

# **DESIRE II LDAP Indexing System and Metadata Enhanced Data Management**

**GRID Work Shop  
Bologna 26.6.2000**

**Peter Gietz, DFN Directory Services**

**[Peter.Gietz@directory.dfn.de](mailto:Peter.Gietz@directory.dfn.de)**

## **Table of contents**

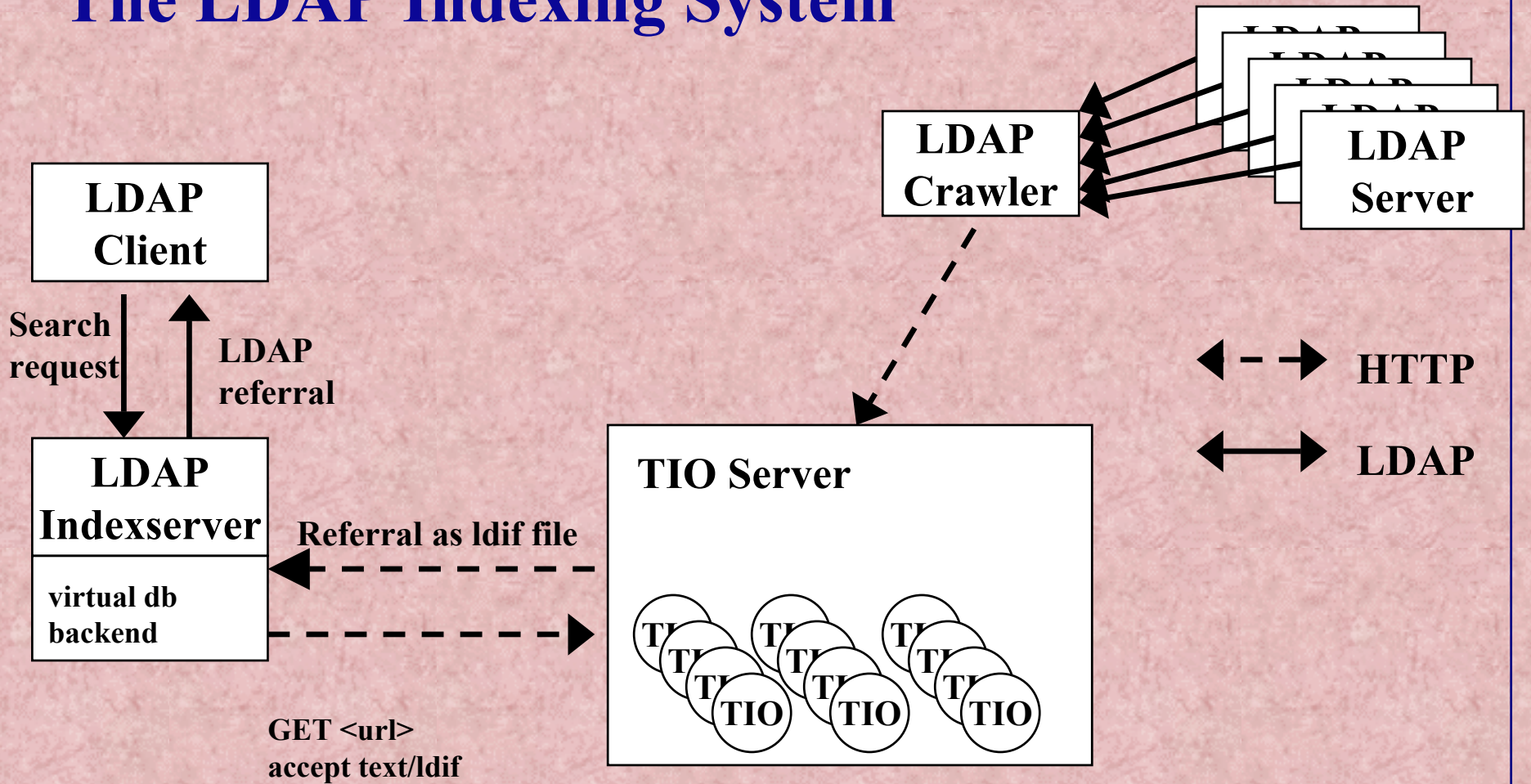
- **I. The DESIRE II Indexing system**
  - **Distributed Indexing System**
  - **Gathering and distribution of Index Objects**
  - **Query Routing**
  - **Architecture of the Referral Server**
  - **Security Considerations**
- **II. Usage for metadata enhanced web indexing**
  - **Requirements**
  - **Metadata formats**
  - **Other LDAP based Projects**
  - **DSML**
  - **Architecture proposal**

