



# Le pôle de recherche sur la conservation des données numériques GIS-SPADON

Jean-Dominique Polack

Institut Jean Le Rond d'Alembert – équipe LAM

UPMC / CNRS UMR 7190 /

Ministère de la Culture et de la Communication

[jean-dominique.polack@upmc.fr](mailto:jean-dominique.polack@upmc.fr)

5 novembre 2014

PREDONx Workshop on Scientific Data Preservation

# Groupement d'Interêt Scientifique

## Supports Pérennes d'Archivage des DOnnées Numériques

- créé en 2004 à l'initiative du LNE,
- Objectifs :
  - étudier tout support numérique permettant de conserver l'information pendant plusieurs décennies voire siècles,
  - améliorer la conservation des informations enregistrées,
  - proposer des solutions d'archivage stables
  - guider les acteurs du domaine (instituts patrimoniaux, fabricants, utilisateurs) dans le choix des supports en précisant leurs conditions.

# Les 6 partenaires du GIS

- LAM (Institut Jean Le Rond d'Alembert, Equipe Lutherie, Acoustique et Musique) Université Pierre et Marie Curie
- CEA-LETI (Laboratoire d'Electronique de Technologie de l'Information), du Commissariat à l'Energie Atomique
- LPMM (Laboratoire de Photochimie Moléculaire et Macromoléculaire), Université Blaise Pascal de Clermont-Ferrand
- LNE (Laboratoire national de métrologie et d'essais),
- INA (Institut national de l'audiovisuel - Inathèque de France)
- IMS (Intégration du Matériau au Système), Université Bordeaux 1

# Plan de l'exposé

1. Historique
2. La problématique de l'archivage
3. Quelques résultats du GIS SPADON
4. Perspectives

# 1. Historique



## Le LAM et les DON

- 1980 création d'une équipe de recherche (Ministère de la Culture/  
Université P&M Curie/CNRS) sur la *conservation des enregistrements  
sonores*.
- 1995 hébergement de l'équipe au Laboratoire d'Acoustique Musicale (LAM,  
UMR 7604)
- 2000 numérisation des collections d'enregistrements par le Ministère de la  
Culture - choix *disque optique enregistrable* comme support et début  
des études sur ce support.

# Historique (2)

## Le LAM et les DON

**A** partir de 1990, **études expérimentales** sur :

- la qualité initiale des disques,
- le comportement de disques soumis à des environnements sévères,
- la problématique du vieillissement accéléré.

**E**n 2004, formation du GIS-DON autour de l'Université P&M Curie, du Laboratoire National d'Essais, de l'Université Blaise Pascal et du CNRS.

**T**ravaux **théoriques** récents : approche statistique du vieillissement des CD par étude de la dégradation des taux d'erreurs, amélioration du décodage par application de l'algorithme de Viterbi

## Historique (2)

Les CD et DVD sont utilisés pour la sauvegarde de collections patrimoniales:

- la collection de l'INA
  - de 180 000 CD-R
  - enregistrements des émissions radiophoniques depuis 1994 (dépôt légal)
- Archives de France
- collections du Ministère de la Culture

=> forte implication du Ministère de la Culture

## 2. La problématique de l'archivage

### L'archivage à long terme est un processus complexe

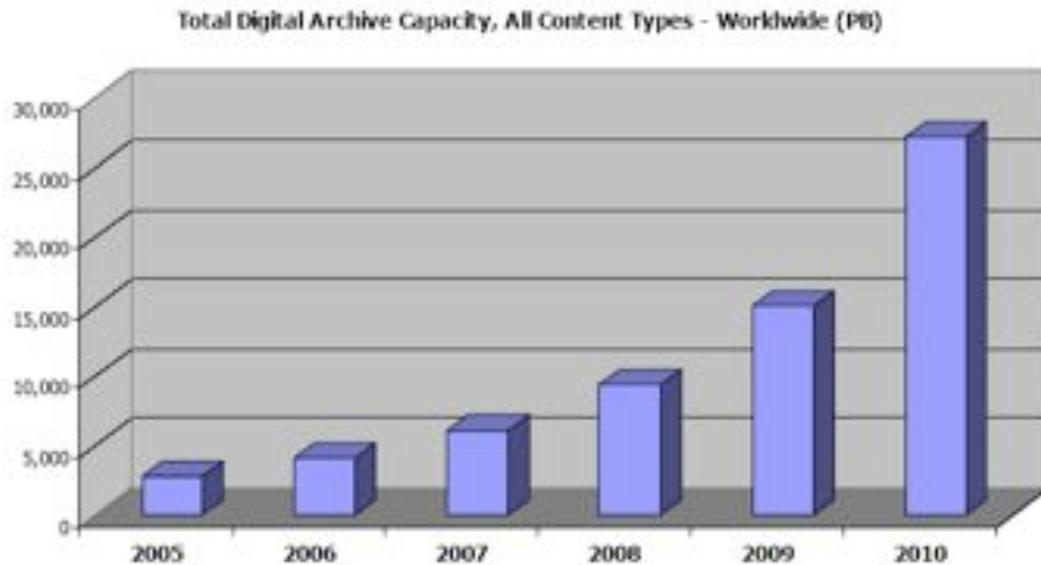


# Les stratégies d'archivage

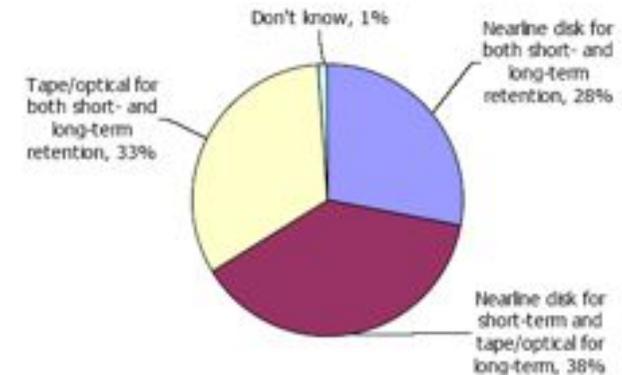
- La stratégie **passive** « Archive et Oublie », traditionnelle, la plus naturelle.
- La stratégie **active** de migration perpétuelle (BNF, CNES, banques, etc.).
- La sous-traitance à un prestataire de service (rôle de Internet); **active**.
- Le retour à l'analogique (rapport du Science and Technology Council); **passive**.

Seule la seconde est appliquée à grande échelle par des institutions dans le monde entier. Elle met généralement en jeu de gros robots (SUN, IBM) manoeuvrant des bandes magnétiques dans des cassettes; elle demande toutefois de gros moyens, en matériel, locaux, et personnel. Est-elle transposable au niveau des individus et des familles?

