



# Conférence-débat : synthèse

« Pérennité de l'information :  
vers une solution et une réalisation »

## e-Babel

info@e-babel.org

**25 novembre 2009**

### Avec la participation de :

Mme <b>Véronique Sordet</b>	: 7ème Programme cadre de recherche européen
M. <b>Andréa Schweizer</b>	: histoire « de Gutenberg à Internet »
M. <b>Raymond Morel</b>	: hardware et software à travers les âges et aujourd'hui
M. <b>François Wollner</b>	: Excusé
M. <b>Johann Sievering</b>	: e-Babel

**L'objectif** était ; d'une part d'explorer les divers champs sur la thématique de la pérennité de l'information dans le temps, et d'autre part de lancer un débat pour tenter de trouver des voies permettant de mettre en place un projet concret.

Ce document n'est pas une approche technique du sujet, car ce qui importe ici, ce sont les réflexions autour de la pérennité de l'information et ses divers aspects afin de dessiner l'architecture générale du projet. Les aspects techniques viendront ultérieurement. Il existe déjà dans ce domaine une très abondante littérature.

*Dans cette phase : c'est au sujet de guider la technologie et non le contraire.*

### Les thématiques abordées furent :

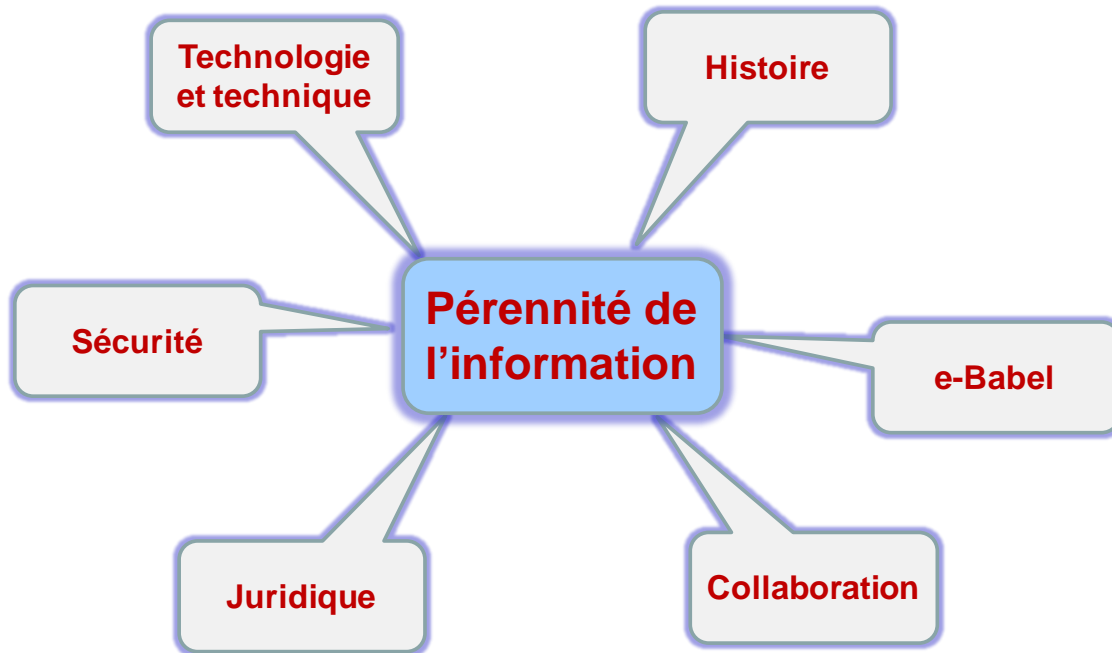
- Histoire de l'informatique et des systèmes d'information ;
- Les besoins des utilisateurs au cours du temps et aujourd'hui ;
- Etat de l'art des systèmes informatiques ;
- Technologie et science : ce qui est possible de faire et pas aujourd'hui ;
- Aspects juridiques, éthiques et politiques;
- Vers une solution concrète.

**Une visite** de l'API (*Association pour le Patrimoine Industriel*), a permis de lancer le sujet dans un contexte concret de processus de pérennisation des données, des objets et des savoir-faire. Cette visite nous a montré les diverses techniques d'impression des premières machines aux monotypes et aux linotypes plus complexes. Un projet de couplage de ces machines avec des systèmes informatiques afin de pérenniser les savoir-faire est actuellement en cours ; c'est la rencontre entre le passé et le présent.

Nous avons constaté de visu la problématique de la persistance de clichés photographiques. Nous avons vu que la conservation seule n'est pas suffisante, il faut également construire un indexe ainsi qu'une systématique de recherche afin de situer chaque objet dans son contexte avec des références croisées entres-eux pour être utile.



