fi yuo cna raed tihs, yuo hvae a sgtrane mnid. Cna yuo raed tihs? Olny 55 plepoe out of 100 can

Making your Metadata Beautiful

13 July 2016

Heather Rosie & Sara Gould

The British Library e-theses online service (EThOS)

What is ‘beautiful’ metadata?

“ACRC” …

- **Accurate**: correct transcription; no duplication; no typos
- **Consistent**: same type of data recorded in the same place; data presented in the same way
- **Rich**: as much relevant data as possible, e.g. abstracts, identifiers, funder, supervisor
- **Current**: material available now needs to be discoverable now

Benefits of beautiful (“ACRC”) metadata

- Optimises resource discovery – and increases usage
- Maximises workflow efficiencies
- Supports accurate and reliable statistics/metrics (e.g. number of theses awarded in the UK by institution, subject or date)
- Increases interoperability - allowing data from multiple sources to be combined

Challenge 1- record sources

1. British Library catalogue records
2. University catalogue records
3. Institutional repository harvested records
4. BL Cataloguer created records
5. BL Imaging Team created records

Challenge 2 - metadata formats

1. MARC Exchange Format (ISO 2709)
2. MARC XML
3. OAI_DC XML
4. UKETD_DC XML
5. RIOXX
6. Tab-delimited (Excel)

Challenge 3 - skills/knowledge of the people creating metadata

1. Cataloguers
2. Students
3. Repository staff (managers, admin, clerical)

Challenge 4 - processing and configuration of data storage and export

1. Programmers
2. Repository Managers
3. Administrative staff
4. EThOS Metadata Manager

Challenge 5 - special characters/diacritics

1. Spelt out (‘alpha’)
2. UTF-8 character set
3. Local convention (e.g. ^ for superscript number)
4. Data represented in $ strings
5. Characters omitted, garbled or replaced by ??

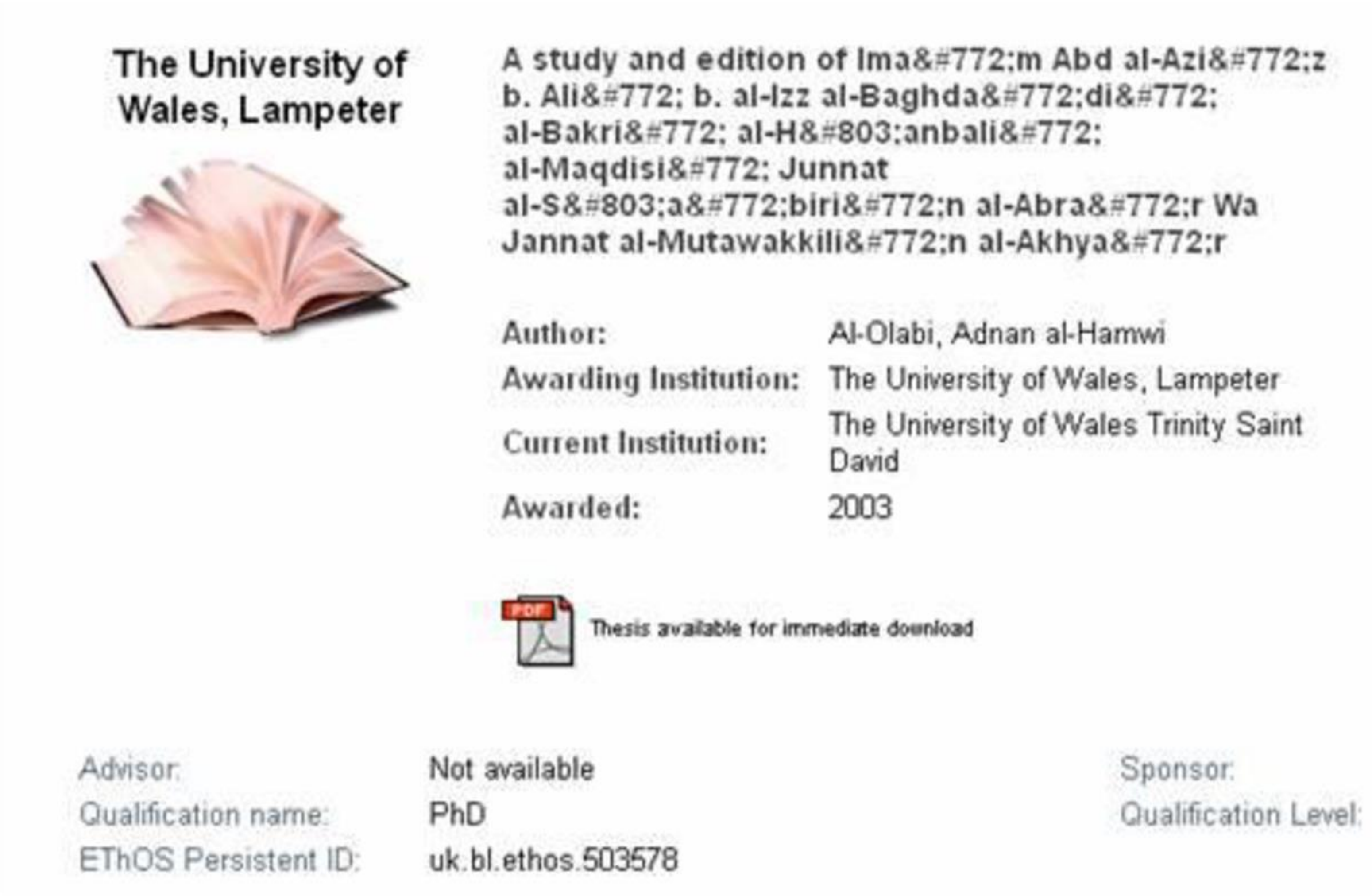
All these challenges together can produce ‘poor’ metadata

- Character encoding (MARC-8 v. UTF-8)
- Typographical errors
- Scanned texts (machine error)
- Missing data (abstracts, subtitles, authors, language, subjects, qualification name)
- No standardisation (e.g. qualification name; author names; representation of words/symbols)
- Lack of identifiers
- Duplication

Character encoding

- Web standard is UTF-8 (Unicode);   MARC character set = MARC-8
- Software applications may not recognise ‘foreign’ character sets and will display as strange characters and/or question mark(s)
- Some characters are ‘invalid’ in XML (and some invalid characters are invisible, e.g. ‘null’); web browsers will not display XML documents containing invalid characters
- Use of invalid characters impacts on resource discovery and display and, in some cases, can cause software failure

