

## TEF: Metadata for French dissertations

Dalia Boudia and Rosa María Gómez de Regil

Institut National des Sciences Appliquées, INSA de Lyon, France

### Abstract

Since 2000, universities and research institutions have been encouraged to promote and preserve scientific and technical productions by disseminating digital theses and dissertations. This national program implies that new practices and new tools impacting thesis identification are going to be established at different levels (administrative circuit, theses production mode, theses submission methods, document processing chain, dissemination platforms, digital preservation systems...). It is important to describe theses and dissertations using metadata in Internet shared formats to gain information visibility. TEF (Thèses Électroniques Françaises) standard objective is to permit quality metadata production, exchange and dissemination for French digital theses and dissertations. TEF defines a metadata set for French digital theses. It proposes two standardization levels: a metadata set and an XML schema (exchange format and validation tool). TEF will be composed of descriptive metadata (bibliographical data) and administrative metadata (administrative circuit, rights and preservation data). The XML schema explicitly defines valid XML documents structure. TEF standard has to facilitate metadata exchange at local level (universities) and at national level (SUDOC, CINES). At local level, it will manage theses production and validation, metadata production and full text access. The descriptive metadata set derives from Dublin Core standard, which guarantees a good interoperability with different kinds of documents or different dissemination environments (cf. OAI-OMH protocol). It is also compatible with NDLTD (Networked Digital Library of Theses and Dissertations) Dublin Core application profile, named ETD-MS (Interoperability Metadata Standard for Electronic Theses and Dissertations).

### Introduction

The French national Department of Education and Research<sup>1</sup> desires to promote the development and long-time preservation of French theses<sup>2</sup> and dissertations, disseminating them in an electronic format<sup>3</sup>. This national position implies that new practices (theses production and validation) and new tools (publishing platforms, metadata management) are going to be established. The objective is to respond to digital environment stakes. This attitude will impact thesis referencing and identification.

Today, French theses are referenced in the SUDOC national catalog (Système Universitaire de Documentation), in library catalogs or in specialized bibliographical tools (subject databases, institutional databases, search engines). If we want to extend theses internet referencing they should be described following a shared metadata schema.

Metadata (any data defining or describing other data) facilitate information sharing, contribute to minimize data loss, help in the searching and data storage processes. There are several metadata models used to describe networked electronic resources. The Dublin Core metadata set is used to make a brief and basic description of an electronic resource through 15 elements. Nevertheless, theses (research official document) need to be described with some other detailed elements. TEF (Thèses Électroniques Françaises) standard will respond to specific thesis description and its administrative needs. A standardized approach is necessary to facilitate systems interoperability and data exchange.

TEF standard is formed of

- a vocabulary (metadata set) that defines description elements and,
- a XML schema, an exchange format and validation tool useful to process data in computer environments.

TEF standard has been developed in two stages and as a consequence, it has two versions :

- TEF 1.0 proposes descriptive metadata (published on May 2005)
- TEF 2.0 is still under development and will propose administrative metadata (publication expected for January 2006).

Descriptive metadata have a bibliographical function in TEF. They include the Dublin Core metadata set. Administrative metadata will manage the administrative submission and publication thesis process (from subject submission to diploma attribution), the legal control associated to the document (copyright) and the long-time preservation.

This paper presents the TEF standard metadata set through descriptive and administrative metadata.

<sup>1</sup> Ministère de l'Éducation Nationale de l'Enseignement Supérieur et de la Recherche

<sup>2</sup> In this article we are going to refer to Doctoral thesis.

<sup>3</sup> Cf. *Circulaire du 21 septembre 2000*.