Incorrect authors' name (A53)

M. Bossant, NEA Data Bank

Principle

The NEA is developing an in-house code that retrieves citations published on the Internet for Springer (www.springerlink.com), Elsevier (www.sciencedirect.com) and Physical Review Online Archive (prola.aps.org) websites. The code stores a list of DOIs extracted from these citations. It associates these DOIs with EXFOR reference information by using a *citation key* which is the triplet (Journal Code,Volume,Start page).

This code enables to quickly determine whether a REFERENCE keyword of type 'J' is correctly coded provided that the publisher offers a citation search website and that our code implements filling of its citation search form. When the search yields no result, the REFERENCE is suspicious and should be further analysed.

Searching citations

EXFOR Coverage

Table of primary references breakdown by reference type:

Reference type	# in database	Percentage
J	16357	79.2%
C	1629	7.9%
R	1152	5.6%
P	689	3.3%
W	373	1.8%
S	271	1.3%
T	125	0.6%
B	60	0.3%
A	1	0.0%
Total	20657	100.0%

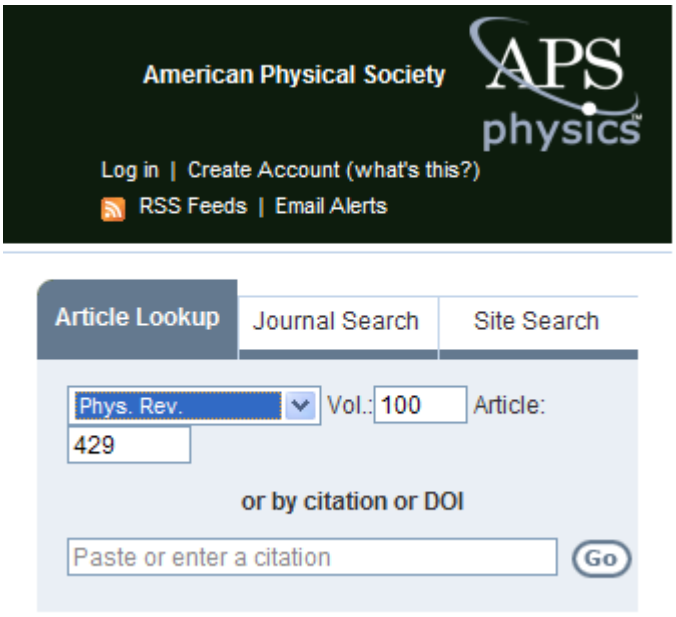
Which journals are covered?

The following table lists the top 10 journal references in the EXFOR database loaded in JANIS. Shaded journals codes are not handled by the code. Note that due to searches returning multiple results (see examples below), there are more citations retrieved from the web than available in EXFOR.

Journal Code	Website	# of citations retrieved from Internet	# of citations in database	# of citations not found
PR/C	prola.aps.org	3007	2985	0
NP/A	www.sciencedirect.com	2627	2318	36
PR	prola.aps.org	1726	1707	38
YF	N/A*	N/A	709	N/A
NSE	www.new.ans.org/pubs Does not use DOI	N/A	650	N/A
NP	www.sciencedirect.com	651	588	1
AE	N/A*	N/A	435	N/A
NIM/B	www.sciencedirect.com	369	366	28
PL/B	www.sciencedirect.com	432	355	0
JIN	www.sciencedirect.com	365	340	0

*: website not found, Springer publishes their English translation (PAN and AE/T journals codes).

