

# Les défis de la préservation numérique



Mark Evans  
Laura Carter

Tessella



## Horaire

9 h 30 – 9 h 40	Accueil et présentations
9 h 40 – 10 h 15	Les documents électroniques – enjeux et défis
10 h 15 – 10 h 30	Pause
10 h 30 – 11 h 05	Revue historique
11 h 05 – 11 h 35	Solutions possibles
11 h 35 – 12 h	Discussion

# Présentations

- Mark Evans – Chef des services professionnels, Archives numériques

[mark.evans@tessella.com](mailto:mark.evans@tessella.com)



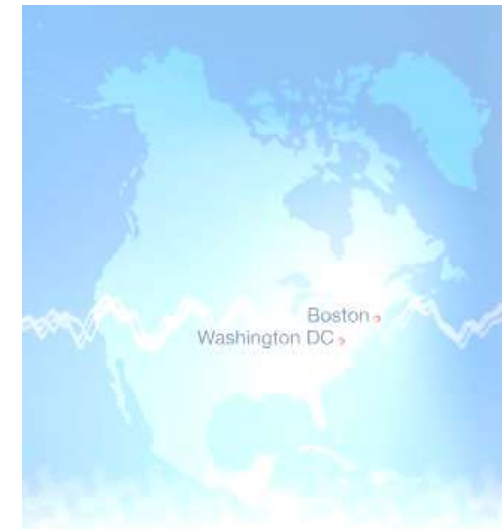
- Laura Carter – VP, Ventas au gouvernement fédéral

[laura.carter@tessella.com](mailto:laura.carter@tessella.com)



## **Expertise-conseil et technologie**

- Société fondée en 1980 : offre des services de qualité depuis 28 ans
  - TI et innovations en R. et D.
  - Croissance constante, segments de marché diversifiés
  - Rentabilité constante, flux positif de trésorerie
  - Société rentable au R.-U., aux É.-U. et aux Pays-Bas
- Services d'archivage et de préservation numériques éprouvés
  - 10 années d'expérience
  - 40 % de ses activités en Amérique du Nord
- Chiffre d'affaires de 14,5 M£ (33,6 M \$CAD) (2007/08)
- Personnel hautement qualifié
  - 250 employés
  - 60 % sont titulaires d'un doctorat
  - Formation poussée



# International Engagements



La British Library



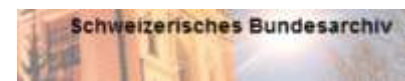
UK National Archives

nationaal archief

Archives nationales des Pays-Bas



US National Archives and Records Administration



Swiss Federal Archives



Archives de la Malaisie

# Étendue de notre expérience



Adoptants précoces :  
Élaboration d'une analyse coûts-efficacité  
pour l'industrie pharmaceutique  
Banc d'essai – Archives des Pays-Bas



Initiatives nationales  
Archive numérique R.-U. 2003 – 2008  
Archive de documents électroniques - NARA  
Datnet de la National Science Foundation  
Archive numérique - British Library  
Archives nationales de la Malaisie  
Archives nationales des Pays-Bas  
Archives fédérales suisses  
*Fournisseur préféré du Wellcome Trust*



Outils de préservation  
Registre de formats techniques PRONOM  
Outil d'identification de format DROID  
Émulateur DIOSCURI

Initiatives internationales  
Membre du consortium PLANETS de l'UE  
• Member of the EU PLANETS Consortium



nationaal archief

DIOSCURI

Search PRONOM >



# Séance 1

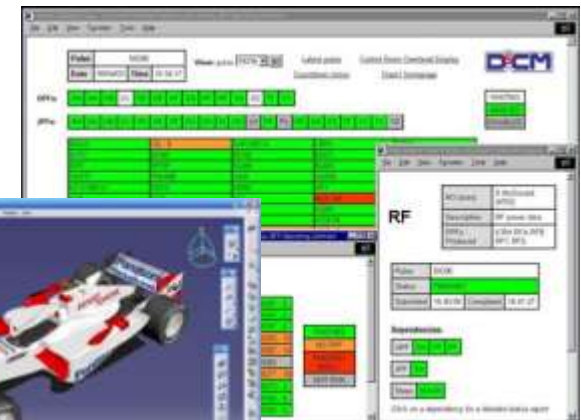
## Les documents électroniques – enjeux et défis

