
	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes

Analyse de l'enquête

Ecole Normale Supérieure de Lyon

préparé par	Emmanuel Quemener	
contrôlé par		
approuvé par	<p>_____ Date: _____</p>	
reference	ENSL-Storage4labs2Descartes-100528	
version	0 draft	
date de version	28 mai 2010	
nom de document	ENSL-Storage4labs2Descartes-100528.odt	

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

Mise à jour

Date	Version	Etat	Pages	Raisons du changement
Initiale	0	Brouillon		
Revue	0.1	Brouillon		Lecture par Véronique Queste




	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

Table des Matières

1 Introduction.....	7
2 Documents applicables.....	7
3 Glossaire.....	7
4 Démarche.....	7
5 Modélisation du circuit des données numériques.....	8
5.1 Les processus.....	8
5.2 Les données.....	8
5.3 Les « formes » d'entreposage de ces données.....	8
6 Questionnaires.....	9
6.1 De la démarche analytique aux questionnaires.....	9
6.2 Une approche en deux temps.....	9
7 Résultats préliminaires.....	10
7.1 Réponses aux questionnaires.....	10
7.1.1 Premier bilan.....	10
7.1.2 Éléments d'analyse.....	10
8 Analyse qualitative.....	10
8.1 Par quoi : quels dispositifs d'acquisition ?.....	10
8.2 Quoi : quelle nature des données numériques ?.....	11
8.2.1 Quelles données en entrée ?.....	11
8.2.2 Quelles données en sortie ?.....	11
8.3 Par qui et pour qui ?.....	12
8.3.1 Acquisition : qui réalise ou exploite ?.....	12
8.3.2 Traitement : qui réalise ou exploite ?.....	12
8.3.3 Exploitation : qui propose ?.....	13
8.3.4 Exploitation : à destination de quel public ?.....	13
8.4 Quelles durées ?.....	13
8.4.1 Quelle durée de conservation ?.....	13
8.4.2 Quelle durée de présentation ?.....	14
8.5 Quelles contraintes ?.....	15
8.6 Quelle infrastructure ?.....	16
9 Analyse quantitative.....	16
9.1 Introduction à l'analyse des données.....	16

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

9.2 Acquisition.....	16
9.3 Traitement.....	17
9.4 Valorisation.....	17
9.5 Exploitation.....	18
10 Cumul de tous les besoins en stockage.....	18
11 Des spécifications fonctionnelles aux spécifications techniques. .	20
11.1 Vers la clôture du triptyque de l'étude.....	20
11.2 Spécifications fonctionnelles.....	20
11.2.1 Pour le stockage.....	20
11.2.2 Pour la sauvegarde.....	20
11.2.3 Pour l'archivage.....	20
12 Retour à la lettre de mission.....	21
13 Conclusion.....	21
14 Annexes.....	22
14.1 Questionnaire « stockage pour les plates-formes d'acquisition».....	22
14.2 Questionnaire « stockage pour les plates-formes de traitement ».....	22
14.3 Questionnaire « stockage pour l'exploitation des résultats ».....	23


	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

Index des illustrations

Illustration 1: Modélisation du circuit de données numériques.....	9
Illustration 2: Quels dispositifs d'acquisition pour les documents ?.....	10
Illustration 3: Quelle nature des données en entrée ?.....	11
Illustration 4: Quelle nature des données en sortie ?.....	11
Illustration 5: Quels acteurs pour l'acquisition ?.....	12
Illustration 6: Quels acteurs pour le traitement ?.....	12
Illustration 7: Qui propose l'exploitation ?.....	13
Illustration 8: A destination de quel public ?.....	13
Illustration 9: Quelle conservation des documents ?.....	14
Illustration 10: Quelle durée en ligne des documents ?.....	14
Illustration 11: Quelles contraintes ?.....	15
Illustration 12: Quelle satisfaction du réseau aux usages ?.....	16
Illustration 13: Besoins en stockage des laboratoires, sur 4 années (en Go).....	19
Illustration 14: Besoins en stockage cumulés pour 4 années (en Go).....	19

Index des Tables

Table 1: Besoins de stockage des laboratoires pour l'acquisition (en Go).....	16
Table 2: Besoins de stockage des laboratoires pour le traitement (en Go).....	17
Table 3: Besoins de stockage pour la valorisation (en Go).....	17
Table 4: Besoins de stockage pour l'exploitation (en Go).....	18
Table 5: Besoins en stockage cumulés des laboratoires (en Go)	18
Table 6: Synthèse des besoins en stockage/sauvegarde/archivage en To.....	21

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-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 au 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_Descartes.pdf
Besoins des laboratoires site Monod	http://perso.ens-lyon.fr/emmanuel.quemener/documents/ENSL-Storage4labs-100415.pdf

