

# Représentation de ressources I: MARC

CERN-UNESCO School on Digital Libraries  
Dakar, Nov 21-25, 2011

Annette Holtkamp  
CERN

# Metadonnées

- données fournissant informations sur autres données
- information structurée et descriptive sur une resource
- clé pour découvrir un objet
- utile pour gestion de notices, archivage
- élément de metadonnées:
  - champ pour stocker information spécifique (e.g. titre)
- valeur d'une metadonnée:
  - contenu d'un élément
  - peut être terme d'un vocabulaire défini

# Types de metadonnées

- descriptive
  - identification et recherche
    - titre, auteur, abstract...
- structurelle
  - présentation
    - chapitres d'un livre,...
- administrative
  - gestion et préservation
    - version, informations techniques, contrôle d'accès

# Schéma de metadonnées

- collection définie d'éléments
- sert a un but spécifique
  - e.g. discipline spécifique, type de resource
- spécifie nom et sens des éléments
- règles optionelles
  - contenu, représentation, valeur des éléments, syntax...
- quelques standards de metadonnées
  - MARC, Dublin Core, MODS...

# MARC

## MAchine Readable Cataloguing

- standard international pour représenter et échanger notices bibliographiques
- développé dans les 60s
- orienté aux cartes de cataloguage
- haut degré of complexité
  - all purpose
- base de la plupart des catalogues bibliothécaires, grande communauté d'utilisateurs

<http://www.loc.gov/marc>


# MARC21

- évolution de MARC
- combinaison de MARC formats de Canada et EE.UU.
- internationalization
- Unicode
  - standard for encoding and representing text in multilingual environments
  - > 100k characters
  - 93 scripts

# Formats

- **bibliographique**
  - livres, journaux, articles, fichiers digitaux, cartes géographiques, musique, matériaux visuels, matériaux combinés
- **autorité**
  - formes autorisées de noms et sujets
- **classification**
  - numérique ou textuel
- **fonds**
  - single-part, multi-part and serial items
  - informations appartenant a un spécifique exemplaire
- **renseignements communautaires**
  - non-bibliographic resources of a community
    - scientifiques, institutions, conférences

# Example d'une notice bibliographique

Information	Discussion	Files
 <b>Article</b>		
Report number	<a href="#">arXiv:0801.1651</a> ; CERN-PH-TH-2008-004 ; FTPI-MINN-2008-01 ; UMN-TH-2008-2630	
Title	<b>Sparticle Discovery Potentials in the CMSSM and GUT-less Supersymmetry-Breaking Scenarios</b>	
Author(s)	<a href="#">Ellis, Jonathan Richard</a> (CERN) ; <a href="#">Olive, Keith A</a> (Univ. Minnesota, Minneapolis, MN, USA) ; <a href="#">Sandick, Pearl</a> (Univ. Minnesota, Minneapolis, MN, USA)	
Imprint	11 Jan 2008. - 20 p.	
In:	<a href="#">J. High Energy Phys. 08 (2008) 013</a>	
Subject category	Particle Physics - Phenomenology	
Abstract	<p>We consider the potentials of the LHC and a linear <math>e^+e^-</math> collider (LC) for discovering supersymmetric particles in variants of the MSSM with soft supersymmetry-breaking mass parameters constrained to be universal at the GUT scale (CMSSM) or at some lower scale <math>M_{\text{in}}</math> (GUT-less models), as may occur in some scenarios with mirage unification. Whereas the LHC should be able to discover squarks and/or gluinos along all the CMSSM coannihilation strip where the relic neutralino LSP density lies within the range favoured for cold dark matter, many GUT-less models could escape LHC detection. In particular, if <math>M_{\text{in}} &lt; 10^{11}</math> GeV, the LHC would not detect sparticles if the relic density lies within the favoured range. For any given discovery of supersymmetry at the LHC, in such GUT-less models the lightest neutralino mass and hence the threshold for sparticle pair production at a LC increases as <math>M_{\text{in}}</math> decreases, and the CMSSM offers the best prospects for measuring sparticles at a LC. For example, if the LHC discovers sparticles with <math>1 \text{ fb}^{-1}</math> of data, within the CMSSM a centre-of-mass energy of 600 GeV would suffice for a LC to produce pairs of neutralinos, if they provide the cold dark matter, whereas over 1 TeV might be required in a general GUT-less model. These required energies increase to 800 GeV in the CMSSM and 1.4 TeV in GUT-less models if the LHC requires <math>10 \text{ fb}^{-1}</math> to discover supersymmetry.</p>	
Copyright/License	Open Access	




