



GEO Grid の情報基盤 現状と今後の展開

小島 功

産業技術総合研究所
情報技術研究部門

GEO Grid & Related Projects (おさらい)

GEO Grid

産総研内の融合(応用研究者とIT研究者)プロジェクト
衛星データ基盤を研究者に提供・運用しながら研究開発も

- 産総研TODAYGEOGrid特集
- http://www.aist.go.jp/aist_j/aistinfo/aist_today/vol09_03/vol09_03_main.html

*For Details, Please have a look our IEEE Systems Journal Paper
"Design Principles and IT Overviews of the GEO Grid" Vol2. No.3,
2008.09 Special Issue for GEOSS*

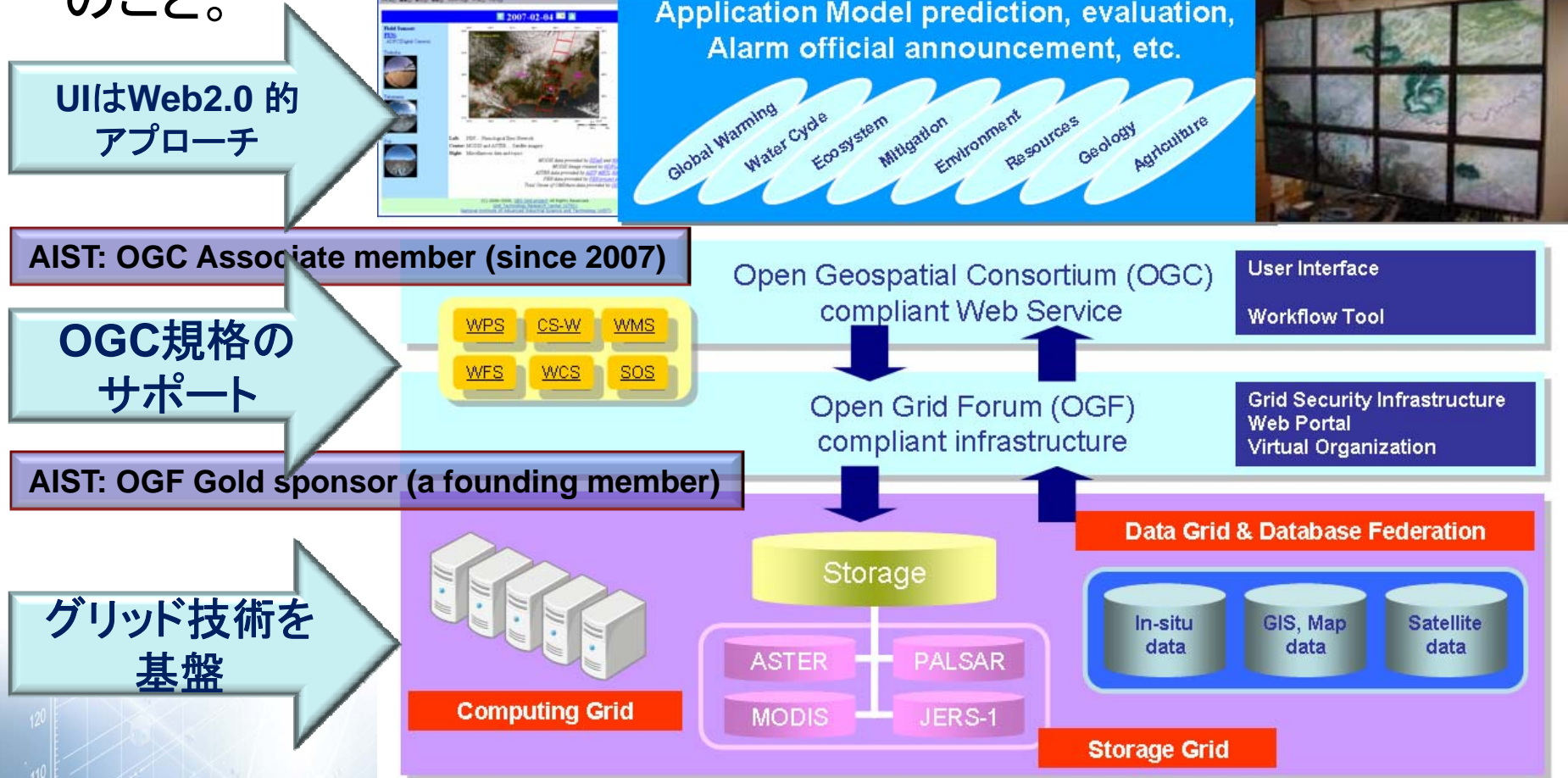
関連研究テーマ

- 文科省「次世代IT基盤構築のための研究開発」RENKEI
 - 「e-サイエンス実現のためのシステム統合・連携ソフトウェアの研究開発」
 - グリッドの環境(NAREGI)におけるEサイエンスの支援
 - 国立情報学研究所など
 - 産総研=セキュリティとDB統合のミドルウェアを担当。
- 科研「不均質なペタバイト級時空間センサーデータ基盤」
GEO Gridの周辺における挑戦的/萌芽的な研究課題
 - 筑波大&静岡大との共同研究





- 衛星画像データベース (ASTERなど) を基礎とした、e-Science基盤のこと。



GEO Grid におけるアプリ例

背景

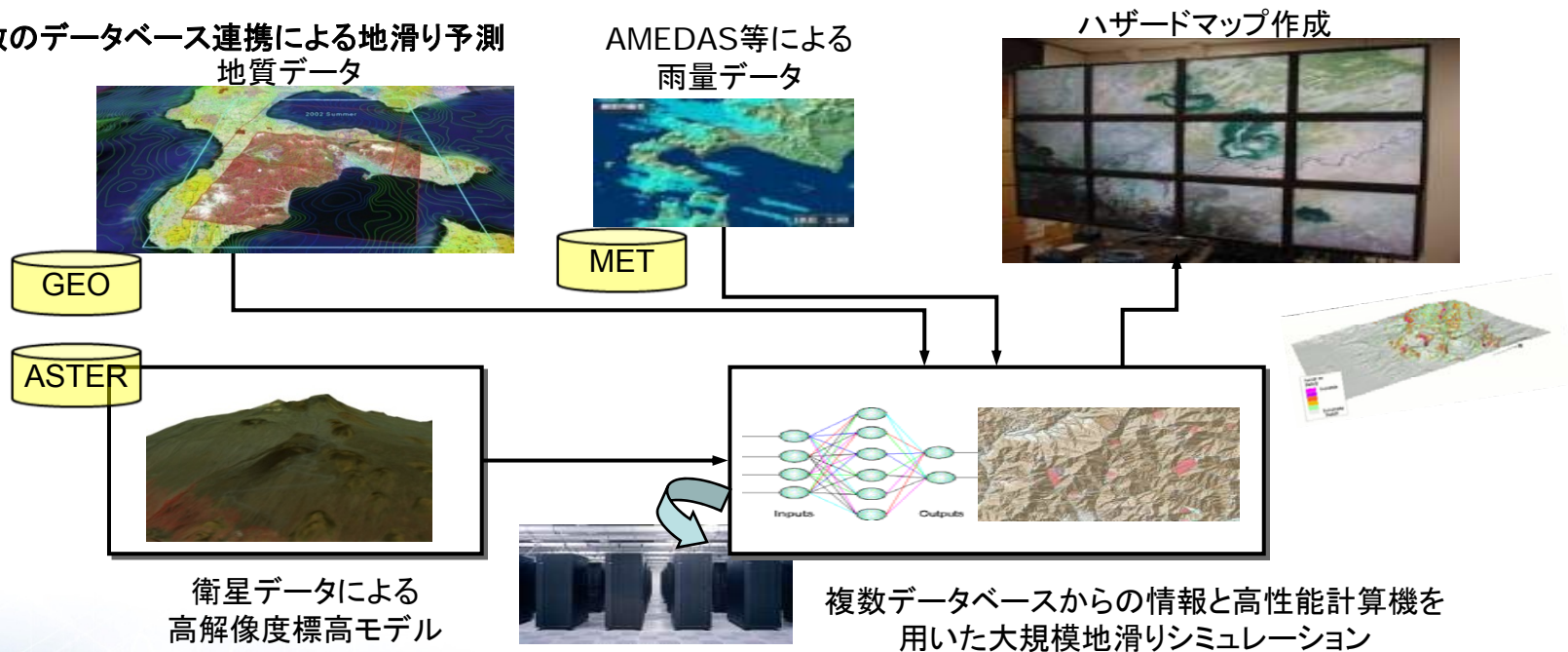
- 異なるデータ格納方法、アクセス方法、スキーマ、データポリシなどが混在したデータベース群を扱う技術が確立されていない。