3 Glossaire

CQQCOQP	Allographe de QQQCCP (Qui fait Quoi ?, Où ? Quand ? Comment ? Combien ? Et Pourquoi ?)
TGE	Très Grand Équipement
TRIANGLE	Action, discours, pensée politique et économique
ICAR	Interactions, Corpus, Apprentissages et Représentations (ICAR) - UMR 5191
CERPHI	Centre d'études en rhétorique, philosophie et histoire des idées
EVS	Environnement, Ville, Société (EVS) - UMR 5600
IAO	Institut d'Asie Orientale
CEP	Centre d'études poétiques (CEP) - UPRES-EA 1633
GRS	Groupe de Recherche sur la Socialisation (GRS) - UMR5040
LARHA	Laboratoire de recherche historique Rhône-Alpes (LARHRA) - UMR 5190
GATE	Groupe d'analyse et de théorie économique (GATE) - UMR 5824
C2SO	Communication, Culture et Société (C2SO)
LIRE	Littérature, idéologies, représentations aux XVIII et XIX siècles (LIRE) - UMR 5611
CIHAM	Histoire et archéologie des mondes chrétiens et musulmans médiévaux (CIHAM) - UMR 5648
CEROR	Centre d'étude et de recherche sur l'occident romain (CEROR) - EA 664
EA 37 08	Institut des langues et cultures slaves - EA 3708
CID	Centre d'ingénierie documentaire
BDD	Bases des Données
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 (Besoins des laboratoires site

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

Monod, page 7) : 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.

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 9) 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 des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

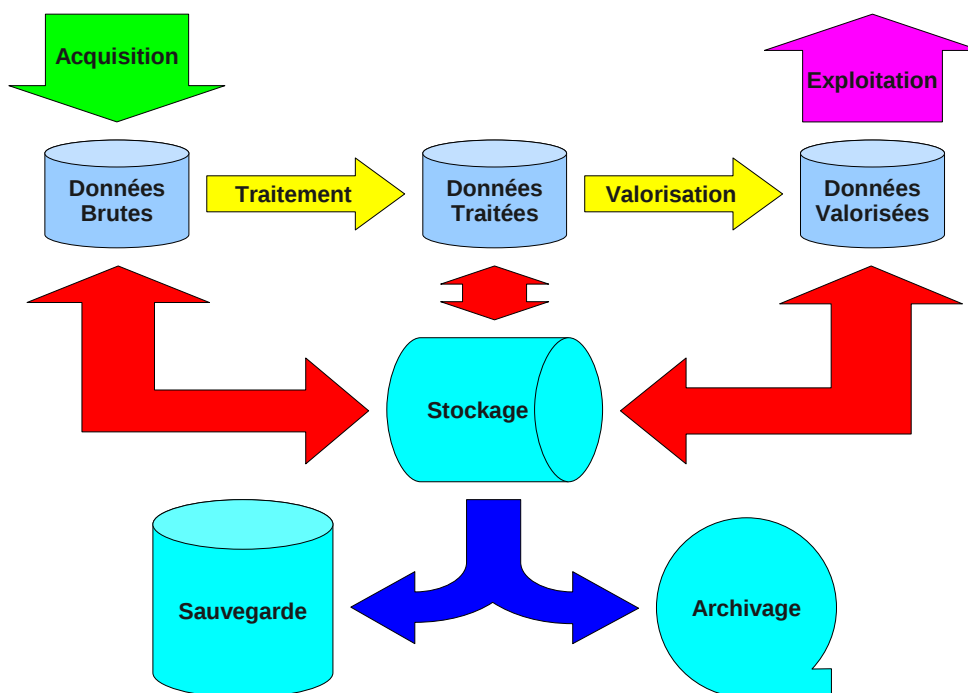


Illustration 1: Modélisation du circuit de données numériques

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 22, chapitre 14.1) ;
- le traitement (page 22, chapitre 14.2) ;
- l'exploitation (page 23, chapitre 14.3).

6.2 Une approche en deux temps

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

La première étape a été d'identifier les individus susceptibles de répondre aux questionnaires. Ces derniers, une fois connus, ont reçu par message électronique les questionnaires mentionnés au paragraphe précédent. Quelques jours de réflexion leur ont permis d'identifier, surtout de manière quantitative, leurs besoins.

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

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

7 Résultats préliminaires

7.1 Réponses aux questionnaires

7.1.1 Premier bilan

Au 12 mai 2010, date de démarrage de rédaction de ce document, sur 15 laboratoires identifiés, 7 avaient exprimé leurs besoins, dont 6 de manière assez constructive pour extraire des données autant qualitatives et quantitatives.

7.1.2 Éléments d'analyse

L'analyse se déroule en deux temps.

D'abord, un premier temps est consacré à l'examen des questions connexes liées aux contextes dans la gestion des données numériques : l'analyse qualitative. Elle comprend cependant un certain nombre de données quantitatives.

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

8 Analyse qualitative

8.1 Par quoi : quels dispositifs d'acquisition ?

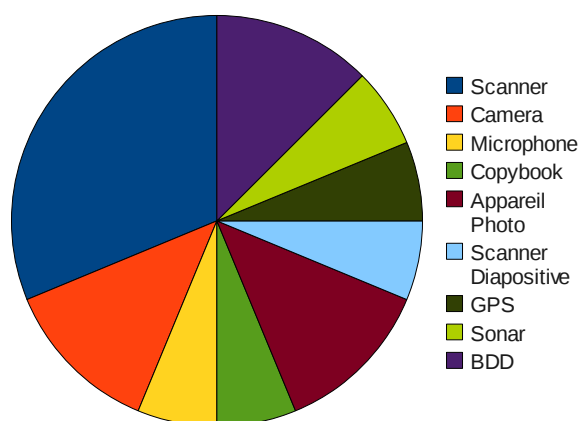



Illustration 2: Quels dispositifs d'acquisition pour les documents ?