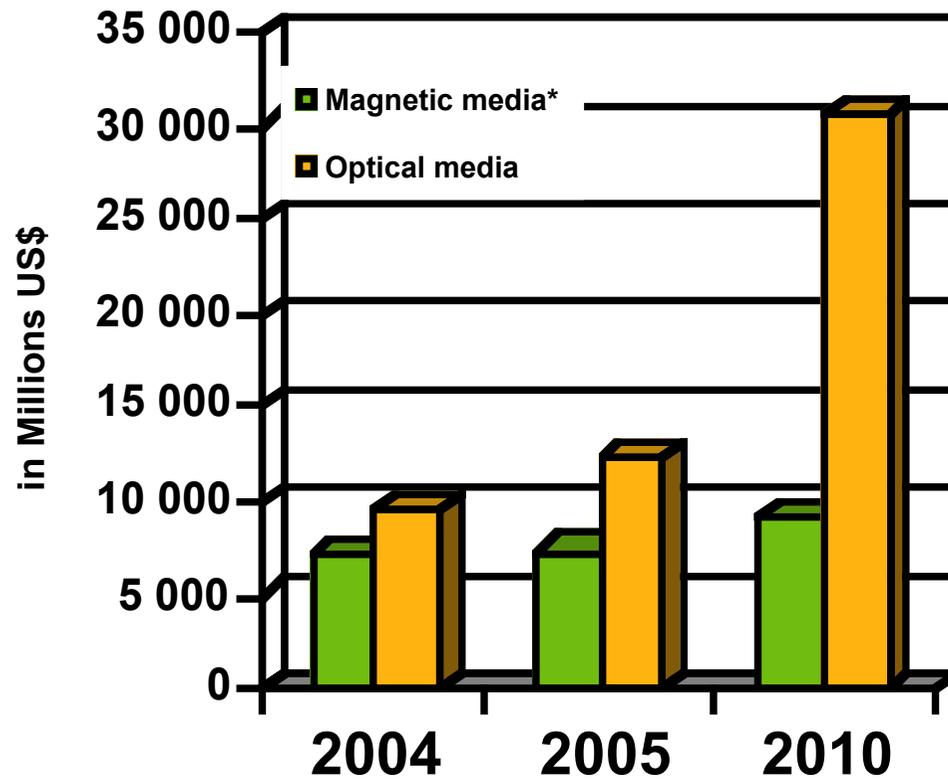
# Archivage et volume



Which of the following most closely describes your organization's current archival storage infrastructure? (Percent of respondents, N = 144)

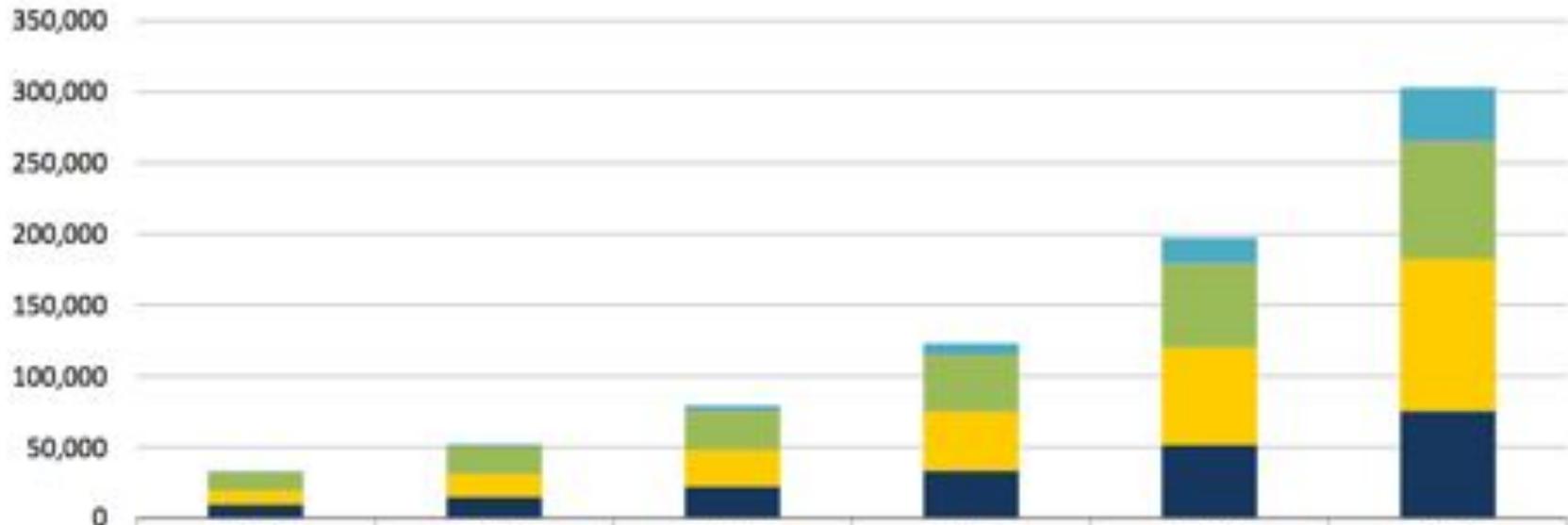


*Prédiction du volume total des archives numériques, et sa répartition par type de support*  
(Source: ESG Survey Digital Archiving: End-User Survey & Market Forecast 2006-2010 - janvier 2006)



*Evolution du marché global pour les supports magnétiques et optiques*  
 (Source: BCC Research Group 2005)

**Total Worldwide Digital Archive Capacity, by Media Type, 2010-2015 (Petabytes)**



	2010	2011	2012	2013	2014	2015
Cloud	768	1,708	3,676	8,033	17,908	37,846
Optical	303	410	537	714	968	1,247
Tape	12,784	18,939	27,116	39,362	58,220	81,562
External disk	9,712	16,053	25,637	41,501	68,430	106,830
Internal disk	9,650	14,881	22,185	33,547	51,707	75,510

*Prédiction du volume total des archives numériques, et sa répartition par type de support*  
 (Source: ESG-NERSC Report: Proving Tape as Cost-Effective and Reliable  
 Primary Data Storage - décembre 2010)

## Archival Methods

	Capacity	Rewriteable	WORM*	Data retrieval time (in library/online)	Cost per GB	Estimated life (yrs)
Spinning disk	TB	Y	Y (special array)	NA/15 ms	\$2-30	10
LTO3 tape	400 GB	Y	Y	10 min/1 min	\$0.25	30
CD-R/RW	700 MB	Y	Y	3 min/150 ms	\$0.25	3-100
DVDsR/RW	4.7 GB, single layer; 8.5 GB, dual layer	Y	Y	3 min/150 ms	\$0.125	10-100
DVD-RAM	9.4 GB	Y	N	3 min/50 ms	\$0.75	10-100
HD DVD	15 GB, single layer; 30 GB, dual layer	Y	Announced	3 min/150 ms	\$3	10-100
Blu-Ray	25 GB, single layer; 50 GB, dual layer	Y	Y	3 min/150 ms	\$0.85	10-100
Magneto optical	9.1 GB (double-sided)	Y	Y	3 min/25 ms	\$9	30
UDO	30 GB (double-sided)	Y	Y	3 min/28 ms	\$1.70	30
PDO	22 GB	Y	Y	3 min/110 ms	\$2.50	30
InPhase/Maxell holographic	300 GB	Expected 2008	Expected 2007	3 min/200 ms	\$1 (est)	30

