DDI3 Uniform Resource Names: Locating and Providing the Related DDI3 Objects

Part of Session: DDI 3 Tools: Possibilities for Implementers

IASSIST Conference, Cornell University, June 1-4, 2010

Joachim Wackerow, GESIS – Leibniz Institute for the Social Sciences
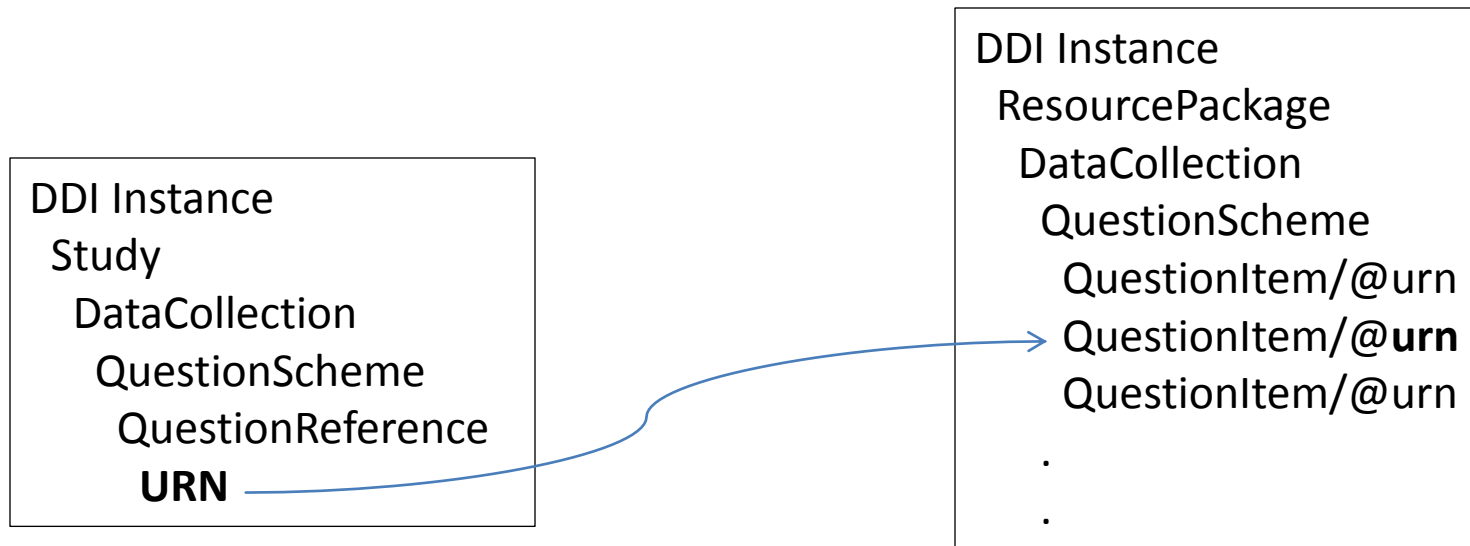
Overview

- Introduction
  - Background in DDI
  - Relationship URI / URN / URL
- URN Resolution
  - DNS-based approach
- Query Protocol Proposal

Introduction: Background in DDI

- DDI is expressed in XML
- 120 elements/objects can be identified by IDs
- This adds reusability of these objects to the hierarchical structure of a DDI instance
- The IDs have a local scope, often related to a DDI scheme
- A DDI scheme is an list of items which is maintained by a DDI agency
  - altogether 31 maintainable objects, the most important ones are 14 DDI schemes
- The IDs and the information about the maintainable object build the basis to construct DDI URNs
- URNs are globally unique identifiers and can be seen as persistent identifiers
- DDI URNs add reusability of DDI objects in a network of DDI instances

Use cases of distributed DDI resources

- Examples of possible main usage as reusable resource package
  - Question bank
  - Standard demographic variables

DDI Instance
 Study
  DataCollection
   QuestionScheme
    QuestionReference
     URN

DDI Instance
 ResourcePackage
  DataCollection
   QuestionScheme
    QuestionItem/@urn
    QuestionItem/@**urn**
    QuestionItem/@urn
  .
  .

