LDAP based repositories for Metadata and Ontologies

NetLab & Friends Conference
Lund, 10. April 2002

Peter Gietz

Peter.gietz@DAASI.de

DAASI International
Directory Applications for Advanced Security and Information Management

Agenda

- An Introduction to LDAP
- LDAP, Common Indexing Protocol and Metadata
- LDAP, Common Information Model and Ontologies

DAASI International

Directory Applications for Advanced Security and Information Management

Directory in German Research environment

- Since 1994 DFN research projects at University of Tübingen:
  - AMBIX an Email directory
  - DFN Directory Services (DDS)
    - Directory competence center
- Since January 2001: DAASI International GmbH
  - Directory Applications for Advanced Security and Information Management
  - Design, implementation and management of directory services
  - Main Customers: Research Institutions in Europe (NRNs, Universities, etc.)

An Introduction to LDAP

Features of a Directory service

- It is a database
  - for storing and retrieving information
- It is a specialized database
  - designed for fast reading, writing is slower
  - static view on the data
  - simple updates without transactions
- It has a network protocol for access
- A Directory Service may include
  - distribution in the net (scalable!)
  - replication of the data (reliable)

DAASI International
Directory Applications for Advanced Security and Information Management

What kind of data can you store?

- Text data
  - names, addresses, descriptions, numbers, etc.
- Pointers
  - URIs, pointers to other data, etc.
- Public key certificates
- Graphics
  - photos, diagrams, etc.
- Other binary data
- Anything else you can think of

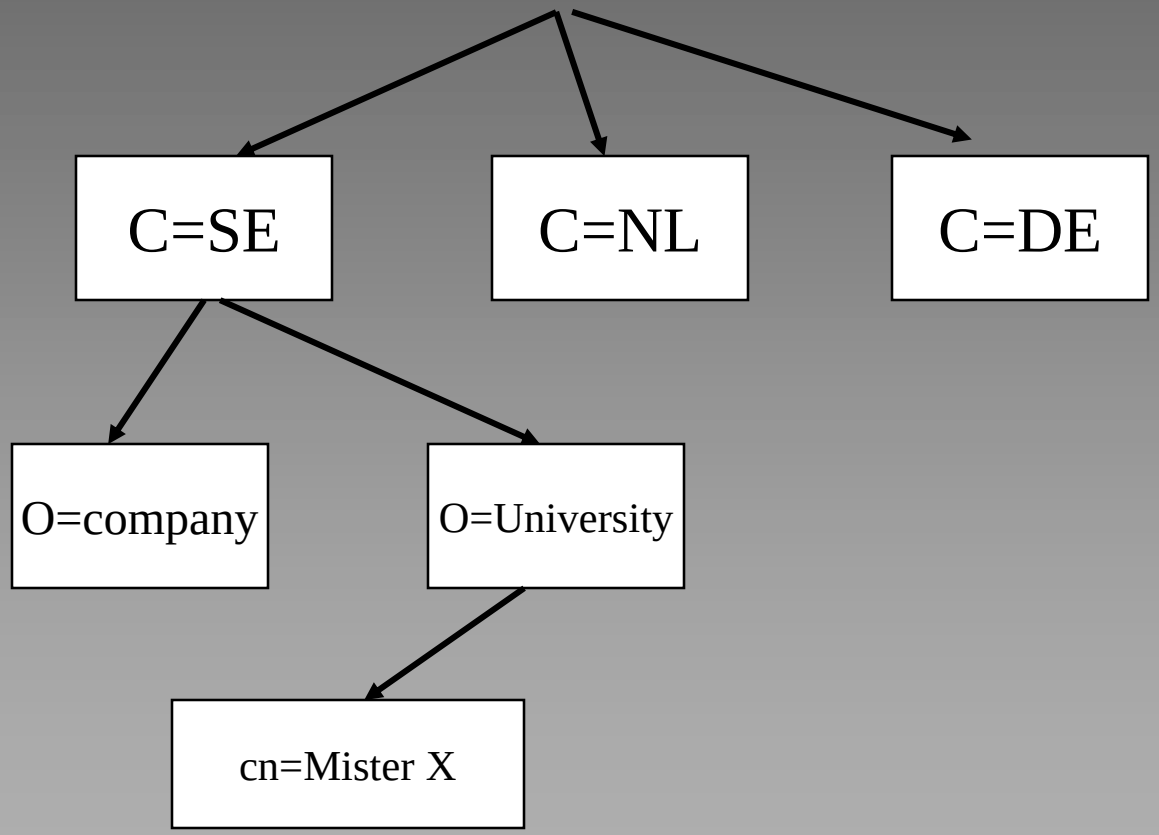
DAASI International
Directory Applications for Advanced Security and Information Management

Directory Information Tree

- Data are stored in entries
- Entries are ordered as tree nodes
- In the Directory Information Tree (DIT)
  - Every node has 0 to n children nodes
  - Every node except root has 1 parent node

