

MICHAEL Project

MICHAEL data model

Version 1.0

Document written by

MICHAEL

Contributions by

Martin Sévigny, AJLSM

Christophe Dessaux, Richard Blandin, Martine Tayeb, MCC,

Stéphane Cagnot, Dédale

Rosella Caffo, Giuliana de Francesco, MiBAC,

David Dawson, Kate Fernie, MLA

Antonella Fresa, Pier Giacomo Sola, Amitié

September 30th 2005

1) Abstract

This document is the reference definition of the MICHAEL data model. The data model can represent XML descriptions of digital collections and related information: institutions, projects or programmes, services or products and physical collections.

Table of contents

1) Introduction	1
2) Definitions	1
3) Entities	2
4) Relations	2
5) Fields	4
6) Annex A : XML schema.....	11

2) Introduction

One the major goal of the European project *MICHAEL* is to build a multilingual inventory of cultural heritage in Europe. To accomplish this task, an approach based on gathering data from national or regional inventories will be used, along with the development of a standard software platform for the creation, aggregation and publication of this data.

The cornerstone of such a distributed information system is always its underlying data structures, the way data is organised and managed in databases or similar environments. For the *MICHAEL* project, inventory and software platform, the rule is the same, but we should mention two important characteristics about these data structures:

- As in all documentary system, the underlying data structures must be understood, to a certain level, by human operators of the systems : people introducing data, searching for it, etc.
- The data is represented using the XML standard, and thus the formal definition of the data structures must follow one of XML's grammar definition languages such as DTD, RelaxNG or XML Schemas.

We introduce these two characteristics to better define the scope of this document. We think that is it very important to have a *human* reference definition for the data model which is not the XML structures itself, but a list of simple definitions for – hopefully – simple concepts. Most of this document is composed of this human reference definition of the data model, although its formal definition as an XML Schema is provided in the first annex, page 12.

Although this reference document can provide basic understanding of the various fields used to describe digital collections and related information, this document should be completed by a user guide in order to make the system and its underlying data structures easily usable and comprehensible for anyone wishing to contribute to *MICHAEL* content. This user guide is out of scope for this reference definition document.

This document is composed of three main parts. The first one introduces some basic definitions of concept used in the following parts. The next part defines the five entities of the data model, followed by a part on the relations between these entities. Finally, the last main part gives definitions for all fields in the data model.

3) Definitions

The *MICHAEL* project focuses primarily on digital collections. But to describe these digital collections we also need information on the institutions that own them, the projects or programmes in which they are created, the physical collections that were digitised and the services or products through which digital collections are made them available

Since one institution, for instance, could be responsible for more than one digital collections, it would be useful to describe only once this institution, and to reuse this description when needed. This example can be extended to cover programmes or projects, physical collections and service or products.

This is why the data model is composed of five **entities**, or five different data structures, one for each *kind of record* we wish to create: digital collections, institutions, programmes or projects, services or products, physical collections. When we talk about the *digital collection entity*, we thus mean the *record type digital collection*, or the *data structure suitable for digital collections*.

A **record** is one manifestation of an entity, to represent an actual thing. For instance, in order to describe the institution *Bibliothèque nationale de France*, we will create one record of type institution, or one concrete manifestation of the entity institution. Entities are thus abstract concepts, and records are real objects in the system.

Surely, records must be linked between each others. For instance, it is not enough to have on one hand a description of the *Bibliothèque nationale de France* and, on the other hand, a description of one of its digital collection. The system – and at the end the user – must know that this particular institution has created the digital collection, for instance. This is called a **relation**, and a relation is always made between two records, and not between two entities; relations are between real objects, not abstract concepts.

Having entities, records and relations is not enough to fully define the data model. We need to know how to describe a specific entity, for instance an institution: name, address, and so on. These characteristics of an entity will be called **fields** in this data model. An entity is thus a collection of fields.

In the human definitions for fields, we will consider that an entity is a flat list of fields. But the reality is more complex and the XML Schema provided at the end will show that an entity is a hierarchical XML document

where some fields are grouped into common concepts, some are repeatable and the occurrences are grouped, together, etc. So the human definitions will be useful for a person wishing to understand the data found in a MICHAEL platform, and the computer definition (the XML schema) will be useful for a programmer who wishes to implement a software component to work with this data component.

4) Entities

For each entity, we give a definition and mention any mandatory relationships between entities. A mandatory relationship means that to publish some information in the system, it must have a specific relationship to another entity (not seeAlso). For instance, if the relationships between *digital collection* and institution are mandatory, it means that a digital collection record is not completed and may not be published without a relation to at least one institution.

