

# Une archive ouverte institutionnelle : OATAO (Open Archive Toulouse Archive Ouverte) à Toulouse



Bordeaux, Mardi 09 février 2010

# Plan

1. Genèse et organisation de l'archive
2. Statistiques
3. Les perspectives

# **1.Genèse et organisation de l'archive**

# Un contexte local favorable

- Une archive ouverte des thèses de l'INP mise en place **dès 2005**
- Des **compétences** acquises par le personnel du SCD
- Des **statistiques** de consultation encourageantes
  - Janvier 2006 : 10 934 thèses téléchargées
  - Décembre 2006 : 19 227 thèses téléchargées
- Un premier « travail » de **sensibilisation** auprès des instances de l'université (CS, CA, etc...) déjà effectué

# Un contexte national et international favorable

- Participation de l'INP aux **groupes de travail** nationaux (GTAO)
- Implication de l'INP dans le projet **ORI-OAI**
- Mise en place du **PRES** toulousain (mars 2007)
- Des **outils matures** de dépôts

# Une archive ouverte : pourquoi faire ?

## 2 arguments :

- Une meilleure diffusion des résultats de la recherche : **augmentation de la visibilité** et de l'accessibilité des travaux, et donc l'impact scientifique
- **Fournir des indicateurs recherche**

# Pourquoi une archive locale ?

- **Archive ouverte institutionnelle** : lien fort avec le SI des établissements.  
L'archive doit devenir l'une des briques de l'entrepôt de ressources (cours, thèses, AO, doc. administratifs...)
- **Un outil pour l'évaluation**
  - Listes d'autorité modérées
  - Contrôle et maîtrise de l'indexation
  - Retrait possible des documents
- **Contrôle** des droits de diffusion par les bibliothèques garanties par les SCD

# Pourquoi une archive locale ?

- Une archive **configurable** au plus proche des besoins des chercheurs
- Au final, deux plateformes de diffusion **interopérables** : une archive locale et une archive nationale...

## **Avantages :**

- complémentaires
- Meilleure visibilité

# Pourquoi une archive multi-établissements ?

- **Premier constat** : la majorité des laboratoires de l'INP sont des laboratoires **mixtes**
- Souhait de plusieurs établissements d'entrer dans l'archive locale

**La future archive sera multi-établissements**

# Quatre établissements d'enseignement supérieur

- École Nationale Vétérinaire de Toulouse (ENVT)
- Institut National Polytechnique de Toulouse (INPT) : 4 écoles d'ingénieurs
- Institut Supérieur de l'Aéronautique et de l'Espace (ISAE)
- École Nationale de Formation Agronomique (ENFA)

# Potentiel recherche

- Disciplines couvertes :
  - Sciences de l'ingénieur
  - Sciences vétérinaires
  - Sciences de l'éducation
  - Sciences de l'information et de la communication
- 900 publications/an
- 1100 chercheurs

