

# L'édition électronique, l'Internet, la « littérature grise » et les archives ouvertes

- Définition de « littérature grise »
  - Thèses, articles ou eprints
  - Serveurs français, internationaux
- Production, diffusion, archivage
  - Formats de fichier
- Le serveur
  - Métadonnées, archives ouvertes
  - Outils utilisables pour la mise en place d'un serveur

# Littérature Grise

- Un document produit par un chercheur:
  - Un article, destiné à la publication ou « eprint »
  - Un rapport de recherche
  - Une thèse ou habilitation à diriger les recherches

# La thèse

- Un diplôme
- Un document spécifique, des règles spécifiques:
  - Exemplaires de soutenance
  - Bordereau et exemplaires de diffusion
  - Catalogage obligatoire par la BU de l'université de soutenance
- Nouvelles dispositions pour thèses en ligne
- <http://www.sup.adc.education.fr/bib>

# L'article

- Destiné à la publication
- Rédaction
- Diffusion à des pairs,
  - Email
  - Serveur de prépublication ou eprint
- Soumission à un journal
- « review » par le comité éditorial
- Publication , ou refus

# Serveurs de prépublications (« eprints »)

- A l'initiative des laboratoires de recherche ou instituts
  - Labos de math ... ex <http://www-fourier.ujf-grenoble.fr>
- Initiative nationale - le CCSD
  - <http://www.ccsd.cnrs.fr>
- Niveau International: LE serveur ArXiv
  - <http://www.arxiv.org>
- Autres initiatives: sciences cognitives...

# Technique: formats de fichier

## 1) Le document original

- 2 familles d'outils de composition de documents:
  - Famille « word » (inclut word, wordperfect, open office, star office...)
    - > document « binaire », comporte du texte des codes
    - > Pour avoir un document « réutilisable », nécessité d'enregistrer au format « rtf »
  - Famille « TeX » (inclut LaTeX, TeXMacS...)
    - Textes comportant des formules de math
    - > document « texte », comporte du texte structuré par des balises.

# Famille word...

- Le texte proprement dit est mélangé à des codes binaires servant à la mise en page.
- Nécessité d'avoir le logiciel de création pour lire le document.
- Le format d'enregistrement RTF (rich text format) est un format d'échange qui ne contient que de l'ascii (texte seul) et des balises de formatage physique (caractères gras, paragraphes...). produit et lu par une multitude de programmes.

# Famille TeX

- Créé par Donald Knuth dans les années 1970
- Le but: produire des textes scientifiques, contenant des formules mathématiques, avec une haute qualité typographique.
- Un fichier (La)TeX est un fichier ascii contenant des balises de mise en page ET de structuration logique, ainsi que différents classes (types) de documents (article, book, letter, etc...)
- Le fichier source TeX est traité par un "moteur" pour obtenir un document imprimable(dvi -> postscript ou pdf)
- (La)TeX est largement utilisé pour les textes en mathématiques et physique.
- Seule méthode pour produire un texte contenant des formules compliqués.
  - Pour un exemple, voir le fichier <http://tel.ccsd.cnrs.fr/documents/archives0/00/00/26/71/tel-00002671-01/tel-00002671.pdf>

# Un morceau de LaTeX

- `% Article pour le colloque Bony (Asterisque)`
- `\documentclass[reqno,english]{smfart}`
- `\usepackage[english]{babel}`
- `\usepackage{smfthm}`
- `\NumberTheoremsIn{subsection}`
  
- `\author{Yves Laurent}`
- `\address{ Institut Fourier Math\`ematiques \\  
UMR 5582 CNRS/UJF\  
BP 74\  
38402 St Martin d'H`eres Cedex}`
- `\email{Yves.Laurent@ujf-grenoble.fr}`
- `\urladdr{http://www-fourier.ujf-grenoble.fr/~laurenty}`
- `\title{Regularity of  $\mathcal{D}$ -modules associated to a symmetric pair}`
- `\alttitle{R`egularit`e des  $\mathcal{D}$ -modules associ`es `a une paire sym`etrique} \begin{document}`
- `\frontmatter`
- `\begin{abstract}`
- The invariant eigendistributions on a reductive Lie algebra are solutions of a holonomic  $\mathcal{D}$ -module which has been proved to be regular by Kashiwara-Hotta. We solve here a conjecture of Sekiguchi saying that in the more general case of symmetric pairs, the corresponding module are still regular.
- `\end{abstract}`

