

Riviste scientifiche nel settore ICAR17: analisi quantitativa delle *keywords* e dei temi di ricerca

Alessandro Luigini

Abstract

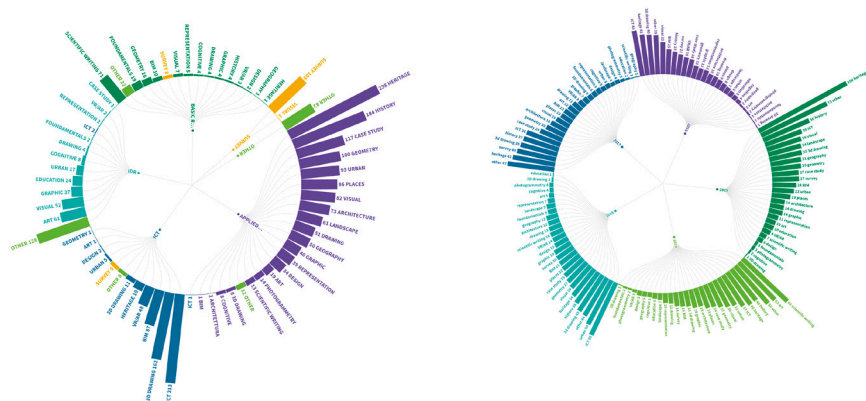
Il paper ha lo scopo di presentare una analisi quantitativa della frequenza, della distribuzione e della caratterizzazione delle *keywords* presenti negli articoli pubblicati sulle riviste scientifiche riconducibili al settore scientifico disciplinare ICAR17 nel periodo dal 2016 al 2020. Le analisi sono eseguite su un dataset realizzato a partire dalle *keywords* e successivamente integrato tramite l'associazione delle *keywords* a categorie tematiche e a cluster che definiscono differenti aree di ricerca.

Il contributo ha principalmente valore metodologico, in quanto presenta una modalità di sviluppo della analisi delle *keywords* ad hoc per il settore disciplinare del Disegno, ed è passibile di ulteriori affinamenti soprattutto per i criteri di determinazione delle categorie e dei cluster, oltre che per l'ampliamento del periodo di osservazione.

Le analisi quantitative sviluppate consentono di avanzare alcune valutazioni sull'uso delle *keywords* per ogni rivista a supporto degli autori che volessero identificare il contesto editoriale in cui pubblicare le proprie ricerche e di identificare possibili trend di sviluppo nella ricerca del settore.

Parole chiave

analisi keywords, metodologia di ricerca, riviste scientifiche, scrittura scientifica



Distribuzione delle
categorie di *keyword*
secondo i cluster e
secondo i 5 anni di
osservazione

Introduzione

La bibliografia statistica, o bibliometria, è una applicazione nota fin dai primi decenni del XX secolo, e già in origine è stata applicata ai più svariati contesti: dallo studio dell'andamento dei brevetti in UK [Hulme 1923] all'obsolescenza delle pubblicazioni acquisite dalle biblioteche [Gosnell 1944], a titolo esemplificativo.

La bibliometria ha lo scopo di "far luce sui processi di comunicazione scritta e sulla natura e il corso dello sviluppo di una disciplina [...] attraverso il conteggio e l'analisi dei vari aspetti della comunicazione scritta" o "l'assemblaggio e l'interpretazione di statistiche relative a libri e periodici [...] per dimostrare movimenti storici, per determinare l'uso nazionale o universale di ricerca di libri e riviste, e per accertare in molte situazioni locali l'uso generale di libri e riviste" [Pritchard 1969, p. 348]. Gli elementi che si possono utilizzare nella indagine bibliometrica sono molti, tra cui le citazioni, gli *abstract* o le *keywords*, ognuno dei quali richiede metodologie specifiche di analisi.

L'uso delle *keywords* nell'editoria scientifica ha lo scopo di facilitare l'identificazione dei topic di ricerca contenuti in un paper o in un articolo e sono la base della metadattazione per la indicizzazione in SCOPUS, WOS e altri database. Le *keywords* sono, in estrema sintesi, la condensazione degli argomenti trattati nella pubblicazione, e insieme al titolo e all'*abstract* servono agli autori per comunicare sinteticamente e chiaramente ad altri ricercatori gli elementi distintivi del proprio lavoro. In molti contesti della più recente editoria scientifica, è possibile reperire vademecum per una corretta scrittura scientifica e in molti casi (*Elsevier*, *Springer*, ecc.) sono dedicate proprio alle *keywords* indicazioni specifiche: oltre alle regole base, alcune delle quali valuteremo di seguito, molti offrono anche servizi di editing linguistico, comprensivo di scelta delle *keywords*, oppure la loro ottimizzazione per i motori di ricerca (SEO, *Search Engine Optimization*). Questo interesse particolare dimostra la rilevanza che le *keywords* hanno nella identificazione della pubblicazione di una ricerca e, soprattutto, nella sua accessibilità e nella sua diffusione.

L'analisi delle *keywords* è utilizzata frequentemente nelle discipline bibliometriche anche per l'analisi dei trend della ricerca, perché capaci di fotografare sinteticamente i focus delle pubblicazioni in una determinata area disciplinare o in un preciso contesto di pubblicazione. Il settore scientifico disciplinare ICAR17 Disegno non è compreso nei settori bibliometrici nella classificazione ANVUR, ma lo *status quo* nell'editoria scientifica nazionale e internazionale si è rapidamente conformata ai meccanismi e ai criteri bibliometrici: in particolare le riviste scientifiche sembrano aver in larga parte accolto questi criteri, e per tale ragione è opportuno applicare sistemi di analisi della ricerca tipici delle aree disciplinari bibliometriche, anche se con la consapevolezza che in un settore scientifico le cui ragioni sono da rintracciare nell'alveo delle scienze umane [1] i soli dati bibliometrici non possono essere esaustivi nella lettura dello status della ricerca.

Questo paper presenta una revisione del modello di analisi proposto al convegno UID2020 *Connessioni* [Luigini 2020] in cui si sono analizzate le *keywords* dei precedenti due convegni annuali della società scientifica Unione Italiana del Disegno (2018 e 2019), con il risultato di aver potuto tracciare alcune considerazioni sulla diffusione di temi di ricerca interdisciplinare. In questa sede si analizzeranno le riviste scientifiche disciplinari, con l'obiettivo di una mappatura generale degli interessi di ricerca del settore e di una caratterizzazione delle riviste prese in considerazione.

Metodologia

La selezione delle riviste da mappare è stata determinata secondo tre criteri:

1. iscrizione negli elenchi ANVUR come riviste scientifiche o riviste di classe A;
2. direzione di uno studioso afferente al settore scientifico disciplinare ICAR17;
3. la presenza sistematica delle *keywords* negli articoli pubblicati.

Questi criteri hanno portato alla selezione di sei riviste:

denominazione	ISSN	classificazione AN-VUR	editore	direttore	open access	URL
Dienne building information modeling, data & semantics	2610-8755	Scientifica	DEI Tipografia del Genio Civile	Tommaso Empler	No	http://www.dienne.org
Disegnare. Idee Immagini	1123-9247 (cartaceo) 2704-7105 (online)	Classe A	Dipartimento di Storia, Disegno e Restauro dell'Architettura / Gangemi editore	Carlo Bianchini	No	https://www.gangemieditore.com/periodici/disegnare-idee-immagini/29
Disegnarecon Scientific Journal on Architecture and Cultural Heritage	1828-5961	Scientifica	Università degli Studi dell'Aquila	Stefano Brusaporci	Si	http://disegnarecon.univaq.it
Diséño Biannual Journal of the UID - Unione Italiana per il Disegno Scientific Society	2533-2899	Scientifica	Unione Italiana per il Disegno	Francesca Fatta	Si	https://disegno.unioneitalianadisegno.it
Scires-IT SCientific RESearch and Information Technology	2239-4303	Classe A	CASPUR-CIBER Publishing	Virginia Valzano, Michela Cigola	Si	http://www.sciresit.it
XY dimensioni del disegno	2499-8338 (cartaceo) 2499-8346 (online)	Scientifica	Università degli Studi di Trento / Officina Edizioni	Roberto de Rubertis	Si	http://rivista.xydigitale.it

Tabella 1. Schedatura delle riviste mappate nel presente paper, elencate in ordine alfabetico

rivista	fascicoli mappati	articoli	articoli keywords con	keywords	% su totale
Dienne	7	55	54	253	9,75%
Disegnare Idee Immagini	10	65	65	328	12,64%
Disegnarecon	10	211	208	935	36,04%
Diséño	6	101	72	294	11,33%
Scires-IT	10	112	110	562	21,67%
XY	8	79	78	222	8,56%
TOTALE	51	623	587	2594	100%

Tabella 2. Dati complessivi su cui è realizzato il dataset

Sono stati presi in considerazione tutti i numeri pubblicati dalle riviste in tabella 1 nell'intervallo di tempo 2016-2020, salvo *Dienne* e *diséño* che sono state avviate entrambe nel 2017. In totale si sono analizzati 51 fascicoli e 623 articoli, di cui 587 corredati da *keywords*. Complessivamente le *keywords* che formano il dataset sono 2594. Per le riviste che pubblicano anche in italiano (*Disegnare. Idee Immagini*, *diséño* e *XY*) sono state prese in considerazione solo le *keywords* in lingua inglese.