# Intérêt

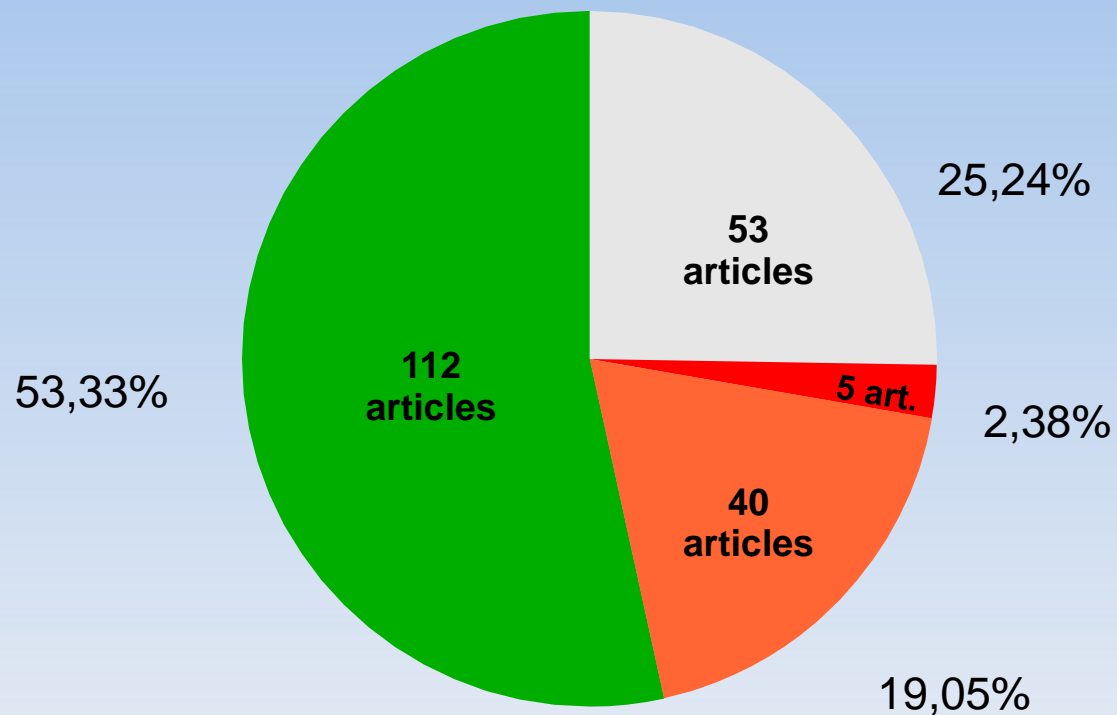
- Renforce la **cohérence** des équipes autour d'un projet transversal
- Valorise le **savoir faire des SCD** au sein des établissements
- Favorise les **contacts** entre chercheurs et bibliothèques : une archive ouverte de proximité
- Renforce les **coopérations** inter-établissements
- **Soude les composantes** et les équipes de recherche autour d'un projet d'établissement

# Convaincre...

## Une étude préalable du SCD

- **Juin 2006** : analyse des conditions juridiques d'archivage institutionnel selon les éditeurs
- 210 articles parus en 2005 dans des revues commerciales à comité de lecture (6 laboratoires)

# Résultat de l'étude



■ Diffusion autorisée

■ Diffusion payante

■ Diffusion non autorisée

■ Pas d'information sur le site de l'éditeur

# Informer...

- **Les instances** : CS, CA, directeurs labos
- **Les décideurs** : Président, VP CS, les responsables de la recherche
- **Les chercheurs** publiants
- **Les professionnels** de la documentation

# Piloter...

- **Projet à l'initiative du SCD et de la direction de la recherche : validation en juin 2006**
- **Pilotage** du projet par le SCD
- **Pas de postes dédiés**
- **Financement** demandé dans le contrat quadriennal 2007-2010

Objectif attendu : **1500 dépôts fin 2010...**

# Un groupe de travail

**embre 2006.** Constitution d'un groupe de travail mixte constitué de :

Correspondants laboratoires

Correspondants bibliothèques

Représentants d'autres universités intéressées par le projet

D'un représentant du SICD

# Les objectifs

nions du groupe en 1 an ...

**Définir** les besoins et les contraintes des laboratoires

**Décrire** les différents scénarios des dépôts possibles

S'accorder sur la **typologie** des documents diffusés dans l'archive ouverte.

**Respecter** les normes techniques et juridiques en vigueur

# Les grands principes retenus

**Un dépôt unique** pour le chercheur dans UNE seule archive. Pas de double dépôt dans HAL : incompatibilité avec l'archive nationale

**Absence d'obligation** de dépôt mais...

Une archive ouverte **doit avoir un contenu** :

le texte intégral doit être obligatoirement déposé

Exception pour les documents confidentiels

**Contrôle qualité rigoureux** et à plusieurs

# Les grands principes retenus

responsabilités **bien réparties** : un workflow simple

**posant** :

est garant de la **validité scientifique**

dispose de l'**autorisation** de tous les auteurs

dispose des **droits** concernant les documents tiers (photos, etc..)

