XML Based Networking Method for Connecting Distributed Anthropometric Databases

24 October 2006

Huaining Cheng

Dr. Kathleen M. Robinette

Human Effectiveness Directorate

Air Force Research Laboratory

Overview

- Integration issues
- Solution options with pros and cons
- WEAR short-term solutions
- WEAR long-term solution
- Summary

WEAR Group

- **WEAR stands for World Engineering Anthropometry Resources**
  - **A nonprofit organization based in France to promote sharing and using of its members' anthropometry resources**
  - **A dozen members located around world**
  - **Members have huge collection of anthropometry survey data over large span of time**

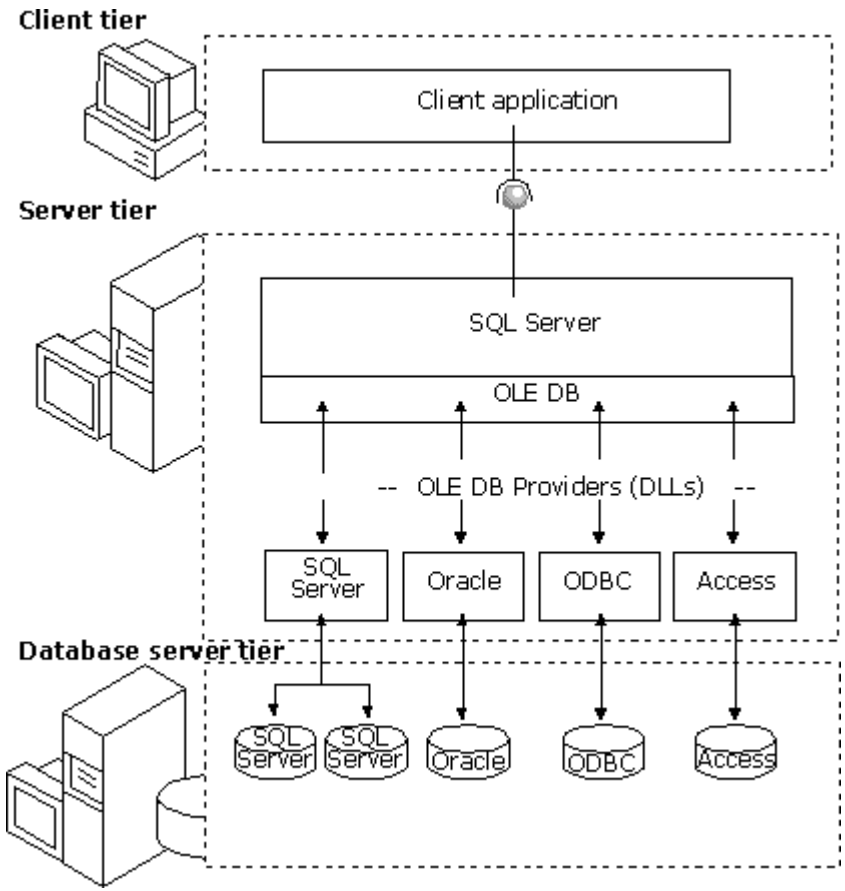
Integration Issues

- Data extraction and search
- Network accessibility
- Data representation
- Data quality
- Security
- Data analysis
- Data autonomy
- Financial and manpower constraints

Integration Solution Options

- Examine three possible integration solutions for the WEAR group integration
  - Linked servers
  - Data warehouse
  - Service-oriented architecture (XML Web services)
- Examine how the solutions can address the above integration issues
- Envision potential anthropometry applications for the integrated WEAR

Linked Server Structure



- **Use four-part name syntax in a query instead of only table name**
  - **LinkedServerName**
  - **DatabaseName**
  - **Owner**
  - **TableName**

REF 1: "Configuring Linked Servers," Microsoft MSDN Library, http://msdn.microsoft.com/library/default.asp?url=/library/en-us/adminsql/ad_1_server_4uuq.asp

Linked Server Pros & Cons

- **Advantages – conceptually easy**
  - **The ability to issue distributed queries and transactions on heterogeneous data sources across the enterprise**
- **Disadvantages – tight integration and direct access**
  - **Maintenance and update nightmare**
  - **No data autonomy**
  - **Require direct access and reliable connection**

