

I si deixem de parlar de preservació digital?

**Variacions sobre un tema de David Rosenthal,
amb un poutpourri final amb la CDLib Orchestra**



<http://www.flickr.com/photos/rickheath/72041533/>

Ferran Jorba
Universitat Autònoma de Barcelona

2n Espai CBUC d'intercanvi de coneixements i experiències
Barcelona, 30 de juny de 2010

Spring CNI Plenary: The Video



Encara no l'heu vist?

<http://blog.dshr.org/2009/07/spring-cni-plenary-video.html>

Començarem pels crèdits, per variar

- David Rosenthal
 - *Spring CNI Plenary: The Video (How are we ensuring the longevity of digital documents?)* <http://blog.dshr.org/2009/07/spring-cni-plenary-video.html>
 - *Spring CNI Plenary: The Remix*, <http://blog.dshr.org/2009/04/spring-cni-plenary-remix.html>
 - *Sheila Morrissey's comment*, <http://blog.dshr.org/2009/05/sheila-morrisseys-comment.html>
- Chris Rusbridge, "*Digital Preservation*" term considered harmful? <http://digitalcuration.blogspot.com/2008/07/digital-preservation-term-considered.html>
- Stephen Abrams, Patricia Cruse, John Kunze, *Preservation Is Not a Place*, International Journal of Digital Curation, Vol 4, No 1 (2009) <http://ijdc.net/index.php/ijdc/article/view/98>
- California Digital Library, *Curation Micro-Services*, <http://www.cdlib.org/services/uc3/curation/>

"Digital preservation" term considered harmful?

The outcome of successful digital preservation is that digital resources remain *accessible and usable over the long term*. This outcome-related notion is immediately understandable to users and decision makers [...].

In contrast, digital preservation has been *over-sold as difficult, complex and expensive* over the long term, while the term itself contains no notion of its own value.

"long term accessibility" or "usability over time"

<http://digitalcuration.blogspot.com/2008/07/digital-preservation-term-considered.html>



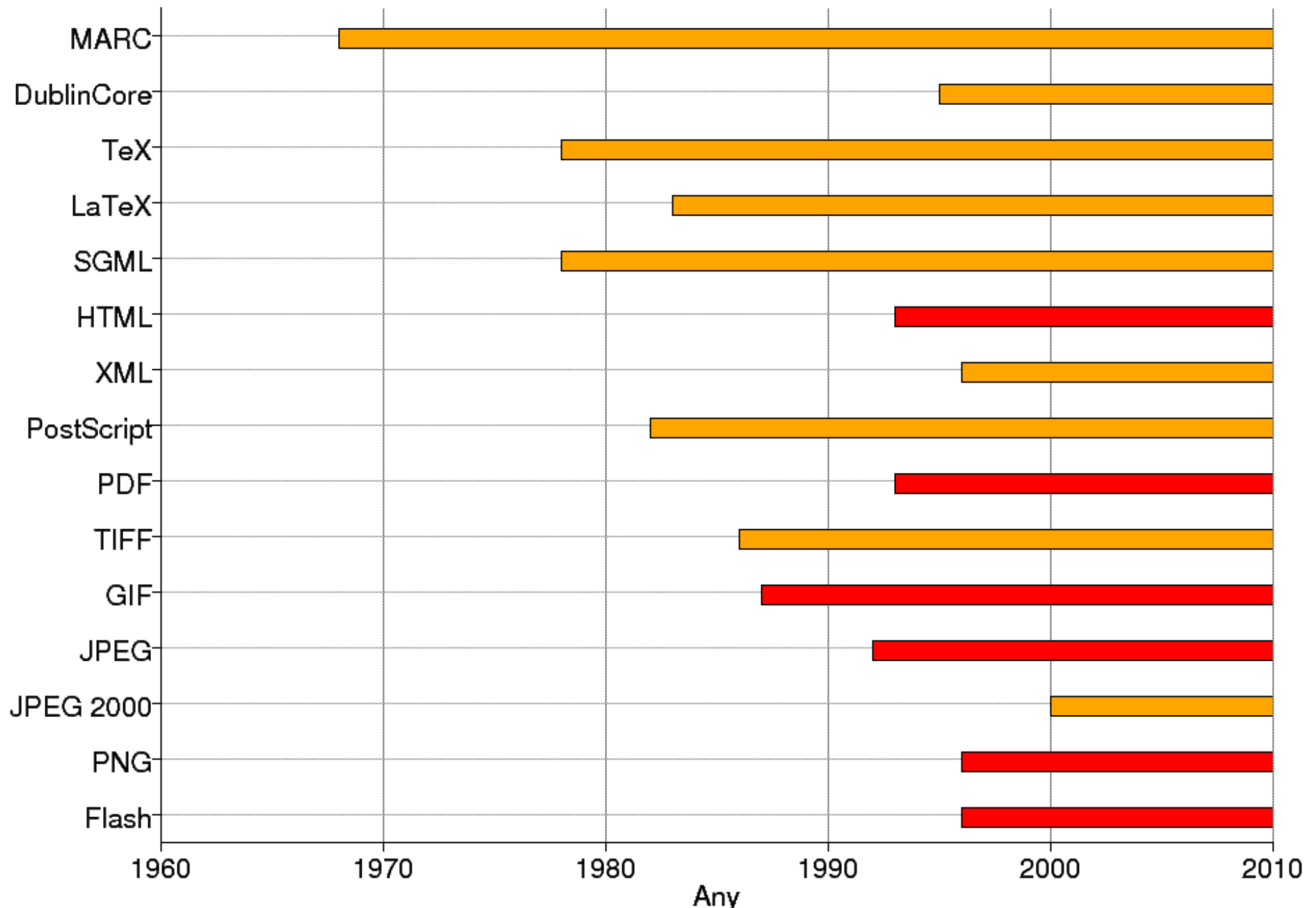
Obsolescència dels formats?

