



Specification

European Affiliation Database

D. Porte
March 2011
Version 1.0

Foreword

This is a draft specification for the implementation of a European Affiliation Database (EAD), essential to building a European Virtual User Office (EVUO). This document must be considered as a working document, outlining possibilities.

Glossary

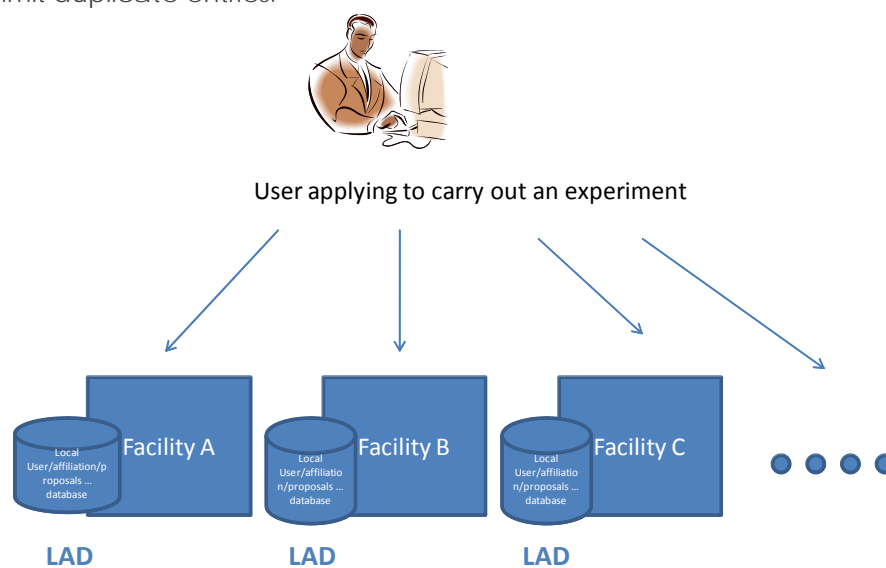
- User Office** Service in charge of managing the data workflow between scientist and the facility.
- User** Scientist proposing an experiment to a facility.
- Proposal** Document describing the experiment proposed by a scientist to a facility.
- EVUO** European Virtual User Office. Software application managing the data workflow between scientists and facilities. It includes a User database, an affiliation database, an authentication mechanism and features to deal with the data workflow. It could also include several general purpose training courses or simulators.
- WUO** Web-based User Office. Software application managing the data workflow between scientists and a facility. This web based software application is located in the facility. It is technically supported by the facility. (e.g. SMIS ESRF, DUO PSI ...).
- EAD** European Affiliation Database. This is the purpose of this specification.
- LAD** Local Affiliation Database. Component of the WUO managing affiliations for a facility.
- EAA** European Authentication and Authorization mechanism. Mechanism to allow the access of a software application based on a login/password. The mechanism provides a subset of all features regarding the user authorization. The activity is recorded thanks to an accounting database.

1. Introduction

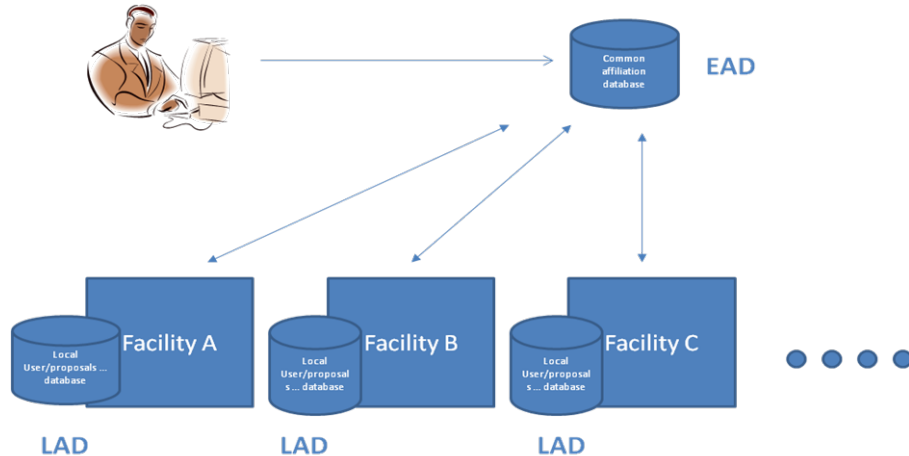
What does an affiliation database mean?

