



## Le Web Sémantique:

une solution fédératrice pour le traitement de l'information

### Sébastien LABORIE

Soutenance en vue de l'obtention de l'Habilitation à diriger des Recherches

Spécialité Informatique

Jeudi 29 août 2024





### Mon parcours académique & scientifique

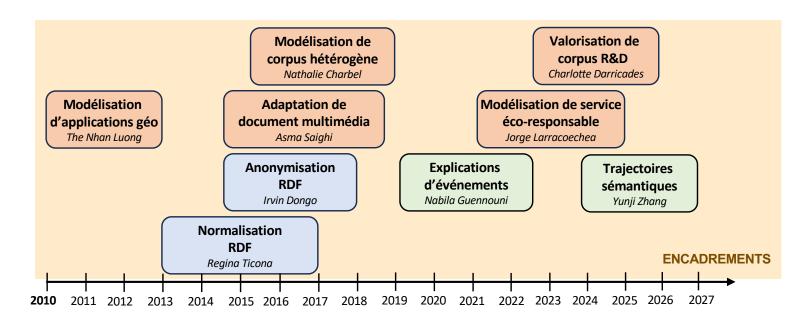
#### **ROSCO B&V Landes MOANO** NobaSearch BISE2 **PROJETS** ANR **BQR** Industriel UPPA / E2S Région **Depuis 2004: IWAAT MOJITO** 102 publications Modélisation de Valorisation de Région Région corpus hétérogène corpus R&D Nathalie Charbel Charlotte Darricades 50 comités de lecture Modélisation de service Modélisation Adaptation de 34 comités d'organisation d'applications géo document multimédia éco-responsable The Nhan Luona Asma Saighi Jorge Larracoechea 3 codirections de thèses **Anonymisation Explications Trajectoires Évaluation HCERES 2020 RDF** d'événements sémantiques Irvin Dongo Nabila Guennouni Yunji Zhana RIPEC C3 en 2022 **Normalisation RDF ENCADREMENTS** Reaina Ticona 2025 2026 2027 2021 2022 2023 2004 2005 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2024 2006 2007 2008 2009 Maître de conférences Hors Classe **Doctorat Postdoc** Maître de conférences Voie nationale - CNU 27 MINRIA LIUPPA TUT BAYONNE PAYS BASQUE Directeur des études Resp. équipe de Recherche **ADMINISTRATION** XITEA2 Chef de IUT BPB - GEA LIUPPA - T2I Département Chargé de Mission Chargé de la Com' IUT BPB - GEA IUT BPB - Recherche IUT BPB - GEA Resp. Spé. Info Commission RTI IUT BPB – LP Numérique ADIUT Membre du Bureau de l'ACD GEA **Comité Experts Comité Experts** UPPA - Section 27 UPPA - Section 27



### Le Web Sémantique : mon fil conducteur

#### > Trois catégories de contributions scientifiques :

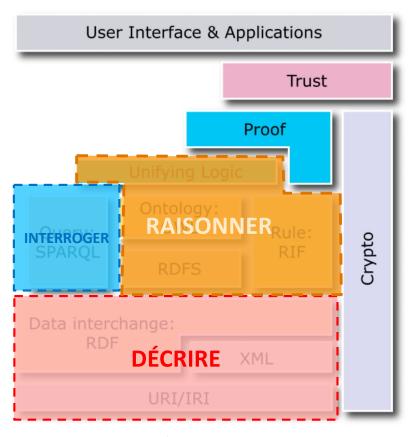
- 1 Les contributions centrées sur les technologies du Web Sémantique.
- Les contributions qui exploitent les technologies du Web Sémantique pour renforcer un domaine applicatif dédié.
- Les contributions qui permettent de fédérer des données de différentes communautés au travers des technologies du Web Sémantique.



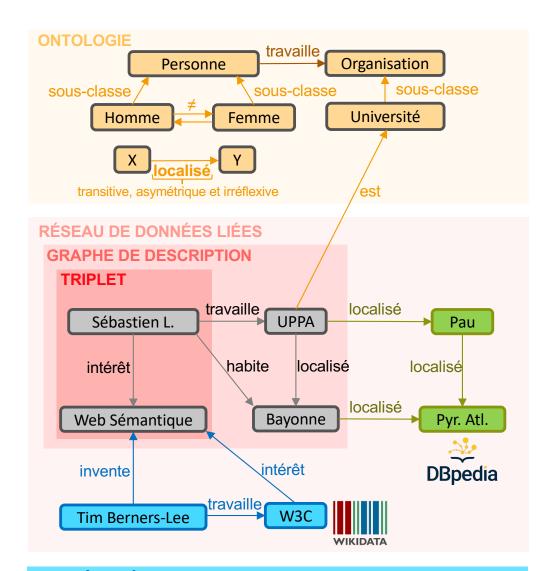


### Qu'est-ce que le Web Sémantique ?









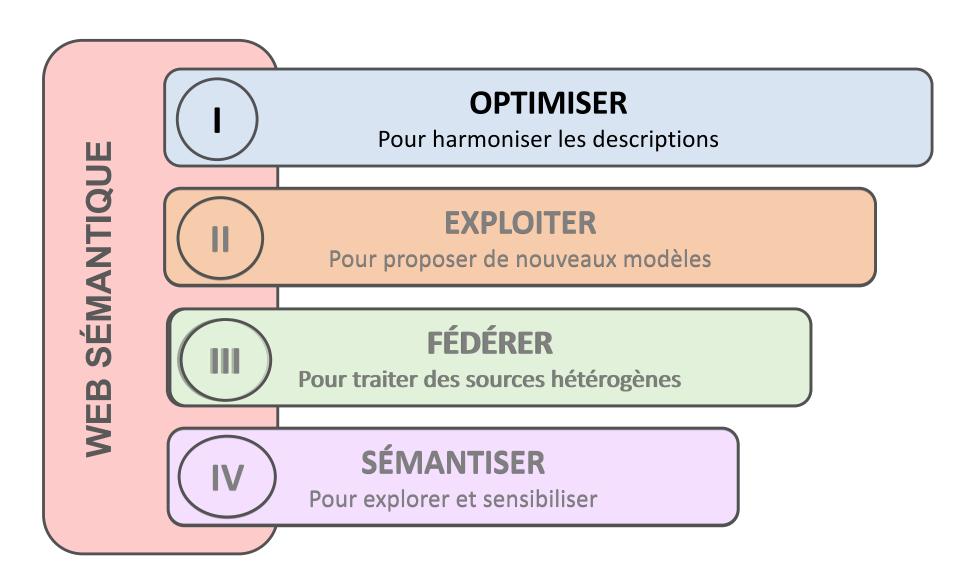
#### **REQUÊTE SÉMANTIQUE**

Quelles sont les organisations dans la description ?