Come nella applicazione precedente ai *proceedings* dei convegni UID 2018 e 2019 [Luigini 2020] e già consolidato in letteratura [Pesta 2018, p. 2], le *keywords* di tutti gli articoli sono state inserite in un dataset e sono state successivamente 'normalizzate' e 'aggregate': 'normalizzate', perché i dati sono risultati non sufficientemente 'nitidi' presentando numerose declinazioni di temi particolarmente diffusi (es: 3D model, 3D models, 3D modeling, etc.), mantenendo le quali si sarebbe ottenuta una alterazione delle analisi, e aggregate perché la grande variabilità delle *keywords*, anche dopo la loro normalizzazione, non avrebbe consentito una lettura significativa della diffusione di ambiti di ricerca e temi prettamente disciplinari. Questo processo, centrale nella impostazione di questo lavoro, necessita un ulteriore approfondimento metodologico.

Aggregazioni successive in categorie e cluster

Il data set è stato strutturato su tre livelli: il più esterno (fig. 2) è il livello che contiene tutte le *keywords* normalizzate; il livello intermedio indica il macro-tema, la categoria a cui ogni singola *keywords* viene stabilito faccia riferimento (*3D drawing, architecture, art, BIM, case study, cognitive, design, drawing, education, fundamentals, geography, geometry, graphic, heritage, history, ICT, landscape, other, photogrammetry, places, representation, scientific writing, survey, urban, visual, VR/AR*); il livello più interno invece indica alcuni cluster (*basic research, applied research, survey, ICT, IDR [2], other*), che vengono assegnati alla singola *keywords* e non alle categorie. Quest'ultima precisazione implica che due *keywords* appartenenti alla stessa categoria possono appartenere a due cluster differenti (es: la *keyword visual culture* fa parte della categoria *visual* e del cluster *basic research*, la *keyword display* fa parte sempre della categoria *visual* ma del cluster *IDR*). A differenza di quanto accaduto in precedenza, in questo dataset è stato mantenuto anche l'informazione del titolo dell'articolo che ha consentito una maggiore precisione nella associazione delle singole *keywords* alle categorie e ai cluster. Alcune altre considerazioni a questo punto sono necessarie.

La prima è che nonostante i cluster *basic research* e *applied research* sembrerebbero poter contenere tutto il dataset, si è preferito identificare in modo distinto le *keywords* riconducibili univocamente a tre cluster (*survey, ICT e IDR*) particolarmente significativi nelle valutazioni conclusive e, soprattutto, perché in molti casi l'eventuale assegnazione di queste *keywords* ai cluster *basic* o *applied research*, è incerto. Inoltre, anche se potrebbe rischiare di apparire pleonastico, è necessario anche definire chiaramente cosa si intenda per ricerca di base e ricerca applicata.

La 'ricerca di base' (o 'fondamentale') "è un'attività sperimentale o teorica sviluppata al fine dell'ampliamento delle conoscenze e di cui non si prevede una specifica applicazione o utilizzazione diretta, mentre La ricerca applicata è quella ricerca originale svolta per ampliare le conoscenze, ma principalmente allo scopo di una specifica applicazione pratica" [3]. Inoltre, quando Stokes nel 1997 propone un modello per la classificazione della ricerca, ipotizza una categorizzazione delle ricerche in base a due fattori, ovvero quanto le ricerche siano proiettate a: 1) ampliare le conoscenze e 2) applicare le conoscenze.

		Relevance for immediate application	
		No	Si
Relevance for the advancement of knowledge	Si	Quadrante di Bohr PURE BASIC RESEARCH	Quadrante di Pasteur USE-INSPIRED BASIC RESEARCH
	No	uninteresting research	Quadrante di Edison PURE APPLIED RESEARCH

Fig. 1. Schema della distribuzione della ricerca secondo Stokes (1997).

Stokes usa come simbolicamente i nomi di tre grandi studiosi del passato, come esemplificazione dei tre approcci alla ricerca: il "quadrante di Bohr" comprende la pura ricerca di base, che ha come obiettivo l'avanzamento della conoscenza senza porsi un obiettivo applicativo immediato, il "quadrante di Edison" comprende la ricerca applicata, che ha come obiettivo la concreta applicazione della conoscenza già disponibile, mentre il "quadrante di Pasteur" – *Use-Inspired Basic Research* – comprende la ricerca che ha come obiettivo l'avanzamento della conoscenza con una immediata possibilità di applicazione a casi studio concreti.

È facile capire come sarebbe possibile determinare la differenza tra un prodotto scientifico

Analisi dei dati

La figura 2 mostra la distribuzione di quantità delle singole *keywords*, nella corona esterna, dove si possono notare che risultano leggibili solo alcune *keywords* la cui frequenza è uguale o superiore a 10. Tali *keywords* sono:

cultural heritage	47	landscape	16
BIM	42	perspective	15
3D modeling	38	virtual reality	14
drawing	36	communication	13
HBIM	28	interoperability	11
representation	25	open science	11
architecture	20	digital survey	10
survey	19	geodesign	10
street art	17		

Tabella 3. Frequenza delle *keywords* con valore uguale o superiore a 10

Lo schema in tabella 3 ci mostra che – a differenza di quanto accadeva nei convegni UID 2018 e 2019 dove le *keywords* più frequenti erano *drawing*, *survey*, e *representation* – sulle riviste scientifiche tre *keywords* più frequenti sono *cultural heritage*, BIM e *3D modeling*. Considerando inoltre che *drawing* e *representation* sono comunque tra le più utilizzate, con 28 ripetizioni troviamo HBIM, ovvero una *keyword* che identifica l'applicazione della tecnologia BIM al patrimonio.

Nella seconda colonna troviamo tre *keywords* non presenti, se non occasionalmente, nei convegni UID 2018 e 2019: *street art*, *open science* e *geodesign*. La ragione dell'alta frequenza di queste tre *keywords* è la presenza di alcuni fascicoli monografici, al di fuori dei quali questi temi non sono osservabili: *street art* e *geodesign* derivano da due fascicoli di *Disegnarecon*, rispettivamente vol. 11, n. 20 *Co-Design: digital tools for knowledge-building and decision-making* e vol. 13, n. 24 *Street art. Drawing on the walls*, mentre *open science* dal fascicolo speciale 2020 di *Scires-IT Open Science: New models of scientific communication and research evaluation*. Questo fenomeno, chiaramente, è tanto più significativo quanto più ridotto è il campione: 5 anni di pubblicazioni per un totale di 51 fascicoli sono un campione attendibile, ma la presenza di un solo fascicolo monografico è ancora capace di spostare significativamente i risultati.

Anche *Interoperability* ha una motivazione specifica che va rintracciata entrando nel merito della collocazione delle *keywords* all'interno delle riviste: 10 delle 11 ricorrenze sono infatti contenute in diversi fascicoli della rivista *Dienne* che, come vedremo di seguito, è il contesto di pubblicazione più numeroso di *keywords* riferite al cluster ICT. Le restanti (*perspective*, *Virtual Reality*, *communication* e *digital survey*) sono sostanzialmente distribuite su tutte le riviste.

Quantità keywords	Frequenza	Quantità keywords	Frequenza	Quantità keywords	Frequenza	Quantità keywords	Frequenza
1	47	1	20	1	13	6	7
1	42	1	19	1	12	13	6
1	38	1	17	2	11	10	5
1	36	1	16	2	10	26	4
1	28	1	15	4	9	42	3
1	25	1	14	8	8	138	2

Tabella 4. Indicazione delle quantità di *keywords* per ogni singolo valore di frequenza superiore a 1

Analizzando le frequenze è utile capire anche quante *keywords* possono ritenersi significative perché ricorrenti, e nella tabella 4 sono elencate tutte le frequenze superiori a 1 con l'indicazione di quante *keywords* presentano una specifica frequenza: si può notare che la somma delle *keywords* con frequenza uguale o superiore a 10, riportate in tabella 3, sia di 384, mentre la somma delle *keywords* con frequenza compresa tra 2 e 9 sia di 776, e le *keywords* con una sola ricorrenza sono 1434.

