

Introduction au web sémantique

- Historique
- Structure de la pile sémantique
- L'implémentation Stanbol

Le sémantique prévu dès la naissance du web

- Dès 1994 Tim Berners-Lee montre les limitations de l'url et de l'hyperlien
- En 1997 et 1988 il publie un premier draft et son layer cake
- 1999 [Weaving the web](#)
- Dans les années 2000-2010 le web 2.0 puis 3.0

Le sémantique prévu dès la naissance du web

- Dès 1994 Tim Berners-Lee montre les limitations de l'url et de l'hyperlien
- En 1997 et 1988 il publie un premier draft et son layer cake
- 1999 [Weaving the web](#)
- Dans les années 2000-2010 le web 2.0 puis 3.0

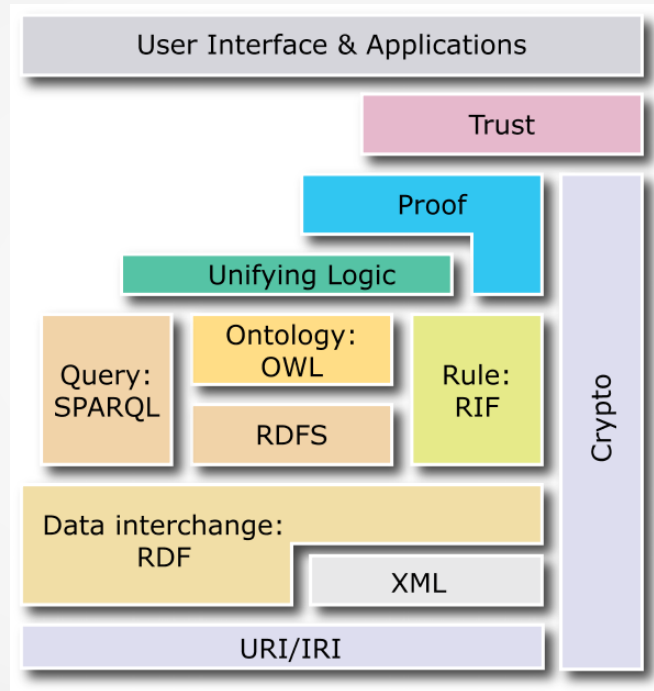
De la complexité de la mise en œuvre 1/2

- W3C : Une norme ne s'impose que si elle est implémentée
 - 2001 – projet [Annotea](#)
 - XHTML 1.0, 1.1, 2.0
 - HTML5 et ses sous spécifications
- Initiatives hors W3C
 - [WHATWG](#)
 - [Schema.org](#)

De la complexité de la mise en œuvre 2/2

- Le layer cake contient 7 couches et autant de langages
- Le XML sert de brique de possède lui aussi ses langages de transformation et de requête
- Tout ces langages ont hérité du XML une syntaxe très verbeuse pour pouvoir être traité au moins en mode semi-automatique
- Les principaux vecteurs de création de documents numériques ne se pas accordé pour unifier les vocabulaires (ou ontologies)
- Beaucoup d'implémentations partielles sans grande visibilité

Layer Cake - 1998



Quels sont ces ingrédients ?

- URI/IRI : identifiant et adresse d'une ressource
- XML : Langage d'échange de flux d'information
- RDF : Framework de description de ressources
- RDFS ou RDF Schema : extension de RDF pour structurer les informations
- SPARQL : langage de requête
- OWL : description des vocabulaires complexes
- RIF : langage d'échange de règles

Que représentent les layers restant ?

- L'unification des flux de données pour présenter une Web API
- La possibilité de pouvoir vérifier la cohérence interne des données en construisant des preuves
- Une couche de sécurité et d'authentification
- De la cryptographie pour protéger les données sensibles (médicales...)