Y=Yes, N=No \*Write-once, read-many

### *Quelques données sur les supports d'archivage*

(source: <http://www.networkcomputing.com/channels/storageandservers/showArticle.jhtml?articleID=19260060>)

# Capacité vs. coûts des supports

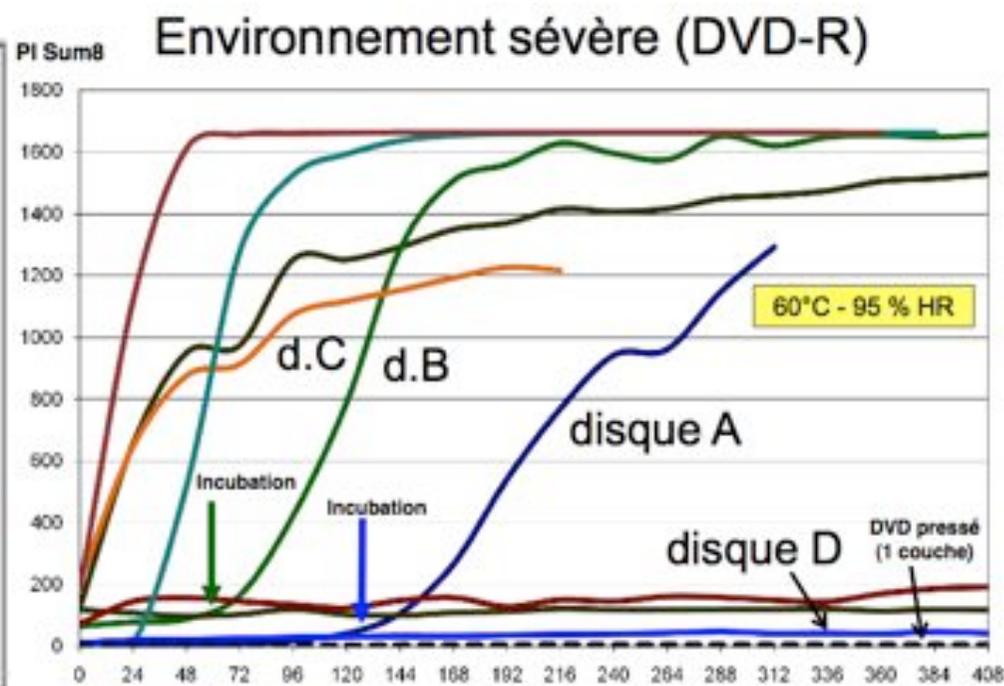
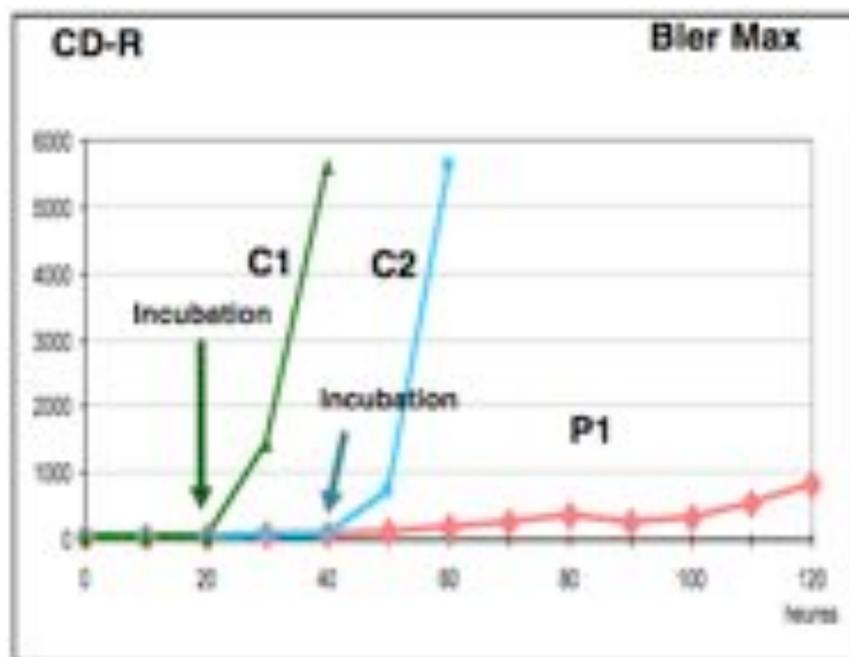
<b>SUPPORTS</b>	DON CD-R	DON DVD-R (SL)	DON BD-R (SL)	BM LTO	DD (SATA)
Capacité par support (Go)	0,7	4,7	25,0	800,0	500,0
Coût unitaire support (*) (€)	1,5	2,0	4,0	50,0	80,0
Coût support / Go (€)	2,14	0,4	0,16	0,06	0,16
Fréquence Migration (**) (An)	10	5	5	4	3
Coût annuel / Go (€)	0,2	0,08	0,032	0,015	0,05

(\*) Prix très variables selon quantité, vendeurs et marques, Mai 2010

(\*\*) Hypothèse de travail...

Si l'estimation du coût des supports peut-être tentée (avec prudence compte-tenu des paramètres très variables), la part des supports relativement au coût du système d'archivage (développement logiciel, équipement, maintenance, personnels...), ne peut guère être évaluée dans ses généralités.

# Vieillissement des disques optiques



*Mise en évidence de la latence d'incubation sur les CD et DVD*  
(Source : Jean-Marc Fontaine)

# Loi d'Arrhenius et d'Ering

On suppose que le processus de vieillissement résulte **d'un seul processus** physico-chimique (inconnu), et qui possède une énergie d'activation bien définie.

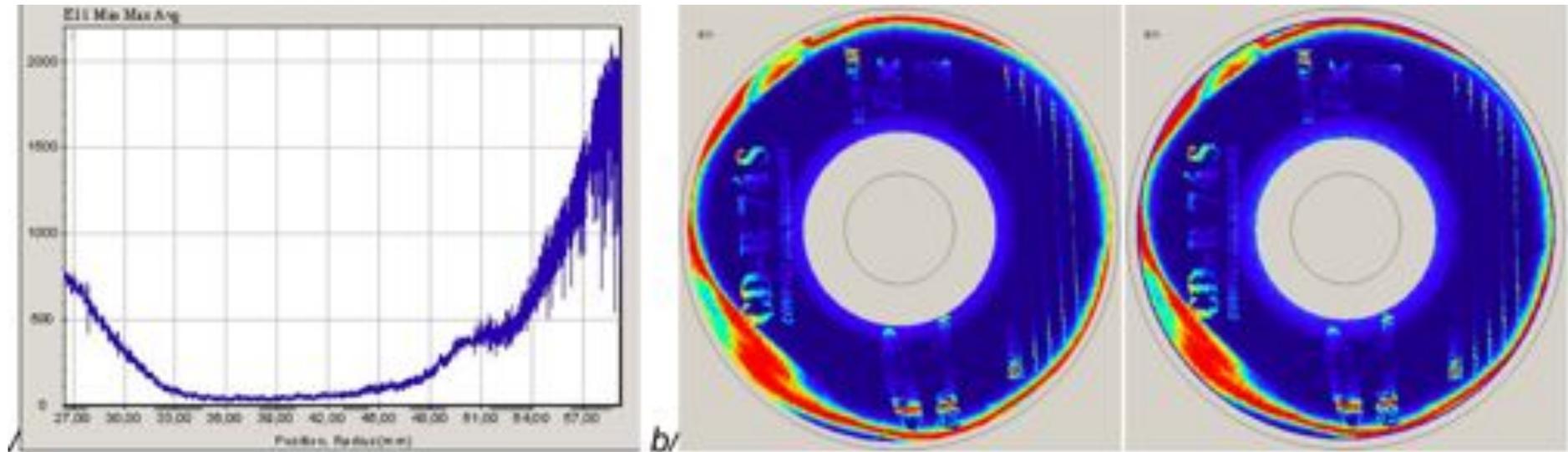
$$t = AT^a e^{\Delta H/kT} e^{(B+C/T) \cdot RH}$$