Diacritics in title



The University of Wales, Lampeter

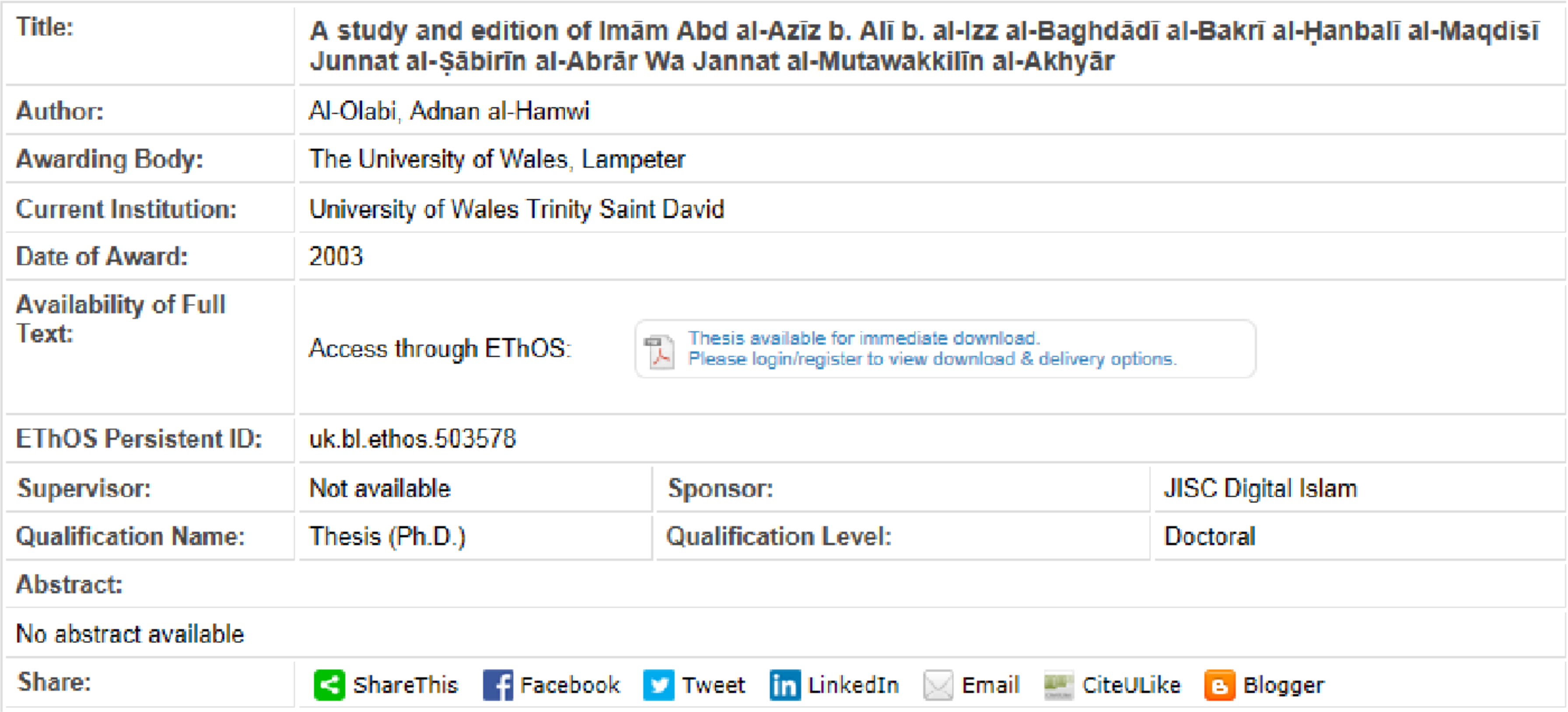
A study and edition of Ima&#772;m Abd al-Azi&#772;z b. Ali&#772; b. al-Izz al-Baghda&#772;di&#772; al-Bakri&#772; al-H&#803;anbali&#772; al-Maqdisi&#772; Junnat al-S&#803;a&#772;biri&#772;n al-Abra&#772;r Wa Jannat al-Mutawakkili&#772;n al-Akhya&#772;r

Author:	Al-Olabi, Adnan al-Hamwi
Awarding Institution:	The University of Wales, Lampeter
Current Institution:	The University of Wales Trinity Saint David
Awarded:	2003

Thesis available for immediate download

Advisor:	Not available	Sponsor:	
Qualification name:	PhD	Qualification Level:	
EThOS Persistent ID:	uk.bl.ethos.503578		

Amended diacritics in title



Title:	**A study and edition of Imām Abd al-Azīz b. Alī b. al-Izz al-Baghdādī al-Bakrī al-Ḥanbalī al-Maqdisī Junnat al-Ṣābirīn al-Abrār Wa Jannat al-Mutawakkilīn al-Akhyār**		
Author:	Al-Olabi, Adnan al-Hamwi		
Awarding Body:	The University of Wales, Lampeter		
Current Institution:	University of Wales Trinity Saint David		
Date of Award:	2003		
Availability of Full Text:	Access through EThOS:	Thesis available for immediate download. Please login/register to view download & delivery options.	
EThOS Persistent ID:	uk.bl.ethos.503578		
Supervisor:	Not available	Sponsor:	JISC Digital Islam
Qualification Name:	Thesis (Ph.D.)	Qualification Level:	Doctoral
Abstract:			
No abstract available			
Share:	ShareThis Facebook Tweet LinkedIn Email CiteULike Blogger		

Typographical errors (human)

Causes include:

- Focusing on quantity rather than quality due to work pressures/targets
- Lack of interest/boredom
- Seeing what you think you see (SWYTYS) rather than what is actually there

fi yuo cna raed tihs, yuo hvae a sgtrane mnid too. Cna yuo raed tihs? Olny 55 plepoe out of 100 can

i cdnuolt blveiee taht I cluod aulaclty uesdnatnrd waht I was rdanieg. The phaonmneal pweor of the hmuan mnid, aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it dseno't mtaetr in waht oerdr the lttere s in a wrod are, the olny iproamtnt tihng is taht the frsit and lsat ltteer be in the rghit pclae.

The rset can be a taotl mses and you can sitll raed it whotuit a pboerlm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe. Azanmig huh? Yaeh, and I awlyas tghuhot slpeling was ipmorantt! If you can raed this bceome a fna.

The ‘SWYTYS’ effect

special reference	specific reference
private	pirate
conservation	conversation
hypertensive	hypersensitive
isoform	inform
economic aspects	economic analysis
biochemical	biomechanical
experiential	experimental

Title errors

Awarenees in ageing

Awareness in ageing

Efficient parallel genetic algorithms applied to numerical optimisation

Efficient Parallel Genetic Alogrithms Applied to Numerical Optimisation

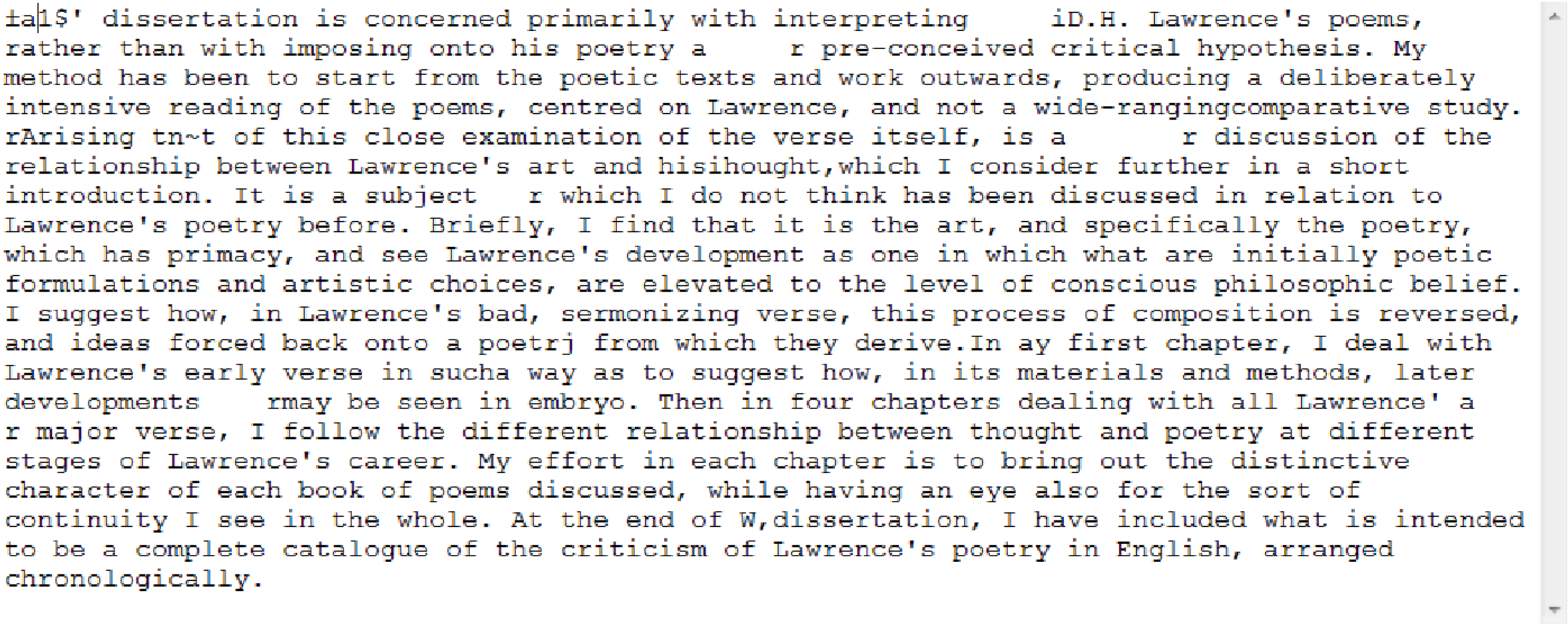
The doll : the figure of the doll in culture and theory