# **I. The DESIRE II Indexing system**

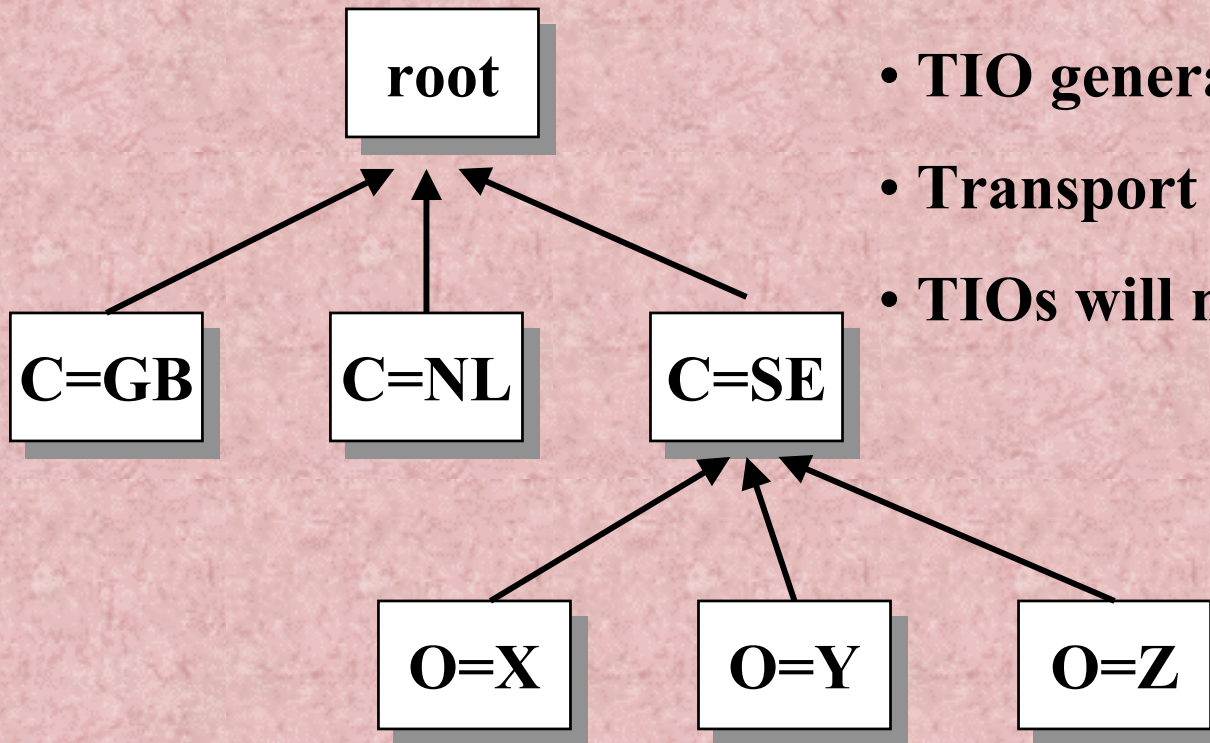
## **Distributed Index system**

- **Hierarchical topology**
- **LDAP v3 technology**
- **Managed by the server side**
- **Index server registration**
- **Subset of CIP**
  - **Dataset Identifier (DSI)**
  - **Base URI for generating referrals**
- **Usage of the Tagged Index Object (TIO)**
  - **Tag identifies common attributes of an entry**