Data Warehouse

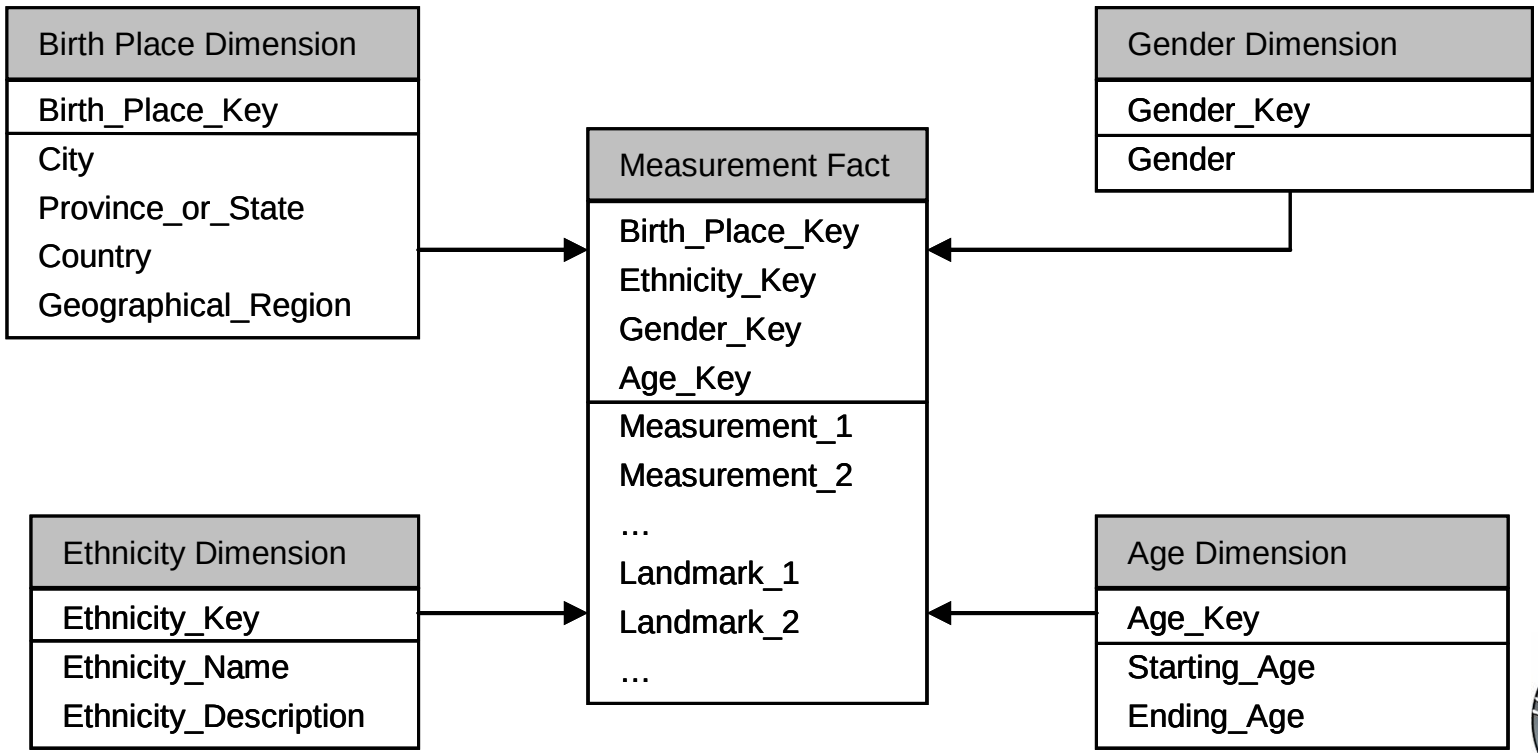
- **Main objective of data warehouse is business intelligence analysis of integrated data over time**
  - **Business trend and variance analysis through OLAP (On-Line Analytical Processing)**
  - **Data mining – automated discovery of implicit patterns and interesting knowledge hidden in the large amounts of data**

Data Warehouses

- **Characteristics of data warehouse**
  - **Specialized database built on top of operational databases.**
  - **Integration process through ETL (Extract, Transform, and Load)**
  - **Optimized for over time data analysis**
  - **Highly normalized structure**

Anthropometry Star Schema

- **Anthropometry database application of star schema**
  - **Find groups of anthropometry measurements as biometrics identifier**