Remarques générales :

- les principaux dispositifs d'acquisition sont des appareils fournissant des médias traditionnels : images, vidéo et son ;
- le contexte de l'acquisition peut être très différent : directement sur les équipements de l'établissement, indirectement en se déplaçant où se trouvent les documents à numériser. Ces usages entraînent souvent une adaptation de la procédure de traitement : la numérisation d'un livre sur le *copybook* par le personnel idoine n'offre pas les mêmes garanties de qualité de prise de vue que le déplacement chez le propriétaire des archives avec un scanner à main par la personne seulement habilitée à consulter les documents.

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

8.2 Quoi : quelle nature des données numériques ?

8.2.1 Quelles données en entrée ?

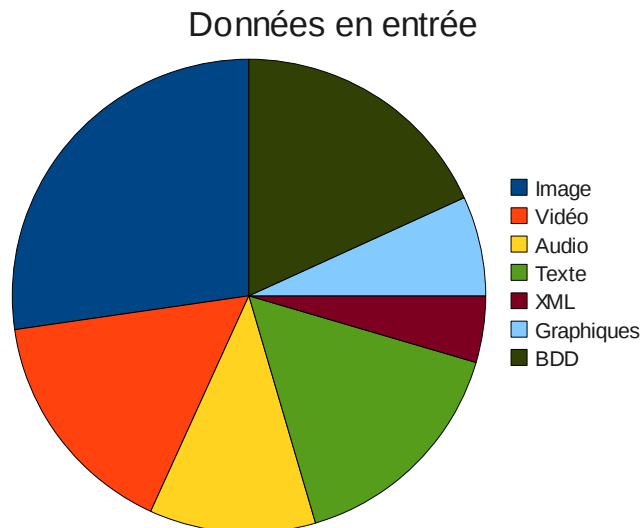


Illustration 3: Quelle nature des données en entrée ?

Remarques générales :

- se retrouvent les principaux types de données issus des dispositifs d'acquisition ;
- l'usage de textes, de bases de données, de langages de représentation (XML) est également très répandu, même si leur volume est incomparable face aux données issues de vidéos ou d'images issues de numérisation de documents papiers.

8.2.2 Quelles données en sortie ?

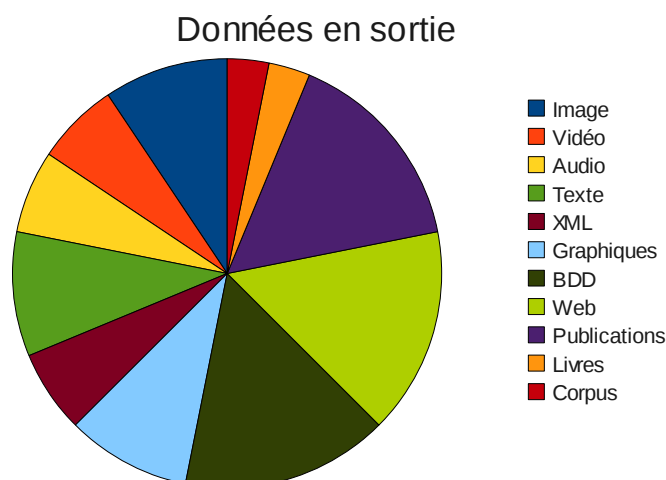



Illustration 4: Quelle nature des données en sortie ?

Remarques générales :

- l'enrichissement des données se réalise souvent par la mise à disposition en ligne (sur le Web ou via des corpus) ;
- les bases de données demeurent aussi une destination prisée des processus de traitement et d'exploitation.

