



UNIVERSITY OF OSLO
FACULTY OF HUMANITIES

The Unit for Digital Documentation (EDD) system for storing coref information

A short overview of a system under development

by Øyvind Eide



Two different views on coref

1. The curator, researcher or exhibition professional using coref as an internal system. Information is added to their local system about external resources co-referring with internal resources
2. The researcher using a network of co-reference information sources to research relations between real-world entities.

These two views can be held by the same individual in different situations, of course.

This presentation is a preliminary report of an implementation of a coref system covering view 1 above. The system is being developed by the Unit for Digital Documentation (EDD) at the University of Oslo.

The examples shown here is related to persons only.



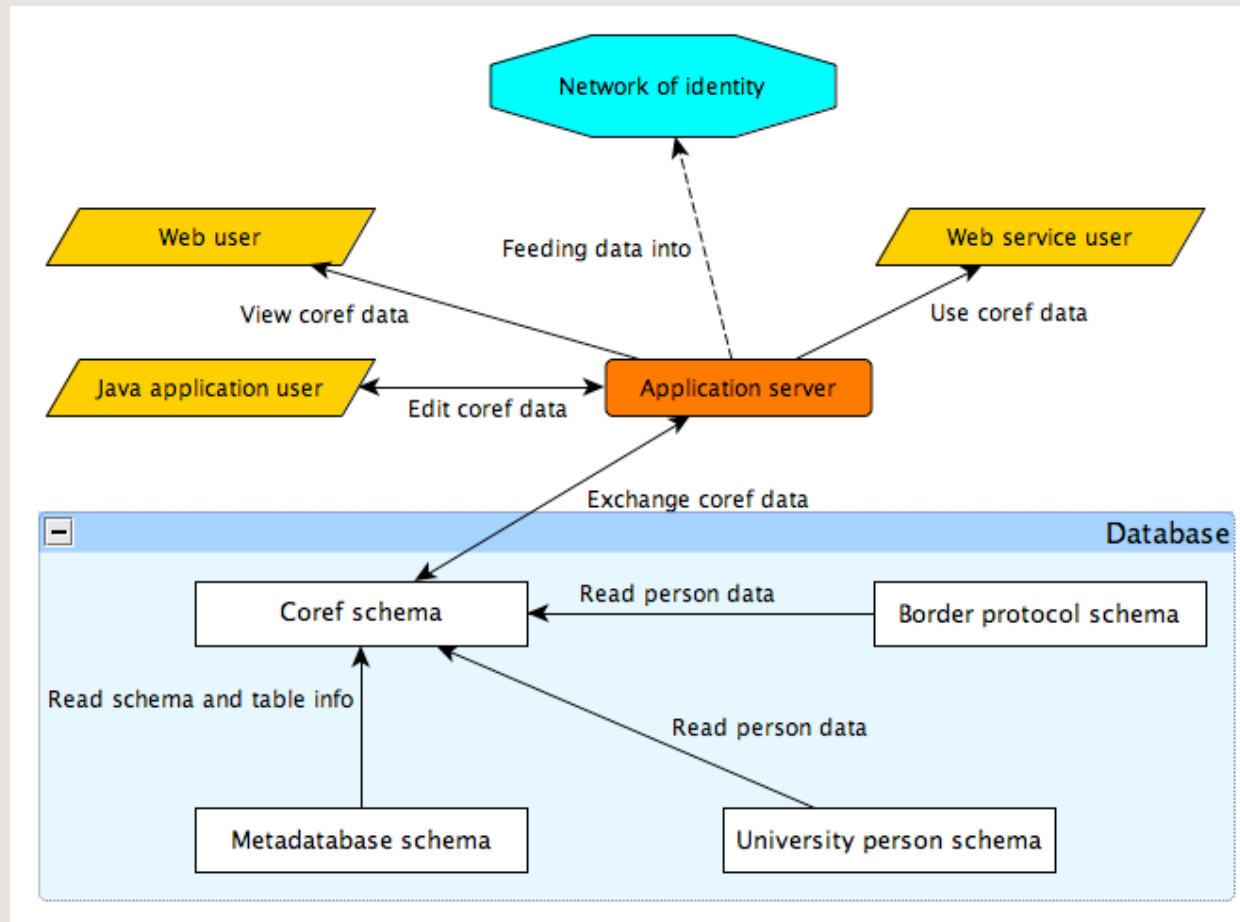
Example databases

1. A database of persons mentioned in a set of border protocols based on examinations of the border between Denmark–Norway and Sweden–Finland in the 1740' s.
2. A database of students and employees of the University of Oslo in the 19th century, based on university protocols.