The figure of the doll in culture and theory

OCR errors (machine)

- Scanning texts is imperfect
- Special characters are often not recognised
- Words/sentences may be concatenated
- Review by humans is subject to the "SWYTYS" effect

OCR errors in abstract



±a1$' dissertation is concerned primarily with interpreting iD.H. Lawrence's poems, rather than with imposing onto his poetry a r pre-conceived critical hypothesis. My method has been to start from the poetic texts and work outwards, producing a deliberately intensive reading of the poems, centred on Lawrence, and not a wide-rangingcomparative study. rArising tn~t of this close examination of the verse itself, is a r discussion of the relationship between Lawrence's art and hisihought,which I consider further in a short introduction. It is a subject r which I do not think has been discussed in relation to Lawrence's poetry before. Briefly, I find that it is the art, and specifically the poetry, which has primacy, and see Lawrence's development as one in which what are initially poetic formulations and artistic choices, are elevated to the level of conscious philosophic belief. I suggest how, in Lawrence's bad, sermonizing verse, this process of composition is reversed, and ideas forced back onto a poetrj from which they derive.In ay first chapter, I deal with Lawrence's early verse in sucha way as to suggest how, in its materials and methods, later developments rmay be seen in embryo. Then in four chapters dealing with all Lawrence' a r major verse, I follow the different relationship between thought and poetry at different stages of Lawrence's career. My effort in each chapter is to bring out the distinctive character of each book of poems discussed, while having an eye also for the sort of continuity I see in the whole. At the end of W,dissertation, I have included what is intended to be a complete catalogue of the criticism of Lawrence's poetry in English, arranged chronologically.

Special characters (e.g. beta)

Many ways to write special characters:

- Spelt out (beta)
- Roman alphabet letter rather than Greek alphabet (b)
- Image rather than text (β)
- Character entity reference (&#946;)
- UTF-8 value (β)

Encodings for Greek beta (fileformat.info)

HTML Entity (decimal)	&#946;
HTML Entity (hex)	&#x3b2;
HTML Entity (named)	&beta;
How to type in Microsoft Windows	Alt +3B2
UTF-8 (hex)	0xCE 0xB2 (ceb2)
UTF-8 (binary)	11001110:10110010
UTF-16 (hex)	0x03B2 (03b2)
UTF-16 (decimal)	946
UTF-32 (hex)	0x000003B2 (3b2)
UTF-32 (decimal)	946
C/C++/Java source code	"\u03B2"
Python source code	u"\u03B2"
	More...

Example of special character errors

Title loaded to EThOS (from Aleph):

Growth and characterisation of terrace graded virtual substrates with SiÃ¢â€šÂÃ¢â€šâ€¹Ã‹Â£GeÃ‹Â£ 0.15 Ã¢â€°Â¤ x Ã¢â€°Â¤ 1

Harvested title:

Growth and characterisation of terrace graded virtual substrates with Si[subscript 1-x]Ge[subscript x] 0.15 ? x ? 1

Correct title (i.e. utf-8 values):

Growth and characterisation of terrace graded virtual substrates with $Si_{1-x}Ge_x$ $0.15 \leq x \leq 1$

How to create beautiful (ACRC) metadata

1. Quality control is essential
2. Combination of human and machine review
3. Software helps with the task

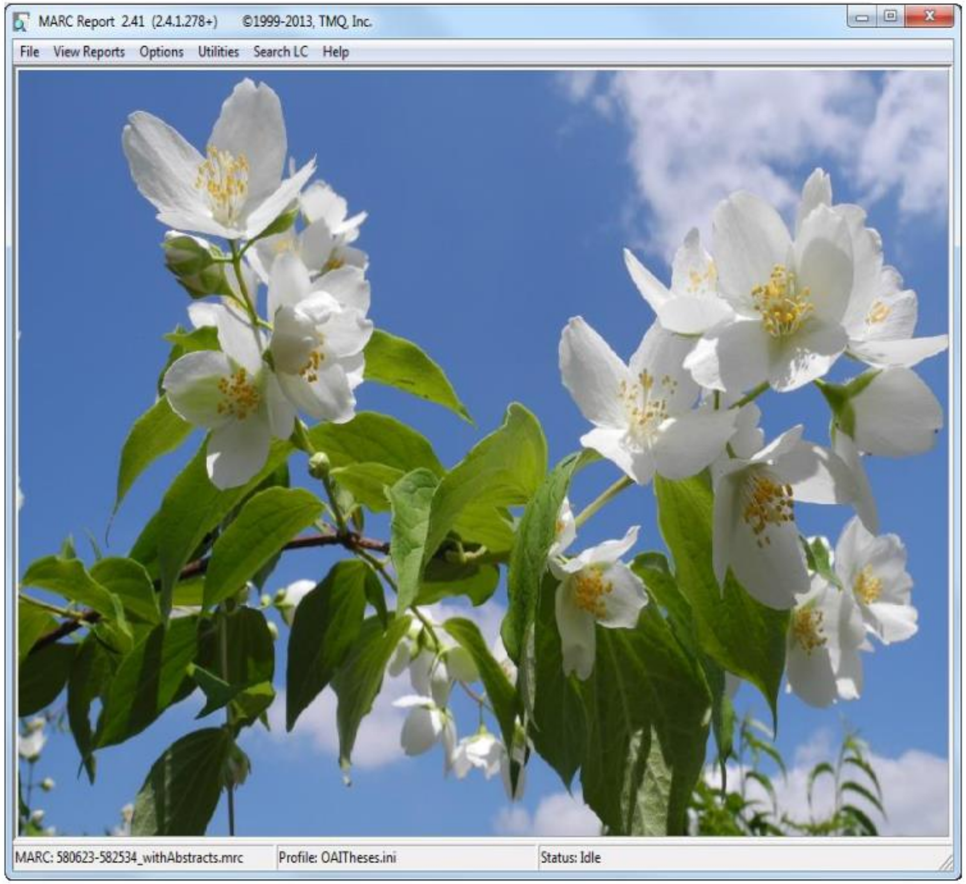
e.g. MARC Report

http://www.marcofquality.com/soft/softindex.html

MARC Report/MARC Global

MARC Report can be used for:

- Checking errors (including invalid utf-8 characters)
- Creating a ‘match key’ and title list for de-duplication
- Converting data from XML to MARC and vice versa
- Lots of other applications!

