International

Virtual

Observatory

Alliance

Astronomical Keywords in the era of the Virtual Observatory

Version 1.0

IVOA Note 2007 April 25

This version:
AstroKeys-20070425
Latest version:

Previous version(s):

Author(s):
Andrea Preite Martinez,
Soizick Lesteven

Abstract

In the framework of the on-going activities within the Semantics working group of the International Virtual Observatory Alliance (IVOA) in this paper we report on the usage of standard keywords in the field of astronomy and astrophysics. In order to monitor the usage of keywords in the literature we have scanned all the issues of Astronomy and Astrophysics (A&A) and of The Astrophysical Journal (ApJ) in the years 1995, 2000 and 2005 and the issues of MNRAS in the years 2000 and 2005.

Scope of the study is: (i) to compare the lists of keywords that the three journals A&A, ApJ and MNRAS propose to the authors. The three lists (at the date of September 2006) are slightly different and we analyze these differences; (ii) to understand how authors use keywords to synthetically describe their work. To do this we derive the frequency of use of the various keywords, its evolution with time and possible anomalies in the use of the keywords.
As a result of the first part of this study, we define a merged, standard list of keywords and we propose to the editors of the three journals to adopt this standard list of keywords as a very short term common action.
Starting from the analysis of the statistics of usage, we propose to the editors amendments to the standard list of keywords, in order to better suit the needs of the authors.
Finally, we suggest an action to transform the list of keywords into an IVOA standard vocabulary.

Status of This Document

This is a IVOA Note. The first release of this document was 2007 April 25.

This is an IVOA Note expressing suggestions from and opinions of the authors. It is intended to share best practices, possible approaches, or other perspectives on interoperability with the Virtual Observatory. It should not be referenced or otherwise interpreted as a standard specification.

A list of current IVOA Recommendations and other technical documents can be found at http://www.ivoa.net/Documents/.

Acknowledgements

Contents

1 Introduction

During a work on the definition of an astronomical standard vocabulary in progress within the IVOA working group on Semantics, we came across different lists of astronomical concepts and keywords used in different contexts (the old astronomical thesaurus, various flavours of keywords, lists of astronomical object types, etc.).

In particular, an interesting list of astronomical concepts, processes and object types is provided by the editors of astronomical journals to help authors of astronomical papers to class their works: **the astronomical keywords** (we will use in this work the term *astronomical* rather than *astronomy and astrophysics* to avoid confusion with the journal of the same name).

In the 70s some astronomical journals introduced keywords with the articles in order to index their papers. Different subject indexes or keyword lists were established. It was in 1992 that publishers of the main astronomical journals (A&A, ApJ, MNRAS) adopted for the first time a common lists of keywords to allow for consistent indexing and search capabilities across the astronomical literature. Since that time some other journals joined in, as AJ in 1998 and ApJS in 2001. Since then the list of keywords is updated on a regular basis to reflect changes in the literature. With the increasing use of electronic journals and the development of more powerful search engines, this set of keywords keeps a important role in the actual context of exchange standards.

We started our study comparing the lists of keywords proposed by the editors of the major astronomical journals, then we moved on in order to analyze the use of those keywords by the authors, check the frequency of use of the different keywords and its evolution with time, with the aim to eventually detect possible obsolete terms and missing important new term.

We limited our study to papers published in the Main Journal section of Astronomy and Astrophysics (A&A), The Astrophysical Journal (ApJ) and Monthly Notices of the Royal Astronomical Society (MNRAS). In order to test the evolution of usage, we used all papers published in the years 1995, 2000 and

2005 for A&A and ApJ, and in the years 2000 and 2005 for MNRAS (because for this journal the files for the year 1995 were not available at the CDS).
Since many years the scientific editors of astronomical journals have put a great effort in the development, standardization and maintenance of keywords. The aim of this paper is to examine how the present corpus of astronomical keywords could be optimized for an efficient and easy use in the context of the Virtual Observatory.

2 Comparison: is there only one list?

The latest lists of keywords were found at: A&A 445, A5-A6 (2006) or http://aanda.u-strasbg.fr:2002/docinfos/AA/keywords.html for A&A, at http://www.journals.uchicago.edu/ApJ/instruct.key.html for ApJ and at Mon. Not. R. Astron. Soc. 320 (2001) or http://www.blackwellpublishing.com/pdf/mnraskey.pdf for MNRAS.

In all the documents above the editors state that the list presented to the authors is common to the three major Astronomical and Astrophysical Journals (A&A, ApJ and MNRAS).
We checked the three lists and we realized that actually the lists differ in (i) number, (ii) order, (iii) format and (iv) spelling of keywords. All the four kind of differences are of some nuisance if one wants to use keywords e.g. for search/classification purposes.

A detailed list of all the differences found between the merged list of keywords and the individual lists of each journal is presented in Table 1. A summary by kind of difference is presented in Table 2.

2.1 Differences in number

If one merges the three lists (A&A, ApJ and MNRAS) and extracts the list of different keywords - regardless of format and spelling - one gets a total of 325 keywords, while the number of keywords in the individual lists are 322, 318 and 317 for A&A, ApJ and MNRAS, respectively.
More significant are the differences between any two individual lists. The lists of A&A and ApJ differ by 10 keywords, ApJ and MNRAS differ by 5, A&A and MNRAS differ by 11 keywords.