**estionnaires** de l'archive assurent :

la **validation technique** des documents et des métadonnées,

# Les grands principes retenus

dépôt exclusif de **postprints** dans l'archive verte

dépôt **rapide** : pas plus de 5 minutes

format de diffusion **unique** : le format PDF

interface web en **langue anglaise**

s de classification thématique

typologie de documents définie avec les chercheurs compatible avec l'archive nationale

# Typologie

Article in a journal, magazine, newspaper. Not necessarily peer-reviewed. May be electronic-only medium, such as an online journal or news website. AUTHOR'S PERMISSION please : Many "Green" publishers do not allow the use of the published pdf

## **Conference or Workshop Item**

Paper, poster, speech, lecture or presentation given at a conference, workshop or event.

## **Conference**

Communication expressed at the time of a congress and answering an invitation. May be given in opening of the event (plenary conference).

## **Conference**

Communication (with support) made as a specialist (except congress) and answering an invitation launched by a university, a school, an organization, a company,...

## **Publication**

# Typologie

**Thesis**  
d thesis.

tation à diriger des recherches.

University work (except thesis): master, maitrise, DEA, DESS, CNAM.

lished patent.

## **scientific article**

ment allowing the largest audience to reach the culture scientific, technical,  
trial and environmental

al reports (only for archive).

# Browse by

ar

rson

boratory

stitution

Université de Toulouse

French research institution

Other partners

# Plateforme de diffusion

x « provisoire » : **Eprints**, logiciel libre OAI

**Déjà utilisée** pour l'archive des thèses

Facilement **configurable**

Une communauté **active**

Principales fonctionnalités :

- Autocomplete-as-you-type
- Export simple de ma bibliographie vers des logiciels bibliographiques
- Alertes spécifiques par mail ou fil RSS

## **groupe de travail « pro doc »**

**septembre 2007.** Création d'un groupe de travail « pro doc ». 7 réunions en 3 mois

**travail sur les métadonnées.** Assurer la **compatibilité** avec :

archive nationale

recommandations du GTA0

<http://gtao.wikidot.com>)

## groupe de travail « pro doc »

rendre **compatible** l'archive locale avec HAL  
signifie d'associer :

un auteur à un laboratoire connu de l'archive  
nationale

à chaque Eprints, la classification de l'archive  
nationale

## groupe de travail « pro doc »

Définition des **métadonnées obligatoires, visibles ou non** au déposant, pour chaque type de document

Définition des **règles de saisie** (largement reprises des normes de saisie de catalogage)

Travail sur **l'interface** web du site

**Rédaction** de la FAQ, des aides en ligne, de la charte de dépôt, etc...

Mise en place **technique** de la maquette suivant les recommandations. Connexion au LDAP, etc

ouverture officielle de l'archive ouverte OATAO  
(en Archive Toulouse/Toulouse Archive Ouverte)

**le 17 janvier 2008**

**Nom de domaine PRES :**

<http://oatao.univ-toulouse.fr>

# Organisation

**une équipe de pilotage** constituée des différents SCD

**une équipe staff-oatao** (~ 10 personnes)

Modération des articles (double, ...)

Contrôle des métadonnées

Contrôle technique des documents

Contrôle des droits de diffusion

Formation des utilisateurs

**une équipe tech-oatao** (3 personnes)

Contrôle qualité

# Organisation

## es outils

Un wiki de travail

Google documents

Des listes (alias) pour :

- Les gestionnaires (staff-oatao@)
- Les administrateurs (tech-oatao@)
- Les politiques (oatao@)

es réunions régulières

# Organisation

## **Un peu de technique**

Un serveur virtuel dédié : Red Hat Enterprise Linux Server

2Go de RAM/38 Go de disque

Sauvegarde assurée par la DSI

Eprints version 3.1.2.1 (Chocolate-coated Coffee Bean)

**Temps moyen par document : gestion**

20 à 45 min

# Faire savoir

## Communication interne

Bilans réguliers en **Conseil Scientifique**

Réunions/formation des chercheurs

Rédaction de « **flyers** »

Articles dans le journal de l'INP

## Communication externe

Déclaration d'OATAO dans les différents **sites de références** : ROAR, DRIVER, Eprints, OAISTER...

