
Lattice QCD
Data Grid Middleware:
Meta Data Catalog (MDC)
-- CCS (tsukuba) proposal --

M. Sato, for ILDG Middleware WG

ILDG Workshop, May 2004

Web Service Architecture

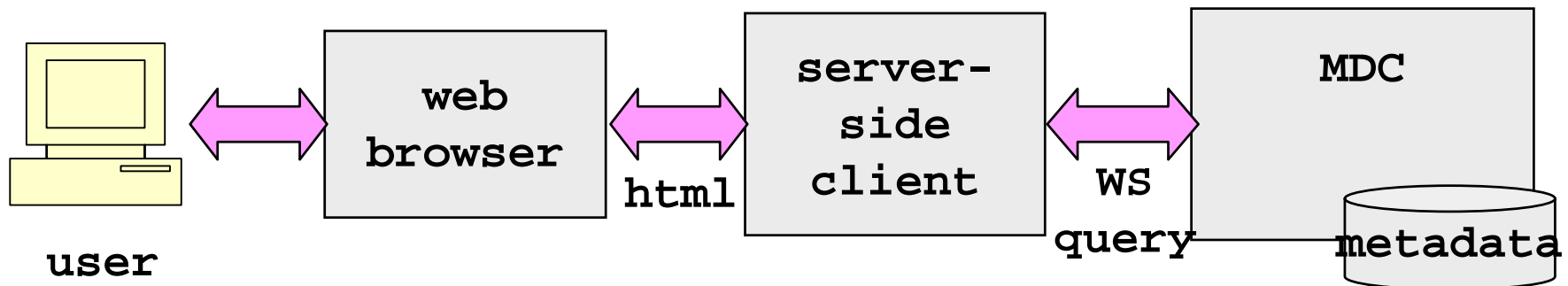
- ◆ Most of the components in ILDG middleware architecture will be stateless web services.
 - They will be standard SOAP web services to which messages are sent, and from which replies are received.

- ◆ The mission of ILDG middleware WG is to define standard and common web service for MDC (and also for SRM, RC)
 - Design and implementation of tools are open to everyone.
 - WG is responsible to provide a reference implementation
 - to validate
 - to show how to use.
 - Contributions such as developing tools will be welcome!

Web (browser) interface

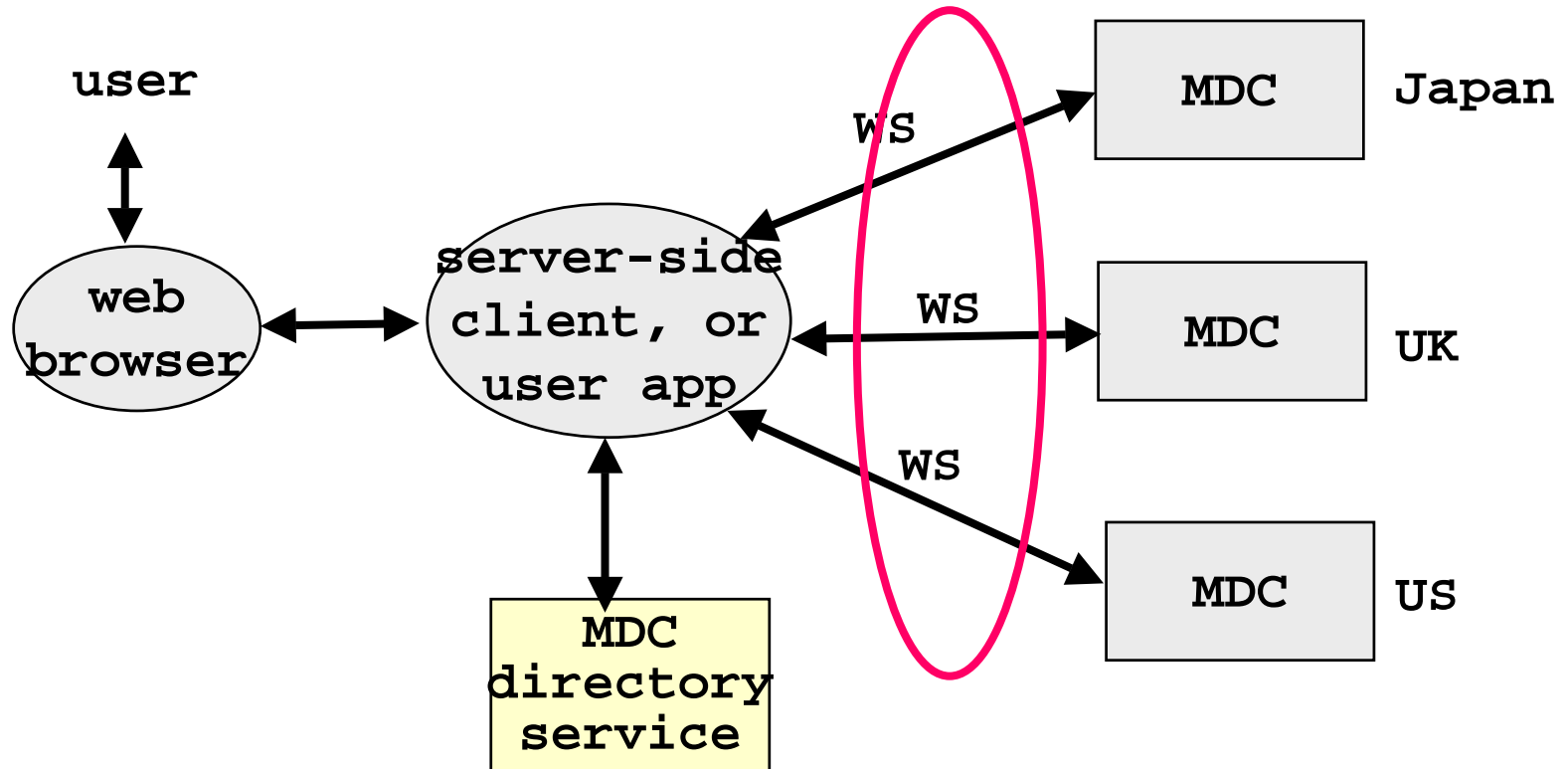
- ◆ In many cases, users will interact with the ILDG using a standard web browser.
 - Some web browsers may support issuing SOAP requests, but it is planned that the ILDG collaboration will support interacting with the web services via dynamic web pages.
 - These pages may be implemented using server-side client such as servlets, and other techniques.