目的

- 分散環境での異種多様なデータベースを、簡便かつ統一したインターフェースで安全に連携させる技術を開発し、e-サイエンスの利用を推進し、科学技術の発展に貢献する。

応用例

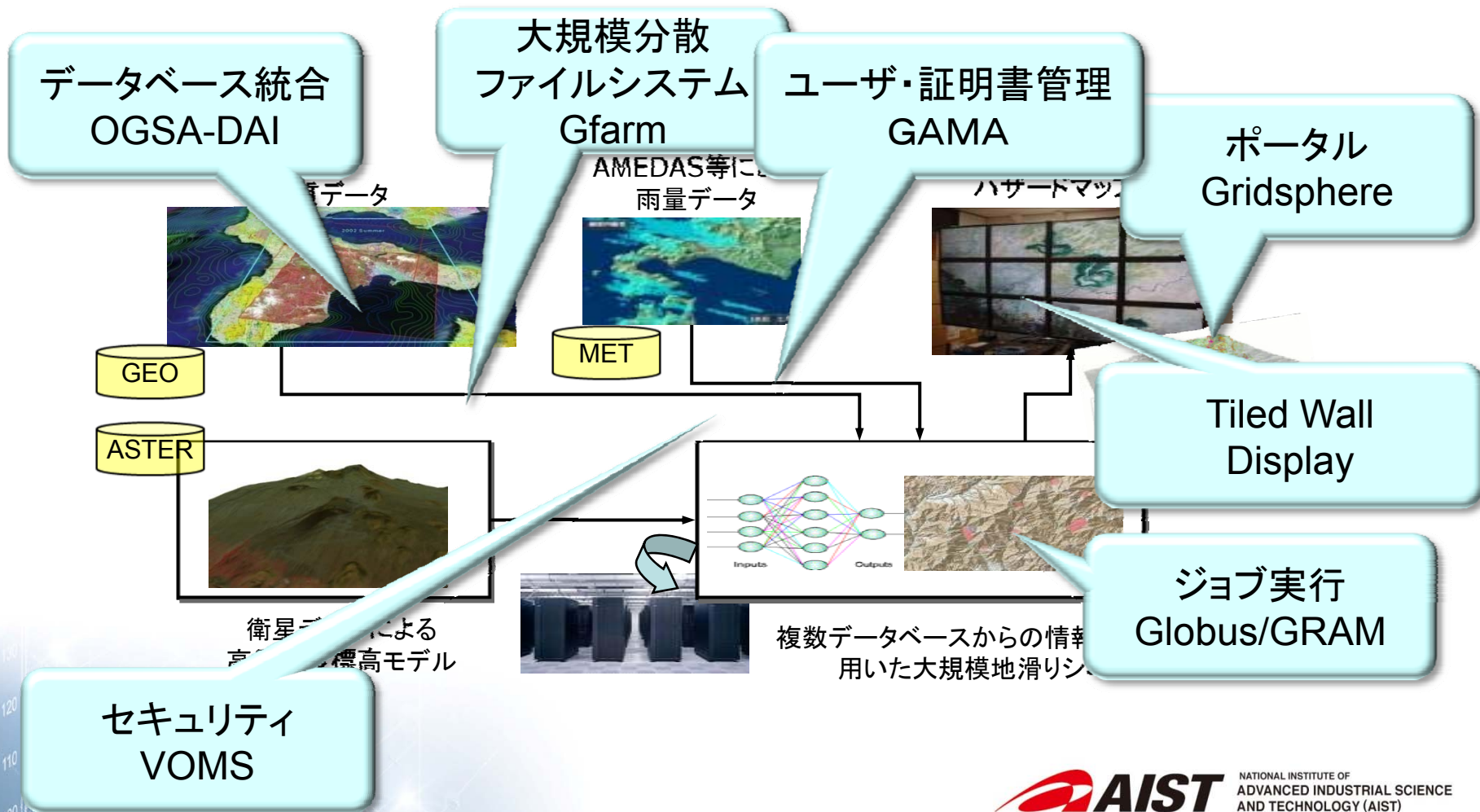
- 複数のデータベース連携による地滑り予測
地質データ



高性能並列計算からデータベース統合からセキュリティから何でもあり

GEO Gridで使われるソフトウェア

Public Domain/ Open Source によるソリューションと研究開発による改良



課題など

• セキュリティ

- OGCサービスとグリッド認証との連携
 - RESTサービスと証明書ベースのグリッド認証との連携

• パフォーマンス

- 3次元DEMの作成：
 - ASTERで現在10分→めざせ数秒！
- 大規模な検索の高速化

• 高機能化

- セマンティック・Web
- ワークフロー・Webアプリの連携？

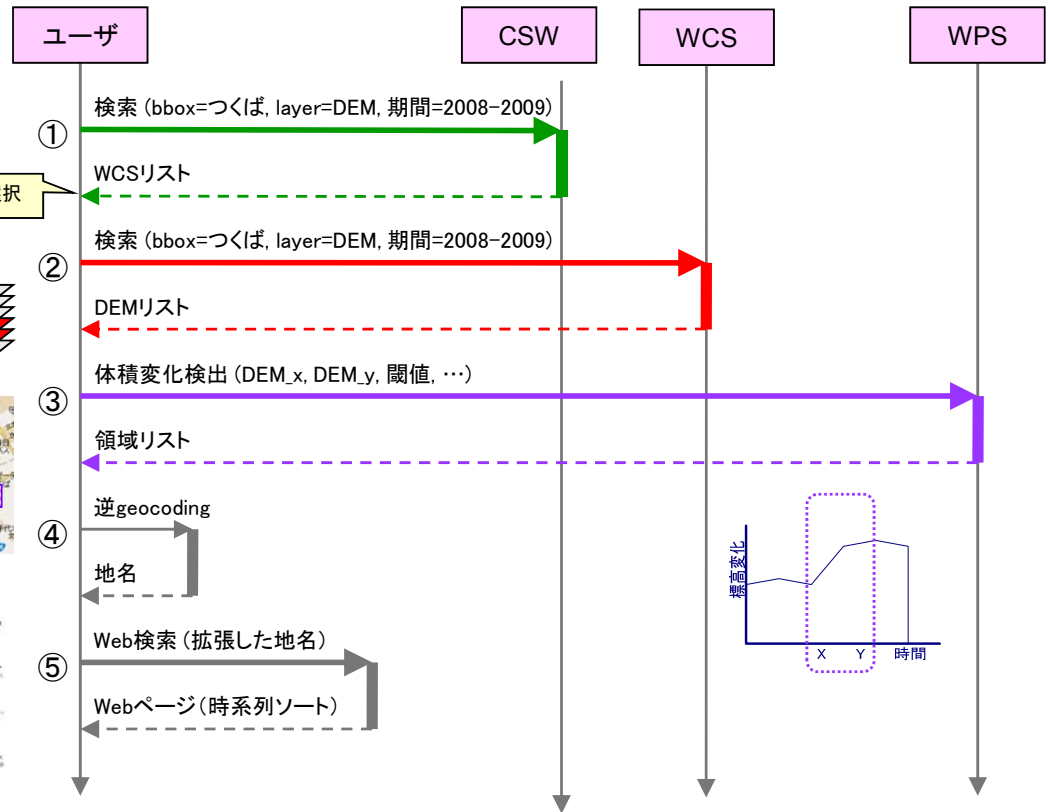
OGC Catalog Service

大規模・高速な実装

メタデータ[カタログ]検索

• カタログ検索

- サービスやデータのありかを発見
- 次に該当するサービスやデータを検索



- これをOGC規格をベースに、全文検索ソフトウェアにポリゴン検索を拡張して実現した(DBMSを使わない)。

メタデータの検索要求

スキーマに依存しない検索をしたい

期間内のデータがほしい

「東京」に関するデータがほしい

雲の量は出来るだけ少ない方がいい

ファイル(E) 編集(E) 表示(V) 履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)

http://etm41.aperid.org:16080/gsjvo-cur/index.html

GMS Gmail - Compose Mail - kojima.isao...

MENU

カタログ検索

検索(最小化)

検索対象: JMP

検索開始日:

検索終了日:

キーワード: 20万分の1

雲量:

検索

全解除 追加 << >>

100 of 46

- 20万分の1数値地質 図幅「高山」 M
- 20万分の1数値地質 図幅「長野」 M
- 20万分の1数値地質 図幅「宇都宮」 M
- 20万分の1数値地質 図幅「水戸」 M
- 20万分の1数値地質 図幅「七尾・富山」 M
- 20万分の1数値地質 図幅「富山」 M
- 20万分の1数値地質 図幅「高田」 M
- 20万分の1数値地質 図幅「日光」 M
- 20万分の1数値地質 図幅「輪島」 M
- 20万分の1数値地質 図幅「長岡」 M
- 20万分の1数値地質 図幅「新潟」 M
- 20万分の1数値地質 図幅「相川及び長岡」 M

WMSレイヤー追加
ベースマップ
選択レイヤー

Permalink

LOG

この範囲のデータがほしい

データ数
120万件

衛星観測メタデータ

- ISO19115 :
 - JMP(Japan Metadata Profile,国土地理院)など、日本も含めたISO標準スキーマ
 - 地質情報、GIS情報など
- ebRim EO Profile
 - 衛星分野で標準になろうとしているスキーマ
 - スキーマ間の階層関係アリ
 - さまざまな衛星のメタデータ
- これらを[都合に応じて]統合して検索したい

```
<?xml version="1.0" encoding="UTF-8" ?>
<getRecordsResponse xmlns="http://www.opengis.net/cat/csw/2.0.2" xmlns:ns2="http://www.opengis.net/ogc"
xmlns:ns3="http://www.opengis.net/gml" xmlns:ns4="http://www.w3.org/1999/xlink" xmlns:ns5="http://www.opengis.net/ows"
xmlns:ns6="http://purl.org/dc/elements/1.1/" xmlns:ns7="http://purl.org/dc/terms/" xmlns:ns8="http://www.isotc211.org/2005/gco"
xmlns:ns9="http://www.isotc211.org/2005/gmd" xmlns:ns10="http://www.opengis.net/gml/3.2"
xmlns:ns11="http://www.isotc211.org/2005/gts" xmlns:jmp20="http://zgate.gsi.go.jp/ch/jmp/">
<SearchResults numberOfRecordsReturned="100" numberOfRecordsMatched="98"> <jmp20:MD_Metadata>
  <jmp20:identificationInfo>