DAASI International
Directory Applications for Advanced Security and Information Management

Directory Information Tree (DIT)



DAASI International
Directory Applications for Advanced Security and Information Management

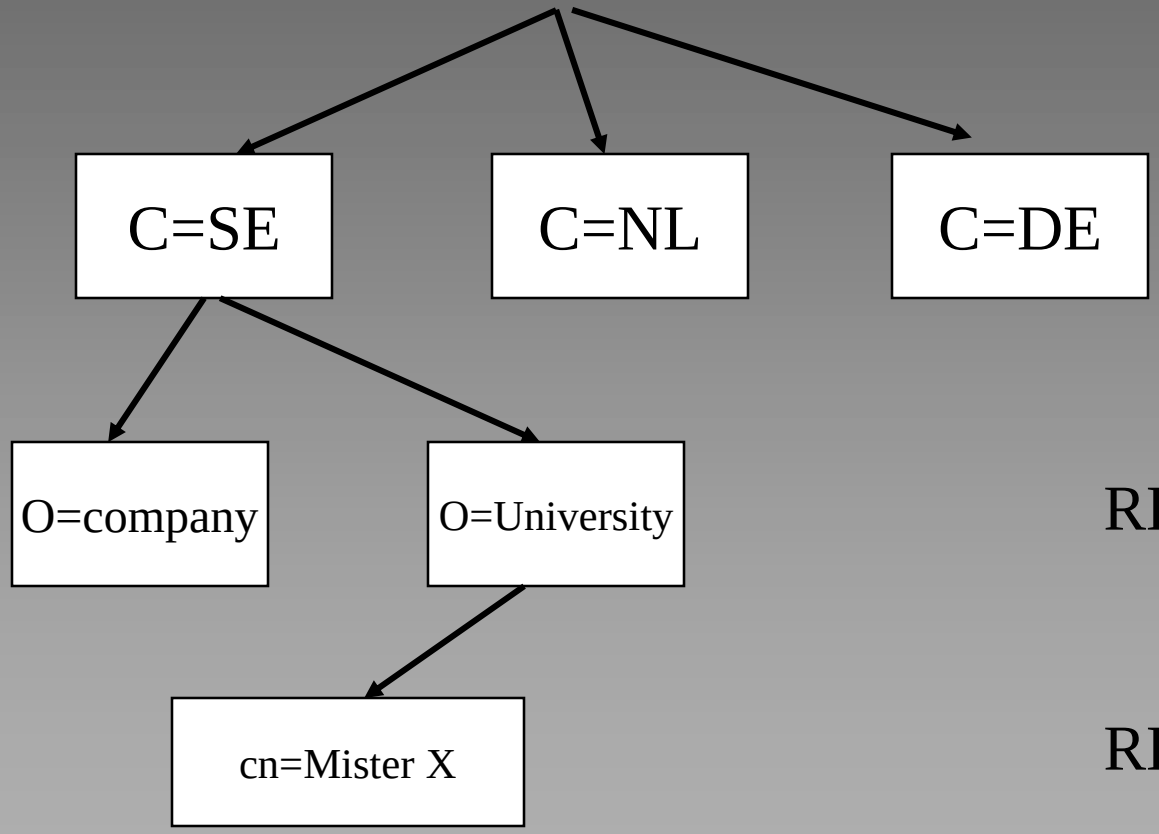
DN Distinguished Name

- An entry has a distinguished name
  - **in its hierarchy level: Relative Distinguished Name (RDN)**
  - **all RDNs from root onwords build the Distinguished Name (DN)**
- No two entries in one hierarchy level can have the same RDN
- Thus no two entries in the whole Directory can have the same DN

DAASI International

Directory Applications for Advanced Security and Information Management

Directory Information Tree (DIT)



RDN: C=NL

RDN: o=University

RDN: cn=Mister X

DN: c=NL,o=University,cn=Mister X

DAASI International
Directory Applications for Advanced Security and Information Management

OIDs

- An Entry is an information object
- The mechanisms for representing the data are objects as well, identified by an OID (Object Identifier)
  - E.g.: 1.234.567.8.123
- OIDs are again represented in an hierarchical tree
- OIDs are world wide unique

DAASI International
Directory Applications for Advanced Security and Information Management

X.500 Information Model

- An Entry contains a number of Attributes
- An Attribute consists of:
  - Attribute Type
  - Attribute Value
- An Attribute Type has an associated Attribute Syntax
- The Attribute Value has to conform to that syntax
- To compare Attributes there are Matching Rules

