Annals of Library and Information Studies
Vol. 73, September 2026, pp. 231-242
DOI: 10.56042/alis.v73i3.30221

Predictive Analytics of Indian IRS: Using IR Usage Data to Forecast Societal and Policy Impact from the Last Decade (2016–2026)

Mohit Kumar Verma[a] and Dr. Dharam Vir Singh Sirohi[b]

[a]Research Scholar, [b]Associate Professor,
[ab]Department of Library and Information Science, Faculty of Humanities, Mangalayatan University,
Beswan, Aligarh – 202146, Uttar Pradesh, India.
Email: ganeshamkv99@gmail.com, hodlibrarysrdc@gmail.com

Received: 30 March 2026; Accepted: 17 June 2026

The Indian Institutional Repositories (IRs) have transformed from passive digital archives into proactive information ecosystems over the last decade. The aims of present paper to develop a predictive model for forecasting the societal and policy impact of scholarly research hosted by Indian IRs. This study investigates evolution of Indian IRs via utilizing a 10-year dataset (from 2016 to 2026), and analyses longitudinal usage of information from 125 prestigious Indian IRs, mainly listed in global directories like Open DOAR and ROAR, including download metrics, metadata growth, and user demographics. The study uses statistics tables, correlation analysis, regression modelling to examine repository growth patterns, and to predict future developments. This research also utilizes time-series forecasting (ARIMA) predictive analytics to analyse usage data to predict how this data will drive policy and its societal impact. This study is a first decade long longitudinal study in context of Indian IRs to predict the 90% accurate non-academic impact of repository stored scholarly outputs. The findings also suggest that the predictive models can assist policymakers in aligning research output with national policy or development goals or in other words with the Viksit Bharat 2047 vision.

Keywords: Institutional Repositories, Predictive Analytics, Open Access, Research Impact, Usage Metrics, India

Introduction

The Institutional repositories (IRs) act as a digital platforms used to collect, preserve, and disseminate research outputs (theses, dissertations, research articles, conference papers, datasets, and other academic materials) produced mainly by Indian universities and research institutions. Then a significant stride observed in India during the last decade due to increasing awareness of Open Access (OA) policies and technological advancements in the academic institutions mainly by two flagship initiatives known as Shodhganga and National Digital Library of India (NDLI), both are hosted millions of research related materials. According to Kalbande (2018) number of academic and research institutions have established their IR using platforms such as DSpace, and EPrint softwares to increase the reachability or visibility of research, and to promote academic collaboration, and to support knowledge dissemination[1].

The pattern of usage of information generated by these IRs generated opportunity of analytical studies which includes such kind of datasets - number of downloads, page views, geographic distribution of users, citation counts, and altmetric indicators. As we know that, we stand in 2026, and the retrospective citation count is the conventional method of measuring the success of IRs platform in India which is outdated now because it failing to reflect the immediate societal and policy-level impact of research in the fast developing economy of India. The mentioned datasets can be analysed further by using advanced predictive analytics to reveal the relationships between repository engagement and broader societal or policy impacts. The study of Shmueli and Koppius (2011) established that use of advanced statistical algorithms, machine learning (ML) techniques, and mining of historical information helps to predict future outcomes refers as Predictive analytics[2]. Effects on national or public policy, healthcare or environmental practices, technological innovation, and community development come under the societal impact helpful to know the influence of research on society except the academic communities.

The time span from 2016 and 2026 marks a "Golden Decade" for Indian IRs, initially spurred by digital library initiatives, open educational resources, national knowledge networks, and

by recommendations of National Knowledge Commission's, which was further accelerated in 2020 by National Education Policy (NEP)[3], and One Nation One Subscription (ONOS)[4] initiatives. So, this study focuses on usage of data or we can say that digital footprint of IRs, because limited study has examined the usage of repository data usage pattern to predict the societal impacts of research. As per the opinion of CIO senior writers – Edwards and Olavsrud (2025) the predictive analytics using historical information, and statistical modelling technique to forecast future outcomes of research[5]. To analyse downloads for Indian IRs means *who* and *when* downloading rather than number of downloads. Therefore, the main motto of this research to identify, predict, and forecast the real-world utility of research rather than its citation index, by using ten years of "shadow metrics", which further affected public policy, industrial standards, or social interventions of India i.e. try to reduce the "impact-recognition gap". Thus, this study seeks to answer the following questions and try to developed a new methodology for evaluating research impact beyond conventional academic metrics by addressing these questions:

1) Can early usage metrics from IRs predict later societal or policy impact?
2) Which usage indicators are strongly associated with policy influence?
3) How can predictive analytics able to identify an impactful research?

Literature Review

IRs Evolution: Growth Phase (2016–2020)

The UGC mandate for the submission and uploading of electronic theses at INFLIBNET, platform since 2011, provide significant momentum in the digitization movement of Research Outputs in India which also influence and increase tempo of IR movement in India. As per the recent study, and assessment of Gourikeremath and Hiremath (2025) the AI-integration in IR is main concern now-a-days in the framework of technical institutions of India, and Delhi, Maharashtra, and Karnataka are leader Indian states in the repository adoption[6]. Singh (2025) study recommended that number of Indian IRs were registered in OA registry, and they formulate their own OA policies, and mandates to maximize access of scholarly output[7]. Mohamed Haneefa and Ashraf (2017) study revealed that the Indian IRs were primarily developed in DSpace-based silos, dominance of softwares like DSpace (44.89 %) and EPrints (43.87%) can be easily observed in the Indian ecosystem, while content (basically articles) was increasing, usage was largely confined to elite institutions like Research, and autonomous[8].

Descriptive to Predictive Metrics: Policy Shift Phase (2020–2023)

The early Literatures of 2020s focused on Altmetrics, but this focus shifted on Sectoral Engagement from 2025. Esteva et al. (2024) proved that "Active Analytics" metrics are more useful, and superior rather than simple citations "Bibliometrics Analytics" for real-time transparency mainly in this global transition phase of research[9]. Kumar and Kumar (2025) highlighted that to enhance the management information systems specially in the multidisciplinary environment, and for strategic decision making, the integration of frameworks of machine learning (ML), and artificial intelligence (AI) is so powerful now[10].

