Informetric analyses and support for open science

January 24, 2025

Due to the exponential growth in the number of scientific works, traditional methods of scientific publications no longer meet the needs in several respects. On the one hand, tools are needed to select those of interest from many publications. On the other hand, in thinking about scientific publications, we rely too much on the frameworks set by the printing. When a publication is published online, it allows direct links to other publications or data. On the other hand, corrections and additions can be dynamic (earlier a special "Errata" contribution was required). Of course, such innovations bring new challenges. The scientific publication system has degenerated anyway. Therefore, it calls for a thorough and well-thought-out overhaul [15]. The answers to the problems are offered by the open science movement.

23.1. Scientific background, problem identification, and objective of the proposed research

From the beginning, human communities accumulated and transmitted knowledge from generation to generation. The invention of drawing and the writing system based on it (3500-3000 BC) enabled the "storage" of knowledge and its dissemination in space and time. Euclid's Elements represents the first known attempt to summarize and organize the accumulated geometric knowledge based on logic. Most of the knowledge of the Mediterranean and Near Eastern peoples was collected in the Library of Alexandria.

After the collapse of the Roman Empire, Christianity in Europe "censored" the ancient legacy. A good part of it was preserved by the Arabs and enriched with the knowledge of India and Central Asia. Unfortunately, much of what was collected was destroyed by the Mongols during the occupation of the eastern Arab lands (Baghdad Library).

A major problem in the accessibility and durability of accumulated knowledge was the slow and expensive reproduction - copying. Therefore, a very important step in the development of the preservation and dissemination of knowledge was the invention of printing (Gutenberg, 1450). This greatly accelerated and reduced the cost of reproducing works, thereby enabling much wider accessibility and greater permanence of knowledge. Books and other forms of works (leaflets, newspapers, etc.) began to be printed. Scientific societies appeared and began to publish their journals (French, English).

During the Enlightenment, in France, they began publishing an Encyclopedia, which was supposed to summarize all human knowledge. Libraries also expanded and the number of works increased. Over time, the number of works outgrew the memory capabilities of librarians. To review the works collected in the library, a list of works was prepared - often in book form. The problem with this solution was the introduction of changes. These were written in the appropriate places. As a more appropriate solution, card catalogs (author, subject) - collections of cards began to be used. Each card contained information about an individual work. The cards were stored in drawers in the chosen order (for quick search). The peak of this approach is the Mundaneum catalog (Otlet & La Fontaine, Brussels 1895-1939), which contained 18 million cards in 15 thousand drawers with information about most of the works published up to that time [46].

In his article "As We May Think" (1945), Vannevar Bush suggested a solution (then based on microfilm)

> "Consider a future device ... in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory." [17]

which began to be realized in the 1980s with personal computers.

The development of computers began before World War II (Zuse). They began to enter general use in the 1950s. Initially, large devices were installed in computer centers. Their capacity grew rapidly. In computers, data is stored in digital form (sequences of 0 and 1). Computers have combined media (numbers, text, (moving) images, sound, programs, etc.) in a single device, enabling the storage and processing of large amounts of data. Reproduction (copying) of data in digital form is easy and cheap.

Scientometrics is a subfield of informometrics that studies the quantitative aspects of scientific literature. While the sociology of science focused on the behavior of scientists, scientometrics focused on the analysis of publications. Modern scientometrics is largely based on the work of Derek J. de Soll Price (Science since Babylon, 1961; Little Science, Big Science, 1963) and Eugene Garfield. The latter founded the Institute for Scientific Information (ISI, 1960; now Clarivate Analytics) and created the Science Citation Index, which is often a source of data for scientometric analyses. A dedicated academic journal, Scientometrics, was founded in 1978. The International Society for Scientometrics and Informetrics (ISSI) was founded in 1993 [41, 40].

Informetrics is the study of quantitative aspects of information. It is an extension and refinement of traditional bibliometrics and scientometrics. Informetrics uses bibliometric and scientometric methods to study primarily problems of resource information management and the evaluation of science and technology [41]. Several other related fields also consider other media and forms of scientific materials (notes, reports, plans, patents, specimens, data, programs, recordings, etc.) [30]. The relationships between them are summarized in the figure below by Alexander Doria

Excessive reliance on informetrics in evaluating scientific works is the basis for the "publish or perish" principle, which leads to low-quality research. This principle is one of the manifestations of Goodhart's Law: When a measure becomes a goal, it ceases to be a good measure [31].