Considerando le numerose declinazioni di *keywords* riconducibili a temi analoghi (si pensi alle declinazioni di *survey*, *3d survey*, *integrated survey*, *digital survey*, *architectural survey* ecc.) si è proceduto a una analisi della frequenza di singole parole all'interno del dataset, riportata nella tabella di seguito:

heritage / HBIM	145	representation	64
model	118	image / visual	58
architecture	114	virtual	51
city/urban	104	data	50
survey	98	landscape	47
3D	97	cartography / map / territory	36
graphic	95	archaeology	33
BIM / HBIM	77	geometry	25
drawing	69		

Tabella 5. Frequenza di singole parole contenute nelle *keywords* con frequenza uguale o superiore a 25

La parola *heritage* anche in questo caso risulta la più frequente, seguita dalla parola *model*, che richiama principalmente alla modellazione tridimensionale più che al 'modello' in termini geometrici o concettuali, e che quindi si potrebbe aggregare con il termine 3D, segnalandosi come il tema di ricerca più esteso nel settore.

Seguono da due ambiti di applicazione della ricerca del settore del Disegno evidentemente centrali: *architecture* e *city / urban*. Il termine *survey*, con tutte le sue possibili declinazioni, si ferma poco al di sotto delle 100 ricorrenze ma che sicuramente ha una stretta correlazione sia con *heritage* che con *architecture* e *city / urban*. Di un certo rilievo risulta la frequenza della parola *graphic*, anche questa distribuita tra tutte le riviste e in tutto il lasso di tempo osservato, utilizzata tanto quanto *Survey* e 3D e più di BIM / HBIM, Drawing o Representation. Successivamente alle analisi sulle singole *keywords* e sulle parole che compongono le *keywords*, si è passato a una analisi delle frequenze delle categorie, con una distribuzione annuale, come mostrato in figura 3. In questo grafico osserviamo l'andamento ondineo ma sostanzialmente costante della distribuzione delle categorie nei cinque anni osservati, con gli evidenti contributi di alcuni numeri monografici che incidono sull'andamento generale delle sei riviste scientifiche in momenti specifici.

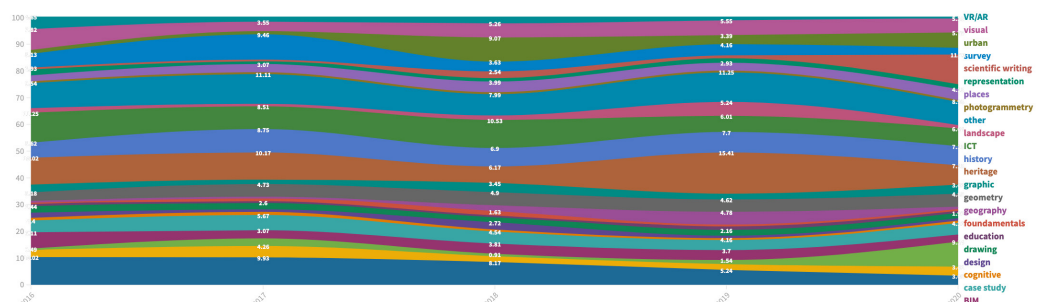


Fig. 3. Distribuzione temporale della frequenza delle *keyword* associate alle categorie.

La figura 4 mostra il dettaglio della distribuzione temporale di ogni singola categoria e si leggono chiaramente i contributi di singoli numeri monografici, in corrispondenza delle cuspidi e dei massimi valori nei singoli diagrammi.

La figura 5 mostra invece la distribuzione in termini assoluti delle *keywords* organizzate in *cluster*, per ogni singola rivista e si evince anche il contributo che ogni rivista ha dato per la definizione del dataset, in quanto la dimensione della barra su cui è riportato il nome della rivista sull'estremo destro del grafico, è proporzionato al numero di *keywords* registrate nel dataset rintracciabile nell'ultima colonna della tabella 2.

Nella figura 6 è possibile osservare le caratteristiche di distribuzione delle *keywords*, delle categorie e dei *cluster* di ogni singola rivista.

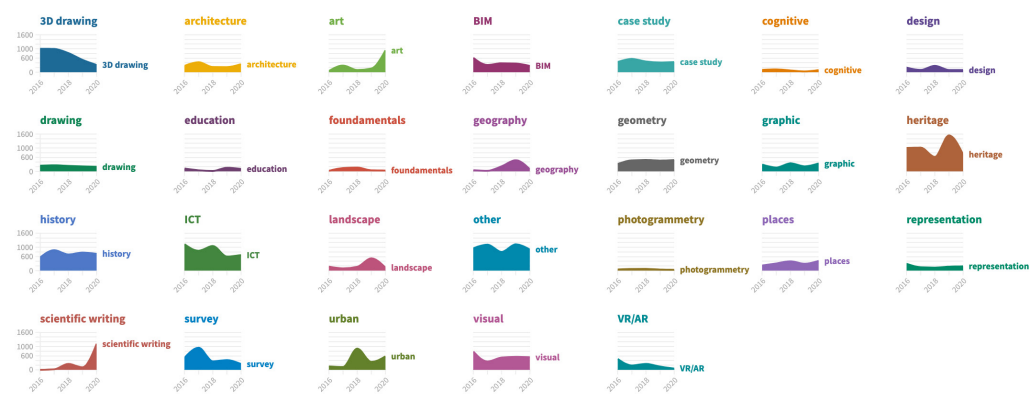


Fig. 4. Distribuzione temporale della frequenza delle keyword associate alle singole categorie.

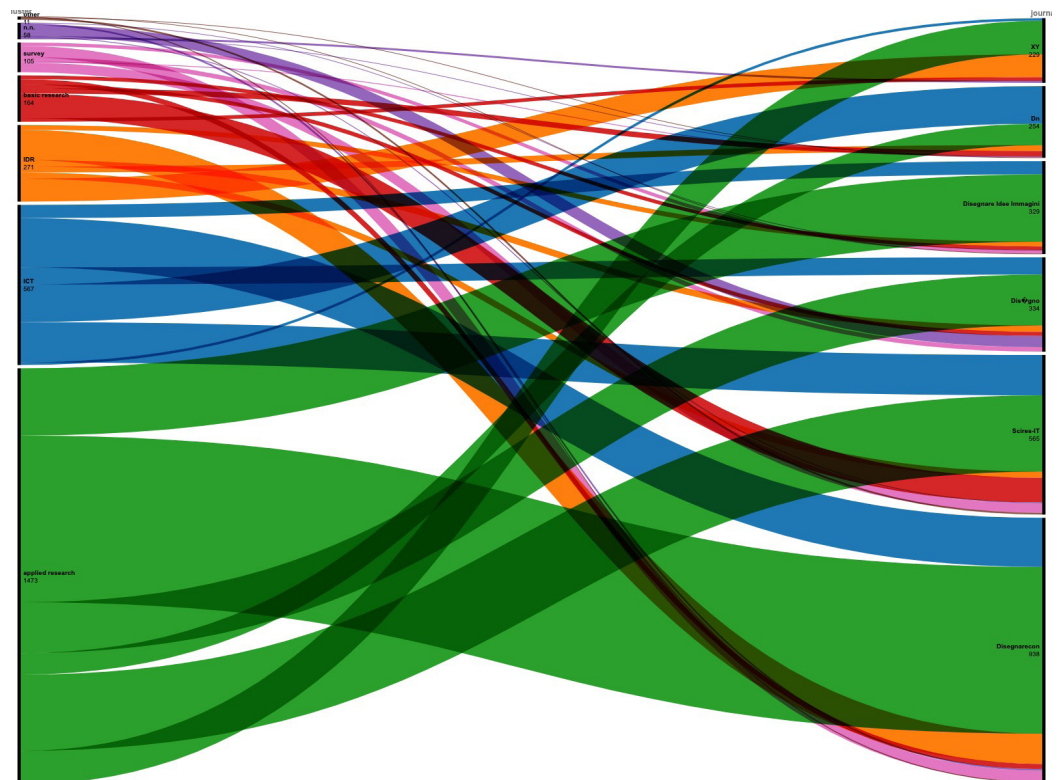


Fig. 5. Distribuzione delle categorie appartenenti ai sei cluster in base alle riviste.

Conclusioni

Abbiamo definito cosa siano le *keywords* nella scrittura scientifica, e abbiamo determinato le differenze tra ricerca di base e ricerca applicata, aggiungendo un modello di ricerca – *used-inspired basic research* – che probabilmente è più adatto per indicare una parte significativa della ricerca svolta nel settore scientifico disciplinare del Disegno. Dopo aver selezionato le riviste secondo i tre criteri sopraelencati, le abbiamo schedate e abbiamo raccolto tutte le *keywords* contenute negli articoli del quinquennio 2016-2020.