	Merged	A&A	ApJ	MNRAS
keywords	325	322	318	317
missing		3	7	8

2.2 Differences in order

In a few cases new keywords were added to older lists without inserting them in the right alphabetic order. This is not a great problem if the author is searching the appropriate keyword(s) via a search engine, but our experience is that on short lists the search is made preferentially by eye inspection (and, by the way, not necessarily on the list provided by the journal chosen for the publication of the paper). This mixing up of keywords and lists can be misleading for the author. Keywords out of order are also those present in all the three lists, but placed under different sections. Displacements of keywords (two cases) only concern the ApJ list. The key “Celestial mechanics” is not listed under the section “Astrometry and celestial mechanics”, but in the section “Astronomical instrumentation, methods and techniques”.
The other displaced keyword is “gravitational lensing”: it is listed in section 2 “Physical data and processes” by all three journals, but **all ApJ authors** (and all A&A authors publishing in the year 2000) have moved it *de facto* to section 12, modifying the text in “cosmology: gravitational lensing”.

2.3 Differences in format

In the text accompanying the individual lists, editors explain the meaning of the different format elements present in their lists: words in italics should be omitted, do not include any part of a heading that is in parentheses, or both (omit if both in italics and in parentheses). Abbreviated forms of longer text is also in parentheses, as HR for Hertzsprung-Russell, but no indication of usage is given to the authors.

2.4 Differences in text or spelling

There are small differences in spelling, concerning the use of hyphenation and the spelling of “delta Scuti”.
The main difference in text is the lack of reference to the Colour-Magnitude diagram in the keyword “*Stars:* HR and C-M diagrams”, in both ApJ and MNRAS lists.

2.5 Table 1

In Table 1 we show the differences between the merged (A&A, ApJ and MNRAS) list of keywords and individual lists. The total number of keywords found in each section is also given.

	A&A	ApJ	MN
1. General	**9**	**8**	**6**
Publications, bibliography	+	-	-
Sociology of Astronomy	+	+	-
Standards	+	+	-
2. Physical data and processes	**41**	**40**	**40**
Astrobiology	nao	+	+
Astrochemistry	nao	+	+
Chaos	+	-	-
Gravitational lensing	+	d.12	+
Magnetohydrodynamics MHD	df	+	+
3. Astronomical instrumentation, methods and techniques	**32**	**30**	**31**
Light pollution	+	-	+
Methods: observational	+	-	+
Space vehicles	+	nao	-
4. Astronomical data bases	**4**	**4**	**4**
5. Astrometry and celestial mechanics	**7**	**7**	**7**
Celestial mechanics	+	d.3	+
6. The Sun	**29**	**30**	**30**
Sun: atmospheric motions	-	+	+
Sun: X-rays, gamma rays	+	+	ds
7. Solar system	**15**	**15**	**15**
8. Stars	**61**	**59**	**59**
Stars: binaries: symbiotic	+	nao	+
Stars: Hertzsprung-Russell (HR) and C-M diagrams	+	dt	dt
Stars: mass-loss	+	ds	+
Stars: peculiar (*except chemically peculiar*)	+	-	-
Stars: starspots	+	dt	+
Stars: variables: delta Scuti	ds	+	ds
Stars: variables: general	+	-	-
Stars: variables: other	-	+	+
Stars: variables: RR Lyr	+	-	-
Stars: winds, outflows	+	nao	+
9. Interstellar medium (ISM), nebulae	**21**	**22**	**22**
ISM: evolution	df	+	+
ISM: globules	-	+	+
ISM: Herbig-Haro objects	nao	+	nao
ISM: individual objects: ... (*except planetary nebulae*)	+	dt	dt
10. The Galaxy	**18**	**18**	**18**
Galaxy: stellar content	df	+	+
Galaxy: structure	df	+	+
11. Galaxies	**42**	**42**	**42**
Galaxies: halos	+	+	ds
Galaxies: irregular	+	nao	nao
Galaxies: star clusters	+	nao	+

12. Cosmology	**10**	**10**	**10**
13. Sources as a function of wavelength	**33**	**33**	**33**

Notes to Table 1:
+ = present
- = absent
nao=not in the alphabetic order listed, but in the same section
ds= different spelling, same words
dt= different text
df= different format (short form suggested to authors)
d.n= present, but displaced in section n

2.6 Table 2

In Table 2 we show the summary of differences between the merged list of keywords and individual lists. For the kind of differences, see note to Table 1.

	A&A	ApJ	MNRAS
Notes:			
missing(-)	3	7	8
nao	3	5	2
ds	1	1	3
dt	0	3	2
df	4	0	0
d.n	0	2	0

3 Statistics on the use of astronomical keywords

In order to study the actual usage of keywords by the scientific community, we proceeded in the following way. We extracted the keywords section from all the papers published by A&A and ApJ in 1995, 2000 and 2005, and by MNRAS in 2000 and 2005.
The resulting 8 files were handled separately, and number of occurrences were found for each keyword per year and per journal. The results are shown in Table 3 and Table 4. In the first column of Table 3 we list all the different keywords found, separating “standard” keywords (i.e.: keywords present in the lists proposed by the editors) from “non-standard” ones. In columns 2, 3 and 4 we show the total number of occurrences over the periods of reference indicated for A&A, ApJ and MNRAS, respectively. The great difference in total number of keywords used for the three journals is entirely due to the different number of papers published in the years indicated by the three journals. Indeed, the number

of keywords per paper (year-averaged) is practically constant (over time and journals) and is about 4 keywords/paper.

In Table 4 we show the evolution of use of the keywords of different sections along the reference years and across journals. Numbers are percentages of the total number of keywords belonging to the given section that were used in the year indicated. One can clearly see from the results shown: (i) what is the scientific production of the different fields, (ii) the evolution with time (10-year span) of their relative importance and (iii) how the different fields are distributed among journals. We can see that papers in the field of "Stars" are more frequent in A&A than in ApJ or MNRAS, while extra-galactic subjects (galaxies, cosmology) are more frequent in MNRAS.