Computers can also be connected to each other via data. The development of connectivity began as early as 1970 (ARPANET) and entered wider academic use in the second half of the 1980s (MAIL, FTP, Telnet). General computer connectivity took off in the 1990s in the form of the Internet (World Wide Web (WWW), Tim Berners-Lee, 1990).

Publishing an article in a journal takes time – from a few weeks to a few years. To speed up research in the second half of the last century, authors often printed an article in the form of a report or preprint and sent it to their colleagues. In August 1991, Paul Ginsparg created a preprint repository called **arXiv** on a computer at Los Alamos National Laboratory (LANL), which was accessible from computers connected to the Internet. arXiv is an open-access repository for preprints or post-prints of accepted (non-peer-reviewed) articles in digital form (e-prints). It accepts articles in mathematics, physics, astronomy, electrical engineering,

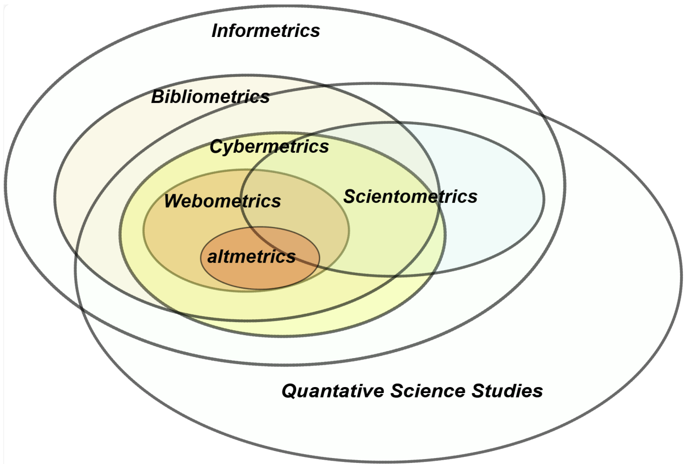

Figure 1: iMetrics

computer science, quantitative biology, statistics, mathematical finance, and economics. It is imitated by several similar repositories: TechRxiv, ChemRxiv, bioRχiv, medRχiv, PsyArXiv, SocArXiv, HAL, RUL, zenodo, etc. Repositories have also emerged for other forms of work: software (c—Net, CPAN, CTAN, CRAN, PyPI (Python), SourceForge), articles (CiteSeerX, ResearchGate, Google Scholar, Academia, DBLP), books (Project Gutenberg, Open Library), data (Kaggle, Microsoft Research Open Data, Google Dataset Search, UCI Machine Learning Repository, CERN Open Data, Dryad, Harvard Dataverse, PEW research, European Social Survey), media (YouTube, Flickr, Google photos, Slideshare, NASA), projects (GitHub, Figshare, Zenodo).

The World Wide Web (WWW) consists of pages with content that can be linked to each other. Each page has its address by which we can reach its content. Initially, the use of the Internet was based on indexes - pages with collections of addresses of interesting pages (usually on a selected topic). Soon the index solution could no longer keep up with the rapid growth of the Internet. It was replaced by collection programs (crawlers) that scan the Internet and add unknown pages to their list. At the user's request, they return the addresses of the pages closest to the request from this list (Lycos, Google). Advertising ("free" service) and the collection of user data (answers and advertisements according to the user's interests) played an important role in improving these search engines. Web 2.0 has enabled (co)creation of online content even by ordinary, non-computer-savvy users. Increasingly, pages are not permanent but are created on the fly at the user's request. [13]

To solve problems on a computer, we need appropriate tools – programs. Most of the time, we have to buy them or write them ourselves. In the late 1960s, Niklaus Wirth developed the Pascal programming language. The Pascal compiler code was available for free. This is one of the earliest examples of free and open software. Openness is very important for the user, because he can (in principle) check what the program does (security) and also find out how (learning). If necessary, he can also adapt such a program to his specific needs (interface language, additional capabilities, etc.). Several such programs appeared, which in the 1980s turned into the GNU free software movement that was formalized by the GNU Manifesto (Richard Stallman, 1985). Free access to collections (libraries, packages) of (sub)programs and also data is very important for the rapid development of science – researchers do not have to start from scratch, but simply build on existing solutions. Towards the end of the 1990s, a related open-source movement emerged, which also included the possibility of commercial programs. The principle "what was developed with public funds should also be publicly available" has been often applied in American research projects (NSF). This enables the rapid transfer of scientific results into use. A good example of this is the development of the first successful GUI-based web browser, Mosaic, which was the basis for both Netscape and Microsoft's Internet Explorer.