avec

$t$	Durée de vie
$A$	Constante de temps
$T^a$	Facteur pré-exponentiel de température
$\Delta H$	Energie d'activation par molécule
$k$	Constante de Boltzmann ( $1,3807 \times 10^{-23}$ J/molécule degré K)
$T$	Température (Kelvin)
$B, C$	Constantes $RH$
$RH$	Humidité relative

Vieillissement en étuve, 80 C, 85% d'humidité; norme ISO 10955

# Augmentation du nombre d'erreurs



*Taux d'erreurs corrigées (E11) d'un CD-R: a/ représentation radiale ; b/ carte de la surface (deux analyses, avec 1.3° de rotation entre elles ; taille des pixels 0.15x0.4mm<sup>2</sup> env.)*

# Un sujet “orphelin”

- **Programme non thématique 2005: *ARCHIVAGEDON***
  - Conservation des données enregistrées sur les supports optiques et numériques
  - 36 mis, 1 M€ LAM / LPMM / LNE
    - *nature physique du vieillissement des disques*
    - *notion de qualité de gravure*
    - *en vue de contribuer à l'amélioration possible de ces produits*
- **Programme blanc 2006 : *ARCHIDON***
  - Conservation des archives sur supports optiques numériques
  - 36 mois, 0,5 M€ LAM / LPMM / LNE
    - *caractérisation chimique du processus de vieillissement*
    - *caractérisation opto-mécanique des disques optiques*
    - *étude et modélisation du vieillissement: quantification de la loi d'évolution*

- **Matetpro Edition 2007 : *DONISTOS***
  - Disques optiques numériques enregistrables à durée de vie optimisée pour l'archivage
  - 36 mois, 1 M€ LAM / LNE / LPMM / CEA-LETI / SCR / CIRIMAT / Plasmon OMS
    - *développement d'un support optique de grande capacité, pérenne (>50 ans), et de coût modéré*
    - *certification de la durée de vie de ces supports*
  
- **Programme MatetPro 2008 : *HYBRIDON***
  - **Disques Optiques Numériques enregistrables Hybrides**, à durée de vie optimisée pour archivage longue durée
  - 36 mois, 1,2 M€ CIRIMAT / LPMM / CEA-LETI / Plasmon OMS / LNE / DPG / SCR
    - *développer des prototypes de DON pour l'archivage à long terme*
    - *définir des procédures de vieillissement rigoureuses (durée de vie)*
    - *impact sur le marché et les habitudes de stockage*

- **Programme BLANC Edition 2008 :      *ESPADON***
  - Espérance de vie des disques optiques numériques enregistrables
  - 36 mois, 0,4 M€    LAM / LNE / LPMM
    - *étude de la latence d'incubation des DON*
    - *en trouver des précurseurs et les relier à l'état chimique des DON*
    - *prédire la durée de vie des DON de manière fiable*
  
- **Programme BLANC Edition 2009 :      *ESPADON***
  - Espérance de vie des disques optiques numériques enregistrables
  - 36 mois, 0,15 M€  LAM / LNE / LPMM
    - *comprendre le phénomène de vieillissement*
    - *établir des lois d'évolution basées sur les processus chimiques*
    - *développer des procédures de vieillissement accéléré*
    - *proposer leur normalisation internationale*

## et le développement durable ?



ce centre de données, d'une puissance de 14 MW, engloutit chaque année l'équivalent de la consommation d'une ville de 20 000 à 50 000 habitants

*in* Les centres de données informatiques avalent des quantités croissantes d'énergie  
LE MONDE | 01.07.2013

## 2 % de l'énergie mondiale

- aucune statistique officielle
- A l'échelle européenne,
  - 56 milliards de kilowatts estimés en 2008,
  - 104 milliards en 2020.
- Greenpeace :
  - les centres de données consomment 2 % de l'énergie mondiale
  - Ils entraînent de larges émissions de CO2
  - dans de nombreux pays, l'électricité vient du charbon et du
  - épingle Microsoft, Amazon et Apple dans un rapport de 2012

### 3. Quelques résultats du GIS SPADON

- Etudes de vieillissement
- Rapport PSN des académies
- Caractérisation chimique du vieillissement

# Etude de vieillissement accéléré : une étude du LNE

- Constat : les disques en verre Century sont réputés très résistants de par leur conception mais il n'y a pratiquement pas eu de validation.
- Etudier le comportement des disques en verre Century dans des conditions sévères (vieillissement accéléré)
- Les comparer aux autres disques réputés pour leur stabilité dans le temps

- Échantillonnage:
  - CD :
    - 2 CD Century Or (1994)
    - 4 CD-R :Delkin, MPO, Taiyo yuden, Sony
    - 5 CD pressés : CD audio achetés dans le commerce
  
  - DVD:
    - 4 DVD Century Argent (2010)
    - 5 DVDR : MPO +R, Verbatim-R, Verbatim archival -R, FUJI -R, Maxell-R
    - 10 DVD pressés : 10 DVD Vidéo - 5 DVD 9 - 4DVD 5 et 1 DVD 10.



## Étuve de vieillissement LNE

# Descriptif de l'étude

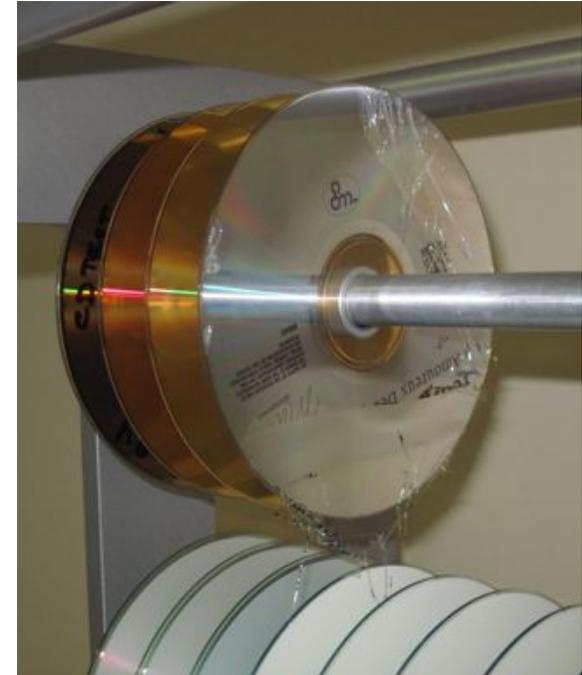
- Conditions climatiques:
  - Température : 90°C
  - Humidité relative : 85%
- Durée d'étuvage:
  - Par période de 125 heures
  - Total actuel : 250 heures



