Methods for measuring social and conceptual dimensions of Convergence Science

Alexander Michael Petersen, Felber Arroyave, and Ioannis Pavlidis

SI Appendix

Appropriateness of CIP and MeSH for measuring convergence. The *Classification of Instructional Programs* ontology is developed and maintained by the US National Center for Educational Statistics, with the 2020 version representing its sixth version since its origination in 1980. The objective of this ontology is "to facilitate the organization, collection, and reporting of fields of study and program completions" (National Center for Education Statistics, 2022). Since educational programs tend to be highly aligned with the faculty departments that deliver them, we expect a very suitable mapping of faculty departments onto educational programs.

While it is possible that program omissions exist, given the depth and breadth of this ontology which encompasses more than 2,100 traditional (e.g. the L_3 category "26.0807 Genome Sciences/Genomics") and technical programs (e.g. "49.0104 Aviation/Airway Management and Operations."), we expect these omissions to be negligible, occurring at the margins of research that meets the indexing standards of PubMed and other publication indices. Given that many international higher education institutions are modeled after the organizational structure found in the US and UK, it is likely that international bias is relatively small and that omissions represent inconsequential corner cases. As an example of its broad inclusivity, the CIP ontology includes "51.3301 Acupuncture and Oriental Medicine", along with several other variants within the "51.33 Alternative and Complementary Medicine and Medical Systems" category.

There is also the question of novel program inclusion. Drawing from its regular updates, CIP indeed includes a number of recent developments in the organizational landscape, many of which belong to the L_1 category "30 Multi/Interdisciplinary Studies" (not shown in **Fig. 3** but rather represented by the ellipses). At the L_2 level within the multidisciplinary category are a number of relevant convergence programs, including "30.43 Geobiology", "30.33 Sustainability Studies" and "30.70 Data Science", among others, which together accommodate the increasing variety of hybrid faculty departments.

A similar question of inclusivity can be raised with respect to the conceptual ontology: how disciplinarily comprehensive are the journals indexed by PubMed relative to other large research publication indices? While PubMed was designed for the core domains of biology, biomedical, and health science, its scope has increased over time, responding to the same forces underlying the convergence science paradigm itself, as nearly all engineering, natural, social sciences have some intersection with the core domains. Nevertheless, the breadth of both journals indexed by PubMed and the MeSH keywords used to classify them are widely under-appreciated.

To address the question of PubMed’s disciplinary breadth, we compared the disciplinary classification of all the journals featured in PubMed using their corresponding Web of Science (WOS) journal classifications. WOS classifies journals according to a flat (non-hierarchical) classification system comprised of 256 Category (WC) tags, which are journal-specific, meaning that all articles published by a given journal will be classified by that WC, independent of the actual subject area content of the research. Also note that the vast majority of journals indexed within WOS are classified according to just one WC tag (i.e., 72% have one WC, 21% have two, 5% have three); and a relatively disproportionate number of journals have WC = ``Multidisciplinary Sciences'', corresponding to a significant wildcard or missing data problem – see **Fig. 7**.[1] The results of our comparison indicate that 236 (corresponding to 92%) of the total 256 WOS Subject Categories are spanned by journals indexed by PubMed.[2]

[1] It is worth reiterating that “Multidisciplinary Sciences” WC is used for high-impact journals such as *Nature*, *PNAS* and *Science*, and so even though these journals publish works from all domains of science, all articles published in these journals lack subject area specificity, and so methods that use WC to classify research vastly underestimate the subject area representation of the highest impact research – see **Fig. 7**. This issue has been exacerbated by the disproportional growth of multidisciplinary journals in the last 20 years. Comparing the last two decades, the percentage of articles published from 2000-2009 that are indexed by WOS with WC = “Multidisciplinary Sciences” was 1.4%; in the last decade, this percentage more than doubled to roughly to 3.0% (2010-2019). Articles with wildcard WC distort efforts to measure disciplinary diversity by way of measuring variation and disparity of the articles cited according to their WC. By way of example, analysis of interdisciplinarity based upon 12 bio-nanoscience articles finds that roughly 1 in 4 articles cited by these twelve belonged to the “Multidisciplinary Sciences” WC (Rafols & Meyer, 2010). This issue is likely to be relatively pronounced in convergence science, which by its very impetus and nature tends to draw on multidisciplinary research. To further emphasize this point, consider again the exemplary bio-mechatronics convergence science by (Hochberg et al., 2012), representing a team of three distinct CIP domains (neuroscience, medicine and biotechnology) that spans all six MeSH SA domains tailored around human brain science (Petersen et al., 2021). This publication cites 37 articles, and 22% of these belong to journals classified by the “Multidisciplinary Sciences” WC, consistent with the frequencies noted in (Rafols & Meyer, 2010), Conversely, of the 1427 follow-up publications citing this article, 8.6% belong to journals classified by the “Multidisciplinary Sciences”, which indicates a non-negligible level of misattributed classification information associated with WC, as well as the inconsistency in these frequencies when comparing the cited and citing WC, partly attributable to the fact that highly-cited articles are more likely to be published in high-impact journals, which are also likely to be classified as “Multidisciplinary Sciences”, representing a selection bias that would be challenging to ameliorate.