    <jmp20:MD_DataIdentification> <jmp20:citation>
      <jmp20:title>20万分の1数値地質図幅「鹿児島」</jmp20:title> <jmp20:date>
        <jmp20:date>2004-12-01</jmp20:date> <jmp20:dateType>002</jmp20:dateType>
        </jmp20:date> </jmp20:citation>
      <jmp20:abstract>20万分の1地質図幅は、既存の地質資料に基づいて、国土地理院発行の20万分の1地勢図のそれぞれの区画毎に編集
      した地質図です。この数値地質図は、原資料である地質図を数値化したものです。</jmp20:abstract>
      <jmp20:pointOfContact>
        <jmp20:individualName>CD-ROM編集 総括:鹿野和彦・長谷川 功、表示システム作成: 巖谷敏光・川畑 晶、データ編集:宮崎純
        一・巖谷敏光、ベクトルデータ作成:宮崎純一・中島和敏、ラスター画像データ作成:川畑 晶、メタデータ作成:渡辺和明、装丁:中島和敏
        </jmp20:individualName>
        <jmp20:organisationName>産業技術総合研究所地質調査総合センター</jmp20:organisationName> <jmp20:contactInfo>
        <jmp20:onlineResource> <jmp20:linkage>http://www.gsj.jp/HomePageJP.html</jmp20:linkage> </jmp20:onlineResource>
        </jmp20:contactInfo> <jmp20:role>009</jmp20:role> </jmp20:pointOfContact>
      <jmp20:descriptiveKeywords> <jmp20:MD_Keywords>
        <jmp20:keyword>20万分の1、ベクトル、数値地質図、鹿児島、中国西部、九州及び南西諸島、日本</jmp20:keyword>
        </jmp20:MD_Keywords> </jmp20:descriptiveKeywords>
      <jmp20:language> <jmp20:isoCode>jpn</jmp20:isoCode> </jmp20:language>
      <jmp20:characterSet>023</jmp20:characterSet>
      <jmp20:topicCategory>008</jmp20:topicCategory> <jmp20:extent> <jmp20:geographicElement>
      <jmp20:EX_CoordinateBoundingBox> <jmp20:extentReferenceSystem> <jmp20:authority> <jmp20:title>測量法</jmp20:title>
      <jmp20:date> <jmp20:date>1960-07-01</jmp20:date> <jmp20:dateType>003</jmp20:dateType> </jmp20:date> </jmp20:authority>
      <jmp20:code>TD / (B, L)</jmp20:code> </jmp20:extentReferenceSystem>
        <jmp20:westBoundCoordinate>130</jmp20:westBoundCoordinate>
        <jmp20:eastBoundCoordinate>131</jmp20:eastBoundCoordinate>
        <jmp20:southBoundCoordinate>31.333333</jmp20:southBoundCoordinate>
        <jmp20:northBoundCoordinate>32</jmp20:northBoundCoordinate>
      </jmp20:EX_CoordinateBoundingBox>
```

(以下略)

ISO/JMP

ebRim EO Profile

gml:AbstractFeature

衛星センサ
共通のスキーマ要素

hma:EarthObservationProduct

緯経度、日付、などなど

sar:EarthObservationProduct

合成開口レーダー(SAR)
(Synthetic Aperture Radar)

PolarisationMode など

ohr:EarthObservationProduct

光学(OHR)
(Optical High Resolution)

雲量・雪など

atm:EarthObservationProduct

大気(ATM)
(Atmospheric)