### **PLAN**





#### > Contexte:

- Production de descriptions sémantiques par différents acteurs.
- Possibilité d'agréger ces descriptions entre elles.

#### > Problèmes :

- Flexibilité de spécification et d'encodage des descriptions.
- Impact sur le stockage, le chargement et les traitements.

#### > Hypothèses :

- Identifier différentes règles de nettoyage d'une description.
- Appliquer selon les besoins les règles de traitement.

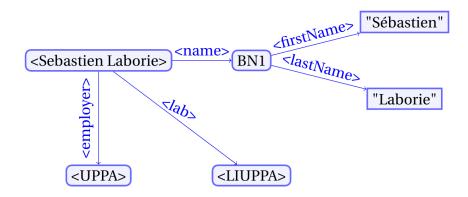
#### > Contribution:

Une normalisation de descriptions RDF (structure + sérialisation) en fonction de critères fournis par un utilisateur.



#### > Un exemple de flexibilité d'écriture d'une description RDF :

**❖** Différentes sérialisations RDF/XML possibles pour une même description



#### Sérialisation 1

**Optimiser** 

```
<?xml version="1.0"?>
2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"</pre>
     xmlns:dc="http://purl.org/dc/elements/1.1/"
     xmlns:ex="http://example.org/"
     xmlns:ex2="http://example.org/">
     <rdf:Description rdf:about="ex:Sebastien-Laborie">
          <ex:name>
              <rdf:Description rdf:nodeID="BN1">
                  <ex:firstName>Sebastien</ex:firstName>
                  <ex:lastName>Laborie</ex:lastName>
              </rdf:Description>
         </ex:name>
          <ex:employer>
              <rdf:Description rdf:about="ex:UPPA" />
          </ex:employer>
          <ex2:lab>
              <rdf:Description rdf:about="ex2:LIUPPA" />
          </ex2:lab>
      </rdf:Description>
20 </rdf:RDF>
```