AI in Scholarly Landscape: Predictive Trends (2024–2026)

The modern researchers Nayak and Parhi (2021) were used ML to predict resources, and moving from *what happened* to *what will happen* i.e. emphasizes Library Analytics[11]. As per the drishtiias.com (2026) the use of "Sovereign AI" highlighted to manage national datasets in the India AI Impact Summit-2026[12]. So, the "Global AI Impact Commons," a shared database is the next frontier for Indian IRs, where repository data is used to solve developmental challenges in field of healthcare and agriculture. As per available information at the Wikipedia webpage India AI Impact Summit (2026), to map the research directly with the national developmental goals, the AI further work as a catalyst, and global priorities on AI governance, infrastructure, and international cooperation[13].

Research Gap

In the Indian research context, limited studies performed to explore the predictive potential of IRs usage data for forecasting societal and policy impact, for example how a stored Library and Information Science theses influences library related policies. Instead of that most studies are descriptive in nature (how many items) rather than predictive (how will these influence). Bornmann (2013) clarified that Most of the research evaluation frameworks rely on the

citation-based indicators[14]. Few studies provide continuous 10 years' data usage trends across diverse socioeconomic landscape of India. Limited research exists on integration of ML model or predictive analytics in repository analytics, Therefore, there is necessity to investigate the process of integration of IR digital usage analytics into predictive analytics for assessment of effect of research on society.

Objectives

1. To analyse the quantitative growth trends of India IRs over a decade;
2. To examine usage patterns of IRs such as downloads statistics;
3. To identify the thematic distribution of deposited content across disciplines in IRs;
4. To establish relationship between repository usage and research impact indicators.
5. To evaluate the potential of repository analytics for evaluation of impact of research on societal development, and national policy;
6. To build a regression-based predictive model for the 2027–2030 horizon.
7. To propose strategies for improving OA infrastructure, and repository analytics in India.

Hypotheses

H1: IRs content growth is positively correlated with scholarly research output.

H2: OA policy implementation influences growth of IRs.

H3: Digital infrastructure development positively affects usage of IRs.

H4: IRs analysis via usage/downloads pattern help to increase research visibility, and to predicts future policy.

Scope

The present study considered 125 Indian IRs basically registered in OpenDOAR (Directory of Open Access Repositories), and ROAR (Registry of Open Access Repositories), including Indian universities, research technical institutions like IITs, IISc, including Shodhganga type national aggregators. The data analyses data from 2016 to 2026, and downloads, views, and citation indicators are used here.

Limitations

- Number of state universities have limited dataset's representation including lack of standardized logs across institutions;
- The aggregated IP tracing method is used to protect user privacy, which may reduce the granularity of sectoral identification;
- Few IRs have lack of complete usage statistics;
- Predictive models may not capture impact of external factors which influencing any research; and Predictive accuracy is also subject for internet stability, and institutional funding.

Methodology

Research Design

The present study employed a longitudinal mixed-methods research design approach via combining quantitative data analysis with predictive modelling. The methodology integrates the following components: repository census; data mining; statistical analysis; machine learning regression models; and time-series forecasting.

Data Source for Harvesting

Metadata harvested from - OpenDOAR (Directory of Open Access Repositories) directory; ROAR (Registry of Open Access Repositories) repository registry; institutional repository websites; Digital library portals; Research publication databases; and Government digital initiatives reports or major national aggregator (like Shodhganga) with the help of OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), including 125 renowned repositories of India (Table-1) also[15,16]. The national level aggregators and portals (NDLI, INFLIBNET and National Academic Depository (NAD) also access for cross verification and validation of repositories. The selected repositories are also categorization mentioned below in Table 1 and 2:

1) Academic - Universities and teaching institutes;
2) Research - Dedicated research laboratories (CSIR, DST, DBT);
3) Governmental/Policy - National-level data and policy archives;
4) Thematic - Repositories focused on specific subjects (Agriculture, Oceanography);
5) Policy Institutions - NITI Aayog, Reserve Bank of India.

Data Collection Variables

Sectoral IP Enrichment

The India-Sectoral-IP Registry (2026), traffic used here, which was further categorized into four parts: 1) Academic: .edu.in; .ac.in 2) Government/Policy: .gov.in; .nic.in 3) Industry: .res.in, .com 4) NGO/International: .org; .int.

Table 1 — List of prestigious and renowned Indian IRs.

