Annals of Library and Information Studies
Vol. 70, March 2023, pp. 1-9
DOI: 10.56042/alis.v70i1.68713

A study of 'calf-path' in file naming in institutional repositories in India

Husna Jabeen[a] and N. S. Harinarayana[b]

[a]Research Candidate, Department of Studies in Library and Information Science, University of Mysore, Mysuru,
E-mail: husnalisc@gmail.com

[b]Professor and Chairman, Department of Studies in Library and Information Science, University of Mysore, Mysuru,
E-mail: harinarayana@lisc.uni-mysore.ac.in

Received: 15 November 2022; *accepted: 19 March 2023*

This study examined the file naming practices in 39 institutional repositories. There is evidence that calf-path exists in file naming among the curators of institutional repositories in India. The study showed that no standard or logic seems to have been followed by repositories in the naming of the files, except by the National Digital Library of India (NDLI) and the CSIR-National Institute of Science Communication and Policy Research (CSIR-NIScPR). The study also examined the composition of the filenames, which shows that the author names (11.2%), titles (11.9%), journal title with volume and issue numbers (21.1%) form the basis for the formation of filenames. It is suggested that digital repository managers have to give more attention to name files in the institutional repository in the interest of uniformity and consistency.

Keywords: Filenames, Metadata, Information retrieval, Information management, Institutional repositories

Introduction

Discoverability and retrieval of files are crucial in a data-driven environment. Specific nomenclature may play a deciding role in discovering files. Hundreds of thousands of files are created and made globally accessible by top-ranked journals and other scholarly outlets. Academic institutions also engage in these activities through their institutional repositories. The academic community as a whole creates a large number of resources (files).

A filename is a component of a file's metadata that helps retrieve it. File names should be simple to recall, short and straightforward to express (e.g. mnemonic)[1]. A filename is usually perceived to indicate to humans what a file contains and how it relates to other files. In this context, it is natural to expect an efficient file naming practice to be used. However, it is a matter of general observation that file naming convention is one of the neglected areas of data management. This study examines the conventions among the curators of institutional repositories.

It is generally observed that the elements of metadata components, such as author, article, journal name, imprint information, etc., are used in forming the filenames. The filenames need to be studied for a better understanding of their formation.

The study primarily aims at studying the 'calf-path' practice that might exist in file naming. The term 'calf-path' relates to that of a wobbling calf that makes its route back home from the forests and which is followed by others. An organization's early digitization efforts may leave a legacy, i.e., calf-path, in organizing digital collections. In such a situation calf route evolves in unexpected ways[2]. The absence of the file naming standard leads to the 'calf-path'. The file naming convention evolved out of the convenience of people managing the resources. From this scenario, this study assumes importance.

A well-managed repository would adopt consistent and descriptive guidelines in all activities, including file naming. The present work aims to study the filenames assigned for scholarly resources in the repositories. It also examines consistency, strategies, and naming practices/conventions of filenames.

Review of literature

To understand the file naming practices of organizations that employed different systems, such as those used for DOS directories, personal computers and institutional repositories, Chapman[3] surveyed file naming habits of people who use and do not use personal computers. There exist methods of constructing a system of categories to organize the various conceptual and structural patterns used by participants to create their filenames. These include truncation, all full words, all abbreviations, all lowercase, all capitals and lowercase, spaces, packing, blending and letter doubling.

Carroll[4] examined more than 2500 filenames by 22 professional computer users at a scientific research centre. In a study, participants were instructed to construct lists of filenames from their folders, annotate each filename with a brief explanation, and submit the lists electronically to the researcher. Each participant created the list independently using their computers without any time constraints. A total of 2507 filenames were investigated according to rule schemes that are recurrent patterns of structural arrangements. The framework is based on the relationship between the three sub-processes of naming, abbreviating and describing. The use of parts of speech and abbreviations was also examined.

An observation that Carroll[5] repeatedly made was the deciding factor in whether to research file naming by PC users qualitatively or quantitatively. A brief review of the findings from the numerous other studies on naming conducted by many researchers[6-9] assume that the larger context in which the act of naming is performed may have a significant bearing on the naming strategies that are most frequently found in the findings.