4.1 Digital Collections

The digital collection entity is the main focus of the Michael project as it aims at being an inventory of digital and digitized cultural heritage.

A digital collection may be a set (or group) of digital items or a set of records describing digital items. A digital collection may be a set of digital images, texts, structured data, sound files, virtual reality models, multi-media or other resources. The collection may be aggregated on one server or distributed across several servers. In MICHAEL all digital collections relate to aspects of the European cultural heritage,

Digital collections are at the heart of the MICHAEL project. The other entities are offered to provide for additional descriptive or reference information for the digital collection, without duplication.

Digital collections have mandatory relationship with an institution or one service/product.

4.2 Institutions

An institution is an agent that owns digital collections, is responsible for digitisation projects, funding programmes, or for developing and managing information services or products. Institutions may include museums, libraries, archives and others with an interest in the digital cultural heritage.

Institutions have a mandatory relationship with either a digital collection, or a physical collection, or a project or a service/product in the MICHAEL system.

4.3 Services or Products

A service (or product) is a point of access to a digital collection or collections. It may consist of an online or an offline service (such as an application that enables users to select and order copies from a collection on demand) or it may consist of a packaged product that presents all or part of one or more collections (such as an electronic learning resource). Services or Products are included in the MICHAEL system to offer users with information about how and where to digital collections.

This entity is recommended; A service or product has a mandatory relationship to at least one digital collection or one institution in the MICHAEL system.

4.4 Projects or Programmes

A project or programme involves one or more institutions and directly or indirectly results in the creation of either digital collections or services/products.

This entity is included to provide a useful reference for institutions and is optional in the MICHAEL system. A project or programme has a mandatory relationship to at least one institution or one digital collection. It is recommended practice that there should be a relationship to an institution.

4.5 Physical Collections

A physical collection is a set of physical items, for example a set of museum objects, or a set of physical archives or a library collection.

The aim of MICHAEL is not to build inventories of physical collections. This entity is provided for reference to physical collections that have given birth to digital collections, for example through digitisation. This is an optional entity in the MICHAEL system. A physical collection has a mandatory relationship to at least one institution and either a digital collection or a project/programme.

5) Relations

As explained previously, there will be relations between records in a MICHAEL inventory. Basically, a relation is a link between two records. This link has some special characteristics:

- It has a role, either chosen from a predefined list or an ad hoc role.
- It may have a description, explaining to humans the exact nature of this role.

In MICHAEL, we decided that link doesn't have a specific direction. If, for instance, there is a relation between an institution and a digital collection, it also means that there is a link from the institution to the digital collection, and a link from the digital collection to the institution. For instance, if an institution is responsible for a digital collection, then the reverse is also true: the digital collection is under the responsibility of the institution.

The MICHAEL data model defines a set of specific relations between records of certain types. In this document, we will define these relations using the five entities, but please recall that relations are really between records, not between entities, in an actual system. Mandatory relationships between entities are defined in section 4.

5.1 Relations involving digital collections

A digital collection...

- May be under the responsibility of an institution
- May be created in the context of a project or programme
- May be created by an institution
- May be a representation of all or part of a physical collection
- May be made available via a product or service
- May be part of another (larger) digital collection

5.2 Relations involving institutions

An institution...

- May be responsible for digital collection(s)
- May be responsible for physical collection(s)
- May be the location of a physical collection
- May be responsible for or may contribute to programmes / projects
- May be responsible for products / services
- May create a product or service
- May be part of another (larger) institution
- May create a digital collection

5.3 Relations involving projects or programmes

A project or programme...

- May be the responsibility of an institution

- May be contributed to by an institution(s)
- Project may be part of a programme
- Programme may be concerned with funding projects
- Project may be part of another project
- May create a digital collection
- May create a product or service

5.4 Relations involving services or products

A service or product...

- Makes available (part of) one or more digital collections
- May be created by an institution
- May be under the responsibility of an institution

5.5 Relations involving physical collections

A physical collection...

- May be under the responsibility of an institution
- May be located at an institution
- May be created in the context of a project or programme
- May be created by an institution
- May be the source of part or all of a digital collection or collections
- May be part of another (larger) physical collection

6) Fields

The following sections contain reference definitions of fields for each entity. For each field, we provide first its name, then its code – a computer name used in the XML schema – and then we give information whether the field is mandatory or not. As a matter of fact, three values are permitted:

1. **mandatory**: the field must be present in order to validate a record
2. **optional**: the field may be present in a record
3. **recommended**: the field should be present, although in some circumstances it can be omitted if not relevant

Following these basic information, you will find a short description of the field, complete enough to understand its role and purpose.