# Résultats à 250 Heures

- CD

REF LNE	Marque	T0	125H	250H
CEN 1	CENTURY	52,5	75,5	76,0
CEN2	CENTURY	42,8	49,5	48,5
CD1	BRASSENS	4,8	Fail	
CD2	PUCCINI	30,4	52,2	996,2
CD3	ZOUK n°1	9,3	8,0	21,8
CD4	CELTIC	8,7	9,4	10,1
CD5	SALSA	2,5	2,6	2,7
CD6	MOUSTAKI	19,7	1046,0	
162-1	SONY	5,5	19,7	23,4
163-1	Taiyo	2,7	10,5	68,9
170-1	MPO	4,9	9,0	5,2
177-1	DELKIN	5,2	5,4	10,1



- 4 CD sont fortement dégradés
- Le BLER des Century a légèrement augmenté à 125 H puis est stable à 250 H.

# Résultats à 250 Heures

- DVD

REF LNE	Marque	T0	125H	250H
576-1	MPO	2	4,8	165
577-1	VERBATIM	6,8	162,4	429,3
582-1	MAXELL	1,9	40,1	
585-1	FUJIFILM	2,1		
587-1	VERBATIM archival	3	34,2	
DVD1	JURASSIC PARK	15,8	297	
DVD2	LA BEUZE	4,4	8,8	42,6
DVD3	TONIGHT	216	307	
DVD4	BEETHOVEN	33,6	413	598
DVD5	ARNAQUE 2	10,5		
DVD6	BEAN 16/9 NOIR	27,9	109	
DVD7	VIVA MARIA	1,8	718	
DVD 10	VOLVER A EMPEZAR	4	354	1194
DVD13	SPINALTAP	14,8	16,3	316
DVD16	SLEEPWALKER	19	201	1486
CEN 3	CENTURY	3,7	5,2	4,5
CEN 4	CENTURY	5,2	7,3	6,1
CEN 5	CENTURY	5,6	8,9	10
CEN 6	CENTURY	2,7	4	4

- Après 250 heures, il reste 3 DVD en bon état : 1 DVD vidéo, 1DVDR et le CENTURY (qui a le taux d'erreur le plus faible)

# Rapport PSN des académies

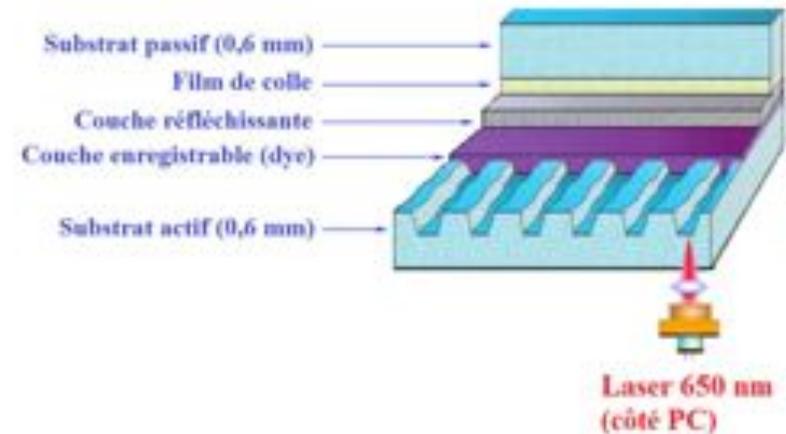
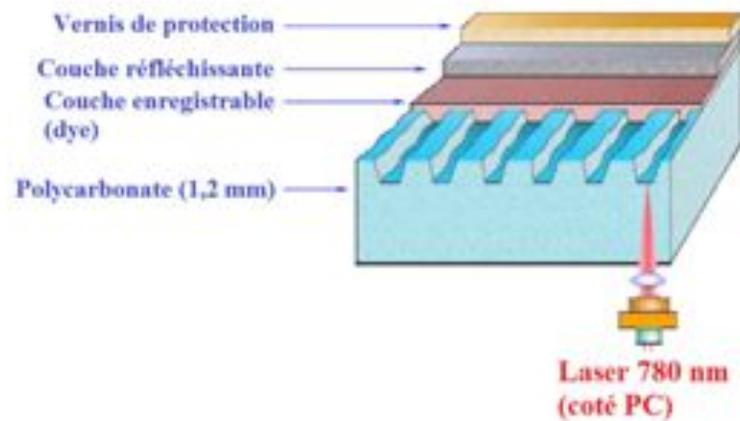
- Groupe de travail commun aux deux Académies, des Sciences et des Technologies
- Membres du groupe:
  - Erich Spitz, Académie des Sciences et Académie des Technologies, président
  - Jean-Charles Hourcade, Académie des Technologies
  - Franck Laloë, LKB/ENS, rapporteur
- 16 auditions:
  - experts, laboratoires, industriels, utilisateurs
  - De septembre 2008 à juin 2009
- Publication : "Longévité de l'information numérique - Les données que nous voulons garder vont-elles s'effacer ?" (2010)

# Recommandations du rapport PSN

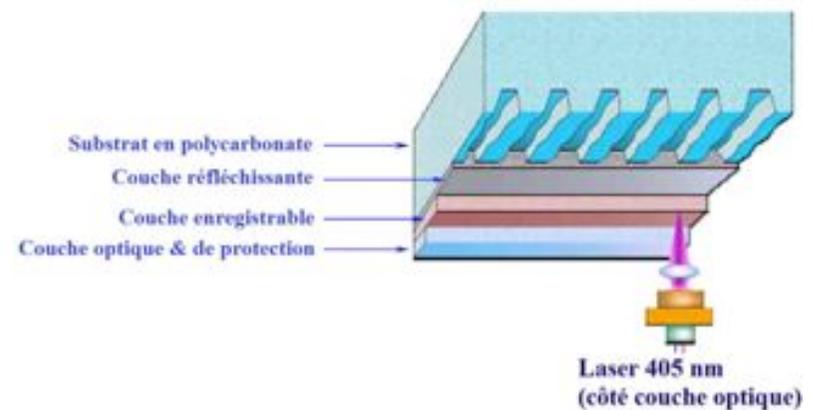
Un message d'alerte au grand public, aux établissements et à l'administration, doit être lancé. Le problème n'est pas spécifique à la France, mais mondial. Une action au niveau européen, ou dans le cadre de l'UNESCO, semblerait souhaitable, afin de faire prendre conscience de l'urgence d'une politique concertée dans ce domaine. A l'échelon national et européen, nous proposons quatre recommandations:

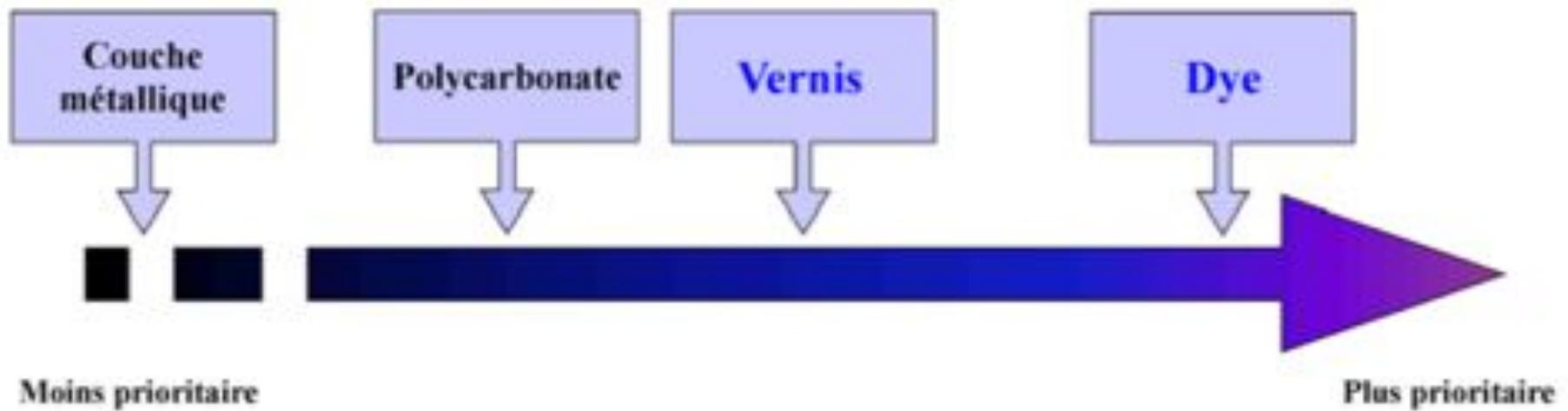
- 1. Débloquent les études sur le sujet.**
- 2. Éviter la perte des compétences dans le privé et le public**
- 3. Favoriser l'innovation et l'apparition d'une offre industrielle de qualité**
- 4. Élaborer une véritable politique d'archivage numérique**

# Caractérisation chimique du vieillissement

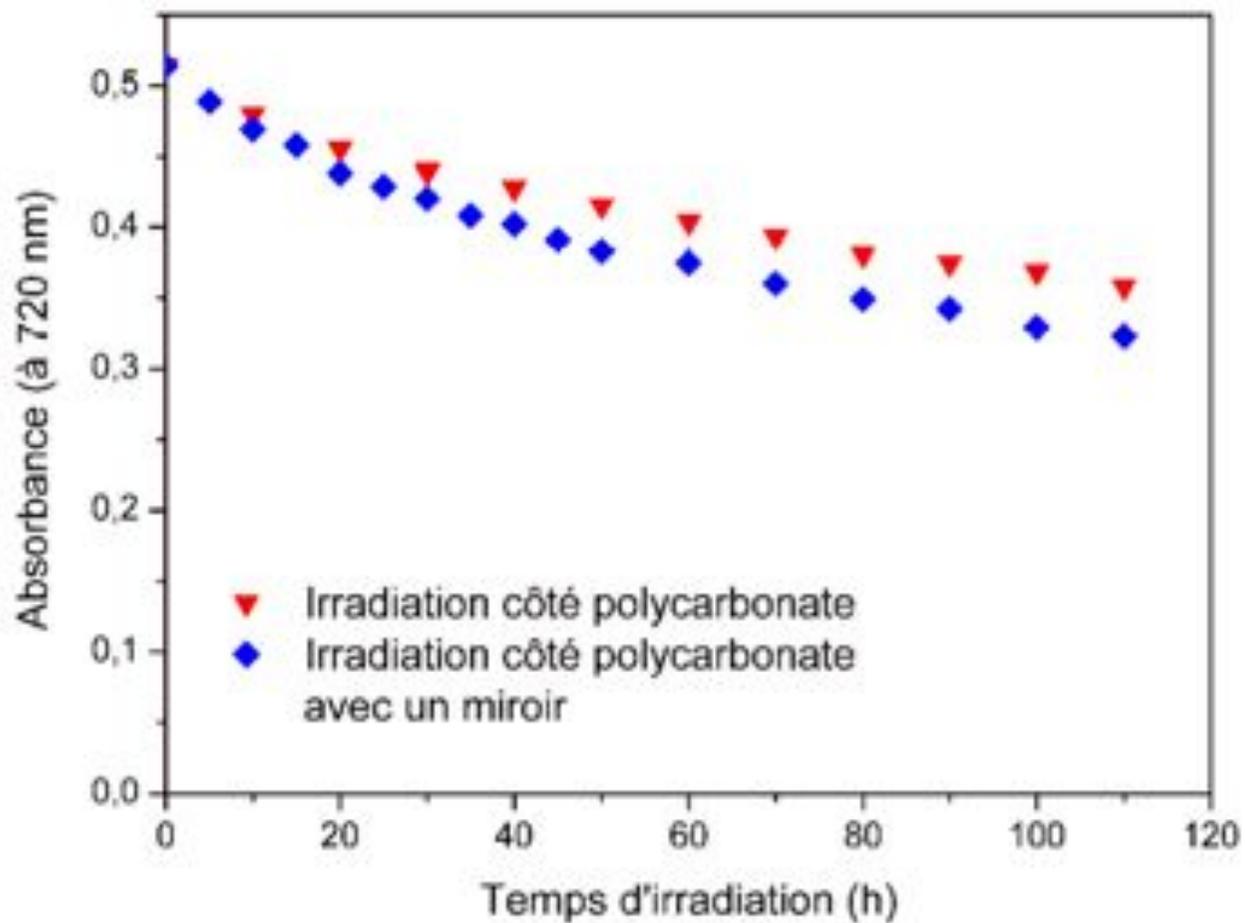


Structure des CD-R,  
DVD-R, et BD-R  
(These S. Collin, 2013)





*priorité des études de photovieillissement des différents constituants de disques optiques*  
(thèse de S. Collin, UBP 2013)



*Variation de l'absorbance à 720 nm du dye azoïque par photovieillissement*  
(thèse de S. Collin, 2013)