Google was also created based on a research project. Google showed that with a suitable query, we can quickly get to the relevant content of pages that it has collected on the Internet. We get the answer in the form of a list of hits - web addresses of pages that match the query. Recently, using artificial intelligence, this list has been replaced with a summary that the user can improve in a conversation with the service.

Amazon showed that with descriptions – data about individual things from the real world, these things can be included (made accessible) on the Internet. The online solution can also be used to organize scientific publications. In this case, the online framework is too general. It makes sense to limit ourselves to scientific publications. For an individual work to be accessible, it must be included in the service collection. This is achieved by digitizing works, which has two basic forms:

- full: the work is already created in digital form or has been converted into it, a description of the work is added
- partial: only its description is created for the work, which also contains data on access to the work itself (copyright)

The development of the digitization of historical and cultural resources is the responsibility of the TEI (Text Encoding Initiative). The preparation of online descriptions of works is the responsibility of the Dublin Core Metadata Initiative (DCMI). In the field of librarianship, we have BIBFRAME: Bibliographic Framework Initiative.

Descriptions of works contain various units (author, work, source, institution, country, etc.). It becomes complicated to identify these units when compiling a description. Namely, the data can be ambiguous (different units can have the same name (a string of characters)) or synonyms can occur (the same unit has several names). It is best to solve the problem of identification when entering data into the collection and use unique identifiers for the units (ORCID, DOI, ISSN, ISBN, ISO alpha 2, URL, URI, etc.). This provides high-quality data for informetric analyses. The task could be made much easier by the authors themselves if they used these unique identifiers when preparing their papers – something that journal editors (or paper submission forms) can take care of.

One of the important characteristics of printed works is their "finality". IT greatly alleviates this. We can correct a digital article, and collect opinions (content comments, ratings) about it, citations to it, the number of hits, etc. For example, we can include 3-dimensional representations [10], very detailed images in vector description [11], or a video clip, which can be viewed with an appropriate viewer. A geography textbook could include tables containing current data about a geographical unit.

From archaeology, we know of records carved in stone, imprinted in clay, carved in wood, written on papyrus, parchment, or paper, etc., which have been preserved in whole or in part for centuries or even millennia and have outlived the civilizations that created them. However, there is a problem with digital records – most recording media (tapes, floppy disks, discs, CDs, DVDs, etc.) have a guaranteed durability of about 5 years [34]. The exception is M-disc discs, which are supposed to last at least 1000 years [43]. An additional problem is access to older reading/writing devices. To ensure the persistence of records, it is recommended to store them in multiple copies on different media and occasionally copy them to newer media. Records are prepared in different formats. Record formats also change. For storage, it makes sense, if possible, to rely on character files and use record formats that preserve the data structure (TeX, SGML, XML, JSON, etc.).

A knowledge graph is a formalism for representing data and knowledge in the form of a graph. Formally, a knowledge graph is defined as a labeled directed graph consisting of a set of triples of the form (Subject, Property, Object). Subject and Object represent the nodes of the graph, and the triples represent the named links of the graph that define the relations between the nodes.

The language for representing knowledge graphs, RDF (Resource Description Framework), was proposed as part of the development of the Semantic Web (Web 3.0). RDF allows for the easy representation of graphs with a text file. Each line of the file contains one triple. For example, the triple (Cankar, Place of Birth, Vrhnika) says that Cankar was born in Vrhnika.

In the last decade, knowledge graphs have been one of the most widely used media for presenting and exchanging data and knowledge in science and industry. In science, RDF is often used to exchange data, such as the results of experiments. In computer science, knowledge graphs are a medium for presenting knowledge from the field of operation of a given information system. For example, all major companies, such as Google, Microsoft, Amazon, LinkedIn, Facebook, and others, develop and maintain knowledge graphs for the needs of implementing their information systems.

23.2. State-of-the-art in the proposed field of research and survey of the relevant literature

Due to the rapid (currently exponential) growth in the number of new works, the current approach to publishing is becoming inadequate (growth in the number of journals, mega-journals, predatory journals, the problem of obtaining reviewers, fraud, etc.).

According to OpenAlex, the number of published scientific papers doubled in the 19 years between 1971 (881,943 papers) and 1989 (1,847,109 papers), then in the next 12 years to 2001 (3,705,036 papers), then in 8 years to 2009 (7,275,504). In 2020, 11,017,156 papers were published. The data for recent years is probably incomplete.