# The LDAP Indexing System



# Index Gathering



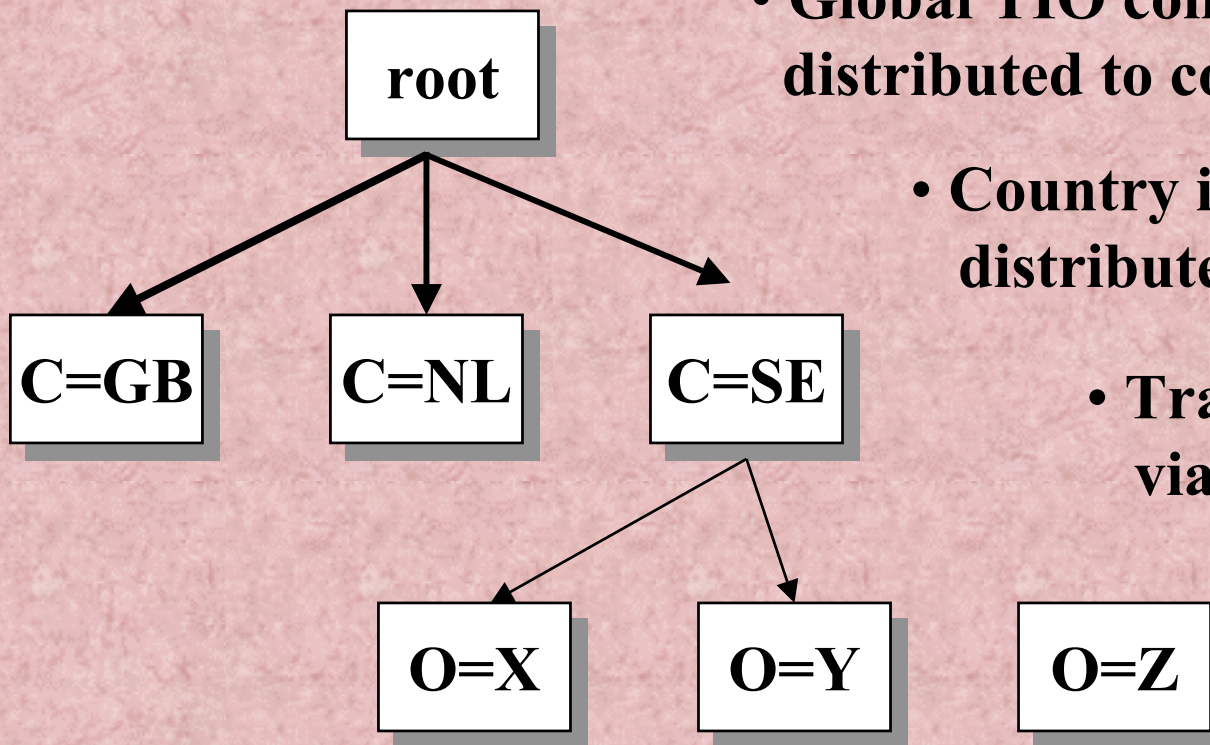
- TIO generated by crawlers
- Transport encrypted via HTTP
- TIOs will not be aggregated