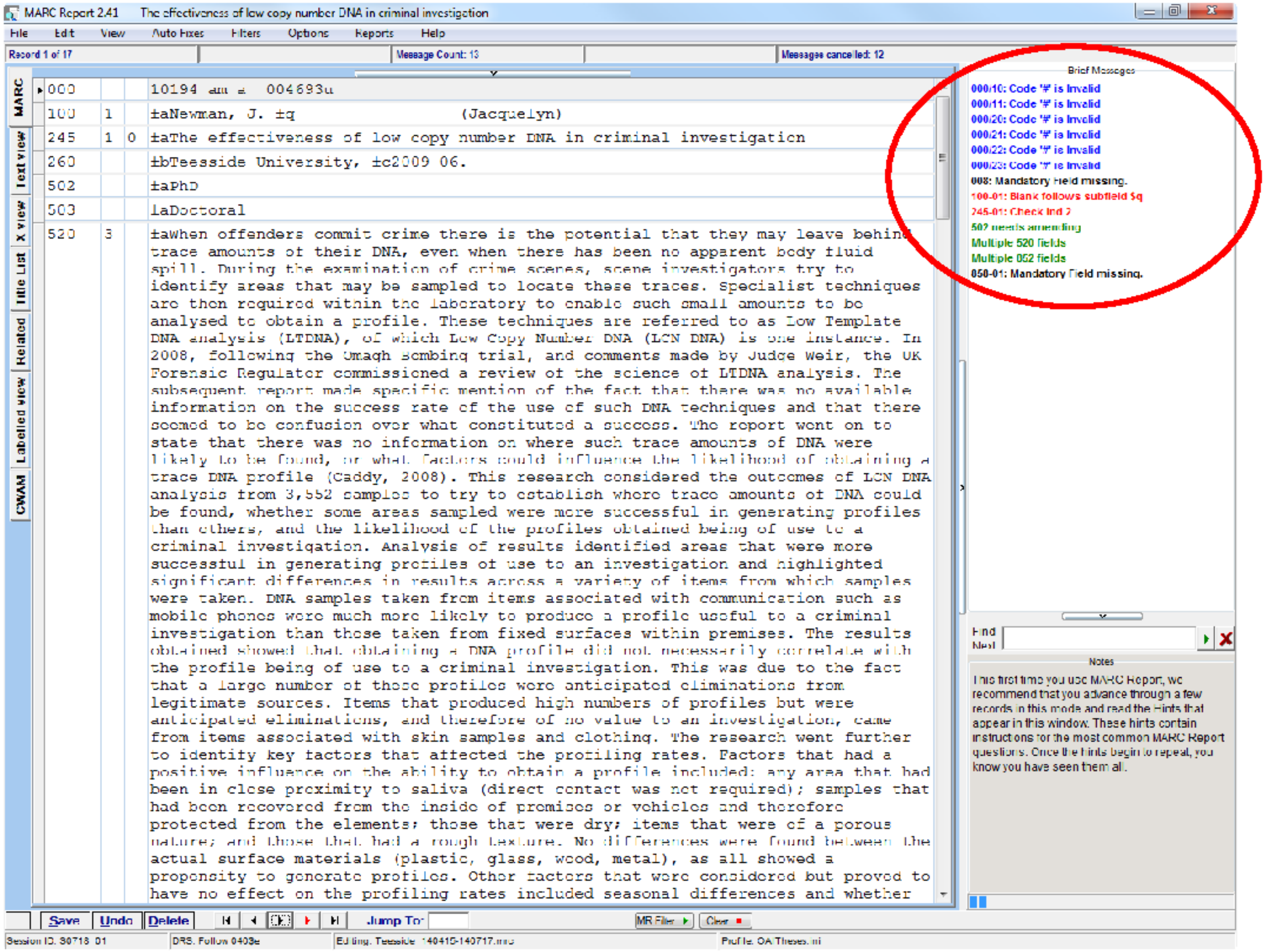


MARC: 580623-582534_withAbstracts.mrc

Profile: OAITheses.ini

Status: Idle

Data correction (errors highlighted)



The effectiveness of low copy number DNA in criminal investigation

Record 1 of 17 | Message Count: 13 | Messages cancelled: 12

```
000          10194 am a   004693u
100   1      ‡aNewman, J. ‡q                    (Jacquelyn)
245   1  0   ‡aThe effectiveness of low copy number DNA in criminal investigation
260          ‡bTeesside University, ‡c2009 06.
502          ‡aPhD
503          ‡aDoctoral
520   3      ‡aWhen offenders commit crime there is the potential that they may leave behind
             trace amounts of their DNA, even when there has been no apparent body fluid
             spill. During the examination of crime scenes, scene investigators try to
             identify areas that may be sampled to locate these traces. Specialist techniques
             are then required within the laboratory to enable such small amounts to be
             analysed to obtain a profile. These techniques are referred to as Low Template
             DNA analysis (LTDNA), of which Low Copy Number DNA (LCN DNA) is one instance. In
             2008, following the Omagh Bombing trial, and comments made by Judge Weir, the UK
             Forensic Regulator commissioned a review of the science of LTDNA analysis. The
             subsequent report made specific mention of the fact that there was no available
             information on the success rate of the use of such DNA techniques and that there
             seemed to be confusion over what constituted a success. The report went on to
             state that there was no information on where such trace amounts of DNA were
             likely to be found, or what factors could influence the likelihood of obtaining a
             trace DNA profile (Caddy, 2008). This research considered the outcomes of LCN DNA
             analysis from 3,552 samples to try to establish where trace amounts of DNA could
             be found, whether some areas sampled were more successful in generating profiles
             than others, and the likelihood of the profiles obtained being of use to a
             criminal investigation. Analysis of results identified areas that were more
             successful in generating profiles of use to an investigation and highlighted
             significant differences in results across a variety of items from which samples
             were taken. DNA samples taken from items associated with communication such as
             mobile phones were much more likely to produce a profile useful to a criminal
             investigation than those taken from fixed surfaces within premises. The results
             obtained showed that obtaining a DNA profile did not necessarily correlate with
             the profile being of use to a criminal investigation. This was due to the fact
             that a large number of those profiles were anticipated eliminations from
             legitimate sources. Items that produced high numbers of profiles but were
             anticipated eliminations, and therefore of no value to an investigation, came
             from items associated with skin samples and clothing. The research went further
             to identify key factors that affected the profiling rates. Factors that had a
             positive influence on the ability to obtain a profile included: any area that had
             been in close proximity to saliva (direct contact was not required); samples that
             had been recovered from the inside of premises or vehicles and therefore
             protected from the elements; those that were dry; items that were of a porous
             nature; and those that had a rough texture. No differences were found between the
             actual surface materials (plastic, glass, wood, metal), as all showed a
             propensity to generate profiles. Other factors that were considered but proved to
             have no effect on the profiling rates included seasonal differences and whether
```