◆ Three tier architecture



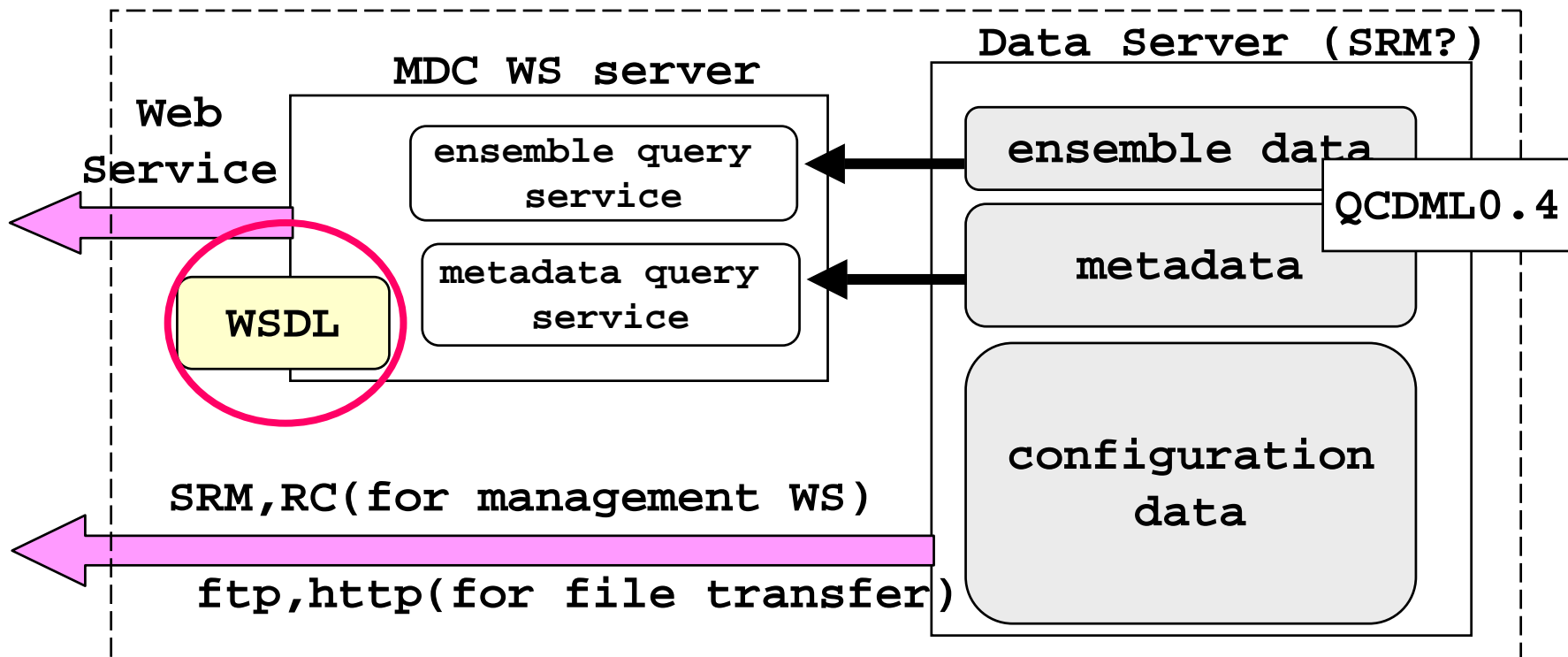
Grid-of-Grids on MDC

- ◆ Clients can access multiple MDC's at different sites
- ◆ Directory service tells the locations of MDC's, which may be managed by LDAP



Components in MDC

- ◆ In QCDML0.4, ensemble and metadata are separated.
- ◆ Ensemble data, metadata, configuration data are stored in data server (SRM?). These data are identified by GFN (global file name) , and transferred via protocol such as SRM, RC, ftp, http,
- ◆ MDC WS server provide the query service to search ensemble and metadata. WS is defined by WSDL (web service description language).



WSDL

◆ Web Service Description Language

- XML document defines a web service interface (service names, message format, binding, protocol)
- Enable the interface to be generated automatically for many prog. lang. bindings.
 - Java (Axis/JAX-RPC)
 - PHP
 - Python
 - C++ (gSoap) , C# , .NET ,

◆ Design the interface == Write WSDL for the MDC

```
<?xml version="1.0" encoding="utf-8"?>
<definitions
  xmlns:s="http://www.w3.org/2001/XMLSchema"
  xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
  xmlns:tns="uri:ildg" targetNamespace="uri:ildg"
  xmlns="http://schemas.xmlsoap.org/wsdl/">

  <types/>
  <message name="ensembleQuery">
    <part ... "/>
  </message>
  <message name="EnsembleQueryReuslts">
    <part name="..."/>
  </message>
  <portType name="ildgMDCPort">
    <operation name="ensembleQuery">
      <input message="tns:ensembleQeury"/>
      <output message="tns:ensembleQueryResults"/>
    </operation>
  </portType>
  <binding name="ildgSoap" type="tns:ildgMDCPort">