DDI URN Example

urn:ddi:de.gesis:VariableScheme.vs1786.4.2.3:Variable.age.1.0.0

- The DDI element Variable with the ID "age" and the version "1.0.0"
- is contained in the VariableScheme with the ID "vs1786" and the version "4.2.3"
- which is maintained by the DDI agency identified by "de.gesis"
- in the URN namespace "ddi"

Relationship URI / URN / URL

- The Uniform Resource Identifier (URI) identifies a name or a resource on the Internet
- The Uniform Resource Name (URN) defines an item's identity
- An URN is a persistent, location-independent resource identifier
- The Uniform Resource Locator (URL) specifies where an identified resource is available and the mechanism for retrieving it.

DDI URN Resolution

- A DDI object is identified by a DDI URN

- The DDI URN is a globally unique identifier

- The DDI URN must be resolved to an URL to find the identified object on the Internet

- A DDI object with an unique URN can have multiple locations identified by multiple URLs

URN Resolution Service
Different approaches

- Specialized resolution services for persistent identifiers
  - Examples Handle, DOI, PURL
  - Not URN compliant, can only be used by application on top of it
  - Dependency from additional framework, possible costs
- DNS-based resolution
  - hierarchical naming system for computers on the Internet, "phone book" for the Internet
  - existing, well maintained infrastructure

DNS-based URN Resolution Service

- Approach focuses on simplicity and uses existing infrastructure

- DNS can be used for URN resolution with additional preparation steps
  - No out-of-the-box resolution for URNs available

- Assumption: all DDI objects of a DDI agency or sub-agency are provided by services with a single entry point
  - Example: HTTP-based service

DNS-based URN Resolution Service Structure

urn:ddi:de.gesis:VariableScheme.vs1786.4.2.3:Variable.age.1.0.0

- Focusing just on the agency id
- Application queries DNS: which services are available for DDI objects maintained by a specific agency?
- Response from DNS: list of available services for this agency
- Application selects a service (e.g. a DDI repository) and queries this service
  - http://ddirepository.gesis.org/
  - http://ddirepository.gesis.org/?URN=urn:ddi:de.gesis:VariableScheme.vs1786.4.2.3: Variable.age.1.0.0

Algorithm

- Input is complete URN. Example: urn:ddi:de.gesis:VariableScheme.vs1786.4.2.3:Variable.age.1.0.0

- Extraction of the maintaining agency id. Example: de.gesis

- Transformation of the agency id to an Internet domain name. Example: gesis.de.ddi.urn.arpa. (URN is below "arpa")

- Sending the agency id (in this format) as request to the DNS.

- The DNS response is a list of available services for DDI objects of this agency. Example: DDI repository providing DDI objects by a RESTful interface.

- The response should be cached by the resolution middleware.

- The application selects an appropriate service from the list of services.