3.1 Frequencies of use

Apart from the "General" section, all standard keywords are used at least once over the period of time considered, but there are keywords rarely used. Take for instance the number of keywords used ≤ 3 times over the period of reference: they are 13 (A&A), 20 (ApJ) and 60 (MNRAS). One could interpret this saying that there is an increasing degree of specialization going from A&A through ApJ to MNRAS, A&A being the most "generalist" journal, MNRAS being used for a more limited number of subjects.

On the other hand there are keywords that are used very frequently, e.g. more than 100 times over the period of reference: there are 50 of them in A&A, 94 in ApJ and 16 in MNRAS or, respecting the proportion of total number of keywords used in this study (18518:27881:8801, i.e.: 100:150:48): 60 in ApJ (used more than 150 times) and 59 in MNRAS (used more than 48 times).

This can be interpreted either as an indication that the corresponding subjects are very popular, or that there is some sort of "degeneracy" in those concepts, a detail that could be made more explicit, or most probably both.

3.2 Typos

We include in this category: real typos, non-conform use of capital letters and hyphens, errors in punctuation or separation of words, the use of tex formatting characters.

Typos are rare in the ApJ and in MNRAS, while relatively frequent in earlier issues of A&A. The situation in A&A started improving in 2000, and in 2005 the number of typos is similar in all three journals: there are 73 misprinted keywords in MNRAS (1.4% of the number of keywords used in 2005), 162 in the ApJ (1.5%, but mostly due to capitalization errors) and 266 in A&A (3.6%). In this latter case the situation is still improvable. An interesting example is to see one of the keywords with the most complex spelling "(*Stars:*) Hertzsprung-Russell (HR) and (C-M) diagrams" being reported by the 20 papers published in 2005 with 10 different spellings! For the records, only 4 out of these 20 papers got it right.

Furthermore, editors should explicitly discourage the use of tex formatting characters within the keywords section, e.g. “{\em N}-body” (see MNRAS) should be replaced by “N-body”.

3.3 Non-standard keywords

Non-standard keywords are keywords not present (totally or partially) in the list provided by the editor of the journal. Non-standard keywords are listed at the end of each section in Table 3. Their use represent 1.4% of the total number of keywords used in A&A over the three years of reference, 1.5% in the ApJ (reduced to 0.6% if the displacement of “gravitational lensing” is not taken into consideration) and they are practically absent (0.1%) in MNRAS.
We can classify them into three different categories: (i) really new concept, (ii) finer subdivision of an existing concept, and (iii) un-required creation of a keyword due to lack of care by the author(s).
An example of category (iii) is the fusion of two keywords into one, as in “ISM: HII regions: abundances” instead of “ISM: abundances – (ISM:) HII regions”, or in: “Infrared: ISM: lines and bands” instead of “Infrared: ISM – ISM: lines and bands”
An example of category (ii) can be “supernova remnants: plerions”, and “Asteroseismology” for category (i).

3.4 A bottom-up search for new keywords

Of course a statistical study on keywords actually used is not the most efficient way to collect new concepts or sub-concepts, because most of the authors feel constrained by the list of standard keywords (and this is *per se* a good thing!).
On the other hand we can let statistics on usage of the present list(s) of keywords guide us in the search of new concepts or sub-concepts. The method we adopted was to use the most frequently used keywords (> 100 times in the reference period) to guide us in this search. As we noted in §2.1 above, a high frequency of citation could hide a higher semantic richness than that carried by the keyword.
For a number of these “best cited” keywords we selected the corresponding papers, in most cases only from A&A and the ApJ, and we analysed the text written by the authors for the Abstract section of their paper. We selected only the Abstract for two obvious reasons: the conciseness of the concepts imposed to authors and the smaller size, simpler to handle. We then extracted all the sentences containing the keyword under study. From those sentences we selected in a very subjective way, depending on the scientific “sensitivity” of the examiner (APM), groups of words containing the given keyword and at the same time indicating a concept more specific than the original keyword (the zero-order concept). In Table 5 we list some of those “best cited” keywords, the number of sentences examined for each of them and the number of sub-concepts found. As

an illustrative example of the kind of sub-concepts we found, in Table 6 we show for 4 keywords (zero-order concepts) the list of sub-concepts at the first-order of detail. This means that in most cases the authors go further in the specification of the concept. If we take for instance the case of the concept “accretion” we find 57 different second-order specifications of the first-order sub-concept “accretion disk”, and 26 for “accretion flow”.
There are of course also top-down methods for updating the standard list of keywords. Among these we can count: (i) the definition of a new object class by the community (e.g.: the recent “dwarf planet”), (ii) the forthcoming launch of a space mission or the first light of a new instrument that are expected to open new interesting and hopefully productive fields (as for, e.g.: “Asteroseismology” or “Exoplanets”) or (iii) synchronization with other lists used by the scientific community, as the Simbad’s object types or the list of object types and processes for astronomical events used by the VO-Event IVOA community (see the IVOA working group web page http://www.ivoa.net/twiki/bin/view/IVOA/IvoaVOEvent).
We think that the updating of the standard list of astronomical keywords is a process that needs a continuous and organized effort, bearing in mind the importance of the issues of interoperability between the publishing journals, seen as the providers of the particular resource “papers”, and all data centers and data providers already integrated in the Virtual Observatory.
Nonetheless, to help editors in their present updating effort and authors to improve the tagging of their papers, we suggest in Table 7 a number of possible new entries that could be added to the standard list of keywords.