Citation search example (prola.aps.org)



Single result found

When the citation key refers to a single article, PROLA displays directly the citation details webpage. This is the common case.
For example, entry C1152.001 is:

```
ENTRY              C1152    20050125                 20050926        0000
SUBENT          C1152001    20050125                 20050926        0000
BIB                    9           9
INSTITUTE   (1USATAM)
REFERENCE   (J,PR/C,61,067307,2000)
AUTHOR      (Y.W.Lui,H.L.Clark,D.H.Youngblood)
TITLE       Giant monopole strength in 58Ni
```

The citation search gives the following page:

Citation Search: Phys. Rev. Lett. Vol. 61 Page/Article 067307 Go

APS » Journals » Phys. Rev. C » Volume 61 » Issue 6

Phys. Rev. C 61, 067307 (2000) [3 pages]

Giant monopole strength in ^{58}Ni

Abstract | References | Citing Articles (7)

Download: PDF (47 kB) Buy this article **Export:** BibTeX or EndNote (RIS)

Y.-W. Lui, H. L. Clark, and D. H. Youngblood
Cyclotron Institute, Texas A&M University, College Station, Texas 77843

Received 3 November 1999; published 19 May 2000

The strength distribution of the giant monopole resonance in ^{58}Ni has been measured from E_x=10 to 35 MeV using sm
$_{0.14}^{+1.69}$MeV.

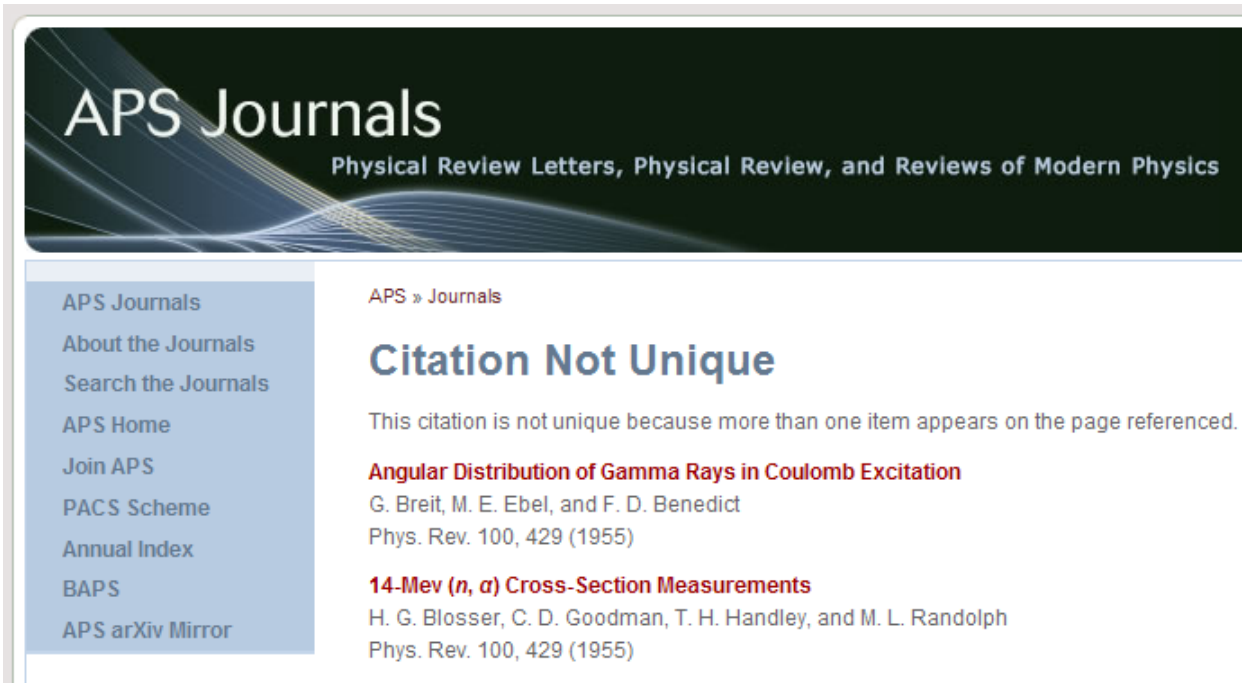
URL: http://link.aps.org/doi/10.1103/PhysRevC.61.067307
DOI: 10.1103/PhysRevC.61.067307
PACS: 24.30.Cz, 25.55.Ci, 27.40.+z, 24.30.Cz, 25.55.Ci, 27.40.+z