- Un mite exagerat provocat per les incompatibilitats gratuïtes en algunes aplicacions *ofimàtiques* per forçar actualitzacions *i pagar*.
- S'havia donat en un entorn pre-Internet (*Incunables digitals*, segons Clifford Lynch).
- Qualsevol format llegible amb software lliure té garantida la seva accessibilitat.
- Especialment els *formats Web*.

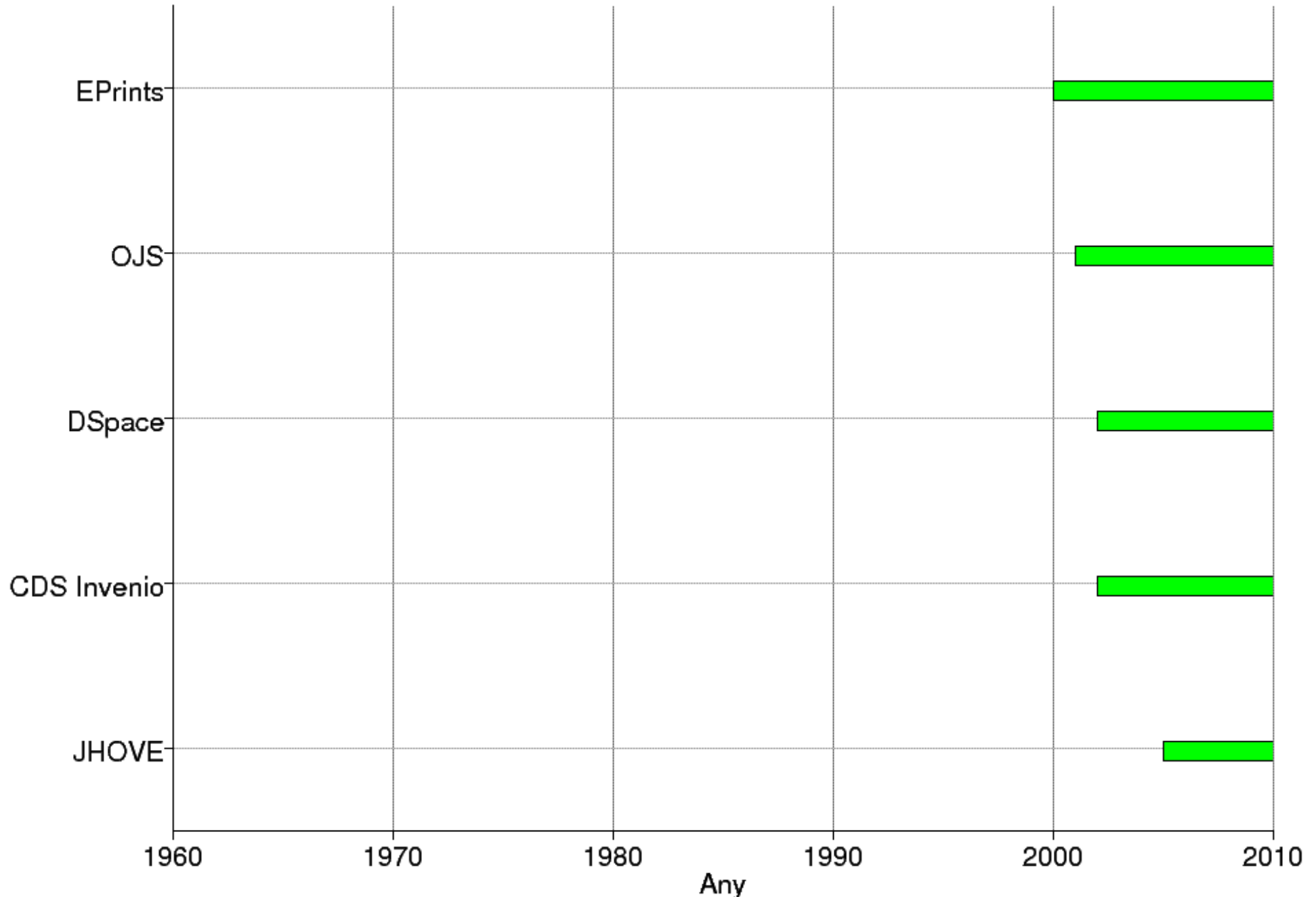


<http://desborregando.wordpress.com/2008/02/07/canon-mentira-y-cintas-de-video/>

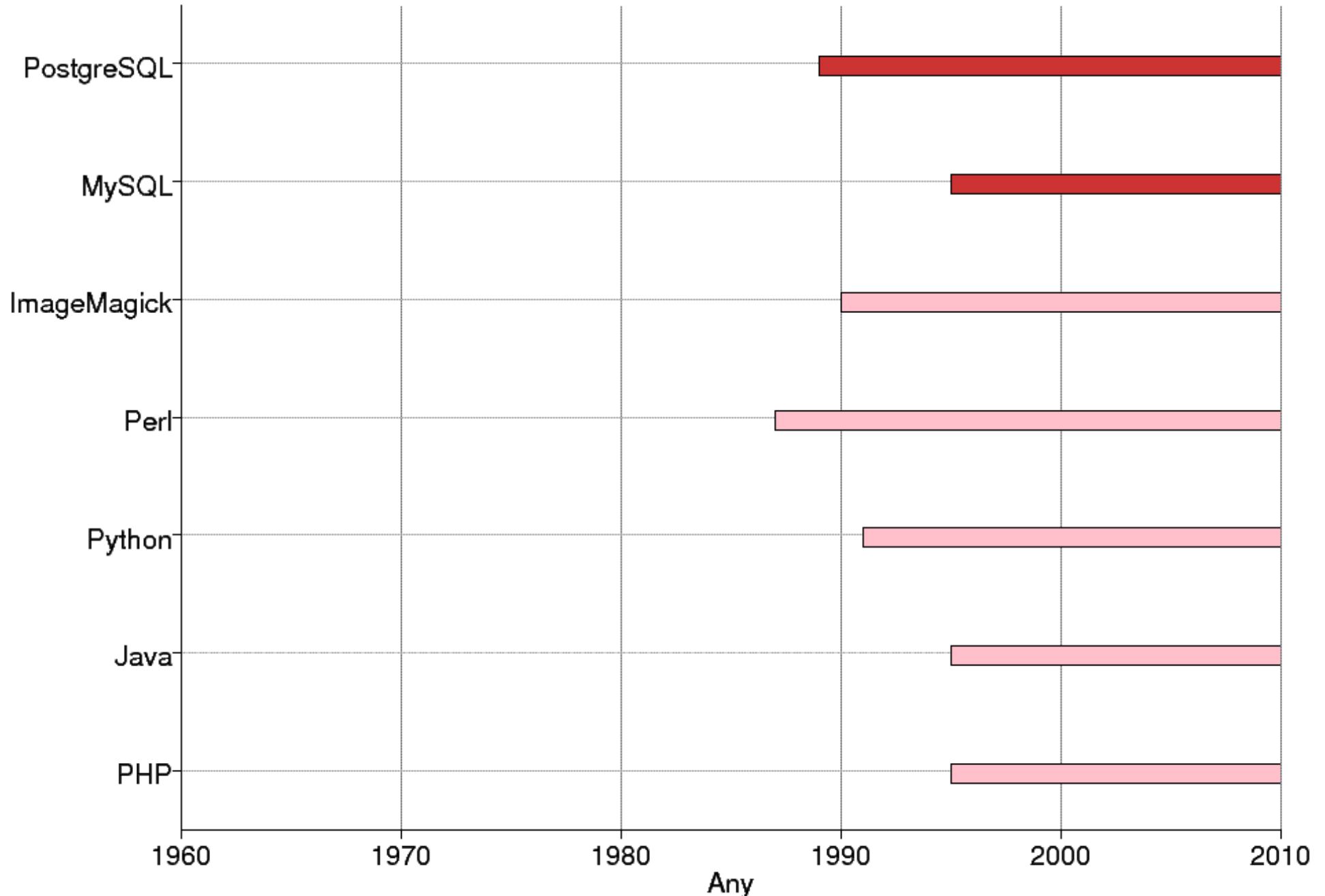
Cronologia d'alguns formats (1/5)