The affiliation of people (users) using a facility for an experiment corresponds to the laboratory from which they are authorized to carry out experiments. This information gives the facility the “professional” address of the users. It also opens up the possibility make statistics. It is important to notice that users may “work” for several laboratories at any given time same.

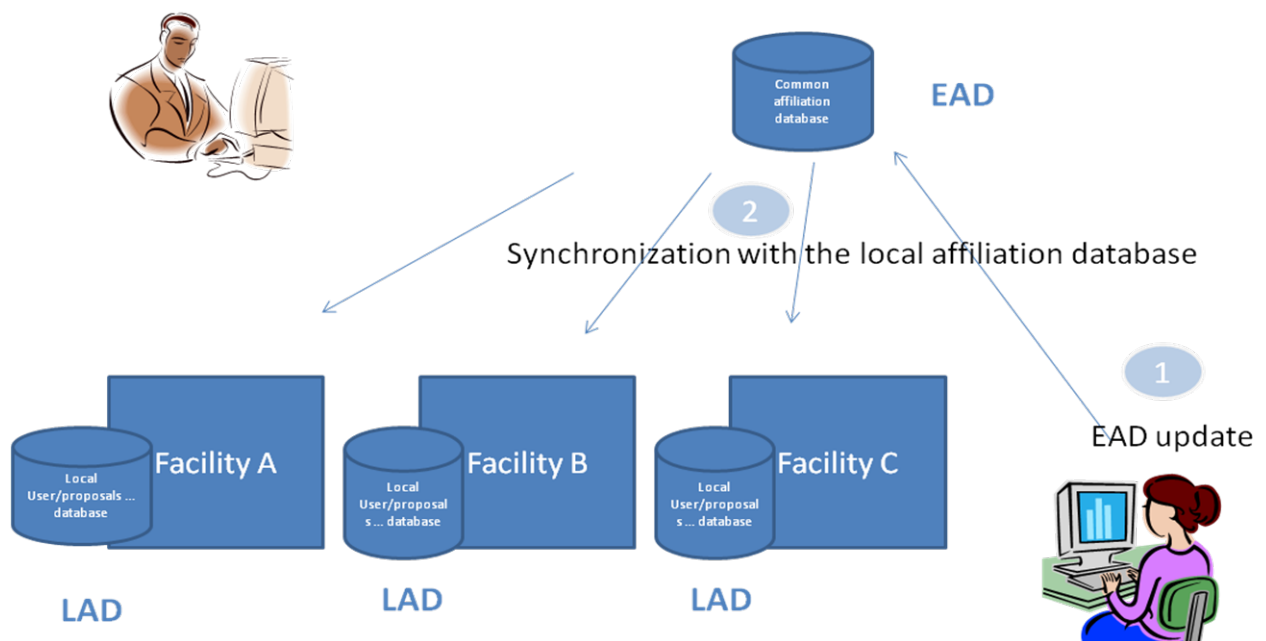
At present, each facility has a local affiliation database (LAD). New affiliations are registered by scientists themselves when they don’t find their own affiliation in the local database. Before being considered as operational, most often these new entries are checked by the User Office; the goal of which is to limit duplicate entries.



Sharing the same affiliation database between facilities will help scientists a lot. The user will no longer have to register the same information several times. In addition, an affiliation record will make sure that the spelling is the same at all facilities.



Even if technologies such as the grid or the cloud have democratized the fact that a significant amount of information located outside the facility, the affiliation database is a sensitive component of a WUO. Consequently, facilities prefer to keep a copy on site. In case of failure of the EAD, the facility would like to have a backup solution on site. A synchronization mechanism will replicate the EAD update to local affiliation databases. A reverse mechanism would be difficult to implement and is not part of this specification. Therefore, when a facility wants to update its local database, it has to make the relevant modification in the EAD. This then activates the synchronization mechanism to update its local database. This mechanism is detailed in the Section 7.