Brief Messages

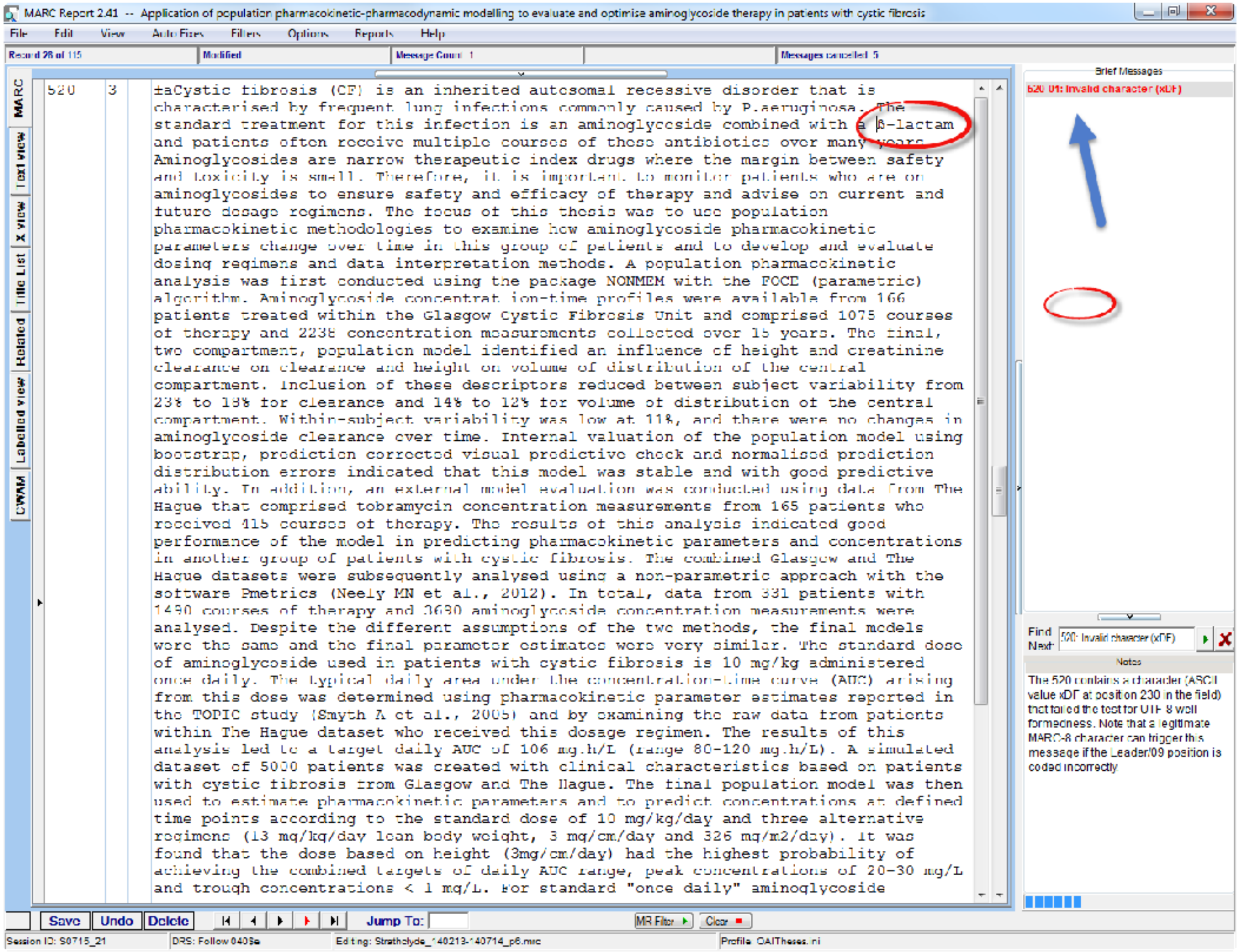
000/10: Code '#' is Invalid
000/11: Code '#' is Invalid
000/20: Code '#' is Invalid
000/21: Code '#' is Invalid
000/22: Code '#' is Invalid
000/23: Code '#' is Invalid
008: Mandatory Field missing.
100-01: Blank follows subfield $q
245-01: Check Ind 2
502 needs amending
Multiple 520 fields
Multiple 052 fields
050-01: Mandatory Field missing.

Notes

This first time you use MARC Report, we recommend that you advance through a few records in this mode and read the Hints that appear in this window. These hints contain instructions for the most common MARC Report questions. Once the hints begin to repeat, you know you have seen them all.

Session ID: 30718 01 | DRS. Follow 0403e | Editing: Teesside 140415-140717.mrc | Profile: OAITheses.ini

Unicode errors: Invalid Greek letter


MARC Report 2.41 -- Application of population pharmacokinetic-pharmacodynamic modelling to evaluate and optimise aminoglycoside therapy in patients with cystic fibrosis
File
Edit
View
Auto Fixes
Filters
Options
Reports
Help
MARC
Text view
X view
Title List
Related
Labelled view
CWAM
Brief Messages
Find Next
Notes
Save
Undo
Delete
Jump To:
MR Filter
Clear

```
Record 28 of 115    Modified    Message Count 1    Messages cancelled 5
```

```
520  3  ‡aCystic fibrosis (CF) is an inherited autosomal recessive disorder that is
        characterised by frequent lung infections commonly caused by P.aeruginosa. The
        standard treatment for this infection is an aminoglycoside combined with a ß-lactam
        and patients often receive multiple courses of these antibiotics over many years.
        Aminoglycosides are narrow therapeutic index drugs where the margin between safety
        and toxicity is small. Therefore, it is important to monitor patients who are on
        aminoglycosides to ensure safety and efficacy of therapy and advise on current and
        future dosage regimens. The focus of this thesis was to use population
        pharmacokinetic methodologies to examine how aminoglycoside pharmacokinetic
        parameters change over time in this group of patients and to develop and evaluate
        dosing regimens and data interpretation methods. A population pharmacokinetic
        analysis was first conducted using the package NONMEM with the FOCE (parametric)
        algorithm. Aminoglycoside concentration-time profiles were available from 166
        patients treated within the Glasgow Cystic Fibrosis Unit and comprised 1075 courses
        of therapy and 2238 concentration measurements collected over 15 years. The final,
        two compartment, population model identified an influence of height and creatinine
        clearance on clearance and height on volume of distribution of the central
        compartment. Inclusion of these descriptors reduced between subject variability from
        23% to 18% for clearance and 14% to 12% for volume of distribution of the central
        compartment. Within-subject variability was low at 11%, and there were no changes in
        aminoglycoside clearance over time. Internal valuation of the population model using
        bootstrap, prediction corrected visual predictive check and normalised prediction
        distribution errors indicated that this model was stable and with good predictive
        ability. In addition, an external model evaluation was conducted using data from The
        Hague that comprised tobramycin concentration measurements from 165 patients who
        received 415 courses of therapy. The results of this analysis indicated good
        performance of the model in predicting pharmacokinetic parameters and concentrations
        in another group of patients with cystic fibrosis. The combined Glasgow and The
        Hague datasets were subsequently analysed using a non-parametric approach with the
        software Pmetrics (Neely MN et al., 2012). In total, data from 331 patients with
        1490 courses of therapy and 3630 aminoglycoside concentration measurements were
        analysed. Despite the different assumptions of the two methods, the final models
        were the same and the final parameter estimates were very similar. The standard dose
        of aminoglycoside used in patients with cystic fibrosis is 10 mg/kg administered
        once daily. The typical daily area under the concentration-time curve (AUC) arising
        from this dose was determined using pharmacokinetic parameter estimates reported in
        the TOPIC study (Smyth A et al., 2005) and by examining the raw data from patients
        within The Hague dataset who received this dosage regimen. The results of this
        analysis led to a target daily AUC of 106 mg.h/L (range 80-120 mg.h/L). A simulated
        dataset of 5000 patients was created with clinical characteristics based on patients
        with cystic fibrosis from Glasgow and The Hague. The final population model was then
        used to estimate pharmacokinetic parameters and to predict concentrations at defined
        time points according to the standard dose of 10 mg/kg/day and three alternative
        regimens (13 mg/kg/day lean body weight, 3 mg/cm/day and 326 mg/m2/day). It was
        found that the dose based on height (3mg/cm/day) had the highest probability of
        achieving the combined targets of daily AUC range, peak concentrations of 20-30 mg/L
        and trough concentrations < 1 mg/L. For standard "once daily" aminoglycoside
```

