilisées ?

	Etude sur les besoins de Stockage Informatique du Pôle de Diffusion des Savoirs Analyse de l'enquête	Référence	:	ENSL-Storage4labs2PDS-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

6. Quel est le type d'exploitation ?
7. Quelle personne propose l'exploitation ?
 1. Stagiaire
 2. Étudiant
 3. Doctorant
 4. Chercheur
 5. Enseignant-Chercheur
 6. Personnel Technique
 7. Prestataire externe
 8. Autre :
8. A qui est destiné cette exploitation ?
 1. Chercheur
 2. Enseignant
 3. Étudiant
 4. Collégien ou Lycéen
 5. Enseignant du secondaire
 6. Tout public
 7. Autre :
9. Quelle est la durée moyenne d'exploitation ?
10. Combien de publications sont-elles réalisées par semaine ?
11. Par combien comptez-vous multiplier votre taux de publication sur ces 3 prochaines années ?
12. Combien de temps les données doivent-elles être conservées ?
13. Où se situent les données sources ?
14. Quelle est la nature de l'exploitation ?
15. Quel est le volume de données des exploitations ?
16. Le réseau actuel (100 Mb/s) est-il satisfaisant en cas de transfert de données (dépassé le 1/4 d'heure) ?