[2] Namely, the only 18 WOS Subject Categories (WC) that are not represented by any journal indexed within PubMed are: “Dance”, Engineering, Geological“, ”Engineering, Manufacturing“, ”Engineering, Marine“, ”Engineering, Ocean“, ”Engineering, Petroleum“, ”Literature, African, Australian, Canadian“, ”Literature, Slavic“, ”Logic“, ”Materials Science, Ceramics“, ”Materials Science, Characterization, Testing“, ”Materials Science, Composites“, ”Materials Science, Paper & Wood“, ”Materials Science, Textiles“, ”Metallurgy & Metallurgical Engineering“, ”Mining & Mineral Processing“, ”Ornithology“, ”Transportation Science & Technology".

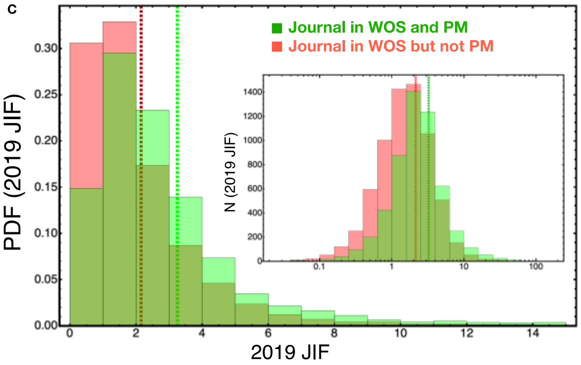
Figure 6(A) shows the number of journals in PubMed associated with the 50 most and 50 least frequent WOS categories across all journals indexed by PubMed. Whereas the top-50 WC largely correspond to the primary focus of biomedical and health sciences, there are also strong indications of other distinct domains that are well-represented, namely various social science journals specializing in "Economics", "Law", "Sociology" and "Political Sciences". **Figure 6**(B) addresses a complementary question – what types of journals are not indexed by PubMed that are indexed by WOS? Results indicate this omitted journal set is mostly populated by the following domains: the management, social sciences, humanities and arts; computer and information sciences; mathematics; physics; and engineering. Regarding the prominence of these omitted journals, **Fig. 6**(C) compares the 2019 JCR Impact Factors (JIF) calculated by WOS, which shows that PubMed is significantly more selective, with the average JIF for journals indexed within PubMed (3.25) roughly 50% larger than the average for those indexed by WOS that are not indexed by PubMed (2.15). This difference in distribution persists beyond the location of the characteristic values (mean and median), to the distribution level as well, with more than half (52%) of PubMed journals featuring JIF above 2.15. **Figure 6**(D) shows the notable omissions from PubMed, ranked by JIF, which are dominated by core physics, chemistry and other STEM journals.

Interestingly, the least-common WC appearing in PubMed identify some extremely distant domains relative to the core, but close inspection reveals the pervasive nature of multidisciplinary intersections. By way of example, the article indexed by PubMed published by a journal with WC = "Folklore" is "Richard III's disfigurement: a medical postscript" (Jones, 1980) (PMID 11619652), which is assigned the MeSH "Congenital Abnormalities", "History, Early Modern 1451-1600, "History, Modern 1601-", and "United Kingdom".