# Technique: formats de fichier

## 2) Le document affichable

- Il s'agit d'un format conforme aux possibilités techniques du moment, destiné à l'affichage sur le net à un instant T.
- Actuellement, dans la grande majorité des cas, on proposera le format PDF, généré par différents outils:
  - Acrobat ou autre outil pour les documents « word »
  - PdfTeX pour les documents écrits en TeX.
- Ou on peut proposer le html (plus rare)

# Technique: formats de fichier

## 3) Le document structuré

- Intérêt de **conserver** un document pérenne où la structure logique apparaît clairement.
- Ce document est indépendant des plateformes matériels et des outils logiciels et doit être lisible.
- A partir de ce document structuré:
  - différentes vues pourront être générées (PDF, html, autre format non encore connu)
  - des informations (métadonnées) pourront éventuellement être extraites. (titre, tables des matières, auteur, date, université de soutenance, discipline, ...)

# Technique: formats de fichier

??? XML ???

- XML est une « recommandation » défini par le W3C. ([www.w3c.org](http://www.w3c.org))
- XML sert à:
  - Structurer un document « primaire », tel qu'une thèse, une monographie, un rapport, un article, ...
  - Echanger facilement des données:
    - Données bibliographiques souvent
    - Autres données factuelles, (ex: les étudiants inscrits à telle formation universitaire)
  - Structurer un site web, ou partie de site, ...

# Technique: formats de fichier

## ??? XML ???

- Un document XML:
  - Texte balisé (entre balise ouvrante et fermante), comme l'html.
  - L'utilisateur a le droit de définir « ses » balises. (écriture d'une DTD ou d'un schéma)
  - Un document XML doit être:
    - Conforme à une DTD ou schéma
    - « bien formé » (ouverture, fermeture de balises sans imbrication)
  - Si on connaît la dtd ou le schéma, on peut lire le document.

# Un document XML ??

- `<article type="normal">`
- `<idart>AIF_2000__50_6_1709_0</idart>`
- `<pagedeb systnum="arabe" pagination="normal" typepag="normal">1709</pagedeb>`
- `<pagefin systnum="arabe" pagination="normal" typepag="normal">1744</pagefin>`
- `<ordre>0</ordre>`
- `<auteur>`
- `<nom>Mignon</nom>`
- `<prenom>Thierry</prenom>`
- `</auteur>`
- `<titre xml:lang="fr">Courbes lisses sur les surfaces rationnelles génériques : un lemme d'Horace différentiel.</titre>`
- `<langue>fr</langue>`
- `</article>`

# Affichages possibles pour ce document...

- Article n° AIF\_2000\_\_50\_6\_1709\_0
  - Mignon, Thierry .- *Courbes lisses sur les surfaces rationnelles génériques: un lemme d'H* .- publié en français dans les Annales de l'Institut Fourier: vol 50, fasc 6, 2000, pp 1709-1744
- **Identifiant:** AIF\_2000\_\_50\_6\_1709\_0
  - **Auteur:** Mignon, Thierry
  - **Titre:** Courbes lisses sur les surfaces rationnelles génériques: un lemme d'H
  - **Langue:** Français
  - **Année de Publication:** 2000
  - **Journal:** Annales de l'Institut Fourier, 50, 6

# L'XML ne comporte aucune notion de présentation

- Un document XML comporte des balises de structuration et du contenu.
  - La présentation d'un document XML peut se faire à travers une feuille de style (fichier .xsl), ou un document de transformation (fichier .xslt), associé à un « moteur xslt »
  - Selon les transformations à effectuer, on utilisera l'une ou l'autre méthode.
  - La réf: <http://www.w3.org/XML/>
- (nombreux stages et cours dispo sur le marché)

# Le serveur de documents

- Dans sa plus simple expression, une collection de documents sur un site web, comportant une ou plusieurs pages contenant des liens vers le document final.
- Un peu plus compliqué: on propose plusieurs possibilités de navigation: par date, par auteur, etc...
- Très vite on se rend compte de la nécessité d'une base de données et de métadonnées.