Multiple results found

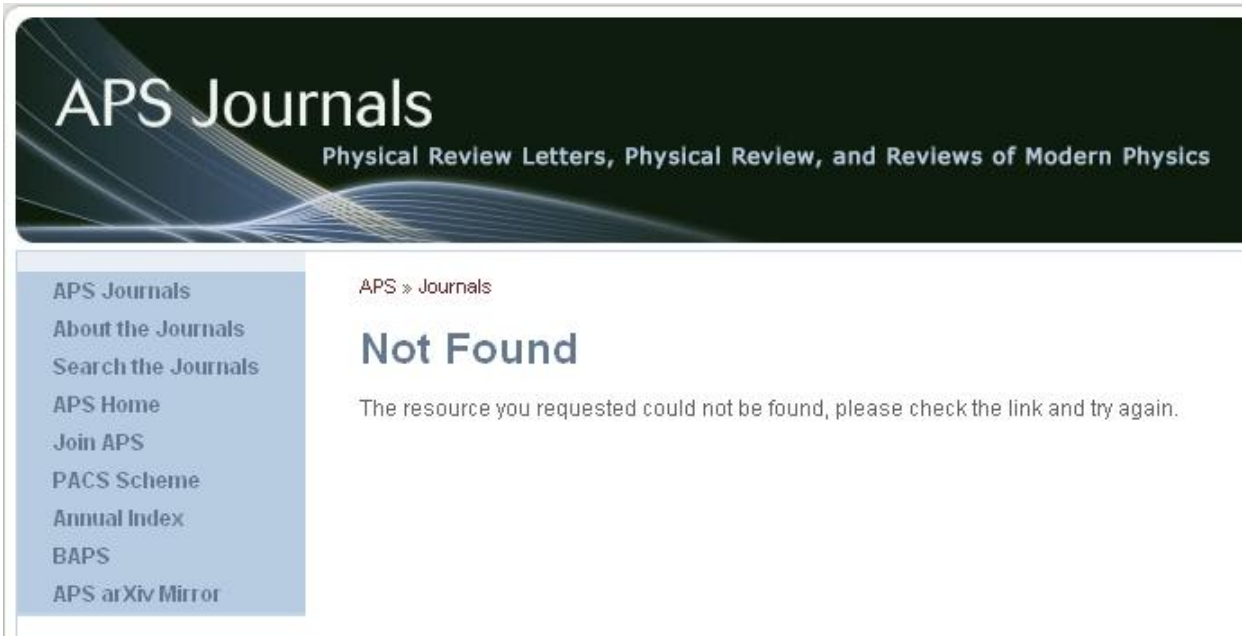
Some articles were published on the same page, searches for this citation key gives multiple results. For example entry 11803.001:

```
ENTRY            11803      860522              20050926        0000
SUBENT        11803001      860522              20050926        0000
BIB                  8          10
INSTITUTE   (1USAORL)
REFERENCE   (J,PR,100,429,55)
AUTHOR      (H.G.BLOSSER,C.D.GOODMAN,T.H.HANDLEY,M.L.RANDOLPH)
TITLE       14-MEV (N,ALPHA) CROSS-SECTION MEASUREMENTS.
```

The search gives the following page:



No result found



When no result could be found, additional investigations are necessary to determine the reason:

- Not available in website?
- Wrong journal code?
- Wrong volume?
- Wrong page?
- Etc.