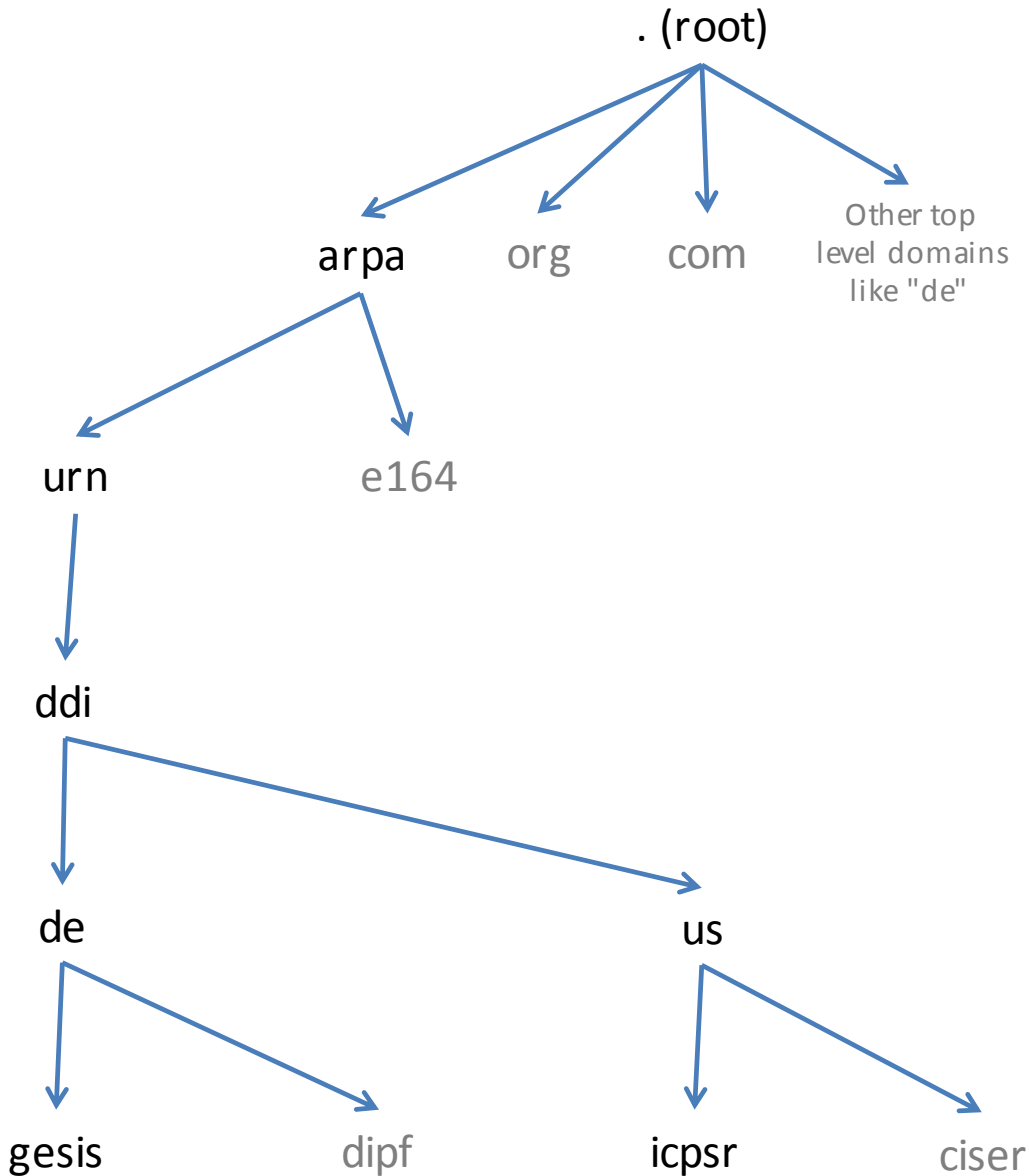- The application queries the service.