# Métadonnées ?

- Des données bibliographiques.
- « champs », auteur, titre, date, mots clés...
- Une norme « minimum » le « dublin core »
  - <http://dublincore.org/documents/1998/09/dces>
- Comment les créer, et ou les ranger ?
- La meilleure solution: une base de données permettant la génération de « pages dynamiques ».
  - <http://www.univ-orleans.fr/SCIENCES/MAPMO/publications/index.php>

# Open Archives Initiative

- Toujours du nouveau sur le web... ☺
- Un groupe de gens responsables d'importants services de documents en ligne (prépublications, ou « eprints ») s'est constitué pour réfléchir au moyen de partager les données bibliographiques, et aller vers la constitution du « web sémantique » (content based)
- Rendre aux scientifiques « leur » web qui est pollué par des services et vendeurs divers et variés.

# OAI: l'initiative pour les archives ouvertes

- **L'idée:** Mettre en place des archives de documents numériques c'est bien, mais pas très utile si personne ne connaît l'existence de l'archive, et s'il faut aller sur plusieurs sites pour trouver le document recherché.
- **Le moyen:** un protocole pour « exposer » et « récolter » les données bibliographiques des différents serveurs.

# L'OAI comporte deux groupes de gens:

- Les « data providers », qui gèrent un service d'archivage.
- Les « service providers », qui ajoutent de la valeur aux métadonnées qu'ils trouvent chez les « data providers ».
- Le data provider doit fournir des métadonnées conformes au protocole OAI
- Le « service provider » doit savoir les lire et les exploiter.
- Un gain de temps pour l'échange d'information bibliographique.

# L'index national des prépublications, thèses et habilitations de mathématiques: un petit exemple français sur l'utilisation d'OAI

- L'index des prépublications et thèses de mathématiques a toujours été basé sur le principe de « récolte de métadonnées ».
- Un moteur de recherche « web » parcourt une 30aine de sites et y ramasse des fichiers contenant des métadonnées. (ancien système)
- Un programme « récolte » les métadonnées concernant les thèses de mathématiques des serveurs TEL et HAL.
- L'ensemble est « fusionné » dans une seule base.

# Le CCSD et MathDoc

- Le CCSD (Centre pour la Communication Scientifique directe, <http://www.ccsd.cnrs.fr>) joue le rôle de « data provider », c'est sur ce serveur que sont stockés les documents et les métadonnées
- MathDoc (Cellule de Coordination Documentaire Nationale pour les Mathématiques, <http://mathdoc.emath.fr>) joue le rôle de « service provider » thématique, avec les métadonnées récoltées il construit différents services annexes: tris par auteur, par date, etc..., à partir de sa base de données « fusionnée ».
- Index des prépublications, index des thèses

# Exemple d'OAI

- [http://hal.ccsd.cnrs.fr/oai/oai.php?verb=ListRecords&set=MATH&metadataPrefix=ccsd\\_hal](http://hal.ccsd.cnrs.fr/oai/oai.php?verb=ListRecords&set=MATH&metadataPrefix=ccsd_hal)
- Je veux tous les enregistrements dans le domaine « math » avec l'ensemble de métadonnées « ccsd\_hal »
- Je récupère un fichier xml que je traite par programme pour insérer les données dans ma base.

# Liste thèses 2002 sur TEL

CCSD thèses-EN-ligne: Liste par domaine : Mathématiques <br>2002 - Microsoft Internet Explorer

Fichier Edition Affichage Favoris Outils ?

Précédente

Rechercher Favoris Média

Adresse [http://tel.ccsd.cnrs.fr/view-thes-math\\_2002\\_fr.html](http://tel.ccsd.cnrs.fr/view-thes-math_2002_fr.html) OK Liens

Favoris

Ajouter... Or

Achats

Cartes

chorale

Liens

Loisirs

Media

meteo

voyages

annuaire univers...

Banque Populaire...

Calendrier CMD

Captaine-doc, le...

Catalogue fusion...

Droits d'auteur ...

Guide des statio...

INIST-CNRS 2èm...

L'éditeur JavaSc...

Les Pages Blanc...

META HTTP-EQU...