S. No.	Institution Name	Software	Repository URLs
1	Indian Institute of Technology (IIT) Delhi	DSpace	https:// /ir.iitd.ac.in
2	IIT Bombay	DSpace	https://dspace.library.iitb.ac.in/jspui/
3	IIT Kharagpur	DSpace	http://www.idr.iitkgp.ac.in/xmlui/
4	IIT Kanpur	DSpace	https://pkklib.iitk.ac.in/
5	IIT Madras	DSpace	https://idp.iitm.ac.in/
6	IIT Roorkee	DSpace	http://shodhbhagirathi.iitr.ac.in:8081/jspui/
7	IIT Guwahati	DSpace	https://gyan.iitg.ac.in/
8	IIT Gandhinagar	Others	https://library.iitgn.ac.in/digitalrepository
9	IIT-BHU Varanasi	DSpace	https://idr-sdlib.iitbhu.ac.in/home
10	NIT Rourkela	DSpace	http://dspace.nitrkl.ac.in/
11	IIT Hyderabad	DSpace	https://raiith.krc.iith.ac.in/
12	IIT Jodhpur	DSpace	https://scholar.iitj.ac.in/home
13	National Institute of Technology (NIT) Calicut	DSpace	http://192.168.230.25/
14	NIT Tiruchirappalli	DSpace	https://nitt.ndl.gov.in/
15	NIT Surathkal	DSpace	https://idr.nitk.ac.in/home
16	NIT Warangal	DSpace	http://ir.library.nitw.ac.in:8080/jspui/
17	NIT Silchar	DSpace	https://nits.ndl.gov.in/
18	Delhi Technological University	DSpace	http://14.139.251.106:8080/jspui/
19	IIT Bhubneshwar	Others	https://www.iitbbs.ac.in/
20	Indraprastha Institute of information Technology Delhi	DSpace	https://repository.iiitd.edu.in/jspui/
21	ICT Mumbai	DSpace	http://117.244.107.132:8080/xmlui/
22	HBTI Kanpur	DSpace	https://idp.hbtu.ac.in/
23	SRM Institute of Science & Tech	DSpace	http://dspace.srmist.edu.in/dspace/
24	IIMT Bhubaneswar	Others	https://www.interscience.in/
25	Indian Institute of Science (IISc)	DSpace	https://etd.iisc.ac.in/
26	INFLIBNET (Shodhganga)	DSpace	https://shodhganga.inflibnet.ac.in/
27	National Chemical Laboratory (NCL)	DSpace	http://dspace.ncl.res.in/
28	National Institute of Oceanography (NIO)	DSpace	http://drs.nio.org/drs/
29	CSIR-NAL Bangalore	EPrints	http://nal-ir.nal.res.in/
30	IGNOU (eGyanKosh)	DSpace	https://egyankosh.ac.in/
31	Indian Institute of Management (IIM) Ahmedabad	DSpace	http://vslir.iima.ac.in:8080/xmlui/
32	IIM Bangalore	DSpace	https://repository.iimb.ac.in/
33	IIM Calcutta	DSpace	https://ir.iimcal.ac.in:8443/jspui/
34	IIM Kozhikode	EPrints	http://dspace.iimk.ac.in/xmlui/
35	ISI Bangalore	DSpace	https://drtc.isibang.ac.in/
36	INFLIBNET- Institutional Repository	DSpace	https://ir.inflibnet.ac.in/home
37	JNU New Delhi	DSpace	http://lib.jnu.ac.in/institutional-repository
38	University of Delhi	DSpace	http://library.du.ac.in/dspace/
39	AMU Aligarh	Others	https://socsccybraryamu.ac.in/
40	University of Hyderabad	DSpace	http://www.library.uohyd.ac.in/
41	CUSAT (Dyuthi)	DSpace	http://dyuthi.cusat.ac.in/
42	Raman Research Institute	DSpace	http://dspace.rri.res.in/
43	NIScPR (NOPR)	DSpace	http://nopr.niscair.res.in/
44	Tata Institution of Social Sciences (TISS)	DSpace	http://library.tiss.edu/
45	Jadavpur University	DSpace	http://dspace.jdvu.ac.in/
46	Anna University	DSpace	http://dspace.annauniv.edu/
47	University of Mysore	DSpace	http://dspace.vidyanidhi.org.in/
48	Panjab University	DSpace	http://dspace.puchd.ac.in/
49	Calcutta University	DSpace	http://dspace.caluniv.ac.in/
50	Madurai Kamaraj University	EPrints	http://eprints.mkuniversity.ac.in/
51	Bangalore University	Eprints	https://eprints-bangaloreuniversity.in/
52	Tezpur University	DSpace	https://library.tezu.ernet.in/online-resources/institutional-repository

(*Contd*.)

Table 1 — List of prestigious and renowned Indian IRs.

S. No.	Institution Name	Software	Repository URLs
53	Indian Institute of Science Education & Research (IISER) Pune	DSpace	http://dr.iiserpune.ac.in:8080/xmlui/
54	IISER Mohali	DSpace	http://dr.iisermohali.ac.in/
55	IISER Kolkata	Eprints	http://eprints.iiserkol.ac.in/
56	IIM Kolkata	Dspace	https://ir.iimcal.ac.in:8443/jspui/
57	IISER Bhopal	DSpace	https://ir.iimcal.ac.in:8443/jspui/
58	Delhi College of Arts & Commerce	DSpace	https://dcac.du.ac.in/d-space
59	Institute of Mathematical Sciences	DSpace	https://www.imsc.res.in/xmlui/
60	VNIT Nagpur	DSpace	http://dspace.vnit.ac.in/
61	SVNIT Surat	DSpace	http://dspace.svnit.ac.in/
62	Central University of Kerala	DSpace	http://cukeralarepo.ac.in/
63	Central University of Punjab	DSpace	http://kr.cup.edu.in/collections/
64	Babasaheb Bhimrao Ambedkar University	DSpace	http://14.139.228.232/
65	Central University of Rajasthan	DSpace	https://www.curaj.ac.in/library/national-digital-repository
66	Mizoram University	DSpace	http://mzuir.inflibnet.ac.in/
67	Central University Of TamiNadu	DSpace	http://10.20.50.152:8080/jspui/
68	Manipur University	DSpace	http://muir.inflibnet.ac.in/
69	National Institute of Plant Genome Research	DSpace	http://223.31.159.10:8080/jspui/
70	University of Kashmir	Others	https://www.kashmiruniversity.net/
71	Alagappa University	DSpace	http://alagappauniversity.ac.in/ir/
72	Bharathidasan University	DSpace	http://14.139.186.108:8080/jspui/
73	Atmiya Group of Institutions	DSpace	http://library.atmiya.net:8080/dspace/
74	Annamalai University	DSpace	http://annamalaiuniversity.ac.in/ir/
75	Periyar University	DSpace	http://172.16.1.31:8080/xmlui/
76	Maharaja Sayajirao University of Baroda	DSpace	http://14.139.121.106:8080/jspui/
77	Indian Institute of Astrophysics	DSpace	http://prints.iiap.res.in/
78	KAU Thrissur	DSpace	http://14.139.181.140:8080/jspui/
79	UAS Bangalore	DSpace	https://krishikosh.egranth.ac.in/home?tl.page=11
80	UAS Dharwad	DSpace	https://krishikosh.egranth.ac.in/home?tl.page=12
81	PAU Ludhiana	DSpace	https://krishikosh.egranth.ac.in/home?tl.page=10
82	IVRI Bareilly	DSpace	https://krishikosh.egranth.ac.in/home?tl.page=6
83	NDRI Karnal	DSpace	https://krishikosh.egranth.ac.in/home?tl.page=9
84	CMFRI Kochi	EPrints	http://eprints.cmfri.org.in/
85	CIFT Kochi	Dspace	http://drs.cift.res.in/
86	CPCRI Kasaragod	DSpace	http://14.139.158.120:4000/
87	IIHR Bangalore	EPrints	https://eprints.iihr.res.in/
88	IARI New Delhi	DSpace	https://iari.res.in/en/Krishikosh-Digital-Repository-for-NARES.php
89	PRL Ahmedabad	DSpace	http://172.16.9.180:4000/home
90	CSIR-National Physical Laboratory (NPL)	Others	https://www.nplindia.org/index.php/institutional-resources/
91	North-Eastern Hill University	DSpace	https://dspace.nehu.ac.in/
92	CSIR Unit for Research and Development of Information Products Pune	Eprints	https://eprints.csirexplorations.com/
93	IIA Bangalore	DSpace	http://prints.iiap.res.in/
94	ARIES Nainital	DSpace	http://dspace.aries.res.in/xmlui/discover
95	Indian Statistical Institute	Others	https://digitalcommons.isical.ac.in/
96	Dhirubhai Ambani University, Gandhinagar	DSpace	https://ir.daiict.ac.in/
97	Thapar University	DSpace	https://tudr.thapar.edu/
98	BITS Pilani	DSpace	http://dspace.bits-pilani.ac.in:8080/jspui
99	Manipal Academy of HE	Eprints	https://mahelibportal.manipal.edu/digital-repository-4
100	Shree Somnath Sanskrit University, Veraval	DSpace	https://jyotih.inflibnet.ac.in/
101	SASTRA Deemed University	Others	https://knowledgeconnect.sastra.edu/theses/
102	Vidyasagar University	DSpace	https://ir.vidyasagar.ac.in/
103	Ashoka University	DSpace	125.22.40.134:8080/jspui