Regarding the multidisciplinary scope of MeSH keywords, the conceptual space spanned by the L_1 branches "Anthropology, Education, Sociology, and Social Phenomena [I]", "Technology, Industry, and Agriculture [J]", "Humanities [K]" and "Information Science [L]" offer plenty of detail relevant to various convergence frontiers beyond the bio-medical domains it was designed to cover. This includes, but is not limited to, the domains of "Sustainable Development" [MeSH tree number I01.655.500.608.700, N06.230.080.900], "Sociology" [F04.096.879.757, I01.880], and "Government" [I01.409, N03.540.348]. In summary, by comparing with the list of 10 NSF *Convergence Accelerator* challenge areas (NSF, accessed 2/2021), the convergence frontiers that are definitively not suitable for MeSH classification are Quantum Technology (Track C), Securely Operating Through 5G Infrastructure (Track G). As previously noted, we anticipate this issue could be easily ameliorated by integrating the recently developed PhySH ontology by way of advances in ontology-alignment techniques (Wang et al., 2018).

A WOS Category	#	WOS Category	#
Public, Environmental & Occupational Health	298	...	...
Biochemistry & Molecular Biology	253	Agriculture, Multidisciplinary	7
Pharmacology & Pharmacy	248	Biodiversity Conservation	7
History	243	Chemistry, Inorganic & Nuclear	7
Neurosciences	231	Computer Science, Theory & Methods	7
Oncology	198	Crystallography	7
Psychiatry	198	Engineering, Industrial	7
Clinical Neurology	189	Geology	7
Surgery	189	Operations Research & Management Science	7
Nursing	181	Physics, Condensed Matter	7
Medicine, General & Internal	171	Soil Science	7
Cell Biology	169	Horticulture	6
Economics	146	Literary Theory & Criticism	6
Cardiac & Cardiovascular System	143	Physics, Nuclear	6
Genetics & Heredity	141	Psychology, Mathematical	6
Immunology	137	Robotics	6
Radiology, Nuclear Medicine & Medical Imaging	131	Andrology	5
Endocrinology & Metabolism	126	Computer Science, Hardware & Architecture	5
Pediatrics	125	Electrochemistry	5
Medicine, Research & Experimental	123	Engineering, Mechanical	5
Microbiology	110	Limnology	5
Biotechnology & Applied Microbiology	109	Literature, British Isles	5
Rehabilitation	108	Literature, German, Dutch, Scandinavian	5
Veterinary Sciences	106	Mineralogy	5
Law	104	Oceanography	5
Health Care Sciences & Services	102	Physics, Fluids & Plasmas	5
Dentistry, Oral Surgery & Medicine	101	Physics, Mathematical	5
Environmental Sciences	100	Energy & Fuels	4
Obstetrics & Gynecology	89	Hospitality, Leisure, Sport & Tourism	4
Psychology, Multidisciplinary	88	Imaging Science & Photographic Technology	4
Sociology	87	Literature, American	4
Health Policy & Services	84	Physics, Particles & Fields	4
Psychology, Clinical	84	Agricultural Economics & Policy	3
Infectious Diseases	81	Folklore	3
Toxicology	79	Forestry	3
Gastroenterology & Hepatology	78	Music	3
Pathology	77	Telecommunications	3
History & Philosophy Of Science	76	Theater	3
Orthopedics	76	Transportation	3
Physiology	76	Agricultural Engineering	2
Plant Sciences	76	Architecture	2
Political Science	76	Engineering, Civil	2
Urology & Nephrology	76	Mathematics	2
Engineering, Biomedical	70	Remote Sensing	2
Hematology	69	Thermodynamics	2
Nutrition & Dietetics	67	Construction & Building Technology	1
Peripheral Vascular Diseases	67	Film, Radio, Television	1
Biology	66	Green & Sustainable Science & Technology	1
Psychology	65	Materials Science, Coatings & Films	1
Biophysics	64	Poetry	1
...	...	Quantum Science & Technology	1

B WOS Category	#
Education & Educational Research	684
Mathematics	467
Economics	424
Materials Science, Multidisciplinary	388
Management	366
Language & Linguistics	356
Humanities, Multidisciplinary	334
Engineering, Electrical & Electronic	320
Law	312
Mathematics, Applied	308
History	279
Religion	276
Business	269
Linguistics	259
Philosophy	257
Computer Science, Information Systems	236
Political Science	230
Environmental Sciences	226
Geosciences, Multidisciplinary	213
Business, Finance	212
Social Sciences, Interdisciplinary	206
Communication	198
Plant Sciences	181
Engineering, Civil	174
Computer Science, Artificial Intelligence	170
Engineering, Mechanical	170
Literature	169
Chemistry, Multidisciplinary	167
Engineering, Multidisciplinary	163
Mechanics	155
Physics, Applied	154
Ecology	151
Medicine, General & Internal	149
Engineering, Chemical	145
Energy & Fuels	140
Information Science & Library Science	140
Archaeology	135
Computer Science, Theory & Methods	134
Chemistry, Physical	130
Computer Science, Interdisciplinary Applications	129
Statistics & Probability	129
Art	127
Environmental Studies	127
Hospitality, Leisure, Sport & Tourism	126
Computer Science, Software Engineering	122
International Relations	122
Sociology	121
Psychology, Multidisciplinary	120
Public, Environmental & Occupational Health	120
Food Science & Technology	116



