

# **Unifier mondes sémantique et non sémantique : enjeux et cas d'application □**

Mohamed Gaha. PhD.

Chercheur Institut de Recherche d'Hydro-Québec

IREQ

# Plan

## 1. Contexte Hydro-Québec

## 2. Base de connaissance

- Technologie traditionnelle
- Technologie sémantique
  - Définition
  - Force et faiblesse

## 3. Combiner données sémantique et non sémantique

- Pourquoi
- Cas d'utilisation

## 4. Résultats et conclusion

# Introduction Hydro-Québec



# Institut de recherche d'Hydro-Québec

- > Plus grand centre de recherche de l'Amérique du nord
  - 500 employés:
    - PhD, Ingénieurs, Techniciens
    - Gestionnaires
    - Avocats
    - Etc.
  - +1000 brevets
  - Publications, normes, comités standardisations
  - Partenariats multiples avec universités, entreprises et organismes



# Projet de recherches

## > Projets informatique:

- CISRI: concept d'interopérabilité sémantique
- ODEMA: outils décisionnels économique pour la maintenance de l'actif
- PSR-IA: positionnement en Intelligence artificielle

**→ Projets à haute valeur technologique ou la connaissance est un élément central**

# Plan

## 1. Contexte Hydro-Québec

## 2. Base de connaissance

- Technologie traditionnelle
- Technologie sémantique
  - Définition
  - Force et faiblesse

## 3. Combiner données sémantique et non sémantique

- Pourquoi
- Cas d'utilisation

## 4. Résultats et conclusion

# Connaissance et informatique

## > Définition connaissance:

- Données → information → connaissance

## > Connaissance élément central de l'entreprise:

- Pérenniser, partager, utiliser, combiner, etc.
- Capitaliser sur la connaissance

## > Connaissance informatique:

- Connaissance procédurale:
  - Comment faire, traitements (Java, C++, .net, etc.)
- Connaissance déclarative:
  - Information factuelle (Prolog, SQL, XML, RDF, etc.)

# Base de connaissance & informatique

## > Exemple de bases de connaissance en informatique :

- Une base de données avec des règles\traitement représente une base de connaissance:
  - **BD relationnelle** (Ex: Oracle)
- Une Ontologie du domaine représente une base de connaissance:
  - **BD sémantique** (RDF, OWL)



# BD relationnelle VS sémantique

## > **BD relationnelle:**

- La donnée est hautement structurée (peu de duplication, liens de composition entre les données, données tabulaires)
- Monde fermé
- Haute performance

## > **BD sémantique:**

- Donnée peu structurée (duplication, plusieurs types de liens prédéfinis)
- Monde ouvert
- Pouvoir d'inférence
- Performance moyenne