2. Affiliation

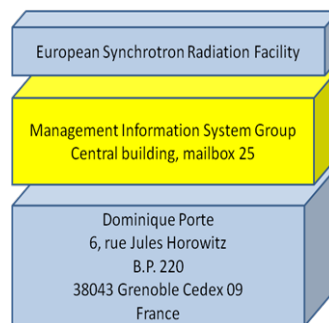
Affiliation is composed of two sets of fields:

- ▶ A set of fields, the so-called S1, common to all facilities. This includes information such as the name and address of the laboratory. Some elements are mandatory (name of

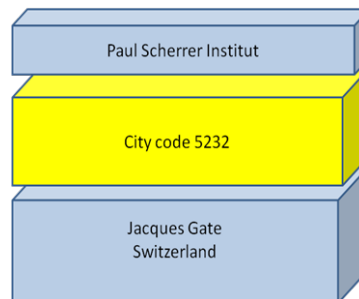
the laboratory, address), others are optional (acronym, main email of the facility, phone number of the reception desk...).

- ▶ A set of customizable fields, the so-called S2. Some facilities have requested that data specific to their WUO is registered. The necessity of recording this in a common database is debatable. In fact, this will be implemented if the fields are used by at least two of the participating facilities. For example, if two facilities need to record the GPS location of an affiliation, this will be stored in the common database.

In the following example, the local facility A is used to sending mail by the traditional postal service. Consequently, it builds the address with several fields of the set S1 (blue) and several fields of the set S2 t (yellow).



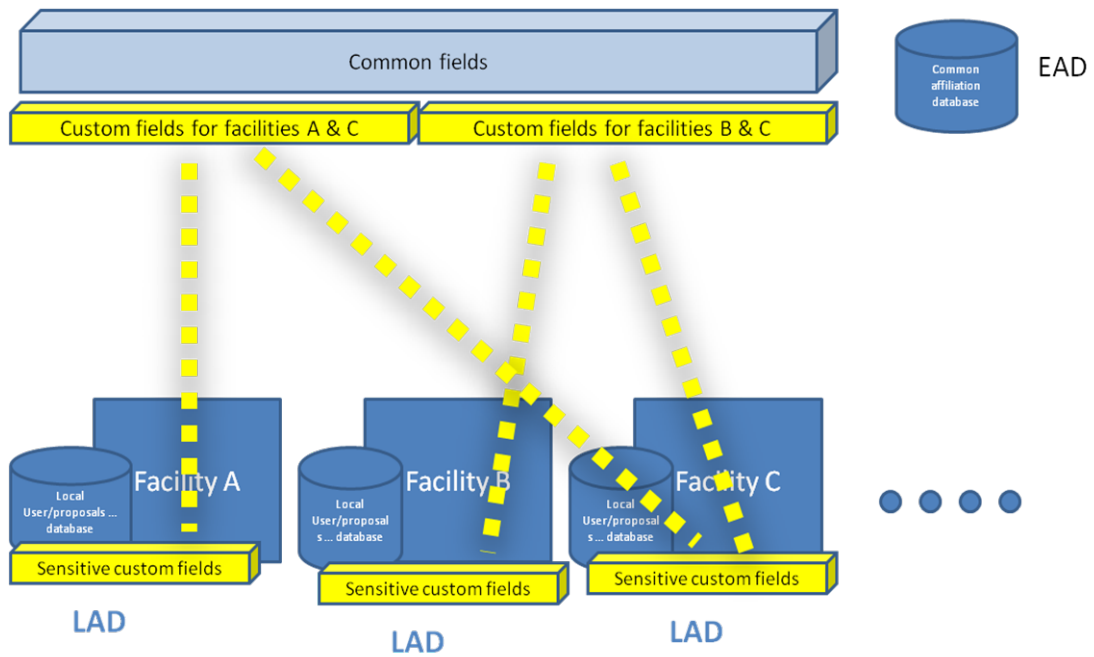
In the following example, the customizable field is used to record the affiliation's city code. This will be used by the facility to apply the relevant rate when reimbursing travel expenses.



