ENCODING AFFILIATION ELEMENT

INTRODUCTION

In the context of rapid growth of institutional repositories, the PEER project (www.peerproject.eu) has been set up to monitor the effects of systematic archiving of author's final peer-reviewed manuscript over time.
To conduct this research, interoperability between editor's data and online repositories is an essential asset. It is therefore important to specify the encoding scheme of required metadata elements, particularly the ones related to author's identification (as it is particularly important for citations, e.g. databases such as the ones accessed through Web of Science).

A high granularity of metadata is preferable. Metadata with a high granularity allows for deeper structured information and enables precise information to be identified and shared. It offers indeed greater flexibility in treating each data field in isolation if required.

As far as representation of metadata is concerned, the PEER project has adopted the Text Encoding Initiative encoding scheme (www.tei-c.org). This document describes the specifications for the encoding of the affiliation element in accordance with TEI 's Guidelines for Text Encoding and Interchange (P5).

Only the encoding of the element "author" and its affiliation will be dealt with here.

ELEMENT DESCRIPTION

Element <author>

It contains the information about the author, personal or corporate, of an article.
It groups three elements: the name of the author (<persName>), the affiliation (<affiliation>) and the email address.

Ex.:

```
<author>
        <persName>
                <forename>J.</forename>
                <surname>Kwela</surname>
        </persName>
        <affiliation>
                <orgName type="department">Institute of Experimental Physics</
        orgName>
                <orgName type="institution">University of Gdansk</orgName>
                <address>
                        <addrLine>ul. Wita Stwosza 57</addrLine>
                        <postCode>80-952</postCode>
                        <settlement>Gdansk</settlement>
                        <country key="PL">Poland</country>
```

```
            </address>
        </affiliation>
        <email>fizjk@univ.gda.pl</email>
</author>
```

Element <affiliation>

It contains an informal description of the author's present or past affiliation with some organization, for example an employer or a sponsor.
It groups two elements: the name of the organization (<orgName>) and its postal address (<address>).

Ex.:

```
<affiliation>
        <orgName type="institution">Technische Universität Darmstadt</orgName>
        <orgName type="department">Institute of Materials Science</orgName>
        <address>
            <addrLine>Petersenstraße 23</addrLine>
            <postCode>D-64287</postCode>
            <settlement>Darmstadt</settlement>
            <country key="DE">Germany</country>
        </address>
</affiliation>
```

We have identified three types of organizations, which correspond to the three-tiered system of WoS.

- Institution: corresponds to the global structure that hosts the author (can be a university or an institute (e.g. MIT, INRIA)) – the largest scale of organization type.
- Department: corresponds to a specialized division of the institution mentioned above – intermediate structure of organization type (department, faculty, institute) if there is one.
- Laboratory: corresponds to the research team or group, which the author belongs to (e.g. Joint Research Laboratory Nanomaterials) – the smallest scale of organization type.

Element <address>

It contains the postal address of the organization, which the author is affiliated to.

Ex.:

```
<address>
        <addrLine>Chemin du Solarium</addrLine>
        <addrLine>BP20</addrLine>
        <postCode>33175</postCode>
        <settlement>Gradignan</settlement>
        <country key="FR">France</country>
</address>
```

Here are the different elements that can be used to describe the address.

- <addrLine> contains one line of a postal address. It can be used as many times as needed (in case of a multiple line address)

- <postCode> contains a numerical or alphanumeric code used as part of a postal address to simplify sorting or delivery of mail.
- <settlement> contains the name of a settlement such as a city, town, or village identified as a single geo-political or administrative unit.
- <region> contains the name of an administrative unit such as a state, province, or county, larger than a settlement, but smaller than a country.
- <country> contains the name of a geo-political unit, such as a nation, country, colony, or commonwealth, larger than or administratively superior to a region and smaller than a bloc. The key attribute must be used to identify the country, according to ISO 3166 a2.

COMPLEX CASES:

- **Multiple institutions or multiple departments**

a) This case typically corresponds to the affiliation of a joint laboratory.
Use the "key" attribute to identify the different institutions (or departments), which the joint laboratory belongs to.

Example: Joint Research Laboratory Nanomaterials, which is a joint laboratory of the Technische Universität Darmstadt and the Karlsruhe Institute of Technology.

```
<affiliation>
        <orgName type="laboratory">Joint Research Laboratory Nanomaterials</
orgName>
        <orgName type="institution" key="instit1">Technische Universität Darmstadt</
        orgName>
        <orgName type="institution" key="instit2">Karlsruhe Institute of Technology</
        orgName>
        <address>
                <addrLine> Petersenstrasse 23</addrLine>
                <postCode>D-64287</postCode>
                <settlement>Darmstadt</settlement>
                <country key="DE">Germany </country>
    </address>
</affiliation>
```

b) In the case of a laboratory described by several names (that is often the case for the UMR of French CNRS), use the same procedure.

Example: GREMI, also named UMR 6606 (a joint laboratory of CNRS and Université d'Orléans)

```
<affiliation>
        <orgName type="laboratory" key="lab1">GREMI</orgName>
        <orgName type="laboratory" key="lab1">UMR 6606</orgName>
        <orgName type="institution" key="instit1">CNRS</orgName>
        <orgName type="institution" key="instit2">Université d'Orléans</orgName>
        <address>
                <addrLine>14, rue d'Issoudun</addrLine>
                <addrLine>BP 6744</addrLine>
                <postCode>45067</postCode>
```

```
            <settlement>Orléans cedex 2</settlement>
            <country key="FR">France</country>
        </address>
</affiliation>
```

- **Several affiliations**

The case is encountered when an author provides two affiliations (the current one and the one he wants to see appearing on the paper, which often corresponds to his past affiliation when he conducted his research that led to the article).
Use the attribute “current” to identify the current affiliation.

Example:

```
<author>
        <persName>
            <forename>B.</forename>
            <surname>Bastin</surname>
        </persName>
        <affiliation>
            <orgName type="laboratory">LPC Caen</orgName>
            <orgName type="institution" key="instit1" >ENSICAEN</orgName>
            <orgName type="institution" key="instit2" >Université de Caen Basse-
            Normandie</orgName>
            <orgName type="institution" key="instit3" >CNRS/IN2P3</orgName>
            <address>
                <postCode>14050</postCode>
                <settlement>Caen</settlement>
                <country key="FR">France</country>
            </address>
        </affiliation>
        <affiliation type="current">
            <orgName type="laboratory">GANIL</orgName>
            <orgName type="institution" key="instit4" >CEA/DSM</orgName>
            <orgName type="institution" key="instit3" >CNRS/IN2P3</orgName>
            <address>
                <postCode>14076</postCode>
                <settlement>Caen</settlement>
                <country key="FR">France</country>
            </address>
        </affiliation>
</author>
```

- **Several geographical addresses**

For lack of encountered examples, this case will not be treated as such here.
Cases exist though and relate to laboratories, which are bilocalized and therefore have a double address for a single laboratory.

REFERENCES:

- Holmes M., Romary L.: “Encoding models for scholarly literature” http://hal.archives-ouvertes.fr/hal-00390966/fr/
- PEER Final report on the provision of usage data and manuscript deposit procedures for publishers and repository managers (D2.2) http://www.peerproject.eu/fileadmin/media/reports/PEER_D2_2_20091028_v5.pdf
- TEI P5 Guidelines http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html