Special Attributes

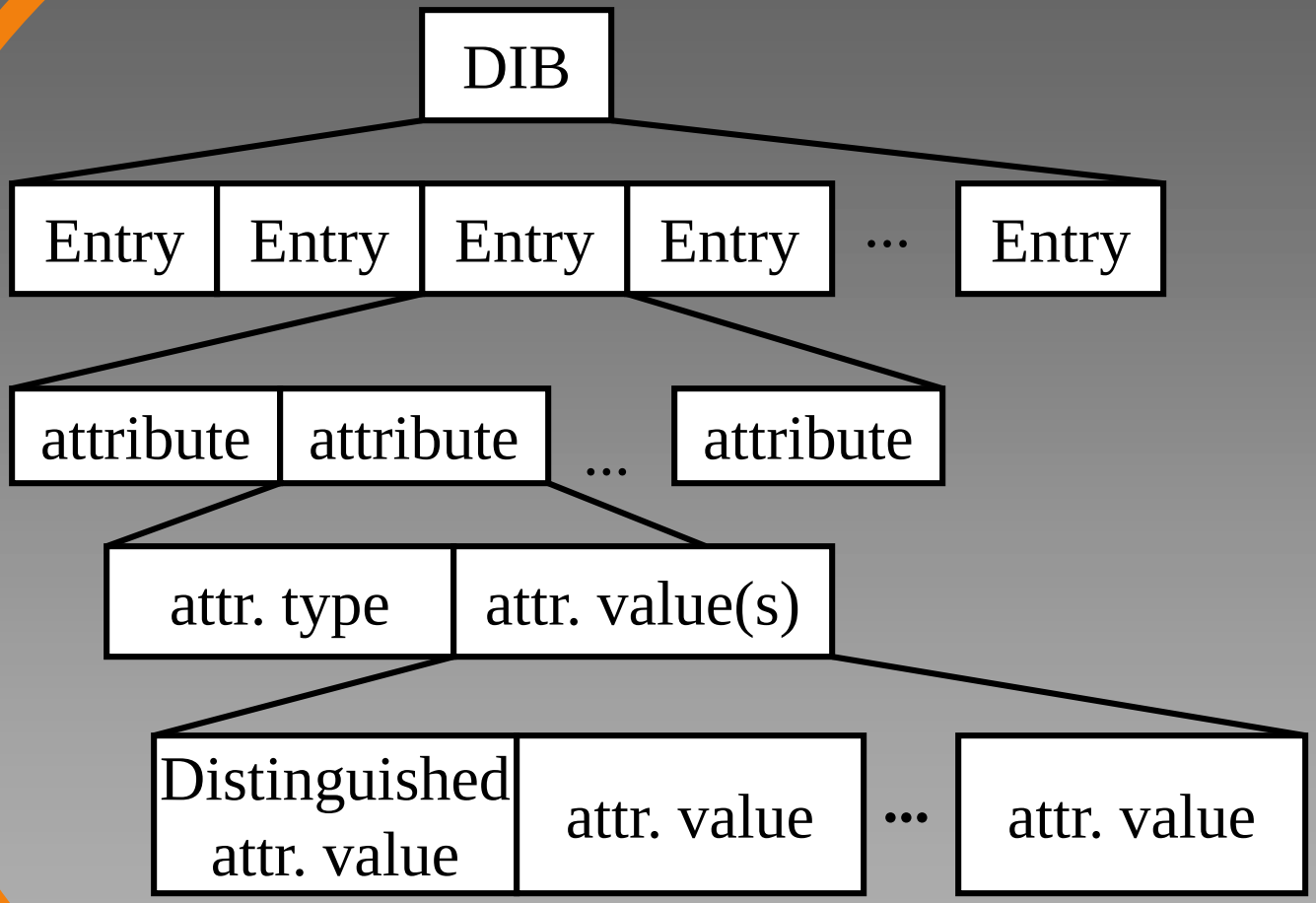
- One or more Attribute Types form the RDN
  - The Naming Attributes or
  - The Distinguished Attributes
- An Entry must have one or more Objectclass Attributes which:
  - Characterizes the Entry, e.g. Person
  - Defines a set of usable Attributes the entry may contain and must contain
- Objectclasses can inherit Attributes from other Objectclasses
- A set of Objectclasses, Attributes and Syntaxes for a special purpose are called schema

DAASI International
Directory Applications for Advanced Security and Information Management

Special Attributes contd.