Customizable field sets are under the responsibility of the local facility that requires such extensions.

3. Confidentiality

The confidentiality of data related to scientific activity is a very important issue. Even if research is carried out in an open-minded way, facilities closely protect their data. Nevertheless, there is no need for this high level protection for most fields forming an affiliation. The address of the affiliation is freely accessible to the public (internet, maps, and yellow pages). The name of a laboratory can also be easily found by surfing the internet. On the other hand, customizable fields may contain private or sensitive information. To avoid a complex implementation of the affiliation database, EAD will only manage non-sensitive information. When supplementary sets of customizable sensitive fields are deemed necessary, by a facility, then they will be managed locally.



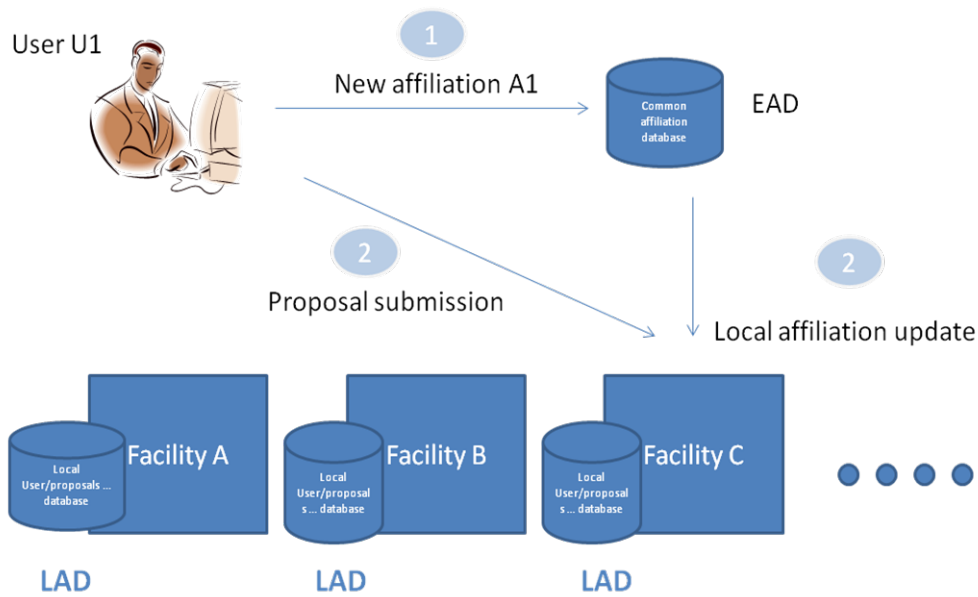
4. Roles & Assignments

The two key players in charge of managing the EAD are the user and the user office of the facility concerned by the user's experiment. As long as the affiliation is not used in a proposal that has already been submitted to a facility, update or deletion of the affiliation stored in the EAD can be done by either player. On the other hand, as soon as this affiliation has been used in a proposal submitted to a facility, modifications can only be done by User Offices. User Offices will have to check the potential consequences of modifying an affiliation already used by at least one proposal. For example, if the street name of an affiliation changes, the User Office will first have to inform all other facilities using this affiliation that it will be updated. Then, they will either have to update the existing affiliation or create a clone including the new street name. The role of the User Offices is important in the control of new affiliation entries. To limit the number of duplicate affiliation entries, User Offices will have to check if a new record does not already exist in the EAD. If this is the case, they request the user to use the existing affiliation.