DNS Delegation and Resolution for DDI URNs
Hierarchy and Example Configuration
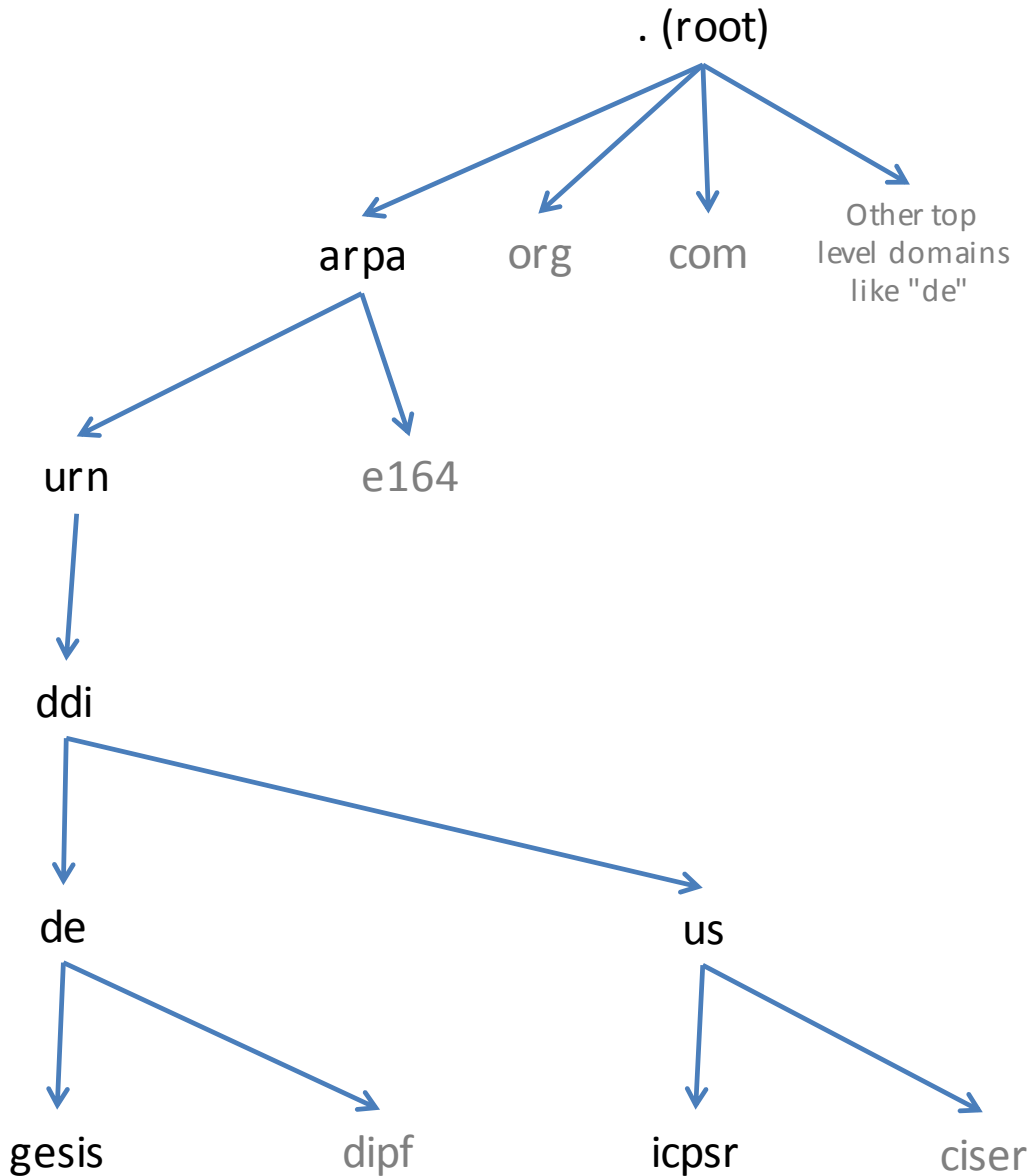
DNS Delegation and Resolution for DDI URNs
Hierarchy and Example Configuration