They do not co-refer with each other, they are both to be connected to external resources.

1 will be used as the example in this presentation.

Implementation overview





The coref schema

This schema contains one core table, in which pairs are kept. Each pair consists of:

- A reference to one column is a schema somewhere in the database
- An external reference identifying an information object somewhere external to the database

The schema also contains stored procedures to be used in communication with the application server.



The meta database

The meta database is an core component of EDD' s database system. It consists of several tables and stored procedures keeping track of the schema, tables and columns of the entire database.

Using the information from the meta database, the records in the person tables of the various schema can be accessed, and unambiguous references to them can be stored.



The border protocol and university person schema

These are schema with different types of information based on printed and hand-written sources.

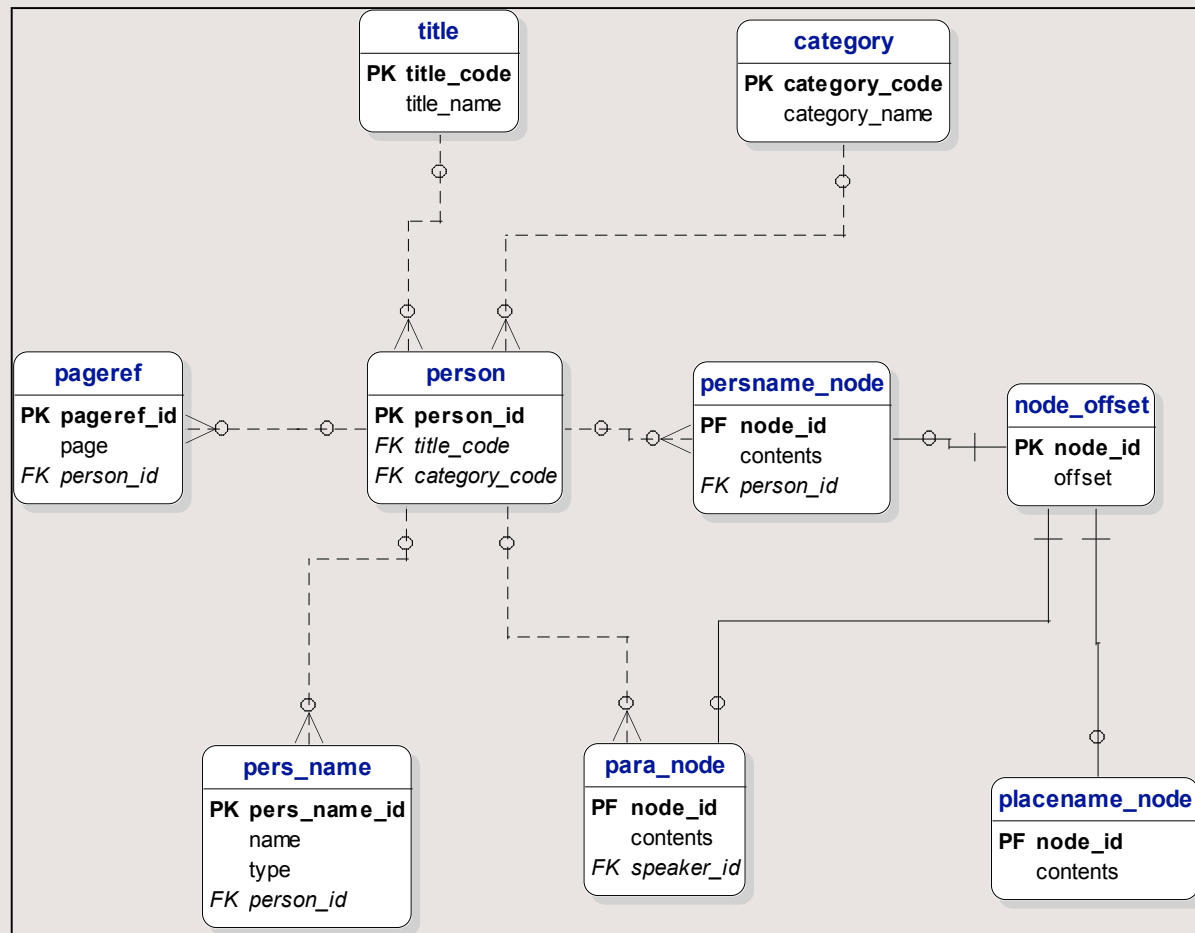
They both have tables with person information.