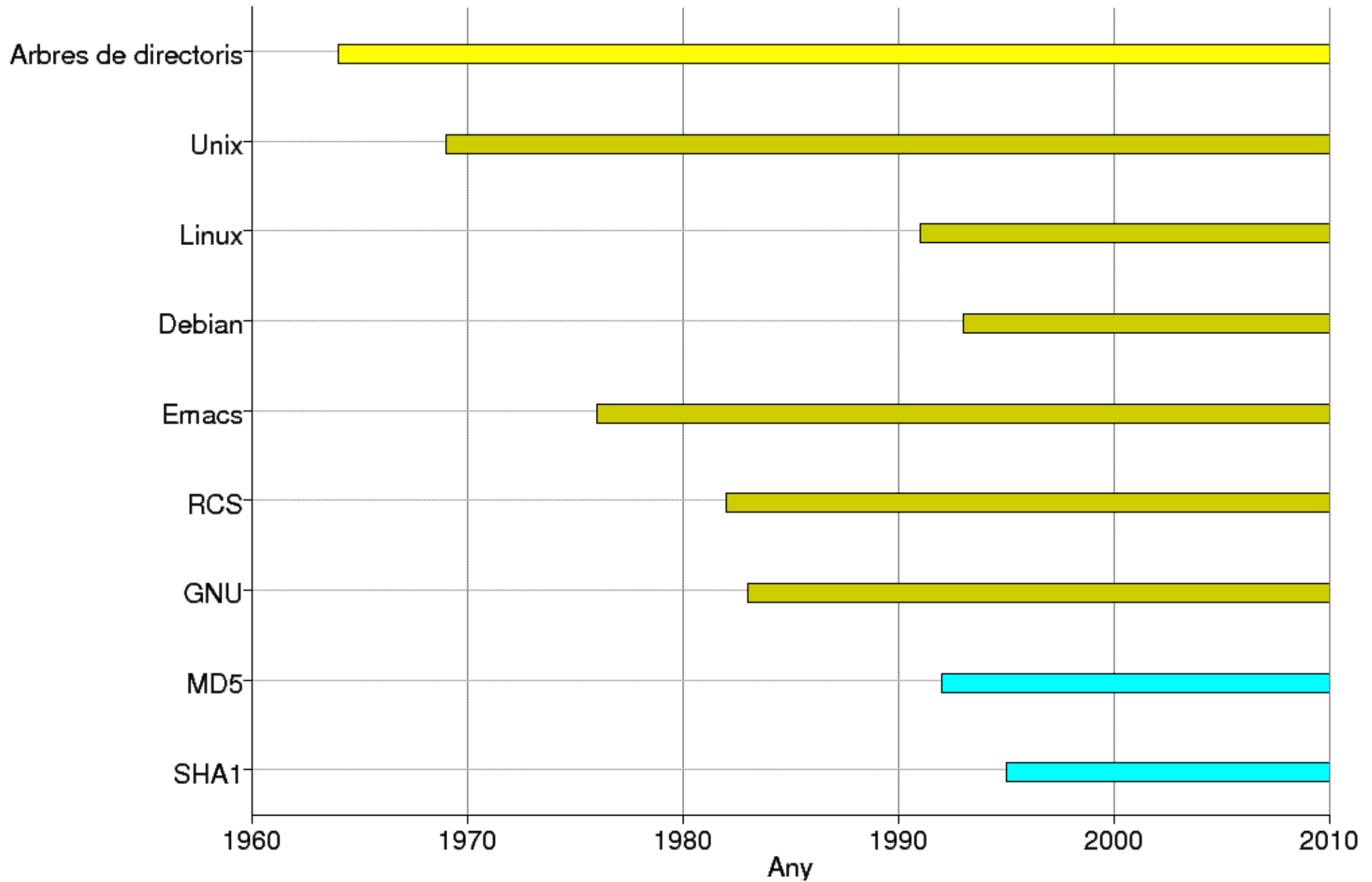
Cronologia d'algunes aplicacions (2/5)