Halbert, Skinner, and Mcmillan[2] describe a phenomenon known as calf-path syndrome this phenomenon arises whenever an organization undertakes early digitization initiatives that are harmful to the preservation preparedness of their expanding digital holdings. The study by these authors offered ideas and prescriptions for the future successful implementation of a distributed digital preservation system.

Rogers[10] investigated a set of guidelines for file naming conventions used by various institutions digital collections. The way to discover what is included within the millions of files is by their names. Special characters (%,*, %, #, @ etc.) are not allowed in filenames in any library. The recommended practices for file naming are; spaces should be removed and replaced with an underscore or dash in a filename. The length of the characters should be 34 instead of 276, as permitted by Hamilton college library.

Methodology

A list of Indian institutional repositories was sampled from the ROAR repository. Incidentally, it was found that DSpace and Eprints are the most often used software to maintain the repositories. The ROAR repository has registered 114 repositories. Only 33 of them were accessible online. Furthermore, in addition to those listed in ROAR, six additional repositories were incorporated into the study. The researchers considered these repositories to be well-known and anticipate that they will be added to the ROAR repository in the future. Consequently, the overall number of repositories examined was 39, as indicated in the Appendix I. The inclusion of these additional repositories helped to broaden the scope of the research and provided a more comprehensive understanding of the topic being investigated. It also allowed for a more diverse sample, which increased the reliability and generalizability of the study's findings.

Each of these accessible repositories was visited and the data was collected. The repositories were manually searched subject-wise, and only the first fifty files that gave full-text access to their collection were selected for the study. The filenames assigned to the documents or publications indexed in these repositories were examined for their accessibility characteristics. File name characteristics studied in this work include the filename length, author name, title, journal and conference names, year of publications, institutions' names, keywords and a few other elements.

Data analysis

The file name is characterized by its length and the characters used. The filename has two parts: a base name, including the path in the file system and an extension or suffix. A base name consists of characters and additional characters such as underscore, dashes, hyphens, etc. Three/four-character extensions indicate the file format, e.g., pdf, html, doc, txt. Long filenames are used to make the filenames more descriptive and content indicative. Generally speaking, the institutions do not insist on any limitations on the filename length. The length of the files, however, is limited by the file systems used. For instance, a FAT (File Allocation Table) fixes the maximum path to nine characters, six character base names, and three character file extensions. Windows 95 uses the shorter '8.3' format filename. On the other hand, VFAT (Virtual File Allocation Table) allows filenames with a maximum of 255 characters.

A filename plays an essential role in an information retrieval system. An efficient retrieval system is made possible by filenames. It is essential to have a consistent naming convention when creating digital files since, without a filename, it is impossible to find the files in a search.

Length of the filenames

The study analysed 3121 filenames from 39 institutional repositories. The length of the filenames

was calculated using the "len function" in Microsoft Excel. The length was calculated, including the characters used for file extension. From each repository, 50 resources were selected. By and large, each resource in these repositories was found to be stored in a separate single file. However, there were instances of a resource stored in multiple related files. For example, a thesis was represented as a collection of chapters and each chapter was stored in separate files. In other words, one resource was represented by multiple files in the computer system. All such files were considered as separate units in this study.

Table 1 resents the number of files and the length of the filenames. Database managers have assigned 30 characters or less to 83% of files. It was also observed that there is a habit of giving unduly long filenames with more than 100 characters. The longest filename with 162 characters (including spaces) recorded in the study is 'Sulfonoquinovosyl diacyl glyceride selectively targets acute lymphoblastic leukaemia cells and exerts potent anti-leukemic effects in vivo _ Scientific Reports.html'. This filename was given to one of the resources in the institutional repository maintained by the Indian Institute of Chemical Biology (IICB), Kolkata.

The study also recorded a few instances of short file names. Thirty-seven files with just one character in their base name, 35 files with just two characters and 58 files with three characters were found in the study. Examples of such files found in the digital institutional repositories were 'a.pdf', '9.pdf', '46.pdf', 'a2.pdf', 'EA2.pdf', 'ch1.pdf', and so on.

Even in the instances where the filenames longer, it was found to be not-so-useful names in terms of understanding the content of the files. This shows that there is a need to evolve some good practices to give meaningful filenames to digital resources.

