

## Experiència ICC en la preservació de dades digitals.

..O com no morir en l'intent de guardar-ho tot, per sempre mes..

Sergi Racero  
Unitat de Tecnologies de la informació  
Gener 2014



# Índex

1. Dades. Visió Global. Escenari Exponencial.
2. Dades. Experiència general ICC.
3. Preservació a l'ICC. Reptes i oportunitats.

# 1. Dades. Visió Global. Escenari Exponencial.

- 2.5 Exabytes es creen diàriament (2.5 Milions de Terabytes). Aquest número es dobla cada mes.. **ESCALA DE BYTES ,KB ,MB,GB,TB,PB,EB,ZB,YB Y XB**
- 90% de les dades de tot el món, van ser creades als darrers 2 anys.
- Ens enfrontem a la triple “v” :
  1. Volum (32Gb per Smartphone, 8Gb per USB, 500Gb per PC/laptop, 9.57 Zettabytes tot els servidors al 2012..)
  2. Varietat (estructurades, no estructurades, IoT sensors, logs dataset). *Gener2014: Google compra Nets !!*
  3. Velocitat (d'ingesta, de procés, de transformació, d'accés, de descarrega, visualització..)

source: <http://www.ni.com/newsletter/51649/es/>

<http://www.businessinsider.com/infographic-how-big-would-all-the-worlds-data-be-2012-8>

[http://visual.ly/big-data-bigger-opportunities?qback=.qde\\_4332669\\_member\\_276002628#!](http://visual.ly/big-data-bigger-opportunities?qback=.qde_4332669_member_276002628#!)

## 1. Dades. Visió Global. Escenari Exponencial.

### ***ILM o el desafiament de les dades !!***



- Només es compleixen 3 dels 4 estats possibles definits per “cucal”: Les dades *neixen, creixen i es reproduïxen ..però no moren (ni desapareixen) !!*

## 2. Experiència general ICC.

### Trets Generals :

- Mes de 30 anys emmagatzemant dades (1982-2014).
- Migració permanent del suport (evolució del media).
- Estratificació / Estructuració de les dades, en funció d'una serie de criteris:
  - Tipus d'accés, nivells de criticitats.
  - Comercialització de les dades (gran volum de dades - HSM. no caben en un disc i recuperació desatesa).
  - Organització interna d'usuaris/projectes (cintotecari amb intervenció humana) i Servidors de fitxers.
- Des de finals de l'any 2004 (primera càmera digital), entrem en un procés de digitalització general.

## 2. Experiència general ICC.

### Objectiu ICC\_TI envers l'ILM :

- Definició ILM de l'SNIA :

*"[Data Management] The policies, processes, practices, services and tools used to align the business value of information with the most appropriate and cost-effective infrastructure from the time information is created through its final disposition.*

*Information is aligned with business requirements through management policies and service levels associated with applications, metadata and data."*

- Presència de l'ILM, en diferents marcs de definició formals de funcions/programes d'actuació de l'ICC (seguretat de la informació):

*"Preservació de les dades digitals de l'ICC. Establiment de les tècniques, metodologies i sistemes adequats per garantir l'accés, conservació, tractament i difusió de la informació emmagatzemada a l' Institut"*

source: [http://www.snia.org/education/dictionary/i#information\\_lifecycle\\_management](http://www.snia.org/education/dictionary/i#information_lifecycle_management)