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

8.3 Par qui et pour qui ?

8.3.1 Acquisition : qui réalise ou exploite ?

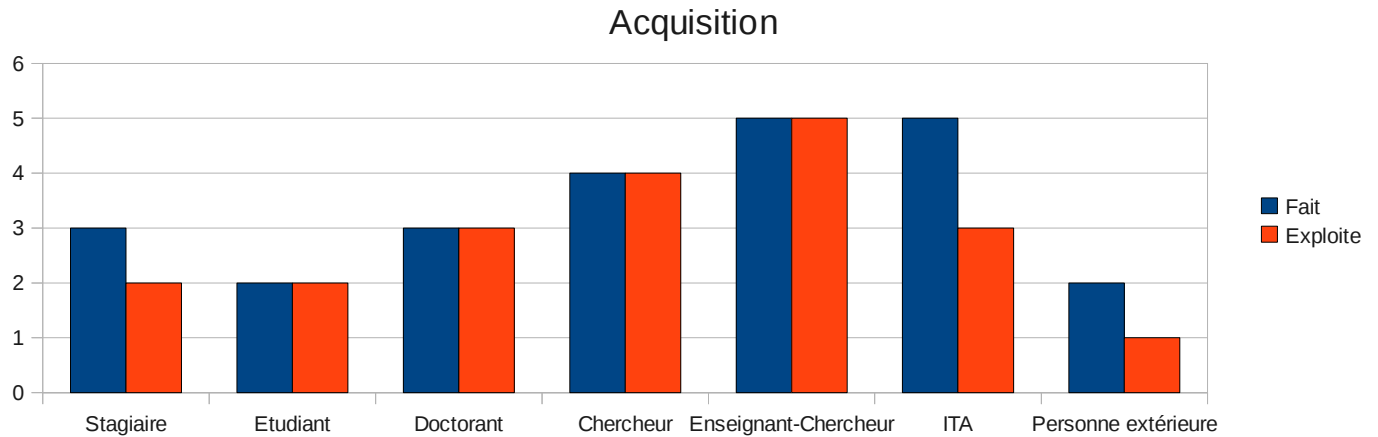


Illustration 5: Quels acteurs pour l'acquisition ?

Remarques générales :

- tous participent au processus parfois fastidieux de l'acquisition ;
- contrairement aux sciences dures, les doctorants semblent moins sollicités sur ces tâches.

8.3.2 Traitement : qui réalise ou exploite ?

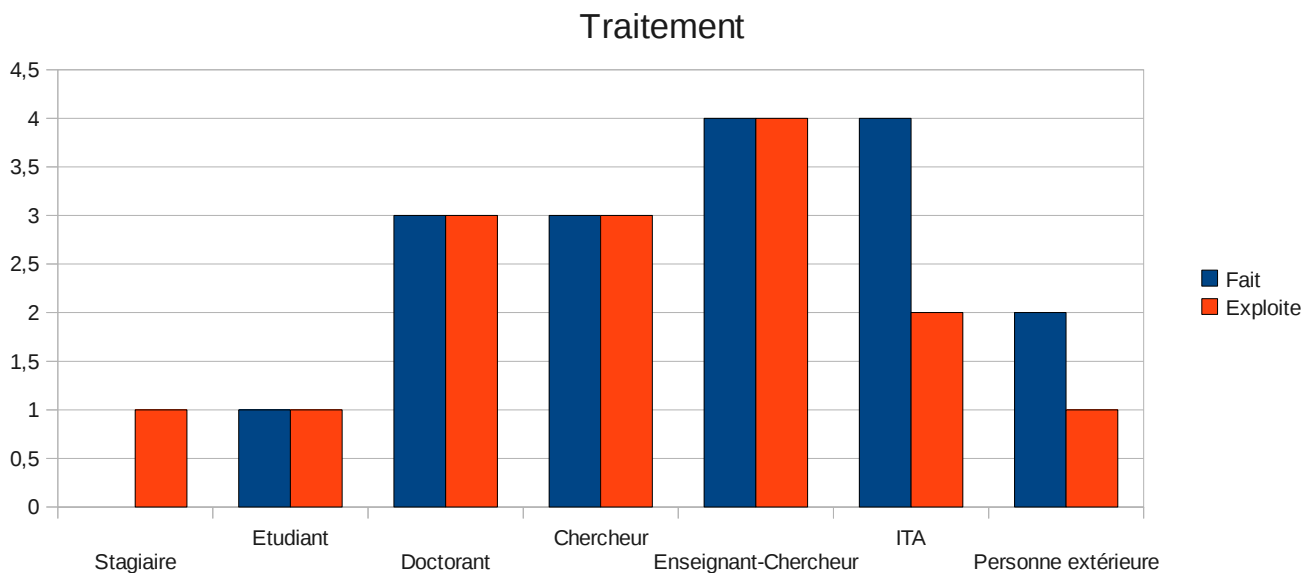



Illustration 6: Quels acteurs pour le traitement ?

Remarques générales :

- ici encore, tous sont sollicités, presque de manière équivalente, dans le processus de traitement ;
- l'exploitation reste cependant l'apanage de ceux dont c'est l'activité principale (les ITA y sont moins actifs).

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

8.3.3 Exploitation : qui propose ?



Illustration 7: Qui propose l'exploitation ?

Remarque générale :

- large équité là encore sur la proposition d'exploitation et la mise en ligne : tous participent de manière équivalente

8.3.4 Exploitation : à destination de quel public ?




Illustration 8: A destination de quel public ?

Remarques générales :

- l'ouverture vers tous les publics ou l'enseignement secondaire forme le tiers de l'activité ;
- les publications auprès des pairs de la recherche en forment plus de la moitié de l'exploitation.