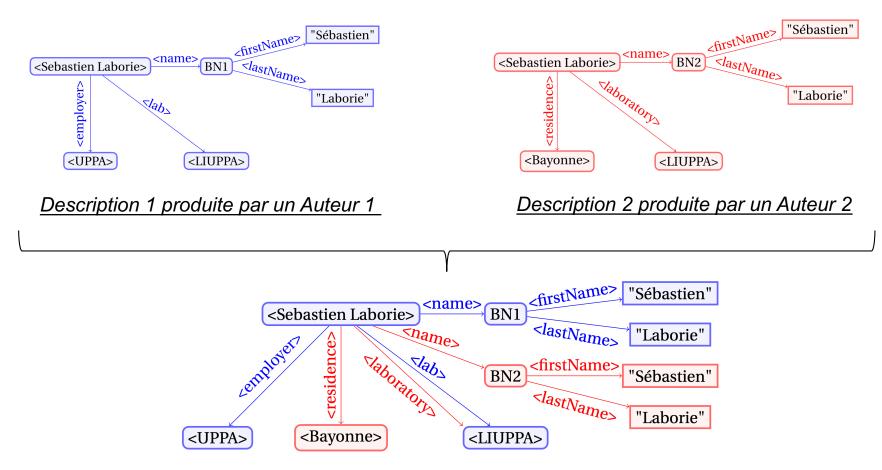
#### Sérialisation 2



#### Un exemple de motivation vis-à-vis de la structure :

**Optimiser** 

Combiner des descriptions RDF peut provoquer des redondances d'informations



Description agrégée avec redondances

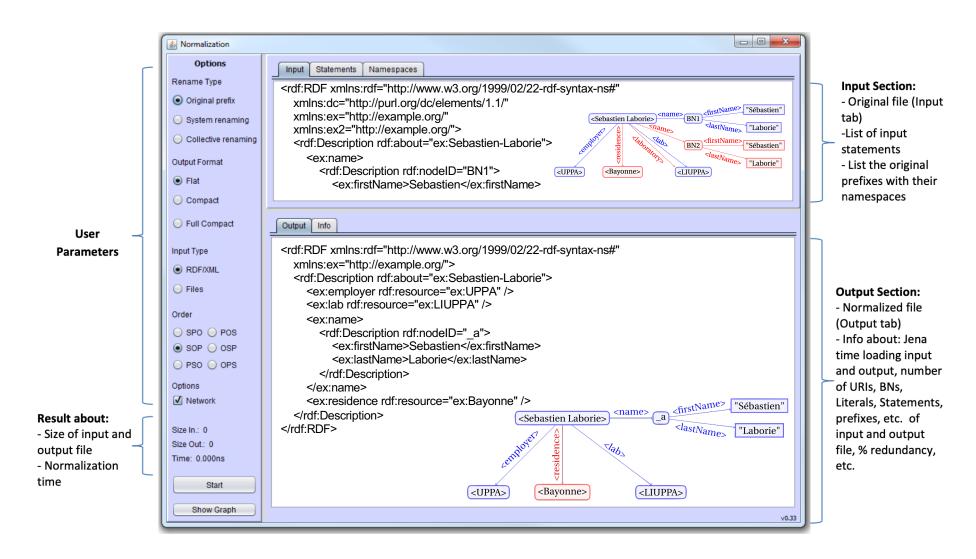




#### Regina Paola Ticona Herrera (2013 – 2016) : RDF2NormRDF

Co-encadrement avec R. Chbeir et J. Tekli

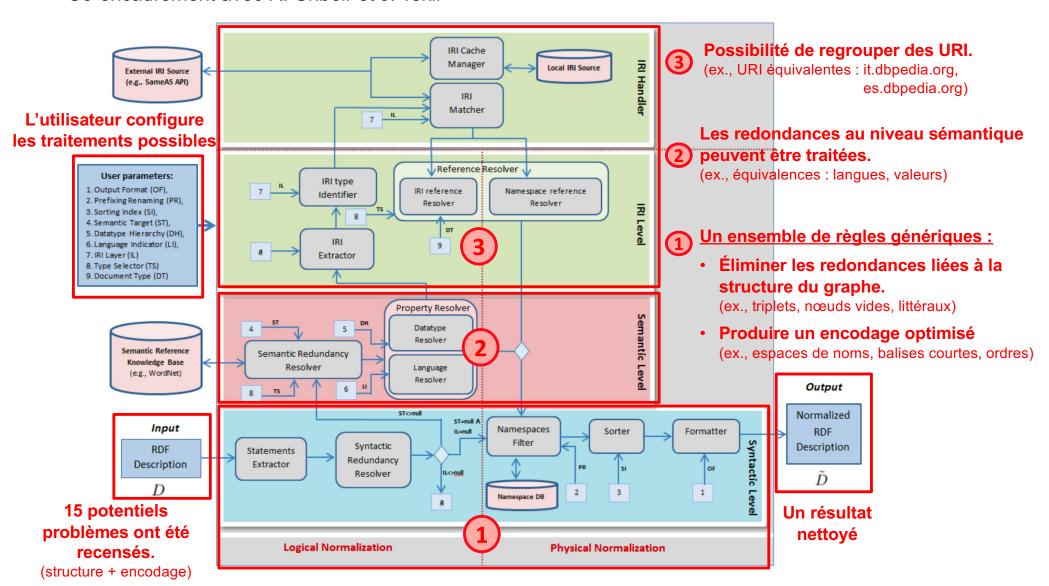
**Optimiser** 





#### Regina Paola Ticona Herrera (2013 – 2016) : RDF2NormRDF

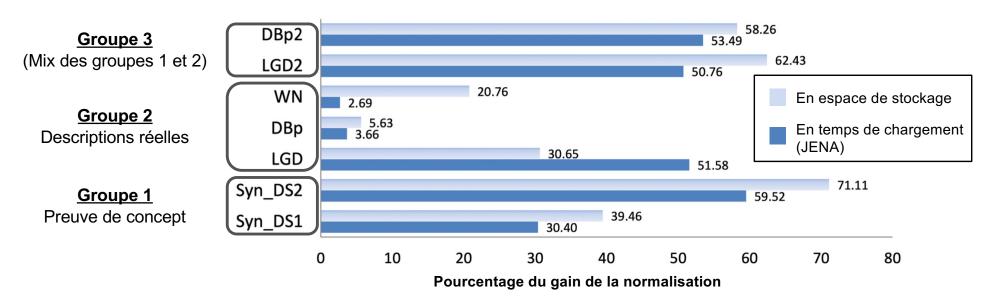
Co-encadrement avec R. Chbeir et J. Tekli



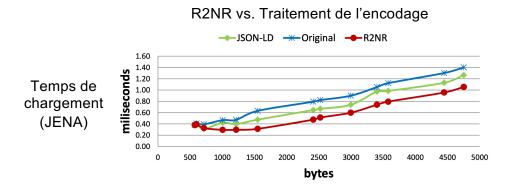


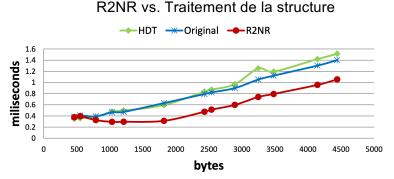
### Expérimentations RDF2NormRDF (R2NR) :

#### Optimisation du volume et du chargement des descriptions



#### Comparaison avec d'autres approches existantes







# **SYNTHÈSE**



#### **OPTIMISER**

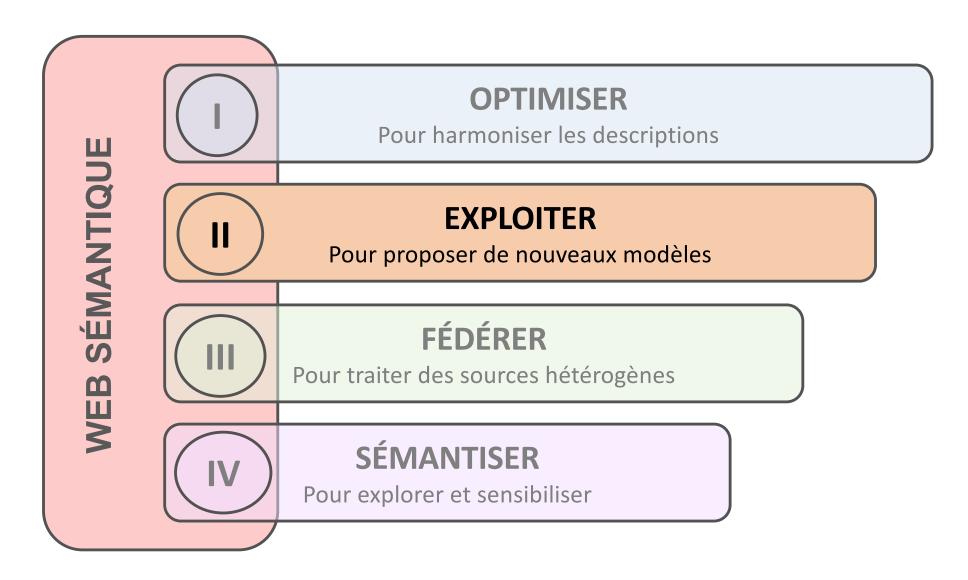
Pour harmoniser les descriptions

- > Un regard critique sur une technologie de base du Web Sémantique : RDF
- > Une normalisation complète de descriptions RDF (structure + sérialisation)
- > Un processus de normalisation qui s'adapte en fonction d'un contexte cible
- > Des résultats qui démontrent l'efficacité de notre proposition
- Une publication de rang A :

Regina Ticona-Herrera, Joe Tekli, Richard Chbeir, **Sébastien Laborie**, Irvin Dongo et Renato Guzman. *Towards RDF Normalization*. In Proceedings of the 34th International Conference on Conceptual Modeling (**ER**), LNCS 9381, Springer International Publishing, pp. 261–275, 2015.



### **PLAN**



#### **Contexte:**

- L'équipe T2i traite de la donnée de son extraction jusqu'à sa valorisation.
- > Différents domaines applicatifs métiers sont considérés (ex., tourisme, bâtiment).

#### Problème :

Un manque de descriptions sémantiques dans les travaux de l'équipe.

#### Hypothèse:

Le Web Sémantique offre une plus-value pour exprimer et exploiter toute la richesse des données des domaines applicatifs visés.