In this reference document, technical fields are omitted. These fields are, for instance, the filename or URL of an image, the last modification date, etc. They are part of the XML schema, but are omitted here since they do not bring anything to the understanding of the data model for humans.

We also omitted the fields used to represent relations between records. These are special fields, technical, and are defined in the XML schema.

6.1 Entity Digital Collection

Section Identification

Identifier [identifier – mandatory]

This is the identifier for the digital collection. This identifier plays an essential role in the MICHAEL system. All entities should have a unique identifier, and this identifier should also be unique in the scope of the European instance.

Title [title – mandatory]

The title of the digital collection.

The title should provide a meaningful point of reference to the digital collection, and preferably it should be unique, at least within an institution. Acronyms and abbreviations should not be used in this field.

Section Description

Description [description – mandatory]

A free text description of the digital collection. This should add to the information provided in the title.

Language [language – recommended]

The language of the material contained in the digital collection.

Digital format [digital-format – recommended]

The digital characteristics of the collection. More specific than the digital type.

Digital type [digital-type – recommended]

This is the general type of digital collection, for example: a collection of texts, a collection of images, a collection of interactive resources, a collection of sound files.

N.B. Includes virtual collections.

Content type [content-type – optional]

This is the type of content in the digital collection. For example, maps, music scores or manuscripts.

Size [size – optional]

An evaluation of the size of the collection. This is intended to provide information for users of the MICHAEL system about the size of the digital collection.

Accrual [accrual – optional]

A statement of accrual policy (closed, passive, active, partial/selective), accrual method (purchase, deposit) and accrual periodicity (closed, irregular, periodic).

This information is especially important for harvesting purposes, to foresee the evolution of aggregated resources.

Standard [standard – optional]

A statement of any descriptive or terminology standards that were used in creating the item level metadata associated with the digital collection.

Legal status [legal-status – mandatory]

A statement of the legal status of the digital collection.

Access control [access-control – optional]

A statement of any access restrictions that are placed on the collection.

This would be used where for example access to the collection is closed for a period of time or where access is restricted to a certain category of users. The information is not intended for public use, but for reference by the institution that owns or manages the collection.

Related database [database – optional]

Database describing the objects in the collection.

Section Subject

Subject [subject – mandatory]

The subject or concept of the items in the digital collection. Lists of subject keywords will be provided.

Spatial coverage [spatial-coverage – recommended]

The spatial coverage of the items in the digital collection. An international list of countries and regions will be used in the MICHAEL system to provide the basis for multi-national searches. More detailed lists of regional and local administrative areas and places will be provided for each national instance.

Period [period – mandatory]

The general period(s) spanned by the content of the digital collection, for example Neolithic or Bronze Age.

Start date [start-date – optional]

This is the approximate date of the earliest item in the digital collection expressed as a year. For example, this would be 1400 for collections that date from the 'Fifteenth Century'.

End date [end-date – optional]

This is the approximate date of the latest (or most recent) item in the digital collection expressed as a year. For example, this would be 1699 for collections that date from the 'Seventeenth Century'.

Culture [culture – optional]

Culture or people that are the subject of the digital collection, for example Islamic or Jewish.

Famous people [famous-people – optional]

Famous (or interesting) people specially concerned with the digital collection.

The aim of this field is to record outstanding individuals. It should not be used to record all known people who are associated with the collection.

Famous event [famous-event – optional]

Famous (or interesting) events specially concerned with the digital collection, for example Waterloo.

The aim of this field is to record outstanding events. It should not be used to record all known events that are associated with the collection.

Famous place [famous-place – optional]

Famous (or interesting) places specially concerned with the digital collection, for example Mont Fuji.

The aim of this field is to record outstanding places. It should not be confused with the spatial-coverage field and must not be used to record all known places that are associated with the collection.

Famous object [famous-object – optional]

Famous objects or star items specially concerned with the digital collection.

This field is particularly intended for use by small and medium cultural institutions. It should be used where there are one or two exceptional items in the collection that would not otherwise be found by subject indexing. It should not be used to provide a list of all items in the collection.

Section Illustration

Illustrations provide sample data, usually images but could be any media. We only provided here non-technical field definitions. Illustrations are optional; when we say that, for instance, the illustration title is mandatory, it is only mandatory if there is an illustration.

Title [title – mandatory]

The title of the illustration.

Creator [creator – optional]

The creator or originator of the illustration, such as a photographer who captured an image.

Legal status [legal-status – mandatory]

A statement of the rights associated with the illustration.