MSN.com

Nouvel Accès PP...

OAI (Open Archi...

Open Archives I...

Open Digital Libr...

phpMyAdmin

PPP UJF

RealPlayer Home...

thèses-EN-ligne

## Liste par domaine : Mathématiques 2002

Page d'accueil

Lister

Rechercher

S'enregistrer

S'abonner

Déposer

Vos documents

Aide

A propos

- [\(thèses-EN-ligne\)](#)
  - [Thèses](#)
    - Mathématiques (137)

[Liste par auteur](#)

[2003](#) - [2002](#) - [2001](#) - [2000](#) - [1999](#) - [1998](#) - [1990](#)

Ce domaine contient **76** entrées pour 2002.

[TCHOUDJEM, Alexis. Représentations d'algèbres de Lie dans des groupes de cohomologie à support. Thèse de doctorat \(20 décembre 2002\). Institut Fourier, UNIVERSITE JOSEPH-FOURIER - GRENOBLE I.](#)

[RICARD, Hélène. Résolution avec régularité jusqu'au bord de l'équation de Cauchy-Riemann dans des domaines à coins et de l'équation de Cauchy-Riemann tangentielle en codimension quelconque. Thèse de doctorat \(20 décembre 2002\). Institut Fourier, UNIVERSITE JOSEPH-FOURIER - GRENOBLE I.](#)

[Verquinox, Roland. KK-théorie équivariante et opérateur de Julg-Valette pour les groupes quantiques. Thèse de doctorat \(19 décembre 2002\). Mathématiques, UNIVERSITE DENIS DIDEROT - PARIS VII.](#)

[Obeid-El Hamidi, Amira. Sur une équation elliptique non linéaire dégénérée. Thèse de doctorat \(19 décembre 2002\). Département de Mathématiques/Laboratoire de Mathématiques Appliquées, UNIVERSITE DE PAU ET DES PAYS DE L'ADOUR. Il existe 2 versions de cette thèse.](#)

gestion

le diapositive

des diapos

en page des diap

exte

ontenu

exte et du

Insertion de nouvelle

Démarrer

Excel

Terminal

D:\Documents...

Eudora

Microsoft Pow...

Thèses de Mat...

CCSD thèses...

FR

16:31

# Liste thèses 2002 MathDoc

Bonjour, nous sommes le 01 avril 2003

Accueil | Nouveautés | Lettre d'information | Plan du site | Contactez-nous

## Liste des thèses et habilitations mathématiques - 2002

*List of mathematical theses and dissertations - 2002*

Légende: *T* = thèse ou non précisé, *H* = habilitation à diriger des recherches  
Legend: *T* = PHD thesis or not specified, *H* = habilitation (post doctoral dissertation)

2003 2002 2001 2000 1999 1998 1997 1996 1995 1994 1993 and before

2002-12-20	TCHOUDJEM, Alexis	<a href="#">Représentations d'algèbres de Lie dans des groupes de cohomologie à support</a>
2002-12-20	RICARD, Hélène	<a href="#">Résolution avec régularité jusqu'au bord de l'équation de Cauchy-Riemann dans des domaines à coins et de l'équation de Cauchy-Riemann tangentielle en codimension quelconque</a>
2002-12-19	EL MACHKOURI, Mohamed	<a href="#">Théorèmes limite pour les champs et les suites stationnaires de variables aléatoires réelles</a>
2002-12-19	BRASSART, Matthieu	<a href="#">Limite semi-classique de transformées de Wigner dans des milieux périodiques ou aléatoires</a>
2002-12-19	Obeid-EI Hamidi, Amira	<a href="#">Sur une équation elliptique non linéaire dégénérée</a>
2002-12-19	Hoernel, Jean-David	<a href="#">Etudes théorique et numérique d'un modèle non-stationnaire de catalyseurs à passages cylindriques</a>
2002-12-19	EL MACHKOURI, Mohamed	<a href="#">Theoremes limite pour les champs et les suites stationnaires de variables aleatoires reelles</a>
2002-12-19	EL MACHKOURI, Mohamed	<a href="#">Theoremes limite pour les champs et les suites stationnaires de variables aleatoires reelles</a>