Table 1 — Length of the filenames

Length	Number of files	Percentage
1-15	1337	42.84
16-30	1253	40.15
31-45	272	8.72
46-60	124	3.97
61-75	73	2.34
76-90	30	0.96
91-105	15	0.48
106-120	6	0.19
121-135	6	0.19
136-150	1	0.03
151-165	4	0.13
Total	3121	100%

Elements of filenames

Filenames are constructed in many ways. Table 2 gives an idea of the basis used in the institutional repositories in naming their resources

Often, the authors, title, journal title, volume/issue number, and so on are used as part of filenames. The usage percentages indicate that there is no significant preference or pattern found, and other miscellaneous terms/abbreviations used in the filenames.

Table 3 shows the use of the name of the author(s) in filenames. The use of complete names is in vogue in some digital repositories but in different ways. For example, the filename 'sandip_madal.pdf' was used as the filename for one of its digital resources in the NIT Rourkela repository. There were filenames formed out of the full name of their first author (3.00%). Some filenames (2.1%) have been formed using the inverted format, i.e., surname followed by forename/initials. For example, the filename from CUSAT reads as 'PREVENTIVE DETENTION AND THE LAW Ajith Kumar J…PDF'. The other practices of using the name of the authors include using only the forename (2.14%), surname (2.00%), and so on.

The first author's names are used in forming the filenames. However, subsequent authors (second, third, and so on) were also used in a few instances. Table 3 depicts a list of ways the authors' names were used to form filenames for digital resources.

It is seen from Table 3 that no consistency was observed while using the authors' names in forming filenames either within the institutions or across them. A lack of policy decisions on the part of the institutions could be the reason for such inconsistencies.

Of the 3121 files, around 12% files used the title as a part of their filenames. Out of which, 4.5% of files

Table 2 — The elements in the names of the files

Sl. No.	Features	Occurrences	Percentage N=3121
1	Author name	349	11.2
2	Title	370	11.9
4	Journal title with volume and issue numbers	660	21.1
5	Convention/Conference name	74	2.4
6	Publication Year	286	9.2
7	Institution name	302	9.6
8	keyword and Conjunctions/Prepositions/Interjections in filenames	480	15.4
9	Miscellaneous in filenames	600	19.2
	Total	3121	100%

use the main title (excluding subtitles) as filenames. The study found more than 50 filenames (1.2%) that adopted partial titles drawn from titles and subtitles in the filenames (Table 4).

Figure 1 presents the different ways the journal names are used as part of filenames. The use of the abbreviation of the journals as part of the filenames was found in (4.8%) of cases. In the remaining cases, either the full name or the partial/truncated names of the journals were used. Surprisingly, adding the volume and issue number was a practice used in naming the files. In the present study, 331 files were found to have adopted this practice. Of the two components, using volume number is predominant (6.6%). These components, volume and issue numbers, help to individualize digital resources. For example, the filename 'IJFTR 14(3) 138-140.pdf' observed in the study individualizes the resource. Journal names alone were not frequently used in filenames to avoid redundancy, which is not allowed in computer file naming systems.

The names of the conventions/conferences were also used as part of the filenames. The file naming practice observed shows that 1.31% of names contained truncated names, whereas 0.8% contained the abbreviation. However, no standard pattern was found in shortening the names of conventions/ conferences in constructing the filenames within the institutions or across them (Table 5).

The date/year of publication is used in filenames to indicate the currency of the resource. Of the three components of a date, the year of publication is predominately used (7.5%) as part of filenames. Months were used in numerical (e.g., 05, 06) and textual formats

Table 3 — Author name in filenames

Sl. No.	Features	Occurrences	Percentage N=3121
1.	Full name of the author (as it is)	93	3.00
2.	Only the forename of the author	67	2.14
3.	Full name of the author in inverted format	63	2.01
4.	Only the surname of the author	62	2.00
5.	Subsequent author initials only	11	0.35
6.	Subsequent author - full name	10	0.32
7.	Initials only	9	0.3
8.	Subsequent author - forename	9	0.3
9.	Truncated name of the author	8	0.25
10.	Subsequent author - surname only	8	0.25
11.	Forename of two authors (first and subsequent author name)	5	0.16
12.	Surname of two authors (first and subsequent author name)	2	0.06
13.	Forename and surname (first and subsequent authors)	1	0.03
14.	Other names	1	0.03
	Total	349	11.2%