## Index Distribution

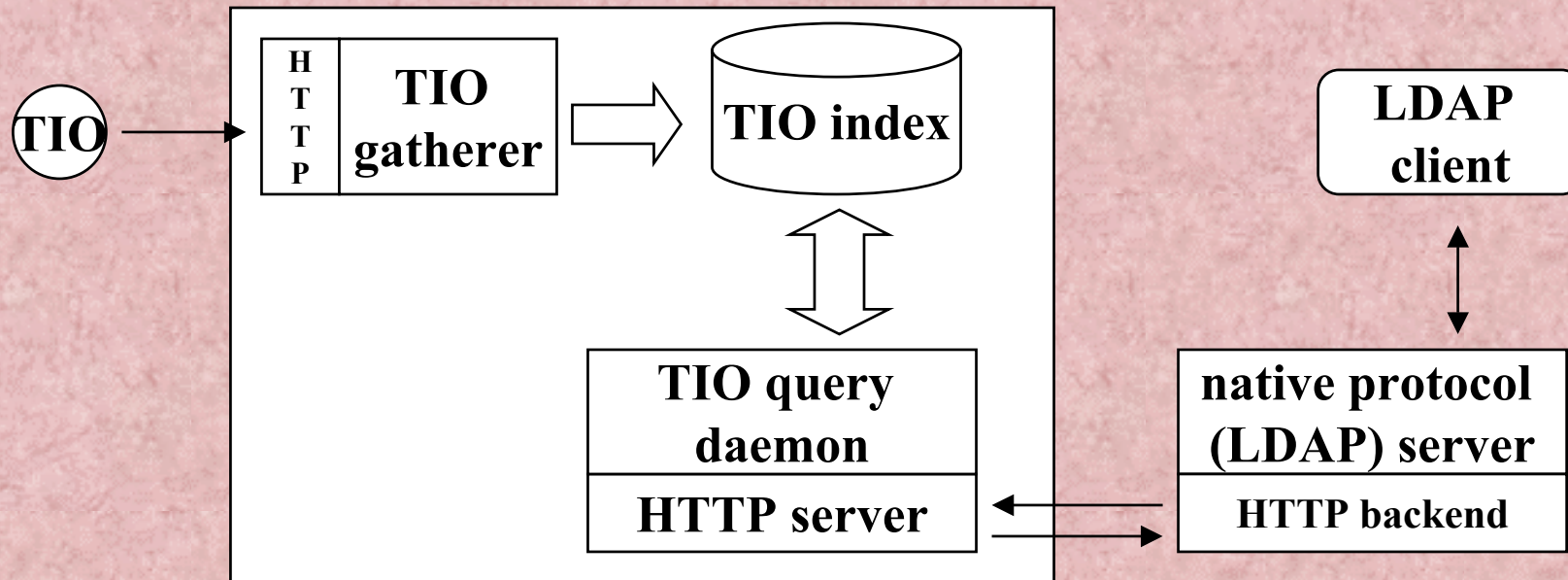
- **Global TIO collection distributed to country level**