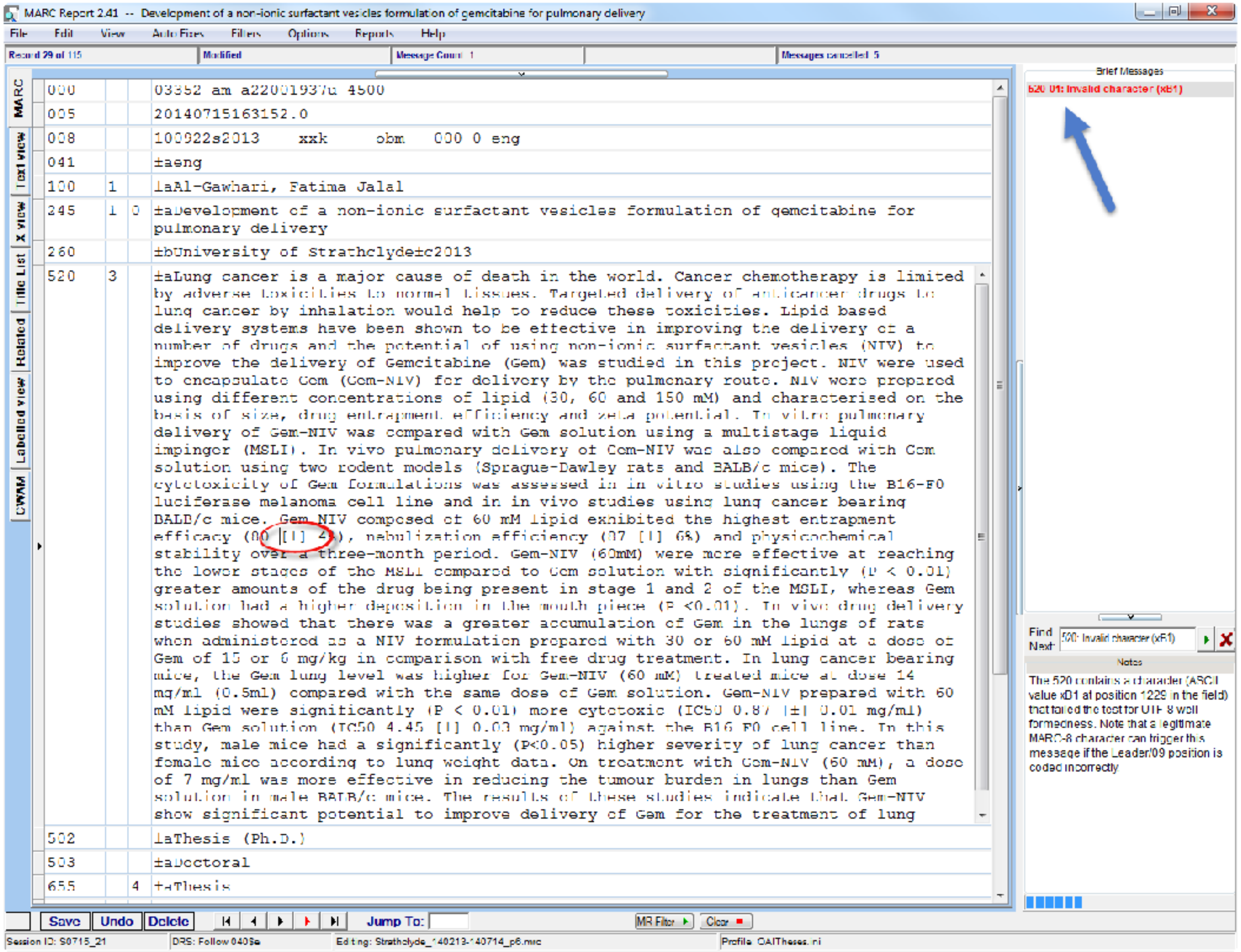
520 01: Invalid character (xDF)

Find Next: 520: Invalid character (xDF)

The 520 contains a character (ASCII value xDF at position 230 in the field) that failed the test for UTF 8 well formedness. Note that a legitimate MARC-8 character can trigger this message if the Leader/09 position is coded incorrectly.

```
Session ID: S0715_21    DRS: Follow 0408a    Editing: Strathclyde_140213-140714_p6.mrc    Profile: OAITheses.ini
```

Unicode errors: Invalid maths symbol



MARC Report 2.41 -- Development of a non-ionic surfactant vesicles formulation of gemcitabine for pulmonary delivery

Record 29 of 115 | Modified | Message Count 1 | Messages cancelled 5

000			03352 am a22001937u 4500
005			20140715163152.0
008			100922s2013 xxk obm 000 0 eng
041			‡aeng
100	1		‡aAl-Gawhari, Fatima Jalal
245	1	0	‡aDevelopment of a non-ionic surfactant vesicles formulation of gemcitabine for pulmonary delivery
260			‡bUniversity of Strathclyde‡c2013
520	3		‡aLung cancer is a major cause of death in the world. Cancer chemotherapy is limited by adverse toxicities to normal tissues. Targeted delivery of anticancer drugs to lung cancer by inhalation would help to reduce these toxicities. Lipid based delivery systems have been shown to be effective in improving the delivery of a number of drugs and the potential of using non-ionic surfactant vesicles (NIV) to improve the delivery of Gemcitabine (Gem) was studied in this project. NIV were used to encapsulate Gem (Gem-NIV) for delivery by the pulmonary route. NIV were prepared using different concentrations of lipid (30, 60 and 150 mM) and characterised on the basis of size, drug entrapment efficiency and zeta potential. In vitro pulmonary delivery of Gem-NIV was compared with Gem solution using a multistage liquid impinger (MSLI). In vivo pulmonary delivery of Gem-NIV was also compared with Gem solution using two rodent models (Sprague-Dawley rats and BALB/c mice). The cytotoxicity of Gem formulations was assessed in in vitro studies using the B16-F0 luciferase melanoma cell line and in in vivo studies using lung cancer bearing BALB/c mice. Gem NIV composed of 60 mM lipid exhibited the highest entrapment efficacy (80 [I] 4%), nebulization efficiency (87 [I] 6%) and physicochemical stability over a three-month period. Gem-NIV (60mM) were more effective at reaching the lower stages of the MSLI compared to Gem solution with significantly (P < 0.01) greater amounts of the drug being present in stage 1 and 2 of the MSLI, whereas Gem solution had a higher deposition in the mouth piece (P <0.01). In vivo drug delivery studies showed that there was a greater accumulation of Gem in the lungs of rats when administered as a NIV formulation prepared with 30 or 60 mM lipid at a dose of Gem of 15 or 6 mg/kg in comparison with free drug treatment. In lung cancer bearing mice, the Gem lung level was higher for Gem-NIV (60 mM) treated mice at dose 14 mg/ml (0.5ml) compared with the same dose of Gem solution. Gem-NIV prepared with 60 mM lipid were significantly (P < 0.01) more cytotoxic (IC50 0.87 [±] 0.01 mg/ml) than Gem solution (IC50 4.45 [I] 0.03 mg/ml) against the B16 F0 cell line. In this study, male mice had a significantly (P<0.05) higher severity of lung cancer than female mice according to lung weight data. On treatment with Gem-NIV (60 mM), a dose of 7 mg/ml was more effective in reducing the tumour burden in lungs than Gem solution in male BALB/c mice. The results of these studies indicate that Gem-NIV show significant potential to improve delivery of Gem for the treatment of lung
502			‡aThesis (Ph.D.)
503			‡aDoctoral
655		4	‡aThesis