8.4 Quelles durées ?

8.4.1 Quelle durée de conservation ?

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

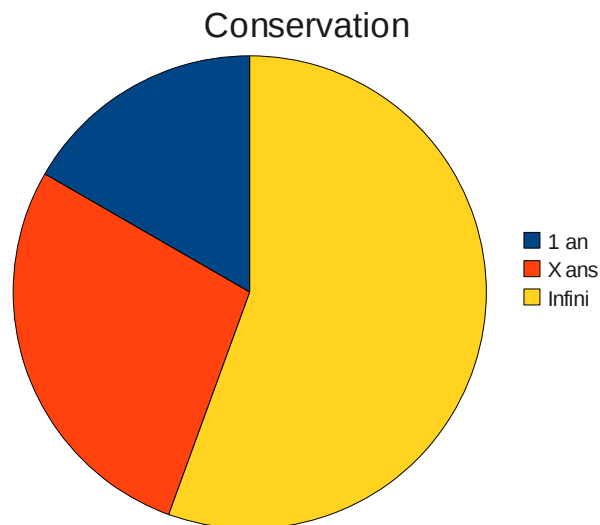


Illustration 9: Quelle conservation des documents ?

Remarques générales :

- la conservation sur de longues voire très longues durées est la règle ;
- la volatilité des données est très faible, ce qui impose que leur pérennité doit être assurée sur de larges périodes.

8.4.2 Quelle durée de présentation ?

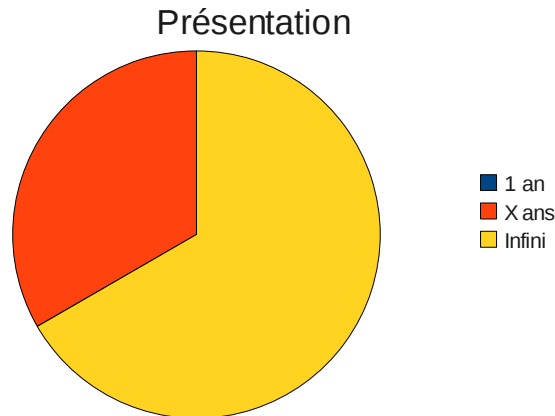



Illustration 10: Quelle durée en ligne des documents ?

Remarque générale :

- personne n'a exprimé de voir ses données seulement accessibles sur de courtes périodes : la diffusion permanente demeure la règle.

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

8.5 Quelles contraintes ?

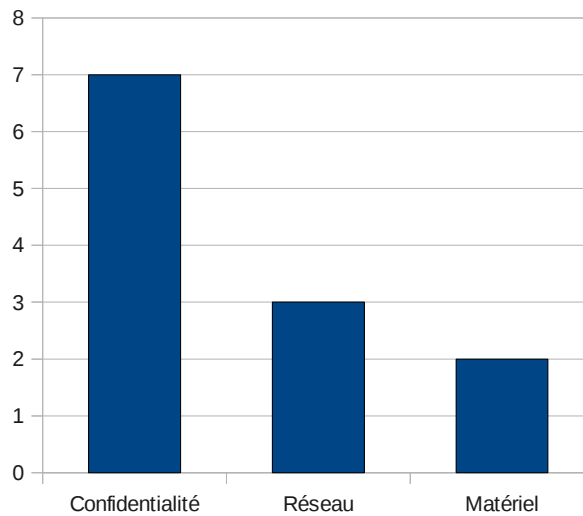



Illustration 11: Quelles contraintes ?

Remarques générales :

- la confidentialité sur certaines données a été la surprise, contrastant avec la réelle volonté de partager le plus possible et le plus longtemps possible : elle est conditionnée par la nature même des données (par exemple, des éléments très personnels de la vie de Desanti comme des ordonnances médicales) ou par les limitations juridiques dans l'exploitation d'œuvres soumises au droit d'auteur (donc généralement cadrée dans des contrats de nature juridiques) ;
- le réseau reste un élément primordial, notamment dans l'accès aux données (la question suivante reprend ce point) ;
- le matériel et son exploitation, notamment dans sa phase d'acquisition, conditionnent le passage dans le monde numérique : la facilité d'acquisition est liée aux contraintes d'accès aux originaux. Cet accès est parfois difficile : le matériel qu'il est alors possible d'utiliser est rarement équivalent à celui utilisable dans un local spécialisé (par exemple, l'accès à des archives chez un particulier distillées au goutte-à-goutte, à quelques personnes).

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

8.6 Quelle infrastructure ?

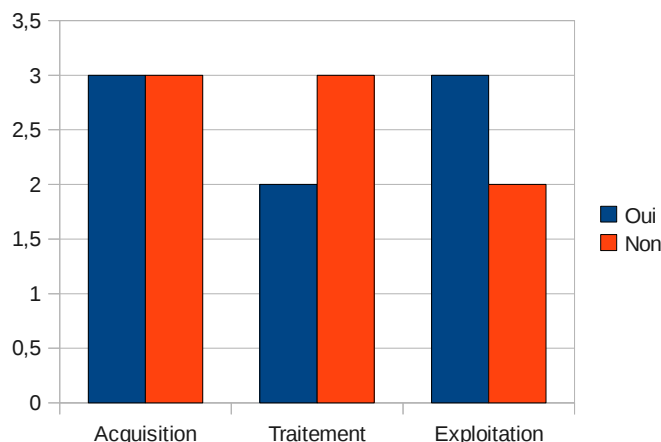


Illustration 12: Quelle satisfaction du réseau aux usages ?