# Calendrier (en résumé)

**in 2006** : validation en conseil scientifique de l'INP

**ptembre 2006** : rédaction cahier des charges,  
définition des besoins

**ptembre 2007**: développements et paramétrages

**nvier 2008** : démarrage opérationnel

**nvier 2010** : connexion avec l'archive nationale

# Les difficultés

**Complexité** des tutelles des laboratoires

Des discours contradictoires aux chercheurs

Des oppositions factices : archive nationale/archive locale

Une archive nationale qui ne « moissonne »

pas. **Obligation** de passer par les **webservices**

Des **peurs** concernant les droits de diffusion

Un **argumentaire** à développer

# Les difficultés

**Information** des gestionnaires : de nouvelles compétences à acquérir

Manque du **PDF** auteur

Modification des **politiques** des éditeurs (Blackwell racheté par Wiley)

Clarifier les **droits** de certains types de documents

**Temps de travail** : ½ heure par article en moyenne

# Les stratégies de contournement

Les « **stagiaires** » des laboratoires

Double rôle : dépôt et formation

Financement par le laboratoire

Recrutement et encadrement par le SCD

Prendre en place d'**alias mail**

**Correspondants** de laboratoires : à l'ISAE,  
Information des secrétaires de départements

# Convaincre : les points clés

Faciliter la **rédaction** des rapports  
annuels

Exportation des données (divers formats)

Rendre **visible** le travail des laboratoires

sur le site web du laboratoire (affichage adapté)

[www.envt.fr/Documentation/ArchivesOuvertes.php?choix=xenobiotiques](http://www.envt.fr/Documentation/ArchivesOuvertes.php?choix=xenobiotiques)

sur les pages personnelles des chercheurs

[personnel.isae.fr/laurent-joly/article/liste-de-publications.html](http://personnel.isae.fr/laurent-joly/article/liste-de-publications.html)

[www.enfa.fr/recherche/laboratoires/toulouse-educagro/isabelle-fabre/](http://www.enfa.fr/recherche/laboratoires/toulouse-educagro/isabelle-fabre/)



## **2.Statistiques**

# Contenu par institution

**2710 documents au 4 février 2010**

ENFA	23	1%
ENVV	865	32%
INPT	1206	44%
ISAE	616	23%

# Contenu par institution

## Autres institutions toulousaines :

JT1	58	2%
JT2 (Mirail)	55	2%
JT3 (Paul Sabatier)	1170	43%
NSA	107	4%

# ENVY

ATAO est utilisé pour la diffusion des thèses

à faible taux de couverture de la production  
de thèses : moins de 7 %

# INPT

4 unités de recherche (toutes UMR) :

3 déposent la totalité de leur production  
2 sont entrés courant 2009

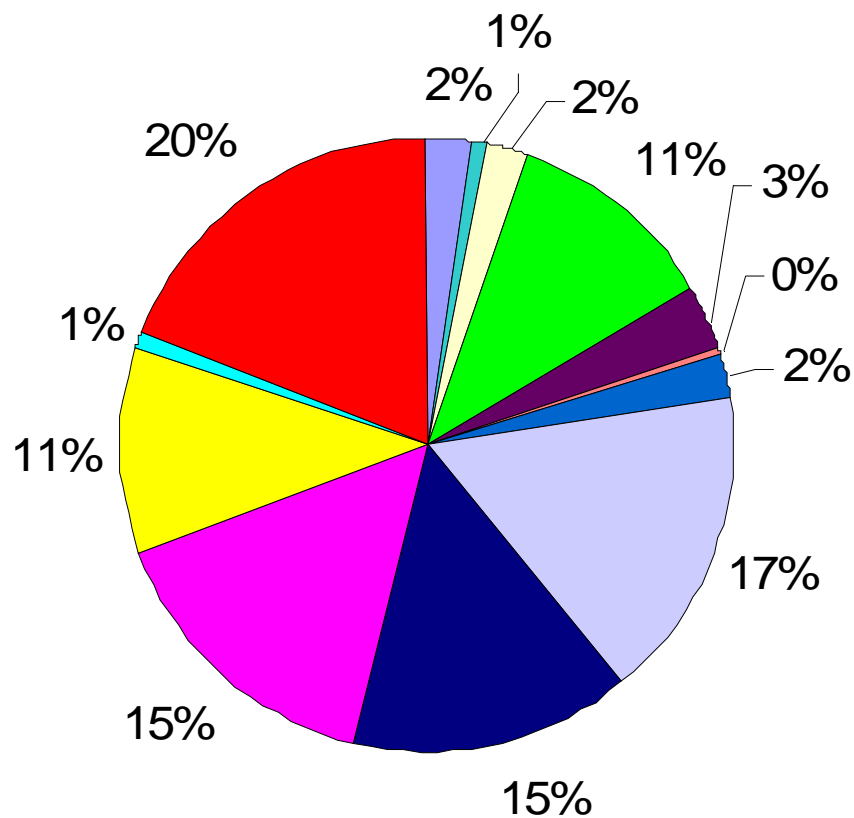
équipes (45 % de la production) déposent de manière  
collective