4 Suggestions

From the analysis performed in this study and from our experience in the framework of the Virtual Observatory and of interoperability techniques, we can derive the following conclusions and suggestions for the editors of the main astronomical journals:

- First of all, we are grateful to the scientific editors of the journals for the delicate and time consuming task of defining and maintaining a list of keywords.
- Tagging papers with keywords is useful, we should keep the system and improve it;
- There should be one and only one list of keywords, available on the web, maintained by a single organization under mandate of the editors;
- The list of keywords should evolve rapidly, not lagging too much behind the needs of the scientific community;
- A new, more efficient updating process needs to be organized;

- It is time for bibliography to be inserted in the Virtual Observatory as a vital resource. The presence of standard keywords is essential;
- As a consequence, keywords should become one of the interoperability standards of the IVOA and of the IAU.

Appendix A: Table 3

Table 3. Statistics on the use of astronomical keywords in the years 1995, 2000 and 2005. The use of non-standard keywords is also reported at the end of each section.

Journal:	A&A	ApJ	MNRAS
years:	95-00-05	95-00-05	00-05
Total no. of keywords used:	**18518**	**27881**	**8801**
Section / keyword			
1. General			
Editorials, notices	1	1	
Errata, addenda	18	53	17
Extraterrestrial intelligence		1	
History and philosophy of astronomy	1	5	
Miscellaneous			
Obituaries, biographies		3	
Publications, bibliography	1		
Sociology of Astronomy			
Standards	2		
Non-standard			
Book reviews	21		
Videotapes		11	
2. Physical data and processes			
Acceleration of particles	77	139	17
Accretion, accretion disks	242	493	180
Astrobiology	5	19	
Astrochemistry	29	53	6
Atomic data	66	59	22
Atomic processes	57	78	16
Black hole physics	60	269	86
Chaos	20	3	4
Conduction	3	9	
Convection	48	74	18
Dense matter	31	30	7
Diffusion	26	53	5
Elementary particles	13	58	4
Equation of state	26	17	5
Gravitation	33	65	34
Gravitational lensing	68		85
Gravitational waves	11	24	12
Hydrodynamics	174	319	75

Instabilities	**70**	**150**	**41**
Line: formation	**99**	**85**	**18**
Line: identification	**35**	**53**	**10**
Line: profiles	**135**	**95**	**33**
Magnetic fields	**41**	**105**	**50**
Magnetohydrodynamics MHD	**177**	**310**	**71**
Masers	**53**	**75**	**22**
Molecular data	**27**	**58**	**11**
Molecular processes	**56**	**132**	**29**
Neutrinos	**5**	**21**	**5**
Nuclear reactions, nucleosynthesis, abundances	**61**	**158**	**9**
Plasmas	**77**	**127**	**13**
Polarization	**107**	**160**	**50**
Radiation mechanisms: general	**11**	**10**	**3**
Radiation mechanisms: non-thermal	**88**	**234**	**49**
Radiation mechanisms: thermal	**19**	**44**	**7**
Radiative transfer	**124**	**167**	**39**
Relativity	**34**	**112**	**25**
Scattering	**47**	**51**	**18**
Shock waves	**97**	**188**	**27**
Stellar dynamics	**39**	**47**	**45**
Turbulence	**82**	**157**	**22**
Waves	**54**	**78**	**20**
Non-standard			
Asteroseismology	**1**		**1**
disk evolution	**1**		
Accretion disks: stationary	**1**		
Accretion induced collapse	**1**		
boundary layers	**2**		
disks, shear flow, expansion opacity	**1**		
electron density	**1**		
dynamos	**2**		
heat transfer	**1**		
Line: broadening	**1**		
Line: intensities	**1**		
microlensing	**2**		
Recombination di-electronic	**1**		
Relativistic hydrodynamics	**1**		
superluminal motions	**1**		
tidal disruption	**1**		
Hydromagnetics		**1**	
Shock heating		**1**	
Contribution function	**1**		
3. Astronomical instrumentation, methods and techniques			
Atmospheric effects	**7**	**18**	**8**
Balloons	**1**	**7**	
Instrumentation: adaptive optics	**9**	**15**	**8**
Instrumentation: detectors	**5**	**12**	**7**
Instrumentation: high angular resolution	**6**	**7**	**2**
Instrumentation: interferometers	**10**	**14**	**3**
Instrumentation: miscellaneous	**1**	**4**	**12**
Instrumentation: photometers	**5**	**1**	**3**
Instrumentation: polarimeters	**6**	**1**	**1**
Instrumentation: spectrographs	**15**	**14**	**6**
Light pollution			**1**