Dalla tabelle 3 e 5 si evince che le *keywords* più frequenti indicano come il centro della disciplina, almeno in termini quantitativi, sia la modellazione tridimensionale del patrimonio, con particolare attenzione all'utilizzo del *Building Information Modeling*, e che per sua natura la ricerca sul patrimonio ha un carattere applicativo, anche se è possibile presumere che una parte della ricerca qui indicata come *applied research* possa far riferimento al "quadrante di Pascal", ovvero a un modello di ricerca di base sviluppata su un caso applicativo.

L'alto numero di *keywords* con una singola ricorrenza (1434, ovvero il 54% del totale) è riconducibile al fatto che, come abbiamo già anticipato, il carattere applicativo di ampia parte della ricerca del settore comporta una presenza di numerose *keywords* legate alla determinazione dei luoghi, degli autori o delle denominazioni dei monumenti nei singoli casi di studio.

Una osservazione entrando nel merito delle singole *keywords*, è relativa alla loro scelta: sappiamo che le *keywords* servono all'autore per rendere il proprio lavoro più visibile nelle banche dati, e che quindi è necessario adottare strategie adeguate a far sì che il paper non si smarrisca in mezzo a mille altri. Nei vademecum per la scrittura scientifica, dove ci sono indicazioni specifiche sulla scelta delle *keywords*, viene spesso consigliato di scegliere *keywords* specifiche, con una sintassi intermedia tra la singola parola e il linguaggio naturale, e, soprattutto, di evitare *keywords* eccessivamente generiche (chiamate *poor keywords* perché caratterizzano troppo marginalmente il paper): in questo senso, considerando che le riviste analizzate hanno tutte come dominio il campo del disegno e della rappresentazione grafica e visuale, si può concludere che le *keywords* *drawing*, *representation*, o *survey* per quanto centrali per il settore, sono da ritenersi delle *poor keywords*, così come *architecture*, *digital survey* (*keywords* composta ma ridondante), *communication*, ecc.

Queste sei appena elencate, che sono le *poor keywords* con frequenza uguale o superiore a 10, rappresentano circa il 5% delle *keywords* totali, e circa l'11% di quelle caratterizzanti (ovvero con frequenza superiore a 1). Lo studioso che cercherà, ad esempio, la ricostruzione digitale di un progetto non realizzato, difficilmente utilizzerà le parole chiave *drawing*, *representation* o *architecture*, ma probabilmente utilizzerà, oltre al nome del progetto e dell'autore, *keywords* come *unbuilt project*, *digital reconstruction* o, più in generale, *3D modeling*.

Infine, si può valutare l'indirizzo nell'uso delle *keywords* nelle sei riviste catalogate, con la precisazione che le osservazioni che seguono non possono essere estese alla rivista nel suo complesso in quanto l'uso delle *keywords* è un parametro significativo ma parziale dei contenuti di una rivista.

In questo contesto è possibile paragonare la distribuzione delle *keywords* complessive, con la distribuzione delle *keywords* in ogni rivista, e sottolineare che:

1. Le riviste che hanno una distribuzione percentuale dei cluster più simile all'intero settore sono *diségno* – come era facile ipotizzare essendo la rivista della società scientifica UID – e *Disegnarecon*;
2. Anche *Scires-IT* non si discosta molto dalla distribuzione dei cluster nell'intero settore, se non per una accentuazione del cluster *ICT* – focus principale della rivista – e del cluster *basic research*, alimentato dal numero monografico del 2020 sulla scrittura scientifica e la valutazione della ricerca;
3. Il cluster *ICT* è preponderante solo nella rivista *Dienne*, dove supera il 50% delle *keywords*, che quindi si segnala come il principale contesto editoriale per ricerche in questo ambito;
4. La ricerca interdisciplinare identificabile dalle *keywords* inserite nel cluster *IDR* è prerogativa della rivista *XY*, nella quale sono limitate le *keywords* riconducibili al cluster *ICT* e quasi completamente assenti quelle riconducibili al cluster *survey*;

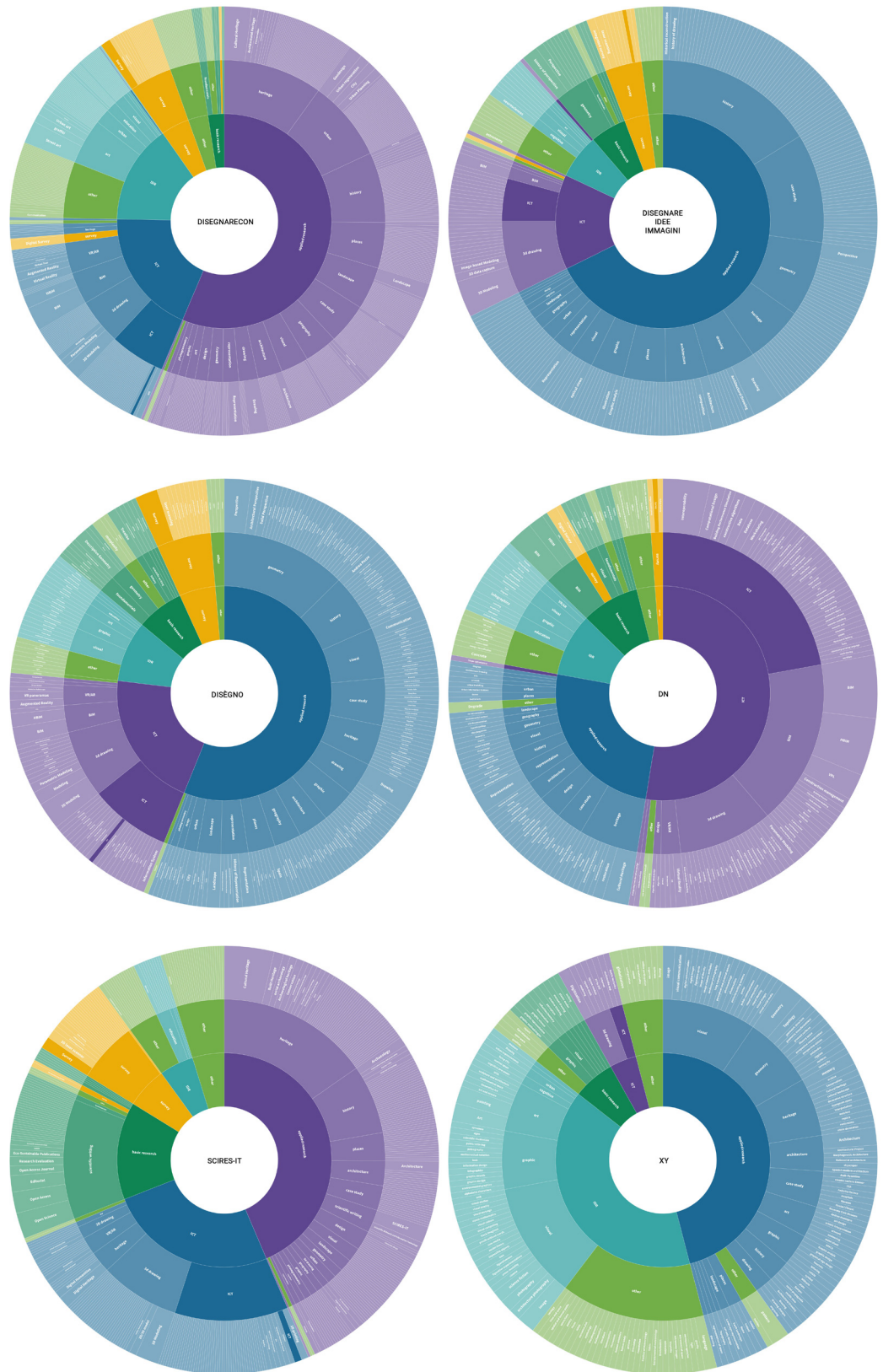


Fig. 6. Distribuzione delle keyword, delle categorie e dei cluster per ogni singola rivista.

5. Per quanto concerne la rivista *Disegnare. Idee Immagini* si può notare un ampio ventaglio di *keywords* riconducibili alla categoria *history* e i numerosi *case studies*. Inoltre, si sottolinea la presenza oltre la media di *keywords* riconducibili alla categoria *geometry*;

6. La categoria *geometry*, a cui vanno ascritte tutte le *keywords* riconducibili allo studio della geometria descrittiva e proiettiva, oltre che su *Disegnare. Idee Immagini* sono presenti in modo significativo su *disegno* e *XY*.