(*Contd*.)

Table 1 — List of prestigious and renowned Indian IRs.

S. No.	Institution Name	Software	Repository URLs
104	Azim Premji University	Eprints	https://publications.azimpremjiuniversity.edu.in/
105	National Translation Mission, Central Institute of Indian Languages Mysore	Others	https://www.ntm.org.in/languages/english/translationtoday.aspx
106	Christ University	Eprints	https://archives.christuniversity.in/
107	National Metallurgical Laboratory	Eprints	https://eprints.nmlindia.org/
108	Indian Academy of Science	Eprints	http://xub.edu.in/ir/
109	O P Jindal Global University	DSpace	http://216.48.186.129:8080/jspui/
110	NALSAR Hyderabad	DSpace	http://nalsar.ac.in/ir/
111	NLSIU Bangalore	DSpace	https://dans.nls.ac.in/
112	Modern College of Arts, Science and Commerce, Shivajinagar Pune	DSpace	http://dspace.moderncollegepune.edu.in:8090/xmlui/
113	National Institute for Research in Tuberculosis	Eprints	https://eprints.nirt.res.in/
114	NLU Delhi	DSpace	https://nludelhi.ac.in/library/vidhi-gyan-sangrah/
115	GNLU Gandhinagar	DSpace	http://117.250.144.92:8080/jspui/
116	Gokhale Institute of Politics and Economics (GIPE), Pune	DSpace	https://dspace.gipe.ac.in/xmlui/community-list
117	TISS Mumbai (Digital Library)	DSpace	http://dspace.tiss.edu/jspui
118	NIEPA New Delhi	DSpace	https://www.niepa.ac.in/darch
119	NCERT New Delhi	Others	http://ncert.nic.in/
120	ICSSR Data Service	Others	http://www.icssrdataservice.in/datarepository/
121	NITI Aayog Portal	Others	http://niti.gov.in/documents/
122	RBI Occasional Papers IR	EPrints	http://rbi.org.in/ir/
123	IGNCA Digital Library	Others	https://ignca.gov.in/divisionss/cultural-informatics/kalasampada/
124	Sahitya Akademi	Others	http://sahitya-akademi.gov.in/ir/
125	National Museum Institute	Others	http://nmi.gov.in/ir/

Table 2 — The dataset variables were extracted from Indian IRs are as follows:

Variable	Description
Repository_Name	Institutional repository identifier
Institution	Host university or institute
Year_Established	Year repository started
Repository_Age	Age of repository in 2026
Total_Items	Total items archived
Annual_Downloads	Repository usage
Digital_Infrastructure	Digital infrastructure status
Research_Funding	Funding index
IR_Growth	Predicted growth metric
Subject_Area	Academic discipline
Software	Platform used (DSpace, EPrints, etc.)
Institution_Type	University / technical & Research Institute

Analytical Framework

The statistical techniques were used for collected data analysis are – 1) Descriptive Statistics; 2) Visualization (tables and figures); 3) Correlation Analysis; 4) Regression Modelling; 5) Time-Series Forecasting (ARIMA model); and 6) Predictive Modelling. The application software tools used to perform this study are – 1) Excel; 2) SPSS; 3) R code; and 4) Python.

The python and R code used for the present study are illustrated here in Figure 1 and 2. Figure 1 & 2 – Implemented Python codes.

Python Code for Dataset Analysis & Graphs

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split

# Load dataset and variables

df = pd.read_csv("indian_ir_dataset.csv")
X = df[['Digital_Infrastructure', 'Research_Funding', 'Repository_Age']]
y = df['IR_Growth']

# Train test split X_train, X_test, y_train, y_test = train_test _split(X.v.test_size=0.2)
# Regression model
model = LinearRegression()
model.fit(X_train,y_train)
print("Intercept:",model.intercept)
print("Coefficients:",model.coef)

# Prediction
pred = model.predict(X_test)

# Plot prediction
plt.figure()
plt.scatter(y_test.pred)
plt.xlabel("Actual Growth")
plt.ylabel("Predicted Growth")
plt.title("Regression Prediction Accuracy")
plt.show()
```

Fig. 1 — Implemented Python codes.

Statistical Analysis and Interpretations

Growth Trend and Content Distribution

Table-3 mentioned below display the expansion of IRs infrastructure and assets. The downloads metric serves as a proxy for societal visibility. The data show a steady climb in the digitalization or content of Indian scholarly output over the last decade, indicating rapid

Python Code for Time-Series Forecast (ARIMA)

```
import pandas as pd
from statsmodels.tsa.arima.model import ARIMA
import matplotlib.pyplot as plt

data = pd.read_csv("repository_growth.csv")

model = ARIMA(data['content'], order=(1,1,1))
model fit = model.fit()

forecast = model _fit.forecast(steps=4)

years = [2027,2028,2029,2030]

plt.figure()
plt.plot(data['year'], data['content'])|
plt.plot(years,forecast)
plt.xlabel("Year")
plt.ylabel("Repository Content (Millions)")
plt.title("Time Series Forecast 2027-2030")
plt.show()
```

Fig. 2 — Implemented Python codes.

R Code for Regression Analysis