Of course, growth is not unlimited. If we assume a raised logistic model (in the initial part, growth is almost exponential)

$$f(x; L_0, L_1, x_0, k) = L_0 + \frac{L1}{1 + e^{-k\cdot(x-x_0)}}, \qquad L = L_0 + L_1$$

we obtain as local extrema the following fitting curves

blue: $L_0 = 914793.1$, $L_1 = 12310320$, $k = 0.1323116$, $x_0 = 2009$, $L = 13225113$
red: $L_0 = 798003.8$, $L_1 = 18594720$, $k = 0.1064136$, $x_0 = 2016$, $L = 19392724$.

which fit the data quite well, but due to the lack of complete data for recent years, they give very different predictions.

Scopus data from 1996 to 2011 show that 15 million scientists published at least one article during this period, but only 150,608, less than 1%, each year. Their names appear in 41% of all articles and as co-authors of 87% of the most cited articles.

Anomalies and fraud [42]

1. articles with a very large number of co-authors. The article "COVIDSurg Collaborative and GlobalSurg Collaborative: Timing of surgery following SARS-CoV-2 infection: an

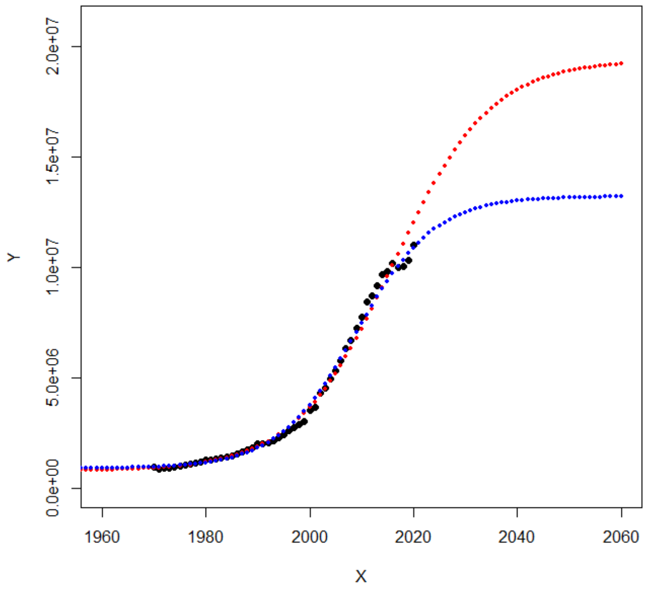

Figure 2: Logistic fitting

international prospective cohort study. Anaesthesia 2021, 76, 748–758" has 16162 co-authors.

2. authors who co-authored a very large number of articles in the first ten months of 2024: Wiwanitkit, V. (492), Daungsupawong, H. (346), Bruze, M. (336), etc.

3. degeneration of peer review [4, 29].

4. increasing the importance of institutions by "buying" top scientists [2].

5. predatory journals and conferences [12]: journals with several thousand annual publications (in 2024: Scientific Reports (27528), Heliyon (16271), PLoS ONE (14912), IEEE Access (11726), etc.), thematic issues of journals [44].

6. requests from reviewers or editors to include references to unnecessary sources in the article [32].

7. author fraud (plagiarism, purchased articles, ChatGPT, fabricated/modified data, etc.) [21].

Goodhart (rough draft)

- There are indications that in the informetric monitoring of publications, there is a change of goals: instead of quality, other quantitative parameters are rewarded: e.g. the impact of publications or even the impact of journals in which the articles are published. Users are adapting to this.

- As a result, publishers and journals are multiplying, trying in every way to increase the popularity of publications, regardless of quality. To publish in predatory journals, it is important to have access to public money.

- There is a drain of public money into the private pockets of multinational publishing houses.

Open science is a modern approach to scientific research and the dissemination of its results (knowledge) transparently and collaboratively. Opening up science through the fastest possible exchange of information and knowledge between researchers increases the chances for faster progress in science [3]. [36, 23, 1, 27, 28, 16, 20]

In 1942, Robert K. Merton, in his article "A Note on Science and Democracy" (reprinted in the book The Sociology of Science), wrote four "pillars" of science ("communism", universality, work for the common good (selflessness, ethics), skepticism). Later, originality was added [45]. Chubin, D. E. also refers to them in his article "Open Science and Closed Science: Tradeoffs in a Democracy" (1985) [18]. In the age of the Internet, the open science movement has made free access to scientific publications its main pillar, as reflected in the declaration of the Budapest conference in 2001. This is legally regulated by the Creative Commons licenses created in 2002. Subsequently, the requirement for openness has been extended to data and software. Openness can be implemented with online repositories.