# Notice d'une conference

Information Discussion Files

## Conference

Conference title	<b>2010 European School of High-energy Physics</b>
Related conference title(s)	ESHEP 2010
Date(s), location	20 Jun – 3 Jul 2010, Raseborg, Finland
Conference contact	Helene Haller, CERN Schools of Physics, CH-1211 Geneva 23, Switzerland email: <a href="mailto:Physics.School@cern.ch">Physics.School@cern.ch</a> fax: 00-41-22-767-9575 00-41-22-767-3632
Imprint	2010
Series	<a href="#">(European School of High-Energy Physics)</a> <a href="#">(CERN School of Physics)</a>
Subject category	Particle Physics

External link:

 [Physics School home page](#)

---

**Contributions to this conference in CDS**

[Hard Probes in Heavy-Ion Physics](#)  
*by Renk, Thorsten*

[QCD for Collider Physics](#)  
*by Skands, Peter*

# Notice d'une institution

**Cheikh Anta Diop of Dakar U.** [Future INSPIRE ID:U. Cheikh Anta Diop, Dakar]

Université Cheikh Anta Diop (UCAD)

BP 5005 Dakar-Fann

Dakar

Senegal

<http://www.ucad.sn/>

→ [9 Papers from Cheikh Anta Diop of Dakar U.](#)

# Notice d'un scientifique

**Roy Aleksan** ([Saclay](#)) [[Publication list](#)] [[Google](#)] [[Students](#)] [[arXiv](#)] [[ADS](#)]

**PhD Advisor:** [Ducros, Yves](#)

**PhD Institution:** [Paris U., VI-VII](#)

**Undergrad:** [Paris U., VI-VII](#)

**Email:** [aleksan@hep.saclay.cea.fr](mailto:aleksan@hep.saclay.cea.fr)

**Field:** ACC-PHYS, HEP-PH

## Institutional History:

Institution	Rank	Start Date	End Date
-------------	------	------------	----------

[Saclay](#)

<a href="#">Paris U., VI-VII</a>	PHD		1986
----------------------------------	-----	--	------

<a href="#">Paris U., VI-VII</a>	UG		
----------------------------------	----	--	--

# Offre d'emploi

## Lecturer in Nuclear Physics

JOBS-94234

[Witwatersrand U.](#) - Junior

**Field of Interest:** nucl-ex, nucl-th

**Region:** AF

**Job description:**

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG FACULTY OF SCIENCE

LECTURER in NUCLEAR PHYSICS

SCHOOL OF PHYSICS

A post of Lecturer in Nuclear Physics is available immediately.

Research fields in the School include theoretical and experimental condensed matter and materials physics, theoretical and experimental high energy physics, nuclear physics, radio astronomy and astrophysics. The DST/NRF Centre of Excellence in Strong Materials, hosted by the University of the Witwatersrand, is well established, in which the School of Physics plays a leading role. The School hosts the Materials Physics Research Institute, the Gauteng node of the National Institute for Theoretical Physics and three DST/NRF Chairs/Associate Chairs.

### Qualifications

Preference will be given to applicants with a background in low and medium-energy experimental nuclear physics together with knowledge of radiation protection, and will be expected to teach courses in these areas at both undergraduate and postgraduate levels. A PhD is a requirement, and the successful candidate should have had relevant postdoctoral experience and have a corresponding publication record. Teaching experience in these areas would be an advantage. There is a very active nuclear structure physics group and the appointee will be encouraged to join and enhance the activities at iThemba LABS, a national facility running both an EN Tandem accelerator and K200 cyclotron.

### Duties

Lecturers are expected to pursue research, teach and participate in the administration of the School, Faculty and University.