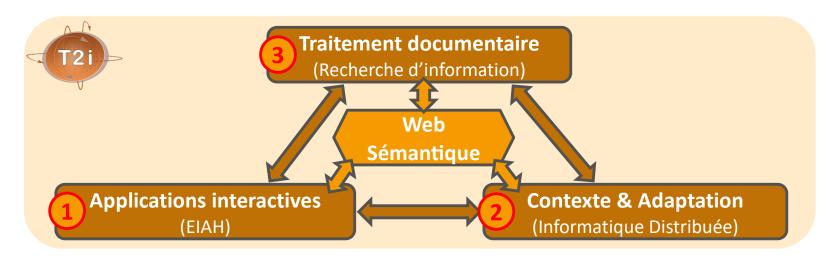
#### **Contributions:**

- Différents modèles de description dans 3 domaines scientifiques :
  - Conception d'applications interactives
  - L'adaptation à un contexte d'exécution
  - > Recherche d'information au sein d'un corpus documentaire





#### Convaincre de l'apport du Web Sémantique :



- > Des contributions dans 3 axes de recherche de l'équipe :
  - 2010 2012 : Conception d'applications interactives en géographie

    The Nhan Luong

    Co-encadrement avec T. Nodenot, C. Marquesuzaà et P. Etcheverry
  - 2014 2018 : Adaptation de documents multimédias
     Asma Saighi
     Co-encadrement avec P. Roose et N. Ghoualmi (Univ. Badji Mokhtar)
  - 2015 2018 : Modélisation de corpus de documents hétérogènes Nathalie Charbel Codirection avec C. Sallaberry







▶ Projet CG64 - The Nhan Luong (2010 – 2012) : WINDMash

Co-encadrement avec T. Nodenot, C. Marquesuzaà et P. Etcheverry

❖ Objectif : Permettre à des non-informaticiens de créer des applications interactives

#### Discovering cities and departments in France French course for foreigners. Basel Angers Tours o Cet été je souhaiterais aller au festival de Cannes. Zür Nantes Habitant à Biarritz, j'ai réservé un billet d'avion pour Bourges Lyon. Pour découvrir la Savoie et les O Cholet Sch Bouches-du-Rhône, je prendrai un bus qui passera par Châteauroux O e-sur-Yon O France Albertville et Aix-en-Provence. Mon hôtel sera aux Poitiers alentours de Cannes. La Rochelle Limoges Clermont-Ferrand Cannes NICE Saint-Etienne Brive-la-Gaillarde Biarritz PAU Bordeaux Bergerac 0 LYON Lyon Albertville CHAMBERY Aix-en-Provence MARSEILLE Montauban Alb Plateforme de conception Toulon Savoie Bouches-du-Rhône Contenu Interface Interaction Scale = 1:7M



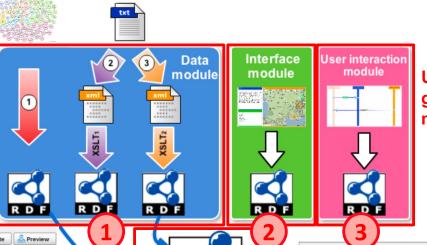
### ➤ Projet CG64 - The Nhan Luong (2010 – 2012) : WINDMash

Co-encadrement avec T. Nodenot, C. Marquesuzaà et P. Etcheverry



- Des services et des opérateurs qui exploitent la sémantique.

- Possibilité d'intégrer d'autres informations issues du Web de données liées (LOD).

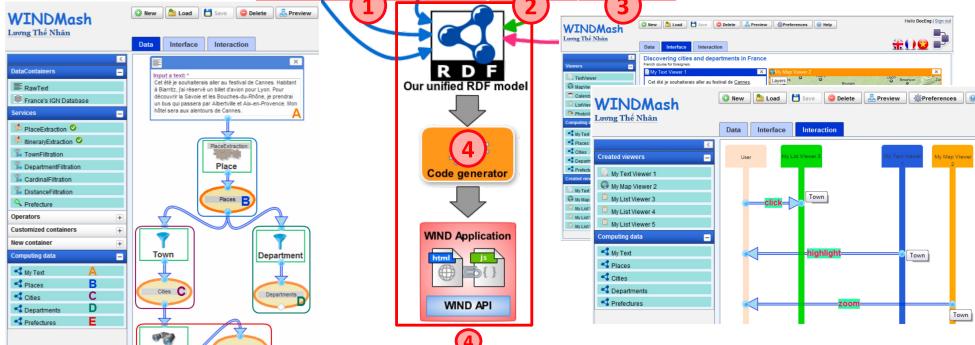


**(2)** 

Une description en RDF de l'interface graphique qui fait référence de façon modulaire aux données identifiées.



Une vérification de la cohérence des interactions définies







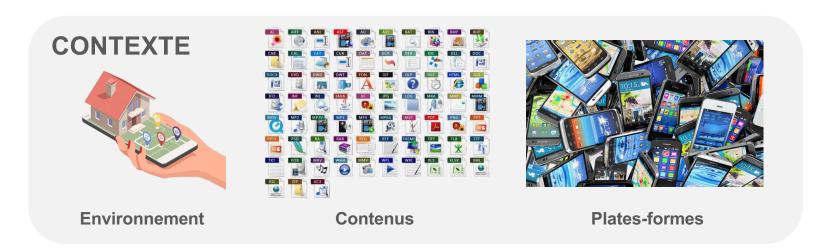
Prefecture







- Projet PNR Algérie Asma Saighi (2014 2018) : HaMA
  - Co-encadrement avec P. Roose et N. Ghoualmi
  - \* Objectif : Adapter des documents multimédias selon un contexte d'exécution











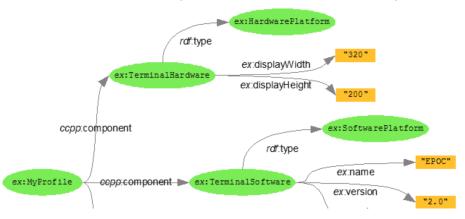


- Projet PNR Algérie Asma Saighi (2014 2018) : HaMA
  - Co-encadrement avec P. Roose et N. Ghoualmi

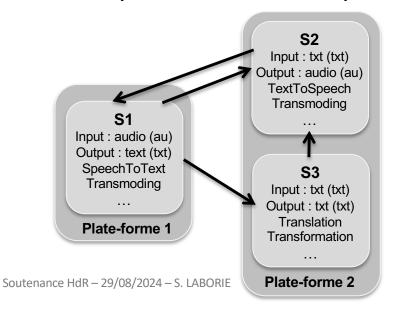
Optimiser

Objectif: Adapter des documents multimédias selon un contexte d'exécution

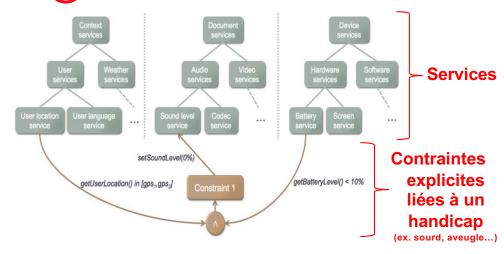
#### Profil CC/PP (Recommandation du W3C)



