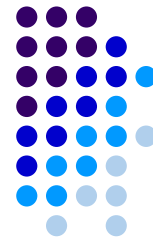


グリッドにおける 研究情報データベースの 統合・連携

小島 功
産総研グリッド研究センター
kojima@ni.aist.go.jp



National Institute of Advanced Industrial Science and Technology



概要その1

- グリッド研究センターの研究
 - O G S AベースのDB関連ツール・ミドルウェア
 - 大規模クラスタ上でのDB処理
- 産総研研究データベースの技術監理
 - 産総研の有する研究データベースの統合(組織内)、
- つくばWANデータ共有委員会
 - つくば地域の公開データベースの連携・統合



National Institute of Advanced Industrial Science and Technology



概要

- **グリッドにおけるデータベース(DB)の統合・連携**
 - 現状と方向
 - 研究課題
- **産総研でのDB統合・連携**
 - 基本的な考え方
 - (研究)開発内容紹介x2



National Institute of Advanced Industrial Science and Technology



グリッドにおけるDB処理の背景

- **計算の大規模化・分散化**
データの大規模化・分散化と密接に関連する。
- **グリッド上での大規模データ応用**
いわゆるデータ・グリッド応用の発展あるいは一部データベースのスケールアウトが、必要不可欠である。
さまざまな形でDBMSがグリッドのツールや開発環境、仕様に埋め込まれつつある。



National Institute of Advanced Industrial Science and Technology



グリッドでのDB統合・連携における、 いずれ解決されようとしているが、 現在はそれなりの課題



1. いわゆるグリッドの機能をどう実現するか。
DBにおける分散・並列処理・リソース管理など
2. いわゆる大規模・分散計算処理と、どう結びつけるか。
分散計算環境・ツールと、DB処理の連携・統合
3. GGFを中心とした標準化や、Globusを中心としたオープンな開発プロジェクトとどう連携するか
特に、OGSAで代表されるグリッドサービス基盤との連携・統合



National Institute of Advanced Industrial Science and Technology



グリッドにおけるデータベース統合

1



- 分散データベース技術の発展形としての、DBのグリッド
 - 分散問い合わせ処理
 - クラスタリング
 - 分散トランザクション
 - リソース管理・最適化
- 既存のDB技術をベースに、完成度の高い分散環境を提供できる
- × 他のDBプロダクト、分散計算ツールなどとの連携



National Institute of Advanced Industrial Science and Technology



グリッドにおけるデータベース統合 2



- GGF (Global Grid Forum)を中心とした標準化仕様
 - OGSi (Open Grid Service Infrastructure)などで成果
- DAIS (Database Access and Integration Service)
 - OGSAベースでのデータベース連携・統合
オープンプラットフォーム・かつ標準になる可能性。
Uk-OGSA-DAIなど開発プロジェクトとの密接な連携。
 - × 仕様がまだ議論中
 - × DBの周辺処理(例えばファイル処理とか)で、他のWG/RGなどとの調整も必要。



National Institute of Advanced Industrial Science and Technology



課題と現状の方向性(まとめ)



DBMSのグリッド対応

- 分散DB技術の延長としてのグリッド機能
 - 完成度が高いが、プロダクト依存性が高くなりかねない。
- GGFをベースとした標準化仕様
 - (OGSAベースでは)まだ議論すべき点が多い。
 - Globus2.X系環境での議論はない。
 - プロトタイプあるいは参照開発プロジェクトが少ない。
OGSA-DAI程度



National Institute of Advanced Industrial Science and Technology



産総研での DB研究開発としてのアプローチ



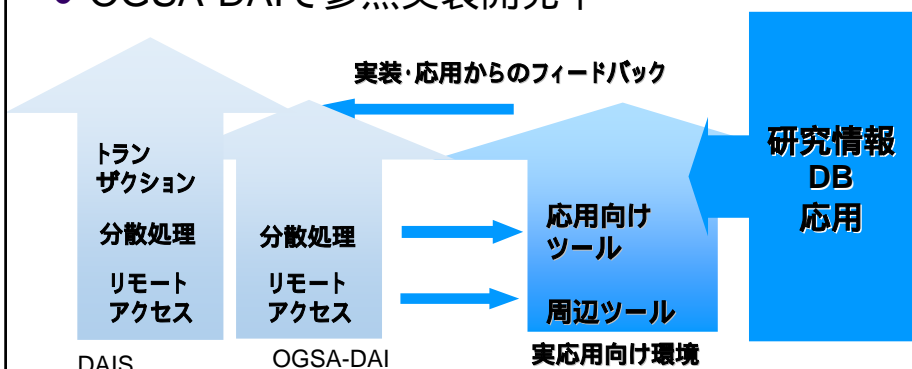
- 100%OGSAベース
Globus Toolkit 3.0
OGSA-DAI
× Globus 2.X 台
- 研究情報データベースの統合・連携
- グリッドおよびDB統合として、よい応用