## 2. Experiència general ICC.

### Distribució de les dades a l'ICC (finals 2013)

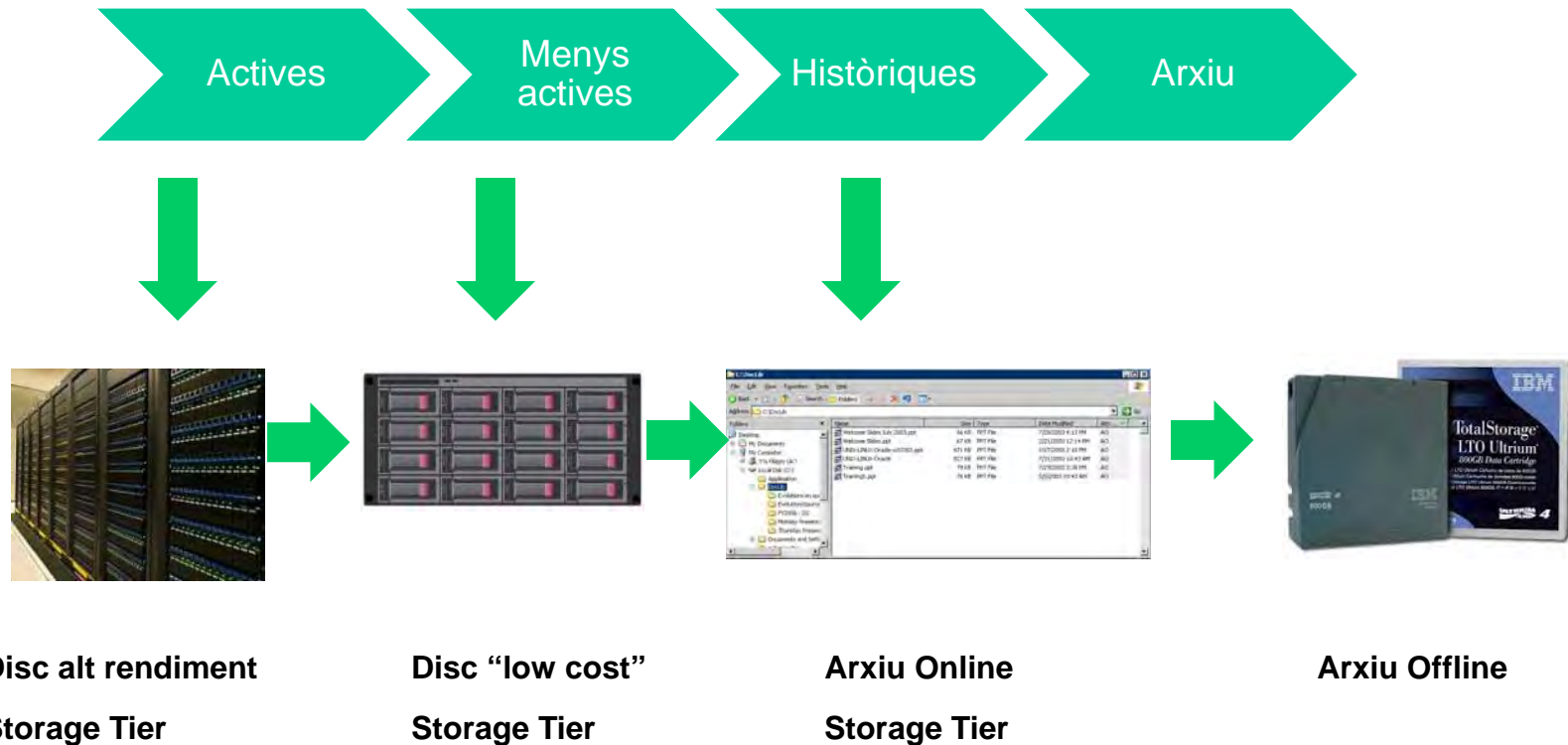
- Dades online: 350TB - Storage Array Disk Systems
- Dades near-online: 421TB (x2) - Hierachy Storage System.
- Dades offline: 1108TB (x2) - Cintotecari ICC (Tape catalogued)

**Total Dades ICC 2013: 1,5 Petabytes (+- aproximat).**

**Total capacitat ocupació d'emmagatzematge (inclou dobles còpies i Backup): 4,16 Petabytes**

## 2. Experiència general ICC

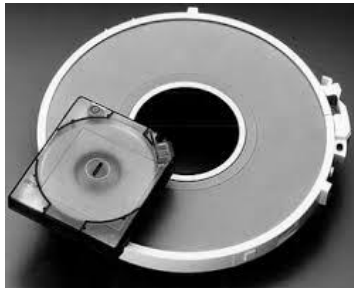
### Cicle de vida de les Dades (ILM)





## 2. Experiència general ICC

### Evolució del suport magnètic a l'ICC



Migració contant de les dades (aprox. cada 4-5 anys)

Actualment estem al voltant de les 1700 cintes x 2 (700GB ocupació mitja)

*En l'actualitat suporten més del 90% de l'informació emmagatzemada (+- 3,81 Petabytes) !!*