Table 4 — Title in filenames

Sl. No.	Convention and Conference names	Occurrences	Percentage N=3121
1.	Partial title	192	6.2
2.	Full title (without subtitle)	142	4.5
3.	Title with subtitle	36	1.2
	Total	370	11.9%

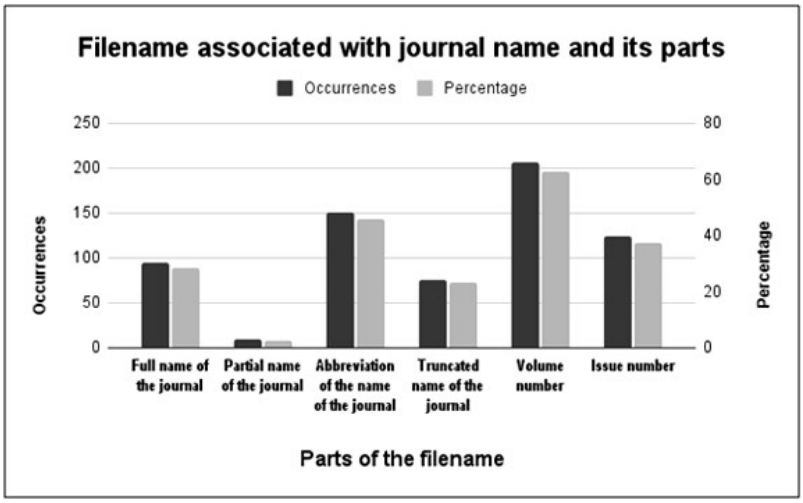

Fig. 1 — File names associated with journal names and their parts

Table 5 — Convention and conference name in filenames

Sl. No.	Convention and Conference names	Occurrences	Percentage N=3121
1.	Truncated name of the convention/conference	41	1.31
2.	Abbreviation of the name of the convention/conference	24	0.8
3.	Full name of the convention/conference	9	0.3
	Total	74	2.4%

Table 6 — Year of publication in filenames

Sl. No.	Imprint Information	Occurrences	Percentage N=3121
1.	Year of publication	233	7.5
2.	Truncated year of publication	37	1.2
3.	Month	10	0.3
4.	Date of publication	6	0.2
	Total	286	9.2%

(e.g., May, June). Sometimes (0.3%), truncated years were represented in the filenames as two-digit numbers (e.g., chacal95.pdf). Surprisingly, the full date of publication was rarely found (0.2%) (Table 6).

It is not unusual for filenames to include the name of the institution. The majority of the times (8.0%), institutions' names were represented in the filenames by their abbreviation/acronyms such as NDL, HBNI, GIPE, NAARM, etc. Using the institutions' acronym as the main part in the filename is a short-sighted approach. Giving merely the institution's name does not serve the purpose of indicating the contents in the files. Even the institution's location was used as a part of a filename in 1.6% of cases. No calf path was found in these filenames while using the institutions' names. For instance, NAARM 10.pdf, NAARM 24.pdf, NAARM 24.pdf, NAARM 4.pdf, NAARM 3.pdf, NAARM 2.pdf (Table 7).

Subject keywords and conjunctions/prepositions/ interjections for example, agriculture.pdf, Afforestation.pdf, Accounting.pdf were found in filenames, as shown in (Table 8), were also considered by the managers of institutional repositories as a part of filenames in some instances (15.2%). Keywords were used in filenames in different forms – full keywords (1.8%), abbreviated format (1.2%) and truncated form (0.4%). This study could not ascertain the basis for the selection of keywords by the managers. Conjunctions/prepositions were also seen in the filenames, 376 (11.8%) conjunctions/prepositions were recorded in the study. A majority of the conjunctions/ prepositions were found in

Table 7 — Institution names in filenames

Sl. No.	Institutional names	Occurrences	Percentage N=3121
1.	Abbreviation of institution's name	250	8.0
2.	Location of institution's name	51	1.6
3.	Abbreviations of divisions	1	0.03
	Total	302	9.63%

Table 8 — Keyword and conjunction/preposition/interjections in filenames

Sl. No.	Keywords assigned in file names	Occurrences	Percentage N=3121
1.	Conjunctions/Prepositions /Interjections	376	11.8
2.	Keyword	56	1.8
3.	Abbreviated keyword	36	1.2
4.	Truncated keyword	12	0.4
	Total	480	15.2%