Methods: analytical	**46**	**37**	**42**
Methods: data analysis	**122**	**82**	**75**
Methods: laboratory	**26**	**71**	**8**
Methods: miscellaneous	**5**	**1**	**3**
Methods: N-body simulations	**14**	**54**	**41**
Methods: numerical	**135**	**234**	**136**
Methods: observational	**24**	**7**	**13**
Methods: statistical	**56**	**83**	**68**
Site testing	**3**	**2**	**5**
Space vehicles	**6**	**9**	**1**
Space vehicles: instruments	**5**	**5**	**1**
Techniques: high angular resolution	**15**	**33**	**2**
Techniques: image processing	**22**	**26**	**16**
Techniques: interferometric	**56**	**63**	**16**
Techniques: miscellaneous	**7**	**1**	**8**
Techniques: photometric	**71**	**22**	**30**
Techniques: polarimetric	**24**	**15**	**9**
Techniques: radar astronomy			**2**
Techniques: radial velocities	**30**	**17**	**5**
Techniques: spectroscopic	**70**	**73**	**24**
Telescopes	**9**	**14**	**8**
Non-standard			
artificial satellites, space probes	**4**		
Telescopes: liquid mirrors	**2**		
synthetic spectra	**1**		
Hipparcos	**1**		
Hubble Space Telescope	**1**		
ISO	**1**		
Metods: binaries: eclipsing			**1**
4. Astronomical data bases			
Astronomical data bases: miscellaneous	**12**	**1**	**6**
Atlases	**2**	**2**	
Catalogs	**46**	**12**	**19**
Surveys	**71**	**107**	**85**
5. Astrometry and celestial mechanics			
Astrometry	**103**	**47**	**17**
Celestial mechanics	**59**	**64**	**38**
Eclipses	**7**	**2**	**3**
Ephemerides	**12**		
Occultations	**8**	**3**	**1**
Reference systems	**33**	**1**	**1**
Time	**7**		**1**
6. The Sun			
Sun: abundances	**20**	**39**	**4**
Sun: activity	**71**	**118**	**11**
Sun: atmosphere	**50**	**37**	
Sun: atmospheric motions		**10**	
Sun: chromosphere	**58**	**50**	
Sun: corona	**154**	**250**	**3**
Sun: coronal mass ejections (CMEs)	**12**	**44**	**1**
Sun: evolution	**4**	**10**	**2**
Sun: faculae, plages	**18**	**4**	
Sun: filaments	**13**	**20**	

Sun: flares	**93**	**181**	**6**
Sun: fundamental parameters	**5**	**3**	**3**
Sun: general	**8**	**8**	**2**
Sun: granulation	**32**	**14**	**3**
Sun: helioseismology	**7**	**29**	**6**
Sun: infrared	**8**	**10**	
Sun: interior	**20**	**65**	**8**
Sun: magnetic fields	**166**	**239**	**12**
Sun: oscillations	**77**	**84**	**13**
Sun: particle emission	**15**	**48**	**1**
Sun: photosphere	**70**	**32**	**3**
Sun: prominences	**17**	**29**	**1**
Sun: radio radiation	**43**	**47**	**2**
Sun: rotation	**13**	**28**	**3**
Sun: solar-terrestrial relations	**3**	**12**	**1**
Sun: solar wind	**47**	**79**	**4**
Sun: sunspots	**47**	**55**	**10**
Sun: transition region	**37**	**30**	**1**
Sun: UV radiation	**60**	**69**	**3**
Sun: X-rays, gamma rays	**51**	**88**	
Non-standard			
Sun: doppler shifts	**1**		
heliosphere	**1**		
Sun: magnetohydrodynamics	**2**		
Sun: polarization	**1**		
spicules	**1**		
7. Solar system			
Comets: general	**51**	**36**	**12**
Comets: individual	**39**	**25**	**8**
Earth	**8**	**6**	**3**
Interplanetary medium	**30**	**47**	**2**
Kuiper Belt	**5**	**18**	**3**
Meteors, meteoroids	**21**	**19**	**12**
Minor planets, asteroids	**56**	**14**	**15**
Moon	**2**	**4**	**1**
Oort Cloud	**3**	**5**	**2**
Planets: rings	**1**	**3**	**1**
Planets and satellites: formation	**29**	**10**	**6**
Planets and satellites: general	**38**	**29**	**8**
Planets and satellites: individual	**51**	**49**	**12**
Solar system: formation	**16**	**66**	**8**
Solar system: general	**21**	**13**	**8**
Non-standard			
Comets: abundances	**2**		
Comets: tensile strength	**1**		
TNOs (Trans Neptunian Objects)	**1**		
8. Stars			
Stars: abundances	**240**	**143**	**46**
Stars: activity	**111**	**63**	**27**
Stars: AGB and post-AGB	**158**	**102**	**33**
Stars: atmospheres	**149**	**96**	**27**
Stars: binaries (*including multiple*): close	**175**	**217**	**107**
Stars: binaries: eclipsing	**79**	**70**	**34**
Stars: binaries: general	**89**	**88**	**65**

Stars: binaries: spectroscopic	**87**	**33**	**30**
Stars: binaries: symbiotic	**36**	**13**	**5**
Stars: binaries: visual	**37**	**25**	**5**
Stars: blue stragglers	**6**	**15**	**2**
Stars: carbon	**42**	**30**	**5**
Stars: chemically peculiar	**90**	**14**	**20**
Stars: chromospheres	**26**	**31**	**5**
Stars: circumstellar matter	**229**	**296**	**61**
Stars: coronae	**48**	**53**	**5**
Stars: distances	**45**	**26**	**9**
Stars: dwarf novae	**5**	**5**	**9**
Stars: early-type	**132**	**101**	**35**
Stars: emission-line, Be	**68**	**29**	**22**
Stars: evolution	**226**	**195**	**58**
Stars: flare	**38**	**35**	**4**
Stars: formation	**221**	**389**	**102**
Stars: fundamental parameters	**186**	**86**	**34**
Stars: general	**21**	**6**	**13**
Stars: Hertzsprung-Russell (HR) and C-M diagrams	**66**	**15**	**5**
Stars: horizontal-branch	**31**	**27**	**5**
Stars: imaging	**18**	**11**	**8**
Stars: individual	**773**	**906**	**270**
Stars: interiors	**63**	**104**	**18**
Stars: kinematics	**39**	**28**	**14**
Stars: late-type	**156**	**112**	**39**
Stars: low-mass, brown dwarfs	**72**	**110**	**27**
Stars: luminosity function, mass function	**36**	**53**	**15**
Stars: magnetic fields	**121**	**117**	**41**
Stars: mass-loss	**154**	**161**	**30**
Stars: neutron	**148**	**322**	**70**
Stars: novae, cataclysmic variables	**106**	**129**	**69**
Stars: oscillations (*including pulsations*)	**148**	**106**	**64**
Stars: peculiar (*except chemically peculiar*)	**5**	**8**	
Stars: planetary systems	**50**	**158**	**27**
Stars: planetary systems: formation	**10**	**44**	**6**
Stars: planetary systems: protoplanetary disks	**9**	**50**	**9**
Stars: Population II	**48**	**33**	**10**
Stars: pre-main sequence	**151**	**215**	**35**
Stars: pulsars: general	**65**	**168**	**62**
Stars: pulsars: individual	**66**	**215**	**42**
Stars: rotation	**98**	**114**	**45**
Stars: starspots	**21**	**7**	**3**
Stars: statistics	**20**	**31**	**5**
Stars: subdwarfs	**26**	**16**	**7**
Stars: supergiants	**25**	**25**	**3**
Stars: supernovae: general	**80**	**160**	**41**
Stars: supernovae: individual	**37**	**124**	**22**
Stars: variables: Cepheids	**26**	**31**	**9**
Stars: variables: delta Scuti	**25**	**6**	**8**
Stars: variables: general	**87**		**2**
Stars: variables: other	**2**	**44**	**69**
Stars: variables: RR Lyr	**10**	**3**	
Stars: white dwarfs	**80**	**146**	**49**
Stars: winds, outflows	**56**	**69**	**20**
Stars: Wolf-Rayet	**65**	**50**	**18**
Non-standard			