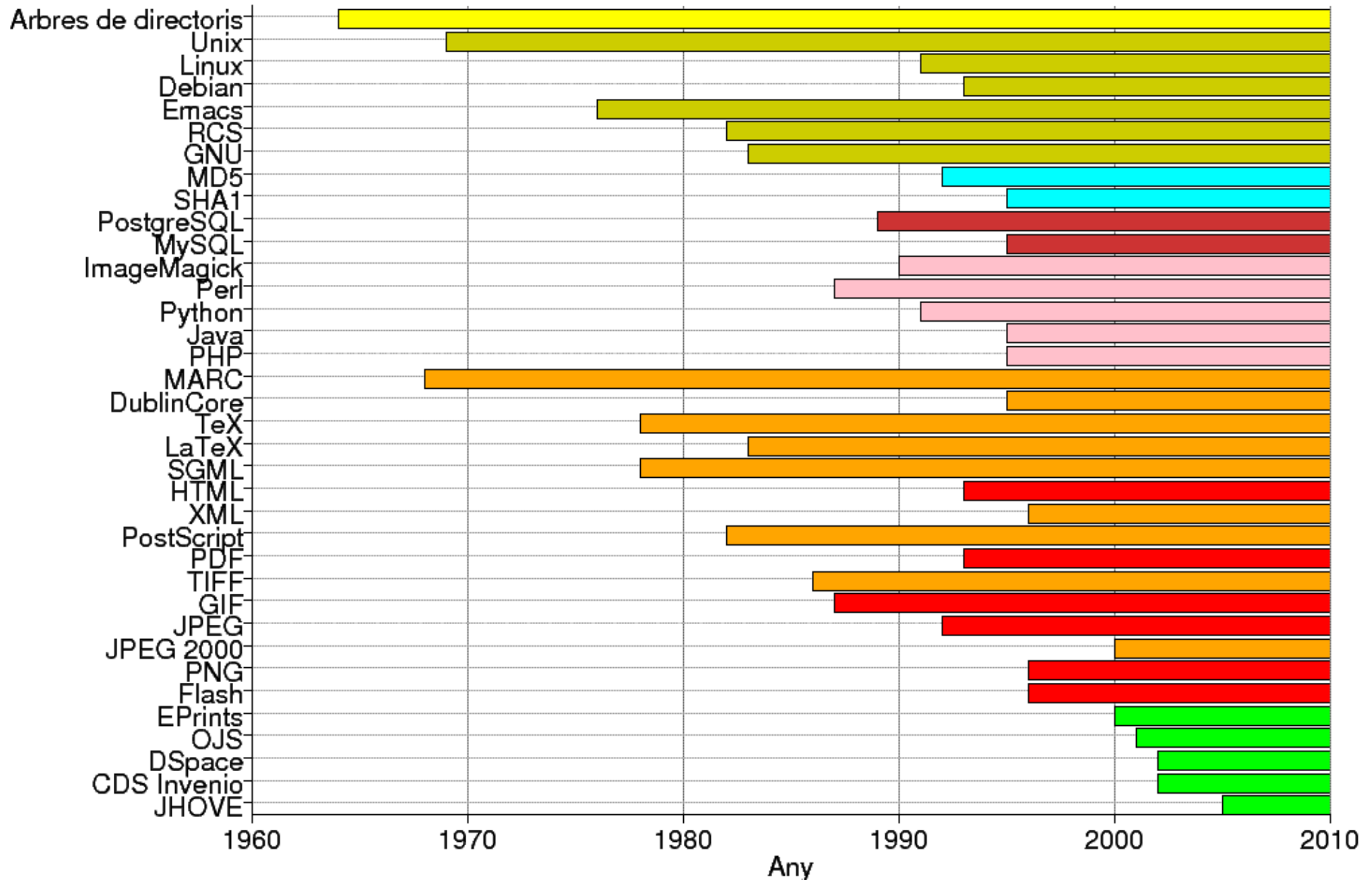
Cronologia d'algunes eines (3/5)



Cronologia d'algunes infrastructures (4/5)

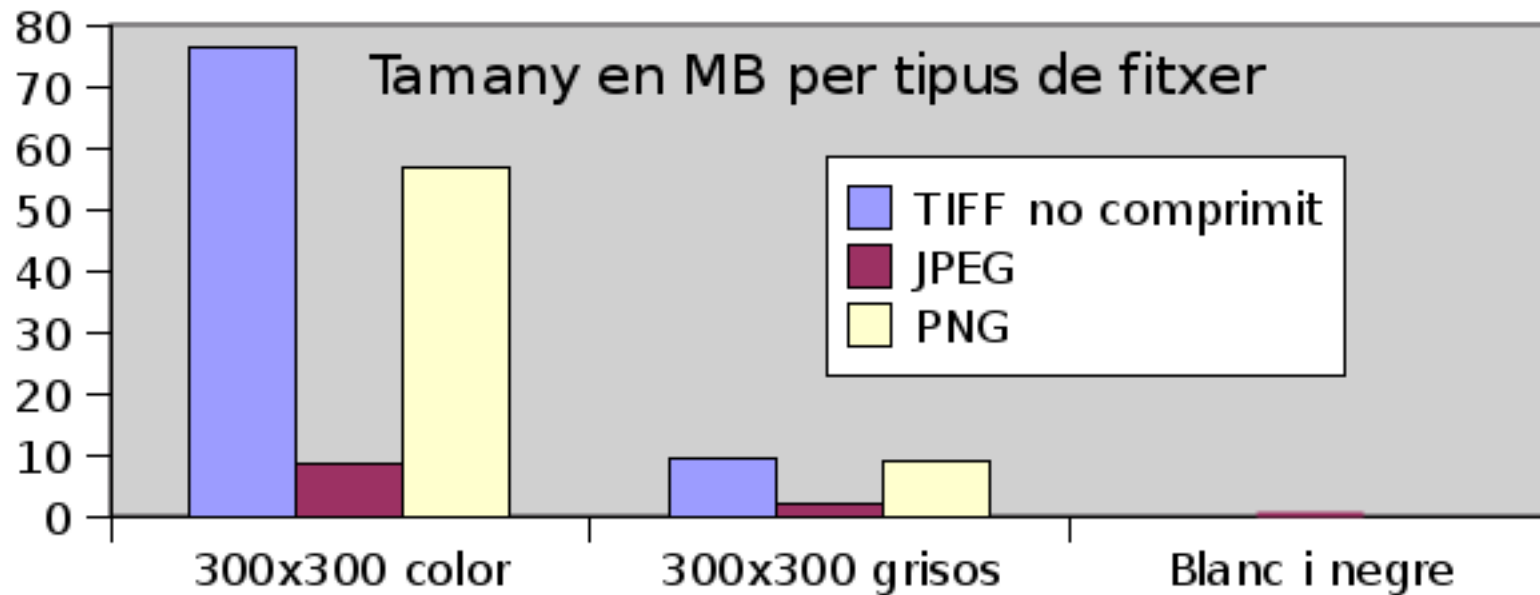


Cronologia del nostre context (5/5)