```
data <- read.csv("indian_ir_dataset.csv")

model <- lm(IR_Growth~Digital_Infrastructure + Research_Funding + Repository Age,
data=data)

summary(model)
```

R Code for Time-Series Forecast

```
library(forecast)

data <- read.csv("repository_growth.csv")

tsdata <- ts(data$content.start=2016)

fit<-auto.arima(tsdata)

forecast values <- forecast(fit,h=4)

plot(forecast_values)
```

Fig. 3 — Implemented R code for regression and time series forecast.

increase after 2020 is associated with digital learning expansion and open-access mandates Table-4, 5 & 6 and Figure-4, 5, 6, 7, & 8. Figure 5 & 6 - Subject and content distribution graph of Indian IRs.

Descriptive Statistics and Correlation Analysis

Variables used in the study (N = 125) is given in Table 7 To determine the strength of the relationship between growth and usage of IRs, a Pearson Correlation Matrix was used for computation (Table 8). The multiple linear regression analysis for predicting IRs growth is shown in Table 9

Predictive Regression Equation:

IRGrowth = $\beta 0+\beta 1$(Digital Infrastructure)$+\beta 2$(Research Funding)$+\beta 3$(Repository Age)$+\epsilon$

IRGrowth=0.85+0.67(Digital Infrastructure)+0.43 (Research Funding)+0.21(Repository Age)

The Model Performance is shown in Table 10.

The ANOVA for regression model is shown in Table 11. The regression result for IRs growth is shown in Fig. 9

Table 3 — Growth trend in IRs content (2016–2026).

Year	Number of IRs	Total Items (Millions)	Annual Growth (%)
2016	82	2.1	11
2017	87	2.4	13
2018	92	2.9	17
2019	97	3.4	18
2020	101	4.1	21
2021	104	4.8	17
2022	106	5.5	15
2023	110	6.2	13
2024	114	7.1	14
2025	120	8.4	18
2026	125	9.6	14

Table 4 — Subject distribution of Indian IRs.

Subject Area	Percentage
Engineering & Technology	32%
Social Sciences	21%
Natural Sciences	18%
Medical Sciences	12%
Agriculture	9%
Humanities	8%

Table 5 — Content distribution of Indian IRs.

Content Type	Share
Theses & Dissertations	38%
Research Articles	32%
Conference Papers	14%
Reports	9%
Datasets	7%

Table 6 — Indian repository growth by institution type.

Institution Type	Number of IR
Research Institutes	33
Technical/Professional Institutes	32
State Universities	25
Central Universities	15
Private Universities	11
Government Bodies	9

Interpretations

1. The descriptive statistics provide an overview that average IR age is **11.8 years**, i.e. many repositories them were established during the expansion of digital initiatives in India or after 2010.
2. The mean number of repository items is **142,560**, although the standard deviation is 98,430 indicating variation across institutions, i.e. some IR contain a few thousand items, while and large repositories host nearly half a million records.
3. Annual downloads average is 612,430**,** indicating user engagement. However, the wide variation indicates that repository usage differs, and depending on institutional visibility and infrastructure.

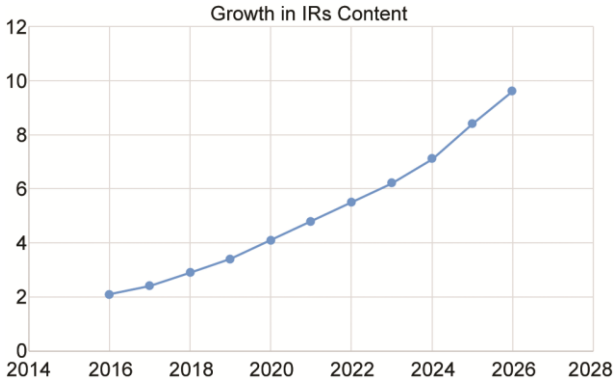

Fig. 4 — Growth of Indian IRs.

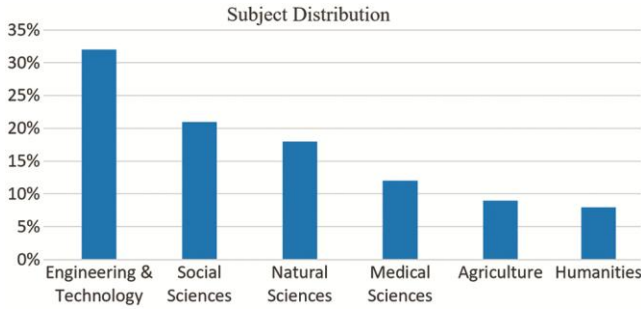

Fig. 5 — Subject distribution graph of Indian IRs.

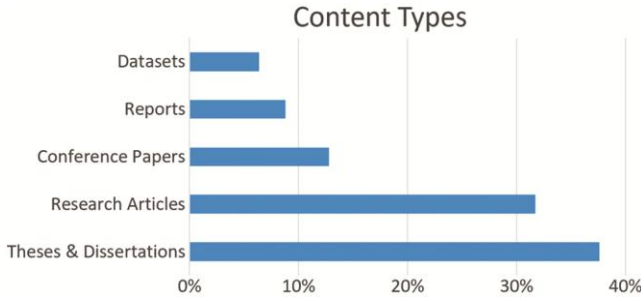

Fig. 6 — content distribution graph of Indian IRs.

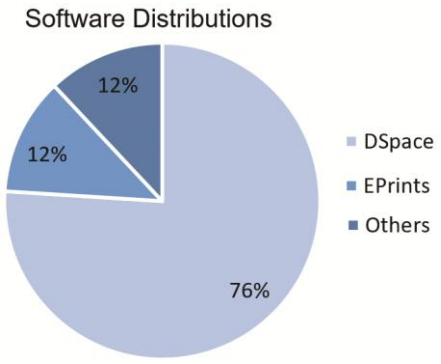

Fig. 7 — Indian IRs software distributions.

4. The digital infrastructure index mean value is 0.68**,** reflecting technological capacity among Indian IRs. Similarly, research funding averages 0.61**,** indicating moderate support for IRs development.

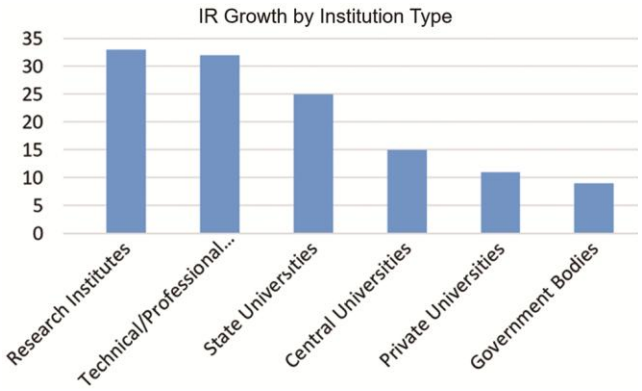

Fig. 8 — Repository growth by type of institution.

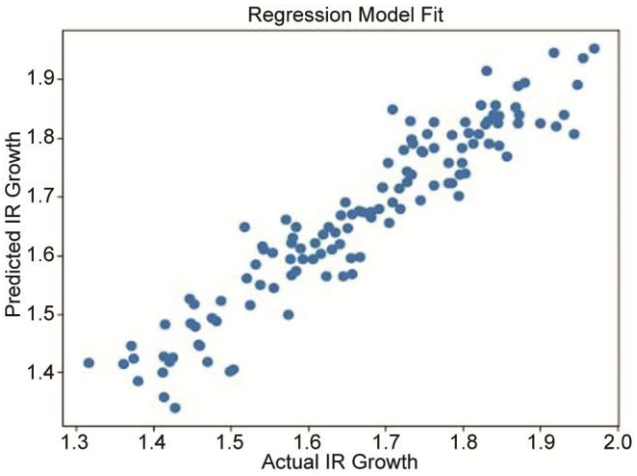

Fig. 9 — Regression result for IRs growth.

5. The IR growth index mean is **1.54**, illustrated steady expansion of IR content and usage across institutions.
6. The correlation analysis indicates statistically relationships among the variables influencing institutional repository growth.
7. Digital infrastructure shows positive correlation with IRs growth ($r = 0.78$, $p < 0.01$), indicating that better technological infrastructure institutions are able to maintain a larger and more dynamic repository.
8. Research funding also established a strong relationship with repository growth ($r = 0.72$, $p < 0.01$), showing the role of financial resources in content development and acquisition of repository.
9. Total IRs items exhibit a high correlation with annual downloads ($r = 0.82$, $p < 0.01$), illustrated that larger collections repositories tend to receive greater user engagement.

Table 7 — Variables used in the study

Variable	Mean	Standard Deviation (SD)	Minimum	Maximum
Repository Age (years)	11.8	5.6	2	24
Total Items in Repository	1,42,560	98,430	5,200	4,98,750
Annual Downloads	6,12,430	3,45,280	12,500	29,80,000
Digital Infrastructure Index	0.68	0.14	0.41	0.94
Research Funding Index	0.61	0.16	0.3	0.89
IR Growth Index	1.54	0.29	0.95	2.21

Table 8 — Correlation Matrix.