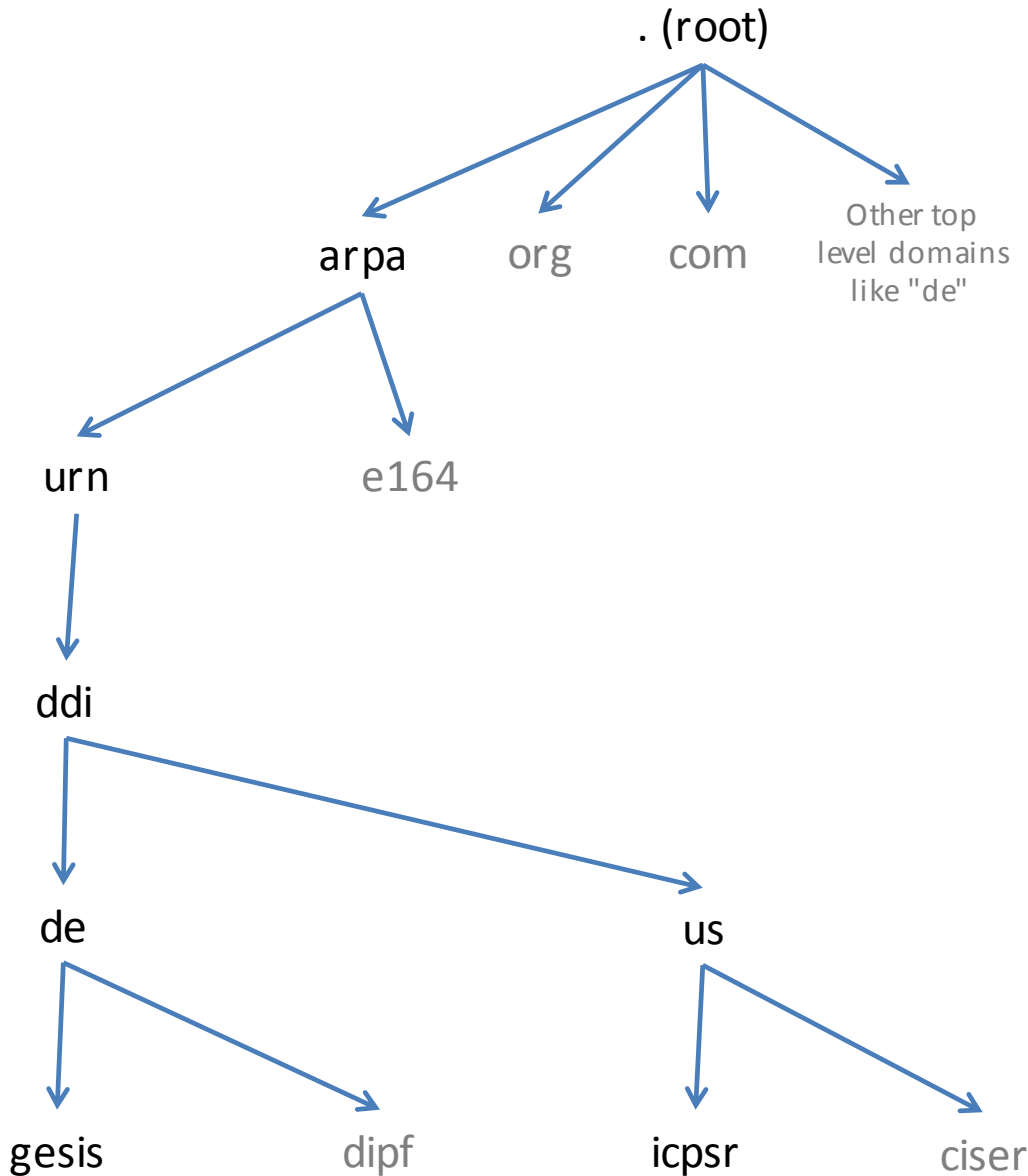# DNS Delegation and Resolution for DDI URNs
Hierarchy and Example Configuration

DNS Delegation and Resolution for DDI URNs
Hierarchy and Example Configuration

. (root)

arpa   org   com   Other top level domains like "de"

urn   e164

ddi

de   us

gesis   dipf   icpsr   ciser

a.iana-servers.net
Delegation
ddi.urn.arpa.
⟶ dns.ddialliance.org.