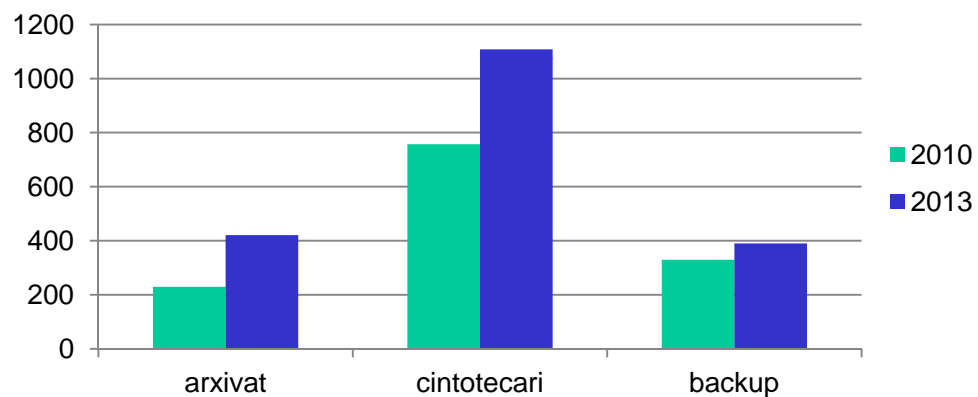
## 2. Experiència general ICC

### Evolució del creixement de les dades (període 2010-2013)

- **Any 2010** *TOTAL ESPAI EMMAGATZEMATGE ICC 2010 : 2,43 Petabytes*  
Espai en disc: 300 TBs  
Espai en cinta: 2304 TBs (Arxivat: 230 TB(x2), Cintotecari: 757 TB(x2), Backup/DR: 330TB)
  
- **Any 2013** *TOTAL ESPAI EMMAGATZEMATGE ICC 2013 : 4,16 Petabytes (+ 71,1% respecte al 2010)*  
Espai en disc: 350 TBs (+ 16.6 % creixement)  
Espai en cinta: 3810 TBs (+ 65.6 % creixement) (Arxivat: 421TB (x2), Cintotecari: 1108 TBs (x2), Backup/DR: 390TB)

## 2. Experiència general ICC

### Increment dades emmagatzemades a cinta (període 2010-2013)



-Arxivat + 83% creixement respecte al 2010

-Cintotecari + 46.3% creixement respecte al 2010

-Backup + 18% increment respecte al 2010

30 TBs setmanal (x4 x2 cicles setmanals) + (x 4 cicles mensuals) = Total 390TB any 2013

### **3. Preservació a l'ICC. Reptes i oportunitats.**

- ❑ Dades externes. Adequació i adaptació (a l'entorn web i a l'accés a les dades) al moment tecnològic.
  - ❑ Disminuint el temps de resposta i l'accés a l'informació útil, "viva" (HD més barat i més lent que la RAM).
  - ❑ Dades que generem vs dades que ens generen: Obtenir informació vàlida per al negoci.
  - ❑ Les dades han de ser "mòbils" (accés permanent, des de qualsevol lloc, amb qualsevol dispositiu i amb varietat de protocols)
  - ❑ Les dades han de ser "socials" (visibilitat, accés, generar serveis/productes, rebre feedback)
  - ❑ Les dades han de ser/estar al "cloud" (independents de d'infraestructura, del medi físic)
  
- ❑ Dades Internes. Definició de metadades que permetin la indexació, cerca i recuperació ràpida de la informació utilitzant sistemes avançats de fitxers (imatges vs document vs wikiportal)
  
- ❑ Protecció de dades ( NoBackup, DR Plan : RTO i RPO)

## Experiència ICC en la preservació de dades digitals.

Preguntes, dubtes, comentaris .. ?

Moltes gràcies.

### Institut Cartogràfic de Catalunya

Parc de Montjuïc,  
E-08038 Barcelona

41°22'12" N, 2°09'20" E (ETRS89)

 [www.icc.cat](http://www.icc.cat)

 [sergi.racero@icc.cat](mailto:sergi.racero@icc.cat)

 [@icc\\_cat](https://twitter.com/icc_cat)

 [icc.cat](https://www.facebook.com/icc.cat)

Tel. (+34) 93 567 15 00

Fax (+34) 93 567 15 67