DatalayerSpeciesなど
(NO2,O3,,,))

```
<rim:ExtrinsicObject id="urn:uuid:bce71bb1-d71b-40a9-ae91-201cbfdc61e7" objectType="urn:x-ogc:specification:csw-ebrim-cim:ObjectType:DataMetadata">
```

```
<rim:Slot name="modified" slotType="dateTime">
```

```
<rim:ValueList> <rim:Value>2006-06-15T15:00:00Z</rim:Value>
```

```
</rim:ValueList>
```

```
</rim:Slot>
```

```
<rim:Slot name="envelope" slotType="geometry">
```

```
<wrs:AnyValue xmlns:gml="http://www.opengis.net/gml" <gml:Polygon srsName="EPSG:4326">
```

```
<gml:outerBoundaryIs>
```

```
<gml:LinearRing>
```

```
<gml:coordinates>139.4282,35.4882 140.2614,35.4882 140.2614,36.1379 139.4282,36.1379 139.4282,35.4882</gml:coordinates>
```

```
</gml:LinearRing>
```

```
</gml:outerBoundaryIs>
```

```
</gml:Polygon>
```

```
</wrs:AnyValue>
```

```
</wrs:ValueList>
```

```
</rim:Slot>
```

```
<rim:Slot name="title" slotType="string">
```

```
<rim:ValueList> <rim:Value>URI: {
```

```
"uri": "http://maps.geogrid.org/mapserv/ms_aster.pl?",
```

```
"option": {
```

```
"LAYERS": "ASTL1A_0606161247510612129002.dat",
```

```
"SERVICE": "WMS",
```

```
"VERSION": "1.1.1"
```

```
}
```

```
<rim:Value>THUMBNAIL-URI: {
```

```
"uri": "http://www.geogrid.org/cgi-bin/thumb.pl?",
```

```
"option": {
```

```
"res": "small",
```

```
"type": "jpeg",
```

```
"filename": "ASTL1A_0606161247510612129002.dat"
```

```
}
```

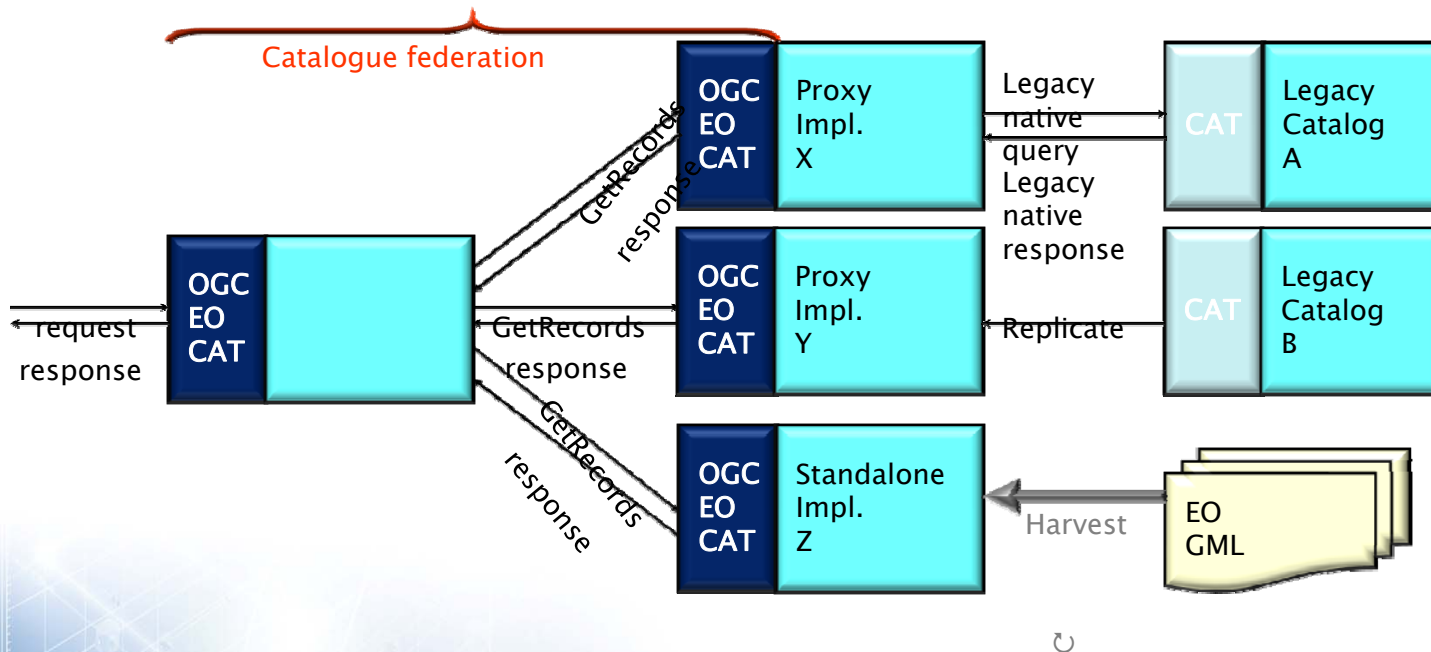
```
</rim:Value>
```

ebRim Object (部分)

ISOと統合して検索したい、
Polygon,
Date
etc

Catalog Service Web

- GEO Grid でサポートする一連のOGC規格の一つ
- REST(HTTP GET/PUT) & SOAP
- 分散カタログの検索を支援



Catalog Service Web 仕様

- ★GetCapabilities :
 - レジストリが支援するインターフェイス情報
 - レジストリが支援するスキーマ情報など、レジストリそのもののサービスに関わる情報の提供
- △DescribeRecord :
 - レジストリが支援しているリソース記述形式の提供
 - (XMLスキーマとか、SQL DDLとか)
- ? GetDomain :
 - レジストリが管理しているリソースのあるフィールドの値の上限と下限
 - (どのクラスタもメモリは4GBから20GBとか。利用法は要検討だが必然性は?)
- × GetRecordById
 - IDによる資源情報の検索(実装予定なし)
- ★GetRecords :
 - 資源の検索 :
 - フォームによるものと問い合わせ言語によるもの
- ★Harvest :
 - 資源を記述するメタデータの収集の指定
- ★Transaction
 - 資源を記述するメタデータの登録・更新(Insert, Delete, Update)

処理の流れ

1) Service (getCapabilities)