In ultima istanza è possibile concludere che mancando una lunga tradizione bibliometrica nel settore, l'uso delle *keywords* può essere ancora affinato, ad esempio con una scelta più mirata e meno generica rispetto ai contenuti dei singoli paper, che questo studio vuole supportare. Inoltre, anche se con i limiti esposti in precedenza, è stato possibile identificare quantitativamente gli orientamenti delle singole riviste, per consentire agli autori di scegliere in modo più efficiente la rivista in cui sottoporre i propri contributi aumentandone la visibilità e qualificando ulteriormente i contesti editoriali di pubblicazione.

Note

[1] È significativa la recente introduzione del settore ERC SH5_12 *Computational Modelling and Digitization in the Cultural Sphere* che chiarisce definitivamente la collocazione delle attività di ricerca centrali del settore scientifico disciplinare ICAR17 nelle *Human Science*.

[2] *InterDisciplinary Research*.

[3] Cfr. voce *ricerca scientifica*, App. IV, III, p. 217, Enciclopedia Treccani.

Riferimenti bibliografici

Gosnell C. (1944). *Obsolescence of Books in College Libraries*. In *College & Research Libraries*, 5(2), pp. 115-125.

Hulme E.W. (1923). *Statistical Bibliography in Relation to the Growth of Modern Civilization: Two Lectures delivered in the University of Cambridge in May 1922*. In *Nature* 112, pp. 585-586.

Luigini A. (2020). *Ricerca interdisciplinare e ICAR17: una proposta per la definizione di un modello condiviso/Interdisciplinary research and ICAR17: a proposal for the definition of a shared model*. In Arena A. et al. (a cura di). *Connettere. Un disegno per annodare e tessere. Atti del 42° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione*, pp. 567-584. Milano: FrancoAngeli.

Pesta B., Fuerst J., Kirkegaard E. (2018). *Bibliometric Keyword Analysis across Seventeen Years (2000–2016) of Intelligence Articles*. In *Journal of Intelligence* 6(4), 46, pp. 1-12.

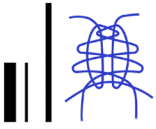
Pritchard A. (1969). *Statistical Bibliography or Bibliometrics?* In *Journal of Documentation* 25, pp. 348-349.

Stokes D. E. (1997). *Pasteur's Quadrant – Basic Science and Technological Innovation*. Washington: Brookings Institution Press.

Autore

Alessandro Luigini, Libera Università di Bolzano, alessandro.luigini@unibz.it

Per citare questo capitolo: Luigini Alessandro (2021). *Riviste scientifiche nel settore ICAR17: analisi quantitativa delle keyword e dei temi di ricerca/Scientific Journals in ICAR17: Quantitative Analysis of Keywords and Research Topics*. In Arena A., Arena M., Medati D., Raffa P. (a cura di). *Connettere. Un disegno per annodare e tessere. Linguaggi Distanze Tecnologie. Atti del 42° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione/Connecting. Drawing for weaving relationship. Languages Distances Technologies. Proceedings of the 42nd International Conference of Representation Disciplines Teachers*. Milano: FrancoAngeli, pp. 879-900.



Scientific Journals in ICAR17: Quantitative Analysis of Keywords and Research Topics

Alessandro Luigini

Abstract

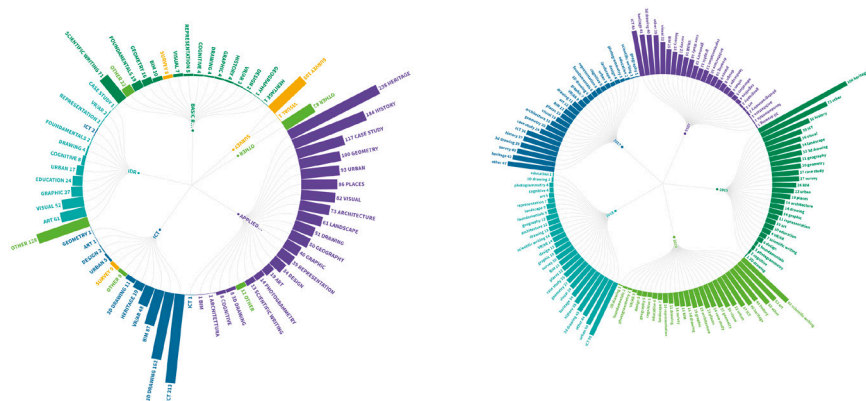
The aim of the paper is to present a quantitative analysis of the frequency, distribution and classification of keywords in articles published in scientific journals related to the disciplinary scientific sector ICAR17 in the period between 2016 and 2020. The analyses are carried out on a dataset based on the keywords and subsequently integrated by assigning the keywords to thematic categories and clusters that define different research areas.

The contribution is mainly of methodological value, in that it shows a way of developing keyword analysis ad hoc for the disciplinary field of Drawing, and is open to further refinement, especially as regards the criteria for determining the categories and clusters, as well as for extending the period of observation.

The quantitative analyses carried out allow us to evaluate the use of keywords for each journal in order to support authors who wish to identify the editorial context in which to publish their research and to identify possible development trends in research in the sector.

Keyword

keyword analysis, research methods, scientific journal, scientific writing.



Distribution of keywords
categories according to
clusters and according to
5 years of observation.

Introduction

Statistical bibliography, or bibliometrics, is an application that has been known since the early decades of the 20th century and was originally applied to a wide variety of contexts: from the study of patent trends in the UK [Hulme 1923] to the obsolescence of publications acquired by libraries [Gosnell 1944], for example.

Bibliometrics aims at “to shed light on the processes of written communication and of the nature and course of development of a discipline [...], by means of counting and analysing the various facets of written communication” or “the assembling and interpretation of statistics relating to books and periodicals [...] to demonstrate historical movements, to determine the national or universal research use of books and journals, and to ascertain in many local situations the general use of books and journals” [Pritchard 1969, p. 348]. The elements that can be used in bibliometric investigation are many, including citations, abstracts or keywords, each of which requires specific methodologies of analysis.

The use of keywords in scientific publishing aims to identify research topics contained in a paper or article and are the basis of metadata for indexing in SCOPUS, WOS and other databases. Keywords are, in a nutshell, the condensation of the topics covered in the publication, and together with the title and abstract serve authors to communicate the distinctive elements of their work concisely and clearly to other researchers. In many contexts of the most recent scientific publishing, it is possible to find vademecums for correct scientific writing, and in many cases (*Elsevier, Springer*, etc.) specific indications are dedicated to keywords: in addition to the basic rules, some of which we will evaluate below, many also offer language editing services, including the choice of keywords, or their optimisation for search engines (SEO, Search Engine Optimisation). This focus proves the relevance of keywords in identifying the publication of a search and, above all, in its accessibility and dissemination. Keyword analysis is frequently used in bibliometric disciplines also for the analysis of research trends, because it is able to synthetically capture the focus of publications in a specific disciplinary area or in a specific publication context. The scientific disciplinary sector ICAR17 Drawing is not included in the bibliometric sectors in the ANVUR classification, but the status quo in national and international scientific journals has rapidly adapted to bibliometric mechanisms and criteria: in particular, scientific journals seem to have largely accepted these criteria, and for this reason it is appropriate to apply research analysis systems typical of bibliometric disciplinary areas, although with the awareness that in a scientific sector whose reasons are to be found in the humanities, bibliometric data alone cannot be exhaustive in reading the status of research.

This paper presents a revision of the analysis model proposed at the UID2020 conference *Conessioni* [Luigini 2020] where the keywords of the previous two annual conferences of the scientific society *Unione Italiana del Disegno* (2018 and 2019) were analysed, with the result of being able to outline some considerations on the diffusion of interdisciplinary research topics. Disciplinary scientific journals will be analysed here, with the aim of a general mapping of research interests in the field and a characterisation of the journals taken into consideration.

Methodology

The selection of journals to be mapped was determined according to three criteria:

1. inclusion in the ANVUR lists as scientific journals or class A journals;
2. direction of a scholar affiliated with the scientific disciplinary sector ICAR17;
3. systematic presence of keywords in published articles.