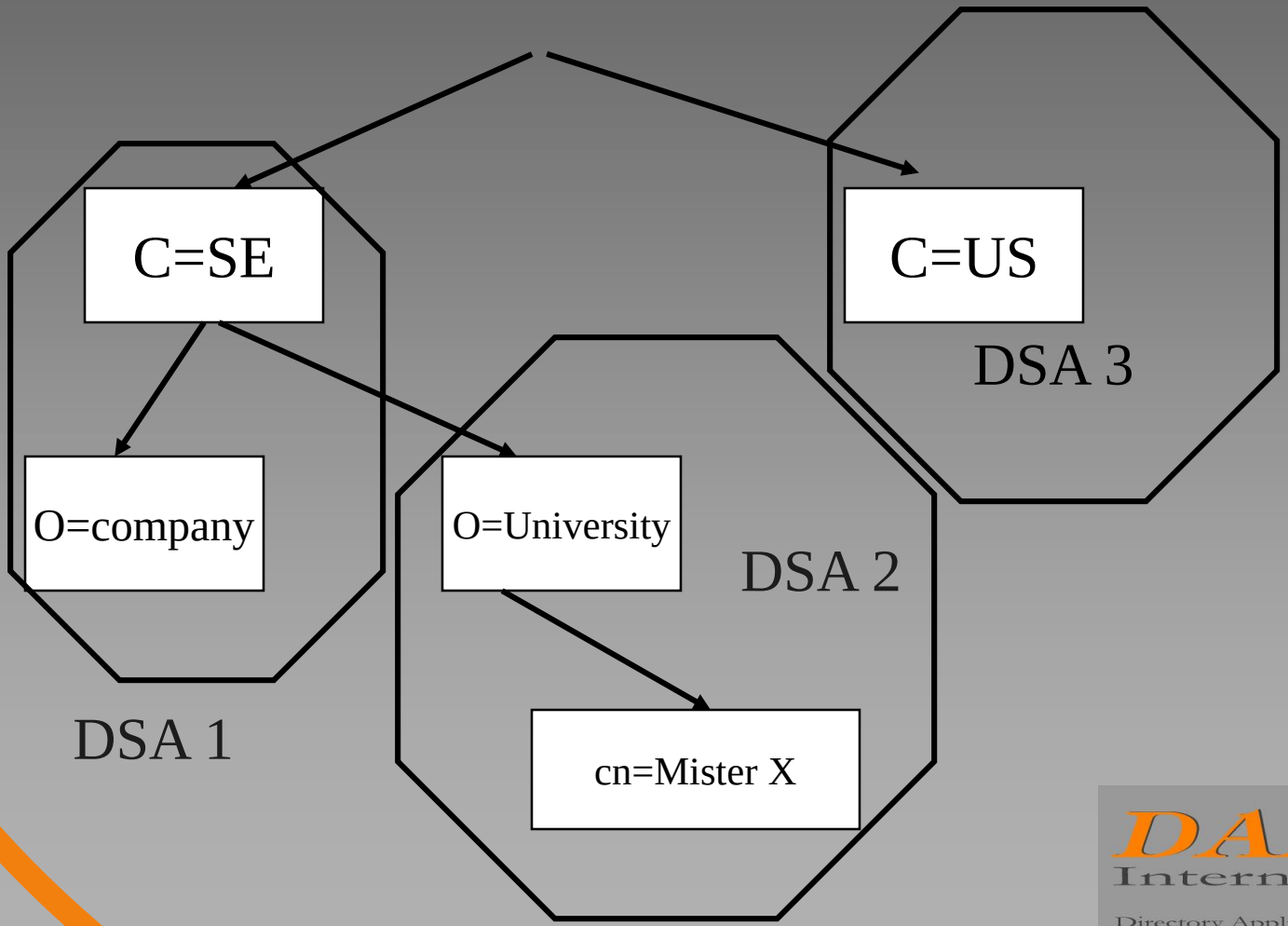
- aliasObjectName Attribute
  - **Alias Entries have a DN and point to another DN via aliasObjectName Attribute**
- seeAlso Attribute
  - **Entry contains data and a seeAlso pointer to another DN with related data**

Directory Information Base



DAASI International
Directory Applications for Advanced Security and Information Management

Distribution of the data among DSAs



DAASI International
Directory Applications for Advanced Security and Information Management

Client Server System

- Originally (v1,v2) LDAP was just a client access protocol for X.500
- LDAP v3 is a whole client server system
  - LDAP does not provide a chaining mechanism
  - Instead server can send referrals to clients
  - Referral is part of LDAPresult structure to indicate that the server does not have the requested data but the servers referred to might have it
- Implementations have server replication mechanisms

DAASI International
Directory Applications for Advanced Security and Information Management

Security Mechanisms

- Several Authentication mechanisms
  - Bind with password
  - SASL mechanisms
- Session encryption
  - TLS
- Access control mechanism
  - On subtree, entry and attribute level
  - Different identifications
    - AuthenticationID, IP address, ...
  - Not yet standardized

LDAP Functional Model

- Authentication and control operations:
  - **bind**
  - **unbind**
  - **abandon**
- Interrogation operations:
  - **search**
  - **compare**
- Update operations:
  - **add**
  - **delete**
  - **modify**
  - **modifyDN**

Search Filter Operators

- **Equality**
  - **e.g.: (cn=Mister X) only entries with common name equals “Mister X”**
- **Negation operator**
  - **e.g. (!(cn=Mister X)) all entries but the one with cn equals “Mister X”**
- **Substring**
  - **e.g. (cn=Mister*) all entries with cn beginning with “Mister”**
- **Approximate**
  - **e.g.: (cn~=Mister) all entries with cn sounding similiar to “Mister”**

DAASI International

Directory Applications for Advanced Security and Information Management

Search Filter Operators (contd.)