Remarques générales :

- la moitié des personnes estiment que le réseau n'est pas satisfaisant pour le transfert des données
- lors de l'étude, il a été constaté que, effectivement, le réseau était extrêmement lent pour le transfert d'images situées sur un dossier partagé : près d'une minute pour afficher une image de 80 Mo. Lors de l'analyse de la métrologie embarquée, pourtant, rien ne laissant transparaître un réseau ou une machine saturée. D'après l'utilisatrice, ce problème est plus que récurrent ! Une analyse donc de toute la chaîne reliant le poste d'exploitation à son espace de stockage est indispensable pour estimer l'approche technique la plus rationnelle : le passage au Gigabit Ethernet serait un atout, mais inutile si le cœur de réseau ou l'équipement de stockage n'assurent pas un transfert optimal.

9 Analyse quantitative

9.1 Introduction à l'analyse des données

Seuls six laboratoires se sont directement exprimés sur les besoins en stockage concernant les processus d'acquisition, de traitement et d'exploitation.


9.2 Acquisition

Laboratoire	Année 0	Année 1	Année 2	Année 3	Total
TRIANGLE	2000	4666.67	7333.33	10000	24000
ICAR	50000	50000	50000	50000	200000
CERPHI	3375	5625	7875	10125	27000
EVS	1000	1666.67	2333.33	3000	8000
GRS	10	10	10	10	40
CID	250	250	250	250	1000
Total	56635	62218.33	67801.67	73385	260040

Table 1: Besoins de stockage des laboratoires pour l'acquisition (en Go)

Remarques générales :

- Le laboratoire ICAR est un gros consommateur d'espace disque, pour l'instant stocké sur des disques durs amovibles (donc sans sécurité). Cette consommation tient à l'activité du laboratoire sollicitant largement l'usage de la vidéo sur des durées très longues ;
- le laboratoire Cerphi est également un gros consommateur : son activité de numérisation, notamment pour la publication de corpus, exige des volumes importants. Il est à noter que, pour le Cerphi, le montant présenté est

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

une extrapolation sur les 9 équipes sont trois seulement s'étaient exprimées. Un facteur 3 a donc été appliqué à leurs besoins.

9.3 Traitement

Laboratoire	Année 0	Année 1	Année 2	Année 3	Total
TRIANGLE	1000	2333,33	3666,67	5000	12000
ICAR	5000	5000	5000	5000	20000
CERPHI	19302	32170	45038	57906	154416
EVS	3000	5000	7000	9000	24000
CID	50	66,67	83,33	100	300
Total	28352	44570	60788	77006	210716

Table 2: Besoins de stockage des laboratoires pour le traitement (en Go)

Remarques générales :

- le laboratoire ICAR reste un gros consommateur d'espace disque. Malgré le traitement réduisant les espaces, il reste d'un volume équivalent au dixième du volume initial ;
- le laboratoire Cerphi devient ici le plus gros consommateur : les traitements associés aux données, la multiplication des formats et des résolutions, destinées à la publication, augmentent très sensiblement les espaces nécessaires. Là encore, les besoins du laboratoire ont été multiplié par trois pour intégrer les équipes ne s'étant pas exprimées ;
- le laboratoire EVS présente également ici de manière significative.

9.4 Valorisation


Les besoins en stockage pour la valorisation n'ont pas été directement exprimés par les utilisateurs.

Ils ont été estimés à partir du nombre de personnes dans chaque unité multiplié par un quota annuel considéré comme confortable : 10 Go, uniquement destiné au travail de recherche proprement dit.

Sur la progression de la consommation, il a été estimé que, à 3 ans, le quota avait doublé.

Valorisation	Année 0	Année 1	Année 2	Année 3	Total
TRIANGLE	1400	1866,67	2333,33	2800	8400
ICAR	1170	1560	1950	2340	7020
CERPHI	970	1293,33	1616,67	1940	5820
EVS	640	853,33	1066,67	1280	3840
IAO	560	746,67	933,33	1120	3360
CEP	470	626,67	783,33	940	2820
GRS	380	506,67	633,33	760	2280
LARHA	260	346,67	433,33	520	1560
GATE	250	333,33	416,67	500	1500
C2SO	220	293,33	366,67	440	1320
LIRE	210	280	350	420	1260
CIHAM	160	213,33	266,67	320	960
CEROR	40	53,33	66,67	80	240
EA 37 08	10	13,33	16,67	20	60
Total	6740	8986,67	11233,33	13480	40440

Table 3: Besoins de stockage pour la valorisation (en Go)

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

9.5 Exploitation

Dans ces données, le volume destiné à l'archivage pérenne a été exclu pour le laboratoire Cerphi.

En effet, l'archivage pérenne peut faire l'objet d'une demande auprès du TGE Adonis.

Exploitation	Année 0	Année 1	Année 2	Année 3	Total
TRIANGLE	550	1283.33	2016.67	2750	6600
ICAR	500	500	500	500	2000
CERPHI	10392	17320	24248	31176	83136
EVS	3000	4000	5000	6000	18000
CID	400	533.33	666.67	800	2400
Total	14842	23636.67	32431.33	41226	112136

Table 4: Besoins de stockage pour l'exploitation (en Go)

10 Cumul de tous les besoins en stockage

Le tableau 5, page 18 présente les besoins cumulés en stockage pour les laboratoires du site Descartes. Ce cumul comprend les processus d'acquisition, de traitement, de valorisation et d'exploitation.