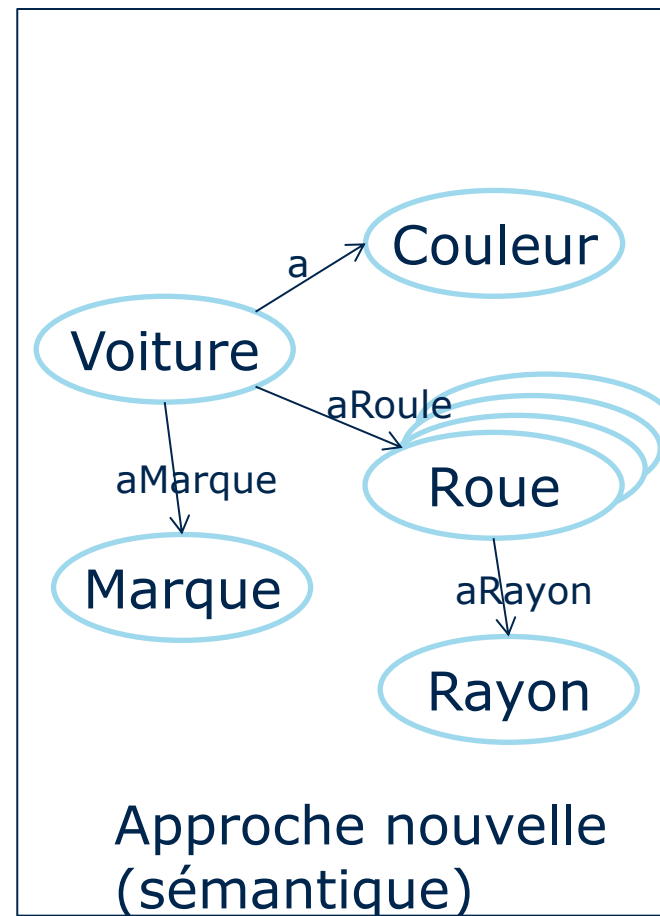
# BD relationnelle VS sémantique

Approche traditionnelle  
(relationnelle)

Voiture  
ID-Voiture  
Marque  
Couleur  
...



Roue  
Rayon  
Largeur  
Hauteur  
ID-Voiture



# Industrie et Sémantique

## Technologie sémantique et industrie électrique: enjeux?

# Sémantique: force et faiblesse

- > **Modéliser facilement un réseau... électrique**
- > **Fusionner différentes données/connaissances**
- > **Inférer de nouvelles connaissances (ex: raisonnement à base de cas)**
- > **Décrire facilement les données et métadonnées**
- > **Attention: performance moyenne (vs SQL)**