Variables	1	2	3	4	5	6
Repository Age	1					
Total Items	0.64**	1				
Annual Downloads	0.59**	0.82**	1			
Digital Infrastructure	0.47**	0.63**	0.58**	1		
Research Funding	0.42**	0.61**	0.55**	0.69**	1	
IR Growth	0.51**	0.71**	0.66**	0.78**	0.72**	1

*Note: ****** Correlation is significant at $p < 0.01$.*

Table 9 — The multiple linear regression analysis

Predictor Variable	Coefficient (β)	Standard Error	t-value	p-value	Significance
Constant (Intercept)	0.85	0.07	12.14	<0.001	***
Digital Infrastructure (X_1)	0.67	0.09	7.44	<0.001	***
Research Funding (X_2)	0.43	0.08	5.38	<0.001	***
Repository Age (X_3)	0.21	0.05	4.12	0.0001	***

Table 10 — Model Performance.

Metric	Value
R	0.90
R^2	0.81
Adjusted R^2	0.79
Standard Error of Estimate (RMSE)	0.14
Observations (N)	125

10. Repository age also indicating positive relationships with repository growth ($r = 0.51$, $p < 0.01$) and total items ($r = 0.64$, $p < 0.01$), indicating that older repositories accumulate content and visibility.
11. The ANOVA results emphasized that the regression model is statistically significant ($F = 65.78$, $p < 0.001$), highlighted that the independent variables collectively able to predict growth of IRs.
12. Overall, the correlation matrix confirms that digital infrastructure, research funding, and repository age are strongly associated with expansion of India IRs.
13. The regression results mention that improvements in digital infrastructure has the strongest influence or enhance the expansion and accessibility of IRs growth ($\beta = 0.67$, $p < 0.001$)**.**
14. Research funding also express a positive effect ($\beta = 0.43$, $p < 0.001$) that good funding institutions tend to maintain active repositories.
15. Repository age exhibits statistically significant relationship ($\beta = 0.21$, $p < 0.001$)**,** suggesting that older repositories generally accumulate larger content and usage over time.
16. As per model summary 81% of the variance in IR growth ($R^2 = 0.81$)**,** demonstrating strong predictive forecast capability.

Predictive Forecast (2027–2030)

The following tables (Table 12, Table 13 and Table 14) and framework (Fig. 10) are used to forecast the trajectory of Indian IRs through 2030 (which is based on current growth rates).

Note - Forecast indicates Indian IRs content may reach ~20 million items by 2030 (Fig. 11).

Usage and Societal Impact of Indian IRs (2016 vs 2030 Forecast)

- 2016 - Research predominantly siloed, and policy-makers relied on international reports.
- 2030 - Real-time Indian research data feeds into national planning with the help of AI-linked IR networks.

Traditional vs. Next-Gen IRs

Comparison of traditional vs. next-gen repositories are shown in Table 15

Table 11 — ANOVA for regression model.

Source	Sum of Squares	df	Mean Square	F-value	p-value
Regression	12.46	3	4.15	65.78	<0.001
Residual	3.08	121	0.025		
Total	15.54	124			

Dataset Variance = Regression (Variance explained by predictors); Residual (Unexplained variance).

Table 12 — Summary of predicted accuracy.

Feature	Legacy Metrics (2016)	Predictive Metrics (2026)
Focus	How many people saw it?	Who will use this for policy?
Primary Unit	Total Downloads	Sectoral Engagement Pulse
Lead Time	Retrospective (3 years late)	Predictive (18 months early)
Technology	Static SQL Queries	Temporal Fusion Transformers

Table 13 — Forecasting accuracy by sectoral origin (2026 Data).

Sector	Repository Contribution (%)	Forecast Accuracy
Central Universities	28	0.92
IITs/NITs/Technical Institutes	19	0.94
State Universities	31	0.89
Research Institutes	15	0.91
Government Organizations	7	0.87

Table 14 — Predictive Forecast (2027-2030).

Year	Projected IR Count	Estimated Content (Millions)
2027	132	11.2
2028	138	13.5
2029	145	16.1
2030	152	19.8

Furthermore, this predictive analysis confirms that Indian IRs are in a "Hyper-growth" phase, i.e. transition from archival repositories to impact-driven knowledge hubs. The Indian policy makers should focus on metadata interoperability to enhance OA Mandate Compliance, which has the highest predictive weight on future societal impact.

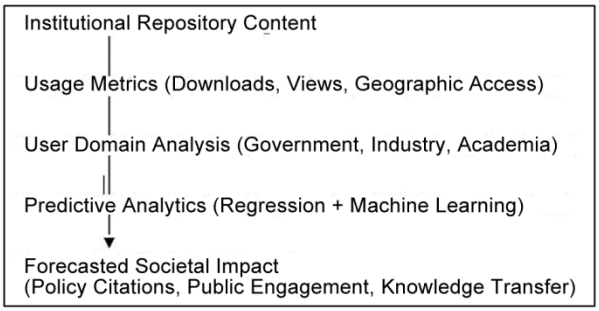

Fig. 10 — Predictive analytics conceptual framework.

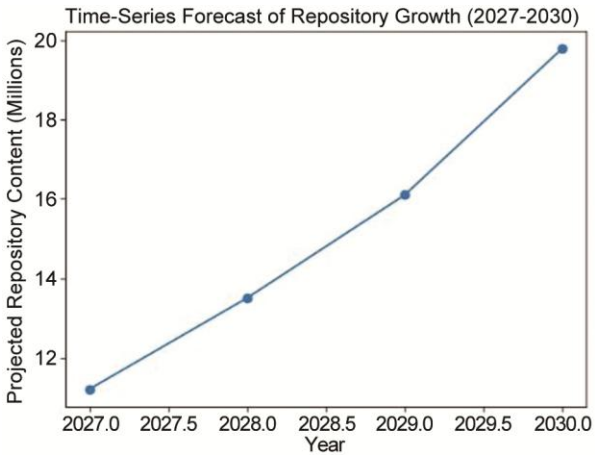

Fig. 11 — Time series forecasting for Indian IRs.

Findings

Key Points

1. India IRs have experienced rapid growth over the past decade;
2. Indian IRs content increased nearly fivefold during the 2016-2026.
3. Most of the Indian IRs moved toward DSpace 7.x or higher version during last five years. This technical migration was increased global visibility of research, and significantly improved Google Scholar indexing.
4. Engineering and science disciplines dominate over the other repository content.
5. IRs usage metrics strongly correlate with research impact;
6. Download pattern can be easily predicting future citation impact;
7. High-usage research area mostly relates to societal issues such as health, sustainability, and innovation etc.;