National Institute of Advanced Industrial Science and Technology

OGSAへのアプローチ

- GGFで仕様検討中
- OGSA-DAIで参照実装開発中



研究情報に関連する問題に対する周辺ツールや応用ツールの開発仕様へのフィードバック

National Institute of Advanced Industrial Science and Technology

研究情報データベースとは



研究活動の結果として整備される情報や、
研究活動の効率化のために必要な情報

実験・計算データ・成果資料・アーカイブなど

- データが多様で、変化しやすい。
 - 型がデータの量に比して多く、動的
- 研究グループ独自の視点など、ボトムアップ的な組織内分散DBに近い。
 - スキーマが統一的でない。変更多い。
- 計算と密接に結びつく
 - シミュレーション結果・測定データなど



National Institute of Advanced Industrial Science and Technology



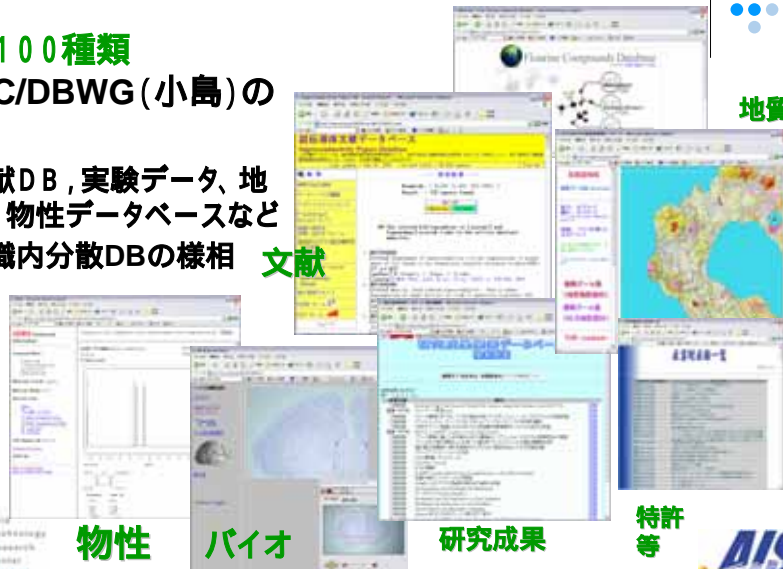
例 産総研の研究関連WWW/DB

<http://www.aist.go.jp/RIODB/> など



80-100種類
TACC/DBWG(小島)の
把握

- 文献DB, 実験データ, 地質, 物性データベースなど
- 組織内分散DBの様相 **文献**



National Institute of Advanced Industrial Science and Technology



つくば地域の 研究情報データベース

- 農林系研究機関
- 大学
- 高エネ研、防災研、物質・材料研
- 宇宙センター
などなど。
- Webで公開されているもの
 - 現在 > 800URLs (小島所属、データ共有委員会調べ)
 - 全文検索ポータルを構築中
<http://www.aist.go.jp/infobase/tsukubawan/>



研究関連情報

DBのグリッドを構築するにはよい素材



National Institute of Advanced Industrial Science and Technology

課題

- 無数のWWWデータリソースを、どうOGSAの中で統合・連携するか。
- グリッド外にあるデータベースの取り込み
 - WebDB
 - WWWサービスなど
- グリッドに持ち込まれた(ボトムアップ的に作られた)情報資源を、どう対応付け、統合するか。
 - 意味的に統合可能なリソース
 - データ型や表現形式が違う
 - スキーマ変換・統合



National Institute of Advanced Industrial Science and Technology



OGSA-DAIベースのDB統合システム開発



- 外部データベースリソースの統合システム
 - Eコマースサイトなど、DBがバックエンドにあるWebサイトやWebサービスサイトのOGSA-DAI上での統合。
- メタデータ統合検索システム
 - 複数・分散したデータベースをXML化しながら統合して、その上で異なるスキーマにまたがった変換を可能にする問い合わせ言語(XQueryのサブセット)を支援する。