For informetric analyses, free access to descriptions (metadata) of published scientific works – including those published in subscription journals – is important. They are striving to make free access also apply to abstracts [38]. An important step in this direction is OpenAlex.

OpenAlex is a completely open catalog of the global research system. It is named after the ancient Library of Alexandria. It was established by the non-profit organization OurResearch. It became accessible in January 2022 via a user interface, a free API, or a snapshot of all the data that can be downloaded to your computer. It is considered a replacement for the Microsoft Academic Graph service, which was discontinued on December 31, 2021. OpenAlex is based on 7 types of units (entities): work, author, resource, institution, concept, publisher or funder. It solves several important issues for the analysis of bibliographic data:

1. identification of bibliographic units (ID, resolution)
2. free access (sharing of obtained data, transfer to the user's computer)
3. content improvement with user participation (the user submits a request for correction)

OpenAlex can be used via a web user interface or programmatically with API calls. For more complex processing, the problem is the limitation to a maximum of one hundred thousand calls per day. This limitation can be avoided by setting up a copy of OpenAlex on a local computer. OpenAlex contains data on most of the works found in the Web of Science and Scopus commercial services. In addition, there is a variety of other works – for example, data on publications in the arXiv repository. This allows us to monitor what are the current "hot" topics in science. Data from commercial services are limited to established journals and contain a time lag caused by the publication process of articles. OpenAlex also contains data on the publication mode (openness) of each work

To use data from OpenAlex, it is necessary to develop appropriate software support that collects and converts the desired data into a format suitable for data (statistical, network, AI) analysis. Support for data cleaning and quality control is particularly important. Sometimes the data also needs to be refined – for example, determining the category of references (agreement, comparison, definition, difference, disagreement, hypothesis, method, position, result, similarity) [14].

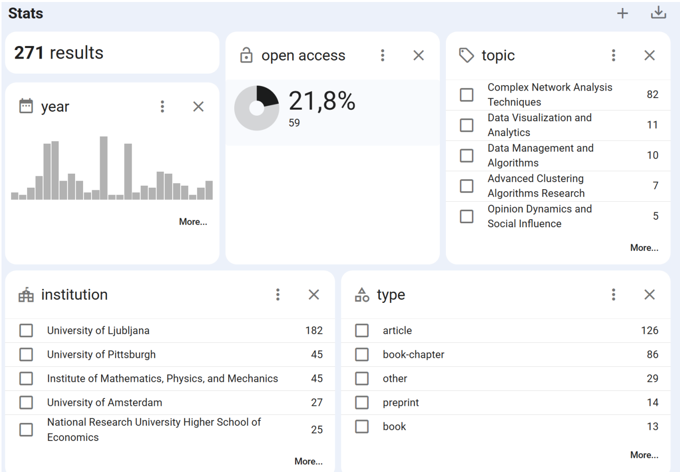

Figure 3: Part of OpenAlex report

Project

Scientific Publishing - Expectations of the Author-Scientist

1. immediate publication and access,
2. "registration" – timestamp,
3. easy to find,
4. unlimited access,
5. permanent storage (longevity, durability),
6. reader (community) response, acknowledgments.

This can be achieved with repositories of works. The evaluation of the work is not the main concern of the author – it is important for his/her employers, funders, etc.

The evaluation of scientific results places too much emphasis on the number of articles and the number of citations of articles. There are other forms of work: books, patents, plans, films, data, programs, websites, etc. Recently, in a webinar, the famous statistician John Bailer said: what are my few thousand citations compared to the millions using Wickham's Tidyverse. Several movements and recommendations attempt to enable and implement a more comprehensive evaluation of scientific research work, such as ORCID, CRediT (Contributor Roles Taxonomy), Leiden Manifesto for Research Metrics, San Francisco Declaration on Research Assessment (DORA), The Vancouver Recommendations, Coalition for Advancing Research Assessment (CoARA). The possibility of repeating experiments and analyses is becoming important, as advocated by the FAIR (Findability, Accessibility, Interoperability, and Reuse) movement [48, 22].

There are indications that in informetric monitoring of publications, goals are being replaced: instead of quality, other quantitative parameters are being rewarded: e.g. the impact of publications or even the impact of journals in which articles are published. Users are adapting to this. Therefore, publishers and journals are multiplying, trying in every way to increase the popularity of publications, regardless of quality. Access to public money is important for publishing in predatory journals. Public money is flowing into the private pockets of multinational publishing houses. Publishing models are changing radically from subscription to open access. APC (Article Processing Charge) and hybrid publishing models are dominant.