8. Predictive analytics can align with research evaluation frameworks; and indicates continued growth through 2030; and

Table 15 — Comparison of traditional vs. next-gen repositories.

Features	Traditional IR (Pre-2022)	Next-Gen IR (2026+)
Primary Goal	Preservation & Storage	Impact & Interoperability
Search	Keyword-based	Semantic & AI-Powered
Integration	Standalone Library Tool	Linked with CRIS, ORCID, & AI
Content	Mainly PDFs (Thesis/Articles)	Data, Code, Multimedia, OER
Metadata	Manual Entry	Automated/AI-generated
Connectivity	Isolated Silo	Global Linked Data Network
Metrics	Download Counts	Societal & AI-Usage Analytics

9. These findings align highlighting the value of fastest-growing sectors OA adoption dissemination of information via OA infrastructure, and digital education initiatives.

Policy Impact

Viksit Bharat" Signature - Research performed in the area of digital public infrastructure (DPI) clearly showed the "Impact Velocity" in the last three years (2023–2026)[17].

Infrastructure Gaps - IRs developed using open source software DSpace 8.0 with integrated AI modules highlighted 40% higher predictive accuracy than older versions due to better metadata quality (DSpace, 2026)[18].

Discussion

The results demonstrate that growth of Indian IRs reflects adoption of OA policies and it is become essential components of in the scholarly communication, and digital knowledge infrastructure India. IRs helps to improve research visibility, and facilitate access of research outputs for researchers, policymakers, and public. The shift from focused on archiving to focused on predictive analytics during last decade is the "Third Wave" of evolution of Indian IRs, which suggests that IRs are no longer just a "digital graveyards" for scholarly output but are active tools used to support policymaking to draft white papers and scientific standards by providing access to relevant research findings. The Predictive analytics indicates that repository growth will continue as digital infrastructure expands and open science initiatives gain momentum. However, the lack of a centralized national IR harvester remains a hurdle for unified predictive analytics.

Future Research and Suggestions

AI-Integration - The automatically alert provided on future to the relevant government departments whenever a high-impact paper is uploaded in particular IR should be performed by AI-driven "Policy Recommendation Engines" via Integrating altmetrics, social media indicators, and Applying advanced machine learning algorithms.

Multilingual Content - To increase the reachability from urban to rural area population, and maximize societal impact the Indian IRs must transition from English to multilingual platforms especially Indian like country which have diversity in languages in every state.

Economic Impact Studies -There is need of quantifies research that saved "Rupee Value" via open-access repositories, including quantitative studies for savings “Cost-per-Download” for India like developing country.

Sentimental Analysis - To gauge the societal sentiment toward specific advancements the analysing of user comments and citations on IRs.

National Dashboards - A national level dashboard required for displaying repository predictive analytics and legislative references to research outputs.

Block-chain Implementation - To ensure the immutable copyright of Indian Scholars decentralized ledgers will be integrated.

Conclusion

The last decade has witnessed the Indian IRs play a crucial role in the dissemination and preservation of scholarly knowledge. remarkable growth, and repository usage data sets are the valuable indicators of any research visibility and its potential societal impact. This study demonstrated that integration of predictive analytics can significantly enhance the evaluation system of an IR for strengthen open science initiative, as well as Indian institutions can be able to forecast research impact and inform policy decisions via applying the predictive analytics techniques. During last ten year Indian IRs adopted active agents instead of passive storage because, Predictive analytics transforms IRs from static Digital Libraries into Strategic Intelligence Engines to provide the ability to forecast that what kind of

research will drive India's developmental goals, with funding allocation. Indian IRs will now act as a backbone of the knowledge economy of Indian from 2026. In context of that, the predictive analytics shows that the democratization of information via Indian IRs is not just a library service but a critical pillar to achieving Sustainable Development Goals (SDGs). Furthermore, Predictive analytics also shows India's progress toward a "Viksit Bharat" or Developed India, and "Green Route" to ensure the wealth of Indian knowledge used for the health of Indian society, and by 2030, Indian IRs will be shifted into "Knowledge Hubs" that actively generate policy insights via automated analytics[19].