## 4. perspectives

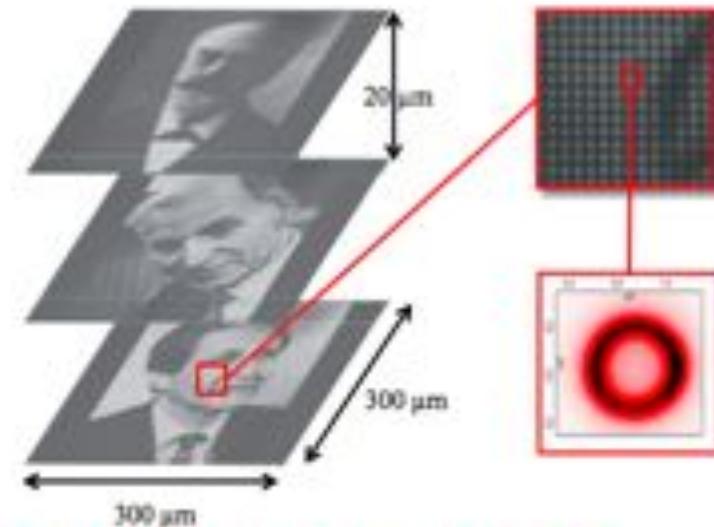
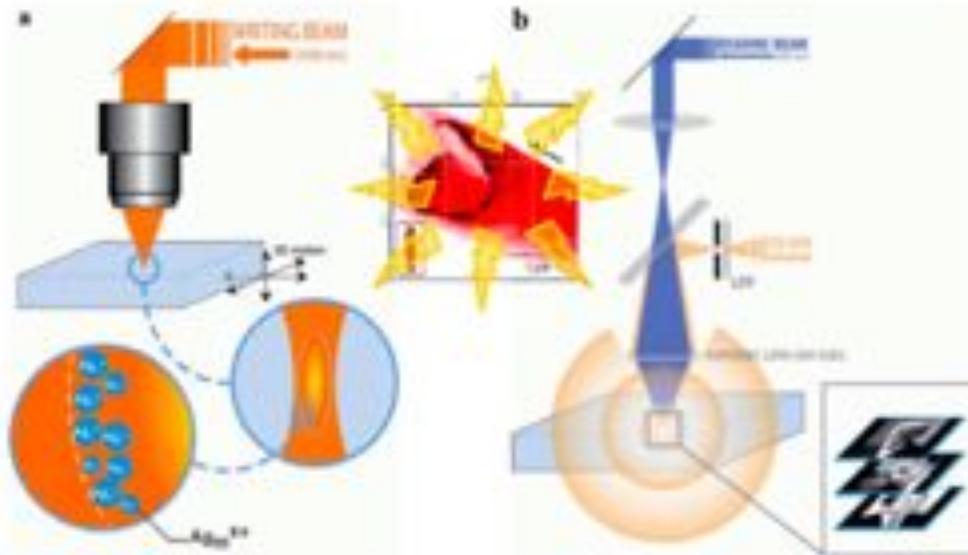
- A court terme:
  - CD-DVD en verre trempé (Century disc, Syylex)



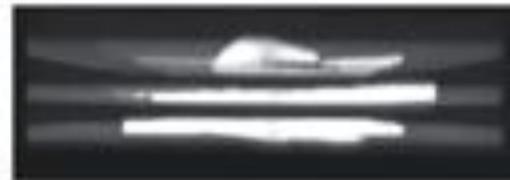
Vieillessement accéléré 90°C 85% HR		Durée de vie (heures)				
Marque	Modèle	0	250	500	750	1000
MPO	DVD+R Gold	[Bar chart showing failure at 250h]				
Verbatim	DVD-R	[Bar chart showing failure at < 250h]				
Maxell	DVD+R	[Bar chart showing failure at < 250h]				
Verbatim	DVD-R Archival	[Bar chart showing failure at < 250h]				
Syylex	DVD GlassMasterDisc	[Bar chart showing failure at > 1000h]				
Millenniata	DVD+R M-Disc	[Bar chart showing failure at < 250h]				
Northern Star	DVD+R Data tesor disc	[Bar chart showing failure at 250h]				

## 4. perspectives

- A court terme:
  - CD-DVD en verre
- A long terme:
  - Gravure dans le verre (Bordeaux, Lyon)
  - Mémoire holographique (Institut d'optique)



Royon *et al.*, *Adv. Mat.* **22**, 5282-5286 (2010).



- Pas de diaphonie entre couches.
- Pas de photo-blanchiment.
- Capacité de stockage élevée.
- Lecture Blu-ray.
- Vitesse de lecture max : 500 Mbits.s<sup>-1</sup>.

• Images d'origine JPEG :

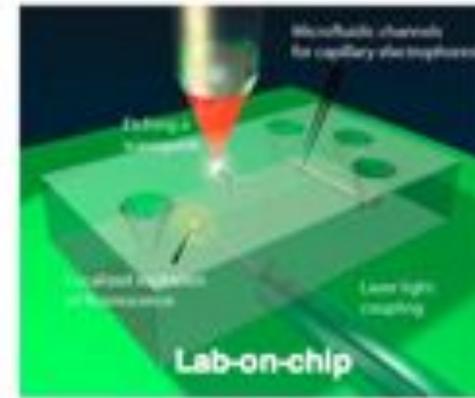
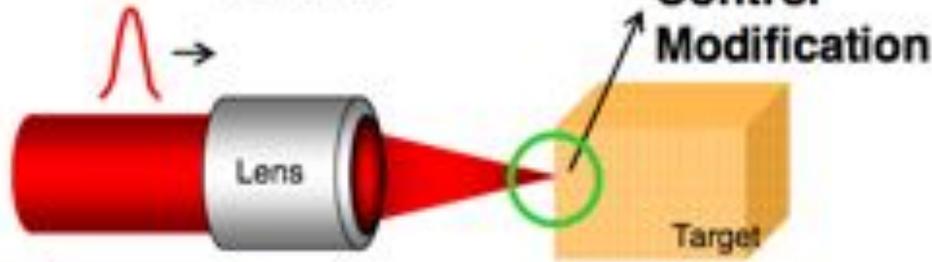
- Taille : 100 × 100 pixels.
- Dynamique d'encodage : 8 bits (256 niveaux).
- Gabriel Lippmann (procédé photographique en couleur, 1908).
- Alfred Kastler (pompage optique, 1966).
- Claude Cohen-Tannoudji (refroidissement laser, 1997).

• Images inscrites dans le verre :

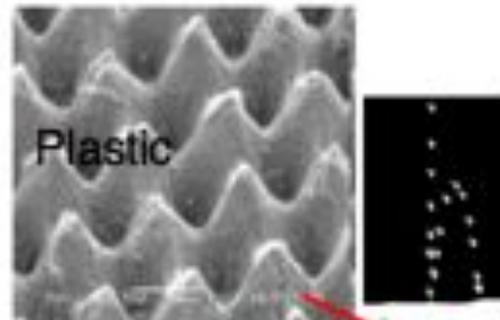
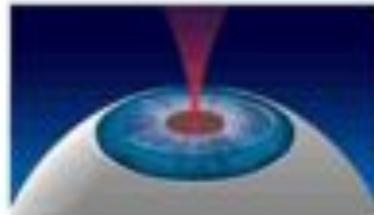
- Taille : 100 × 100 pixels (300 μm × 300 μm).
- Ecart entre pixels : 3 μm. Diamètre pixel : 2 μm.
- Ecart entre couches : 20 μm.
- Dynamique effective : 4 bits (16 niveaux de gris).
- Capacité de stockage : 20 Gbits.cm<sup>-3</sup>.