Table 9 — Miscellaneous in filenames

Sl. No.	Other Words used in filenames	Occurrences	Percentage N=3121
1.	First - last page	179	5.7
2.	Resource type name	132	4.2
3.	Truncation of resource type name	80	2.6
4.	First page	62	1.9
5.	Repetition	46	1.5
6.	Software name	25	0.8
7.	Course name	25	0.8
8.	Spelling mistake	19	0.6
9.	Irrelevant filename	14	0.4
10.	Salutation words	7	0.22
11.	File extension in the filename	5	0.2
12.	Incorrect filename	4	0.12
13.	Zeros(00)	2	0.06
	Total	600	19.2%

title-based filenames. In many instances, files were named with the full title of the document, e.g., 'Seasonal investigation on prediction accuracy of atmospheric turbulence.pdf', 'On the development of abnormally large postsunset upward_____.pdf'. This kind of filename undoubtedly creates noise in retrieving documents.

Table 9 shows that miscellaneous that can be found in filenames. The published source's starting and last pages were found in 179 files. Resource types of the documents, such as articles, papers, conference, and thesis, were also used as a part of the filenames. 4.2% of filenames contained resource names, for example: 'ARINT Article November 2013. pdf was found in

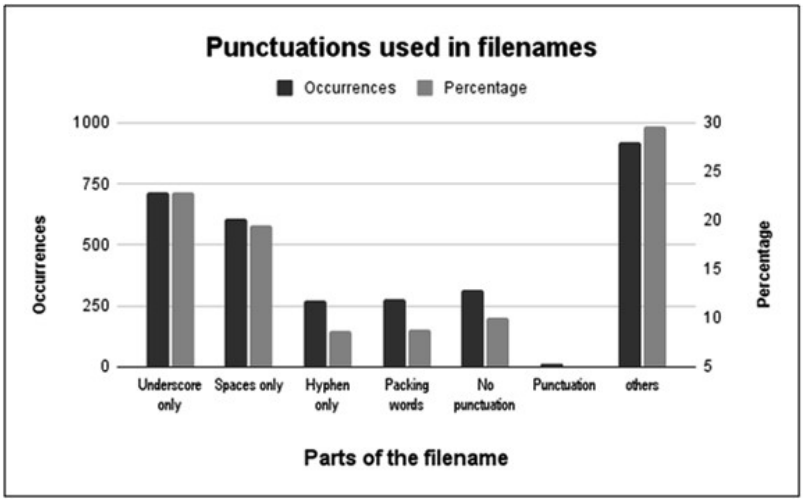

Fig. 2 — Punctuations used in filenames

Table 10 — Words, Numbers and Letters in filenames

Sl. No.	Words, Numbers and Letters in filenames	Occurrences	Percentage N=3121
1.	Words and numbers	1762	56.4
2.	Only words	701	22.5
3.	Letters and numbers	440	14.1
4.	Numbers only	205	6.6
5.	Letters only	13	0.4
	Total	3121	100%

Table 11 — Full words, Single or multiple words, numbers and letters in filenames

Sl. No.	Word and letters in File names	Occurrences	Percentage N=3121
1.	Single word	1745	55.9
2.	Multi words	707	22.6
3.	Two words	669	21.4
	Total	3121	100%

the repository. Further, truncation of resource type names was observed in 2.6% of filenames. E.g. 'art27.pdf' and 'JPS-v20-art11.pdf' are seen in the repositories. Course name and software name were also used in 0.8% of filenames. The other cases include filenames with starting zeros and salutations words (Dr., Prof., etc.). Typos were also noticed in filenames in some cases. Some of the filenames were irrelevant and inappropriate. The page numbers were also used as part of the filenames.

Figure 2 presents the use of punctuation in filenames. Almost 23% of files have used the underscore '(_)' followed by only space, which was used in 19.48% of the total files. Other types of punctuation used while naming the files can be found in Figure 2. Most of the files have used different punctuation marks while assigning filenames for documents. A few notable examples were found in the filenames, such as '2013(4.2-13).pdf' and 'Arrears in Courts: Measures to Contain them.PDF'. It is not uncommon to see the underscore as a part of filenames. The use of underscore in filenames is one of the calf paths that resulted from earlier operating systems which did not allow spaces to be used as a part of the filenames. It could also be attributed to being influenced by the general practice of using the underscore '(_)' extensively used for email or other environments. From one of the repositories, the unexpected component was found in filenames. The filenames read 'art%3A10.1208%2Fs12249-016-0652-6.pdf', '(589326778) NANOTECHNOLOGY IN RICE PRODUCTION SYSTEMS 17.7.08.1.pdf' in one repository, and 'Synthesis, spec ... 4)-phenylsemicarbazone.pdf' in another.