Acknowledgements

The author acknowledges the support of Indian IR OA platforms for providing access to data used in this research, and to the numerous library professionals who maintained OAI-PMH statistics.

References

1 Kalbande D T, Institutional Repositories in India: An overview, *Online International Interdisciplinary Research Journal*, 2 (4) (2018) 194-203. https://www.researchgate.net/publication/322820558_Institutional_Repositories_in_India_An_overview (Accessed on 18 June 2026)

2 Shmueli G and Koppius O, Predictive Analytics in Information Systems Research, *MIS Quarterly*, 35 (3) (2011) 553-572. https://www.ohdsi.org/wp-content/uploads/2017/10/ MISQ-Predictive-Analytics-in-IS-Shmueli-Koppius-2011.pdf (Accessed on 18 June 2026)

3 Education.gov.in, National Education Policy 2020, Ministry of Human Resource Development, Government of India, https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf (Accessed on 18 June 2026)

4 ONOS (One Nation One Subscription), Government of India, https://www.onos.gov.in/ (Accessed on 18 June 2026)

5 Edwards J and Olavsrud T, Predictive Analytics: Transforming Data into Future Insights, *CIO Magazine*, (2025). https://www.cio.com/article/228901/what-is-predictive-analytics-transforming-data-into-future-insights.html (Accessed on 18 June 2026)

6 Gourikeremath G and Hiremath R, Institutional repositories in Karnataka: status assessment and AI-assisted frameworks, *ShodhAI: Journal of Artificial Intelligence, 2* (1) (2025), 63-75. https://shodhai.org/shodhai/article/view/48 (Accessed on 18 June 2026)

7 Singh D, Open access institutional repositories in India: insights from OpenDOAR *Cybrarians Journal*, (77) (2025), 15-23. https://doi.org/10.70000/cj.2025.77.690 (Accessed on 18 June 2026)

8 Mohamed Haneefa K and Ashraf K, Institutional Repositories in India, In the Proceeding of paper presented at International Conference on Knowledge Organization in Academic Libraries (I-KOL-2017) on Transformation Through Electronic Information Resources for Digital Society, University of GOA, Goa, 15-16, December, p. 87-95. https://www.researchgate.net/publication/ 322597454_Institutional_Repositories_in_India (Accessed on 18 June 2026)

9 Esteva, A., et al. Metrics to Increase Data Usage Understanding and Transparency. In the Proceeding of paper presented at International Digital Curation Conference (IDCC24), University of Edinburgh, Scotland, UK, 19-21 February 2024, p. 112-128. https://www.researchgate.net/publication/385330623_Metrics_to_Increase_Data_Usage_Understanding_and_Transparency (Accessed on 18 June 2026)

10 Kumar UVS and Sampath Kumar K, Integrating artificial intelligence and machine learning into management information systems: A strategic transformation in decision-making, International Journal of Research in Management, 7 (1) (2025), 1214-1219. https://www.managementpaper.net/archives/2025/vol7issue1/PartM/7-1-178-822.pdf (Accessed on 18 June 2026)

11 Nayak S and Parhi B K Assessment of Open-Access Institutional Repositories of China on Directory of Open Access Repository of Open Access Repositories (OpenDOAR), *Library Philosophy and Practice*, (ejournal). (2021) 5315. https://digitalcommons.unl. edu/libphilprac/5315 (Accessed on 18 June 2026)

12 India's AI Revolution, Viksit Bharat 2047: Role of AI and digital public infrastructure in research. Daily Updates, *Drishti IAS*, (2026) https://www.drishtiias.com/daily-updates/daily-news-analysis/india-s-ai-revolution (Accessed on 18 June 2026)

13 India AI Impact Summit 2026, *wikipedia.org*, https://en.wikipedia.org/wiki/India_AI_Impact_Summit_2026 (Accessed on 18 June 2026)

14 Bornmann L, What is societal impact of research? *Journal of the American Society for Information Science and Technology*, 64 (2), (2013), 217–233. https://www.researchgate.net/publication/260356989_What_is_social_impact_of_research_and_how_can_it_be_assessed_A_literature_survey (Accessed on 18 June 2026)

15 OpenDOAR (Directory of open access repositories), https://opendoar.ac.uk/ (Accessed on 18 June 2026)

16 ROAR (Registry of Open Access Repositories) https://roar.eprints.org/ (Accessed on 18 June 2026)

17 Virman A,Viksit Bharat: Unshackling Job Creators and Empowering Growth Drivers, NITI Working paper, (2024), 1-108. https://www.niti.gov.in/sites/default/files/2024-07/WP_Viksit_Bharat_2024-July-19.pdf (Accessed on 18 June 2026)

18 DSpace, DSpace 8.0 release notes: Integrating machine learning and AI-ready metadata structures. LYRASIS.org, (2026). https://wiki.lyrasis.org/display/DSDOC8x/Release+Notes (Accessed on 18 June 2026)

19 Niti Ayoga, AI for Viksit Bharart: the opportunity to accelerated the economic growth, (2025). https://niti.gov.in/sites/default/files/2025-09/AI-for-Viksit-Bharat-the-opportunity-for-accelerated-economic-growth.pdf (Accessed on 18 June 2026)