These criteria led to the selection of six journals:

name	ISSN	ANVUR classification	publisher	Editor-in-chief	open access	URL
Dienne building information modeling, data & semantics	2610-8755	Scientifica	DEI Tipografia del Genio Civile	Tommaso Empler	No	http://www.dienne.org
Disegnare. Idee Immagini	1123-9247 (cartaceo) 2704-7105 (online)	Classe A	Dipartimento di Storia, Disegno e Restauro dell'Architettura / Gangemi editore	Carlo Bianchini	No	https://www.gangemieditore.com/periodici/disegnare-idee-immagini/29
Disegnarecon Scientific Journal on Architecture and Cultural Heritage	1828-5961	Scientifica	Università degli Studi dell'Aquila	Stefano Brusaporci	Si	http://disegnarecon.univaq.it
Diségno Biannual Journal of the UID - Unione Italiana per il Disegno Scientific Society	2533-2899	Scientifica	Unione Italiana per il Disegno	Francesca Fatta	Si	https://disegno.unioneitalianadisegno.it
Scires-IT SCientific RESearch and Information Technology	2239-4303	Classe A	CASPUR-CIBER Publishing	Virginia Valzano, Michela Cigola	Si	http://www.sciresit.it
XY dimensioni del disegno	2499-8338 (cartaceo) 2499-8346 (online)	Scientifica	Università degli Studi di Trento / Officina Edizioni	Roberto de Rubertis	Si	http://rivista.xydigitale.it

Table 1. Index of journals mapped in this paper; listed in alphabetical order

journal	mapped issues	articles	articles with keywords	keywords	% on total
Dienne	7	55	54	253	9,75%
Disegnare Idee Immagini	10	65	65	328	12,64%
Disegnarecon	10	211	208	935	36,04%
Diségno	6	101	72	294	11,33%
Scires-IT	10	112	110	562	21,67%
XY	8	79	78	222	8,56%
TOTAL	51	623	587	2594	100%

Table 2. Overall data set

All issues published by the journals in Table 1 in the time interval 2016-2020 were taken into account, except for *Dienne* and *diségno* which were both started in 2017.

A total of 51 issues and 623 articles were analysed, of which 587 were keyworded. The total number of keywords in the dataset is 2594. For the journals that also publish in Italian (*Disegnare Idee Immagini*, *Diségno* and *XY*), only the keywords in English were taken into account.

As in the previous application to the proceedings of the 2018 and 2019 UID conferences [Luigini 2020] and already consolidated in the literature [Pesta 2018, p. 2], the keywords of all the articles were included in a dataset and were then 'normalized' and 'aggregated': 'normalized', because the data were not sufficiently 'sharp' presenting numerous declinations of particularly popular themes (e.g.: 3D model, 3D models, 3D modeling, etc.), keeping them would have altered the analysis, and 'aggregated' because the great variability of the keywords, even after their normalization, would not have allowed a meaningful reading of the spread of research areas and purely disciplinary themes.

This process, which is central to the setting up of this work, requires further methodological investigation.

Follow-up aggregation into categories and clusters

The data set was organized on three levels: the outermost one (fig. 2) is the level that contains all the normalised keywords; the intermediate level indicates the macro-theme, the category to which each single keyword is established to refer (3D drawing, architecture, art, BIM, case study, cognitive, design, drawing, education, fundamentals, geography, geometry, graphic, heritage, history, ICT, landscape, other, photogrammetry, places, representation, scientific writing, survey, urban, visual, VR/AR); the innermost level instead indicates some clusters (basic research, applied research, survey, ICT, IDR, other), which are assigned to the single keyword and not to the categories. This last clarification implies that two keywords within the same category can belong to two different clusters (e.g.: the keyword visual culture is part of the category visual and of the cluster basic research, the keyword display is also part of the category visual but of the cluster IDR). In contrast to what happened previously, in this dataset the information of the title of the article was also maintained, which allowed for greater precision in the association of the individual keywords to the categories and clusters.

Some other considerations are necessary at this point.

The first is that although the clusters 'basic research' and 'applied research' seem to be able to include the whole dataset, we preferred to identify separately the keywords that can be univocally attributed to three clusters (survey, ICT and IDR) that are particularly significant in the conclusive evaluations and, above all, because in many cases the possible assignment of these keywords to the clusters basic or applied research is uncertain. Moreover, even though it might risk appearing pleonastic, it is also necessary to clearly define what is meant by basic and applied research.

'Basic (or fundamental) research' is an experimental or theoretical activity developed for the purpose of extending knowledge and for which no specific application or direct use is envisaged, whereas 'applied research' is that original research carried out to extend knowledge, but mainly for the purpose of a specific practical application. Furthermore, when Stokes in 1997 proposes a model for the classification of research, he assumes a categorisation of research according to two factors, namely how much research is projected to: 1) expanding knowledge and 2) applying knowledge.

		Relevance for immediate application	
		No	Si
Relevance for the advancement of knowledge	Si	Quadrante di Bohr PURE BASIC RESEARCH	Quadrante di Pasteur USE-INSPIRED BASIC RESEARCH
	No	uninteresting research	Quadrante di Edison PURE APPLIED RESEARCH

Fig. 1. Diagram of the distribution of research according to Stokes (1997).

Stokes uses the names of three great scientists of the past as symbolic examples of the three approaches to research: "Bohr's quadrant" concerns pure basic research, the aim of which is to advance knowledge without any immediate application, "Edison's quadrant" concerns applied research, the aim of which is to apply the knowledge already available, and "Pasteur's quadrant" –Use-Inspired Basic Research– concerns research that aims to advance knowledge with an immediate possibility of application to concrete case studies.

It is easy to understand how it would be possible to determine the difference between a scientific product belonging to “Pasteur’s quadrant” and another belonging to “Edison’s quadrant” only by entering into the content dimension of the article, while from the keywords it is only possible to distinguish, and even in this case with a certain amount of uncertainty, between research of a basic nature (basic research) and research applied to case studies (applied research). This leads to a high number of keywords identified as belonging to applied research, but ideally distributed between pure applied research and applied research promoting an advancement of basic knowledge, i.e. belonging to “Pasteur’s quadrant”. And it is also easy to observe how much of the basic research of the scientific disciplinary sector ICAR17 is carried out precisely in an applied perspective, for example, but not only, in the field of Digital Heritage.

Moreover, it should be noted that the set of clusters is not entirely homogeneous, given that many –if not almost all– of the keywords attributable to applied research can also be associated with other clusters (e.g.: the keyword cultural heritage was included, by analogy with the more generic term heritage, in the category history or applied research, but due to its broad nature it could often have been included among the IDR themes) but for the purposes of our work –a methodological proposal and not an exhaustive formulation– we felt that such a distinction could yield sufficiently significant considerations.

Once defined, the dataset was graphed through the web app Flourish (<https://app.flourish.studio/>) in order to visually observe the incidence of each disciplinary area, of each macro-theme and the frequency of the single keywords.



Fig. 2. Diagram of the complete dataset of standardised, categorised and clustered keywords.

Data Analysis

Figure 2 shows the quantitative distribution of the individual keywords, in the outer ring, where it can be seen that only a few keywords whose frequency is equal to or higher than 10 are readable:

cultural heritage	47		landscape	16
BIM	42		perspective	15
3D modeling	38		virtual reality	14
drawing	36		communication	13
HBIM	28		interoperability	11
representation	25		open science	11
architecture	20		digital survey	10
survey	19		geodesign	10
street art	17			

Table 3. Frequency of keywords with a value equal to or greater than 10

Table 3 shows us that –in contrast to the 2018 and 2019 UID conferences where the most frequent keywords were drawing, survey, and representation– on the scientific journals three most frequent keywords are cultural heritage, BIM and 3D modeling. Considering that drawing and representation are still among the most used, with 28 repetitions we find HBIM, a keyword that identifies the application of BIM technology to heritage.

In the second column we find three keywords not present, if not occasionally, in the 2018 and 2019 UID conferences: street art, open science and geodesign. The reason for the high frequency of these three keywords is the presence of some monographic issues, outside of which these themes are not observable: street art and geodesign come from two issues of *Disegnarecon*, respectively vol. 11, n. 20 Co-Design: digital tools for knowledge-building and decision-making and vol. 13, n. 24 Street art. Drawing on the walls, while open science from the special issue 2020 of *Scires-IT* Open Science: New models of scientific communication and research evaluation. This phenomenon is clearly all the more significant the smaller the sample: 5 years of publications for a total of 51 issues is a reliable sample, but the presence of only one monographic issue is still capable of significantly shifting the results.

Interoperability also has a specific motivation, which can be traced by looking at the location of the keywords within the journals: 10 of the 11 occurrences are in fact contained in various issues of the journal *Dienne* which, as we shall see below, is the publication context with the highest number of keywords referring to the ICT cluster. The remaining ones (perspective, Virtual Reality, communication and digital survey) are substantially distributed over all the journals.

Keywords Quantity	Frequency	keywords Quantity	Frequency	keywords Quantity	Frequency	keywords Quantity	Frequency
1	47	1	20	1	13	6	7
1	42	1	19	1	12	13	6
1	38	1	17	2	11	10	5
1	36	1	16	2	10	26	4
1	28	1	15	4	9	42	3
1	25	1	14	8	8	138	2

Table 4. Keyword quantities for each individual frequency value above.

