

RFC 4810 : Long-Term Archive Service Requirements

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 30 mars 2007

Date de publication du RFC : Mars 2007

<https://www.bortzmeyer.org/4810.html>

L'archivage de données à long terme est un enjeu essentiel et qui soulève d'innombrables difficultés techniques et sociales. Ce RFC fait le point sur le cahier des charges d'une solution IETF.

C'est une banalité, mais une banalité tristement vraie, que de dire que notre époque ne laissera que peu de traces. Si je veux écrire un livre sur la diplomatie française au XVIIIème siècle (je recommande la lecture du « Secret du Roi » de Gilles Perrault), j'ai à ma disposition une impressionnante quantité d'archives, chacun des protagonistes de l'époque écrivait abondamment et ce qu'il écrivaient est toujours lisible aujourd'hui. Au XXIème siècle, ils envoient des "emails" en Word et, dans cinquante ans, on aura perdu les sauvegardes, on ne pourra plus lire les clés USB ou les DVD et, de toute façon, Word sera oublié depuis longtemps <<https://www.bortzmeyer.org/afterword.html>>.

Bref, un service d'archivage à long terme nécessite plus que des graveurs de DVD. Il nécessite une organisation sociale solide (cf. la BNF en France, BNF qui d'ailleurs travaille sur le dépôt légal du Web <http://www.bnf.fr/pages/infopro/depotleg/dl-internet_quest.htm>). Et il nécessite un certain nombre de services techniques.

Notre (court) RFC ne décrit pas en détail les solutions techniques (on peut consulter un article du CINES <http://www.cines.fr/IMG/pdf/Archivage_perenne_Gazette20.pdf> à ce sujet) mais les exigences du service.

Celui-ci devra donc permettre de gérer des autorisations (de dépôt et d'accès), de spécifier des métadonnées associées à chaque objet numérique archivé, de prouver (par exemple grâce à la cryptographie) que les objets n'ont pas été modifiés, etc.

Les considérations de sécurité sont particulièrement détaillées : car les données archivées peuvent être sensibles et nécessiter, en plus de l'intégrité, de la confidentialité.

Parmi les articles déjà publiés sur le sujet (comme celui du CINES cité plus haut), Nicolas Krebs recommande Archiver le web, il devient temps d[Caractère Unicode non montré ¹]y penser <http://bulles.agora.eu.org/20070405_archiver_le_web.html> qui résume bien le problème.

1. Car trop difficile à faire afficher par L^AT_EX