# Tout sur OAI...

- Une seule adresse:
- <http://www.openarchives.org/>

Tout service de documents scientifiques « en ligne » devrait comporter un serveur OAI

- des logiciels « tout fait » existent pour gérer des sites

# Outils pour un serveur de documents

- Chaîne de production « Lyon 2 »  
<http://sourcesup.cru.fr/cybertheses/>
- ETD-DB (thèses) de virginia Tech
  - <http://scholar.lib.vt.edu/ETD-db/>
- Le logiciel eprints (southampton)
  - <http://software.eprints.org/>
  - Basé sur MySql et perl, compatible OAI
- Le logiciel D-Space (MIT)
  - Ensemble logiciel plus conséquent
  - <http://dspace.org/>

# Bibliographie (sites)

- **Serveurs littérature grise en france**
  - <http://theses.univ-lyon2.fr/>
  - <http://tel.ccsd.cnrs.fr> et <http://hal.ccsd.cnrs.fr>
  - <http://mathdoc.emath.fr/Prepub/> et <http://mathdoc.emath.fr/Theses/>
  - <http://bibliotheques.univ-lille1.fr/default.asp?bustl/grisemine>
  - ...
- **Ministère (directive sur les thèses):**
  - <http://www.sup.education.fr/bib/>
- **Serveurs Internationaux**
  - <http://fr.arxiv.org/>
  - <http://www.ndltd.org/>
  - <http://www.eprints.org>
  - <http://repec.org/>
  - ...
- **Serveur OAI**
- <http://www.openarchives.org/>

# A partir d'ici c'est facultatif...

- Ce sont des diapos concernant les thèses en ligne, faits pour un autre cours.

# La thèse – le diplôme (extraits de la charte des thèses)

- **Sujet précisé à l'inscription**
  - L'inscription en thèse précise le sujet, le contexte de la thèse et l'unité d'accueil.
- **Novateur et d'actualité**
  - Le choix du sujet de thèse repose sur l'accord entre le doctorant et le directeur de thèse, formalisé au moment de l'inscription. Le directeur de thèse, sollicité en raison d'une maîtrise reconnue du champ de recherche concerné, doit aider le doctorant à dégager le caractère novateur dans le contexte scientifique et s'assurer de son actualité ; il doit également s'assurer que le doctorant fait preuve d'esprit d'innovation
- **Durée 3 ans:**
  - La durée de référence de préparation d'une thèse est de trois ans. A la fin de la seconde année, l'échéance prévisible de soutenance devra être débattue, au vu de l'avancement du travail de recherche. .

# La thèse – le document

- **Documents officiels encore en vigueur aujourd'hui:**
- **L'Arrêté du 25 septembre 1985**
  - définit les modalités de dépôt, signalement et reproduction des thèses
- **Circulaire du 6 novembre 1985**
  - Précise les modalités
    - Le service de doctorat de l'établissement fournit le bordereau vierge au candidat et attribue un numéro d'enregistrement à la thèse.
    - C'est ce numéro, reporté sur le bordereau, qui donnera la référence de la thèse dans l'inventaire annuel.
    - Le candidat établit le bordereau en deux exemplaires, avec le concours de la bibliothèque de l'établissement.
    - Le candidat, après avoir éventuellement sollicité les conseils de son directeur de thèse, s'adresse à la bibliothèque qui lui apporte un soutien professionnel. Cette intervention de la bibliothèque - pièce maîtresse du dispositif de signalement - suppose les mesures appropriées : service d'accueil, réunion, notes d'information ...
    - Le candidat remet au service de doctorat les deux exemplaires du bordereau rempli, en même temps que les exemplaires de la thèse.
- **Circulaire du 11 mars 1996.**
  - Définit les modalités de catalogage
    - La bibliothèque de l'université de soutenance effectue le catalogage et l'indexation.
    - L'ABES se chargera du suivi de la saisie, des corrections et de l'intégration dans la base TELETHESES.
- **Depuis 2000 (?)**, Les thèses soutenues dans toutes les disciplines sont signalées dans le catalogue collectif des établissements d'enseignement supérieur (Système Universitaire de Documentation - [Sudoc](#)).