    <soap:binding style="rpc"
      transport="http://schemas.xmlsoap.org/soap/http">
    <operation name="helloWorld">
      <soap:operation style="rpc"/>
      <input>
        <soap:body use="encoded" namespace="uri:ildg"
          encodingStyle="http://schemas.xmlsoap.org/s
        </input>
      <output>
        <soap:body use="encoded" namespace="uri:ildg"
          encodingStyle="http://schemas.xmlsoap.org/s
        </output>
      </operation>
    </binding>
  <service name="ildgMDC">
    <document>
      This is a test for ildg.
    </document>
    <port name="ildgMDCPort1" binding="tns:ildgSoap">
      <soap:address location="http://127.0.0.1:5335/"
    </port>
    <port name="ildgMDCPort2" binding="tns:ildgSoap">
      <soap:address location="http://127.0.0.1:5334/"
```

Interface for MDC (1)

- ◆ Query to ensemble query service (in Java)
 - returns a set of GFNs in EnsembleQueryResults

EnsembleQueryResults

```
doEnsembleQuery(String queryFormat, /* ALL, Xpath, SQL, ...*/
  String queryString, /* string for query */
  int startIndex, /* the start index to be returned */
  int maxResults); /* maximum # to be returned */
```

```
class EnsembleQueryResults {
  String queryFormat; /* same as input */
  String queryString; /* same as input */
  int totalRresults; /* total number of matched ensemble */
  int startIndex; /* same as input */
  int numberOfGFNs; /* # of GFNs returned */
  String GFNs[]; /* the array of GFNs matched */
  String QueryTime; /* query time (optional) */
}
```

- If queryFormt == “ALL”, retrieve all data.
- SQL-style query may be useful and efficient (as suggested by Eric)

Interface for MDC (2)