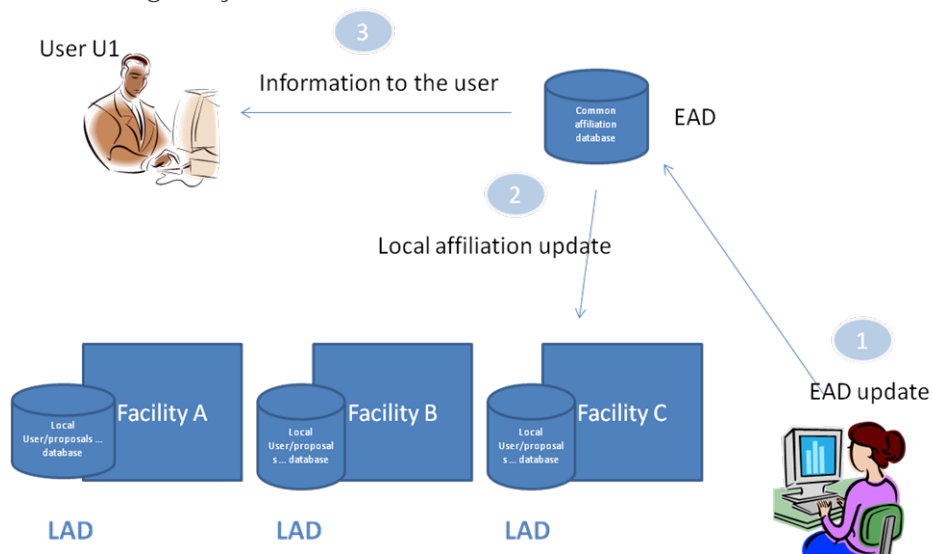
Below are several examples of the affiliation management process.

Example 1

User "U1" wants to carry out an experiment in facility "C". As his affiliation does not yet exist in the EAD, he has to record it before applying for an experiment. Then, he submits his proposal to facility C. The submission triggers a synchronization of the EAD and the local affiliation database.



The User Office of facility C is informed that a new entry in the EAD has been registered and synchronized with its local affiliation database. As it is the first facility concerned by this new entry, it must check that the information input is correct - spelling. If it appears that notices that user U1 made a mistake, it updates the EAD and informs user U1 of the modification. The local affiliation database is again synchronized with the EAD.

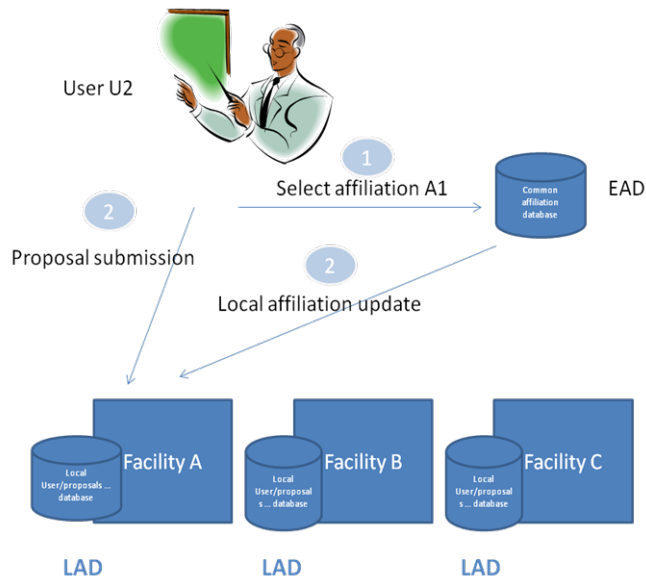


Remark: LADs have been in existence for a couple of years. Consequently, several thousands of affiliations have already been recorded in databases. If a local affiliation entry corresponds to the new EAD entry, a link from the local entry and the EAD entry will be recorded in the LAD.