# Example d'une notice bibliographique - MARC

001\_\_ 1080272  
003\_\_ SzGeCERN  
005\_\_ 20081003111503.0  
0247\_ \$\$a10.1088/1126-6708/2008/08/013\$\$2DOI  
0248\_ \$\$aoai:cds.cern.ch:1080272\$\$pcerncds:CERN  
035\_\_ \$\$9arXiv\$\$aoai:arXiv.org:0801.1651  
035\_\_ \$\$9SPIRES\$\$a7620977  
037\_\_ \$\$aarXiv:0801.1651  
041\_\_ \$\$aeng  
088\_\_ \$\$aCERN-PH-TH-2008-004  
088\_\_ \$\$aFTPI-MINN-2008-01  
100\_\_ \$\$aEllis, Jonathan Richard\$\$uCERN  
245\_\_ \$\$aParticle Discovery Potentials in the CMSSM and GUT-less Supersymmetry-Breaking  
Scenarios  
269\_\_ \$\$c11 Jan 2008  
300\_\_ \$\$a20 p  
520\_\_ \$\$aWe consider the potentials of the LHC and a linear  $e^+e^-$  collider (LC) for discovering ...  
595\_\_ \$\$aOA  
65017 \$\$2arXiv\$\$aParticle Physics - Phenomenology  
690C\_ \$\$aARTICLE  
690C\_ \$\$aCERN  
700\_\_ \$\$aOlive, Keith A\$\$uUniv. Minnesota, Minneapolis, MN, USA  
773\_\_ \$\$c013\$\$pJ. High Energy Phys.\$\$v08\$\$y2008  
8564\_ \$\$uhttp://arxiv.org/pdf/0801.1651.pdf\$\$yFulltext  
8564\_ \$\$uhttp://cdsweb.cern.ch/record/1080272/files/jhep082008013.pdf\$\$ySISSA/IOP OA article

# Notice bibliographique - structure

- Guide (leader)
  - information basique sur l'objet  
e.g. type de matériau
  - renseignements sur le traitement de la notice  
longueur de notice, schéma d'encoder les caractères...
  - longueur fixe, premiers 24 caractères d'une notice
  - important pour l'échange entre différents systèmes
- répertoire (directory)
  - Index des locations des champs de contrôle et données  
12 caracteres a position 24

zones de longueur variable:

- zones de contrôle 00x
- zones de données

# Zones de contrôle 00x

001 – numéro de contrôle / numéro de système

003 – control number identifier, MARC code de l'organisation

005 – date et temps de la dernière transaction, pour identifier version

001\_\_ 1080272

003\_\_ SzGeCERN

005\_\_ 20081003111503.0

# Zones de données - structure

- étiquette numérique a 3 caractères
  - x9y, 9xy: réservé à usage local
- souvent répétable
- jusqu'à 2 indicateurs
  - interprètent ou complètent les données dans ce zone
  - caractères numériques ou alphabétiques en minuscule
- plusieurs sous-zones
  - caractères numériques ou alphabétiques en minuscule
  - indépendemment définies pour chaque zone
  - parfois répétable

65017\$\$2arXiv\$\$aParticle Physics - Phenomenology



# Zones de données - classes

0xx – zones de numéros et codes, classification

1xx – vedettes principales

2xx – zones de titre/publication

3xx – description matérielle

4xx – mentions de collection

5xx – notes

6xx – zones des vedettes-matieres

7xx – vedettes secondaires, zones de liaison

8xx – vedettes de location, fonds, location...

9xx – réservé pour implémentation locale

liste complète at <http://www.loc.gov/marc/bibliographic/>

# 01x-04x – zones de numeros et codes

010 – Library of Congress control number

020 – ISBN

\$a – ISBN

\$u – medium (non-standard)

020\_\_ \$\$a9783540632931\$\$uprint version, paperback

022 – ISSN

024 – autres identifiants standards (e.g. DOI)

041 – code de langage

e.g. eng pour Anglais

# 05x-08x – indices de classification et cotes

050 – Library of Congress call number

080 – UDC Universal Decimal Classification number

080\_\_ \$\$a514.763

082 – DDC Dewey Decimal Classification number

084 – other classification number

088 – report series number

088\_\_ \$\$aCERN-PH-TH-2010-240

# 1xx – vedettes principales

## 100 – Personal name

\$a - personal name

\$e – relator term

\$u – affiliation

\$i – author id (undefined subfield, used by Inspire)

100\_\_ \$\$aClerboux, Barbara\$\$e. \$\$iINSPIRE-00314890\$\$uBrussels U.

## 110 – Corporate name

\$a – corporate name

\$b – subordinate unit

\$g – acronym

110\_\_ \$\$aCentre des Recherches Nucleaires\$\$gCERN

# 2xx – titre + publication

## 245 – Titre

\$a – Titre

\$b – sous-titre

245\_\_ \$\$aRemoving The Haystack\$\$bThe CMS Trigger and Data Acquisition Systems

## 246 – variante du titre

## 242 – titre traduit

## 250 – édition statement

\$a – édition

## 260 – publication

\$a – lieu de publication

\$b – nom de l'éditeur

\$c – date de publication

260\_\_ \$\$aLondon\$\$bImperial College Press\$\$c2010

# 3xx – description matérielle

## 300 – description matérielle

\$a – nombre de pages, duree en minutes...