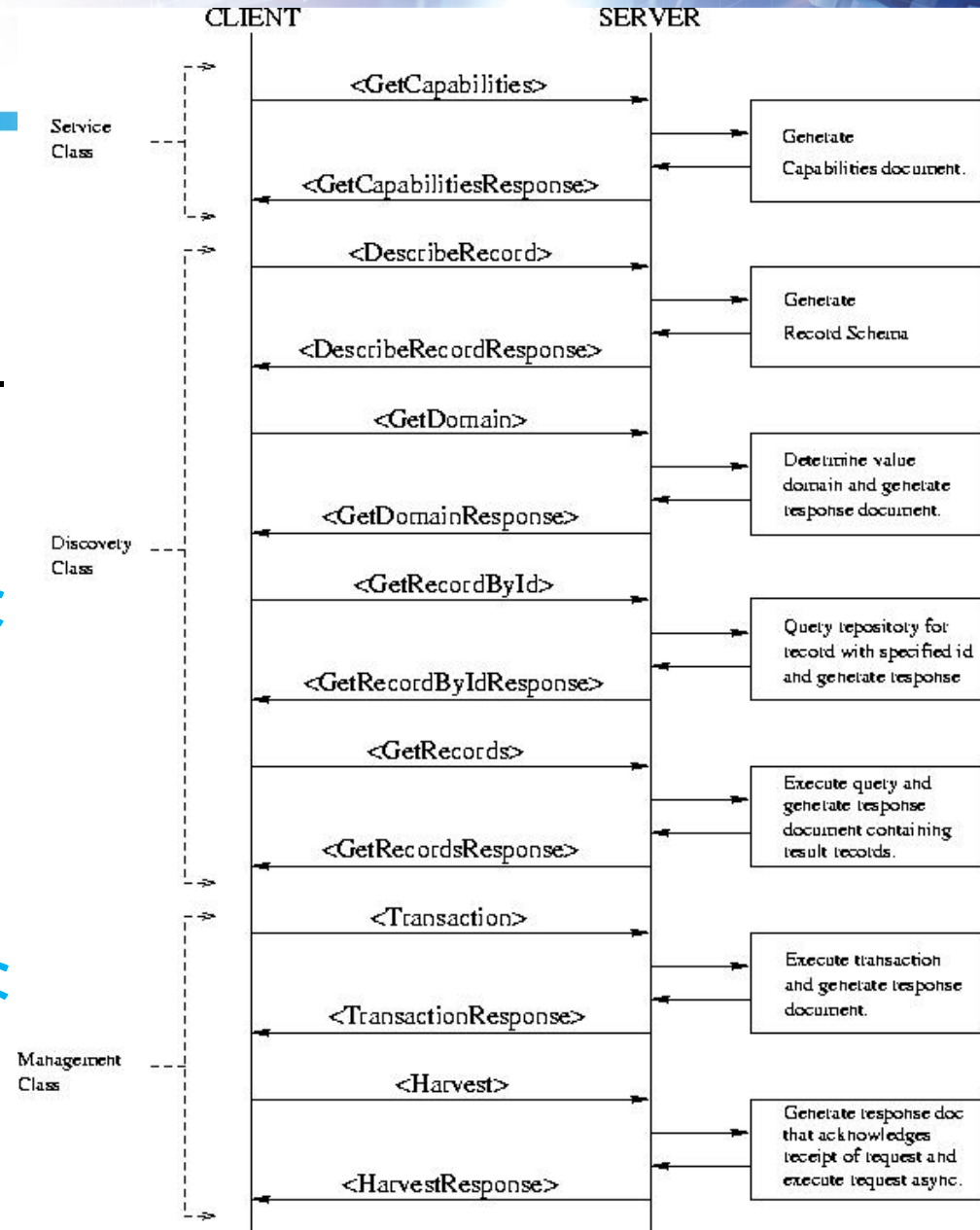
このサービスでどのようなスキーマが処理でき、どのようなメッセージが支援されるかを知る

2) Discovery (getRecordsなど)

資源情報の検索・発見を行う

3) Management (Transactionなど)

資源情報の管理[更新追加)を行う



GetCapabilities (HTTP GET & POST)

システムの構成情報や格納データのスキーマなどのメタ情報の取得メッセージ。
スキーマや利用法が固定している場合にはほとんど使われないと思われる。

HTTP GETでは、<http://www.aist.go.jp/registry/wrs?Service=CSW&Request=getCapabilities>のような形式でOK.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<GetCapabilities
  xmlns="http://www.opengis.net/cat/csw/2.0.2"
  xmlns:ows="http://www.opengis.net/ows"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.opengis.net/cat/csw/2.0.2
  ../../csw/2.0.2/CSW-discovery.xsd" service="CSW">
<ows:AcceptVersions>
  <ows:Version>2.0.2</ows:Version>
  支援するバージョン[の列]。本レジストリは2. 0. 2に従うが、応用が全然違うため無視でもよい。
  <ows:Version>2.0.0</ows:Version>
  <ows:Version>0.7.2</ows:Version>
</ows:AcceptVersions>
<ows:AcceptFormats>
  <ows:OutputFormat>application/xml</ows:OutputFormat> 出力フォーマット 標準はXML
</ows:AcceptFormats>
</GetCapabilities>
```


GetRecords

HTTP/GET(KVP)

- [http://www.someserver.com/csw/csw.cgi?request=GetRecords&version=2.0.2&outputFormat=application/xml&outputSchema=http://www.opengis.net/cat/csw/2.0.2&namespace=csw:http://www.opengis.org/cat/csw&ResponseHandler="mailto:pvretano@cubewerx.com"&typeName=csw:Record&elementSetName=brief&constraintlanguage=CQLTEXT&constraint="csw:AnyText Like '%pollution%'"](http://www.someserver.com/csw/csw.cgi?request=GetRecords&version=2.0.2&outputFormat=application/xml&outputSchema=http://www.opengis.net/cat/csw/2.0.2&namespace=csw:http://www.opengis.org/cat/csw&ResponseHandler=)

GetRecords Request

```
<?xml version="1.0" encoding="UTF-8" ?>
<csw:GetRecords xmlns:csw="http://www.opengis.net/csw" service="CSW"
  version="2.0.0" resultType="results" outputSchema="JMP"
  outputFormat="text/xml" startPosition="1" maxRecords="100">
<csw:Query typeName="csw:GetRecord">
  <csw:Constraint version="1.0.0">
<Filter xmlns:ogc="http://www.opengis.net/ogc">
  <And>
<!-- ISO|RIM|JMP -->
<FileType>JMP</FileType>
<!-- 検索開始日 -->
<PropertyIsGreaterThanOrEqualTo>
  <PropertyName>createDate</PropertyName> <Literal>2000-12-01</Literal> </PropertyIsGreaterThanOrEqualTo>
<!-- 検索終了日 -->
<PropertyIsLessThanOrEqualTo>
  <PropertyName>createDate</PropertyName> <Literal>2008-12-01</Literal> </PropertyIsLessThanOrEqualTo>
<!-- 雲量(JMPの場合はなし) -->
<PropertyIsLessThanOrEqualTo>
  <PropertyName>cloudCoverage</PropertyName> <Literal>20</Literal> </PropertyIsLessThanOrEqualTo>
<!-- 全文検索 -->
<PropertyIsLike wildCard="%" singleChar="_" escape="¥">
<PropertyName>AnyText</PropertyName> <Literal>鹿児島</Literal> </PropertyIsLike>
<!-- Bounding Box -->
<BBOX> <PropertyName>ows:BoundingBox</PropertyName>
  <gml:Envelope xmlns:gml="http://www.opengis.net/gml">
    <gml:lowerCorner>128.0287109375 30.370703125</gml:lowerCorner>
    <gml:upperCorner>146.793359375 47.0259765625</gml:upperCorner>
  </gml:Envelope>
</BBOX>
</And> </Filter> </csw:Constraint> </csw:Query> </csw:GetRecords>
```

Filter

```
<Filter xmlns=http://http://www.opengis.net/ogc xmlns:foo="http://www.someverser.com/foo/">
```

```
<And>
```

```
<Or>
```

```
<PropertyIsLessThan>
```

```
<PropertyName>foo:FIELD1</PropertyName>
```

```
<Literal>10</Literal>
```

```
</PropertyIsLessThan>
```

```
<PropertyIsEqualTo>
```

```
<PropertyName>foo:FIELD1</PropertyName>
```

```
<Literal>20</Literal>
```

```
</PropertyIsEqualTo>
```

```
</Or>
```

```
<PropertyIsEqualTo>
```

```
<PropertyName>foo:STATUS</PropertyName>
```

```
<Literal>VALID</Literal>
```

```
</PropertyIsEqualTo>
```

```
<PropertyIsBetween>
```

```
<PropertyName>/csw:Record/dct:modified</PropertyName>
```

```
<LowerBoundary>
```

```
<Literal>2006-01-01</Literal>
```

```
</LowerBoundary>
```

```
<UpperBoundary>
```

```
<Literal>2006-05-30</Literal>
```

```
</UpperBoundary> z
```

```
</PropertyIsBetween>
```

```
</And>
```

```
</Filter>
```

他、SortBy[予定]
Function定義(一部実装のみ)
など

CQL

SQLのWhere節のみに近い形式 Xpathを最小限支援