- **Country index can be distributed downwards**

- **Transport encrypted via HTTP**



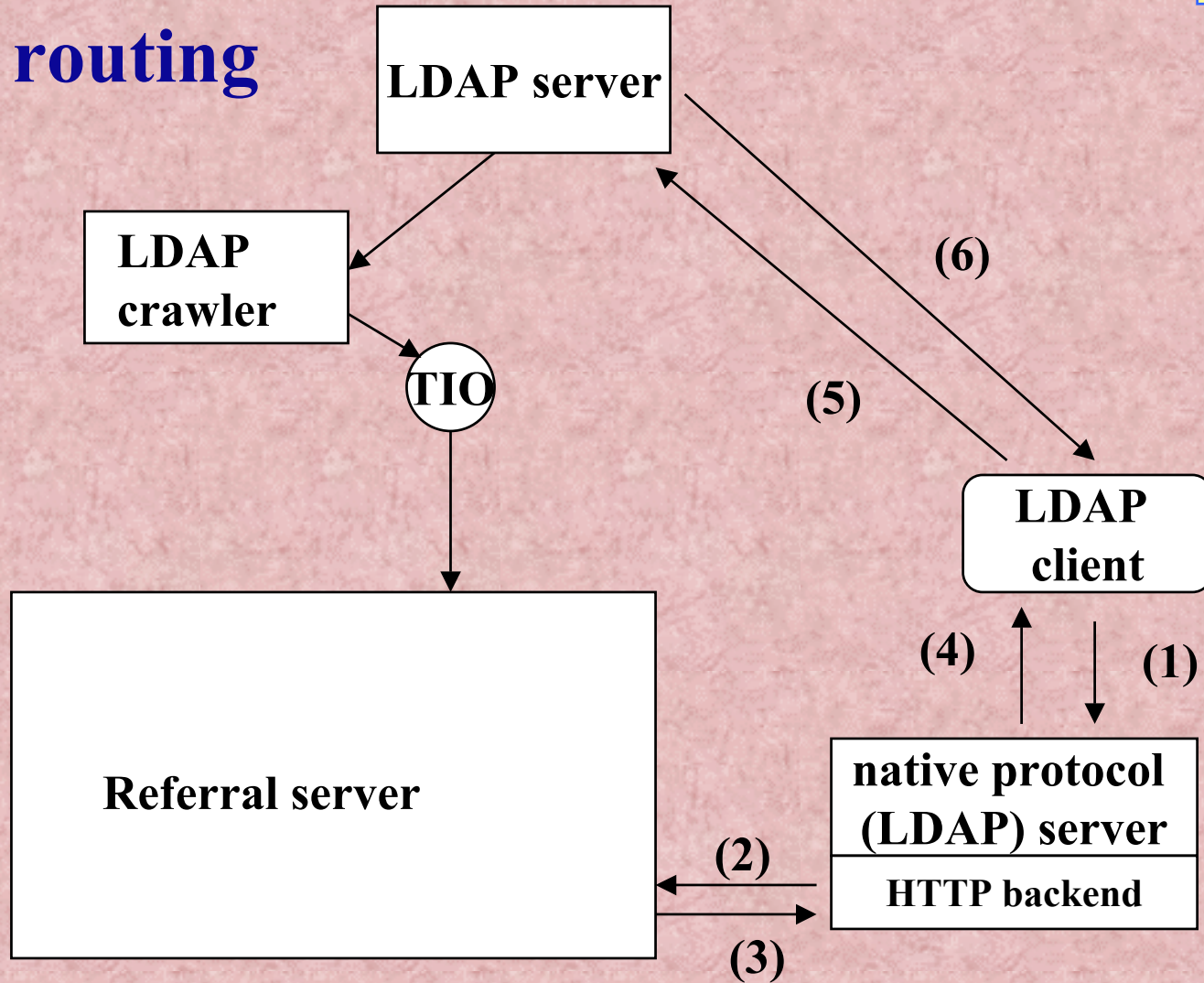
# Referral Server Architecture



**Http request: GET ldap://hostport/c=nl??sub?(cn=\*pers\*) Accept:text/ldif**  
**Http response: Content-Type:text/ldif dn:ref=ldap://host/o=abc,c=nl**



# Query routing



## **Security Requirements**

- **Personal Data are subject to privacy legislation**
- **Public data have different status in collections**
- **We don't want to serve spammers**
- **Participating applications should be known**

## **Security Solutions**

- **All Index objects will be encrypted while on the net**
- **PGP encrypted S/MIME RFC 2015**
  - **Transport protocol independent**
- **Data server registration**
- **Crawler policy stored in the data server**
- **Crawler registration**
- **Referral Server will give back a limited amount of referrals**

## Current State

- **Implementation ready:**
  - **LDAP crawler that outputs LDIF and starts a LDIF2TIO converter and a MIME wrapper to send the TIO to the TIO index server**
  - **HTTP Server with SQL database to store and retrieve TIO information**
  - **TIO-backend for the openLDAP SLAPD that queries the TIO server**
  - **LDAPv3 client**
- **A ad hoc working group is about to define a CIP/TIO based service, which will include:**
  - **This DESIRE II implementation**
  - **TISDAG implementations**
  - **Implementation by Catalogix**

## **Results of the LDAP Deployment BoF on Directory-Indexing**

- **Meeting held on May 12, 2000 TERENA office, Amsterdam**
- **Agreement to set up a TERENA Task Force**
- **One of the Tasks: Directory-Indexing Service**
  - **As follow up on DESIRE II work and the SUNET TISDAG Project**
  - **Pilot: European Academics White and Yellow pages Service**
  - **Agreement on an index object schema**
  - **Pilot will use DESIRE software and the GIDS of Catalogix**
  - **DANTE will provide the pilot's top level index object repository**
- **Other subjects of the meeting:**
  - **DIRECT Project on LDAP DIT and PKI and LDAP**

## **II. Usage for metadata enhanced web indexing**

## **Requirements for a distributed metadata index**

- **Data maintained de-central**
- **Variety of metadata formats**
  - **DC, MARC, SOIF, GILS**
- **Variety of representation of metadata formats**
  - **RDF, RDM, LDIF, HTML-header**
- **Publishing of schemas via metadata registries**
- **Conversion of XML based schemas to LDAP (DSML)**
- **LDAP schemas for the metadata formats**
- **CIP and TIO**