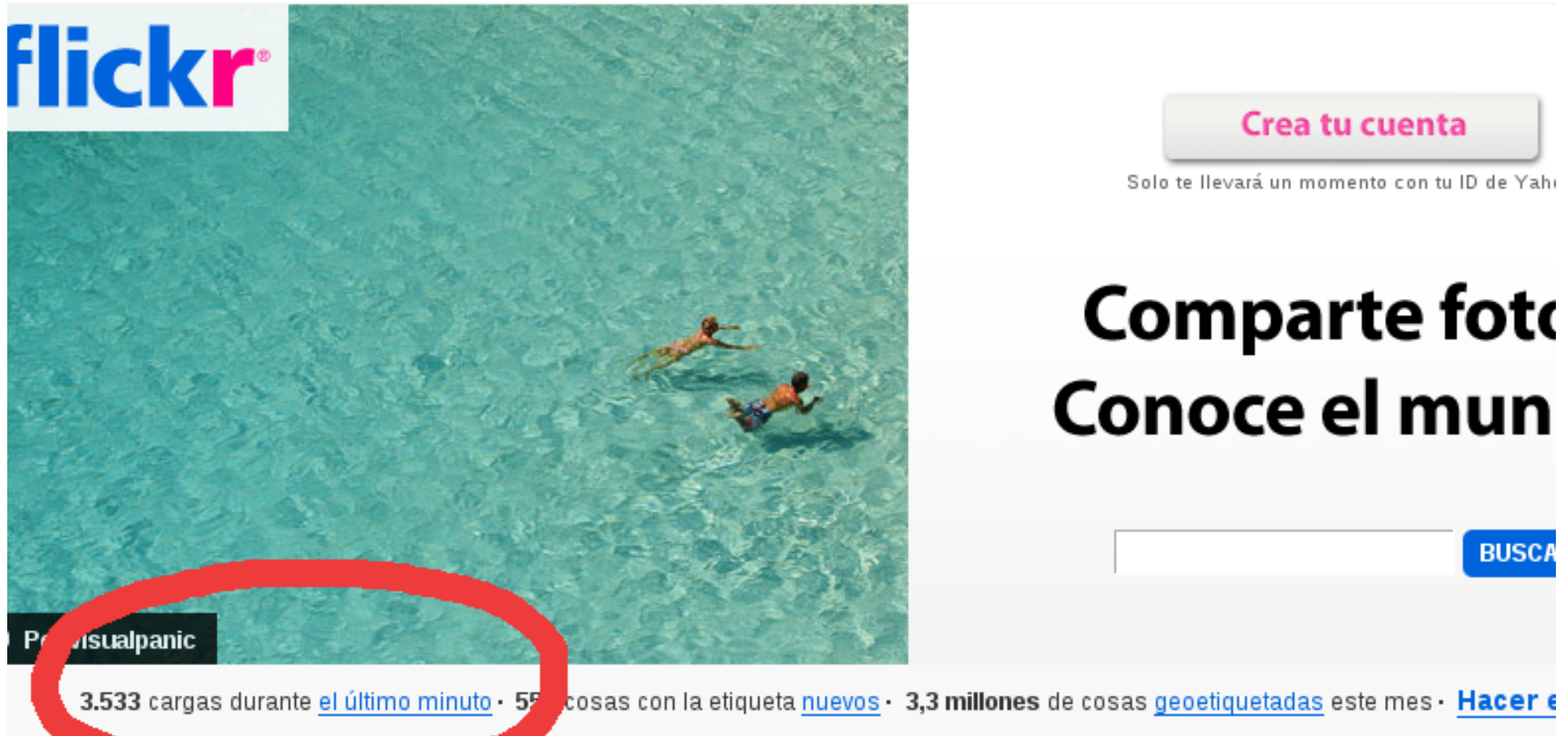
Per què digitalitzem en TIFF? (1/5)

- De debò és 10 vegades millor que el JPEG?
- Què és millor, tenir 1 TIFF o 10 JPEGs?
 - 100.000 TIFFs o 1.000.000 JPEGs?
- Què és millor, blanc i negre en TIFF o color en JPEG?



Per què digitalitzem en TIFF? (2/5)

Segur que el TIFF té més garanties de futur que el JPEG?

A screenshot of the Flickr homepage. The left side features a large, high-resolution photograph of two people swimming in clear, turquoise water. The Flickr logo is in the top left corner. On the right side, there is a 'Crea tu cuenta' button, a search bar with a 'BUSCAR' button, and promotional text. At the bottom, a navigation bar contains various links and statistics. A red hand-drawn oval highlights the text '3.533 cargas durante el último minuto' in the bottom navigation bar.

flickr®

Crea tu cuenta

Solo te llevará un momento con tu ID de Yahoo!