certains chercheurs déposent de manière isolée

absence d'obligation de dépôt

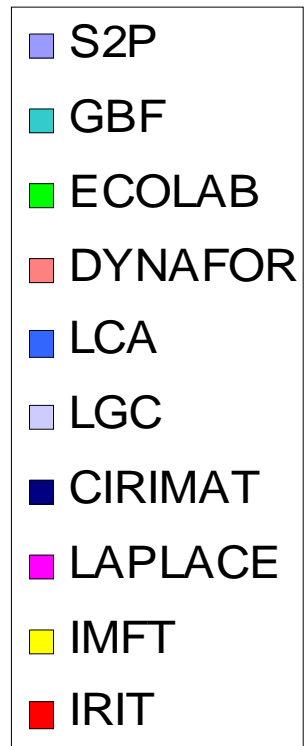
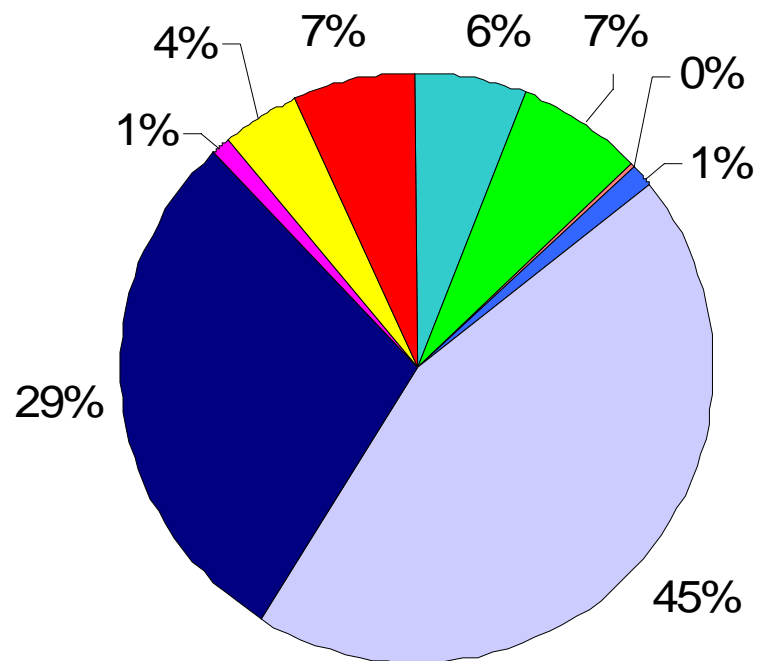
taux de couverture : 26,8% (2008)

# INPT / Publications par labo (2007)



- S2P
- GBF
- TANDEM
- ECOLAB
- AGIR
- DYNAFOR
- LCA
- LGC
- CIRIMAT
- LAPLACE
- IMFT
- LOSE

# PT / part des labo dans OATAO



# ISAE

citation forte de l'institution à déposer :  
tion d'un indicateur à partir des données  
AO

aux de couverture 70 % (ensemble des  
ications depuis 2005)

# Contenu : documents

55 % datent de **3 ans ou moins**

83 % sont en **langue anglaise** ; **90 %** si l'on  
exclut les thèses

22 % des documents sont des **articles**

25 % sont des conférences

29 % sont des thèses

# Contenu

## ENVV

date de publication 2007 +	28%
articles	8%
dont internationaux à comité de lecture	90%
thèses	92%