# Séance 1 : Programme

- Nature des documents électroniques
- Comprendre le document électronique
- Principaux enjeux
- Le grand défi
- Études de cas
  - Domesday Book
  - NASA
- Autres considérations

# La nature des documents électroniques

- Très différents des documents traditionnels
- Le contenu et l'aspect dépendent d'un large éventail de technologies
- Ont des propriétés comportementales
- L'interaction avec l'utilisateur est très différente
- Permet de multiples vues des mêmes données
- Multiples structures
- Restrictions d'accès complexes



# Comprendre le document électronique

- Le document électronique n'est pas directement lisible
  - Essentiellement composé d'une série de 1
  - Dépend des formats de fichiers
  - Nécessite un logiciel d'application
  - Nécessite un système d'exploitation
  - Nécessite un matériel informatique
- Il faut davantage que le seul fichier de données pour comprendre complètement l'information
  - Métadonnées contextuelles, structurelles, descriptives et techniques



# Un exemple concret : Interpréter le flux binaire

Métadonnées  
descriptives

La seule partie de l'édifice du Centre qui a survécu à l'incendie de 1916, la Bibliothèque, a été préservée grâce à un employé - M. MacCormac – qui a fermé les portes de fer à l'entrée du bâtiment.

Métadonnées techniques

Nom de fichier : CLP.jpg

Sortie de l'application



Format de données

Format d'échange de fichiers JPEG

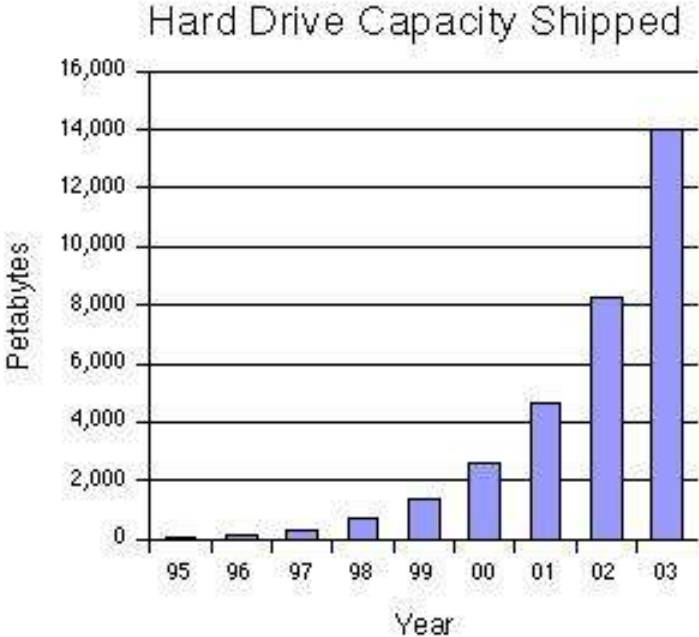
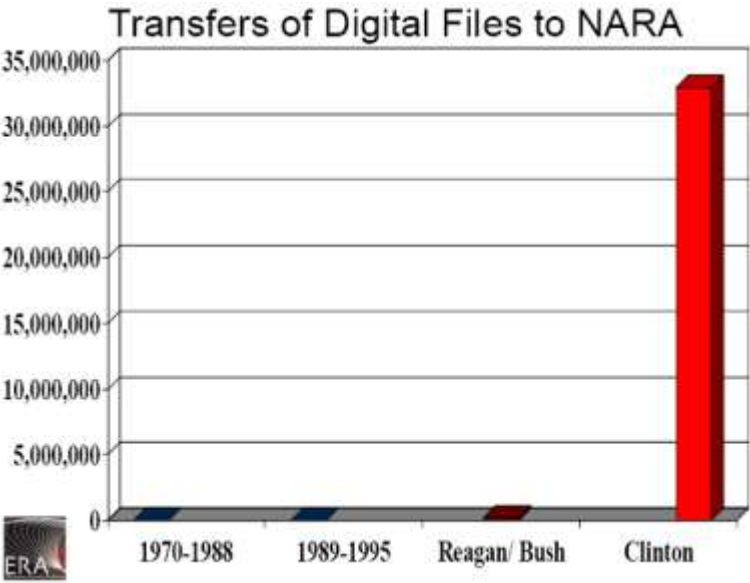
Séquence binaire