## **Existing LDAP based metadata projects**

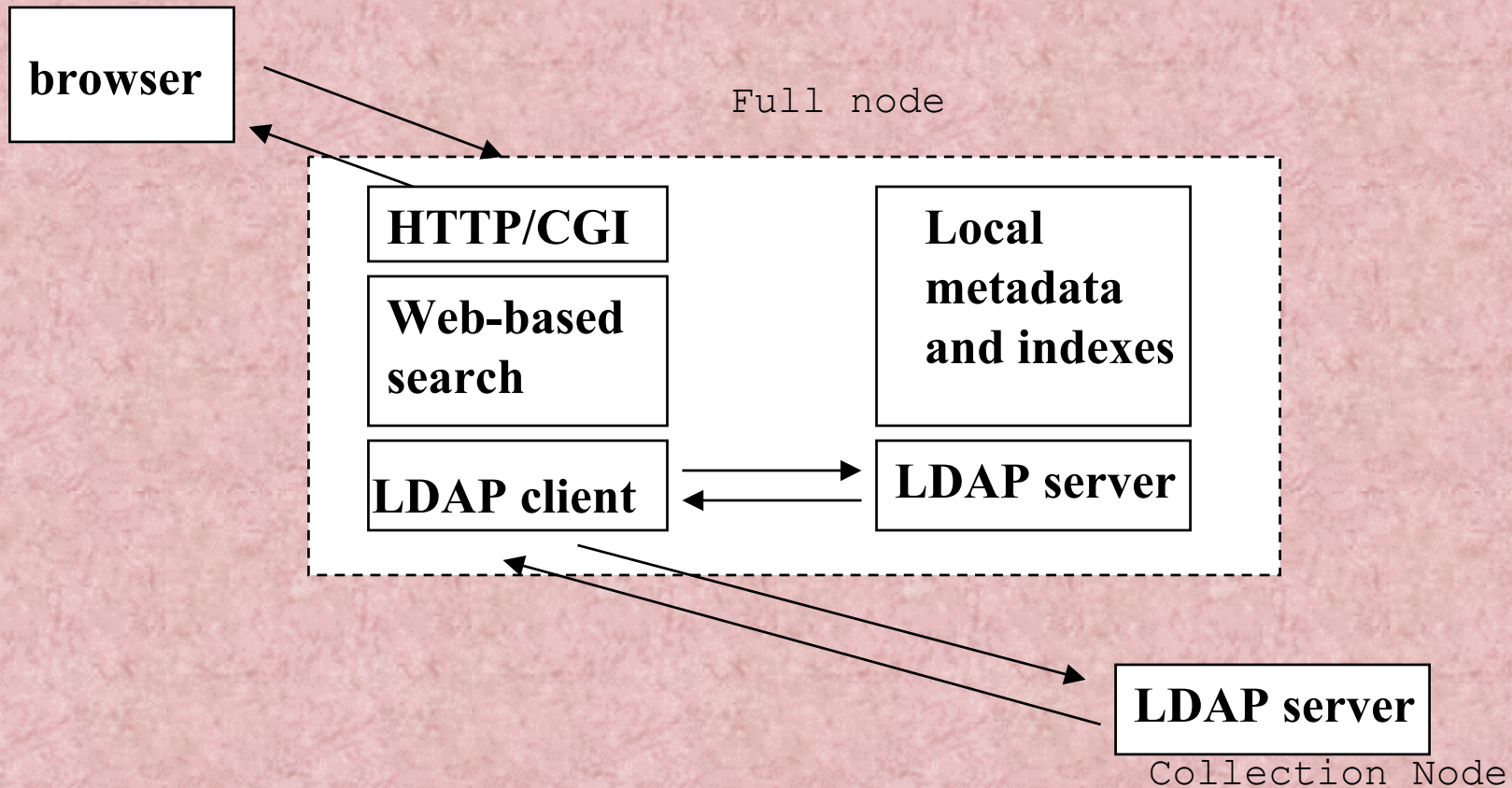
- **Isaak Network**
- **IMesh**



## **Isaac Network of the Internet Scout Project**

- **Current status unknown**
- **Distributed architecture for resource discovery using metadata**
- **Metadata standard DC as common base**
- **Metadata repository based on LDAP servers**
- **Indexing service based on CIP with TIO**
- **Search interface web based (HTTP/HTML)**
- **Includes a Whois++ Gateway to include ROADS based information**