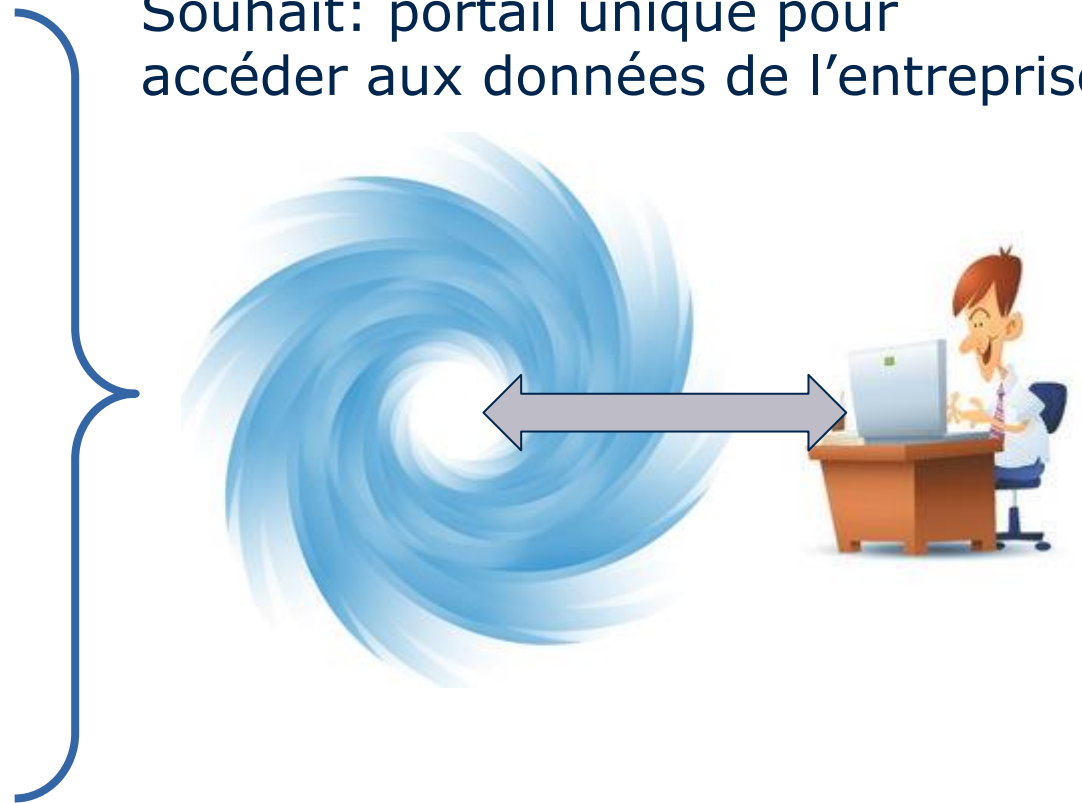
# Plan

1. Contexte Hydro-Québec
2. Base de connaissance
  - Technologie traditionnelle
  - Technologie sémantique
    - Définition
    - Force et faiblesse
3. **Combiner données sémantique et non sémantique**
  - Pourquoi
  - Cas d'utilisation
4. Résultats et conclusion

# Nos données sont distribuées



Souhait: portail unique pour accéder aux données de l'entreprise



# Approche sémantique

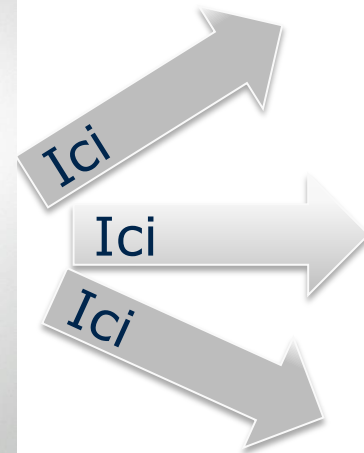
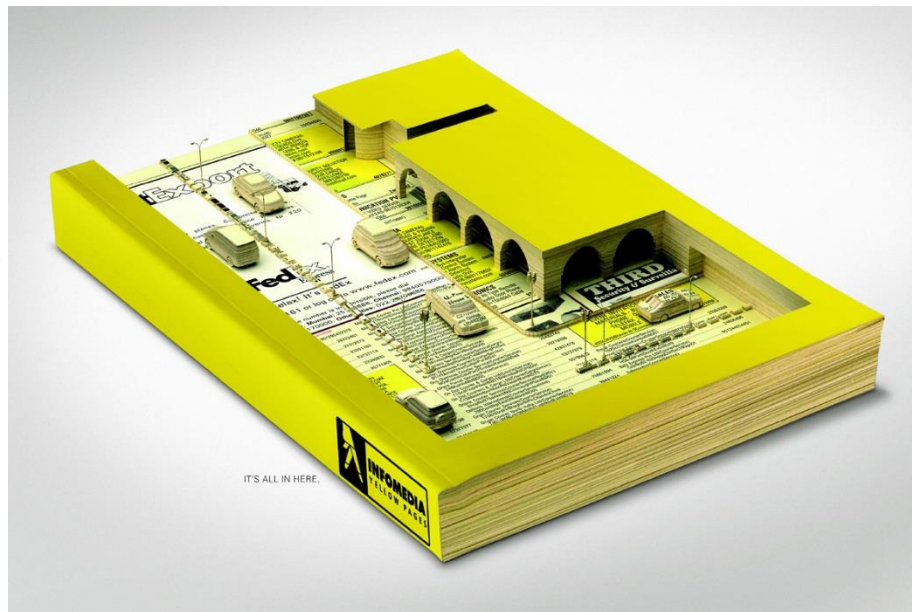


*Utiliser la sémantique pour décrire  
les **données** et **métadonnées***

# La sémantique pour supporter les données distribuées

- > **Garder les données dans leur systèmes de base de données respectifs (si possible)**
  - Car c'est performant!
- > **Utiliser la technologie sémantique pour décrire où se trouve la donnée**