- **Greater than or equal to and less than or equal to**
  - **e.g. (sn<=Smith) all entries where sn equals "Smith" or is lexicographically above "Smith" (from sn=Adam to sn=smirnow)**
  - **e.g. (age>21) is not possible, use (!(age<=21)) instead**
- **Presence**
  - **e.g. (telephoneNumber=*) all entries that contain a telephone number**
  - **e.g. (objectclass=*) all entries, since every entry contains at least one objectclass**

DAASI International
Directory Applications for Advanced Security and Information Management

Search Filter Extensions

- LDAPv3 defines an extensible matching filter
  - syntax: attr [“:dn”] [“:” matchingrule] “:=“ value
    - attr is an attribute name
    - “:dn” says that also the attribute in the dn should be searched as well
    - matching rule given by an OID or associated descriptive name
  - examples:
    - (cn:1.2.3.4.5.6:=Mister X) use matching rule 1.2.3.4.5.6 for comparision
    - (o:dn:=company) search for o=company in attributes and also in DN

DAASI International
Directory Applications for Advanced Security and Information Management

Search filter combinations

- Filters can be combined
  - **AND operator: &**
    - **e.g. (& (cn=Mister X) (mail=*dot.com)) only entries that have both cn=Mister X and a mail address ending with dot.com**
  - **OR operator: |**
    - **e.g.: (| (cn=Mister X) (sn=Xerxes)) all entries that have cn=Mister X or sn=Xerxes**

DAASI International
Directory Applications for Advanced Security and Information Management

LDAP URL (RFC 2255)

- Format:
  - ldap://<host>:<portnumber>/<basedn>?<attrlist>?<scope>?<filter>?<extensions>
- Example:
  - ldap://myhost.org:9999/c=SE,o=University?cn,telephonenumber?subtree?(cn=Mister X)
- LDAP URLs are used as referral

DAASI International
Directory Applications for Advanced Security and Information Management

LDAP Data Interchange Format LDIF

- RFC 2849:
  - The LDAP Data Interchange Format (LDIF) - Technical Specification, G. Good, June 2000
- Format for exchanging data
- Example:

```
dn: cn=Mister X, o=University, c=CE
objectclass=top
objectclass=person
objectclass=organizationalPerson
cn=Mister X
cn=Xavier Xerxes
mail=X@dot.com
mail=Mister.X@dot.com
telephoneNumber=1234567

dn: cn=next entry, ...
```

DAASI International
Directory Applications for Advanced Security and Information Management

Who talks LDAP?

- Big number of LDAP implementations
  - OpenLDAP (open source)
  - Implementations e.g. by Sun, IBM, Syntegra, …
- All other directory implementations have an LDAP interface:
  - all X.500(93) implementations
  - Novell Directory Service (NDS)
  - Microsoft Active Directory (AD)
- Many client applications have an LDAP interface:
  - Mail agents
  - Browser
  - PGP clients

DAASI International
Directory Applications for Advanced Security and Information Management

1997: LDAP v3 Proposed Standard

- RFC 2251:
  - Lightweight Directory Access Protocol (v3), M. Wahl, T. Howes, S. Kille. December 1997
- RFC 2252:
  - Lightweight Directory Access Protocol (v3) - Attribute Syntax Definitions, M. Wahl, A. Coulbeck, T. Howes, S. Kille. December 1997
- RFC 2253:
  - Lightweight Directory Access Protocol (v3) - UTF-8 String Representation of Distinguished Names, M. Wahl, S. Kille, T. Howes. December 1997

1997 LDAPv3 contd.

- RFC 2254:
  - The String Representation of LDAP Search Filters, T. Howes. December 1997
- RFC 2255:
  - The LDAP URL Format, T. Howes, M. Smith. December 1997
- RFC 2256:
  - A Summary of the X.500(96) User Schema for use with LDAPv3, M. Wahl. December 1997

IETF WG LDAPbis

- Revision of all LDAP core RFCs
- With references to mandatory security mechanism of RFC 2829 and 2830 possible to go for Draft Standard
- No changes in the data definitions
- Some clarifications in wording
- Some SHOULDS to MUST etc.
- Some additional documents, e.g.:
  - IANA considerations
  - UTF-8 matching

DAASI International
Directory Applications for Advanced Security and Information Management