extra-solar planets	**1**		
black holes	**5**		
Stars: accretion disks	**1**		
Stars: bynaries: pulsars	**2**		
Stars: bynaries: low-mass X-Ray	**1**		
Stars: faint blue	**1**		
Stars: giant	**2**	**1**	
Stars: Herbig Ae/Be	**1**		
Stars: polarization	**1**		
Stars: variables: Other: long-period		**6**	
Stars: variables: Other: luminous blue		**5**	
Stars: OH/IR	**2**		
Stars: variables: Other: beta Cep	**1**		
Stars: variables: Other: AM Her		**1**	
Stars: variables: Other: FU Ori		**1**	
Stars: variables: Other: GW Vir		**1**	
Stars: variables: Other: RS CVn	**1**		
Stars: variables: Other: RV Tau		**1**	
Stars: pulsar wind nebulae	**1**		
Stars: structure			**1**
Stars: X-Ray		**2**	
Stars: substellar companions	**1**		
9. Interstellar medium (ISM), nebulae			
ISM: abundances	**90**	**170**	**23**
ISM: atoms	**25**	**33**	**4**
ISM: bubbles	**35**	**34**	**6**
ISM: clouds	**161**	**265**	**59**
ISM: cosmic rays	**51**	**119**	**15**
ISM: dust, extinction	**179**	**250**	**62**
ISM: evolution	**7**	**12**	**4**
ISM: general	**53**	**59**	**17**
ISM: globules		**10**	**4**
ISM: Herbig-Haro objects	**12**	**9**	**1**
ISM: HII regions	**92**	**151**	**43**
ISM: individual objects: (*except planetary nebulae*)	**201**	**412**	**51**
ISM: jets and outflows	**108**	**162**	**67**
ISM: kinematics and dynamics	**59**	**107**	**20**
ISM: lines and bands	**21**	**33**	**6**
ISM: magnetic fields	**29**	**77**	**10**
ISM: molecules	**196**	**359**	**65**
ISM: planetary nebulae: general	**50**	**52**	**24**
ISM: planetary nebulae: individual	**58**	**54**	**22**
ISM: reflection nebulae	**14**	**18**	**5**
ISM: structure	**50**	**87**	**12**
ISM: supernova remnants	**71**	**164**	**4**
Non-standard			
ISM: abundances: depletion	**1**		
ISM: cosmic ray electrons	**1**		
diffuse medium	**1**		
gas dynamics	**1**		
sublimation	**1**		
ISM: HI gas	**1**		
ISM: HII regions: abundances	**1**		
ISM: molecules: PAH	**2**		
ISM: photon escape	**1**		

ISM: supernova remnants: plerions	**1**		
ISM: theory			**1**
10. The Galaxy			
Galaxy: abundances	**52**	**32**	**5**
Galaxy: bulge	**6**	**7**	**7**
Galaxy: center	**30**	**109**	**15**
Galaxy: disk	**18**	**3**	**4**
Galaxy: evolution	**48**	**28**	**4**
Galaxy: formation	**12**	**3**	**12**
Galaxy: fundamental parameters	**5**	**3**	**2**
Galaxy: general	**17**	**13**	**1**
Galaxy: globular clusters: general	**55**	**84**	**26**
Galaxy: globular clusters: individual	**62**	**87**	**19**
Galaxy: halo	**57**	**74**	**16**
Galaxy: kinematics and dynamics	**36**	**24**	**16**
Galaxy: nucleus		**5**	
Galaxy: open clusters and associations: general	**50**	**24**	**16**
Galaxy: open clusters and associations: individual	**109**	**58**	**49**
Galaxy: solar neighbourhood	**35**	**9**	**8**
Galaxy: stellar content	**38**	**32**	**9**
Galaxy: structure	**49**	**64**	**28**
Non-standard			
Galaxy: thick disk	**1**		
diffuse galactic gamma rays	**1**		
11. Galaxies			
Galaxies: abundances	**66**	**82**	**28**
Galaxies: active	**246**	**448**	**224**
Galaxies: BL Lacertae objects: general	**30**	**35**	**9**
Galaxies: BL Lacertae objects: individual	**35**	**71**	**11**
Galaxies: bulges	**3**	**4**	**5**
Galaxies: clusters: general	**131**	**280**	**170**
Galaxies: clusters: individual	**78**	**176**	**58**
Galaxies: cooling flows	**12**	**69**	**23**
Galaxies: distances and redshifts	**54**	**121**	**27**
Galaxies: dwarf	**40**	**52**	**20**
Galaxies: elliptical and lenticular, cD	**58**	**136**	**60**
Galaxies: evolution	**159**	**375**	**181**
Galaxies: formation	**66**	**273**	**148**
Galaxies: fundamental parameters	**45**	**98**	**53**
Galaxies: general	**43**	**25**	**29**
Galaxies: halos	**22**	**100**	**67**
Galaxies: high-redshift	**26**	**96**	**30**
Galaxies: individual	**347**	**669**	**195**
Galaxies: interactions	**75**	**103**	**56**
Galaxies: intergalactic medium	**41**	**192**	**77**
Galaxies: irregular	**20**	**21**	**7**
Galaxies: ISM	**160**	**218**	**45**
Galaxies: jets	**80**	**132**	**42**
Galaxies: kinematics and dynamics	**102**	**178**	**104**
Galaxies: Local Group	**25**	**30**	**12**
Galaxies: luminosity function, mass function	**19**	**42**	**25**
Galaxies: Magellanic Clouds	**85**	**79**	**26**
Galaxies: magnetic fields	**54**	**14**	**3**
Galaxies: nuclei	**108**	**249**	**69**