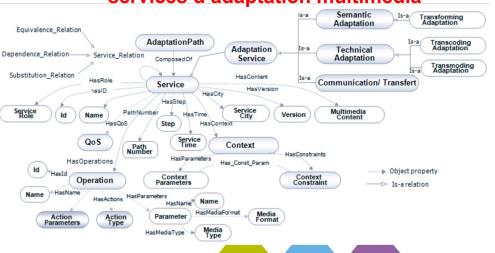
#### Composition des services d'adaptation



### Une nouvelle modélisation des profils

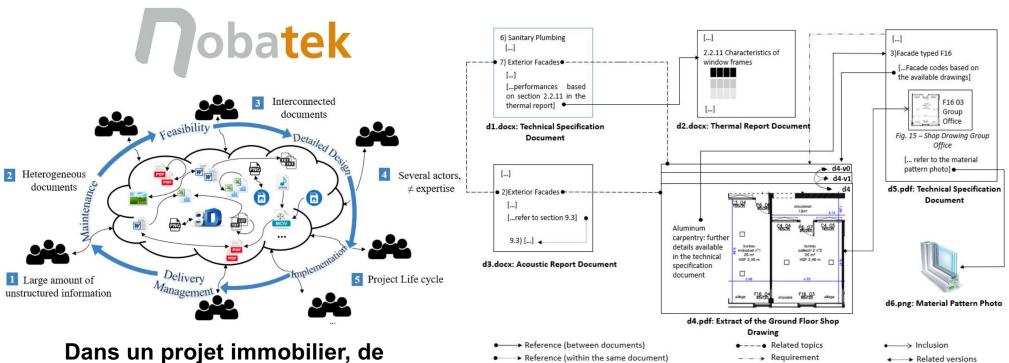


ASQ: une ontologie qui décrit les services d'adaptation multimédia





- Projet industriel Nathalie Charbel (2015 2018): FEED2SEARCH Codirection avec C. Sallaberry
  - Objectif: Faciliter la recherche d'information dans des documents hétérogènes



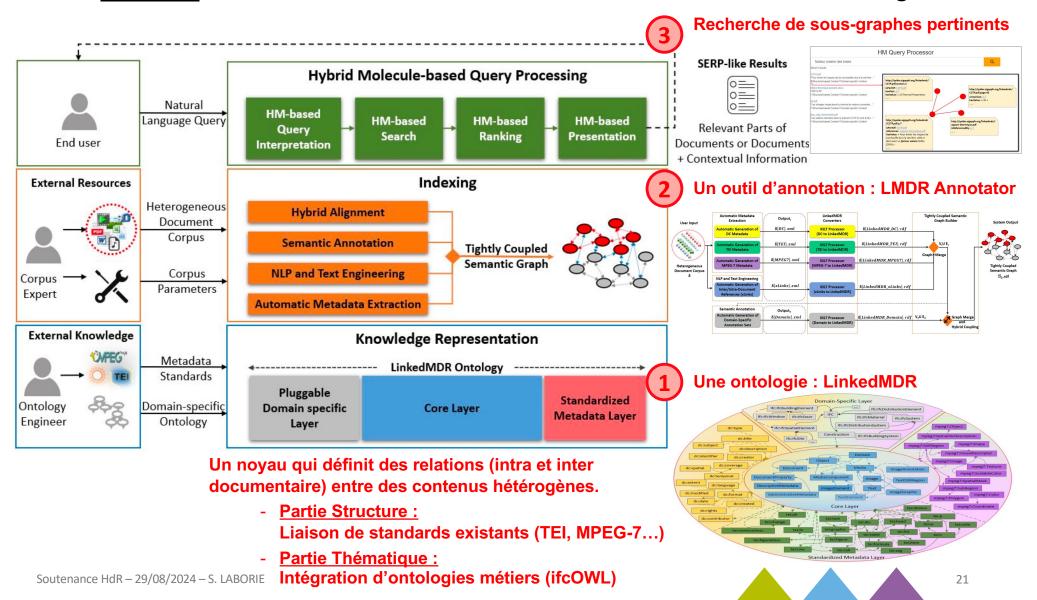
multiples acteurs collaborent avec des documents hétérogènes



Je désire des informations au sujet des façades extérieures d'un projet immobilier.



- Projet industriel Nathalie Charbel (2015 2018): FEED2SEARCH Codirection avec C. Sallaberry
  - \* Objectif: Faciliter la recherche d'information dans des documents hétérogènes





# **SYNTHÈSE**



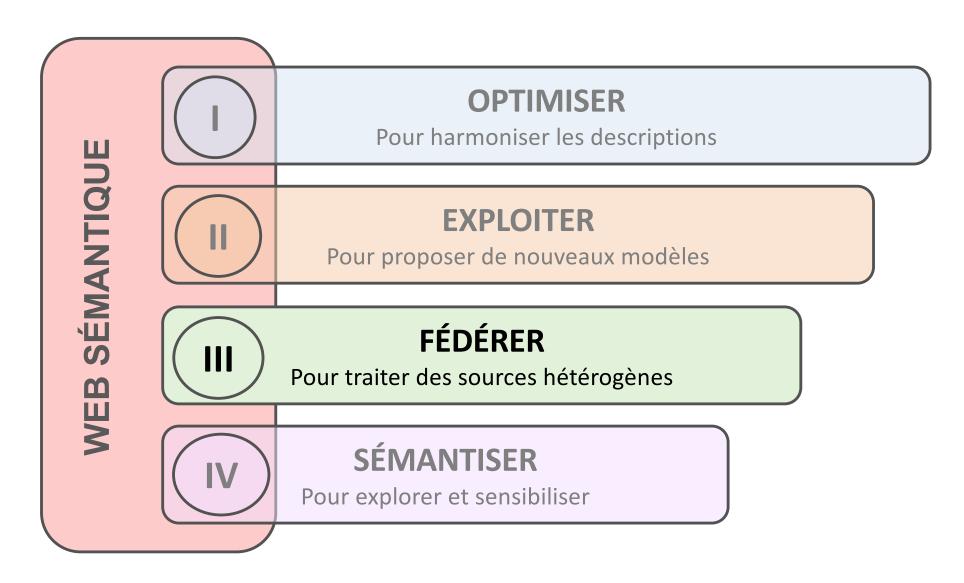
#### **EXPLOITER**

Pour proposer de nouveaux modèles

- > Des contributions dans les trois axes de recherche de l'équipe T2i :
  - (1) EIAH, (2) Informatique distribuée et (3) Recherche d'information
- Des prototypes réalisés dans le cadre de thèse pour chaque contribution
- Le Web Sémantique a offert une plus-value aux différents travaux de l'équipe Cela a permis de lever des verrous scientifiques dans différentes communautés
- > Publications dans différentes communautés :
  - ➤ EIAH : 2 conférences (DocEng + FDSE) + 2 journaux
  - Informatique distribuée : une trentaine de publications
  - Recherche d'information : 3 conférences (SAC + DEXA + INFORSID) & 1 journal (Q2)