LDAP, Common Indexing Protocol and Metadata

Common Indexing Protocol CIP

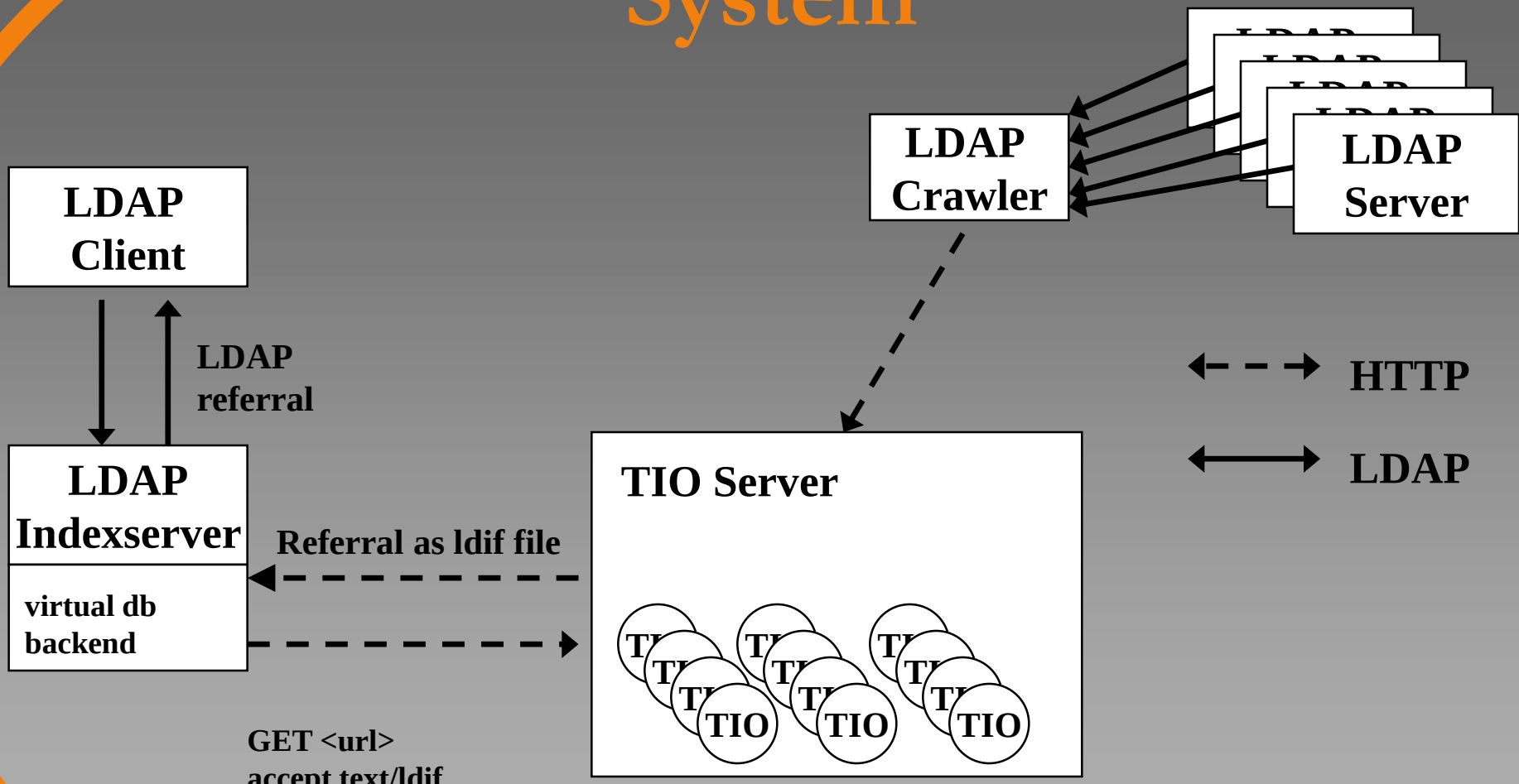
- RFC 2651 – 2655
- Index definitions for any directory technology
- Based on Whois++ Index mesh
  - Server server communication
  - Multiple topologies possible
- MIME wrapper
- Transport protocol

CIP contd.

- Different index object formats
  - SOIF (Summary Object Interchange Format)
  - TIO (Tagged Index Object)
    - Tag identifies common attributes of an entry
- Dataset Identifier (DSI)
  - Identifies server
- Base URI for generating referrals
  - Identifies server and baseDN

DAASI International
Directory Applications for Advanced Security and Information Management

The LDAP Indexing System



DAASI International
Directory Applications for Advanced Security and Information Management

What can the index system be used for?