# La thèse: diffusion et archivage traditionnelle

- BU de soutenance (archivage exemplaires papier), PIB.
- Micro fiches: (conservation, diffusion)
  - **Pour les thèses de lettres, de droit, de sciences humaines**  
Atelier national de reproduction des thèses (ANRT)  
9, rue Auguste Angellier  
59046 LILLE CEDEX
  - Pour les thèses de sciences, économie, gestion**  
Atelier national de reproduction des thèses (ANRT)  
BP 47  
38040 GRENOBLE CEDEX 9
- Depuis le 1er janvier 1997, en partenariat avec l'ANRT de Lille, les Presses universitaires du Septentrion impriment et commercialisent les thèses des docteurs qui en font la demande. Leur [catalogue de thèses imprimées](#) est consultable.  
<http://www.septentrion.com/theses/choixtheses.html>

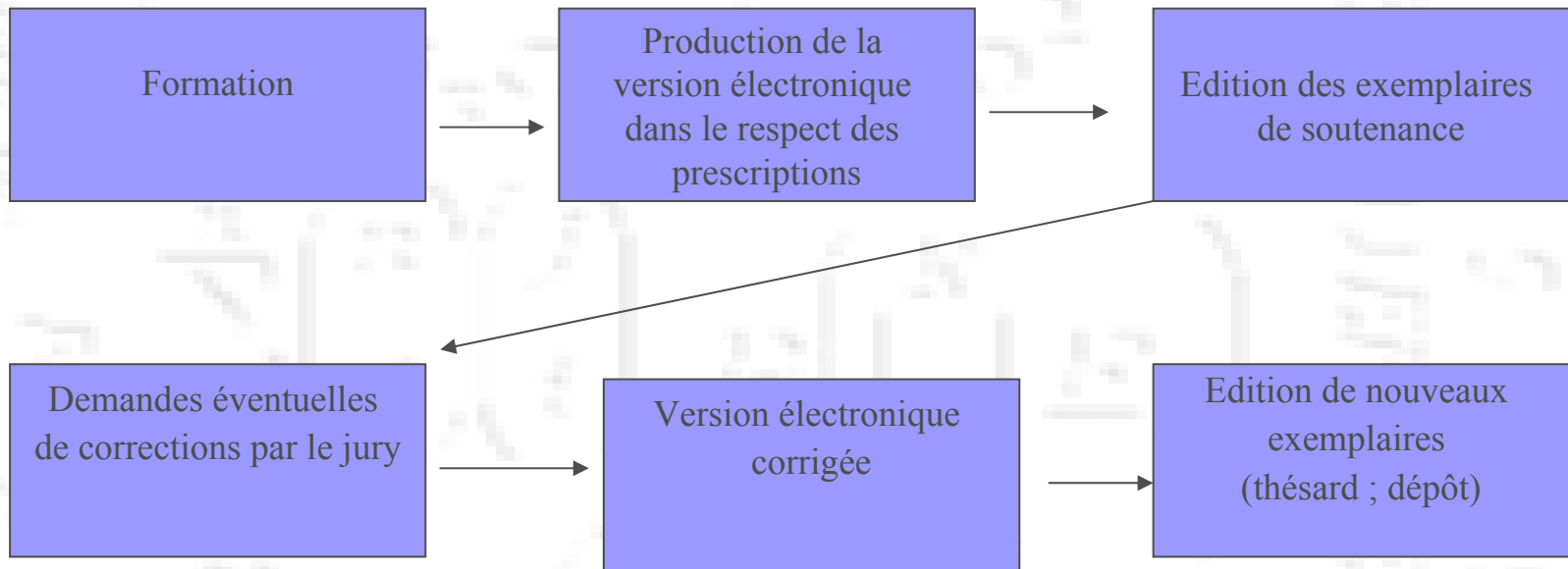
# Archivage et diffusion numériques

- Le thésard saisit son document à l'aide d'un ordinateur, et cela depuis longtemps.
- Il semble logique et économiquement intéressant de conserver et diffuser ces documents sous forme numérique.
- Le ministère de l'EN (SDBIB) a réuni à partir de 1999 un groupe de réflexion sur la diffusion et l'archivage électronique des thèses.
- Un texte «Guide pour la diffusion et l'archivage électroniques des thèses à l'attention des établissements d'enseignement supérieur » issu des travaux de ce groupe de travail a été publié en 2001.