Data Warehouse Pros & Cons

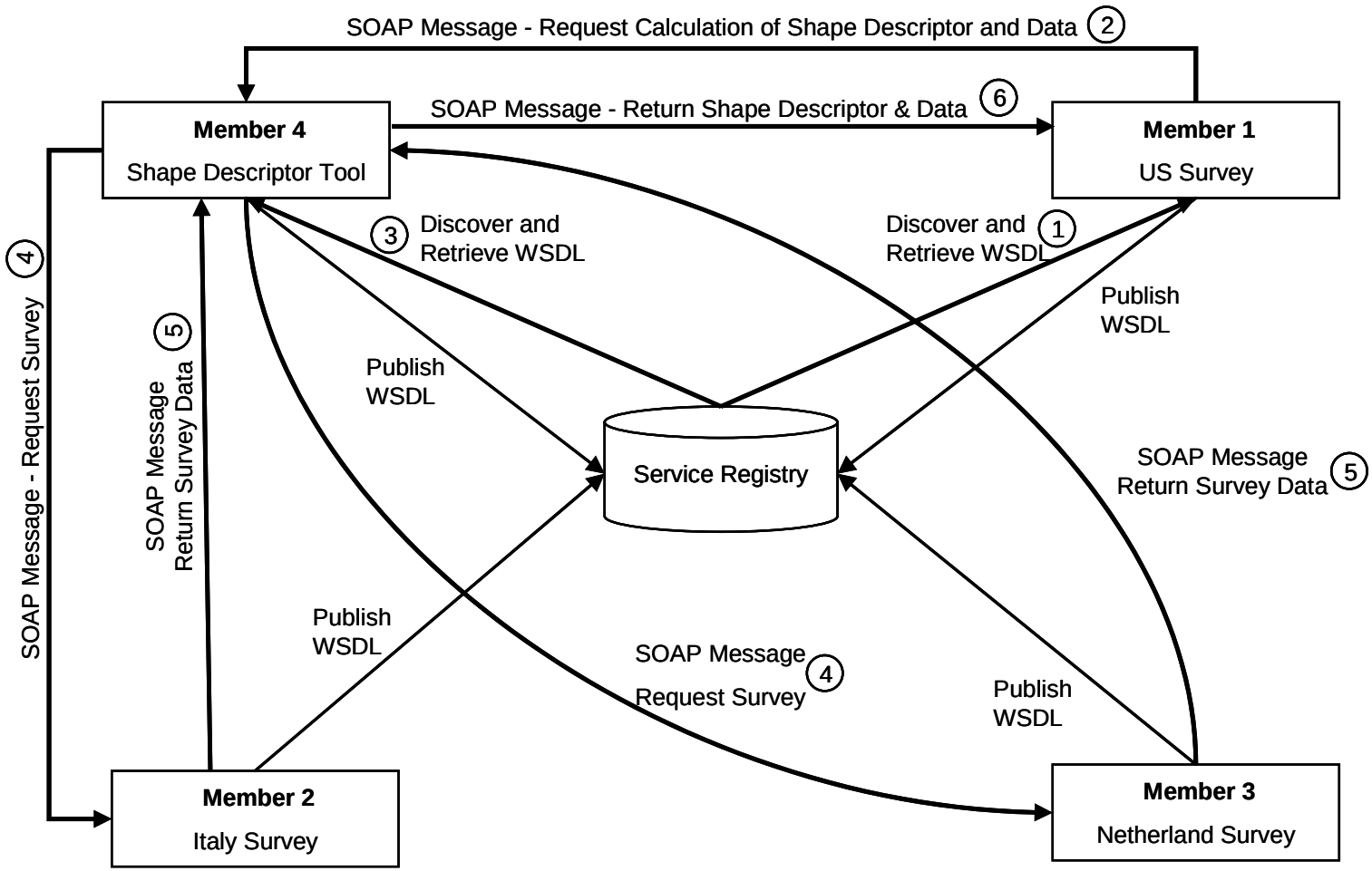
- **Advantages – strong analytical capability**
  - **A platform for many potential anthropometry applications of OLAP and data mining**
- **Disadvantages – tight integration and high cost**
  - **Difficult to build ETL processes with various systems platforms and database structures as well as different locales**
  - **Lack of data autonomy**

Web Service Architecture

- **Web service architecture adheres to the principles of service-orientation**
  - **Services are loosely coupled, autonomous, stateless, and discoverable.**
- **Consists of three basic types of entities**
  - **Service requestor, service provider, and service registry**
  - **Communicate through TCP/IP**

WEAR Application of Web Services

- **Member 1 requests survey data & shape descriptors**



Loosely Coupled Integration

- XML Web service architecture is a loosely coupled integration
  - Integration is done through the service contract (WSDL) instead of open connection
  - Work is requested/delivered as payloads in the SOAP messages
  - Messaging mechanism brings
    - Autonomy
    - Statelessness

Web Services Pros & Cons

- **Advantages – autonomy and scalability**
  - **Solve the problems that are difficult to handle by a tight integration**
- **Disadvantages – security and performance**
  - **Challenge in performing and propagating user authentication and authorization**
  - **XML documents are slow to create and process**
  - **Evolving standards and specifications**

WEAR Integration Objectives Short-Term

- Real world – two types of integrations
  - Enterprise systems integration
  - Business to business (B2B) integration
- Short-term WEAR integration objectives
  - Integrate and share anthropometry survey data automatically
  - WEAR members maintain and control independently their databases and existing web application

WEAR Integration Solution Short-Term

- WEAR has to be treated as a federation
- Web service architecture is inherently federated because of its loosely coupled nature
  - Best solution to make the WEAR integration satisfy the short-term objectives
    - Universal anthropometry data sets – XML
    - Autonomous