- White Pages Service
- Metadata indexing service
- Certificate indexing service
  - Based on Internet Draft on X.509certificate object class (draft-klasen-x509certificate-schema-00.txt)
- Web Services repository (with or without a UDDI frontend)
- …

DAASI International
Directory Applications for Advanced Security and Information Management

DSML

- Directory Service Markup Language v1
- Means for representing directory information as an XML document
- Directory enhancement for XML based applications
- Can be used to convert XML data to directory data
- A DSML document can describe directory entries, directory schema or both
- DSML v2 will define LDAP operational model

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:directory-schema>
```

DAASI International

Directory Applications for Advanced Security and Information Management

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="mail">
      <dsml:value>damy@brunel.gb</dsml:value>
    </dsml:attr>
  </dsml:entry>
</dsml:directory-entries>
</dsml:dsml>
```

DAASI International
Directory Applications for Advanced Security and Information Management

Distributed Metadata

- Requirements:
  - 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

DAASI International

Directory Applications for Advanced Security and Information Management

Isaac Network

- Part 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)

LDAP, Common Information Model and Ontologies

Current WWW

- Mere publishing medium
- Huge amount of information
- Designed for human access only
- Lack of structure and organization
- Insufficiant access methods
- Ambiguous:
  - **bank (finance institute) the same as**
  - **Bank (river bank)**

Visions for the future

- Web Services
- Accessed by humans *and* programs
- Quality content
- Better structured
- Knowlegde enhanced
- Disambigued:
  - **Bank (finance institute) is not the same as**
  - **Bank (river bank)**

Buzwords for the new visions

- „Semantic Web“ (Tim Berners-Lee)
- Grid
  - Computational Grid (Foster/Kesselman)
    - Computing power out of the wall
  - Information Grid
    - Information about resources, data and the rest
  - Knowledge Grid
    - Knowledge is relations between concepts and information

How to achieve knowledge

- Metadata
  - Data about information
- Ontologies
  - Concepts and relations between them
  - Computer knows more than inputed

```
Input: Parents have children
Input: Mother = female parent
Output: Mothers have children
```

DAASI International
Directory Applications for Advanced Security and Information Management

Ontology Description

- **E.g.: DAML+OIL (predecessor of WebOnt):**

```
<daml:Class rdf:ID="xxx" rdf:about="#xxx" >
    <rdfs:label>xxx</rdfs:label>
    <rdfs:comment>xxx</rdfs:comment>
    <rdfs:subClassOf rdf:resource="#xxx"/>
    <daml:disjointWith rdf:resource="#yyy"/>
    <daml:Restriction>
            <daml:onProperty rdf:resource="#xxx"/>
            <daml:toClass rdf:resource="#xxx"/>
    </daml:Restriction>
</daml:Class>
```

DAASI International
Directory Applications for Advanced Security and Information Management

Ontology Description 2

```
<daml:UniqueProperty rdf:ID="xxx">
    <rdfs:domain rdf:resource="#xxx"/>
    <rdfs:subPropertyOf rdf:resource="#xxx"/>
    <rdfs:range rdf:resource="#xxx"/>
    <daml:inverseOf rdf:resource="#hasParent"/>
</daml:UniqueProperty>
```

DAASI International

Directory Applications for Advanced Security and Information Management

Ontologie Storage Proposal

- Combined repository for metadata and ontologies based on LDAP technology and thus accessible with the same protocol
- Large scalability by setting up an Indexing system based on Common Indexing Protocol (CIP)
- Ontologie data model based on CIM which provides a model for associations that can be used for mapping the relations between objects

DAASI International

Directory Applications for Advanced Security and Information Management

What could you store?

- Multiple ontologies with links between different ontologies
- General ontologies (e.g. WordNet)
- Special ontologies (e.g. on special subjects)