# Contenu

## INPT

date de publication 2007 +	56%
articles	72%
dont internationaux à comité de lecture	97%
conférences	25%

# Contenu

## ISAE

date de publication 2007 +	53%
articles	27%
dont internationaux à comité de lecture	87%
conférences	66%

# Usage : année 2009

116 visites

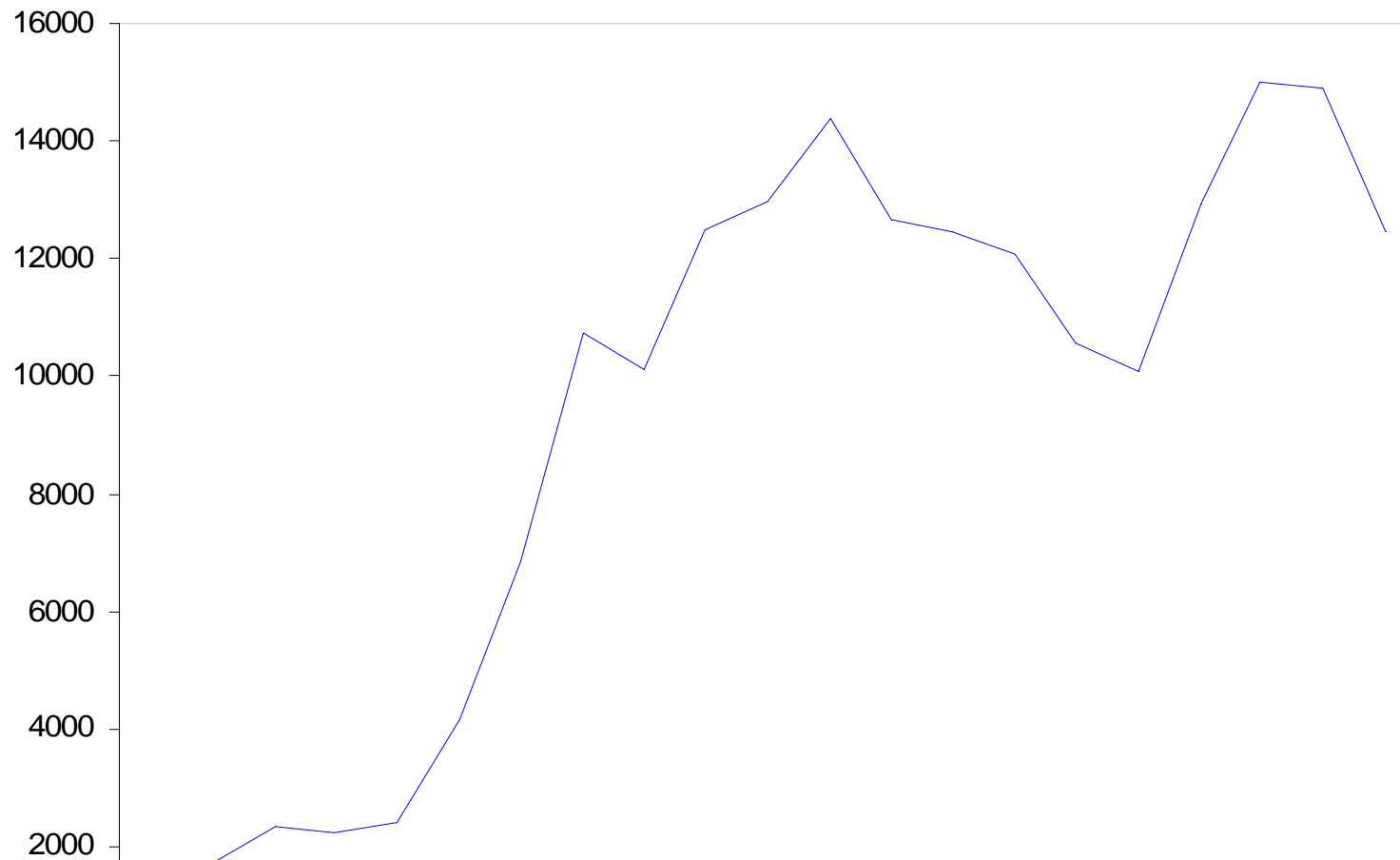
Taux de rebond : 66%

7059 visites « actives »