# Isaak Project (contd.)



## **IMesh Toolkit**

- **UK/US Project on distributed subject gateways (SG)**
- **Sept. 1999 - August 2002**
- **Aims are:**
  - **Develop overall framework for SGs**
  - **Framework for interoperating between various SGs**
  - **Create favorable environment for implementing systems and services**
- **Specific objectives:**
  - **Subject gateway architecture, APIs, tools for SG management**
  - **Integrated development environment for metadata**
  - **Metadata registry**

## **IMesh (contd.)**

- **Conversion between different metadata formats (DC, MARC, ROADS/IAFA)**
- **Support for different retrieval protocols (LDAP, Whois++, Z39.50)**
- **Forwarding of knowledge via CIP using the TIO**

## **DSML (Directory Service Markup Language)**

- **Means for representing directory information as an XML document**
- **Directory enhancement for XML based applications**
- **Similiar XMLDir from Novell used as Data interchange format**
- **Can be used to convert RDF data to directory data**
- **A DSML document can describe:**
  - **directory entries**
  - **directory schema**
  - **both**

# DSML Example

```
<dsml:dsml xmlns:dsml="http://www.dsml.org/DSML">
<dsml:directory-schema>
<dsml:class id="person" superior="#top" type="structural">
  <dsml:name>person</dsml:name>
  <dsml:description>objectclass for Person</dsml:description>
  <dsml:object-identifier>2.5.6.6</dsml:object-identifier>
  <dsml:attribute ref="#cn" required="true"/>
  ...
  <dsml:attribute ref="#description" required="false"/>
  ...
</dsml:class>
```

## DSML Example (contd.)

```
<dsml:directory-entries>
<dsml:entry dn="cn=Damy Mahl, o=Brunel University,c=GB">
  <dsml:objectclass>
    <dsml:oc-value>top</dsml:oc-value>
    <dsml:oc-value>person</dsml:oc-value>
  </dsml:objectclass>
  <dsml:attr name="cn"><dsml:value>Damy Mahl</dsml:value></dsml:attr>
<dsml:attr name="cn"><dsml:value>Damy Mahl</dsml:value></dsml:attr>
<dsml:attr name="mail">
  <dsml:value>damy@brunel.gb</dsml:value>
  <dsml:value>damy@xyz.com</dsml:value>
</dsml:attr>
</dsml:entry>
</dsml:directory-entries> </dsml:dsml>
```

## **TIO Index for metadata**

- **Web crawler crawls webpages and stores metadata in local LDAP server including URI.**
- **XML encoded Metadata can also be stored into LDAP server**
- **Agent could convert other XML based metadata repositories into LDIF via DSML to include the data in the LDAP server.**
- **LDAP crawler crawls all LDAP servers holding metadata information, produces LDIF files and starts the following:**
  - **Tools convert LDIF data to TIOs (or SOIF?).**
  - **A MIME wrapper sends TIOs to the TIO referral server.**
- **Referral server stores TIOs into SQL data base.**