プロトタイプが動作中。11月に1版の公開予定

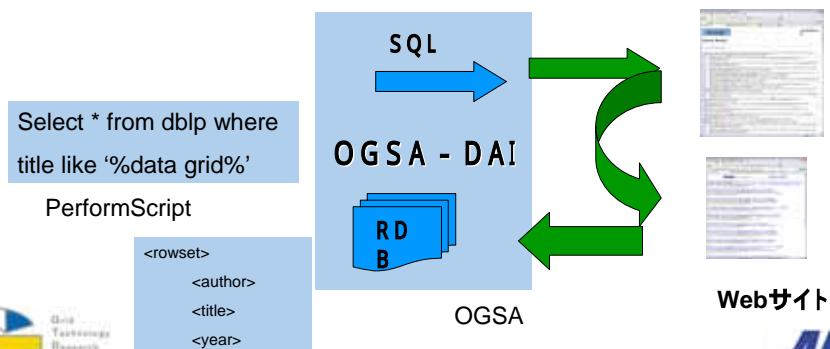


National Institute of Advanced Industrial Science and Technology

1: 外部DBの統合システム



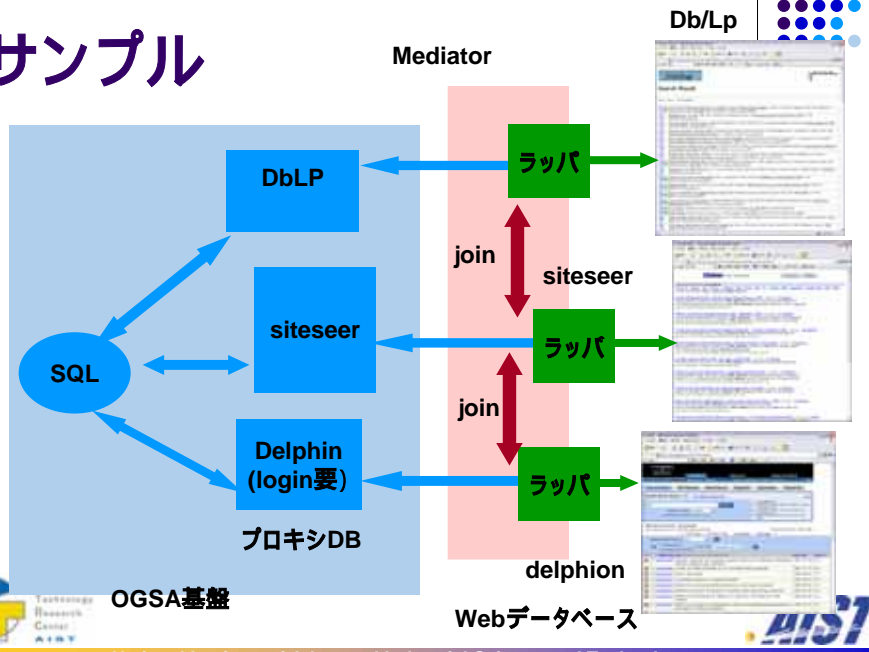
- Eコマースサイトや現在のWeb上での検索システムを、グリッド上のDBとしてアクセス可能とする。
- 各サイトに対するラッパ
- サイト間処理(Joinとか)を行うメディエータ



National Institute of Advanced Industrial Science and Technology



サンプル



OGSA基盤

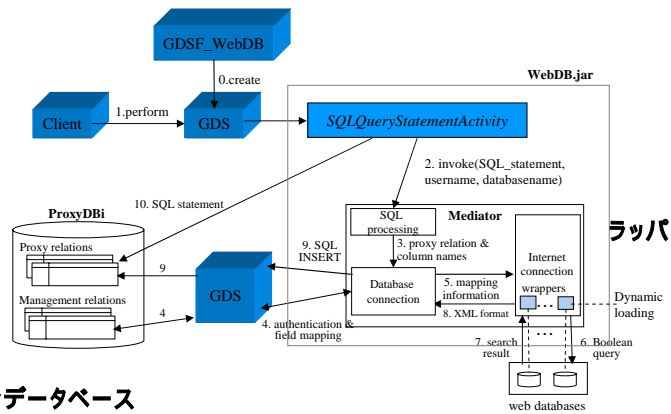
Webデータベース

National Institute of Advanced Industrial Science and Technology



構成

OGSA-DAIのSQL処理を拡張して、ラッパ・メディエータ処理と連動させる



プロキシデータベース (GDSからアクセス可能)