518 documents téléchargés

Téléchargements / visite active en moyenne

# Visites 2008-2009



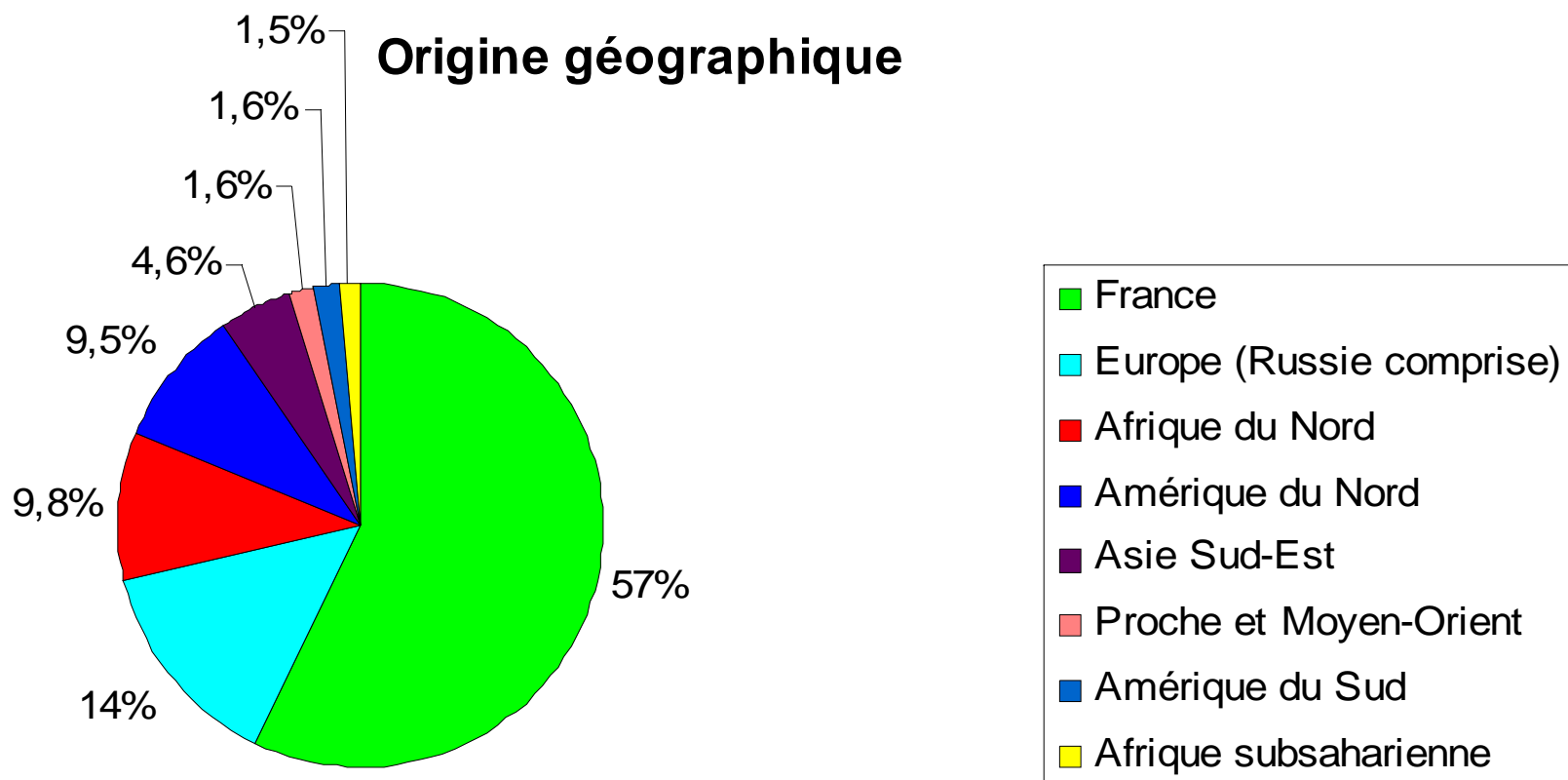
# Visites

Visites directes : 5 %

Visites via site référent : 9 %

Visites via moteurs de recherche : 86 %  
(Google : 80 %)

# visites : origine géographique



The image features a light blue gradient background. At the top, there are two horizontal bars: a wider, lighter blue bar and a narrower, darker blue bar below it. The text '3.Perspectives' is centered in the lower half of the page.