Le laboratoire Cerphi exprime les besoins les plus importants : il représente la moitié la première année et près des deux tiers sur une projection de 4 années de consommation : dans le cadre de cette étude, pour le Cerphi, seules trois équipes sur neuf s'étaient exprimées. Les volumes ont donc été construits par extrapolation en considérant que les besoins exprimés pour ces trois équipes représentaient seulement le tiers

Cumul	Année 0	Année 1	Année 2	Année 3	Total
TRIANGLE	4950	10150	15350	20550	51000
ICAR	56670	57060	57450	57840	229020
CERPHI	34039	56408,33	78777,67	101147	270372
EVS	7640	11520	15400	19280	53840
IAO	560	746,67	933,33	1120	3360
CEP	470	626,67	783,33	940	2820
GRS	390	516,67	643,33	770	2320
LARHA	260	346,67	433,33	520	1560
GATE	250	333,33	416,67	500	1500
C2SO	220	293,33	366,67	440	1320
LIRE	210	280	350	420	1260
CIHAM	160	213,33	266,67	320	960
CEROR	40	53,33	66,67	80	240
EA 37 08	10	13,33	16,67	20	60
CID	700	850	1000	1150	3700
Total	106569	139411,67	172254,33	205097	623332

Table 5: Besoins en stockage cumulés des laboratoires (en Go)

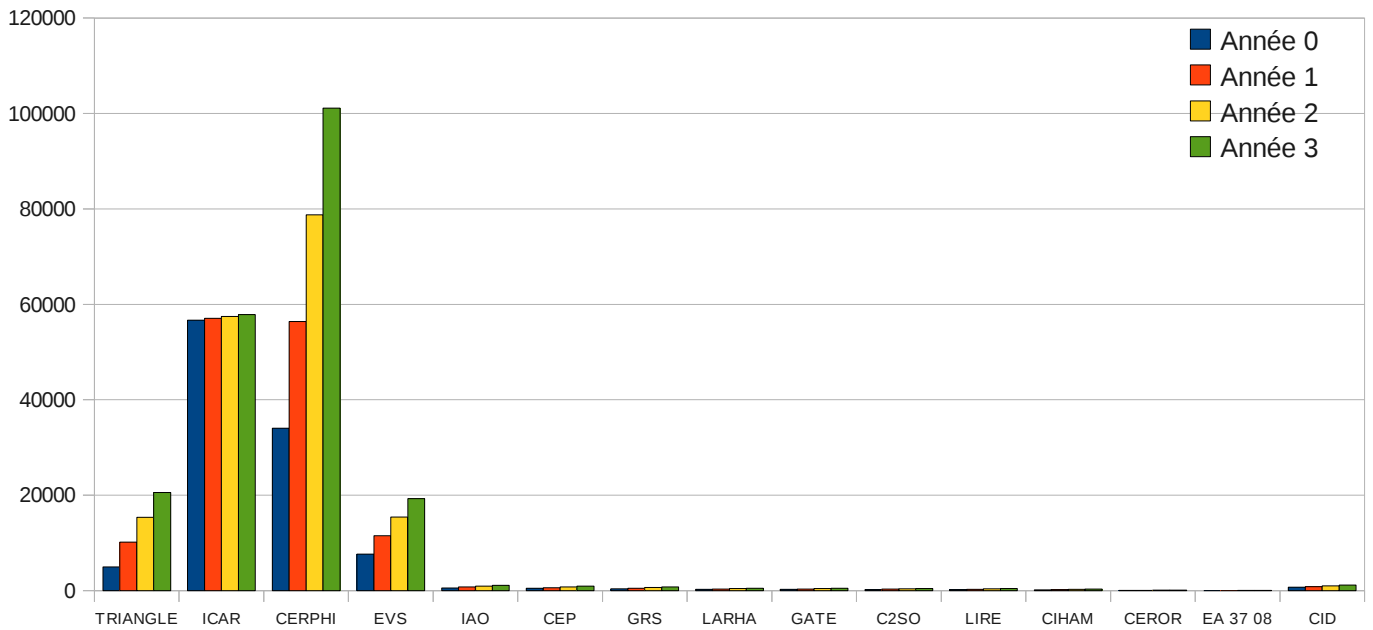


Illustration 13: Besoins en stockage des laboratoires, sur 4 années (en Go)