Webリソース



National Institute of Advanced Industrial Science and Technology



特徴

- **全部OGSA化**
 - データベースアクセス
 - データベースの定義
 - Webリソースの定義とラッパの登録
 - ラッパのクラス名など管理データベースもGDS化されている。
 - GSIサポート、外部サイトのアカウントとの対応
- **商用ラッパ開発ツールと連携(XFetch,WebIなど)**
- **プロキシDBをグリッド内に保持**
 - 結果のキャッシングと検索結果のリファイン、ローカル処理
- **パブリックドメインソフトをベース**



● GlobusToolkit3.0, OGSA-DAI3.0, MySQL, Oracle



National Institute of Advanced Industrial Science and Technology

基礎となる考え方。

- オープンプラットフォームの組み合わせ。
 - プラットフォーム独立
- ラッパ開発労力の軽減を狙う。
 - 汎用のラッパ開発ツールとの連携
- 効率的な統合というよりは、ともあれ外部リソースや異なるDBをグリッドの中に取り込む。
 - 実体(プロキシDB)をOGSA-DAI内に保持。
 - 最適化や性能では不利。
実際には、外部リソースへのアクセスが支配的で、単純な最適化でOK。
 - 統合や効率化はOGSAの中でやる。
(例えば、OracleとDB2の統合など)



● 統合や効率化はOGSAの中でやる。
(例えば、OracleとDB2の統合など)



National Institute of Advanced Industrial Science and Technology

研究課題

- 外部DBリソースと、内部のグリッドDBを含めた問い合わせ処理の効率化
 - 結合を含む演算
 - 外部DBリソースの提供する演算子とSQLとの対応
 - プラットフォーム独立を仮定するから問題が難しい。
- このツールが目的ではない。
 - OGSA-DAISプラットフォームへの移行手段の提供
 - 大量に外部のDBリソースがあった場合のリソース管理
 - 分散情報サービス(もともとのねらい)



National Institute of Advanced Industrial Science and Technology



2:メタデータ統合検索システム

- ボトムアップ的に作られた(スキーマが異なる)分散メタデータのDBをXMLDBとして統合する。

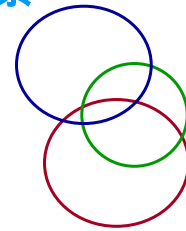


National Institute of Advanced Industrial Science and Technology



構成

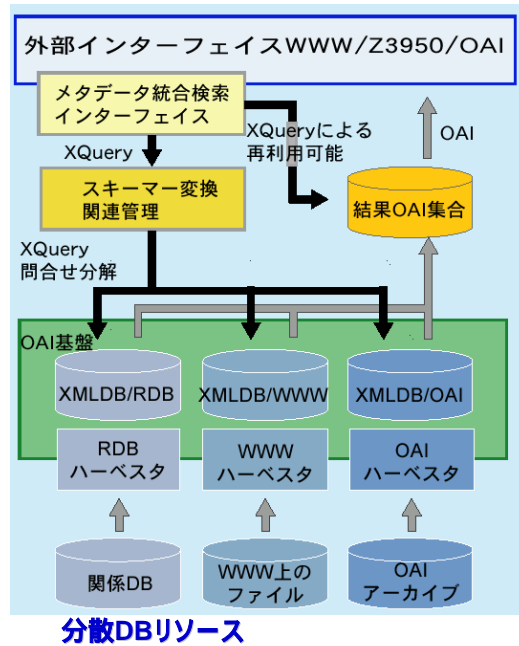
- XMLDBによる実装
 - XQueryのサブセット実装
 - Oracleをバックエンド
- 異なるスキーマに対する横断的な検索
 - 異なるXMLスキーマ
 1. 意味的に共通性の高い要素
 2. 独自のスキーマ要素
 応用の視点により、見方が異なる。
- 異種情報源
 - RDB/OAI/Webなどのハーベスタ