As journal codes are mostly coded on three characters, a possible error might be introduced by using a journal code for another one, e.g.:

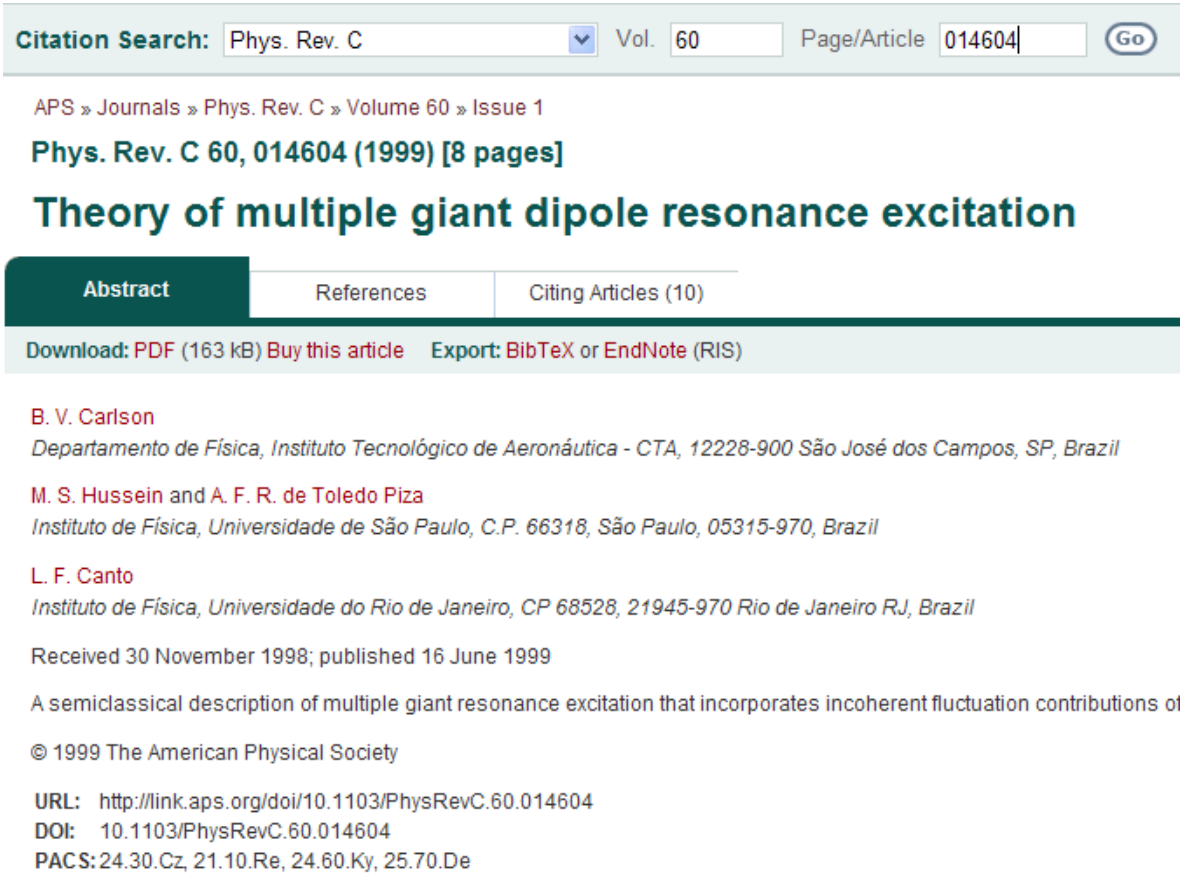
Entry	Keyword	Citation Key	Citation Key (correct)
A0146.003	REL-REF	CJ**P**,41,2194	J,CJ**C**,41,2194
C0498.001	REFERENCE	CJ**P**,47,1667	J,CJ**C**,47,1667
E1986.004	REL-REF	J**M**M,85,467	J,J**N**M,85,467

Comparing citations

Checking existence of a citation is not sufficient, e.g. entry C1156:

```
ENTRY            C1156    20050126                20050926        0000
SUBENT         C1156001   20050126                20050926        0000
BIB                   9         10
INSTITUTE   (1USATAM)
REFERENCE   (J,PR/C,60,014604,1999)
AUTHOR      (D.H.Youngblood,Y.W.Lui,H.L.Clark)
TITLE       Giant monopole strength in 58Ni
```