Brief Messages

520 01: Invalid character (xB1)

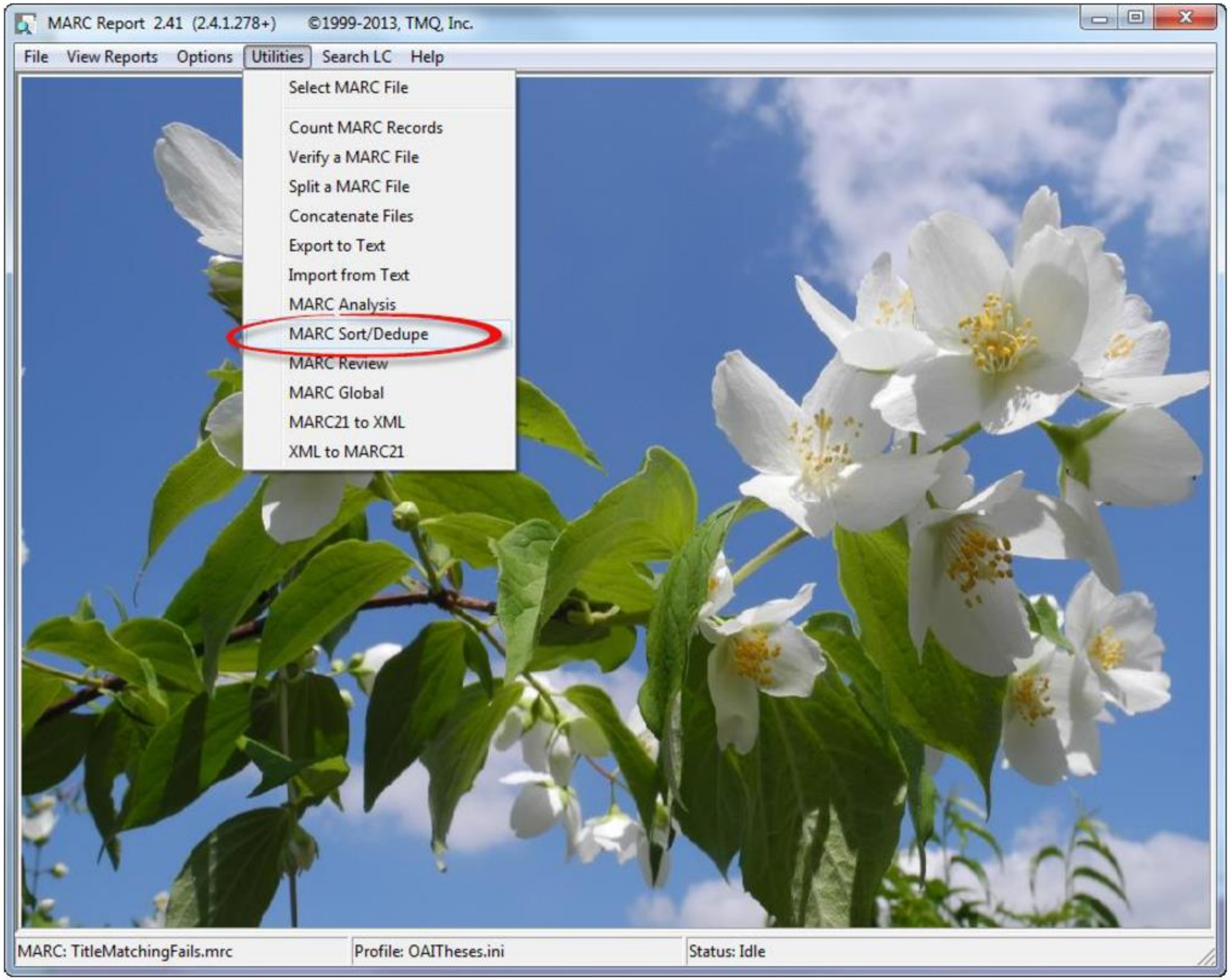
Find Next: 520: Invalid character (xB1)

Notes

The 520 contains a character (ASCII value xB1 at position 1229 in the field) that failed the test for UTF-8 well formedness. Note that a legitimate MARC-8 character can trigger this message if the Leader/09 position is coded incorrectly.

Session ID: S0715_21 | DRS: Follow 040$e | Editing: Strathclyde_140213-140714_p6.mrc | Profile: OAITheses.ini