Comparte fotos
Conoce el mundo

3.533 cargas durante [el último minuto](#) • 55 cosas con la etiqueta [nuevos](#) • 3,3 millones de cosas [geoetiquetadas](#) este mes • [Hacer e](#)

Per què digitalitzem en TIFF? (3/5)

Per guardar aquesta informació necessitem 71 MB *de més, per fitxer?*

[...]

ReferenceBlackWhite: 0, 255, 0, 255, 0, 255

DateTimeCreated: 2007-02-28T15:03:50

ScannerManufacturer: i2s Digibook Scanner

ScannerModelName: Suprascan 10000 RGB

ScanningSoftware: Digibook Version :5, 4, 8, 12

Orientation: normal

SamplingFrequencyUnit: inch

XSamplingFrequency: 300

YSamplingFrequency: 300

BitsPerSample: 8, 8, 8

BitsPerSampleUnit: integer

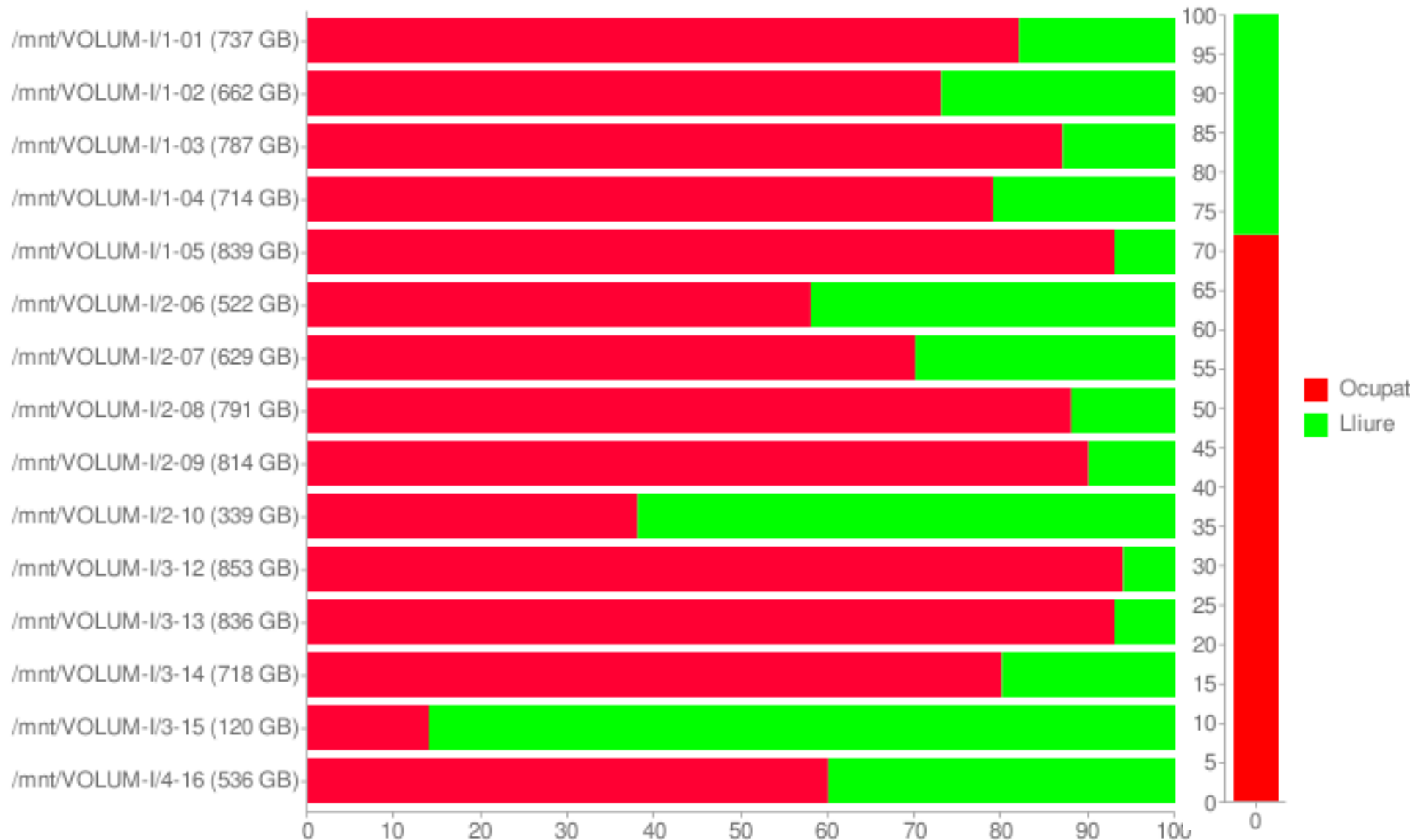
SamplesPerPixel: 3

[...]