## Les avenues de la conférence



De ces avenues, découlent les thèmes de la conférence :

- *Histoire de l'informatique* et des systèmes d'information ;
- *Etat de l'art et des technologies* des systèmes informatiques;
- *Sécurité* : ce qui est possible de faire et ne pas faire aujourd'hui ;
- Aspects *juridiques*, éthiques et politiques;
- Les projets et la *recherche et de collaboration*;
- Vers une *solution concrète* : *e-Babel*.

### Le fil conducteur des réflexions.

Le fil conducteur est construit sur cinq considérations qui furent les bases de la réflexion. Les réponses à ces considérations ont permis de former un contour du domaine du futur projet. e-Babel avait déjà mené un certain nombre d'éléments qui ont été discutés lors du débat.

#### Considération 1 : histoire, évolution

Dans la réflexion sur la pérennité de l'information:  
Comment l'évolution des technologies et de la représentation de l'information peut-elle *s'impliquer dans notre présent* et nous aider à *construire l'avenir* ?



## Considération 2 : technologie

L'évolution observable des technologies nous montre que nous développons des supports de l'information *de plus en plus virtuels* (de plus en plus indépendants des technologies et de plus en plus communicants).  
*Jusqu'où* pourrons (devons)-nous aller ?  
*Qu'advientra-t-il* des données du passé ?

## Considération 3 : sécurité

Une *donnée*, dans notre contexte, est un état interprétable d'un support accessible par nos sens.  
Ce support peut être plus ou moins *sensible au temps et aux bruits ambiants*, plus ou moins *fiable* et plus ou moins *facile d'accès*.  
Cet état peut être *modifié* intentionnellement ou non.  
Les données peuvent être codées, ne concerner que certains acteurs et n'être diffusables que sous certaines conditions.  
*Quelles sont les règles* techniques, sociales, politiques, commerciales à mettre en œuvre aujourd'hui pour accéder aux données du passé et prévoir celles de demain ?

## Considération 4 : juridique éthique

Les données sont techniquement là et disponibles.  
Peut-on, a-t-on le droit, de *les utiliser directement* selon ses besoins propres ?  
Peut-on les *croiser*, les *diffuser* ?  
Quels sont les droits des auteurs ?  
*Quelles règles et quelles lois* s'appliquent aujourd'hui, que doit-on *prévoir* pour l'avenir ?

## Considération 5 : collaboration

Le temps de l'inventeur créant dans sa cuisine a évolué vers la participation à des *réseaux sociaux*.  
Quelles collaborations et quels grands projets sont actuellement présents pour soutenir les nouvelles idées ?



## Les points clés des réflexions

### Considération 1 : histoire, évolution

Il est nécessaire de retracer l'histoire de l'informatique pour pouvoir en dégager les évolutions et intégrer chaque étape dans le processus de pérennisation.

Lorsque l'on parle de pérennisation, il ne faut pas se focaliser uniquement sur les aspects technologiques, mais il faut également intégrer :

- le contexte pour l'interprétation des contenus ;
- les savoir-faire pour utiliser et gérer les données ;
- les règles et les lois contemporaines en vigueur à la donnée pour fixer les modalités d'utilisation des contenus ou des processus dans l'intention des auteurs ;
- les règles et lois actuelles pour assurer que les traitements réalisés répondent aux contraintes du moment.

Un autre aspect de la pérennisation est la localisation historique des équipements, des supports ou des contenus. La visite de l'API nous a bien montré que, même en possession du matériel (par exemple les plaques photographiques), le manque d'indexage pose un problème majeur et peut même invalider le travail d'archivage physique des supports des données. Une analogie directe peut être faite avec les supports informatiques. Comme un intervenant l'a souligné « *un document mal, ou pas, classé est un document perdu* ». Il faut donc mettre à disposition un accès organisé et systématique pour circuler parmi les données, les supports et les relations entre tous ces objets.

Un projet de présentation de l'historique de l'informatique dans une interface graphique et utilisant des hyperliens permettrait de synthétiser l'ensemble des équipements, des supports ainsi que des logiciels. Une mise en relation dynamique (par exemple à l'aide de la souris), permettrait de créer des zones de pertinence qui pourraient prendre en charge la pérennité d'objets et simultanément les munirait d'informations permettant de les archiver correctement.

### Considération 2 : technologie

Cette question a été abordée sous plusieurs angles dont celui de la production en masse des données. Notre époque dispose de technologies où même une machine à café peut produire des informations. En soit, il est anodin de laisser l'automate passer des commandes dès qu'il en a « besoin ». Faut-il conserver ces informations ? En effet, elles pourraient devenir plus tard pertinentes, par exemple dans certaines situations comptables.

Les communications entre machines sont de plus en plus nombreuses, et permettent de maintenir une qualité de service. Faut-il les conserver ? Par exemple, elles peuvent permettre de retracer des faits dans le cadre judiciaire.

La virtualisation passe par la communication et génère de plus en plus de données, qu'en faire ?