In response to these challenges, new publishing models have emerged, such as "publish, review, curate" (PRC). This model reverses the traditional review-then-publish approach by first publishing the work online and then subjecting it to peer review. The aim of this approach is to increase transparency and accelerate the dissemination of research [19].

Building a knowledge graph for the OpenAlex data environment. The OpenAlex knowledge graph will contain an ontology that includes article types, scientific fields, institutions, and other entities related to articles. The construction of the OpenAlex knowledge graph is possible by connecting to existing knowledge graphs, such as DBpedia, WikiData, Yago, and others. These knowledge graphs mostly represent common-sense knowledge about the world.

Newer big data analytics systems, such as the open-source Apache Spark, allow for the implementation of complex analytics on data that is either in text format (JSON) or in graph format. The analytics can be performed using interactive SQL queries or programs in general-purpose programming languages such as Java, Python, and R.

23.3. Detailed description of the work programme

Topic groups

- OpenAlex (local, support development, connectivity with Slovenian databases, study of graph databases, methods of use)

- open publishing (PRC, study, monitoring and involvement in developments, AMC support for editors, ethical issues)

- support for open science (informatics analyses, theoretical and software support)

- development and software support of methods for bibliometric analyses and services

Divide into groups

- local OpenAlex; creation of software support - KajDogaja? use in lectures from databases (UP)

- collect and complete software support for the review process at the open journal Ars Mathematica Contemporanea (AMC) (selection of reviewers, checking language, plagiarism, relevance of keywords, relevance of sources, etc.)

- overview of the state of open science in Slovenian libraries; inclusion of this content in the education of librarians (UL FF)

- overview of theoretical starting points and websites for experiments with the PRC approach (publish, review, curate); attempt to establish cooperation with the selected one(s).

- PRC support - automatic review procedures: author recognition and connection to ORCID, adding DOI to individual sources, assessment of the connection of sources with the topic of the article, assessment of the quality of an individual source, analysis of the connection between authors and "reviewers", identification of unethical activities, related articles, ... trial use in AMC, ... standards ... OpenAlex, ArXiv, ...

- support for analytical procedures of data from OpenAlex; graph databases (Neo4j, ...); inclusion of these contents in the course of network analysis in the study of statistics (UL)

- Goodhart: Formal description of Goodhart's law; Application of Goodhart's law in informetrics; Development of methods for detecting potentially unethical behavior in scientific publications (authors, editors, reviewers, publishers); Preparation of recommendations for correction of the situation.

- possible implicit and explicit measures of article quality; Goodhart

- further development of methods based on bibliometric networks and higher-level bibliographic services

Analyses

- We are trying to assess the speed of change in the field of scientific publishing, the increase in numbers in recent years (no. of publications, no. of journals, APC + subscriptions).

- Development of FJN (Free Journal Network) Free journal network unites some diamond open access journals.

Hacker attacks

The emergence of unethical publications and mechanisms for their elimination.

- The problem of misclassification of journals and researchers. In many places, more than one classification of a research field appears for a unit. (e.g. SICRIS, SCIMAGO). Due to different citation cultures, serious errors can occur in ranking within a single field. We will develop tools to find such anomalies.

- The problem of distinguishing potentially predatory methods in journals with a high impact factor. It is not possible to determine from the impact factor alone whether a journal is suitable for publication. We will define methods that will identify potentially controversial journals with the help of other data.

- A survey or surveys to gather feedback from scientists and other stakeholders on the changes brought about by open science.

Suggested research questions related to open science:

- How does the data on publications (and OA) between OpenAlex and COBISS match? How could the quality of both sources be improved? What types of data are missing?

- Is the share of publications in OA increasing and how fast? Are there differences between different research fields and research organizations (according to type, size and other characteristics, if available)?

- How is the volume of different types of OA (diamond, gold, green) increasing and are there any differences between research fields or organizations?

- What are the costs of publications in gold access? How do the costs differ between different research fields and organizations? (Data on APC are available via publishers' websites, e.g. for Elsevier at this link: APC

- How are different types of open access related to the quality of publications according to selected quality indicators?

- Is it possible to link publications in repositories (in COBISS category 2.20) to data publications? What is the proportion of publications that cite databases?