The search gives the following page:



Citation Search: Phys. Rev. C Vol. 60 Page/Article 014604 Go

APS » Journals » Phys. Rev. C » Volume 60 » Issue 1

Phys. Rev. C 60, 014604 (1999) [8 pages]

Theory of multiple giant dipole resonance excitation

Abstract | References | Citing Articles (10)

Download: PDF (163 kB) Buy this article Export: BibTeX or EndNote (RIS)

B. V. Carlson
Departamento de Física, Instituto Tecnológico de Aeronáutica - CTA, 12228-900 São José dos Campos, SP, Brazil

M. S. Hussein and A. F. R. de Toledo Piza
Instituto de Física, Universidade de São Paulo, C.P. 66318, São Paulo, 05315-970, Brazil

L. F. Canto
Instituto de Física, Universidade do Rio de Janeiro, CP 68528, 21945-970 Rio de Janeiro RJ, Brazil

Received 30 November 1998; published 16 June 1999

A semiclassical description of multiple giant resonance excitation that incorporates incoherent fluctuation contributions of

URL: http://link.aps.org/doi/10.1103/PhysRevC.60.014604
DOI: 10.1103/PhysRevC.60.014604
PACS: 24.30.Cz, 21.10.Re, 24.60.Ky, 25.70.De

Citations published on the Internet contain the following extractable information:

- Journal name
- Volume
- Issue
- Start Page – End page

- Publication date
- Title
- Author(s)
- DOI

This information can be extracted from web pages to be compared against information coming from EXFOR, resulting in a second level of checking.

Provided that there is a single DOI associated with a citation key, our code compares information extracted from web citation against EXFOR content for the following keywords:

- REFERENCE
- REL-REF
- MONIT-REF
- TITLE*
- AUTHOR*

*: applies to *primary* references only.

Results for American Physical Society publications

The following tables detail the results of exercising the code on citations from American Physical Society publications retrieved from the PROLA website for *primary references* only.

The following table lists citation search results for journals published by American Physical Society present in EXFOR (dic #5):

EXFOR Journal code	**# Primary reference**	**# citations not found**
PR	1610	38
PR/A	1	0
PR/B	3	0
PR/C	2976	0
PR/D	11	0
PRL	317	0
RMP	2	0
Total	4920	38

No citation found

The checking program was not able to find 38 citations. These concern old Physical Review articles published before 1955 that do not seem to be available in PROLA website for the time being.

Ambiguous search results

42 primary references could not be checked: there were multiple DOIs for their *citation key*.

Publication year results

For the time being, the code compares publication's year.

Year difference between EXFOR and web citation (absolute value)	**Number of primary references**
0	4860
1	18
2	4
TOTAL	4882

The publication year may change between a preprint version and the published one explaining a majority of errors detected.

First author results

For the time being, the code compares family names. The code applies the following rules on the information retrieved from the web page in order to determine the family name:

- removes junior, senior, I, II, III suffixes
- transcodes all characters not valid in EXFOR to their equivalent (see LEXFOR Author)
- ignores all characters appearing before
  - the rightmost dot (to ignore initials)
  - the rightmost quote (ignore particles like d', O', …)
  - the rightmost dash (for compound names)
  - the rightmost blank character (to ignore particles like van der, de, …)
- removes 'Mc' and 'Mac' prefixes
- changes all remaining letters to lowercase

Then web and EXFOR family names are compared using the Longest Common Subsequence (LCS) algorithm to produce a list of differences. For example, when comparing 'fergerson' with 'fergison' it detects that the two letters 'er' were replaced with an 'i'.