Galaxies: peculiar	**7**	**8**	
Galaxies: photometry	**75**	**101**	**38**
Galaxies: quasars: absorption lines	**46**	**170**	**41**
Galaxies: quasars: emission lines	**15**	**43**	**16**
Galaxies: quasars: general	**113**	**166**	**94**
Galaxies: quasars: individual	**105**	**140**	**36**
Galaxies: Seyfert	**91**	**193**	**61**
Galaxies: spiral	**89**	**96**	**55**
Galaxies: starburst	**109**	**189**	**84**
Galaxies: star clusters	**51**	**72**	**26**
Galaxies: statistics	**22**	**45**	**37**
Galaxies: stellar content	**59**	**115**	**53**
Galaxies: structure	**66**	**161**	**65**
Non-standard			
Galaxies: active galactic nuclei (AGN)	**1**		
Galaxies: associaton	**1**		
Galaxies: clustering	**5**	**12**	**1**
Galaxies: compact	**10**	**11**	**3**
Galaxies: dark halo population	**1**		
Galaxies: lenticular	**1**		
Galaxies: quasars: redshift	**2**		
Galaxies: quasars: UV radiation	**1**		
Galaxies: spheroidal	**1**		
Galaxies: star formation			**1**
Galaxies: theory	**1**		
12. Cosmology			
Cosmology: cosmic microwave background	**39**	**147**	**72**
Cosmology: cosmological parameters	**17**	**27**	**22**
Cosmology: miscellaneous	**21**	**18**	**20**
Cosmology: observations	**124**	**303**	**102**
Cosmology: theory	**60**	**387**	**188**
Cosmology: dark matter	**95**	**220**	**131**
Cosmology: diffuse radiation	**23**	**64**	**20**
Cosmology: distance scale	**16**	**85**	**10**
Cosmology: early Universe	**19**	**64**	**21**
Cosmology: large-scale structure of Universe	**100**	**298**	**191**
Non-standard			
Cosmology: cosmic microwave background: anisotropies		**1**	
Cosmology: gravitational lensing	**40**	**250**	
Cosmic strings		**6**	
Cosmic background radiation		**1**	
Cosmology: black holes			**1**
13. Sources as a function of wavelength			
Gamma rays: bursts	**82**	**310**	**46**
Gamma rays: observations	**87**	**129**	**11**
Gamma rays: theory	**35**	**99**	**14**
Infrared: galaxies	**95**	**167**	**76**
Infrared: general	**19**	**21**	**5**
Infrared: ISM	**29**	**37**	**22**
Infrared: solar system	**14**	**19**	**1**
Infrared: stars	**113**	**164**	**44**
Radio continuum: galaxies	**70**	**117**	**48**
Radio continuum: general	**18**	**28**	**10**
Radio continuum: ISM	**16**	**42**	**9**

Radio continuum: solar system	**4**	**1**	
Radio continuum: stars	**46**	**67**	**18**
Radio lines: galaxies	**93**	**46**	**15**
Radio lines: general	**6**	**7**	**4**
Radio lines: ISM	**105**	**129**	**24**
Radio lines: solar system	**9**	**3**	
Radio lines: stars	**19**	**17**	**1**
Submillimeter	**25**	**46**	**22**
Ultraviolet: galaxies	**20**	**113**	**11**
Ultraviolet: general	**6**	**32**	**5**
Ultraviolet: ISM	**10**	**73**	**2**
Ultraviolet: solar system	**2**	**8**	
Ultraviolet: stars	**41**	**135**	**14**
X-rays: binaries	**83**	**93**	**56**
X-rays: bursts	**20**	**48**	**4**
X-rays: diffuse background	**5**	**11**	**7**
X-rays: galaxies	**138**	**336**	**134**
X-rays: galaxies: clusters	**22**	**43**	**22**
X-rays: general	**70**	**98**	**36**
X-rays: individuals	**39**	**49**	**21**
X-rays: ISM	**26**	**86**	**6**
X-rays: stars	**215**	**346**	**76**
Non-standard			
gamma rays: general	**3**	**1**	
gamma rays: individuals	**1**		
gamma rays: lines	**1**		
Infrared: ISM: continuum	**27**	**32**	
Infrared: ISM: lines and bands	**46**	**57**	
Infrared: lines	**1**		
millimeter	**1**		
Radio lines: atomic	**3**		
Radio lines: molecular	**4**		
Radio sources: general	**1**	**1**	
Radio sources: lines		**1**	
Ultraviolet: quasars	**1**		
Ultraviolet: spectra		**3**	
X-Rays: comets	**1**		
X-Rays: novae	**1**		
X-Rays: quasars	**2**		
X-Rays: transients	**2**		
X-Rays: supersoft sources	**1**		

Appendix B: Table 4

Table 4. Distribution of the use of astronomical keywords in the journals A&A and ApJ in the years 1995, 2000 and 2005 and MNRAS in 2000 and 2005 (all values are percentages).