dns.ddialliance.org
Delegation
gesis.de.ddi.urn.arpa.
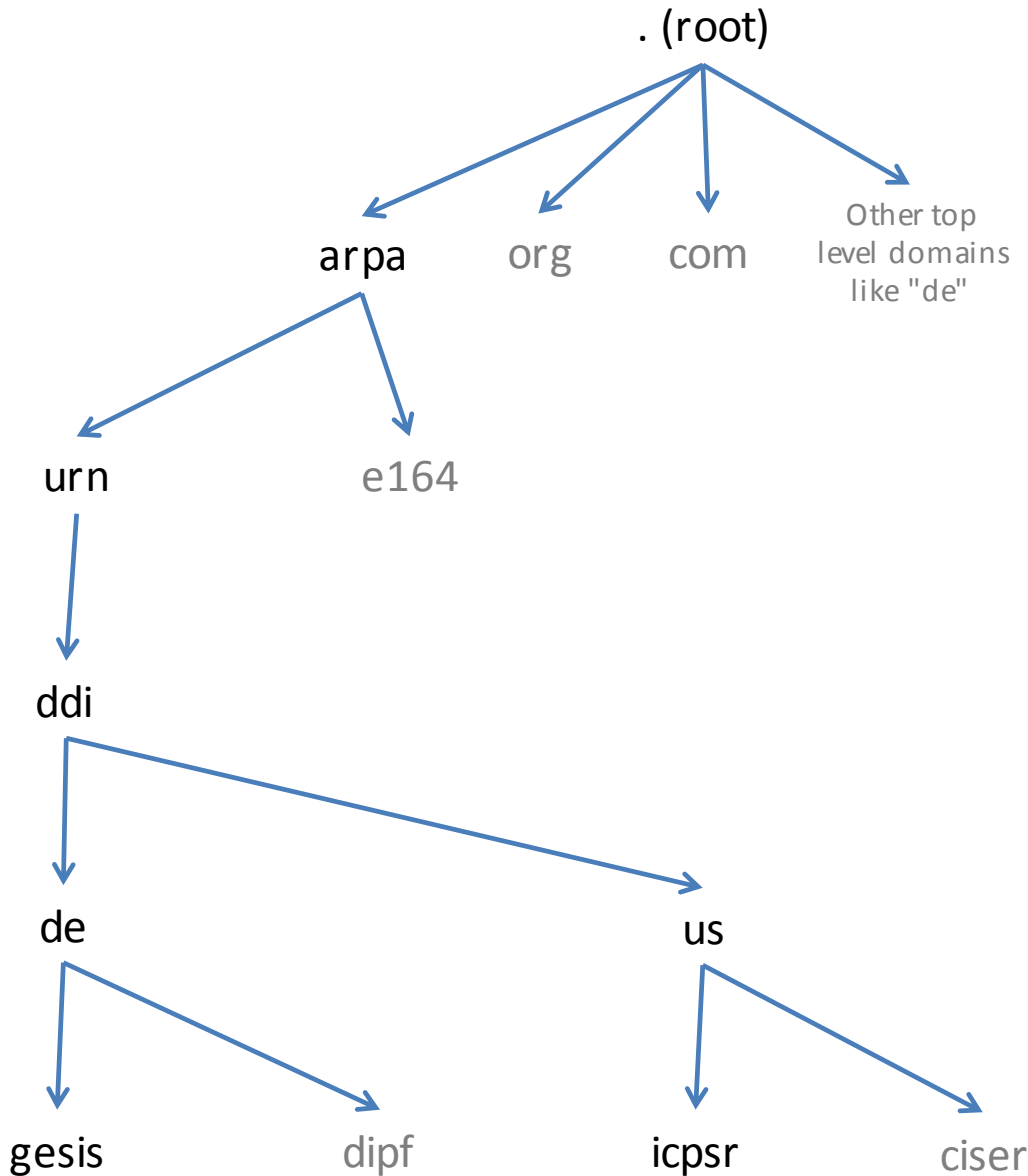⟶ dns.gesis.org.
icpsr.us.ddi.urn.arpa.
⟶ dns.icpsr.umich.edu.
Resolution
*.ddi.urn.arpa.
⟶ http://centralrepository.ddialliance.org/

dns.gesis.org
Resolution
gesis.de.ddi.urn.arpa.
⟶ http://repository.gesis.org/
*.de.ddi.urn.arpa.
⟶ http://centralrepository.gesis.org/