# Guide pour l'archivage et la diffusion électronique des thèses

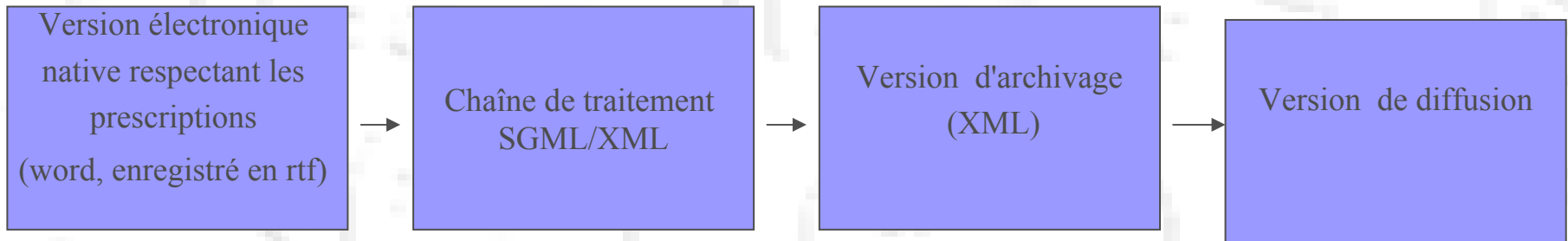
- Obligations du doctorant
  - l'utilisation d'un logiciel largement diffusé, en tout état de cause compatible avec les formats RTF ou TeX.
  - l'usage d'un *modèle de document* permettant de *structurer* la thèse.
  - diverses règles de présentation.
- Obligations de l'école doctorale
  - Formation des doctorants à l'utilisation d'outil de création d'un document structuré
- Les établissements sont laissés libres pour organiser le dépôt, stockage, catalogage...

# Circuit de la thèse « électronique »



# Chaîne de traitement de la thèse « électronique » préconisée par le ministère

---



# Et les thèses scientifiques?

☹ pas encore de solution  
d'archivage en XML ☹

- Il n'est pas actuellement possible de générer un document en XML à partir d'un document TeX ou LaTeX, à cause des formules de mathématiques.
- Le langage MathML existe mais ne comprend pas toutes les possibilités d'expression des mathématiques.
- On va conserver le format original (LaTeX)

# Initiatives françaises/francophones et internationales sur les thèses en ligne

- Françaises/francophones
  - L'université Lyon 2 et le projet « cyberthèses »
  - Le serveur multidisciplinaire « TEL »
- Internationales
  - NDLTD Networked Digital Library of Theses and Dissertations
- Nationales et/ou disciplinaires
  - DissOnline et MathDiss (sous ensemble)
  - Index des thèses de mathématiques en France

# L'Université Lyon 2

- A l'origine du groupe de travail ministériel, et de la directive.
- Application de la directive ☺
- Mise en place d'une chaîne de production des documents
- Mise à la disposition d'autres universités de leurs outils. (?)
- Compatible OAI
- Pour en savoir plus... <http://theses.univ-lyon2.fr/>

# Le serveur multidisciplinaire, et multi établissements « TEL »

- Créé par le CCSD (Centre pour la communication scientifique directe) et la Cellule MathDoc.
- L'idée: mettre en place une structure permettant le dépôt de thèses sans attendre la mise en place par les établissements. (manque de moyens)
- Offre la possibilité de générer des vues spécifiques par établissement ou même par laboratoire.
- Ne respecte pas la directive à cause de l'impossibilité de créer des fichiers XML d'archivage sur les textes scientifiques (math physique).
- Compatible OAI
- Pour en savoir plus <http://tel.ccsd.cnrs.fr/>

# NDLTD Networked Digital Library of Theses and Dissertations

- La base de thèses la plus importante
- Initiative lancée avant le protocole OAI
- Récolte de métadonnées facilitée par OAI
- En pleine mutation actuellement
- Cf le site <http://www.ndltd.org/>
- Search and browse for ETDs in the [NDLTD Union Catalog](#) which contains entries from multiple members institutions [hosted by [VTLS](#)]
  - Search and browse and ETDs in the experimental [OAI-based Union Catalog](#)