## **3.Perspectives**

# Evolution technique : une nouvelle plateforme

ation à terme vers ORI-OAI (Outil de  
rérencement et d'Indexation) :

<http://www.ori-oai.org>

antages :

ure intégration au SI (ENT en particulier)

s aux données des briques du SI. Ex. : Webservice  
ogée

s aux WebServices de la plateforme STAR et de

chines nationales

# amélioration du contenu et des services

Vers un **dépôt systématique** : adoption de chartes d'établissement

Des **services** toujours plus proches des chercheurs : la question des **statistiques** et la nécessité d'une norme internationale

La **sauvegarde** externe et la **conservation** : CINES

Le **web sémantique** (notamment pour les collaborations)

# Une meilleure visibilité

## **Visibilité nationale :**

et dans l'archive nationale via les webservice

## **Visibilité locale :**

et le directeur du PRES toulousain

et le directeur à d'autres établissements  
scientifiques toulousains

# Une visibilité internationale

Visibilité internationale :

Groupement OATAO – EPFL – UPC

(Polytechnique Fédérale de Lausanne

et Institut Politècnica de Catalunya )

Mise en place d'un moissonneur pour un accès unique aux trois entrepôts














Meilleure visibilité de nos publications

Accès facilité pour le chercheur

# Classement Webometrics 2010

24e

200e

1	Smithsonian/NASA Astrophysics Data System*		2	1	507	1
2	Hal CNRS		8	3	3	3
3	Institut National de Recherche en Informatique et en Automatique Archive Ouverte		17	4	19	8
4	CERN Document Server		4	8	40	5
5	École Polytechnique Federale de Lausanne Infoscience		3	6	44	84
6	MIT Dspace		5	21	5	4
7	University of Michigan Deep Blue		23	19	10	13
8	Hyper Article en Ligne Sciences de l'Homme et de la Société		80	5	109	21
9	Tufts University Perseus Digital Library		1	2	409	524
10	University of Southampton: Department of Electronics and Computer Science		20	11	25	57
11	Igitur Archive Universiteit Utrecht		22	46	1	19
12	Universitat Stuttgartelektronische Hochschulschriften		33	7	28	225
13	Dissertations of the University of		42	37	2	24



# Une visibilité internationale

une archive thématique en Sciences de l'Ingénieur

Président du consortium européen ECATA (aerospace)

Élargissement à l'échelle européenne

Financement UE dans le cadre du programme *E-Content*

# Conclusion : les clés du succès

l'archive ouverte doit constamment évoluer :

proposer en permanence des services adaptés aux besoins de l'institution et de ses chercheurs

être visible

proposer un accès simplifié et lisible à l'information

implication de la direction de l'établissement

indicateur décisif : taux de couverture

effort de communication auprès des

# Conclusion

archive ouverte institutionnelle locale,  
forme à l'esprit de la LLRU, permet à  
institution autonome de maîtriser sa politique de  
sion

archive ouverte institutionnelle locale permet  
accès unifié à l'information par le biais de  
sonneur nationaux ([Australie](#), [Pays-Bas](#)) ou  
atiques ([Economists Online](#))

# Contacts

Sandrine Malotaux ([sandrine.malotaux@inp-toulouse.fr](mailto:sandrine.malotaux@inp-toulouse.fr))

Jean-Marie Le Behec ([Jean-Marie.LeBehec@inp-toulouse.fr](mailto:Jean-Marie.LeBehec@inp-toulouse.fr))

OATAO : <http://oatao.univ-toulouse.fr>