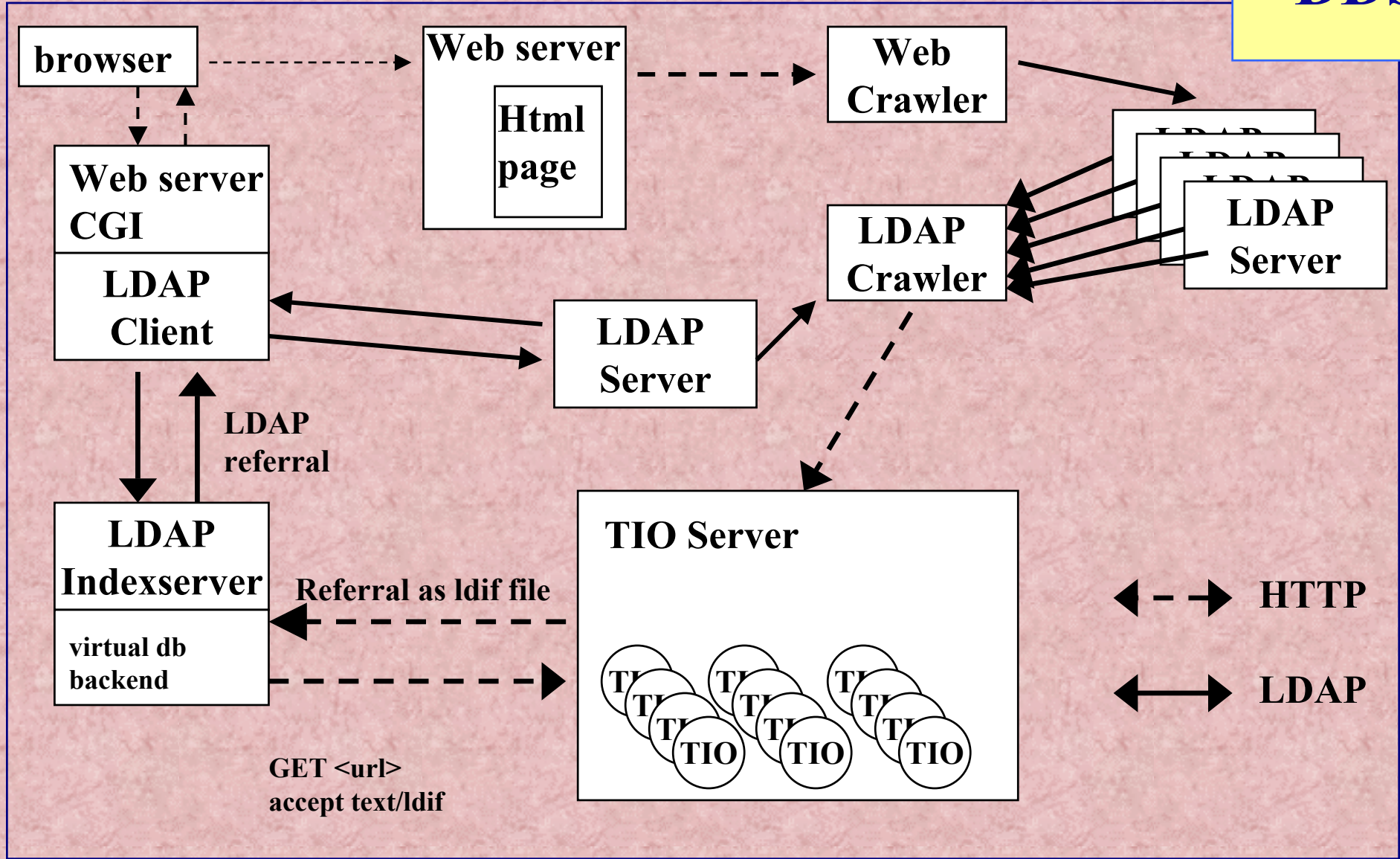


## **TIO Index for metadata (contd.)**

- **Web based query-interface has LDAP client as backend**
- **Backend LDAP client:**
  - **Performs search at LDAP server with TIO backend.**
  - **Follows the referral and retrieves the data.**
  - **Converts the data into HTML.**
- **Interface displays result as hyperlinks to the indexed web page.**

# TIO Index for metadata

**DDS**



## Partners in DESIRE II / More Information

- **Partners**
  - SURFnet
  - DANTE, Cambridge
  - University of Brunel
- **More Information:**
  - <http://www.desire.org>
  - [Peter.Gietz@directory.dfn.de](mailto:Peter.Gietz@directory.dfn.de)
  - [draft-gietz-ldapindex-00.txt](#)
  - <http://www.directory.dfn.de>