11111111 11011000 11111111 11100000 00000000  
00010000 01001010 01000110 01001001 01000110

# Principaux enjeux

# Gérer des volumes sans cesse croissants

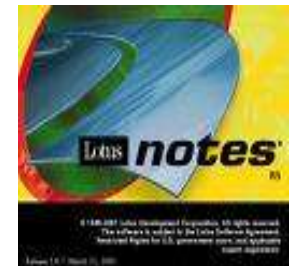
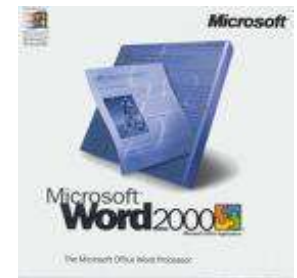
- Le volume de documents électroniques connaît une croissance exponentielle



1970 –1988 = 6 000, 1989-1995 = 15 000, Reagan/Bush = 200 000  
 GW Bush : 300 millions seulement pour les courriels

# Composer avec une grande variété de formats

- Il existe de nombreuses façons différentes d'enregistrer de l'information dans une variété de formats et de structures de données
  - Documents de bureautique
  - Bases de données
  - Courriels
  - Sites Web
  - Images
  - Dessins d'ingénierie
  - Ensembles de données scientifiques
- Peu ou pas de contrôle sur les formats
- Large utilisation de formats « propriétaires »



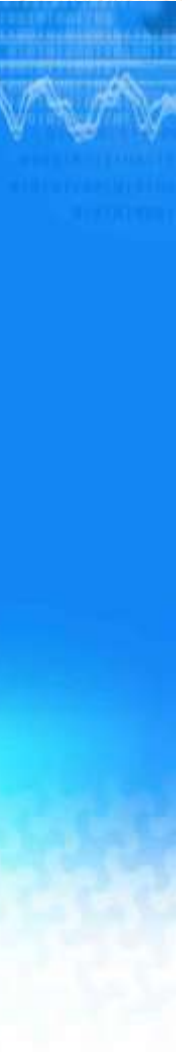
# Lutter contre l'obsolescence

- Obsolescence des supports
  - Qui se rappelle des disquettes?
  - Quand les CD disparaîtront-ils?
  - Les supports aussi se dégradent
- Obsolescence des formats
  - Qui utilise encore Wordstar?
  - Les formats ont une durée de vie beaucoup plus courte que les documents
- Obsolescence des technologies
  - À quelle fréquence changez-vous de PC?
  - Utilisez-vous Windows 3.1?



# « Assurer l'accès continu et à perpétuité à une copie authentique d'un document électronique »

- La technologie évolue à un rythme effarant, entraînant l'obsolescence des matériels, des logiciels et des supports
- La période de conservation des documents électroniques peut dépasser la durée de vie des matériels et des logiciels qui ont été utilisés pour les créer et pour y accéder
- Il faut assurer l'intégrité du flux binaire tout au long de sa période de conservation
- Les restitutions futures du document **DOIVENT** être fidèles à l'original
- Les archives doivent être vues comme une source fiable



# Études de cas

# Le Domesday Book – Une leçon de l'histoire



Mark.evans@tessella.com

laura.carter@tessella.com

## L'original – 1086

- Grand recensement de l'Angleterre commandé par Guillaume le Conquérant
- Documents manuscrits sur papier
- Conservé aux Archives nationales du Royaume-Uni
- Les documents papier peuvent être conservés longtemps...