Comparison results for AUTHOR keyword are classified according to the following table:

CODE	Number	Percentage	Explanation	Sample
EXACT	4677	95.8%	Family names matches exactly	
CANNOT_COMPARE	104	2.1%	AUTHOR code could not be parsed by JANIS, e.g.: [AUTHOR] : Blanks are not permitted following initials	22958.001
ONE_LETTER_MISSING	43	0.9%	A single letter is missing in EXFOR author name, e.g.: ' kopa**t**ch' / ' kopach'	23099.001
ONE_LETTER_WRONG	20	0.4%	A single letter is wrong, e.g.: 'g**o**lonsky' / 'g**a**lonsky'	11090.001
ONE_LETTER_MORE	11	0.2%	A single letter has been added to EXFOR author name, e.g: 'sonzogni' / ' sonzo**n**gni'	C0609.001
ONE_LETTER_EXCHANGED	4	0.1%	A single letter has been exchanged, e.g.: 'kr**ie**sler' / kr**ei**sler''	11134.001
OTHER	23	0.5%		See below
TOTAL	4882			

The following tables show detailed analysis of the 23 errors in first author classified as 'OTHER'.

The first table lists 6 errors because the code did not consider the particle 'De' as a special case.

Entry	First Author in EXFOR	First Author from Web citation	Explanation
10166.001	A. De Volpi	A.DeVolpi	Particle 'De' should be ignored from the family name. The checking code should ignore this type of error.
11176.001	J.DE JUREN	James DeJuren	
11177.001	J.DE JUREN	J. DeJuren	
11411.001	J.A.DE JUREN	James A. DeJuren	
11415.001	J.A.DE JUREN	J. A. DeJuren	
L0073.001	L.J.DeBever	L. J. de Bever	

Then, there are 4 references for which the first author was exchanged with the second one:

Entry	Author(s) in EXFOR	Author(s) from Web citation	Explanation
14068.001	J.D.Brandenberger, M.T.McEllistrem, ...	M. T. McEllistrem, J. D. Bradenberger, ...	First and second author exchanged
22445.001	P.V.Sedyshev, H.Beer, ...	H. Beer, P. V. Sedyshev,	
C0677.001	M.L.Barlett, G.W.Hoffmann,...	G. W. Hoffmann, M. L. Barlett, ...	

D0493.001	K.J.Hofstetter, J.D.Stickler	J. D. Stickler, K. J. Hofstetter	

First author code is misspelled in 3 references:

Entry	First Author in EXFOR	First Author from Web citation	Explanation
A0405.001	R.Branch	R. Brandt	Spelling mistake?
C0066.001	R.W.FERGISON	R. W. Fergerson	Spelling mistake?
C1300.001	S.M.Mullinsand	S. M. Mullins	'and' appended to first author family name

Finally, the 10 remaining errors have been detected for various reasons:

Entry	First Author in EXFOR	First Author from Web citation	Explanation
13453.001	B.P.BAYHURST	Los Alamos Radiochemistry Group	Bayhurst was part of the 'Los Alamos Radiochemistry Group', collaboration names were not legal in AUTHOR keyword before 2008.
13942.001	M.Kadi	Minfang Yeh	Following authors are correct
C0597.001	S.Choi	E. J. Stephenson	First author on PROLA website is listed as 3rd author
C1156.001	D.H.Youngblood	B. V. Carlson	Wrong Volume and Page: PR/C,**76**,014303 should be: PR/C,**78**,014303
E2118.001	S.Nakayama	M. Axiotis	Wrong Volume: PR/C,**76**,014303 should be PR/C,**78**,014303
F0047.001	F.E.CECIL	B. L. Berman	Wrong Page: PR/C,27,(1),**1** should be PR/C,27,**6**
M0555.001	H.SCHMIEDEN	F. De Smet	Requires further analysis
O0632.001	H.M.STEINER	Herbert M. Jungerman	Second author seems correct

Entries shaded have wrong reference information coded in EXFOR.