```
<csw:Constraint>
```

```
<CqlText>/csw:Record/dct:modified BETWEEN 2006-01-01 AND 2006-05-30</CqlText>
```

```
</csw:Constraint>
```

どこまで言語スペックを実装するかは現時点で未定。

関連研究 & システム

- **関係データベースを用いた実装が一般的**

(後述のredspider cswや、パブリックドメインのdeegree cswなど)

- データはDBに格納
- Webなどの検索フォームをSQLに変換。
- **ポリゴン型の検索や全文検索などを拡張・追加。**

- **問題： 利用者の意図に反する膨大な結果が返る場合**

結果が膨大だが、そのうちの数個をとにかく見たい、、

- 全文検索エンジンでは一般的な答えの返し方
- データの様子を知らずにGUIなどを使って検索するとよく発生する。
- 関係DBベースの実装は、全部の結果を求めてしまうまでの時間がきわめて大きく、答えが膨大な場合の実装に問題がある。

関連システム

	特徴
Redspider CSW	<ul style="list-style-type: none">• 検索要件は満たすしている• 大規模データの検索は不可
GiST/GIN	<ul style="list-style-type: none">• Postgresqlに全文検索機能を付加• 基本機能は満たしているが、複雑なSQLを書く必要がある• スキーマに準拠する必要がある
GeoLucene LocalLucene	<ul style="list-style-type: none">• 全文検索エンジンLuceneに幾何演算が出来るように• 拡張にはLucene本体の改造が必要
全文検索エンジン	<ul style="list-style-type: none">• 全文検索以外に、数値、日付なども検索できる• 独自索引の追加はできない

RDBベース: 商用もパブリックドメインもデータが大規模な状況で×
台数に対するスケーラビリティが望まれる

全文検索エンジン に基づく実装

- 検索: 特定のフィールドに対する検索条件のAND・OR:
 - SQLのような記述力は不要。
- 様々なXML形式の格納:
 - スキーマレスの必要性
- 全文検索: 検索結果の数に対する応答性能が高い。
 - ページングなど。
- 地球観測メタデータ
 - 更新がほとんどなく追加のみ。複雑な索引構造でもOK。
- 全文検索
 - やっぱり便利

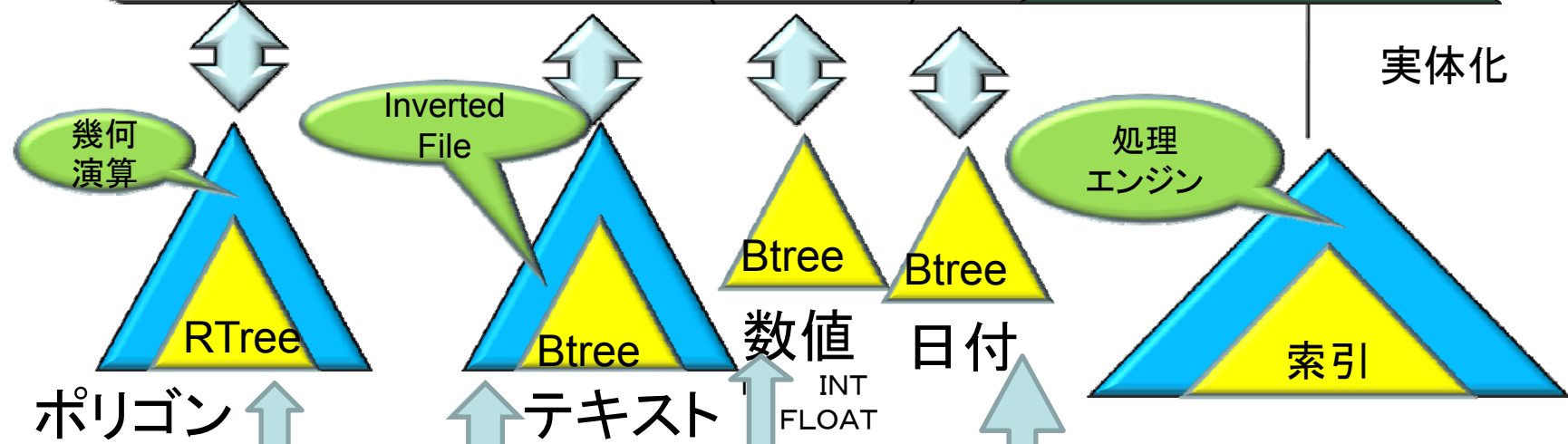
問合せ
Form形式

OGC Filter条件式:
XMLの各フィールドに対する比較条件の
AND&OR
(SQLのような結合はない)

メイン部分

- クエリーの解析
- 適切な索引へ検索要求
- 戻った結果セットの演算(AND/OR)

Index
抽象クラス



<polygon>

XML