D Journal	2019 JIF	Journal	2019 JIF
NATURE REVIEWS MATERIALS	71.20	WHO TECHNICAL REPORT SERIES	59.00
NATURE ENERGY	46.50	REVIEWS OF MODERN PHYSICS	45.00
WORLD PSYCHIATRY	40.60	LIVING REVIEWS IN RELATIVITY	35.40
NATURE REVIEWS CHEMISTRY	35.00	PROGRESS IN MATERIALS SCIENCE	31.60
NATURE PHOTONICS	31.20	NATURE CATALYSIS	30.50
ENERGY & ENVIRONMENTAL SCIENCE	30.30	PROGRESS IN ENERGY AND COMBUSTION SCIENCE	28.90
NATURE ELECTRONICS	27.50	JOULE	27.10
MATERIALS SCIENCE & ENGINEERING R–REPORTS	26.60	MATERIALS TODAY	26.40
ADVANCED ENERGY MATERIALS	25.20	IEEE COMMUNICATIONS SURVEYS AND TUTORIALS	23.70
PROGRESS IN POLYMER SCIENCE	22.60	NATURE CLIMATE CHANGE	20.90
LIVING REVIEWS IN SOLAR PHYSICS	20.00	CHEM	19.70
ANNUAL REVIEW OF PLANT BIOLOGY	19.50	NATURE PHYSICS	19.30
ACS ENERGY LETTERS	19.00	SCIENCE ROBOTICS	18.70
APPLIED PHYSICS REVIEWS	17.10	NANO TODAY	16.90
ADVANCED FUNCTIONAL MATERIALS	16.80	WILEY INTERDISCIPLINARY REVIEWS–COMPUTATIONAL MOLECULAR SCIENCE	16.80
NATIONAL SCIENCE REVIEW	16.70	APPLIED CATALYSIS B–ENVIRONMENTAL	16.70
NANO ENERGY	16.60	ADVANCES IN PHYSICS	16.40
ENERGY STORAGE MATERIALS	16.30	ASTRONOMY AND ASTROPHYSICS REVIEW	16.00
ADVANCED SCIENCE	15.80	FUNGAL DIVERSITY	15.40
COORDINATION CHEMISTRY REVIEWS	15.40	JOURNAL OF EXTRACELLULAR VESICLES	15.00

Figure 6: Subject Category coverage of PubMed. (A) Top and bottom 50 WOS Categories (WC) represented by journals indexed by PubMed (each count indicated in the # column represents a single journal). All but 18 of the 256 WOS WC are represented by journals indexed by PubMed. (B) Top 50 WC for journals not indexed by PubMed, which identifies the core areas (mathematics, humanities and social sciences, physics) that are under-represented in PubMed with respect to WOS. And while 'History' occurs in both panels A and B, this merely indicates that there are relatively large number of journals indexed by WOS with this WC, and only a fraction of those appear in PubMed, but in total numbers this is still a large number of distinct journals. (C) Distribution of 2019 JIF for journals indexed by PubMed and those missing from PubMed. Vertical dashed bars indicate the corresponding distribution mean. Journals missing from PubMed are of relatively lower JIF. (D) Top 40 journals by 2019 JIF missing from PubMed, which are primarily core physics and chemistry journals. Comparison with panel A shows that these WC are nevertheless spanned by PubMed, just in smaller proportions, and also includes the main multidisciplinary journals (*Nature*, *PNAS* and *Science*) where the highest impact research in these core STEM areas are frequently published. In summary, while the coverage of PM is not as extensive as WOS, it spans nearly the same topical range as WOS, the journals it does include are of generally higher research impact. Thus, the principal advantage of PM is the article-level topical annotation by way of MeSH embedded in a hierarchical thesaurus-based ontology.