Deduplication



MARC: TitleMatchingFails.mrc

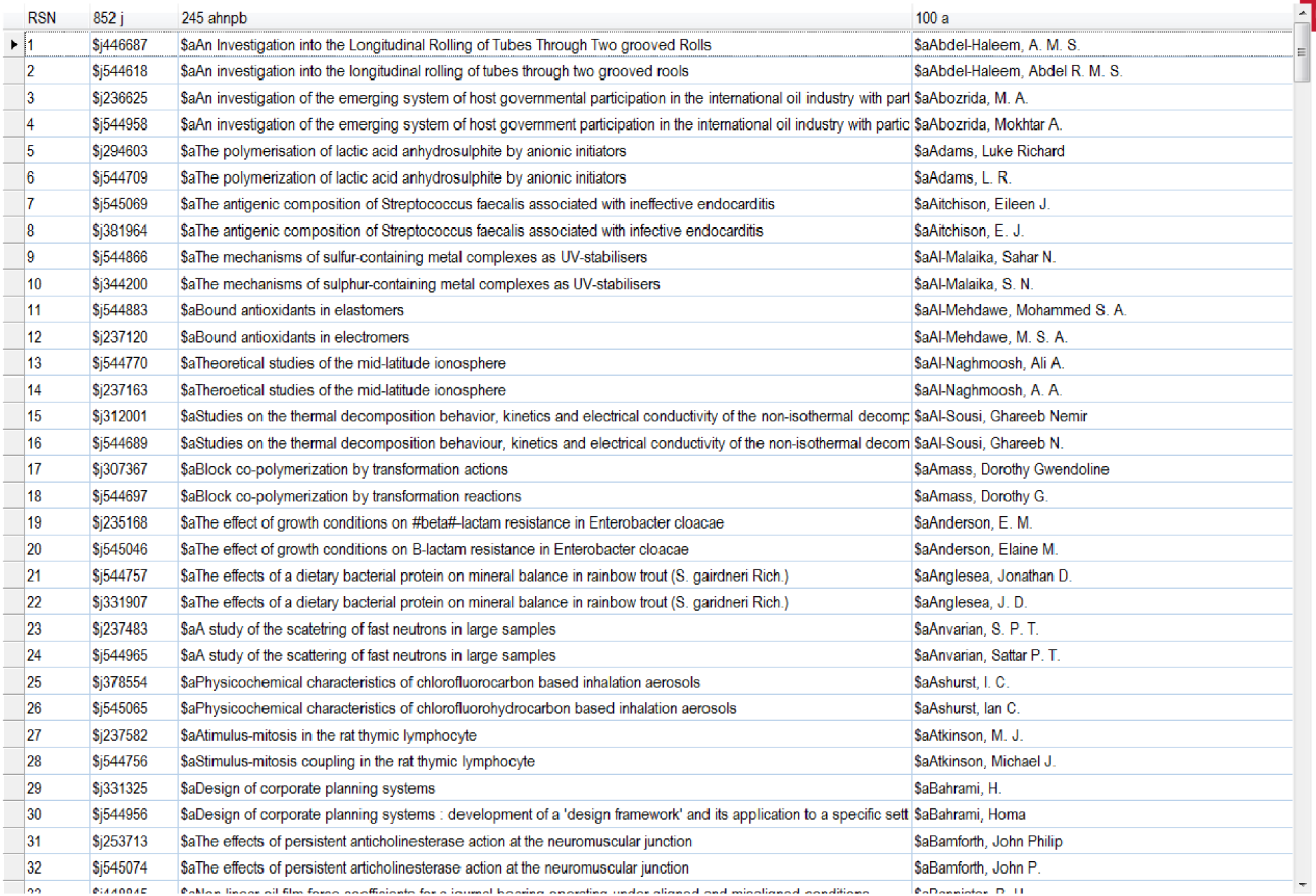
Profile: OAITheses.ini

Status: Idle

Deduplication

- Using MARC Report Sort/deduplication - create a ‘match key’ that excludes punctuation, spacing, and diacritics.
- Author match on surname only (due to differences in recording forenames) and sort by title match key

Author match title list



RSN	852 j	245 ahnpb	100 a
1	$j446687	$aAn Investigation into the Longitudinal Rolling of Tubes Through Two grooved Rolls	$aAbdel-Haleem, A. M. S.
2	$j544618	$aAn investigation into the longitudinal rolling of tubes through two grooved rools	$aAbdel-Haleem, Abdel R. M. S.
3	$j236625	$aAn investigation of the emerging system of host governmental participation in the international oil industry with part	$aAbozrida, M. A.
4	$j544958	$aAn investigation of the emerging system of host government participation in the international oil industry with partic	$aAbozrida, Mokhtar A.
5	$j294603	$aThe polymerisation of lactic acid anhydrosulphite by anionic initiators	$aAdams, Luke Richard
6	$j544709	$aThe polymerization of lactic acid anhydrosulphite by anionic initiators	$aAdams, L. R.
7	$j545069	$aThe antigenic composition of Streptococcus faecalis associated with ineffective endocarditis	$aAitchison, Eileen J.
8	$j381964	$aThe antigenic composition of Streptococcus faecalis associated with infective endocarditis	$aAitchison, E. J.
9	$j544866	$aThe mechanisms of sulfur-containing metal complexes as UV-stabilisers	$aAl-Malaika, Sahar N.
10	$j344200	$aThe mechanisms of sulphur-containing metal complexes as UV-stabilisers	$aAl-Malaika, S. N.
11	$j544883	$aBound antioxidants in elastomers	$aAl-Mehdawe, Mohammed S. A.
12	$j237120	$aBound antioxidants in electromers	$aAl-Mehdawe, M. S. A.
13	$j544770	$aTheoretical studies of the mid-latitude ionosphere	$aAl-Naghmoosh, Ali A.
14	$j237163	$aTheroetical studies of the mid-latitude ionosphere	$aAl-Naghmoosh, A. A.
15	$j312001	$aStudies on the thermal decomposition behavior, kinetics and electrical conductivity of the non-isothermal decomp	$aAl-Sousi, Ghareeb Nemir
16	$j544689	$aStudies on the thermal decomposition behaviour, kinetics and electrical conductivity of the non-isothermal decom	$aAl-Sousi, Ghareeb N.
17	$j307367	$aBlock co-polymerization by transformation actions	$aAmass, Dorothy Gwendoline
18	$j544697	$aBlock co-polymerization by transformation reactions	$aAmass, Dorothy G.
19	$j235168	$aThe effect of growth conditions on #beta#-lactam resistance in Enterobacter cloacae	$aAnderson, E. M.
20	$j545046	$aThe effect of growth conditions on B-lactam resistance in Enterobacter cloacae	$aAnderson, Elaine M.
21	$j544757	$aThe effects of a dietary bacterial protein on mineral balance in rainbow trout (S. gairdneri Rich.)	$aAnglesea, Jonathan D.
22	$j331907	$aThe effects of a dietary bacterial protein on mineral balance in rainbow trout (S. garidneri Rich.)	$aAnglesea, J. D.
23	$j237483	$aA study of the scatetring of fast neutrons in large samples	$aAnvarian, S. P. T.
24	$j544965	$aA study of the scattering of fast neutrons in large samples	$aAnvarian, Sattar P. T.
25	$j378554	$aPhysicochemical characteristics of chlorofluorocarbon based inhalation aerosols	$aAshurst, I. C.
26	$j545065	$aPhysicochemical characteristics of chlorofluorohydrocarbon based inhalation aerosols	$aAshurst, Ian C.
27	$j237582	$aAtimulus-mitosis in the rat thymic lymphocyte	$aAtkinson, M. J.
28	$j544756	$aStimulus-mitosis coupling in the rat thymic lymphocyte	$aAtkinson, Michael J.
29	$j331325	$aDesign of corporate planning systems	$aBahrami, H.
30	$j544956	$aDesign of corporate planning systems : development of a 'design framework' and its application to a specific sett	$aBahrami, Homa
31	$j253713	$aThe effects of persistent anticholinesterase action at the neuromuscular junction	$aBamforth, John Philip
32	$j545074	$aThe effects of persistent articholinesterase action at the neuromuscular junction	$aBamforth, John P.
33	$j448845	$aNon linear oil film force coefficients for a journal bearing operating under aligned and misaligned conditions	$aBannister, R. H.

Top 10 tips for metadata creation

1. Ensure accurate transcription
2. Use controlled vocabularies / standardised data entry
3. Use full names for authors, with correct entry element
4. Use the UTF-8 (Unicode) character set; avoid html markup, especially in the title:

   Preparation of <span style='font-family:Symbol'>h</span>³-allymolybdenum complexes using <i>cis</i>-Mo(CO)<sub>4</sub

5. Use sentence casing not ALL CAPS:

   CodeZebraOS (CODEZEBRAOS)

   GaAs (GAAS)

   MARS (Mars)

Top 10 tips for metadata creation

6. Use correct punctuation and spacing:

   I like cooking my family and my pets

   I like cooking, my family, and my pets

7. Subject classify all theses

8. Always review OCR’d text

9. Use identifiers where possible

10. Use repeated fields for data of the same type (e.g. subject keywords, multiple authors)

Test !

- A
- C
- R
- C

Test !

- **Accurate**
- **C**
- **R**
- **C**

Test !

- Accurate
- Consistent
- R
- C

Test !

- **Accurate**
- **Consistent**
- **Rich**
- **C**

Test !

- **Accurate**
- **Consistent**
- **Rich**
- **Current**

"I liked it better before big data and metadata when we just had good old regular data."

Heather.Rosie@bl.uk

EThOS.bl.uk