# Le Domesday Book moderne – 1986

- Commandé par la BBC pour célébrer le 900<sup>e</sup> anniversaire
- Compilation de documents historiques
- Édition électronique
  - Logiciel personnalisé
  - Matériel personnalisé
  - Vidéodisques laser 12 pouces



**On a oublié les besoins de préservation!!!**

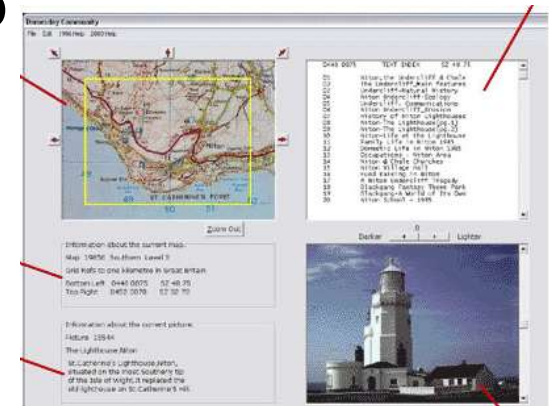
# Le Domesday Book moderne – Problèmes

- Support obsolète :
  - Format non standard
  - Fin du soutien 5 ans plus tard
- Logiciel obsolète :
  - Format personnalisé
  - Mal documenté
- Matériel obsolète :
  - Micro-ordinateur 8 bits populaire dans les écoles
  - Remplacé par les PC moins de 5 ans plus tard
  - Petit nombre de lecteurs produits



# Domesday Book moderne – Solution

- Rétro-ingénierie
  - Il a fallu faire appel à la mémoire de ses artisans!
  - Et faire des expériences
  - Et utiliser eBay pour retrouver l'équipement original
- 2001 : conversion dans un format lisible
  - Migration vers la technologie Web
  - Effort d'émulation
- Coût : 2,5 millions de livres
  - Coût initial : 4 000 livres



# Domesday Book – Leçons tirées de l'expérience

- Le numérique est moins facile à préserver que le papier :
  - 20 ans par opposition à 900 ans!
- Effort monumental pour un petit ensemble de documents
- Solution non évolutive :
  - Il a fallu faire appel aux connaissances de ses artisans
  - Il a fallu expérimenter
  - Beaucoup d'efforts...
- Nous devons faire mieux!



# NASA – Préservation partielle

## Données du Viking Lander

- La NASA a recueilli une importante quantité de données télémétriques
- A réussi la migration des supports
- N'a pas préservé la capacité d'accès
- N'a récupéré qu'une série de 1 et de 0
- Réelle possibilité réelle de perte d'information
- Coût réel de la récupération - Inconnu



*Courtesy NASA/JPL-Caltech*

# Autres considérations

Mark.evans@tessella.com

laura.carter@tessella.com

## Autres considérations

- Transformation des processus opérationnels
  - Les documents électroniques nous obligent à transformer nos processus opérationnels
  - Ce qui a une incidence sur l'évaluation, le transfert, la préservation et la livraison
- Transformation de la main-d'œuvre
  - La façon de travailler changera
  - De nouvelles compétences seront nécessaires
  - Il existe un besoin réel de « spécialistes de la préservation numérique »

# Résumé

- Les documents électroniques sont d'une nature très différente
- Nous devons faire face à de nombreux enjeux et défis nouveaux
- Nous ne devons pas répéter les erreurs du passé
- Nous devons nous éduquer et peut-être modifier nos façons de faire
- C'est difficile, mais nous sommes tous aux prises avec les mêmes problèmes

# Questions?

[Mark.evans@tessella.com](mailto:Mark.evans@tessella.com)

[laura.carter@tessella.com](mailto:laura.carter@tessella.com)