Title results

The following table lists some differences for TITLE keyword found by the code:

Entry	Citation key	Title (1st line from web, 2nd line from EXFOR)	Explanation
10337.001	PR,115,961	Polarization in the D(d,n)3He Reaction and in 12C(n,n)12C Elastic Scattering	Title of the second reference given: J,PR/C,6,1106,197209
		Scattering of 2- to 4-Mev Polarized Neutrons by Carbon	
14068.001	PR/C,9,670	The elastic and inelastic scattering of neutrons from even A Mo isotopes	Title of second reference? First and second author are also exchanged
		Discovery and assignment of excited 0+ levels in even-A Mo isotopes	
22052.001	PR/C,35,1646	Nuclear structure of 208Pb from 207Pb+n resonances	Requires further analysis
		-High Resolution Neutron Resonance Spectroscopy-	
22445.001	PR/C,52,3442	Cross section of 36S(n,γ)37S	Words have been reordered, 37S has been omitted.
		.Measurement of the 36S(N,gamma) cross section	
22831.001	PR/C,70,014607	.Experimental double-differential cross-sections for protons and light charged particle emission in neutron induced reactions at 96 MeV incident neutron energy on three targets: Fe-nat, Pb-nat, et U-nat. - New data at 96 MeV	Title of the second reference? Authors are correct
		Nucleon-induced reactions at intermediate energies: New data at 96 MeV and theoretical status	
22901.001	PR/C,70,044610	Thermal-neutron-induced fission of 243-Cm: light-peak data from the Lohengrin mass separator	Title of second reference (conference proceeding): C,2004SANTA,,605
		Fission-product formation in the thermal-neutron-induced fission of odd Cm isotopes	
22950.001	PR/C,74,054604	.Indirect and direct measurements of thermal neutron acceleration by inelastic scattering on the Lu-177m isomer.	Title of 2nd reference: C,2007NICE,1,463
		Evidence for inelastic neutron acceleration by the 177Lu isomer	
22978.001	PR/C,77,035802	Neutron capture cross section of 90Zr: Bottleneck in the s-process reaction flow	A sentence has been added to the original

		.Neutron capture cross section of 90Zr: Bottleneck in the s-process reaction flow. In addition, measurement of the 90,91,92,93,94,96Zr(n,g), 139La(n,g) cross-sections at n-TOF.	title
23027.001	PR/C,82,015804	.Capture cross section measurements of 186,187,188 -Os at n-TOF: The resolved resonance region .	Title of the 2nd reference: C,2007NICE,1,599
		Neutron physics of the Re/Os clock. III. Resonance analyses and stellar (n,γ) cross sections of 186,187,188Os	
30079.001	PR,152,1055	Measurement of neutron capture cross sections by activation at 24 keV and comparison with statistical theory	Requires further analysis
		Test of Statistical Theory of Nuclear Reactions at 24 keV	
E2118.001	PR/C,76,014303	High-spin gamma-ray spectroscopy in 52Mn	Wrong Volume: PR/C,**76**,014303 should be PR/C,**78**,014303
		Analogs of the giant dipole and spin-dipole resonances in 4He and in alpha clusters of 6,7Li studied by the 4He, 6,7Li(7Li, 7Be gamma) reactions	
M0555.001	PR/C,47,652	16O(γ,p0)15N cross section from 33 to 69 MeV	Requires further analysis
		SCALING OF THE 4HE(G,P) REACTION IN THE EG = 80 - 160 MEV REGION.	
M0680.001	PR/C,72,044004	Simultaneous measurement of the photodisintegration of 4He in the giant dipole resonance region. The title "Photonuclear reactions of light nuclei studied with high-intensity real photon beams" is presented in C,2006CERN,176,2006 (identical data).	A sentence has been added to the original title.
		Simultaneous measurement of the photodisintegration of 4He in the giant dipole resonance region	

Conclusion

This working paper presented the results obtained with the information extracted from PROLA web site. Detection of coding mistakes from journal articles abstracts published on the Internet is possible. However, experience shows that additional analysis is always required to exploit the results and identify actual coding mistakes.