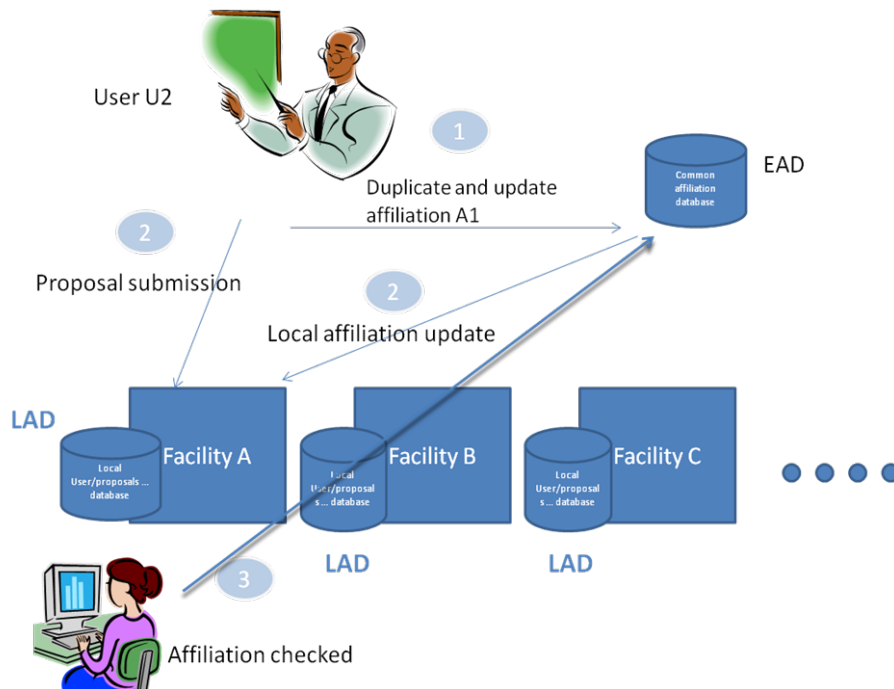
Example 2

User "U2" wants to carry out an experiment in facility "A". Users U1 and U2 are from the same laboratory. Consequently, they have the same affiliation A1.

First case: User U2 finds his affiliation A1 in the EAD. He submits his proposal to facility A. The submission triggers a synchronization of the EAD and the local affiliation database.



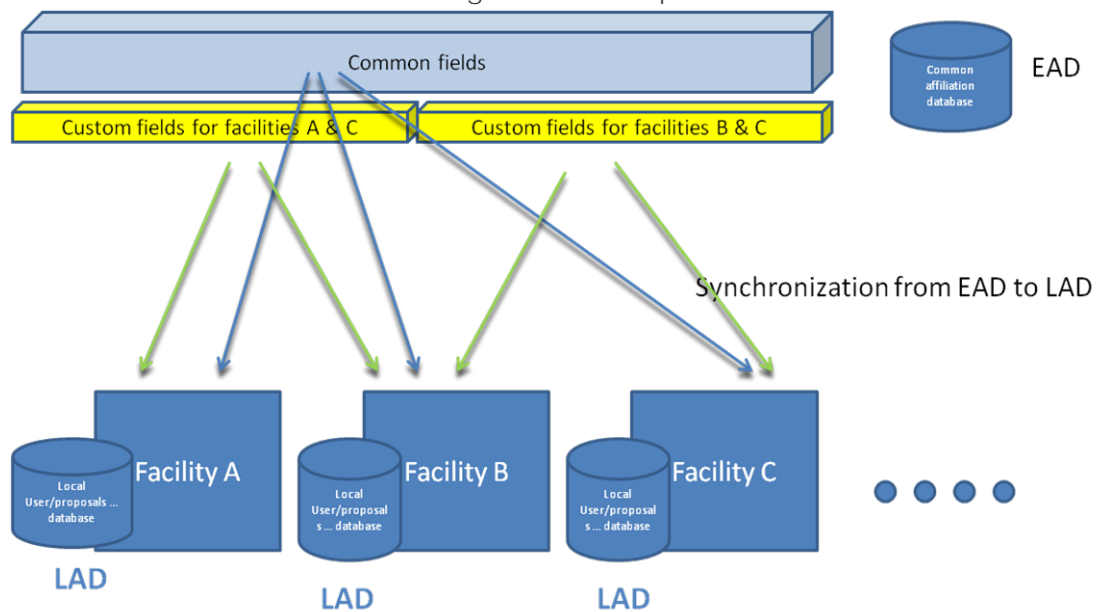
2nd case: User U2 finds his affiliation in the EAD but he considers that the spelling is incorrect. He duplicates the affiliation entry and modifies it.