Expected results

Expected results Goodhart

- Article based on the paper at IS 2024
- Article on the statistical-theoretical foundations of Goodhart's law
- Contributions to IS 2025,2026,2027
- Article on the results of surveys
- Article on determining potential predators.
- Article on improving classifications
- Contribution to Open Science Days 2026 and 2027 and to the Applied Statistics 2027 conference
- Published research data in the FAIR repository

Detailed description of the work program (23.3)

WP1: Project management

Task 1.1 Project administration and coordination

Management activities will take place during the entire project and will include administrative work related to work performance, quality assurance, risk assessment and mitigation, compliance with the proposal and contractual obligations with the Slovenian Research and Innovation Agency. In parallel, coordination activities will include fostering agreement on responsibilities, scheduling and coordinating team meetings and internal briefing on outcomes of individual activities.

Task 1.2 Data management

A data management plan (DMP) will be prepared in the first six months of the project (D1.1.) following the FAIR principles of making data findable, accessible, interoperable, and reusable. The DMP will detail the purpose of data collection, relevance to the objectives of the project, specify the types and formats of data and metadata to be collected, the expected size, and how data will be curated and preserved.

WP2: Bibliographic data

Task 2.1 Acquisition, cleaning, and preparation of bibliographic data (from OpenAlex and COBISS and administrative data from the University of Primorska and/or other institutions)?

Task 2.2 Development of knowledge graph

Task 2.3 Development of software for bibliographic analysis and services

Task 2.3 Analysis of bibliographic data

Task 2.5 Computation of measures of publication quality

Task 2.6 Detection of non-ethical publication practices

WP3: PRC publication model data

Task 3.1 Acquisition, cleaning, and preparation of data on PRC (i.e. data on APCs, editorial process, open peer reviews, etc.)

Task 3.2 Analysis of PRC data

WP4 Survey data

Task 3.1 Review of previous survey data on publishing

Task 3.2. Development of survey questionnaires

Task 3.3 Survey of researchers (at the University of Priomorska/in all Slovenian research institutions?)

Task 3.4 Survey of journal editors (for a sample of journals from all research fields or for just one research field as a case study?)

Task 3.5 Analysis of suvey data

WP5: Exploitation and dissemination of results?

23.4. Available research equipment over 5.000 €

23.5. Project management: Detailed implementation plan and timetable

References

[1] Ahmed, Abubakari; Al-Khatib, Aceil; Boum II, Yap; Debat, Humberto; Dunkelberg, Alonso Gurmendi; Hinchliffe, Lisa Janicke; Jarrad, Frith; Mastroianni, Adam; Mineault, Patrick; Pennington, Charlotte R.; Pruszynski, J. Andrew: The future of academic publishing. Nature Human Behaviour volume 7, pages 1021–1026 13 July 2023. PDF

[2] Ansede, Manuel: Dozens of the world's most cited scientists stop falsely claiming to work in Saudi Arabia. El Pais, Dec 05, 2024. WWW

[3] ARIS: Odprta znanost. WWW

[4] Bal, Mieke: Let's Abolish the Peer-Review System. 2018. WWW

[5] Batagelj, Vladimir. On fractional approach to analysis of linked networks. Scientometrics. May 2020, vol. 123, iss. 2, p. 621-633. DOI:[10.1007/s11192-020-03383-y, WWW

[6] Batagelj, Vladimir : OpenAlex2Pajek – an R Package for converting OpenAlex bibliographic data into Pajek networks. COLLNET 2024, Strasbourg, December 12-14. V: Jain, Praveen Kumar (ur.), et al. Innovations in webometrics, informetrics, and scientometrics: AI-driven approaches and insights. Delhi: Bookwell, cop. 2024. p. 66-77. ISBN 978-93-86578-65-5. PDF

[7] Batagelj, Vladimir, Maltseva, Daria. Temporal bibliographic networks. Journal of informetrics: an international journal. Feb. 2020, vol. 14, iss. 1, art. 101006 (14 p). DOI

[8] Batagelj, Vladimir, Ferligoj, Anuška, Squazzoni, Flaminio. The emergence of a field: a network analysis of research on peer review. Scientometrics. 2017, vol. 113, iss. 1, str. 503-532, DOI

[9] Batagelj, Vladimir, Cerinšek, Monika. On bibliographic networks. Scientometrics. 2013, vol. 96, iss. 3, p. 845-864. DOI

[10] Batagelj, Vladimir: European Airports core of order 13. X3D

[11] Batagelj, Vladimir: Slovenski politiki - matrični prikaz / vse relacije; obdobje do april 2022. WWW/SVG

[12] Beall's List of Potential Predatory Journals and Publishers. WWW

[13] Berners-Lee, Tim, Hendler, James: Publishing on the semantic web. Nature volume 410, pages1023–1024 (2001). WWW