The border protocol schema have sets of names connected to each person, based on a printed version of the protocols.

There are also links from each name to the place(s) they are used to refer to the specific person in the source text.



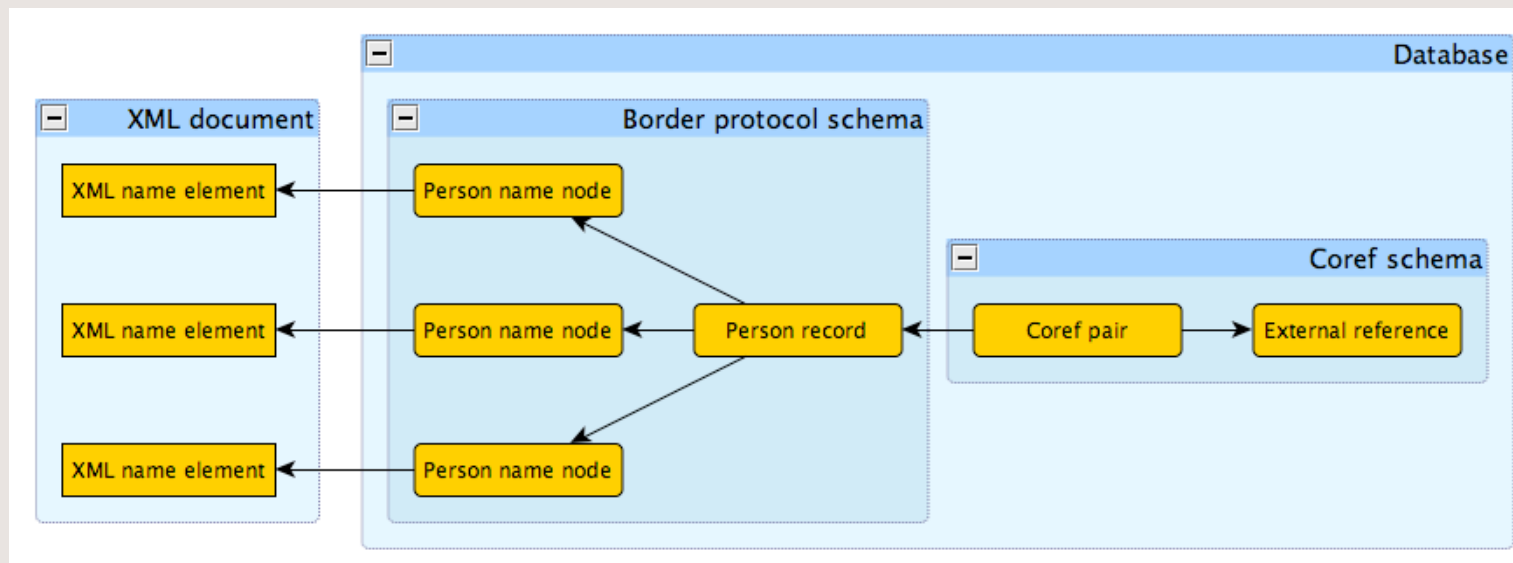
The border protocol core data model



The connection from the coref schema to the border protocol

The person table in fact acts as a local coref storage for the border protocol schema, storing corefs in the source.

By using a person record in the border protocol schema as one side of the pair in the coref schema, the set of co-refered person name strings in the source for the data in the border protocol schema is connected to the external side of the coref pair.





The application server

Handles database connection and access rights.

Translates between the XML used in client connection and database statements



The java client

List co-references

Edit co-references

- Keeps track of person responsible (user), time, and optionally stores a free text describing e.g. the reason for changing this coref.

Create a new co-reference

- Search for person records in the chosen local namespace
- Choose a person for coref
- Add external reference
- Keeps track of person responsible (user), time, and optionally stores a free text describing e.g. the reason for entering this coref.



The web client

List all corefs

List the corefs of a specific namespace

Search for corefs



Coref as a web service

List all corefs

List corefs in a specific namespace

Search for corefs

Obtain global identifier for the corefs

- A global identifier is our database identifier + the unambiguous reference to the record in our database



A network of identity

Such a network should be built on top of our systems, using the web service to connect to the coref information.