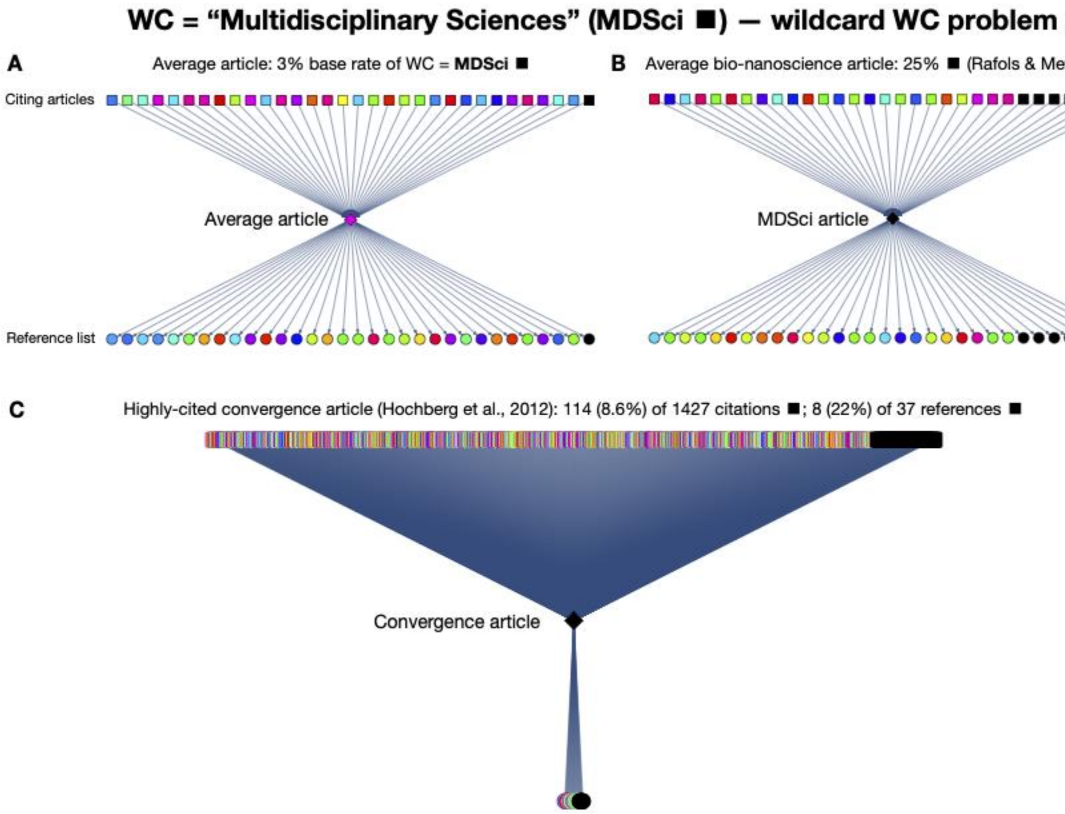

WC = "Multidisciplinary Sciences" (MDSci ■) — wildcard WC problem
A
Average article: 3% base rate of WC = MDSci ■
Citing articles
Average article
Reference list
B
Average bio-nanoscience article: 25% ■ (Rafols & Meyer ,2010)
MDSci article
C
Highly-cited convergence article (Hochberg et al., 2012): 114 (8.6%) of 1427 citations ■; 8 (22%) of 37 references ■
Convergence article

Figure 7: Illustration of the WC ``Multidisciplinary'' wildcard problem. (A) The local citation network schematic shows the average paper based upon the base rate of WC = "Multidisciplinary Sciences", such that roughly 1 in 30 citing articles, and 1 in 30 referenced articles are labelled as such. (B) However, using the articles in the case study of Rafols & Meyer (2010), interdisciplinary research is likely to have a higher frequency of WC = MDSci, and so it's not clear how extant methods deal, or have dealt with this issue when measuring IDR. Are the WC = MDSci discarded, i.e. treated in the same way as ``missing information''. Or are they included, such that MDSci articles are assumed to be similar just by their very classification as such? (C) This issue is likely to be even further exacerbated in highly-cited research, and/or convergence science which is typically performed in new frontiers of science, such that a larger number of citing documents are classified as wildcards – ie, the is no useful information provided in the classification of a research article as ``Multidisciplinary'' for the intensive purpose of measuring IDR, or convergence for that matter.