[14] Bertin, M., Atanassova, I. Linguistic perspectives in deciphering citation function classification. Scientometrics 129, 6301–6313 (2024). DOI

[15] Boulton, Geoffrey; Koley, Moumita: More is not better: the developing crisis of scientific publishing. July 2, 2024. WWW

[16] Brembs, Björn; Huneman, Philippe; Schönbrodt, Felix; Nilsonne, Gustav; Susi, Toma; Siems, Renke; Perakakis, Pandelis; Trachana, Varvara; Ma, Lai; Rodriguez-Cuadrado, Sara. (2021). Replacing academic journals, version 5. DOI; Zenodo

[17] Bush, Vannevar: As We May Think. The Atlantic, July 1945 PDF

[18] Chubin, D. E. (1985). Open science and closed science: Tradeoffs in a democracy. Science, Technology, and Human Values, 10(2), 73-80. PDF

[19] Corker, K. S., Waltman, L., & Coates, J. A. (2024). Understanding the Publish-Review-Curate (PRC) model of scholarly communication. MetaArXiv

[20] Council of the European Union: High-quality, transparent, open, trustworthy and equitable scholarly publishing (approved on 23 May 2023). PDF

[21] Day, Adam; Cucen, Ebru: Searching for Research Fraud in OpenAlex with Graph Data Science. 2022. WWW

[22] FAIR Principles. WWW

[23] Guédon, Jean-Claude: In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing. ARL 2001. PDF

[24] Hardt, Moritz; Megiddo, Nimrod; Papadimitriou, Christos; Wootters, Mary: Strategic classification. Association for Computing Machinery (ACM), New York, 2016, 111–122. ISBN: 978-1-4503-4057-1. MR3629816.

[25] Hennessy, Christopher A.; Goodhart, Charles A. E.: Goodhart's law and machine learning: a structural perspective. Internat. Econom. Rev. 64 (2023), no. 3, 1075–1086. MR4624804.

[26] Horbach, S.P.J.M.(.S., Halffman, W.(.W. The changing forms and expectations of peer review. Res Integr Peer Rev 3, 8 (2018). DOI. PDF

[27] International Science Council: The future of scientific publishing. WWW

[28] Kochetkov, D. (2024, March 21). Evolution of Peer Review in Scientific Communication. DOI; SocArXiv

[29] Levy, Neil: Philosophy, Bullshit, and Peer Review. Cambridge (UK): Cambridge University Press; 2023 Dec. ISSN: 2398-0567; 2514-3832. WWW, PDF

[30] Maltseva, Daria, Batagelj, Vladimir. iMetrics: the development of the discipline with many names. Scientometrics. Oct. 2020, iss. 1, vol. 125, p. 313-359, DOI: 10.1007/s11192-020-03604-4,

[31] Pisanski, T., Batagelj, V., Pisanski, J.: Open Science and Goodhart's Law. IS2024-Cognitive Science, Ljubljana, October 11 2024. p. 20-23. DOI

[32] Marquez, Ronald: LinkedIn: What should #Editors do? WWW

[33] Matt, Christian; Hoerndlein, Christian; Hess, Thomas: Let the crowd be my peers? How researchers assess the prospects of social peer review. 2017. PDF

[34] Mediafix: Durability of storage media. WWW

[35] Nooy, Wouter De, Mrvar, Andrej, Batagelj, Vladimir. Exploratory social network analysis with Pajek. 3rd ed., revised and expanded ed. for updated software. Cambridge [etc.]: Cambridge University Press, 2018. XXX, 455 p, ilustr. Structural analysis in the social sciences, 46. ISBN 978-1-108-47414-6, ISBN 978-1-108-46227-3. DOI

[36] Ouvrir la science! WWW

[37] Stokstad E. The 1% of scientific publishing: Only a handful of researchers manage to publish one or more papers per year. Science. 2014 Jul;11. WWW

[38] The Initiative for Open Abstracts. WWW

[39] Wikipedia: arXiv. WWW

[40] Wikipedia: Derek J. de Solla Price. WWW

[41] Wikipedia: Informetrics. WWW

[42] Wikipedia: List of scientific misconduct incidents. WWW

[43] Wikipedia: M-disc. WWW; Verbatim

[44] Wikipedia: Mega journal. WWW

[45] Wikipedia: Mertonian norms. WWW

[46] Wikipedia: Mundaneum. WWW

[47] Wikipedia: Open science. WWW WWW

[48] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). DOI, [WWWWWW