# Leitmotiv

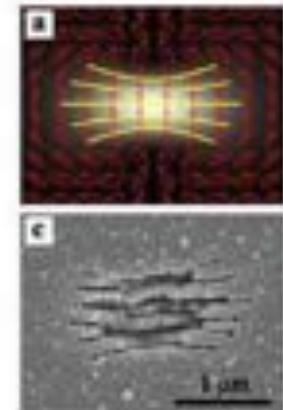
Control  $E(x,y,z,t)$



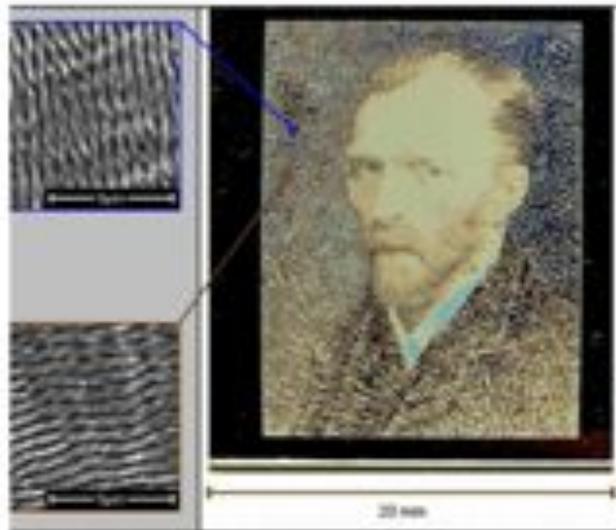
Neite et al. App. Phys. A 77 (2000)



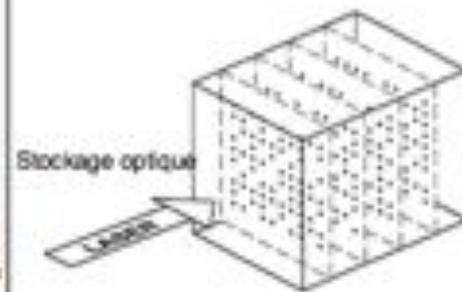
Plastic  
Groenendijk M. Laser Technik Journal, 5 (2008)



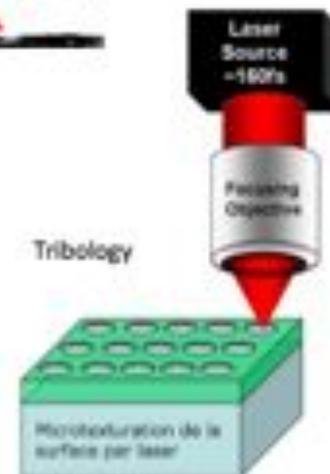
Hnatovsky et al. Phys. Rev. Lett. 106 (2011)



Dusser et al. Opt. Express (2010)

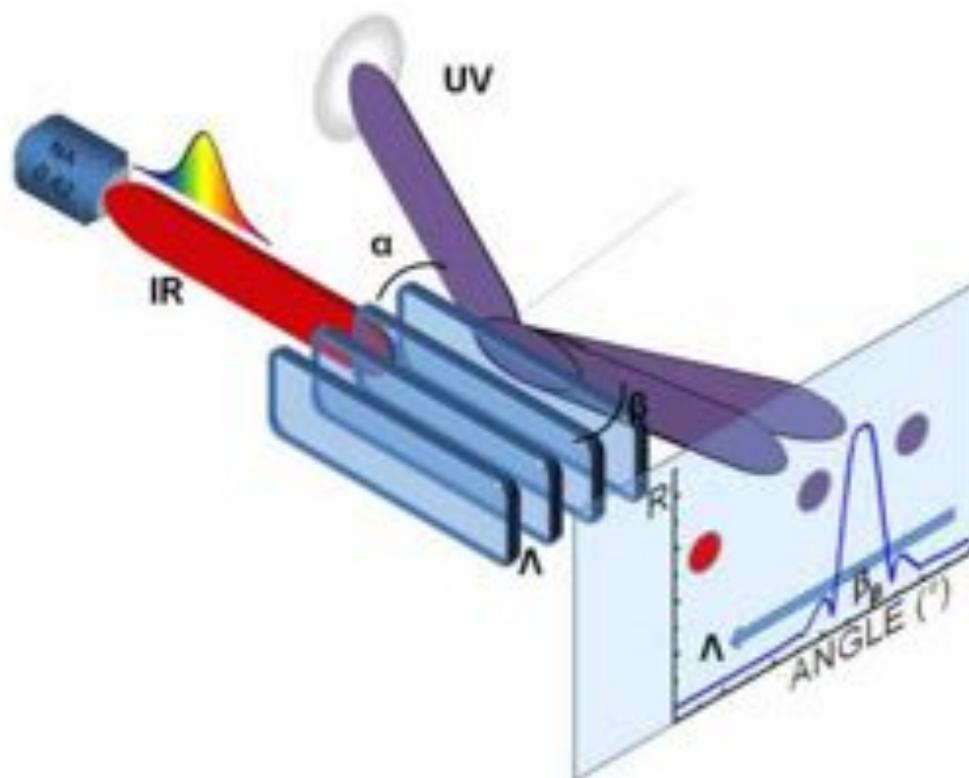
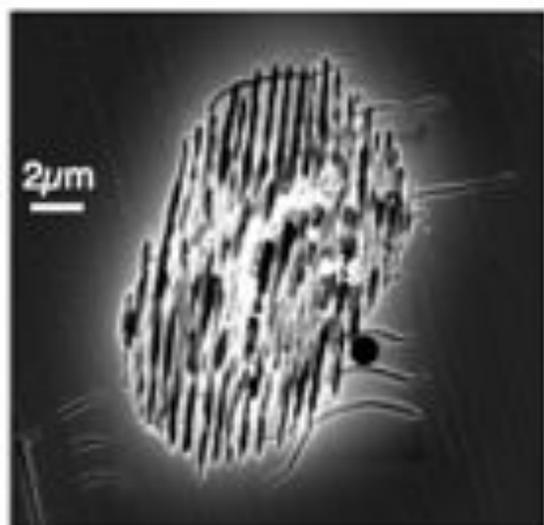


Glezer et al. Opt. Lett. Vol.21 (1996)



- Temporal beam shaping

Control of nanostructuration in fused silica



# Conclusions

- Il existe des solutions à court terme:
  - CD-DVD en verre
  - Sélection des CD-DVD (INA, LNE)
- A plus long terme:
  - soutenir la recherche sur les supports
- Étudier les mémoires FLASH

# CD-R, DVD-R, BD-R

## Modèles pour une conservation de longue durée

Type de support	Anné de l'étude	Modèles recommandés	Code d'identification du fabricant	Vitesse de gravure	
				Maximale	Recommandée
CD-R	2008	MPO Gold White Inkjet	97 25 07 (MPO)	32x	16x
		DELKIN Archival Gold	97 27 58 (Mitsui Chemicals)	52x	16x
DVD-R	2012	FTI Gold Archival et Platinum Archival	TTG02 (TDK)	8x	8x
		VERBATIM "standard"	MCC 03RG20 (Mitsubishi)	16x	8x
		JVC Archival Grade	TYG03 (Taiyo Yuden)	16x	8x
BD-R (HTL)	2012	PANASONIC	MEI T02 (001)	4x	4x
		SONY	SONY NN3 (002)	6x	6x

Merci pour votre attention