Demande données





# Mode de fonctionnement

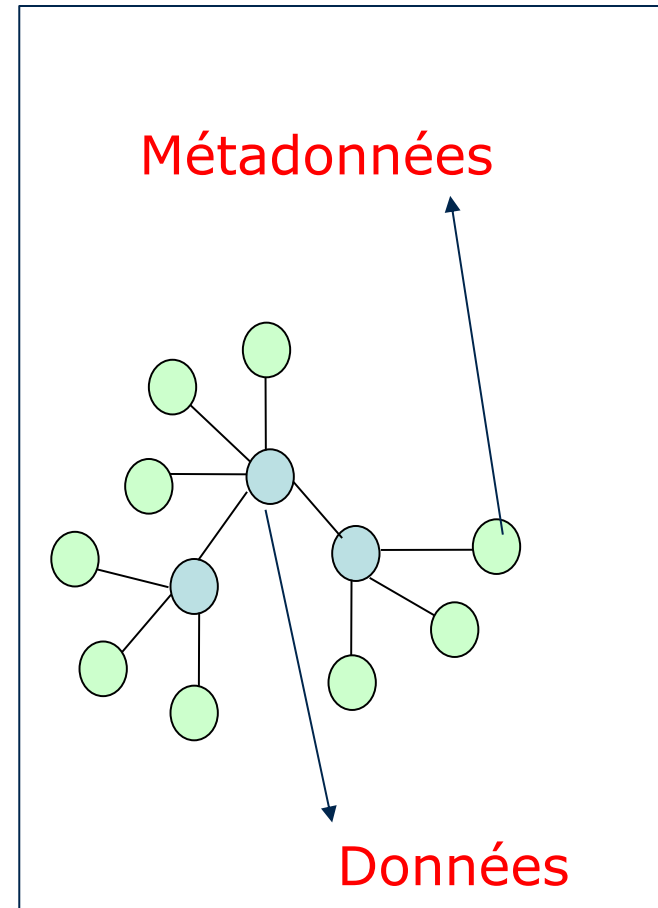
## Source de données multiples:

- Relationnel: Mysql, Oracle
- NoSQL: PI
- Fichier: XML, ...
- ...

Introduire dans les métadonnées **sémantique** les paramètres de connexion REST:

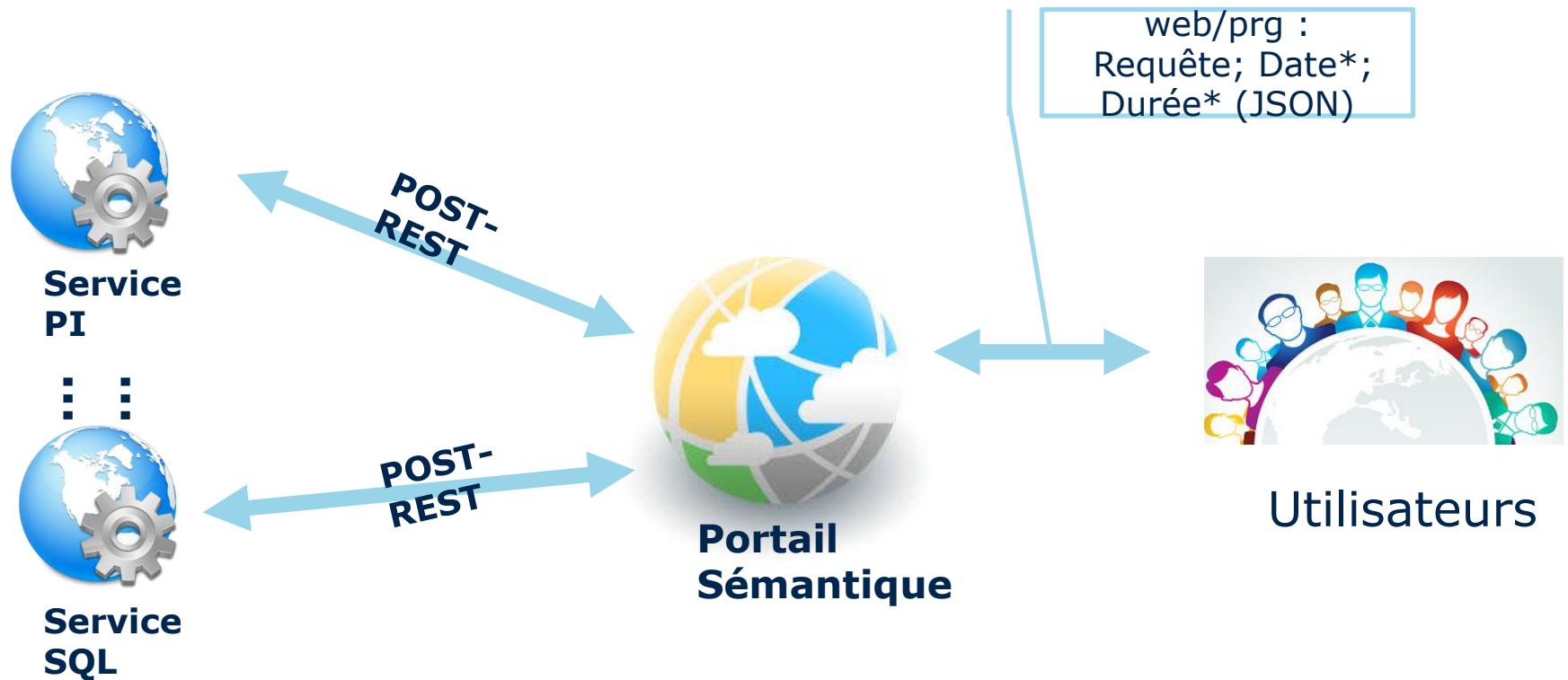
- URL du service
- Clés
- Attributs

**REST permet de garder les métadonnées simples!**

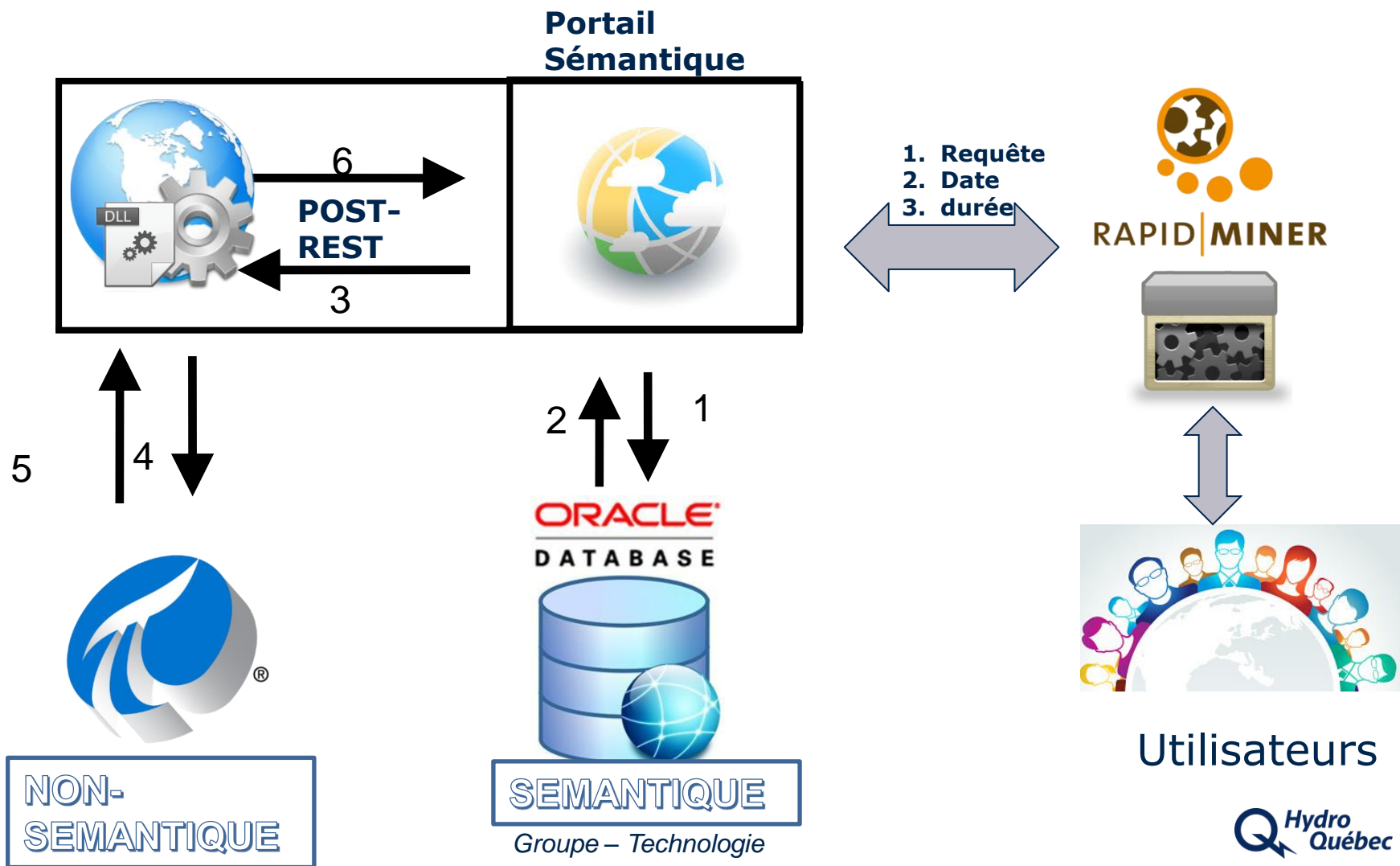


Approche sémantique

# Architecture: vue haut niveau



# Implémentation



# Interface graphique

Parameters Context

Extract\_OST

Host: jdbc:oracle:thin:@

Username: o

Model: OS

Password: \*\*\*\*\*

Begin date: 2014.06.24 11:22:26 -0400

End date: 2014.06.25 11:22:26 -0400

Query: Build SQL Query...

SEMANTIQUE

NON SEMANTIQUE

Result Overview ExampleSet (Extract\_OST)

Data View Meta Data View Plot View Advanced Charts Annotations

ExampleSet (194924 examples, 0 special attributes, 6 regular attributes)

Row No.	subject	usagePoint	intervalBlock	intervalRea...	val	time
194895	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	247.8679	6/23/2014 5:
194896	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	246.1445	6/23/2014 10:
194897	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	247.3711	6/23/2014 4:
194898	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	244.1064	6/22/2014 11:
194899	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	247.5195	6/23/2014 5:
194900	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	245.2929	6/23/2014 10:
194901	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	241.8662	6/23/2014 9:
194902	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	243.957	6/23/2014 6:
194903	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	244.0312	6/22/2014 7:
194904	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	244.625	6/22/2014 2:
194905	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	244.5507	6/22/2014 3:
194906	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	242.5468	6/23/2014 6:
194907	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	243.7343	6/22/2014 7:
194908	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7	243.8828	6/22/2014 8:
194909	SEI-G7M7M	2474765181	2474765181	ima_SEI_G7		

# Plan

1. Contexte Hydro-Québec
2. Base de connaissance
  - Technologie traditionnelle
  - Technologie sémantique
    - Définition
    - Force et faiblesse
3. Combiner données sémantique et non sémantique
  - Pourquoi
  - Cas d'utilisation
4. Résultats et conclusion

# Résultats et conclusion

- > **Hydro-Québec est à l'avant-garde des technologies sémantique**
  - Publications et présentations
  - Travaux cités par Oracle
- > **Développer un portail unique pour accéder à la données**
  - Rapidité d'exécution
  - Requête transparente coté client
  - Code réutilisable et polyvalent (REST)
- > **Réduire le « *gap* » entre technologie sémantique et non sémantique**
  - Résistance au changement
  - Performance moyenne du sémantique