◆ Query to metadata query service

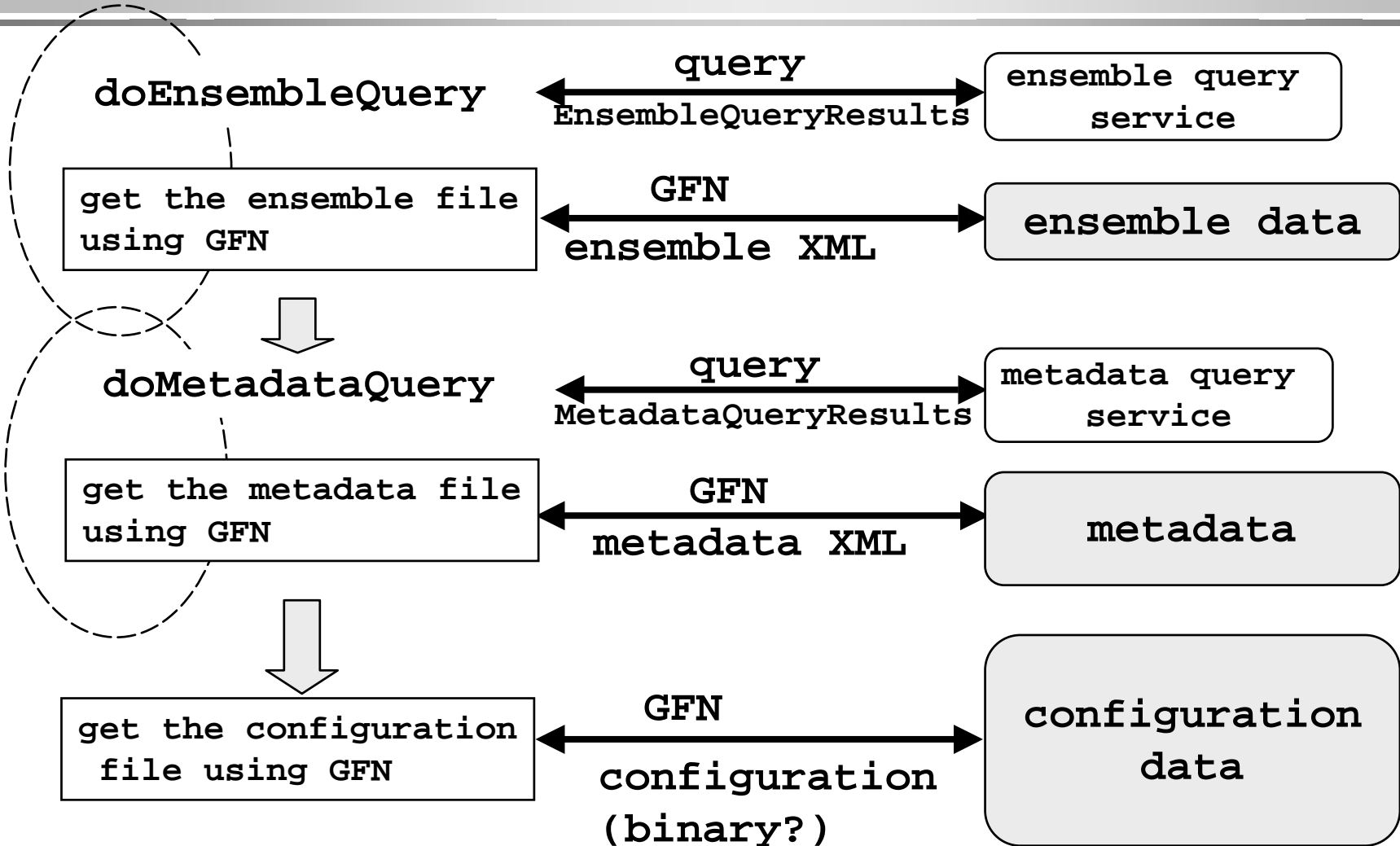
- returns a set of metadata GFNs
- Almost the same to doEnsambleQuery.

MetadataQueryResults

```
doMetadataQuery(String queryFormat, /* ALL, Xpath, SQL, ...*/
    String queryString, /* string for query */
    int startIndex, /* the start index to be returned */
    int maxResults); /* maximum # to be returned */
```

```
class MetadataQueryResults {
    String queryFormat; /* same as input */
    String queryString; /* same as input */
    int totalRresults; /* total number of matched metadata */
    int startIndex; /* same as input */
    int numberOfGFNs; /* # of GFNs returned */
    String GFNs[]; /* the array of GFNs matched */
    String QueryTime; /* query time (optional) */
}
```


A Use Case



Discussion

- ◆ Data and metadata export tools
 - Insertion (export) operation is not mandatory?
 - insertion operation may be done locally at each site, in each collaboration.
 - Coherence between MDC and SRM&RC
 - Insert data into SRM, then insert metadata into MDC.
 - It is valid if data exists in SRM and no metadata exists. Is it OK?

- ◆ File format issue
 - who and when pack metadata and dataset?

- ◆ Search for multiple sites
 - flat or hierarchical (recursive) ?
 - directory service.

Plan

- ◆ **Finish the definition of MDC interface in WSDL (for single MDC)**
- ◆ **Provide reference single Grid MDC**

- ◆ **Multiple MDCs**
- ◆ **Directory service of MDCs**