Table 10 indicates the use of only words, words in combination with numbers or words and letters in filenames. Only words were used in 22.5% of the total documents as filenames. A combination of words and numbers was found in 56.4%. In comparison, 14.1% of filenames consisted of alphabets and numbers. Only numbers appeared in 6.6% of filenames. Words in combination with numbers were predominantly used for assigning filenames. It is also a general practice that words and numbers are frequently used while assigning a filename for a resource.

Table 11 shows that single words were used in more than 50% of filenames. In 22.6% of filenames, there were three or more words, compared to two words in 21.4% of filenames. This table shows that single-word filenames were used more frequently than those with two or more words. Both short and long filenames have their advantages. While the short

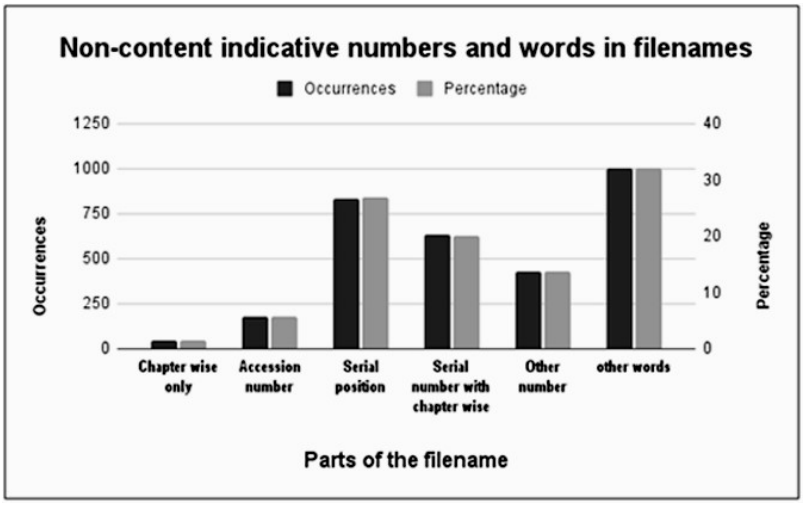

Fig. 3 — Non-content indicative numbers and words in filenames

Table 12 — File extension in filename

Sl. No.	File Extensions	Occurrences	Percentage N=3121
1	.pdf	2976	95.4
2	.html	101	3.2
3	.doc	30	0.96
4	.txt	13	0.41
5	.jpg	1	0.03
	Total	3121	100%

Table 13 — Sentence case in filenames

Sl. No.	Sentence case in filenames	Occurrences	Percentage N=3121
1.	Alpha-numerical	2219	71.09
2.	Capitals & lowercase letters	453	14.51
3.	Numbers	204	6.53
4.	All lowercase letters	177	5.7
5.	All capitals letters	67	2.14
6.	Camel Case	1	0.03
	Total	3121	100%

filenames facilitate easy retrieval, the long filenames are content indicative. It is noticed that too-long filenames make it difficult to retrieve files. In order to make them easier to find, filenames should be brief, straight forward, and meaningful.

Table 12 indicates the filename extension used. PDF file extension was found in more than 95% of documents. Other types of file extensions can be seen in (Table 12). The study revealed that the Indian repositories are only limited to depositing research publications in PDF format. This format is another calf path revealed in the study.

The 'sentence case' of filenames is shown in (Table 13). Alpha-numerical files were discovered in more than 50%, while 14.51% per cent filenames, upper and lower case, were used together. 5.7% of files use all-lowercase letters. Only 6.53% of filenames were found to be numerical. These types of sentence cases were used in filenames. For example, Light Induced Electron-Phonon Scattering Mediated Resistive Switching in Nanostructured Nb Thin Film Superconductor.pdf, Studies In Aerobic Digestion Of Waste Activated Sludge.pdf, 28912.pdf, slender.pdf, AIAA-5708-482.pdf).