```
<polygon>...</>
<cloud>20</>
<date>2005</>
```

<cloud>

<date>

実装

- 現時点では商用の全文検索エンジンをベース
 - 検索対象のタグに対して、データ型ごとの索引を作る。
 - <観測日付><登録日付>一>日付
 - <雲量><雪>一>整数型
 - 個々の属性はタグで区別
 - ジオメトリ処理用の索引を別途構築
 - Polygon/BBOXの検索を支援
 - PostgreSQL & PostGISと同じ関係

現実装における処理の概要

1. 各索引ごとの部分検索に分解
2. 各索引への問い合わせ発行
3. 各索引から戻ってくる結果を識別子で集合演算 (AND/OR)
4. 識別子列から、取得件数だけの実データ取得

最適化の余地多し

各索引に対する部分問い合わせの結果の個数が膨大に

=>遅くないか？

集合演算のコストが馬鹿にならないか？

実験環境

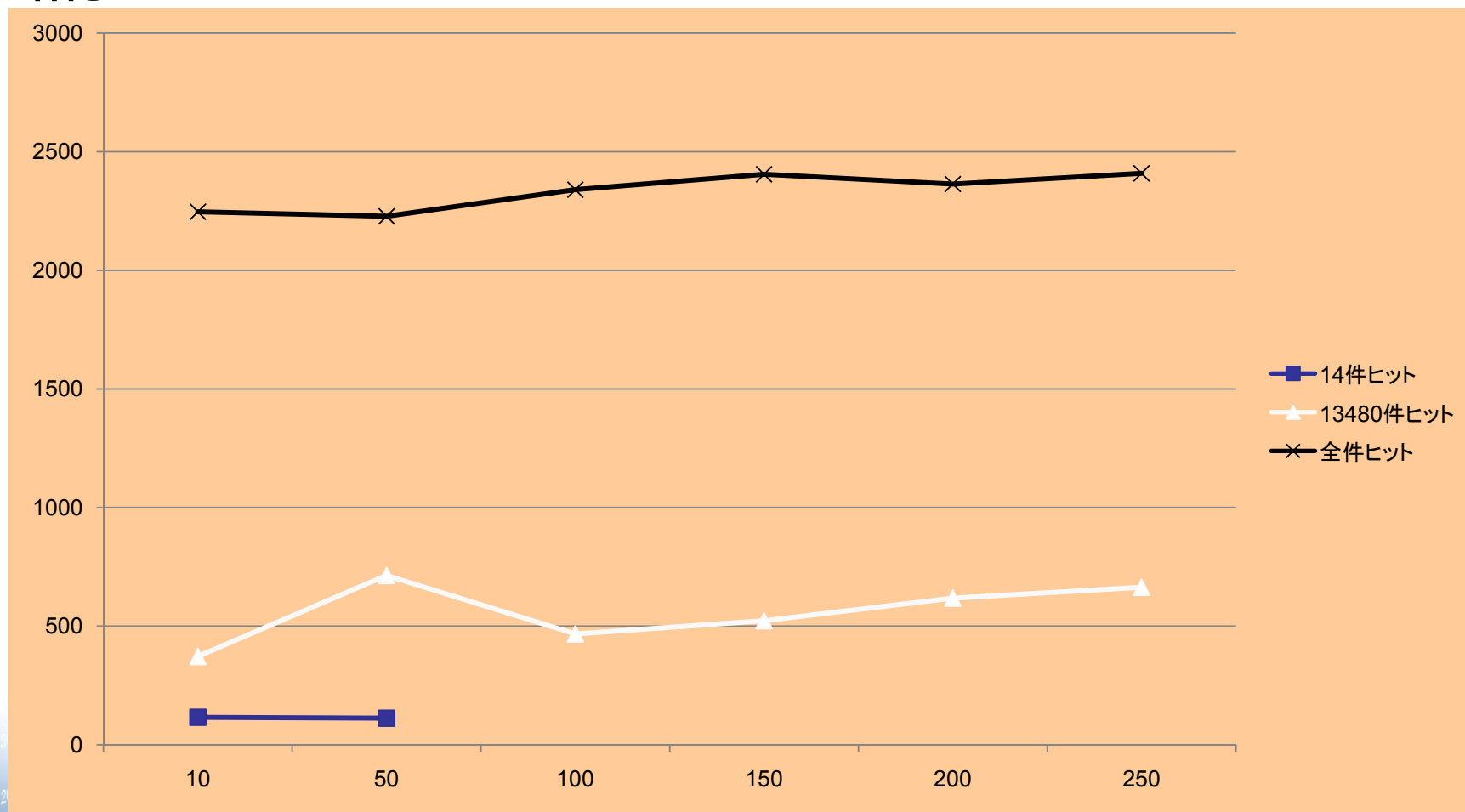
- マシン
 - CPU: Xeon 2.66GHz、メモリ: 8G、HDD: 250GB × 4 (RAID0)
 - OS: CentOS 5.2 (64ビット)
- ソフトウェア
 - 全文検索エンジン: WiSE 2.3
 - (幾何検索: JTS 1.8)
- データ。約10kバイト平均のXMLテキスト 約120万個
 - 詳細データ(インデクスのサイズ、ストレージ総点有量)は未整理
 -

実験概要 1

- 大規模なデータ集合に対して現実的な応答性能が提供できるか？
 - 検索結果の個数に対する応答時間の評価
- 全文検索エンジンを数値比較などに使えるか？
 - 全文検索以外のデータ型の検索が遅いかどうか。