### **PLAN**



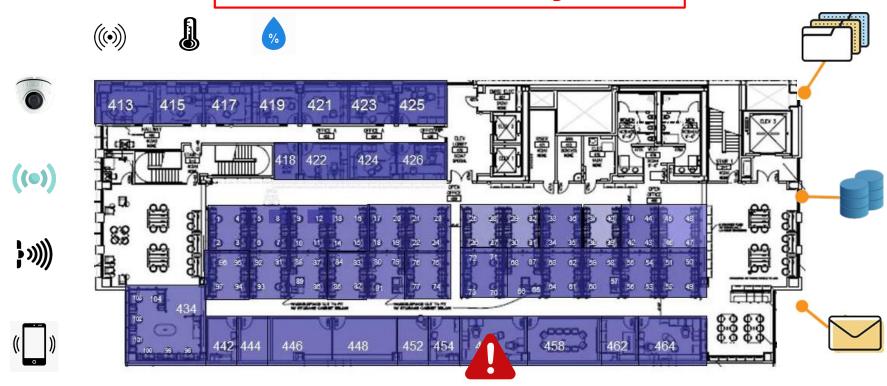
Depuis 2020, le LIUPPA porte un projet autour des Systèmes Cyber-Physiques

Fédérer

> Projet E2S/UPPA : Solutions pour l'Énergie et l'Environnement en renforçant notre savoir-faire en Recherche d'Information dans le domaine du bâtiment.

### **Contexte:**

### **Environnement hybride**



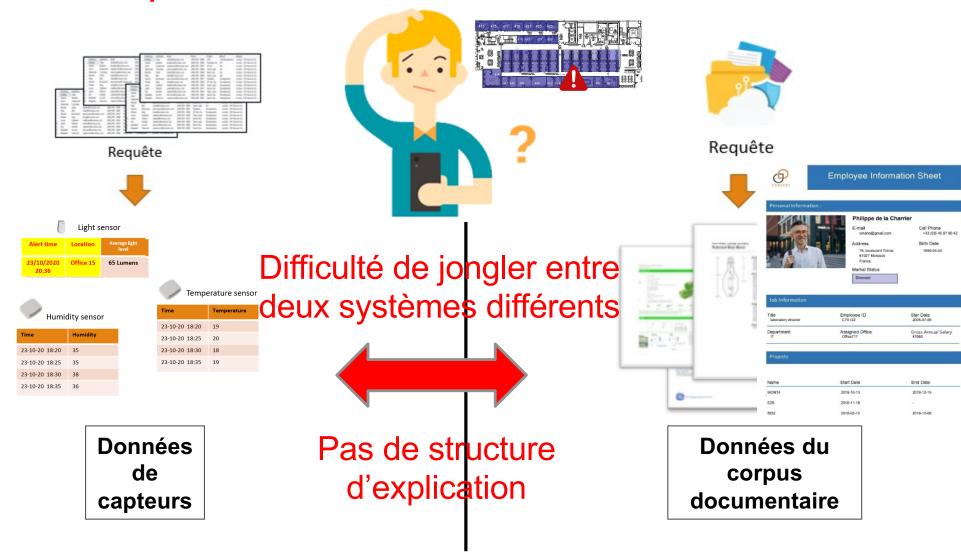
Données de capteurs

Événement gaspillage de lumière

Données du corpus documentaire



**Défi : > Comment guider un utilisateur pour lui fournir des pistes** d'explications?



#### Contexte:

- Environnement hybride = Réseau de capteurs + corpus documentaire.
- Des événements peuvent se déclencher.

#### > Problèmes :

- Difficulté pour un utilisateur de bâtir de lui-même des pistes d'explications.
- Pas de structure permettant de définir un cadre d'explication.

#### > Hypothèses:

- Une représentation sémantique permet de lier des données issues de capteurs et de documents.
- ➤ Le 5W1H issu des systèmes questions-réponses en RI offre un cadre que l'on propose d'adapter pour expliquer un événement qui se déclenche.

#### > Contribution:

> Un processus de recherche d'explications 5W1H qui exploite un graphe de description avec des données issues d'un environnement hybride.







#### Projet E2S/UPPA - Nabila Guennouni (2019 – 2022) : ISEE

Codirection avec C. Sallaberry

Adapter l'approche 5W1H pour structurer les explications d'événements au sein d'un environnement hybride.



#### Un exemple d'une référence du 5W1H en Recherche d'Information :

F. Hamborg, C. Breitinger, M. Schubotz, S. Lachnit and B. Gipp. *Extraction of main event descriptors from news articles by answering the journalistic five W and one H questions*. Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries, JCDL 2018, Fort Worth, TX, USA, June 03-07, 2018, pp. 339–340.