The User Office of Facility A is informed that a cloned affiliation has been recorded in the EAD and synchronized with its local affiliation database. As it is the facility first concerned by this new clone, the User Office checks if this clone creation was really due to a misspelling of the original record. If it turns out that user U2 has made a mistake, the User Office informs user U2 that the original affiliation was correctly spelt and asks him to modify the proposal. On the other hand, if the User Office agrees with User U2 that the original affiliation was not correctly spelled, they informed facilities already using this affiliation.

5. Synchronization mechanism

The goal of the synchronization mechanism is to propagate EAD updates to local affiliation databases. As mentioned in the chapter §2 "Affiliation", there are two sets of fields; that common to all facilities and a set of customizable fields specific to some facilities. In both sets, some fields are mandatory whereas others are optional. It is important to note that a field can be mandatory for some facilities and optional for others. The synchronization mechanism must take due account of these characteristics. In the following example, the LAD of facility C is updated with common fields and the set of custom fields on the right side of the picture.



Remarks

Sensitive custom fields directly managed by facilities are not concerned by this synchronization. To simplify the synchronization process, it is not planned to have a reverse synchronization mechanism going from the LAD to the EAD.

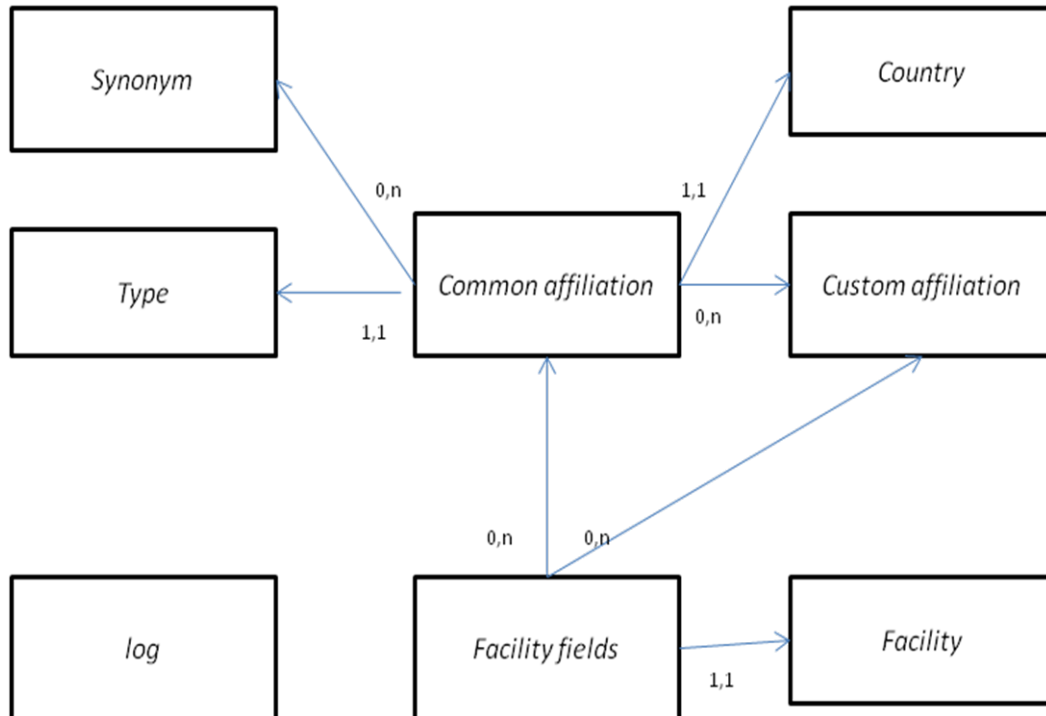
6. EAD database design

The EAD database is based on the following tables:

- ▶ *Common Affiliation* stores fields common to all facilities.
- ▶ *Custom Affiliation* stores complementary custom fields required by some facilities.
- ▶ *Facility Fields* stores what are the fields necessary for facilities. In addition, for each field, it states if the field for a facility is mandatory.
- ▶ *Facility* stores the list of facilities involve in the EAD.
- ▶ *synonym* provides the link between an affiliation and its clones.
- ▶ *Type of Laboratory* stores the list affiliation categories (university, research center, commercial company ...).