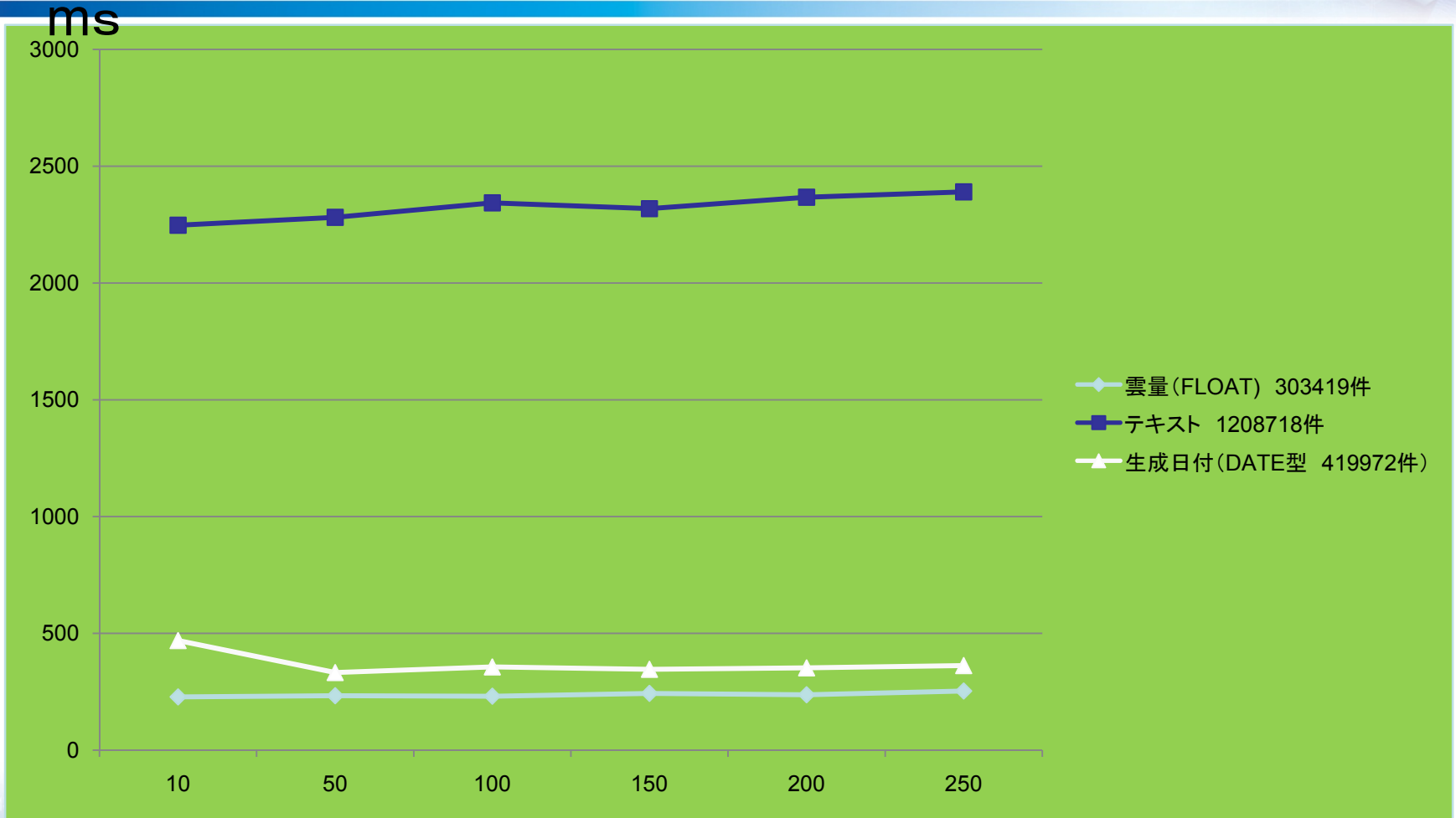
検索結果個数に対する応答時間 (全件=約120万件) アプリ実行時間

ms



結果のうち一度に取得する個数

データ型毎の処理

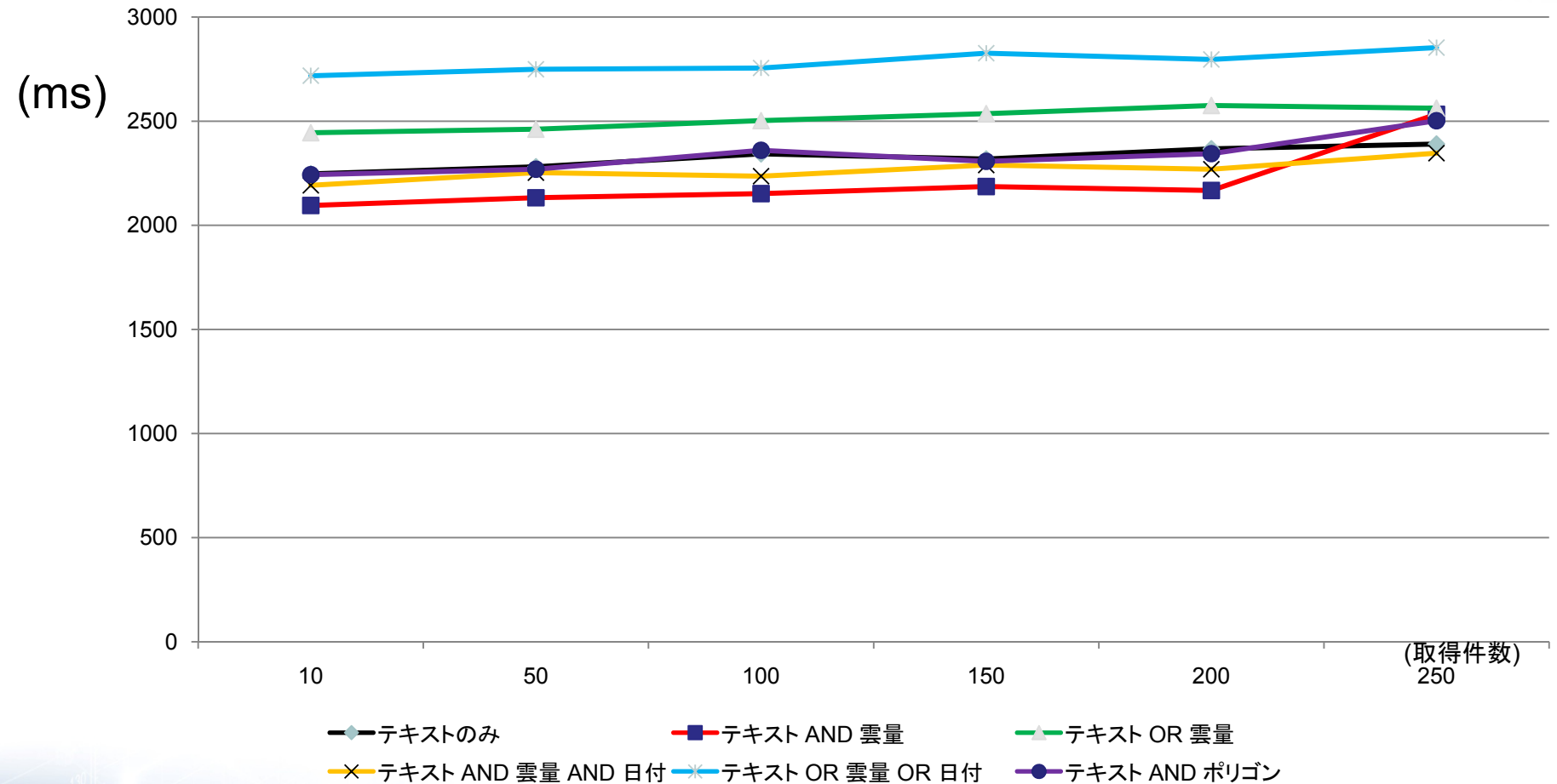


テキストのようなマッチングが不要 = むしろ速い

論理演算

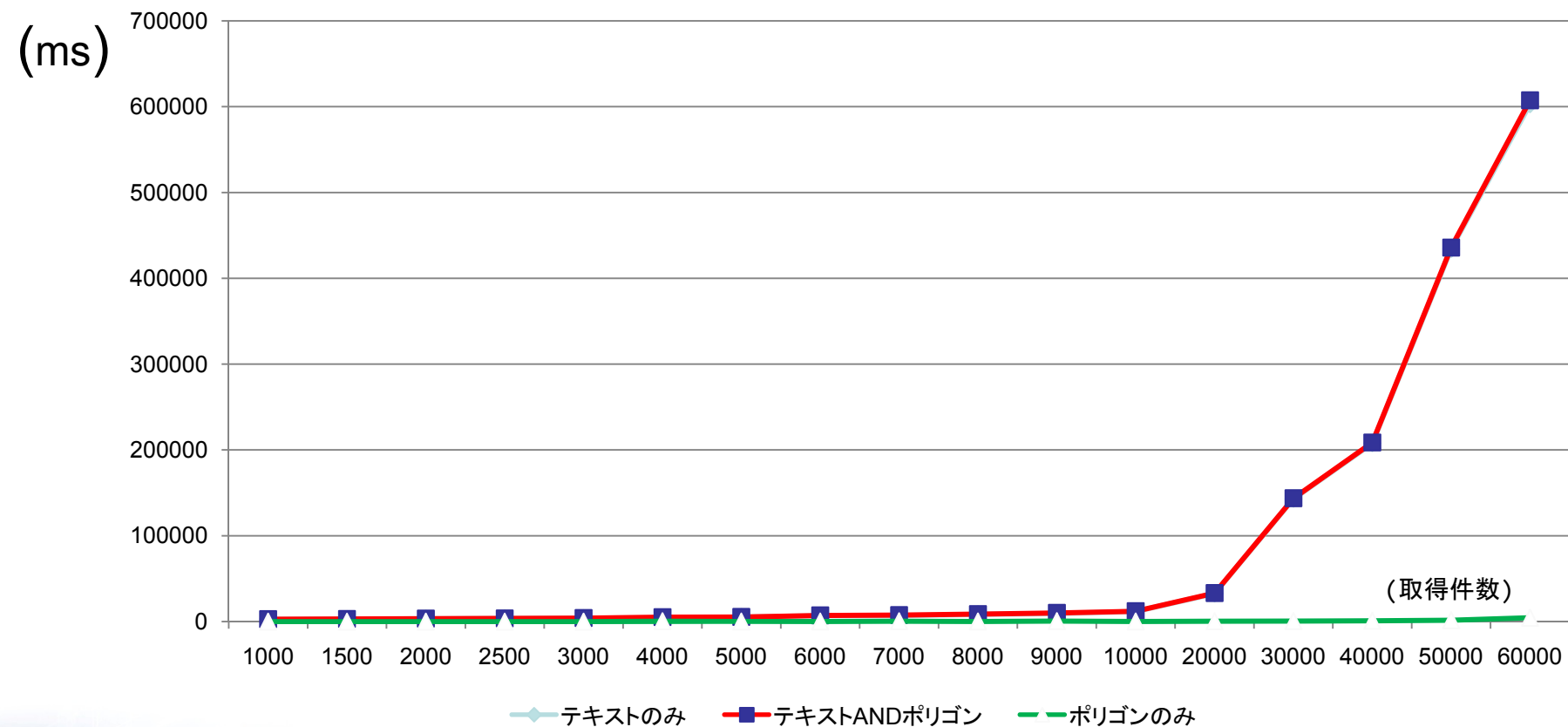
- 検索条件：
 - 全部の索引から全部のデータが部分結果として返る場合の検索（テキストなら120万件）
 - 上記検索に対して集合演算AND,ORかける
 - 上記部分結果間の集合処理が実用的かどうか。
- いずれの場合も、検索結果（答）の総数＝データ数
 - 一度に取得する結果の個数を動かして評価
 - ＝サーチエンジンの検索における、1回（1ページ分）の取得データ数

実験3 論理演算



AND/OR演算の影響は小さい。

実験4 大規模な取得結果



結果の取得件数が多くなるとメモリに入りきれなくなってしまうが、数万件を一度にクライアント側で取得する可能性は低い。

特徴

- 全文検索エンジンの特徴をそのまま引き継ぐ
 - 部分処理による高い応答性能(ページング等)
 - スキーマ非依存
 - 問合せがシンプルである
- 高い拡張性
 - 任意の索引とその処理エンジンの追加
 - 複数の索引間での論理演算が可能

まとめ

- **カタログ(メタデータ)検索**
 - RDBでなくて全文検索エンジンの高機能化
 - 規模に対する応答性能
 - スキーマレス
- **ワークフロー**
 - UIレベルの(コピペに近いイメージでの)連携
 - ブラウザ同士がメッセージ通信
 - 本当にこれが正解か？