By analysing the frequencies it is also useful to understand how many keywords can be considered significant because they recur, and Table 4 lists all the frequencies above 1 with an indication of how many keywords present a specific frequency: it can be seen that the sum of the keywords with a frequency equal to or greater than 10, shown in Table 3, is 384, while the sum of the keywords with a frequency between 2 and 9 is 776, and the keywords with a single recurrence are 1434.

Considering the numerous declinations of keywords referable to similar themes (e.g. survey, 3d survey, integrated survey, digital survey, architectural survey, etc.), an analysis of the frequency of single words within the dataset was carried out, as shown in the table below:

heritage / HBIM	145	representation	64
model	118	image / visual	58
architecture	114	virtual	51
city/urban	104	data	50
survey	98	landscape	47
3D	97	cartography / map / territory	36
graphic	95	archaeology	33
BIM / HBIM	77	geometry	25
drawing	69		

Table 5. Frequency of individual words in keywords with a frequency of 25 or more

The word heritage is again the most frequent, followed by the word model, which mainly refers to three-dimensional modelling rather than to the “model” in geometric or conceptual terms, and which could therefore be combined with the term 3D, marking itself out as the most extensive research topic.

These are followed by two clearly central areas of application of research in the Drawing sector: architecture and city/urban. The term survey, with all its possible declinations, stops just short of 100 occurrences but certainly has a close correlation with both heritage and architecture and city/urban. Of some significance is the frequency of the word graphic, also distributed among all the magazines and throughout the time period observed, used as much as Survey and 3D and more than BIM / HBIM, Drawing or Representation.

Following the analyses on the individual keywords and the words composing the keywords, we moved on to an analysis of the frequencies of the categories, with an annual distribution, as shown in figure 3:

In this graph, we observe the fluctuating but essentially constant distribution of the categories over the five years observed, with the obvious contributions of some monographic issues affecting the overall performance of the six scientific journals at specific times.

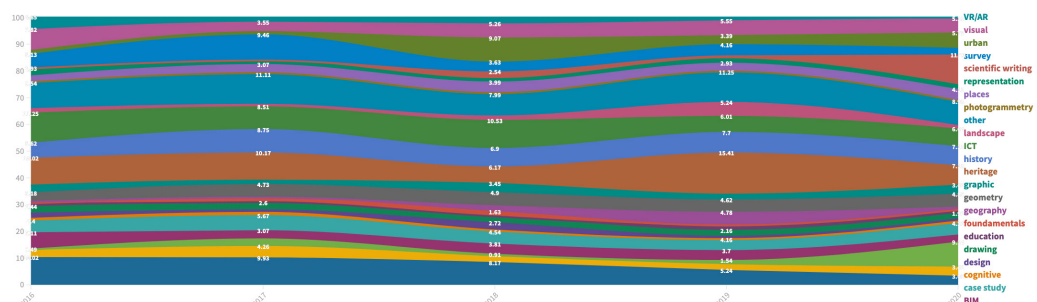


Fig. 3. Time distribution of the frequency of keywords associated with categories.

Figure 4 shows the detail of the temporal distribution of each single category and one can clearly read the contributions of single monographic issues, in correspondence of the cusps and the maximum values in the single diagrams.

Figure 5 shows instead the distribution in absolute terms of the keywords organised in clusters, for each single journal and it is also evident the contribution that each journal gave for the definition of the dataset, since the size of the bar on which the name of the journal is reported on the far right of the graph, is proportional to the number of keywords recorded in the dataset traceable in the last column of table 2.

In figure 6 it is possible to observe the distribution characteristics of the keywords, categories and clusters of each single journal.

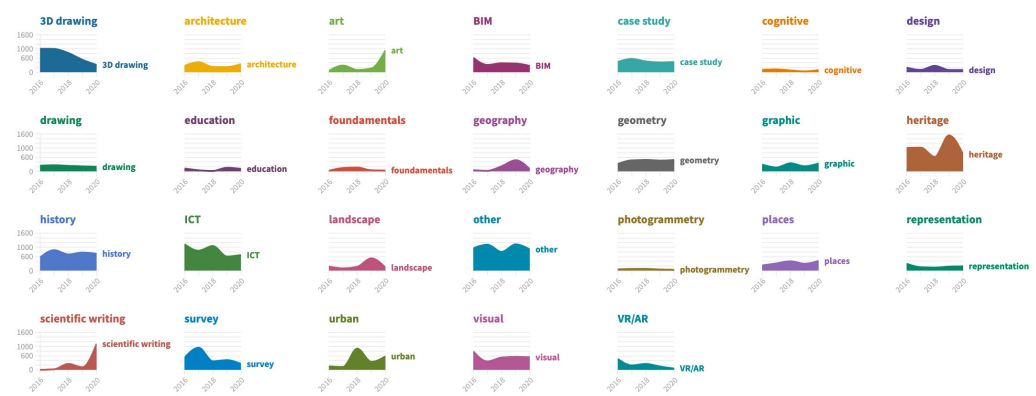


Fig. 4. Time distribution of the frequency of keywords associated with the individual categories.

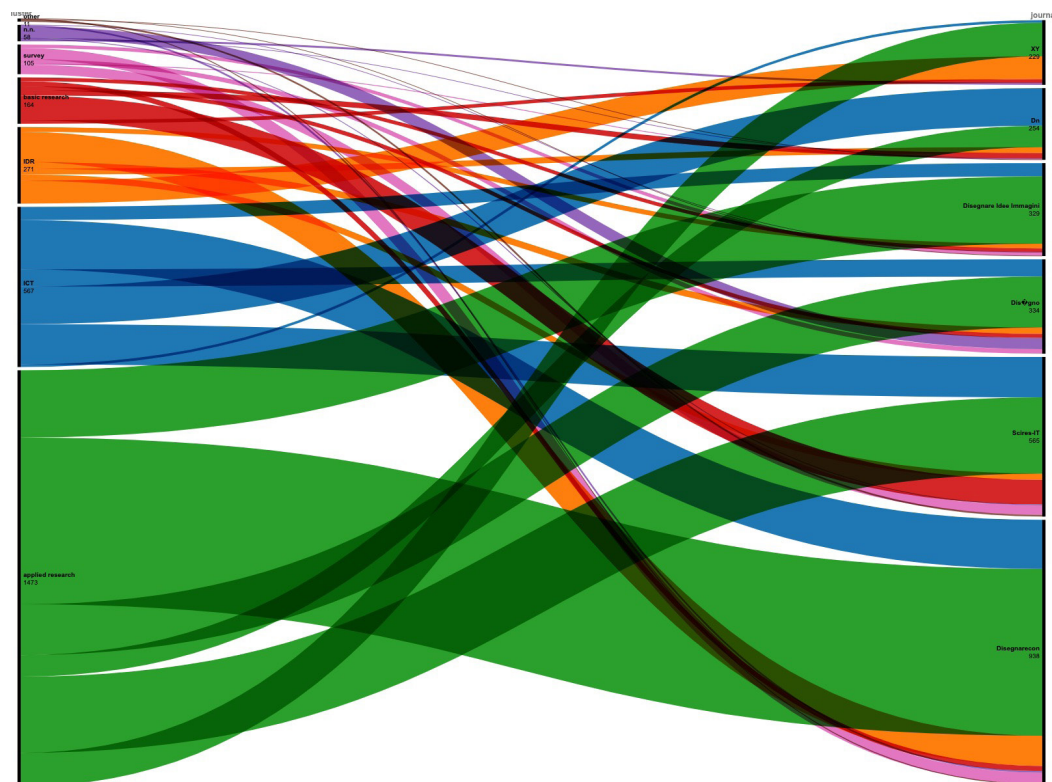


Fig. 5. Distribution of the categories belonging to the six clusters according to journals.

Conclusions

We defined what keywords are in scientific writing, and we determined the differences between basic and applied research, adding a research model - 'used-inspired basic research' - which is probably more suitable to indicate a significant part of the research carried out in the scientific disciplinary field of Drawing. After selecting the journals according to the three criteria listed above, we filed them and collected all the keywords contained in the articles from the five-year period 2016-2020.

Tables 3 and 5 show that the most frequent keywords indicate that the focus of the discipline, at least in quantitative terms, is the three-dimensional modelling of heritage, with particular attention to the use of Building Information Modeling, and that by its nature heritage research has an applied character; although it is possible to assume that some of the research indicated here as applied research may refer to "Pascal's quadrant", that is, to a basic research model developed on an application case.

The high number of keywords with a single occurrence (1434, i.e. 54% of the total) can be attributed to the fact that, as we have already mentioned, the applied character of a large part of the research in the field entails the presence of numerous keywords linked to the determination of places, authors or names of monuments in individual case studies.