	A&A				ApJ				MNRAS		
	1995	**2000**	**2005**		**1995**	**2000**	**2005**		**2000**	**2005**	
				Total				***Total***			***Total***
General	0,00	0,34	0,30	***0,24***	0,57	0,00	0,27	***0,27***	0,11	0,25	***0,19***
Physical data and processes	14,90	13,02	13,67	***13,75***	16,51	16,37	14,53	***15,71***	14,42	12,98	***13,57***
Astronomical instrumentation, methods and techniques	3,31	3,35	6,09	***4,43***	2,48	3,58	3,87	***3,38***	4,96	7,43	***6,42***
Astronomical data bases	0,37	0,56	1,06	***0,71***	0,50	0,30	0,52	***0,44***	1,14	1,32	***1,25***
Astrometry and celestial mechanics	1,74	0,94	1,18	***1,24***	0,60	0,40	0,30	***0,42***	0,89	0,56	***0,69***
The Sun	4,99	6,60	7,65	***6,62***	5,77	6,22	6,53	***6,21***	1,61	0,86	***1,17***
Solar system	2,24	1,50	2,36	***2,03***	0,80	1,04	1,73	***1,23***	1,22	1,09	***1,15***
Stars	31,35	31,44	27,92	***30,01***	23,53	20,79	21,59	***21,87***	22,86	21,31	***21,94***
Interstellar medium (ISM), nebulae	9,27	9,12	7,46	***8,49***	11,31	9,42	8,11	***9,46***	6,43	5,64	***5,97***
The Galaxy	3,59	3,46	3,93	***3,68***	2,45	2,20	2,44	***2,36***	2,28	2,98	***2,69***
Galaxies	15,61	17,75	16,59	***16,75***	18,61	21,20	22,93	***21,13***	27,17	27,63	***27,44***
Cosmology	2,86	3,15	2,94	***2,99***	6,43	7,51	6,20	***6,71***	8,02	9,41	***8,84***
Sources as a function of wavelength	9,76	8,78	8,87	***9,06***	10,44	10,97	10,95	***10,81***	8,88	8,54	***8,68***

Appendix C: Table 5

Table 5. Sentences analyzed and sub-concepts found for a sample of "best cited" keywords.

Keyword	Analysed sentences	Extracted sub-concepts
abundances	2792	45
accretion	2672	130
circumstellar	592	129
galaxies active	854	41
galaxies cluster	4481	120
GRB	2205	37
hydrodynamics	192	8
mass-loss	261	30
neutron stars	1486	52
numerical	1170	76
stars binary close	409	37
stars formation	1769	129
supernova	1969	34

Appendix D: Table 6

Table 6. Examples of sub-concepts extracted (only at first order).

Zero-order concept (keyword)	first-order sub-concepts
accretion	column, disk, efficiency, flow, rate, stream, torque, torus
circumstellar	absorption, accretion, atmosphere, bubble, cavities, clouds, cocoon, debris, density, disk, dust, emission, envelopes, environment, expansion, features, gas, grains, HI, interaction, layers, magnetic field, maser, mass, mass-loss, material, matter, medium, nebula, nebulosity, opacity, outflow, plasma, properties, radius, region, ring, shell, structures, torus, winds
neutron star	atmosphere, bare, binaries, center, core, crust, isolated, magnetosphere, polar caps, surface cold, hot, radio-quiet, transient

	accreting, bursting, coalescing, cooling, glitches, magnetar, magnetized, mergers, millisecond, non-rotating, oscillations, precessing, pulsations, rotating, spin, spinning, superfluid, unmagnetized, wind-accreting evolution, model, new-born, old, progenitor, proto
supernova	bubble, core, debris, ejecta, envelope, shell enrichment, high-redshift, light-curve, radio, spectrum, typeIa, typeIb, typeIc, typeII, typeIIn, typeIIp, typeV blast-wave, core-collapse, ejection, explosion, fallback, kick, shock, yields first-generation, models, Population III, pre-SN, pre-SN wind, progenitor, rate

Appendix E: Table 7

Table 7. Proposed new concepts.

Section	**Proposed new concept**
Physical data and processes	Asteroseismology
Astronomical instrumentation, methods and techniques	Virtual Observatory
Solar system	Dwarf planet
Stars	Exoplanet (*Stars:*) variables: long-period
Interstellar medium (ISM), nebulae	(*ISM:*) clouds: molecular
The Galaxy	MACHO
Galaxies	AGN Blazar compact
Sources as a function of wavelength	Gamma rays: general

References

[1] R. Hanisch, *Resource Metadata for the Virtual Observatory* , http://www.ivoa.net/Documents/latest/RM.html
[2] R. Hanisch, M. Dolensky, M. Leoni, *Document Standards Management: Guidelines and Procedure* , http://www.ivoa.net/Documents/latest/DocStdProc.html
[3] A&A 445, A5-A6 (2006)
[4] Mon. Not. R. Astron. Soc. 320 (2001)