National Institute of Advanced Industrial Science and Technology



構成図



National Institute of Advanced Industrial Science and Technology



特徴 異なるスキーマにまたがる検索



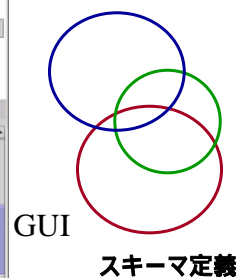
- 複数のXMLスキーマにまたがった検索
問い合わせ実行時にXQuery-XQuery変換を行う。
- 関連・対応関係の保持
名前の異なる同一データ型
著者 = Author
- 複数のタグの組み合わせへの対応
氏 | 名 = 氏名
- 簡単なデータ変換
Date型<->文字列
- XSLT変換(未実装)



National Institute of Advanced Industrial Science and Technology



メタデータスキーマ定義間の 関連・対応関係の定義・提供サービス



スキーマ要素の対応・変換関係の管理
応用に依存する 応用や分野固有の視野として提供



National Institute of Advanced Industrial Science and Technology



目標: グリッドサービスとして提供



- スキーマ間の対応・変換関係の管理サービス
 - DBリソースのXMLスキーマ登録
 - スキーマ間の対応・変換関係の登録・記述
- XQuery問い合わせ変換サービス
 - 個別設定、個別スキーマに応じた問い合わせ変換
- サービス化: 他のDB環境からも利用可能
 - 先のアプローチに基づく
 - とにかくOGSA内に取り込んで、その中で最適化・統合



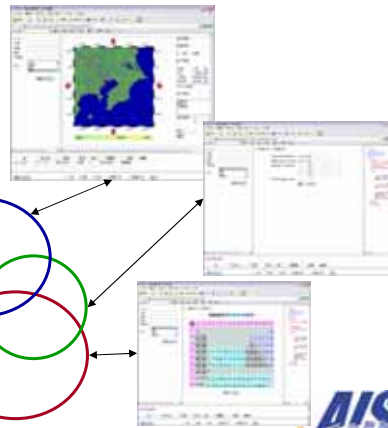
National Institute of Advanced Industrial Science and Technology



状況



- OGSA非対応版が公開中
 - 一部OGSA-DAI実験済 JDBC->OGSA-DAI
- 開発作業
 - 完全なOGSA-DAI化
 - 速度向上
 - インターフェイスの改善
 - 視野に応じた
インターフェイスの提供



視野の意味的な
対応・包含関係



National Institute of Advanced Industrial Science and Technology



研究課題

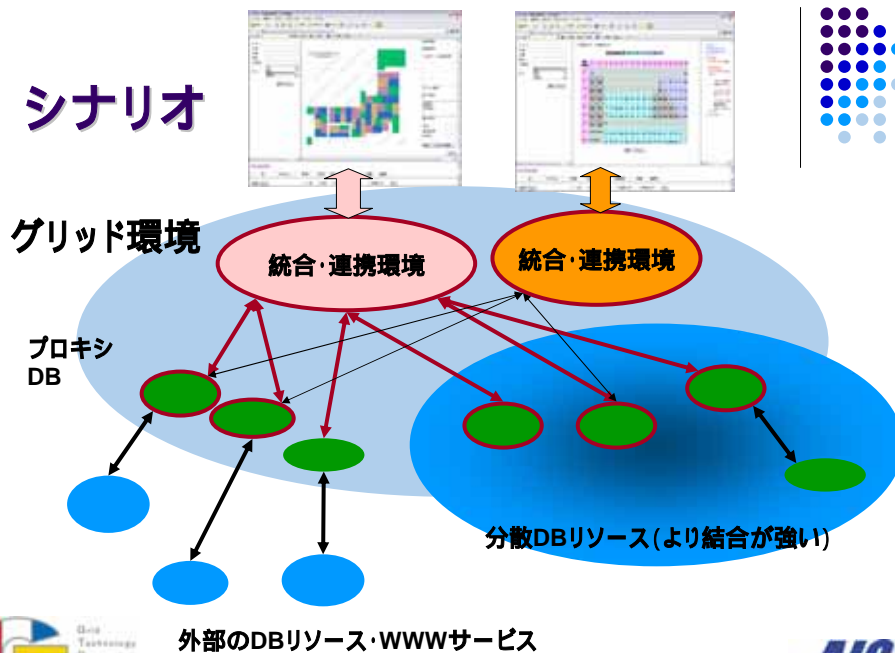
- いわゆる、メタデータプロファイルの研究
 - DublinCore, JPM.GILS, MARC, OAIなど
- 異種メタデータ間の変換・対応
 - 一種のオントロジー
 - メタデータ(それなりに意味的な共通性がある)ので、複雑な問題を扱わない。
 - 詳細化(Qualification) 要素の分解による詳細化
 - 拡張(Extension) 要素の追加
 - 型の変換