DNS Details

- Delegation to name servers of DDI agencies by NS records

- Resolution of an DDI agency id to a DDI service by
    - NAPTR records (base URL can be specified)
    - Combination of NAPTR and SRV records (flexible protocol specification)

- Properties of DDI service can be specified in a detailed way
    - host name, Internet protocol, port, base URL, type of service, priority, replication of services, load balancing

Requirements for DNS-based DDI URN Resolution

- Application for the URN namespace "ddi" by a formal Request for Comments (RFC) document

- DNS servers at ddialliance.org as central entry point for DDI URN resolution. Few configuration records (ca. 3) for each DDI agency

- DNS configuration for DDI services in DNS server of each DDI agency

Extensibility

- Delegation to DNS servers of sub-agencies is possible
  - For DDI objects below urn:ddi:project1.de.gesis: dns.gesis.org can delegate to dns.project1.gesis.org
- An additional delegation level can be introduced on the country level, when the amount of DDI agencies increases
  - Agency ids must have a country code like "de.gesis", international organizations use "int".
- For specific purposes, a resolution for the URN of single DDI objects can be configured
  - The planned DNS-based resolution is actually providing services for DDI objects of a DDI agency, it is not a URN resolution

DNS-based DDI URN Resolution Summary

- Lightweight approach
- Main focus is the level of the DDI agency
- Can point to different DDI services in a flexible way
- Existing DNS infrastructure is used
- Efficient processing possible, because DNS cache structure is used, and the resolution middleware can additionally cache the query results.
- Extension possible: additional delegation on country level, resolution for single DDI objects

DDI Services

- Different DDI services will be available
- Simple repository serving DDI objects
- Full registry with index and search
- Major use case is probably the simple DDI repository
  - Standard query protocol should be available

Query Protocol Proposal

- REST-based approach, i.e. an URL represents a DDI object
  - REpresentational State Transfer (REST) can be understood as a "simple web service"
  - REST is an architecture style not a standard
- Query uses only HTTP GET and the HTTP error codes, e.g. "404 not found"
- REST is strong in infrastructure reusability
  - HTTP Framework with features like access control, encryption, compression, response caching

Query Protocol: Structure

- <URL of service> (like http://ddirepos.gesis.org/)
- Usage of query parameters for all properties of requested object
  - Name/value pairs are robust, no positional parameters like in a path
  - Query parameters have exact meaning, no ambiguity like with HTTP content negotiation
  - Query parameters can be easily processed by client and server software.
  - Query String is extensible, additional parameters can be added in future

Query Protocol: Parameters
Single DDI object

- urn: URN of the requested object in DDI URN syntax

- ddiVersion: <Version of DDI>

- resolveReferences: yes | no | asIs

- view: complete, index, …

- mimeType: <MIME type of output format> (can make sense for proxy service)

Query Protocol: Response

- DDI instance wrapped in DDIInstance
- Valid DDI
    - At least valid according to DDI XML Schemas
    - Preferable valid according to secondary validation tools
    - DDI instance is valid according to a DDI profile related to a specific purpose

Query Protocol: Parameters Repository-specific

- Parameters for indexing and harvesting purposes (loosely related to OAI-PMH)
  - repository:
    - listObjects (list of available DDI objects)
    - listVersions (list of available DDI versions)
  - elementType: <DDI element name>
- Response can be represented as a DDI instance with the answer items as variables and the data (list of items) as DataSet in-line

Acknowledgements

- Peter Koch from DENIC (central registry for all domains under the top level Domain .de)
- Ad-hoc group at IASSIST 2009 in Tampere
- Dan Smith from Algenta