\$c – dimensions

300\_\_ \$\$aStreaming video ; 2 DVD video\$\$c720x576 4/3, 25

# 4xx – Mention de collection

490 – mention de collection

\$a – serie

\$v – volume

490\_\_ \$\$aLecture Notes in Mathematics\$\$v1358

# 5xx – note fields

500 – note generale

502 – note de thèses

\$b – genre de diplôme (e.g. PhD)

\$c – institution conferant le grade

\$d – annee du diplôme

502\_\_\$bPhD\$cDakar U.\$d2011

506 – restrictions d'utilisation

indicator 1

0 – no restriction

1 – restrictions apply

\$a – conditions régissant la consultation

\$d – usagers autorisés

5061\_ \$aRestricted\$dais-users [CERN]

520 – résumé

\$a – résumé

540 – modalités d'utilisation et de reproduction

\$a – modalités d'utilisation et de reproduction, e.g. CC license

\$b – personne ou organization imposant la restriction

\$u – indentificateur de ressource uniforme (URI)

542 – statut du droit d'auteur



# 6xx – Zones des vedettes-matière

## 650 – vedette-matière (nom commun)

indicateur 1: niveau de la vedette-matière

1 – principal

2 – secondaire

indicateur 2: thésaurus

0 – vedettes-matière de Library of Congress

7 – source spécifiée dans la sous-zone \$2

\$a – sujet ou nom géographique

\$2 – source

65017 \$\$2arXiv\$aParticle Physics - Theory

## 653 – terme d'indexation – vedette non contrôlée

\$a – terme non contrôlée (e.g. donne par l'auteur)

\$9 – source (e.g. auteur) (non standard, CDS/Inspire usage)

6531\_ \$\$9CERN\$acomputer networks

## 69x – zones d'accès local aux sujets

# 7xx – vedettes secondaires

700 – auteurs additionnels

710 – nom de collectivite additionnel

# 76x-78x – zones de liaison

spécifie la relation avec un autre objet

773 – liaison au document hôte

relation vertical (chapitres d'un livre, articles d'un journal)

773\_\_ \$\$c75-78\$\$pPhys.Lett.B\$\$v700\$\$y2011

775 – liaison a une autre edition

787 – autre relation

\$w – numéro de contrôle de la notice reliée

\$i – information sur la relation

exemple: lien entre projections d'une présentation et le rapport de la conférence

787\_\_ \$\$w1234567\$\$islides

# 85x – fonds, location

852 – emplacement

856 – emplacement et accès électronique

indicator 1: méthode d'accès

4: http

\$q – type de format électronique (html, pdf, jpeg...)

\$u – URI

\$y – texte du lien

8564\_ \$\$u<http://arxiv.org/pdf/1011.1200.pdf>\$\$yPreprint

# 9xx – zones locales

exemple de CDS:

999 – références

\$o numéro de la référence

\$h auteurs

\$a DOI

\$u Uniform Resource Identifier

\$r report number

\$s journal publication

\$m reste non identifié

R. W. Robinett and J. L. Rosner, Phys. Rev. D 25, 3036 (1982)

999C5\$\$o1\$\$hR.W. Robinett and J.L. Rosner\$\$sPhys. Rev. D 25  
(1982) 3036\$\$a10.1103/PhysRevD.25.3036

# MARC XML

- XML schéma basé sur MARC21
- développé par Library of Congress
- XML: Extensible Markup Language
  - collection de règles pour encoder des structures de données arbitraires
  - contenu (metadonnées) indépendant de la présentation

# MARC XML: elements

- <collection>
  - file of several records
- <record>
  - delineates records within a collection
- <leader>
  - MARC leader data string
- <control field>
  - MARC control field data string
- <data field>
- <subfield>

# MARC XML: datafield

- MARC balises et indicateurs représentés comme attributs d'un élément "datafield"

```
<datafield tag="100" ind1="1" ind2=" " >
```

- chaque sous-zone un element

- code de sous-zone comme attribut

```
<subfield code="a">...</subfield>
```

Exemple: book editor

```
100__ $$aClerbaux, Barbara$$eed.$$iINSPIRE-00314890$$uBrussels U.
```

```
<datafield tag="100" ind1=" " ind2=" " >
```

```
<subfield code="a">Clerbaux, Barbara</subfield>
```

```
<subfield code="e">ed.</subfield>
```

```
<subfield code="i">INSPIRE-00314890</subfield>
```

```
<subfield code="u">Brussels U.</subfield>
```

```
</datafield>
```



# MARC XML

- aim: easy sharing of bibl info
- easy access at subfield level
- lossless conversion from MARC21
- manipulated and transformed via XSL stylesheets
  - Extensible Stylesheet Language
- “bus” for conversion between different standards