National Institute of Advanced Industrial Science and Technology



シナリオ



National Institute of Advanced Industrial Science and Technology



DAIS仕様へのフィードバック



能動性・自律性の高いデータベース仕様

- 開発・利用の経験から検討中

計算の自律性に対応した

データベース処理の能動性・自律性

- 現在のDBMSは、どれも受動的な仕様
- トランザクション処理や、イベント処理との融合



National Institute of Advanced Industrial Science and Technology



まとめ



- OGSA/OGSA-DAI周辺でのツールの研究開発
研究情報データベースの応用
- 外部DBリソースのOGSA-DAIへの統合環境
 - ラッパ・メディエータの処理
- 分散メタデータの統合システム
 - 異種のスキーマに対応する問い合わせ変換



National Institute of Advanced Industrial Science and Technology

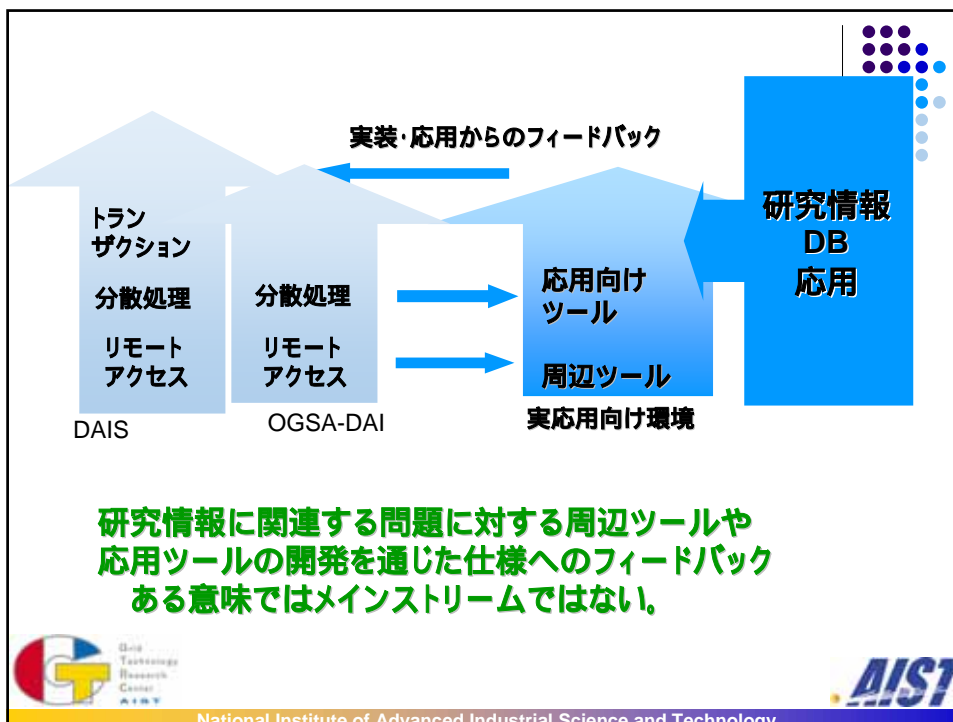


実装をベースとした研究

- 能動的・自律性の高いDB処理の記述
 - ECAルールなどの再評価
 - 特に、イベントの扱い。
- リソースが膨大な場合の管理
 - DBが数万個 それだけで一種のDBになる
 - 分散情報サービス



National Institute of Advanced Industrial Science and Technology



National Institute of Advanced Industrial Science and Technology



OGSA-DAI プロジェクト(英国)から

大学・研究所・会社 (IBM, Oracleなど) などが
加わった共同研究・開発。

- ・ ソフトウェア開発(OGSA-DAI)や研究(分散処理)を含む

日本でも、(協議会を中心に?)、
よりいっそうの連携・協働が重要

- ・ 大手 独自のロードマップと強い開発力
日本独自の研究開発・国際的な貢献
- ・ 中小 標準でオープンなプラットフォームの策定
マーケット・参入のいい機会



National Institute of Advanced Industrial Science and Technology