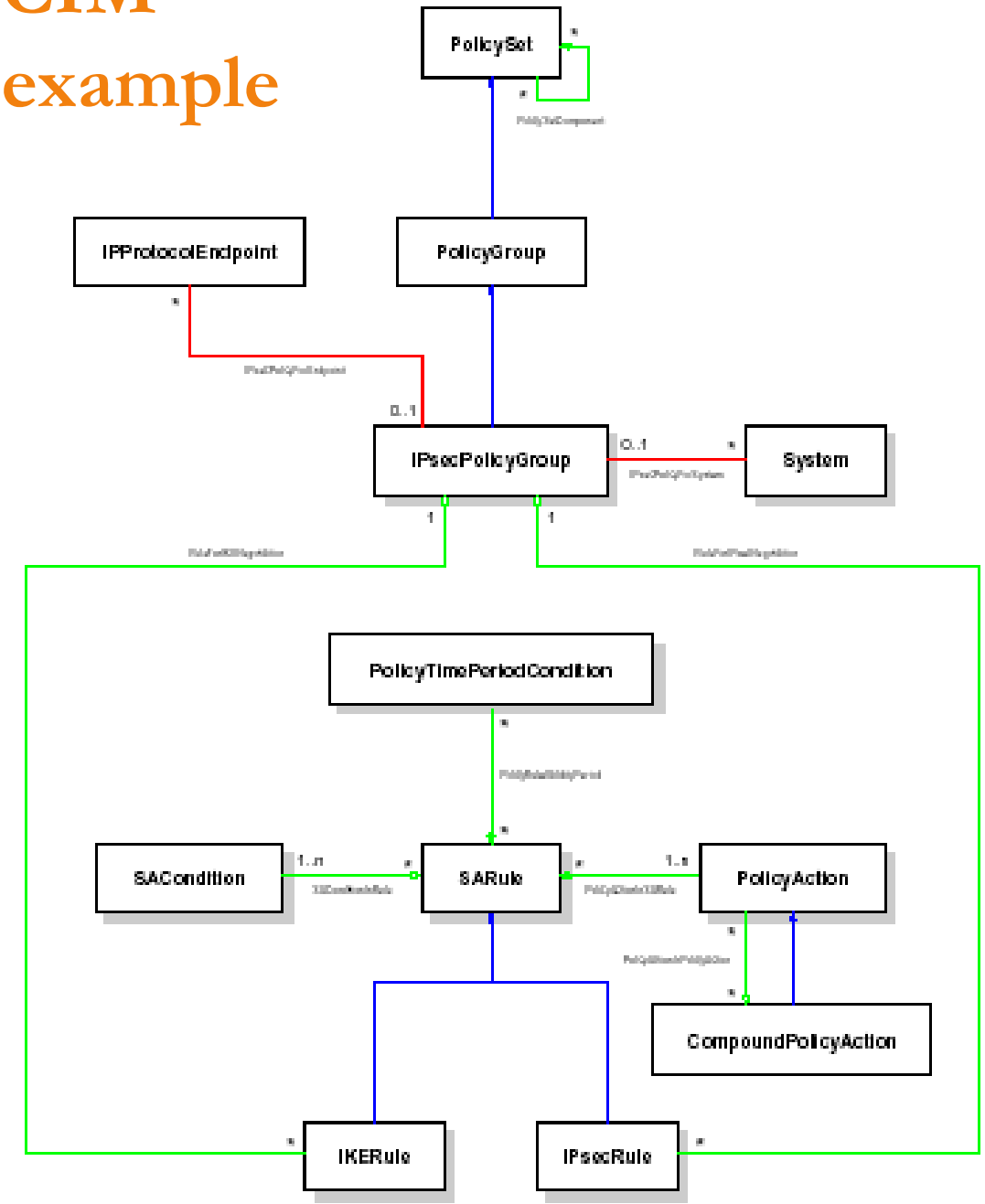
DAASI International
Directory Applications for Advanced Security and Information Management

Common Information Model

- Object oriented meta model for structuring information technology independantly
- Capable of describing the whole computer world
- Basically an Ontology
- Three layers
  - Core: the basic lego bricks
  - Common: standardized descriptions
  - Extesion: vendor‘s extras

DAASI International
Directory Applications for Advanced Security and Information Management

CIM example



objects

inheritance

aggregation

association

DAASI International

Directory Applications for Advanced Security and Information Management

CIM mapped to LDAP 1

- **objectClass** (1.3.6.1.4.1.412.100.2.1.3.60 NAME ' dlm1MemberOfCollection ,
  DESC ' MemberOfCollection is an aggregation used to establish membership of ManagedElements in a Collection .,
  SUP top ABSTRACT)

CIM mapped to LDAP 2

- **attributetype** (1.3.6.1.4.1.412.100.2.2.186 NAME ' dlmMemberOfCollectionCollectionRef ,
  DESC ' The Collection that aggregates members . Values of this attribute point to entries of class dlmCollection .,
  SYNTAX 1.3.6.1.4.1.1466.115.121.1.12
  EQUALITY distinguishedNameMatch)
- **attributetype** (1.3.6.1.4.1.412.100.2.2.187 NAME ' dlmMemberOfCollectionMemberRef ,
  DESC ' The aggregated member of the collection . Values of this attribute point to entries of class dlmManagedElement .,
  SYNTAX 1.3.6.1.4.1.1466.115.121.1.12
  EQUALITY distinguishedNameMatch)

DAASI International
Directory Applications for Advanced Security and Information Management

CIM mapped to LDAP 3

- **objectClass** (1.3.6.1.4.1.412.100.2.1.3.61 NAME ' dlm1MemberOfCollectionAuxClass , DESC ' MemberOfCollection is an aggregation used toestablish membership of ManagedElements in a Collection ., SUP dlm1MemberOfCollection AUXILIARY MAY (dlmMemberOfCollectionCollectionRef $ dlmMemberOfCollectionMemberRef))

DAASI International
Directory Applications for Advanced Security and Information Management

CIM, LDAP and Ontologies

- Any kind of relations can be defined with CIM and mapped to LDAP
- LDAP provides:
  - Object Class inheritance
  - Attribute inheritance
- Associations and aggregations can be mapped by object classes

Questions?

- DFN Directory Services
  - peter.gietz@directory.dfn.de
  - www.directory.dfn.de
- DAASI International GmbH
  - Info@daasi.de
  - www.daasi.de