### > Projet E2S/UPPA - Nabila Guennouni (2019 – 2022) : ISEE

Codirection avec C. Sallaberry

- ❖ Rechercher des explications 5W1H à l'aide d'une approche par filtrage successif au sein d'un graphe de description.
- On propose de suivre le cycle de vie d'un événement

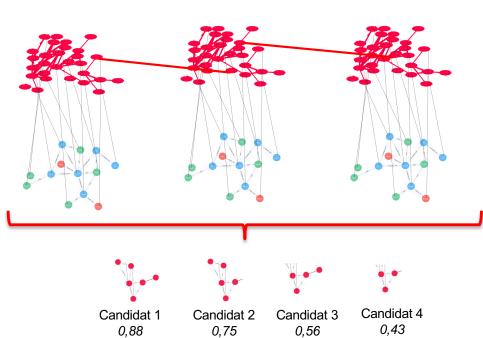


Étape 1 : Instanciation & Alignement

Étape 2 : Définition d'un événement (Filtrage des concepts)

<u>Étape 3</u>: Déclenchement d'un événement (Filtrage des instances)

Étape 4 : Sélection de candidats 5W1H

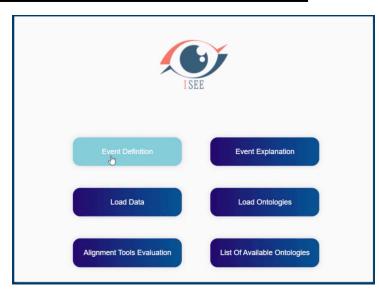




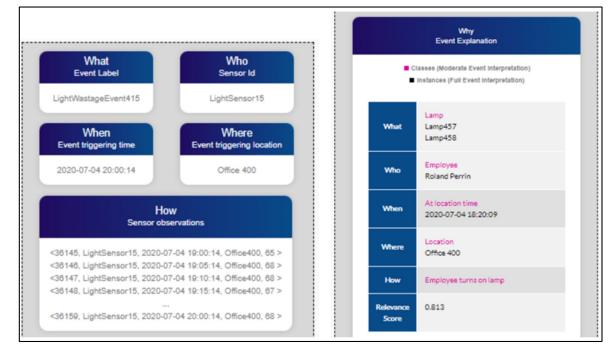


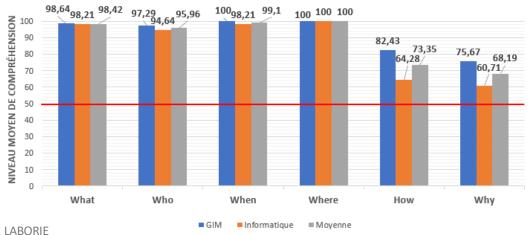
Codirection avec C. Sallaberry

### **Expérimentations:**



Un exemple de réponse pour expliquer le déclenchement d'un événement









# **SYNTHÈSE**



- > Contribution à un domaine de recherche récent (RI réseau de capteurs).
- ➤ Le Web Sémantique a permis de fédérer des données provenant d'un environnement hybride qui mêle des données issues de capteurs avec un corpus documentaire hétérogène.
- 4 publications (3 conférences & 1 journal de rang Q1) :

[2020] Nabila Guennouni, Christian Sallaberry, **Sébastien Laborie** et Richard Chbeir. *A Novel Framework for Event Interpretation in a Heterogeneous Information System*. In Proceedings of the 12th International Conference on Management of Digital EcoSystems (**MEDES**), Nov. 2020, Virtual Event United Arab Emirates, France. pp.140-148.

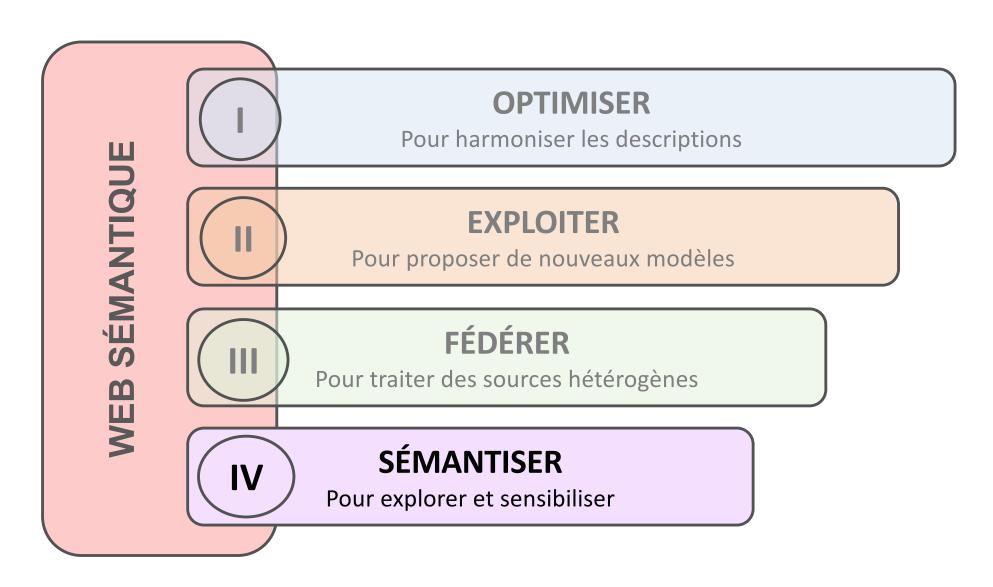
[2021] Nabila Guennouni, Christian Sallaberry, **Sébastien Laborie** et Richard Chbeir, Elio Mansour. *ISEE: A heterogeneous information system for event explainability in smart connected environments. Internet of Things.* 

[2022] Nabila Guennouni, **Sébastien Laborie**, Christian Sallaberry, Richard Chbeir, Elio Mansour. *ISEEapp: An Event Explanation Prototype bridging the gap between sensor network and document corpora data.* In Proceedings of the 18th International Conference on Intelligent Environments (**IE**), Jun. 2022, Biarritz, France.

[2023] **Sébastien Laborie**, Nabila Guennouni, Christian Sallaberry. *ISEE : Un système pour l'explication d'événements dans les environnements hybrides*. Congrès National de la Recherche des IUT (CNRIUT), Jun. 2023, Saint Pierre (Réunion), France



### **PLAN**



#### Travaux en cours & futurs : (1) Explorer le thème du Bien Vivre & du Bien Vieillir

#### > Contexte:

- Une Chaire de recherche industrielle sur le Bien Vivre et le Bien Vieillir.
- Des acteurs locaux dans les Landes souhaitent être guidés avec des données de terrain pour leurs prises de décision.

#### > Problèmes :

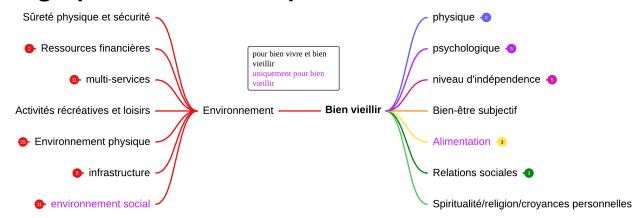
- > De multiples sources de données avec des informations variées :
  - Sources plus ou moins complètes et volumineuses.
  - Sources hétérogènes, plus ou moins structurées.
- > Différentes thématiques (ex., transport, logement, loisirs, santé, éducation)
- Différents points de vue sur un même thème.
- > Différents niveaux de granularité.