6.2 Entity Institution

Section Identification

Identifier [identifier – mandatory]

This is the identifier for the institution. This identifier plays an essential role in the MICHAEL system. All entities should have a unique identifier, and this identifier should also be unique in the scope of the European instance.

Name [name – mandatory]

The name of the institution.

The name of the institution should be provided in full. Acronyms and abbreviations must not be used.

Acronym [acronym – optional]

This is the acronym (or abbreviated form of an institution's name) that is commonly used to identify an institution. For example, the 'Museums, Libraries and Archives Council' is commonly abbreviated to the acronym 'MLA'.

Acronyms should be recorded where they exist and are commonly used. They should not be created for the purpose of entering data in the MICHAEL system. Acronyms may or may not form part of the Institution Identifier.

Jurisdiction [jurisdiction – optional]

The organisation the institution is affiliated to or sponsored by. For example, the ministry that funds an institution.

Logo [logo – optional]

A link to an image file that contains the Institution's logo.

Section Description

Institution type [institution-type – recommended]

This is the general activity or sector in which an institution operates, for example archive, museum, library, local community and other.

Administrative status [administrative-status – recommended]

The general administrative status of the institution, for example public, commercial or non-profit making.

Section Location

This section contains address information. An address is an aggregation of the fields that follow, and at least one address must be provided.

Street [street – recommended]

This is a generic field used for the street, the name of a building, the name of a department, room number etc.

PO Box [pobox – optional]

This is the PO box number.

Locality [locality – recommended]

The locality is the smallest administrative area, which is usually a town, city, ward, village or a *commune*, etc.

Postal code [postal-code – recommended]

This is the post-code., for example in the United Kingdom W1A 1AA.

Region [region – recommended]

Region is a general concept, and can be any administrative area that is larger than a locality but smaller than a country. In the United Kingdom, region may include a devolved administration, a government region or a county. In France it may include a region or a *département* and in Italy a region or province.

Country [country – mandatory]

This is the country part of the address.

*Section Communications***Telephone number [telephone – optional]**

This is the general telephone number for the institution, for example the number of the switchboard or for a help/information desk.

The international dialling code must be included in the telephone number to support international use of the MICHAEL system.

Fax number [fax – optional]

This is the general fax number for the institution, for example the number of the switchboard or for a help/information desk.

The international dialling code must be included in the telephone number to support international use of the MICHAEL system.

Email [email – recommended]

This is the general email address for the institution, for example the address for a help/information desk.

URL [url – recommended]

This is the URL of the home page for an institution.

*Section Contact Person***Agent name [agent-name – optional]**

This is the name of the agent or service (for example a department within an organisation).

Agent telephone [telephone – optional]

This is the contact telephone number for the agent or service (for example a department within an organisation).

The international dialling code must be included in the telephone number to support international use of the MICHAEL system.

Agent fax number [fax – optional]

This is the contact fax number for the agent or service (for example a department within an organisation).

The international dialling code must be included in the telephone number to support international use of the MICHAEL system.

Agent email [email – optional]

This is the contact email address for the agent or service (for example a department within an organisation).

6.3 Entity Service or Product*Section Identification***Identifier [identifier – mandatory]**

This is the identifier for the service or product. This identifier plays an essential role in the MICHAEL system. All entities should have a unique identifier, and this identifier should also be unique in the scope of the European instance.

The identifier could consist of a country code followed by a unique identifier for the product or service.

Title [title – mandatory]

This is the name of a service or product.

This should provide reference to the type of service or product that is being provided. It may consist of:

- title by which the service is known, e.g. the title of a website, or
- the name of the collection plus the type of access that is being provided, for example: ‘The works of Caravaggio: CD-ROM’, or
- the type of access, for example a print on demand service.

Section *Description*

Description [description – recommended]

This is a short description of the content of a product or service and the functions that are being provided. It should add to the information provided in the title.

For example “this web-site allows users to search and browse the collections database of the Museum”. It should not replicate the description of the content of the collection.

Language [language – mandatory]

The language(s) in which the product or service is made available.

Maintenance [maintenance – optional]

This is a general indication of the maintenance status of the product or service, for example ongoing (live), complete, regular update and so on.

Audience [audience – recommended]

The intended audience for whom the product or service has been designed.

Legal status [legal-status – recommended]

A statement of the legal status of the product or service.

Section *Access Conditions*

Access type [access-type – mandatory]

This is the general type of service or product that is available.

This should include online, offline, hard-copy, print-on-demand and so on.

Accessibility [accessibility – optional]

A statement of the characteristics of the service or product that make it accessible to users.

For example the availability of a speech enabled browser (for an offline service).

