

# Service de Cache pour les Grilles de Calcul <sup>1</sup>

Yonny Cardenas, Jean-Marc Pierson et Lionel Brunie  
LIRIS, CNRS UMR 5205  
INSA de Lyon, Bât. B. Pascal, 7 av. Jean Capelle, 69621 Villeurbanne cedex,  
France  
<prenom>.<nom>@liris.cnrs.fr  
<http://liris.cnrs.fr>

**Résumé.** Nous proposons un système fédérateur de caches pour les grilles que les applications de la grille utilisent comme un service de cache uniforme. Le système est fondé sur le concept de l'activité de données où les applications partagent et réutilisent l'information sémantique liée à l'activité des données sous la forme de métadonnées. Ces métadonnées représentent la connaissance sur les données et sur leur gestion. Elles permettent d'optimiser, suivant le contenu et l'utilisation de ces données, leur placement, leur recherche, leur durée de vie et leur pertinence vis-à-vis de leur exploitation.

## 1 Introduction

Les grilles de calcul sont la conséquence de l'évolution des systèmes distribués. Leur principale caractéristique est la capacité à réunir dynamiquement des ressources distribuées pour offrir des moyens de calcul et de stockage à grande échelle. Dans ce cadre, plusieurs institutions peuvent établir une « organisation virtuelle » à partir de ces infrastructures informatiques (Foster et al. 2003). Ainsi, les grilles de calcul mettent en oeuvre la technologie permettant l'intégration et le partage de données entre différentes organisations.

Les mécanismes d'accès aux données qui sont disponibles aujourd'hui dans les intergiciels de grille OGSA-WSRF (Globus 2005) sont assez rudimentaires pour le traitement et la gestion de données complexes, ce qui est le cas par exemple dans le domaine médical, domaine cible du projet RagTime <sup>1</sup>. Cela amène potentiellement à une prolifération de copies des données alors même que les mécanismes d'utilisation des métadonnées sur la structure, la disponibilité et la localisation de ces données ne sont pas assez avancés. La gestion des métadonnées sur la grille souffre de l'absence de mécanismes de haut niveau pour le partage d'information. Ce partage permettrait aux applications d'exploiter efficacement (en particulier pour des applications d'extraction des connaissances) la disponibilité de certaines données présentes sur la grille. Le service de cache pour grilles présenté ici a comme double objectif de permettre un accès global aux données dispersées dans la grille et d'optimiser l'usage des ressources de données (stockage et communication des données).

---

<sup>1</sup> Ce travail est soutenu par la Région Rhône-Alpes (project RagTime), et l'ACI Masse de Données.