We know that keywords are used by the author to make his work more visible in databases, and that it is therefore necessary to adopt appropriate strategies to ensure that the paper does not get lost among a thousand others. In the *vademecum* for scientific writing, where there are specific indications on the choice of keywords, it is often advised to choose specific keywords, with a syntax intermediate between the single word and natural language, and, above all, to avoid excessively generic keywords (called poor keywords because they characterise the paper too marginally). In this sense, considering that the journals analysed all have the field of drawing and graphic and visual representation as their domain, it can be concluded that the keywords drawing, representation, or survey, although central to the sector, are to be considered poor keywords, as are architecture, digital survey (a compound but redundant keyword), communication, etc.

These six keywords, which are the poor keywords with a frequency of 10 or more, represent about 5% of the total keywords, and about 11% of the characterising keywords (i.e. with a frequency of more than 1). The researcher who is looking, for example, for the digital reconstruction of an unbuilt project, is unlikely to use the keywords drawing, representation or architecture, but will probably use, in addition to the name of the project and the author, keywords such as unbuilt project, digital reconstruction or, more generally, 3D modeling.

Finally, the direction of the use of keywords in the six catalogued journals can be assessed, with the clarification that the following observations cannot be extended to the journal as a whole since the use of keywords is a significant but partial parameter of the contents of a journal.

In this context it is possible to compare the overall keywords distribution with the distribution of keywords in each journal and to point out that:

1. The journals that have a percentage distribution of clusters most similar to the sector as a whole are *diségno* - as was easy to assume as the journal of the scientific society UID - and *Disegnarecon*;
2. *Scires-IT* also does not deviate much from the distribution of clusters in the entire sector, except for an accentuation of the ICT cluster - the journal's main focus - and the basic research cluster, fuelled by the 2020 monographic issue on scientific writing and research evaluation;
3. The ICT cluster is preponderant only in the journal *Dienne*, where it exceeds 50% of the keywords, thus marking itself out as the main editorial context for research in this field;
4. The interdisciplinary research identifiable by the keywords included in the IDR cluster is the prerogative of the journal *XY*, in which the keywords attributable to the ICT cluster are limited and those attributable to the survey cluster almost completely absent;
5. As far as the magazine *Disegnare Idee e Immagini* is concerned, there is a wide range of

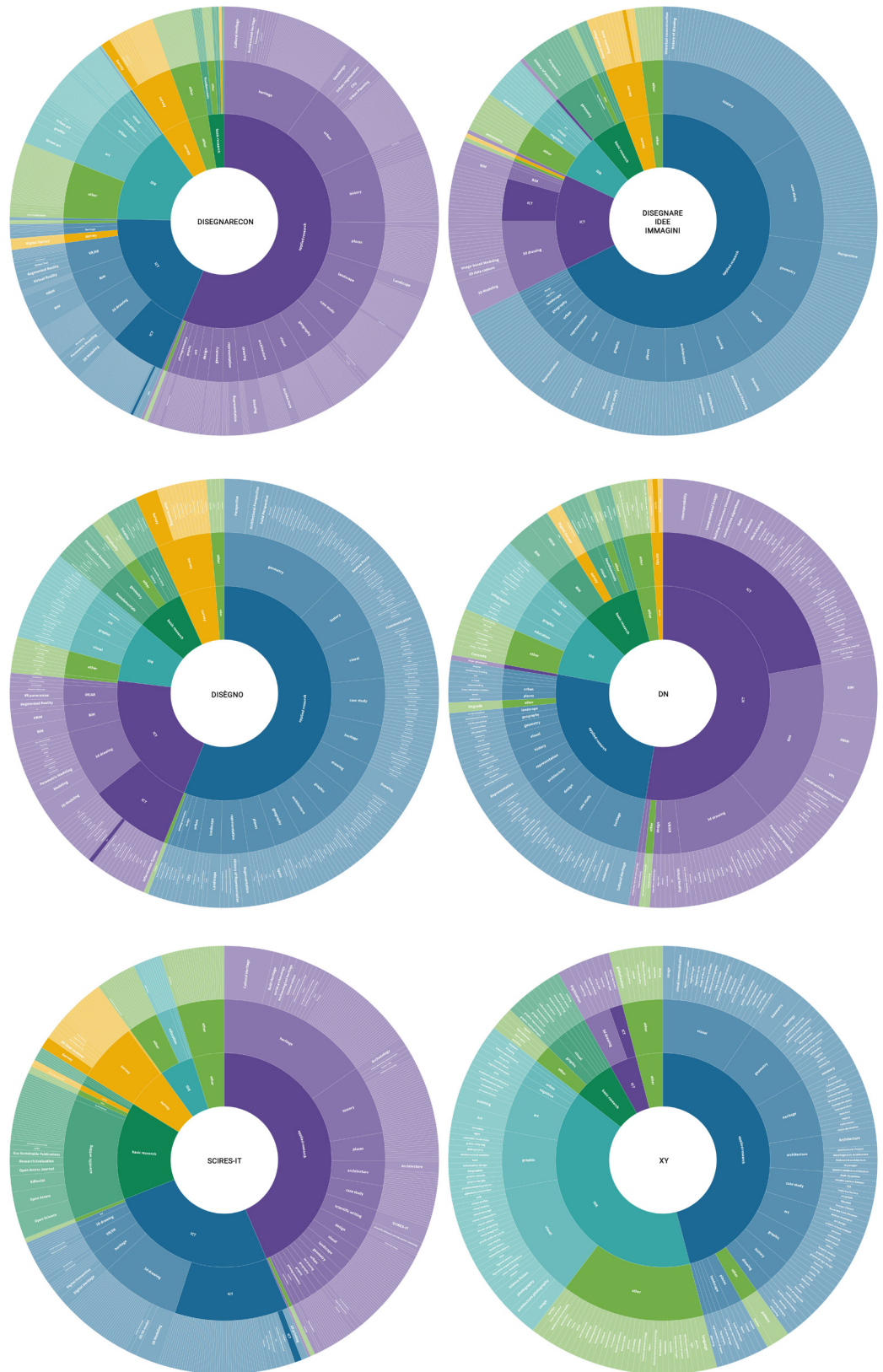


Fig. 6. Distribution of keywords, categories and clusters for each journal separately.

keywords from the history category and numerous case studies. In addition, the above-average presence of keywords in the geometry category is also noteworthy;

6. The geometry category, which includes all the keywords related to the study of descriptive and projective geometry, is present not only in *Disegnare Idee Immagini* but also significantly in *Diségno* and *XY*.

Ultimately, it can be concluded that in the lack of a long bibliometric tradition in the field, the use of keywords can still be refined, e.g. by a more targeted and less generic choice with respect to the content of individual papers, which this study aims to support. In addition, although with the limitations outlined above, it was possible to quantitatively identify the orientations of individual journals, in order to allow authors to more efficiently choose the journal in which to submit their contributions, increasing their visibility and further qualifying the editorial contexts of publication.

References

- Gosnell C. (1944). Obsolescence of Books in College Libraries. In *College & Research Libraries*, 5(2), pp. 115-125.
- Hulme E. W. (1923). Statistical Bibliography in Relation to the Growth of Modern Civilization: Two Lectures delivered in the University of Cambridge in May 1922. In *Nature* 112, pp. 585-586.
- Luigini A. (2020). Ricerca interdisciplinare e ICAR17: una proposta per la definizione di un modello condiviso/Interdisciplinary research and ICAR17: a proposal for the definition of a shared model. In Arena A. et al. (a cura di). *Connettere. Un disegno per annodare e tessere. Atti del 42° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione*, pp. 567-584. Milano: FrancoAngeli.
- Pesta B., Fuerst J., Kirkegaard E. (2018). Bibliometric Keyword Analysis across Seventeen Years (2000–2016) of Intelligence Articles. In *Journal of Intelligence* 6(4), 46, pp. 1-12.
- Pritchard A. (1969). Statistical Bibliography or Bibliometrics? In *Journal of Documentation* 25, pp. 348-349.
- Stokes D. E. (1997). *Pasteur's Quadrant – Basic Science and Technological Innovation*. Washington: Brookings Institution Press.

Author

Alessandro Luigini, Libera Università di Bolzano, alessandro.luigini@unibz.it

To cite this chapter: Luigini Alessandro (2021). Riviste scientifiche nel settore ICAR17: analisi quantitativa delle keyword e dei temi di ricerca/Scientific Journals in ICAR17: Quantitative Analysis of Keywords and Research Topics. In Arena A., Arena M., Mediatì D., Raffa P. (a cura di). *Connettere. Un disegno per annodare e tessere. Linguaggi Distanze Tecnologie. Atti del 42° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione/Connecting. Drawing for weaving relationship. Languages Distances Technologies. Proceedings of the 42nd International Conference of Representation Disciplines Teachers*. Milano: FrancoAngeli, pp. 879-900.