Per què digitalitzem en TIFF? (4/5)

Estat dels discs al Volum-I

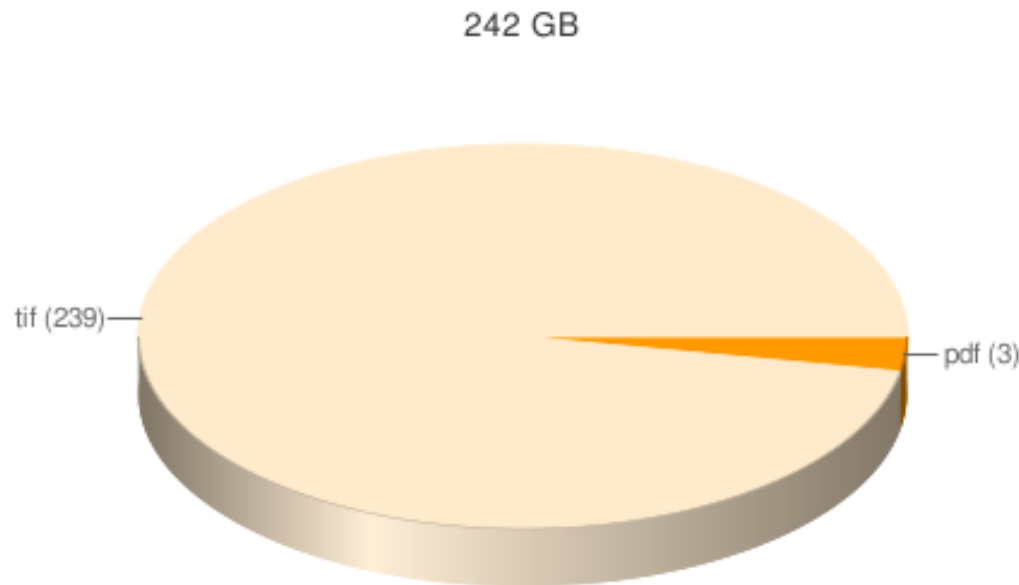
9.9 TB



Per què digitalitzem en TIFF? (5/5)

Butlletí del Centre Excursionista de Catalunya, 1891-1938

- 538 números
- 19.749 pàgines
- 242 GB



Pèrdua de bits? (*Rotten bit?*)



No és



- A un disquet li pots tocar la superfície i fer-li malbé *un trosset*.
- Els sistemes actuals tenen tantes capes d'abstracció que no és possible ni tan sols saber on és físicament el fitxer.

Els bits perduts de debò són els dels documents no preservats.

Compte amb els *experts* en preservació! (1/2)

[Portico is] an organization whose business model depends on persuading libraries that the problems of digital preservation are so complex and difficult that they cannot and should not attempt to tackle them for themselves [...]



PORTICO

Your content.
Preserved here.

<http://blog.dshr.org/2009/05/sheila-morrisseys-comment.html>

Compte amb els *experts* en preservació! (2/2)

- *Denigration*: downplaying libraries' capabilities,
- *Exaggeration*: making the problems look more difficult than they are, primarily by claiming that many tasks are essential which are in fact optional,
- *Obfuscation*: confusing the uninitiated reader by using undefined jargon and failing to cite definitions and sources, thus making the arguments accessible only to experts.



PORTICO
Your content.
Preserved here.

<http://blog.dshr.org/2009/05/sheila-morrisseys-comment.html>

Preservation is not a place (1/2)

Stephen Abrams, Patricia Cruse, John Kunze
International Journal of Digital Curation, Vol. 4, N. 1, 2009
<http://ijdc.net/index.php/ijdc/article/view/98>

S. R. Ranganathan,
The Five Laws of Library Science
Madras, 1931

1. Els llibres són per ser usats.
2. A cada lector el seu llibre.
3. A cada llibre el seu lector.
4. Estalvieu-li temps al lector.
5. La biblioteca és un organisme en creixement.



En cas de dubte, apliquem-los també als nostres documents digitals.

Preservation is not a place (2/2)

Stephen Abrams, Patricia Cruse, John Kunze

International Journal of Digital Curation, Vol. 4, N. 1, 2009

<http://ijdc.net/index.php/ijdc/article/view/98>

- Lots of copies keep stuff safe
- Lots of services keeps stuff useful
- Lots of uses keep stuff valuable

CDLib Curation Micro-Services

- Providing security through redundancy
Però que les còpies siguin visibles i accessibles! Els «arxius foscos» no són gaire útils...
- Maintaining meaning through description
És a dir, amb catalogació (d'acord, metadades...)
- Facilitating utility through service
És a dir, amb una aplicació que ho gestioni (organitzi, indexi, etc.)
- Adding value through use
Notificacions, alertes, comentaris, etc.

<http://www.cdlib.org/services/uc3/curation/>



Algunes lliçons apreses en els darrers anys

(Una llista personal i provisional)

- Estàndards, estàndards! *Busque, compare*, i si no en trobem cap, crear-ne de locals. Hem de pensar a llarg termini.
- Identificar bé els fitxers, que en definitiva és el que queda. Triar noms i convencions que permetin una automatització posterior.
- Conèixer eines, que n'hi ha moltes i de molt bones (Debian és un tresor!)
- Fer els controls de qualitat com més aviat millor.
- Consensuar bé els circuits de treball, perquè cadascú dóna per suposat una cosa diferent (encara no en sabem prou).

Moltes gràcies!