Une des pistes évoquée est de doter chaque donnée d'un degré de pertinence et d'un temps de vie. Ce tableau à double entrée permet d'évaluer si la données doit être



conservée et combien de temps. Une fois la pertinence ou le délai passé elle est simplement supprimée.

Cependant, comment décider de la pertinence d'une donnée et comment estimer son temps de persistance. En effet, ces deux paramètres sont contextuels. A leur production, ces données ont une certaine évaluation qui peut évoluer en fonction du nouveau contexte (technologique ou social), des nouveaux équipements à disposition ou des besoins du moment.

Une solution possible est de joindre un ensemble de règles de modération qui pourraient agir directement sur ces valeurs. Ces fonctions seraient contextualisées. Elles seraient définies dès la production de la donnée et pourraient être ensuite complétées au fur et à mesure de l'évolution du domaine. Ces règles feraient partie intégrante de la donnée et seraient exprimées sous une forme logique (pour assurer leur pérennité indépendamment de la technologie).

### Considération 3 : sécurité

Le domaine de la sécurité recoupe plusieurs approches (même si elle est communément vue sous les angles uniques de la confidentialité et de la protection des accès).

La sécurité des données couvre en substance :

- *Intégrité* : les données restent exactes et complètes au cours du temps;
- *Disponibilité* : accès au moment voulu par des acteurs autorisés;
- *Accès* : possibilité d'accéder aux données par au moins un vecteur de recherche (par exemple un index);
- *Confidentialité* : seuls les acteurs autorisés ont un accès aux données selon leur profil;
- *Traçabilité* : conservation d'un historique consultable des opérations réalisées sur les données locales ou en itinérances;
- *Diffusion / divulgation* : maîtrise de l'évolution de la donnée dans le temps et dans l'espace ;
- *Juridique* : utilisation dans la loi et dans les règles. Protection des droits d'auteur et d'exploitation.

### Considération 4 : juridique éthique

Hors mis la problématique des aspects technologiques, il demeure des questions concernant les contenus et leurs implications sur les individus, sociales, politiques, institutionnelles, etc. En effet, un contenu peut révéler des aspects inattendus du passé. Et ce phénomène est amplifié avec les croisements et les recoupements de données en masse. *Le droit à l'oubli n'est plus assuré.*

De plus, les droits affiliés à des données et aux processus de gestion des données peuvent devenir complexes surtout dans les cas de disparition des auteurs ou des bénéficiaires.

Ces aspects sont actuellement dans un vide juridique qu'il est nécessaire de combler au plus vite, surtout dans ce contexte où les technologies évoluent de plus en plus rapidement. Ces questions ne feront que de s'amplifier au cours du temps.

Les aspects éthiques sont également à intégrer dans la réflexion. Une donnée impliquant l'avenir d'un individu ou d'une entité morale doit-elle être



systématiquement portée à la connaissance de l'intéressé ? Ces considérations sont similaires à celles qu'ont les médecins face à leur patient.

### Considération 5 : collaboration

Il n'est plus question aujourd'hui de développer des applications locales, « dans son garage », pour un public restreint. L'informatique moderne est résolument orientée collaboration et communication, par exemples les réseaux sociaux, le web2, les twitters, les forums, etc.

Un projet a besoin de s'intégrer dans un contexte plus large lui permettant d'une part de bénéficier de l'expérience d'autres projets similaires et d'autre part de se diffuser à une plus large échelle et ainsi rencontrer des critiques constructives. Cette démarche permet d'enrichir le projet et de l'amener plus loin.

### En résumé

Si la technique permet relativement aisément de récupérer les données ainsi que les processus du passé, il ne faut pas négliger l'ensemble des autres aspects évoqués ci-dessus. Tous ces aspects, historiques, technologiques, de sécurités, juridiques et sociales doivent *faire partie intégrante* de l'architecture du système de pérennisation *quel qu'en soit sa nature*.

### Exprimer l'implicite

Un autre aspect à ne pas négliger est la mise à jour d'informations implicites. Ce sont ces informations qui sont présentes, directement visibles ou non, mais *qui font sens que lorsque l'on en connaît la clé*. Très souvent, la clé n'est pas externe au système et détenue par une tierce partie, mais *elle se forge en croisant les informations* déjà connues. Les recoupements et analyses automatiques permettent d'explicitier ces informations (implicites dans le système) et de mettre en lumière certains aspects de l'objet d'information considéré voir de profil de personne.

Il faut donc prendre en charge cette problématique dès le départ pour éviter la réalisation d'un système pouvant traiter des informations contraires au *droit*, à la *morale*, à l'*éthique*, au *droit à l'oubli*, au *droit d'auteur* ou simplement aux *souhaits* des acteurs impliqués.

e-Babel s'emploie à réaliser une architecture intégrant l'ensemble de ces contraintes. (C.f. les documents annexés).