#### Hypothèse :

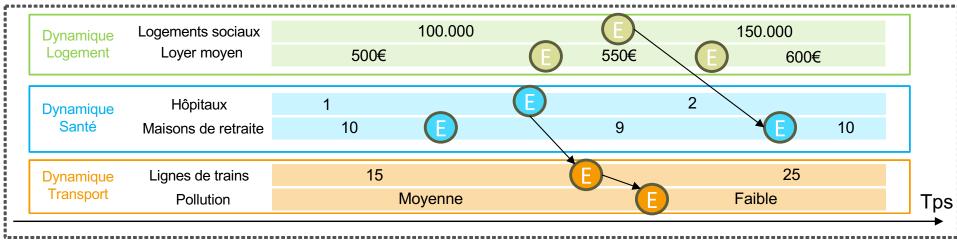
Un modèle sémantique qui exploite le temps et l'espace permettrait d'intégrer, d'analyser et de visualiser la dynamique d'un territoire.



- Projet Région Nouvelle-Aquitaine Yunji Zhang (2023 2026)
  - Une cartographie des thématiques du Bien Vivre et du Bien Vieillir



Définir et exploiter un modèle sémantique de trajectoires territoriales



- Zone géographique
- Instancier ces trajectoires à partir de données ouvertes.
- Offrir des outils de visualisation de ces trajectoires.



#### \* Travaux en cours & futurs : (2) Sensibiliser à l'éco-responsabilité

#### Contexte:

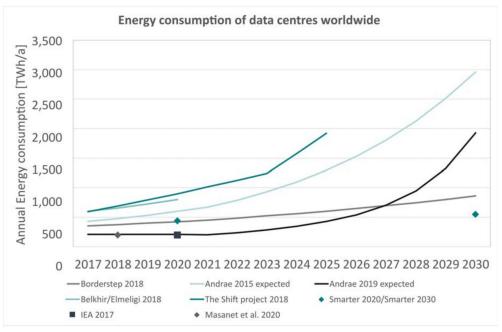
- Les analyses de données consomment des quantités d'énergie nonnégligeables.
- > L'UPPA est engagée dans une démarche éco-responsable (Label i-Site).
- > Thématique du Green-IT émergente au LIUPPA.
- > C'est une thématique de la Chaire industrielle « Bien Vivre & Bien Vieillir ».

#### > Problème :

Non-connaissance de la consommation d'énergie par les usagers.

#### > Hypothèse:

Qualifier des consommations énergétiques pour sensibiliser les utilisateurs.



Montevecchi, F., Stickler, T., Hintemann, R., Hinterholzer, S. (2020). Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market. Final Study Report. Vienna





▶ Projet E2S/UPPA - Jorge Larracoechea (2021 – 2024) : RADIANCE

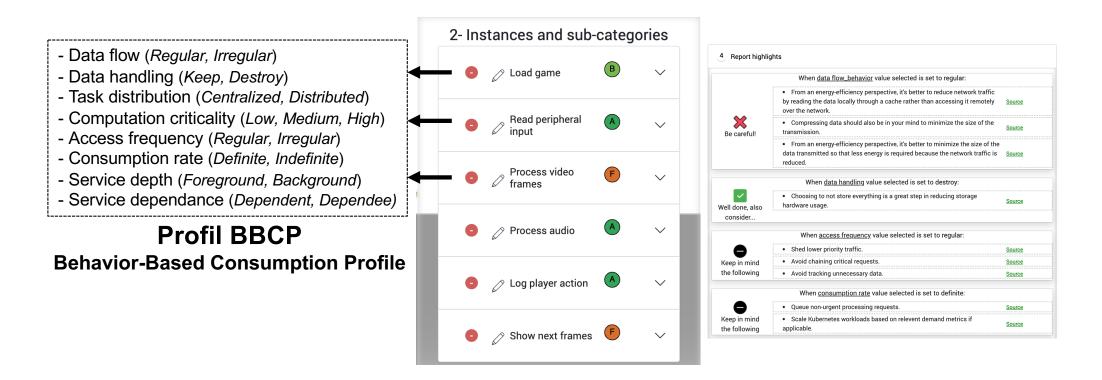
Co-encadrement avec P. Roose, S. Ilarri, Y. Cardinale



- RADIANCE = softwaRe behAvior DesIgn And eNergy Consumption assEssment (Conception logiciel et évaluation de la consommation d'énergie)
- Engendrer un profil de consommation énergétique d'une application logicielle dès sa phase de conception pour sensibiliser les développeurs.



#### ➢ Projet E2S/UPPA - Jorge Larracoechea (2021 – 2024): RADIANCE Co-encadrement avec P. Roose, S. Ilarri, Y. Cardinale



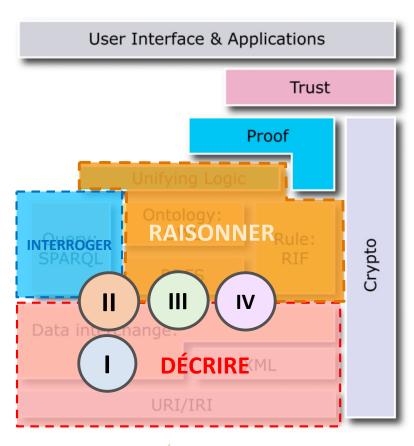
- « Sémantiser » une consommation énergétique d'une application au sein d'une approche holistique : normaliser un vocabulaire, produire une ontologie...
- Raisonner et interroger des descriptions sémantiques au sujet de l'écoresponsabilité : recommander des services alternatifs vertueux.





### CONCLUSION

#### Positionnement de mes contributions & conduite à tenir :





- **Contributions sur les fondamentaux** 
  - Prise de recul.

- **Contributions dans différents domaines** 
  - Principe gagnant-gagnant.

#### Intégration de domaines émergents

Fédération de différentes données.

(IV)

#### Prise en compte de sujets d'actualités

Ouverture scientifique.



# Le Web Sémantique : une solution fédératrice pour le traitement de l'information

#### Une communauté toujours très active :

- ❖ W3C Rec. 21/05/2024 : RDF Dataset Canonicalization.
- ❖ W3C Working Draft : RDF-star et SPARQL-star.



- ❖ Scientifiques : EIAH, Informatique distribuée, RI...
- ❖ Applicatifs : Tourisme, Bâtiment, Bien Vivre/Bien Vieillir...

#### Des collaborations humaines :

- ❖ Faire connaître et former au Web Sémantique.
- Partager notre savoir-faire et collaborer avec des acteurs variés.