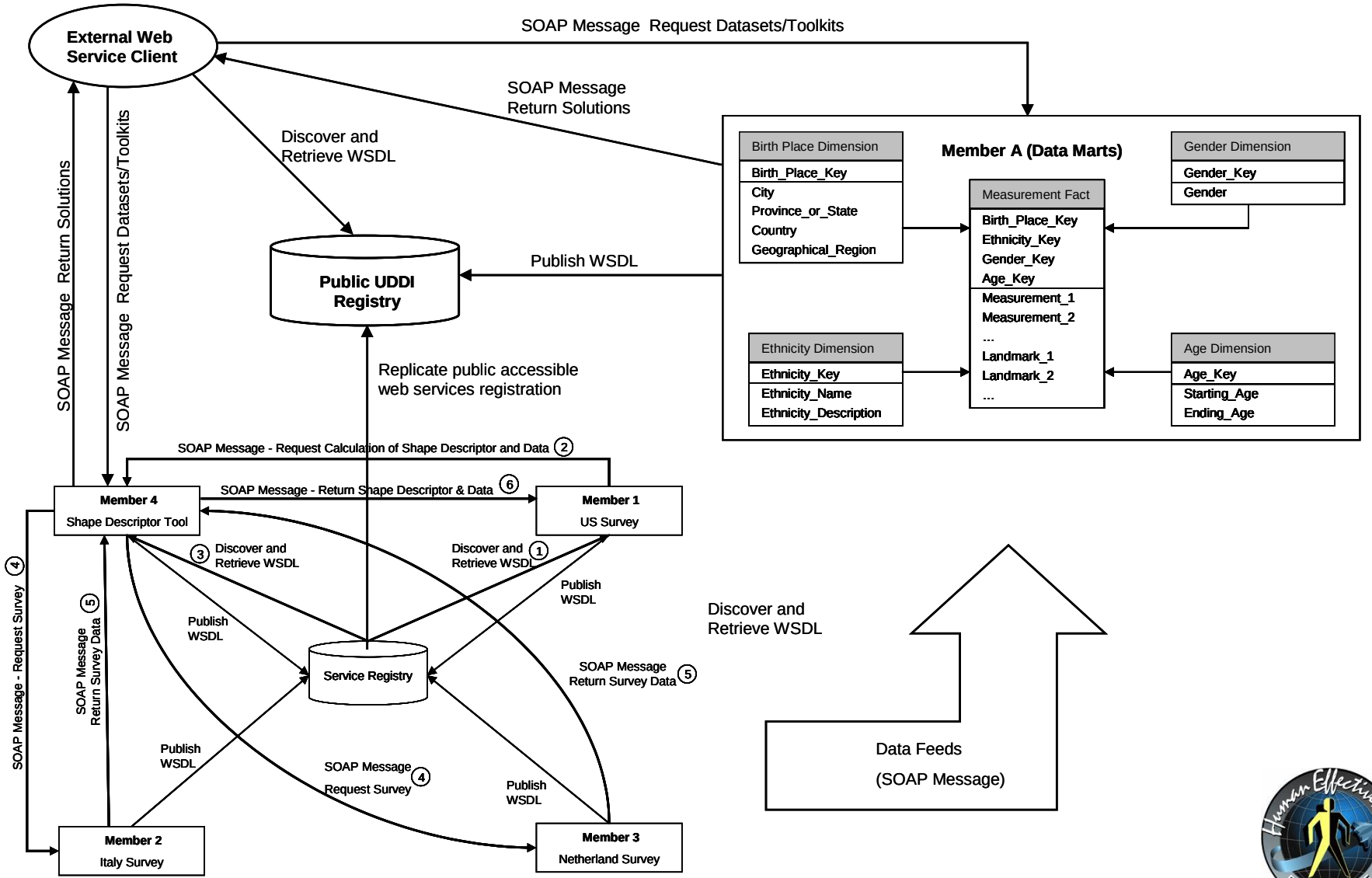
WEAR Integration Solution Implementation Issues

- Limit services to data sets only without RPC (remote procedure call)
- Real-time performance is not a concern
- Use restricted UDDI registry to increase security
- Implement user authentication using X509 digital client certificate
- Implement user authorization through SOAP header

WEAR Integration Objectives Long-Term

- **WEAR integration long-term objectives**
  - **Build analytical models to produce anthropometry solution toolkits**
  - **Offer these toolkits as Web services accessible by the public and special industry groups**

WEAR Hybrid Web Service / Data Mart Model



Conclusions

- **WEAR integration is a type of federated integration**
- **XML Web service is the best solution due to its loosely coupling nature and service orientation**
- **Data marts have great potential for discovery of anthropometry data**
- **Hybrid Web service/data mart model is a solution to combine analytical models and XML web services**
- **XML is the foundation of the entire integration solution**