Figure 3 displays the pattern of use of ordinal numbers associated with the resources as part of the filenames. Serial position (26.8%) and Chapter numbers (20.2%) were the most widely used. Some other words used for naming the files were as follows: 'Online Thesis Search Results.html', 'file 4 (chapter 1).pdf.' These file names are assigned to resource documents in repositories.

Conclusion

The file naming practices (calf path) in the selected repositories is a common phenomenon in the digital environment. The present study examined the file naming practices seen in the institutional repositories. The study revealed the naming practices by curators of digital resources, such as archivists, digital librarians, records managers and others. National policies need to be developed regarding file naming practices for resources in digital repositories.

References

1 Comer DE, and Murtagh TP, The Tilde file naming scheme. Department of Computer Science and Technical Reports, (No. 85-507), Purdue University, West Lafayette, Indiana, 1985, https://docs.lib.purdue.edu/cstech/428/.

2 Halbert M, Skinner K, and Mcmillan G, Avoiding the calf-path: digital preservation readiness for growing

collections and distributed preservation networks, In *Archiving Conference*, Society for Imaging Science and Technology, 2008, p. 86–91. https://educopia.org/wp-content/uploads/2018/07/Archiving2009-Halbert-Skinner-McMillan.pdf.

3 Chapman HP, The file naming habits of personal computer users, PhD thesis, University of North Carolina, 1999.

4 Carroll JM, Creative names for personal files in an interactive computing environment, *International Journal of Man-Machine Studies*, 16 (4) (1982) 405–438.

5 Carroll JM, Creating names for things, *Journal of Psycholinguistic Research*, 10 (4) (1981) 441–455.

6 Collantes LY, Degree of agreement in naming objects and concepts for information retrieval, *Journal of the American Society for Information Science*, 46 (2), (1995) 116–132.

7 Hodge MH, and Pennington F M, Some studies of word abbreviation behavior, *Journal of Experimental Psychology*, 98 (2) (1973) 350–361.

8 Katzenberg B and Piela P, Work language analysis and the naming problem, *Communications of the ACM*, 36 (4) (1993) 86–93.

9 Teasley BE, The effects of naming style and expertise on program comprehension, *International Journal of Human-Computer Studies*, 40 (5) (1994) 757–770.

10 Rogers J, File naming standards for digital collections, *B Sides: Fieldwork*, (1), (2014) 1–15. file:///tmp/mozilla_hari0/bsides-27901-rogers.pdf.

Appendix I

List of institutional repositories considered for the study

Sl. No.	Name of the Institution / University	Name of the Institutional repository	Type of repository collections	Software name	URL
1	National Digital Library of India	NDL India	Articles	Virtual repository	https://ndl.iitkgp.ac.in/
2	Bangalore University	Digital repository of Bangalore university	Articles, books and book section	E-prints	http://eprints-bangaloreuniversity.in/view/subjects/
3	Indian Institute of Management Kozhikode	Institutional repository of IIM	Articles and thesis	Dspace	http://dspace.iimk.ac.in/
4	INFLIBNET	INFLIBNET Institutional repository	Articles	Dspace	http://ir.inflibnet.ac.in/
5	National Institute of Technology Rourkela	Institutional repository of NIT Rourkela	Thesis and Dissertation	E-prints	http://ethesis.nitrkl.ac.in/view/subjects/
6	Raman Research Institute	Digital repository of RRI	Thesis	Dspace	http://dspace.rri.res.in/
7	Indian Institute of Astrophysics	Digital repository of IIA	Articles and thesis	Dspace	http://prints.iiap.res.in/
8	Institute of Mathematical Science	Digital repository of IMSc	Thesis	Dspace	https://www.imsc.res.in/xmlui/handle/123456789/1
9	Indian Institute of Technology Delhi	Institutional repository of Central library, IIT Delhi	Articles and thesis	E-prints	http://eprint.iitd.ac.in/
10	Cochin University of Science and Technology	Digital library of CUSAT	Article, Book Chapter and conference article	Dspace	http://dspace.cusat.ac.in/jspui/
11	National Institute of Science Communication and Information Resources (NISCAIR) (now NIScPR)	NISCAIR online periodicals repository	Articles	Dspace	http://nopr.niscair.res.in/
12	National Aerospace Laboratories (NAL)	Institutional repository of NAL	Articles and Conference paper	E-prints	https://nal-ir.nal.res.in/
13	eGyankosh	National digital repository	Book chapter	Dspace	http://egyankosh.ac.in/browse?type=subject
14	Central Marine Fisheries Research Institute (CMFRI)	Open access Institutional repository	Article, Book section and thesis (Master's)	E-prints	http://eprints.cmfri.org.in/view/subjects/
15	Indian Institute of Technology Roorkee (IITR)	Shodh Bhagirathi	Thesis and Dissertation	Dspace	http://shodhbhagirathi.iitr.ac.in:8081/jspui/
16	Indian Institute of Petroleum (IIP)	Institutional repository of IIP	Articles	Dspace	http://library.iip.res.in:8080/dspace
17	Gokhale Institute of Politics and Economics (GIPE)	Digital repository of GIPE	Books	Dspace	https://dspace.gipe.ac.in/xmlui/