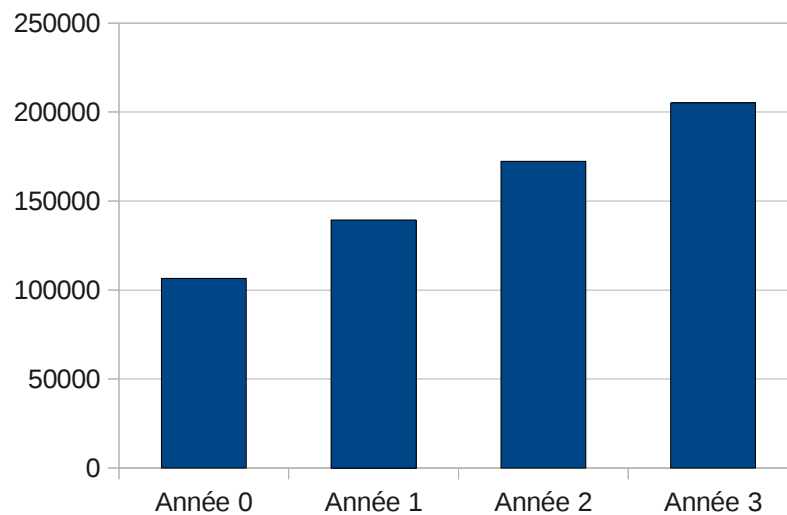



Illustration 14: Besoins en stockage cumulés pour 4 années (en Go)

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

11 Des spécifications fonctionnelles aux spécifications techniques

11.1 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 ? ».

Les deux premières questions ont été exprimées par les deux enquêtes sous forme de questionnaires et leur analyse. De plus, une sommaire modélisation, à partir de conversations et de courriels, a permis d'établir les besoins courants de stockage informatique pour l'activité quotidienne des personnes des laboratoires.

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.

11.2 Spécifications fonctionnelles

11.2.1 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 laboratoire :
 - espace accessible de l'extérieur de manière simple et sécurisée ;
- une indexation indispensable des expériences et des traitements pour assurer la pérennité des données ;
- une mise à disposition rapide pour faire face aux demandes urgentes et limiter le risque de perte par casse ;

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 entre plateformes (de l'expérience à son traitement) notamment par la généralisation d'une connectivité haut débit (GE minimum) ;
- 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.

11.2.2 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.

11.2.3 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 ;

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon

- 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.

12 Retour à la lettre de mission

La lettre de mission du 21 avril 2010 de mener une nouvelle mission d'étude des besoins de stockage, sauvegarde et archivage de données informatiques pour les laboratoires du site Descartes, d'une part, et pour le pôle de diffusion des savoirs, d'autre part.

Ce document présente les besoins, en terme de stockage, des laboratoires du site Descartes : ces derniers sont de **106 To** la première année et forment un volume global de **615 To** en cumulant 4 années de besoins.

Pour une sauvegarde équivalente au stockage, les besoins pour la première année se montent à des montants identiques à ceux présentés précédemment : **106 To** la première année et **615 To** pour 4 années.

Pour un archivage directement intégré au stockage et à la sauvegarde, avec l'hypothèse de 20% d'espace nécessaire complémentaire, les besoins d'espace d'archivage se montent à **44 To** la première année et **246 To** pour 4 années d'exploitation.

Ainsi, pour conclure, avec les hypothèses ci-dessus, les besoins en stockage, sauvegarde et archivage se montent à **234 To** la première année et **1476 To** pour 4 années cumulées. Le montant de 1,5 Po de données à 4 ans est comparable à celui exprimé, voici quelques mois, dans l'enquête concernant les laboratoires du site Monod.

Usage	Besoin immédiat	Besoin sur 4 années
Stockage	106	615
Sauvegarde	106	615
Archivage	44	246
Total	234	1476

Table 6: Synthèse des besoins en stockage/sauvegarde/archivage en To

13 Conclusion

L'expression des besoins a donc été menée pour les laboratoires de l'Ecole Normale Supérieure de Lyon, site Descartes. Ce travail n'aurait pu être réalisé sans l'investissement très lourd, sur une période aussi brève, d'Aurélien Antonio, de la DSI. Il n'aurait pas non plus abouti sans la motivation clairement affichée des personnels ITA en charge des questions informatiques dans leur laboratoire ou gros utilisateurs de données.


De plus, il est intéressant de constater que les personnes interrogées se sont largement retrouvées dans l'enchaînement des quatre processus, de l'acquisition à l'exploitation.

Ainsi, l'analyse des besoins, actuels et sur une projection de 3 années, a permis de dégager un volume de stockage nécessaire.

Il convient maintenant, à partir de ces volumes, de s'entendre sur une stratégie de sauvegarde et d'archivage pour ces volumes importants.

Une fois cette stratégie établie, l'ensemble des spécifications fonctionnelles et les quelques spécifications techniques serviront à compléter le cahier des charges dans l'achat des équipements nécessaires.

Cette étude a également permis de mettre en lumière plusieurs aspects, liés notamment à la sécurité informatique et la conservation du patrimoine scientifique.

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-100528	
		Date	:	28/05/2010	
		Version	:	0	brouillon


14 Annexes

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

1. Nom du Laboratoire
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) ?

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


1. Nom du Laboratoire

	Etude sur les besoins de Stockage Informatique des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-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) ?

14.3 Questionnaire « stockage pour l'exploitation des résultats »

1. Quel est votre nom ?
2. Quel est le laboratoire 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 des Laboratoires Site Descartes Analyse de l'enquête	Référence	:	ENSL-Storage4labs2Descartes-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 (dépasse le 1/4 d'heure) ?