

# Application BASTRI

## Fiches Equipes

### CEDAR (SR0837TR)

Analyse et traitement de données complexes à grande échelle  
CEDAR (SR0734UR) □ CEDAR

**Statut:** Décision signée

**Responsable :** Ioana Manolescu

**Mots-clés de "A - Thèmes de recherche en Sciences du numérique - 2024" :** A3.1.1. Modélisation, représentation, A3.1.2. Gestion, interrogation et stockage, A3.1.3. Données distribuées, A3.1.6. Optimisation de requêtes, A3.1.7. Données ouvertes, A3.1.8. Données massives (production, stockage, acheminement), A3.1.9. Bases de données, A3.2.1. Bases de connaissances, A3.2.3. Inférence, A3.2.4. Web sémantique, A3.2.5. Ontologies, A3.3. Analyse de données et de connaissances, A3.3.1. Analyse en ligne, A3.3.2. Fouille de données, A3.3.3. Analyse de données massives, A3.4.1. Apprentissage supervisé, A3.4.6. Réseaux de neurones, A3.4.8. Apprentissage profond, A9.1. Connaissances, A9.2. Apprentissage

**Mots-clés de "B - Autres sciences et domaines d'application - 2024" :** B6.5. Systèmes d'information, B8.5.1. Démocratie participative, B9.5.6. Science des données, B9.7.2. Données ouvertes, B9.10. Confidentialité, vie privée

**Domaine :** Perception, Cognition, Interaction

**Thème :** Représentation et traitement des données et des connaissances

**Période :** 01/04/2018 -> 30/06/2025

**Dates d'évaluation :** 03/10/2019 ,

**Etablissement(s) de rattachement :** IP-PARIS, CNRS

**Laboratoire(s) partenaire(s) :** LIX (UMR7161)

**CRI :** Centre Inria de Saclay

**Localisation :** Centre de recherche Inria de Saclay

**Code structure Inria :** 111074-1

**Numéro RNSR :** 201622056j

**N° de structure Inria:** SR0837TR

### Présentation

Les applications à usage intensif de données de présentent un très grand degré d'hétérogénéité, dû au fait que différentes applications sont mieux servies par différents types de données. Ainsi, les applications de commerce online sont déployées au dessus de bases de données relationnelles, des plateformes de gestion de contenus Web gèrent des documents semi-structurés, les capteurs fournissent des flots de données numériques, les applications scientifiques manipulent des vecteurs, les logs (journaux) des systèmes tels que des serveurs Web ou moniteurs transactionnels produisent du texte structuré etc.; souvent, des données très hétérogènes sont exploitées sous la forme de graphes RDF etc.

Le rythme de production et de consommation des données Big Data de nos jours rend l'unification de telles sources de données hétérogènes dans une architecture unique (par les processus de Extract-Transform-Load ou ETL typiques des entrepôts de données relationnels) infaisable. L'équipe CEDAR s'attèle aux défis posés par cette variété de données produites et consommées à très grande échelle, en s'orientant sur des modèles expressifs et des outils efficaces pour le traitement de telles données. Nos outils sont conçu pour être déployés dans le cloud, afin de passer à l'échelle des très grands volumes de données.

### Axes de recherche

Nos travaux de recherche s'organisent dans deux grandes catégories, comme suit:

A l'intérieur du cloud, dans des plateformes massivement distribuées de traitement de données, nous nous intéressons à:

1. Plate-formes de stockage de données hétérogènes à très grande échelle

### Contact

- **Responsable :** Ioana Manolescu
- **Tél :** 01.72.92.59.20
- **Secrétariat Tél :** 01.74.85.42.84

### En savoir plus

- Site de l'équipe
- Site sur [inria.fr](http://inria.fr)
- Site du [responsable](#)
- Derniers Rapports d'Activité : [2016](#), [2017](#), [2018](#), [2019](#), [2020](#), [2021](#), [2022](#), [2023](#), [2024](#)

### Documents sur la structure

- [Intranet](#)
- [Privés](#)

### Décisions

- [12794](#) (03/04/2018) : création
- [14331](#) (30/07/2020) : prolongation
- [16994](#) (26/04/2024) : prolongation
- [17478](#) (27/11/2024) : prolongation

### Localisation

- **Adresse postale :** Centre de recherche Inria de Saclay  
Campus de l'École Polytechnique - Bâtiment Alan Turing  
1 rue Honoré d'Estienne d'Orves 91120 Palaiseau France
- **Coordonnées GPS :** 48.714, 2.206

## 2. Algorithmes efficace pour répondre aux requêtes sur des données sémantiques

A l'interface entre le cloud et les utilisateurs, nous visons à définir des nouveaux modes d'interaction et exploration des données par les utilisateurs, en particulier en équipant les serveurs de bases de données de nouvelles fonctionnalités pour simplifier l'expérience des utilisateurs. On s'intéresse en particulier à:

4. L'interrogation exploratoire des grands graphes sémantiques
5. Réponses représentatives aux requêtes sémantiques

### Relations industrielles et internationales

A l'étranger, nous collaborons avec: UCSD (Alin Deutsch), AT&T (D. Srivastava), U. Madison Wisconsin (D. DeWitt) and U. Berkeley (M. Franklin), U. Bolzano (D. Calvanese).

Nos partenaires industriels actuels sont Alibaba, Swift, RadioFrance and Le Monde.