- *Country* stores the list of countries.
- *Log* is used to record all modifications.

The diagram below gives an overview of the database model. A more detailed view of table structure is given in annex 1.



7. EAD software components and implementation

The technical components of the EAD are:

- An open source relational database.
- A web portal
- An authentication mechanism provided by the EVUO infrastructure.
- Several web applications to interact with the database and the portal.
- A set of services including the synchronization mechanism.

People using the portal would fall into one of the three following categories:

- The user
- A User Office member
- The portal administrator

The user will use a web application to search, create, modify or delete its affiliation. The web application will be flexible enough to be pluggable into another portal using a technology such as IFRAME.

The User Office member will have the same features as the user. In addition he/she will have the possibility to run the synchronization mechanism.

The portal administrator will have a full access to the EAD.

The database schema is described in the chapter §6 and in annex1. It could be implemented using the MySQL tool.

The synchronization mechanism will be based on web applications and web services. As soon as this specification is validated, a complementary specification will developed. It will be the annex2 of this document.

The initialization of the EAD will be a sensitive task. Either the EAD will start from scratch (with no data), or it will be fed with affiliations already stored in LAD. As local affiliation databases already have several thousands of records, merging and checking all this data sounds unrealistic. Consequently, the first approach will be selected.

8. Conclusion

Setting-up a common affiliation database will help a lot users and facilities. Even if the EAD initialization will take a couple of months, as the workload will be shared by the different facilities, it would not be too time consuming for the User Offices. It is difficult to predict how long this process will take. Nevertheless, if all facilities play their part game, we may consider that two years will be necessary to have most of LAD into the EAD.

Table Common affiliation

| Field name | Description | Example |
|-------------------|----------------------------------|---|
| ID | Identifier | 00002346 |
| ID_Synonym | Synonym list ID | 4 |
| Name | Name of the laboratory | European Radiation Facility |
| Acronym | Acronym of the laboratory | ESRF |
| Department | Name of the department | Technical Infrastructure Division |
| Service | Name of the service | Management Information System |
| Type | Type of the laboratory | U |
| Postal address | Postal address of the laboratory | ESRF, BP 220, 38043 GRENOBLE CEDEX 9, FRANCE |
| Address | Full address | ESRF, Polygone Scientifique Louis Néel, 6 rue Jules Horowitz |
| Postal code | Postal or City code | 38000 |
| Town | Town | Grenoble |
| Country | Country | France |
| Web | Laboratory web site | www.esrf.eu |
| Created | Date of creation | 2011-04-01 |
| Last modification | Date of the last modification | 2011-04-02 |

Table custom affiliation

| Field name | Description | Example |
|-------------|------------------------|-----------|
| Common_ID | ID of the common field | 00002346 |
| Field_ID | ID of the field | 1 |
| Description | Description | City code |
| Value | Value | 5232 |

Table Facility fields

| Field name | Description | Example |
|------------|---------------------------------|--------------------|
| Facility | Name of the facility | ESRF |
| Table | Common or custom affiliation | Common affiliation |
| Field_name | Name of the field in this table | Web |
| Mandatory | Mandatory: yes or no | N |

Table facility

| Field name | Description | Example |
|------------|----------------------|---------|
| Name | Name of the facility | ESRF |

Table synonym

| Field name | Description | Example |
|------------|------------------------|---------|
| ID_Synonym | ID of the synonym | 4 |
| ID_Synonym | Affiliation identifier | 0002336 |

Table type of laboratory

| Field name | Description | Example |
|-------------|------------------------|------------|
| Type | Type of the laboratory | U |
| Description | Description | University |

Table country (based on the ISO 3166 norm)

| Field name | Description | Example |
|-------------|---------------------|---------|
| ID_Synonym | Codet ISO 3166 | FR |
| Description | Name of the country | France |

Table log

| Field name | Description | Example |
|----------------|------------------|--------------------|
| User | User or facility | ESRF |
| Date | Date | 2011-04-01 |
| Table | Table name | Common affiliation |
| Action | Action | Update |
| Field | Field concerned | Type |
| Previous value | Previous value | U |
| New value | New Value | R |