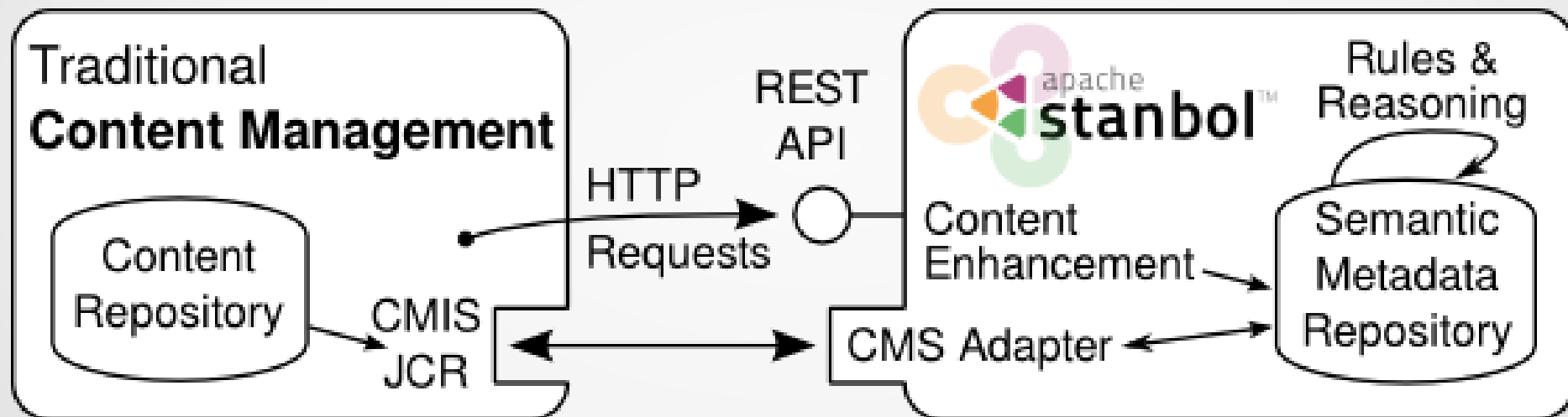
ExistDB : une base XML

- Fonctionnalités phares
 - Stocke et indexe directement des fichiers XML
 - Protocoles : ReST, SOAP, XML-RPC, WebDAV
 - Requête/transformation : XQuery 3.0 et XSLT
 - Formats de sortie : XML, XHTML, JSON
- Avantage
 - Traite directement du XML
- Inconvénient
 - Il faut construire les autres couches (mais parfait pour un microframework)

IKS – l'UE se lance dans le sémantique

- Après le rachat par Dassault du dernier moteur de recherche sémantique 'ouvert' l'UE lance un programme pour réduire la dépendance de son économie aux technologies américaines
- Le but : créer un projet implémentant toute la pile sémantique en 5 ans avec les enjeux de le rendre autonome et flexible
- Le choix de l'opensource n'était évident mais il a été vu comme une solution de pérennité
- En 2012, le projet devient un projet phare de la fondation apache : mission accomplie

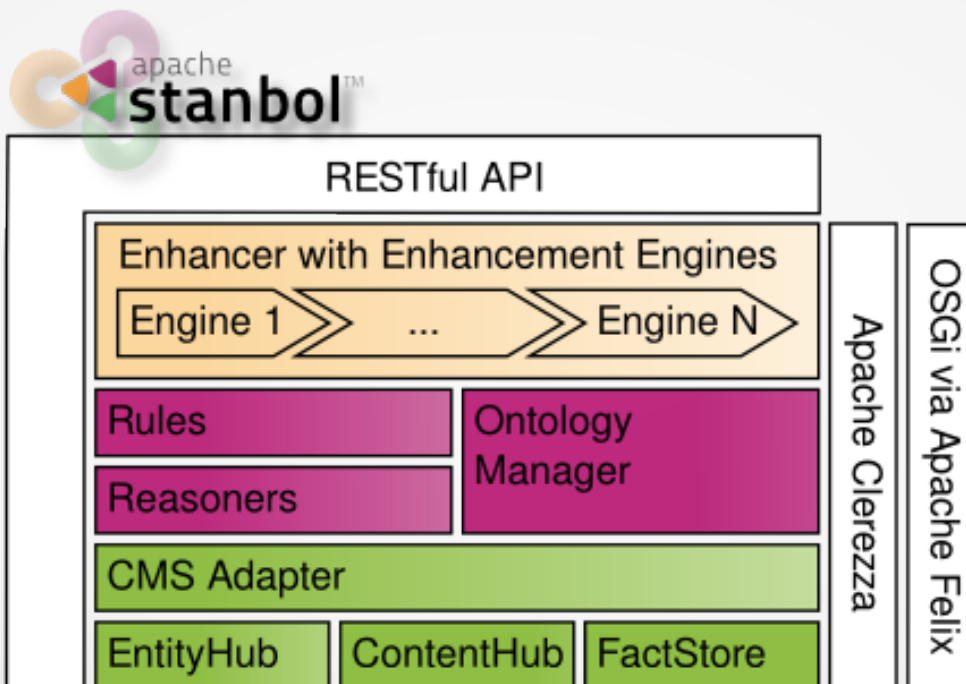
L'interconnexion avec les CMS



Pourquoi se tourner vers les CMS

- Ils sont le principal vecteur de stockage de documents et leur enrichissement est souvent pauvre
- Impliquer tôt les CMS permet d'avoir des retours concrets tôt dans le projet
- Attirer des développeurs de tous les horizons

Le schéma flexible



Une architecture de services

- Chacune des fonctionnalités possède un système d'API de service REST
- Chacune des fonctionnalités possède un système d'extension
- Un fournisseur de service peut implémenter qu'une partie du totale
- Les extensions peuvent être ouvertes ou privatives
- Les données peuvent être ouvertes ou privées

VIE, l'éditeur sémantique

- Concept venu du projet IKS
- Permet d'ajouter les améliorations en asynchrone dans les ressources en construction
- Intégré dans [Create.js](#)

Les problématiques

- Le web sémantique, un web de données ouvertes ?
 - L'appropriation des données par les fournisseurs de services
 - Des ontologies fermées
- Comment construire une communauté d'intérêt commun ?

Références

- [Web sémantique \(Wikipedia\)](#)
- [Introduction au web sémantique \(developpez.com\)](#)
- [Blog de Christian Fauré](#)
- [Microdata et microformats \(Alsacreations\)](#)
- [eXistDB](#)
- [Stanbol](#)
- [VIE](#)
- Et encore beaucoup d'autres.

Le conférencier

- Encolpe DEGOUTE – Bearstech
- Membre de l'ALDIL depuis 1998
- Président de l'ALDIL depuis 2010
- Développeur et disséminateur opensource
- Python/Zope/Plone/Django, mais aussi LUA, C++ et Java à l'occasion
- Membre de la fondation Plone depuis 2007
- Early adopter Stanbol en 2010