WAI [wai – recommended]

This is the level of conformance to WAI specifications of the service or product, for example A, AA, AAA.

Access conditions [access-conditions – mandatory]

This is a general indication of any conditions on access to the service or products, for example free, charged for, restricted etc.

Comment on access conditions [comment-access-conditions – optional]

This is a brief note, or comments, providing more information about conditions for access to the service or product.

Section *Technical Information*

Technical requirements [technical-requirement – optional]

This is a brief description on the technical requirements for accessing a service or product. For example the plug-ins that a user would require to use a service.

Technical description [technical-description – optional]

This is a link to an external description that provides information about how a remote access service is configured, such as the available inputs and outputs. For example, a link to the specification for an institution's Z target.

Protocol [protocol – optional]

This is the communication protocol, for example Z39.50, OAI-PMH, ZING, etc.

Output format [output – optional]

This is the output format from a service, for example the output format of an OAI target might be XML.

Section *Access Location*

Description of access location [access-location/description – optional]

This is a short description of the means of accessing a service or product.

In the case of an analogue product or an off-line service, it is recommended that all of the relevant information should be provided in this field.

Access locator [access-location/locator – recommended]

A locator, such as an URL, of the access-point for the service or product.

6.4 Entity Project or Programme

Section *Identification*

Identifier [identifier – mandatory]

This is the identifier for the project or programme. This identifier plays an essential role in the MICHAEL system. All entities should have a unique identifier, and this identifier should also be unique in the scope of the European instance.

This could consist of a country code followed by a unique identifier for the project.

Title [title – mandatory]

This is the title of a project or programme.

This should be written in full. Acronyms and abbreviations must not be used as the aim is to provide information that is understandable to a general audience.

Acronym [acronym – optional]

This is the acronym (or abbreviated form) of the project or programme title.

This acronym should be commonly used, it must not be created simply for the MICHAEL system.

Logo [logo – optional]

A link to an image file that contains the project or programme's logo.

Section *Description*

Description [description – recommended]

This is a short description of the project or programme, which should add to the information that is provided in the title.

Digitisation process [digitisation-process – optional]

This is a general indication of the technical features of the project or programme, and of the digitisation processes used. This may be direct or indirect etc.

Funding type [funding-type – optional]

This is a general indication of the type of funding that supports the project or programme, for example internal, external and so on.

Section *Communications*

Email [email – optional]

This is the general email address for the project or programme, for example the address for a help/information desk.

URL [url – optional]

This is the URL of the home page for a project or programme.

Section *Progress*

Start date [start-date – recommended]

This is the approximate starting date of the project or programme. This may be expressed as a year.

Completion date [completion-date – optional]

This is the approximate date when the project or programme ended, where this is known. This may be expressed as a year.

Project status [project-status – optional]

This is a general indication of the status of the project or programme, for example planned, on-going, completed.

Section *Contact Person*

Agent name [agent-name – optional]

This is the name of the agent.

Agent telephone [telephone – optional]

This is the contact telephone number for the agent.

The international dialling code must be included in the telephone number to support international use of the MICHAEL system.

Agent fax number [fax – optional]

This is the contact fax number for the agent.

The international dialling code must be included in the telephone number to support international use of the MICHAEL system.

Agent email [email – optional]

This is the contact email address for the agent.

6.5 Entity Physical Collection

Section *Identification*

Identifier [identifier – mandatory]

This is the identifier for the physical collection. This identifier plays an essential role in the MICHAEL system. All entities should have a unique identifier, and this identifier should also be unique in the scope of the European instance.

Title [title – mandatory]

The title of the physical collection.

The title should provide a meaningful point of reference to the physical collection, and preferably it should be unique, at least within an institution. Acronyms and abbreviations should not be used in this field.

Section *Description*

Description [abstract – recommended]

A free text description of the physical collection. This should add to the information provided in the title.

Language [language – optional]

The language of the material contained in the physical collection.

Physical format [physical-format – recommended]

The physical characteristics of the collection.

Size [size – optional]

An evaluation of the size of the collection. This is intended to provide information for users of the MICHAEL system about the size of the physical collection.

Accrual [accrual – optional]

A statement of accrual policy (closed, passive, active, partial/selective), accrual method (purchase, deposit) and accrual periodicity (closed, irregular, periodic).

This information is especially important for harvesting purposes, to foresee the evolution of aggregated resources.

Standard [standard – optional]

A statement of any descriptive or terminology standards that were used in creating the item level metadata associated with the physical collection.

7) Annex A : XML schema

An XML schema will be provided as a reference definition once the data model is at version 1.0.