(*Contd.*)

Appendix I

List of institutional repositories considered for the study (*Contd*.)

18	International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)	Open access repository of ICRISAT	Articles, Book section, Abstract and conference paper	E-prints	http://oar.icrisat.org/
19	Indian Institute of Science Bangalore (IISc)	Digital repository of Etd	Thesis and Dissertation	Dspace	http://etd.iisc.ac.in/
20	National Metallurgical Laboratory Jamshedpur (NML)	Institutional repository of NML	Article, thesis, books and conference article	E-prints	http://eprints.nmlindia.org/view/subjects/
21	National Institute of Immunology	Institutional repository of NII	Articles	Dspace	http://202.54.249.144:8090/dspace/
22	Maharaja Sayajirao University of Baroda	Digital library of Hansa Mehta Library	Thesis and Book	Dspace	http://14.139.121.106:8080/jspui/
23	Cochin University of Science & Technology	Dyuthi at CUSAT	Article and thesis	Dspace	https://dyuthi.cusat.ac.in/xmlui/
24	Indian Institute of Chemical Biology	Open access repository of IICB	Articles, thesis and conference paper	E-prints	http://www.eprints.iicb.res.in/
25	Indian Academy of Sciences	Open access repository of IAS	Articles	E-prints	http://repository.ias.ac.in/
26	Anjuman-I-Islam's Kalsekar Technical Campus (AIKTC),	Institutional repository of AIKTC	Articles	Dspace	http://www.aiktcdspace.org:8080/jspui/
27	National Institute Of Oceanography	Digital repository of NIO	Articles, Thesis and Book chapter	Dspace	http://drs.nio.org/drs/
28	Vidyasagar University	Institutional repository of Vidyasagar University	Article and thesis	Dspace	http://inet.vidyasagar.ac.in:8080/jspui/
29	Central Food Technology Research Institute	Institutional repository of CSIR-CFRTRI	Article and conference paper	E-prints	http://ir.cftri.com/
30	National Academy of Agricultural Research Management	Digital repository	Article, book, book section and conference paper	E-prints	http://eprints.naarm.org.in/
31	National Physical Laboratory	Institutional repository of NPL	Articles and Conference paper	E-prints	http://npl.csircentral.net/
32	Indian Institute of Technology Hyderabad	Research Archive of Indian Institute of Technology Hyderabad	Articles, Thesis and dissertation and Book chapter	E-prints	http://raiith.iith.ac.in/
33	Indian Council of Agricultural Research Institutes	Krishikosh : Institutional repository of ICARI	Thesis	Dspace	http://krishikosh.egranth.ac.in/
34	Digital Repository of Ministry of Earth Sciences	Open access digital repository of MOES	Articles and Conference paper	E-prints	http://moeseprints.incois.gov.in/
35	Aligarh Muslim University	Knowledge repository	Thesis and Dissertation	E-prints	http://ir.amu.ac.in/
36	Saurastra University	Online archive of PhD theses of SU	Theses	E-prints	http://etheses.saurashtrauniversity.edu/
37	Shodhganga	Shodhganga : Indian ETD Repository	Theses	Dspace	http://shodhganga.inflibnet.ac.in/
38	Mahatma Gandhi University	Online theses search repository	Theses	NityaD' Arch	http://mgutheses.org/
39	CSIR Unit for Research and Development of Information Products	Open access repository of Indian theses	Theses	E-prints	http://eprints.csirexplorations.com/